



DESCUBRIMIENTO DE PATRONES DE DESEMPEÑO ACADÉMICO

SILVIO RICARDO TIMARÁN PEREIRA
ISABEL HERNÁNDEZ ARTEAGA
SEGUNDO JAVIER CAICEDO ZAMBRANO
ARSENIO HIDALGO TROYA
JUAN CARLOS ALVARADO PÉREZ

Descubrimiento de patrones
de desempeño académico con árboles
de decisión en las competencias
genéricas de la formación profesional

Discovery of Academic Performance
Patterns Using Decision Trees in Generic
Skills of Professional Education

Resumen

El objetivo de este estudio fue descubrir patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales en las pruebas Saber Pro 2011-2, a partir de los datos sociodemográficos, económicos, académicos e institucionales almacenados en las bases de datos del Instituto Colombiano para la Evaluación de la Educación (ICFES), y utilizando técnicas de minería de datos. Los estudios realizados hasta el momento con respecto al análisis de los resultados de las pruebas Saber Pro 2011-2 se basan en información procesada mediante una investigación estadística, en la cual, fundamentalmente se consideran variables y relaciones primarias, sin tener en cuenta las verdaderas interrelaciones, que por lo general están ocultas y que únicamente se pueden descubrir utilizando un tratamiento más complejo de los datos, que es posible con la mencionada minería de datos. Por otro lado, se aplicó la metodología CRISP-DM y con los repositorios de datos, limpios y transformados para cada competencia, se descubrieron patrones asociados al buen o mal desempeño académico, utilizando la técnica de clasificación basada en árboles de decisión. En los patrones descubiertos se destacan la acreditación institucional y la modalidad de estudio como dos atributos importantes asociados al desempeño académico de los estudiantes en las pruebas Saber Pro 2011-2, en las cuatro competencias genéricas. El conocimiento descubierto se incorporará al existente y se podrá integrar a los procesos de toma de decisiones del ICFES y de las instituciones gubernamentales y académicas que velan por la calidad de la educación superior.

Palabras clave: calidad de la educación superior, competencias genéricas, desempeño académico, minería de datos, patrones de desempeño académico, pruebas Saber Pro.

Abstract

The objective of this study aimed to discover academic performance patterns in generic competences of the students in professional programs in the 2011-12 Saber Pro Exam, based on sociodemographic, economic, academic and institutional data stored in the databases of the Instituto Colombiano para la Evaluación de la Educación (ICFES), using data mining techniques. The studies carried out until now to analyze the 2011-12 Saber Pro test results are based on information processed using statistical research, in which fundamentally primary variables and relationships are considered, without taking into account the true interrelationships that are generally hidden and can only be discovered using a more complex data treatment, which is possible with the above-mentioned data. Therefore, the CRISP-DM methodology was applied and through the use of clean data repositories that were transformed for each competence, patterns associated with good or poor academic performance were discovered, using the classification technique based on decision trees. In the patterns that were found, institutional accreditation and study modality stand out as two important attributes associated with the academic performance of the students in the 2011-12 Saber Pro tests, in the four generic competences. The findings will be incorporated into what already exists and can be integrated into decision-making processes of the ICFES and of the governmental and academic institutions that oversee the quality of higher education.

Keywords: quality of higher education, generic competences, academic performance, data mining, academic performance patterns, Saber Pro tests.

¿Cómo citar este libro?/ How to cite this book?

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*. Bogotá: Ediciones Universidad Cooperativa de Colombia. DOI: <http://dx.doi.org/10.16925/9789587600490>

Autores

Silvio Ricardo Timarán-Pereira. Doctor en Ingeniería con énfasis en Ciencias de la Computación. Director e investigador del Grupo de Investigación Aplicada en Sistemas (Grias) y profesor titular del Departamento de Sistemas de la Facultad de Ingeniería, de la Universidad de Nariño, Colombia. Correo electrónico: ritimar@udenar.edu.co

Isabel Hernández-Arteaga. Doctora en Ciencias de la Educación. Investigadora del Centro de Investigaciones en Docencia Universitaria (CIDU) y profesora de la Escuela de Postgrados, de la Universidad Cooperativa de Colombia, sede Bogotá. Correo electrónico: isabel.hernandez@ucc.edu.co

Segundo Javier Caicedo-Zambrano. Doctor en Ciencias de la Educación. Investigador del Grupo de Investigación Aplicada en Sistemas (Grias). Profesor adscrito a la Facultad de Ciencias Exactas y Naturales de la Universidad de Nariño, Asesor de Desarrollo Académico de la Universidad de Nariño, Colombia. Correo electrónico: jacaza1@gmail.com

Arsenio Hidalgo-Troya. Magíster en Ciencias. Miembro del Grupo de Investigación en Salud Pública y docente del Departamento de Matemáticas y Estadística de la Facultad de Ciencias Exactas y Naturales, de la Universidad de Nariño, Colombia. Correo electrónico: arsenio.hidalgo@gmail.com

Juan Carlos Alvarado-Pérez. Máster en Sistemas Inteligentes. Miembro del Grupo Interdisciplinario de Investigación en Salud-Enfermedad (GISSE) y docente de la Facultad de Medicina, de la Universidad Cooperativa de Colombia, sede Pasto. Correo electrónico: juan.alvarado@campusucc.edu.co

Authors

Silvio Ricardo Timarán-Pereira. Ph.D. in Engineering with emphasis on Computer Sciences. Head and researcher of the Systems Applied Research Group (Grias) and associate professor of the Systems Department in the School of Engineering at Universidad de Nariño, Colombia. E-mail: ritimar@udenar.edu.co

Isabel Hernández-Arteaga. Ph.D. in Education Sciences. Researcher of the Research Center for University Teaching (CIDU) and professor in the School of Graduate Studies at Universidad Cooperativa de Colombia, Bogotá. E-mail: isabel.hernandez@ucc.edu.co

Segundo Javier Caicedo-Zambrano. Ph.D. in Education Sciences. Researcher of the Systems Applied Research Group (Grias) and professor of the School of Exact and Natural Sciences at Universidad de Nariño, Colombia. E-mail: jacaza1@gmail.com

Arsenio Hidalgo-Troya. Master's in Sciences. Member of the Public Health Research Group and professor of the Mathematics and Statistics Department in the School of Exact and Natural Sciences at Universidad de Nariño, Colombia. E-mail: arsenio.hidalgo@gmail.com

Juan Carlos Alvarado-Pérez. Master's in Smart Systems. Member of the Interdisciplinary Research Group in Health and Disease (GISSE) and professor in the School of Medicine at Universidad Cooperativa de Colombia, Pasto. E-mail: juan.alvarado@campusucc.edu.co

Descubrimiento de patrones
de desempeño académico con árboles
de decisión en las competencias
genéricas de la formación profesional

Discovery of Academic Performance
Patterns Using Decision Trees in Generic
Skills of Professional Education

Silvio Ricardo Timarán Pereira

Isabel Hernández Arteaga

Segundo Javier Caicedo Zambrano

Arsenio Hidalgo Troya

Juan Carlos Alvarado Pérez



Universidad Cooperativa
de Colombia



EDICIONES
Universidad Cooperativa
de Colombia

Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional / Silvio Ricardo Timarán Pereira y otros. -- Bogotá : Fondo Editorial Ediciones Universidad Cooperativa de Colombia, 2016.

194 páginas ; 24 cm.
Incluye índice analítico
ISBN 978-958-760-049-0

1. Educación superior 2. Educación por competencias 3. Calidad de la educación superior 4. Formación profesional 5. Competencias profesionales I. Timarán Pereira, Silvio Ricardo, autor.
378 cd 21 ed.
A1531883

CEP-Banco de la República-Biblioteca Luis Ángel Arango

Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional

© Ediciones Universidad Cooperativa de Colombia, Bogotá, abril de 2016

© Silvio Ricardo Timarán Pereira

© Isabel Hernández Arteaga

© Segundo Javier Caicedo Zambrano

© Arsenio Hidalgo Troya

© Juan Carlos Alvarado Pérez

ISBN (digital): 978-958-760-050-6

DOI: <http://dx.doi.org/10.16925/9789587600490>

Colección Investigación en Educación

ISSN: 2500-6649

Proceso de arbitraje doble ciego

Recepción: abril de 2015

Evaluación propuesta de obra: junio de 2015

Evaluación de contenidos: octubre de 2015

Correcciones de autor: enero de 2016

Aprobación: febrero de 2016

Fondo Editorial

Director Nacional Editorial, Manfred Acero-Gómez

Coordinadora Editorial, Ruth Elena Cuasialpud-Canchala

Producción editorial de libros, Camilo Cuéllar-Mejía

Producción editorial de revistas, Daniel Urquijo-Molina

Proceso editorial

Corrección de estilo, Nathalie de La Cuadra

Lectura de pruebas, Hernando Sierra

Traducción al inglés, Nathalie Barrientos

Elaboración de índice analítico, Sebastián Montero

Diseño y diagramación, Pedro C. Gutiérrez-Jiménez

Diseño de portada, Ivonne Carolina Cardozo-Pachón

Impresión, Xpress Estudio Gráfico y Digital

Impreso en Bogotá, Colombia. Depósito legal según el Decreto 460 de 1995.

El Fondo Editorial Ediciones Universidad Cooperativa de Colombia se adhiere a la filosofía del acceso abierto y permite libremente la consulta, descarga, reproducción o enlace para uso de sus contenidos, bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional. <http://creativecommons.org/licenses/by-nc-nd/4.0/>



Contenido

Págs.

15 **Introducción**

1

Desempeño académico y competencias genéricas en la formación de profesionales

Academic Performance and Generic Skills in Professional Training

20 Introducción

20 Aseguramiento de la calidad de la educación superior:
tarea permanente de la universidad del siglo XXI

27 Formación profesional mediante el enfoque de competencias

41 Desempeño académico

51 Contextualización de las pruebas Saber Pro

57 Referencias

2

El proceso de descubrimiento de conocimiento en bases de datos

The Process of Knowledge Discovery on Databases

64 Introducción

64 Etapas del proceso KDD

67 Tareas de minería de datos

74 Áreas relacionadas con el proceso KDD

75 Metodología CRISP-DM

84 Referencias

3

Tendencias de desempeño académico en competencias genéricas - pruebas Saber Pro

Academic Performance Trends in Generic Skills - Saber Pro Exam

88 Introducción

88 Análisis de correlación entre las competencias genéricas

- 89 Desempeño en las competencias genéricas según variables sociodemográficas, económicas, académicas e institucionales
- 99 Referencias

4

Descubrimiento de patrones de desempeño académico en las competencias genéricas

Discovery of Academic Performance Patterns in Generic Skills

- 102 Introducción
- 102 Comprensión del negocio o problema
- 102 Comprensión de los datos
- 120 Preparación de los datos
- 134 Modelado
- 146 Evaluación
- 146 Implementación
- 148 Referencias

5

Resultados y discusión

Results and discussion

- 152 Resultados
- 161 Discusión

- 169 **Conclusiones**
- 173 **Bibliografía general**
- 185 **Índice analítico**

Índice de tablas

Págs.

26	Tabla 1. Factores y características que se consideran en la evaluación de la calidad institucional en Colombia
35	Tabla 2. Lectura crítica: categorías, concepto y criterios de evaluación
37	Tabla 3. Comunicación escrita: categorías, concepto y criterios de evaluación
38	Tabla 4. Razonamiento cuantitativo: categorías, concepto y criterios de evaluación
39	Tabla 5. Inglés: categorías, concepto y criterios de evaluación
54	Tabla 6. Cambios en las pruebas Saber Pro en el período 2009-2012
55	Tabla 7. Listado de grupos de referencia de los programas de educación superior en Colombia
88	Tabla 8. Matriz de correlaciones de competencias genéricas
90	Tabla 9. Desempeño académico en competencias genéricas según género. Pruebas Saber Pro 2011-2
91	Tabla 10. Desempeño académico en competencias genéricas según estado civil. Pruebas Saber Pro 2011-2
93	Tabla 11. Desempeño académico en competencias genéricas según nivel educativo de los padres. Pruebas Saber Pro 2011-2
94	Tabla 12. Desempeño académico en competencias genéricas según estrato socioeconómico. Pruebas Saber Pro 2011-2
96	Tabla 13. Desempeño académico en competencias genéricas según modalidad de estudio. Pruebas Saber Pro 2011-2
97	Tabla 14. Desempeño académico en competencias genéricas según tipo de institución. Pruebas Saber Pro 2011-2
98	Tabla 15. Desempeño académico en competencias genéricas según acreditación institucional. Pruebas Saber Pro 2011-2
99	Tabla 16. Desempeño académico en competencias genéricas según acreditación del programa. Pruebas Saber Pro 2011-2

103	Tabla 17. Diccionario de datos del repositorio inicial T97068A94
108	Tabla 18. Estado civil
108	Tabla 19. Origen institución
109	Tabla 20. Carácter académico
109	Tabla 21. Código jornada
110	Tabla 22. Área del conocimiento
110	Tabla 23. Núcleo pregrado
112	Tabla 24. Grupo de referencia
113	Tabla 25. Porcentaje de créditos cursados y aprobados
113	Tabla 26. Valor anual de la matrícula año anterior
113	Tabla 27. Educación del padre o madre
114	Tabla 28. Ocupación del padre o madre
115	Tabla 29. Estrato socioeconómico
115	Tabla 30. Educación de los hermanos mayores
116	Tabla 31. Nivel de Sisben
116	Tabla 32. Material de los pisos
117	Tabla 33. Ingresos familiares mensuales
117	Tabla 34. Trabajo del estudiante
117	Tabla 35. Lee o habla otro idioma
118	Tabla 36. Etnia estudiante
119	Tabla 37. Clasificación de atributos en dimensiones
120	Tabla 38. Atributos con porcentaje alto de datos nulos
121	Tabla 39. Atributos con valores constantes y atributos identificadores
121	Tabla 40. Nuevos atributos del repositorio de datos
125	Tabla 41. Valores discretizados del atributo estu_edad_examen
126	Tabla 42. Valores del atributo inst_programa_zona
127	Tabla 43. Valores discretizados del atributo num_estudiantes_zona
127	Tabla 44. Valores discretizados del atributo num_instituciones_zona
128	Tabla 45. Valores del atributo eco_condicion_hogar
128	Tabla 46. Cálculo del índice de condición_hogar
129	Tabla 47. Cálculo del índice de condición_tic
129	Tabla 48. Asignación de valores del atributo área_grupo_referencia
130	Tabla 49. Diccionario de datos del repositorio final T97068A35
133	Tabla 50. Repositorios por cada competencia

Índice de figuras

Págs.

65	Figura 1. Etapas del proceso KDD
76	Figura 2. Esquema de los cuatro niveles de CRISP-DM
77	Figura 3. Ciclo de vida de CRISP-DM
78	Figura 4. Fase de comprensión del negocio
79	Figura 5. Fase de comprensión de los datos
80	Figura 6. Fase de preparación de los datos
82	Figura 7. Fase de modelado
83	Figura 8. Fase de evaluación
84	Figura 9. Fase de implementación
137	Figura 10. Árbol para lectura crítica
138	Figura 11. Precisión y matriz de confusión del árbol de lectura crítica
139	Figura 12. Reglas más representativas de lectura crítica
140	Figura 13. Árbol para comunicación escrita
141	Figura 14. Precisión y matriz de confusión del árbol de comunicación escrita
141	Figura 15. Reglas más representativas de comunicación escrita
142	Figura 16. Árbol para razonamiento cuantitativo
143	Figura 17. Precisión y matriz de confusión del árbol de razonamiento cuantitativo
144	Figura 18. Reglas más representativas de razonamiento cuantitativo
145	Figura 19. Árbol para inglés
146	Figura 20. Precisión y matriz de confusión del árbol de inglés
147	Figura 21. Reglas más representativas de inglés

| Agradecimientos

Los autores expresan su agradecimiento a las instituciones que hicieron posible esta obra resultado de un proceso de investigación:

Al Instituto Colombiano para la Evaluación de la Educación (ICFES), por financiar el proyecto de investigación “Descubrimiento de patrones de desempeño académico en las competencias genéricas de la Pruebas Saber Pro 2011-2”, una de las nueve investigaciones seleccionadas dentro de la convocatoria para Grupos de Investigación del año 2013.

A la Universidad de Nariño y a la Universidad Cooperativa de Colombia, por aportar los recursos de contrapartida y posibilitar el trabajo interdisciplinario e interinstitucional entre los investigadores participantes en este proyecto.

Introducción

Uno de los objetivos del examen de Estado de calidad de la educación superior Saber Pro, según el Ministerio de Educación Nacional (MEN, 2009), a través del Decreto 3963 del 14 de octubre del 2009, es comprobar el grado de desarrollo de competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las instituciones de educación superior. El examen está compuesto por pruebas que evalúan competencias genéricas y específicas. De acuerdo con los lineamientos Saber Pro del Instituto Colombiano para la Evaluación de la Educación (ICFES, 2011a), todos los estudiantes deben presentar los módulos de competencias genéricas—sin importar el programa de formación que cursen—, que incluyen competencias de lectura crítica, comunicación escrita, razonamiento cuantitativo e inglés.

Según el ICFES (2012a), en la competencia de lectura crítica se evalúan los desempeños asociados a lectura, pensamiento crítico y entendimiento interpersonal; en comunicación escrita, la competencia para comunicar ideas por escrito referidas a un tema dado; en razonamiento cuantitativo, los desempeños relacionados con uso de lenguaje cuantitativo y solución de problemas; y en inglés, la competencia del estudiante para comunicarse efectivamente en inglés.

A pesar que en la prueba Saber Pro no se pretende que los estudiantes de todas las carreras profesionales desarrollen las competencias genéricas a un mismo nivel, ni aún las comunes a grupos de programas, sí es importante para el sector de la educación superior conocer cómo se asocian los factores correspondientes a las dimensiones sociodemográficas, económicas, académicas e institucionales del estudiante en relación con el nivel de desempeño en estas competencias según las pruebas Saber Pro.

Los estudios que se han realizado hasta el momento, según el ICFES (2010, 2011b y 2012b), Zapata (2011) y la Universidad Nacional de Colombia (2012), con respecto al análisis de los resultados de las pruebas Saber Pro, se basan en información procesada mediante un análisis estadístico, en el que fundamentalmente se consideran variables y relaciones primarias, sin tener en cuenta las verdaderas

interrelaciones, que por lo general están ocultas y que únicamente se pueden descubrir utilizando un tratamiento más complejo de los datos, lo cual es posible a través de la minería de datos. Mientras que la estadística plantea hipótesis que deben ser validadas a partir de los datos disponibles, la minería de datos descubre patrones no previstos con la estadística. Teniendo en cuenta a Febles y González (2002), los resultados que se obtienen con las técnicas de minería de datos no se pueden utilizar para confirmar o rechazar hipótesis, porque esto puede conducir a errores críticos. Su función es explorar datos, darles sentido, convertir en conocimiento un volumen de datos que poco o nada aportan a la descripción con el fin de interpretar fenómenos y adoptar decisiones de acuerdo con las necesidades.

La minería de datos en la educación no es un tema nuevo, su estudio y aplicación ha sido muy relevante en los últimos años; se pueden utilizar sus técnicas para explicar o predecir cualquier fenómeno dentro del campo educativo (Timarán, Calderón y Jiménez, 2013a y 2013b). Por ejemplo, utilizando las técnicas de minería de datos, se puede predecir, con un porcentaje muy alto de confiabilidad, la probabilidad de deserción de cualquier estudiante (Valero, 2009; Valero, Salvador y García, 2010). Las instituciones de educación pueden usar la minería de datos para hacer análisis comprensivos de las características de sus estudiantes, en métodos evaluativos, y develar así procesos exitosos o, por el contrario, detectar fraudes o inconsistencias (Orea, Vargas y Alonso, 2005).

El objetivo de este estudio fue descubrir patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales en las pruebas Saber Pro 2011-2, a partir de los datos sociodemográficos, económicos, académicos e institucionales almacenados en las bases de datos del ICFES, utilizando técnicas de minería de datos. Se aplicó la metodología CRISP-DM y con los repositorios de datos —limpios y transformados para cada competencia— se descubrieron patrones asociados al buen o mal desempeño académico utilizando la técnica de clasificación basada en árboles de decisión. En los patrones descubiertos se destacan la acreditación institucional y la modalidad de estudio como dos atributos importantes asociados al desempeño académico de los estudiantes en las pruebas Saber Pro, en las cuatro competencias genéricas. El conocimiento descubierto se incorpora al existente y se puede integrar a los procesos de toma de decisiones del ICFES y de las instituciones gubernamentales y académicas que velan por la calidad de la educación superior.

La investigación que se reporta pretendió dar respuesta a la siguiente pregunta: ¿cuáles son los patrones sociodemográficos, económicos, académicos e

institucionales asociados al desempeño académico de los estudiantes en las competencias genéricas de los programas profesionales en las pruebas Saber Pro 2011-2?

En relación a la estructura, el libro está organizado por capítulos. En el capítulo I, se conceptualiza acerca de la calidad de la educación superior, particularmente sobre la acreditación institucional y la acreditación de programas, por ser uno de los patrones que muestran los resultados del estudio; igualmente se conceptualiza sobre el desempeño académico y las competencias genéricas en la formación de profesionales. En el capítulo II, se describen las diferentes etapas del proceso de descubrimiento de conocimiento en bases de datos y las fases de la metodología para proyectos de minería de datos CRISP-DM, utilizada en ese proyecto de investigación. Por su parte, en el capítulo III se hace un análisis exploratorio de los datos y se muestran algunas tendencias sobre el desempeño académico con relación a las competencias genéricas. El capítulo IV refiere el desarrollo del proyecto de investigación aplicando cada una de las fases de la metodología CRISP-DM, que orientó el desarrollo del proyecto. En el capítulo V se evalúan, interpretan y discuten los resultados obtenidos. Finalmente, en el capítulo VI, los investigadores presentan las conclusiones de la investigación con base en los resultados obtenidos y hacen recomendaciones tanto para el campo de la educación superior en lo relacionado con la política educativa de calidad de la misma, como para nuevas investigaciones, haciendo uso de la información recopilada en las bases de datos del ICFES y teniendo en cuenta las técnicas de la minería de datos.

1 Desempeño académico y competencias genéricas en la formación de profesionales

Academic Performance and Generic Skills in Professional Training

Resumen

Este capítulo presenta las tendencias conceptuales sobre las variables y categorías que sustentan el proyecto de investigación titulado *Descubrimiento de patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales en las pruebas Saber Pro 2011-2*, que dio origen a este libro. La finalidad de las pruebas Saber Pro, que lleva a cabo el Instituto Colombiano para la Evaluación de la Educación (ICFES), es comprobar el grado de desarrollo de las competencias de estudiantes próximos a culminar el pregrado; en estas se evalúan competencias genéricas y específicas. En dichas pruebas los estudiantes deben presentar evaluación de los módulos de competencias genéricas sin importar el programa de formación que cursen. El período señalado para esta investigación incluye competencias de lectura crítica, comunicación escrita, razonamiento cuantitativo e inglés. La teoría pone en evidencia complejas relaciones subjetivas, sociales, académicas, económicas y culturales entre las temáticas que abordó la investigación y que afectan el desempeño académico del estudiante, el cual, en este caso, establece la relación entre el proceso de aprendizaje y el logro de competencias genéricas, que identifican elementos comunes a cualquier profesión.

Palabras clave: calidad de la educación, competencias genéricas, desempeño académico, pruebas Saber Pro.

Abstract

This chapter discusses conceptual trends on the variables and categories that support the research project titled *Discovery of Academic Performance Patterns in Generic Skills of Undergraduates in the Saber Pro Exam 2011-2*, which gave rise to this book. The objective of the Saber Pro Exam, applied by the Instituto Colombiano para la Evaluación de la Educación (ICFES), is to check the degree of development of both generic and specific skills of undergraduates who are about to get their degree. In such exam, students must take the test on the generic skills modules regardless of the undergraduate program they belong to. The period set for this research includes critical reading, written communication, quantitative reasoning and English skills. The theory brings to light complex subjective, social, academic, economic and cultural relationships among the topics addressed by the research that affect the academic performance of the student, who in this case establishes the relationship between the learning process and the achievement of generic skills that identify common elements to all professions.

Keywords: education quality, generic skills, academic performance, Saber Pro Exam.

¿Cómo citar este capítulo?/How to cite this chapter?

Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, S. J., Hidalgo Troya, A. y Alvarado Pérez, J. C. (2016). Desempeño académico y competencias genéricas en la formación de profesionales. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 19-62). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>



Introducción

Diversas investigaciones buscan identificar factores que determinan el desempeño exitoso en la formación profesional. El rastreo bibliográfico permite dar cuenta de una variedad de dichos factores que dan respuesta a preguntas de investigación de los estudios realizados y que presentan resultados pertenecientes a factores que corresponden a las dimensiones sociodemográfica, económica, académica e institucional. Por ejemplo, género, edad, trayectoria académica, características familiares, ocupación, nivel educativo de los padres, relaciones entre familia y la institución, cultura, infraestructura, modalidad de enseñanza, calidad de las instituciones de educación superior (IES) y formas de evaluación. Estos elementos ponen en evidencia relaciones subjetivas, sociales, académicas y económicas complejas que afectan de alguna manera el desempeño académico del estudiante.

Esta publicación, correspondiente a los resultados de la investigación sobre descubrimiento de patrones de desempeño académico en las competencias genéricas de las pruebas Saber Pro 2011-2, contribuye indiscutiblemente a la reflexión sobre la formación de profesionales en las competencias genéricas que se evaluaron en el período 2011-2, y cuya finalidad en este nivel de la educación superior es la formación de profesionales críticos, creativos y propositivos.

Uno de los objetivos de las pruebas Saber Pro es comprobar el grado de desarrollo de las competencias de estudiantes próximos a culminar el pregrado; son pruebas que evalúan competencias genéricas y específicas. Los estudiantes deben presentar los módulos de competencias genéricas sin importar el programa de formación que cursen, lo cual para las pruebas Saber Pro 2011-2 incluye competencias de lectura crítica, comunicación escrita, razonamiento cuantitativo e inglés.

El objetivo de este libro es presentar los patrones de desempeño académico en las competencias genéricas de los estudiantes de carreras profesionales, en relación con los factores de las dimensiones sociodemográfica, económica, académica e institucional, cuyos datos se encuentran almacenados en las bases de datos del ICFES, particularmente en las que corresponden al período 2011-2.

Aseguramiento de la calidad de la educación superior: tarea permanente de la universidad del siglo XXI

El discurso de la calidad de la educación superior se sitúa en la agenda de la universidad en el mundo a partir de la década de los noventa, al crearse organismos

propios para su desarrollo. Se ha avanzado considerablemente y se han superado obstáculos y tensiones, pero aún no se ha consolidado plenamente el sistema de aseguramiento de la calidad, pues implica superar concepciones dominantes, referidas a estándares y mediciones que promueven una funcionalidad elemental de la universidad, poniéndola a órdenes de un mercado indefinido, ambiguo e incierto. Se asume que la calidad en la educación superior es prioritaria y se convierte en valor determinante para los Estados, quienes ven la necesidad de establecer su regulación con el fin de afrontar las grandes diferencias entre las IES, y propender así a la superación de la heterogeneidad existente.

Conceptualización de calidad

Las diferentes definiciones de calidad coinciden en un concepto de construcción colectiva y gradual, que integra y articula visiones y demandas diferentes con los valores y propósitos de las IES. Desde el enfoque sociocultural, Rodríguez (2010) invita a mirar más allá de lo evidente, a situar las acciones humanas en su contexto histórico y a examinar las condiciones que llevan a la producción de conocimientos. Es por esta razón que el concepto de calidad de la educación superior transforma sus contenidos en cada época, caracterizándose por su inestabilidad, ya que es un concepto eminentemente histórico que depende de las demandas de la sociedad enmarcadas en su proceso de evolución.

En general, se ha establecido que se tiene o se posee calidad reconocida, cuando es posible acreditarla; por lo tanto, es preciso una definición que permita asumir adecuadamente este proceso. Al respecto, Aguila (2005) argumenta que la calidad tiene múltiples dimensiones e interpretaciones, lo importante es llegar a determinar aquel concepto que más convenga a la evaluación en las condiciones de la realidad de la educación. Es preciso tener en cuenta que la calidad está íntimamente relacionada con la pertinencia y el impacto, pues no se puede concebir una institución de educación superior de calidad que no sea pertinente en el entorno del que forma parte. Así, el concepto de calidad de la educación superior varía y se transforma, y lo afectan diversas dimensiones, siendo un reto para la universidad conocer permanentemente las relaciones que dinamizan dichos procesos para adecuarse a las nuevas exigencias de la sociedad y mantenerse en ella prestando el servicio educativo.

En la Conferencia Mundial sobre Educación Superior (Unesco, 1998) se menciona que la calidad es la acomodación del ser y quehacer de la educación

superior a su deber ser. Se desarrolla entonces esta idea en la cual se señala que cada uno de los elementos institucionales que componen la definición es evaluado, predominantemente, como una categoría específica. En palabras de Restrepo (2008), se manejan dos enfoques de calidad: el primero, es de carácter interno a las IES, el que está ligado a estándares cuantitativos e indicadores objetivos; el segundo, de carácter externo, es más cualitativo y está basado en las respuestas que la universidad presenta frente a las expectativas de la sociedad. De acuerdo con Fernández (2006), la calidad tiene diferentes connotaciones según los contextos donde se aplique: para los académicos se refiere a los saberes; para los empleadores, a las competencias; para los estudiantes, a la empleabilidad; para la sociedad, a ciudadanos respetables y competentes; para el Estado, según la concepción que asuma, varía en los aspectos vinculados con el desarrollo social y humano a la eficiencia, a los costos y a los requerimientos de capital humano.

Para Tünnermann (2008) la calidad de la educación superior es un concepto pluridimensional, que requiere comprender todas sus funciones sustantivas y que implica un proceso permanente de autoevaluación, definiendo normas comparativas de calidad, reconocidas en el plano internacional, con miras a tener en cuenta la diversidad y evitar la uniformidad; por lo tanto, las IES deben ser protagonistas y contribuir al proceso de autoevaluación como parte integral de ellas.

En este contexto, en los últimos años, se observa interés por la búsqueda de calidad, y en los debates universitarios surge este tema como inherente al servicio educativo que presta la universidad. No obstante, resulta complicado definir el concepto de calidad, siendo que proviene de la empresa, por lo cual no es fácil su traslado al ámbito educativo. El Informe de Bricall (2000) indica que este concepto ha sufrido un proceso evolutivo, que va desde la perspectiva cuantitativa hasta una vertiente que trata de satisfacer las demandas de los usuarios. El usuario (estudiante) y el productor (universidad) son piezas claves en el proceso de calidad; el primero, porque es el destinatario último del programa y es él quien define en primer lugar la calidad; el segundo, porque su objetivo es satisfacer la demanda de un producto de calidad, cuya financiación es eminentemente de naturaleza pública y, como tal, debe ser un servicio público de calidad.

Por su parte, Fernández (2006) hace una reflexión en torno a la calidad de la educación de América Latina, donde, durante las dos últimas décadas se han creado IES para atender la demanda creciente de educación superior, la mayoría de carácter privado y sin criterios previos de calidad y pertinencia, lo que genera diversificación y heterogeneidad en la calidad y origina problemas comunes en

los sistemas de este nivel educativo. Esta situación demanda establecer procesos de regulación para hacer frente a este problema, en particular de las IES privadas. Por otra parte, se presenta una creciente demanda sobre la relación IES y sociedad. En este orden de ideas, la imprescindible rendición de cuentas se convierte en componente principal de programas de docencia e investigación que demandan una relación efectiva entre Estado, sociedad y universidad.

La misión de la universidad colombiana debe considerar su vocación de servicio de calidad, pero según Luna y Hernández (2010) desafortunadamente hay universidades “de garaje”, que no cumplen con requisitos ni exigencias mínimas para brindar educación superior; estas perjudican la formación de profesionales y frenan el desarrollo investigativo de los pueblos. Dada la importancia de su labor, que implica gran responsabilidad y compromiso social, la universidad debe rendir cuentas a la sociedad con base en los procesos de autoevaluación y autorregulación, ya que la educación es un servicio público. Enfatizan los autores que el carácter de bien público social que tiene la educación superior se reafirma en la medida en que se establecen condiciones para favorecer el acceso a una formación de calidad, mediante estrategias consecuentes, tanto en lo público como en lo privado; por esta razón, es menester fortalecer mecanismos de certificación y acreditación que garanticen la transparencia y la condición de servicio público.

Una educación superior de calidad debe articularse con todo el sistema educativo a través del apoyo a la formación de sólidas bases cognitivas, y así promover y fomentar el desarrollo de valores, habilidades, destrezas y capacidades para adquirir, construir y transferir conocimientos en beneficio de la sociedad. Por esto es indudable la responsabilidad en la formación de profesionales que intervengan en la producción de contenidos educativos en la perspectiva de favorecer la enseñanza e investigación.

La calidad de la educación superior es la razón de ser del Sistema Nacional de Acreditación; reconocerla, velar por su incremento y fomentar su desarrollo otorga sentido a las acciones del Consejo Nacional de Acreditación, quien ha adoptado una aproximación integral al concepto de alta calidad al desarrollar su modelo, el cual ha permanecido a lo largo de su historia y ha sido validado nacional e internacionalmente. El Consejo ha procurado tener presentes todos los factores que inciden en ella y la manera como éstos pesan en un momento y en un contexto específico. (Colombia, Sistema Nacional de Acreditación [SNA], 2015, p. 13).

La universidad colombiana se identifica permanentemente con procesos de calidad, cuyos propósitos se sintetizan en fomentar y promover una cultura de mejoramiento continuo, con aumento en la satisfacción de la comunidad universitaria, mejoramiento de la calidad de cada proceso y puesto de trabajo, para alcanzar, inicialmente, la certificación, luego la acreditación de los programas y finalmente la acreditación institucional. La prioridad es el fortalecimiento del sistema de autorregulación, con la definición de indicadores y metas que perfilen el direccionamiento de las actividades para mejorar en un estrecho nexo con la calidad. Así pues, uno de los cambios fundamentales en la universidad consiste en comprender, asimilar e incorporar la dinámica de la autoevaluación y autorregulación en todas sus funciones sustantivas de docencia, investigación e interacción social, para lo cual se requieren reflexiones filosóficas y académicas, así como características que le permitan operacionalizar sus procesos en el marco de la evaluación de sus funciones misionales con pertinencia social.

Acreditación

De acuerdo con la Ley 30 (1992), en Colombia,

El Sistema Nacional de Acreditación (SNA), es el conjunto de políticas, estrategias, procesos y organismos, cuyo objetivo fundamental es garantizar a la sociedad que las instituciones de educación superior que hacen parte del sistema, cumplan con los más altos requisitos de calidad y que realicen sus propósitos y objetivos (artículo 53).

Hernández, Alvarado y Luna (2015), en la responsabilidad social de la universidad, conciben la acreditación como el acto por el cual el Estado adopta y hace público el reconocimiento que los pares académicos presentan sobre la comprobación que realizan en una institución acerca de la calidad de sus programas académicos, organización, funcionamiento y el cumplimiento de su función social. Ahora bien, la acreditación no es solo una oportunidad para el reconocimiento por parte del Estado sobre la calidad de un programa o una institución, es también una posibilidad para comparar la formación que se imparte con aquella, que se reconoce como válida y deseable.

El Consejo Nacional de Acreditación (CNA), creado como organismo académico por la Ley 30 de 1992 y conformado por académicos de las más altas calidades, se encarga de llevar a cabo el proceso de acreditación y recomienda al Ministerio de Educación Nacional (MEN) acreditar los programas e instituciones que lo merezcan.

La política pública sobre calidad educativa, basada en procesos de autoevaluación, certificación y acreditación, inició tarde, pues aunque se reglamentó en la década de los noventa, solo se ejecutó a inicios del siglo XXI. Así pues, presionada por la normatividad, la universidad ajusta su proceso de mejoramiento continuo. En opinión de Restrepo (2008), la autonomía motivó la necesidad de dar cuenta a la sociedad, sobre la calidad en relación con la misión, propósitos y objetivos que cumplen las instituciones y los programas de educación superior. Según los lineamientos del CNA, en los procesos de certificación y acreditación de programas e instituciones se estandarizan factores, características y aspectos, y se da libertad para plantear los indicadores que den cuenta de la calidad de los programas y de las instituciones, por lo cual los deben ajustar a su contexto. Concretamente, el CNA define la acreditación de la siguiente manera:

La Acreditación es el acto por el cual el Estado adopta y hace público el reconocimiento que los pares académicos hacen sobre la calidad de un programa o institución con base en un proceso previo de evaluación en el cual intervienen la institución, las comunidades académicas y el Consejo Nacional de Acreditación. El acto administrativo mediante el cual se expresa y hace pública la acreditación es expedido por el Ministerio de Educación. (Colombia, SNA, 2015, p. 57).

La universidad deriva la justificación de su existencia de las relaciones que establece con la sociedad; en consecuencia, al prestar servicios educativos en el nivel superior debe responder a necesidades que provienen de las instancias y niveles de la sociedad. Es por ello que está obligada a rendir cuentas sobre la calidad de dichos servicios. Como fruto de esta tendencia, aparecen los temas de cambios económicos, educativos y pedagógicos, centrados en la formación integral, la investigación y la evaluación. La educación superior hace énfasis en tendencias ligadas a la ciencia, la tecnología, el talento humano y el desarrollo sostenible, como frentes prioritarios de los contextos exigidos en el desarrollo de individualidades, en la reconstitución de las estructuras sociales basadas en acuerdos y nuevas reglas de convivencia. Según Chaparro (2008), el CNA adelanta un análisis de la evolución de la acreditación de alta calidad de programas académicos en Colombia; el autor analiza tres aspectos complementarios: (a) el número total de programas acreditados; (b) la demanda por acreditación medida en términos del número de solicitudes que se reciben semestralmente para acreditación de programas, y (c) el número de programas en proceso de evaluación.

En Colombia, según el MEN (2015), entre 1998 y 2014, las cifras que hablan cuantitativamente de la calidad de la educación son preocupantes para este sector. Existen muy pocos programas académicos acreditados en alta calidad; por ejemplo, de los 7079 programas de IES oficiales, solo están acreditados 443. La pregunta sería entonces, ¿cuál es el nivel de formación profesional que ofrecen los restantes 6636 programas académicos que no están acreditados en alta calidad? Igual panorama se presenta en las IES privadas, ya que de los 3647 programas académicos se han acreditado 437, lo que significa que 3210 programas no son de alta calidad; por lo cual, de los 10 726 programas académicos existentes en el país, solo han obtenido la acreditación de alta calidad un mínimo del 8,2 %, porcentaje que corresponde a 880 programas. Las cantidades expuestas anteriormente son el fiel reflejo de la calidad en la formación profesional, la cual es discriminatoria porque tiene mayores posibilidades de formación de excelencia solamente un mínimo de jóvenes que pueden acceder a los programas de alta calidad que permiten mayores posibilidades en el campo laboral; mientras que la mayoría de los estudiantes universitarios se ubican en instituciones y en programas que no cumplen con todas las condiciones que los procesos de calidad exigen.

Tabla 1
Factores y características que se consideran en la evaluación de la calidad institucional en Colombia

N.º	Aspectos	Características
1	Misión y proyecto institucional	<ul style="list-style-type: none"> - Coherencia y pertinencia de la misión - Orientaciones y estrategias del proyecto educativo institucional - Formación integral y construcción de la comunidad académica en el proyecto educativo institucional
2	Estudiantes	<ul style="list-style-type: none"> - Deberes y derechos de los estudiantes - Admisión y permanencia de estudiantes - Sistemas de estímulos y créditos para estudiantes
3	Profesores	<ul style="list-style-type: none"> - Deberes y derechos del profesorado - Planta profesoral - Carrera docente - Desarrollo profesoral - Interacción académica de los profesores
4	Procesos académicos	<ul style="list-style-type: none"> - Políticas académicas - Pertinencia académica y relevancia social - Procesos de creación modificación y extensión de programas académicos

(continúa)

(viene)

N.º	Aspectos	Características
5	Visibilidad nacional e internacional	- Inserción de la institución en contextos académicos nacionales e internacionales - Relaciones externas de profesores y estudiantes
6	Investigación y creación artística y cultural	- Formación para la investigación - Investigación
7	Pertinencia e impacto social	- Institución y entorno - Graduados e institución
8	Autoevaluación y autorregulación	- Sistemas de autoevaluación - Sistemas de información - Evaluación de directivas, profesores y personal administrativo
9	Bienestar institucional	- Estructura del bienestar institucional
10	Organización, gestión y administración	- Administración y gestión - Procesos de comunicación - Capacidad de gestión
11	Recursos de apoyo académico e infraestructura física	- Recursos de apoyo académico - Infraestructura física
12	Recursos financieros	- Recursos, presupuesto y gestión financiera

Nota. Tomado de *Lineamientos para la acreditación institucional*. Consejo Nacional de Acreditación –CNA, por Sistema Nacional de Acreditación [SNA], 2015.

Según el MEN (2015), en Colombia existen pocas instituciones acreditadas en alta calidad. En efecto, de las 71 IES oficiales, solo 13 están acreditadas, y de las 186 IES privadas se han acreditado 22; por lo tanto, de los 257 IES existentes en Colombia, solo ha obtenido la acreditación de alta calidad el 13,6 %, porcentaje que corresponde a 35 IES. Esta realidad amerita reflexión y acción de la política pública de educación superior, pero ante todo de quienes prestan el servicio educativo en la formación de profesionales.

Formación profesional mediante el enfoque de competencias

Según la misión de la universidad, esta tiene un reto frente a la cultura de globalización, en la cual la formación de profesionales se sitúa en el marco de una sociedad global y en el núcleo del desarrollo de la persona y las comunidades; de tal modo que, como lo plantea Delors (1999), “la educación tiene la misión de permitir a todos, sin excepción hacer fructificar todos sus talentos y todas sus capacidades

de creación, lo que implica que cada uno pueda responsabilizarse de sí mismo y realizar su proyecto personal” (p. 13).

La responsabilidad que la universidad tiene con la sociedad le impone el desafío de formar profesionales competentes, lo que ha de evidenciarse en los procesos académicos como parte de la formación integral, cuyos fines y propósitos se circunscriben en una educación superior de calidad, de tal suerte que su impacto debe reflejarse en los distintos sectores de la sociedad. Teniendo en cuenta a Hernández y Hernández (2013), es en la universidad —entidad en la que se crea y se recrea el conocimiento— donde se deben proponer las soluciones a los problemas de la sociedad, y generar las propuestas para el aprovechamiento óptimo de las fortalezas y potencialidades existentes; por tanto, se requieren profesionales competentes. En este contexto, la formación por competencias surge como una alternativa para responder a las demandas sociales en procura de disponer de modelos educativos con mayor pertinencia y calidad, en los que, según Umaña (2008), una de las mayores exigencias está en la articulación entre educación y sociedad, en particular en el ámbito laboral; es decir, en el hecho que los estudiantes obtengan una formación acorde a los requerimientos sociales y con una participación efectiva en su desempeño laboral.

Cabe señalar que el contexto de acción de la universidad se encuentra atravesado por relaciones reales de mercado, que estipulan la dinámica interna de la institución; no obstante, esta debe responder a su esencia misma, con la generación del saber, a partir de la formación de profesionales. Según Hernández y Hernández (2013), para cumplir con esta responsabilidad el quehacer de la universidad debe estar encaminado a propiciar una formación de actualidad, relevante y pertinente, mediada por el conocimiento y un desarrollo cultural trascendente en escenarios y condiciones, donde el estudiante tenga la posibilidad de convertirse en un profesional responsable y ejerza su profesión de manera competente.

En el siglo XXI, la formación por competencias se ha constituido en el eje articulador de la formación profesional en Colombia. El MEN (2013) define la competencia como el conjunto de conocimientos, habilidades, actitudes, comprensiones y disposiciones cognitivas, socioafectivas y psicomotoras, relacionadas entre sí para hacer posible el desempeño eficaz y eficiente de una actividad en contextos relativamente nuevos y retadores; enfoque que está en sintonía con las tendencias mundiales de la educación. Según Hernández y Luna (2014), los horizontes y los lindes educativos se abrieron y cambiaron así las responsabilidades y los roles de estudiantes y docentes, los cuales optan por moverse hacia metodologías de

trabajo que permitan el logro de las competencias, integrando nuevas herramientas, actividades, espacios y formas de aprender.

Las exigencias del modelo de competencias cuestiona el rol de las IES, a las que les asigna la responsabilidad de propiciar espacios de aprendizaje, donde los estudiantes aprendan a aprender y a crear nuevos conocimientos. En este sentido, Salinas (2006 y 2008) señalan que las IES se encuentran en transición; los permanentes cambios en la sociedad del conocimiento demandan procesos de enseñanza y de aprendizajes flexibles y accesibles.

El modelo de formación profesional vigente en la universidad, según Mora (2004), responde a necesidades de una sociedad y de un mercado laboral que están desapareciendo; de modo que el contexto al que debe dar respuesta la educación superior está en transformación y, por tanto, es necesario que también se modifique el modelo de formación profesional, para responder favorablemente a los requerimientos del contexto. A nivel global, se configura un discurso académico, alimentado por reflexiones y discusiones sobre indicadores y valoración de resultados, que reclaman por parte de la universidad compromiso y ajuste curricular hacia un mercado laboral imprevisible e impredecible, en el cual las competencias no pueden confundirse con habilidades y destrezas que, por ser mecánicas, son incapaces de afrontar la incertidumbre y la complejidad de la vida contemporánea.

Fundamentación conceptual

Competencia es un término que en el campo de la educación ha generado una amplia gama de criterios respecto a su interpretación; en algunas elaboraciones teóricas presenta significados diversos y en ocasiones abstractos, pues se definen desde un enfoque conceptual, más que científico. Refiere Álvarez (2008) que las competencias son un tema de actualidad en el debate educativo, pero aún carecen de un campo semántico preciso, por lo cual se considera un término polisémico. El Proyecto Tuning para América Latina (Beneitone et al., 2007), concibe que el término competencia no es simple, conlleva la noción de producción y transmisión de conocimiento, relación entre educación y sociedad, misión y valores del sistema educativo, prácticas de enseñanza, evaluación y procesos de aprendizaje” (p. 35). Tobón, Pimienta y García (2010) afirman que las competencias son actuaciones integrales que permiten identificar, interpretar, argumentar y resolver problemas con idoneidad y compromiso ético, movilizandolos saberes ser, hacer y conocer.

Los autores Hernández y Luna (2013) entienden las competencias como el encadenamiento de saberes, orientados y articulados a la concepción del ser, del saber, saber hacer y del saber convivir. Por lo tanto, el desarrollo de competencias constituye un proceso de formación integral y global del profesional en sus dimensiones como persona, ciudadano y profesional, a la hora de enfrentarse a problemas reales. Esta concepción implica cambios sustanciales en los procesos de formación profesional que llevan a cabo las universidades, desde lo curricular, lo didáctico y, sobre todo, desde lo actitudinal, para asumir el aprendizaje cooperativo, significativo, autónomo, orientado por el pensamiento complejo, crítico, creativo e innovador.

Por su parte, Martínez, Cegarra y Rubio (2012) puntualizan que la formación por competencias pretende que las personas desarrollen capacidades amplias, aprendiendo y desaprendiendo permanentemente y adaptándose a situaciones de un mundo en constante transformación. En este sentido, según Schunk (2008), la adopción de este modelo supone mover el vértice de la enseñanza al aprendizaje, entendiéndolo como un proceso que permite captar, codificar, relacionar y almacenar nueva información con aquella que existe; dicho proceso forma parte de todas las actividades cognitivas y significa un cambio de cultura en el campo educativo.

Según el ICFES (2013), las competencias son el componente articulador de los diferentes niveles de formación, se constituyen en los referentes comunes sobre lo que las personas deben saber y saber hacer durante su trayectoria en el sistema educativo. Se reconoce que toda persona, independientemente de su profesión, debe desarrollar capacidades para aprender durante toda la vida; por ejemplo, comprender contextos y situaciones, comunicarse con los otros, seleccionar y analizar información, solucionar problemas de distinta naturaleza en un mundo incierto y cambiante. De esta manera, el enfoque de las competencias es el discurso necesario por excelencia en el debate de la calidad de la educación superior, específicamente en el de la formación profesional, que según Lorente (2012) implica una nueva ortodoxia en la pedagogía y en el quehacer educativo, al concebirse como un adecuado dispositivo para la reestructuración de las instituciones y de la misión de la universidad en las sociedades del conocimiento.

La formación basada en competencias se origina en la interrelación de distintos elementos, pero fundamentalmente, según Ruiz (2009), por la convergencia de dos acontecimientos claves: primero, el replanteamiento de la educación en términos de facilitar el aprendizaje coligado al concepto de aprendizaje como un

fenómeno del individuo que aprende; segundo, la formación de profesionales capaces de tomar decisiones acertadas y plantear de manera eficaz la solución de problemas complejos en situaciones concretas y reales, sin menoscabo del saber en lo conceptual, procedimental y actitudinal. Este modelo de la formación por competencias demanda cambiar una lógica basada en discurso teórico a otra que implica la solución de problemas reales que vivencia la sociedad en su contexto. Según Climent (2014), este enfoque está relacionado con la influencia de la tendencia globalizante; concretamente, por la economía global. Por esta razón, la formación por competencias tiende hacia relaciones de causa y efecto con los procesos globalizadores.

Los cambios de paradigmas en la formación profesional implican cambios epistemológicos, axiológicos y actitudinales. Según Hernández y Luna (2013), incorporar a la formación profesional el concepto de competencias implica acciones muy particulares, no como consecuencia de un proceso de aprendizaje superficial desde lo memorístico, sino de un aprendizaje significativo en el cual se desarrollen y acrecienten las capacidades humanas mediante el proceso integral de las dimensiones cognoscitivas, en las cuales la investigación, la creación y la innovación sean la base para la toma de decisiones y la solución de problemas en contextos determinados. Desde las competencias, la educación es un servicio pertinente a las demandas sociales; en este punto toman significado el *ser* y *saber cómo*, en una formación profesional capaz de ofrecer aprendizajes que habilitan para desempeñarse con eficacia en el contexto específico de las dificultades y los retos.

Es necesario precisar que en la formación profesional bajo el modelo de competencias, según intereses, predominan tres enfoques identificados por Achaerandio (2010): primero está el *enfoque educativo*, implementado por universidades innovadoras responsables de formar a los profesionales del siglo XXI, los centros de investigación en educación, con la intención de crear conocimiento que fortalezca el desempeño de los nuevos profesionales; también están los ministerios de educación de países interesados en transformar y reorientar sus procesos de desarrollo a partir de una educación de calidad. El segundo es el *enfoque empresarial*, que tiene como finalidad formar el capital humano para incentivar y generar productividad. El tercero es el *enfoque humanista social*, representado en pensadores de la educación, teóricos y prácticos interesados y comprometidos con el desarrollo humano sustentable, quienes marcan la tendencia de la formación por competencias e impactan, desde el conocimiento, al sector productivo.

Competencias genéricas en la formación de profesionales

Las competencias genéricas identifican los elementos comunes a cualquier profesión. Entre dichas competencias, señalan González y Wagenaar (2003), están la capacidad de aprender, la toma de decisiones, el diseño de proyectos, las habilidades interpersonales, las cuales se complementan con las competencias específicas de cada área de estudio. En un mundo caracterizado por la incertidumbre y una sociedad en permanente transformación, las competencias genéricas son relevantes; su elección es un punto de referencia dinámico que aporta ventajas a la formación de profesionales.

Según Pinilla (2011), el Proyecto Tuning en Europa agrupó las competencias genéricas en *instrumentales*, *interpersonales* y *sistémicas*. Las primeras se refieren a la capacidad de análisis y síntesis, la comunicación oral y escrita, el conocimiento de idiomas, el manejo del computador y gestión para buscar información; las segundas se relacionan con el trabajo en equipo interdisciplinario, comunicación con expertos de otras áreas, reconocimiento de la diversidad y multiculturalidad y compromiso ético; y las terceras corresponden a la aplicación del conocimiento en la práctica, en investigación, creatividad, liderazgo, trabajo autónomo, diseño y gestión de proyectos.

En América Latina, la importancia asignada a las competencias genéricas es destacable y coincide con los ejes del debate de la educación universitaria para el siglo XXI, los cuales González y Wagenaar (2003) los relacionan así: una educación centrada en el estudiante, una pedagogía transversal y transdisciplinar, una educación de calidad, pertinente y transparente; además, se debate sobre la duración de las carreras y modalidades de formación profesional y sobre los requerimientos para el desarrollo de un espacio universitario de Latinoamérica, el Caribe y Europa. En este sentido, Beneitone et al. (2007) afirman que para América Latina se tomó la lista de las 30 competencias genéricas identificadas en Europa, como base del trabajo del Proyecto Tuning de los 18 países participantes; esta reflexión dio como resultado 85 competencias genéricas que, posterior a los estudios realizados, se concretaron en 27 competencias genéricas, de las cuales se destacan, según el concepto de académicos, expertos, estudiantes y empresarios, las cinco más importantes, a saber: capacidad para aplicar los conocimientos en la práctica, capacidad para aprender y actualizarse, capacidad para identificar, plantear y resolver problemas, compromiso ético y compromiso con la calidad. Teniendo en cuenta lo expuesto por el ICFES (2012a), estas competencias son evaluadas en varios países; así es como la Organización para la Cooperación y el

Desarrollo Económico (OCDE) las incluyó en su proyecto de Evaluación Internacional de Competencias en Educación Superior.

Las competencias genéricas, bajo presupuestos del Proyecto Tuning para América Latina, según Beneitone et al. (2007), serán las que en virtud de su carácter como habilidades transferibles dotarán a los estudiantes de una formación más versátil para ejercer su función social como personas, ciudadanos y profesionales en el contexto social; por tanto, la universidad aborda retos interesantes, entre los cuales está el modelo pedagógico centrado en el estudiante. Esto implica que la formación en competencias genéricas comunes a todo tipo de carreras profesionales sea el eje central por medio del cual se construya ciudadanía, identidad y trabajo.

Las competencias genéricas dotan de relevancia y capacidad de transferencia a la universidad, al programa y al profesional en su ideario e inserción social. Al respecto, Tobón et al. (2006) clasifican las competencias genéricas de acuerdo con el grado de amplitud de las competencias así: competencias que tienen un carácter global, es decir, son amplias y se relacionan con áreas de desempeño; y las competencias que, a diferencia de las anteriores, son concretas y se refieren a actividades generales dentro de las mismas competencias y no a áreas de desempeño.

En los estudios de González y González (2008), se destaca la importancia de las competencias genéricas en las condiciones del desempeño profesional que exige el siglo XXI, a saber: ejercer eficientemente la profesión en contextos diversos, con autonomía, flexibilidad, ética y responsabilidad. Por lo tanto, la versatilidad, la flexibilidad mental, la capacidad para adaptarse a nuevos desafíos, el saber en cuanto a cómo resolver problemas y situaciones problemáticas, y la preparación para la incertidumbre son habilidades mentales que requieren los profesionales y en las que es necesario formarlos para propiciar equilibrio ante los cambios y capacidad de inserción en contextos diversos. En este marco, la importancia de las competencias genéricas en el contexto profesional se produce en la medida en que estas se vinculan al desempeño profesional.

Las competencias genéricas, refiere Peña (2008), son transversales a todos los niveles educativos, son transferibles a diversidad de funciones y se constituyen en una respuesta a las necesidades de la sociedad. En Colombia existe un consenso en la comunidad educativa en torno a la importancia de las competencias genéricas, considerándolas transversales en la formación de profesionales, mas son indispensables para el desempeño académico y laboral. En el período 2008-2009, el MEN (2009) identificó las siguientes competencias genéricas por desarrollar en

la educación superior: comunicación en lengua materna y en otra lengua internacional, pensamiento matemático, ciudadanía y ciencia, y tecnología y manejo de la información.

Al respecto, complementa Beltrán (2011) que desde el 2008 el MEN emprendió acciones tendientes a la formulación de competencias genéricas a todos los núcleos de formación en educación superior, que posibiliten un monitoreo de la calidad de la educación superior en el país y que puedan constituirse en un elemento articulador de todos los niveles educativos: inicial, básica, media y superior. Este autor clasifica en tres grupos las competencias genéricas de acuerdo con las necesidades educativas del país. Primero, competencias abstractas del pensamiento, en el que incluye razonamiento crítico, entendimiento interpersonal, pensamiento creativo, razonamiento analítico y solución de problemas. Segundo, conocimientos y competencias prácticas necesarias para el despliegue de las competencias abstractas; en este grupo relaciona el conocimiento del entorno, la comunicación, el trabajo en equipo, la alfabetización cuantitativa, el manejo de información, la comunicación en inglés y las tecnologías de la información y la comunicación (TIC). Tercero, dinamizadores para el desarrollo de las competencias genéricas, grupo en el cual relaciona el saber aprender y el saber recontextualizar.

Expone Beltrán (2011) que el desarrollo e implementación de competencias genéricas transversales en la formación de profesionales, en todos los niveles y en los diferentes énfasis y programas de formación, se configuran como respuesta a las necesidades de la sociedad actual; esto teniendo en cuenta que las competencias genéricas no pueden suplantar las competencias específicas que se forman en los distintos programas de educación superior, pues su desarrollo con respecto a estas últimas tiene un carácter complementario e integrador. En efecto, las circunstancias del mercado actual, al que todo estudiante universitario se debe enfrentar, hacen que la universidad no limite su papel a proporcionar únicamente formación académica, sino también una educación integral que viabilice que los retos y avances contemporáneos se articulen al saber ser en sociedad.

Según lo establecido por el MEN, en el Decreto 3963 (2009), en Colombia las competencias genéricas se definen como aquellas necesarias para el adecuado desempeño profesional, independientemente del programa que se curse. En coherencia con esta concepción, las pruebas Saber Pro se estructuran en torno a módulos concebidos como instrumentos que evalúan no solo las competencias consideradas fundamentales para los futuros egresados de programas de formación, sino también aquellas que deben desarrollar todos los estudiantes de educación

superior, aparte del programa o nivel en el que se encuentren. Para el 2011, la evaluación de competencias genéricas incluye competencias de razonamiento cuantitativo, lectura crítica, escritura e inglés, las cuales a nivel internacional, en el campo de la educación superior, son calificadas como “atributos” necesarios para cualquier egresado.

Las competencias genéricas para el siglo XXI, según el ICFES (2013), son transversales a las áreas curriculares y del conocimiento; en el caso de la formación profesional, los estudiantes las desarrollan y las logran a través del trabajo en una o más áreas del conocimiento, y lo hacen en un determinado momento y circunstancias, pudiendo ser transferidas a otros contextos. Este tipo de competencias se consideran transversales, es decir, están presentes en todos los niveles de formación y su complejidad es progresiva a lo largo de toda su formación; por lo tanto, se entiende que las deben desarrollar todas las personas, aunque no en el mismo nivel.

Los módulos de competencias genéricas determinadas por el ICFES (2011a) para las pruebas Saber Pro 2011-2 son las siguientes:

- **Lectura crítica.** Evalúa competencias relacionadas con la capacidad para leer de manera analítica y reflexiva; además, implica comprender los planteamientos expuestos en un texto e identificar sus perspectivas y juicios de valor. Esto exige que el lector identifique y recupere información presente en uno o varios textos, que construya su sentido global, establezca relaciones entre enunciados y evalúe su intencionalidad. Aborda procesos como ubicar información, relacionar, construir la representación global del texto, analizar y evaluar la relación entre procesos discursivos y contexto sociocultural.

Tabla 2

Lectura crítica: categorías, concepto y criterios de evaluación

Categoría	Concepto	Criterios de evaluación
Ubicar información	Es la capacidad de encontrar información específica en los textos según un criterio de búsqueda dado. La información que se pide hallar puede ser más o menos evidente y solicitarse empleando las mismas expresiones que aparecen en el texto o utilizando otras equivalentes.	<ul style="list-style-type: none"> - Encontrar la información requerida en los textos que se leen (verbales o no verbales), sea esta prominente o marginal. - Encontrar información dentro del texto, solicitada utilizando las mismas expresiones del texto. - Encontrar información dentro del texto, solicitada utilizando expresiones equivalentes a las del texto.

(continúa)

(viene)

Categoría	Concepto	Criterios de evaluación
Relacionar	Es la identificación de las relaciones entre distintas partes de los textos. Las relaciones pueden ser de implicación, inclusión, pertenencia, causalidad, orden, ejemplificación, categorización, equivalencia, complementariedad, oposición, contradicción o contraste, analogía o contra argumentación.	<ul style="list-style-type: none"> - Analizar cómo se relacionan lógicamente las proposiciones en un texto. - Caracterizar las relaciones que se presentan entre dos o más proposiciones. - Establecer la validez de las relaciones lógicas propuestas. - Analizar las relaciones de cohesión entre segmentos del texto. - Reconocer marcadores que dan cohesión al texto y sus funciones. - Usar adecuadamente los marcadores de cohesión para establecer las relaciones que hay entre proposiciones.
Construir la representación global del texto	Se refiere a la comprensión del texto como un todo y la construcción del sentido global a partir de la interpretación de sus componentes implícitos y explícitos.	<ul style="list-style-type: none"> - Identificar los significados presentes en el texto y reconocer cómo se configuran las relaciones entre ellos. - Reconocer cómo la estructura del texto contribuye a la elaboración del sentido global del mismo. - Reconocer cómo se configuran las relaciones de quienes participan en la construcción de significados en un texto, y cómo estas relaciones configuran sus roles en el mismo. - Comprender cómo las relaciones entre los significados presentes en un texto, y las relaciones que se dan entre personajes, interlocutores, emisores, receptores, contribuyen a la configuración del sentido global del texto.
Analizar y evaluar la relación entre procesos discursivos y contexto sociocultural	Es la capacidad para tomar distancia del texto y rastrear las concepciones de mundo subyacentes, mediante la identificación de las estrategias discursivas utilizadas y el reconocimiento del rol de quienes participan en la materialización de los discursos.	<ul style="list-style-type: none"> - Analizar el texto en su nivel profundo, estableciendo relaciones entre su contenido, el propósito comunicativo y la audiencia a la que se dirige. - Reconocer perspectivas y juicios de valor presentes en el texto. - Establecer relaciones entre el uso de un tipo específico de lenguaje (técnico, periodístico, científico), o de un tipo de texto en particular, su propósito comunicativo y su audiencia. - Evaluar la forma en que se relacionan prácticas socioculturales y discursos. - Reconocer cómo los discursos configuran prácticas socioculturales. - Reconocer cómo las prácticas socioculturales configuran discursos.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

- **Comunicación escrita.** Evalúa la competencia para comunicar por escrito ideas referidas a un tema dado. Cada estudiante debe elaborar un escrito argumentativo según un tema planteado. Incluye las siguientes acciones: expresar adecuadamente la intención comunicativa, dar coherencia y cohesión al texto, expresarse mediante un lenguaje apropiado, aplicando las reglas que rigen el lenguaje escrito.

Tabla 3

Comunicación escrita: categorías, concepto y criterios de evaluación

Categorías	Nivel I	Nivel II	Nivel III
Intención y propósito	- Respuesta no definida sobre el tema. - Respuesta incoherente. - Texto vago o impreciso.	Respuesta adecuada sobre el tema propuesto.	Respuesta apropiada al tema propuesto.
Estructura y organización	- Texto desorganizado. - Texto desarticulado.	Texto organizado que evidencia alguna estructura argumentativa.	Texto organizado y con clara evidencia de una estructura argumentativa.
Manejo del lenguaje	Texto limitado en cuanto a convenciones, reglas y normas de redacción y gramática.	Manejo adecuado de convenciones, reglas y normas de redacción y gramática. Texto comprensible y claro.	Manejo apropiado de las convenciones, reglas y normas de redacción y gramática. Estilo apropiado y consistente.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

- **Razonamiento cuantitativo.** Evalúa la comprensión de conceptos básicos de matemáticas para analizar, modelar y resolver problemas, aplicando métodos y procedimientos cuantitativos y esquemáticos, basados en las propiedades de los números y en las operaciones de las matemáticas. Incluye interpretación de datos, formulación y ejecución de problemas, evaluación y validación de resultados.

Tabla 4

Razonamiento cuantitativo: categorías, concepto y criterios de evaluación

Categoría	Concepto	Criterios de evaluación
Interpretar datos	Consiste en la comprensión e interpretación de datos presentados de diferentes formas (tablas, gráficas, esquemas, símbolos, expresiones verbales), así como la generación de diversas representaciones a partir de datos dados.	<ul style="list-style-type: none"> - Comprender y manipular la información presentada en distintos formatos. - Reconocer y obtener piezas de información a partir de descripciones, series, gráficas, tablas y esquemas. - Comparar distintas formas de representar una misma información. - Relacionar los datos disponibles con su sentido o significado dentro de la información.
Formular y ejecutar	Involucra procesos relacionados con la identificación del problema, la construcción y proposición de estrategias adecuadas para su solución además del tratamiento de datos, la modelación y el uso de herramientas cuantitativas (aritméticas, métricas, geométricas, algebraicas elementales y de probabilidad y estadística).	<ul style="list-style-type: none"> - Plantear procesos y estrategias adecuados para resolver un problema. - Seleccionar la información relevante y establecer relaciones entre variables pertinentes durante el proceso de análisis. - Proponer soluciones pertinentes a las condiciones presentadas en la información. - Diseñar planes, estrategias y alternativas para la solución de problemas. - Utilizar herramientas cuantitativas (tratamiento de datos). - Realizar cálculos sencillos para la ejecución de un plan.
Evaluar y validar	Incluye procesos relacionados con la verificación de resultados, hipótesis o conclusiones que se derivan de la interpretación y de la modelación de situaciones.	<ul style="list-style-type: none"> - Validar procedimientos y estrategias matemáticas utilizadas para dar solución a problemas. - Identificar las fallas o limitaciones de la información que se le presenta. - Identificar fortalezas y debilidades de un proceso propuesto para resolver un problema. - Aplicar estrategias cuantitativas y esquemáticas orientadas a validar, corregir, o descartar soluciones obtenidas a problemas propuestos.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

- **Inglés.** Evalúa la competencia para comunicarse efectivamente en inglés y se alinea con el Marco Común Europeo; permite además clasificar a los examinados en cuatro niveles de desempeño A1, A2, B1 y B2.

Tabla 5

Inglés: categorías, concepto y criterios de evaluación

Categoría	Criterios de evaluación
Nivel A1	<ul style="list-style-type: none"> - Comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. - Presentarse a sí mismo y a otros, pedir y dar información básica sobre domicilio, pertenencias y personas conocidas. - Relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad.
Nivel A2	<ul style="list-style-type: none"> - Comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia especialmente relevantes (información básica sobre él mismo y su familia, compras, lugares de interés, ocupaciones, etc.). - Comunicarse cuando se llevan a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones conocidas o habituales. - Describir en términos sencillos aspectos del pasado y del entorno, así como cuestiones relacionadas con las necesidades inmediatas del estudiante.
Nivel B1	<ul style="list-style-type: none"> - Comprender los contenidos principales de textos claros y en lengua estándar, si tratan cuestiones conocidas. - Desenvolverse en la mayoría de situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua. - Producir textos sencillos y coherentes sobre temas que son familiares al estudiante o en los que tiene un interés personal. - Describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones o explicar sus planes.
Nivel B2	<ul style="list-style-type: none"> - Entender las ideas principales de textos complejos que traten de temas concretos y abstractos, incluso si son de carácter técnico, siempre que estén dentro de su especialización. - Relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad, de modo que la comunicación se realice sin esfuerzo por los interlocutores. - Producir textos claros y detallados en torno a temas diversos, así como defender un punto de vista sobre temas generales.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Evaluación por competencias

El modelo de formación por competencias incluye el proceso de evaluación por competencias, un nuevo paradigma en relación con las demandas del sistema educativo y las necesidades de la sociedad contemporáneas; una experiencia integradora que procura al estudiante ampliar sus fortalezas y superar sus dificultades, mediante el logro de conocimientos, habilidades, destrezas, actitudes, aptitudes y valores en consonancia con la estructura cognitiva. En este nuevo enfoque de la evaluación, la retroalimentación es una fase esencial y se debe procurar que sea oportuna y con asertividad. Para Gómez et al. (2013), la evaluación de competencias genera la formación del aprendizaje autónomo y significativo, e impone la necesidad de reconceptualizar los fundamentos y reelaborar las prácticas educativas en el ámbito universitario. Este tipo de evaluación, orientada al aprendizaje toma como fundamento la implicación de los estudiantes en el proceso evaluativo, la realización de tareas realistas y la importancia de la retroalimentación, propiciando la evolución de la evaluación, del simple control a una evaluación sostenible que favorezca la adquisición de estrategias que sirvan para la autorregulación del aprendizaje.

Tobón et al. (2010) consideran que la evaluación de las competencias es una experiencia significativa de aprendizaje, que tiene en cuenta los logros y los aspectos por mejorar en cierta competencia, según criterios acordados y evidencias pertinentes, en el marco del desempeño del estudiante en la realización de actividades o el análisis, comprensión o resolución de problemas del contexto profesional, social, disciplinar e investigativo, considerando el saber, el saber conocer, el saber hacer y el saber convivir.

Al respecto, Cano (2008) considera que la evaluación por competencias obliga al docente a utilizar instrumentos y a implicar a diferentes agentes; en este punto se toman muestras de las ejecuciones de los estudiantes y se utiliza la observación como estrategia de recogida de información sistemática. El autor afirma que la evaluación debe ser coherente con el resto de elementos del diseño formativo; además, hace más conscientes a los estudiantes de cuál es su nivel de competencias, de cómo resuelven las tareas y de qué puntos fuertes deben potenciar y qué puntos débiles deben corregir para enfrentarse a situaciones de aprendizaje futuras. Teniendo en cuenta a Fernández (2010), la evaluación de competencias informa permanentemente al estudiante sobre el progreso en su aprendizaje; se basa en el acceso a fuentes múltiples y variadas de información con el fin de determinar si ha alcanzado el nivel esperado de desarrollo en las competencias, así como un grado

suficiente de dominio de los recursos vinculados a cada competencia. Para ello, la evaluación deberá realizarse en una situación lo más auténtica posible; dicho de otra manera, al contexto profesional donde los estudiantes podrán evolucionar una vez hayan terminado los estudios.

La evaluación de competencias, según Pérez (2013), es un proceso de descripción e interpretación del desarrollo del estudiante, sus cualidades, sus fortalezas y debilidades, el grado de configuración actual de cada uno de los componentes de las competencias, conocimientos, habilidades, emociones, actitudes y valores, así como el funcionamiento de las competencias como conjuntos, como sistemas de comprensión, toma de decisiones y actuación. La evaluación, concluye en un informe comprensible y accesible a todos, pero principalmente al evaluado, para que la información aportada pueda ayudar a tomar decisiones y autorregularse en el futuro a corto, medio y largo plazo.

Desempeño académico

Antecedentes

El tema del desempeño académico es relevante en el marco de la formación profesional. Tras revisar diferentes estudios, se precisa que no es un fenómeno reciente dentro del ámbito universitario, su literatura es extensa y de larga data. El estudio pionero en el campo es el de Coleman (1966), en el que se trabajó con más de 500 mil estudiantes y que está encaminado a examinar la relación entre los insumos del proceso de formación y su producto final; está relacionado además con la política educativa. En el estudio se concluye que el desempeño académico en los Estados Unidos estaba influenciado en gran medida por las características socioeconómicas de los estudiantes y que las variables asociadas a la institución educativa tenían poco o ningún efecto sobre las diferencias en el desempeño académico. Desde entonces, estos resultados animaron innumerables estudios que han reafirmado o refutado tal hipótesis.

En la investigación titulada *El rendimiento académico: concepto, investigación y desarrollo*, realizada por Navarro (2003), se consideran distintas perspectivas teórico-metodológicas, desde las que conceptualiza el desempeño académico como un constructo susceptible de adoptar valores cuantitativos y cualitativos, a través de los cuales se encuentra que existe una aproximación a la evidencia y dimensión del perfil de habilidades, conocimientos, actitudes y valores desarrollados

por el estudiante en el proceso de formación. El autor destaca que el desempeño académico forma una red de articulaciones cognitivas, considera que es dinámica, cuyos rasgos característicos distinguen los resultados de cualquier proceso de enseñanza aprendizaje.

Por otra parte, Jano y Ortiz (2005) realizan en España la investigación *Determinación de los factores que afectan al rendimiento académico en la educación superior*, por medio de la cual llegan a la conclusión de que uno de ellos es el esfuerzo efectivo realizado por el estudiante, como también las habilidades y conocimientos previos a su ingreso a la universidad. Las variables que utilizaron para el análisis se clasifican en cuatro dimensiones: variable por explicar, que mide el desempeño académico; las que hacen referencia al esfuerzo del estudiante; las que reflejan sus habilidades y conocimientos previos, y las características sociodemográficas. Tras la estimación del modelo se observó que tanto el esfuerzo efectivo realizado durante la formación del estudiante, como las habilidades y conocimientos previos de este influyen de manera significativa en el rendimiento académico; por el contrario, las variables sociodemográficas analizadas en el estudio no parecen tener un efecto relevante en este. El resultado no significa que otro tipo de variables como los antecedentes familiares y sociales de los individuos y los recursos que proporciona la familia para facilitar el desempeño académico del estudiante pudieran ser relevantes, el caso es que en el estudio dichos autores no dispusieron de tal información.

En Argentina, Di Gresia et al. (2005) realizan un estudio cuyo objetivo es analizar la transición de los estudiantes universitarios entre su ingreso a la universidad y la salida por graduación o abandono. En dicho estudio se analizaron los determinantes del desempeño utilizando información de un censo realizado en 1994, que reúne información de los estudiantes de casi todas las universidades públicas. La medida de desempeño fue la cantidad de materias aprobadas por período académico. Como factores explicativos se emplearon cinco grupos de variables, así: la universidad y la carrera; las características del estudiante y su familia; el tipo de escuela secundaria a la que asistió; el proceso de formación y las horas de estudio. Entre los resultados están que las mujeres tienen un rendimiento superior a los varones; asimismo, ser argentino, no soltero, haber concurrido a una escuela de enseñanza media privada, haberse mudado para asistir a la universidad, tener padres con mayor instrucción, o destinar más horas a los estudios, son elementos también asociados a resultados superiores.

Los investigadores Birch y Miller (2006) utilizan el enfoque de cuartiles para analizar el desempeño de los estudiantes durante su primer año de universidad. Las variables explicativas incluidas son la nota obtenida en un examen final de la escuela secundaria, género, lugar de residencia, estatus socioeconómico y tipo de escuela secundaria a la que asistieron. En el estudio encontraron que el efecto de la mayoría de las variables independientes consideradas se transforma a lo largo de la distribución de rendimiento. Además, se evidencia que ser mujer es favorable especialmente para los estudiantes ubicados en la cola inferior de la distribución. En los niveles superiores el género pareció no ser importante.

Autores como Montero, Villalobos y Valverde (2007) desarrollaron el estudio multinivel, titulado *Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica*, en el cual utilizaron como variables independientes un conjunto de factores en las dimensiones institucionales, sociodemográficas, psicosociales y pedagógicas. Los resultados encontrados señalan que el mejor predictor ha sido la calificación obtenida en el proceso de admisión, la cual combina las notas obtenidas en su educación secundaria y una prueba de habilidades de razonamiento. También resultaron explicativas las siguientes variables: puntaje obtenido en la prueba de inteligencia emocional y modalidad de enseñanza. Por su parte, Moreno y Santín (2010) señalan que los factores del desempeño académico son diversos y que se centran, fundamentalmente, en la institución educativa y el estudiante; la primera, según el sector al que pertenece (oficial o privado) y su calidad. No obstante, los resultados encontrados en esta investigación ponen en evidencia que el desempeño en las competencias genéricas no depende únicamente de la institución y el estudiante, sino también de los aspectos sociodemográficos, socioeconómicos y académicos de estos.

En la Universidad de Salamanca, Tejedor y García-Valcárcel (2007) estudian las causas del bajo rendimiento del estudiante universitario como propuestas de mejora en el marco del Espacio Europeo de Educación Superior (EEES). Dichas causas las clasifican en tres categorías: institucionales, del docente y del estudiante. Los autores ofrecen pautas de actuación para mejorar el desempeño académico del estudiante, las cuales podrán ser tenidas en cuenta en la planificación y gestión de las instituciones en el marco del EEES.

La investigación *Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública*, realizada en la Universidad de Costa Rica por Garbanzo (2007), deja claro que estudiar

los factores asociados al rendimiento académico permite conocer y analizar la complejidad de este; además, que es posible que las relaciones que se presentan entre las variables, por más incidencia que parezcan tener, no permiten generalizar, dado que los resultados pueden variar según el conjunto de interacciones entre sí, por las poblaciones que se estudien y sus respectivos contextos. Refiere la autora, además, que el desempeño académico comparte su complejidad conceptual con la calidad de la educación, al intervenir distintos factores que según sus enfoques de estudio y contextos despliegan disímiles resultados.

En México, Izar, Ynzunza y López (2011), realizan la investigación *Factores que afectan el desempeño académico de los estudiantes de nivel superior en Rioverde, San Luis Potosí*, cuyo objetivo es estudiar algunos factores relacionados con el desempeño académico de los estudiantes de educación superior. Incluye la descripción del impacto de los cambios demográficos en la educación y los principales problemas que hay en torno a ella. En este estudio, las variables que resultaron significativas son: el género, la trayectoria académica y el interés de formarse profesionalmente. Uno de los resultados de interés es que las mujeres tienen mejor desempeño que los hombres, lo que da lugar a investigaciones para buscar las causas de este hallazgo. Se encontró además que no hay asociación entre desempeño académico y nivel socioeconómico, aspecto que corrobora el hecho que no por contar con mayores recursos económicos, se alcanza un mejor desempeño académico. Tampoco resultó significativa su relación con la escolaridad de sus padres.

En Colombia, Vélez y Roa (2005), en la investigación *Factores asociados al rendimiento académico en estudiantes de medicina de la Universidad del Rosario*, en Bogotá, describen que el hecho de no leer como pasatiempo, violencia intrafamiliar, consumo de drogas psicoactivas, provenir de un colegio mixto, padres sin estudios profesionales, no tener hábitos de estudio y utilizar más tiempo en actividades sociales, la dependencia económica, entre otras, ocasionan un regular desempeño académico. En los resultados se destaca que los factores sociodemográficos no tienen incidencia significativa sobre el desempeño académico.

En el Sistema de Universidades Estatales del Caribe Colombiano (SUE-Caribe), los autores Villalba y Salcedo (2008) realizaron la investigación *Rendimiento académico en el nivel de educación media, como factor asociado al rendimiento académico en la universidad*. Se buscó establecer la relación entre el desempeño académico en el nivel medio y el desempeño académico en la universidad. Se trata de un estudio descriptivo correlacional, en el cual se consideraron las siguientes variables: desempeño académico en educación media, indicadores de desempeño y el

desempeño en la universidad, asumido en este estudio como *trayectoria escolar*. Los resultados permitieron determinar que existe una relación estadísticamente positiva ($p=0,417$) entre el rendimiento académico en educación media y la trayectoria escolar en la universidad. Si bien no puede afirmarse que el rendimiento en el nivel precedente tiene un valor predictivo, a partir de los resultados obtenidos en este estudio puede establecerse que hay una tendencia a que los estudiantes mantengan o mejoren el rendimiento académico en la universidad con relación al obtenido en educación media.

Cabe destacar el trabajo adelantado por Chicaiza, Galvis y Ramírez (2010), titulado *Determinantes del rendimiento académico en Colombia: pruebas ICFES Saber 11º, 2009*, en el que se utilizan los resultados obtenidos por los estudiantes en las áreas de matemáticas y lenguaje como medio para identificar los determinantes del desempeño académico en Colombia, utilizando el modelo Logit Ordenado Generalizado. Los resultados obtenidos presentan la relevancia de la dimensión socioeconómica en el desempeño para ambas áreas; particularmente, las variables relacionadas con el nivel de ingreso, que presentan impacto positivo y significativo en el resultado de las pruebas. Los estudiantes con nivel socioeconómico alto pueden acceder a mejores condiciones de alimentación, transporte, infraestructura y TIC; además, tener computador en el hogar implica una mayor probabilidad de obtener buen desempeño académico. Por su parte, la variable sobre el nivel educativo de los padres tiene gran significancia, pues a medida que aumenta el nivel educativo de los padres se evidencia mayor desempeño académico en las áreas objeto de estudio. Este hecho refleja que padres con mayor nivel educativo pueden apoyar favorablemente el proceso de aprendizaje de sus hijos de manera directa.

Dicha investigación encontró que los hombres tienen mejor desempeño en matemáticas y las mujeres en lenguaje; sobre la variable edad, en la medida que esta aumenta existe mayor probabilidad de obtener menor desempeño. Igualmente, incide la jornada de estudio; los de jornada completa obtienen puntajes más altos que los estudiantes de otras jornadas. Conviene tener en cuenta estos resultados a la hora de formular políticas educativas.

La Secretaría de Educación Distrital de Bogotá adelantó el Estudio liderado por Salinas (2010), sobre los factores que influyen en el rendimiento escolar, cuyo objetivo es discutir sobre los factores asociados al desempeño académico exitoso, los cuales se han encontrado en diferentes estudios a nivel internacional, con el fin de buscar cuáles influyen en los resultados del examen de Estado para el ingreso a la educación superior. El estudio se realizó con estudiantes de grado once

de 55 colegios públicos y 71 colegios privados de Bogotá, donde se contemplan variables endógenas y exógenas. Dentro de las primeras está el género, la edad, la frecuencia de estudio y hábitos, y la trayectoria de la vida académica; en las segundas se relacionan la comunidad, es decir, el entorno inmediato, la familia, su composición, clima familiar, nivel de ingresos económicos, la ocupación y el nivel educativo de los padres, seguridad, infraestructura, recursos disponibles para el aprendizaje, entre otros. Entre las variables que sobresalen por su asociación al desempeño académico se mencionan el estrato social del estudiante, los ingresos económicos, el nivel educativo de los padres, las características profesionales de los docentes y su vinculación a la institución.

En el contexto colombiano es importante la investigación adelantada por Montes y Lerner (2012), titulada *Rendimiento académico de los estudiantes de pregrado de la Universidad EAFIT. Perspectiva cuantitativa*, la cual analizó los factores que influyen sobre el desempeño, en la perspectiva de disponer de una mirada cuantitativa de las variables observables y no observables, relacionadas con el proceso de aprendizaje y los logros académicos. El estudio agrupó las variables en las dimensiones académica, económica y familiar, personal e institucional, y se desarrolló en dos etapas: la primera es exploratoria, en la que se caracterizaron los 840 estudiantes de la cohorte 2009-1; la segunda es focalización, la cual se realizó con la información recolectada de 467 estudiantes de la misma cohorte.

El estudio señala que, en general, las condiciones económicas de los estudiantes son buenas: sus familias los sostienen totalmente o, en su lugar, lo hace una institución; tienen padres activos laboralmente y viven principalmente en zonas residenciales que pertenecen a los tres estratos más altos. La financiación de los estudios universitarios la realizan en mayor proporción los padres; no obstante, existe un porcentaje bajo de estudiantes becados. La familia, su composición y clima familiar también contribuyen a la explicación del desempeño académico. El clima familiar, caracterizado por la libertad de hablar francamente en la familia y ser solidarios, resultó determinante, así como el grado de escolaridad de los padres; los hijos de padres universitarios tienen una mayor participación en los intervalos correspondientes a mayores promedios.

Los resultados obtenidos por Montes y Lerner (2012) sugieren asistir académicamente a la población que ingresa a la universidad proveniente de colegios de niveles bajos según la clasificación del ICFES y con bajas calificaciones del bachillerato; recomiendan también promover la discusión académica, el pensamiento analítico y crítico más allá del aula; incentivar el aumento del nivel de concentración

de los estudiantes; motivar la organización de actividades de los estudiantes y distribución del tiempo; promover el trabajo grupal de los estudiantes; dar continuidad al programa de becas de la universidad; preservar el reconocimiento de calidad institucional y de programas académicos; mantener oportunidades en actividades extracurriculares.

Conceptualización sobre desempeño académico

Lo complejo del desempeño académico se inicia a partir de la misma conceptualización, pues recibe y se conoce a través de diferentes denominaciones, entre ellas la aptitud escolar, el rendimiento escolar y el desempeño académico; no obstante, las diferencias que se presentan entre los conceptos tienen explicación desde la semántica, puesto que en el acontecer de los procesos académicos y en las relaciones que comprenden los procesos de enseñanza y de aprendizaje, tales denominaciones se utilizan como sinónimos.

El desempeño académico, como un fenómeno complejo y multidimensional, se encuentra documentado teóricamente por diferentes autores; por ejemplo, Nieto (2008) señala que este tema amerita ser estudiado a través de diferentes tipos y modelos de investigación, como el exploratorio, descriptivo y explicativo, en los cuales es imprescindible relacionar diferentes dimensiones que agrupan muchas variables. Advierte el autor que a pesar del gran cúmulo de información existente sobre el tema, esta no es concluyente. Dadas sus implicaciones en el proyecto educativo de los estudiantes, en su formación profesional y posterior desempeño en el campo laboral, el desempeño académico es un tema determinante, complejo y multifactorial en el campo de la educación superior; igualmente, lo es en la calidad de los servicios educativos que ofrecen las IES. Se entiende el desempeño académico como la relación que existe entre el proceso de aprendizaje y los resultados que se obtienen en valores predeterminados. Montes y Lerner (2012) consideran que el desempeño académico no se explica únicamente por este tipo de resultados, es decir, por las calificaciones o puntajes que obtienen los estudiantes, suponen la existencia de otros factores que aportan a su comprensión y los agrupan en las dimensiones: académica, económica, familiar, personal e institucional.

El desempeño académico visto como efectividad en la formación del estudiante es definido por Reyes (2003) como el grado de logro de los objetivos establecidos en los programas de formación profesional, siendo este un indicador de tipo multidimensional al involucrar variables de orden cognitivo, emocional,

fisiológico, cultural, social, demográfico, económico, entre otras. Por su parte, Tonconi (2010) señala que el desempeño académico concebido como resultado no siempre da cuenta de las competencias logradas por los estudiantes en el proceso de formación profesional; el esfuerzo del estudiante y la calidad del proceso de su formación no son directamente proporcionales con el resultado obtenido; por lo tanto, se hace necesario, concebir un concepto de desempeño académico que corresponda e incluya el proceso del estudiante y sus condiciones socioeconómicas.

El concepto de desempeño académico que enmarca la investigación *Descubrimiento de patrones de desempeño académico de los estudiantes de carreras profesionales en las competencias genéricas de las Pruebas Saber Pro 2011-2*, que se reporta en este libro, establece la relación entre el proceso de aprendizaje y el logro de competencias genéricas, el cual comprende factores extrínsecos e intrínsecos al estudiante, y el producto derivado de él, expresado tanto en valores predeterminados por el contexto sociocultural como en las decisiones y acciones del estudiante en relación con las competencias genéricas logradas en el proceso.

Factores asociados al desempeño académico

Uno de los temas de mayor relevancia en los procesos de enseñanza y de aprendizaje lo constituye el desempeño académico del estudiante; cuando se trata de evaluarlo y de buscar cómo mejorarlo, se precisa analizar los factores que se asocian.

Para el análisis de los factores que pueden influir en el desempeño académico, Benítez, Giménez y Osicka (2000) consideran que se deben tener en cuenta los socioeconómicos, la complejidad de los programas de estudio, las modalidades de formación, las metodologías de enseñanza y los conceptos previos que tienen los estudiantes. Sobre este tema, también Mizala y Romaguera (2002) analizan los factores que inciden en el rendimiento escolar, identificando factores asociados a los escolares, a sus características familiares y a la institución educativa. Estos autores clasifican las variables en sociales, socioeconómicas e institucionales.

Respecto a lo anterior, Castaño, Gallón, Gómez y Vasquéz (2004) encuentran cuatro conjuntos de factores como determinantes del desempeño académico. Primero, las características sociodemográficas del estudiante; segundo, el factor académico, asociado a variables que contemplan la educación y la orientación del estudiante y su desempeño en la institución; tercero, el socioeconómico, que considera aspectos económicos laborales del estudiante y su familia, así como la dependencia económica e indicadores económicos; el cuarto factor es

el institucional, que contempla las variables asociadas con la adaptación del estudiante al ambiente universitario. Del mismo modo, Porto, Di Gresia y López (2004) toman como variables de su estudio las características del estudiante y de su familia, y encuentran de esta manera que existen factores explicativos del desempeño académico, así: el género (las mujeres obtienen un mejor desempeño); la edad (los jóvenes presentan el mejor desempeño académico); la educación de los padres (cuanto más educado el padre, mejor el rendimiento).

Con base en la conceptualización de desempeño académico, los antecedentes teóricos sobre los factores que intervienen en él, los cuales se asocian y colindan unos con otros, y de acuerdo con las variables encontradas en las bases de datos de las pruebas Saber Pro 2011-2 del ICFCES, para el estudio que se reporta en este libro dichos factores se agrupan en cuatro dimensiones, a saber: sociodemográfica, económica, académica e institucional. Se considera que una dimensión es una tipología que admite límites flexibles entre las agrupaciones de variables, teniendo como referencia los objetivos de la investigación. Vale advertir que para identificar tendencias y determinar patrones de desempeño académico se pueden asociar entre sí las cuatro dimensiones referidas y las variables agrupadas en cada una de ellas.

Dimensiones asociadas al desempeño académico

Las dimensiones para el estudio del desempeño académico corresponden a la magnitud que, junto o en asocio con otras, sirven para definir el fenómeno. Como se dijo en el párrafo anterior, para el desarrollo de la presente investigación se conceptualizaron y utilizaron las dimensiones sociodemográfica, económica, académica e institucional, que se explican a continuación.

- **Dimensión sociodemográfica.** Corresponde a las variables que forman parte de los estudios de la población en relación con su estructura, evolución y características generales, según estado, distribución y variaciones en un momento determinado, relacionando las características del individuo y su grupo familiar. Implica una serie de interacciones que reflejan comportamientos directamente relacionados con su desempeño académico. Se puede fundamentar en Hidalgo de Paz, Gil y Rodríguez (2001) y Salanova, Martínez, Bresó, Llorens y Grau (2005), quienes consideran que el género, el estado civil, la edad, la ocupación y la tipología familiar, el nivel educativo de los padres y las interrelaciones con otras dimensiones y variables se asocian al desempeño académico.

Teniendo en cuenta las variables de las bases de datos del ICFES de las pruebas Saber Pro 2011-2, se consideran dos grupos: primero, el que concentra variables del individuo, tales como género, edad, estado civil, fecha de nacimiento, ocupación y discapacidad; y el segundo, aquel que reúne variables del grupo familiar, como características de la residencia, personas que conforman la familia, ocupación de los padres y nivel educativo de los padres.

- **Dimensión económica.** Corresponde a las condiciones que tiene el estudiante para satisfacer sus necesidades mientras cursa la carrera profesional. Se sustenta esta dimensión en estudios realizados por Mizala y Romaguera (2002), Valdivieso et al. (2004), Castaño et al. (2004), Porto et al. (2004), Tonconi (2010) y Garbanzo (2014), quienes señalan que las condiciones económicas o comodidades materiales que posee el estudiante o su familia inciden significativamente en su desempeño académico. No obstante, puntualizan Espindola y León (2002) que la condición económica como tal no es un elemento de riesgo, sino más bien un resultado derivado de la presencia de otras variables asociadas que explican el desempeño académico.

En esta investigación se clasificó el desempeño académico en dos grupos: el primero, tiene que ver con su estrato socioeconómico, en el cual se consideran variables que se encuentran en las bases de datos del ICFES, como ingresos del estudiante o de su familia, nivel Sisben, condiciones de la vivienda, medios de transporte, uso de electrodomésticos, manejo de equipos electrónicos y medios de comunicación; el segundo considera el valor de la matrícula y el origen de este (estudiante, familia, crédito, beca, entre otros).

- **Dimensión académica.** Hace referencia al proceso de desarrollo en la formación académica del estudiante, es decir, al qué y al cómo en su formación profesional. Montes y Lerner (2012) señalan que es recurrente que los investigadores consideren el desempeño académico previo al ingreso a la universidad, el cual se considera predictor de buen desempeño académico en el proceso de formación profesional. Esta dimensión se fundamenta en investigaciones de Barraza (2007) y Contreras, Caballero, Palacios y Pérez (2008), donde se identifican el campo y área del conocimiento, los semestres y créditos académicos cursados y los saberes previos del bachillerato, como aspectos que se deben tener en cuenta para el análisis del desempeño académico en la formación profesional. Con base en estas consideraciones y las variables de las bases de datos del ICFES de las pruebas Saber Pro 2011-2, en esta investigación se pueden definir dos grupos, uno referido al bachillerato, en el que se incluyen variables como tipo

de bachillerato y otro referido a la formación profesional, que agrupa variables relacionadas con el área de conocimiento, el grupo de referencia, el programa y su metodología, el semestre y el número de créditos cursados.

• **Dimensión institucional.** Las IES deben cumplir con la responsabilidad y compromisos sociales misionales; entre otros, formar profesionales de calidad que requiere la sociedad. Son las encargadas del proceso de aprendizaje y del desarrollo de las competencias necesarias que necesita un profesional para su ingreso al mundo del trabajo y que le permitan hacerse responsable de su vida y desempeñarse con solvencia y autonomía como persona, ciudadano y profesional, responsable de su proyecto de vida y el desarrollo de su entorno. Esta dimensión corresponde a las características estructurales y funcionales de la institución. Se fundamenta en estudios realizados por Galand, Frena y Bourgeois (2004), Garbanzo (2007) y Artunduaga (2008), quienes consideran que el funcionamiento de una institución educativa, en su complejidad, está asociado al desempeño académico de los estudiantes; destacan además su proyecto educativo, su infraestructura física, tecnológica y las características particulares de su función administrativa, al igual que su reconocimiento en el sector académico. Para el presente estudio, la dimensión institucional agrupa variables relacionadas con el tipo de institución (pública, privada, régimen especial), el carácter académico (institución universitario o universidad) y la zona a la que pertenece (municipio, departamento, zona); variables que resultaron de la depuración de las bases de datos suministrada por el ICFES de las pruebas Saber Pro 2011-2.

Contextualización de las pruebas Saber Pro

La educación es un derecho y un servicio público que tiene una función social, a través de la cual se busca el acceso al conocimiento, a la ciencia, a la técnica y a los demás bienes y valores de la cultura (Constitución Política de Colombia, 1991, artículo 67). De modo que corresponde al Estado regular y ejercer la inspección y vigilancia de la educación, con el fin de velar por su calidad, por el cumplimiento de sus objetivos, con el propósito de mejorar la formación moral, intelectual y física de los colombianos, en la perspectiva de garantizar el adecuado cubrimiento del servicio educativo y asegurar las condiciones necesarias para su acceso y permanencia en el sistema educativo.

En tal sentido, la Ley 30 de 1992 establece que es obligación de las IES prestar un servicio con calidad, lo cual implica estar atentas a los resultados académicos, a los medios y procesos empleados, a la infraestructura institucional, a las dimensiones cualitativas y cuantitativas de estos y a las condiciones con las que se desarrolla cada institución.

En la primera década del siglo XXI, las nuevas políticas del Estado en materia de logro y evaluación de competencias y desempeño académico de los estudiantes de programas de pregrado a nivel nacional han provocado el surgimiento de una serie de cuestionamientos acerca del alcance de comprensión real y aprehensión pertinente de los aprendizajes, específicamente en las competencias genéricas, lo que ha causado la inquietud por indagar sobre los factores implicados en los resultados del desempeño académico en este campo, con el fin de contribuir a la formación integral de profesionales idóneos, competentes y pertinentes con las demandas de la sociedad.

En este contexto, uno de los grandes retos del ICFES es realizar las mediciones a través de las cuales se pueda mostrar el estado de la educación y su evolución, y realizar las investigaciones sobre su calidad; esto se convierte en una de las principales funciones de las pruebas, en la vía del mejoramiento de la política pública de la educación superior, para que la prestación del servicio educativo por parte de las IES sea de calidad. Así pues, las pruebas Saber Pro se constituyen en un instrumento estandarizado para la evaluación externa de la calidad de la educación superior. Según Zapata (2011), estas pruebas forman parte, con otros procesos y acciones, de un conjunto de instrumentos que el Estado dispone para evaluar la calidad del servicio público educativo y ejercer su inspección y vigilancia.

Reglamentación de las pruebas Saber Pro

El MEN de Colombia expide el Decreto 1781 del 2003, en el cual se reglamentan los Exámenes de Estado de Calidad de la Educación Superior (ECAES). Dicho decreto estableció la obligatoriedad de la presentación de los ECAES por parte de los estudiantes de programas de pregrado. Además, los definió como pruebas académicas de carácter oficial, cuyos propósitos fundamentales son comprobar el grado de desarrollo de las competencias de los estudiantes que cursan el último año de los programas académicos de pregrado que ofrecen las IES y servir de fuente de información para la construcción de indicadores de evaluación del servicio público educativo, que fomenten la cualificación de los procesos institucionales,

la formulación de políticas y faciliten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo.

A través de la Ley 1324 de julio del 2009, el Estado fija parámetros y criterios para organizar el sistema de evaluación de resultados de la calidad de la educación; por ejemplo, en el artículo 7°, se establece que los ECAES se constituyen en un requisito de grado en la educación superior, por lo cual es responsabilidad de las IES reportar la totalidad de los estudiantes que hayan aprobado por lo menos el 75% de los créditos académicos, o que, según el cronograma, tengan previsto graduarse el año siguiente a la aplicación de una determinada prueba. Además, estas pruebas se constituyen en un instrumento estandarizado para la evaluación externa de la calidad de la educación superior en Colombia, y forman parte, junto con otros procesos y acciones, de un conjunto de instrumentos que el Gobierno Nacional dispone para inspección y vigilancia.

A través de la Resolución 782 del 2010, expedida por el ICFES, se adopta el nombre de pruebas Saber Pro para el examen de evaluación de la educación superior, a partir del segundo semestre del 2010. En el primer semestre del 2011 se aplicaron 31 exámenes de competencias específicas, se aumentaron las competencias genéricas y se establecieron las pruebas de competencias comunes por áreas de formación: ciencias de la educación, ingenierías y ciencias básicas. En el segundo semestre del 2011 se crearon los 30 grupos de referencias que agrupan los programas académicos afines con el propósito de generar resultados comparables y diferenciar los niveles universitario, técnico, tecnológico y de ciclo complementario de escuelas normales de educación superior.

La implementación de las pruebas Saber Pro se debe principalmente a la necesidad de contar con una visión independiente y objetiva sobre la calidad de la educación superior, con miras a formular e implementar políticas públicas pertinentes a las necesidades de una formación integral del individuo y las demandas de la sociedad globalizada, en la búsqueda de la calidad del humano que contribuya al desarrollo de las regiones y el país.

Evolución de las pruebas Saber Pro

Los ECAES, ahora denominados pruebas Saber Pro, adquieren transformaciones significativas desde su creación en el 2002. Lo más significativo en estos cambios es que se pasó de medir conocimientos específicos a medir competencias, tanto genéricas como específicas.

Una síntesis de los cambios de las pruebas Saber Pro, desde el 2009 hasta el primer semestre del 2012, se presenta en la tabla 6.

Tabla 6
Cambios en las pruebas Saber Pro en el período 2009-2012

Años	Novedades	Pruebas de competencias genéricas
2009	<ul style="list-style-type: none"> - La Ley 1324 establece la obligatoriedad de la prueba Saber Pro para obtener un título de pregrado y define que su estructura deberá mantenerse por períodos no menores a 12 años. - El Decreto 3963 determina la gradualidad en la adopción de la estructura de las pruebas de competencias genéricas y específicas. 	<ul style="list-style-type: none"> - Todos los programas académicos presentaron dos pruebas de competencias genéricas sobre inglés y comprensión lectora. - Los programas académicos que no aplicaban ECAES sobre competencias específicas, presentaron el examen de Competencias Genéricas sobre Pensamiento Crítico, Solución de Problemas, Entendimiento Interpersonal y Escritura.
2010- 1		<ul style="list-style-type: none"> - Todos los programas presentaron dos pruebas de competencias genéricas: inglés y comprensión lectora. - Los programas académicos que no aplicaban ECAES sobre competencias específicas, presentaron el examen de competencias genéricas: pensamiento crítico, solución de problemas, entendimiento interpersonal y escritura.
2010- 2	<ul style="list-style-type: none"> - Con la Resolución 782 del 2010 se cambia el nombre de Pruebas ECAES por Saber Pro. - Los estudiantes de administración y medicina se evalúan, en la jornada de la mañana, en competencias genéricas y en la de la tarde, en competencias específicas. 	<ul style="list-style-type: none"> - Todos los programas presentaron dos pruebas de competencias genéricas: inglés y comprensión lectora. - Los programas de administración, medicina y los programas que no aplican Saber Pro de competencias específicas, presentaron examen de competencias genéricas: pensamiento crítico, solución de problemas, entendimiento interpersonal y escritura.
2011- 1	<ul style="list-style-type: none"> - Aumentan los programas que presentan pruebas de competencias genéricas. - Aparecen las pruebas de competencias comunes. 	<ul style="list-style-type: none"> - Todos los programas presentaron dos pruebas de competencias genéricas: inglés y comprensión lectora. - Programas de administración, medicina, ingenierías, ciencias de la educación, ciencias naturales y exactas, técnicos y tecnólogos del área de ingeniería, más los programas que no tenían Saber Pro de competencias específicas, presentaron examen de competencias genéricas: pensamiento crítico, solución de problemas, entendimiento interpersonal y escritura.

(continúa)

(viene)

Años	Novedades	Pruebas de competencias genéricas
2011- 2	<ul style="list-style-type: none"> - Desaparecen las pruebas específicas. - Todos los programas deben presentar pruebas de competencias genéricas. - Se crean grupos de referencias de programas. - MEN y comunidad académica define pruebas comunes para grupos de referencia. 	Todos los programas presentarán una prueba de competencias genéricas en razonamiento cuantitativo, lectura crítica, inglés y escritura.
2012- 1	Se puede seleccionar la combinación de módulos para conformar el examen.	La prueba de Competencias Genéricas se complementa con un nuevo módulo: competencias ciudadanas.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Grupos de referencia

Es una agrupación de programas con características similares para los cuales se presenta una oferta de combinaciones de módulos de competencias específicas. En efecto, para las pruebas Saber Pro 2011-2 se diseñaron cuatro módulos para la evaluación de competencias genéricas, a saber: lectura crítica, comunicación escrita, razonamiento cuantitativo e inglés, pruebas que se aplicaron a los 15 grupos de referencia en los que se agrupan los programas profesionales que se ofrecen en Colombia. La tabla 7 muestra la organización de los grupos de referencia.

Tabla 7

Listado de grupos de referencia de los programas de educación superior en Colombia

Nombre del grupo de referencia	Programas
Bellas artes y diseño	Diseño Industrial, Diseño Gráfico, Música, Artes Plásticas, Diseño de Modas, Bellas Artes y afines.
Ciencias naturales y exactas	Biología, Matemáticas, Química, Física, Estadística, Geología, Microbiología, Ecología y afines.
Ciencias sociales	Psicología, Trabajo Social, Sociología, Antropología, Ciencia Política, Relaciones Internacionales, Geografía y afines.

(continúa)

(viene)

Nombre del grupo de referencia	Programas
Humanidades	Filosofía, Historia, Teología, Literatura, Filosofía y Letras, Traducción Simultánea, Español y Filología Clásica, Lenguas Modernas y afines.
Derecho	Derecho, Criminalística y afines.
Comunicación, Periodismo y Publicidad	Comunicación Social, Mercadeo y Publicidad, Comunicación Social Publicidad, Ciencias de la Información y la Documentación y afines.
Ciencias militares y navales	Ciencias Militares, Ciencias Navales para Oficiales de Infantería de Marina, Ciencias Navales para Oficiales Navales y afines.
Ciencias agropecuarias	Zootecnia, Medicina Veterinaria y Zootecnia, Administración de Empresas Agropecuarias, Medicina Veterinaria, Agronomía y afines.
Ciencias económicas y administrativas	Administración de Empresas, Contaduría Pública, Economía, Comercio Internacional, Negocios Internacionales, Mercadeo, Administración de Negocios Internacionales, Administración de Hotelería y Turismo y afines.
Educación	Licenciatura en Educación Preescolar, Licenciatura en Educación Básica con énfasis en Ciencias Naturales y Educación Ambiental, Licenciatura en Pedagogía Infantil, Licenciatura en Ciencias Sociales, Licenciatura en Educación Básica con énfasis en Ciencias Sociales, Licenciatura en Matemáticas y Física, Licenciatura en Español y Literatura, y afines.
Arquitectura y urbanismo	Arquitectura y afines.
Ingeniería	Ingeniería de Sistemas, Ingeniería Industrial, Ingeniería Electrónica, Ingeniería Civil, Ingeniería Mecánica, Ingeniería Ambiental, Ingeniería Eléctrica, Ingeniería de Alimentos, Ingeniería Agroindustrial, Ingeniería Agronómica, Ingeniería Química y afines.
Salud	Enfermería, Odontología, Fisioterapia, Fonoaudiología, Instrumentación Quirúrgica, Terapia Ocupacional, Nutrición y Dietética, Salud Ocupacional, Bacteriología, Optometría, Química Farmacéutica y afines.
Medicina	Medicina.
Recreación y deportes	Tecnología Deportiva, Cultura Física, Deporte y Recreación, Recreación, Ciencias del Deporte y afines.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Referencias

- Achaerandio, L. (2010). *Competencias fundamentales para la vida*. Guatemala: Universidad Rafael Landívar.
- Aguila, V. (2005). El concepto: calidad en la educación universitaria, clave para el logro de la competitividad institucional. *Revista Iberoamericana de Educación*, 35. Recuperado de <http://www.rieoei.org/deloslectores/880Aguila.PDF>
- Álvarez, J. (2008). Evaluar el aprendizaje en una enseñanza centrada en las competencias. En J. Gimeno (Comp.), *Educación por competencias. ¿Qué hay de nuevo?* (pp. 206-234). Madrid: Morata.
- Artunduaga, M. (2008). Variables que influyen en el rendimiento académico en la Universidad. Recuperado de <http://es.slideshare.net/1234509876/variables-del-rendimiento-academico-universidad>
- Barraza M. (2007). El Estrés de Examen. *Revista Electrónica de Psicología Científica*, 9(32). Recuperado de <http://www.psicologiacientifica.com/bv/psicologia-306-1-el-estres-de-examen.html>
- Beltrán, M. (2011). *Propuesta de lineamientos para la formación por competencias en educación superior*. Bogotá: Editorial Norma.
- Beneitone, P., Esquetine, C., González, J., Marty, M., Siufi, G. y Wagenaar, R. (eds.) (2007). *Reflexiones y perspectivas de la educación superior en América. Informe final Proyecto Tuning - América Latina 2004-2007*. Bilbao: Universidad de Deusto y Universidad de Groningen.
- Benítez, M., Giménez, M. y Osicka, R. (2000). Las asignaturas pendientes y el rendimiento académico: ¿existe alguna relación? Disponible en: <http://www1.unne.edu.ar/cyt/humanidades/h-009.pdf>
- Birch, E. y Miller, P. (2006). Student Outcomes at University in Australia: a Quantile Regression Approach. *Australian Economic Papers*, 45(1), 1-17.
- Bricall, J. (2000). *Informe Universidad 2 mil: Comisión de Rectores de las Universidades Españolas*. Madrid: CRUE.
- Cano, M. (2008). La evaluación por competencias en la educación superior. *Revista del currículum y formación del profesorado*, 12(3), 1-16.
- Castaño, E., Gallón, S., Gómez, K. y Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de economía*, 60(60), 39-65.
- Chaparro, F. (2008). *Evolución de la acreditación de alta calidad en Colombia 1998-2008*. Bogotá: Consejo Nacional de Acreditación [CNA].
- Chicaiza, S., Galvis, D. y Ramírez, A. (2010). Determinantes del rendimiento académico en Colombia: pruebas ICFES Saber 11°, 2009-1. *Revista Universidad EAFIT*, 46(160), 48-72.

Recuperado de <http://publicaciones.eafit.edu.co/index.php/revista-universidad-eafit/article/view/754>

- Climént, J. (2014). Origen, desarrollo y declive de las competencias individuales en tiempos de incertidumbre. *Educación y Educadores*, 17(1), 149-168.
- Coleman, J. (1966). *Equality of educational opportunity*. Washington: US Government Printing Office.
- Colombia. Sistema Nacional de Acreditación SNA. (2015). *Lineamientos para la acreditación institucional*. Consejo Nacional de Acreditación –CNA. Disponible en: http://www.cna.gov.co/1741/articles-186359_Lin_Ins_2014.pdf
- Contreras, K., Caballero, C., Palacios, J. y Pérez A. (2008). Factores asociados al fracaso académico en estudiantes universitarios de Barranquilla. *Psicología desde el Caribe*, 22. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-417X2008000200008
- Constitución Política de Colombia de 1991. Reforma de 1997. Recuperado de http://www.unesco.org/culture/natlaws/media/pdf/colombia/colombia_constitucion_politica_1991_spa_orof.pdf
- Delors, J. (1999). *La educación encierra un tesoro: informe a la Unesco de la Comisión Internacional sobre la Educación para el siglo XXI*. París: Santillana, Unesco.
- Di Gresia, L., Fazio, A., Porto, L., Ripani, W. y Sosa, E. (2005). Rendimiento y productividad de los estudiantes. El caso de las universidades públicas argentinas. En A. Porto (ed.), *Economía de la Educación Universitaria: Argentina-Brasil-Perú* (pp. 8-27). La Plata: Editorial de la Universidad Nacional de La Plata.
- Espíndola, E. y León, A. (2002). La deserción escolar en América Latina: un tema prioritario para la agenda regional. *Revista Iberoamericana de Educación*, 30. Recuperado de <http://www.rieoei.org/rie30a02.htm>
- Febles, J. y González, A. (2002). Aplicación de la minería de datos en la bioinformática. *Revista Cubana de los Profesionales de la Información y la Comunicación en Salud Acimed*, 10(2). Recuperado de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352002000200003#cargo
- Fernández, N. (2006). La evaluación y la acreditación de la calidad: situación, tendencias y perspectivas. En IESALC. *Informe sobre la educación superior en América Latina y el Caribe 2000–2005: la metamorfosis de la educación superior* (pp. 33-42). Venezuela: Metrópolis C.A.
- Galand, B., Frenay, M. y Bourgeois, E. (2004). Facteurs de réussite en première de candidature. *Journée d'étude Chaire Unesco de Pédagogie Universitaire*.

- Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde calidad de la educación superior pública. *Revista Educación*, 31(1), 43-63.
- Garbanzo, G. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare*, 18(1), 119-154.
- Gómez, M., Rodríguez, G. e Ibarra, M. (2013). Compes: auto informe sobre las competencias básicas relacionadas con la evaluación de estudiantes universitarios. *Estudios sobre Educación*, 24, 197-224. doi: 10.15581/004.24.197-224
- González, V. y González, R. (2008). Competencias genéricas y formación Profesional: un análisis desde la docencia Universitaria. *Revista Iberoamericana de Educación*, 47. Recuperado de <http://www.rieoei.org/rie47a09.htm>
- González, J. y Wagenaar, R. (eds.) (2003). Structures in Europe. Informe final Proyecto Piloto-Fase 1. Bilbao: Universidad de Deusto, Universidad de Groningen y Comisión.
- Hernández, I. y Hernández, I. (2013). Responsabilidad de la universidad en la trilogía: universidad-empresa-Estado. En I. Hernández-Arteaga y L. S. Pemberthy-Gallo (comps.), *Universidad-Empresa-Estado: hacia la cultura de la investigación y la innovación. Segunda Rueda de Negocios de Innovación en Cauca y Nariño* (pp. 23-43). Bogotá: Editorial Universidad Cooperativa de Colombia, Red URIEL.
- Hernández, I., Alvarado, J. C. y Luna, J. A. (2015). Responsabilidad social en la relación universidad-empresa-Estado. *Educación y Educadores*, 18(1), 95-110. doi: 10.5294/edu.2015.18.1.6
- Hidalgo de Paz, A., Gil, M. y Rodríguez, E. (2001). Capacidad de atención y resultados docentes de estudiantes de primer año de medicina. *Revista Cubana Educación Médica Superior*, 15(3), 273-278.
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2011a). Examen saber pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior. (Internet). Disponible en: <http://acofartes.org.co/docsweb/documento/ICFES%202011,%20M%C3%93DULOS%20COMPETENCIAS%20GEN%C3%89RICAS%20Y%20ESPEC%C3%8DFICAS.pdf>
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2012a). *Examen Saber Pro, junio de 2012-I. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. ICFES. Recuperado de www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias

- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2012b). Saber Pro: Principales resultados en Competencias Genéricas. Santa Marta: ICFES. Disponible en: www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2013). Estructura del Saber Pro. Disponible en: http://ead.uis.edu.co/empresarial/images/stories/doc/SABER_PRO_2013_ESTRUCTURA.pdf
- Hernández Arteaga, I. y Luna, S. M. (2014). Creation and Innovation: A Strategy for Regional Competitiveness. En I. Hernández Arteaga y L. S. Pemberthy-Gallo (Comps.), *University-Company-State: Towards a Culture of Research and Innovation. Second Conference of Business Innovation in Cauca and Nariño* (pp. 192-204). Bogotá: Editorial Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600261>
- Izar, J., Ynzunza, C. y López, H. (2011). Factores que afectan el desempeño académico de los estudiantes de nivel superior en Rioverde, San Luis Potosí, México. *CPU-e Revista de Investigación Educativa*, 12. Recuperado de <https://www.uv.mx/cpue/num12/opinion/completos/izar-desempeno%20academico.html>
- Jano, D. y Ortiz, S. (2005). Determinación de los factores que afectan al rendimiento académico en la educación superior. *XII Jornadas de la Asociación de Economía de la Educación*. Oviedo, Universidad de Oviedo.
- Lorente, R. (2012). *La formación profesional según el enfoque de las competencias. La influencia del discurso europeo en España*. Barcelona: Octaedro.
- Luna, J. y Hernández, I. (2010). TIC: una tendencia que transforma el quehacer de la Universidad. *Revista Memorias*, 8(14), 166-184.
- Martínez, A., Cegarra, J. y Rubio, J. (2012). Aprendizaje basado en competencias: una propuesta para la autoevaluación del docente. *Profesorado. Revista del currículum y formación del Profesorado*, 16(2), 373-386.
- Mizala, A. y Romaguera, T. (2002). *Factores que inciden en el rendimiento escolar en Bolivia*. Santiago de Chile: Universidad de Chile.
- Montero, E., Villalobos, J. y Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico y a la repetición estudiantil en la Universidad de Costa Rica: un estudio multinivel. *Revista Relieve*, 13(2), 215-234.
- Montes, I. y Lerner, J. (2012). *Rendimiento académico de los estudiantes de pregrado de la Universidad EAFIT. Perspectiva cuantitativa*. Medellín: Universidad EAFIT.
- Mora, J. (2004). La necesidad del cambio educativo para la sociedad del conocimiento, en calidad y acreditación universitaria. *Revista de Educación Iberoamericana*, núm. 35. Disponible en: <http://www.rieoei.org/rie35a01.htm>

- Moreno, J. y Santín, D. (2010). Los determinantes de la eficiencia educativa en la Unión Europea. Instituto de Estudios Fiscales. *Revista de Economía Pública*, 193, 131-156.
- Navarro, R. (2003). El rendimiento académico: concepto, investigación y desarrollo. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2), 1-15. Recuperado de <http://www.ice.deusto.es/rinace/reice/vol1n2/Edel.pdf>
- Nieto, S. (2008). Hacia una teoría sobre el rendimiento académico en enseñanza empírica. *Teoría de la Educación*, 20, 249-274.
- Peña, L. (2008). La competencia oral y escrita en la educación superior. Bogotá: Ministerio de Educación Nacional. Recuperado de http://mineducacion.gov.co/1621/articulos-189357_archivo_pdf_comunicacion.pdf
- Pérez, A. (2013). Reválidas, evaluación de competencias y calidad de los aprendizajes. *Revista Currículum*, 26, 11-25.
- Pinilla, A. (2011). Modelos pedagógicos y formación de profesionales en el área de la salud. *Acta Médica Colombiana*, 36(4), 204-218.
- Porto A., Di Gresia L. y López, M. (2004). *Mecanismos de admisión a la Universidad y rendimiento de los estudiantes*. Recuperado de <http://www.aep.org.ar/espa/anales/resumen04/04/Porto-DiGresia-Armengol.pdf>
- Restrepo, B. (2008). *Política pública sobre calidad de la educación superior, y retos de la educación superior hoy*. Medellín: Asiesda.
- Reyes, Y. (2003). Relación entre el rendimiento académico, la ansiedad ante los exámenes, los rasgos de personalidad, el autoconcepto y la asertividad. Ministerio de Educación del Perú. Universidad Nacional Mayor de San Marcos. Disponible en: http://ciberdocencia.gob.pe/?id=493&a=articulo_completo
- Rodríguez, W. (2010). El concepto de calidad educativa: una mirada crítica desde el enfoque históricocultural. *Revista Electrónica Actualidades Investigativas en Educación*, 10(1), 1-28.
- Ruiz, G. (2009). El enfoque de la formación profesional en torno a la generación de competencia: ¿ejercicio impostergable o “lo que sucedió a un rey con los burladores que hicieron el paño”. *Estudios pedagógicos*, 34(1), 287-299.
- Salanova S., Martínez, I., Bresó, E., Llorens, S. y Grau, R. (2005). Bienestar psicológico en estudiantes universitarios. Facilitadores y obstaculizadores del desempeño académico. *Rev. Anales de Psicología*, 21(1), 170-180.
- Salinas, J. (2006). Flexibilidad en el currículo de la educación superior en el ámbito de las competencias. *II Encuentro Académico, Comisión de Currículo de la Comisión Nacional de Rectores (CONARE)*. Costa Rica.
- Salinas, J. (2008). *Modelos didácticos en los campus virtuales universitarios: patrones metodológicos generados por los profesores en procesos de enseñanza aprendizaje en entornos virtuales*. Informe

final del proyecto EA2007-0121. Recuperado de <http://tecnologiaedu.us.es/tecnoedu/images/stories/EA2007-0121-memoria.pdf>

- Salinas, F. (2010). *Estudio sobre los factores que influyen en el rendimiento escolar*. Bogotá: Secretaría de Educación Distrital. Recuperado de: <http://evaluacion.educacionbogota.edu.co/files/Factores%20que%20influyen%20en%20el%20rendimiento%20escolar.pdf>
- Schunk, D. (2008). *Learning theories: an educational perspective*. New York: Prentice Hall.
- Tejedor, F. y García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario. Propuestas de mejora en el marco del EEEE. *Revista de Educación*, 342, 443-473.
- Tobón, S., Pimienta, J. y García, J. (2010). *Secuencias didácticas: aprendizaje y evaluación de competencias*. México: Pearson.
- Tobón, S. Sánchez, A. Carretero, M. García, J. (2006). *Competencias calidad y educación superior*. Bogotá: Magisterio.
- Tonconi, J. (2010). Factores que influyen en el rendimiento académico y la deserción de los estudiantes de la Facultad de Ingeniería Económica de la UNA-Puno, Periodo 2009. *Cuadernos de Educación y Desarrollo*, 2(1). Recuperado de <http://www.eumed.net/rev/ced/11/jtq.pdf>
- Tünnermann, C. (2008). Introducción. En *La educación superior en América Latina y el Caribe: Diez años después de la Conferencia Mundial de 1998*. Bogotá: Pontificia Universidad Javeriana. Unesco e IESALC.
- Umaña, A. (2008). Reflexiones sobre el diseño curricular por competencias en la Universidad Estatal a Distancia de Costa Rica. *Revista Cognición*, 13. Edición especial. II congreso Cread Andes y II Encuentro Virtual Educa UTP, Loja.
- Unesco (1998). *Conferencia mundial sobre la educación superior. La educación superior en el siglo XXI. Visión y acción*. París: Unesco. Recuperado de <http://unesdoc.unesco.org/images/0011/001163/116345s.pdf>
- Valdivieso, M., Monar, K. y Granda, M. (2004). Análisis de los determinantes del rendimiento de los estudiantes de ESPOL –2002. *Revista Tecnológica*, 17(1), 213-218.
- Vélez, A. y Roa, C. (2005). Factores asociados al rendimiento académico en estudiantes de medicina. *Educación Médica*, 8(2), 74-82.
- Villalba, A. y Salcedo, M. (2008). *El rendimiento académico en el nivel de educación media, como factor asociado al rendimiento académico en la universidad*. (Tesis de maestría en educación). Sistema de Universidades Estatales del Caribe – Colombia.
- Zapata, L. (2011). *Factores Académicos asociados al bajo rendimiento en Inglés en la Pruebas ECAES presentada por los estudiantes de la Facultad de Educación en el año 2009*. Medellín: Fundación Universitaria Luis Amigó.

2 El proceso de descubrimiento de conocimiento en bases de datos

The Process of Knowledge Discovery on Databases

Resumen

En este capítulo se describen las etapas del proceso KDD y se hace énfasis en la etapa de minería de datos y en las técnicas más comúnmente utilizadas, como son la clasificación, la asociación, el agrupamiento y los patrones secuenciales. Se detalla además una de las metodologías de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico e industrial, como CRISP-DM, que está compuesta por seis fases: análisis del problema, análisis de los datos, preparación de los datos, modelado, evaluación y explotación.

Palabras clave: CRISP-DM, bases de datos, minería de datos, proceso KDD.

Abstract

The stages of the KDD process are described herein, emphasizing the data mining stage and more commonly used techniques, such as classification, association, grouping and sequential patterning. Additionally, one of the most used reference methodologies in the implementation of data mining projects in academic and industrial fields, such as CRISP-DM, is detailed. It consists of six phases: problem analysis, data analysis, data preparation, modeling, assessment and exploitation.

Keywords: CRISP-DM, databases, data mining, KDD process.

¿Cómo citar este capítulo?/How to cite this chapter?

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. DOI: <http://dx.doi.org/10.16925/9789587600490>



Introducción

El proceso de extraer conocimiento a partir de grandes volúmenes de datos ha sido reconocido por muchos investigadores como un tópico de investigación clave en los sistemas de bases de datos, y por muchas compañías industriales como una importante área y una oportunidad para obtener mayores ganancias (Timarán, 2009). Autores como Fayyad, Piatetsky-Shapiro y Smith (1996, p. 89) lo definen como “El proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente entendibles al usuario a partir de los datos”.

El Descubrimiento de conocimiento en bases de datos (KDD, del inglés *Knowledge Discovery in Databases*) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos (*data mining*) y presentar resultados (Agrawal y Srikant, 1994) (Chen, Han y Yu, 1996) (Piatetsky Shapiro, Brachman y Khabaza, 1996) (Han y Kamber, 2001). KDD se puede aplicar en diferentes dominios, por ejemplo, para determinar perfiles de clientes fraudulentos (evasión de impuestos), para descubrir relaciones implícitas existentes entre síntomas y enfermedades, entre características técnicas y diagnóstico del estado de equipos y máquinas, para determinar perfiles de estudiantes “académicamente exitosos” en términos de sus características socioeconómicas y para determinar patrones de compra de los clientes en sus canastas de mercado.

Etapas del proceso KDD

El proceso KDD que se muestra en la figura 1 es interactivo e iterativo, involucra numerosos pasos con la intervención del usuario en la toma de muchas decisiones. Se resume en las siguientes etapas:

- Selección.
- Preprocesamiento/limpieza.
- Transformación/reducción.
- Minería de datos (*data mining*).
- Interpretación/evaluación.

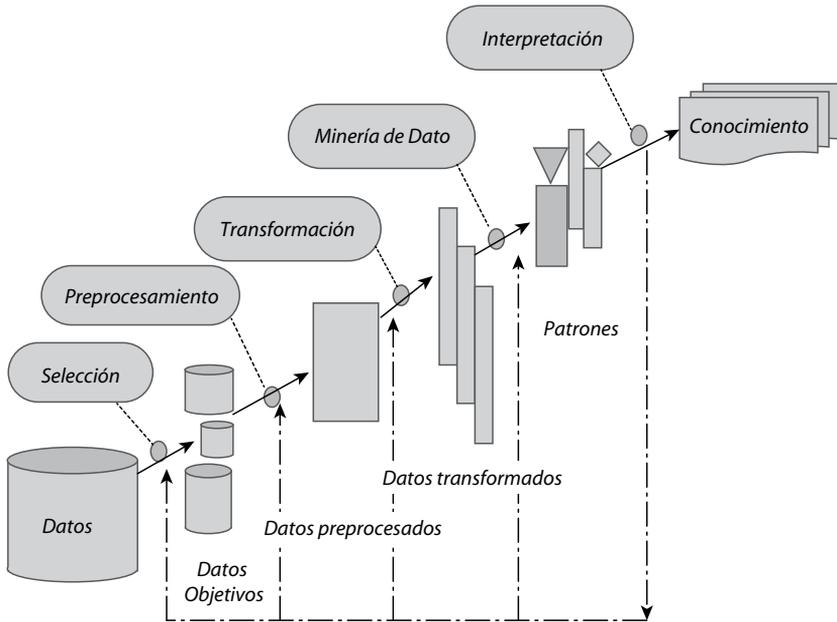


Figura 1. Etapas del proceso KDD. Elaboración propia.

Etapa de selección

En la etapa de selección, una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso KDD, desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La selección de los datos varía de acuerdo con los objetivos del negocio.

Etapa de pre-procesamiento/limpieza

En la etapa de preprocesamiento/limpieza (*data cleaning*) se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos (*missing* y *empty*), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista.

Los datos ruidosos (*noisy data*) son valores que están significativamente fuera del rango de valores esperados; se deben principalmente a errores humanos, a cambios en el sistema, a información no disponible a tiempo y a fuentes

heterogéneas de datos. Los datos desconocidos *empty* son aquellos a los cuales no les corresponde un valor en el mundo real y los *missing* son aquellos que tienen un valor que no fue capturado. Los datos nulos son datos desconocidos que son permitidos por los sistemas gestores de bases de datos relacionales (SGBDR). En el proceso de limpieza todos estos valores se ignoran, se reemplazan por un valor por omisión, o por el valor más cercano, es decir, se usan métricas de tipo estadístico como media, moda, mínimo y máximo para reemplazarlos.

Etapa de transformación/reducción

En la etapa de transformación/reducción de datos, se buscan características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos (Fayyad et al., 1996).

Los métodos de reducción de dimensiones pueden simplificar una tabla de una base de datos horizontal o verticalmente. La reducción horizontal implica la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo por otro de alto nivel, en una jerarquía definida de valores categóricos o por la discretización de valores continuos (por ejemplo, edad por un rango de edades). La reducción vertical implica la eliminación de atributos que son insignificantes o redundantes con respecto al problema, como la eliminación de llaves, la eliminación de columnas que dependen funcionalmente (por ejemplo, edad y fecha de nacimiento). Se utilizan técnicas de reducción como agregaciones, compresión de datos, histogramas, segmentación, discretización basada en entropía, muestreo, entre otras (Han y Kamber, 2001).

Etapa de minería de datos

El objetivo de la etapa minería de datos es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación (Quinlan, 1986) (Wang, Iyer y Scott, 1998), clustering (Ng y Han, 1994), (Zhang, Ramakrishnan, Livny, 1996), patrones secuenciales (Agrawal y Srikant, 1995) y asociaciones (Agrawal y Srikant, 1994), (Srikant y Agrawal, 1996), entre otras.

Las técnicas de minería de datos crean modelos que son predictivos o descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos

de variables de interés, que se denominan variables objetivo, dependientes o clases, usando otras variables denominadas independientes o predictivas, como por ejemplo predecir para nuevos clientes si son buenos o malos basados en su estado civil, edad, género y profesión, o determinar para nuevos estudiantes si desertan o no en función de su zona de procedencia, facultad, estrato, género, edad y promedio de notas. Entre las tareas predictivas están la clasificación y la regresión. Los modelos descriptivos identifican patrones que explican o resumen los datos; sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos, como identificar grupos de personas con gustos similares o identificar patrones de compra de clientes en una determinada zona de la ciudad. Entre las tareas descriptivas se cuentan las reglas de asociación, los patrones secuenciales, los *clustering* y las correlaciones.

Por lo tanto, la escogencia de un algoritmo de minería de datos incluye la selección de los métodos por aplicar en la búsqueda de patrones en los datos, así como la decisión sobre los modelos y los parámetros más apropiados, dependiendo del tipo de datos (categóricos, numéricos) por utilizar.

Etapas de interpretación/evaluación de datos

En la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para incorporarlo en otro sistema para posteriores acciones o, simplemente, para documentarlo y reportarlo a las partes interesadas; también para verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

Tareas de minería de datos

Dentro de la minería de datos se encuentran diferentes tipos de tareas, las cuales pueden considerarse como un tipo de problema para ser resuelto por un algoritmo de minería de datos (Hernández, Ramírez y Ferri, 2005). Entre las tareas de minería de datos más importantes están la clasificación, segmentación o *clustering*, asociación y patrones secuenciales.

Clasificación

La clasificación de datos permite obtener resultados a partir de un proceso de aprendizaje supervisado. Es, además, el proceso por medio del cual se encuentran propiedades comunes entre un conjunto de objetos de una base de datos y se los cataloga en diferentes clases, de acuerdo con el modelo de clasificación (Agrawal, Ghosh, Imielinsky, Iyer y Swami, 1992).

Este proceso se realiza en dos pasos: en el primero se construye un modelo, en el cual cada tupla de un conjunto de tuplas de la base de datos tiene una clase conocida (etiqueta), determinada por uno de los atributos de la base de datos llamado *atributo clase*. El conjunto de tuplas que sirve para construir el modelo se denomina *conjunto de entrenamiento* y se escoge randómicamente del total de tuplas de la base de datos. A cada tupla de este conjunto se denomina *ejemplo de entrenamiento* (Han y Kamber, 2001). En el segundo paso se usa el modelo para clasificar. Inicialmente, se estima la exactitud del modelo utilizando otro conjunto de tuplas de la base de datos, cuya clase es conocida, denominado *conjunto de prueba*. Este conjunto es escogido randómicamente y es independiente del conjunto de entrenamiento. A cada tupla de este conjunto se denomina *ejemplo de prueba* (Han y Kamber, 2001).

La exactitud del modelo, sobre el conjunto de prueba, es el porcentaje de ejemplos de prueba que son correctamente clasificadas por el modelo. Si la exactitud del modelo se considera aceptable, se puede usar para clasificar futuros datos o tuplas para los cuales no se conoce la clase a la que pertenecen. Se han propuesto varios métodos de clasificación: *rough sets*, árboles de decisión, redes neuronales, Bayes, algoritmos genéticos entre otros.

El modelo de clasificación basado en árboles de decisión es probablemente el más utilizado y popular por su simplicidad y facilidad para entender (Han y Kamber, 2001), (Sattler y Dunemann, 2001). Este modelo tiene su origen en los estudios de aprendizaje de máquina. Este es un método de aprendizaje supervisado que construye árboles de decisión a partir de un conjunto de casos o ejemplos denominados *conjunto de entrenamiento (training set)* extraídos de la base de datos. También se escoge un conjunto de prueba, cuyas características son conocidas, con el fin de evaluar el árbol.

La calidad del árbol depende de la precisión de la clasificación y del tamaño del árbol (Chen, Han y Yu, 1996). El método primero escoge un subconjunto del conjunto de entrenamiento y forma un árbol de decisión. Si el árbol no da la respuesta correcta para todos los objetos del conjunto prueba, una selección

de excepciones se adiciona al conjunto de entrenamiento y el proceso continúa hasta que se encuentra el conjunto de decisiones correctas. El eventual resultado es un árbol en el cual cada hoja lleva un nombre de la clase y cada nodo interior especifica un atributo con una rama correspondiente a cada posible valor del atributo.

Entre los algoritmos de clasificación para árboles de decisión se cuentan ID-3 (Quinlan, 1986), C4.5 (Quinlan, 1993), Sprint (Shafer, Agrawal y Metha, 1996), SLIQ (Metha, Agrawal y Rissanen, 1996) y J48 (Hall, Frank y Witten, 2011). La idea básica de estos algoritmos es construir los árboles de decisión en los que:

- Cada nodo no terminal está etiquetado con un atributo.
- Cada rama que sale de un nodo está etiquetada con un valor de ese atributo.
- Cada nodo terminal está etiquetado con un conjunto de casos, los cuales satisfacen todos los valores de atributos que etiquetan el camino desde ese nodo al nodo inicial.

La aplicación de un atributo A como criterio de selección clasifica los casos en distintos conjuntos (tantos como valores discretos del atributo). Se trata de construir el árbol de decisión más simple que sea consistente con el conjunto de entrenamiento T . Para ello hay que ordenar los atributos relevantes, desde la raíz a los nodos terminales, de mayor a menor poder de clasificación. El poder de clasificación de un atributo A es su capacidad para generar particiones del conjunto de entrenamiento que se ajuste en un grado dado a las distintas clases posibles; de esta forma se introduce un orden en dicho conjunto. El orden o el desorden (ruido) de un conjunto de datos son medibles. El poder de clasificación de un atributo se mide de acuerdo con su capacidad para reducir la incertidumbre o entropía (grado de desorden de un sistema). Esta métrica se denomina *ganancia de información*. El atributo con la más alta ganancia de información se escoge como el atributo que forme un nodo en el árbol (Quinlan, 1993) (Agrawal et al., 1992).

El árbol de decisión se construye de la siguiente forma:

- Calcular la entropía que puede reducir cada atributo.
- Ordenar los atributos de mayor a menor capacidad de reducción de entropía.
- Construir el árbol de decisión siguiendo la lista ordenada de atributos.

La ganancia de información obtenida por el particionamiento del conjunto T , de acuerdo con el atributo A se define como:

$$Gain(T, A) = I(T) - E(A)$$

Donde, $I(T)$ es la entropía del conjunto T , compuesto de s ejemplos y m distintas clases C_i ($i=1, m$) y se calcula:

$$I(T) = - \sum p_i \log_2(p_i)$$

Donde, $p_i = s_i/s$ es la probabilidad que un ejemplo cualquiera pertenezca a una clase C_i y s_i es el número de ejemplos de T de la clase C_i .

$E(A)$ es la entropía del conjunto T si es particionado por los n diferentes valores del atributo A en n subconjuntos, $\{S_1, S_2, \dots, S_n\}$, donde S_j contiene esos ejemplos de T que tienen el valor a_j en A y s_{ij} el número de ejemplos de la clase C_i en el subconjunto S_j .

$E(A)$ se calcula:

$$E(A) = \sum s_{ij}/s * I(S_{ij})$$

Donde, s_{ij} el número de ejemplos de la clase C_i en el subconjunto S_j

$$I(S_{ij}) = - \sum p_{ij} \log_2(p_{ij})$$

Donde $p_{ij} = s_{ij} / |s_j|$ es la probabilidad de que un ejemplo de S_j pertenezca a la clase C_i .

En otras palabras, $Gain(T, A)$ es la reducción esperada de la entropía causada por el particionamiento de T de acuerdo con el atributo A .

Finalmente, las reglas de clasificación se obtienen recorriendo cada rama del árbol desde la raíz hasta el nodo terminal. El antecedente de la regla es la conjunción de los pares recogidos en cada nodo y el consecuente es el nodo terminal.

Segmentación o clustering

El proceso de agrupar objetos físicos o abstractos en clases de objetos similares se llama segmentación o *clustering* o clasificación no supervisada (Chen, Han y Yu,

1996). Básicamente, el *clustering* agrupa un conjunto de datos (sin un atributo de clase predefinido) basado en el principio de maximizar la similitud intraclase y minimizar la similitud interclase. El análisis de *clustering* ayuda a construir particiones significativas de un gran conjunto de objetos basado en la metodología divide y conquista, la cual descompone un sistema de gran escala en pequeños componentes para simplificar el diseño y la implementación.

La meta del *clustering* en una base de datos es la partición de esta en segmentos o *clusters* de registros similares que comparten un número de propiedades y son considerados homogéneos. Los registros en diversos *clusters* son diferentes y estos últimos tienen una alta homogeneidad interna (dentro del *cluster*) y una alta heterogeneidad externa (entre *clusters*). Por homogeneidad se entiende que los registros en un *cluster* están próximos unos a otros; allí la proximidad se expresa por medio de una medida, dependiendo de la distancia de los registros al centro del segmento. Por heterogeneidad se entiende que los registros en diferentes segmentos no son similares de acuerdo con una medida de similaridad (Cabena, Hadjinian, Stadler, Verhees y Zanasi, 1998).

La segmentación, típicamente, permite descubrir subpoblaciones homogéneas: por ejemplo, se aplica a una base de datos de clientes, para mejorar la exactitud de los perfiles, identificando subgrupos de clientes que tienen un comportamiento similar al comprar.

El algoritmo de *clustering* segmenta una base de datos sin ninguna indicación por parte del usuario sobre el tipo de *clusters* que va a encontrar en la base de datos, y desecha cualquier sesgo o intuición por parte del usuario; así potencia el verdadero descubrimiento de conocimiento. Por esta razón, al método de segmentación o *clustering* se lo denomina *aprendizaje no supervisado*. Algunos de los algoritmos utilizados para *clustering* son: K-Means (Han y Kamber, 2001), Clarans (*Clustering Large Applications based upon Randomized Search*) (Ng y Han, 1994), y Birch (*Balanced Iterative Reducing and Clustering using Hierarchies*) (Zhang, Ramakrishnan y Livny, 1996).

El *clustering* se utiliza por ejemplo en el análisis de flujo de efectivo para un grupo de clientes que paga en un período del mes en particular, para hacer segmentación de mercado y para descubrir grupos de afinidades. También se utiliza para descubrir subpoblaciones homogéneas de consumidores en bases de datos de *marketing*.

Asociación

La tarea de asociación descubre patrones en forma de reglas, que muestran los hechos que ocurren frecuentemente juntos en un conjunto de datos determinado. El problema fue formulado por Agrawal et al. (1992), y a menudo se referencia como el problema de canasta de mercado (*market-basket*). En este problema se da un conjunto de ítems y una colección de transacciones que son subconjuntos (canastas) de estos ítems. La tarea es encontrar relaciones entre los ítems de esas canastas para descubrir reglas de asociación que cumplan unas especificaciones mínimas dadas por el usuario, expresadas en forma de soporte y confianza. Las cantidades de ítems comprados en una transacción no se toman en cuenta, lo que significa que cada ítem es una variable binaria que representa si un ítem está presente o no en una transacción.

Formalmente, sea $I = \{i_1, i_2, \dots, i_m\}$ un conjunto de literales, llamados ítems; sea D un conjunto de transacciones, donde cada transacción T es un conjunto de ítems tal que $T \subseteq I$. Cada transacción se asocia con un identificador llamado $\pi(T)$. Sea X un conjunto de ítems. Se dice que una transacción T contiene a X si y solo si $X \subseteq T$.

Una regla de asociación es una implicación de la forma $X \Rightarrow Y$, donde X y Y son conjuntos de ítems que $X \subset I$, $Y \subset I$ y $X \cap Y = \emptyset$.

El significado intuitivo de tal regla es que las transacciones de la base de datos que contienen X tienden a contener Y . La regla $X \Rightarrow Y$ se cumple en el conjunto de transacciones D con una confianza c si el $c\%$ de las transacciones en D que contienen X también contienen Y . La regla $X \Rightarrow Y$ tiene un soporte s en el conjunto de transacciones D si el $s\%$ de las transacciones en D contienen $X \cup Y$.

La confianza denota la fuerza de la implicación y el soporte indica la frecuencia de ocurrencia de los patrones en la regla. Las reglas con una confianza alta y soporte fuerte son referidas como reglas fuertes (*strong rules*) (Agrawal et al., 1992). El problema de encontrar reglas de asociación se descompone en los siguientes pasos:

- Descubrir los *itemsets* frecuentes, i.e., el conjunto de ítems que tienen el soporte de transacciones por encima de un predeterminado soporte s mínimo.
- Usar los *itemsets* frecuentes para generar las reglas de asociación para la base de datos.

Después de que los *itemsets* frecuentes son identificados, las correspondientes reglas de asociación se pueden derivar de una manera directa. Un ejemplo de una regla de asociación es “el 30% de las transacciones que contienen cerveza también contienen pañales; el 2% de todas las transacciones contienen a ambos ítems” (Agrawal et al., 1996, p. 244). Aquí el 30% es la confianza de la regla y el 2%, el soporte de la regla.

Según Han y Kamber (2001), existen varios criterios para clasificar las reglas de asociación, uno de estos es el de las dimensiones que estas abarcan. De acuerdo con este criterio, las reglas de asociación pueden ser unidimensionales y multidimensionales. Una regla de asociación es unidimensional, si los ítems o atributos de la regla hacen referencia a un solo predicado o dimensión. Por ejemplo, se tiene la siguiente regla de asociación:

Cerveza \wedge papas fritas \Rightarrow pañales, que se puede reescribir como:

Compra (cerveza) \wedge compra (papas fritas) \Rightarrow compra (pañales), hace referencia a una sola dimensión: compra.

Una regla de asociación es multidimensional, si los ítems o atributos de la regla hacen referencia a dos o más criterios o dimensiones. Por ejemplo, está la siguiente regla de asociación:

Edad (30...39) \wedge ocupación (ingeniero) \Rightarrow compra (laptop), contiene tres predicados: *edad*, *ocupación* y *compra*.

Un uso clásico de asociaciones es el análisis de la canasta de mercado, en la cual la asociación es una lista de afinidades de productos. Por ejemplo, observar los pedidos individuales de clientes para suministros de oficina puede generar una regla: el 70% de los clientes que ordenan plumas y lápices también ordenan libretas.

Otras aplicaciones de reglas de asociación son los análisis de demandas médicas para determinar procedimientos médicos que se realizan al mismo tiempo o a lo largo de un periodo, para un diagnóstico en particular. También se aplican para el análisis de textos, diseño de catálogos, segmentación de clientes basado en patrones de compra, en mercadeo, entre otros.

Patrones secuenciales

Los patrones secuenciales buscan ocurrencias cronológicas. El problema de descubrimiento de patrones secuenciales se trata en Agrawal y Srikant (1995). Se aplica principalmente en el análisis de la canasta de mercado y su objetivo es descubrir en los clientes ciertos comportamientos de compra en el tiempo. El dato de entrada es un conjunto de secuencias llamado *data-secuencia*. Cada una de estas últimas es una lista de transacciones, en las que cada transacción es un conjunto de ítems (literales). Típicamente, hay un tiempo asociado con cada transacción.

Un patrón secuencial también se compone de una lista de conjuntos de ítems. El problema es encontrar todos los patrones secuenciales que cumplan con un soporte mínimo especificado por el usuario, en el cual el soporte es el porcentaje de *data-secuencias* que contiene el patrón. Por ejemplo, en una base de datos de una librería, cada *data-secuencia* puede corresponder a todas las selecciones de libros de un cliente y cada transacción, a los libros seleccionados por el cliente en un orden.

Un patrón secuencial puede ser “El 5% de los clientes que compran ‘*Foundation*’, después ‘*Foundation y Empire*’ y luego ‘*Second Foundation*’ (Agrawal y Srikant, 1995, p. 3). La *data-secuencia* correspondiente al cliente, quien compró otros libros conjuntamente o después de estos libros, contiene este patrón secuencial. La *data-secuencia* puede también tener otros libros en la misma transacción, así como uno de los libros del patrón. Elementos de un patrón secuencial pueden ser conjuntos de ítems; por ejemplo, “‘*Foundation*’ y ‘*Ringworld*’, seguido por ‘*Foundation y Empire*’ y ‘*Ringworld Engineers*’”. Sin embargo, todos los ítems en un elemento de un patrón secuencial deben estar presentes en una transacción simple para que la *data-secuencia* soporte al patrón (Agrawal y Srikant, 1995).

Los patrones secuenciales, en el dominio de la medicina, se pueden utilizar por ejemplo para ayudar a identificar síntomas y enfermedades que preceden a otras enfermedades.

Áreas relacionadas con el proceso KDD

KDD se ha desarrollado y continúa desarrollándose con base en las investigaciones realizadas en los campos del aprendizaje de máquina, reconocimiento de patrones, bases de datos, estadística, inteligencia artificial, sistemas expertos, visualización

de datos y computación de alto rendimiento. La meta común es la extracción de conocimiento de los datos en el contexto de grandes bases de datos.

KDD se relaciona con aprendizaje de máquina y reconocimiento de patrones en el estudio de teorías y algoritmos de minería de datos para modelamiento de datos y extracción de patrones. Así mismo, se enfoca en la extensión de estas teorías y algoritmos al problema de encontrar patrones entendibles que puedan ser interpretados como conocimiento útil o interesante, y hace un fuerte énfasis sobre el trabajo con grandes conjuntos de datos del mundo real.

KDD tiene que ver con la estadística, particularmente con el análisis exploratorio de datos. La inferencia de conocimiento a partir de los datos tiene un componente estadístico fundamental (Elder y Pregibon, 1996). La estadística provee un lenguaje y una estructura para cuantificar el grado de certeza de los resultados cuando se trata de inferir patrones generales sobre una muestra particular de toda una población.

Data warehousing es otra área con la que se relaciona KDD y se refiere a la actual tendencia de los negocios de coleccionar y limpiar datos transaccionales con el fin de que se encuentren disponibles para el análisis en línea y el soporte de decisiones. Un popular método para el análisis de bodegas de datos (*data warehouse*) es OLAP (*On-line Analytical Processing*) (Gill y Rao, 1996). Las herramientas de OLAP se enfocan en proveer análisis de datos multidimensional y están destinadas hacia la simplificación y soporte interactivo de análisis de datos, mientras que el objetivo de las herramientas de DCBD es automatizar el proceso, tanto como sea posible.

Metodología CRISP-DM

Generalidades

En 1993, líderes de la industria como Daimler Benz, SPSS de Inglaterra, OHRA de Holanda, NCR de Dinamarca, consorcio de empresas Europeas, y AG de Alemania construyeron el acrónimo CRISP-DM (*Cross-Industry Standard Process for Data Mining*), el cual tiene como finalidad proporcionar nuevas ideas a los que decidan trabajar con minería de datos. Esta metodología tiene la ventaja de que no ha sido construida de manera teórica y académica, sino que se basa en experiencias reales de cómo la gente hace proyectos (Moro, Laureano y Cortez, 2011) (Martínez y Podestá, 2014) (Raus, Vegega, Pytel y Pollo-Cattaneo, 2014).

Este modelo es uno de los más utilizados como guía de referencia en el desarrollo de proyectos de minería de datos. La metodología CRISP-DM consiste en un conjunto de tareas que están organizadas en cuatro niveles de abstracción: fases, tareas generales, tareas especializadas e instancias de proceso (véase figura 2). Dichos niveles están establecidos respetando jerarquías en tareas; inician en el nivel más general hasta llegar, finalmente, a los casos más específicos (Chapman et al., 2000).

Ciclo de vida de la metodología CRISP-DM

El modelo provee una representación completa del ciclo de vida de un proyecto de minería de datos. El proceso es dinámico e iterativo, por lo que la ejecución de los procesos no es estricta y con frecuencia se puede pasar de uno a otro proceso, de atrás hacia delante y viceversa. Estos dependen del resultado de cada fase o la planeación de la siguiente tarea por ejecutar.

Estas fases ayudan a las organizaciones a entender el proceso y proveen de un “mapa del camino” que se debe seguir, así: conocimiento del negocio, conocimiento de los datos, preparación de los datos, modelado, evaluación, despliegue (véase figura 3).

Cada fase se estructura en varias tareas generales, las tareas generales se proyectan en tareas específicas, en las cuales finalmente se describen las acciones que deben ser desarrolladas para situaciones definidas (Larose y Larose, 2014).

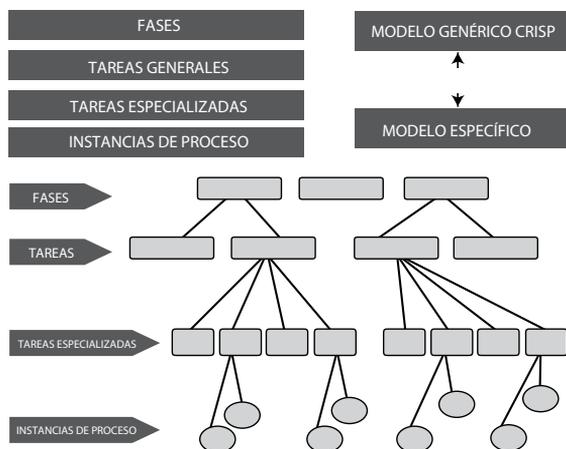


Figura 2. Esquema de los cuatro niveles de CRISP-DM. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

Fases de la metodología

Fase 1. Comprensión del negocio o problema. Comprende los requisitos y objetivos del proyecto desde una perspectiva empresarial o institucional para convertirlos en objetivos técnicos y en un plan de proyecto, para lo cual es necesario comprender de manera completa el problema por resolver (ver figura 4).

- **Determinar los objetivos.** Se determina cuál es el problema que se quiere resolver y por qué se usa minería de datos para dicho propósito; también se deben fijar los criterios de éxito. En cuanto a estos últimos, pueden ser de tipo cualitativo o de tipo cuantitativo; por ejemplo, si el problema es detectar fraude en el uso de tarjetas de crédito, el criterio de éxito cuantitativo sería el número de detecciones de fraude.
- **Evaluar la situación actual.** En esta tarea se evalúan antecedentes y requisitos del problema, tanto en términos del negocio como en términos de la minería de datos. Algunos de los aspectos por tener en cuenta pueden ser el conocimiento previo acerca del tema, la cantidad de datos requeridos para resolver el problema, ventajas de aplicar minería de datos al problema, entre otros.

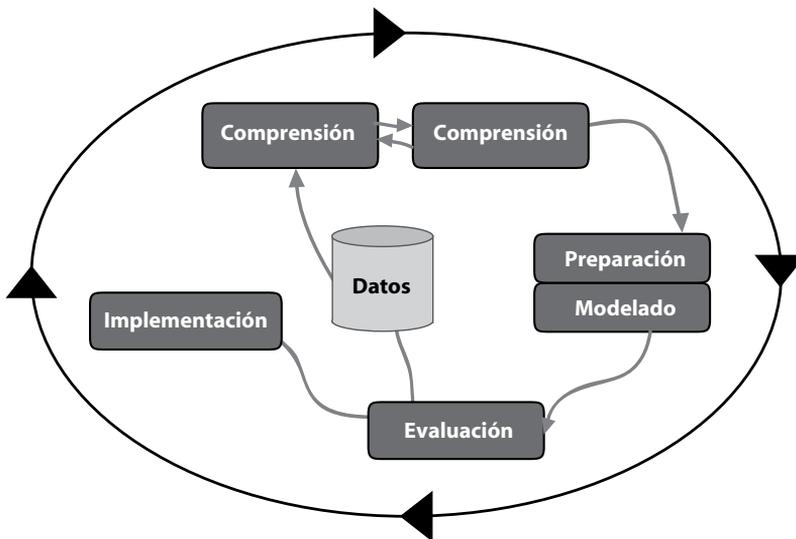


Figura 3. Ciclo de vida de CRISP-DM. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

- **Determinar los objetivos de la minería de datos.** El objetivo de esta tarea es representar los objetivos del negocio en términos de las metas del proyecto de minería de datos. Por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar asignación de créditos hipotecarios, la meta de la minería de datos sería determinar el perfil de los clientes respecto a su capacidad de endeudamiento.
- **Producir un plan de proyecto.** La última tarea de esta fase tiene como objetivo desarrollar el plan de proyecto considerando los pasos que se deben seguir y los métodos por emplear en cada paso.

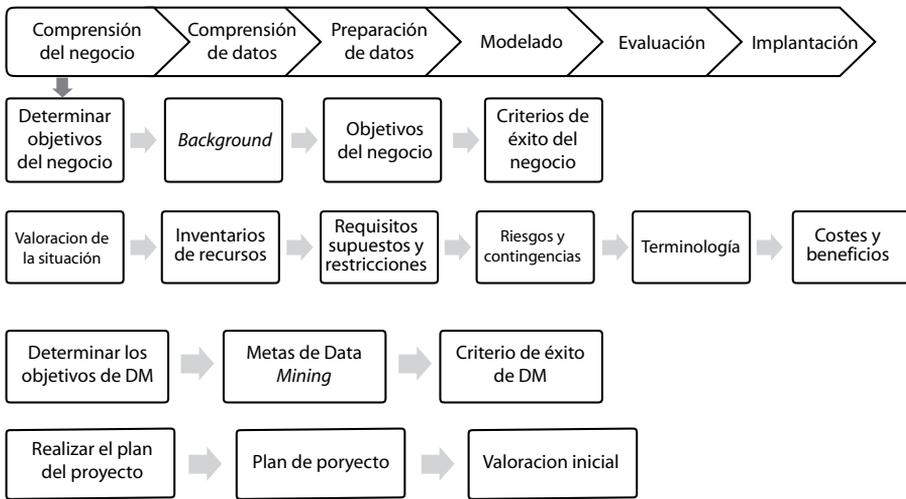


Figura 4. Fase de comprensión del negocio. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

Fase 2. Comprensión de los datos. Corresponde a la recolección inicial de los datos para establecer un primer contacto con el problema; esta fase, junto con la fase 3 y la fase 4, demanda mayor esfuerzo y tiempo (véase figura 5).

Las principales tareas que se deben desarrollar en la fase de comprensión de los datos son: recolectar datos iniciales, describir los datos, explorar los datos y verificar la calidad de los datos.

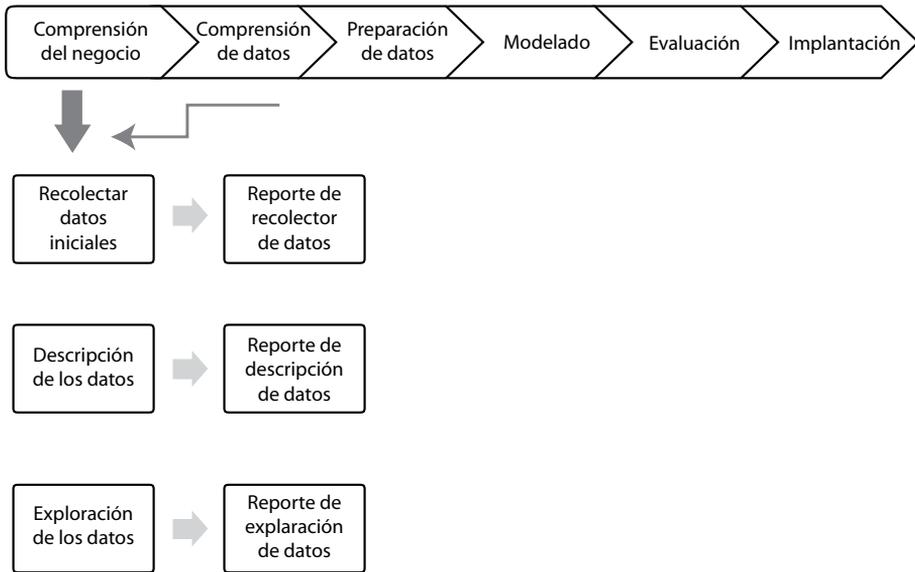


Figura 5. Fase de comprensión de los datos. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

- **Recolectar datos iniciales.** Tiene como objetivo principal la recolección de datos iniciales y su adecuación para su posterior procesamiento. Se deben elaborar informes con una lista de los datos adquiridos, su localización, las técnicas utilizadas en su recolección y los problemas y soluciones inherentes a este proceso.
- **Describir los datos.** Se deben describir los datos iniciales obtenidos, tales como número de registros y campos por registro, su identificación, el significado de cada campo y la descripción del formato inicial.
- **Explorar los datos.** Su finalidad es descubrir una estructura general para los datos. Involucra la aplicación de pruebas estadísticas básicas, que revelen propiedades en los datos, se crean tablas de frecuencia y se construyen gráficos de distribución. Se crea un informe de exploración de datos.
- **Verificar la calidad de los datos.** Se realiza la verificación de los datos para determinar la consistencia de los valores de los campos, la cantidad y distribución de los valores nulos, encontrar valores fuera de rango que pueden

ser ruido para el proceso. Se tiene como objetivo asegurar la completitud y corrección de los datos.

Fase 3. Preparación de los datos. Se usa para adaptarlos a la técnica de minería de datos, mediante la visualización de los datos y la búsqueda de relaciones entre las variables. Esta fase es la de modelado, ya que los datos requieren ser procesados de diferentes formas; por ende, las fases de preparación y modelado interactúan permanentemente (véase figura 6).

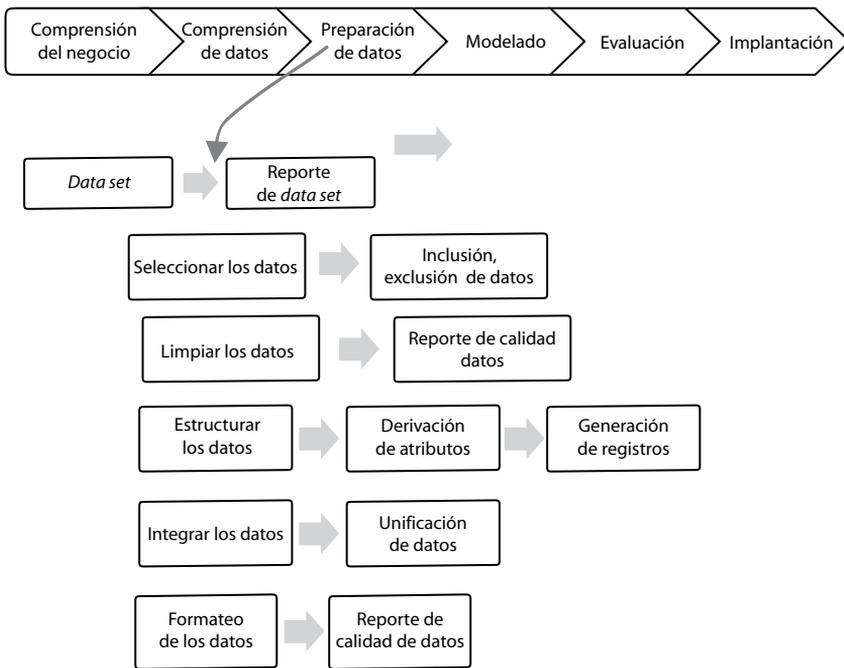


Figura 6. Fase de preparación de los datos. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

Los pasos que se consideran para la preparación de los datos son: seleccionar, limpiar, estructurar, integrar y formatear los datos.

- **Seleccionar los datos.** Se selecciona un subconjunto de datos considerando la calidad de los datos, la limitación en el volumen o en los tipos de datos que están relacionados con las técnicas de minería de datos seleccionados.

- **Limpiar los datos.** Existe una diversidad de técnicas aplicables a esta tarea con el fin de optimizar la calidad de los datos en la perspectiva de prepararlos para la fase de modelación. Algunas de las técnicas son: normalización de los datos, discretización de campos numéricos, tratamiento con valores vacíos, reducción del volumen de datos.
- **Estructurar los datos.** Algunas de las operaciones por realizar en esta tarea son la generación de nuevos atributos a partir de atributos ya existentes, integración de nuevos registros o transformación de valores para atributos existentes.
- **Integrar los datos.** Involucra la creación de nuevas estructuras; por ejemplo, crear nuevos campos, nuevos registros, fusión de tablas o nuevas tablas.
- **Formatear los datos.** Consiste principalmente en transformar sintácticamente los datos sin modificar su significado con el fin de permitir o facilitar, en particular, el empleo de alguna técnica de minería de datos; por ejemplo, eliminar comas, tabuladores, caracteres especiales, espacios, máximos y mínimos para las cadenas de caracteres, etc.

Fase 4. Modelado. Corresponde a la selección de un modelo adecuado y específico; para ello se usan técnicas que cumplan los siguientes criterios (véase figura 7):

- Ser apropiada para el problema.
- Disponer de datos adecuados.
- Cumplir con los requisitos del problema.
- Técnica adecuada para obtener un modelo.
- Conocimiento pleno de la técnica.

Por ejemplo, si el problema es de clasificación, podemos elegir entre árboles de decisión, *k-nearest neighbour* o razonamiento basado en caos (CBR).

- **Generar plan de prueba.** Se debe generar un plan para probar la calidad y validez del modelo construido; por ejemplo, en una tarea como la clasificación es posible usar la razón de error como medida de la calidad. Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba.
- **Construir el modelo.** Se ejecuta la técnica seleccionada sobre los datos preparados para generar uno o más modelos. Todas las técnicas del modelado tienen un conjunto de parámetros que determinan características del modelo

por generar. La tarea de selección de los mejores parámetros es iterativa, basada en los resultados generados. Estos deben ser interpretados y su rendimiento justificado.

- **Evaluar el modelo.** Se deben interpretar los modelos de acuerdo con el conocimiento del dominio y los criterios de éxitos preestablecidos.

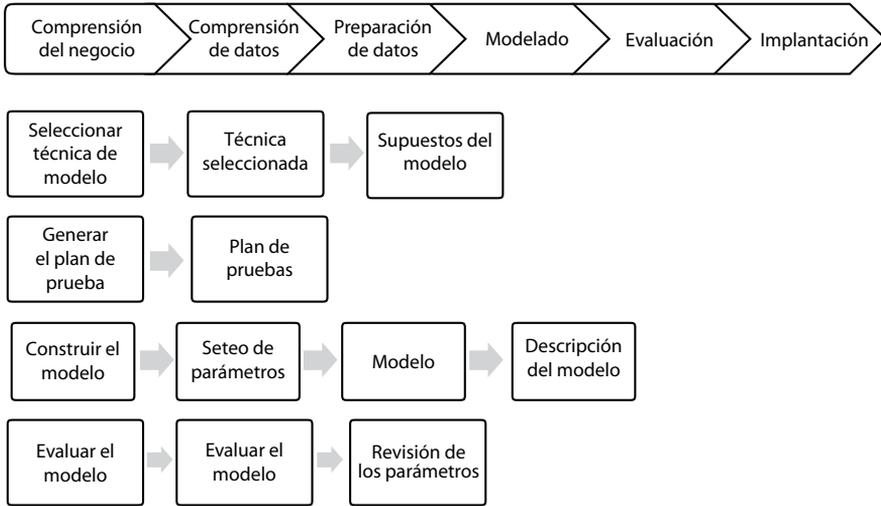


Figura 7. Fase de modelado. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

Fase 5. Evaluación. Evalúa el modelo teniendo en cuenta el cumplimiento de los criterios de éxito del problema; para ello se emplean múltiples herramientas para la interpretación de los resultados, entre ellas matrices de confusión Edelstein 1999, que es una tabla que indica cuántas clasificaciones se han hecho para cada tipo. La diagonal de la tabla representa las clasificaciones correctas (figura 8).

Si es válido lo anterior, se procede a la explotación del modelo, que es el mantenimiento de la aplicación y la posible difusión de los resultados.

Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio; la retroalimentación generada por la monitorización y mantenimiento puede indicar si el modelo está siendo utilizado apropiadamente.

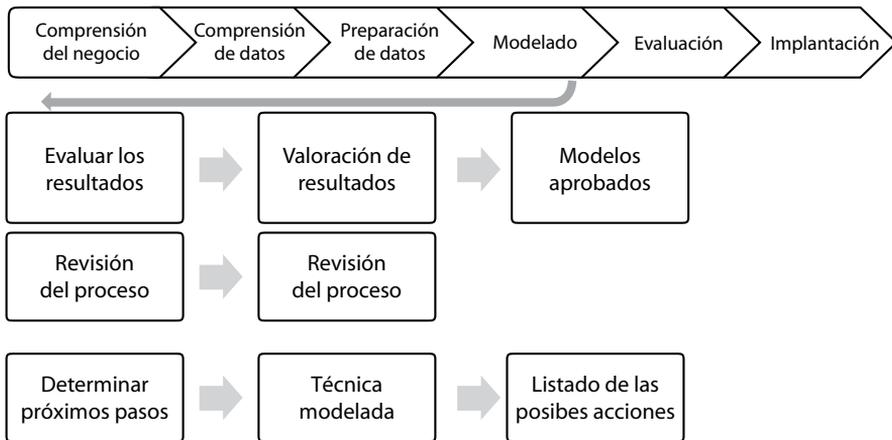


Figura 8. Fase de evaluación. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

Fase 6. Implementación. Es aquí donde el conocimiento obtenido se transforma en acciones dentro del proceso de negocio, ya sea observando el modelo y resultados, o aplicándolo a múltiples grupos de datos o como parte del proceso. Las tareas que se efectúan son: planear la implementación, monitorizar y mantener, informe final y revisar el proyecto.

- **Planear la implementación.** Esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación. Si un procedimiento general se ha identificado para crear el modelo, debe estar documentado para su posterior implementación (véase figura 9).
- **Monitorizar y mantener.** Se deben preparar estrategias de monitorización y mantenimiento para ser aplicadas sobre los modelos.
- **Informe final.** Dependiendo del plan de implementación, este puede ser un resumen de los puntos importantes del proyecto y la experiencia lograda, o puede ser una presentación final que incluya y explique los resultados logrados con el proyecto.
- **Revisar el proyecto.** Se evalúa lo correcto y lo incorrecto.

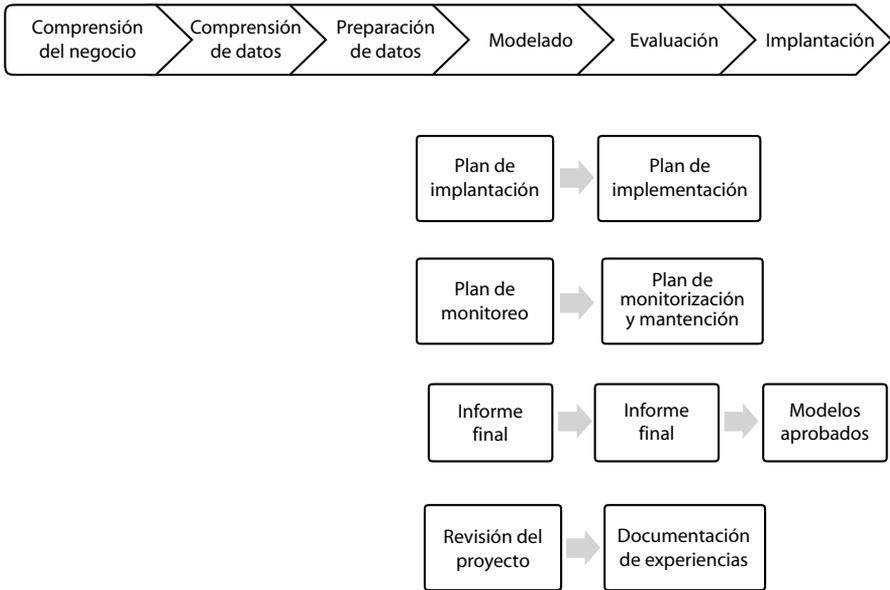


Figura 9. Fase de implementación. Tomado de *CRISP-DM 1.0 Step-by-Step Data Mining Guide* (P. Chapman et al.), 2000.

En este capítulo se describieron los conceptos fundamentales del proceso de descubrimiento de conocimiento en bases de datos, haciendo énfasis en la etapa de minería de datos donde se especificaron las tareas y técnicas de minería de datos más importantes. De igual forma se describió la metodología CRISP-DM que sirvió de base para esta investigación.

Referencias

Agrawal, R. y Srikant, R. (1994). Fast Algorithms for Mining Association Rules. *Vldb Conference*, Santiago de Chile.

Agrawal, R. y Srikant, R. (1995). Mining Sequential Patterns. *The 11th International Conference on Data Engineering ICDE*, Taipei, República de China.

Agrawal, R., Ghosh S., Imielinski, T., Iyer, B. y Swami, A. (1992). An Interval Classifier for Database Mining Applications. *Proceedings VDLB Conference*, Vancouver.

- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. y Zanasi A. (1998). *Discovering Data Mining from Concept to Implementation*, Prentice Hall. Recuperado de <http://dl.acm.org/citation.cfm?id=270298>
- Chapman, P., Clinton, J., Randy, K., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R. (2000). *CRISP-DM 1.0 Step-by-Step Data Mining Guide*. Recuperado de <http://www.crisp-dm.org/CRISPPWP-0800.pdf>
- Chen, M., Han, J. y Yu, P. (1996). Data Mining: An Overview from Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*.
- Elder, J. y Pregibon, D. (1996). A Statistical Perspective on Knowledge Discovery in Databases. En *Advances in Knowledge Discovery and Data Mining, AAAI Pres/ The MIT Press*.
- Fayyad, U., Piatestky-Shapiro, G. y Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(11), 27-34.
- Gill, H. y Rao, P. (1996). *Data warehousing: la integración de información para la mejor toma de decisiones*. Prentice-Hall.
- Hall, M., Frank, E., y Witten, I. (2011). *Practical Data Mining: Tutorials*. University of Waikato. Disponible en: www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf
- Han, J. y Kamber, M. (2001). *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Hernández, J., Ramírez, M. y Ferri, C. (2005). *Introducción a la Minería de Datos*. Madrid: Editorial Pearson Educación SA.
- Larose, D. y Larose, Ch. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (2da. ed.). New Jersey: John Wiley & Sons.
- Martínez, D, y Podestá, C. (2014). Metodología de estudio del rendimiento académico mediante la minería de datos. *Campus Virtuales*, 3(1), 56-73.
- Metha M., Agrawal R., Rissanen J. (1996). SLIQ: A Fast Scalable Classifier for Data Mining. *Proceedings EDBT96*. Avignon, France.
- Moro, S., Laureano, R. y Cortez, P. (2011). Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. *Proceedings of European Simulation and Modelling Conference -ESM'2011*, 117–121. Recuperado de http://sci2s.ugr.es/docencia/in/pdf/MoroCortezLaureano_DMApproach4DirectMKT.pdf
- Ng, R. y Han, J. (1994). Efficient and Effective Clustering Method for Spatial Data Mining. *VLDB Conference*. Santiago de Chile, Chile.
- Piatetsky-Shapiro, G., Brachman, R. y Khabaza, T. (1996). *An Overview of Issues in Developing Industrial Data Mining and Knowledge Discovery Applications*. Association for the

Advancement of Artificial Intelligence [AAAI], MIT Press. Recuperado de <http://www.aaai.org/Papers/KDD/1996/KDD96-015.pdf>

Quinlan, J. (1986). Induction of Decision Trees. *Machine Learning Journal*, 1(1), 81-106.

Raus, N., Vegega, C., Pytel, P. y Pollo-Cattaneo, M. (2014). Metodología propuesta para la predicción de deserción universitaria mediante explotación de información (pp. 1014-1158). *WICC 2014 XVI Workshop de investigadores en ciencias de la computación*. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/43835/Documento_completo.pdf?sequence=1

Srikant R. y Agrawal, R. (1996). *Mining quantitative association rules in large relational tables*, ACM SIGMOD, Montreal. Recuperado de <http://rakesh.agrawal-family.com/papers/sigmod-96qassoc.pdf>

Sattler, K. y Dunemann, O. (2001). SQL database primitives for decision tree classifiers. En *Proceedings of the tenth international conference on Information and knowledge management* (pp. 379-386). Atlanta: CIKM. Recuperado de <http://dl.acm.org/citation.cfm?id=502650>

Shafer J., Agrawal R., Metha M. (1996). SPRINT: A Scalable Parallel Classifier for Data Mining. *Proceedings of the VLDB Conference*. Bombay, India.

Timaran, R. (2009). Una mirada al descubrimiento de conocimiento en bases de datos. *Revista Ventana Informatica*, 20, 39-58.

Wang, M., Iyer, B. y Scott, J. (1998). Scalable Mining for Classification Rules in Relational Databases. *International Database Engineering and Application Symposium - Ideas*. Cardiff, Wales.

Zhang, T., Ramakrishnan, R. y Livny, M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM SIGMOD International Conference on Management of Data*. Montreal, Canada.

3 Tendencias de desempeño académico en competencias genéricas - pruebas Saber Pro

Academic Performance Trends in Generic Skills - Saber Pro Exam

Resumen

Se realiza en este capítulo un análisis estadístico preliminar de los datos con el fin de obtener tendencias de desempeño académico de los estudiantes, en las cuatro competencias genéricas de las pruebas Saber Pro 2011-2, en relación con las variables sociodemográficas, económicas, académicas e institucionales de los estudiantes que presentaron esta prueba. Para tal efecto, se cruzan los puntajes de las competencias genéricas obtenidos en la prueba con las variables mencionadas y se utilizan para el análisis las siguientes medidas estadísticas: coeficiente de correlación, promedio, desviación estándar, intervalos de confianza y el tamaño del efecto de la diferencia estandarizada de las medias con el estadístico d de Cohen.

Palabras clave: desempeño académico, dimensiones, promedios, tendencias de desempeño académico, variables.

Abstract

In this chapter, a preliminary statistical analysis of data is conducted to obtain students' academic performance trends in four generic skills of the Saber Pro Exam 2011-2 in relation to socio-demographic, economic, academic and institutional variables of students taking this test. For such purpose, the scores of generic skills obtained in the exam are matched to the variables mentioned and used for the analysis of the following statistical measures: correlation coefficient, average, standard deviation, confidence intervals and the effect size of the standardized mean difference with Cohen's d .

Keywords: academic performance, dimensions, averages, academic performance trends, variables.

¿Cómo citar este capítulo?/How to cite this chapter?

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). Tendencias de desempeño académico en competencias genéricas - pruebas Saber Pro. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 87-99). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>



Introducción

Con el fin de tener una comprensión preliminar de los datos, se realizó un análisis de correlación entre las cuatro competencias genéricas y se estableció la tendencia estadística del desempeño en las cuatro competencias en relación con variables sociodemográficas, económicas, académicas e institucionales.

Análisis de correlación entre las competencias genéricas

Utilizando la base de datos de las pruebas Saber Pro 2011-2 del ICFES, de la cual se seleccionaron únicamente los registros de los estudiantes de programas profesionales y se almacenaron en un repositorio de datos denominado T97068A94 (este proceso se describe completamente en el capítulo siguiente), a través del coeficiente de correlación de Pearson, se establece cómo se asocian linealmente, entre sí, las cuatro competencias genéricas. Los resultados se presentan en la tabla 8.

De acuerdo con esta tabla, se observa que la lectura crítica con razonamiento cuantitativo y la lectura crítica con inglés presentan correlaciones altas ($r > 0,5$) (Cohen, 1988); la lectura crítica con comunicación escrita y el razonamiento cuantitativo con inglés presentan correlaciones moderadas ($0,3 < r \leq 0,5$); mientras que las competencias en comunicación escrita con razonamiento cuantitativo y comunicación escrita con inglés presentan correlaciones bajas ($r \leq 0,3$). Por lo tanto, se espera que los estudiantes con buen desempeño en lectura crítica también tengan buen desempeño en razonamiento cuantitativo y en inglés.

Tabla 8
Matriz de correlaciones de competencias genéricas

Competencias	Coficiente	Lectura crítica	Comunicación escrita	Razonamiento cuantitativo	Inglés
Lectura crítica	Coficiente de Correlación	1	0,355**	0,535**	0,508**
	p valor		0,000	0,000	0,000
	N		95337	97068	97068
Comunicación escrita	Coficiente de Correlación		1	0,233**	0,295**
	p valor			0,000	0,000
	N			95337	95337

(continúa)

(viene)

Competencias	Coefficiente	Lectura crítica	Comunicación escrita	Razonamiento cuantitativo	Inglés
Razonamiento cuantitativo	Coefficiente de Correlación p valor N			1	0,481** 0,000 97068
Inglés	Coefficiente de Correlación p valor N				1

**La correlación es significativa al nivel 0,01 (bilateral).

Nota. Elaboración propia.

Desempeño en las competencias genéricas según variables sociodemográficas, económicas, académicas e institucionales

Con el fin de conseguir tendencias de desempeño académico de los estudiantes en las competencias genéricas de las pruebas Saber Pro 2011-2, en relación con las variables sociodemográficas, económicas, académicas e institucionales, incluidas en el estudio, se cruzaron los puntajes de las competencias genéricas obtenidos en la prueba con las variables mencionadas y se utilizaron las medidas estadísticas: promedio, desviación estándar, intervalos de confianza y el tamaño del efecto de la diferencia estandarizada de las medias d de Cohen.

Para establecer si las diferencias en puntajes en las cuatro competencias son estadísticamente significativas, se calcularon los intervalos de confianza de medias al 95%. Aquellas categorías de las variables analizadas que muestran rangos que no se traslapan presentan verdaderas diferencias; en caso contrario, las diferencias no son estadísticamente significativas. Para su interpretación se agrupan las categorías de las variables que no presentan diferencias formando grupos simbolizados con las letras A, B, C, etc., para denotar puntajes promedios que van de los valores más altos a los más bajos. Igualmente, para determinar el tamaño de las diferencias de los promedios entre los diferentes grupos de las variables analizadas se utilizó el estadístico d de Cohen con la siguiente escala propuesta por Cohen (1998): [0.0, 0.2] trivial o muy pequeña, [0.2, 0.5] pequeña, [0.5, 0.8] moderada y [0.8, infinito] grande. Este estadístico se obtiene dividiendo la diferencia de medias (en valor absoluto) de los dos grupos por comparar, entre la desviación estándar

conjunta de estos. Para su cálculo se utiliza como referencia el grupo de más alto promedio en cada competencia.

A continuación se presenta el análisis de desempeño de los estudiantes en las pruebas Saber Pro 2011-2 según género, estado civil, nivel educativo de los padres, estrato socioeconómico, tipo de institución, modalidad de estudio, acreditación institucional y de programas.

Género y desempeño académico en competencias genéricas

En la tabla 9 se muestra el desempeño en las competencias genéricas según el género. A partir de esta tabla se concluye que en las pruebas de competencias genéricas Saber Pro 2011-2 los hombres presentan mejor desempeño que las mujeres en todas las competencias, excepto en la de escritura, en la cual las mujeres superan a los hombres, aunque según el estadístico *d* de Cohen estas diferencias son de magnitud trivial, tanto en lectura crítica como en comunicación escrita, y pequeña en razonamiento cuantitativo e inglés.

Tabla 9

Desempeño académico en competencias genéricas según género. Pruebas Saber Pro 2011-2

Competencia genérica	Género	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Femenino	56824	10,199	0,0041	10,191	10,207	B	0,11
	Masculino	40231	10,308	0,0049	10,298	10,318	A	
Comunicación escrita	Femenino	55948	10,216	0,0041	10,208	10,225	A	0,16
	Masculino	39389	10,058	0,0052	10,048	10,068	B	
Razonamiento cuantitativo	Femenino	56825	9,994	0,0039	9,987	10,002	B	0,46
	Masculino	40232	10,457	0,0055	10,447	10,468	A	
Inglés	Femenino	56755	10,222	0,0049	10,213	10,232	B	0,23
	Masculino	40191	10,507	0,0066	10,494	10,520	A	

Nota. Elaboración propia.

Estado civil y desempeño académico en competencias genéricas

La tabla 10 muestra el desempeño en las competencias genéricas según el estado civil; además, se puede observar que los solteros presentan el mejor desempeño en

todas las pruebas (A), mientras que casados y separados o divorciados pertenecen al segundo grupo (B), igualmente en las cuatro competencias. Los estudiantes viudos y en unión libre presentan el más bajo desempeño en lectura crítica y escritura, mientras que en el módulo de razonamiento cuantitativo los estudiantes en unión libre están en el grupo intermedio B y los viudos en el último grupo C. Finalmente, en el módulo de inglés, los viudos pertenecen al grupo B y los de unión libre están en el último grupo C. Según el estadístico d de Cohen todos los grupos presentan diferencias pequeñas frente a los solteros, excepto los viudos que alcanzan un tamaño moderado en los módulos de lectura crítica y razonamiento cuantitativo. En comunicación escrita las diferencias son muy pequeñas (casados, separados) y pequeñas (unión libre y viudos) frente al grupo de solteros con mejor desempeño. En el módulo de inglés los estudiantes casados, en unión libre y viudos alcanzan diferencias moderadas en los puntajes promedios frente a los solteros, mientras que las diferencias de los separados/divorciados son pequeñas frente a estos últimos.

Tabla 10

Desempeño académico en competencias genéricas según estado civil. Pruebas Saber Pro 2011-2

Competencia genérica	Estado civil	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura Crítica	Soltero	77951	10,328	0,0035	10,321	10,335	A	
	Casado	10482	9,939	0,0089	9,922	9,957	B	0,40
	Unión libre	7175	9,846	0,0112	9,824	9,868	C	0,49
	Separado/divorciado	1283	9,931	0,0255	9,881	9,980	B	0,41
	Viudo	164	9,707	0,0751	9,559	9,856	C	0,64
Comunicación escrita	Soltero	76671	10,194	0,0036	10,187	10,201	A	
	Casado	10257	10,023	0,0095	10,004	10,042	B	0,17
	Unión Libre	6993	9,896	0,0118	9,873	9,919	C	0,30
	Separado/divorciado	1262	10,056	0,0269	10,003	10,109	B	0,14
	Viudo	154	9,830	0,0765	9,679	9,981	C	0,37

(continúa)

(viene)

Competencia genérica	Estado civil	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Razonamiento cuantitativo	Soltero	77951	10,267	0,0037	10,259	10,274	A	
	Casado	10484	9,876	0,0087	9,859	9,894	B	0,38
	Unión libre	7175	9,847	0,0107	9,826	9,867	B	0,41
	Separado/divorciado	1283	9,822	0,0251	9,773	9,871	B	0,43
	Viudo	164	9,543	0,0665	9,411	9,674	C	0,70
Inglés	Soltero	77875	10,474	0,0046	10,465	10,483	A	
	Casado	10463	9,831	0,0086	9,815	9,848	B	0,52
	Unión Libre	7163	9,730	0,0093	9,711	9,748	B	0,60
	Separado/divorciado	1282	9,864	0,0263	9,813	9,916	B	0,48
	Viudo	163	9,726	0,0738	9,580	9,871	B	0,58

Nota. Elaboración propia.

Nivel educativo de los padres y desempeño académico en competencias genéricas

En la tabla 11 se muestra el desempeño en las competencias genéricas según el nivel educativo de los padres. La tendencia en el desempeño académico en las cuatro competencias genéricas está asociada directamente con el nivel educativo de los padres; a mayor nivel educativo de estos, mejor desempeño académico del estudiante. Como se observa en la mencionada tabla, el promedio del desempeño en las cuatro competencias genéricas crece en la medida que aumenta el nivel de escolaridad de los padres, pues cada uno de los seis niveles constituye un grupo, así: posgrado (A), profesional (B), técnico (C), secundaria (D), primaria (E) y sin escolaridad (F). Estos se presentan en la tabla en orden consecutivo según la media.

El valor *d* de Cohen indica que en lectura crítica y comunicación escrita las diferencias en el promedio de puntajes de los estudiantes cuyos padres alcanzan niveles de posgrado frente a los estudiantes cuyos padres tienen estudios profesionales y técnicos son pequeñas, mientras que los estudiantes con padres con niveles de secundaria y primaria las diferencias son moderadas y frente a los

estudiantes con padres sin ningún nivel educativo, son grandes en el módulo de lectura crítica y moderadas en comunicación escrita. En razonamiento cuantitativo, frente al grupo de mejor desempeño, los estudiantes con padres que tienen estudios profesionales presentan diferencias muy pequeñas; los estudiantes con padres con estudios técnicos y de secundaria, diferencias pequeñas, y los de padres con estudios de primaria o ninguno, diferencias moderadas. En inglés, las diferencias son pequeñas cuando los padres tienen estudios profesionales; moderadas si tienen estudios técnicos y grandes si alcanzan máximos estudios de secundaria, frente a los estudiantes con padres que tienen estudios de posgrado.

Tabla 11

*Desempeño académico en competencias genéricas según nivel educativo de los padres.
Pruebas Saber Pro 2011-2*

Competencia genérica	Nivel educativo padres	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Posgrado	10964	10,675	0,0091	10,657	10,693	A	
	Profesional	17915	10,459	0,0073	10,445	10,473	B	0,22
	Técnico	10207	10,331	0,009	10,313	10,349	C	0,37
	Secundaria	26765	10,218	0,0057	10,207	10,229	D	0,49
	Primaria	21664	10,017	0,0063	10,004	10,029	E	0,70
	Ninguno	9538	9,842	0,0096	9,823	9,861	F	0,88
Comunicación escrita	Posgrado	10817	10,399	0,01	10,379	10,419	A	
	Profesional	17668	10,247	0,0077	10,232	10,262	B	0,15
	Técnico	10022	10,205	0,0099	10,185	10,224	C	0,19
	Secundaria	26272	10,124	0,0061	10,112	10,136	D	0,27
	Primaria	21250	10,042	0,0067	10,029	10,055	E	0,36
	Ninguno	9307	9,949	0,0101	9,929	9,969	F	0,45
Razonamiento cuantitativo	Posgrado	10964	10,599	0,0108	10,577	10,62	A	
	Profesional	17914	10,383	0,0082	10,367	10,399	B	0,19
	Técnico	10207	10,253	0,0099	10,233	10,272	C	0,32
	Secundaria	26768	10,165	0,0061	10,153	10,177	D	0,42
	Primaria	21663	9,979	0,0063	9,966	9,991	E	0,62
	Ninguno	9539	9,803	0,0092	9,785	9,821	F	0,77

(continúa)

(viene)

Competencia genérica	Nivel educativo padres	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Inglés	Posgrado	10958	11,231	0,0147	11,203	11,26	A	
	Profesional	17898	10,818	0,0106	10,798	10,839	B	0,28
	Técnico	10196	10,379	0,0112	10,357	10,401	C	0,63
	Secundaria	26736	10,214	0,0066	10,201	10,227	D	0,83
	Primaria	21642	9,901	0,0061	9,889	9,913	E	1,15
	Ninguno	9514	9,730	0,0083	9,713	9,746	F	1,20

Nota. Elaboración propia.

Estrato socioeconómico y desempeño académico en competencias genéricas

Según la tabla 12, el desempeño académico en las cuatro competencias se asocia con el estrato socioeconómico del estudiante; esto se marca en la tendencia de que a mayor estrato, mejor desempeño académico. Los resultados muestran que el promedio del desempeño asciende en la medida que sube el estrato social del estudiante. Sin embargo, el estadístico *d* de Cohen sugiere que las diferencias del promedio de puntajes, frente al estrato 6, van desde muy pequeñas (estratos 5 y 4) hasta moderadas (estrato 1 y rural) en el módulo de lectura crítica; desde muy pequeñas (estrato 5) hasta grandes (estrato 1) en los módulos de comunicación escrita y razonamiento cuantitativo (estrato 1), y en el módulo de inglés los estudiantes de todos los estratos presentan diferencias grandes, excepto los de estrato 5 que tienen diferencias pequeñas.

Tabla 12

Desempeño académico en competencias genéricas según estrato socioeconómico. Pruebas Saber Pro 2011-2

Competencia genérica	Estrato social	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Estrato 6	2605	10,777	0,0193	10,739	10,815	A	
	Estrato 5	5067	10,646	0,0134	10,62	10,673	B	0,14
	Estrato 4	13617	10,510	0,0082	10,494	10,527	C	0,28

(continúa)

(viene)

Competencia genérica	Estrato social	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Estrato 3	36353	10,301	0,0049	10,291	10,31	D	0,51
	Estrato 2	29271	10,101	0,0055	10,09	10,111	E	0,72
	Zona rural	112	9,996	0,1014	9,795	10,196	F	0,79
	Estrato 1	10042	9,746	0,0099	9,726	9,765	G	1,04
Comunicación escrita	Estrato 6	2571	10,463	0,0196	10,424	10,501	A	
	Estrato 5	5004	10,363	0,0145	10,335	10,391	B	0,10
	Estrato 4	13430	10,309	0,0088	10,292	10,326	C	0,15
	Estrato 3	35789	10,192	0,0053	10,181	10,202	D	0,27
	Estrato 2	28705	10,070	0,0058	10,058	10,081	E	0,40
	Estrato 1	9730	9,836	0,0098	9,817	9,855	F	0,64
	Zona rural	108	9,791	0,0885	9,615	9,966	F	0,68
Razonamiento cuantitativo	Estrato 6	2605	10,747	0,0229	10,702	10,792	A	
	Estrato 5	5067	10,573	0,0159	10,541	10,604	B	0,15
	Estrato 4	13617	10,426	0,0095	10,407	10,444	C	0,29
	Estrato 3	36353	10,232	0,0053	10,221	10,242	D	0,50
	Zona rural	112	10,024	0,0867	9,852	10,196	E	0,62
	Estrato 2	29271	10,049	0,0056	10,038	10,06	E	0,71
Inglés	Estrato 6	2605	12,219	0,031	12,158	12,28	A	
	Estrato 5	5067	11,455	0,022	11,412	11,498	B	0,49
	Estrato 4	13617	10,913	0,0124	10,888	10,937	C	0,89
	Estrato 3	36353	10,306	0,0061	10,294	10,318	D	1,60
	Zona rural	112	10,232	0,1284	9,978	10,487	D	1,26
	Estrato 2	29271	9,953	0,0059	9,942	9,965	E	2,12
	Estrato 1	10042	9,644	0,0095	9,625	9,662	F	2,32

Nota. Elaboración propia.

Modalidad de estudio y desempeño académico en competencias genéricas

Se observa en la tabla 13 que los estudiantes que se forman en modalidad presencial (grupo A) obtienen mejor desempeño académico que aquellos que estudian en programas a distancia (grupo B).

Las diferencias entre estos dos grupos, en todos los módulos son moderadas, excepto en comunicación escrita, en la cual la diferencia es pequeña, según el estadístico *d* de Cohen.

Tabla 13

Desempeño académico en competencias genéricas según modalidad de estudio. Pruebas Saber Pro 2011-2

Competencia genérica	Modalidad de estudio	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Presencial	84272	10,326	0,0033	10,32	10,333	A	0,66
	A distancia	12783	9,702	0,0079	9,686	9,718	B	
Comunicación escrita	Presencial	82888	10,193	0,0035	10,186	10,2	A	0,32
	A distancia	12449	9,870	0,0085	9,853	9,887	B	
Razonamiento cuantitativo	Presencial	84272	10,259	0,0036	10,252	10,266	A	0,54
	A distancia	12785	9,706	0,0074	9,692	9,721	B	
Inglés	Presencial	84202	10,451	0,0044	10,442	10,459	A	0,69
	A distancia	12744	9,613	0,0063	9,601	9,625	B	

Nota. Elaboración propia.

Tipo de institución y desempeño académico en competencias genéricas

La tabla 14 muestra dos grupos: las IES oficiales (grupo A), privadas y de régimen especial (grupo B). Se puede observar que los estudiantes del grupo A tienen mejor desempeño que los del grupo B. En la competencia de comunicación escrita no hay diferencia significativa en los desempeños académicos de estudiantes de IES oficiales, privadas y de régimen especial. En razonamiento cuantitativo, se distinguen dos grupos: IES oficiales y de régimen especial (grupo A) y privadas (grupo B); en estas el mejor desempeño lo muestran los estudiantes del grupo A. Finalmente, en inglés se notan dos grupos: IES privadas (grupo A) y oficiales y de régimen especial (grupo B), en las cuales presentan un mejor desempeño los del grupo A, lo cual puede ocurrir por tener mejores posibilidades de acceso a materiales y cursos adicionales.

Sin embargo, teniendo en cuenta el estadístico *d* de Cohen, en todas las competencias genéricas las diferencias de los promedios de los puntajes son muy pequeñas o pequeñas.

Tabla 14

*Desempeño académico en competencias genéricas según tipo de institución.
Pruebas Saber Pro 2011-2*

Competencia genérica	Tipo de institución	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Oficial	38864	10,269	0,0051	10,259	10,279	A	
	Privada	58144	10,228	0,0039	10,220	10,236	A	0,04
	Régimen especial	47	9,996	0,1197	9,755	10,237	B	0,27
Comunicación escrita	Oficial	38099	10,137	0,0052	10,126	10,147	A	
	Privada	57191	10,161	0,0042	10,152	10,169	A	0,02
	Régimen especial	47	10,155	0,1508	9,852	10,459	A	0,02
Razonamiento cuantitativo	Oficial	38867	10,251	0,0055	10,241	10,262	A	
	Privada	58143	10,143	0,0041	10,135	10,151	B	0,11
	Régimen especial	47	10,296	0,134	10,026	10,565	A	0,04
Inglés	Oficial	38810	10,219	0,0059	10,207	10,230	B	0,16
	Privada	58089	10,422	0,0054	10,412	10,433	A	
	Régimen especial	47	9,847	0,0973	9,651	10,043	C	0,44

Nota. Elaboración propia.

Acreditación institucional y desempeño académico en competencias genéricas

En la tabla 15 se muestra el desempeño en las competencias genéricas según la acreditación institucional. Se evidencia que estudiantes de IES acreditadas (grupo A) tienen mejor desempeño académico en las cuatro competencias objeto estudio, en comparación con estudiantes que se forman profesionalmente en IES no acreditadas (grupo B).

Según el estadístico *d* de Cohen, las diferencias de medias entre los dos grupos anteriores son pequeñas en comunicación escrita, moderadas en lectura crítica y razonamientos cuantitativos y grandes en el módulo de inglés.

Tabla 15

Desempeño académico en competencias genéricas según acreditación institucional. Pruebas Saber Pro 2011-2

Competencia genérica	Acreditación institucional	N	Media	Error típico	IC media al 95%		Grupo	d
					LI	LS		
Lectura crítica	Acreditada	16306	10,835	0,0072	10,821	10,849	A	
	No acreditada	80749	10,125	0,0033	10,118	10,131	B	0,76
Comunicación escrita	Acreditada	16152	10,547	0,0082	10,531	10,563	A	
	No acreditada	79185	10,070	0,0035	10,063	10,077	B	0,48
Razonamiento cuantitativo	Acreditada	16307	10,833	0,009	10,815	10,850	A	
	No acreditada	80750	10,056	0,0034	10,049	10,062	B	0,78
Inglés	Acreditada	16300	11,508	0,0118	11,485	11,531	A	
	No acreditada	80646	10,104	0,0037	10,097	10,112	B	1,23

Nota. Elaboración propia.

Acreditación del programa y desempeño académico en competencias genéricas

El mismo comportamiento anterior en el desempeño académico se encuentra en los programas acreditados (grupo A) y no acreditados (grupo B), tal como se muestra en la tabla 16; en este punto, el mejor desempeño lo presentan los estudiantes de programas acreditados.

Según el estadístico *d* de Cohen, las diferencias de medias entre los dos grupos anteriores son pequeñas en lectura crítica, comunicación escrita y razonamientos cuantitativos, y moderadas en inglés.

Tabla 16

*Desempeño académico en competencias genéricas según Acreditación del Programa.
Pruebas Saber Pro 2011-2*

Acreditación del programa	N	Media	Error típico	IC media al 95%		Grupo	d	
				LI	LS			
Lectura crítica	Acreditado	14036	10,628	0,0081	10,612	10,644	A	0,47
	No acreditado	83019	10,179	0,0033	10,173	10,186	B	
Comunicación escrita	Acreditado	13839	10,383	0,0087	10,366	10,400	A	0,27
	No acreditado	81498	10,112	0,0035	10,105	10,118	B	
Razonamiento cuantitativo	Acreditado	14037	10,615	0,0098	10,596	10,634	A	0,50
	No acreditado	83020	10,114	0,0034	10,107	10,120	B	
Inglés	Acreditado	14021	11,027	0,0124	11,003	11,051	A	0,65
	No acreditado	82925	10,224	0,0041	10,216	10,232	B	

Nota. Elaboración propia.

Referencias

Cohen, J. (1988). *Análisis de poder estadístico para las Ciencias del comportamiento* (2da. ed.). Nueva Jersey: Lawrence Erlbaum.

4 Descubrimiento de patrones de desempeño académico en las competencias genéricas

Discovery of Academic Performance Patterns in Generic Skills

Resumen

De acuerdo con los lineamientos Saber Pro del ICES, todos los estudiantes de programas de pregrado deben presentar los módulos de competencias genéricas sin importar el programa de formación que cursen, el cual incluye competencias de lectura crítica, razonamiento cuantitativo, escritura e inglés. En este capítulo se presentan los resultados de cada una de las fases de la metodología CRISP-DM que se utilizó, con el fin de detectar patrones de desempeño académico en las competencias genéricas de los estudiantes colombianos de programas profesionales que presentaron las pruebas de Estado Saber Pro 2011-B. Con este objetivo, se describe cómo se selecciona, de las bases de datos del ICES, la información sociodemográfica, económica, académica e institucional de estos estudiantes, que sirve de soporte para las posteriores fases de la metodología. Se detalla además cómo se construye, limpia y transforma un repositorio de datos y cómo a partir de él se descubren patrones asociados al buen o mal desempeño académico de los estudiantes en cada competencia, utilizando un modelo de clasificación basado en árboles de decisión.

Palabras clave: árboles de decisión, competencias genéricas, patrones de desempeño académico, metodología CRISP-DM.

Abstract

According to the ICES Saber Pro guidelines, all undergraduates must take the generic skills modules, regardless of the undergraduate program they belong to, including critical reading, quantitative reasoning, writing and English skills. This chapter discusses the results of each phase of the CRISP-DM methodology used in order to detect academic performance patterns in generic skills of Colombian undergraduates who took the state exam Saber Pro 2011-B. To this end, it describes how the socio-demographic, economic, academic and institutional information of these students is selected from the ICES databases to support subsequent phases of the methodology. It is also detailed how data repositories are built, purged and transformed and, based on this, how patterns associated with good or poor academic performance of students in each skill are discovered using the decision tree-based classification model.

Keywords: decision trees, generic skills, academic performance patterns, CRISP-DM methodology.

¿Cómo citar este capítulo?/How to cite this chapter?

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). Descubrimiento de patrones de desempeño académico en las competencias genéricas. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 101-150). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>



Introducción

La investigación fue de tipo descriptivo bajo enfoque cuantitativo, aplicando un diseño no experimental. Para el descubrimiento de patrones de desempeño académico, se aplicó la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) utilizada, principalmente, en los ambientes académico e industrial; metodología que se ha constituido en la guía de referencia más ampliamente utilizada en el desarrollo de este tipo de proyectos.

Comprensión del negocio o problema

En esta fase se realizaron las actividades que permitieron profundizar y apropiarse de una manera completa el problema objeto de estudio, como también los objetivos y requisitos de esta investigación. Esto posibilitó la recolección y producción de los datos necesarios para efectuar una interpretación adecuada de los resultados.

De igual manera, la fase se fundamentó teóricamente sobre temas de calidad de la educación superior, desempeño académico, formación profesional desde el enfoque de competencias y competencias genéricas; además, se abordó el objeto de la investigación, es decir, la concepción y estructura de las pruebas Saber Pro 2011-2; todo lo cual se encuentra de forma detallada y precisa en el capítulo 1.

Comprensión de los datos

En esta fase, el grupo de investigación identificó, recopiló y se familiarizó con la información disponible en las bases de datos del ICFES, sobre las competencias genéricas en las pruebas Saber Pro 2011-2 y sobre los datos sociodemográficos, económicos, académicos e institucionales de los estudiantes que presentaron esta prueba, pertenecientes a programas profesionales.

A partir de las bases de datos del ICFES, se construyó un repositorio inicial en el cual se encuentran almacenados los datos de todos los estudiantes de programas técnico profesionales, tecnológicos y profesionales que presentaron las pruebas Saber Pro 2011-2. Este repositorio inicial, que se denomina T153123A94, cuenta con 153 123 registros y 94 atributos. De este repositorio se seleccionaron únicamente los registros de estudiantes de programas profesionales objeto de este

estudio; de esta manera, quedó un repositorio con 97 068 registros y 94 atributos, identificado como T97068A94, el cual sirvió de base para las fases subsiguientes. En la tabla 17 se encuentra el diccionario de datos de este repositorio.

Tabla 17

Diccionario de datos del repositorio inicial T97068A94

N.º	Atributo	Descripción	Valores
1	estu_consecutivo	Código identificador del estudiante	
2	estu_cod_aplicacion	Nombre de la aplicación	EK+año-semestre
3	estu_genero	Género	M - Masculino F - Femenino
4	estu_nacimiento_dia	Día de nacimiento del estudiante	
5	estu_nacimiento_mes	Mes de nacimiento del estudiante	
6	estu_nacimiento_anno	Año de nacimiento del estudiante	
7	estu_pais_reside	País de residencia	
8	estu_estado_civil	Estado civil	Ver tabla 18
9	estu_disc_invidente	Discapacidad invidente	
10	estu_disc_sordo_con_interprete	Discapacidad sordo - requiere intérprete	
11	estu_disc_sordo_sin_interprete	Discapacidad sordo - no requiere intérprete	
12	estu_disc_motriz	Discapacidad motriz	
13	estu_disc_sordo_ceguera	Discapacidad sordo-ceguera	
14	estu_zona	Código identificador de la zona de la ciudad donde reside	
15	estu_anno_egreso	Año en que salió de bachiller	
16	econ_area_vive	Área donde vive	1. Cabecera municipal 2. Área rural
17	estu_reside_codmpio	Código DANE del municipio de residencia	Códigos DANE
18	estu_exam_codmpio_presentacion	Código del municipio donde presentó el examen	Códigos DANE
19	estu_exam_mpio_presentacion	Nombre municipio donde presentó el examen	
20	estu_exam_dpto_presentacion	Nombre del departamento donde presentó el examen	

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
21	estu_exam_cod	Código identificador del examen/prueba que está presentando el estudiante	Código 548
22	estu_exam_nombre	Nombre del examen que presentó el estudiante	Examen 2011-2
23	inst_cod_institucion	Código SNP de la institución de educación superior	
24	inst_nombre_institucion	Nombre de la institución de educación superior	
25	inst_origen	Origen institución educativa	Ver tabla 19
26	inst_caracter_academico	Carácter académico	Ver tabla 20
27	estu_prac_id_prgrm_academico	Código identificador programa académico	
28	estu_prgrm_academico	Nombre programa académico	
29	estu_nivel_prgrm_academico	Nivel del programa académico que cursa el estudiante	Universitario
30	estu_metodo_prgrm	Metodología del programa académico bajo la cual cursa el estudiante	Presencial Semi-presencial Distancia (tradicional) A distancia (virtual)
31	dipo_codigomunicipio	Código municipio del programa	Código DANE
32	inst_cod_jornada	Código de la jornada de clases	Ver tabla 21
33	estu_area_conoc	Nombre del área de conocimiento a la que pertenece el programa académico del estudiante	Ver tabla 22
34	estu_nucleo_pregrado	Nombre del núcleo de pregrado al que pertenece el programa académico del estudiante	Ver tabla 23
35	estu_grupo_referencia	Nombre del grupo de referencia al que pertenece el programa académico del estudiante	Ver tabla 24
36	estu_cod_grupo_ref	Código del grupo de referencia al que pertenece el programa académico del estudiante	Ver tabla 24
37	estu_semestre_cursa	Semestre que cursa actualmente el estudiante	1 a 12 semestres 12 significa 12 o más.

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
38	estu_pje_creditos	Porcentaje de créditos cursados y aprobados por el estudiante	Ver tabla 25
39	inst_vlr_matricula_ant	Valor anual de la matrícula el año anterior	Ver tabla 26
40	estu_sn_matricula_propio	Si canceló la matrícula año anterior con recursos propios	Sí (1) No (0)
41	estu_sn_matricula_padres	Si canceló la matrícula año anterior con recursos de los padres o familiares	Sí (1) No (0)
42	estu_sn_matricula_beca	Si canceló la matrícula año anterior con recursos de beca o subsidio	Sí (1) No (0)
43	estu_sn_matricula_credito	Si canceló la matrícula año anterior con recursos de crédito	Sí (1) No (0)
44	estu_titulo_bto	Título de bachiller obtenido por el estudiante	A-Académico N-Normalista T-Técnico
45	estu_exam_semestre_pretacion	Semestre en el que presentó el examen de Estado	1 o 2 semestre
46	estu_exam_anno_presentacion	Año en el que presentó el examen de Estado	
47	estu_hogar_actual	Tipo de hogar actual donde reside el estudiante	1. Es el habitual-permanente 2. Es temporal por razones de estudio u otra razón
48	fami_num_pers_grup_fam	Número total de personas que conforman el grupo familiar	
49	estu_sn_cabeza_fmliia	Si el estudiante es cabeza de familia	Sí (1) No (2)
50	fami_num_pers_cargo	Número de personas a cargo (cuando es cabeza de familia)	0 a 5 personas 5 representa 5 o más
51	fami_cod_educa_padre	Código nivel educativo más alto alcanzado por el padre del estudiante	Ver tabla 27
52	fami_cod_educa_madre	Código nivel educativo más alto alcanzado por la madre del estudiante	Ver Tabla 27
53	fami_sn_lee_escribe_padre	Si puede leer y escribir el padre del estudiante	Sí (1) No (2)

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
54	fami_sn_lee_escribe_madre	Si puede leer y escribir la madre del estudiante	Sí (1) No (2)
55	fami_cod_ocup_padre	Ocupación actual del padre del estudiante (o última si falleció)	Ver Tabla 28
56	fami_cod_ocup_madre	Ocupación actual de la madre del estudiante (o última si falleció)	Ver tabla 28
57	estu_estrato	Estrato socioeconómico de la vivienda donde reside actualmente el estudiante según el recibo del servicio de energía eléctrica	Ver tabla 29
58	fami_num_hermanos	Número de hermanos del estudiante	
59	fami_nivel_hermano	Nivel educativo más alto alcanzado por los hermanos mayores del estudiante	Ver tabla 30
60	econ_cuartos	Número total de cuartos que dispone la residencia del estudiante	
61	fami_nivel_sisben	Nivel de clasificación del Sisben	Ver Tabla 31
62	econ_material_pisos	Material de los pisos que predomina en la vivienda donde reside el hogar habitual o permanente del estudiante	Ver Tabla 32
63	econ_sn_televisor	Si hay televisor en el hogar habitual del estudiante	Sí (1) No (0)
64	econ_sn_motocicleta	Si hay una motocicleta en el hogar habitual del estudiante	Sí (1) No (0)
65	econ_sn_energia	Si posee el servicio de energía eléctrica en el hogar habitual del estudiante	Sí (1) No (0)
66	econ_sn_acueducto	Si posee el servicio de acueducto en el hogar habitual del estudiante	Sí (1) No (0)
67	econ_sn_alcantarillado	Si posee alcantarillado en el hogar habitual del estudiante	Sí (1) No (0)
68	econ_sn_aseo	Si posee el servicio de aseo en el hogar habitual del estudiante	Sí (1) No (0)
69	econ_sn_estufa	Si posee estufa a gas o eléctrica en el hogar habitual del estudiante	Sí (1) No (0)
70	econ_sn_telefonia	Si posee el servicio de telefonía en el hogar habitual del estudiante	(1): Sí (0): no
71	econ_sn_internet	Si posee el servicio de Internet en el hogar habitual del estudiante	(1): Sí (0): no

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
72	econ_sn_servicio_tv	Si posee el servicio de servicio de televisión por cable, satélite o parabólica en el hogar habitual del estudiante	(1): Sí (0): no
73	econ_sn_computador	Si hay un computador en el hogar habitual del estudiante	(1): Sí (0): no
74	econ_sn_celular	Si hay un teléfono celular en el hogar habitual del estudiante	(1): Sí (0): no
75	econ_sn_dvd	Si hay un reproductor DVD en el hogar habitual del estudiante	(1): Sí (0): no
76	econ_sn_lavadora	Si hay una máquina lavadora de ropas en el hogar habitual del estudiante	(1): Sí (0): no
77	econ_sn_microondas	Si hay un horno microondas en el hogar habitual del estudiante	(1): Sí (0): no
78	econ_sn_automovil	Si hay un automóvil particular en el hogar habitual del estudiante	(1): Sí (0): no
79	econ_sn_horno	Si hay un horno eléctrico o a gas en el hogar habitual del estudiante	(1): Sí (0): no
80	econ_sn_nevera	Si hay una nevera o enfriador en el hogar habitual del estudiante	(1): Sí (0): no
81	infa_dormitorios	Número de dormitorios en el hogar habitual del estudiante	1 a 10 dormitorios 10 representa 10 o más
82	fami_ing_fmliar_mensual	ingresos mensuales familiares en salarios mínimos	Ver tabla 33
83	estu_trabaja	Si estudiante trabaja o no	Ver tabla 34
84	estu_horas_trabajo	Nro. de horas / semanal si trabaja	
85	estu_otro_idioma_lee	Si lee otro idioma	Ver tabla 35
86	estu_otro_idioma_habla	Si habla otro idioma	Ver tabla 35
87	estu_nivel_postgrado	Si el estudiante posee un postgrado	05 -Especialización 06 -Maestría 07 -Doctorado 08 -Ninguno
88	estu_etnia	Si el estudiante pertenece a una etnia	Ver tabla 36
89	mod_lectura_critica	Puntaje asignado al módulo de lectura crítica	
90	mod_comunica_escrita_punt	Puntaje asignado al módulo de comunicación escrita	

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
91	mod_comunica_escrita_desem	Desempeño asignado al módulo de comunicación escrita	
92	mod_razona_cuantitativo_punt	Puntaje asignado al módulo de razonamiento cuantitativo	
93	mod_ingles_punt	Puntaje asignado al módulo de inglés	
94	mod_ingles_desem	Desempeño asignado al módulo de inglés	

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 18
Estado Civil

Código	Descripción
01	Soltero(a)
02	Casado(a)
03	Viudo(a)
04	Separado(a) o divorciado
05	Unión libre

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 19
Origen institución

Descripción
No oficial - corporación
No oficial - fundación
No oficial
Oficial
Oficial departamental
Oficial municipal
Oficial nacional
Régimen especial

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 20
Carácter académico

Descripción
Académico
Escuela tecnológica
Institución tecnológica
Institución universitaria
Normalista
Técnica profesional
Técnica profesional
Universidad

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 21
Código jornada

Código	Descripción
1	Diurno
2	Nocturno
3	Educación a distancia
4	Mixta
5	Registro calificado
6	Registro calificado - acreditación voluntaria
9	No aplica
10	Registro simple
11	Registro alta calidad
12	Jornada única

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 22
Área del conocimiento

Descripción
Agronomía, veterinaria y afines
Bellas artes
Ciencias de la educación
Ciencias de la salud
Ciencias sociales y humanas
Economía, administración, contaduría y afines
Ingeniería, arquitectura, urbanismo y afines
Matemáticas y ciencias naturales

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 23
Núcleo pregrado

Código	Descripción
1	Administración
2	Agronomía
3	Antropología, artes liberales
4	Arquitectura
5	Artes representativas
6	Artes plásticas, visuales y afines
7	Bacteriología
8	Bibliotecología, otros de ciencias sociales y humanas
9	Biología, microbiología y afines
10	Ciencia política, relaciones internacionales
11	Comunicación social, periodismo y afines
12	Contaduría pública
13	Deportes, educación física y recreación
14	Derecho y afines
15	Diseño
16	Diseño
17	Economía
18	Educación

(continúa)

(viene)

Código	Descripción
19	Enfermería
20	Filosofía, teología y afines
21	Física
22	Formación relacionada con el campo militar o policial
23	Geografía, historia
24	Geología, otros programas de ciencias naturales
25	Ingeniería administrativa y afines
26	Ingeniería agrícola, forestal y afines
27	Ingeniería agroindustrial, alimentos y afines
28	Ingeniería ambiental, sanitaria y afines
29	Ingeniería biomédica y afines
30	Ingeniería civil y afines
31	Ingeniería de minas, metalurgia y afines
32	Ingeniería de sistemas, telemática y afines
33	Ingeniería eléctrica y afines
34	Ingeniería electrónica, telecomunicaciones y afines
35	Ingeniería industrial y afines
36	Ingeniería mecánica y afines
37	Ingeniería química y afines
38	Instrumentación quirúrgica
39	Lenguas modernas, literatura, lingüística y afines
40	Matemáticas, estadística y afines
41	Matemáticas, estadística y afines
42	Medicina
43	Medicina veterinaria
44	Música
45	Nutrición y dietética
46	Odontología
47	Optometría, otros programas de ciencias de la salud
48	Otras ingenierías
49	Otros programas asociados a bellas artes
50	Psicología
51	Publicidad y afines
52	Química y afines

(continúa)

(viene)

Código	Descripción
53	Salud pública
54	Sin clasificar
55	Sin especificar
56	Sociología, trabajo social y afines
57	Terapias
58	Zootecnia

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 24
Grupo de Referencia

Código	Detalle
1	Arquitectura y urbanismo
2	Artes, diseño, comunicación
3	Bellas artes y diseño
4	Ciencias agropecuarias
5	Ciencias económicas y administrativas
6	Ciencias militares y navales
7	Ciencias naturales y exactas
8	Ciencias sociales
9	Comunicación, periodismo y publicidad
10	Derecho
11	Educación
12	Grupo referencia nacional
13	Humanidades
14	Ingeniería
15	Medicina
16	Militar y policial
17	Recreación y deportes
18	Salud

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 25
Porcentaje de créditos cursados y aprobados

Código	Descripción
0	No sigue el sistema de créditos
1	Menos del 40% (años 2008-1, 2008-2 y 2009-2)
1	Menos del 75% (año 2010-2)
2	Entre el 40% y el 59% (años 2008-1, 2008-2 y 2009-2)
2	Entre el 75% y el 80% (año 2010-2)
3	Entre el 60% y el 79% (años 2008-1, 2008-2 y 2009-2)
3	Entre el 81% y el 90% (año 2010-2)
4	El 80% o más (años 2008-1, 2008-2 y 2009-2)
4	Más del 90% (año 2010-2)

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 26
Valor anual de la matrícula año anterior

Código	Descripción
1	No pagó matrícula
2	Menos de 500 mil
3	Entre 500 mil y menos de 1 millón
4	Entre 1 millón y menos de 3 millones
5	Entre 3 millones y menos de 5 millones
6	Más de 5 millones

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 27
Educación del padre o madre

Código	Descripción
0	Ninguno
9	Primaria incompleta
10	Primaria completa

(continúa)

(viene)

Código	Descripción
11	Bachillerato incompleto
12	Bachillerato completo
13	Educación técnica o tecnológica sin título
14	Educación técnica o tecnológica con título
15	Educación profesional sin título
16	Educación profesional con título
17	Postgrado

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 28
Ocupación del padre o madre

Código	Descripción
13	Empresario(a) dueño(a) de empresa industrial, comercial, agropecuaria o de servicios de más de 10 trabajadores.
14	Pequeño empresario(a) dueño(a) de microempresa o pequeño negocio familiar, de finca-parcela, que vive de su explotación.
15	Empleado con cargo como director(a) o gerente general de empresa privada -entidad pública.
16	Empleado(a) de nivel directivo (con personas a cargo) en empresa privada -entidad pública.
17	Empleado(a) de nivel técnico/profesional de empresa privada -entidad pública.
18	Empleado(a) de nivel auxiliar -administrativo de empresa privada -entidad pública.
19	Obrero u operario empleado(a) de empresa privada -entidad pública.
20	Profesional independiente ejerce su profesión sin vinculación laboral permanente en empresa privada -entidad pública.
21	Trabajador por cuenta propia ejerce un oficio sin vinculación laboral: comerciantes por cuenta propia; obreros u operarios independientes; jornaleros; trabajadores independientes que prestan servicios.
22	Hogar personas dedicadas principalmente a las labores del hogar.
23	Pensionado(a) persona que vive de una pensión por concepto de jubilación.
24	Rentista persona que vive de ingresos —rentas— generados de su patrimonio.

(continúa)

(viene)

Código	Descripción
25	Estudiante persona que la mayor parte del tiempo se dedica al estudio.
26	Otra actividad u ocupación.
99	No sabe.

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 29
Estrato socioeconómico

Código	Descripción
1	Estrato 1
2	Estrato 2
3	Estrato 3
4	Estrato 4
5	Estrato 5
6	Estrato 6
8	Vive en una zona rural donde no hay estratificación socioeconómica

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 30
Educación de los hermanos mayores

Código	Descripción
0	Ninguno
9	Primaria incompleta
10	Primaria completa
11	Bachillerato incompleto
12	Bachillerato completo
13	Educación técnica o tecnológica sin título
14	Educación técnica o tecnológica con título

(continúa)

(viene)

Código	Descripción
15	Educación profesional sin título
16	Educación profesional con título
17	Postgrado
88	No tiene hermanos mayores
99	No sabe

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 31
Nivel de Sisben

Código	Descripción
1	Nivel 1
2	Nivel 2
3	Nivel 3
4	Está clasificada en otro nivel
5	No está clasificada por el Sisben

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 32
Material de los pisos

Código	Descripción
1	Tierra, arena
2	Cemento, gravilla, ladrillo
3	Madera burda, tabla -tablón
4	Madera pulida, baldosa, tableta, mármol, alfombra
5	Madera pulida, mármol, alfombra -tapete de pared a pared

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 33
Ingresos familiares mensuales

Código	Descripción
1	Menos de 1 SM
2	Entre 1 y menos de 2 SM
3	Entre 2 y menos de 3 SM
4	Entre 3 y menos de 5 SM
5	Entre 5 y menos de 7 SM
6	Entre 7 y menos de 10 SM
7	10 o más SM

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 34
Trabajo del estudiante

Código	Descripción
0	No
3	Sí, para contribuir a pagar su matrícula y/o los gastos del hogar
4	Sí, por ser práctica obligatoria del programa de estudios
5	Sí, para adquirir experiencia y/o recursos para sus gastos personales

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 35
Lee o habla otro idioma

Código	Descripción
01	Inglés
02	Francés
03	Alemán
04	Italiano
05	Portugués
06	Japonés
99	Otro

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior.* Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Tabla 36
Etnia estudiante

Código	Descripción
01	Comunidades negras
02	Raizal (isleño)
03	Paez
04	Sikuani
05	Arhuaco
06	Emberá
07	Guambiano
08	Pijao
09	Wayúu
10	Zenú
11	Pasto
12	Cancuamo
13	Inga
14	Tucano
15	Huitoto
16	Cubeo
17	Comunidad Rom (gitana)
99	Otro

Nota. Tomado de *Examen Saber Pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Instituto Colombiano para la Evaluación de la Educación [ICFES], 2011a.

Con base en la conceptualización de desempeño académico y los antecedentes teóricos sobre los factores que intervienen en él —que se asocian y colindan unos con otros—, los 94 atributos del conjunto de datos T97068A94 se clasificaron en cuatro dimensiones: sociodemográfica, económica, académica e institucional. La tabla 37 presenta esta clasificación.

Tabla 37
Clasificación de atributos en dimensiones

Dimensión	Atributo
Sociodemográfica	estu_consecutivo, estu_cod_aplicación, estu_genero, estu_nacimiento_dia, estu_nacimiento_mes, estu_nacimiento_anno, estu_pais_reside, estu_estado_civil, estu_disc_invidente, estu_disc_sordo_con_ interprete, estu_disc_sordo_sin_interprete, estu_disc_motriz, estu_disc_sordo_ ceguera, estu_zona, econ_area_vive, estu_reside_codmpio, estu_hogar_actual, fami_num_pers_grup_fam, estu_sn_cabeza_fmliia, fami_num_pers_cargo, fami_cod_educa_padre, fami_cod_educa_madre, fami_sn_lee_escribe_padre, fami_sn_lee_escribe_madre, fami_cod_ocup_padre, fami_cod_ocup_madre, fami_num_hermanos, fami_nivel _hermano, estu _etnia
Económica	inst_vlr_matricula_ant, estu_sn_matricula_propio, estu_sn_matricula_padres, estu_sn_matricula_beca, estu_sn_matricula_credito, estu_estrato, econ_cuartos, fami_nivel_sisben, econ_material_pisos, econ_sn_televisor, econ_sn_motocicleta, econ_sn_energía, econ_sn_acueducto, econ_sn_alcantarillado, econ_sn_aseo, econ_sn_estufa, econ_sn_telefonia econ_sn_internet, econ_sn_servicio_tv, econ_sn_computador, econ_sn_celular , econ_sn_dvd, econ_sn_lavadora, econ_sn_microondas, econ_sn_automovil, econ_sn_horno, econ_sn_nevera, infa_dormitorios, fami_ing_fmliar_mensual, estu_trabaja, estu_horas_trabajo
Académica	estu_anno_egreso, estu_exam_codmpio_presentacion, ,estu_exam_mpio_ presentacion, estu_exam_dpto_presentacion, estu_exam_cod, estu_exam_nombre, estu_prac_id_ prgrm_academico, estu_prgm_academico, estu_nivel_prgm_ academico, estu_metodo_progr, dipo_codigomunicipio, inst_cod_jornada, estu_area_conoc, estu_nucleo_pregrado, estu_grupo _referencia, estu_cod_grupo_ref, estu_semestre_cursa, estu_pje_ creditos, estu_titulo_bto, estu_exam_semestre_pretacion, estu_exam_anno_ presentación, estu_otro_idioma _lee, estu_otro _idioma _habla, estu_nivel_postgrado
Institucional	inst_cod_institucion, inst_nombre_institucion, inst_origen, inst_caracter_academico, dipo_codigo_municipio
Clases	mod_lectura _critica, mod_comunica_escrita_punt, mod_comunica _escrita_desem, mod_razona_cuantitativo_punt, mod_ingles_punt, mod_ingles_desem

Nota. Elaboración propia.

Además de lo descrito, se realizó una exploración de los datos del repositorio T97068A94 y, mediante un análisis preliminar, se determinó qué variables son potencialmente importantes para el estudio. Este proceso se describe en el capítulo IV.

Preparación de los datos

En esta fase, los 94 atributos del repositorio base T97068A94, considerados por el ICFES como los más importantes para capturar la información de las pruebas Saber Pro 2011-2, fueron depurados teniendo en cuenta la calidad de los datos y las técnicas de minería de datos por aplicar; además, se limpiaron (eliminación de datos nulos y valores constantes) e integraron los datos; se generaron atributos adicionales a partir de los existentes por ganancia de información y se realizaron transformaciones o cambios de formato a los valores de los atributos que se consideraron necesarios. Como resultado de esta fase se obtuvo un repositorio de datos limpio y transformado, listo para aplicarle las técnicas de minería de datos.

En la tabla 38 se describen los atributos del repositorio que presentaron un porcentaje alto de valores nulos y en la tabla 39 se muestran los atributos con valores constantes y los atributos identificadores.

Tabla 38
Atributos con porcentaje alto de datos nulos

Atributo	Nulos (%)
estu_disc_sordo_ceguera, econ_area_vive, fami_sn_lee_escribe_padre, fami_sn_lee_escribe_madre, fami_num_hermanos, fami_nivel_hermano, econ_cuartos, econ_sn_televisor, econ_sn_motocicleta, econ_sn_energia, econ_sn_acueducto, econ_sn_alcantarillado, econ_sn_aseo, econ_sn_estufa, estu_otro_idioma_lee, estu_otro_idioma_habla, estu_nivel_postgrado, estu_etnia	100
estu_disc_invidente, estu_disc_sordo_con_interprete, estu_disc_sordo_sin_interprete, estu_zona	99,99
estu_disc motriz	99,92
estu_anno_egreso	97,77

Nota. Elaboración propia.

Tabla 39

Atributos con valores constantes y atributos identificadores

Atributo	Valor
estu_consecutivo	SBPRO+identificador
estu_cod_aplicacion	EK20112
estu_pais_reside	Colombia
estu_exam_nombre	Examen específico 2011-2
estu_exam_cod	548
estu_nivel_prgm_academico	Universitario

Nota. Elaboración propia.

Se efectuó una primera selección de atributos y se descartaron aquellos que presentaron un alto porcentaje de valores nulos dada la imposibilidad de encontrar sus valores a través de fuentes externas de datos (véase tabla 12). De igual manera, se descartaron los atributos con valores constantes y aquellos que solo servían de identificadores de cada estudiante (véase tabla 13).

Se sabe que la alta dimensionalidad es un problema para el descubrimiento de patrones con minería de datos (Hernández et al., 2005). Uno de los criterios utilizados para resolver este problema es la reducción del número de atributos por analizar, a través de su transformación en nuevos atributos que generalicen los datos y que ofrezcan mayor información. Teniendo en cuenta este criterio, en el repositorio de datos T97068A94 se seleccionaron aquellos atributos que por sí mismos no tenían mayor significado, pero que al integrarlos en uno nuevo, adquirirían mayor semántica. En este sentido, se crearon nuevos atributos y se redujo el número de variables para la investigación. En la tabla 40 se describe el proceso de construcción de nuevos atributos y se muestran aquellos que se reemplazan. Los atributos reemplazados fueron eliminados del repositorio.

Tabla 40

Nuevos atributos del repositorio de datos

Atributo	Descripción	Acción	Valores
estu_nacimiento_fecha	Fecha de nacimiento del estudiante.	Reemplaza a los atributos: estu_nacimiento_día estu_nacimiento_mes estu_nacimiento_año	Valores con formato DD/MM/YYYY

(continúa)

(viene)

Atributo	Descripción	Acción	Valores
estu_edad	Edad del estudiante en el momento de presentar la prueba.	Reemplaza al atributo: estu_nacimiento_fecha	Valores numéricos
estu_discapacidad	Tipo de discapacidad del estudiante.	Reemplaza a los atributos: estu_disc_invidente estu_disc_sordo_con_interprete estu_disc_sordo_sin_interprete estu_disc_motriz estu_disc_sordo_ceguera	I= invidente R= sordo con interprete S= sordo sin interprete M=discapacidad motriz C= sordo ceguera N = sin discapacidad
estu_reside_coddpto	Departamento de residencia del estudiante.	Reemplaza al atributo: estu_reside_codmpio	Códigos DANE
estu_pers_cargo	Si el estudiante tiene personas a cargo.	Reemplaza al atributo: fami_num_pers_cargo	Sí No
estu_financiacion_matricula	Forma de financiar el pago de la matrícula.	Reemplaza a los atributos: estu_sn_matricula_propio estu_sn_matricula_padres estu_sn_matricula_beca estu_sn_matricula_credito	Combinaciones en código binario de 4 Bits
fami_nivel_educacion_padres	Máximo nivel educativo completo entre el padre y la madre.	Reemplaza a los atributos: fami_cod_educacion_padre fami_cod_educacion_madre	Primaria secundaria Técnico/ tecnológico Profesional Postgrado Ninguno
fami_ocupacion_padre	Ocupación del padre.	Reemplaza al atributo: fami_cod_ocupacion_padre	Directivo Empleado Empresario Hogar Independiente Otra Pensionado Profesional
fami_ocupacion_madre	Ocupación de la madre.	Reemplaza al atributo: fami_cod_ocupacion_madre	Los mismos valores de la ocupación del padre

(continúa)

(viene)

Atributo	Descripción	Acción	Valores
inst_tipo	Tipo de institución del estudiante.	Reemplaza a los atributos: inst_origen	Oficial Privada Régimen especial
inst_acreditada	Determina si la institución a la que pertenece el estudiante es acreditada o no según CNA.	Reemplaza a los atributos: inst_cód_institución inst_nombre_institucion	Acreditada No acreditada
inst_prog_acreditado	Determina si el programa al cual pertenece el estudiante es acreditado o no según CNA.	Reemplaza a los atributos: estu_prac_id_progr_ académico estu_prog_académico	Acreditado No acreditado
area_grupo_referencia	Áreas de los grupos de referencia de los programas.	Reemplaza al atributo: estu_grupo_referencia	Ciencias humanas Ciencias sociales Ciencias naturales y técnicas
inst_cod_dpto	Código DANE del departamento donde se ofrece el programa.	Reemplaza al atributo: dipo_cod_municipio	Códigos departamentos según DANE
inst_dpto_programa	Nombre del departamento donde se ofrece el programa académico.	Reemplaza al atributo: inst_cod_dpto	Nombres departamentos de Colombia según DANE
inst_programa_zona	Zona geográfica donde se ofrece el programa.	Reemplaza al atributo: inst_dpto_programa	Bogotá Eje Cafetero Caribe Centro Oriente Pacífico Centro Sur Llano
num_estudiantes_zona	Número de estudiantes por zona geográfica.	Nuevo atributo generado a partir del atributo: inst_programa_zona	Alto Medio Bajo
num_instituciones_zona	Número de IES por zona geográfica.	Nuevo atributo generado a partir del atributo: inst_programa_zona	Alto Medio Bajo

(continúa)

(viene)

Atributo	Descripción	Acción	Valores
eco_condicion_vivienda	Condición de la vivienda del estudiante	Reemplaza al atributo: econ_material_pisos	Buena Mala Regular
eco_condicion_hogar	Condición del hogar del estudiante	Reemplaza a los atributos: econ_sn_dvd econ_sn_lavadora econ_sn_microondas econ_sn_horno econ_sn_nevera	Buena Regular Mala
eco_condicion_transporte	Condición de transporte en el hogar del estudiante	Reemplaza al atributo: econ_sn_automovil	Particular Público
eco_condicion_tic	Condición de uso de TIC en el hogar del estudiante	Reemplaza a los atributos: econ_sn_telefonia econ_sn_celular econ_sn_internet econ_sn_servicio_tv econ_sn_computador	Buena Regular Mala
eco_condicion_vive	Condición de vida del estudiante	Reemplaza a los atributos: infa_dormitorios fami_num_pers_fam	Sin hacinamiento Hacinamiento medio Hacinamiento crítico
mod_lectura_critica_desemp	Desempeño del estudiante en el módulo de lectura crítica.	Reemplaza al atributo: mod_lectura_critica	Sobre la media (>= la media) Bajo la media (< la media)
mod_comunica_escrita_desemp	Desempeño del estudiante en el módulo de comunicación escrita.	Reemplaza a los atributos: mod_comunica_escrita_punt mod_comunica_escrita_desem	Sobre la media (>= la media) Bajo la media (< la media)
mod_razona_cuantitativo_desemp	Desempeño del estudiante en el módulo de razonamiento cuantitativo.	Reemplaza al atributo: mod_razona_cuantitativo_punt	Sobre la media (>= la media) Bajo la media (< la media)
mod_ingles_desemp	Desempeño del estudiante en el módulo de inglés	Reemplaza a los atributos: mod_ingles_punt mod_ingles_desem	Sobre la media (>= la media) Bajo la media (< la media)

Nota. Elaboración propia.

Con el fin de facilitar la detección de patrones de rendimiento académico, se discretizaron los valores numéricos de ciertos atributos, para lo cual se tuvo en cuenta un rango de valores y la proporcionalidad de las frecuencias por cada valor, a fin de evitar sesgos en la construcción de los modelos de minería de datos. En la tabla 41 se detalla este proceso para el atributo `estu_edad`.

Tabla 41
Valores discretizados del atributo `estu_edad_examen`

Valor	Número de estudiantes
Edad hasta 21	17 785
Edad 22	15 602
Edad 23	12 133
Edad de 25 a 26	12 022
Edad de 27 a 29	10 766
Edad de 30 a 34	9915
Edad mayor igual que 35	9863
Edad 24	8900
Sin dato	82
Total	97 068

Nota. Elaboración propia.

En esta investigación se consideró conveniente crear un nuevo atributo denominado `inst_programa_zona`, con el cual se reorganizarían las IES en zonas geográficas, de manera que la información suministrada aportara al descubrimiento de patrones de desempeño académico, ya que el atributo `estu_zona` de la base de datos del ICFES (véase tabla 1), que ubica la zona cardinal de la ciudad de residencia del estudiante, no aporta información relevante para este estudio.

Durante el proceso se consideraron varias opciones para la clasificación de dichas zonas, entre las cuales están la reorganización por zonas geográficas como tal, zonas político-administrativas, zonas culturales y zonas de los Órganos Colegiados de Administración y Decisión (OCAD); estas últimas son creadas y utilizadas en el marco del Sistema General de Regalías (SGR), que busca beneficiar a todos los departamentos y municipios del país, no únicamente a aquellos en los cuales se lleva a cabo la explotación de recursos naturales no renovables; por esta razón, se constituyen en una propuesta que se fundamenta en argumentos

de justicia social y que tiene la finalidad de distribuir de manera equitativa los recursos de regalías en todas las zonas del país. Todos los recursos del SGR financian proyectos de inversión presentados a los OCAD por las entidades territoriales, según lo reglamentado en el Decreto 1075 del 2012. Los OCAD se constituyen a nivel municipal, departamental, regional, nacional; además, está el OCAD de Ciencia Tecnología e Innovación, al cual pertenecen las universidades.

Estas zonas OCAD, en la actualidad, constituyen un reordenamiento territorial del país frente a las regiones geográficas que ha manejado históricamente y que ha permitido marcar diferencias significativas entre el centro y la periferia de Colombia. La ventaja de esta organización es que cada zona OCAD agrupa departamentos completos, con sus respectivos municipios. Gracias a esto es posible predecir cuál sería el rendimiento académico en las competencias genéricas en estas zonas con base en la información de las pruebas Saber Pro y, en futuros estudios, se puede analizar cómo han influido las regalías en el desarrollo de las IES y en el rendimiento académico de los estudiantes de estas zonas.

Los OCAD se encuentran distribuidos en seis zonas que cubren la totalidad del país. Teniendo en cuenta que Bogotá es la ciudad donde se concentran la gran mayoría de universidades, para esta investigación se consideró conveniente tomarla como una nueva zona. Así pues, para este estudio las zonas geográficas en las que se agrupan las IES son siete, como se detallan en la tabla 42.

Tabla 42
Valores del atributo *inst_programa_zona*

Zonas	Departamentos
Caribe	Atlántico, Bolívar, Cesar, Córdoba, Guajira, Magdalena, San Andrés, Providencia y Santa Catalina y Sucre.
Centro Sur	Amazonas, Caquetá, Huila, Putumayo y Tolima.
Centro Oriente	Boyacá, Cundinamarca, Norte de Santander y Santander.
Eje Cafetero	Antioquia, Caldas, Risaralda y Quindío.
Llano	Arauca, Casanare, Guainía, Guaviare, Meta, Vaupés y Vichada.
Pacífico	Cauca, Chocó, Nariño y Valle del Cauca.
Bogotá	Distrito Capital de Bogotá

Nota. Adaptado de Colombia. Departamento Nacional de Planeación. Sistema General de Regalías SGR. (2012). Bogotá: DNP.

Clasificadas las IES por zonas, se procedió a contar el número de estudiantes e IES por zonas y a crear dos nuevos atributos, a saber: num_estudiantes_zona y num_instituciones_zona, los cuales fueron discretizados.

Para los valores del atributo num_estudiantes_zona, se considera alto si el número de estudiantes es mayor que 20 000; medio si el número de estudiantes está entre 10 000 y 20 000, y bajo si es menor que 10 000. En la tabla 43 se muestran los datos en referencia.

Tabla 43
Valores discretizados del atributo num_estudiantes_zona

Zona	N.º de estudiantes	Valor
Bogotá	33 544	Alto
Eje Cafetero	16 500	Medio
Caribe	15 312	Medio
Centro Oriente	13 519	Medio
Pacífico	11 766	Medio
Centro Sur	5283	Bajo
Llano	1144	Bajo
Total de estudiantes	97 068	

Nota. Elaboración propia.

Para los valores del atributo num_instituciones_zona, se considera alto si el número de IES es mayor que 70; medio si el número de IES está entre 60 y 70, y bajo si es menor que 60. En la tabla 44 se muestran los resultados de tal discretización.

Tabla 44
Valores discretizados del atributo num_instituciones_zona

Zona	Número de Instituciones	Valor
Bogotá	73	Alto
Eje Cafetero	78	Alto
Caribe	64	Medio
Centro oriente	56	Bajo
Pacífico	64	Medio
Centro sur	21	Bajo
Llano	12	Bajo
	368	

Nota. Elaboración propia.

Para la nueva variable `eco_condicion_hogar` se tuvieron en cuenta los materiales de los pisos de la vivienda. En la tabla 45 se muestran estos valores y el material de los pisos.

Tabla 45

Valores del atributo `eco_condicion_hogar`

Material de los Pisos de la Vivienda	Condición Vivienda
Tierra, arena	Mala
Cemento, gravilla, ladrillo	Mala
Madera burda, tabla – tablón	Regular
Madera pulida, baldosa, tableta, mármol, alfombra	Buena
Madera pulida, mármol, alfombra - tapete de pared a pared	Buena

Nota. Elaboración propia.

Para asignar los valores a la nueva variable `eco_condicion_hogar` se creó un índice con los valores de los atributos de los electrodomésticos que esta variable reemplazó. El índice es el resultado de la sumatoria de los valores de la presencia (1) o ausencia (0) de cada electrodoméstico en la vivienda. Si el índice está entre 4 y 5, se asume que la condición del hogar es buena; si está entre 2 y 3 la condición del hogar es media, y si está entre 0 y 1 la condición del hogar es mala. En la tabla 46 se muestran las variables que intervienen en el cálculo del índice `condicion_hogar`.

Tabla 46

Cálculo del índice de `condicion_hogar`

Electrodomésticos	Sí	No
<code>econ_sn_dvd</code>	1	0
<code>econ_sn_lavadora</code>	1	0
<code>econ_sn_microondas</code>	1	0
<code>econ_sn_horno</code>	1	0
<code>econ_sn_nevera</code>	1	0

Nota. Elaboración propia.

De manera semejante se procedió para la discretización de los valores de la nueva variable `eco_condicion_tic` para que tome los valores buena, media y mala, con base en los servicios que dispone el estudiante. En la tabla 47 se muestran los atributos que intervienen en el cálculo del índice de `condicion_tic`.

Tabla 47
Cálculo del índice de condición_tic

Servicios	Sí	No
econ_sn_internet	1	0
econ_sn_servicio_tv	1	0
econ_sn_telefonia	1	0
econ_sn_celular	1	0
econ_sn_computador	1	0

Nota. Elaboración propia.

Para los valores de la variable *eco_condicion_transporte* se consideraron dos opciones: particular si el estudiante dispone de automóvil y público en caso contrario, para lo cual se tiene en cuenta la variable *econ_sn_automovil* que reemplaza.

Para asignar los valores a la nueva variable *eco_condicion_vive* se calculó el índice de hacinamiento. El hacinamiento refiere la relación entre el número de personas que habitan una vivienda o casa y el espacio o número de cuartos disponibles (Spicker, Alvarez y Gordon, s.f.).

Generalmente se aceptan los siguientes valores: sin hacinamiento, hasta 2,4; hacinamiento medio, de 2,5 a 4,9, y hacinamiento crítico, de 5,0 o más.

Teniendo en cuenta estos conceptos, el índice de hacinamiento para cada estudiante se obtiene dividiendo los valores de los atributos *fami_num_pers_fam* entre *infa_dormitorios*, variables que fueron reemplazadas por el atributo *eco_condicion_vive*. Los valores de este nuevo atributo se asignaron teniendo en cuenta los valores aceptados para el hacinamiento.

Para obtener los valores de la nueva variable *area_grupo_referencia*, se agruparon los valores del atributo *estu_grupo_referencia* en ciencias humanas, ciencias sociales y ciencias naturales y técnicas. Los datos se muestran en la tabla 48.

Tabla 48
Asignación de valores del atributo área_grupo_referencia

Grupo de referencia programas	Área grupo referencia programas
Ciencias económicas y administrativas	Ciencias sociales
Ingeniería	Ciencias naturales y técnicas
Educación	Ciencias sociales
Derecho	Ciencias humanas

(continúa)

(viene)

Grupo de referencia programas	Área grupo referencia programas
Salud	Ciencias humanas
Ciencias sociales	Ciencias sociales
Comunicación, periodismo y publicidad	Ciencias sociales
Medicina	Ciencias humanas
Bellas artes y diseño	Ciencias humanas
Ciencias agropecuarias	Ciencias naturales y técnicas
Ciencias naturales y exactas	Ciencias naturales y técnicas
Arquitectura y urbanismo	Ciencias sociales
Humanidades	Ciencias humanas
Militar y policial	Ciencias sociales
Recreación y deportes	Ciencias humanas
Ciencias militares y navales	Ciencias naturales y técnicas
Artes-diseño-comunicación	Ciencias sociales

Nota. Elaboración propia.

Como resultado de todos los procesos descritos, se obtuvo un nuevo repositorio de datos denominado T97068A35, limpio y transformado, listo para aplicarle las técnicas de minería de datos. La descripción del repositorio T97068A35 con 97068 registros y 35 atributos se muestra en la tabla 49, y están organizados por los factores sociodemográficos, económicos, institucionales y clases.

Tabla 49
Diccionario de datos del repositorio final T97068A35

N.º	Atributo	Descripción	Valores
Sociodemográficos			
1	estu_genero	Género	M, F
2	estu_edad	Edad del estudiante en el momento de presentar la prueba	Ver tabla 15
3	estado_civil	Estado civil del estudiante	Soltero, casado Separado/divorciado, Unión libre, viudo
4	estu_hogar_actual	Tipo de hogar actual donde reside el estudiante	Habitual/permanente Temporal
5	estu_sn_cabeza_fmliia	Si el estudiante es cabeza de familia o no	Sí, no

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
6	estu_pers_cargo	Si el estudiante tiene personas a cargo o no	Sí, no
7	fami_nivel_educacion_padres	Máximo nivel educativo completo entre el padre y la madre	Primaria, secundaria Técnico/tecnológico Profesional, postgrado Ninguno
8	fami_ocup_padre	Ocupación del padre	Directivo, empleado Empresario, hogar Independiente, otra Pensionado, profesional
9	fami_ocup_madre	Ocupación de la madre	Los mismos valores de la ocupación del padre
Económicos			
10	estu_financiacion_matricula	Forma de financiar el pago de la matrícula	Propios, crédito, padres Beca y todas las combinaciones posibles entre estos valores
11	estu_estrato	Estrato socioeconómico del estudiante	Estratos 1 a 6 Zona rural sin estrato
12	fami_nivel_sisben	Nivel de clasificación en el Sisben al que pertenece el estudiante	Niveles 1, 2, 3, Otro nivel, No está en Sisben
13	econ_condicion_vivienda	Condición de la vivienda del estudiante	Buena, mala, regular
14	eco_condicion_hogar	Condición del hogar del estudiante	Buena, mala, regular
15	eco_condicion_transporte	Condición de transporte en el hogar del estudiante	Particular, público
16	eco_condicion_tic	Condición de uso de TIC en el hogar del estudiante	Buena, regular, mala
17	eco_condicion_vive	Condición de vida del estudiante	Sin hacinamiento Hacinamiento medio Hacinamiento crítico
18	fami_ing_fmiliar_mensual	Ingresos mensuales familiares en salarios mínimos	Ver tabla 33
19	estu_trabaja	Si estudiante trabaja o no	Ver tabla 34
Académicos			
20	estu_metodo_prm	Metodología del programa académico bajo la cual cursa el estudiante	A distancia Presencial

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
21	estu_area_conoc	Nombre del área de conocimiento a la que pertenece el programa académico del estudiante	Ver tabla 22
22	area_grupo_referencia	Área del grupo de referencia de los programas	Ciencias humanas Ciencias sociales Ciencias naturales y técnicas
23	estu_pje_creditos	Porcentaje de créditos cursados por el estudiante al realizar la prueba	Más del 90 Entre 81 y 90 Entre el 75 y el 80 Menos del 75 No sigue el sistema de créditos
24	estu_titulo_bto	Tipo de título de bachillerato obtenido	Académico, normalista Técnico
Institucionales			
25	inst_tipo	Tipo de institución del estudiante	Oficial, privada Régimen especial
26	inst_caracter_academico	Carácter académico de la IES	Escuela tecnológica Institución tecnológica Institución universitaria Técnica profesional Universidad
27	inst_acreditada	Si la institución donde pertenece el estudiante es acreditada o no según CNA	Acreditada, No acreditada
28	inst_prog_acreditado	Si el programa al cual pertenece el estudiante es acreditado o no según CNA	Acreditado, No acreditado
29	inst_programa_zona	Zona OCAD donde se ofrece el programa	Bogotá, eje cafetero Caribe, centro oriente Pacífico, centro sur Llano
30	num_estudiantes_zona	Número de estudiantes por zona OCAD	Alto, medio, bajo
31	num_instituciones_zona	Número de IES por zona OCAD	Alto, medio, bajo
Clases			
32	mod_lectura_critica_desemp	Desempeño del estudiante en el módulo de lectura crítica	Sobre la media (>= la media) Bajo la media (< la media)

(continúa)

(viene)

N.º	Atributo	Descripción	Valores
33	mod_comunica_ escrita_desemp	Desempeño del estudiante en el módulo de comunicación escrita	Sobre la media (>= la media) Bajo la media (< la media)
34	mod_razona_ cuantitativo_desemp	Desempeño del estudiante en el módulo de razonamiento cuantitativo	Sobre la media (>= la media) Bajo la media (< la media)
35	mod_ingles_desemp	Desempeño del estudiante en el módulo de inglés	Sobre la media (>= la media) Bajo la media (< la media)

Nota. Elaboración propia.

A partir del repositorio T97068A34 y con el fin de descubrir patrones asociados al rendimiento académico en competencias genéricas, se construyó por cada competencia (clase) un repositorio de datos. Para cada repositorio se tomaron los 31 atributos descritos en la tabla 49 y el respectivo atributo clase, dependiendo de la competencia. En cada repositorio se eliminaron los registros con valores nulos en el atributo clase (competencia). En la tabla 50 se presenta la relación de repositorios y competencias.

Tabla 50
Repositorios por cada competencia

Repositorio	Descripción
T97055A32LEC	Repositorio para análisis de la competencia de lectura crítica. Contiene 97055 registros y 32 atributos.
T95337A32ESC	Repositorio para análisis de la competencia de comunicación escrita. Contiene 95337 registros y 32 atributos.
T97057A32CUA	Repositorio para análisis de la competencia de razonamiento cuantitativo. Contiene 97057 registros y 32 atributos.
T96946A32ING	Repositorio para análisis de la competencia de inglés. Contiene 96946 registros y 32 atributos.

Nota. Elaboración propia.

Modelado

En esta fase se seleccionó la tarea de clasificación con árboles de decisión, como la técnica de minería de datos más adecuada para solucionar el problema objeto de la investigación. Con esta tarea se pretende obtener un modelo que permita predecir para los nuevos casos de estudiantes de programas profesionales los factores socio-demográficos, económicos, académicos e institucionales asociados a un probable buen o mal desempeño académico en las cuatro competencias genéricas, evaluadas en las pruebas Saber Pro 2011-2.

Descubrimiento de patrones de desempeño académico con árboles de decisión

La técnica de clasificación que se utilizó para el descubrimiento de patrones de desempeño académico en las competencias genéricas de las pruebas Saber Pro 2011-2 fue árboles de decisión. El modelo de clasificación basado en árboles de decisión es probablemente el más utilizado y popular por su simplicidad y facilidad para entender (Han y Kamber, 2001), (Sattler y Dunemann, 2001), (Timarán y Millán, 2006). La importancia de los árboles de decisión se debe a su capacidad de construir modelos interpretables, siendo este un factor decisivo para su aplicación. La clasificación con árboles de decisión considera clases disjuntas, de forma que el árbol conducirá a una y solo una hoja, asignando una única clase a la predicción (Hernández y Lorente, 2009).

El algoritmo de la herramienta Weka (Hall, Frank y Witten, 2011) utilizado para obtener el modelo de clasificación con árboles de decisión fue J48, el cual implementa al algoritmo C.45 (Quinlan, 1993). El algoritmo J48 se basa en la utilización del criterio del coeficiente de ganancia de información (*information gain ratio*). De esta manera, se consigue evitar que las variables con mayor número de posibles valores salgan beneficiadas en la selección. Además, el algoritmo incorpora una poda del árbol de clasificación una vez que este ha sido inducido (Hernández y Lorente, 2009). El parámetro más importante que se debe tener en cuenta para la poda es el factor de confianza *C* (*confidence level*), que influye en el tamaño y capacidad de predicción del árbol construido. Cuanto más baja se haga esa probabilidad, más se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor por defecto de este factor es del 25%, y conforme va bajando este valor se permiten más operaciones

de poda; por lo tanto, se puede llegar a árboles cada vez más pequeños (García y Álvarez, 2010). Otra forma de variar el tamaño del árbol es a través del parámetro M que especifica el mínimo número de instancias o registros por nodo del árbol; es menos importante puesto que depende del número absoluto de instancias en el conjunto de datos de partida (Hall, Frank y Witten, 2011).

Antes de construir un modelo se debe definir un procedimiento para probar la calidad del modelo y su validez. Por tanto, para entrenar y probar un modelo de clasificación, el diseño de prueba específica divide los datos en dos conjuntos: entrenamiento y prueba. Existen diferentes medidas de evaluación del clasificador en Weka (Hall, Frank y Witten, 2011):

- Usar el conjunto de datos de entrenamiento (*Use training set*): se emplea todo el conjunto de datos para entrenar el modelo y después se prueba (esta técnica puede ser muy buena para ese conjunto de datos, pero puede ser poco precisa para nuevos datos).
- Proveer un conjunto de datos de prueba (*Supplied test set*): se emplea un conjunto de datos para entrenar y otro conjunto independiente al universo de los datos con los que se está trabajando para prueba (se corre el riesgo de que el conjunto de prueba no refleje o se corresponda con las características de los datos que se emplean para entrenar el modelo).
- Porcentaje de Partición (*Percentage Split*): se emplea un porcentaje aleatorio de datos para entrenar y otro porcentaje para probar; este método difiere del anterior por cuanto ambos conjuntos pertenecen al universo de datos con el que se está trabajando, por lo cual se elimina el riesgo que corre el anterior.
- Validación cruzada (*Cross validation*): este mecanismo permite reducir la dependencia del resultado del experimento en el modo como se realiza la partición (Hernández, Ramírez y Ferri, 2005). Para este caso, se utiliza el método de evaluación validación cruzada con n pliegues (*n-fold cross validation*). Esta es la opción por defecto y la más comúnmente utilizada. Este método consiste en dividir el conjunto de entrenamiento en n subconjuntos disjuntos de similar tamaño, llamados pliegues (*folds*) de forma aleatoria. El número de subconjuntos se puede introducir en el campo *Folds*. Posteriormente se realizan n iteraciones (igual al número de subconjuntos definido), donde, en cada una se reserva un subconjunto diferente para el conjunto de prueba y los restantes $n-1$ (uniendo todos los datos) para construir el modelo (entrenamiento). En cada iteración se calcula el error de muestra parcial del modelo. Por último,

se construye el modelo con todos los datos y se obtiene su error promediando los obtenidos anteriormente en cada una de las iteraciones. Otra ventaja de la validación cruzada es que la varianza de los n errores de muestra parciales permite estimar la variabilidad del método de aprendizaje con respecto al conjunto de datos. Comúnmente, se suelen utilizar 10 particiones (*10-fold cross validation*) (Hernández, Ramírez y Ferri, 2005).

Por otra parte, si se dispone de la matriz de confusión, es bastante sencillo evaluar o estimar el coste de un clasificador para un determinado conjunto de ejemplos. La matriz de confusión (*Confusion Matrix*) representa de forma detallada el número de instancias que son predichas por clase. La suma de los registros que se representan en cada fila i , $i = 1 \dots n$ constituyen el número de instancias que realmente pertenecen a la clase i . Similarmente, la sumatoria de los ejemplos o registros en cada columna j , $j = 1 \dots n$ son las instancias que ha predicho el algoritmo al valor j de la clase. Los valores en la diagonal son los aciertos y el resto son los errores de clasificación (ejemplos que pertenecían a la clase i de la fila i y que fueron clasificados incorrectamente en otra) (Fernández, 2009).

Teniendo en cuenta los parámetros de evaluación anteriores y los repositorios de datos descritos en la tabla 43, se procedió a construir los diferentes árboles de decisión con el algoritmo J48. Se escogió como clase el desempeño en la competencia respectiva, y para evaluar la calidad del modelo y su validez, el método de validación cruzada, específicamente la validación cruzada con 10 pliegues por los mejores resultados que se obtienen.

Con el fin de obtener diferentes modelos de árboles por competencia y reglas de clasificación generalizadas, hasta reglas más detalladas, se establecieron cuatro porcentajes de prepoda del árbol para el factor M igual a 10%, 5%, 1% y 0,5% del total de registros del repositorio de datos, manteniendo constante el factor confianza C en el 25%. Por los mejores resultados obtenidos y por la facilidad de análisis de los patrones, se escogió el árbol construido con los parámetros $M=1\%$ y $C=25\%$. Una vez construidos los árboles, se aplicó un proceso de pospoda para dejar las ramas y, por ende, las reglas más representativas, que son aquellas que sobrepasan un mínimo soporte del 0,5% y una confianza del 60%.

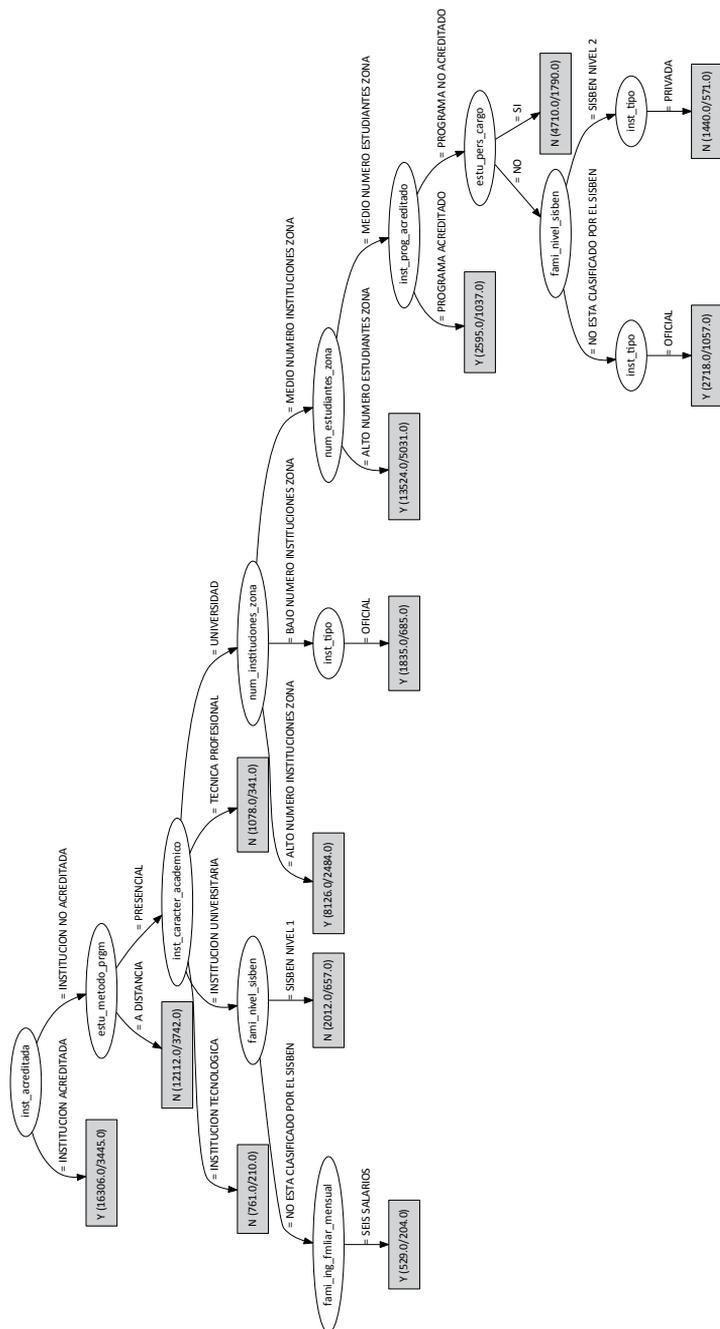


Figura 10. Árbol para lectura crítica

Nota. Elaboración propia.

```

==== 10 Fold Cross Validation ====
==== Summary ====
Correctly Classified Instances          62781          64.686 %
Incorrectly Classified Instances       34274          35.314 %
Kappa statistic                        0.2831
Mean absolute error                    0.442
Root mean squared error                0.4704
Relative absolute error                89.1798 %
Root relative squared error            94.4938 %
Coverage of cases (0.95 level)        99.999 %
Mean rel. region size (0.95 level)    99.9995 %
Total Number of Instances              97055

==== Confusion Matrix ====
a          b <-- classified as
37498     15548 | a = Y
18726     25283 | b = N
    
```

Figura 11. Precisión y matriz de confusión del árbol de lectura crítica

Nota. Elaboración propia.

Descubrimiento de patrones de desempeño en lectura crítica

Para la construcción del árbol de decisión para el descubrimiento de patrones de desempeño en la competencia de lectura crítica en las pruebas Saber Pro 2011-2, se utilizó el conjunto de datos T97055A32LEC. El árbol construido con los parámetros $M=971$ (1%) y $C=0,25$ para la prepoda y confianza mayor o igual al 60% y soporte mayor o igual al 0,5%, se muestra en la figura 10. En la figura 11 se puede ver la precisión del árbol y su matriz de confusión, y en la figura 12 están las reglas más representativas con un soporte $\geq 0,5\%$ y confianza $\geq 60\%$.

Descubrimiento de patrones de desempeño en comunicación escrita

Para la construcción del árbol de decisión para el descubrimiento de patrones de desempeño en la competencia de comunicación escrita, en las pruebas Saber Pro 2011-2, se utilizó el conjunto de datos T95337A32ESC. El árbol construido con los parámetros $M=954$ (1%) y $C=0,25$ para la prepoda y confianza mayor o igual al 60% y soporte mayor o igual al 0,5% se muestra en la figura 13. En la figura 14 puede verse la precisión del árbol y su matriz de confusión, y en la figura 15 se presentan las reglas más representativas con un soporte $\geq 0,5\%$ y confianza $\geq 60\%$.

Antecedente	Consecuente	Confianza (%)	Soporte (%)
1 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = INSTITUCION TECNOLOGICA	N	72.40	0.78
2 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = A DISTANCIA	N	69.11	12.48
3 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = TECNICA PROFESIONAL	N	68.37	1.11
4 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = INSTITUCION UNIVERSITARIA y fami_nivel_sisben = SISBEN NIVEL 1	N	67.35	2.07
5 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = MEDIO NUMERO INSTITUCIONES ZONA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y inst_prog_acreditado = PROGRAMA NO ACREDITADO y estu_pers_cargo = SI	N	62.00	4.85
6 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = MEDIO NUMERO INSTITUCIONES ZONA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y inst_prog_acreditado = PROGRAMA NO ACREDITADO y estu_pers_cargo = NO y fami_nivel_sisben = SISBEN NIVEL 2 y inst_tipo = PRIVADA	N	60.35	1.48
7 inst_acreditada = INSTITUCION ACREDITADA	Y	78.87	16.80
8 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = ALTO NUMERO INSTITUCIONES ZONA	Y	69.43	8.37
9 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = MEDIO NUMERO INSTITUCIONES ZONA y num_estudiantes_zona = ALTO NUMERO ESTUDIANTES ZONA	Y	62.80	13.93
10 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = BAJO NUMERO INSTITUCIONES ZONA y inst_tipo = OFICIAL	Y	62.67	1.89
11 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = INSTITUCION UNIVERSITARIA y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y fami_ing_fmliar_mensual = SEIS SALARIOS	Y	61.44	0.55
12 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = MEDIO NUMERO INSTITUCIONES ZONA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y inst_prog_acreditado = PROGRAMA NO ACREDITADO y estu_pers_cargo = NO y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y inst_tipo = OFICIAL	Y	61.11	2.80
13 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y inst_caracter_academico = UNIVERSIDAD y num_instituciones_zona = MEDIO NUMERO INSTITUCIONES ZONA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y inst_prog_acreditado = PROGRAMA ACREDITADO	Y	60.04	2.67

Figura 12. Reglas más representativas de lectura crítica

Nota. Elaboración propia.

Descubrimiento de patrones de desempeño en razonamiento cuantitativo

Para la construcción del árbol de decisión para el descubrimiento de patrones de desempeño en la competencia de razonamiento cuantitativo en las pruebas Saber Pro 2011-2, se utilizó el conjunto de datos T97057A32CUA. El árbol construido con los parámetros $M=971$ (1%) y $C=0,25$ para la prepoda y confianza mayor o igual al 60% y soporte mayor o igual al 0,5% se muestra en la figura 16. En la figura 17 está la precisión del árbol y su matriz de confusión, y en la figura 18 se ven las reglas más representativas con un soporte $\geq 0,5\%$ y confianza $\geq 60\%$.

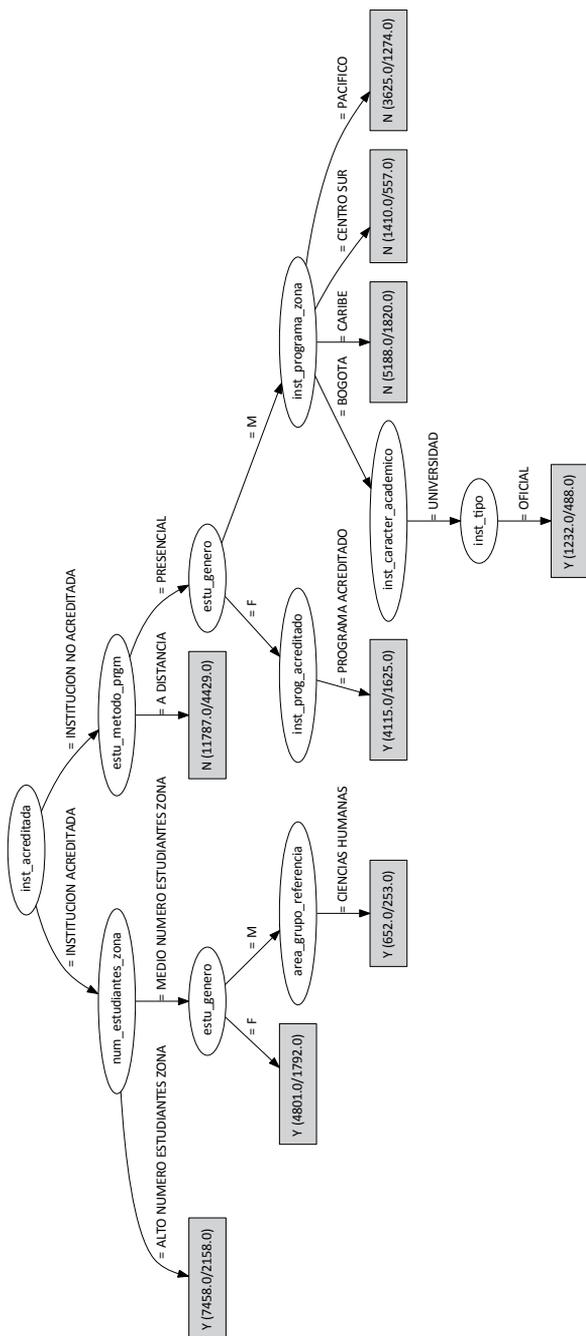


Figura 13. Árbol para comunicación escrita

Nota. Elaboración propia.

```

==== 10 Fold Cross Validation ====
==== Summary ====
Correctly Classified Instances      56669      59.4407 %
Incorrectly Classified Instances    38668      40.5593 %
Kappa statistic                    0.1876
Mean absolute error                0.4762
Root mean squared error            0.4882
Relative absolute error            95.2592 %
Root relative squared error        97.6493 %
Coverage of cases (0.95 level)    100 %
Mean rel. region size (0.95 level) 99.9969 %
Total Number of Instances          95337

==== Confusion Matrix ====
a      b <-- classified as
31234  17117 | a = N
21551  25435 | b = Y
    
```

Figura 14. Precisión y matriz de confusión del árbol de comunicación escrita

Nota. Elaboración propia.

Antecedente	Consecuente	Confianza (%)	Soporte (%)
1 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y estu_genero = M y inst_programa_zona = CARIBE	N	64.92	5.44
2 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y estu_genero = M y inst_programa_zona = PACIFICO	N	64.86	3.80
3 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = A DISTANCIA	N	62.42	12.36
4 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y estu_genero = M y inst_programa_zona = CENTRO SUR	N	60.50	1.48
5 inst_acreditada = INSTITUCION ACREDITADA y num_estudiantes_zona = ALTO NUMERO ESTUDIANTES ZONA	Y	71.06	7.82
6 inst_acreditada = INSTITUCION ACREDITADA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y estu_genero = F	Y	62.67	5.04
7 inst_acreditada = INSTITUCION ACREDITADA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y estu_genero = M y area_grupo_referencia = CIENCIAS HUMANAS	Y	61.20	0.68
8 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y estu_genero = F y inst_prog_acreditado = PROGRAMA ACREDITADO	Y	60.51	4.32
9 inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y estu_genero = M y inst_programa_zona = BOGOTA y inst_caracter_academico = UNIVERSIDAD y inst_tipo = OFICIAL	Y	60.39	1.29

Figura 15. Reglas más representativas de comunicación escrita

Nota. Elaboración propia.


```

==== 10 Fold Cross Validation ====
==== Summary ====
Correctly Classified Instances      64549      66.5063 %
Incorrectly Classified Instances    32508      33.4937 %
Kappa statistic                    0.3295
Mean absolute error                 0.4273
Root mean squared error            0.4626
Relative absolute error            85.4605 %
Root relative squared error        92.5244 %
Coverage of cases (0.95 level)    100 %
Mean rel. region size (0.95 level) 99.9923 %
Total Number of Instances         97057

==== Confusion Matrix ====
a      b <-- classified as
29093  19051 | a = Y
13457  35456 | b = N
    
```

Figura 17. Precisión y matriz de confusión del árbol de razonamiento cuantitativo

Nota. Elaboración propia.

Descubrimiento de patrones de desempeño en inglés

Para la construcción del árbol de decisión para el descubrimiento de patrones de desempeño en la competencia de inglés en las pruebas Saber Pro 2011-2, se utilizó el conjunto de datos T96946A32ING. El árbol construido con los parámetros $M=970$ (1%) y $C=0,25$ para la prepoda y confianza mayor o igual al 60% y soporte mayor o igual al 0,5%, se muestra en la figura 19. En la figura 20 se presenta la precisión del árbol y su matriz de confusión, y en la figura 21 pueden observarse las reglas más representativas con un soporte $\geq 0,5\%$ y confianza $\geq 60\%$.

Antecedente	Consecuente	Confianza (%)	Soporte (%)
1 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = A DISTANCIA	N	73.88	10.23
2 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = PRESENCIAL y estu_area_conoc = ECONOMIA, ADMINISTRACION, CONTADURIA Y AFINES y estu_genero = F y estu_edad = EDAD MAYOR IGUAL A 35	N	72.50	0.84
3 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS HUMANAS y inst_tipo = PRIVADA y fami_ing_fmliar_mensual = UN SALARIO	N	71.12	0.52
4 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS HUMANAS y inst_tipo = PRIVADA y fami_ing_fmliar_mensual = DOS SALARIOS	N	70.87	3.69
5 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = PRESENCIAL y estu_area_conoc = CIENCIAS DE LA EDUCACION	N	70.66	7.24
6 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = PRESENCIAL y estu_area_conoc = CIENCIAS SOCIALES Y HUMANAS	N	68.35	6.13
7 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = CIENCIAS DE LA EDUCACION	N	66.85	1.14
8 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS HUMANAS y inst_tipo = PRIVADA y fami_ing_fmliar_mensual = TRES SALARIOS	N	66.54	4.29
9 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = PRESENCIAL y estu_area_conoc = ECONOMIA, ADMINISTRACION, CONTADURIA Y AFINES y estu_genero = F y estu_edad = EDAD DE 30 A 34	N	65.11	1.39
10 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = PRESENCIAL y estu_area_conoc = ECONOMIA, ADMINISTRACION, CONTADURIA Y AFINES y estu_genero = F y estu_edad = EDAD DE 27 A 29	N	62.28	1.61
11 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS HUMANAS y inst_tipo = PRIVADA y fami_ing_fmliar_mensual = CUATRO SALARIOS	N	60.03	3.92
12 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = INGENIERIA, ARQUITECTURA, URBANISMO Y AFINES	Y	90.05	4.21
13 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = MATEMATICAS Y CIENCIAS NATURALES	Y	87.52	0.83
14 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = CIENCIAS DE LA SALUD	Y	78.24	1.43
15 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS NATURALES Y TECNICAS y estu_metodo_prgm = PRESENCIAL y inst_prog_acreditado = PROGRAMA ACREDITADO	Y	78.24	2.22
16 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = ECONOMIA, ADMINISTRACION, CONTADURIA Y AFINES	Y	76.37	3.91
17 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS NATURALES Y TECNICAS y estu_metodo_prgm = PRESENCIAL y inst_prog_acreditado = PROGRAMA NO ACREDITADO y estu_genero = M	Y	70.52	9.26
18 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS HUMANAS y inst_tipo = OFICIAL y estu_pers_cargo = NO y num_instituciones.zona = ALTO NUMERO INSTITUCIONES ZONA	Y	69.92	1.13
19 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = CIENCIAS SOCIALES Y HUMANAS y num_estudiantes.zona = ALTO NUMERO ESTUDIANTES ZONA	Y	68.04	2.15
20 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS NATURALES Y TECNICAS y estu_metodo_prgm = PRESENCIAL y inst_prog_acreditado = PROGRAMA NO ACREDITADO y estu_genero = F y num_estudiantes.zona = ALTO NUMERO ESTUDIANTES ZONA	Y	65.97	1.77
21 inst_acreditada = INSTITUCION ACREDITADA y estu_area_conoc = BELLAS ARTES	Y	64.48	1.00
22 inst_acreditada = INSTITUCION NO ACREDITADA y area_grupo_referencia = CIENCIAS SOCIALES y estu_metodo_prgm = PRESENCIAL y estu_area_conoc = ECONOMIA, ADMINISTRACION, CONTADURIA Y AFINES y estu_genero = M y num_estudiantes.zona = ALTO NUMERO ESTUDIANTES ZONA	Y	63.46	2.03

Figura 18. Reglas más representativas de razonamiento cuantitativo

Nota. Elaboración propia.

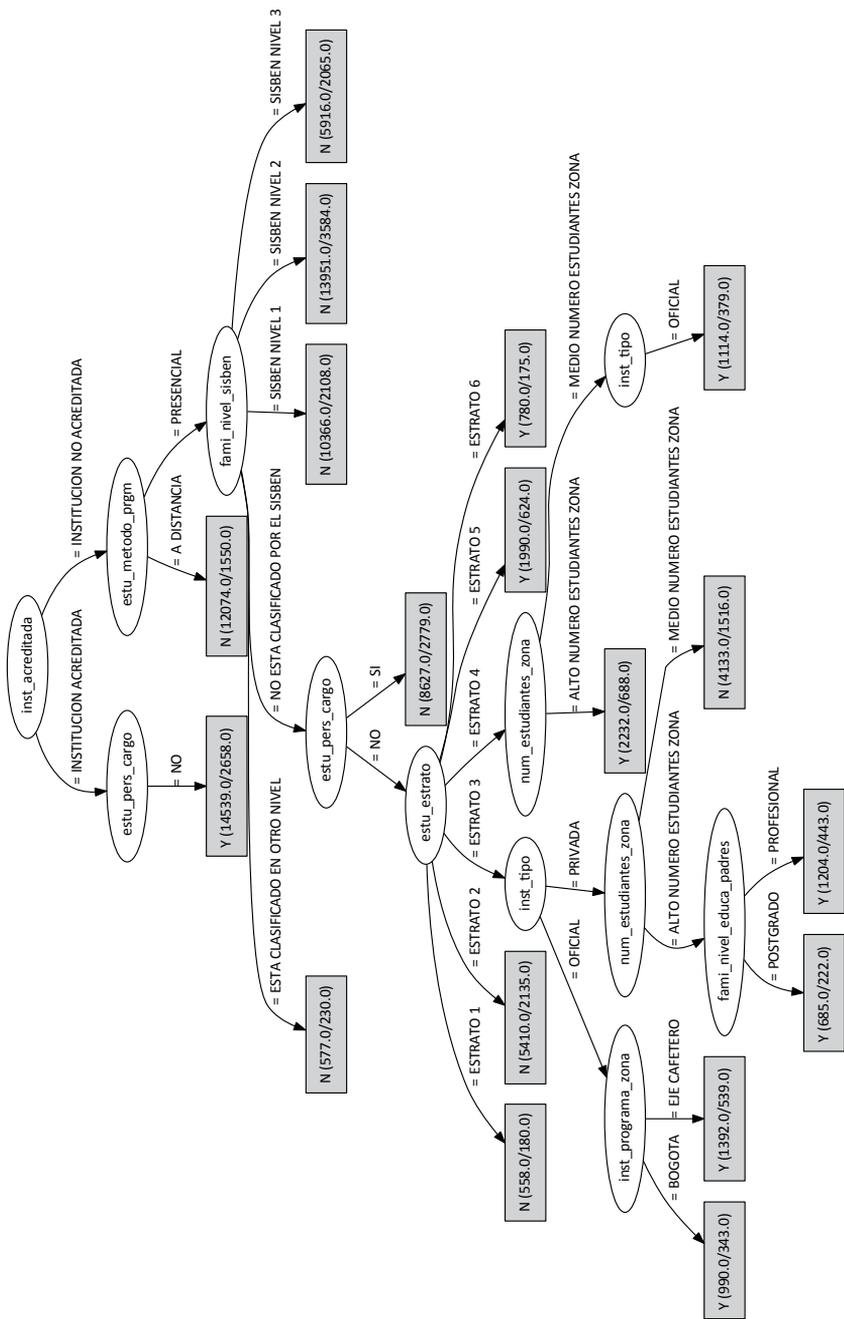


Figura 19. Árbol para inglés

Nota. Elaboración propia.

```

==== 10 Fold Cross Validation ====
==== Summary ====
Correctly Classified Instances          69600          71.7925 %
Incorrectly Classified Instances       27346          28.2075 %
Kappa statistic                        0.3931
Mean absolute error                    0.3815
Root mean squared error                0.437
Relative absolute error                78.6852 %
Root relative squared error           88.7538 %
Coverage of cases (0.95 level)       99.9979 %
Mean rel. region size (0.95 level)   99.999 %
Total Number of Instances            96946

==== Confusion Matrix ====
a          b <-- classified as
48836     8073 | a = N
19273     20764 | b = Y
    
```

Figura 20. Precisión y matriz de confusión del árbol de inglés

Nota. Elaboración propia.

Evaluación

En esta fase se evaluaron los patrones descubiertos con el fin de determinar su validez, remover los patrones redundantes o irrelevantes y traducir los patrones útiles en términos que sean entendibles por el usuario. La evaluación e interpretación de los patrones descubiertos se describe en el capítulo VI de resultados.

Implementación

En esta fase, el conocimiento descubierto se incorpora al existente y se podrá integrar a los procesos de toma de decisiones del ICES y de las instituciones gubernamentales y académicas que velan por la calidad de la educación superior. Una vez estas instituciones intervengan los factores asociados al desempeño académico de los estudiantes de programas profesionales en las pruebas Saber Pro, será posible analizar los resultados y determinar sus efectos.

	Antecedente	Consecuente	Confianza (%)	Soporte (%)
1	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = A DISTANCIA	N	87.16	12.45
2	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = SISBEN NIVEL 1	N	79.66	10.69
3	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = SISBEN NIVEL 2	N	74.31	14.39
4	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = SI	N	67.79	8.90
5	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 1	N	67.74	0.58
6	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = SISBEN NIVEL 3	N	65.09	6.10
7	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 3 y inst_tipo = PRIVADA y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA	N	63.32	4.26
8	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 2	N	60.54	5.58
9	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = ESTA CLASIFICADO EN OTRO NIVEL	N	60.14	0.60
10	inst_acreditada = INSTITUCION ACREDITADA y estu_pers_cargo = NO	Y	81.72	15.00
11	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 6	Y	77.56	0.80
12	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 4 y num_estudiantes_zona = ALTO NUMERO ESTUDIANTES ZONA	Y	69.18	2.30
13	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 5	Y	68.64	2.05
14	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 3 y inst_tipo = PRIVADA y num_estudiantes_zona = ALTO NUMERO ESTUDIANTES ZONA y fami_nivel_educa_padres = POSTGRADO	Y	67.59	0.71
15	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 4 y num_estudiantes_zona = MEDIO NUMERO ESTUDIANTES ZONA y inst_tipo = OFICIAL	Y	65.98	1.15
16	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 3 y inst_tipo = OFICIAL y inst_programa_zona = BOGOTA	Y	65.35	1.02
17	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 3 y inst_tipo = PRIVADA y num_estudiantes_zona = ALTO NUMERO ESTUDIANTES ZONA y fami_nivel_educa_padres = PROFESIONAL	Y	63.21	1.24
18	inst_acreditada = INSTITUCION NO ACREDITADA y estu_metodo_prgm = PRESENCIAL y fami_nivel_sisben = NO ESTA CLASIFICADO POR EL SISBEN y estu_pers_cargo = NO y estu_estrato = ESTRATO 3 y inst_tipo = OFICIAL y inst_programa_zona = EJE CAFETERO	Y	61.28	1.44

Figura 21. Reglas más representativas de inglés

Nota. Elaboración propia.

Referencias

- Artunduaga, M. (2008). *Variables que influyen en el rendimiento académico en la Universidad*. Recuperado de <http://es.slideshare.net/1234509876/variables-del-rendimiento-academico-universidad>
- Colombia, Consejo Nacional de Acreditación [CNA] (2013). *Lineamientos para la acreditación de programas de pregrado*. Bogotá.
- Colombia, Ministerio de Educación Nacional [MEN] (s. f.). Propuesta de lineamientos para la formación por competencias en educación superior. Recuperado de http://www.mineducacion.gov.co/1621/articles-261332_archivo_pdf_lineamientos.pdf
- Davidson, M. y McKinney, G. (2001). Quantitative Reasoning: An Overview. *Western Washington University*. Recuperado de <http://www.wwu.edu/vpue/documents/issue8.pdf>
- Fernández, G. (2009). *Extracción de Información de la Web usando Técnicas de Minería de Datos*. Recuperado de <http://www.tdg-seville.info/Download.ashx?id=48>
- Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde calidad de la educación superior pública. *Revista Educación*, 31(1), 43-63.
- García, M. y Álvarez, A. (2010). Análisis de Datos en WEKA –Pruebas de Selectividad. Recuperado el 5 de mayo de 2013, de <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>
- Gómez, G. y Soares, A. (2013). Diferencias de género con relación al desempeño académico en estudiantes de nivel básico. *Alternativas en Psicología*, xvii(28), 106-118.
- Hall, M., Frank, E. y Witten, I. (2011). Practical Data Mining: Tutorials. *University of Waikato*. Recuperado de www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf
- Han, J. y Kamber, M. (2001). *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Hernández, J., Ramírez, M. y Ferri, C. (2005). *Introducción a la Minería de Datos*. Madrid: Pearson Educación SA.
- Hernández, E. y Lorente, R. (2009). *Minería de datos aplicada a la detección de Cáncer de Mama*. Madrid: Universidad Carlos III. Recuperado de <http://tps5to-utn-frre.googlecode.com/svn/trunk/BI/Cancer%20de%20Mama/14.pdf>
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2011a). *Examen saber pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Recuperado de <http://acofartes.org.co/docsweb/documento/ICFES%202011,%20M%C3%93DULOS%20COMPETENCIAS%20GEN%C3%89RICAS%20Y%20ESPEC%C3%8DFICAS.pdf>

- Moncada, L. y Rubio M. (2011). Determinantes inmediatos del rendimiento académico en los nuevos estudiantes matriculados en el sistema de educación superior a distancia del Ecuador: caso Universidad Técnica Particular de Loja. *Red Internacional de Educación Docente, RIED*, 14(2), 77-95.
- Montero, E. y Villalobos, J. y Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico y a la repetición estudiantil en la Universidad de Costa Rica: un estudio multinivel. *Revista Relieve*, 13(2), 215-234.
- Parra, C., Mejía, L., Valencia, A., Castañeda, E. (2012). *Rendimiento académico de los estudiantes de primer semestre de pregrado de la Facultad de Ingeniería de la Universidad de Antioquia: cohorte 2012-2*. Medellín: Ingeniería y Sociedad. Recuperado de file:///C:/Users/Aspire/Downloads/16537-56603-1-PB%20(1).pdf
- Pineda, C. y Pedraza, A. (2011). *Persistencia y graduación. Hacia un modelo de retención estudiantil para Instituciones de Educación Superior*. Bogotá: Arfo Editores e Impresores Ltda.
- Quinlan, J. (1993). *C4. 5: programs for machine learning. Vol. 1*. Baltimore: Morgan Kaufmann Publishers. Recuperado de <http://books.google.com.co/books?hl=es&lr=&id=HExncp-jbYroC&oi=fnd&pg=PR7&dq=Programs+for+Machine+Learning&ots=nLkbbRq2Y-j&sig=Y5h5CQUdtbZjs1Fjd8ilbJfyRLE>
- Sarramona, J. (2002). *Evaluación de programas de educación a distancia*. Barcelona: Universidad Autónoma de Barcelona.
- Sattler, K. y Dunemann, O. (2001). SQL database primitives for decision tree classifiers. En *Proceedings of the tenth international conference on Information and knowledge management* (pp. 379-386). Atlanta: CIKM. Recuperado de <http://dl.acm.org/citation.cfm?id=502650>
- Seibold, J. (2000). La calidad integral en educación. Reflexiones sobre un nuevo concepto de calidad educativa que integre valores y equidad educativa. *Revista Iberoamericana de Educación*, 23. Recuperado de <http://www.rieoei.org/rie23a07.htm>
- Spicker, P., Alvarez, S. y Gordon, D. (s. f.). *hH: Hacinamiento Hambruna*. Recuperado de <http://biblioteca.clacso.edu.ar/gsd/collect/clacso/index/assoc/D9393.dir/h.pdf>
- Timarán, R. y Millán, M. (2006). New algebraic operators and SQL primitives for mining classification rules. En *Computational Intelligence* [pp. 61-65]. Recuperado de <http://www.actapress.com/PaperInfo.aspx?PaperID=29048&reason=500>
- Toro, J. y Villaveces, J. (2008). *El pensamiento matemático: una competencia genérica emergente*. Recuperado de http://www.mineducacion.gov.co/1621/articles-189357_archivo_pdf_matematica_1B.pdf

Vásquez, C. y Rodríguez, M. (2007). La deserción estudiantil en educación superior a distancia: perspectiva teórica y factores de incidencia. *Revista Latinoamericana de Estudios Educativos*, XXXVII(3 y 4), 107-122.

5 Resultados y discusión

Results and discussion

Resumen

Teniendo en cuenta que la investigación que se reporta se centró en la búsqueda de patrones de desempeño académico en las competencias genéricas de las pruebas Saber Pro 2011-2, en este capítulo se presentan los patrones más destacados en cada una de las competencias genéricas: lectura crítica, comunicación escrita, razonamiento cuantitativo e inglés. Dichos patrones se presentan como reglas para cada competencia y se listan en orden jerárquico según el soporte y la confianza de cada regla. En cuanto a la discusión, este capítulo presenta una reflexión sobre los patrones de mayor confianza, de los que se ubican por encima de la media y los que se sitúan por debajo de ella; este procedimiento se realizó para cada competencia genérica. Vale decir que entre los patrones descubiertos se destacan la acreditación institucional y la modalidad de estudio como dos atributos importantes asociados al desempeño académico.

Palabras clave: competencias genéricas, comunicación escrita, inglés, lectura crítica, razonamiento cuantitativo, patrones de desempeño.

Abstract

Considering that the reported research centered on the search for academic performance patterns in generic skills of the Saber Pro Exam 2011-2, this chapter shows the most outstanding patterns in each generic skill: critical reading, written communication, quantitative reasoning and English. Such patterns are presented as rules for each skill and listed in hierarchical order according to the support and reliability of each rule. Concerning the discussion, this chapter presents a reflection on the most reliable patterns of those above and those below the average; this procedure was performed for each generic skill. It should be noted that, among the patterns discovered, institutional accreditation and mode of study stand out as two important attributes associated with academic performance.

Keywords: generic skills, written communication, English, critical reading, quantitative reasoning, performance patterns.

¿Cómo citar este capítulo?/How to cite this chapter?

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado-Pérez, J. C. (2016). Resultados y discusión. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 151-167). Bogotá: Ediciones Universidad Cooperativa de Colombia. DOI: <http://dx.doi.org/10.16925/9789587600490>



Resultados

En esta sección, por cada competencia se evalúan los modelos de clasificación basados en árboles de decisión y se interpretan los resultados obtenidos en la etapa de modelamiento.

Evaluación e interpretación de resultados para la competencia lectura crítica

Se analizaron los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T97055A32LEC (véase figura 10), en el cual se almacenan los datos válidos sobre los factores sociodemográficos, económicos, académicos e institucionales de 97 055 estudiantes de programas profesionales que presentaron las pruebas Saber Pro 2011-2, en la competencia genérica de lectura crítica, en la que se escogió el atributo *mod_lectura_critica_desemp* como clase. Teniendo en cuenta lo anterior, se puede observar que este último clasifica correctamente a 62 781 instancias, que corresponde a un porcentaje de precisión del 64,7% y 34 274 instancias incorrectamente clasificadas, correspondientes a un porcentaje del 35,3% (véase figura 11).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_lectura_critica_desemp* del repositorio T97055A32LEC que contiene 53 046 registros para el valor “sobre la media” y 44 009 registros para el valor “bajo la media”, y evaluando el modelo con la matriz de confusión de la figura 11, dicho atributo clasifica correctamente a 37 498 casos de estudiantes cuyos resultados en lectura crítica están sobre la media y a 25 283 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 15 548 casos de estudiantes cuyos resultados en lectura crítica están sobre la media y 18 726 casos que están bajo la media. Esto significa que el modelo clasifica correctamente al 70,7% de los estudiantes que están sobre la media en lectura crítica y al 57,4% de las estudiantes que están bajo la media en esta competencia.

De acuerdo con las reglas de clasificación obtenidas con este modelo (véase figura 12), los patrones más representativos descubiertos, teniendo en cuenta el soporte y la confianza, son los siguientes:

- **Regla 2.** Si la IES no está acreditada y la metodología de formación es a distancia, el desempeño de los estudiantes en lectura crítica tiene mayor

probabilidad de estar bajo la media. El 12,48% de los 97 055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 69,11% de los 12 112 (12,48%) estudiantes se clasifica correctamente y el 27,52% de los 44 009 estudiantes que están bajo la media cumple este patrón.

- **Regla 5.** Si la IES no está acreditada, la metodología de formación es presencial, el carácter académico de la institución es de universidad, el número de IES en la zona OCAD a la que pertenece la institución está entre 60 y 70, el número de estudiantes de programas profesionales en la zona OCAD en la que se ofrece el programa está entre 10 000 y 20 000, los estudiantes pertenecen a programas no acreditados y además tienen personas a cargo, el desempeño de los estudiantes en lectura crítica tiene mayor probabilidad de estar bajo la media. El 4,85% de los 97 055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 62% de los 4710 (4,85%) estudiantes se clasifica correctamente y el 10,70% de los 44 009 estudiantes que están bajo la media cumplen este patrón.

- **Regla 4.** Si la IES no está acreditada, la metodología de formación es presencial, el carácter académico de la institución es de institución universitaria y los estudiantes están clasificados en el nivel 1 del Sisben, el desempeño de los estudiantes en lectura crítica tiene mayor probabilidad de estar bajo la media. El 2,07% de los 97 055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifican de esta manera. El 67,35% de los 2012 (2,07%) estudiantes se clasifican correctamente y el 4,57% de los 44 009 estudiantes que están bajo la media cumplen este patrón.

- **Regla 7.** Si la IES está acreditada, el desempeño de los estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 16,80% de los 97 055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 78,87% de los 16 306 (16,80%) estudiantes se clasifica correctamente y el 30,74% de los 53 046 estudiantes que están sobre la media cumple este patrón.

- **Regla 9.** Si la IES no está acreditada, la metodología de formación es presencial, el carácter académico de la institución es de universidad, el número de IES en la zona a la que pertenece la institución está entre 60 y 70, y el número de estudiantes de programas profesionales en la zona OCAD en la que se ofrece el programa es mayor que 20 000, el desempeño de los estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 13,93% de los 9

055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 62,80% de los 13 524 (13,93%) estudiantes se clasifica correctamente y el 25,49% de los 53 046 estudiantes que están sobre la media cumplen este patrón.

- **Regla 8.** Si la IES no está acreditada, la metodología de formación es presencial, el carácter académico de la institución es de universidad y el número de IES en la zona OCAD a la que pertenece la institución es mayor que 70, el desempeño de los estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 8,37% de los 97 055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 69,43% de los 8126 (8,37%) estudiantes se clasifica correctamente y el 15,32% de los 53 046 estudiantes que están sobre la media cumple este patrón.

- **Regla 12.** Si la IES no está acreditada, la metodología de formación es presencial, el carácter académico de la institución es de universidad, el número de IES en la zona a la que pertenece la institución está entre 60 y 70, el número de estudiantes de programas profesionales en la zona OCAD en la que se ofrece el programa está entre 10 000 y 20 000, los estudiantes pertenecen a programas no acreditados, no tienen personas a cargo, no están clasificados en el Sisben y la institución es de tipo oficial, el desempeño de los estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 2,80% de los 97 055 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 61,11% de los 2718 (2,80%) estudiantes se clasifica correctamente y el 5,12% de los 53,046 estudiantes que están sobre la media cumple este patrón.

Evaluación e interpretación de resultados para la competencia comunicación escrita

Se analizaron los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T95337A32ESC (véase figura 13), en el cual se almacenan los datos válidos sobre los factores sociodemográficos, económicos, académicos e institucionales de 95 337 estudiantes de programas profesionales que presentaron las pruebas Saber Pro 2011-2 en la competencia genérica de comunicación escrita, en la que se escogió el atributo *mod_comunica_escrita_desemp* como clase. Teniendo en cuenta esto, se puede observar que dicho atributo clasifica correctamente a 56 669 instancias, que corresponden a un porcentaje de precisión del 59,4% y 38

668 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 40,5% (véase figura 14).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_comunica_escrita_desemp* del repositorio T95337A32ESC que contiene 46 986 registros para el valor “sobre la media” y 48 351 registros para el valor “bajo la media”, y evaluando el modelo con la matriz de confusión de la figura 14, se encuentra que este atributo clasifica correctamente a 25 435 casos de estudiantes cuyos resultados en comunicación escrita están sobre la media y a 31 234 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 21 551 casos de estudiantes cuyos resultados en comunicación escrita están sobre la media y 17 117 casos que están bajo la media. Esto significa que el modelo clasifica correctamente al 54,13% de los estudiantes que están sobre la media en comunicación escrita y al 64,6% de las estudiantes que están bajo la media en esta competencia.

De acuerdo con las reglas de clasificación obtenidas con este modelo (ver figura 15), los patrones más representativos descubiertos, teniendo en cuenta el soporte y la confianza, son los siguientes:

- **Regla 3.** Si la IES no está acreditada y la metodología de formación es a distancia, el desempeño de los estudiantes en comunicación escrita tiene mayor probabilidad de estar bajo la media. El 12,36% de los 95 337 estudiantes evaluados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 62,42% de los 11 787 (12,36%) estudiantes se clasifica correctamente y el 24,38% de los 48 351 estudiantes que están bajo la media cumplen este patrón.
- **Regla 1.** Si la IES no está acreditada, la metodología de formación es presencial, los estudiantes son de género masculino y la zona OCAD en la que se ofrece el programa es Caribe, el desempeño de los estudiantes en comunicación escrita tiene mayor probabilidad de estar bajo la media. El 5,44% de los 95 337 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 64,92% de los 5188 (5,44%) estudiantes se clasifica correctamente y el 10,73% de los 48 351 estudiantes que están bajo la media cumple este patrón.
- **Regla 5.** Si la IES está acreditada y el número de estudiantes de programas profesionales en la zona OCAD en la que se ofrece el programa es mayor que 20 000, el desempeño de los estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 7,82% de los 95 337 estudiantes

analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 71,06% de los 7458 (7,82%) estudiantes se clasifica correctamente y el 15,87% de los 46 986 estudiantes que están sobre la media cumple este patrón.

- **Regla 6.** Si la IES está acreditada y el número de estudiantes de programas profesionales en la zona OCAD en la que se ofrece el programa está entre 10 000 y 20 000 y los estudiantes son de género femenino, el desempeño de los estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 5,04% de los 95 337 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 62,67% de los 4801 (5,04%) estudiantes se clasifica correctamente y el 10,22% de los 46 986 estudiantes que están sobre la media cumple este patrón.

- **Regla 8.** Si la IES no está acreditada, la metodología de formación es presencial, los estudiantes son de género femenino y el programa es acreditado, entonces, el desempeño de los estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 4,32% de los 95 337 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 60,51% de los 4115 (4,32%) estudiantes se clasifica correctamente y el 8,76% de los 46 986 estudiantes que están sobre la media cumple este patrón.

Evaluación e interpretación de resultados para la competencia razonamiento cuantitativo

Se analizaron los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T97057A32CUA (véase figura 16), en el cual se almacenan los datos válidos sobre los factores sociodemográficos, económicos, académicos e institucionales de 97 057 estudiantes de programas profesionales que presentaron las pruebas Saber Pro 2011-2 en la competencia genérica de razonamiento cuantitativo, en la cual se escogió el atributo *mod_razona_cuantitativo_desemp* como clase. A partir de lo anterior, se puede observar que el atributo mencionado clasifica correctamente a 64 549 instancias, que corresponde a un porcentaje de precisión del 66,5% y 32 508 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 33,5% (véase figura 17).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_razona_cuantitativo_desemp* del repositorio T97057A32CUA que contiene 48 144 registros para el valor “sobre la media” y 48 913 registros para el valor “bajo la

media”, y evaluando el modelo con la matriz de confusión de la figura 17, este último clasifica correctamente a 29 093 casos de estudiantes cuyos resultados en razonamiento cuantitativo están sobre la media y a 35 456 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 19 051 casos de estudiantes cuyos resultados en razonamiento cuantitativo están sobre la media y 13 457 casos que están bajo la media. Esto significa que el modelo clasifica correctamente al 60,43% de los estudiantes que están sobre la media en razonamiento cuantitativo y al 72,49% de los estudiantes que están bajo la media en esta competencia.

De acuerdo con las reglas de clasificación obtenidas con este modelo (véase 37 figura), los patrones más representativos descubiertos, teniendo en cuenta el soporte y la confianza, son los que se presentan a continuación:

- **Regla 1.** Si la IES no está acreditada, el área del grupo de referencia del programa es ciencias sociales y la metodología de formación es a distancia, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 10,23% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 73,88% de los 9927 (10,23%) estudiantes se clasifica correctamente y el 20,30% de los 48 913 estudiantes que están bajo la media cumple este patrón.
- **Regla 5.** Si la IES no está acreditada, el área del grupo de referencia del programa es ciencias sociales, la metodología de formación es presencial y el área del conocimiento al que pertenece el programa es ciencias de la educación, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 7,24% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 70,66% de los 7029 (7,24%) estudiantes se clasifican correctamente y el 14,37% de los 48 913 estudiantes que están bajo la media cumplen este patrón.
- **Regla 6.** Si la IES no está acreditada, el área del grupo de referencia del programa es ciencias sociales, la metodología de formación es presencial y el área del conocimiento al que pertenece el programa es ciencias sociales y humanas, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 6,13% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 68,35% de los 5953 (6,13%) estudiantes se clasifica correctamente y el 12,17% de los 48 913 estudiantes que están bajo la media cumple este patrón.

- **Regla 8.** Si la IES no está acreditada, el área del grupo de referencia del programa es ciencias humanas, la institución es de carácter privado y los ingresos familiares de los estudiantes es de tres salarios mínimos, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 4,29% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 66,54% de los 4166 (4,29%) estudiantes se clasifica correctamente y el 8,52% de los 48 913 estudiantes que están bajo la media cumple este patrón.
- **Regla 17.** Si la IES no está acreditada, el área del grupo de referencia del programa es ciencias naturales y técnicas, la metodología de formación es presencial, el programa no es acreditado y el género de los estudiantes es masculino, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar sobre la media. El 9,26% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 70,52% de los 8988 (9,26%) estudiantes se clasifica correctamente y el 18,67% de los 48 144 estudiantes que están sobre la media cumple este patrón.
- **Regla 12.** Si la IES está acreditada y el área del conocimiento al que pertenece el programa es ingeniería, arquitectura, urbanismo y afines, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar sobre la media. El 4,21% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 90,05% de los 4090 (4,21%) estudiantes se clasifica correctamente y el 8,50% de los 48 144 estudiantes que están sobre la media cumple este patrón.
- **Regla 16.** Si la IES está acreditada y el área del conocimiento al que pertenece el programa es economía, administración, contaduría y afines, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar sobre la media. El 3,91% de los 97 057 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 76,37% de los 3791 (3,91%) estudiantes se clasifica correctamente y el 7,87% de los 48 144 estudiantes que están sobre la media cumple este patrón.

Evaluación e interpretación de resultados para la competencia inglés

Se analizaron los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T96946A32ING (véase figura 38), en el cual se almacenan los

datos válidos sobre los factores sociodemográficos, económicos, académicos e institucionales de 96 946 estudiantes de programas profesionales que presentaron las pruebas Saber Pro 2011-2 en la competencia genérica de inglés, en la cual se escogió el atributo *mod_ingles_desemp* como clase. A partir de esto se puede observar que dicho atributo clasifica correctamente a 69 600 instancias, que corresponden a un porcentaje de precisión del 71,8% y 27 346 instancias incorrectamente clasificadas, correspondientes a un porcentaje del 28,2% (véase figura 39).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_ingles_desemp* del repositorio T96946A32ING que contiene 40 037 registros para el valor “sobre la media” y 56 909 registros para el valor “bajo la media”, y evaluando el modelo con la matriz de confusión de la figura 39, este clasifica correctamente a 20 764 casos de estudiantes cuyos resultados en inglés están sobre la media y a 48 836 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 19 273 casos de estudiantes cuyos resultados en inglés están sobre la media y 8073 casos que están bajo la media. Esto significa que el modelo clasifica correctamente al 51,86% de los estudiantes que están sobre la media en inglés y al 85,81% de los estudiantes que están bajo la media en esta competencia.

De acuerdo con las reglas de clasificación obtenidas con este modelo (véase figura 40), los patrones más representativos descubiertos, teniendo en cuenta el soporte y la confianza, son los siguientes:

- **Regla 3.** Si la IES no está acreditada, la metodología de formación es presencial y los estudiantes están clasificados en el nivel 2 del Sisben, el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 14,39% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 74,31% de los 13 951 (14,39%) estudiantes se clasifica correctamente y el 24,51% de los 56 909 estudiantes que están bajo la media cumple este patrón.
- **Regla 1.** Si la IES no está acreditada y la metodología de formación es a distancia, el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 12,45% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 87,16% de los 12 074 (12 45%) estudiantes se clasifica correctamente y el 21,22% de los 56 909 estudiantes que están bajo la media cumple este patrón.
- **Regla 2.** Si la IES no está acreditada, la metodología de formación es presencial y los estudiantes están clasificados en el nivel 1 del Sisben, el desempeño de

los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 10,69% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 79,66% de los 10 366 (10,69%) estudiantes se clasifica correctamente y el 18,22% de los 56 909 estudiantes que están bajo la media cumple este patrón.

- **Regla 4.** Si la IES no está acreditada, la metodología de formación es presencial, si los estudiantes no están clasificados en el Sisben y tienen además personas a cargo, el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 8,90% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 67,79% de los 8627 (8,90%) estudiantes se clasifica correctamente y el 18,22% de los 56 909 estudiantes que están bajo la media cumple este patrón.

- **Regla 10.** Si la IES está acreditada y los estudiantes no tienen personas a cargo, el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 1,0% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 81,72% de los 14 539 (15,0%) estudiantes se clasifica correctamente y el 36,31% de los 40 037 estudiantes que están sobre la media cumple este patrón.

- **Regla 12.** Si la IES no está acreditada, la metodología de formación es presencial, los estudiantes no están clasificados en el Sisben, no tienen personas a cargo, son de estrato 4 y el número de estudiantes de la zona en la cual se ofrece el programa es mayor que 20 000, el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 2,30% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifica de esta manera. El 68,18% de los 2232 (2,30%) estudiantes se clasifica correctamente y el 5,57% de los 4 037 estudiantes que están sobre la media cumplen este patrón.

- **Regla 13.** Si la IES no está acreditada, la metodología de formación es presencial, los estudiantes no están clasificados en el Sisben, no tienen personas a cargo y son de estrato 5, el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 2,05% de los 96 946 estudiantes analizados en esta competencia en la prueba Saber Pro 2011-2 se clasifican de esta manera. El 68,64% de los 1990 (2,05%) estudiantes se clasifica correctamente y el 4,97% de los 40 037 estudiantes que están sobre la media cumple este patrón.

Discusión

El objetivo de esta investigación fue descubrir patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales en las pruebas Saber Pro 2011-2, a partir de los datos sociodemográficos, económicos, académicos e institucionales almacenados en las bases de datos del ICFES, utilizando la técnica de clasificación basada en árboles de decisión. Dicho objetivo se logró en su totalidad.

Para efectos de la discusión de los resultados, se escogieron los patrones de mayor confianza, tanto los que se ubican por encima de la media, como aquellos que se sitúan por debajo de ella; este procedimiento se realizó para cada competencia genérica. En los patrones descubiertos, se destacan la acreditación institucional y la modalidad de estudio como dos atributos importantes asociados al desempeño académico.

En efecto, los modelos de clasificación basados en árboles de decisión, contruidos para cada competencia genérica, muestran que el patrón de IES acreditadas se asocia con el buen desempeño académico de los estudiantes de programas profesionales en las cuatro competencias genéricas objeto de estudio. Este resultado está en línea con la fundamentación conceptual expresada en la política pública sobre calidad de la educación superior, particularmente sobre el tema de acreditación institucional, pues de hecho el cumplimiento de los doce factores de calidad señalados por el CNA (2013) da cuenta de las condiciones de favorabilidad de todos los procesos en el interior de las IES, que permiten un mejor desempeño académico de los estudiantes.

En la investigación sobre compromiso estudiantil y desempeño académico universitario, realizada en siete universidades colombianas con acreditación institucional de alta calidad, llevada a cabo por Pineda y Pedraza (2011), desde la perspectiva de los estudiantes, señala que estas universidades se preocupan por proponer actividades y tareas que implican la aplicación de conceptos y teorías en problemas prácticos; le asignan importancia al trabajo en equipo, utilización de medios electrónicos con fines académicos y a procesos de realimentación del aprendizaje de los estudiantes; fomentan espacios de socialización que fortalecen la calidad de las relaciones estudiante y comunidad educativa. Estas estrategias favorecen el desempeño académico de los estudiantes en el desarrollo de las competencias genéricas.

En el mismo orden de ideas, si la IES no está acreditada, se encontró que este hecho junto con la modalidad de estudio a distancia se constituyen en patrón que se asocia con el bajo desempeño académico de los estudiantes de carreras profesionales en las competencias genéricas de las pruebas Saber Pro 2011-2.

La modalidad del programa, según resultados de la investigación de Parra et al. (2012), realizada con estudiantes de primer semestre de pregrado de la Facultad de Ingeniería de la Universidad de Antioquia, cohorte 2012-2, hace referencia a la metodología mediante la cual se oferta y desarrolla el programa, que puede ser: presencial, semipresencial, a distancia y virtual. En dicho estudio se encontró que la modalidad del programa se asocia de manera importante al desempeño académico. Por ejemplo, en la condición de insuficiente se ubica el 17% de los matriculados en la modalidad presencial y el 50% de los de la modalidad virtual; cifras que indican que los estudiantes de programas presenciales tienen mejor desempeño que aquellos que están en la modalidad a distancia.

Por su parte, Artunduaga (2008), en relación con la modalidad de formación, considera que es factor clave del desempeño académico la forma como se orienta al estudiante en el aprendizaje. Al respecto, Vásquez y Rodríguez (2007) evidencian el papel del estudiante de modalidad a distancia y aseguran que es él el responsable de su éxito educativo, que le corresponde autodirigirse y autoorganizarse, con el fin de alcanzar los objetivos propuestos en el plan de estudio. Dichos autores argumentan que si no se tienen desarrolladas estas capacidades de independencia, tendería al bajo desempeño, pues su característica principal está dada por la separación entre estudiante y docente, que puede conllevar de manera paradójica, las ventajas y desventajas de este modelo y potenciar igualmente efectos desfavorables si los programas no están bien estructurados y diseñados.

En el mismo sentido, Rubio (2009) considera que la modalidad a distancia, a diferencia de la presencial, tiene sus propias características: autonomía en el aprendizaje, utilización de tecnología, apoyo-tutorial, comunicación masiva, entre otros, que influyen en el desempeño académico. Sarramona (2002) señala que la adaptación del material para la educación a distancia con todos los principios pedagógicos que esta requiere constituye una tarea difícil, dada la falta de especialistas en el tema. A menudo se confunde un recurso para la autoformación con un simple material informativo, lo cual posteriormente se convierte en factor negativo en el desempeño académico del estudiante y fuente de fracaso de los programas. El autor afirma que el bajo desempeño académico en la modalidad a distancia puede provenir de varias fuentes: capacidad intelectual insuficiente,

débil compromiso académico y otros factores provenientes de las dimensiones sociodemográfica y económica, que caracterizan al estudiante de modalidad a distancia. En los sociodemográficos destaca la edad, el grupo familiar, las personas a cargo, el nivel cultural, el estado civil, la ubicación geográfica, entre otros; en los factores económicos subraya el estrato social, los ingresos económicos individuales y familiares. Por su parte, Moncada y Rubio (2011) exponen, además, que estos estudiantes tienen características particulares que dependen de la dispersión geográfica, edad, estado civil, motivación, intereses profesionales, falta de tiempo, aislamiento, falta de recursos y manejo de la tecnología; la mayoría posee obligaciones laborales y familiares.

Particularmente, para cada competencia se encuentra que la acreditación institucional y la modalidad de estudio se asocian al buen o bajo desempeño académico de los estudiantes que presentaron las pruebas Saber Pro 2011-2.

Analizando el modelo de la figura 10, en la competencia de lectura crítica, el buen desempeño se asocia al hecho que la IES esté acreditada; el hecho de que la IES no lo esté, junto con la modalidad presencial, constituye un patrón de buen desempeño. Siguiendo el modelo jerárquicamente desde el nodo de modalidad presencial, se encuentra que el buen desempeño académico en esta competencia está asociado, además, con altos ingresos familiares (dimensión económica), número alto y medio de IES, número de estudiantes en la zona y con programas acreditados (dimensión institucional).

Por su parte, el bajo desempeño en lectura crítica tiene relación con IES no acreditadas y la modalidad a distancia. En el caso de la modalidad presencial, el bajo desempeño se asocia a la dimensión económica, bajo nivel de Sisben y con la dimensión institucional, las instituciones tecnológicas o técnicas profesionales. En este contexto, Garbanzo (2007) indica que factores como la pobreza y la falta de apoyo social están relacionados con el desempeño académico; no obstante, advierte que no hay correspondencia estricta entre las desigualdades sociales y educativas; indica que existen otros factores como la familia, el sistema educativo o la institución, que pueden incidir en tal desigualdad. Seibold (2000) señala que si bien el contexto socioeconómico afecta el nivel de calidad educativa, de ningún modo lo determina; así mismo, reconoce que se presenta asociación entre estas variables, pero no configura una relación causa-efecto.

Analizando el modelo de la figura 13, en la competencia de comunicación escrita, el buen desempeño se asocia al hecho de que la IES esté acreditada y que haya un alto número de estudiantes en la zona geográfica de la IES. En un orden

inferior de jerarquía, se encuentra que también tienen buen desempeño las mujeres y los hombres que estudien carreras de ciencias humanas. En el caso de que la IES no esté acreditada y la modalidad sea presencial, también es un patrón de buen desempeño el hecho de que el programa esté acreditado.

Por su parte, el bajo desempeño en comunicación escrita tiene relación con IES no acreditadas y con la modalidad a distancia. En el caso de la modalidad presencial, el bajo desempeño se asocia con el género masculino de las zonas geográficas Caribe, Centro sur y Pacífico.

En este contexto, están los hallazgos de Gómez y Soares (2013), quienes en su análisis sobre la diferencia de género en relación con el desempeño académico expresan que tales diferencias no son significativas y que no se puede afirmar de modo definitivo que exista una relación directa entre el rendimiento académico y el sexo; sin embargo, según Garbanzo (2007), las mujeres sí tienen un ligero mejor desempeño académico que los hombres. En el mismo sentido, Montero y Villalobos (2004) señalan que existe una relación significativa entre el sexo y el desempeño académico que está a favor de las mujeres.

Analizando el modelo de la figura 16, en la competencia de razonamiento cuantitativo, el buen desempeño se asocia al hecho de que la IES esté acreditada y que el área del grupo de referencia sea ciencias naturales y técnicas. En un orden inferior de jerarquía, se encuentra que también tienen buen desempeño los estudiantes de IES no acreditadas, de modalidad presencial y de ciencias naturales y técnicas. Por su parte, el bajo desempeño en la competencia de razonamiento cuantitativo tiene relación con IES no acreditadas y el área del grupo de referencia de ciencias sociales y humanas.

En efecto, los estudiantes de las áreas de ciencias sociales y humanas, y ciencias de la educación presentan bajo desempeño en esta competencia; realidad que es reconocida por las IES, a tal punto que como alternativa de solución vienen implementando cursos adicionales a sus planes de estudio con el fin de proporcionar una formación básica en matemáticas, en programas que al margen de las pruebas Saber Pro habría sido impensable su implementación. No obstante, lo ideal sería que aquellos cursos fueran constitutivos de los currículos de todos los programas de formación profesional, en la perspectiva de recuperar su aporte a la formación y estructuración del pensamiento matemático, de modo que, como señalan Davidson y McKinney (2001), se desarrolle la competencia para la interpretación e inferencia a partir de modelos; la representación de la información matemática de diversas formas; la utilización de métodos aritméticos, algebraicos,

geométricos y estadísticos para solucionar problemas; la estimación y chequeo de las respuestas a problemas matemáticos con miras a establecer su razonabilidad, identificar alternativas y seleccionar los mejores resultados; y el reconocimiento de los límites de los métodos matemáticos y estadísticos, para contribuir al fomento y desarrollo de la competencia de razonamiento cuantitativo.

Se entiende que es en este sentido como el Ministerio de Educación Nacional (s. f), concibe la alfabetización cuantitativa como el dominio del lenguaje y los métodos matemáticos indispensables para lograr la comprensión de problemas y discusiones cotidianas contemporáneas, sin lo cual las personas se encontrarían parcialmente incomunicadas. En la misma línea, haciendo referencia al pensamiento matemático como una competencia genérica emergente, los autores Toro y Villaveces (2008) plantean que “podemos pensar que hay un sustrato mínimo de competencias matemáticas, sin el cual un profesional puede quedar parcialmente incomunicado del mundo moderno que lo rodea” (p. 1); afirmación que muestra que la competencia del razonamiento cuantitativo es indispensable y, por lo mismo, inaplazable para todo profesional.

A pesar del reconocimiento de la importancia del razonamiento cuantitativo en la formación de todo profesional, se encuentran currículos de programas de humanidades, ciencias sociales y afines que atienden poco o nada su fomento y desarrollo, hecho que podría asociarse con el bajo desempeño académico en la competencia de razonamiento cuantitativo, tal como se encontró en esta investigación.

Analizando el modelo de la figura 38, en la competencia de inglés el buen desempeño se asocia al hecho de que la IES esté acreditada y que el estudiante no tenga personas a cargo. En el caso de que la IES no esté acreditada y la modalidad sea presencial, también constituye un patrón de buen desempeño el hecho de que los estudiantes pertenezcan a estratos altos (4, 5 y 6).

Por su parte, el bajo desempeño en inglés tiene relación con una IES no acreditada y con modalidad a distancia. En el caso de la modalidad presencial, el bajo desempeño se asocia con niveles de Sisben bajos (1 y 2) y con estratos socioeconómicos bajos (1 y 2). Estos hechos coinciden con Garbanzo (2007), Seibold (2000) y Montero y Villalobos (2004), en el sentido de que un resultado generalmente aceptable en el desempeño académico es la existencia de una asociación significativa entre el nivel socioeconómico del estudiante y su desempeño académico, hecho que coincide con lo encontrado en la presente investigación, en cuanto que forma parte de los patrones descubiertos, particularmente en la competencia genérica de inglés.

En resumen, entre los patrones de desempeño académico de las competencias genéricas, descubiertos y considerados en esta discusión, la acreditación institucional se encuentra asociada con otros atributos, tanto en el desempeño por encima de la media como por debajo de ella, excepto en lectura crítica, en la cual constituye en sí mismo un patrón de desempeño académico por encima de la media. Entre los atributos se destacan la modalidad del programa, el área del conocimiento, el número de instituciones en la zona, el número de estudiantes por zona, el estrato socioeconómico, el número de personas a cargo del estudiante y el género, entre otros; atributos que, en relación con el desempeño académico, de alguna manera, son considerados en los resultados de investigaciones de diversos autores.

Por otra parte, también se encontró que existe asociación para el buen desempeño académico en las competencias genéricas de lectura crítica y comunicación escrita, un alto número de estudiantes e instituciones por zona y otros atributos. En este contexto, conviene destacar que las variables denominadas número de IES y número de estudiantes por zona se crearon en el marco de esta investigación y se convierten en atributos importantes para el desarrollo de la educación superior en las zonas y regiones a las cuales pertenecen; por lo cual deberían ser consideradas en las políticas públicas de mejoramiento de la calidad en la formación de profesionales y su cobertura.

Así mismo, el área de los grupos de referencia de los programas académicos está asociada con el buen y mal desempeño en la competencia genérica de razonamiento cuantitativo. Otros atributos en los patrones descubiertos en esta investigación, asociados al buen y mal desempeño académico en las competencias de lectura crítica e inglés son el estrato socioeconómico y el nivel de Sisben al que pertenece el estudiante.

Finalmente, otro de los hallazgos de este estudio es la identificación del atributo género en los patrones de desempeño académico. En efecto, el género femenino se asocia, junto con otros atributos, al buen desempeño en la competencia de comunicación escrita, y el género masculino, con un bajo desempeño en dicha competencia; no obstante, se destaca que tales diferencias no son significativas.

Estos resultados no se pueden obtener a través de consultas simples a las bases de datos, ya que se desconocen las relaciones de los atributos con las clases. Únicamente, aplicando las técnicas y algoritmos de minería de datos, que en el caso de esta investigación es clasificación por árboles de decisión con el algoritmo J48, se logra obtener patrones asociados con el buen a mal rendimiento académico en las pruebas Saber Pro 2011-2.

Es necesario reconocer que la exactitud de los modelos de clasificación por árboles de decisión obtenidos para las cuatro competencias genéricas no es muy alta, pero sí satisfactoria para poder obtener en una primera aproximación, con esta técnica, los patrones que asocian los factores sociodemográficos, económicos, académicos e institucionales con el desempeño académico de los estudiantes que presentaron las pruebas Saber Pro 2011-2. Esto se debe a que se utilizaron como fuentes de datos únicamente las bases de datos del ICFES, en las cuales la llave principal para identificar cada estudiante es ficticia, lo que imposibilita el cruce de valores con otras fuentes de datos externas como las bases de datos del Sisben, de la Registraduría Nacional y del MEN, entre otras, que seguramente hubiesen mejorado la calidad de los datos y por ende la exactitud de los modelos.

De todas formas, se debe continuar investigando en este tema, seguramente identificando nuevos atributos o reevaluando los utilizados, mejorando las técnicas de limpieza y transformación utilizadas o usando otras técnicas de minería de datos que nos permitan mejores resultados en cuanto a precisión.

Conclusiones

En este estudio la utilización de la técnica clasificación por árboles de decisión generó modelos predictivos para cada competencia genérica, que permitieron descubrir conocimiento para predecir tendencias sobre el desempeño académico de los actuales y futuros estudiantes de los programas profesionales que presentan las pruebas Saber Pro. Esto con el fin de identificar riesgos y oportunidades que ayuden a las instituciones gubernamentales y de educación superior a tomar decisiones para el mejoramiento de la educación superior en Colombia.

Los resultados obtenidos en esta investigación no se podrían conseguir a través de consultas simples SQL a las bases de datos, ya que se desconocen las relaciones o asociaciones de los 32 atributos seleccionados con las clases. Únicamente aplicando las técnicas y algoritmos de minería de datos, que en el caso de esta investigación es clasificación por árboles de decisión con el algoritmo J48, se logró obtener patrones asociados con el buen o mal rendimiento académico en las pruebas Saber Pro 2011-2

En los patrones de desempeño académico descubiertos en las cuatro competencias genéricas de las pruebas Saber Pro 2011-2, se encontró que la acreditación institucional se constituye en factor importante asociado al desempeño académico de los estudiantes de programas profesionales en las pruebas Saber Pro 2011-2. Esto debido a que la acreditación de la calidad de las IES implica la calidad integral en el proceso de formación profesional, en el cual el logro de las competencias genéricas y específicas constituye uno de los aspectos fundamentales que impacta el desempeño académico de los estudiantes. El resultado obtenido en el descubrimiento de los patrones de desempeño académico en las competencias genéricas está en línea directa con la fundamentación conceptual expresada en la política pública sobre calidad de la educación superior, particularmente sobre el tema de acreditación institucional; de hecho, el cumplimiento de los doce factores de calidad señalados por el CNA (2015) da cuenta de las condiciones de favorabilidad de todos los procesos que permiten un mejor desempeño académico de los estudiantes, en las competencias genéricas previstas por el ICFES para el período 2011-2. En el

mismo orden de ideas, la carencia de acreditación institucional se constituye en un factor que se asocia con bajo desempeño académico de estudiantes de carreras profesionales.

En este sentido, es imperativa la implementación de la política pública de calidad de la educación superior, en la que el Estado regule y vigile de manera permanente la calidad de la educación superior, de modo que las IES adquieran una dinámica integradora de los procesos relacionados con su acreditación, en una visión sistémica e interdependiente, teniendo en cuenta que en el proceso de acreditación de programas e instituciones se estandarizan factores, características y aspectos. Mejorar la calidad de la educación implica que las IES autoevalúen sus procesos con el fin de que identifiquen aquellos que sean susceptibles de mejoras, pues a través de su vigilancia y control se crean mayores probabilidades para atender las múltiples relaciones de formación que requiere la persona como sujeto inserto en la dinámica global del proceso educativo y productivo.

Así mismo, se requiere que las IES creen unidades de apoyo académico, cuyo objetivo sea la orientación de estrategias y programas que fortalezcan el desempeño académico del estudiante en las competencias genéricas; siendo de vital importancia para estas unidades ahondar en el conocimiento de los factores asociados al desempeño académico en las dimensiones sociodemográficas, económicas, académicas e institucionales que atañen a los estudiantes y que afectan su rendimiento académico y su bienestar en general.

Las IES, en la consecución global de los factores de acreditación institucional, deben fortalecer las experiencias educativas que representen reto a nivel cognitivo y creativo para el estudiante. Es decir, invertir y propender al diseño e implementación de propuestas curriculares que beneficien la lectura crítica, comunicación escrita, razonamiento cuantitativo y la comunicación efectiva en inglés, de tal suerte que estimulen al estudiante y lo impulsen a movilizar sus saberes. Esta afirmación destaca la importancia del desarrollo de estrategias didácticas creativas e innovadoras, que exigen al docente en particular y a la comunidad académica en general estar a la vanguardia en el manejo de pedagogías apropiadas.

Por otra parte, en los patrones descubiertos de desempeño académico en las pruebas Saber Pro, la acreditación institucional se encuentra asociada con otros factores, tanto en el desempeño por encima de la media como por debajo de ella. Se observó que los estudiantes que se forman en modalidad presencial obtienen mejor desempeño académico que aquellos que estudian en programas de modalidad a distancia, y que dichos resultados son ocasionados por la asociación de

múltiples factores provenientes de las dimensiones sociodemográfica, económica, académica e institucional. Entre dichos atributos se destacan el área del conocimiento, el número de instituciones en la zona, el número de estudiantes por zona, el estrato socioeconómico, el número de personas a cargo del estudiante y el género, entre otros; atributos que, de alguna manera, son considerados en los resultados de investigaciones de diversos autores en relación con el desempeño académico.

Entre las dificultades presentadas en el desarrollo de la investigación están la mala calidad de los datos de las bases de datos del ICFES, a tal punto que se tuvieron que descartar ciertos atributos por la imposibilidad de obtener sus valores en otras fuentes, y que de alguna manera podrían haber influido en el descubrimiento de los patrones objeto de este estudio, además del gran consumo de recursos que implicó el proceso de limpieza y transformación de los datos y que de alguna manera incidieron en la exactitud de los modelos obtenidos para cada competencia.

Los procesos de acreditación de IES y programas fortalecen la formación profesional por competencias y favorecen así el logro en los estudiantes, tanto de las competencias genéricas como específicas, necesarias para su adecuado desempeño profesional, teniendo en cuenta que ellas llevan finalmente a una actividad profesional altamente productiva y de calidad, y que a esto se suman cualidades de la personalidad como los conocimientos, destrezas y aptitudes que permiten resolver problemas propios de la profesión. Los resultados de la evaluación de las competencias se convierten entonces en un sistema que puede valorar diversos elementos del proceso educativo que permite determinar el grado en el cual estos contribuyen al logro del propósito de la formación profesional. Estos elementos incluyen aspectos de distintas dimensiones, entre ellas: sociodemográfica, económica, académica y de infraestructura; cuya convergencia de resultados motiva la mejora en los procesos de calidad tanto de las IES como de sus programas académicos en busca de óptimos desempeños académicos de sus estudiantes.

Los resultados se constituyen en un llamado de atención a las autoridades educativas del país para afinar la política pública sobre formación profesional, orientando a las IES en el corto y mediano plazo al mejoramiento de sus procesos de calidad, particularmente de la acreditación, cuyo logro incentiva la obtención de competencias de sus estudiantes y un alto desempeño tanto académico como en la parte laboral.

Como trabajos futuros se plantea complementar este estudio utilizando técnicas descriptivas de minería de datos, como asociación y agrupación, que permitan identificar qué atributos se presentan juntos asociados al desempeño académico

en las pruebas Saber Pro y cómo se agrupan los individuos de acuerdo con su rendimiento en dichas pruebas; así mismo, queda pendiente el estudio sobre el desempeño académico en las pruebas Saber Pro en programas de formación tecnológica con técnicas de minería de datos.

Además, sería recomendable realizar estudios sobre análisis relacional entre las pruebas Saber Once, el desempeño académico en las IES en la formación profesional y las pruebas Saber Pro; así como analizar la relación que pudiera existir entre el número de estudiantes y el número de IES en una zona geográfica con el desempeño académico. De igual manera, quedan pendientes investigaciones comparadas sobre desempeño académico entre diferentes modalidades de estudio, entre otras iniciativas.

Bibliografía general

- Achaerandio, L. (2010). *Competencias fundamentales para la vida*. Guatemala: Universidad Rafael Landívar.
- Agrawal, R., Mehta, M., Safer, J. y Srikant, R. (1996). The Quest Data Mining System. *Second Conference KDD y Data Mining*. Portland, Oregon, Estados Unidos.
- Agrawal, R. y Srikant, R. (1994). Fast Algorithms for Mining Association Rules. *VLDB Conference*, Santiago de Chile.
- Agrawal, R. y Srikant, R. (1995). Mining Sequential Patterns. *The 11th International Conference on Data Engineering ICDE*, Taipei, República de China.
- Agrawal, R., Ghosh S., Imielinski, T., Iyer, B. y Swami, A. (1992). An Interval Classifier for Database Mining Applications. *Proceedings VLDB Conference*, Vancouver.
- Aguila, V. (2005). El concepto: calidad en la educación universitaria, clave para el logro de la competitividad institucional. *Revista Iberoamericana de Educación*, 35. Recuperado de <http://www.rieoei.org/deloslectores/880Aguila.PDF>
- Álvarez, J. (2008). Evaluar el aprendizaje en una enseñanza centrada en las competencias. En J. Gimeno (Comp.). *Educación por competencias. ¿Qué hay de nuevo?* (pp. 206-234). Madrid: Morata.
- Artunduaga, M. (2008). *Variables que influyen en el rendimiento académico en la Universidad*. Recuperado de <http://es.slideshare.net/1234509876/variables-del-rendimiento-academico-universidad>
- Azevedo, A. I. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. Recuperado de <http://hdl.handle.net/10400.22/136>
- Barraza M. (2007). El estrés de examen. *Revista Electrónica de Psicología Científica*, 9(32). Recuperado de <http://www.psicologiacientifica.com/bv/psicologia-306-1-el-estres-de-examen.html>
- Beltrán, M. (2011). *Propuesta de lineamientos para la formación por competencias en educación superior*. Bogotá: Norma.
- Beneitone, P., Esquetine, C., González, J., Marty, M., Siufi, G. y Wagenaar, R. (eds.) (2007). *Reflexiones y perspectivas de la educación superior en América. Informe final Proyecto Tuning - América Latina 2004-2007*. Bilbao: Universidad de Deusto y Universidad de Groningen.

- Benítez, M., Giménez, M. y Osicka, R. (2000). *Las asignaturas pendientes y el rendimiento académico: ¿existe alguna relación?* Recuperado de <http://www1.unne.edu.ar/cyt/humanidades/h-009.pdf>
- Birch, E. y Miller, P. (2006). Student Outcomes at University in Australia: a Quantile Regression Approach. *Australian Economic Papers*, 45(1), 1-17.
- Bodensiek, A. (2010). *Estudio sobre los factores que influyen en el rendimiento escolar*. Bogotá: Secretaria de Educación Municipal.
- Bricall, J. (2000). *Informe Universidad 2 mil: Comisión de Rectores de las Universidades Españolas*. Madrid: Centro regulatorio de Urgencias y Emergencias [CRUE].
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. y Zanasi A. (1998). *Discovering Data Mining from Concept to Implementation*, Prentice Hall. Recuperado de <http://dl.acm.org/citation.cfm?id=270298>
- Cano, M. (2008). La evaluación por competencias en la educación superior. *Revista del currículum y formación del profesorado*, 12(3), 1-16.
- Cantaluppi, F. (2006). Rendimiento académico y abandono en la educación superior a distancia. *Servicio de Educación a Distancia al Exterior* [Seade]. Recuperado de http://www.colegiomilitar.mil.ar/rediu/pdf/ReDiU_0312_art1-Rendimiento_Academico_Parte2.pdf
- Castaño, E., Gallón, S., Gómez, K. y Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de economía*, 60(60), 39-65.
- Climént, J. (2014). Origen, desarrollo y declive de las competencias individuales en tiempos de incertidumbre. *Educación y Educadores*, 17(1), 149-168.
- Cohen, J. (1988). *Análisis de poder estadístico para las Ciencias del comportamiento* (2da. ed.). Nueva Jersey: Lawrence Erlbaum.
- Coleman, J. (1966). *Equality of educational opportunity*. Washington: US Government Printing Office.
- Colombia. Departamento Nacional de Planeación. Sistema General de Regalías SGR (2012). Bogotá: DNP.
- Colombia, Congreso de la República. (29 de diciembre de 1992). Ley Reglamentaria de la Educación Superior, “Ley 30 de 1992”, Por la cual se organiza el servicio público de la educación superior. Diario Oficial N.º 40.700, Bogotá, Colombia.
- Colombia, Consejo Nacional de Acreditación [CNA]. (2013). *Lineamientos para la acreditación de programas de pregrado*. Bogotá: autor.
- Colombia, Consejo Nacional de Acreditación [CNA]. (s. f.). *Acreditación Institucional*. Recuperado de <http://www.cna.gov.co/1741/article-186371.html>

- Colombia, Ministerio de Educación Nacional [MEN] (2009). “Decreto 3963”, Por el cual se reglamenta el Examen de Estado de la Educación Superior. Recuperado de http://www.mineduacion.gov.co/1621/articles-205955_archivo_pdf_decreto3963.pdf
- Colombia, Ministerio de Educación Nacional [MEN] (2009). Competencias genéricas en la educación superior. *Revolución Educativa. Colombia Aprende. Boletín Informativo*, (13). Recuperado de http://www.mineduacion.gov.co/1621/articles-92779_archivo_pdf_Boletin13.pdf
- Colombia, Ministerio de Educación Nacional [MEN] (s. f.). *Propuesta de lineamientos para la formación por competencias en educación superior*. Recuperado de http://www.mineduacion.gov.co/1621/articles-261332_archivo_pdf_lineamientos.pdf
- Colombia, Ministerio de Educación Nacional [MEN] (2013). *Competencias TIC para el desarrollo profesional docente*. Bogotá: Ministerio de Educación Nacional, Colombia Aprende.
- Colombia, Ministerio de Educación Nacional [MEN] (2015). La acreditación en Colombia: cifras y estado actual. Disponible en: http://www.mineduacion.gov.co/cvn/1665/articles-348747_archivo_pdf.pdf
- Colombia, Presidencia de la República. (2012). “Decreto 1075-2012”, Por el cual se reglamenta la organización y funcionamiento de los Órganos Colegiados de Administración y Decisión y las secretarías técnicas, de acuerdo con lo establecido en el artículo 6° de la Ley número 1530 de 2012. Recuperado de <http://www.alcaldiabogota.gov.co/sisjur/normas/Normal.jsp?i=47530>
- Colombia, Sistema Nacional de Acreditación [SNA]. (2015). Lineamientos para la acreditación institucional. *Consejo Nacional de Acreditación [CNA]*. Recuperado de http://www.cna.gov.co/1741/articles-186359_Lin_Ins_2014.pdf
- Constitución Política de Colombia de 1991. *Reforma de 1997*. Recuperado de http://www.unesco.org/culture/natlaws/media/pdf/colombia/colombia_constitucion_politica_1991_spa_orof.pdf
- Contreras, K., Caballero, C., Palacios, J. y Pérez A. (2008). Factores asociados al fracaso académico en estudiantes universitarios de Barranquilla. *Psicología desde el Caribe*, 22. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-417X2008000200008
- Chaparro, F (2008). *Evolución de la acreditación de alta calidad en Colombia 1998-2008*. Bogotá: Consejo Nacional de Acreditación [CNA].
- Chapman, P, Clinton, J., Randy, K., Khabaza, T., Reinartz, T. et al. (2000). CRISP-DM 1.0 Step-by-Step Data Mining Guide. Recuperado de <http://www.crisp-dm.org/CRISPWP-0800.pdf>

- Chen, M., Han, J. y Yu, P. (1996). Data Mining: An Overview from Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*.
- Cohen, J. (1988). *Análisis de poder estadístico para las Ciencias del comportamiento* (2da. ed.). Nueva Jersey: Lawrence Erlbaum.
- Chicaiza, S., Galvis, D. y Ramírez, A. (2010). Determinantes del rendimiento académico en Colombia: pruebas ICFCES Saber 11°, 2009-1. *Revista Universidad EAFIT*, 46(160), 48-72. Recuperado de <http://publicaciones.eafit.edu.co/index.php/revista-universidad-eafit/article/view/754>
- Davidson, M. y McKinney, G. (2001). Quantitative Reasoning: An Overview. *Western Washington University*. Recuperado de <http://www.wwu.edu/vpue/documents/issue8.pdf>
- Delors, J. (1999). *La educación encierra un tesoro: Informe a la Unesco de la Comisión Internacional sobre la Educación para el siglo XXI*. París: Santillana, Unesco.
- Di Gresia, L., Fazio, A., Porto, L., Ripani, W. y Sosa, E. (2005). Rendimiento y productividad de los estudiantes. El caso de las universidades públicas argentinas. En A. Porto (ed.), *Economía de la Educación Universitaria: Argentina-Brasil-Perú* (pp. 8-27). La Plata: Editorial de la Universidad Nacional de La Plata.
- Elder, J. y Pregibon, D. (1996). A Statistical Perspective on Knowledge Discovery in Databases. En *Advances in Knowledge Discovery and Data Mining, AAAI Pres/ The MIT Press*.
- El Proyecto Tuning-América Latina 2004-2007. (2009). Reflexiones y perspectivas de la educación superior en América Latina. Informe Final. *Universidad de Deusto y Universidad de Groningen*. Recuperado de <http://tuning.unideusto.org/tuningal>.
- Espindola, E. y León, A. (2002). La deserción escolar en América Latina: un tema prioritario para la agenda regional. *Revista Iberoamericana de Educación*, 30. Recuperado de <http://www.rieoei.org/rie30a02.htm>
- Fayyad, U., Piasteky-Shapiro, G. y Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(11), 27-34.
- Febles, J. y González, A. (2002). Aplicación de la minería de datos en la bioinformática. *Revista Cubana de los Profesionales de la Información y la Comunicación en Salud Acimed*, 10(2). Recuperado de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352002000200003#cargo
- Fernández, A. (2010). La evaluación orientada al aprendizaje en un modelo de formación por competencias en la educación universitaria. *Revista de Docencia Universitaria*, 8(1), 11-34.
- Fernández, G. (2009). *Extracción de Información de la Web usando Técnicas de Minería de Datos*. Recuperado de <http://www.tdg-seville.info/Download.ashx?id=48>
- Fernández, N. (2006). La evaluación y la acreditación de la calidad: situación, tendencias y perspectivas. En *IESALC. Informe sobre la educación superior en América Latina y*

- el Caribe 2000-2005: la metamorfosis de la educación superior* (pp. 33-42). Venezuela: Metrópolis C.A.
- Galand, B., Frenay, M. y Bourgeois, E. (2004). Facteurs de réussite en première de candidature. *Journée d'étude Chaire Unesco de Pédagogie Universitaire*.
- Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde calidad de la educación superior pública. *Revista Educación*, 31(1), 43-63.
- Garbanzo, G. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare*, 18(1), 119-154.
- García, M. y Álvarez, A. (2010). Análisis de Datos en WEKA –Pruebas de Selectividad. Recuperado el 5 de mayo de 2013, de <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>
- García, M. y Aránzazu, A. (2010). Análisis de datos en WEKA–pruebas de selectividad. Recuperado de <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>
- Gill, H. y Rao P. (1996). *Data warehousing: la integración de información para la mejor toma de decisiones*. México: Prentice-Hall Hispanoamericana.
- Gómez, G. y Soares, A. (2013). Diferencias de género con relación al desempeño académico en estudiantes de nivel básico. *Alternativas en Psicología*, xvii(28), 106-118.
- Gómez, M., Rodríguez, G. e Ibarra, M. (2013). Compes: auto informe sobre las competencias básicas relacionadas con la evaluación de estudiantes universitarios. *Estudios sobre Educación*, 24, 197-224. doi: 10.15581/004.24.197-224.
- González, J. y Wagenaar, R. (eds.) (2003). Structures in Europe. Informe final Proyecto Piloto-Fase 1. Bilbao: Universidad de Deusto, Universidad de Groningen y Comisión.
- González, V. y González, R. (2008). Competencias genéricas y formación Profesional: un análisis desde la docencia Universitaria. *Revista Iberoamericana de Educación*, 47. Recuperado de <http://www.rieoei.org/rie47a09.htm>
- Hall, M., Frank, E. y Witten, I. (2011). Practical Data Mining: Tutorials. *University of Waikato*. Recuperado de www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf
- Han, J. y Kamber, M. (2001). *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Hernández, E. y Lorente, R. (2009). *Minería de datos aplicada a la detección de Cáncer de Mama*. Madrid: Universidad Carlos III. Recuperado de <http://tps5to-utn-frre.googlecode.com/svn/trunk/BI/Cancer%20de%20Mama/14.pdf>
- Hernández Arteaga, I., Alvarado, J. C. y Luna, J. A. (2015). Responsabilidad social en la relación universidad-empresa-Estado. *Educación y Educadores*, 18(1), 95-110. doi: 10.5294/edu.2015.18.1.6

- Hernández Arteaga, I. y Hernández Ramírez, I. (2013). Responsabilidad de la universidad en la trilogía: Universidad-Empresa-Estado. En I. Hernández-Arteaga y L.S. Pemberthy-Gallo (comps.), *universidad-empresa-Estado: hacia la cultura de la investigación y la innovación. Segunda Rueda de Negocios de Innovación en Cauca y Nariño* (pp. 23-43). Bogotá: Editorial Universidad Cooperativa de Colombia.
- Hernández Arteaga, I. y Hernández Ramírez, I. (2014). Responsibility of the University in the University-Company-State Triad. En I. Hernández Arteaga y L. S. Pemberthy-Gallo (Comps.), *University-Company-State: Towards a Culture of Research and Innovation. Second Conference of Business Innovation in Cauca and Nariño* (pp. 39-58). Bogotá: Editorial Universidad Cooperativa de Colombia. DOI: <http://dx.doi.org/10.16925/9789587600261>
- Hernández Arteaga, I. y Luna, S.M. (2013). Creación e innovación: estrategia de competitividad regional. En I. Hernández-Arteaga y L. S. Pemberthy-Gallo (Comps.), *Universidad-Empresa-Estado: hacia la cultura de la investigación y la innovación. Segunda Rueda de Negocios de Innovación en Cauca y Nariño* (pp. 187-199). Bogotá: Editorial Universidad Cooperativa de Colombia.
- Hernández Arteaga, I. y Luna, S. M. (2014). Creation and Innovation: A Strategy for Regional Competitiveness. En I. Hernández Arteaga y L. S. Pemberthy-Gallo (Comps.), *University-Company-State: Towards a Culture of Research and Innovation. Second Conference of Business Innovation in Cauca and Nariño* (pp. 192-204). Bogotá: Editorial Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600261>
- Hernández, J., Ramírez, M. y Ferri, C. (2005). *Introducción a la Minería de Datos*. Madrid: Pearson Educación SA.
- Hidalgo de Paz, A., Gil, M. y Rodríguez, E. (2001). Capacidad de atención y resultados docentes de estudiantes de primer año de medicina. *Revista Cubana Educación Médica Superior*, 15(3), 273-278.
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2010). *Exámenes de Estado de calidad de la educación superior. Saber Pro (ECAES). Análisis de resultados del período 2004-2008*. ICFES. Recuperado de <http://54.208.2.57/datos/Informe%20final%20resultados%20SABER%20PRO%202004%22008.pdf>
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2011a). *Examen saber pro noviembre de 2011-II. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. Recuperado de <http://acofartes.org.co/docsweb/documento/ICFES%202011,%20M%C3%93DULOS%20COMPETENCIAS%20GEN%C3%89RICAS%20Y%20ESPEC%C3%8DFICAS.pdf>

- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2011b). *Lineamientos Saber Pro, Noviembre de 2011*. Recuperado de http://www2.icfes.gov.co/examenes/component/docman/doc_view/66-lineamientos-saber-pro-noviembre-2011?Itemid=
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2012a). *Examen Saber Pro, junio de 2012–I. Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior*. ICFES. Recuperado de www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2012b). *Saber Pro: principales resultados en competencias genéricas*. Santa Marta: Recuperado de www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2012c). *Resultados agregados de Saber Pro en los módulos de competencias genéricas, año 2012 programas universitarios*. Recuperado de <http://www.icfes.gov.co/resultados/saber-pro-resultados-individuales/resultados-agregados-saber-pro>
- Instituto Colombiano para la Evaluación de la Educación [ICFES] (2013). *Estructura del Saber Pro*. Recuperado de http://ead.uis.edu.co/empresarial/images/stories/doc/SABER_PRO_2013_ESTRUCTURA.pdf
- Izar, J., Ynzunza, C. y López, H. (2011). Factores que afectan el desempeño académico de los estudiantes de nivel superior en Rioverde, San Luis Potosí, México. *CPU-e Revista de Investigación Educativa*, 12. Recuperado de <https://www.uv.mx/cpue/num12/opinion/completos/izar-desempeno%20academico.html>
- Jano, D. y Ortiz, S. (2005). Determinación de los factores que afectan al rendimiento académico en la educación superior. *XII Jornadas de la Asociación de Economía de la Educación*. Oviedo, Universidad de Oviedo.
- Larose, D. y Larose, Ch. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (2da. ed.). New Jersey: John Wiley & Sons.
- Lorente, R. (2012). *La formación profesional según el enfoque de las competencias. La influencia del discurso europeo en España*. Barcelona: Octaedro.
- Luna, J. y Hernández Arteaga, I. (2010). TIC: una tendencia que transforma el quehacer de la Universidad. *Revista Memorias*, 8(14); 166-184.
- Martínez, A., Cegarra, J. y Rubio, J. (2012). Aprendizaje basado en competencias: una propuesta para la autoevaluación del docente. *Profesorado. Revista del currículum y formación del Profesorado*, 16(2), 373-386.
- Martínez, D. y Podestá, C. (2014). Metodología de estudio del rendimiento académico mediante la minería de datos. *Campus Virtuales*, 3(1), 56-73.

- Metha M., Agrawal R., Rissanen J. (1996). SLIQ: A Fast Scalable Classifier for Data Mining. *Proceedings EDBT96*. Avignon, France.
- Mizala, A. y Romaguera, T. (2002). *Factores que inciden en el rendimiento escolar en Bolivia*. Santiago de Chile: Universidad de Chile.
- Moncada, L. y Rubio M. (2011). Determinantes inmediatos del rendimiento académico en los nuevos estudiantes matriculados en el sistema de educación superior a distancia del Ecuador: caso Universidad Técnica Particular de Loja. *Red Internacional de Educación Docente, RIED*, 14(2), 77-95.
- Montes, I. y Lerner, J. (2012). *Rendimiento académico de los estudiantes de pregrado de la Universidad EAFIT. Perspectiva cuantitativa*. Medellín: Universidad EAFIT.
- Montero, E., Villalobos, J. y Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico y a la repetición estudiantil en la Universidad de Costa Rica: un estudio multinivel. *Revista Relieve*, 13(2), 215-234.
- Mora, J. (2004). La necesidad del cambio educativo para la sociedad del conocimiento, en calidad y acreditación universitaria. *Revista de Educación Iberoamericana*, (35). Recuperado de <http://www.rieoei.org/rie35a01.htm>
- Moreno, J. y Santín, D. (2010). Los determinantes de la eficiencia educativa en la Unión Europea. Instituto de Estudios Fiscales. *Revista de Economía Pública*, 193, 131-156.
- Moro, S., Laureano, R. y Cortez, P. (2011). Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. *Proceedings of European Simulation and Modelling Conference -ESM'2011*, 117-121. Recuperado de http://sci2s.ugr.es/docencia/in/pdf/MoroCortezLaureano_DMAApproach4DirectMKT.pdf
- Navarro, R. (2003). El rendimiento académico: concepto, investigación y desarrollo. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2), 1-15. Recuperado de <http://www.ice.deusto.es/rinace/reice/vol1n2/Edel.pdf>
- Ng, R. y Han, J. (1994). Efficient and Effective Clustering Method for Spatial Data Mining. *VLDB Conference*. Santiago de Chile, Chile.
- Nieto, S. (2008). Hacia una teoría sobre el rendimiento académico en enseñanza empírica. *Teoría de la Educación*, 20, 249-274.
- Orea, S., Vargas, A. y Alonso, M. (2005). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. *Ene*, 779(73), 33.
- Parra, C., Mejía, L., Valencia, A., Castañeda, E. (2012). *Rendimiento académico de los estudiantes de primer semestre de pregrado de la Facultad de Ingeniería de la Universidad de Antioquia:*

- cohorte 2012-2. Medellín: Ingeniería y Sociedad. Recuperado de file:///C:/Users/Aspire/Downloads/16537-56603-1-PB%20(1).pdf
- Peña, L. (2008). *La competencia oral y escrita en la educación superior*. Bogotá: Ministerio de Educación Nacional [MEN]. Recuperado de http://mineduacion.gov.co/1621/articulos-189357_archivo_pdf_comunicacion.pdf
- Pérez, A. (2013). Reválidas, evaluación de competencias y calidad de los aprendizajes. *Revista Curriculum*, 26, 11-25.
- Piatetsky-Shapiro, G., Brachman, R. y Khabaza, T. (1996). *An Overview of Issues in Developing Industrial Data Mining and Knowledge Discovery Applications*. Association for the Advancement of Artificial Intelligence [AAAI], MIT Press. Recuperado de: <http://www.aaai.org/Papers/KDD/1996/KDD96-015.pdf>
- Pinilla, A. (2011). Modelos pedagógicos y formación de profesionales en el área de la salud. *Acta Médica Colombiana*, 36(4), 204-218.
- Porto A., Di Gresia L. y López, M. (2004). *Mecanismos de admisión a la Universidad y rendimiento de los estudiantes*. Recuperado de <http://www.aep.org.ar/espa/anales/resumen04/04/Porto-DiGresia-Armengol.pdf>
- Porto, A. y Di Gresia, L. (2000). *Características y rendimiento de estudiantes universitarios. Caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata*. Documento de Trabajo. N°. 24.
- Pineda, C. y Pedraza, A. (2011). *Persistencia y graduación. Hacia un modelo de retención estudiantil para Instituciones de Educación Superior*. Bogotá: Arfo Editores e Impresores Ltda.
- Quinlan, J. (1993). *C4.5: programs for machine learning. Vol. 1*. Baltimore: Morgan Kaufmann Publishers. Recuperado de <http://books.google.com.co/books?hl=es&lr=&id=HEXncpjbYroC&oi=fnd&pg=PR7&dq=Programs+for+Machine+Leraning&ots=nLkbbRq2Yj&sig=Y5h5CQUdtbZjs1Fjd8ilbJfyRLE>
- Quinlan, J. (1986). Induction of Decision Trees. *Machine Learning Journal*, 1(1), 81-106.
- Raus, N., Vegega, C., Pytel, P. y Pollo-Cattaneo, M. (2014). Metodología propuesta para la predicción de deserción universitaria mediante explotación de información (pp. 1014-1158). *WICC 2014 xvi Workshop de investigadores en ciencias de la computación*. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/43835/Documento_completo.pdf?sequence=1
- Restrepo, B. (2008). *Política pública sobre calidad de la educación superior, y retos de la educación superior hoy*. Medellín: Asiesda.
- Reyes, Y. (2003). *Relación entre el rendimiento académico, la ansiedad ante los exámenes, los rasgos de personalidad, el autoconcepto y la asertividad*. Ministerio de Educación del Perú.

- Universidad Nacional Mayor de San Marcos. Recuperado de http://ciberdocencia.gob.pe/?id=493&a=articulo_completo
- Rodríguez, W. (2010). El concepto de calidad educativa: una mirada crítica desde el enfoque historicocultural. *Revista Electrónica Actualidades Investigativas en Educación*, 10(1), 1-28.
- Ruiz, G. (2009). El enfoque de la formación profesional en torno a la generación de competencia: ¿ejercicio impostergable o “lo que sucedió a un rey con los burladores que hicieron el paño”. *Estudios pedagógicos*, 34(1), 287-299.
- Rubio, M. (2009). Nuevas orientaciones y metodología para la educación a distancia. Loja: Editorial de la Universidad Técnica Particular de Loja.
- Salanova S., Martínez, I., Bresó, E., Llorens, S. y Grau, R. (2005). Bienestar psicológico en estudiantes universitarios. Facilitadores y obstaculizadores del desempeño académico. *Revista Anales de Psicología*, 21(1), 170-180.
- Salas, R. (2000). La calidad en el desarrollo profesional: avances y desafíos. *Educación Médica Super*, 14(2), 136-147.
- Salinas, F. (2010). *Estudio sobre los factores que influyen en el rendimiento escolar*. Bogotá: Secretaría de Educación Distrital. Recuperado de: <http://evaluacion.educacionbogota.edu.co/files/Factores%20que%20influyen%20en%20el%20rendimiento%20escolar.pdf>
- Salinas, J. (2008). *Modelos didácticos en los campus virtuales universitarios: patrones metodológicos generados por los profesores en procesos de enseñanza aprendizaje en entornos virtuales. Informe final del proyecto EA2007-0121*. Recuperado de <http://tecnologiaedu.us.es/tecnoedu/images/stories/EA2007-0121-memoria.pdf>
- Salinas, J. (2006). Flexibilidad en el currículo de la educación superior en el ámbito de las competencias. II *Encuentro Académico, Comisión de Currículo de la Comisión Nacional de Rectores (CONARE)*. Costa Rica.
- Sarramona, J. (2002). *Evaluación de programas de educación a distancia*. Barcelona: Universidad Autónoma de Barcelona.
- Sattler, K. y Dunemann, O. (2001). SQL database primitives for decision tree classifiers. En *Proceedings of the tenth international conference on Information and knowledge management* (pp. 379-386). Atlanta: CIKM. Recuperado de <http://dl.acm.org/citation.cfm?id=502650>
- Shunk, D. (2008). *Learning theories: an educational perspective*. New York: Prentice Hall.
- Seibold, J. (2000). La calidad integral en educación. Reflexiones sobre un nuevo concepto de calidad educativa que integre valores y equidad educativa. *Revista Iberoamericana de Educación*, 23. Recuperado de <http://www.rieoei.org/rie23a07.htm>
- Shafer J., Agrawal R., Metha M. (1996). SPRINT: A Scalable Parallel Classifier for Data Mining. *Proceedings of the VLDB Conference*. Bombay, India.

- Spicker, P., Alvarez, S. y Gordon, D. (s. f.). *hH: Hacinamiento Hambruna*. Recuperado de <http://biblioteca.clacso.edu.ar/gsd/collect/clacso/index/assoc/D9393.dir/h.pdf>
- Srikant R. y Agrawal, R. (1996). *Mining quantitative association rules in large relational tables*, *ACM SIGMOD, Montreal*. Recuperado de <http://rakesh.agrawal-family.com/papers/sigmod-96qassoc.pdf>
- Tejedor, F. y García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario. Propuestas de mejora en el marco del EEEE. *Revista de Educación*, 342, 443-473.
- Timarán, R. (2009). Una mirada al descubrimiento de conocimiento en bases de datos. *Revista Ventana Informática*, 20, 39-58.
- Timarán, R., Calderón, A. y Jiménez, J. (2013a). Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil. *Ventana Informática*, 28. Recuperado de <http://revistasum.umanizales.edu.co/ojs/index.php/ventanainformatica/article/view/181>
- Timarán, R., Calderón, A. y Jiménez, J. (2013b). La minería de datos como un método innovador para la detección de patrones de deserción estudiantil en programas de pregrado en instituciones de educación superior. En *WEEF 2013 Cartagena*. Recuperado de <http://www.acofipapers.org/index.php/acofipapers/2013/paper/view/211>
- Timarán, R. y Millán, M. (2006). New algebraic operators and SQL primitives for mining classification rules. En *Computational Intelligence* [pp. 61-65]. Recuperado de <http://www.actapress.com/PaperInfo.aspx?PaperID=29048&reason=500>
- Tobón, S. Sánchez, A. Carretero, M. y García, J. (2006). *Competencias calidad y educación superior*. Bogotá: Magisterio.
- Tobón, S., Pimienta, J. y García, J. (2010). *Secuencias didácticas: aprendizaje y evaluación de competencias*. México: Pearson.
- Tobón, S. (2010). *Formación integral y competencias: pensamiento complejo, currículo, didáctica y evaluación*. Bogotá: Ecoe.
- Tonconi, J. (2010). Factores que influyen en el rendimiento académico y la deserción de los estudiantes de la Facultad de Ingeniería Económica de la UNA-Puno, Periodo 2009. *Cuadernos de Educación y Desarrollo*, 2(1). Recuperado de <http://www.eumed.net/rev/ced/11/jtq.pdf>
- Toro, J. y Villaveces, J. (2008). *El pensamiento matemático: una competencia genérica emergente*. Recuperado de http://www.mineducacion.gov.co/1621/articles-189357_archivo_pdf_matematica_1B.pdf
- Tünnermann, C. (2008). Introducción. En *La educación superior en América Latina y el Caribe: diez años después de la Conferencia Mundial de 1998*. Bogotá: Pontificia Universidad Javeriana, Unesco e Iesalc.

- Unesco. (1998). *Conferencia mundial sobre la educación superior. La educación superior en el siglo XXI. Visión y acción*. París: Unesco. Recuperado de <http://unesdoc.unesco.org/images/0011/001163/116345s.pdf>
- Umaña, A. (2008). Reflexiones sobre el diseño curricular por competencias en la Universidad Estatal a Distancia de Costa Rica. *Revista Cognición*, 13. Edición especial. II congreso Cread Andes y II Encuentro Virtual Educa UTP, Loja.
- Universidad Nacional de Colombia. (2012). *Análisis de los resultados obtenidos por la Universidad Nacional de Colombia sede Bogotá en las pruebas Saber Pro 2011-2*. Universidad Nacional de Colombia. Recuperado de www.unal.edu.co/diracad/evaluacion/SaberPro_2012/analisis_de_resultados.pdf
- Valero, S. (2009). *Aplicación de técnicas de minería de datos para predecir deserción*. Matamoros: Universidad Tecnológica de Izúcar.
- Valero, S., Salvador, A. y García, M. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. *Ene*, 779(73), 33.
- Valdivieso, M., Monar, K. y Granda, M. (2004). Análisis de los determinantes del rendimiento de los estudiantes de ESPOL-2002. *Revista Tecnológica*, 17(1), 213-218.
- Vásquez, C. y Rodríguez, M. (2007). La deserción estudiantil en educación superior a distancia: perspectiva teórica y factores de incidencia. *Revista Latinoamericana de Estudios Educativos*, xxxvii(3 y 4), 107-122.
- Vélez, A. y Roa, C. (2005). Factores asociados al rendimiento académico en estudiantes de medicina. *Educación Médica*, 8(2), 74-82.
- Villalba, A. y Salcedo, M. (2008). *El rendimiento académico en el nivel de educación media, como factor asociado al rendimiento académico en la universidad* (Tesis de maestría). Sistema de Universidades Estatales del Caribe, Colombia.
- Wang, M., Iyer, B. y Scott, J. (1998). Scalable Mining for Classification Rules in Relational Databases. *International Database Engineering and Application Symposium - Ideas*. Cardiff, Wales.
- Zapata, L. (2011). *Factores Académicos asociados al bajo rendimiento en Inglés en la Pruebas ECAES presentada por los estudiantes de la Facultad de Educación en el año 2009*. Medellín: Fundación Universitaria Luis Amigó.
- Zhang, T., Ramakrishnan, R. y Livny, M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM SIGMOD International Conference on Management of Data*. Montreal, Canadá.

Índice analítico

A

Acreditación, 23, 24, 25, 26, 169, 170, 171

Acreditación de alta calidad, 27, 161

Acreditación de programas, 17, 24, 25, 90, 98, 99, 170, 171

Acreditación institucional, 3, 16, 17, 24, 25, 27, 90, 97, 98, 151, 161, 166, 169, 170, 171

Actitudes, 28

Actividades extracurriculares, 47

Agrupamiento, 63

Algoritmo de la herramienta Weka, 134, 135

Alimentación, 45

Ambiente académico e industrial, 63, 102

Ambiente universitario, 49

Ámbito laboral (ver Campo laboral)

América Latina, 22, 29, 32, 33

Análisis de los datos, 63, 75

Análisis del problema, 63

Aprendizaje, 29, 30, 31, 40, 46, 52, 68, 71, 136, 161, 162

Aprendizaje de máquina, 68, 74, 75

Aptitudes, 40, 171

Árboles de decisión, 3, 16, 68, 69, 70, 81, 101, 134, 135, 136, 138, 139, 143, 152, 154, 156, 158, 161, 166, 167, 169

Área del conocimiento, 35, 50, 157, 158, 166, 171

Asociación, 44, 46, 63, 66, 67, 72, 73, 163, 165, 166, 169, 170, 171

Atributos, 3, 16, 35, 66, 68, 69, 70, 71, 73, 81, 102, 103, 118, 120, 121, 125, 127, 128, 129, 130, 133, 151, 152, 154, 155, 156, 159, 161, 166, 167, 169, 171

Autoevaluación, 22, 23, 24, 25, 170

Autorregulación, 23, 24, 40

B

Bachillerato, 43, 46, 50, 51

Bases de datos, 3, 16, 17, 20, 49, 50, 51, 63, 64, 66, 71, 74, 75, 84, 101, 102, 161, 166, 167, 169, 171

Becas, 47, 50

Estudiantes becados, 46

Bogotá, 44, 45, 46, 126

C

- Cálculo del índice de condición, 128
 - Calidad de la educación
 - superior, 3, 15, 16, 17, 20, 21, 22, 23, 30, 34, 43, 52, 53, 102, 146, 161, 169, 170
- Calificaciones (ver Notas)
- Campo laboral, 26, 28, 47
- Canasta de mercado, 72, 73, 74
 - Análisis de la canasta de mercado, 73, 74
- Capacidad de aprender, 32
- Capital humano, 22, 25, 31
- Caribe, 32, 44, 155, 164
- Centro sur y Pacífico, 164
- Centros de investigación, 31
- Ciencia, 25, 34, 51, 126
 - Ciencias básicas, 53
 - Ciencias de la educación, 53, 157, 164
 - Ciencias humanas, 129, 157, 158, 164
 - Ciencias naturales, 129, 158, 164
 - Ciencias sociales, 129, 157, 164, 165
 - Ciencias técnicas, 129, 158, 164
- Ciudadanía, 33, 34
- Clasificación, 3, 16, 46, 63, 66, 67, 68, 69, 70, 81, 82, 101, 118, 125, 134, 135, 136, 152, 155, 157, 159, 161, 166, 167, 169
- Coertura, 166
- Coefficiente de correlación, 67, 87, 88
 - Análisis de correlación, 88
 - Coefficiente de correlación de Pearson, 88
- Cognitivo, 47, 170
- Colegios privados, 46
- Colegios públicos, 46
- Competencias, 15, 19, 20, 22, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 41, 48, 51, 52, 53, 88, 89, 90, 91, 94, 97, 102, 133, 171
- Evaluación Internacional de Competencias en Educación Superior, 33
- Evaluación por competencias, 40, 41, 52, 171
- Formación por competencias, 28, 30, 31, 40
- Modelo de competencias, 29, 31
- Modelo de formación por competencias, 40
- Competencias específicas, 15, 19, 20, 32, 34, 53, 55, 169, 171
- Competencias genéricas, 3, 15, 16, 17, 19, 20, 32, 33, 34, 35, 43, 48, 52, 53, 55, 87, 88, 89, 90, 92, 97, 101, 102, 126, 133, 134, 151, 161, 162, 166, 167, 169, 170, 171
- Competencia de lectura crítica (ver Lectura crítica)
- Competencia de comunicación escrita (ver Comunicación escrita)
- Competencia de razonamiento cuantitativo (ver Razonamiento cuantitativo)
- Competencia de inglés (ver Inglés)
- Evaluación de los módulos de competencias genéricas, 19, 35, 55

- Comprensión del negocio o problema, 77, 102
- Computación de alto rendimiento, 75
- Computador, 32, 45
- Comunicación escrita, 15, 19, 20, 37, 55, 88, 90, 91, 92, 93, 94, 96, 97, 98, 138, 151, 154, 155, 156, 163, 164, 166, 170
- Comunidad académica (ver Comunidad educativa)
- Comunidad educativa, 25, 33, 161
- Condición del hogar (ver Vivienda)
- Conferencia Mundial sobre Educación Superior, 21
- Conocimientos, 21, 23, 28, 29, 32, 34, 40, 41, 42, 53, 171
- Consejo Nacional de Acreditación (CNA), 23, 24, 25, 161, 169
- Costa Rica, 43
- CRISP-DM (*cross-industry standard process for data mining*), 3, 16, 17, 63, 75, 84, 101, 102
- Currículos, 164, 165
- D** *Data warehousing* (análisis de bodegas de datos), 75
- Data-secuencia*, 74
- Datos sociodemográficos, económicos, académicos e institucionales, 3, 16, 102, 161
- DCBD, 75
- Derecho, 51
- Desarrollo sostenible, 25
- Desempeño académico, 3, 17, 19, 20, 33, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 87, 89, 90, 92, 94, 95, 96, 97, 98, 101, 102, 118, 125, 134, 146, 151, 161, 162, 163, 164, 165, 166, 167, 169, 170, 171, 172
- Desempeño en las competencias genéricas, 43, 89, 90, 92, 97
- Desempeño laboral, 28
- Desigualdades sociales, 163
- Desviación estándar, 87, 89
- Dimensiones sociodemográficas, económicas, académicas e institucionales, 15, 20, 49, 163, 170, 171
- Discretización, 66, 81, 127, 128
- Diseño de proyectos, 32
- Disposiciones cognitivas, socioafectivas y psicomotoras, 28
- Docencia, 23, 24
- Docente, 28, 40, 43, 46, 162, 170
- E** Economía, 158
- Economía global, 31
- Edad, 20, 45, 46, 49, 50, 66, 67, 73, 125, 163
- Educación de América Latina, 22
- Educación media, 42, 43, 44, 45
- Educación primaria, 92, 93
- Educación secundaria (ver Educación media)
- Educación superior, 3, 15, 16, 17, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 33, 34, 42, 43, 44, 45, 46, 47, 52, 53, 92, 93, 102, 146, 161, 166, 169, 170
- Enfoque educativo, 31
- Enfoque empresarial, 31

Enfoque humanista social, 31
 Enseñanza media (ver Educación media)
 Entendimiento interpersonal, 15, 34
 Entropía, 66, 69, 70
 Escolaridad, 44, 46, 92
 Esfuerzo del estudiante, 42, 48
 Espacios de aprendizaje, 29
 Espacio Europeo de Educación Superior (EEES), 43
 Estadística, 3, 16, 45, 65, 74, 75, 79, 87, 88, 89
 Estadístico *d* de Cohen, 87, 89, 90, 91, 92, 94, 96, 97, 98
 Estado, 21, 22, 23, 24, 25, 45, 49, 51, 52, 53, 64, 170
 Estado civil, 49, 50, 67, 90, 163
 Estados Unidos, 41
 Estrato social del estudiante (ver Estrato socioeconómico)
 Estrato socioeconómico, 46, 50, 90, 94, 166, 171
 Estudiantes de programas profesionales, 3, 16, 19, 26, 34, 42, 43, 88, 102, 134, 146, 152, 153, 154, 155, 156, 159, 161, 169
 Estudiantes universitarios (ver Estudiantes de programas profesionales)
 Estudios de primaria (ver Educación primaria)
 Estudios profesionales (ver Educación superior)
 Estudios universitarios (ver Educación superior)
 Ética, 33

Europa, 32
 Evaluación de instituciones, 20, 24, 25
 Evaluación de la calidad institucional, 26
 Evaluación de la educación, 21, 25, 29, 52, 53
 Evaluación de resultados, 37, 82, 83, 152, 154, 156, 158
 Evaluación e interpretación de los patrones, 146
 Examen de Estado para el Ingreso a la Educación Superior, 45
 Exámenes de Estado de Calidad de la Educación Superior (Ecaes), 52
 Explotación, 63, 82

F

Factor de confianza, 72, 73, 87, 89, 134, 136, 138, 139, 143, 151, 152, 155, 157, 159, 161
 Factores sociodemográficos, económicos, académicos e institucionales, 134, 152, 154, 156, 159, 167
 Facultad, 67, 162
 Familia, 20, 42, 44, 46, 47, 48, 49, 50, 158, 163
 Financiación de los estudios universitarios, 46, 50
 Formación a distancia, 95, 152, 155, 157, 159, 162, 163, 164, 165, 170
 Formación integral, 25, 28, 30, 52, 53
 Formación profesional, 20, 26, 27, 28, 29, 30, 31, 32, 35, 41, 47, 48, 50, 51, 102, 164, 169, 171, 172

Fortalezas y debilidades, 41

Función social, 24, 33, 51

G

Ganancia de información,
69, 70, 120, 134

Género, 20, 43, 44, 46, 49, 50, 67, 90,
155, 156, 158, 164, 166, 169, 171

Grupo familiar, 49, 50, 163

Grupos de afinidades, 71

H

Habilidades, 23, 28, 29, 32,
33, 40, 41, 42, 43

Habilidades interpersonales, 32

I

Identidad, 33

Implementación, 34, 53, 71,
83, 146, 164, 170

Informe de Bricall, 22

Infraestructura, 20, 45, 46, 51, 52,
171

Ingeniería, 53, 162

Inglés, 15, 19, 20, 34, 35, 39, 55, 64,
88, 90, 91, 93, 94, 96, 97, 98, 101,
143, 151, 158, 159, 160, 165, 166,
170

Institución de educación superior

(IES), 20, 21, 22, 23, 27, 29, 47, 51,
52, 53, 125, 126, 127, 153, 154,
161, 163, 164, 166, 169, 170, 171,
172

IES acreditadas, 27, 97, 152, 153,
154, 155, 156, 157, 158, 159,
160, 161, 162, 163, 164, 165

IES oficiales, 26, 27, 96

IES privadas, 23, 26, 27, 96, 158

Instituto Colombiano para la

Evaluación de la Educación (ICFES),
3, 15, 16, 17, 19, 20, 30, 32, 35, 45,
46, 49, 50, 51, 52, 53, 88, 101, 102,
120, 125, 146, 161, 167, 169, 171

Bases de datos del ICFES, 16, 17, 20,
50, 101, 102, 161, 167, 171

Inteligencia artificial, 74

Interacción social, 24

Interpretación de los resultados, 82

Interpretación/evaluación de datos,
37, 63, 64, 67, 76, 89

Intervalos de confianza, 87, 89

Investigación, 3, 16, 17, 19, 20, 23, 24,
25, 31, 32, 41, 42, 43, 44, 45, 46,
47, 48, 49, 50, 52, 64, 74, 84, 102,
121, 125, 126, 134, 151, 161, 162,
165, 166, 169, 171, 172

J

Jornada de estudio, 45

K

K-nearest neighbour o razona-
miento basado en caos (CBR),
81

L

Lectura crítica, 15, 19, 20,
35, 55, 88, 90, 91, 92, 93,
94, 97, 98, 101, 138, 151,
152, 153, 154, 163, 166, 170

Lenguaje, 45, 165

Lenguaje escrito, 37

M

Manejo de la información, 34
 Marco Común Europeo, 39
 Matemáticas, 37, 45, 164, 165
 Matriz de confusión, 136, 138, 139, 143, 152, 155, 157, 159
 Matrices de confusión Edelstein, 82
 Media, 66, 87, 89, 92, 97, 98, 51, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 166, 170
 Medidas estadísticas, 87, 89
 Mejoramiento continuo, 24, 25
 Mercado laboral, 29
 México, 44
 Minería de datos, 3, 16, 17, 63, 64, 66, 67, 75, 76, 77, 78, 80, 81, 84, 120, 121, 125, 130, 134, 166, 167, 169, 171, 172
 Ministerio de Educación Nacional (MEN), 15, 24, 26, 27, 28, 33, 34, 52, 165, 167
 Misión, 23, 25, 27, 29, 30
 Modalidad a distancia (ver Formación a distancia)
 Modalidad de enseñanza, 20, 43
 Modalidad de estudio, 3, 16, 90, 95, 151, 161, 162, 163
 Modalidad de formación, 32, 48, 162, 166
 Modalidad presencial, 65, 153, 154, 155, 156, 157, 158, 159, 160, 162, 163, 164, 165, 170
 Modalidad semipresencial, 162
 Modalidad virtual, 162
 Modelado, 63, 76, 80, 81, 134

Modelo descriptivo, 47, 66, 67
 Modelo Logit Ordenado Generalizado, 45
 Modelo pedagógico, 33
 Modelos de investigación, 47
 Modelos predictivos, 66, 169
 Muestra, 40, 75
 Error de muestra parcial 135, 136
 Muestra representativa, 65
 Muestreo, 66

N

Necesidades de la sociedad, 33, 34, 40
 Nivel cultural, 163
 Nivel de Sisbén, 50, 153, 154, 159, 160, 163, 165, 166, 167
 Nivel educativo de los padres, 20, 45, 46, 49, 50, 90
 Nivel socioeconómico, 44, 45, 165
 Notas, 43, 46, 47
 Número de instituciones en la zona, 166, 171

O

OLAP (*on-line analytical processing*), 75
 Organización para la Cooperación y el Desarrollo Económico (OCDE), 32

P

Pares académicos, 24, 25
 Patrones de desempeño académico, 3, 16, 19, 20, 48, 49, 101, 102, 125, 134, 138, 139, 143, 151, 161, 166, 169

- Patrones de rendimiento académico
(ver “patrones de desempeño académico”)
- Patrones secuenciales, 63, 66, 67, 74
- Pedagogía, 30, 170
Pedagogía transversal y transdisciplinar, 32
- Pensamiento crítico, 15
- Pensamiento matemático, 34, 164, 165
- Personas a cargo, 153, 154, 160, 163, 165, 166, 171
- Política educativa, 17, 41, 45
- Política pública de educación superior, 24, 25, 27, 52, 53, 161, 166, 169, 170, 171
- Posgrado, 92, 93
- Pregrado, 15, 19, 20, 46, 52, 101, 162
- Preparación de los datos, 63, 76, 80, 120
- Preprocesamiento/limpieza de datos, 64, 65, 66, 167, 171
- Problema objeto de la investigación, 102, 134
- Proceso de aprendizaje, 19, 29, 31, 42, 45, 47, 48, 51, 68
- Proceso de calidad, 22
- Proceso KDD, 63, 64, 65, 74, 75
- Producción de conocimientos, 21
- Programa acreditado, 25, 98, 153, 154, 156, 158, 163, 164
Programas acreditados alto calidad, 26
- Promedio, 46, 67, 87, 89, 90, 91, 92, 94, 96
- Promedio de notas, 67
- Proyecto Tuning, 29, 32, 33
- Prueba de inteligencia emocional, 43
- Pruebas Saber Once, 172
- Pruebas Saber Pro, 15, 16, 19, 20, 34, 51, 52, 53, 54, 87, 126, 146, 164, 169, 170, 172
- Pruebas Saber Pro 2011-2, 3, 16, 17, 19, 20, 35, 48, 49, 50, 51, 55, 87, 88, 89, 90, 102, 120, 134, 138, 139, 143, 151, 152, 154, 156, 159, 161, 162, 163, 166, 167, 169
- R**
- Rango de valores esperados, 65
- Razonamiento basado en caos (CBR), 81
- Reconocimiento de patrones, 74, 75
- Regiones geográficas (ver Zonas geográficas)
- Registraduría Nacional, 167
- Reglas de asociación, 67, 72, 73
- Rendimiento académico, 41, 42, 43, 44, 45, 46, 125, 126, 133, 164, 166, 169, 170
- Resultados académicos, 52
- Retroalimentación, 40, 82
- S**
- Saber, 30, 31, 33, 40
Saber aprender, 34
Saber cómo, 31
Saber conocer, 40
Saber convivir, 30, 40
Saber hacer, 30, 40
Saber recontextualizar, 34
Saber ser en sociedad, 34

Secretaría de Educación Distrital de Bogotá, 45

Sector académico, 15, 26, 51

Sector oficial o privado, 43

Sector productivo, 31

Segmentación o *clustering*, 66, 67, 70, 71, 73

Cluster, 66, 67, 70, 71

 Segmentación de mercado, 71

Selección de datos, 64, 65, 67, 80, 88, 101, 102, 121, 169

Servicio educativo, 21, 22, 25, 27, 47, 51, 52

Servicio público, 22, 23, 51, 52

Sexo (ver Género)

Sistema de Universidades Estatales del Caribe Colombiano (SUE-Caribe), 44

Sistema educativo, 23, 29, 30, 40, 51, 53, 163

Sistema General de Regalías (SGR), 125, 126

Sistema Nacional de Acreditación (SNA), 23, 24

Sistemas expertos, 74

Solución de problemas, 15, 31, 34, 40

T

Talento humano (ver “capital humano”)

Tecnología, 25, 34, 126, 162, 163

 Tecnologías de la información y la comunicación (TIC), 34, 45

Tendencia estadística del desempeño, 88

Tipo de institución, 51, 90, 96

Toma de decisiones, 3, 16, 31, 32, 41, 53, 146

Trabajo en equipo, 32, 34, 161

Transformación/reducción de datos, 64, 66, 81, 120, 121, 167, 171

Transporte, 45, 50, 129

U

Universidad, 20, 21, 22, 23, 25, 25, 27, 28, 29, 30, 31, 33, 34, 42, 43, 44, 45, 46, 47, 50, 51, 126, 153, 154, 161, 162

Universidad colombiana, 23, 24

Universidad de Antioquia, 162

Universidad de Costa Rica, 43

Universidad de Salamanca, Tejedor y García-Valcárcel, 43

Universidad del Rosario, 44

Universidad EAFIT, 47

Universidad Nacional de Colombia, 15

Uso de lenguaje cuantitativo, 15

V

Validación cruzada, 135, 136

Valores, 21, 23, 29, 40, 41, 51

Variables sociodemográficas, económicas, académicas e institucionales, 42, 87, 88, 89

Visualización de datos, 75, 80

Vivienda, 50, 128, 129

Z

Zonas culturales, 125

Zonas de los Órganos Colegiados de
Administración y Decisión (OCAD),
125, 126, 153, 154, 155, 156
OCAD de Ciencia Tecnología e
Innovación, 125

Zonas geográficas, 125, 126, 163, 164,
172
Zonas político-administrativas, 125

Este libro se terminó de imprimir y encuadernar
en Xpress Estudio Gráfico y Digital en abril de 2016.

Fue publicado por el Fondo Editorial
de la Universidad Cooperativa de Colombia. Se
emplearon las familias tipográficas ITC Berkeley
Oldstyle Std, Myriad Pro y Zapf Humanist 601 BT.

This book was printed and bound
by Xpress Estudio Gráfico y Digital, in April 2016.
It was published by the Universidad Cooperativa de Colombia Press.
ITC Berkeley Oldstyle Std, Myriad Pro and
Zapf Humanist 601 BT typeface families were used.

Este libro se propone hacer un aporte teórico y práctico que incida en la formulación e implementación de políticas para la medición de la calidad de la educación superior en Colombia, a partir del descubrimiento de patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales en las pruebas Saber Pro 2011-2, utilizando técnicas de minería de datos. Se trata de un estudio transdisciplinario en el que intervinieron investigadores expertos en el campo de la educación, la ingeniería y las ciencias básicas, avalados por la Universidad de Nariño y la Universidad Cooperativa de Colombia. La investigación fue financiada por el Instituto Colombiano para la Evaluación de la Educación (ICFES).



Universidad Cooperativa
de Colombia