

Aplicación de la Minería de Datos en la Detección de Patrones de Desempeño Académico en las Pruebas Saber Pro



Silvio Ricardo Timarán Pereira
Segundo Javier Caicedo Zambrano
Arsenio Hidalgo Troya



Editorial
Universidad de **Nariño**



Editorial
Universidad de **Nariño**

APLICACIÓN DE LA MINERÍA DE DATOS EN LA
DETECCIÓN DE PATRONES DE DESEMPEÑO
ACADÉMICO EN LAS PRUEBAS SABER PRO

Silvio Ricardo Timarán Pereira
Segundo Javier Caicedo Zambrano
Arsenio Hidalgo Troya

APLICACIÓN DE LA MINERÍA DE DATOS EN LA
DETECCIÓN DE PATRONES DE DESEMPEÑO
ACADÉMICO EN LAS PRUEBAS SABER PRO



Editorial
Universidad de **Nariño**

Timarán Pereira, Silvio Ricardo

Aplicación de la minería de datos en la detección de patrones de desempeño académico en las pruebas Saber Pro/ Silvio Ricardo Timarán Pereira, Segundo Javier Caicedo Zambrano, Arsenio Hidalgo Troya. 1ª ed. San Juan de Pasto: Editorial Universidad de Nariño, 2021
200p.: fig, tab.

Incluye Referencias

ISBN: 978-958-5123-43-4

1. Minería de datos - Educación Superior - Udenar 2. Pruebas Saber - Patrones de desempeño - Udenar 3. Minería de datos (metodología CRISP-DM) 4. Saber Pro (pruebas de educación 2012 al 2014) - Udenar.

378.00287 T 582 – SCDD-Ed.22 Biblioteca Alberto Quijano Guerrero

Aplicación de la minería de datos en la detección de patrones de desempeño académico en las Pruebas Saber Pro

© Silvio Ricardo Timarán Pereira
Segundo Javier Caicedo Zambrano
Arsenio Hidalgo Troya

© Editorial Universidad de Nariño

ISBN: 978-958-5123-43-4

Primera edición

Impresión:

Graficolor Pasto SAS

Calle 18 No. 29-67 Tel. 7310652

graficolorpasto@hotmail.com

Marzo de 2021

San Juan de Pasto, Nariño, Colombia

Prohibida la reproducción total o parcial, por cualquier medio o con cualquier propósito, sin autorización escrita de los autores o de la Editorial Universidad de Nariño.

Agradecimientos

Al Sistema de Investigaciones de la Universidad de Nariño por financiar esta investigación y la publicación de este libro.

CONTENIDO

RESUMEN DE CONCEPTOS	15
PRUEBAS DE ESTADO	15
SABER PRO	15
COMPETENCIAS	16
COMPETENCIAS GENÉRICAS	16
COMPETENCIAS ESPECÍFICAS	17
MINERÍA DE DATOS	17
TAREAS DE MINERÍA DE DATOS	18
CLASIFICACIÓN	18
SEGMENTACIÓN O CLUSTERING	18
ASOCIACIÓN	18
PATRONES SECUENCIALES	19
DESEMPEÑO ACADÉMICO	19
CRISP-DM	19
ÁRBOLES DE DECISIÓN	20
EFECTO INSTITUCIONAL EN COMPETENCIAS GENÉRICAS	20
INTRODUCCIÓN	21
CAPÍTULO I	28
ESTUDIOS SOBRE DESEMPEÑO ACADÉMICO Y PRUEBAS SABER PRO	28
CAPÍTULO II	36
LA MINERÍA DE DATOS EDUCATIVA Y CRISP-DM	36
2.1. MINERÍA DE DATOS EDUCATIVA	36
2.2. METODOLOGÍA CRISP-DM	43
CAPÍTULO III	47
EFFECTOS DEL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER PRO	47
3.1. CARACTERÍSTICAS SOCIOECONÓMICAS	49
3.2. CORRELACIÓN ENTRE LAS CINCO COMPETENCIAS GENÉRICAS	52

3.3. EFECTO INSTITUCIONAL EN LAS COMPETENCIAS GENÉRICAS EN EL CONTEXTO NACIONAL	53
3.3.1. Efecto Institucional en Comunicación Escrita	54
3.3.2. Efecto Institucional en Inglés.	61
3.3.3. Efecto Institucional en Lectura Crítica	67
3.3.4. Efecto Institucional en Razonamiento Cuantitativo	73
3.3.5. Efecto Institucional en Competencias Ciudadanas.	79
3.4. EFECTO DE LOS PROGRAMAS EN LAS COMPETENCIAS GENÉRICAS EN EL CONTEXTO INSTITUCIONAL	85
3.4.1. Efecto del programa en Comunicación Escrita	85
3.4.2. Efecto del programa en Inglés	87
3.4.3. Efecto del programa en Lectura Crítica	89
3.4.4. Efecto del programa en Razonamiento Cuantitativo.	91
3.4.5. Efecto del programa en Competencias Ciudadanas	93
3.5. EFECTO DE LAS VARIABLES SOCIECONÓMICAS EN LAS COMPE- TENCIAS GENÉRICAS EN EL CONTEXTO INSTITUCIONAL	95
3.5.1. Efecto del género	96
3.5.2. Efecto de la edad.	96
3.5.3. Efecto del estado civil.	98
3.5.4. Efecto del estrato socioeconómico.	99
3.5.5. Efecto del hogar actual de residencia	101
3.5.6. Efecto de ser cabeza de familia	101
3.5.7. Efecto de tener trabajo	102
CAPÍTULO IV	104
APLICACIÓN DE CRISP-DM EN LA DETECCIÓN DE PATRONES DE RENDI- MIENTO EN LAS PRUEBAS SABER PRO	104
4.1. COMPRENSIÓN DEL NEGOCIO O PROBLEMA	105
4.2. COMPRENSIÓN DE LOS DATOS	105
4.3. PREPARACIÓN DE LOS DATOS	132
4.4. MODELADO.	146
4.4.1. Clasificación con árboles de decisión	146
4.4.1.1. Descubrimiento de patrones de desempeño en lectura crítica	150
4.4.1.2. Descubrimiento de patrones de desempeño en comunicación escrita	150
4.4.1.3. Descubrimiento de patrones de desempeño en razonamiento cuantitativo.	152

4.4.1.4. Descubrimiento de patrones de desempeño en inglés	152
4.4.1.5. Descubrimiento de patrones de desempeño en competencias ciudadanas	152
4.4.1.6. Descubrimiento de patrones de desempeño general en todas las competencias	158
4.5. EVALUACIÓN	160
4.6. IMPLEMENTACIÓN.	160
CAPÍTULO V	161
INTERPRETACIÓN DE RESULTADOS.	161
5.1. Evaluación e interpretación de resultados para la competencia lectura crítica	161
5.2. Evaluación e interpretación de resultados para la competencia comunicación escrita	165
5.3. Evaluación e interpretación de resultados para la competencia razonamiento cuantitativo	167
5.4. Evaluación e interpretación de resultados para la competencia inglés	169
5.5. Evaluación e interpretación de resultados para la competencia ciudadana	173
5.6. Evaluación e interpretación de resultados para el desempeño general en todas las competencias	176
CAPÍTULO VI	181
DISCUSIÓN DE RESULTADOS Y CONCLUSIONES	181
REFERENCIAS BIBLIOGRÁFICAS.	194

ÍNDICE DE FIGURAS

Figura 1. Relación de la minería de datos con otras disciplinas.	38
Figura 2. Fases de la metodología CRISP-DM	44
Figura 3. Tendencia del Rendimiento en Comunicación Escrita de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.	55
Figura 4. Tendencia del desempeño en inglés de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	61
Figura 5. Tendencia del Rendimiento en Lectura Crítica de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	67

Figura 6.	Tendencia del Rendimiento en Razonamiento Cuantitativo de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.	73
Figura 7.	Tendencia del Rendimiento en Competencia Ciudadana de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.	80
Figura 8.	Uso de la herramienta Yachay DTM.	146
Figura 9.	Configuración de la herramienta Yachay DTM para clasificación	149
Figura 10.	Mejor Árbol para la competencia de Lectura Crítica.	151
Figura 11.	Precisión y matriz de confusión del árbol de Lectura Crítica	152
Figura 12.	Mejor Árbol para la competencia de Comunicación Escrita.	153
Figura 13.	Precisión y matriz de confusión del árbol de Comunicación Escrita	154
Figura 14.	Mejor Árbol para la competencia de Razonamiento Cuantitativo	154
Figura 15.	Precisión y matriz de confusión del árbol de Razonamiento Cuantitativo	155
Figura 16.	Mejor Árbol para la competencia Inglés	156
Figura 17.	Precisión y matriz de confusión del árbol de Inglés	156
Figura 18.	Mejor Árbol para Competencias Ciudadanas.	157
Figura 19.	Precisión y matriz de confusión del árbol de Competencias Ciudadanas	158
Figura 20.	Mejor Árbol para el desempeño general en todas las competencias.	159
Figura 21.	Precisión y matriz de confusión del árbol general de Competencias	160

ÍNDICE DE TABLAS

Tabla 1.	Cambios en las pruebas Saber Pro de 2009 a 2012.	24
Tabla 2.	Características Socioeconómicas de estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	49
Tabla 3.	Programas Académicos de estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.	50
Tabla 4.	Matriz de correlaciones de las Competencias genéricas.	52
Tabla 5.	Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Comunicación Escrita pruebas Saber Pro 2012-2014.	55
Tabla 6.	Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Inglés pruebas Saber Pro 2012-2014	62

Tabla 7.	Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Lectura Crítica pruebas Saber Pro 2012-2014.	68
Tabla 8.	Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Razonamiento Cuantitativo pruebas Saber Pro 2012-2014	74
Tabla 9.	Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Competencias Ciudadanas pruebas Saber Pro 2012-2014	80
Tabla 10.	Clasificación por Programas Académicos de la Universidad de Nariño en Comunicación Escrita de las pruebas Saber Pro 2012-2014	86
Tabla 11.	Clasificación por Programas Académicos de la Universidad de Nariño en Inglés de las pruebas Saber Pro 2012-2014	88
Tabla 12.	Clasificación por Programas Académicos de la Universidad de Nariño en Lectura Crítica de las pruebas Saber Pro 2012-2014	90
Tabla 13.	Clasificación por Programas Académicos de la Universidad de Nariño en Razonamiento Cuantitativa de las pruebas Saber Pro 2012-2014	92
Tabla 14.	Clasificación por Programas Académicos de la Universidad de Nariño en Competencias Ciudadanas de las pruebas Saber Pro 2012-2014	94
Tabla 15.	Género y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	96
Tabla 16.	Edad y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	97
Tabla 17.	Estado Civil y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	98
Tabla 18.	Estrato Socioeconómico y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	100
Tabla 19.	Hogar actual y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	101
Tabla 20.	Cabeza de familia y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.	102
Tabla 21.	Trabaja y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014	102
Tabla 22.	Repositorios pruebas Saber Pro en el período 2012-2014	106
Tabla 23.	Análisis de atributos de los repositorios pruebas Saber Pro 2012-2014 .	106
Tabla 24.	Diccionario de datos del repositorio saberpro_udenar_2012_2014	106
Tabla 24a.	Estado Civil	122

Tabla 24b. Origen institución	122
Tabla 24c. Carácter académico	123
Tabla 24d. Código jornada.	123
Tabla 24e. Área del Conocimiento	123
Tabla 24f. Núcleo pregrado	124
Tabla 24g. Grupo de Referencia	125
Tabla 24h. Porcentaje de créditos cursados y aprobados	126
Tabla 24i. Valor anual de la matrícula año anterior	126
Tabla 24j. Educación del padre o madre.	126
Tabla 24k. Ocupación del padre o madre	127
Tabla 24l. Estrato socioeconómico	128
Tabla 24m. Educación de los hermanos mayores.	128
Tabla 24n. Nivel de SISBEN	128
Tabla 24o. Material de los pisos	129
Tabla 24p. Ingresos familiares mensuales	129
Tabla 24q. Trabajo del estudiante	129
Tabla 24r. Lee o habla otro idioma	130
Tabla 24s. Etnia estudiante	130
Tabla 25. Clasificación de atributos en dimensiones	131
Tabla 26. Atributos con porcentaje alto de datos nulos	133
Tabla 27. Atributos con valores constantes y atributos identificadores.	133
Tabla 28. Nuevos atributos del repositorio de datos.	135
Tabla 29. Valores discretizados del atributo estu_rango_edad	138
Tabla 30. Valores del atributo eco_condicion_hogar	139
Tabla 31. Cálculo del índice de condición_hogar	139
Tabla 32. Cálculo del índice de condición_tic	140
Tabla 33. Asignación de valores del atributo area_grupo_referencia.	141
Tabla 34. Diccionario de datos del repositorio final saberpro_minable_udenar	142
Tabla 35. Repositorios por cada competencia.	145
Tabla 36. Patrones destacados en la competencia Lectura Crítica	182
Tabla 37. Patrones destacados en la competencia Comunicación Escrita	184
Tabla 38. Patrones destacados en la competencia Razonamiento Cuantitativo.	185
Tabla 39. Patrones destacados en la competencia inglés	187
Tabla 40. Patrones destacados en Competencias Ciudadanas	1187

RESUMEN DE CONCEPTOS

PRUEBAS DE ESTADO

Según la Constitución Política de Colombia de 1991, en el artículo 67 establece que, 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; así que, corresponde al Estado regular y ejercer la inspección y vigilancia de la educación, con el fin de garantizar su calidad, el cumplimiento de los objetivos, en la perspectiva de mejorar la formación moral, intelectual y física de los colombianos.

La Ley 30 de 1992, establece que es obligación de las Instituciones de Educación Superior, prestar un servicio con calidad. En este contexto, el Ministerio de Educación Nacional - MEN, expide el Decreto 1781 de 2003, que reglamenta los Exámenes de Estado de Calidad de la Educación Superior (ECAES), denominados Saber Pro desde el año 2010.

SABER PRO

El Decreto 1781 de 2003, establece la obligatoriedad de la presentación de los ECAES por parte de los estudiantes de todos los programas de pregrado; 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”.

Desde el año 2008 el MEN emprendió acciones tendientes a la formulación de competencias genéricas o transversales 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 el elemento articulador de todos los niveles educativos.

COMPETENCIAS

Según el Instituto Colombiano para la Evaluación de la Educación -ICFES, las competencias se constituyen en el articulador de los diferentes niveles de formación. 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 diferente naturaleza.

COMPETENCIAS GENÉRICAS

Según el Decreto 3963 de 2009 (MEN, 2009), expedido por el Ministerio de Educación Nacional, las competencias genéricas son aquellas competencias que son necesarias para el adecuado desempeño profesional; 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, las deben desarrollar todas las personas, dependiendo del nivel de formación, sin importar el programa de formación.

Hasta el año 2011, la evaluación Saber Pro incluía cuatro (4) competencias genéricas, a saber: lectura crítica, razonamiento cuantitativo, comunicación escrita, inglés; y a partir de 2012, se adicionó “competencias ciudadanas”; de modo que, ahora, las

competencias genéricas que evalúan las pruebas Saber Pro son: comunicación escrita, razonamiento cuantitativo, lectura crítica competencias ciudadanas y competencias en inglés.

COMPETENCIAS ESPECÍFICAS

Corresponde a competencias que son comunes a grupos de programas que tienen características similares de formación, denominados “grupos de referencia”, que constituyen una clasificación de programas por áreas y núcleos básicos de conocimiento. Por ejemplo, al grupo de referencia “Ingeniería” pertenecen los siguientes programas: Ingeniería de telecomunicaciones, Ingeniería eléctrica, Ingeniería electromecánica, Ingeniería electrónica, Ingeniería en mecatrónica, Ingeniería en nanotecnología, Ingeniería en procesos agroindustriales, Ingeniería en procesos industriales, Ingeniería en seguridad y salud para el trabajo, Ingeniería en software, Ingeniería en tecnologías de la información y las comunicaciones, Ingeniería en telecomunicaciones, Ingeniería geológica, Ingeniería industrial, Ingeniería mecánica, Ingeniería química. Tales pruebas se consideran específicas para los grupos de referencia. El ICFES dispone de 42 pruebas en total, de las cuales, los estudiantes escogen dos (2) o tres (3), y constituyen la evaluación de competencias específicas.

MINERÍA DE 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 pre-procesar los datos, hacer minería de datos (datamining) y presentar resultados. El proceso KDD se puede aplicar en diferentes dominios, por ejemplo, para determinar perfiles de clientes que evaden 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, para determinar patrones de compra de los clientes en sus canastas de mercado.

TAREAS DE MINERÍA DE DATOS

Dentro de la minería de datos se encuentran diferentes tipos de tareas, cada una puede considerarse como un tipo de problema a ser resuelto por un algoritmo de minería de datos. Entre las tareas más importantes de minería de datos están la clasificación, segmentación o clustering, asociación y patrones secuenciales.

CLASIFICACIÓN

Permite obtener resultados a partir de un proceso de aprendizaje supervisado. Es el proceso por medio del cual se encuentra propiedades comunes entre un conjunto de objetos de una base de datos y se los cataloga en diferentes clases, de acuerdo al modelo de clasificación.

SEGMENTACIÓN O CLUSTERING

Es el proceso de agrupar objetos físicos o abstractos en clases de objetos similares. Se agrupa un conjunto de datos, sin considerar un atributo de clase predefinido, tomando como base el principio de maximizar la similitud intraclase y minimizar la similitud interclase. El clustering ayuda a construir particiones significativas de un 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 su partición en segmentos o clusters de registros similares que comparten un número de propiedades; tales registros son considerados homogéneos y los que pertenecen a otros clusters son considerados diferentes.

ASOCIACIÓN

Con la tarea de asociación se descubre patrones en forma de reglas, que muestran los hechos que ocurren frecuentemente juntos, en un conjunto de datos determinado; a menudo se lo 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 de estos ítems. La tarea es encontrar relaciones entre los ítems de esas canastas para descubrir reglas que cumplan unas especificaciones mínimas dadas por el usuario, expresadas en forma de soporte y confianza.

PATRONES SECUENCIALES

Los patrones secuenciales buscan ocurrencias cronológicas. 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. Un patrón secuencial se compone de una lista de conjuntos de ítems.

El problema consiste en encontrar todos los patrones secuenciales que cumplan con un soporte mínimo especificado por el usuario, donde 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 una compra. En el campo de la medicina, los patrones secuenciales se pueden utilizar para ayudar a identificar síntomas y enfermedades que preceden a otras enfermedades.

DESEMPEÑO ACADÉMICO

El desempeño académico, como fenómeno complejo y multi-dimensional, no se explica únicamente por las calificaciones o puntajes que obtienen los estudiantes. El desempeño académico que se concibe en este proyecto de investigación, se relaciona con el proceso de aprendizaje y con el logro de competencias de los estudiantes, en donde están implicados factores internos y externos al estudiante.

CRISP-DM

La metodología CRISP-DM, consiste en un conjunto de tareas que están organizadas en cuatro niveles de abstracción, estable-

cidos en términos de jerarquías de tareas, que inician en el nivel más general, hasta llegar, finalmente, a los casos más específicos. 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, no es estricto el orden de ejecución de los procesos, y con frecuencia se puede pasar de uno a otra, de atrás hacia delante y viceversa. Éstos dependen del resultado de cada fase o de la planeación de la siguiente tarea a ejecutar. Cada fase se estructura en varias tareas generales, que se proyectan en tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones definidas.

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 los patrones. La importancia de los árboles de decisión se debe a su capacidad para 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 conduce a una y sólo una hoja, asignando una única clase a la predicción de las reglas.

EFFECTO INSTITUCIONAL EN COMPETENCIAS GENÉRICAS

A partir de las diferencias estandarizadas entre los promedios en cada una de las instituciones, frente a la institución que alcanza el máximo puntaje promedio en una de las cinco competencias genéricas, se determina un ranking de las instituciones. Aplicando el estadístico d' de Cohen, se calcula el tamaño del Efecto Institucional. Cohen propone la siguiente escala para interpretar las diferencias en unidades estándar, obtenidas con el estadístico referido:

- [0.0, 0.2] diferencia trivial o irrelevante.
- (0.2, 0.5] diferencia pequeña.
- (0.5, 0.8] diferencia moderada.
- (0.8, infinito) diferencia grande.

INTRODUCCIÓN

El desempeño académico de los estudiantes es un tema determinante en la calidad de la educación, ya que, es entendido como el aprendizaje de los alumnos, dependiente de diversas variables tanto objetivas como subjetivas. Las primeras, hacen referencia a los sistemas de evaluación, las calificaciones, las políticas educativas, entre otras; las segundas, involucran aspectos cognitivos, familiares, sociales y socioeconómicos (Erazo, 2012).

La calidad de la educación superior hace referencia a la capacidad que tiene el sistema educativo para lograr que los estudiantes obtengan resultados académicos de calidad, relacionados con el aprendizaje, independientemente de su origen social (Ardila, 2011). En Colombia, el inconveniente principal que se presenta para alcanzar este reto, es que el sistema educativo tiene poca posibilidad de evolucionar, a menos que supere las brechas socioeconómicas, científicas y políticas de los estudiantes.

En Colombia, el Gobierno Nacional se ha propuesto como tarea reducir estas brechas, a través de la gestión de la calidad de la Educación Superior. Para este fin, en el año 2009, el Congreso de la República fijó los parámetros y los criterios para el sistema de calidad de la educación, entre los cuales, se encuentran los Exámenes de Estado (Congreso de la República de Colombia-Ley 1324, 2009). El Examen de Estado de la Educación Superior, definido en el decreto 3963 del 2009, se constituye en el instrumento estandarizado y externo a las Instituciones de Educación Superior

(IES), con el cual, se evalúa las competencias de los estudiantes, próximos a finalizar la Educación Superior. Las IES analizan los resultados de las pruebas y planean acciones de mejoramiento, en la perspectiva de que los estudiantes logren un adecuado dominio de las competencias en su carrera respectiva (Lemos et al., 2014).

Según el Instituto Colombiano para la Evaluación de la Educación -ICFES (2013), las competencias se constituyen en el articulador de los diferentes niveles de formación. 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 diferente naturaleza, entre otros. El Decreto 3963 (2009), expedido por el Ministerio de Educación Nacional, define las competencias genéricas, como aquellas competencias que son necesarias para el adecuado desempeño profesional; 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, las deben desarrollar todas las personas, dependiendo del nivel de formación (Icfes, 2013). En este contexto, las competencias genéricas identifican los aspectos comunes para cualquier profesión, tales como: capacidad para aprender, para la toma de decisiones, el diseño de proyectos, las habilidades interpersonales, entre otras (González & Wagenar, 2003), que se complementan con las competencias específicas de cada carrera.

Entre los objetivos del examen de estado de calidad de la educación superior - Saber Pro, están: constatar el nivel de desarrollo de las competencias de los estudiantes próximos a culminar los programas académicos de pregrado, generar indicadores de valor agregado de la educación superior, y, constituirse como fuente de información para la construcción de indicadores de evaluación de la calidad de los programas e IES (Icfes, 2016). Este examen está compuesto por pruebas que evalúan competencias genéricas y competencias comunes a grupos de programas, denominadas competencias específicas.

De acuerdo a los lineamientos Saber Pro, expedidos por el Icfes (2011a), todos los estudiantes deben presentar los módulos de competencias genéricas (Icfes, 2011b), sin importar el programa de formación que cursen. Desde 2012 (Icfes, 2012A) las competencias genéricas que se evalúan en Saber Pro, son lectura crítica, comunicación escrita, razonamiento cuantitativo, inglés y competencias ciudadanas (Icfes, 2016):

- *Comunicación escrita*: evalúa la competencia para comunicar ideas por escrito 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.
- *Razonamiento Cuantitativo*: evalúa 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.
- *Lectura Crítica*: evalúa competencias relacionadas con la capacidad para leer de manera analítica y reflexiva. 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, construya su sentido global, establezca relaciones entre enunciados y evalúe su intencionalidad. Aborda los siguientes procesos: ubicar información, relacionar, construir la representación global del texto, analizar y evaluar la relación entre procesos discursivos y contexto sociocultural.
- *Competencias ciudadanas*: evalúa los conocimientos y habilidades que posibilitan la construcción de marcos de comprensión del entorno, los cuales promueven el ejercicio de la ciudadanía y la coexistencia inclusiva, según la Constitución política de Colombia.

- *Inglés*: evalúa la competencia de los estudiantes para comunicarse efectivamente en inglés; se alinea con el Marco Común Europeo que clasifica el desempeño en cuatro niveles: A1, A2, B1 y B2.

Vale destacar que, estas pruebas constituyen 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 evaluar la calidad del servicio público educativo, y para ejercer la inspección y vigilancia.

En la tabla 1, se presenta un resumen de los cambios de las Pruebas Saber Pro desde 2009 hasta 2015, en relación con los módulos de competencias genéricas.

Tabla 1. Cambios en las pruebas Saber Pro de 2009 a 2012.

Años	Descripción
2009-2011-1	Todos los programas académicos presentaron dos módulos de competencias genéricas: inglés y comprensión lectora.
2011-2	Todos los programas presentaron cuatro módulos de competencias genéricas: razonamiento cuantitativo, lectura crítica, inglés y escritura.
2012-2015	Todos los programas presentan cinco (5) competencias genéricas: comunicación escrita, razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés

Fuente: Icfes, 2016: Informe nacional de resultados Saber Pro 2012-2015

Por otra parte, diversas investigaciones acerca del desempeño académico de los estudiantes, se orientan a identificar los principales factores que están asociados con el desempeño académico en pruebas estandarizadas. Se ha encontrado que, las variables que inciden en el rendimiento académico, pueden ser externas o internas al estudiante; las primeras, influyen desde el exterior de la institución, entre tales variables se encuentran las de tipo sociodemográfico, económico, académico e institucional; las

segundas, se relacionan directamente con aspectos personales, psicológicos o somáticos del estudiante.

La investigación que se reporta en este libro, busca contribuir a la reflexión sobre la formación de profesionales de primer nivel, críticos, con una estructura de pensamiento que les permita tener una visión amplia; además, que sean profesionales que se caractericen por ser excelentes ciudadanos, seres humanos y profesionales, tal como lo establece el PEI de la Universidad de Nariño (UDENAR-PEI, 2013).

Los estudios que se han realizado hasta el momento (icfes, 2010), (Zapata, 2011), (icfes, 2012b) y (UNAL, 2012), con respecto al análisis de las pruebas Saber Pro, se basan en una interpretación estadística de los resultados donde, fundamentalmente, se consideran variables y relaciones primarias, sin estudiar otras interrelaciones que, por lo general, están ocultas y que se pueden descubrir utilizando un tratamiento más complejo de los datos, tal como sucede con la minería de datos. Tomando como base los resultados de las pruebas Saber Pro 2011-2, Timarán et al (2016), realizaron un estudio, a nivel nacional, para descubrir patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales, que presentaron las pruebas Saber Pro en el segundo semestre del año 2011. En dicha investigación se aplicó la metodología CRISP-DM y clasificación por árboles de decisión, la cual constituye una técnica predictiva de minería de datos. Dentro de los resultados obtenidos en el estudio referido, se destaca la importancia de la acreditación institucional como un factor que influye en el desempeño de los estudiantes en las diferentes competencias genéricas.

Por otra parte, en la Universidad de Nariño no se han planteado investigaciones que analicen el desempeño de los estudiantes, de los diferentes programas profesionales que ofrece esta institución, en las competencias genéricas de las pruebas Saber Pro 2012, 2013 y 2014, utilizando técnicas estadísticas o de minería de datos, por lo cual, el estudio que se reporta en este texto, resultó ser muy pertinente.

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 y predecir cualquier fenómeno dentro del campo educativo (Timarán, Calderón, & Jiménez, 2013a), (Timarán, Calderón, & Jiménez, 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, & García, 2010). Las instituciones de educación pueden usar la minería de datos para hacer análisis comprensivos de las características de sus estudiantes, de métodos evaluativos, develando procesos exitosos o por el contrario, detectando fraudes o inconsistencias (Orea, Vargas & Alonso, 2005).

El objetivo del estudio que se reporta, fue descubrir patrones de desempeño académico en competencias genéricas de los estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro 2012, 2013 y 2014, a partir de los datos socio demográficos, económicos, académicos e institucionales almacenados en las bases de datos del ICFES, utilizando técnicas de minería de datos. Para lograr este objetivo, se requirió apropiarse y analizar los elementos que constituyen cada una de las competencias genéricas; generar un repositorio limpio y transformado de datos a partir de la selección de los atributos sociodemográficos, económicos, académicos e institucionales de las bases de datos del ICFES; generar el modelo de detección de patrones de desempeño académico mediante la aplicación de técnicas de minería de datos, y, finalmente, generar conocimiento a partir de la interpretación de los patrones que se obtuvieron.

Igualmente, la investigación se orientó a dar respuesta a la siguiente pregunta: ¿Cuáles son los factores sociodemográficos, económicos, académicos e institucionales asociados al desempeño académico de los estudiantes de los programas profesionales de la Universidad de Nariño en las competencias genéricas de las pruebas Saber Pro 2012, 2013 y 2014?

Este libro está organizado por capítulos. En el capítulo I, se relacionan los estudios que se han efectuado sobre desempeño académico y las pruebas Saber Pro. En el capítulo II, se conceptualiza sobre la minería de datos educativa y se describen las fases de la metodología para proyectos de minería de datos CRISP-DM, misma que es utilizada en ese proyecto de investigación. En el capítulo III, se presenta un análisis exploratorio de los datos y se muestran algunas tendencias sobre el desempeño académico con relación a las competencias genéricas. En el capítulo IV, se indica cómo se desarrolló el proyecto de investigación aplicando las diversas fases de la metodología CRISP-DM. En el capítulo V, se evalúan, interpretan y discuten los resultados obtenidos en el estudio. Finalmente, en el capítulo VI, se presentan conclusiones y recomendaciones con base en los resultados obtenidos.

CAPÍTULO I

ESTUDIOS SOBRE DESEMPEÑO ACADÉMICO Y PRUEBAS SABER PRO

El caso del desempeño académico no es un tema reciente en el contexto universitario. Se considera un estudio pionero el realizado en Estados Unidos por Coleman (1966), en el cual se trabajó con más de 500 mil estudiantes, buscando determinar la relación entre los insumos del proceso de formación y su producto final, relacionándolo también con la política educativa. Los resultados indicaron que el desempeño académico en los Estados Unidos estaba influenciado, en gran medida, por las características socioeconómicas de los estudiantes y que, las variables referidas a la institución educativa, influían poco o nada en el desempeño académico. Estos resultados indujeron muchos estudios, algunos han reafirmado la hipótesis, otros, la han rechazado.

En relación con los estudios sobre desempeño académico, los autores Benítez, Giménez y Osicka (2000), consideran que se deben tener en cuenta factores socioeconómicos, complejidad de los programas de estudio, modalidades de formación, metodologías de enseñanza y los conceptos previos de los estudiantes. Por su parte, según Castaño et. al. (2004), existen cuatro grupos de variables

que determinan el desempeño académico: socio-demográficas, académicas, socioeconómicas, e institucionales. Para Porto y Di Gresia (2004), constituyen factores explicativos del desempeño académico los siguientes: el género, la edad y el nivel educativo de los padres. Para estos autores, las mujeres tienen mejor desempeño que los hombres; los jóvenes se desempeñan mejor que los adultos, y a mayor nivel educativo de los padres, mejor desempeño de los estudiantes.

En la investigación “El rendimiento académico: concepto, investigación y desarrollo”, realizada por Navarro (2003), se destaca que el desempeño académico constituye una red dinámica de articulaciones cognitivas, cuyos rasgos característicos diferencian los resultados de cualquier proceso de enseñanza y de aprendizaje.

Los autores Jano y Ortiz (2005), realizaron la investigación “Determinación de los factores que afectan el rendimiento académico en la educación superior”, llegando a la conclusión de que, entre tales factores, se encuentran el esfuerzo efectivo que realiza el estudiante, sus habilidades y los conocimientos previos a su ingreso a la universidad. Los autores encontraron que, tanto el esfuerzo efectivo realizado durante la formación del estudiante, como las habilidades y conocimientos previos, influyen de manera significativa en el rendimiento académico, y que, las variables socio-demográficas utilizadas en el estudio, parecen no tener un efecto relevante en el mismo. Los investigadores no utilizaron información relacionada con antecedentes familiares y sociales de los individuos, ni los recursos que proporciona la familia.

En Argentina, Di Gresia, Fazio, Porto, Ripani y Sosa (2005) realizaron un estudio con el objetivo de analizar la transición de los estudiantes universitarios entre su ingreso a la universidad y la salida por graduación o deserción; analizaron los determinantes del desempeño académico utilizando información de un censo realizado en 1994 que provee información de los estudiantes de la mayoría de las universidades públicas de Argentina. Se utilizó como media de desempeño académico la cantidad de materias aprobadas por período académico y como variables independien-

tes se utilizaron cinco grupos de variables: 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. Los resultados muestran que las mujeres tienen un rendimiento superior a los varones; además, que también son factores asociados al buen desempeño, los siguientes: ser argentino, no soltero, haber estudiado en una escuela de enseñanza media privada, haberse mudado para asistir a la universidad, tener padres con mayor nivel educativo, destinar más horas al estudio.

En la Universidad de Salamanca, Tejedor y García (2007) estudian las causas del bajo rendimiento del estudiante universitario con base en opiniones de profesores y estudiantes. Las causas del bajo rendimiento las clasifican en tres categorías: institucionales, del docente y del estudiante. El estudio finaliza ofreciendo orientaciones para mejorar el desempeño académico del estudiante, las cuales pueden ser consideradas en la planificación y gestión de las instituciones educativas.

Garvanzo (2007), en Costa Rica, realizó una investigación sobre los “Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública”, en la cual deja claro que, estudiar los factores asociados al rendimiento académico, permite conocer y analizar la complejidad del mismo; indica que es posible que las relaciones que se presentan entre las variables, por más incidencia que parezcan tener, no permiten generalizar los resultados, ya que pueden variar según el conjunto de interacciones, las poblaciones y sus contextos.

Por su parte, Izar(2011), realiza en México la investigación “Factores que afectan el desempeño académico de los estudiantes de nivel superior en Rioverde, San Luis Potosí”; el objetivo es investigar algunos factores relacionados con el desempeño académico de los estudiantes de educación superior, incluye una descripción del impacto de los cambios demográficos en la educación y los principales problemas que hay en torno a ella. Las variables que resultaron significativas, son las siguientes: género,

trayectoria académica y el interés de formarse profesionalmente. Se reportó que no hay asociación entre desempeño académico y nivel socioeconómico, indicando que contar con mejores recursos económicos no implica mejor desempeño académico; lo mismo se concluyó sobre el nivel de escolaridad de sus padres.

A nivel nacional, los autores Vélez y Roa (2005), en Bogotá, realizaron la investigación “Factores asociados al rendimiento académico en estudiantes de medicina de la Universidad del Rosario”; indican que: no leer como pasatiempo, la violencia intrafamiliar, el consumo de drogas psicoactivas, provenir de un colegio mixto, padres sin estudios profesionales, no disponer de hábitos de estudio, utilizar más tiempo en actividades sociales, la dependencia económica, entre otros factores, producen un desempeño académico regular; adicionalmente, destacan que los factores sociodemográficos no tienen incidencia significativa sobre el desempeño académico.

En el Sistema de Universidades Estatales del Caribe Colombiano, 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, versus, el desempeño académico en la universidad. Es un estudio descriptivo correlacional, donde se estudiaron las siguientes variables: desempeño académico en educación media, indicadores de desempeño y desempeño en la universidad. Los resultados permitieron determinar que existe una relación estadísticamente positiva ($\rho = 0,417$) entre el rendimiento académico en educación media y en la Universidad. Si bien, no se puede afirmar que el rendimiento académico en la educación media, constituye un predictor del desempeño en la Universidad, sí se puede afirmar que hay una tendencia en el sentido que los estudiantes mantengan o mejoren el rendimiento académico en la universidad con relación al obtenido en la educación media.

En el estudio realizado por Chicaiza, Galvis y Ramírez (2010) sobre “Determinantes del rendimiento académico en Colombia:

pruebas ICFES Saber 11^o, 2009”, se utilizan los resultados obtenidos por los estudiantes en las áreas de matemáticas y lenguaje, como un medio para identificar los determinantes del desempeño académico en Colombia. Los resultados muestran la relevancia de la dimensión socioeconómica en el desempeño para las dos áreas; particularmente, las variables relacionadas con el nivel de ingreso, la cual, presenta 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, el hecho de tener computador en el hogar implica una mayor probabilidad de obtener buen desempeño académico. Adicionalmente, se encontró que el nivel educativo de los padres, tiene gran significancia: a medida que aumenta el nivel educativo de los padres se evidencia mejor desempeño académico de los estudiantes; pues, padres con mayor nivel educativo tienen la posibilidad de apoyar, de manera directa, el proceso de aprendizaje.

En relación con el género, el estudio encontró que los hombres tienen mejor desempeño en matemáticas y las mujeres en lenguaje. En cuanto a la variable edad, se encontró que, en la medida que aumenta, existe mayor probabilidad de obtener menor desempeño. Incide igualmente, la jornada de estudio: los de jornada completa obtienen puntajes más altos que los estudiantes de otras jornadas.

La Secretaría de Educación Distrital de Bogotá, adelantó el Estudio liderado por Bodensiek (2010), sobre los factores que influyen en el rendimiento escolar, cuyo objetivo era discutir sobre los factores asociados al desempeño académico exitoso que se han encontrado en diferentes estudios a nivel internacional, con el fin de identificar los que influyen en los resultados del examen de Estado para 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 trayectoria de la vida académica; en las segundas, se relacionan 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, están las siguientes: estrato social del estudiante, ingresos económicos, nivel educativo de los padres, características profesionales de los docentes y su vinculación a la institución.

En Colombia 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 con el objeto 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. El proyecto se desarrolló en dos etapas: primera, exploratoria, donde se caracterizaron los 840 estudiantes de la cohorte 2009-1; segunda, focalización, la cual se realizó con la información recolectada de 467 estudiantes de la misma cohorte.

Es importante señalar que, en general, las condiciones económicas de los estudiantes en la EAFIT, 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, resultaron determinantes; como también, el grado de escolaridad de los padres, pues, los hijos de padres universitarios tienen una mayor participación en los mayores promedios.

Los resultados que obtuvieron Montes y Lerner (2012), sugieren lo siguiente: asistir académicamente a la población que ingresa a la universidad proveniente de instituciones educativas de niveles bajos, según la clasificación del ICFES, y con bajas calificaciones del bachillerato; 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 en grupo; ofrecer continuidad al programa de becas de la Universidad; preservar el reconocimiento de calidad institucional y de programas académicos, y mantener oportunidades en actividades extracurriculares.

Los estudios que se han realizado hasta el momento con respecto al análisis de desempeño en las pruebas Saber Pro, se basan en una interpretación estadística de resultados.

En un estudio de Ramírez y Teichler (2014), analizan los factores socioeconómicos y educativos asociados con el desempeño académico, según nivel de formación y género de los estudiantes que presentaron la prueba Saber Pro 2009, para lo cual, utilizaron el método de regresión lineal múltiple. Se concluyó que, el desempeño académico previo medido en los puntajes obtenidos en las pruebas Saber 11° está asociado fuerte y positivamente con el de desempeño de los estudiantes en las pruebas Saber Pro, sin importar su género o nivel de formación. Por su parte, las variables socioeconómicas y educativas estudiadas, resultaron asociadas débilmente con el desempeño en la educación superior en las pruebas Saber Pro, en comparación con el desempeño en Saber 11°.

En el informe nacional de resultados Saber Pro 2012-2015 (icfes,2016), se analizaron los resultados de los módulos de competencias genéricas del examen Saber Pro, desde tres indicadores: agregado nacional, agregado nacional por género y agregado nacional por sector. Los resultados del agregado nacional denotan ausencia de progreso en el puntaje promedio de los módulos de comunicación escrita y competencias ciudadanas, así como una

leve tendencia de crecimiento en las demás competencias. Los resultados por género indican que, los hombres obtuvieron mejores puntajes en todos los módulos de competencias genéricas, con excepción de comunicación escrita; y en la desagregación por sector, señala que, las IES oficiales denotan superioridad en los módulos de comunicación escrita, razonamiento cuantitativo, lectura crítica y competencias ciudadanas; y que las IES privadas cuentan con mejores resultados en el módulo de inglés.

Los autores Timarán et al. (2016), realizaron un estudio en Colombia en el que buscaban dar respuesta a la 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 el estudio utilizaron la técnica clasificación, de minería de datos, basada en árboles de decisión. En los patrones descubiertos se destaca que, la acreditación institucional y la modalidad de estudio, son dos atributos importantes asociados al desempeño académico de los estudiantes en las pruebas Saber Pro 2011-2, en las competencias genéricas: lectura crítica, composición escrita, razonamiento cuantitativo e inglés.

CAPÍTULO II

LA MINERÍA DE DATOS EDUCATIVA Y CRISP-DM

2.1 MINERÍA DE DATOS EDUCATIVA

La minería de datos, también conocida como Descubrimiento de Conocimiento en Bases de Datos-KDD (del inglés *Knowledge Discovery in Databases*) es la encargada, a través de un conjunto de técnicas, de explorar, seleccionar, preparar y transformar los datos para obtener información oculta y útil en ellos. Es el proceso de identificar y encontrar conocimiento a partir de grandes volúmenes de datos, con el fin de ayudar a la toma de decisiones en las organizaciones (Fayyad, 1996), (Hernández, Ramírez & Ferri, 2005), (Han, Kamber & Pei, 2012), (Timarán et al., 2016).

Según Maya et al. (2016), la minería de datos como área de las Tecnologías de la Información y Comunicación (TIC), integra numerosas técnicas de análisis de datos y extracción de modelos, capaz de obtener patrones, describir tendencias y regularidades, predecir comportamientos y explorar información computarizada proveniente de fuentes heterogéneas y grandes volúmenes, permitiendo a los individuos y a las organizaciones comprender

y modelar de una manera más eficiente y precisa el contexto en que deben actuar y tomar decisiones inteligentes. Como un dominio altamente basado en aplicaciones, la minería de datos ha incorporado muchas técnicas de otros dominios como estadísticas, aprendizaje automático, reconocimiento de patrones, bases de datos y sistemas de almacenamiento de datos, recuperación de información, visualización, algoritmos, computación de alto rendimiento y muchos dominios de aplicación (Figura 1). La naturaleza interdisciplinaria de la investigación y el desarrollo de la minería de datos contribuyen significativamente al éxito de la minería de datos y sus amplias aplicaciones (Han, Kamber & Pei, 2012).

En la minería de datos existen dos tipos de métodos de aprendizaje: el aprendizaje supervisado o predictivo y el aprendizaje no supervisado o descriptivo. El aprendizaje supervisado se basa en entrenar un conjunto de datos de un modelo para que, por medio de diferentes datos, se pueda predecir el comportamiento de una variable (Sanchez-Santillan et al., 2016). Pretende estimar valores futuros o desconocidos de variables de interés, que se denominan variables objetivo, dependientes o clases, usando otras variables o atributos denominadas variables independientes o predictivas. Se puede utilizar, por ejemplo, para predecir nuevos clientes, si son buenos o malos, basado en su estado civil, edad, género, profesión; o determinar para nuevos estudiantes, si desertan o no, en función de su zona de procedencia, facultad, estrato y promedio de notas.

El aprendizaje no supervisado, es el método que permite agrupar datos de forma rápida, también son llamados métodos simétricos o indirectos. Las observaciones se clasifican en grupos que no son conocidos con anticipación; las variables pueden estar conectadas según vínculos desconocidos; así, todos los atributos disponibles son evaluados en el mismo nivel y no hay casualidades (Emtiyaz & Keyvanpour, 2011). Identifican patrones que explican o resumen los datos. Sirven para explorar las propiedades de los datos examinados, no se pueden utilizar para predecir nuevos datos. Se puede utilizar, por ejemplo, para identificar grupos de personas con similares gustos, o identificar patrones de compra de clientes de una zona.

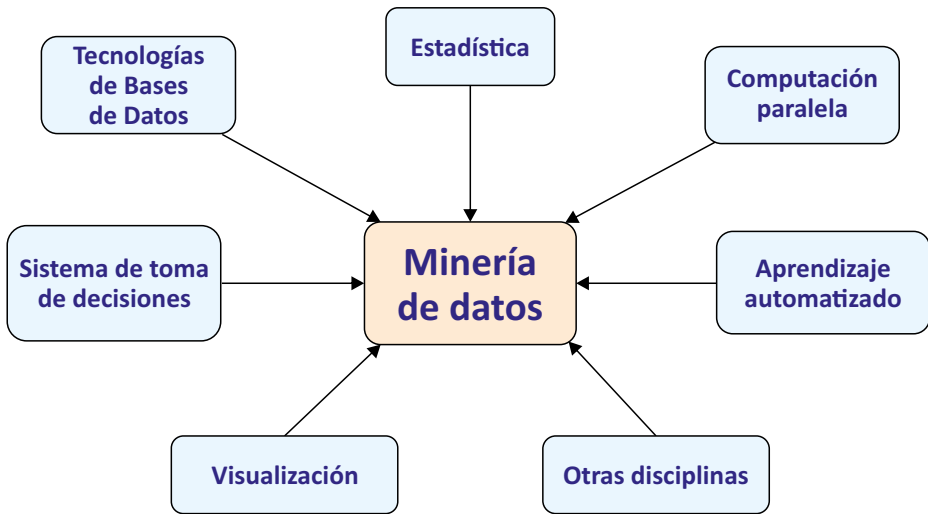


Figura 1. Relación de la minería de datos con otras disciplinas

Las técnicas de minería de datos se aplican según el aprendizaje que se esté aplicando, por lo cual, se agrupan en Predictivas y Descriptivas (Kurgan, 2006). Las primeras, están compuestas por tareas de clasificación y regresión; mientras que las segundas, están conformadas por agrupación, asociación y correlación (Rosado & Vergel, 2017).

La minería de datos se ha empleado en numerosos campos, incluyendo desde los ya conocidos casos de canastas de las compras, hasta la bioinformática e investigaciones contra el terrorismo. Se puede aplicar en diferentes dominios, por ejemplo, para determinar perfiles de clientes que evaden 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, entre otros (Timaran et al., 2016).

La aplicación de las técnicas y herramientas de la minería de datos en los diversos contextos educativos, se conoce como minería de datos en la educación EDM (del inglés Educational Data Mining). El proceso EDM convierte los datos brutos provenientes de los sistemas educativos en información útil que podría tener

un gran impacto en la investigación y las prácticas educativas (Romero & Ventura, 2010). Según Baker (2010), Huapaya et al. (2012) y Algarni (2016), la EDM es el área de la investigación científica centrada en el desarrollo de métodos para realizar descubrimientos dentro de los tipos únicos de datos que provienen de entornos educativos, y usar esos métodos para comprender mejor a los estudiantes y el contexto en el que aprenden. Además, EDM extrae información interesante, interpretable, útil y novedosa de datos educativos. Desde un punto de vista práctico, EDM permite extraer conocimiento de los datos del alumno. Este conocimiento se puede utilizar para validar y evaluar un sistema educativo, mejorar la calidad de los procesos de enseñanza-aprendizaje y sentar las bases para un proceso de aprendizaje más eficaz.

Se puede utilizar sus técnicas para explicar y/o predecir cualquier fenómeno dentro del campo educativo (Timarán, Calderón, & Jiménez, 2013a; Peña, 2014). El EDM es útil en muchas áreas diferentes, como la identificación de estudiantes con alto riesgo académico, las necesidades de aprendizaje prioritarias para diferentes grupos de estudiantes, el aumento de los índices de graduación, la evaluación del desempeño institucional, la maximización de los recursos del campus y la optimización de la renovación del currículo de la asignatura (Algarni, 2016). Usando técnicas de extracción de datos, por ejemplo, se puede predecir, con un porcentaje muy alto de confiabilidad, la probabilidad de deserción de cualquier estudiante (Valero, 2009), (Valero, Salvador, & García, 2010), (Timarán, 2010), (Timarán et al., 2016). En el ámbito educativo, la aplicación de la minería de datos como técnica de análisis se ubica en el entorno del sistema educativo, y en concreto en educación superior. Las instituciones de educación superior pueden usar la EDM para hacer análisis comprensivos de las características de sus estudiantes, métodos evaluativos, develando procesos exitosos, o por el contrario, detectando fraudes o inconsistencias (Valero et al, 2010). Baker (2010) sugirió cuatro áreas clave de la aplicación EDM: mejorar los modelos de los estudiantes, mejorar los modelos de dominio, estudiar el apoyo pedagógico proporcionado por el software de aprendizaje y realizar investigaciones científicas sobre el aprendizaje y los alumnos.

Castro et. al. (2007) clasificó las tareas de EDM en cuatro áreas diferentes: aplicaciones que se ocupan de la evaluación del rendimiento del aprendizaje de los estudiantes, la adaptación del curso y las recomendaciones de aprendizaje para personalizar el aprendizaje de los estudiantes en función de los comportamientos individuales de los estudiantes, desarrollando un método para evaluar materiales en cursos en línea; enfoques que use la retroalimentación de estudiantes y profesores en cursos de aprendizaje electrónico (e-learning) y modelos de detección para descubrir los comportamientos de aprendizaje de los estudiantes.

La EDM, según Jiménez y Álvarez (2010), ofrece numerosas ventajas comparándola con los paradigmas más tradicionales de investigación relativa a la educación, como experimentos de laboratorio, estudios sociológicos o investigación de diseño. En particular, la creación de repositorios públicos de datos educacionales (como los que se utilizan en esta investigación sobre las pruebas Saber Pro almacenados en los repositorios del ICFES) ha creado una base que facilita la minería de datos educacionales. Los datos de estos repositorios son totalmente válidos (ya que son datos reales sobre el rendimiento y aprendizaje de estudiantes reales, en ambientes educacionales, tomados en tareas de aprendizaje) y cada vez más fácilmente accesibles para comenzar una investigación. Estos puntos permiten a los investigadores ahorrar mucho tiempo en tareas como la búsqueda de individuos (tales como escuelas, profesores y alumnos), organización de los estudios y recopilación de datos, ya que estos se encuentran directamente accesibles (Jiménez & Álvarez, 2010).

Hay una gran variedad de métodos empleados habitualmente en el ámbito de la educación en la minería de datos. Estos métodos están clasificados en las siguientes categorías: predicción, agrupamiento, minería de relaciones, inferencia a través de modelos, y destilación de datos para la interpretación por parte de un ser humano (Algarni, 2016). Las tres primeras categorías son universales para distintos tipos de minería de datos (aunque en algunos casos con distintos nombres). Las categorías cuarta y quinta consiguen una particular importancia dentro de la minería de datos educacionales (Jiménez & Álvarez, 2010).

La predicción tiene como objetivo predecir variables desconocidas basadas en datos históricos de la misma variable. En esta tarea, se desarrolla un modelo de modo que, se pueda inferir una variable a partir de alguna combinación de otras variables incluidas en los datos. La efectividad del modelo de predicción depende del tipo de variables de entrada. Se requiere que el modelo de predicción tenga datos etiquetados limitados para la variable de salida. Los datos etiquetados ofrecen algunos conocimientos previos sobre las variables que se necesita predecir. Sin embargo, es importante considerar los efectos de la calidad de los datos de entrenamiento para lograr el modelo de predicción (Algarni, 2016). En general, existen tres tipos de predicción: clasificación, regresión y estimación de densidad. En clasificación, el valor predicho es una variable categórica o binaria. Es la más común y efectiva técnica de minería de datos utilizada en la clasificación y predicción de valores (Abu, 2016). Algunos métodos populares de clasificación incluyen árboles de decisión, regresión logística y máquinas de soporte vector. En regresión, el valor predicho es una variable continua. Algunos métodos populares de regresión en EDM incluyen la regresión lineal, redes neuronales y regresión sobre máquina de soporte vector. En estimación de densidad, la variable predicha es una función de densidad de probabilidad. Estimadores de densidad pueden estar basados en una variedad de funciones de kernel, incluyendo funciones gaussianas (Jiménez & Álvarez, 2010).

La agrupación o clustering, es un método utilizado para separar datos en diferentes grupos o clusters en función de ciertas características comunes (Han, Kamber & Pei, 2012). El agrupamiento es útil en casos donde las categorías más comunes de los datos no son conocidas. A diferencia del método de clasificación, en la agrupación, las etiquetas de datos son desconocidas. El método de agrupamiento ofrece al usuario una visión amplia de lo que está sucediendo en ese conjunto de datos. La agrupación, a veces se conoce como una clasificación no supervisada, porque las etiquetas de clase son desconocidas. Si un conjunto de clusters es óptimo, en cada categoría, cada punto será más similar a los puntos pertenecientes a su cluster que a puntos pertenecientes a

otros grupos. Por ejemplo, se pueden agrupar diferentes escuelas en función de las similitudes y diferencias entre ellas.

La minería de relaciones tiene como objetivo encontrar relaciones entre diferentes variables en conjuntos de datos con una gran cantidad de variables (Agrawal & Srikant, 1994). Una forma de realizar esto es localizando las variables más fuertemente relacionadas con una única variable de interés, o también, mediante el descubrimiento de las relaciones más fuertes entre diferentes variables. Las relaciones encontradas a través de la minería de relaciones deben satisfacer dos criterios: importancia estadística e interés. Grandes cantidades de datos contienen muchas variables y, por lo tanto, tienen muchas reglas asociadas. Los investigadores han desarrollado diferentes medidas de interés a lo largo de los años, incluido el soporte y la confianza. Sin embargo, algunas investigaciones han concluido que el lift y el coseno son los más relevantes utilizados en EDM (Merceron & Yacef, 2008). Se pueden usar muchos tipos de minería de relaciones, como la minería de reglas de asociación, la minería de patrones secuenciales y la minería de patrones frecuentes. La minería de reglas de asociación es el método EDM más común (Algarni, 2016).

En la inferencia a través de modelos, estos generalmente se basan en agrupamiento, predicción o ingeniería del conocimiento utilizando razonamiento humano en lugar de métodos automatizados. El modelo desarrollado se utiliza como parte de otros modelos integrales, como la minería de relaciones (Jiménez & Álvarez, 2010) (Algarni, 2016).

Otra área de interés en EDM es la destilación de datos para la interpretación humana. La destilación de datos para el juicio humano tiene como objetivo hacer que los datos sean comprensibles. Presentar los datos de diferentes maneras ayuda al cerebro humano a descubrir nuevos conocimientos. En algunos casos, los seres humanos pueden realizar inferencias sobre datos cuando éstos son presentados adecuadamente, que se encuentran más allá del punto de mira inmediato de los métodos de minería de datos totalmente automatizados. Los métodos en esta área se aplican en los datos educativos para dos propósitos: clasificación y/o identi-

ficación. La destilación de datos para la clasificación puede ser un proceso de preparación para construir un modelo de predicción y la identificación tiene como objetivo mostrar datos de manera que sean fácilmente identificables a través de patrones bien conocidos que no pueden formalizarse. De todos modos, las visualizaciones más utilizadas en el campo educacional suelen ser distintas a las utilizadas para la resolución de problemas de visualización de la información, debiéndose a la estructura específica y el significado embebido en esa estructura, habitualmente presente en datos educacionales (Romero & Ventura, 2010).

Todas estas técnicas de minería específicas de la educación pueden mejorar en gran medida el diseño instruccional y las decisiones pedagógicas.

2.2 METODOLOGÍA CRISP-DM

En un intento de normalización del proceso de minería de datos, de forma similar a como se hace en ingeniería de software para normalizar el proceso de desarrollo software, surgieron a finales de los años 90, del siglo pasado, dos metodologías principales: CRISP-DM (Cross Industry Standard Process for Data Mining) y SEMMA (Sample, Explore, Modify, Model, and Assess). Ambas especifican las tareas a realizar en cada fase descrita por el proceso, asignando tareas concretas y definiendo lo que es deseable obtener tras cada fase (Villena, 2016). Azevedo y Santos (2008) comparan ambas implementaciones y llegan a la conclusión de que, aunque se puede establecer un paralelismo claro entre ellas, CRISP-DM es más completo porque tiene en cuenta la aplicación al entorno de negocio de los resultados, y por ello, es la que se adoptó popularmente. En encuestas realizadas en KDNuggets en 2002, 2004, 2007 y 2014, se comprobó que CRISP-DM era la principal metodología utilizada, cuatro (4) veces más que SEMMA (Villena, 2016). En la investigación que se reporta en este libro se utilizó CRISP-DM.

La metodología CRISP-DM, para proyectos de minería de datos, no es la “más actual” o “la mejor”, pero es muy útil para comprender esta tecnología o extraer ideas para diseñar o revisar

métodos de trabajo para proyectos de similares características. La metodología CRISP-DM, consiste en un conjunto de tareas que están organizadas en cuatro niveles de abstracción, los cuales, están establecidos respetando jerarquías en tareas; inician en el nivel más general, hasta llegar, finalmente, a los casos más específicos (Chapman et al. 2000). El modelo provee una representación completa del ciclo de vida de un proyecto de minería de datos. El proceso es dinámico e iterativo; no es estricto el orden de ejecución de los procesos, y con frecuencia se puede pasar de uno a otra, de atrás hacia delante y viceversa. Éstos dependen del resultado de cada fase o de la planeación de la siguiente tarea a ejecutar. Cada fase se estructura en varias tareas generales, que se proyectan en tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones definidas (Larose & Larose, 2014).

CRISP-DM está compuesta por seis fases: análisis del problema, análisis de los datos, preparación de los datos, modelado, evaluación y explotación (Gallardo, 2009), tal como se muestra en la Figura 2.

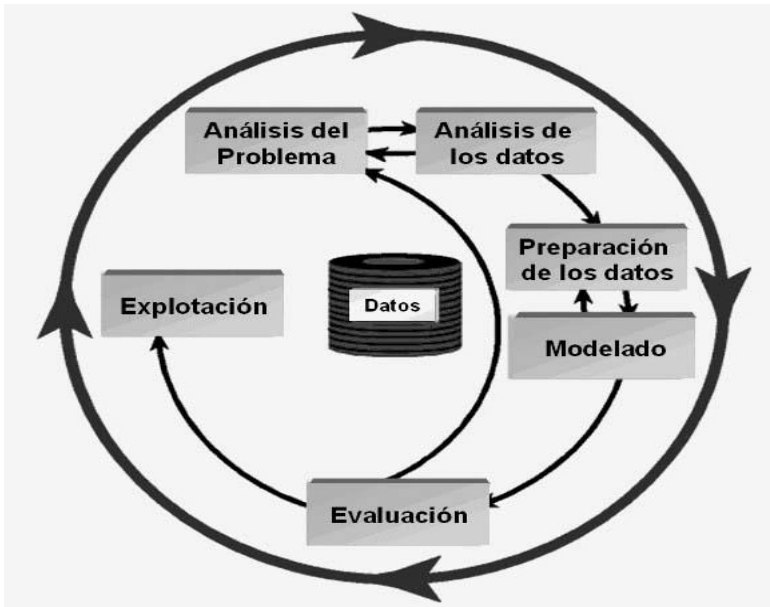


Figura 2. Fases de la metodología CRISP-DM

La fase de Análisis del Problema, se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, luego, en la conversión de este conocimiento en una definición del problema de minería de datos, y un plan preliminar diseñado para alcanzar los objetivos. Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea, permitirá obtener resultados fiables. Para obtener el mejor provecho de la minería de datos, es necesario entender, de la manera más completa, el problema que se desea resolver, esto permitirá recolectar los datos correctos e interpretar correctamente los resultados. En esta fase, es muy importante la capacidad de poder convertir el conocimiento adquirido del negocio, en un problema de minería de datos y en un plan preliminar, cuya meta sea alcanzar los objetivos del negocio (Gallardo, 2009).

La fase de Comprensión de Datos, comienza con una familiarización de los datos; identificación de la calidad de los problemas, para descubrir las primeras señales dentro de los datos y detectar temas interesantes para poder formular hipótesis de información oculta. Esta fase junto a las próximas dos fases, son las que demandan el mayor esfuerzo y tiempo en un proyecto de minería de datos (Chapman et al., 2000).

La fase de Preparación de Datos, cubre todas las actividades para construir un conjunto de datos, con el fin de que se le pueda aplicar posteriormente las técnicas de minería de datos. Estas tareas son ejecutadas en múltiples oportunidades y sin orden. La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato (Chapman et al., 2000) (Gallardo, 2009).

En la fase de Modelado, se seleccionan y aplican varias técnicas de modelado y se calibran los parámetros para obtener óptimos resultados. Hay varias técnicas que tienen requerimientos específicos sobre el formato de los datos, por lo que, frecuentemente es necesario retornar a la fase de preparación de datos. Las técnicas a utilizar en esta fase se eligen en función de que sean las apro-

piadas al problema, que se disponga de los datos adecuados, que cumplan con los requisitos del problema y que tenga un tiempo adecuado para obtener un modelo o conocimiento de la técnica (Chapman et al., 2000) (Gallardo, 2009).

En la fase de Evaluación, se evalúa el modelo construido teniendo en cuenta el cumplimiento de los criterios de éxito, y se interpretan los resultados. Debe considerarse, además, que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que, posiblemente, se haya cometido algún error (Chapman et al., 2000) (Gallardo, 2009).

Finalmente, en la fase de Explotación o Implementación, el conocimiento que se obtiene se transforma en acciones dentro del proceso de negocio. Generalmente, un proyecto de minería de datos no concluye en la implantación del modelo; pues, se deben documentar y presentar los resultados de manera comprensible para el usuario, con el objetivo de lograr un incremento del conocimiento y facilitar la toma de decisiones. Por otra parte, en la fase de explotación, se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados (Chapman et al., 2000) (Gallardo, 2009).

Entre las ventajas de utilizar CRISP-DM, se destaca la posibilidad de replicación de proyectos, su independencia de la industria, aplicación o proyecto, su neutralidad con respecto a las herramientas y su enfoque en las situaciones de negocios y en el análisis técnico (García, 2018).

EFFECTOS DEL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER PRO

En este capítulo se presenta un análisis estadístico preliminar de los datos con el fin de identificar las características sociodemográficas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014, con el fin de establecer un comparativo del desempeño académico de los estudiantes, en las cinco competencias genéricas de dichas pruebas, con las universidades de todo el país, lo que se denominó el “efecto institucional”; y también, analizar el rendimiento específico en dichas competencias en cada uno de los 36 programas que ofrece la Institución, con el fin de establecer lo que se denominó el “efecto programa”. Igualmente, se establece cuál es el efecto que tienen las variables socioeconómicas: edad, género, estado civil, estrato social, hogar actual, cabeza de familia y trabajo laboral, en el rendimiento de las competencias genéricas. Para determinar el tamaño del efecto en las diferentes variables, se establece las diferencias estandarizadas de las medias con el estadístico de Cohen. Con base en este estadístico, se estableció un ranking de las universidades e instituciones de educación superior de todo el país en estas competencias, con el fin de ubicar el lugar que

ocupa la Universidad de Nariño; además, se realizó un ranking por programas de la institución, en las cinco competencias genéricas, en el período analizado.

Mediante análisis estadístico descriptivo, utilizando la calificación media obtenida en las pruebas por estudiantes de cada programa de la Universidad de Nariño, se explica el rendimiento académico que presentan en las competencias genéricas de las pruebas Saber Pro en el período 2012 a 2014; el “efecto institucional”, que en este caso está determinado, en gran parte, por la calidad de la institución, el relativamente débil poder explicativo que tiene el nivel socioeconómico, y las brechas de género encontradas en las diferentes áreas estudiadas. Para la clasificación, se utilizan las diferencias estandarizadas, tomando como referente el promedio nacional más alto de cada competencia, para establecer el “efecto institucional” y ubicar la posición de la Universidad de Nariño en el contexto nacional. Y de otra parte, se toma el promedio de los resultados de en cada uno de los programas en la universidad para establecer el “efecto programa”.

Analizar los resultados del desempeño académico universitario constituye un factor relevante, debido a que, permite disponer de evidencias sólidas sobre la calidad de la educación superior, en la medida que identifican elementos que afectan el rendimiento de los estudiantes, los cuales, son importantes en el momento de diseñar políticas públicas para el mejoramiento de los procesos educativos universitarios.

En este contexto, el presente trabajo aporta evidencia empírica acerca de los factores que determinan el rendimiento académico de los estudiantes de la Universidad de Nariño. Se utilizan datos de corte transversal, pertenecientes a la prueba Saber Pro 2012 a 2014, de las áreas de formación que ofrece la Universidad de Nariño, mismos que están disponibles en las bases de datos del Instituto Colombiano para la Evaluación de la Educación (ICFES).

Inicialmente, se describen las características socioeconómicas, las cuales, posiblemente generan brechas de rendimiento académico de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014, información recogida por el ICFES a través de los formularios de inscripción.

Posteriormente, con el fin de tener una comprensión preliminar de la relación entre los datos, se realiza un análisis de correlación entre las cinco competencias genéricas, y, finalmente, se establece el efecto que tiene en el desempeño de dichas competencias, las variables: institución, programa, edad, género, estado civil, estrato social, hogar actual, cabeza de familia y trabajo laboral.

3.1 CARACTERÍSTICAS SOCIOECONÓMICAS

En el período 2012-2014, un total de 3.823 estudiantes de la Universidad de Nariño presentaron las pruebas Saber Pro. La tabla 2, presenta las características socioeconómicas de dichos estudiantes. Por género, la mayoría son hombres, en un 53.6%; por edad, el mayor porcentaje se encuentra en el rango de 20 a 29 años, con un 86.8%. Igualmente, el estado civil prevalente es de solteros con un 91.3%, observándose porcentajes similares entre estudiantes casados y en unión libre, que suman aproximadamente un 8%. En una altísima mayoría, los estudiantes son de estratos bajos; aproximadamente, las tres cuartas partes de los estudiantes (74.5%) pertenecen a los estratos sociales más bajos (uno y dos), un 20.1% al estrato tres y solo el 4.2% ubican en los estratos medio y alto. Aunque la mayoría de estudiantes vive en su hogar habitual o permanente (57.4%), un número importante, que representa el 42.6%, reside en un hogar temporal por razones de estudio. Un porcentaje destacable de estudiantes (12.1%) es cabeza de familia y un 39.1% trabaja, de los cuales, el 24.3% lo hace para pagar sus estudios.

Tabla 2. Características Socioeconómicas de estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

VARIABLE SOCIOECONÓMICA		N	%
Género	Femenino	1773	46,4%
	Masculino	2046	53,6%
Grupos de edad	Menor de 20 años	18	0,5%
	20-29	3313	86,8%
	30-39	387	10,1%
	40-9 81	81	2,1%
	50 o más años	20	0,5%

VARIABLE SOCIOECONÓMICA		N	%
Estado Civil	Soltero(a)	3485	91,3%
	Casado(a)	164	4,3%
	Viudo(a)	1	0,0%
	Separado(a) Divorciado(a)	21	0,6%
	Unión libre	146	3,8%
Estrato social	Uno	1225	32,1%
	Dos	1616	42,4%
	Tres	799	20,9%
	Cuatro	146	3,8%
	Cinco	14	0,4%
	Seis	0	0,0%
	Rural	14	0,4%
Hogar actual	Habitual o permanente	2195	57,4%
	Temporal por estudio	1628	42,6%
Es cabeza de familia	No	3362	87,9%
	Si	461	12,1%
Trabaja	No	2320	60,9%
	Si, por Pago matrícula	925	24,3%
	Si, por Práctica estudio	177	4,6%
	Si, por Experiencia	387	10,2%

Fuente: elaboración propia

La tabla 3, presenta los programas académicos a los que pertenecen los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro, en los tres años del estudio.

Tabla 3. Programas Académicos de estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

PROGRAMA ACADÉMICO	N	%
ADMINISTRACIÓN DE EMPRESAS	314	8,2%
ARQUITECTURA	94	2,5%
ARTES VISUALES	78	2,0%
BIOLOGÍA	37	1,0%
COMERCIO INTERNACIONAL Y MERCADEO	94	2,5%
DERECHO	338	8,8%
DISEÑO GRÁFICO Y MULTIMEDIAL	107	2,8%
DISEÑO INDUSTRIAL	112	2,9%

PROGRAMA ACADÉMICO	N	%
ECONOMÍA	214	5,6%
FÍSICA	13	0,3%
GEOGRAFÍA APLICADA A LA ORGANIZACIÓN DEL ESPACIO Y PLANIFICACIÓN REGIONAL	49	1,3%
INGENIERÍA AGROFORESTAL	132	3,5%
INGENIERÍA AGROINDUSTRIAL	161	4,2%
INGENIERÍA AGRONÓMICA	106	2,8%
INGENIERÍA CIVIL	213	5,6%
INGENIERÍA DE SISTEMAS	133	3,5%
INGENIERÍA ELECTRÓNICA	147	3,8%
INGENIERÍA EN PRODUCCIÓN ACUÍCOLA	49	1,3%
LICENCIATURA EN INGLÉS Y FRANCÉS	65	1,7%
LICENCIATURA EN ARTES VISUALES	39	1,0%
LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN HUMANIDADES, LENGUA CASTELLANA E INGLÉS	90	2,4%
LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS NATURALES Y EDUCACIÓN AMBIENTAL	175	4,6%
LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS SOCIALES	52	1,4%
LICENCIATURA EN FILOSOFÍA Y LETRAS	68	1,8%
LICENCIATURA EN INFORMÁTICA	95	2,5%
LICENCIATURA EN LENGUA CASTELLANA Y LITERATURA	215	5,6%
LICENCIATURA EN MATEMÁTICAS	50	1,3%
LICENCIATURA EN MÚSICA	91	2,4%
MEDICINA	74	1,9%
MEDICINA VETERINARIA	113	3,0%
PSICOLOGÍA	87	2,3%
QUÍMICA	61	1,6%
SOCIOLOGÍA	79	2,1%
TECNOLOGÍA EN COMPUTACIÓN	2	0,1%
TECNOLOGÍA EN PROMOCIÓN DE LA SALUD	1	0,0%
ZOOTECNIA	77	2,0%

Fuente: elaboración propia

3.2 CORRELACIÓN ENTRE LAS CINCO COMPETENCIAS GENÉRICAS

A través del coeficiente de correlación de Pearson, se establece cómo se asocian linealmente, entre sí, los resultados obtenidos en las cinco competencias genéricas por los estudiantes de la Universidad de Nariño que presentaron las pruebas en el período 2012-2014. Los resultados se indican en la tabla 4.

Tabla 4. Matriz de correlaciones de las Competencias genéricas

Competencias	Coeficiente de Pearson	Inglés	Lectura crítica	Razonamiento cuantitativo	Competencias Ciudadanas
Comunicación escrita	Correlación	0,187**	0,281**	0,127**	0,287**
	valor p	0,000	0,000	0,000	0,000
	N	3807	3807	3807	3807
Inglés	Correlación		0,358**	0,335**	0,314**
	valor p		0,000	0,000	0,000
	N		3824	3824	3824
Lectura crítica	Correlación			0,426**	0,595**
	valor p			0,000	0,000
	N			3824	3824
Razonamiento cuantitativo	Correlación				0,371**
	valor p				0,000
	N				3824

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

Fuente: elaboración propia

De acuerdo a la tabla anterior, todas las correlaciones resultan altamente significativas ($p\text{-valor} < 0,01$) por el tamaño de los datos. Siguiendo la clasificación de Cohen (1998), para la interpretación del coeficiente de Pearson, se observó que lectura crítica presenta una correlación alta ($r > 0,5$) con competencias ciudadanas, y una correlación moderada ($0,3 < r \leq 0,5$) con razonamiento cuantitativo; las correlaciones de inglés con el resto de competencias, excepto con comunicación escrita, son moderadas; también

es moderada la correlación entre razonamiento cuantitativo y competencias ciudadanas. En cuanto a comunicación escrita, se observó correlación baja con todas las competencias ($r < = 0,3$).

3.3 EFECTO INSTITUCIONAL EN LAS COMPETENCIAS GENÉRICAS EN EL CONTEXTO NACIONAL

Para establecer el “Efecto Institucional” de la Universidad de Nariño en el contexto Nacional, se elaboró un ranking de las Universidades e Instituciones de Educación Superior que participaron en las pruebas Saber Pro durante el período 2012 a 2014. Se excluyó de dicha clasificación, las instituciones cuya participación de estudiantes fue inferior a un total de 200 participantes en cada año de la prueba, por considerar dicho tamaño como no representativo.

Para la elaboración del ranking se utilizó el *estadístico d de Cohen* (1998) que permite calcular el Tamaño del Efecto Institucional a partir de las diferencias estandarizadas entre los promedios en cada uno de las instituciones, frente a la IES que alcanzó el máximo puntaje promedio en cada una de las cinco competencias genéricas, la cual se toma como referencia; así:

$$d_i = \frac{\bar{X}_0 - \bar{X}_i}{S_{i0}}$$

Donde:

$$S_{i0} = \sqrt{\frac{(n_0 - 1)S_0^2 + (n_i - 1)S_i^2}{n_0 + n_i - 2}}$$

d_i = tamaño del efecto de la *institución i* con relación al *referente*.

X_0 = Media de la institución de referencia (de mayor puntaje).

X_i = Media de la institución *i*.

S_0^2 = Varianza de la institución de referencia.

S_i^2 = Varianza de la institución de i.

n_o = número de estudiantes que presentaron las pruebas de la institución de referencia

n_i = número de estudiantes que presentaron las pruebas de la institución i

Cohen (1998), propone la siguiente escala para interpretar las diferencias en unidades estándar, obtenidas con el estadístico de la expresión matemática anterior:

- [0.0, 0.2] diferencia trivial o irrelevante.
- (0.2, 0.5] diferencia pequeña.
- (0.5, 0.8] diferencia moderada.
- (0.8, infinito) diferencia grande.

3.3.1 Efecto Institucional en Comunicación Escrita

En el período estudiado, el rendimiento de los estudiantes de la Universidad de Nariño en comunicación escrita presentó resultados por encima del promedio nacional, aunque en los dos últimos años se observó una disminución en dicho promedio (ver figura 3).

La tabla 5 presenta el ranking del desempeño promedio de las 100 primeras Universidades e Instituciones de Educación Superior en Comunicación Escrita en el período 2012-2014; en dicha competencia, la Universidad de Nariño ocupó la posición 23, presentando una diferencia moderada, con 0,578 desviaciones estándar, respecto a la Universidad Externado de Colombia, la cual obtuvo el mayor puntaje en esta competencia.

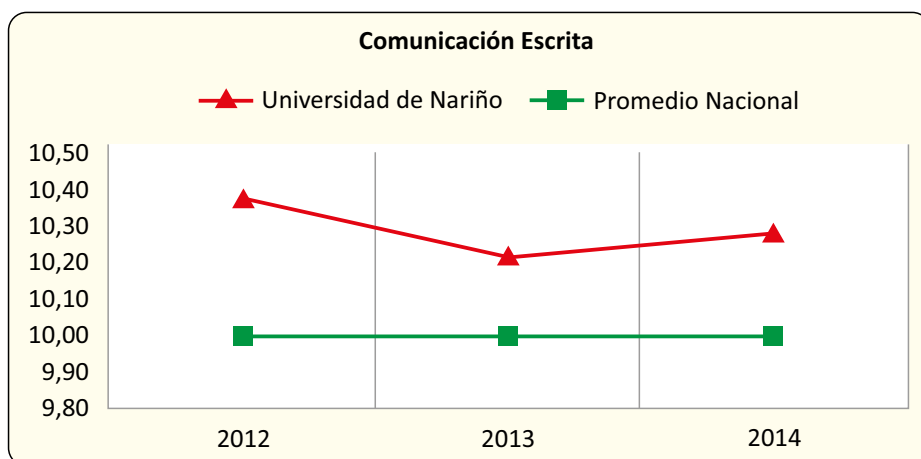


Figura 3. Tendencia del Rendimiento en Comunicación Escrita de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Tabla 5. Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Comunicación Escrita pruebas Saber Pro 2012-2014. No. INSTITUCIÓN N Media Desviación estándar d Cohen

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
1	UNIVERSIDAD EXTERNADO DE COLOMBIA	3037	10,88	0,99	-
2	COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO	4667	10,82	1,00	0,065
3	UNIVERSIDAD DE LOS ANDES	7860	10,80	1,01	0,081
4	UNIVERSIDAD DE LA SABANA	3708	10,70	0,94	0,187
5	UNIVERSIDAD ICESI	2298	10,66	0,95	0,227
6	UNIVERSIDAD NACIONAL DE COLOMBIA	17673	10,57	1,07	0,291
7	UNIVERSIDAD PEDAGÓGICA NACIONAL	4350	10,50	1,07	0,370
8	PONTIFICIA UNIVERSIDAD JAVERIANA	12176	10,49	1,04	0,379
9	UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS	8755	10,43	1,07	0,431

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
10	UNIVERSIDAD DE ANTIOQUIA	13169	10,41	1,07	0,445
11	UNIVERSIDAD SERGIO ARBOLEDA	3193	10,41	1,01	0,472
12	UNIVERSIDAD DEL CAUCA	4372	10,38	1,06	0,492
13	FUNDACIÓN UNIVERSITARIA KONRAD LORENZ	1186	10,39	0,99	0,493
14	ESCUELA DE INGENIERÍA DE ANTIOQUIA	743	10,38	0,98	0,506
15	UNIVERSIDAD DE LA SALLE	7016	10,38	0,98	0,512
16	UNIVERSIDAD-COLEGIO MAYOR DE CUNDINAMARCA	2447	10,37	1,01	0,512
17	FUNDACIÓN UNIVERSIDAD DEL NORTE	4519	10,35	1,05	0,516
18	UNIVERSIDAD EAFIT	4230	10,37	0,99	0,516
19	INSTITUTO DE CIENCIAS DE LA SALUD	981	10,33	1,03	0,555
20	UNIVERSIDAD MILITAR NUEVA GRANADA	5997	10,32	1,00	0,564
21	UNIVERSIDAD DEL VALLE	9995	10,27	1,09	0,576
22	ESCUELA COLOMBIANA DE INGENIERÍA	1551	10,32	0,97	0,576
23	UNIVERSIDAD DE NARIÑO	3825	10,28	1,07	0,578
24	UNIVERSIDAD INDUSTRIAL DE SANTANDER	8747	10,28	1,06	0,578
25	UNIVERSIDAD SANTO TOMÁS	11821	10,28	1,04	0,587
26	UNIVERSIDAD CATÓLICA DE COLOMBIA	5018	10,30	1,00	0,588
27	UNIVERSIDAD DE CALDAS	4707	10,28	1,04	0,590
28	FUNDACIÓN UNIVERSIDAD CENTRAL	3186	10,29	1,02	0,590
29	UNIVERSIDAD AUTÓNOMA LATINOAMERICANA -UNAUULA-	1499	10,29	1,00	0,598
30	UNIVERSIDAD EAN	2462	10,26	1,01	0,620

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
31	UNIVERSIDAD PONTIFICIA BOLÍ- VARIANA	8664	10,23	1,04	0,632
32	UNIVERSIDAD DE MEDELLÍN	4494	10,24	1,01	0,639
33	UNIVERSIDAD TECNOLÓGICA DE PEREIRA	4959	10,22	1,07	0,641
34	UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA -UNAB-	3320	10,21	1,06	0,649
35	UNIVERSIDAD DE BOYACÁ -UNI- BOYACÁ-	1719	10,24	0,98	0,649
36	UNIVERSIDAD CATÓLICA POPU- LAR DEL RISARALDA	1052	10,21	1,15	0,650
37	UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA	10588	10,23	0,99	0,656
38	FUNDACIÓN UNIVERSITARIA LUIS AMIGO FUNLAM	4399	10,23	0,98	0,663
39	FUNDACIÓN UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO	5971	10,18	1,07	0,673
40	UNIVERSIDAD DE MANIZALES	1909	10,20	0,98	0,695
41	INSTITUCIÓN UNIVERSITARIA DE ENVIGADO	1585	10,18	1,00	0,704
42	FUNDACIÓN UNIVERSITARIA MONSERRATE	1077	10,17	1,04	0,705
43	FUNDACIÓN UNIVERSIDAD AU- TÓNOMA DE COLOMBIA -FUAC-	3575	10,16	1,02	0,714
44	UNIVERSIDAD LA GRAN COLOM- BIA	7091	10,17	1,00	0,716
45	FUNDACIÓN UNIVERSIDAD DE AMÉRICA	1693	10,16	0,97	0,731
46	FUNDACIÓN UNIVERSITARIA EM- PRESARIAL DE LA CÁMARA DE COMERCIO DE BOGOTÁ	716	10,16	0,96	0,734
47	UNIVERSIDAD SURCOLOMBIANA	4218	10,12	1,06	0,738
48	UNIVERSIDAD CATÓLICA DE ORIENTE	2022	10,13	0,98	0,764

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
49	INSTITUCIÓN UNIVERSITARIA CENTRO DE ESTUDIOS SUPE- RIORES MARÍA GORETTI	1817	10,09	1,11	0,764
50	UNIVERSIDAD DE CUNDINA- MARCA -UDEC-	4404	10,10	1,02	0,769
51	UNIVERSIDAD LIBRE	15386	10,08	1,05	0,770
52	UNIVERSIDAD DE SAN BUENA- VENTURA	7659	10,08	1,05	0,775
53	FUNDACIÓN UNIVERSITARIA AGRARIA DE COLOMBIA -UNIA- GRARIA-	1588	10,10	1,01	0,782
54	FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD	932	10,11	0,97	0,783
55	UNIVERSIDAD CENTRAL	1858	10,10	0,99	0,789
56	UNIVERSIDAD DE IBAGUÉ -CO- RUNIVERSITARIA-	3105	10,09	1,00	0,793
57	UNIVERSIDAD AUTÓNOMA DE OCCIDENTE	3127	10,08	1,02	0,796
58	UNIVERSIDAD MANUELA BEL- TRÁN -UMB-	2444	10,08	0,98	0,816
59	UNIVERSIDAD INCCA DE CO- LOMBIA	2201	10,06	1,03	0,817
60	COLEGIO MAYOR DE ANTIO- QUIA	1034	10,07	1,00	0,819
61	POLITÉCNICO COLOMBIANO	3830	10,05	1,02	0,824
62	FUNDACIÓN UNIVERSITARIA JUAN N. CORPAS	637	10,05	0,99	0,835
63	FUNDACIÓN UNIVERSITARIA DE SAN GIL -UNISANGIL-	1669	10,06	0,97	0,836
64	UNIVERSIDAD DEL ATLÁNTICO	7912	10,02	1,04	0,841
65	ESCUELA SUPERIOR DE ADMI- NISTRACIÓN PÚBLICA -ESAP-	6845	9,99	1,08	0,843
66	FUNDACIÓN UNIVERSITARIA JUAN DE CASTELLANOS	823	10,03	1,06	0,848

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
67	UNIVERSIDAD DEL QUINDÍO	7447	10,02	1,01	0,851
68	UNIVERSIDAD DE CIENCIAS APLICADAS Y AMBIENTALES -UDCA-	2395	10,05	0,97	0,852
69	UNIVERSIDAD EL BOSQUE	3527	10,03	0,99	0,857
70	UNIVERSIDAD AUTÓNOMA DE MANIZALES	1496	10,02	1,02	0,860
71	FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES	7635	10,02	1,01	0,860
72	CORPORACIÓN ESCUELA DE ARTES Y LETRAS	708	10,02	1,05	0,861
73	CORPORACIÓN UNIVERSITARIA IBEROAMERICANA	1444	10,04	0,96	0,862
74	UNIVERSIDAD MARIANA	3175	9,97	1,12	0,865
75	FUNDACIÓN UNIVERSITARIA -INPAHU-	2075	9,99	1,10	0,865
76	CORPORACIÓN UNIVERSITARIA AUTÓNOMA DEL CAUCA	1086	10,01	1,03	0,868
77	UNIVERSIDAD DE LOS LLANOS	2420	10,02	0,99	0,870
78	FUNDACIÓN UNIVERSITARIA DE POPAYÁN	1058	10,01	1,03	0,871
79	INSTITUCIÓN UNIVERSITARIA FUNDACIÓN EDUCATIVA ESU- MER	2403	10,01	0,99	0,877
80	CORPORACIÓN UNIVERSITARIA REPUBLICANA	1321	10,02	0,97	0,878
81	UNIVERSIDAD CATÓLICA DE MANIZALES	1480	10,01	0,99	0,879
82	UNIDAD CENTRAL DEL VALLE DEL CAUCA	1663	10,01	1,00	0,882
83	FUNDACIÓN UNIVERSITARIA CATÓLICA DEL NORTE	1639	9,97	1,11	0,887
84	FUNDACIÓN UNIVERSITARIA PA- NAMERICANA	4147	9,97	1,05	0,892

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
85	POLITÉCNICO GRANCOLOMBIA- NO	10098	9,99	1,00	0,895
86	UNIVERSIDAD DE CARTAGENA	7568	9,91	1,12	0,901
87	UNIVERSITARIA AGUSTINIANA -UNIAGUSTINIANA-	1895	10,00	0,96	0,902
88	FUNDACIÓN UNIVERSITARIA -CEIPA-	260	9,99	0,96	0,910
89	CORPORACIÓN UNIVERSIDAD PILOTO DE COLOMBIA	4058	9,97	1,01	0,911
90	CORPORACIÓN UNIVERSITARIA MINUTO DE DIOS	772	9,97	1,02	0,914
91	UNIMINUTO - CORPORACIÓN UNIVERSITARIA MINUTO DE DIOS	3246	9,95	1,01	0,927
92	CORPORACIÓN UNIVERSITARIA UNITEC	1742	9,94	1,06	0,928
93	CORPORACIÓN UNIVERSITARIA DE CIENCIA Y DESARROLLO	2946	9,94	1,03	0,935
94	UNIVERSIDAD COOPERATIVA DE COLOMBIA	23533	9,94	1,01	0,936
95	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	8650	9,94	1,01	0,945
96	UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA -UNAD-	13922	9,89	1,06	0,951
97	COLEGIO ODONTOLÓGICO CO- LOMBIANO	690	9,93	1,03	0,956
98	ESCUELA COLOMBIANA DE CA- RRERAS INDUSTRIALES	6530	9,92	1,01	0,959
99	UNIVERSIDAD DE SUCRE	2112	9,91	1,04	0,963
100	UNIVERSIDAD DE PAMPLONA	13241	9,87	1,06	0,970

Fuente: elaboración propia

3.3.2 Efecto Institucional en Inglés

En las pruebas Saber Pro realizadas entre 2012 y 2014, el desempeño en inglés de los estudiantes de la Universidad de Nariño, presentó resultados por encima del promedio nacional, observando un mejoramiento en los promedios de los últimos años (ver figura 4).

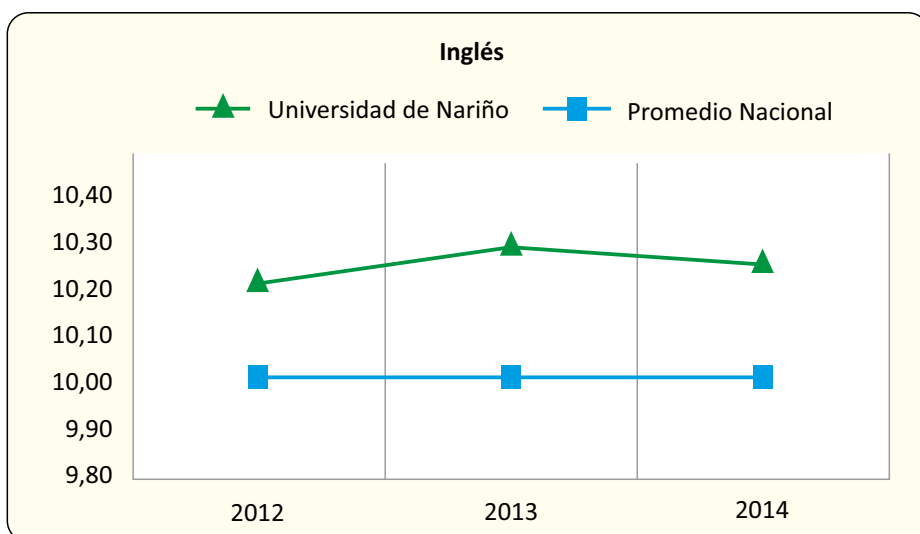


Figura 4. Tendencia del desempeño en inglés de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Según la tabla 6, durante el periodo estudiado, la Universidad de Nariño ocupó la posición 51 en el ranking del desempeño promedio de Universidades e Instituciones de Educación Superior en Inglés. Se detectó una diferencia muy grande de 2,399 desviaciones estándar, con respecto a la Universidad de los Andes, que ocupó el primer lugar.

Tabla 6. Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Inglés pruebas Saber Pro 2012-2014.

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
1	UNIVERSIDAD DE LOS ANDES	7860	13,36	1,34	-
2	ESCUELA DE INGENIERÍA DE ANTIOQUIA	743	12,96	1,40	0,293
3	UNIVERSIDAD EAFIT	4230	12,48	1,40	0,646
4	COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO	4667	12,45	1,43	0,660
5	UNIVERSIDAD DE LA SABANA	3708	12,30	1,38	0,779
6	UNIVERSIDAD ICESI	2298	12,16	1,51	0,867
7	FUNDACIÓN UNIVERSIDAD DEL NORTE	4519	12,05	1,59	0,909
8	UNIVERSIDAD EXTERNADO DE COLOMBIA	3037	12,04	1,48	0,951
9	PONTIFICIA UNIVERSIDAD JAVERIANA	12176	11,99	1,64	0,892
10	INSTITUTO DE CIENCIAS DE LA SALUD	981	11,78	1,69	1,142
11	UNIVERSIDAD NACIONAL DE COLOMBIA	17673	11,53	1,58	1,214
12	ESCUELA COLOMBIANA DE INGENIERÍA	1551	11,40	1,37	1,452
13	FUNDACIÓN UNIVERSIDAD DE BOGOTÁ-JORGE TADEO LOZANO	5971	11,32	1,49	1,444
14	UNIVERSIDAD SERGIO ARBOLEDA	3193	11,25	1,46	1,533
15	FUNDACIÓN UNIVERSIDAD DE AMÉRICA	1693	11,15	1,35	1,645
16	UNIVERSIDAD EAN	2462	11,03	1,60	1,653
17	UNIVERSIDAD PONTIFICIA BOLÍVARIANA	8664	10,95	1,69	1,572
18	UNIVERSIDAD INDUSTRIAL DE SANTANDER	8747	10,89	1,52	1,720
19	UNIVERSIDAD EL BOSQUE	3527	10,87	1,38	1,840
20	UNIVERSIDAD DE MEDELLÍN	4494	10,81	1,45	1,842

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
21	UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA -UNAB-	3320	10,81	1,45	1,851
22	FUNDACIÓN UNIVERSITARIA EMPRESARIAL DE LA CÁMARA DE COMERCIO DE BOGOTÁ	716	10,80	1,26	1,920
23	FUNDACIÓN UNIVERSITARIA KONRAD LORENZ	1186	10,77	1,19	1,955
24	UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS	8755	10,75	1,37	1,922
25	UNIVERSIDAD MILITAR NUEVA GRANADA	5997	10,73	1,33	1,964
26	UNIVERSIDAD AUTÓNOMA DE OCCIDENTE	3127	10,72	1,41	1,939
27	UNIVERSIDAD TECNOLÓGICA DE PEREIRA	4959	10,71	1,53	1,867
28	UNIVERSIDAD DE LA SALLE	7016	10,70	1,29	2,017
29	UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR	2006	10,67	1,44	1,974
30	UNIVERSIDAD DE ANTIOQUIA	13169	10,61	1,50	1,908
31	CORPORACIÓN UNIVERSITARIA UNITEC	1742	10,59	1,24	2,090
32	UNIVERSIDAD DEL VALLE	9995	10,58	1,47	1,966
33	UNIVERSIDAD CATÓLICA POPULAR DEL RISARALDA	1052	10,56	1,36	2,082
34	FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD	932	10,55	1,23	2,110
35	CORPORACIÓN ESCUELA DE ARTES Y LETRAS	708	10,52	1,31	2,119
36	FUNDACIÓN UNIVERSITARIA JUAN N. CORPAS	637	10,50	1,29	2,133
37	UNIVERSIDAD SANTO TOMÁS	11821	10,40	1,27	2,274
38	UNIVERSIDAD CENTRAL	1858	10,40	1,12	2,275
39	UNIVERSIDAD AUTÓNOMA DEL CARIBE	5076	10,38	1,31	2,238
40	UNIVERSIDAD DE SAN BUENAVENTURA	7659	10,34	1,37	2,221

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
41	UNIVERSIDAD DE IBAGUÉ -CORUNIVERSITARIA-	3105	10,33	1,24	2,305
42	UNIVERSIDAD DEL CAUCA	4372	10,32	1,26	2,310
43	UNIVERSIDAD PEDAGÓGICA NACIONAL	4350	10,32	1,43	2,211
44	FUNDACIÓN UNIVERSIDAD CENTRAL	3186	10,31	1,18	2,348
45	UNIVERSIDAD DE CALDAS	4707	10,31	1,36	2,258
46	POLITÉCNICO GRANCOLOMBIANO	10098	10,28	1,28	2,358
47	CORPORACIÓN UNIVERSIDAD PILOTO DE COLOMBIA	4058	10,28	1,27	2,335
48	UNIVERSIDAD DEL ATLÁNTICO	7912	10,28	1,38	2,262
49	UNIVERSIDAD AUTÓNOMA DE MANIZALES	1496	10,28	1,33	2,301
50	UNIVERSIDAD DE MANIZALES	1909	10,26	1,24	2,341
51	UNIVERSIDAD DE NARIÑO	3825	10,26	1,18	2,399
52	INSTITUCIÓN UNIVERSITARIA DE ENVIGADO	1585	10,20	1,27	2,375
53	UNIVERSIDAD-COLEGIO MAYOR DE CUNDINAMARCA	2447	10,18	0,98	2,511
54	INSTITUCIÓN UNIVERSITARIA FUNDACIÓN EDUCATIVA ESUMER	2403	10,17	1,20	2,436
55	UNIVERSIDAD CATÓLICA DE COLOMBIA	5018	10,17	1,07	2,569
56	UNIVERSIDAD DE CIENCIAS APLICADAS Y AMBIENTALES UDCA.	2395	10,14	1,07	2,506
57	INSTITUTO TECNOLÓGICO DE SOLEDAD ATLÁNTICO	908	10,14	0,93	2,468
58	FUNDACIÓN UNIVERSITARIA -CEIPA-	1260	10,13	1,39	2,397
59	FUNDACIÓN UNIVERSIDAD AUTÓNOMA DE COLOMBIA -FUAC-	3575	10,12	1,03	2,581

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
60	UNIVERSIDAD SURCOLOMBIANA	4218	10,11	1,32	2,432
61	UNIVERSIDAD LIBRE	15386	10,10	1,16	2,669
62	UNIVERSIDAD DE CARTAGENA	7568	10,09	1,30	2,475
63	COLEGIO ODONTOLÓGICO COLOMBIANO	690	10,08	1,04	2,483
64	ESCUELA DE ADMINISTRACIÓN Y MERCADOTECNIA DEL QUINDÍO	1449	10,07	1,11	2,512
65	INSTITUCIÓN UNIVERSITARIA SALAZAR Y HERRERA	1340	10,07	1,20	2,486
66	INSTITUTO TECNOLÓGICO METROPOLITANO	6131	10,07	1,02	2,714
67	ESCUELA COLOMBIANA DE CARRERAS INDUSTRIALES	6530	10,06	0,95	2,794
68	UNIDAD CENTRAL DEL VALLE DEL CAUCA	1663	10,06	1,15	2,521
69	UNIVERSIDAD MANUELA BELTRÁN -UMB-	2444	10,05	1,08	2,573
70	FUNDACIÓN ACADEMIA DE DIBUJO PROFESIONAL	669	10,05	1,26	2,480
71	UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA	10588	10,04	1,11	2,729
72	FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES	7635	10,04	1,04	2,760
73	UNIVERSIDAD CATÓLICA DE MANIZALES	1480	10,04	1,16	2,526
74	UNIVERSIDAD LA GRAN COLOMBIA	7091	10,02	1,10	2,704
75	UNIVERSIDAD SANTIAGO DE CALI	6529	10,02	1,25	2,569
76	FUNDACIÓN UNIVERSITARIA AGRARIA DE COLOMBIA -UNIAGRARIA-	1588	10,02	0,99	2,592

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
77	CORPORACIÓN UNIVERSITARIA DE INVESTIGACIÓN Y DESARROLLO	2374	10,01	1,03	2,628
78	FUNDACIÓN UNIVERSITARIA LUIS AMIGO FUNLAM	4399	10,00	1,23	2,579
79	COLEGIO MAYOR DE ANTIOQUIA	1034	10,00	1,10	2,555
80	UNIVERSITARIA AGUSTINIANA-UNIAGUSTINIANA	1895	9,98	0,90	2,661
81	UNIVERSIDAD DEL QUINDÍO	7447	9,98	1,27	2,578
82	POLITÉCNICO COLOMBIANO	3830	9,98	1,11	2,657
83	UNIVERSIDAD DE BOYACÁ UNIBOYACÁ	1719	9,98	0,86	2,663
84	UNIVERSIDAD DE CÓRDOBA	5235	9,97	1,26	2,587
85	UNIVERSIDAD DEL SINÚ ELÍAS BECHARA ZAINUM - UNISINÚ-	3028	9,95	1,03	2,705
86	UNIVERSIDAD AUTÓNOMA LATINOAMERICANA -UNAUULA-	1499	9,94	1,01	2,638
87	INSTITUTO TÉCNICO CENTRAL	1251	9,94	0,87	2,654
88	FUNDACIÓN UNIVERSITARIA-INPAHU	2075	9,94	0,95	2,693
89	UNIVERSIDAD DE SANTANDER	3898	9,93	1,09	2,714
90	FUNDACIÓN UNIVERSITARIA PANAMERICANA	4147	9,93	0,95	2,812
91	FUNDACIÓN UNIVERSITARIA SAN MARTÍN	12850	9,91	1,07	2,916
92	UNIVERSIDAD INCCA DE COLOMBIA	2201	9,91	1,02	2,698
93	CORPORACIÓN POLITÉCNICO DE LA COSTA ATLÁNTICA	1188	9,90	1,30	2,588
94	CORPORACIÓN UNIVERSITARIA DE LA COSTA CUC	3539	9,90	1,08	2,735
95	FUNDACIÓN UNIVERSITARIA TECNOLÓGICO COMFENALCO - CARTAGENA	3575	9,88	0,99	2,798

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
96	FUNDACIÓN PARA LA EDUCACIÓN SUPERIOR SAN MATEO FUNDACIÓN SAN MATEO	874	9,86	0,96	2,670
97	UNIVERSIDAD CATÓLICA DE ORIENTE	2022	9,86	1,26	2,640
98	FUNDACIÓN UNIVERSITARIA DEL AREA ANDINA	6734	9,86	1,02	2,902
99	FUNDACIÓN UNIVERSITARIA DE SAN GIL -UNISANGIL-	1669	9,86	0,86	2,757
100	UNIVERSIDAD DE CUNDINAMARCA -UDEC-	4404	9,85	0,79	2,983

Fuente: elaboración propia

3.3.3 Efecto Institucional en Lectura Crítica

En el período estudiado, el desempeño de los estudiantes de la Universidad de Nariño en Lectura Crítica presentó resultados por encima del promedio nacional, manteniéndose constante su tendencia (ver figura 5).

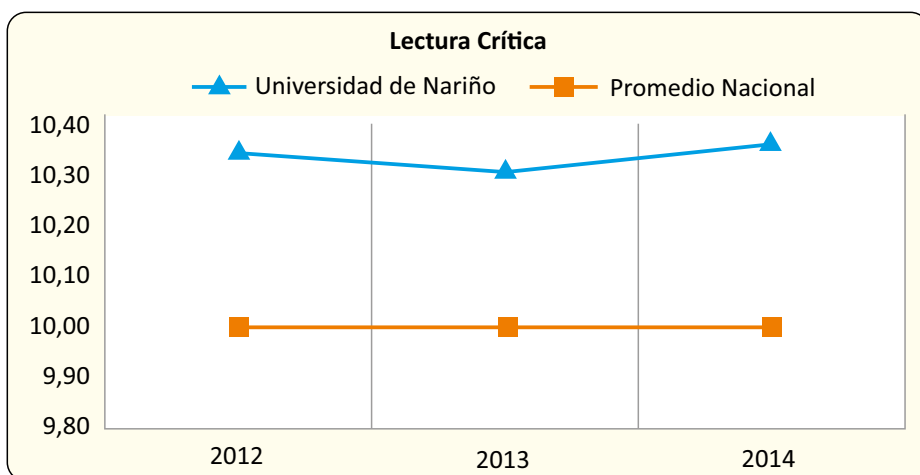


Figura 5. Tendencia del Rendimiento en Lectura Crítica de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Según el ranking que presenta la tabla 7, en esta competencia la Universidad de Nariño ocupó la posición 31, observando un tamaño de efecto grande de 0,868 desviaciones estándares, con respecto a la Universidad de los Andes, institución que, al igual que en la competencia de inglés, ocupó el primer lugar.

Tabla 7. Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Lectura Crítica pruebas Saber Pro 2012-2014.

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
1	UNIVERSIDAD DE LOS ANDES	7860	11,21	1,02	-
2	UNIVERSIDAD NACIONAL DE COLOMBIA	17673	11,07	0,94	0,143
3	COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO	4667	11,04	0,94	0,175
4	UNIVERSIDAD EXTERNADO DE COLOMBIA	3037	10,97	0,89	0,239
5	UNIVERSIDAD ICESI	2298	10,92	0,93	0,285
6	ESCUELA DE INGENIERÍA DE ANTIOQUIA	743	10,91	0,87	0,300
7	INSTITUTO DE CIENCIAS DE LA SALUD	981	10,75	0,91	0,454
8	UNIVERSIDAD DE ANTIOQUIA	13169	10,76	0,94	0,467
9	UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS	8755	10,75	0,88	0,486
10	UNIVERSIDAD EAFIT-	4230	10,67	1,02	0,523
11	PONTIFICIA UNIVERSIDAD JAVERIANA	12176	10,69	0,96	0,528
12	UNIVERSIDAD DE LA SABANA	3708	10,66	0,93	0,549
13	UNIVERSIDAD DEL VALLE	9995	10,66	0,97	0,557
14	ESCUELA COLOMBIANA DE INGENIERÍA	1551	10,64	0,86	0,576
15	UNIVERSIDAD INDUSTRIAL DE SANTANDER	8747	10,62	0,99	0,592
16	FUNDACIÓN UNIVERSIDAD DEL NORTE	4519	10,55	1,08	0,638
17	UNIVERSIDAD PEDAGÓGICA NACIONAL	4350	10,56	0,93	0,658
18	UNIVERSIDAD DE MEDELLÍN	4494	10,47	1,01	0,730

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
19	UNIVERSIDAD-COLEGIO MAYOR DE CUNDINAMARCA	2447	10,49	0,86	0,736
20	UNIVERSIDAD DEL CAUCA	4372	10,44	0,93	0,774
21	FUNDACIÓN UNIVERSITARIA KONRAD LORENZ	1186	10,41	0,91	0,791
22	FUNDACIÓN UNIVERSITARIA JUAN N. CORPAS	637	10,41	0,81	0,797
23	FUNDACIÓN UNIVERSIDAD DE AMÉRICA	1693	10,41	0,86	0,803
24	UNIVERSIDAD PONTIFICIA BOLÍVARIANA	8664	10,38	1,02	0,816
25	FUNDACIÓN UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO	5971	10,39	0,95	0,827
26	FUNDACIÓN UNIVERSITARIA EMPRESARIAL DE LA CÁMARA DE COMERCIO DE BOGOTÁ	716	10,37	0,85	0,830
27	UNIVERSIDAD SERGIO ARBOLEDA	3193	10,38	0,94	0,831
28	UNIVERSIDAD MILITAR-NUEVA GRANADA	5997	10,40	0,90	0,835
29	UNIVERSIDAD DE CALDAS	4707	10,35	0,96	0,855
30	UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA -UNAB-	3320	10,34	1,00	0,860
31	UNIVERSIDAD DE NARIÑO	3825	10,33	0,98	0,868
32	UNIVERSIDAD TECNOLÓGICA DE PEREIRA	4959	10,31	1,01	0,887
33	UNIVERSIDAD CENTRAL	1858	10,30	0,95	0,906
34	UNIVERSIDAD AUTÓNOMA DE OCCIDENTE	3127	10,29	0,94	0,917
35	UNIVERSIDAD AUTÓNOMA LATINOAMERICANA -UNAUULA-	1499	10,28	0,92	0,929
36	CORPORACIÓN ESCUELA DE ARTES Y LETRAS	708	10,27	0,89	0,934
37	UNIVERSIDAD EAN	2462	10,28	0,88	0,935
38	UNIVERSIDAD DE LA SALLE	7016	10,28	0,91	0,952
39	UNIVERSIDAD EL BOSQUE	3527	10,27	0,89	0,960
40	FUNDACIÓN UNIVERSIDAD CENTRAL	3186	10,27	0,86	0,967

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
41	FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD	932	10,23	0,93	0,969
42	UNIVERSIDAD CATÓLICA POPULAR DEL RISARALDA	1052	10,21	0,99	0,988
43	UNIVERSIDAD CATÓLICA DE COLOMBIA	5018	10,23	0,91	1,002
44	UNIVERSIDAD SANTO TOMÁS	11821	10,23	0,93	1,010
45	UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR	2006	10,18	0,99	1,012
46	UNIVERSIDAD SURCOLOMBIANA	4218	10,19	0,96	1,024
47	COLEGIO MAYOR DE ANTIOQUIA	1034	10,16	0,90	1,041
48	UNIVERSIDAD DE CARTAGENA	7568	10,11	1,08	1,046
49	UNIVERSIDAD DEL ATLÁNTICO	7912	10,14	0,99	1,064
50	FUNDACIÓN UNIVERSIDAD AUTÓNOMA DE COLOMBIA -FUAC-	3575	10,15	0,91	1,071
51	UNIVERSIDAD DE MANIZALES	1909	10,12	0,94	1,082
52	UNIVERSIDAD DE LOS LLANOS	2420	10,13	0,90	1,085
53	FUNDACIÓN ACADEMIA DE DIBUJO PROFESIONAL	669	10,10	1,02	1,087
54	UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA	10588	10,15	0,93	1,090
55	UNIVERSIDAD DE SAN BUENAVENTURA	7659	10,13	0,95	1,098
56	UNIVERSIDAD LA GRAN COLOMBIA	7091	10,14	0,92	1,101
57	CORPORACIÓN UNIVERSITARIA UNITEC	1742	10,10	0,91	1,106
58	CORPORACIÓN UNIVERSIDAD PILOTO DE COLOMBIA	4058	10,11	0,92	1,109
59	UNIVERSIDAD AUTÓNOMA DE MANIZALES	1496	10,09	0,92	1,113

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
60	INSTITUCIÓN UNIVERSITARIA DE ENVIGADO	1585	10,09	0,91	1,114
61	UNIVERSIDAD DE CIENCIAS APLICADAS Y AMBIENTALES UDCA	2395	10,10	0,91	1,115
62	INSTITUTO TECNOLÓGICO METROPOLITANO	6131	10,11	0,90	1,129
63	UNIVERSIDAD LIBRE	15386	10,11	0,95	1,131
64	FUNDACIÓN UNIVERSITARIA LUIS AMIGO FUNLAM	4399	10,09	0,93	1,134
65	INSTITUTO TÉCNICO CENTRAL	1251	10,06	0,93	1,139
66	POLITÉCNICO COLOMBIANO	3830	10,09	0,89	1,139
67	UNIVERSIDAD DE IBAGUÉ -CORUNIVERSITARIA-	3105	10,08	0,90	1,140
68	FUNDACIÓN UNIVERSITARIA PANAMERICANA	4147	10,08	0,89	1,153
69	UNIVERSIDAD DE BOYACÁ UNIBOYACÁ	1719	10,06	0,86	1,156
70	ESCUELA COLOMBIANA DE CARRERAS INDUSTRIALES	6530	10,09	0,90	1,159
71	UNIVERSIDAD MANUELA BELTRÁN -UMB-	2444	10,05	0,90	1,165
72	POLITÉCNICO GRANCOLOMBIANO	10098	10,07	0,94	1,166
73	UNIDAD CENTRAL DEL VALLE DEL CAUCA	1663	10,03	0,95	1,167
74	FUNDACIÓN CENTRO DE INVESTIGACIÓN DOCENCIA Y CONSULTORÍA ADMINISTRATIVA -F-CIDCA-	1113	10,01	1,01	1,174
75	FUNDACIÓN UNIVERSITARIA -CEIPA-	1260	10,00	0,90	1,206
76	INSTITUCIÓN UNIVERSITARIA FUNDACIÓN EDUCATIVA -ESUMER-	2403	10,00	0,90	1,216
77	UNIVERSITARIA AGUSTINIANA -UNIAGUSTINIANA-	1895	10,00	0,85	1,220
78	COLEGIO ODONTOLÓGICO COLOMBIANO	690	9,97	0,85	1,226

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
79	FUNDACIÓN CENTRO COLOMBIANO DE ESTUDIOS PROFESIONALES, -F.C.E.C.E.P.	991	9,97	0,93	1,228
80	FUNDACIÓN UNIVERSITARIA AGRARIA DE COLOMBIA -UNIAGRARIA-	1588	9,98	0,88	1,228
81	FUNDACIÓN UNIVERSITARIA -INPAHU-	2075	9,97	0,93	1,236
82	UNIVERSIDAD DE SUCRE	2112	9,96	0,96	1,237
83	UNIVERSIDAD INCCA DE COLOMBIA	2201	9,98	0,89	1,241
84	CORPORACIÓN UNIVERSAL DE INVESTIGACIÓN Y TECNOLOGÍA -CORUNIVERSITEC-	1499	9,96	0,89	1,248
85	FUNDACIÓN UNIVERSITARIA MONSERRATE	1077	9,95	0,85	1,260
86	UNIVERSIDAD DEL QUINDÍO	7447	9,96	0,94	1,267
87	FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES	7635	9,99	0,89	1,272
88	UNIVERSIDAD DE CÓRDOBA	5235	9,93	0,97	1,283
89	FUNDACIÓN PARA LA EDUCACIÓN SUPERIOR SAN MATEO FUNDACIÓN SAN MATEO	874	9,92	0,87	1,284
90	UNIDADES TECNOLÓGICAS DE SANTANDER	4534	9,92	0,95	1,292
91	UNIVERSIDAD DE CUNDINAMARCA -UDEC-	4404	9,95	0,88	1,293
92	INSTITUTO TECNOLÓGICO PASCUAL BRAVO	1168	9,88	1,04	1,301
93	INSTITUCIÓN UNIVERSITARIA SALAZAR Y HERRERA	1340	9,89	0,98	1,304
94	INSTITUTO TECNOLÓGICO MUNICIPAL	1827	9,91	0,88	1,305
95	CORPORACIÓN UNIVERSITARIA DE INVESTIGACIÓN Y DESARROLLO	2374	9,90	0,96	1,306
96	CORPORACIÓN UNIVERSITARIA IBEROAMERICANA	1444	9,90	0,83	1,318

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
97	FUNDACIÓN DE EDUCACIÓN SUPERIOR SAN JOSÉ -FESSANJOSE-	922	9,88	0,94	1,319
98	CORPORACIÓN UNIVERSITARIA MINUTO DE DIOS	772	9,87	0,95	1,323
99	UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA UNAD	13922	9,90	0,96	1,330
100	POLITÉCNICO INTERNACIONAL INSTITUCIÓN DE EDUCACIÓN SUPERIOR	1016	9,87	0,88	1,331

Fuente: elaboración propia

3.3.4 Efecto Institucional en Razonamiento Cuantitativo

Al igual que en las competencias anteriormente descritas, Comunicación Escrita, Inglés y Lectura Crítica, durante los tres años observados, el desempeño de los estudiantes de la Universidad de Nariño en Razonamiento Cuantitativo estuvo siempre por encima del promedio nacional (ver figura 6).

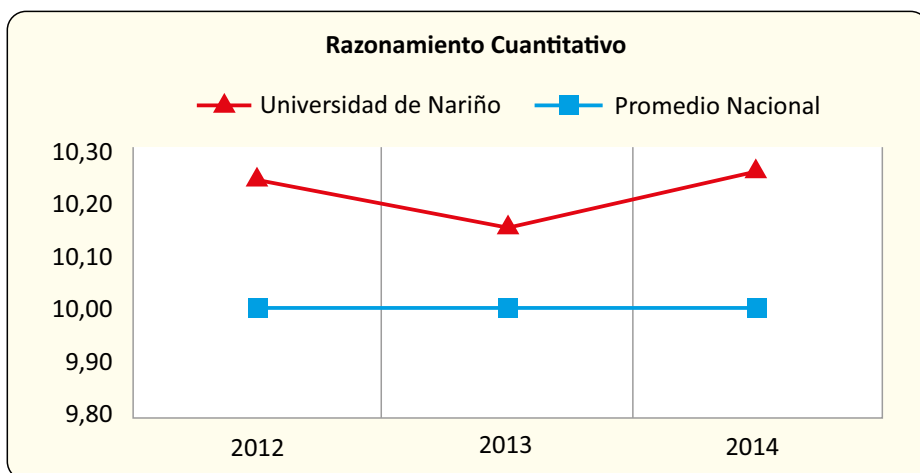


Figura 6. Tendencia del Rendimiento en Razonamiento Cuantitativo de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Según la tabla 8, la Universidad de Nariño ocupó la posición 31 en la competencia de Razonamiento Cuantitativo durante el período 2012 a 2014, con un tamaño del efecto distante en 1,444 desviaciones estándar del promedio alcanzado por la Escuela de Ingeniería de Antioquia, institución que encabezó el ranking a nivel nacional en esta prueba.

Tabla 8. Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Razonamiento Cuantitativo pruebas Saber Pro 2012-2014.

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
1	ESCUELA DE INGENIERÍA DE ANTIOQUIA	743	11,67	1,14	-
2	UNIVERSIDAD DE LOS ANDES	7860	11,62	1,31	0,035
3	UNIVERSIDAD NACIONAL DE COLOMBIA	17673	11,35	1,18	0,274
4	ESCUELA COLOMBIANA DE INGENIERÍA	1551	11,27	1,03	0,377
5	UNIVERSIDAD ICESI	2298	11,04	1,06	0,581
6	UNIVERSIDAD INDUSTRIAL DE SANTANDER	8747	10,86	1,17	0,695
7	UNIVERSIDAD -EAFIT-	4230	10,86	1,13	0,716
8	UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS	8755	10,84	1,07	0,771
9	COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO	4667	10,71	1,09	0,871
10	FUNDACIÓN UNIVERSIDAD DE AMÉRICA	1693	10,76	0,96	0,898
11	UNIVERSIDAD DE ANTIOQUIA	13169	10,62	1,11	0,939
12	UNIVERSIDAD DEL VALLE	9995	10,56	1,15	0,968
13	FUNDACIÓN UNIVERSIDAD DEL NORTE	4519	10,51	1,20	0,977
14	UNIVERSIDAD EXTERNADO DE COLOMBIA	3037	10,61	1,02	1,011
15	PONTIFICIA UNIVERSIDAD JAVERIANA	12176	10,53	1,09	1,042

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
16	INSTITUTO DE CIENCIAS DE LA SALUD	981	10,52	1,03	1,064
17	UNIVERSIDAD DE LA SABANA	3708	10,47	1,03	1,138
18	UNIVERSIDAD TECNOLÓGICA DE PEREIRA	4959	10,34	1,15	1,163
19	FUNDACIÓN UNIVERSITARIA EMPRESARIAL DE LA CÁMARA DE COMERCIO DE BOGOTÁ	716	10,43	0,89	1,208
20	INSTITUTO TÉCNICO CENTRAL	1251	10,41	0,96	1,220
21	UNIVERSIDAD PONTIFICIA BOLÍVARIANA	8664	10,28	1,10	1,269
22	UNIVERSIDAD DEL CAUCA	4372	10,32	1,05	1,270
23	UNIVERSIDAD MILITAR-NUEVA GRANADA	5997	10,41	0,94	1,307
24	UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR	2006	10,25	1,06	1,307
25	FUNDACIÓN UNIVERSITARIA JUAN N. CORPAS	637	10,19	0,90	1,426
26	UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA -UNAB-	3320	10,17	1,04	1,426
27	UNIVERSIDAD DE MEDELLÍN	4494	10,22	0,98	1,442
28	UNIVERSIDAD SURCOLOMBIANA	4218	10,14	1,05	1,442
29	UNIVERSIDAD DE LOS LLANOS	2420	10,19	0,99	1,443
30	UNIVERSIDAD SERGIO ARBOLEDA	3193	10,18	1,01	1,444
31	UNIVERSIDAD DE NARIÑO	3825	10,21	0,98	1,444
32	FUNDACIÓN UNIVERSIDAD CENTRAL	3186	10,29	0,89	1,457
33	ESCUELA COLOMBIANA DE CARRERAS INDUSTRIALES	6530	10,27	0,93	1,467
34	UNIVERSIDAD CENTRAL	1858	10,21	0,93	1,469
35	FUNDACIÓN CENTRO DE INVESTIGACIÓN DOCENCIA Y CONSULTORÍA ADMINISTRATIVA -F-CIDCA-	1113	10,11	1,01	1,473

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
36	UNIVERSIDAD AUTÓNOMA DE OCCIDENTE	3127	10,20	0,96	1,474
37	UNIVERSIDAD DE IBAGUÉ -CORUNIVERSITARIA-	3105	10,18	0,95	1,498
38	UNIVERSIDAD EAN	2462	10,21	0,91	1,508
39	UNIVERSIDAD DE CALDAS	4707	10,13	0,99	1,528
40	INSTITUTO TECNOLÓGICO METROPOLITANO	6131	10,23	0,91	1,538
41	UNIVERSIDAD DE LA SALLE	7016	10,23	0,91	1,546
42	UNIVERSIDAD DE CARTAGENA	7568	10,02	1,06	1,548
43	CORPORACIÓN UNIVERSIDAD PILOTO DE COLOMBIA	4058	10,20	0,91	1,555
44	UNIVERSIDAD-COLEGIO MAYOR DE CUNDINAMARCA	2447	10,25	0,83	1,558
45	POLITÉCNICO COLOMBIANO	3830	10,16	0,93	1,563
46	UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA	10588	10,12	0,97	1,571
47	FUNDACIÓN UNIVERSITARIA KONRAD LORENZ	1186	10,10	0,89	1,574
48	FUNDACIÓN UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO	5971	10,14	0,95	1,575
49	UNIDADES TECNOLÓGICAS DE SANTANDER	4534	10,13	0,95	1,575
50	INSTITUTO TECNOLÓGICO PASCUAL BRAVO	1168	10,01	0,98	1,586
51	UNIVERSIDAD DE SUCRE	2112	10,05	0,96	1,598
52	FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD	932	10,06	0,88	1,608
53	FUNDACIÓN UNIVERSITARIA AGRARIA DE COLOMBIA -UNIAGRARIA-	1588	10,11	0,87	1,623
54	FUNDACIÓN CENTRO COLOMBIANO DE ESTUDIOS PROFESIONALES -F.C.E.C.E.P.	991	10,02	0,90	1,631
55	UNIVERSIDAD EL BOSQUE	3527	10,13	0,90	1,636

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
56	UNIVERSIDAD DE CIENCIAS APLICADAS Y AMBIENTALES UDCA.	2395	10,12	0,86	1,651
57	FUNDACIÓN UNIVERSITARIA INTERNACIONAL DEL TRÓPICO AMERICANO	690	9,98	0,87	1,652
58	UNIVERSIDAD CATÓLICA POPULAR DEL RISARALDA	1052	9,99	0,92	1,653
59	UNIVERSIDAD CATÓLICA DE COLOMBIA	5018	10,13	0,89	1,660
60	UNIVERSIDAD DEL ATLÁNTICO	7912	9,96	1,01	1,677
61	CORPORACIÓN UNIVERSITARIA DE INVESTIGACIÓN Y DESARROLLO	2374	10,04	0,91	1,685
62	UNIVERSIDAD SANTO TOMÁS	11821	10,11	0,91	1,692
63	FUNDACIÓN DE EDUCACIÓN SUPERIOR SAN JOSÉ -FESSANJOSÉ-	922	9,94	0,92	1,694
64	INSTITUTO TECNOLÓGICO DE SOLEDAD ATLÁNTICO	908	9,95	0,89	1,700
65	FUNDACIÓN UNIVERSIDAD AUTÓNOMA DE COLOMBIA -FUAC-	3575	10,11	0,86	1,701
66	UNIVERSIDAD AUTÓNOMA DE MANIZALES	1496	9,98	0,91	1,702
67	FUNDACIÓN UNIVERSITARIA PANAMERICANA	4147	10,08	0,89	1,704
68	FUNDACIÓN UNIVERSITARIA -CEIPA-	1260	9,98	0,89	1,706
69	FUNDACIÓN PARA LA EDUCACIÓN SUPERIOR SAN MATEO FUNDACIÓN SAN MATEO	874	9,95	0,88	1,706
70	UNIVERSIDAD DE SAN BUENAVENTURA	7659	9,98	0,97	1,712
71	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	8650	9,99	0,96	1,725

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
72	INSTITUTO TECNOLÓGICO MUNICIPAL	1827	10,00	0,89	1,726
73	UNIVERSIDAD PEDAGÓGICA NACIONAL	4350	10,04	0,90	1,737
74	CORPORACIÓN POLITÉCNICO DE LA COSTA ATLÁNTICA	1188	9,87	0,96	1,745
75	CORPORACIÓN UNIVERSITARIA DEL META	1308	9,97	0,84	1,769
76	FUNDACIÓN INSTITUTO TECNOLÓGICO COOPERATIVO DE COOMULTRASAN -FITECDECOOM-	711	9,84	0,90	1,778
77	UNIVERSIDAD AUTÓNOMA LATINOAMERICANA -UNAUULA-	1499	9,94	0,88	1,781
78	FUNDACIÓN ACADEMIA DE DIBUJO PROFESIONAL	669	9,73	1,02	1,782
79	CORPORACIÓN ESCUELA DE ARTES Y LETRAS	708	9,90	0,81	1,782
80	INSTITUCIÓN UNIVERSITARIA SALAZAR Y HERRERA	1340	9,93	0,86	1,786
81	COLEGIO MAYOR DE ANTIOQUIA	1034	9,95	0,81	1,789
82	POLITÉCNICO GRANCOLOMBIANO	10098	10,00	0,91	1,799
83	FUNDACIÓN DE ESTUDIOS SUPERIORES COMFANORTE -F.E.S.C.-	622	9,80	0,89	1,804
84	UNIVERSIDAD MANUELA BELTRÁN -UMB-	2444	9,96	0,88	1,805
85	UNIVERSIDAD DE BOYACÁ UNIBOYACÁ	1719	9,98	0,82	1,816
86	CORPORACIÓN UNIVERSITARIA UNITEC	1742	9,95	0,85	1,820
87	UNIDAD CENTRAL DEL VALLE DEL CAUCA	1663	9,90	0,88	1,822
88	CORPORACIÓN UNIVERSAL DE INVESTIGACIÓN Y TECNOLOGÍA -CORUNIVERSITEC-	1499	9,92	0,85	1,830

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
89	UNIVERSIDAD DE CUNDINAMARCA -UDEC-	4404	10,00	0,86	1,844
90	UNIVERSITARIA AGUSTINIANA -UNIAGUSTINIANA-	1895	9,94	0,83	1,854
91	ESCUELA DE ADMINISTRACIÓN Y MERCADOTECNIA DEL QUINDÍO	1449	9,80	0,94	1,854
92	INSTITUCIÓN UNIVERSITARIA DE ENVIGADO	1585	9,87	0,88	1,856
93	INSTITUCIÓN UNIVERSITARIA FUNDACIÓN EDUCATIVA -ESUMER-	2403	9,94	0,86	1,858
94	UNIVERSIDAD INCCA DE COLOMBIA	2201	9,90	0,88	1,864
95	UNIVERSIDAD AUTÓNOMA DEL CARIBE	5076	9,86	0,95	1,866
96	FUNDACIÓN UNIVERSITARIA DE SAN GIL -UNISANGIL-	1669	9,90	0,84	1,879
97	UNIVERSIDAD DE MANIZALES	1909	9,86	0,88	1,889
98	UNIVERSIDAD CATÓLICA DE ORIENTE	2022	9,79	0,94	1,889
99	INSTITUTO TECNOLÓGICO DEL PUTUMAYO	694	9,76	0,84	1,896
100	INSTITUTO DE EDUCACIÓN TÉCNICA PROFESIONAL DE ROLDANILLO	679	9,74	0,86	1,900

Fuente: elaboración propia

3.3.5. Efecto Institucional en Competencias Ciudadanas

Aunque en esta competencia, el desempeño de los estudiantes de la Universidad de Nariño no es el mejor de las cinco competencias, presentando desempeños por debajo del promedio nacional, como se observa en la figura 7, la Universidad de Nariño, en el ranking de las mejores 100 universidades del país, alcanza una meritoria posición, ocupando el lugar 37, a una distancia de 0,892 desviaciones estándar frente a la universidad de mayor puntaje, que en este caso, es la Universidad de los Andes (ver tabla 9).

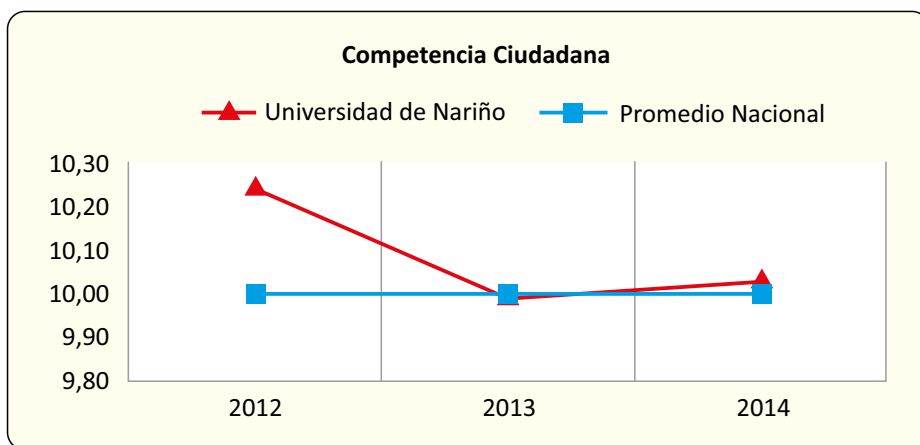


Figura 7. Tendencia del Rendimiento en Competencia Ciudadana de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Tabla 9. Ranking de 100 primeras Universidades e Instituciones de Educación Superior en Competencias Ciudadanas pruebas Saber Pro 2012-2014.

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
1	UNIVERSIDAD DE LOS ANDES	7860	11,07	1,15	-
2	COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO	4667	10,93	1,02	0,123
3	UNIVERSIDAD EXTERNADO DE COLOMBIA	3037	10,87	0,96	0,182
4	UNIVERSIDAD NACIONAL DE COLOMBIA	17673	10,87	1,01	0,183
5	ESCUELA DE INGENIERÍA DE ANTIOQUIA	743	10,78	0,99	0,249
6	UNIVERSIDAD ICESI	2298	10,71	1,03	0,316
7	UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS	8755	10,58	0,91	0,473
8	UNIVERSIDAD EAFIT-	4230	10,50	1,08	0,502
9	UNIVERSIDAD DE ANTIOQUIA	13169	10,51	1,03	0,514
10	INSTITUTO DE CIENCIAS DE LA SALUD	981	10,43	0,98	0,560

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
11	PONTIFICIA UNIVERSIDAD JAVERIANA	12176	10,46	0,99	0,580
12	UNIVERSIDAD DE LA SABANA	3708	10,43	0,95	0,586
13	UNIVERSIDAD DEL VALLE	9995	10,41	1,03	0,610
14	UNIVERSIDAD-COLEGIO MAYOR DE CUNDINAMARCA	2447	10,39	0,89	0,619
15	ESCUELA COLOMBIANA DE INGENIERÍA	1551	10,37	0,86	0,635
16	UNIVERSIDAD DE MEDELLÍN	4494	10,35	1,06	0,646
17	UNIVERSIDAD INDUSTRIAL DE SANTANDER	8747	10,36	1,03	0,647
18	FUNDACIÓN UNIVERSIDAD DEL NORTE	4519	10,27	1,09	0,706
19	UNIVERSIDAD PEDAGÓGICA NACIONAL	4350	10,29	0,92	0,723
20	UNIVERSIDAD MILITAR-NUEVA GRANADA	5997	10,30	0,93	0,727
21	UNIVERSIDAD AUTÓNOMA LATINOAMERICANA -UNAUULA-	1499	10,25	0,97	0,733
22	FUNDACIÓN UNIVERSITARIA KONRAD LORENZ	1186	10,21	0,87	0,771
23	UNIVERSIDAD EAN	2462	10,20	0,88	0,792
24	UNIVERSIDAD SERGIO ARBOLEDA	3193	10,19	0,98	0,794
25	FUNDACIÓN UNIVERSITARIA JUAN N. CORPAS	637	10,17	0,88	0,795
26	UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA -UNAB-	3320	10,17	1,03	0,806
27	UNIVERSIDAD DE CALDAS	4707	10,18	0,99	0,812
28	UNIVERSIDAD CATÓLICA DE COLOMBIA	5018	10,20	0,90	0,822
29	UNIVERSIDAD DEL CAUCA	4372	10,16	1,02	0,825
30	FUNDACIÓN UNIVERSIDAD DE AMÉRICA	1693	10,15	0,90	0,825
31	FUNDACIÓN UNIVERSITARIA EMPRESARIAL DE LA CÁMARA DE COMERCIO DE BOGOTÁ	716	10,13	0,85	0,831

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
32	UNIVERSIDAD TECNOLÓGICA DE PEREIRA	4959	10,14	1,03	0,837
33	UNIVERSIDAD PONTIFICIA BOLÍVARIANA	8664	10,14	1,06	0,846
34	INSTITUTO TÉCNICO CENTRAL	1251	10,10	0,97	0,857
35	UNIVERSIDAD LA GRAN COLOMBIA	7091	10,13	0,97	0,879
36	FUNDACIÓN UNIVERSIDAD CENTRAL	3186	10,10	0,95	0,882
37	UNIVERSIDAD DE NARIÑO	3825	10,08	1,03	0,892
38	FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD	932	10,05	0,97	0,901
39	FUNDACIÓN UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO	5971	10,08	0,97	0,915
40	UNIVERSIDAD DE LA SALLE	7016	10,10	0,93	0,919
41	FUNDACIÓN CENTRO DE INVESTIGACIÓN DOCENCIA Y CONSULTORÍA ADMINISTRATIVA -FCID-CA-	1113	10,03	0,99	0,921
42	UNIVERSIDAD SANTO TOMÁS	11821	10,09	0,96	0,938
43	CORPORACIÓN ESCUELA DE ARTES Y LETRAS	708	10,00	0,90	0,947
44	INSTITUCIÓN UNIVERSITARIA DE ENVIGADO	1585	10,00	1,02	0,951
45	UNIVERSIDAD DE MANIZALES	1909	10,00	0,99	0,952
46	UNIVERSIDAD DE LOS LLANOS	2420	10,00	0,94	0,966
47	FUNDACIÓN UNIVERSIDAD AUTÓNOMA DE COLOMBIA -FUAC-	3575	10,01	0,94	0,968
48	UNIVERSIDAD AUTÓNOMA DE OCCIDENTE	3127	10,00	1,00	0,968
49	UNIVERSIDAD SURCOLOMBIANA	4218	9,99	1,03	0,970
50	UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA	10588	10,05	0,95	0,976
51	UNIVERSIDAD EL BOSQUE	3527	10,02	0,88	0,981

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
52	CORPORACIÓN UNIVERSAL DE INVESTIGACIÓN Y TECNOLOGÍA -CORUNIVERSITEC-	1499	9,97	0,91	0,983
53	FUNDACIÓN PARA LA EDUCACIÓN SUPERIOR SAN MATEO FUNDACIÓN SAN MATEO	874	9,95	0,93	0,989
54	UNIVERSIDAD LIBRE	15386	10,03	0,99	0,991
55	CORPORACIÓN UNIVERSITARIA REPUBLICANA	1321	9,96	0,87	0,993
56	UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR	2006	9,95	0,99	0,996
57	UNIVERSIDAD CENTRAL	1858	9,96	0,91	0,996
58	UNIDAD CENTRAL DEL VALLE DEL CAUCA	1663	9,96	0,92	0,998
59	FUNDACIÓN CENTRO COLOMBIANO DE ESTUDIOS PROFESIONALES -FCECEP-	991	9,93	1,02	1,004
60	FUNDACIÓN UNIVERSITARIA AGRARIA DE COLOMBIA -UNIAGRARIA-	1588	9,95	0,88	1,004
61	FUNDACIÓN UNIVERSITARIA PANAMERICANA	4147	9,99	0,91	1,005
62	UNIVERSIDAD DE CARTAGENA	7568	9,95	1,08	1,006
63	COLEGIO MAYOR DE ANTIOQUIA	1034	9,94	0,89	1,007
64	INSTITUTO TECNOLÓGICO METROPOLITANO	6131	9,99	0,94	1,010
65	UNIVERSIDAD AUTÓNOMA DE MANIZALES	1496	9,94	0,94	1,010
66	UNIVERSIDAD DE CIENCIAS APLICADAS Y AMBIENTALES -UDCA-	2395	9,96	0,90	1,014
67	FUNDACIÓN DE EDUCACIÓN SUPERIOR SAN JOSÉ -FESSANJOSÉ-	922	9,92	0,90	1,017
68	FUNDACIÓN UNIVERSITARIA INPAHU	2075	9,94	0,90	1,020

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
69	POLITÉCNICO COLOMBIANO	3830	9,97	0,91	1,022
70	ESCUELA COLOMBIANA DE CARRERAS INDUSTRIALES	6530	9,99	0,93	1,023
71	UNIVERSIDAD DE BOYACÁ UNIBOYACÁ	1719	9,93	0,90	1,028
72	UNIVERSIDAD CATÓLICA POPULAR DEL RISARALDA	1052	9,91	0,93	1,031
73	UNIVERSIDAD DE IBAGUÉ -CORUNIVERSITARIA-	3105	9,94	0,92	1,031
74	FUNDACIÓN UNIVERSITARIA LUIS AMIGÓ -FUNLAM-	4399	9,94	0,92	1,056
75	FUNDACIÓN ACADEMIA DE DIBUJO PROFESIONAL	669	9,86	1,03	1,059
76	CORPORACIÓN UNIVERSITARIA UNITEC	1742	9,89	0,92	1,062
77	UNIVERSIDAD MANUELA BELTRÁN -UMB-	2444	9,89	0,93	1,070
78	FUNDACIÓN UNIVERSITARIA- CEIPA-	1260	9,87	0,91	1,071
79	UNIVERSIDAD DEL ATLÁNTICO	7912	9,90	1,01	1,080
80	INSTITUTO TECNOLÓGICO PASCUAL BRAVO	1168	9,84	1,05	1,084
81	POLITÉCNICO GRANCOLOMBIANO	10098	9,94	0,95	1,086
82	ESCUELA SUPERIOR DE ADMINISTRACIÓN PÚBLICA -ESAP-	6845	9,86	1,06	1,088
83	CORPORACIÓN UNIVERSIDAD PILOTO DE COLOMBIA	4058	9,88	0,95	1,096
84	INSTITUCIÓN UNIVERSITARIA FUNDACIÓN EDUCATIVA ESUMER	2403	9,86	0,91	1,099
85	UNIVERSIDAD DE SAN BUENAVENTURA	7659	9,90	0,96	1,102
86	UNIVERSITARIA AGUSTINIANA- UNIAGUSTINIANA	1895	9,85	0,85	1,114
87	POLITÉCNICO INTERNACIONAL INSTITUCIÓN DE EDUCACIÓN SUPERIOR	1016	9,81	0,88	1,122

No.	INSTITUCIÓN	N	Media	Des- viación estándar	d Cohen
88	UNIVERSIDAD INCCA DE COLOMBIA	2201	9,84	0,87	1,123
89	CORPORACIÓN UNIVERSITARIA DE CIENCIA Y DESARROLLO	2946	9,83	0,95	1,124
90	UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA UNAD	13922	9,86	1,03	1,127
91	CORPORACIÓN UNIVERSITARIA DEL META	1308	9,80	0,93	1,128
92	INSTITUTO TECNOLÓGICO MUNICIPAL	1827	9,80	0,95	1,141
93	COLEGIO ODONTOLÓGICO COLOMBIANO	690	9,78	0,85	1,142
94	FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES	7635	9,88	0,92	1,142
95	FUNDACIÓN UNIVERSITARIA MONSERRATE	1077	9,78	0,90	1,148
96	UNIDADES TECNOLÓGICAS DE SANTANDER	4534	9,81	1,00	1,150
97	FUNDACIÓN UNIVERSITARIA CATÓLICA DEL NORTE	1639	9,75	1,11	1,152
98	FUNDACIÓN UNIVERSITARIA DE SAN GIL - UNISANGIL -	1669	9,79	0,88	1,153
99	FUNDACIÓN UNIVERSITARIA INTERNACIONAL DEL TROPICO AMERICANO	690	9,76	1,01	1,153
100	UNIVERSIDAD DEL QUINDIO	7447	9,83	0,96	1,164

Fuente: elaboración propia

3.4 EFECTO DE LOS PROGRAMAS EN LAS COMPETENCIAS GENÉRICAS EN EL CONTEXTO INSTITUCIONAL

3.4.1 Efecto del programa en Comunicación Escrita

Como se observa en la tabla 10, en la competencia de Comunicación Escrita, durante el período 2012 a 2014, Sociología alcanzó el mejor desempeño, entre los programas académicos que ofrece la Universidad de Nariño, seguido por Derecho a una diferencia no relevante ($d < 0,2$), situándose luego, Psicología, Licenciatura

en Inglés y Francés, Medicina, Administración de Empresas, Economía y Licenciatura en Educación Básica con énfasis en Humanidades, Lengua Castellana e Inglés, con diferencias pequeñas ($d < 0,5$) respecto al primer programa. Los demás programas alcanzan tamaños de efecto moderado o grande.

Tabla 10. Clasificación por Programas Académicos de la Universidad de Nariño en Comunicación Escrita de las pruebas Saber Pro 2012-2014.

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
1	SOCIOLOGÍA	79	10,93	0,87	-
2	DERECHO	338	10,78	1,03	0,154
3	PSICOLOGÍA	87	10,61	1,21	0,309
4	LICENCIATURA EN INGLÉS Y FRANCÉS	65	10,62	1,03	0,332
5	MEDICINA	74	10,61	0,99	0,347
6	ADMINISTRACIÓN DE EMPRESAS	314	10,48	1,05	0,445
7	ECONOMÍA	214	10,47	1,07	0,451
8	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN HUMANIDADES, LENGUA CASTELLANA E INGLÉS	90	10,52	0,92	0,457
9	COMERCIO INTERNACIONAL Y MERCADEO	94	10,43	1,13	0,499
10	QUÍMICA	61	10,46	0,91	0,532
11	LICENCIATURA EN FILOSOFÍA Y LETRAS	68	10,36	1,09	0,587
12	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS SOCIALES	52	10,38	0,98	0,609
13	LICENCIATURA EN LENGUA CASTELLANA Y LITERATURA	215	10,25	1,11	0,648
14	BIOLOGÍA	37	10,34	0,95	0,660
15	ZOOTECNIA	77	10,21	1,08	0,738
16	INGENIERÍA CIVIL	213	10,18	1,07	0,745
17	LICENCIATURA EN MATEMÁTICAS	50	10,24	1,02	0,745
18	INGENIERÍA AGROINDUSTRIAL	161	10,20	1,01	0,754
19	LICENCIATURA EN MÚSICA	91	10,19	1,02	0,778

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
20	ARQUITECTURA	94	10,17	0,96	0,828
21	DISEÑO GRÁFICO Y MULTIMEDIAL	107	10,07	1,12	0,848
22	MEDICINA VETERINARIA	113	10,15	0,96	0,849
23	INGENIERÍA EN PRODUCCIÓN ACUÍCOLA	49	10,13	1,06	0,851
24	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS NATURALES Y EDUCACIÓN AMBIENTAL	175	10,10	1,00	0,865
25	DISEÑO INDUSTRIAL	112	10,09	1,02	0,875
26	GEOGRAFÍA APLICADA A LA ORGANIZACIÓN DEL ESPACIO Y PLANIFICACIÓN REGIONAL	49	10,01	1,10	0,964
27	LICENCIATURA EN INFORMÁTICA	95	10,02	0,98	0,986
28	INGENIERÍA DE SISTEMAS	133	10,01	0,97	0,995
29	INGENIERÍA AGROFORESTAL	132	10,00	0,96	1,015
30	INGENIERÍA AGRONÓMICA	106	9,93	1,03	1,037
31	FÍSICA	13	9,98	0,92	1,092
32	INGENIERÍA ELECTRÓNICA	147	9,91	0,96	1,108
33	ARTES VISUALES	78	9,76	1,07	1,200
34	LICENCIATURA EN ARTES VISUALES	39	9,57	1,04	1,467
GENERAL		3822	10,28	1,07	0,612

* No se incluyen los programas de Tecnologías en Computación y Promoción de la Salud cuyo número de estudiantes fue solamente de 2 y 1 respectivamente

Fuente: elaboración propia

3.4.2 Efecto del programa en Inglés

El programa de Licenciatura en Inglés y Francés, obtuvo el mejor desempeño en la prueba de Inglés, durante el período estudiado, seguido por la Licenciatura en Educación Básica con énfasis en Humanidades, Lengua Castellana e Inglés, a una diferencia moderada ($d < 0,8$); y las posiciones siguientes, son ocupadas por los programas de Ingeniería Electrónica, Medicina, Psicología y Biología, con diferencias grandes ($d > 0,8$). Los demás programas alcanzan tamaños de efecto superiores (ver tabla 11).

Tabla 11. Clasificación por Programas Académicos de la Universidad de Nariño en Inglés de las pruebas Saber Pro 2012-2014.

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
1	LICENCIATURA EN INGLÉS Y FRANCÉS	65	12,77	1,01	-
2	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN HUMANIDADES, LENGUA CASTELLANA E INGLÉS	90	12,18	1,12	0,542
3	INGENIERÍA ELECTRÓNICA	147	11,03	1,10	1,615
4	MEDICINA	74	10,99	1,13	1,652
5	PSICOLOGÍA	87	10,60	1,32	1,809
6	BIOLOGÍA	37	10,86	1,08	1,838
7	DERECHO	338	10,35	1,32	1,900
8	DISEÑO INDUSTRIAL	112	10,16	1,48	1,962
9	INGENIERÍA DE SISTEMAS	133	10,63	1,05	2,062
10	ARTES VISUALES	78	9,54	1,74	2,214
11	DISEÑO GRÁFICO Y MULTIMEDIAL	107	10,41	1,08	2,235
12	FÍSICA	13	10,62	0,58	2,236
13	QUÍMICA	61	10,70	0,81	2,245
14	ARQUITECTURA	94	10,29	1,16	2,250
15	INGENIERÍA AGRONÓMICA	106	10,05	1,26	2,314
16	COMERCIO INTERNACIONAL Y MERCADEO	94	10,50	0,96	2,316
17	INGENIERÍA CIVIL	213	10,49	0,91	2,425
18	ADMINISTRACIÓN DE EMPRESAS	314	10,17	1,07	2,451
19	LICENCIATURA EN MATEMÁTICAS	50	10,27	0,73	2,771
20	MEDICINA VETERINARIA	113	10,22	0,81	2,858
21	ECONOMÍA	214	10,02	0,94	2,869
22	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS NATURALES Y EDUCACIÓN AMBIENTAL	175	9,52	1,15	2,907

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
23	ZOOTECNIA	77	10,01	0,74	3,145
24	INGENIERÍA AGROINDUSTRIAL	161	9,91	0,85	3,182
25	LICENCIATURA EN LENGUA CASTELLANA Y LITERATURA	215	9,70	0,90	3,312
26	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS SOCIALES	52	9,90	0,61	3,333
27	LICENCIATURA EN FILOSOFÍA Y LETRAS	68	9,79	0,75	3,351
28	LICENCIATURA EN MÚSICA	91	9,73	0,77	3,460
29	SOCIOLOGÍA	79	9,90	0,62	3,491
30	INGENIERÍA AGROFORESTAL	132	9,95	0,68	3,513
31	INGENIERÍA EN PRODUCCIÓN ACUÍCOLA	49	9,70	0,60	3,554
32	LICENCIATURA EN INFORMÁTICA	95	10,04	0,53	3,560
33	GEOGRAFÍA APLICADA A LA ORGANIZACIÓN DEL ESPACIO Y PLANIFICACIÓN REGIONAL	49	9,75	0,54	3,583
34	LICENCIATURA EN ARTES VISUALES	39	9,58	0,57	3,639
	GENERAL	3822	10,26	1,18	2,124

* No se incluyen los programas de Tecnologías en Computación y Promoción de la Salud cuyo número de estudiantes fue solamente de 2 y 1 respectivamente

Fuente: elaboración propia

3.4.3 Efecto del programa en Lectura Crítica

Según la tabla 12, durante el período 2012 a 2014, el programa de Medicina presentó el mejor desempeño en la competencia de Lectura Crítica, seguido por Derecho y Psicología, con diferencias irrelevantes en tamaño ($d < 0,2$). Con diferencias pequeñas ($d < 0,5$) se ubicaron los programas de Biología, Química y Física; los demás alcanzan tamaños de efecto superiores.

Tabla 12. Clasificación por Programas Académicos de la Universidad de Nariño en Lectura Crítica de las pruebas Saber Pro 2012-2014.

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
1	MEDICINA	74	11,04	0,87	-
2	DERECHO	338	11,03	1,02	0,009
3	PSICOLOGÍA	87	10,88	0,83	0,186
4	BIOLOGÍA	37	10,86	0,90	0,207
5	QUÍMICA	61	10,71	0,86	0,381
6	FÍSICA	13	10,65	1,11	0,423
7	SOCIOLOGÍA	79	10,61	0,70	0,548
8	LICENCIATURA EN FILOSOFÍA Y LETRAS	68	10,56	0,77	0,578
9	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS SOCIALES	52	10,56	0,68	0,602
10	INGENIERÍA ELECTRÓNICA	147	10,48	0,92	0,617
11	LICENCIATURA EN INGLÉS Y FRANCÉS	65	10,47	0,74	0,700
12	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN HUMANIDADES, LENGUA CASTELLANA E INGLÉS	90	10,43	0,82	0,715
13	INGENIERÍA CIVIL	213	10,33	1,00	0,733
14	MEDICINA VETERINARIA	113	10,43	0,78	0,746
15	ARQUITECTURA	94	10,39	0,85	0,749
16	COMERCIO INTERNACIONAL Y MERCADEO	94	10,28	1,08	0,764
17	LICENCIATURA EN MATEMÁTICAS	50	10,36	0,87	0,780
18	INGENIERÍA AGRONÓMICA	106	10,37	0,81	0,795
19	DISEÑO GRÁFICO Y MULTIMEDIAL	107	10,30	0,92	0,814
20	INGENIERÍA DE SISTEMAS	133	10,29	0,94	0,820
21	ECONOMÍA	214	10,20	1,00	0,865
22	DISEÑO INDUSTRIAL	112	10,33	0,79	0,869
23	LICENCIATURA EN INFORMÁTICA	95	10,30	0,83	0,869

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
24	ADMINISTRACIÓN DE EMPRESAS	314	10,17	0,97	0,906
25	INGENIERÍA AGROFORESTAL	132	10,25	0,83	0,931
26	LICENCIATURA EN LENGUA CASTELLANA Y LITERATURA	215	10,09	1,00	0,984
27	GEOGRAFÍA APLICADA A LA ORGANIZACIÓN DEL ESPACIO Y PLANIFICACIÓN REGIONAL	49	10,22	0,53	1,077
28	ZOOTECNIA	77	10,04	0,86	1,152
29	LICENCIATURA EN MÚSICA	91	9,92	0,99	1,193
30	INGENIERÍA AGROINDUSTRIAL	161	9,92	0,92	1,238
31	ARTES VISUALES	78	9,90	0,95	1,246
32	INGENIERÍA EN PRODUCCIÓN ACUÍCOLA	49	9,97	0,80	1,262
33	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS NATURALES Y EDUCACIÓN AMBIENTAL	175	9,75	0,96	1,378
34	LICENCIATURA EN ARTES VISUALES	39	9,39	0,89	1,876
GENERAL		3822	10,33	0,98	0,721

* No se incluyen los programas de Tecnologías en Computación y Promoción de la Salud cuyo número de estudiantes fue solamente de 2 y 1 respectivamente

Fuente: elaboración propia

3.4.4 Efecto del programa en Razonamiento Cuantitativo

En Razonamiento Cuantitativo, el programa de Medicina alcanzó el mejor promedio en el período analizado, Ingeniería Electrónica ocupó el segundo lugar con una diferencia en el desempeño no relevante en tamaño ($d < 0,2$). Los programas de Física, Ingeniería Civil y Química, ocupan las siguientes posiciones con diferencias pequeñas ($d < 0,5$) respecto del primer lugar. Los programas de Derecho, Biología, Ingeniería de Sistemas, Licenciatura en Matemáticas y Arquitectura alcanzan diferencias moderadas ($d < 0,8$); los restantes programas presentan diferencias grandes (ver tabla 13).

Tabla 13. Clasificación por Programas Académicos de la Universidad de Nariño en Razonamiento Cuantitativo de las pruebas Saber Pro 2012-2014.

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
1	MEDICINA	74	11,12	0,89	-
2	INGENIERÍA ELECTRÓNICA	147	11,02	0,92	0,111
3	FÍSICA	13	10,88	1,11	0,256
4	INGENIERÍA CIVIL	213	10,76	0,95	0,389
5	QUÍMICA	61	10,73	0,83	0,448
6	DERECHO	338	10,64	0,96	0,509
7	BIOLOGÍA	37	10,66	0,91	0,509
8	INGENIERÍA DE SISTEMAS	133	10,56	1,14	0,532
9	LICENCIATURA EN MATEMÁTICAS	50	10,64	0,80	0,565
10	ARQUITECTURA	94	10,48	0,83	0,746
11	ECONOMÍA	214	10,29	0,98	0,867
12	INGENIERÍA AGRONÓMICA	106	10,38	0,79	0,892
13	ADMINISTRACIÓN DE EMPRESAS	314	10,25	0,98	0,907
14	COMERCIO INTERNACIONAL Y MERCADEO	94	10,26	0,94	0,937
15	LICENCIATURA EN INFORMÁTICA	95	10,31	0,79	0,965
16	PSICOLOGÍA	87	10,27	0,83	0,991
17	MEDICINA VETERINARIA	113	10,26	0,80	1,027
18	INGENIERÍA AGROFORESTAL	132	10,25	0,80	1,043
19	INGENIERÍA AGROINDUSTRIAL	161	10,19	0,87	1,068
20	SOCIOLOGÍA	79	10,26	0,67	1,103
21	ZOOTECNIA	48	10,15	0,78	1,141
22	DISEÑO INDUSTRIAL	112	10,07	0,80	1,255
23	INGENIERÍA EN PRODUCCIÓN ACUÍCOLA	49	10,06	0,74	1,278
24	LICENCIATURA EN MÚSICA	91	9,79	1,04	1,366
25	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN HUMANIDADES, LENGUA CASTELLANA E INGLÉS	90	10,00	0,75	1,372

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
26	DISEÑO GRÁFICO Y MULTIMEDIAL	107	9,99	0,75	1,392
27	GEOGRAFÍA APLICADA A LA ORGANIZACIÓN DEL ESPACIO Y PLANIFICACIÓN REGIONAL	49	9,99	0,57	1,452
28	LICENCIATURA EN FILOSOFÍA Y LETRAS	68	9,86	0,70	1,563
29	LICENCIATURA EN INGLÉS Y FRANCÉS	65	9,81	0,65	1,674
30	LICENCIATURA EN LENGUA CASTELLANA Y LITERATURA	215	9,64	0,88	1,678
31	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS SOCIALES	52	9,71	0,59	1,814
32	ARTES VISUALES	78	9,51	0,85	1,857
33	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS NATURALES Y EDUCACIÓN AMBIENTAL	175	9,42	0,70	2,243
34	LICENCIATURA EN ARTES VISUALES	39	9,20	0,72	2,300
GENERAL		3822	10,21	0,98	0,928

* No se incluyen los programas de Tecnologías en Computación y Promoción de la Salud cuyo número de estudiantes fue solamente de 2 y 1 respectivamente

Fuente: elaboración propia

3.4.5 Efecto del programa en Competencias Ciudadanas

En las Competencias Ciudadanas, el programa de mejor desempeño es el de Derecho, y ocupan las posiciones siguientes los programas de Psicología, Medicina, y Sociología, en su orden, con tamaños de efecto pequeños ($d < 0,05$). Luego se ubican los programas de Biología, Física, Ingeniería Electrónica, Química y Licenciatura en Educación Básica con énfasis en Ciencias Sociales, a diferencias moderadas del primer lugar ($d < 0,8$). Los demás programas presentan diferencias grandes en tamaño del efecto (ver tabla 14).

Tabla 14. Clasificación por Programas Académicos de la Universidad de Nariño en Competencias Ciudadanas de las pruebas Saber Pro 2012-2014.

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
1	DERECHO	338	11,01	0,99	-
2	PSICOLOGÍA	87	10,66	0,82	0,370
3	MEDICINA	74	10,59	0,97	0,428
4	SOCIOLOGÍA	79	10,54	0,78	0,489
5	BIOLOGÍA	37	10,49	0,88	0,536
6	FÍSICA	13	10,37	0,90	0,652
7	INGENIERÍA ELECTRÓNICA	147	10,28	1,02	0,731
8	QUÍMICA	61	10,30	0,69	0,751
9	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS SOCIALES	52	10,27	0,82	0,769
10	MEDICINA VETERINARIA	113	10,26	0,73	0,806
11	ARQUITECTURA	94	10,14	0,93	0,896
12	LICENCIATURA EN FILOSOFÍA Y LETRAS	68	10,15	0,72	0,909
13	COMERCIO INTERNACIONAL Y MERCADEO	94	10,11	1,02	0,910
14	INGENIERÍA AGRONÓMICA	106	10,11	0,93	0,927
15	GEOGRAFÍA APLICADA A LA ORGANIZACIÓN DEL ESPACIO Y PLANIFICACIÓN REGIONAL	49	10,12	0,71	0,935
16	INGENIERÍA CIVIL	213	10,09	0,93	0,954
17	ECONOMÍA	214	10,07	0,97	0,964
18	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN HUMANIDADES, LENGUA CASTELLANA E INGLÉS	90	10,07	0,86	0,982
19	INGENIERÍA DE SISTEMAS	133	10,01	0,95	1,021
20	LICENCIATURA EN INGLÉS Y FRANCÉS	65	9,91	1,43	1,032
21	INGENIERÍA AGROFORESTAL	132	10,00	0,91	1,043
22	ADMINISTRACIÓN DE EMPRESAS	314	9,96	1,02	1,050

No.	PROGRAMA*	N	Media	Des- viación estándar	d Cohen
23	DISEÑO INDUSTRIAL	112	9,97	0,77	1,108
24	ZOOTECNIA	77	9,95	0,79	1,118
25	LICENCIATURA EN MATEMÁTICAS	50	9,83	1,01	1,194
26	DISEÑO GRÁFICO Y MULTIMEDIAL	107	9,83	0,93	1,211
27	LICENCIATURA EN INFORMÁTICA	95	9,81	0,81	1,263
28	INGENIERÍA AGROINDUSTRIAL	161	9,74	0,93	1,313
29	LICENCIATURA EN MÚSICA	91	9,54	1,37	1,364
30	INGENIERÍA EN PRODUCCIÓN ACUÍCOLA	49	9,69	0,80	1,365
31	LICENCIATURA EN LENGUA CASTELLANA Y LITERATURA	215	9,64	0,90	1,436
32	ARTES VISUALES	78	9,54	0,91	1,511
33	LICENCIATURA EN EDUCACIÓN BÁSICA CON ÉNFASIS EN CIENCIAS NATURALES Y EDUCACIÓN AMBIENTAL	175	9,47	0,95	1,576
34	LICENCIATURA EN ARTES VISUALES	39	9,16	0,69	1,931
GENERAL		3822	10,08	1,02	0,915

* No se incluyen los programas de Tecnologías en Computación y Promoción de la Salud cuyo número de estudiantes fue solamente de 2 y 1 respectivamente

Fuente: elaboración propia

3.5 EFECTO DE LAS VARIABLES SOCIECONÓMICAS EN LAS COMPETENCIAS GENÉRICAS EN EL CONTEXTO INSTITUCIONAL

Para establecer si las diferencias en puntajes en las cinco competencias son estadísticamente significativas, en las variables socioeconómicas, se calcularon los intervalos de confianza de medias al 95%. Los grupos que muestran rangos que no se traslapan, presentan verdaderas diferencias. Los grupos se ordenan de acuerdo al nivel de desempeño, de mayor a menor, teniendo en cuenta las diferencias estandarizadas, tomando como referencia

el grupo de más alto puntaje en cada prueba. Para su interpretación, se forman grupos identificados con las letras A, B, C, etc., denotando puntajes promedios que van de los valores más altos a los más bajos.

3.5.1 Efecto del género

De la tabla 15 se concluye que, en las pruebas de competencias genéricas Saber Pro 2012-2014, los hombres presentan mejores desempeños que las mujeres en razonamiento cuantitativo e inglés, y las mujeres superan a los hombres en las tres competencias restantes. Según el estadístico *d* de Cohen, se presentan diferencias de magnitud relevante ($>0,2$) en razonamiento cuantitativo, en favor de los hombres; y en inglés, en favor de las mujeres.

Tabla 15. Género y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Género	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	Femenino	1762	10,45	1,03	10,40	10,50	-	A
	Masculino	2040	10,14	1,08	10,10	10,19	0,29	B
Inglés	Femenino	1773	10,20	1,11	10,15	10,25	0,09	B
	Masculino	2046	10,30	1,24	10,25	10,36	-	A
Lectura crítica	Femenino	1773	10,35	0,96	10,30	10,39	-	A
	Masculino	2046	10,32	1,00	10,28	10,37	0,02	B
Razonamiento cuantitativo	Femenino	1773	10,02	0,88	9,97	10,06	0,38	B
	Masculino	2046	10,38	1,03	10,33	10,42	-	A
Competencias ciudadanas	Femenino	1773	10,09	1,01	10,05	10,14	-	A
	Masculino	2046	10,06	1,04	10,02	10,11	0,03	B

Fuente: elaboración propia

3.5.2 Efecto de la edad

Por grupos de edad, los estudiantes en el rango de 20 a 29 años superan a los demás grupos en todas las competencias, con tamaños de efecto relevantes. Se observó una marcada tendencia negativa, a mayor edad menor rendimiento, casi en la totalidad de las cinco competencias (ver tabla 16).

Tabla 16. Edad y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Grupo de Edad	N	Media	Des- viación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	Menor de 20 años	18	10,08	1,34	9,42	10,75	0,23	C
	20-29	3304	10,32	1,05	10,28	10,36	-	A
	30-39	381	10,05	1,17	9,94	10,17	0,25	D
	40-49	80	10,13	1,13	9,88	10,38	0,18	B
	50 o más años	19	9,77	0,94	9,32	10,22	0,53	E
Inglés	Menor de 20 años	18	10,03	0,84	9,61	10,45	0,25	B
	20-29	3313	10,32	1,17	10,28	10,36	-	A
	30-39	387	9,94	1,08	9,83	10,05	0,33	C
	40-49	81	9,39	1,60	9,03	9,74	0,79	E
	50 o más años	20	9,51	1,31	8,89	10,12	0,70	D
Lectura crítica	Menor de 20 años	18	10,08	0,89	9,63	10,52	0,35	B
	20-29	3313	10,41	0,94	10,37	10,44	-	A
	30-39	387	10,01	1,06	9,90	10,11	0,41	C
	40-49	81	9,37	1,00	9,15	9,59	1,10	D
	50 o más años	20	9,13	0,85	8,73	9,53	1,35	E
Razonamiento cuantitativo	Menor de 20 años	18	9,82	0,72	9,46	10,18	0,49	B
	20-29	3313	10,29	0,96	10,26	10,32	-	A
	30-39	387	9,81	0,96	9,72	9,91	0,50	C
	40-49	81	9,20	0,92	8,99	9,40	1,13	E
	50 o más años	20	9,35	0,56	9,08	9,61	0,98	D
Competencias ciudadanas	Menor de 20 años	18	9,69	0,81	9,29	10,09	0,43	C
	20-29	3313	10,12	1,01	10,09	10,15	-	A
	30-39	387	9,91	1,10	9,80	10,02	0,21	B
	40-49	81	9,44	1,13	9,19	9,69	0,67	E
	50 o más años	20	9,22	1,01	8,75	9,69	0,89	N

Fuente: elaboración propia

3.5.3 Efecto del estado civil

Según la tabla 17, los soltero(a)s presentan mejores desempeños en las competencias de inglés, lectura crítica y razonamiento cuantitativo. En las pruebas de comunicación escrita y lectura crítica, tienen mejor desempeño los separados/divorciados y los viudos respectivamente, aunque en este último grupo aparece un solo estudiante que presentó las pruebas en el período estudiado.

Tabla 17. Estado Civil y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Grupo de Edad	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	Soltero(a)	3470	10,30	1,06	10,26	10,33	0,30	C
	Casado(a)	163	10,30	1,07	10,13	10,47	0,29	B
	Viudo(a)	1	9,90	-	-	-	0,62	E
	Separado(a) / Divorciado(a)	21	10,61	1,14	10,09	11,13	-	A
	Unión libre	145	9,95	1,18	9,76	10,14	0,56	D
Inglés	Soltero(a)	3485	10,30	1,18	10,26	10,33	-	A
	Casado(a)	164	9,94	0,88	9,81	10,08	0,30	C
	Viudo(a)	1	9,50	-	-	-	0,67	E
	Separado(a) / Divorciado(a)	21	10,08	1,04	9,60	10,55	0,19	B
	Unión libre	146	9,71	1,44	9,47	9,94	0,50	D
Lectura crítica	Soltero(a)	3485	10,37	0,97	10,34	10,40	-	A
	Casado(a)	164	10,08	0,97	9,93	10,23	0,30	C
	Viudo(a)	1	10,10	-	-	-	0,28	B
	Separado(a) / Divorciado(a)	21	10,01	0,92	9,59	10,44	0,37	D
	Unión libre	146	9,78	1,06	9,60	9,95	0,61	E

Competencia Genérica	Grupo de Edad	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Razonamiento cuantitativo	Soltero(a)	3485	10,25	0,98	10,21	10,28	-	A
	Casado(a)	164	9,97	0,94	9,83	10,11	0,42	C
	Viudo(a)	1	9,60	-	-	-	0,80	E
	Separado(a) / Divorciado(a)	21	10,06	1,21	9,51	10,61	0,32	B
	Unión libre	146	9,66	0,88	9,52	9,81	0,74	D
Competencias ciudadanas	Soltero(a)	3485	10,11	1,00	10,07	10,14	0,99	D
	Casado(a)	164	9,84	1,30	9,64	10,04	0,97	C
	Viudo(a)	1	11,10	-	-	-	-	A
	Separado(a) / Divorciado(a)	21	10,09	1,18	9,55	10,62	0,86	B
	Unión libre	146	9,64	1,11	9,46	9,82	1,32	E

Fuente: elaboración propia

3.5.4 Efecto del estrato socioeconómico

Los estratos más altos (cuatro y cinco) ocupan las primeras posiciones en rendimiento en la mayoría de las competencias observadas, excepto en comunicación escrita, en la cual, el estrato cinco, presenta bajo rendimiento. Los estratos más bajos (uno y rural) tienen bajos rendimientos en la mayoría de competencias, a excepción de competencia ciudadana, en donde el rural ocupa un tercer lugar, tal como se observa en la tabla 18. En comunicación escrita, las diferencias estandarizadas no son relevantes con relación al referente, con excepción del rural que sí presenta una diferencia importante.

Tabla 18. Estrato Socioeconómico y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Estrato Social	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	Uno	1219	10,19	1,08	10,13	10,25	0,17	D
	Dos	1610	10,31	1,04	10,26	10,36	0,06	C
	Tres	794	10,37	1,09	10,29	10,44	0,01	B
	Cuatro	145	10,38	1,09	10,20	10,55	-	A
	Cinco	14	10,18	1,18	9,49	10,86	0,18	E
	Rural	14	9,73	1,34	8,96	10,50	0,58	F
Inglés	Uno	1225	9,94	0,98	9,89	10,00	1,58	F
	Dos	1616	10,28	1,16	10,22	10,34	1,05	D
	Tres	799	10,54	1,30	10,45	10,63	0,74	C
	Cuatro	146	11,00	1,48	10,75	11,24	0,34	B
	Cinco	14	11,51	1,70	10,53	12,49	-	A
	Rural	14	9,95	0,51	9,65	10,25	1,24	E
Lectura crítica	Uno	1225	10,08	0,98	10,03	10,14	0,62	F
	Dos	1616	10,37	0,93	10,33	10,42	0,34	D
	Tres	798	10,57	0,97	10,50	10,64	0,12	C
	Cuatro	146	10,69	1,00	10,53	10,86	-	A
	Cinco	14	10,59	1,35	9,81	11,36	0,10	B
	Rural	14	10,30	0,75	9,87	10,73	0,40	E
Razonamiento cuantitativo	Uno	1225	10,02	0,92	9,97	10,07	0,96	F
	Dos	1616	10,21	0,95	10,16	10,26	0,74	E
	Tres	798	10,42	1,02	10,35	10,49	0,49	C
	Cuatro	146	10,59	1,12	10,41	10,77	0,27	B
	Cinco	14	10,93	2,24	9,63	12,22	-	A
	Rural	14	10,09	0,91	9,57	10,62	0,49	D
Competencias ciudadanas	Uno	1225	9,84	1,02	9,79	9,90	1,09	F
	Dos	1616	10,12	0,98	10,07	10,17	0,85	E
	Tres	798	10,28	0,98	10,22	10,35	0,68	D
	Cuatro	146	10,33	1,32	10,11	10,54	0,47	B
	Cinco	14	10,96	1,34	10,18	11,73	-	A
	Rural	14	10,15	1,36	9,36	10,94	0,60	C

Fuente: elaboración propia

3.5.5 Efecto del hogar actual de residencia

En todas las competencias genéricas, los estudiantes que viven en un hogar habitual o permanente, superan en rendimiento a aquellos que viven en un hogar temporal por estudio, aunque solo alcanzan una diferencia relevante en las competencias de inglés y lectura crítica (ver tabla 19).

3.5.6 Efecto de ser cabeza de familia

Los estudiantes que son cabeza de familia tienen rendimientos más bajos que aquellos que no lo son, en todas las competencias genéricas, a excepción de comunicación escrita, cuya diferencia es trivial; las diferencias entre estos dos grupos son pequeñas en las pruebas observadas (ver tabla 20).

Tabla 19. Hogar actual y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Hogar actual	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	Habitual o permanente	2186	10,33	1,05	10,28	10,37	-	A
	Temporal por estudio	1619	10,23	1,08	10,18	10,28	0,09	B
Inglés	Habitual o permanente	2195	10,39	1,25	10,34	10,44	-	A
	Temporal por estudio	1628	10,08	1,07	10,03	10,13	0,26	B
Lectura crítica	Habitual o permanente	2194	10,44	0,97	10,40	10,48	-	A
	Temporal por estudio	1628	10,19	0,97	10,14	10,24	0,26	B
Razonamiento cuantitativo	Habitual o permanente	2194	10,26	1,04	10,22	10,30	-	A
	Temporal por estudio	1628	10,14	0,90	10,10	10,19	0,12	B
Competencias ciudadanas	Habitual o permanente	2194	10,16	1,05	10,12	10,21	-	A
	Temporal por estudio	1628	9,96	0,99	9,92	10,01	0,19	B

Fuente: elaboración propia

Tabla 20. Cabeza de familia y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Cabeza de flia.	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	No	3348	10,30	1,06	10,27	10,34	-	A
	Si	457	10,14	1,08	10,04	10,24	0,16	B
Inglés	No	3362	10,31	1,18	10,27	10,35	-	A
	Si	461	9,84	1,13	9,74	9,95	0,40	B
Lectura crítica	No	3361	10,39	0,95	10,36	10,42	-	A
	Si	461	9,94	1,07	9,84	10,04	0,46	B
Razonamiento cuantitativo	No	3361	10,26	0,97	10,23	10,29	-	A
	Si	461	9,86	1,01	9,77	9,96	0,40	B
Competencias ciudadanas	No	3361	10,11	1,00	10,08	10,15	-	A
	Si	461	9,83	1,18	9,72	9,94	0,27	B

Fuente: elaboración propia

3.5.7 Efecto de tener trabajo

En las cinco competencias genéricas, se observó que los estudiantes que se ven obligados a trabajar para pagar su matrícula, tienen menor rendimiento que el resto de estudiantes que no trabajan o que si lo hacen, es por práctica o por ganar experiencia (tabla 21).

Tabla 21. Trabaja y desempeño académico de las Competencias Genéricas de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en el período 2012-2014.

Competencia Genérica	Trabaja	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Comunicación escrita	No	2309	10,29	1,05	10,24	10,33	0,24	C
	Si, por Pago matrícula	921	10,21	1,10	10,14	10,28	0,30	D
	Si, por Práctica estudio	175	10,55	1,13	10,38	10,71	-	A
	Si, por Experiencia	386	10,34	1,05	10,23	10,44	0,19	B

Competencia Genérica	Trabaja	N	Media	Desviación estándar	IC media al 95%		d Cohen	Grupo
					LI	LS		
Inglés	No	2320	10,27	1,21	10,22	10,31	0,20	C
	Si, por Pago matrícula	925	10,10	1,09	10,03	10,17	0,36	D
	Si, por Práctica estudio	177	10,50	1,35	10,30	10,70	-	A
	Si, por Experiencia	387	10,46	1,16	10,35	10,58	0,03	B
Lectura crítica	No	2319	10,38	0,96	10,34	10,42	0,07	C
	Si, por Pago matrícula	925	10,18	1,04	10,11	10,24	0,26	D
	Si, por Práctica estudio	177	10,44	0,88	10,31	10,57	-	A
	Si, por Experiencia	387	10,40	0,91	10,31	10,49	0,04	B
Razonamiento cuantitativo	No	2319	10,23	0,96	10,19	10,26	0,13	B
	Si, por Pago matrícula	925	10,12	1,02	10,06	10,19	0,22	D
	Si, por Práctica estudio	177	10,16	0,87	10,04	10,29	0,19	C
	Si, por Experiencia	387	10,35	1,04	10,25	10,46	0,0	A
Competencias ciudadanas	No	2319	10,12	1,00	10,08	10,16	-	A
	Si, por Pago matrícula	925	9,96	1,05	9,89	10,02	0,16	D
	Si, por Práctica estudio	177	10,10	0,99	9,96	10,25	0,01	B
	Si, por Experiencia	387	10,11	1,12	9,99	10,22	0,01	C

Fuente: elaboración propia

CAPÍTULO IV

APLICACIÓN DE CRISP-DM EN LA DETECCIÓN DE PATRONES DE RENDIMIENTO EN LAS PRUEBAS SABER PRO

La investigación realizada es de tipo descriptivo, bajo el enfoque cuantitativo, aplicando un diseño no experimental. Se tomó como referencia las fases de la metodología CRISP-DM (Cross Industry Standard Process for Data Mining). Se trata de un modelo de proceso de minería de datos que describe los enfoques comunes que utilizan los expertos en minería de datos; se constituye en uno de los modelos utilizados, principalmente, en los ambientes académico e industrial, y es la guía de referencia más ampliamente utilizada en el desarrollo de este tipo de proyectos (Hernández, Ramírez, & Ferri, 2005).

La metodología CRISP-DM contempla seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación, que se describen a continuación.

4.1 COMPRENSIÓN DEL NEGOCIO O PROBLEMA

En esta fase, se realizaron las actividades que permitieron profundizar y apropiarse de una manera completa el conocimiento acerca de las pruebas Saber Pro y de las competencias genéricas que evalúa, que posibilitaron la recolección de los datos correctos para interpretar adecuadamente los resultados.

El examen de estado de calidad de la educación superior, Saber Pro, está compuesto por módulos de competencias genéricas y específicas. Las primeras, son entendidas como aquellas que deben desarrollar todos los estudiantes, sin distinción de su área de conocimiento; mientras que las segundas, son aplicadas según grupos de programas que tienen características de formación similares (Icfes, 2017).

Las competencias genéricas para el siglo XXI, según el ICFES (2016), 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 (Icfes, 2016).

La fundamentación teórica requerida sobre desempeño académico, competencias genéricas y pruebas Saber Pro, se encuentra detallada en el capítulo I del presente texto.

4.2 COMPRENSIÓN DE LOS DATOS

En esta fase, se identificó, recopiló y familiarizó con la información disponible en las bases de datos del ICFES, sobre los resultados de competencias genéricas en las pruebas Saber Pro aplicadas en los años 2012, 2013 y 2014 a nivel nacional, y sobre los datos socioeconómicos, académicos e institucionales de los estudiantes de programas profesionales que presentaron esta prueba.

Dado que los archivos del ICFES tienen diferentes estructuras, inicialmente se construyeron tres repositorios con ayuda del SGBD PostgreSQL, denominados saberpro_2012, saberpro_2013 y saberpro_2014, cuyo número de registros y atributos se muestran en la tabla 22.

Tabla 22. Repositorios pruebas Saber Pro en el período 2012-2014.

Repositorio	No. Atributos	No. Registros
Saberpro_2012	98	253013
Saberpro_2013	98	245877
Saberpro_2014	88	221966
Total		720856

Fuente: elaboración propia

Analizando los atributos de estos repositorios, se encuentran casos en los que la información registrada de un año no se reporta en los demás años, y viceversa. También existen casos en los que se solicita la misma información en todos los años, pero con un formato o un nombre de atributo diferente. Los repositorios de datos de los años 2012 y 2013 tienen exactamente los mismos atributos, y en comparación con el repositorio del año 2014, contienen 67 atributos comunes. Este análisis se encuentra en la tabla 23.

Tabla 23. Análisis de atributos de los repositorios pruebas Saber Pro 2012-2014.

#	Atributo	12	13	14	Observaciones
1	capr_desempeno_comu	✗	✗	✓	El atributo sólo está presente en el 2014
2	capr_desempeno_ingl	✗	✗	✓	Sólo está presente en el 2014
3	capr_desempeno_lect	✗	✗	✓	Sólo está presente en el 2014
4	capr_desempeno_razo	✗	✗	✓	Sólo está presente en el 2014
5	capr_puntaje_comp	✗	✗	✓	Sólo está presente en el 2014
6	capr_puntaje_comu	✗	✗	✓	Sólo está presente en el 2014
7	capr_puntaje_ingl	✗	✗	✓	Sólo está presente en el 2014
8	capr_puntaje_lect	✗	✗	✓	Sólo está presente en el 2014

#	Atributo	12	13	14	Observaciones
9	capr_puntaje_razo	✗	✗	✓	Sólo está presente en el 2014
10	dipo_nombredepartamento	✗	✗	✓	Sólo está presente en el 2014
11	dipo_nombremunicipio	✗	✗	✓	Sólo está presente en el 2014
12	econ_area_vive	✓	✓	✗	NO está presente en el 2014
13	econ_material_pisos	✓	✓	✓	Está presente en todos los repositorios
14	econ_sn_automovil	✓	✓	✓	Está presente en todos los repositorios
15	econ_sn_celular	✓	✓	✓	Está presente en todos los repositorios
16	econ_sn_computador	✓	✓	✓	Está presente en todos los repositorios
17	econ_sn_dvd	!	!	!	A partir del año 2012, el atributo cambió de nombre: de econ_sn_dvd a econ:sn_reproductor_dvd
18	econ_sn_reproductordvd	✓	✓	✓	A partir del año 2012, el atributo cambió de nombre: de econ_sn_dvd a econ:sn_reproductor_dvd
19	econ_sn_horno	✓	✓	✓	Está presente en todos los repositorios
20	econ_sn_internet	✓	✓	✓	Está presente en todos los repositorios
21	econ_sn_lavadora	✓	✓	✓	Está presente en todos los repositorios
22	econ_sn_microondas	✓	✓	✓	Está presente en todos los repositorios
23	econ_sn_nevera	✓	✓	✓	Está presente en todos los repositorios
24	econ_sn_servicio_tv	✓	✓	✓	Está presente en todos los repositorios
25	econ_sn_telefonia	✓	✓	✓	Está presente en todos los repositorios
26	es_especifica	✓	✓	✗	Sólo está presente en los años 2012 y 2013

#	Atributo	12	13	14	Observaciones
27	estu_actividad_no_curso	✓	✓	!	En el año 2012 y 2013 se llamaba “estu_actividad_no_curso” y para el 2014 se llamaba “estu_actividadnocurso”
28	estu_actividadnocurso	!	!	✓	En el año 2012 y 2013 se llamaba “estu_actividad_no_curso” y para el 2014 se llamaba “estu_actividadnocurso”
29	estu_area_conoc	✓	✓	✗	NO está presente en el 2014
30	estu_cod_aplicacion	✓	✓	✓	Está presente en todos los repositorios
31	estu_cod_grupo_ref	✓	✓	✗	NO está presente en el 2014
32	estu_colegiotermino_cod	✓	✓	✓	Está presente en todos los repositorios
33	estu_colegiotermino_dept	✓	✓	✗	Sólo presente en el 2012 y 2013
34	estu_colegiotermino_mpio	✓	✓	✗	Sólo presente en el 2012 y 2013
35	estu_colegiotermino_nombre	✓	✓	✗	Sólo presente en el 2012 y 2013
36	estu_colegiotermino_pais	✓	✓	✗	Sólo presente en el 2012 y 2013
37	estu_consecutivo	!	!	✓	Cambió de nombre en el 2014
38	estuconsecutivo	✓	✓	!	Cambió de nombre en el 2014
39	estu_curso_docentesnies	✓	✓	!	Se llama “estu_curso_docentesnies” en el 2012 y 2013, pero para el 2014 cambió de nombre a “estu_cursodocentesies”
40	estu_cursodocentesies	!	!	✓	Se llama “estu_curso_docentesnies” en el 2012 y 2013, pero para el 2014 cambió de nombre a “estu_cursodocentesies”
41	estu_curso_iesexterno	✓	✓	!	Se llama “estu_curso_iesexterno” en el 2012 y 2013, pero para el 2014 cambió de nombre a “estu_cursoiesexterno”
42	estu_cursoiesexterno	!	!	✓	Se llama “estu_curso_iesexterno” en el 2012 y 2013, pero para el 2014 cambió de nombre a “estu_cursoiesexterno”

#	Atributo	12	13	14	Observaciones
43	estu_curso_pregicfes	✓	✓	!	Se llama “estu_curso_pregicfes” en el 2012 y 2013, pero para el 2014 cambió de nombre a estu_cursopregicfes”
44	estu_cursopregicfes	!	!	✓	Se llama “estu_curso_pregicfes” en el 2012 y 2013, pero para el 2014 cambió de nombre a estu_cursopregicfes”
45	estu_disc_invidente	✓	✓	✓	Está presente en todos los repositorios
46	estu_disc_motriz	✓	✓	✓	Está presente en todos los repositorios
47	estu_disc_sordo_ ceguera	✓	✓	✓	Está presente en todos los repositorios
48	estu_disc_sordo_con_ interprete	✓	✓	✓	Está presente en todos los repositorios
49	estu_disc_sordo_sin_ interprete	✓	✓	✓	Está presente en todos los repositorios
50	estu_dormitorios	✓	✓	✓	Está presente en todos los repositorios
51	infa_dormitorios	✓	✓	✓	Está presente en todos los repositorios
52	estu_estado_civil	✓	✓	✓	Está presente en todos los repositorios
53	estu_estrato	✓	✓	✓	Está presente en todos los repositorios
54	estu_exam_anno_ presentacion	✓	✓	✓	Está presente en todos los repositorios
55	estu_exam_cod	✓	✓	✗	NO está presente en el 2014
56	estu_exam_codmpio_ presentacion	✓	✓	✓	Está presente en todos los repositorios
57	estu_exam_dept_ presentacion	!	!	✓	Está presente en todos los repositorios con un nombre diferente para el 2014
58	estu_exam_dpto_ presentacion	✓	✓	!	Está presente en todos los repositorios con un nombre diferente para el 2014
59	estu_exam_mpio_ presentacion	✓	✓	✓	Está presente en todos los repositorios

#	Atributo	12	13	14	Observaciones
60	estu_exam_nombre	✓	✓	✗	NO está presente en el 2014
61	estu_exam_semestre_presentacion	✓	✓	!	Está presente en todos los repositorios con nombre diferente para cada año
62	estu_exam_semestre_prestacion	!	!	✓	Está presente en todos los repositorios con nombre diferente para cada año
63	estu_exam_semestre_pretacion	!	!	!	Está presente en todos los repositorios con nombre diferente para cada año
64	estu_genero	✓	✓	✓	Está presente en todos los repositorios
65	estu_grupo_referencia	✓	✓	✗	NO está presente en el 2014
66	estu_gruporefprogper_id	✓	✓	✗	Sólo está presente en los años 2012 y 2013
67	estu_gruporefprogper_periodo	✓	✓	✗	Sólo está presente en los años 2012 y 2013
68	estu_hogar_actual	✓	✓	✓	Está presente en todos los repositorios
69	estu_horas_trabajo	✓	✓	✓	Presente en todos los repositorios
70	estu_instituto_preparacion	✓	✓	!	Presente en todos los repositorios pero con nombres diferentes
71	estu_institutopreparacion	!	!	✓	Presente en todos los repositorios pero con nombres diferentes
72	estu_metodo_prgm	✓	✓	✗	NO está presente en el 2014
73	estu_nacimiento_anno	✓	✓	✓	Presente en todos los repositorios
74	estu_nacimiento_dia	✓	✓	✓	Presente en todos los repositorios
75	estu_nacimiento_mes	✓	✓	✓	Presente en todos los repositorios
76	estu_nivel_prgm_academico	✓	✓	✗	NO está presente en el 2014

#	Atributo	12	13	14	Observaciones
77	estu_nucleo_pregrado	✓	✓	✗	NO está presente en el 2014
78	estu_pais_reside	✓	✓	✗	NO está presente en el 2014
79	estu_pje_creditos	✓	✓	✓	Está presente en todos los repositorios
80	estu_prac_id_prgrm_academico	!	!	✗	Presente hasta el 2013 con nombres diferentes, en el 2014 no está.
81	estu_prgm_academico_cod	✓	✓	✗	Presente hasta el 2013 con nombres diferentes, en el 2014 no está.
82	estu_prgm_academico	✓	✓	✗	NO está presente en el 2014
83	estu_refuerzoareas	✓	✓	✓	No está presente en el año 2011
84	estu_refuerzogsa	✓	✓	✓	No está presente en el año 2011
85	estu_reside_codmpio	✓	✓	✓	Presente en todos los repositorios
86	estu_reside_dept	✓	✓	✓	No está presente en el año 2011
87	estu_reside_mpio	✓	✓	✓	No está presente en el año 2011
88	estu_semestre_cursa	!	!	!	Presente en todos los repositorios con nombres diferentes
89	estu_semestre_cursando	✓	✓	✓	Presente en todos los repositorios con nombres diferentes
90	estu_sn_cabeza_fmliia	✓	✓	✓	Presente en todos los repositorios
91	estu_sn_matricula_beca	✓	✓	✓	Está presente en todos los repositorios
92	estu_sn_matricula_credito	✓	✓	✓	Está presente en todos los repositorios
93	estu_sn_matricula_padres	✓	✓	✓	Está presente en todos los repositorios

#	Atributo	12	13	14	Observaciones
94	estu_sn_matricula_propio	✓	✓	✓	Está presente en todos los repositorios
95	estu_titulo_bto	✓	✓	✓	Está presente en todos los repositorios
96	estu_tomo_curso	✓	✓	!	Presente en todos los repositorios pero con nombres diferentes
97	estu_tomocurso	!	!	✓	Presente en todos los repositorios pero con nombres diferentes
98	estu_trabaja	✓	✓	✓	Está presente en todos los repositorios
99	estu_zona	✓	✓	✗	NO está presente en el 2014
100	eval_tipo_documento	✓	✓	✗	Sólo está presente en los años 2012 y 2013
101	fami_cod_educa_madre	✓	✓	✓	Está presente en todos los repositorios
102	fami_cod_educa_padre	✓	✓	✓	Está presente en todos los repositorios
103	fami_cod_ocup_madre	✓	✓	✓	Está presente en todos los repositorios
104	fami_cod_ocup_padre	✓	✓	✓	Está presente en todos los repositorios
105	fami_ing_fmliar_mensual	✓	✓	✓	Está presente en todos los repositorios
106	fami_nivel_sisben	✓	✓	✓	Está presente en todos los repositorios
107	fami_num_pers_cargo	✓	✓	✓	Está presente en todos los repositorios
108	fami_num_pers_grup_fam	✓	✓	✓	Está presente en todos los repositorios
109	grre_codigoicfes	✗	✗	✓	Sólo está presente en el 2014
110	grre_nombre	✗	✗	✓	Sólo está presente en el 2014

#	Atributo	12	13	14	Observaciones
111	inst_caracter_academico	✓	✓	!	Está presente en todos los repositorios con nombre diferente para cada año
112	inst_caracteracademico	!	!	✓	Está presente en todos los repositorios con nombre diferente para cada año
113	inst_cod_institucion	✓	✓	✓	Está presente en todos los repositorios
114	inst_nombre_institucion	✗	✗	!	No está presente en 2012 y 2013, pero sí en el 2014 aunque con nombre diferente
115	inst_nombreinstitucion	✗	✗	✓	No está presente en 2012 y 2013, pero sí está en 2014 aunque con nombre diferente
116	inst_origen	!	!	✗	No está presente en el 2014, para los demás años tiene nombres diferentes
117	inst_origen_cod	✓	✓	✗	No está presente en el 2014, para los demás años tiene nombres diferentes
118	inst_pertenece_evaluado	✓	✓	✗	Sólo está presente en los años 2012 y 2013
119	inst_prac_codigoicfes	✗	✗	✓	Sólo está presente en el 2014
120	inst_prac_consecutivosnies	!	!	✓	Presente en todos los repositorios pero con nombres diferentes
121	inst_prac_conseutivosnies	✓	✓	!	Presente en todos los repositorios, pero con nombres diferentes
122	inst_sector	✗	✗	✓	Sólo está presente en el 2014
123	inst_vlr_matricula_ant	✓	✓	✓	Está presente en todos los repositorios
124	mod_comp_ciudadanas_punt	✓	✓	✗	NO está presente en el 2014
125	mod_comunica_escrita_desem	✓	✓	✗	NO está presente en el 2014
126	mod_comunica_escrita_punt	✓	✓	✗	NO está presente en el 2014

#	Atributo	12	13	14	Observaciones
127	mod_ingles_desem	✓	✓	✗	NO está presente en el 2014
128	mod_ingles_punt	✓	✓	✗	NO está presente en el 2014
129	mod_lectura_critica	✓	✓	✗	NO está presente en el 2014
130	mod_razona_cuantitativo_punt	✓	✓	✗	NO está presente en el 2014
131	plan_codigodaneinstitucion	✗	✗	✓	Sólo está presente en el 2014
132	plan_codigodanesede	✗	✗	✓	Sólo está presente en el 2014
133	plan_municipiosdr	✗	✗	✓	Sólo está presente en el 2014
134	plan_nombrecomun	✗	✗	✓	Sólo está presente en el 2014
135	prac_nombre	✗	✗	✓	Sólo está presente en el 2014
136	prac_titulo	✓	✓	✗	Presente en los años 2012 y 2013
137	prueba	✓	✓	✓	Presente en todos los repositorios

Fuente: elaboración propia

Teniendo en cuenta los atributos comunes, equivalentes (con diferentes nombres pero con los mismos valores), los no comunes, y eliminando los atributos que se presentan en un solo año, estos repositorios se integraron en un único repositorio, denominado SaberPro_2012-2014, el cual quedó compuesto, inicialmente por 720.856 registros y 96 atributos, donde se encuentran almacenados los datos de todos los estudiantes de programas técnico profesionales, tecnológicos y profesionales que presentaron las pruebas Saber Pro 2012,2013 y 2014, a nivel nacional. De este repositorio se seleccionaron únicamente los registros de estudiantes de programas profesionales de la Universidad de Nariño, quedando un repositorio con 3825 registros y 96 atributos, que se lo denominó saberpro_udenar_2012_2014, el cual, se constituyó en insumo para las fases subsiguientes. En la tabla 24 se encuentra el diccionario de datos de este repositorio, donde se describe cada uno de sus atributos.

Tabla 24. Diccionario de datos del repositorio saberpro_udenar_2012_2014

No.	Atributo	Descripción	Valores
1	estu_consecutivo	Código identificador del estudiante	
2	estu_cod_aplicacion	Nombre de la aplicación	EK+año-semester
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 24a
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

No.	Atributo	Descripción	Valores
19	estu_exam_mpio_presentacion	Nombre municipio donde presentó el examen	
20	estu_exam_dpto_presentacion	Nombre del departamento donde presentó el examen	
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 24b
26	inst_caracter_academico	Carácter Académico	Ver tabla 24c
27	estu_prac_id_prgrm_academico	Código identificador programa académico	
28	estu_prgm_academico	Nombre programa académico	
29	estu_nivel_prgm_academico	Nivel del programa académico que cursa el estudiante	Universitario
30	estu_metodo_prgm	Metodología del programa académico bajo la cual cursa el estudiante	Presencial Semipresencial 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 24d

No.	Atributo	Descripción	Valores
33	estu_area_conoc	Nombre del área de conocimiento a la que pertenece el programa académico del estudiante	Ver tabla 24e
34	estu_nucleo_pregrado	Nombre del núcleo de pregrado al que pertenece el programa académico del estudiante	Ver tabla 24f
35	estu_grupo_referencia	Nombre del grupo de referencia al que pertenece el programa académico del estudiante	Ver tabla 24g
36	estu_cod_grupo_ref	Código del grupo de referencia al que pertenece el programa académico del estudiante	Ver tabla 24g
37	estu_semestre_cursa	Semestre que cursa actualmente el estudiante	1 a 12 semestres 12 Significa 12 o más.
38	estu_pje_creditos	Porcentaje de créditos cursados y aprobados por el estudiante	Ver tabla 24h
39	inst_vlr_matricula_ant	Valor anual de la matrícula el año anterior	Ver tabla 24i
40	estu_sn_matricula_propio	Si canceló la matrícula año anterior con recursos propios	Si(1) No(0)
41	estu_sn_matricula_padres	Si canceló la matrícula año anterior con recursos de los padres o familiares	Si(1) No(0)
42	estu_sn_matricula_beca	Si canceló la matrícula año anterior con recursos de beca o subsidio	Si(1) No(0)
43	estu_sn_matricula_credito	Si canceló la matrícula año anterior con recursos de crédito	Si(1) No(0)

No.	Atributo	Descripción	Valores
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_fmilia	Si el estudiante es cabeza de familia.	Si (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 1j
52	fami_cod_educa_madre	Código nivel educativo más alto alcanzado por la madre del estudiante	Ver Tabla 24j
53	fami_sn_lee_escribe_padre	Si puede leer y escribir el padre del estudiante	Si (1) No (2)
54	fami_sn_lee_escribe_madre	Si puede leer y escribir la madre del estudiante	Si (1) No (2)
55	fami_cod_ocup_padre	Ocupación actual del padre del estudiante (o última si falleció)	Ver Tabla 24k
56	fami_cod_ocup_madre	Ocupación actual de la madre del estudiante (o última si falleció)	Ver tabla 24k

No.	Atributo	Descripción	Valores
57	estu_estrato	Estrato socioeconómico de la vivienda donde reside actualmente el estudiante según el recibo del servicio de energía eléctrica	Ver tabla 24l
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 24m
60	econ_cuartos	Número total de cuartos que dispone la residencia del estudiante	
61	fami_nivel_sisben	Nivel de clasificación del SISBEN	Ver Tabla 24n
62	econ_material_pisos	Material de los pisos que predomina en la vivienda donde reside el hogar habitual o permanente del	Ver Tabla 24o
63	econ_sn_televisor	Si hay televisor en el hogar habitual del estudiante	Si (1) No (0)
64	econ_sn_motocicleta	Si hay una motocicleta en el hogar habitual del estudiante	Si (1) No (0)
65	econ_sn_energia	Si posee el servicio de energía eléctrica en el hogar habitual del estudiante	Si (1) No (0)
66	econ_sn_acueducto	Si posee el servicio de acueducto en el hogar habitual del estudiante	Si (1) No (0)
67	econ_sn_alcantarillado	Si posee alcantarillado en el hogar habitual del estudiante	Si (1) No (0)
68	econ_sn_aseo	Si posee el servicio de aseo en el hogar habitual del estudiante	Si (1) No (0)

No.	Atributo	Descripción	Valores
69	econ_sn_estufa	Si posee estufa a gas o eléctrica en el hogar habitual del estudiante	Si (1) No (0)
70	econ_sn_telefonia	Si posee el servicio de telefonía en el hogar habitual del estudiante	(1): Si (0): no
71	econ_sn_internet	Si posee el servicio de Internet en el hogar habitual del estudiante	(1): Si (0): no
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): Si (0): no
73	econ_sn_computador	Si hay un computador en el hogar habitual del estudiante	(1): Si (0): no
74	econ_sn_celular	Si hay un teléfono celular en el hogar habitual del estudiante	(1): Si (0): no
75	econ_sn_dvd	Si hay un reproductor DVD en el hogar habitual del estudiante	(1): Si (0): no
76	econ_sn_lavadora	Si hay una máquina lavadora de ropas en el hogar habitual del estudiante	(1): Si (0): no
77	econ_sn_microondas	Si hay un horno microondas en el hogar habitual del estudiante	(1): Si (0): no
78	econ_sn_automovil	Si hay un automóvil particular en el hogar habitual del estudiante	(1): Si (0): no
79	econ_sn_horno	Si hay un horno eléctrico o a gas en el hogar habitual del estudiante	(1): Si (0): no
80	econ_sn_nevera	Si hay una nevera o enfriador en el hogar habitual del estudiante	(1): Si (0): no

No.	Atributo	Descripción	Valores
81	infa_dormitorios	Número de dormitorios en el hogar habitual del estudiante	1 a 10 dormitorios 10 representa 10 o mas
82	fami_ing_fmliar_mensual	ingresos mensuales familiares en salarios mínimos	Ver tabla 24p
83	estu_trabaja	Si estudiante trabaja o no	Ver tabla 24q
84	estu_horas_trabajo	Nro. de horas / semanal si trabaja	
85	estu_otro_idioma_lee	Si lee otro idioma	Ver tabla 24r
86	estu_otro_idioma_habla	Si habla otro idioma	Ver tabla 24r
87	estu_nivel_postgrado	Si el estudiante posee un postgrado	05 -Especialización 06 -Maestría 07 -Doctorado 98 -Ninguno
88	estu_etnia	Si el estudiante pertenece a una etnia	Ver tabla 24s
89	mod_lectura_critica	Puntaje asignado al Módulo de Lectura Crítica	
90	mod_comunica_escrita_punt	Puntaje asignado al módulo de Comunicación Escrita	
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	

No.	Atributo	Descripción	Valores
95	mod_comp_ciudadanas_punt	Puntaje asignado al módulo de Competencias Ciudadanas	
96	prueba	Periodo en el que se presenta la prueba	20121,20123,20131,20133,20142,20143

Fuente: ICFES

Tabla 24a. Estado Civil

Código	Descripción
01	Soltero(a)
02	Casado(a)
03	Viudo(a)
04	Separado(a) y/o Divorciado
05	Unión Libre

Fuente: ICFES

Tabla 24b. Origen institución

Descripción
No Oficial – Corporación
No Oficial - Fundación
No Oficial
Oficial
Oficial Departamental
Oficial Municipal
Oficial Nacional
Régimen Especial

Fuente: ICFES

Tabla 24c. Carácter académico

Descripción
Académico
Escuela Tecnológica
Institución Tecnológica
Institución Universitaria
Normalista
Técnica Profesional
Técnica Profesional
Universidad

Fuente: ICFES

Tabla 24d. Código jornada

Código	Descripción
1	Diurno
2	Nocturno
3	Educación a Distancia
4	Mixta
5	Registro Calificado
6	Reg. Calif Acred. Voluntaria
9	No aplica
10	Registro Simple
11	Registro Alta Calidad
12	Jornada Única

Fuente: ICFES

Tabla 24e. 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

Fuente: ICFES

Tabla 24f. 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
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

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

Fuente: ICFES

Tabla 24g. 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

Fuente: ICFES

Tabla 24h. 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% ó más (AÑOS 2008-1, 2008-2 y 2009,2)
4	Más del 90% (AÑO 2010-2)

Fuente: ICFES

Tabla 24i. 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

Fuente: ICFES

Tabla 24j. Educación del padre o madre

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
15	Educación profesional sin título
16	Educación profesional con título
17	Postgrado

Fuente: ICFES

Tabla 24k. Ocupación del padre o madre

Código	Descripción
13	Empresario(a), dueño(a) de empresa industrial, comercial, agropecuaria, de servicios, de más de 10 trabajadores.
14	Pequeño empresario(a), dueño(a) de microempresa, pequeño negocio familiar de finca, parcela, que vive de su explotación
15	Empleado con cargo como director(a), 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
25	Estudiante, persona que la mayor parte del tiempo se dedica al estudio
26	Otra actividad u ocupación
99	No sabe

Fuente: ICFES

Tabla 24l. 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 socio económica

Fuente: ICFES

Tabla 24m. 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
15	educación profesional sin título
16	educación profesional con título
17	postgrado
88	No tiene hermanos mayores
99	no sabe

Fuente: ICFES

Tabla 24n. 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

Fuente: ICFES

Tabla 24o. 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

Fuente: ICFES

Tabla 24p. 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

Fuente: ICFES

Tabla 24q. Trabajo del estudiante

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

Fuente: ICFES

Tabla 24r. Lee o habla otro idioma

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

Fuente: ICFES

Tabla 24s. 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

Fuente: ICFES

Con base en la conceptualización de desempeño académico y los antecedentes teóricos sobre los factores que intervienen en él, los cuales se asocian y colindan unos con otros, los 96 atributos del conjunto de datos saberpro_udenar_2012_2014, se clasificaron en tres dimensiones: socioeconómica, académica, e institucional. Los resultados se presentan en la tabla 25.

Tabla 25. Clasificación de atributos en dimensiones

Dimensión	Atributo
Socioeconómica	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_fmliar, 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 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

Dimensión	Atributo
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, prueba
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, mod_comp_ciudadanas_punt

Fuente: esta investigación

4.3 PREPARACIÓN DE LOS DATOS

En esta fase, los 96 atributos del repositorio *saberpro_ude-nar_2012_2014*, fueron depurados, teniendo en cuenta la calidad de los datos y las técnicas de minería de datos a aplicar; 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 aplicar las técnicas de minería de datos.

El proceso de limpieza y transformación de este repositorio se realizó como se describe en seguida. Con el fin de determinar la calidad de los datos, se realizó una exploración inicial de los datos almacenados en el repositorio `saberpro_udenar_2012_2014`. Como resultado de este proceso, se efectuó una primera selección de atributos, y se descartó los que tenían un alto porcentaje de valores nulos, para los que no era posible encontrar sus valores a través de fuentes externas de datos (ver tabla 26). También se descartó los atributos con valores constantes y aquellos que solo servían como identificadores de los estudiantes (ver tabla 27).

Tabla 26. Atributos con porcentaje alto de datos nulos

Atributo	Nulos (%)
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_sordo_ceguera, estu_disc_invidente, estu_disc_sordo_con_interprete, estu_disc_sordo_sin_interprete, estu_disc motriz	99.82
estu_anno_egreso	97.77
econ_area_vive	86.82

Fuente: elaboración propia

Tabla 27. Atributos con valores constantes y atributos identificadores

Atributo	Valor
estu_consecutivo	SBPRO+identificador
estu_cod_aplicacion	EK20121, EK20122,...
estu_pais_reside	COLOMBIA
estu_exam_nombre	EXAMEN ESPECÍFICO

Atributo	Valor
estu_exam_cod	552, 567, 587,593
estu_exam_codmpio_presentacion	52001
estu_exam_mpio_presentacion	PASTO
estu_exam_dpto_presentacion	NARIÑO
inst_cod_institucion	1206
inst_nombre_institucion	UNIVERSIDAD DE NARIÑO
inst_origen	OFICIAL DEPARTAMENTAL
inst_caracter_academico	UNIVERSIDAD
estu_nivel_prgm_academico	UNIVERSITARIO

Fuente: elaboración propia

Debido a que la alta dimensionalidad es un problema para el descubrimiento de patrones con minería de datos (Hernández, Ramirez & Ferri, 2005), uno de los criterios utilizados para resolver este problema, fue reducir el número de atributos a analizar, que se puede realizar mediante la transformación de estos, en nuevos atributos, que generalicen los datos y que ofrezcan mayor información. Con base en este criterio, en el repositorio de datos *saberpro_udenar_2012_2014*, se seleccionó los atributos que por sí mismos no tenían mayor significado, pero que integrados en uno nuevo, adquirirían mayor semántica; en este sentido, se crearon nuevos atributos en reemplazo de éstos, con lo cual se obtuvo atributos más representativos para el estudio y se redujo el número de variables a tener en cuenta en la investigación. En la tabla 28, se describe el proceso sobre cómo se construyeron estos nuevos atributos y los atributos que reemplazan. De hecho, los atributos reemplazados, fueron eliminados.

Tabla 28. 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
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	S N
estu_sn_vive_flia	Si el estudiante vive con la familia o no	Reemplaza al atributo: estu_hogar_actual	S N
estu_pers_cargo	Si el estudiante tiene personas a cargo o no	Reemplaza al atributo: fami_num_pers_cargo	S N
fami_max_nivel_educapadres	Máximo nivel educativo completo entre el padre y la madre	Reemplaza a los atributos: fami_cod_educapadre fami_cod_educamadre	PRIMARIA SECUNDARIA TÉCNICO/TECNOLÓGICO PROFESIONAL POSTGRADO NINGUNO

Atributo	Descripción	Acción	Valores
fami_ocup_padre	Ocupación del padre	Reemplaza al atributo: fami_cod_ocup_padre	DIRECTIVO EMPLEADO EMPRESARIO HOGAR INDEPENDIENTE OTRA PENSIONADO PROFESIONAL
fami_ocup_madre	Ocupación de la madre	Reemplaza al atributo: fami_cod_ocup_madre	LOS MISMOS VALORES DE LA OCUPACIÓN DEL PADRE
estu_facultad	Facultad a la que pertenece el estudiante	Se obtiene a partir del atributo estu_prog_académico	
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_prog_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
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

Atributo	Descripción	Acción	Valores
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)

Atributo	Descripción	Acción	Valores
mod_comp_ciudadanas_desemp	Desempeño del estudiante en el Módulo de competencias ciudadanas	Reemplaza a los atributos: mod_comp_ciudadanas_punt	SOBRE LA MEDIA (\geq la media) BAJO LA MEDIA ($<$ la media)
mod_general_desemp	Desempeño del estudiante en las cinco competencias	Se calcula como el promedio de los puntajes en las cinco competencias	SOBRE LA MEDIA (\geq la media) BAJO LA MEDIA ($<$ la media)

Fuente: esta investigación

Con el fin de facilitar la detección de patrones de rendimiento académico, se discretizaron los valores numéricos de ciertos atributos teniendo en cuenta un rango de valores o que las frecuencias por cada valor fueran proporcionales, con el fin de evitar sesgos al construir los modelos de minería de datos. Por esta razón, para el atributo *estu_edad* se creó un nuevo atributo, denominado *estu_rango_edad*, cuyos valores se muestran en la tabla 29.

Tabla 29. Valores discretizados del atributo *estu_rango_edad*

Valor	No. Estudiantes
[<23)	1718
[23-32)	2538
[32-41)	270
[41-50)	75
[\geq 50]	20

Fuente: esta investigación

Para la nueva variable *eco_condicion_vivienda* (ver tabla 28) se tuvo en cuenta los materiales de los pisos de la vivienda. En la tabla 30, se muestran los valores asignados a este atributo de acuerdo al material de los pisos.

Tabla 30. 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

Fuente: esta investigación.

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, la condición del hogar es BUENA; si el índice está entre 2 y 3, la condición del hogar es MEDIA; y finalmente, si el índice está entre 0 y 1, la condición del hogar es MALA. En la tabla 31, se presenta las variables que intervienen en el cálculo del índice *condición_hogar*.

Tabla 31. Cálculo del índice de condición_hogar

Electrodomésticos	Si	No
econ_sn_dvd	1	0
econ_sn_lavadora	1	0
econ_sn_microondas	1	0
econ_sn_horno	1	0
econ_sn_nevera	1	0

Fuente: esta investigación.

De igual manera, se procedió con los valores de la nueva variable *eco_condicion_tic* para los valores BUENA, MEDIA y MALA, pero esta vez, teniendo en cuenta los servicios con que cuenta el hogar. En la tabla 32, se muestran los atributos que intervienen en el cálculo del índice de *condición_tic*.

Tabla 32. Cálculo del índice de condición_tic

Servicios	Si	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

Fuente: esta investigación.

Para los valores de la variable *eco_condicion_transporte*, se consideraron dos valores: PARTICULAR, si en el hogar se disponía de un automóvil, y PÚBLICO si no se disponía de automóvil; esto teniendo en cuenta la variable *econ_sn_automovil* que reemplazó.

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

Generalmente se aceptan los valores: hasta 2.4 - sin hacinamiento; de 2.5 a 4.9 - hacinamiento medio y de 5.0 o más - hacinamiento crítico.

Teniendo en cuenta estos conceptos, el índice de hacinamiento para cada estudiante se obtuvo 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. Este proceso se muestra en la tabla 33.

Tabla 33. Asignación de valores del atributo *area_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
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

Fuente: esta investigación.

Como resultado de los procesos descritos, se obtuvo un nuevo conjunto de datos, limpio y transformado, listo para aplicar las técnicas de minería de datos; repositorio denominado *saberpro_minable_udenar*, compuesto por 3825 registros y 31 atributos. En la tabla 34 se describe este repositorio, organizado por los factores socioeconómicos, institucionales y clases.

Tabla 34. Diccionario de datos del repositorio final saberpro_minable_udenar

No.	Atributo	Descripción	Valores
Socioeconómicos			
1	estu_genero	Género	M ,F
2	estu_rango_edad	Rango de edades a la que pertenece el estudiante en el momento de presentar la prueba	Ver tabla 29
3	estu_estado_civil_nom	Estado civil del estudiante.	SOLTERO, CASADO SEPARADO/DIVORCIADO, UNION LIBRE, VIUDO
4	estu_sn_vive_flia	Si el estudiante vive con su familia o no	S, N
5	estu_sn_pers_cargo	Si el estudiante tiene personas a cargo o no	S, N
6	fami_max_nivel_educa_padres	Máximo nivel educativo completo entre el padre y la madre	PRIMARIA, SECUNDARIA, TÉCNICO/TECNOLÓGICO, PROFESIONAL, POSTGRADO NINGUNO
7	fami_ocup_padre	Ocupación del padre	DIRECTIVO, EMPLEADO, EMPRESARIO, HOGAR, INDEPENDIENTE, OTRA PENSIONADO, PROFESIONAL
8	fami_ocup_madre	Ocupación de la madre	Los mismos valores de la ocupación del padre
9	estu_estrato	Estrato socioeconómico del estudiante	ESTRATOS 1 a 6, ZONA RURAL SIN ESTRATO
10	fami_nivel_sisben	Nivel de clasificación en el SISBEN al que pertenece el estudiante	NIVELES 1, 2, 3, OTRO NIVEL, NO ESTA EN SISBEN
11	econ_condicion_vivienda	Condición de la vivienda del estudiante	BUENA, MALA, REGULAR

No.	Atributo	Descripción	Valores
Socioeconómicos			
12	eco_condicion_hogar	Condición del hogar del estudiante	BUENA, MALA, REGULAR
13	eco_condicion_transporte	Condición de transporte en el hogar del estudiante	PARTICULAR, PÚBLICO
14	eco_condicion_tic	Condición de uso de TIC en el hogar del estudiante	BUENA, REGULAR, MALA
15	eco_condicion_vive	Condición de vida del estudiante	SIN HACINAMIENTO, HACINAMIENTO MEDIO, HACINAMIENTO CRÍTICO
16	fami_ing_fmliar_mensual	Ingresos mensuales familiares en salarios mínimos	Ver tabla 1p
17	estu_sn_trabaja	Si estudiante trabaja o no	S N
Académicos			
19	estu_prgm_academico	Nombre del programa académico de la Universidad de Nariño al cual pertenece el estudiante	
20	estu_facultad	Nombre de la facultad de la Universidad de Nariño a la cual pertenece el estudiante	
21	estu_area_conoc	Nombre del área de conocimiento a la que pertenece el programa académico del estudiante	Ver tabla 24e

No.	Atributo	Descripción	Valores
Socioeconómicos			
22	estu_nucleo_pregrado	Nombre del núcleo de pregrado al que pertenece el programa académico del estudiante	Ver tabla 24f
23	estu_grupo_referencia	Nombre del grupo de referencia al que pertenece el programa académico del estudiante	Ver tabla 24g
24	area_grupo_referencia	Área del grupo de referencia de los programas	CIENCIAS HUMANAS, CIENCIAS SOCIALES, CIENCIAS NATURALES Y TÉCNICAS
Institucionales			
25	inst_prog_acreditado	Si el programa al cual pertenece el estudiante es acreditado o no según CNA	ACREDITADO, NO ACREDITADO
Clases			
26	mod_lectura_critica_desemp	Desempeño del estudiante en el Módulo de lectura crítica	SOBRE LA MEDIA (>= la media) BAJO LA MEDIA (< la media)
27	mod_comunica_escrita_desemp	Desempeño del estudiante en el Módulo de comunicación escrita	SOBRE LA MEDIA (>= la media) BAJO LA MEDIA (< la media)
28	mod_razona_cuantitativo_desemp	Desempeño del estudiante en el Módulo de razonamiento cuantitativo	SOBRE LA MEDIA (>= la media) BAJO LA MEDIA (< la media)
29	mod_ingles_desemp	Desempeño del estudiante en el Módulo de inglés	SOBRE LA MEDIA (>= la media) BAJO LA MEDIA (< la media)

No.	Atributo	Descripción	Valores
Socioeconómicos			
30	mod_comp_ciudadanas_desemp	Desempeño del estudiante en el Módulo de competencias ciudadanas	SOBRE LA MEDIA (>= la media) BAJO LA MEDIA (< la media)
31	mod_desemp_general	Desempeño promedio del estudiante en las cinco competencias	SOBRE LA MEDIA (>= la media) BAJO LA MEDIA (< la media)

Fuente: esta investigación.

A partir del repositorio *saberpro_minable_udenar* 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 25 atributos correspondientes a las dimensiones socioeconómica, académica e institucional descritos en la tabla 34 y el respectivo atributo clase dependiendo de la competencia. Se tuvo que eliminar, en cada repositorio, los registros con valores nulos en el atributo clase (competencia). La descripción de cada repositorio por competencia, se muestra en la tabla 35.

Tabla 35. Repositorios por cada competencia

Repositorio	Descripción
T3824A26LEC_udenar	Repositorio para análisis de la competencia de lectura crítica. Contiene 3824 registros y 26 atributos.
T3807A26ESC_udenar	Repositorio para análisis de la competencia de comunicación escrita. Contiene 3807 registros y 26 atributos.
T3824A26CUA_udenar	Repositorio para análisis de la competencia de razonamiento cuantitativo. Contiene 3824 registros y 26 atributos.
T3825A26ING_udenar	Repositorio para análisis de la competencia de inglés. Contiene 3825 registros y 26 atributos.
T3824A26CIU_udenar	Repositorio para análisis de la competencia ciudadana. Contiene 3824 registros y 26 atributos.
T3825A26GEN_udenar	Repositorio para análisis del desempeño general en las competencias. Contiene 3825 registros y 26 atributos.

Fuente: esta investigación.

4.4 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. La clasificación es una técnica predictiva que se aplica a problemas en los que hay que predecir nuevos datos para uno o más ejemplos que van acompañados de una salida denominada clase (Hernández et al., 2005). Con la técnica de clasificación se pretende obtener un modelo que permita predecir, para los nuevos casos de estudiantes de programas profesionales de la Universidad de Nariño, los factores socioeconómicos, académicos e institucionales, asociados a un probable buen o mal desempeño académico, en las cinco competencias genéricas evaluadas en las pruebas Saber Pro.

Para el descubrimiento de patrones de desempeño académico, se utilizó *Yachay Datamining Tool - Yachay DTM-*, herramienta de minería de datos que se construyó para ser aplicada en este proyecto de investigación. En la figura 8, se muestra el proceso para obtener el modelo de clasificación con esta herramienta. Para mayor información sobre esta herramienta, se puede consultar el manual del usuario de *Yachay DMT* (GRIAS, 2016).



Figura 8. Uso de la herramienta Yachay DTM

4.4.1 Clasificación con 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 & Kamber, 2001), (Sattler & Dunemann,

2001), (Timarán & Millán, 2006). La importancia de los árboles de decisión se debe a su capacidad para 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 conduce a una y sólo una hoja, asignando una única clase a la predicción (Hernández & Lorente, 2009).

El algoritmo de la herramienta *Yachay DTM*, utilizado para obtener el modelo de clasificación con árboles de decisión es J48, que fue incorporado de la herramienta Weka (Hall et al., 2011), la cual implementa el algoritmo C.45 (Quinlan, 1993). El algoritmo J48 se basa en la utilización del criterio ratio de ganancia (*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 éste ha sido inducido (Hernández & 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, 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 con el fin de llegar a árboles cada vez más pequeños (García & 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; aunque es menos importante, puesto que depende del número absoluto de instancias en el conjunto de datos de partida (Hall, Frank, & Witten, 2011).

Antes de construir un modelo se debe definir un procedimiento para probar la calidad del modelo y su validez. En consecuencia, 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 la herramienta *Yachay DTM*, utilizada en esta investigación, a saber:

- 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 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 en que ambos conjuntos pertenecen al universo de datos con el que se está trabajando, por lo que se elimina el riesgo que corre el anterior.
- Validación cruzada (*Cross validation*): permite reducir la dependencia del resultado del experimento en el modo en que se realiza la partición (Hernández, Ramirez y Ferri, 2004).

Para el caso que se reporta en este libro, 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 de forma aleatoria, el conjunto de entrenamiento, en n subconjuntos disjuntos, de similar tamaño, llamados pliegues (*folds*). 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 que se obtenga en cada una de las iteraciones. Una 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 suele utilizar 10 particiones (*10-fold cross validation*) (Hernández et al., 2004).

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 cada clase. La suma de los registros que se representan en cada fila $i, i = 1, 2, 3, \dots, ni, i = 1, 2, 3, \dots, n$ constituyen el número de instancias que realmente pertenecen a la clase ii . Similarmente, la sumatoria de los ejemplos o registros en cada columna $j, j = 1, 2, 3, \dots, nj, j = 1, 2, 3, \dots, n$ son las instancias que ha predicho el algoritmo al valor jj 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 ii de la fila ii y fueron clasificados incorrectamente en otra) (Fernández, 2009).

Teniendo en cuenta los parámetros de evaluación anteriores y los repositorios de datos descritos en la tabla 35, 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, se escogió el método de validación cruzada; específicamente, la validación cruzada con 10 pliegues, por los mejores resultados que se obtienen. En la figura 9, se muestra esta configuración con *Yachay DMT*.

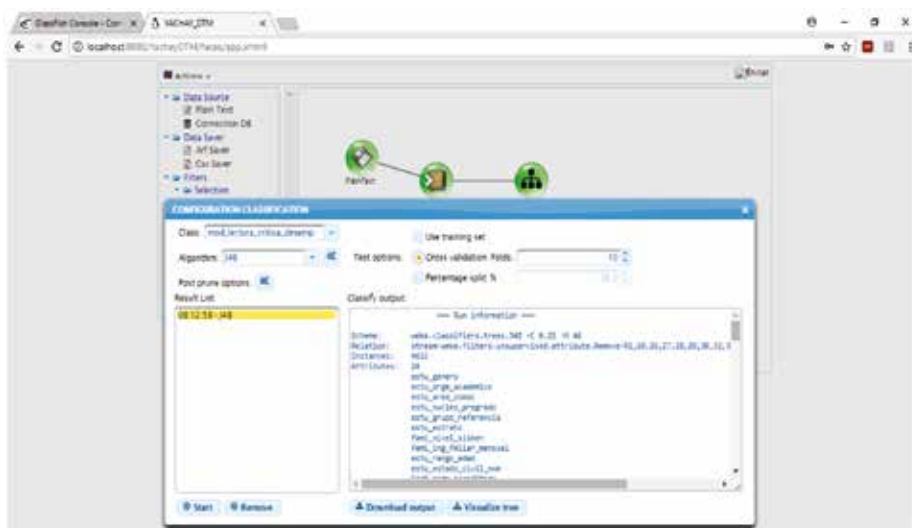


Figura 9. Configuración de la herramienta Yachay DMT para clasificación

Con el fin de obtener por competencia, diversos modelos de árboles y reglas de clasificación generalizadas, hasta reglas más detalladas, se configuraron cuatro valores para el factor de confianza C , con valores de 25%, 20%, 15% y 10%, manteniendo constante el factor M en 1%. Además, se aplicó un proceso de *postpoda* para dejar las ramas y las reglas más representativas; es decir, para dejar aquellas que sobrepasan un mínimo soporte del 1% y una confianza del 65%. Finalmente, para cada competencia se escoge el árbol con la mayor exactitud de clasificación, que sea fácil de interpretar.

En la construcción de los árboles de decisión para el descubrimiento de patrones de desempeño académico en las competencias genéricas de lectura crítica, composición escrita, razonamiento cuantitativo, inglés y competencias ciudadanas, de los estudiantes de programas profesionales de la Universidad de Nariño, que presentaron las pruebas Saber Pro 2012-2014, se utilizó el conjunto de datos de los repositorios descritos en la tabla 35.

4.4.1.1 Descubrimiento de patrones de desempeño en lectura crítica

Se utilizó el conjunto de datos *T3824A26LEC_udenar*. El mejor árbol fue el construido con los parámetros $M=46$ (1%) y $C=0.15$ para la prepoda, y con una confianza mayor o igual al 70% y soporte mayor o igual al 1% para la postpoda. En la figura 10 se presenta el árbol de clasificación obtenido, y en la figura 11, se muestra la precisión del árbol y la matriz de confusión para la competencia de lectura crítica.

4.4.1.2 Descubrimiento de patrones de desempeño en comunicación escrita

Se utilizó el conjunto de datos *T3807A26ESC_udenar*. El árbol construido con los parámetros $M=46$ (1%) y $C=0.15$ para la prepoda, y con una confianza mayor o igual al 70% y soporte mayor o igual al 1% para la postpoda, fue el mejor. En la figura 12 se presenta el árbol de clasificación obtenido, y en la figura 13, se muestra la precisión del árbol y la matriz de confusión para esta competencia.

```

=== Classifier model (full training set) ===
J48 pruned tree
-----
estu_rango_edad = [23-32)
| estu_facultad = FACULTAD DE CIENCIAS PECUARIAS
| | eco_condicion_vivienda = REGULAR: BAJO LA MEDIA (52.0/22.0)
| | eco_condicion_vivienda = BUENA: SOBRE LA MEDIA (107.0/43.0)
| | eco_condicion_vivienda = MALA: BAJO LA MEDIA (2.0)
| estu_facultad = FACULTAD DE ARTES: SOBRE LA MEDIA (318.0/139.0)
| estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES: SOBRE LA MEDIA (146.0/41.0)
| estu_facultad = FACULTAD DE CIENCIAS HUMANAS: SOBRE LA MEDIA (260.0/78.0)
| estu_facultad = FACULTAD DE DERECHO: SOBRE LA MEDIA (171.0/35.0)
| estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: SOBRE LA MEDIA (51.0/4.0)
| estu_facultad = FACULTAD DE EDUCACIÓN: BAJO LA MEDIA (203.0/98.0)
| estu_facultad = FACULTAD DE CIENCIAS ECONOMICAS Y ADMINISTRATIVAS
| | estu_estrato = BAJO
| | | eco_condicion_vivienda = REGULAR: BAJO LA MEDIA (119.0/47.0)
| | | eco_condicion_vivienda = BUENA
| | | | estu_genero = F: BAJO LA MEDIA (46.0/18.0)
| | | | estu_genero = M: SOBRE LA MEDIA (62.0/23.0)
| | | | estu_genero = SIN DATO: SOBRE LA MEDIA (0.0)
| | | eco_condicion_vivienda = MALA: BAJO LA MEDIA (10.0/1.0)
| | estu_estrato = MEDIO: SOBRE LA MEDIA (80.0/24.0)
| | estu_estrato = RURAL: BAJO LA MEDIA (1.0)
| | estu_estrato = ALTO: BAJO LA MEDIA (1.0)
| estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL: BAJO LA MEDIA (91.0/35.0)
| estu_facultad = FACULTAD DE INGENIERÍA
| | estu_sn_vive_flia = S: SOBRE LA MEDIA (117.0/35.0)
| | estu_sn_vive_flia = N
| | | inst_prgm_acreditado = NO ACREDITADO: SOBRE LA MEDIA (47.0/15.0)
| | | inst_prgm_acreditado = ACREDITADO: BAJO LA MEDIA (114.0/50.0)
| estu_facultad = FACULTAD DE CIENCIAS AGRICOLAS: SOBRE LA MEDIA (128.0/45.0)
estu_rango_edad = [32-41)
| fami_nivel_sisben = OTRO NIVEL: BAJO LA MEDIA (0.0)
| fami_nivel_sisben = NIVEL 2: SOBRE LA MEDIA (54.0/22.0)
| fami_nivel_sisben = NIVEL 1: BAJO LA MEDIA (123.0/46.0)
| fami_nivel_sisben = NIVEL 3: BAJO LA MEDIA (5.0/2.0)
| fami_nivel_sisben = SIN NIVEL: SOBRE LA MEDIA (46.0/13.0)
estu_rango_edad = [<23): SOBRE LA MEDIA (1380.0/369.0)
estu_rango_edad = [41-50): BAJO LA MEDIA (70.0/22.0)
estu_rango_edad = [>=50): BAJO LA MEDIA (20.0/3.0)
Number of Leaves : 30
Size of the tree : 39

```

Figura 10. Mejor Árbol para la competencia de Lectura Crítica

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 2494 65.2197%
Incorrectly Classified Instances 1330 34.7803%
Kappa statistic 0.1637
Mean absolute error 0.4339
Root mean squared error 0.4696
Relative absolute error 93.5099%
Root relative squared error 97.4899%
Total Number of Instances 3824
Ignored Class Unknown Instances 1
=== Confusion Matrix ===
a b <-- classified as
391 1008 | a = BAJO LA MEDIA
322 2103 | b = SOBRE LA MEDIA
    
```

Figura 11. Precisión y matriz de confusión del árbol de Lectura Crítica

4.4.1.3 Descubrimiento de patrones de desempeño en razonamiento cuantitativo

Se utilizó el conjunto de datos *T3824A26CUA_udenar*. El mejor árbol fue construido con los parámetros $M=46$ (1%) y $C=0.15$ para la prepoda, y con una confianza mayor o igual al 70% y soporte mayor o igual al 1% para la postpoda. En la figura 14 se presenta el árbol de clasificación obtenido, y en la figura 15, se muestra la precisión del árbol y la matriz de confusión para esta competencia.

4.4.1.4 Descubrimiento de patrones de desempeño en inglés

Se utilizó el conjunto de datos *T3825A26ING_udenar*. El mejor árbol se obtuvo con los parámetros $M=46$ (1%) y $C=0.05$ para la prepoda, y con una confianza mayor o igual al 70% y soporte mayor o igual al 1% para la postpoda. En la figura 16 se presenta el árbol de clasificación obtenido, y en la figura 17, se muestra la precisión del árbol y la matriz de confusión para esta competencia.

4.4.1.5 Descubrimiento de patrones de desempeño en competencias ciudadanas

Se utilizó el conjunto de datos *T3824A26CIU_udenar*. El mejor árbol se obtuvo con los parámetros $M=46$ (1%) y $C=0.05$ para la prepoda, y con una confianza mayor o igual al 70% y soporte mayor o igual al 1% para la postpoda. En la figura 18 se presenta

el árbol de clasificación obtenido, y en la figura 19, se muestra la precisión del árbol y la matriz de confusión para esta competencia.

```
=== Classifier model (full training set) ===
J48 pruned tree
-----
estu_genero = F: SOBRE LA MEDIA (1762.0/594.0)
estu_genero = M
| estu_estado_civil_nom = SOLTERO(A)
| | estu_facultad = FACULTAD DE CIENCIAS PECUARIAS: SOBRE LA MEDIA
(119.0/58.0)
| | estu_facultad = FACULTAD DE ARTES: BAJO LA MEDIA (318.0/129.0)
| | estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES
| | | area_grupo_referencia = CIENCIAS NATURALES Y TÉCNICAS: SOBRE LA
MEDIA (49.0/16.0)
| | | area_grupo_referencia = CIENCIAS HUMANAS: BAJO LA MEDIA
(80.0/37.0)
| | | area_grupo_referencia = CIENCIAS DE LA SALUD: SOBRE LA MEDIA (0.0)
| | | area_grupo_referencia = CIENCIAS SOCIALES: SOBRE LA MEDIA (0.0)
| | estu_facultad = FACULTAD DE CIENCIAS HUMANAS: SOBRE LA MEDIA
(159.0/58.0)
| | estu_facultad = FACULTAD DE DERECHO: SOBRE LA MEDIA (157.0/53.0)
| | estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: SOBRE LA MEDIA
(38.0/13.0)
| | estu_facultad = FACULTAD DE EDUCACIÓN: SOBRE LA MEDIA (100.0/45.0)
| | estu_facultad = FACULTAD DE CIENCIAS ECONOMICAS Y ADMINISTRA-
TIVAS: SOBRE LA MEDIA (270.0/93.0)
| | estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL: SOBRE
LA MEDIA (81.0/40.0)
| | estu_facultad = FACULTAD DE INGENIERÍA: BAJO LA MEDIA (376.0/173.0)
| | estu_facultad = FACULTAD DE CIENCIAS AGRICOLAS
| | | estu_sn_vive_flia = S: SOBRE LA MEDIA (55.0/23.0)
| | | estu_sn_vive_flia = N: BAJO LA MEDIA (69.0/24.0)
| estu_estado_civil_nom = UNION LIBRE: BAJO LA MEDIA (77.0/29.0)
| estu_estado_civil_nom = CASADO(A): SOBRE LA MEDIA (81.0/34.0)
| estu_estado_civil_nom = SEPARADO(A)/DIVORCIADO(A): SOBRE LA MEDIA
(10.0/1.0)
| estu_estado_civil_nom = VIUDO(A): BAJO LA MEDIA (1.0)
estu_genero = SIN DATO: BAJO LA MEDIA (5.0/2.0)

Number of Leaves : 21
Size of the tree : 26
```

Figura 12. Mejor Árbol para la competencia de Comunicación Escrita

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2337      61.3869%
Incorrectly Classified Instances    1470      38.6131%
Kappa statistic                    0.1492
Mean absolute error                0.465
Root mean squared error            0.4841
Relative absolute error            96.1195%
Root relative squared error        98.4261%
Total Number of Instances          3807
Ignored Class Unknown Instances    18

=== Confusion Matrix ===

  a  b  <-- classified as
1827 420 | a = SOBRE LA MEDIA
1050 510 | b = BAJO LA MEDIA
    
```

Figura 13. Precisión y matriz de confusión del árbol de Comunicación Escrita

```

=== Classifier model (full training set) ===

J48 pruned tree
-----
area_grupo_referencia = CIENCIAS NATURALES Y TÉCNICAS: SOBRE LA MEDIA (1244.0/353.0)
area_grupo_referencia = CIENCIAS HUMANAS
|
|  estu_facultad = FACULTAD DE CIENCIAS PECUARIAS: BAJO LA MEDIA (0.0)
|  estu_facultad = FACULTAD DE ARTES: BAJO LA MEDIA (427.0/145.0)
|  estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES: SOBRE LA MEDIA (145.0/44.0)
|  estu_facultad = FACULTAD DE CIENCIAS HUMANAS: BAJO LA MEDIA (362.0/134.0)
|  estu_facultad = FACULTAD DE DERECHO: SOBRE LA MEDIA (338.0/135.0)
|  estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: BAJO LA MEDIA (0.0)
|  estu_facultad = FACULTAD DE EDUCACIÓN: BAJO LA MEDIA (390.0/68.0)
|  estu_facultad = FACULTAD DE CIENCIAS ECONOMICAS Y ADMINISTRATIVAS: BAJO LA MEDIA (0.0)
|  estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL: BAJO LA MEDIA (0.0)
|  estu_facultad = FACULTAD DE INGENIERÍA: BAJO LA MEDIA (0.0)
|  estu_facultad = FACULTAD DE CIENCIAS AGRICOLAS: BAJO LA MEDIA (0.0)
area_grupo_referencia = CIENCIAS DE LA SALUD: SOBRE LA MEDIA (75.0/8.0)
area_grupo_referencia = CIENCIAS SOCIALES
|
|  estu_sn_vive_flia = S
|  |
|  |  estu_facultad = FACULTAD DE CIENCIAS PECUARIAS: SOBRE LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE ARTES: SOBRE LA MEDIA (51.0/9.0)
|  |  estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES: SOBRE LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE CIENCIAS HUMANAS: BAJO LA MEDIA (93.0/38.0)
|  |  estu_facultad = FACULTAD DE DERECHO: SOBRE LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: SOBRE LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE EDUCACIÓN: SOBRE LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE CIENCIAS ECONOMICAS Y ADMINISTRATIVAS: SOBRE LA MEDIA
|  |  (266.0/66.0)
|  |  |
|  |  |  estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL: SOBRE LA MEDIA (0.0)
|  |  |  estu_facultad = FACULTAD DE INGENIERÍA: SOBRE LA MEDIA (0.0)
|  |  |  estu_facultad = FACULTAD DE CIENCIAS AGRICOLAS: SOBRE LA MEDIA (0.0)
|  |  estu_sn_vive_flia = N
|  |  |
|  |  |  estu_genero = F: BAJO LA MEDIA (234.0/78.0)
|  |  |  estu_genero = M: SOBRE LA MEDIA (199.0/89.0)
|  |  |  estu_genero = SIN DATO: BAJO LA MEDIA (0.0)
|
|
Number of Leaves : 27
Size of the tree : 32
    
```

Figura 14. Mejor Árbol para la competencia de Razonamiento Cuantitativo

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2648      69.2469%
Incorrectly Classified Instances    1176      30.7531%
Kappa statistic                    0.3735
Mean absolute error                 0.4132
Root mean squared error             0.4564
Relative absolute error             83.2549%
Root relative squared error         91.6171%
Total Number of Instances          3824
Ignored Class Unknown Instances     1

=== Confusion Matrix ===

  a  b  <-- classified as
1609 468 | a = SOBRE LA MEDIA
 708 1039 | b = BAJO LA MEDIA
    
```

Figura 15. Precisión y matriz de confusión del árbol de Razonamiento Cuantitativo

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

estu_sn_pers_cargo = S: BAJO LA MEDIA (891.0/248.0)
estu_sn_pers_cargo = N
|  estu_facultad = FACULTAD DE CIENCIAS PECUARIAS
|  |  estu_estrato = BAJO: BAJO LA MEDIA (139.0/45.0)
|  |  |  estu_estrato = MEDIO: SOBRE LA MEDIA (53.0/23.0)
|  |  |  |  estu_estrato = RURAL: BAJO LA MEDIA (0.0)
|  |  |  |  |  estu_estrato = ALTO: BAJO LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE ARTES
|  |  |  estu_estrato = BAJO: BAJO LA MEDIA (236.0/76.0)
|  |  |  |  estu_estrato = MEDIO
|  |  |  |  |  eco_condicion_electrodomesticos = REGULAR: BAJO LA MEDIA (50.0/21.0)
|  |  |  |  |  |  eco_condicion_electrodomesticos = MALA: BAJO LA MEDIA (11.0/3.0)
|  |  |  |  |  |  |  eco_condicion_electrodomesticos = BUENA: SOBRE LA MEDIA (53.0/19.0)
|  |  |  |  |  |  |  |  estu_estrato = RURAL: SOBRE LA MEDIA (3.0)
|  |  |  |  |  |  |  |  |  estu_estrato = ALTO: SOBRE LA MEDIA (4.0/1.0)
|  |  estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES
|  |  |  area_grupo_referencia = CIENCIAS NATURALES Y TÉCNICAS: SOBRE LA MEDIA
(95.0/20.0)
|  |  |  |  area_grupo_referencia = CIENCIAS HUMANAS: BAJO LA MEDIA (114.0/51.0)
|  |  |  |  |  area_grupo_referencia = CIENCIAS DE LA SALUD: SOBRE LA MEDIA (0.0)
|  |  |  |  |  |  area_grupo_referencia = CIENCIAS SOCIALES: SOBRE LA MEDIA (0.0)
|  |  estu_facultad = FACULTAD DE CIENCIAS HUMANAS
|  |  |  area_grupo_referencia = CIENCIAS NATURALES Y TÉCNICAS: SOBRE LA MEDIA (0.0)
|  |  |  |  area_grupo_referencia = CIENCIAS HUMANAS: SOBRE LA MEDIA (284.0/100.0)
|  |  |  |  |  area_grupo_referencia = CIENCIAS DE LA SALUD: SOBRE LA MEDIA (0.0)
|  |  |  |  |  |  area_grupo_referencia = CIENCIAS SOCIALES: BAJO LA MEDIA (88.0/27.0)
|  |  estu_facultad = FACULTAD DE DERECHO
|  |  |  estu_sn_vive_flia = S
|  |  |  |  estu_rango_edad = [23-32]: BAJO LA MEDIA (94.0/42.0)
|  |  |  |  |  estu_rango_edad = [32-41]: BAJO LA MEDIA (6.0/3.0)
    
```

```

| | | estu_rango_edad = [<23]: SOBRE LA MEDIA (97.0/28.0)
| | | estu_rango_edad = [41-50]: BAJO LA MEDIA (1.0)
| | | estu_rango_edad = [>=50]: SOBRE LA MEDIA (0.0)
| | estu_sn_vive_flia = N: BAJO LA MEDIA (81.0/27.0)
| estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: SOBRE LA ME-
DIA (70.0/13.0)
| estu_facultad = FACULTAD DE EDUCACIÓN: BAJO LA MEDIA (228.0/47.0)
| estu_facultad = FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRA-
TIVAS
| | estu_sn_vive_flia = S
| | | inst_prgm_acreditado = NO ACREDITADO: SOBRE LA MEDIA
(134.0/39.0)
| | | inst_prgm_acreditado = ACREDITADO: BAJO LA MEDIA (73.0/27.0)
| | estu_sn_vive_flia = N: BAJO LA MEDIA (268.0/72.0)
| estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL: BAJO LA
MEDIA (133.0/46.0)
| estu_facultad = FACULTAD DE INGENIERÍA: SOBRE LA MEDIA
(417.0/120.0)
| estu_facultad = FACULTAD DE CIENCIAS AGRÍCOLAS: BAJO LA MEDIA
(202.0/75.0)

Number of Leaves :      33

Size of the tree : 44

```

Figura 16. Mejor Árbol para la competencia Inglés

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2567      67.1111%
Incorrectly Classified Instances    1258      32.8889%
Kappa statistic                     0.3142
Mean absolute error                 0.4284
Root mean squared error             0.4661
Relative absolute error             87.2263%
Root relative squared error         94.0698%
Total Number of Instances          3825

=== Confusion Matrix ===

  a  b  <-- classified as
1704 464 |  a = BAJO LA MEDIA
 794 863 |  b = SOBRE LA MEDIA

```

Figura 17. Precisión y matriz de confusión del árbol de Inglés

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

estu_facultad = FACULTAD DE CIENCIAS PECUARIAS: SOBRE LA MEDIA
(239.0/104.0)
estu_facultad = FACULTAD DE ARTES
|  estu_genero = F: SOBRE LA MEDIA (142.0/62.0)
|  estu_genero = M
|  |  area_grupo_referencia = CIENCIAS NATURALES Y TÉCNICAS: BAJO LA MEDIA
(0.0)
|  |  area_grupo_referencia = CIENCIAS HUMANAS: BAJO LA MEDIA (311.0/108.0)
|  |  area_grupo_referencia = CIENCIAS DE LA SALUD: BAJO LA MEDIA (0.0)
|  |  area_grupo_referencia = CIENCIAS SOCIALES: SOBRE LA MEDIA (68.0/31.0)
|  estu_genero = SIN DATO: BAJO LA MEDIA (0.0)
estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES
|  inst_prgm_acreditado = NO ACREDITADO: BAJO LA MEDIA (158.0/70.0)
|  inst_prgm_acreditado = ACREDITADO: SOBRE LA MEDIA (98.0/23.0)
estu_facultad = FACULTAD DE CIENCIAS HUMANAS: SOBRE LA MEDIA
(490.0/158.0)
estu_facultad = FACULTAD DE DERECHO: SOBRE LA MEDIA (338.0/38.0)
estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: SOBRE LA MEDIA
(75.0/14.0)
estu_facultad = FACULTAD DE EDUCACIÓN: BAJO LA MEDIA (390.0/150.0)
estu_facultad = FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS
|  estu_sn_vive_flia = S: SOBRE LA MEDIA (266.0/88.0)
|  estu_sn_vive_flia = N
|  |  estu_estrato = BAJO: BAJO LA MEDIA (305.0/128.0)
|  |  estu_estrato = MEDIO: SOBRE LA MEDIA (47.0/19.0)
|  |  estu_estrato = RURAL: SOBRE LA MEDIA (1.0)
|  |  estu_estrato = ALTO: SOBRE LA MEDIA (2.0)
estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL: BAJO LA MEDIA
(161.0/70.0)
estu_facultad = FACULTAD DE INGENIERÍA
|  estu_sn_vive_flia = S: SOBRE LA MEDIA (237.0/81.0)
|  estu_sn_vive_flia = N
|  |  inst_prgm_acreditado = NO ACREDITADO: SOBRE LA MEDIA (78.0/28.0)
|  |  inst_prgm_acreditado = ACREDITADO: BAJO LA MEDIA (180.0/79.0)
estu_facultad = FACULTAD DE CIENCIAS AGRÍCOLAS: SOBRE LA MEDIA
(238.0/112.0)

Number of Leaves :      23

Size of the tree :    31
    
```

Figura 18. Mejor Árbol para Competencias Ciudadanas

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2368      61.9247%
Incorrectly Classified Instances    1456      38.0753%
Kappa statistic                    0.2196
Mean absolute error                 0.4456
Root mean squared error             0.4751
Relative absolute error             90.719 %
Root relative squared error         95.8792%
Total Number of Instances          3824
Ignored Class Unknown Instances     1

=== Confusion Matrix ===

  a  b  <-- classified as
883 775 |  a = BAJO LA MEDIA
    
```

Figura 19. Precisión y matriz de confusión del árbol de Competencias Ciudadanas

4.4.1.6 Descubrimiento de patrones de desempeño general en todas las competencias

Se utilizó el conjunto de datos *T3825A26GEN_udenar*. El árbol construido con los parámetros $M=46$ (1%) y $C=0.25$ para la pre-poda, y con una confianza mayor o igual al 70% y soporte mayor o igual al 1% para la postpoda, fue el mejor. En la figura 20 se presenta el árbol de clasificación obtenido, y en la figura 21, se muestra la precisión del árbol y la matriz de confusión para esta competencia.

```

=== Classifier model (full training set) ===

J48 pruned tree
-----
estu_facultad = FACULTAD DE CIENCIAS PECUARIAS: SOBRE LA MEDIA (239.0/90.0)
estu_facultad = FACULTAD DE ARTES
| fami_nivel_sisben = OTRO NIVEL: BAJO LA MEDIA (7.0/3.0)
| fami_nivel_sisben = NIVEL 2: BAJO LA MEDIA (120.0/50.0)
| fami_nivel_sisben = NIVEL 1: BAJO LA MEDIA (161.0/59.0)
| fami_nivel_sisben = NIVEL 3: SOBRE LA MEDIA (11.0/4.0)
| fami_nivel_sisben = SIN NIVEL: SOBRE LA MEDIA (222.0/89.0)
estu_facultad = FACULTAD DE CIENCIAS EXACTAS Y NATURALES: SOBRE LA MEDIA
(256.0/64.0)
estu_facultad = FACULTAD DE CIENCIAS HUMANAS: SOBRE LA MEDIA (490.0/124.0)
estu_facultad = FACULTAD DE DERECHO: SOBRE LA MEDIA (338.0/45.0)
estu_facultad = FACULTAD DE CIENCIAS DE LA SALUD: SOBRE LA MEDIA (75.0/8.0)
estu_facultad = FACULTAD DE EDUCACIÓN
| estu_rango_edad = [23-32]: BAJO LA MEDIA (203.0/70.0)
| estu_rango_edad = [32-41]: BAJO LA MEDIA (45.0/13.0)
| estu_rango_edad = [<23]: SOBRE LA MEDIA (110.0/49.0)
| estu_rango_edad = [41-50]: BAJO LA MEDIA (27.0/1.0)
| estu_rango_edad = [>=50]: BAJO LA MEDIA (5.0)
estu_facultad = FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS
| estu_sn_vive_flia = S: SOBRE LA MEDIA (267.0/66.0)
| estu_sn_vive_flia = N
| | fami_ocup_madre = OTRA: BAJO LA MEDIA (14.0/6.0)
| | fami_ocup_madre = HOGAR
| | | eco_condicion_vivienda = REGULAR: BAJO LA MEDIA (107.0/52.0)
| | | eco_condicion_vivienda = BUENA: SOBRE LA MEDIA (71.0/26.0)
| | | eco_condicion_vivienda = MALA: BAJO LA MEDIA (11.0/3.0)
| | fami_ocup_madre = EMPLEADO: BAJO LA MEDIA (55.0/19.0)
| | fami_ocup_madre = INDEPENDIENTE: BAJO LA MEDIA (61.0/22.0)
| | fami_ocup_madre = EMPRESARIO: BAJO LA MEDIA (9.0/2.0)
| | fami_ocup_madre = OBRERO: SOBRE LA MEDIA (17.0/5.0)
| | fami_ocup_madre = PENSIONADO: BAJO LA MEDIA (4.0/1.0)
| | fami_ocup_madre = PROFESIONAL: SOBRE LA MEDIA (6.0/2.0)
| | fami_ocup_madre = NO SABE: BAJO LA MEDIA (0.0)
estu_facultad = FACULTAD DE INGENIERÍA AGROINDUSTRIAL
| eco_condicion_vivienda = REGULAR: BAJO LA MEDIA (65.0/25.0)
| eco_condicion_vivienda = BUENA: SOBRE LA MEDIA (90.0/31.0)
| eco_condicion_vivienda = MALA: BAJO LA MEDIA (6.0/2.0)
estu_facultad = FACULTAD DE INGENIERÍA: SOBRE LA MEDIA (495.0/125.0)
estu_facultad = FACULTAD DE CIENCIAS AGRÍCOLAS: SOBRE LA MEDIA (238.0/82.0)

Number of Leaves: 32
Size of the tree: 39
    
```

Figura 20. Mejor Árbol para el desempeño general en todas las competencias

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      2606      68.1307%
Incorrectly Classified Instances    1219      31.8693%
Kappa statistic                    0.2561
Mean absolute error                0.4123
Root mean squared error            0.458
Relative absolute error             89.3194%
Root relative squared error         95.3427%
Total Number of Instances          3825

=== Confusion Matrix ===
  a  b  <-- classified as
544 838 |  a = BAJO LA MEDIA
381 2062 |  b = SOBRE LA MEDIA
    
```

Figura 21. Precisión y matriz de confusión del árbol general de Competencias

4.5 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 para el usuario. La evaluación e interpretación de los patrones descubiertos se describe en el capítulo V, que corresponde a interpretación de resultados.

4.6 Implementación

En esta fase, el conocimiento descubierto se podrá integrar a los procesos de toma de decisiones de la Universidad de Nariño, con el fin de mejorar la calidad de la educación superior que se imparte en la Institución. Una vez la Universidad de Nariño intervenga los factores asociados al desempeño académico de los estudiantes de programas profesionales en las Pruebas SaberPro, será posible analizar los nuevos resultados y determinar sus efectos.

INTERPRETACIÓN DE RESULTADOS

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

5.1 Evaluación e interpretación de resultados para la competencia lectura crítica

Analizando los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T3824A26LEC (ver figura 10), en el cual se almacenan los datos válidos sobre los factores socioeconómicos, académicos e institucionales de 3824 estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro 2012-2014, en la competencia genérica de lectura crítica, donde se escogió el atributo *mod_lectura_critica_desemp* como clase, se puede observar que este clasifica correctamente a 2494 instancias, que corresponde a un porcentaje de precisión del 65,2% y aparecen 1330 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 34.8% (ver figura 11).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_lectura_critica_desemp* del repositorio T3824A26LEC que es de 2425 registros para el valor “sobre la media” y 1399 registros para el valor “bajo la media”, y evaluando el modelo con la matriz de confusión de la figura 11, este clasifica correctamente a 2103 casos de estudiantes cuyos resultados en lectura crítica están sobre la media, y a 391 casos que están bajo la media. Adicionalmente, el modelo clasifica incorrectamente a 322 casos de estudiantes cuyos resultados en lectura crítica están sobre la media y a 1008 casos que están bajo la media. Así pues, el modelo clasifica correctamente al 86.7% de los estudiantes que están sobre la media en lectura crítica y al 27.9% de las estudiantes que están bajo la media en esta competencia.

De acuerdo a las reglas de clasificación obtenidas con este modelo (ver figura 10), los patrones más representativos descubiertos, teniendo en cuenta que superen el soporte mínimo del 1% y una confianza mínima del 65%, son 10 y corresponden a las siguientes reglas:

Regla 1. Si el rango de edad de los estudiantes es menor que 23 años, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 36.1% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 73.3% de los 1380 estudiantes que se clasifican así, están correctamente clasificados; y el 41.7% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 2. Si el rango de edad de los estudiantes está entre 23 y 31 años y son de la facultad de Ciencias Humanas, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 6.8% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 70% de los 260 estudiantes que se clasifican así, están correctamente clasificados; y el 7.5% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 3. Si el rango de edad de los estudiantes está entre 23 y 31 años y son de la facultad de Derecho, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 4.5% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 79.5% de los 171 estudiantes que se clasifican así, son correctamente clasificados; y el 5.6% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 4. Si el rango de edad de los estudiantes está entre 23 y 31 años y son de la facultad de Ciencias Exactas y Naturales, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 3.8% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 71.9% de los 146 estudiantes que se clasifican así, son correctamente clasificados; y el 4.3% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 5. Si el rango de edad de los estudiantes está entre 23 y 31 años, son de la facultad de Ingeniería y viven con la familia, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 3.1% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 70.1% de los 117 estudiantes que se clasifican así, son correctamente clasificados; y el 3.4% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 6. Si el rango de edad de los estudiantes está entre 23 y 31 años, son de la facultad de Ciencias Administrativas y Económicas y de estrato medio (3 y 4), entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 2.1% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 70% de los 80 estudiantes que se clasifican así, son correctamente clasificados; y el 2.3% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 7. Si el rango de edad de los estudiantes está entre 23 y 31 años y son de la facultad de Ciencias de la Salud, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 1.3% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 92.2% de los 51 estudiantes que se clasifican así, son correctamente clasificados; y el 1.9% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 8. Si el rango de edad de los estudiantes está entre 41 y 49 años, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar bajo la media. El 1.8% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 68.6% de los 70 estudiantes que se clasifican así, son correctamente clasificados; y el 3.4% de los 1399 estudiantes que están bajo la media, cumplen este patrón.

Regla 9. Si el rango de edad de los estudiantes está entre 32 y 40 años y no tienen nivel sisben, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 1.2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 71.7% de los 46 estudiantes que se clasifican así, son correctamente clasificados; y el 1.4% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

Regla 10. Si el rango de edad de los estudiantes está entre 23 y 31 años, son de la facultad de Ingeniería, no viven con la familia y el programa al cual están matriculados no es acreditado, entonces, el desempeño de estos estudiantes en lectura crítica tiene mayor probabilidad de estar sobre la media. El 1.2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 68.1% de los 47 estudiantes que se clasifican así, son correctamente clasificados; y el 1.3% de los 2425 estudiantes que están sobre la media, cumplen este patrón.

5.2 Evaluación e interpretación de resultados para la competencia comunicación escrita

Analizando los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T3807A26ESC_udenar (ver figura 12), en el cual se almacenan los datos válidos sobre los factores socioeconómicos, académicos e institucionales de 3807 estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro 2012-2014 en la competencia genérica de comunicación escrita, donde se escogió el atributo *mod_comunica_escrita_desemp* como clase, se puede observar que este clasifica correctamente a 2337 instancias, que corresponde a un porcentaje de precisión del 61,4% y 1470 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 38.6% (ver figura 13).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_comunica_escrita_desemp* del repositorio T3807A26ESC_udenar que es de 2247 registros para el valor “sobre la media” y 1560 registros para el valor “bajo la media” y evaluando el modelo con la matriz de confusión de la figura 13, este clasifica correctamente a 1827 casos de estudiantes cuyos resultados en comunicación escrita están sobre la media y a 510 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 420 casos de estudiantes cuyos resultados en comunicación escrita están sobre la media y 1050 casos que están bajo la media. Por lo tanto, el modelo clasifica correctamente al 81.3% de los estudiantes que están sobre la media en comunicación escrita y al 32.7% de las estudiantes que están bajo la media en esta competencia.

De acuerdo a las reglas de clasificación obtenidas con este modelo (ver figura 12), los patrones más representativos descubiertos, teniendo en cuenta que superen un soporte mínimo del 1% y una confianza mínima del 65%, son cinco (5) y corresponden a las siguientes reglas:

Regla 1. Si los estudiantes son de género femenino entonces el desempeño de estos estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 46.1% de los 3807 estudiantes evaluados en esta competencia en la prueba Saber

Pro 2012-2014 se clasifican de esta manera. El 66.3% de los 1762 estudiantes que se clasifican así, son correctamente clasificados y el 52% de los 2247 estudiantes que están sobre la media cumplen este patrón.

Regla 2. Si los estudiantes son de género masculino, solteros y de la facultad de Ciencias Económicas y Administrativas entonces el desempeño de estos estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 7.1% de los 3807 estudiantes evaluados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 65.6% de los 270 estudiantes que se clasifican así, son correctamente clasificados y el 7.9% de los 2247 estudiantes que están sobre la media cumplen este patrón.

Regla 3. Si los estudiantes son de género masculino, solteros y de la facultad de Derecho entonces el desempeño de estos estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 4.1% de los 3807 estudiantes evaluados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 66.2% de los 157 estudiantes que se clasifican así, son correctamente clasificados y el 4.6% de los 2247 estudiantes que están sobre la media cumplen este patrón.

Regla 4. Si los estudiantes son de género masculino, solteros, de la facultad de Ciencias Exactas y Técnicas y el área del grupo de referencia pertenece a las Ciencias Naturales y Técnicas entonces el desempeño de estos estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 1.3% de los 3807 estudiantes evaluados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 67.3% de los 49 estudiantes que se clasifican así, son correctamente clasificados y el 1.5% de los 2247 estudiantes que están sobre la media cumplen este patrón.

Regla 5. Si los estudiantes son de género masculino, solteros y de la facultad de Ciencias de la Salud entonces el desempeño de estos estudiantes en comunicación escrita tiene mayor probabilidad de estar sobre la media. El 1% de los 3807 estudiantes evaluados en esta competencia en la prueba Saber Pro 2012-2014

se clasifican de esta manera. El 65.8% de los 38 estudiantes que se clasifican así, son correctamente clasificados y el 1.1% de los 2247 estudiantes que están sobre la media cumplen este patrón.

Regla 6. Si los estudiantes son de género masculino, solteros, de la facultad de Ciencias Agrícolas y no viven con su familia entonces el desempeño de estos estudiantes en comunicación escrita tiene mayor probabilidad de estar bajo la media. El 1.8% de los 3807 estudiantes evaluados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 65.62% de los 69 estudiantes que se clasifican así, son correctamente clasificados y el 2.9% de los 1560 estudiantes que están bajo la media cumplen este patrón.

5.3 Evaluación e interpretación de resultados para la competencia razonamiento cuantitativo

Analizando los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T3824A26CUA_udenar (ver figura 14), en el cual se almacenan los datos válidos sobre los factores socioeconómicos, académicos e institucionales de 3824 estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro entre los años 2012 y 2014 en la competencia genérica de razonamiento cuantitativo, donde se escogió el atributo *mod_razona_cuantitativo_desemp* como clase, se puede observar que este clasifica correctamente a 2648 instancias, que corresponde a un porcentaje de precisión del 69,2% y 1176 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 38,8% (ver figura 15).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_razona_cuantitativo_desemp* del repositorio T3824A-26CUA_udenar que es de 2077 registros para el valor “sobre la media” y 1747 registros para el valor “bajo la media” y evaluando el modelo con la matriz de confusión de la figura 15, este clasifica correctamente a 1609 casos de estudiantes cuyos resultados en razonamiento cuantitativo están sobre la media y a 1039 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 468 casos de estudiantes cuyos resultados en razonamiento cuan-

titativo están sobre la media y 708 casos que están bajo la media. Esto significa que el modelo clasifica correctamente al 77.5% de los estudiantes que están sobre la media en razonamiento cuantitativo y al 59.5% de las estudiantes que están bajo la media en esta competencia.

De acuerdo a las reglas de clasificación obtenidas con este modelo (ver figura 14), los patrones más representativos descubiertos, teniendo en cuenta que superen un soporte mínimo del 1% y una confianza mínima del 65%, son seis (6) y corresponden a las siguientes reglas:

Regla 1. Si los estudiantes pertenecen a un programa cuya área de referencia es Ciencias Naturales y Técnicas, entonces el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar sobre la media. El 32.5% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 71.6% de los 1244 estudiantes que se clasifican así, son correctamente clasificados y el 42.9% de los 2077 estudiantes que están sobre la media cumplen este patrón.

Regla 2. Si los estudiantes pertenecen a un programa cuya área de referencia es Ciencias Humanas y son de la facultad de Artes entonces el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 11.2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 66% de los 427 estudiantes que se clasifican así, son correctamente clasificados y el 16.1% de los 1747 estudiantes que están bajo la media cumplen este patrón.

Regla 3. Si los estudiantes pertenecen a un programa cuya área de referencia es Ciencias Humanas y son de la facultad de Educación entonces el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 10.2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 82.6% de los 390 estudiantes que se clasifican así, son correcta-

mente clasificados y el 18.4% de los 1747 estudiantes que están bajo la media cumplen este patrón.

Regla 4. Si los estudiantes pertenecen a un programa cuya área de referencia es Ciencias Sociales, viven con su familia y son de la facultad de Ciencias Económicas y Administrativas entonces el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar sobre la media. El 7% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 75.2% de los 266 estudiantes que se clasifican así, son correctamente clasificados y el 9.6% de los 2077 estudiantes que están sobre la media cumplen este patrón.

Regla 5. Si los estudiantes pertenecen a un programa cuya área de referencia es Ciencias Sociales, no viven con su familia y son de sexo femenino, el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar bajo la media. El 6.1% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 66.7% de los 234 estudiantes que se clasifican así, son correctamente clasificados y el 8.9% de los 1747 estudiantes que están bajo la media cumplen este patrón.

Regla 6. Si los estudiantes pertenecen a un programa cuya área de referencia es Ciencias de la Salud entonces el desempeño de los estudiantes en razonamiento cuantitativo tiene mayor probabilidad de estar sobre la media. El 2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 89.3% de los 75 estudiantes que se clasifican así, son correctamente clasificados y el 3.2% de los 2077 estudiantes que están sobre la media cumplen este patrón.

5.4 Evaluación e interpretación de resultados para la competencia inglés

Analizando los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T3825A26ING_udenar (ver figura 16), en el cual se almacenan los datos válidos sobre los factores socioeconómicos, académicos e institucionales de 3825

estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro entre los años 2012 y 2014 en la competencia genérica de inglés, donde se escogió el atributo *mod_ingles_desemp* como clase, se puede observar que este clasifica correctamente a 2567 instancias, que corresponde a un porcentaje de precisión del 67.1% y 1258 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 32.9% (ver figura 17).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_ingles_desemp* del repositorio T3825A26ING_udenar que es de 1657 registros para el valor “sobre la media” y 2168 registros para el valor “bajo la media” y evaluando el modelo con la matriz de confusión de la figura 17, este clasifica correctamente a 863 casos de estudiantes cuyos resultados en inglés están sobre la media y a 1704 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 794 casos de estudiantes cuyos resultados en inglés están sobre la media y 464 casos que están bajo la media. En consecuencia, el modelo clasifica correctamente al 52.1% de los estudiantes que están sobre la media en inglés y al 78.6% de las estudiantes que están bajo la media en esta competencia.

De acuerdo a las reglas de clasificación obtenidas con este modelo (ver figura 16), los patrones más representativos descubiertos, teniendo en cuenta que superen un soporte mínimo del 1% y una confianza mínima del 65%, son trece (13) y corresponden a las siguientes reglas:

Regla 1. Si los estudiantes tienen personas a cargo entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 23.3% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 72.2% de los 891 estudiantes que se clasifican así, son correctamente clasificados y el 29.7% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 2. Si los estudiantes no tienen personas a cargo y son de la facultad de Ingeniería entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 10.9% de los 3825 estudiantes analizados en esta competencia

en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 71.2% de los 417 estudiantes que se clasifican así, son correctamente clasificados y el 17.9% de los 1647 estudiantes que están sobre la media cumplen este patrón.

Regla 3. Si los estudiantes no tienen personas a cargo, son de la facultad de Ciencias Económicas y Administrativas y no viven con su familia entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 7% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 73.1% de los 268 estudiantes que se clasifican así, son correctamente clasificados y el 9% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 4. Si los estudiantes no tienen personas a cargo, son de la facultad de Artes y son de estrato socioeconómico bajo (1,2) entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 6.2% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 67.8% de los 236 estudiantes que se clasifican así, son correctamente clasificados y el 7.4% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 5. Si los estudiantes no tienen personas a cargo, son de la facultad de Educación entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 6% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 79.4% de los 228 estudiantes que se clasifican así, son correctamente clasificados y el 8.3% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 6. Si los estudiantes no tienen personas a cargo, son de la facultad de Ciencias Agrícolas entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 5.3% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 62.9% de los 202 estudiantes que se clasifican así, son

correctamente clasificados y el 5.9% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 7. Si los estudiantes no tienen personas a cargo, son de la facultad de Ciencias Pecuarias y de estrato socioeconómico bajo (1,2) entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 3.6% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 67.6% de los 139 estudiantes que se clasifican así, son correctamente clasificados y el 4.3% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 8. Si los estudiantes no tienen personas a cargo y son de la facultad de Ciencias Económicas y Administrativas, viven con su familia y el programa al cual están matriculados no está acreditado entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 3.5% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 70.9% de los 134 estudiantes que se clasifican así, son correctamente clasificados y el 5.7% de los 1647 estudiantes que están sobre la media cumplen este patrón.

Regla 9. Si los estudiantes no tienen personas a cargo y son de la facultad de Ingeniería Agroindustrial entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 3.5% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 65.4% de los 133 estudiantes que se clasifican así, son correctamente clasificados y el 4% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 10. Si los estudiantes no tienen personas a cargo, son de la facultad de Ciencias Exactas y Naturales y son de un programa que pertenece al área del grupo de referencia de Ciencias Naturales y Técnicas entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 2.5% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 78.9% de los

95 estudiantes que se clasifican así, son correctamente clasificados y el 4.5% de los 1647 estudiantes que están sobre la media cumplen este patrón.

Regla 11. Si los estudiantes no tienen personas a cargo, son de la facultad de Derecho, viven con su familia y son menores que 23 años entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 2.5% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 71.1% de los 97 estudiantes que se clasifican así, son correctamente clasificados y el 4.2% de los 1647 estudiantes que están sobre la media cumplen este patrón.

Regla 12. Si los estudiantes no tienen personas a cargo, son de la facultad de Derecho y no viven con su familia entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar bajo la media. El 2.1% de los 3825 estudiantes analizados en esta competencia en la prueba Saber Pro 2012-2014 se clasifican de esta manera. El 66.7% de los 81 estudiantes que se clasifican así, son correctamente clasificados y el 2.5% de los 2168 estudiantes que están bajo la media cumplen este patrón.

Regla 13. Si los estudiantes no tienen personas a cargo, son de la facultad de Ciencias de la Salud entonces el desempeño de los estudiantes en inglés tiene mayor probabilidad de estar sobre la media. El 1.8% de los 3825 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 81.4% de los 70 estudiantes que se clasifican así, son correctamente clasificados y el 3.4% de los 1647 estudiantes que están sobre la media cumplen este patrón.

5.5 Evaluación e interpretación de resultados para la competencia ciudadana

Analizando los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T3824A26CIU_udenar (ver figura 18), en el cual se almacenan los datos válidos sobre los factores socioeconómicos, académicos e institucionales de 3824 estudiantes de programas profesionales de la Universidad de Na-

riño que presentaron las pruebas Saber Pro entre los años 2012 y 2014 en la competencia genérica de competencias ciudadanas, donde se escogió el atributo *mod_comp_ciudadanas_desemp* como clase, se puede observar que este clasifica correctamente a 2368 instancias, que corresponde a un porcentaje de precisión del 61.9% y 1456 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 38.1% (ver figura 19).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_comp_ciudadanas_desemp* del repositorio T3824A-26CIU_udenar que es de 2166 registros para el valor “sobre la media” y 1658 registros para el valor “bajo la media” y evaluando el modelo con la matriz de confusión de la figura 19, este clasifica correctamente a 1485 casos de estudiantes cuyos resultados en competencias ciudadanas están sobre la media y a 883 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 681 casos de estudiantes cuyos resultados en competencias ciudadanas están sobre la media y 775 casos que están bajo la media. Por lo tanto, el modelo clasifica correctamente al 68.6% de los estudiantes que están sobre la media en competencias ciudadanas y al 53.3% de las estudiantes que están bajo la media en esta competencia.

De acuerdo a las reglas de clasificación obtenidas con este modelo (ver figura 18), los patrones más representativos descubiertos, teniendo en cuenta que superen un soporte mínimo del 1% y una confianza mínima del 65%, son siete (7) y corresponden a las siguientes reglas:

Regla 1. Si los estudiantes son de la facultad de Ciencias Humanas entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 12.8% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 67.8% de los 490 estudiantes que se clasifican así, son correctamente clasificados y el 15.3% de los 2166 estudiantes que están sobre la media cumplen este patrón.

Regla 1. Si los estudiantes son de la facultad de Ciencias Humanas entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 12.8% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 67.8% de los 490 estudiantes que se clasifican así, son correctamente clasificados y el 15.3% de los 2166 estudiantes que están sobre la media cumplen este patrón.

Regla 2. Si los estudiantes son de la facultad de Derecho entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 8.8% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 88.8% de los 338 estudiantes que se clasifican así, son correctamente clasificados y el 13.9% de los 2166 estudiantes que están sobre la media cumplen este patrón.

Regla 3. Si los estudiantes son de la facultad de Artes, son de sexo masculino y son de un programa que pertenece al área del grupo de referencia de Ciencias Humanas, entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar bajo la media. El 8.1% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 65.3% de los 311 estudiantes que se clasifican así, son correctamente clasificados y el 12.2% de los 1658 estudiantes que están bajo la media cumplen este patrón.

Regla 4. Si los estudiantes son de la facultad de Ciencias Económicas y Administrativas y viven con su familia entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 7% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 66.9% de los 266 estudiantes que se clasifican así, son correctamente clasificados y el 8.2% de los 2166 estudiantes que están sobre la media cumplen este patrón.

Regla 5. Si los estudiantes son de la facultad de Ingeniería y viven con su familia entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 6.2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 65.8% de los 237 estudiantes que se clasifican así, son correctamente clasificados y el 7.2% de los 2166 estudiantes que están sobre la media cumplen este patrón.

Regla 6. Si los estudiantes son de la facultad de Ciencias Exactas y Naturales y son de un programa acreditado entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 2.6% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 76.5% de los 98 estudiantes que se clasifican así, son correctamente clasificados y el 3.5% de los 2166 estudiantes que están sobre la media cumplen este patrón.

Regla 7. Si los estudiantes son de la facultad de Ciencias de la Salud entonces el desempeño de los estudiantes en competencias ciudadanas tiene mayor probabilidad de estar sobre la media. El 2% de los 3824 estudiantes analizados en esta competencia en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 81.3% de los 75 estudiantes que se clasifican así, son correctamente clasificados y el 2.8% de los 2166 estudiantes que están sobre la media cumplen este patrón.

5.6 Evaluación e interpretación de resultados para el desempeño general en todas las competencias

Analizando los resultados obtenidos con el árbol de decisión construido con el conjunto de datos T3825A26GEN_udenar (ver figura 20), en el cual se almacenan los datos válidos sobre los factores socioeconómicos, académicos e institucionales de 3825 estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro entre los años 2012 y 2014 en las cinco competencias, donde se escogió el atributo

mod_general_desemp como clase, se puede observar que este clasifica correctamente a 2606 instancias, que corresponde a un porcentaje de precisión del 68.1% y 1219 instancias incorrectamente clasificadas, correspondiente a un porcentaje del 31.9% (ver figura 21).

Teniendo en cuenta la distribución de los valores del atributo clase *mod_comp_ciudadanas_desemp* del repositorio T3825A-26GEN_udenar que es de 2443 registros para el valor “sobre la media” y 1382 registros para el valor “bajo la media” y evaluando el modelo con la matriz de confusión de la figura 21, este clasifica correctamente a 2062 casos de estudiantes cuyos resultados promedio en las cinco competencias están sobre la media y a 544 casos que están bajo la media. Por otra parte, clasifica incorrectamente a 381 casos de estudiantes cuyos resultados promedio en todas las competencias están sobre la media y 838 casos que están bajo la media. En consecuencia, el modelo clasifica correctamente al 84.4% de los estudiantes que están sobre la media en el desempeño general en todas las competencias y al 39.4% de las estudiantes que están bajo la media en este desempeño general.

De acuerdo a las reglas de clasificación obtenidas con este modelo (ver figura 20), los patrones más representativos descubiertos, teniendo en cuenta que superen un soporte mínimo del 1% y una confianza mínima del 65%, son once (11) y corresponden a las siguientes reglas:

Regla 1. Si los estudiantes son de la facultad de Ingeniería, entonces, el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar sobre la media. El 12.9% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014, se clasifican de esta manera. El 74.7% de los 495 estudiantes que se clasifican así, son correctamente clasificados y el 15.1% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 2. Si los estudiantes son de la facultad de Ciencias Humanas, entonces, el desempeño general de los estudiantes en todas las

competencias tiene mayor probabilidad de estar sobre la media. El 12.8% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 74.7% de los 490 estudiantes que se clasifican así, son correctamente clasificados y el 15% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 3. Si los estudiantes son de la facultad de Derecho entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar sobre la media. El 8.8% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 86.7% de los 338 estudiantes que se clasifican así, son correctamente clasificados y el 12% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 4. Si los estudiantes son de la facultad de Ciencias Económicas y Administrativas y viven con su familia entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar sobre la media. El 7% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 75.3% de los 267 estudiantes que se clasifican así, son correctamente clasificados y el 8.2% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 5. Si los estudiantes son de la facultad de Ciencias Exactas y Naturales entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar sobre la media. El 6.7% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 75% de los 256 estudiantes que se clasifican así, son correctamente clasificados y el 7.9% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 6. Si los estudiantes son de la facultad de Ciencias Agrícolas entonces el desempeño general de los estudiantes en todas las

competencias tiene mayor probabilidad de estar sobre la media. El 6.2% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 65.5% de los 238 estudiantes que se clasifican así, son correctamente clasificados y el 6.4% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 7. Si los estudiantes son de la facultad de Educación y su edad está entre 23 y 31 años entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar bajo la media. El 5.3% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 65.5% de los 203 estudiantes que se clasifican así, son correctamente clasificados y el 9.6% de los 1382 estudiantes que están bajo la media cumplen este patrón.

Regla 8. Si los estudiantes son de la facultad de Artes y son del nivel 1 del SISBEN entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar bajo la media. El 4.2% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 63.4% de los 161 estudiantes que se clasifican así, son correctamente clasificados y el 7.4% de los 1382 estudiantes que están bajo la media cumplen este patrón.

Regla 9. Si los estudiantes son de la facultad de Ingeniería Agroindustrial y la condición de su vivienda es buena entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar sobre la media. El 2.4% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 65.6% de los 90 estudiantes que se clasifican así, son correctamente clasificados y el 2.4% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 10. Si los estudiantes son de la facultad de Ciencias de la Salud entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar sobre la media. El 2% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 89.3% de los 75 estudiantes que se clasifican así, son correctamente clasificados y el 2.7% de los 2443 estudiantes que están sobre la media cumplen este patrón.

Regla 11. Si los estudiantes son de la facultad de Ciencias Económicas y Administrativas, no viven con su familia y la ocupación de su madre es empleada entonces el desempeño general de los estudiantes en todas las competencias tiene mayor probabilidad de estar bajo la media. El 1.4% de los 3825 estudiantes analizados en el desempeño general de las cinco competencias en las pruebas Saber Pro 2012-2014 se clasifican de esta manera. El 65.5% de los 55 estudiantes que se clasifican así, son correctamente clasificados y el 2.6% de los 1382 estudiantes que están bajo la media cumplen este patrón.

DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

El objetivo de la investigación que se reporta en este libro, fue “Detectar patrones de desempeño académico en las competencias genéricas de las pruebas Saber Pro que presentaron los estudiantes de los programas profesionales de la Universidad de Nariño en los años 2012, 2013 y 2014 utilizando técnicas de minería de datos”, el cual se logró en su totalidad. Las fuentes de datos fueron las bases de datos del ICFES, de las cuales se obtuvo los datos socioeconómicos, académicos e institucionales de estos estudiantes. La metodología utilizada fue CRISP-DM, y la técnica de minería de datos aplicada para el descubrimiento de patrones de desempeño académico, fue clasificación, basada en árboles de decisión.

Para el descubrimiento de patrones, en cada competencia se tuvo en cuenta aquellos que sobrepasaron un soporte mínimo del 1% del total de casos evaluados y una confianza del 65% de casos que cumplían el patrón descubierto. El resto de patrones fueron descartados.

Por esta razón, se analizaron cuarenta y una reglas, distribuidas como sigue: lectura crítica, diez reglas; comunicación escrita, seis;

razonamiento cuantitativo, seis; inglés, trece; y finalmente, en competencias ciudadanas se interpretaron siete reglas.

De la figura 11, se obtiene que en las pruebas Saber Pro 2012 a 2014, el porcentaje de estudiantes de la Universidad de Nariño en la competencia de Lectura Crítica que se ubican SOBRE la media es de 63.4%; y el porcentaje de estudiantes BAJO la media es de 36.6%. La Tabla 36 contiene un resumen de las diez reglas con mayor probabilidad de ocurrencia, tanto de puntuaciones SOBRE la media, como BAJO la media.

Tabla 36. Patrones destacados en la competencia Lectura Crítica

Patrones destacados	% SOBRE la media correctamente clasificados	% BAJO la media correctamente clasificados
R1. Edad < 23 años	73.3	
R2. Edad entre 23 y 31 años; Facultad de Ciencias Humanas	70.0	
R3. Edad entre 23 y 31 años; Facultad de Derecho	79.5	
R4. Edad entre 23 y 31 años; Facultad de Ciencias Exactas y Naturales	71.9	
R5. Edad entre 23 y 31 años; Facultad de Ingeniería; vive con la familia	70.1	
R6. Edad entre 23 y 31 años; Facultad de Ciencias Económicas y Administrativas; estrato medio (3 y 4)	70.0	
R7. Edad entre 23 y 31; Facultad de Ciencias de la Salud	92.2	
R8. Edad entre 41 y 49 años;		68.6
R9. Edad entre 32 y 40; NO SISBEN	71.7	
R10. Edad entre 23 y 31 años; Facultad de Ingeniería; NO vive con la familia; programa NO acreditado	68.1	

Fuente: esta investigación.

De la tabla anterior se tiene que, en promedio, el 74.1% de los estudiantes que pertenecen a los descritos por las reglas que más aportan al mejoramiento de la media de la Institución en Lectura Crítica, son menores de 23 años; o, en caso contrario, tienen edades entre 23 y 31 años, y son de las facultades de Ciencias Humanas, Derecho, Ciencias de la Salud, y Ciencias Exactas y Naturales - FACIEN; también pertenecen los estudiantes de estrato medio de la Facultad de Ciencias Económicas y Administrativas - FACEA; los de la Facultad de Ingeniería que viven con la familia y los estudiantes de todas las facultades de edades entre 32 y 40 años que no tengan SISBEN.

Por otra parte, los estudiantes con edades entre 41 y 49 años tienen el 68.6% de probabilidad de que sus puntuaciones estén bajo la media de la Institución; no obstante, no requieren atención especial porque la población estudiantil en ese rango de edad, no representa una proporción importante en la Universidad de Nariño.

De lo anterior se puede señalar que las políticas de la Institución tendientes a mejorar el desempeño de los estudiantes en “Lectura Crítica” se deben enfocar en aquellos cuya edad sea mayor que 23 años; pero, especialmente, en quienes no pertenezcan a las Facultades de Ciencias Humanas, Derecho, FACIEN, FACEA, Ingeniería, y Facultad de Ciencias de la Salud - FACIES.

Por otra parte, de la matriz de confusión de la figura 13, se obtiene que el 59% de estudiantes de la Institución que presentó las pruebas Saber Pro 2012 a 2014, obtuvo puntuaciones que se ubican SOBRE la media en la competencia en “Comunicación Escrita”; por su parte, 41% de tal población obtuvo puntuaciones BAJO la media.

En la Tabla 37 se muestra las seis reglas de los patrones que tienen mayor probabilidad de que la media en “Comunicación Escrita”, esté SOBRE la media o BAJO la media.

Tabla 37. Patrones destacados en la competencia Comunicación Escrita

Patrones destacados	% SOBRE la media correctamente clasificados	% BAJO la media correctamente clasificados
R1. Género femenino	66.3	
R2. Género masculino; solteros; Facultad de Ciencias Económicas y Administrativas	65.6	
R3. Género masculino; solteros; Facultad de Derecho	66.2	
R4. Género masculino; solteros; Facultad de Ciencias Exactas y Naturales; Grupo Ciencias Naturales y Técnicas	67.3	
R5. Género masculino; solteros; Facultad Ciencias de la Salud	65.8	
R6. Género masculino; solteros; Facultad Ciencias Agrícolas; NO vive con la familia		65.6

Fuente: esta investigación.

Según los patrones configurados con las reglas de clasificación del árbol de “Comunicación Escrita”, indicados en la Tabla 37, en promedio el 66.2% de los estudiantes incluidos en las reglas, tienen la posibilidad de obtener puntuaciones sobre la media de la Institución. En general, las mujeres de la Institución representan el 66.3% de la población que tiene puntuaciones por encima de la media, en comunicación escrita; pero también se destacan los aportes sobre la media de los hombres solteros de las facultades FACEA, Derecho, FACIES y FACIEN. Según los patrones, merecen atención importante los estudiantes hombres, solteros de la Facultad de Ciencias Agrícolas – FACIA, que no viven con la familia, pues, ellos tienen una probabilidad del 65.6% de que sus puntuaciones, en esta competencia, estén bajo la media; probabilidad que se considera importante para el aporte negativo de la media.

Si bien los estudiantes descritos en las primeras cinco reglas, tienen mayor probabilidad de que sus puntuaciones en comunicación escrita, estén por encima de la media, la probabilidad es baja, teniendo en cuenta que apenas superan el 50%; por lo cual, es urgente, que la Institución implemente una política para mejorar el desempeño de todos los estudiantes en las competencias de Lectura Crítica.

En la competencia de “Razonamiento Cuantitativo”, en las pruebas Saber Pro 2012 a 2014, se encuentra que el 54.3% de los estudiantes de la Universidad de Nariño obtienen puntuaciones SOBRE la media, y el 45.7 se ubican debajo de ella, datos que se obtienen de la matriz de confusión de la figura 15. La tabla 38 que sigue, resume las seis reglas más importantes de los patrones de estudiantes con mayor probabilidad de obtener puntuaciones SOBRE la media y BAJO la media.

Tabla 38. Patrones destacados en la competencia Razonamiento Cuantitativo

Patrones destacados	% SOBRE la media correctamente clasificados	% BAJO la media correctamente clasificados
R1. Área ciencias Naturales y Técnicas	71.6	
R2. Área ciencias Humanas; Facultad de Artes		66
R3. Área Ciencias humanas; Facultad de Educación		82.6
R4. Área Ciencias Sociales; Facultad de Ciencias Económicas y Administrativas; vive con la familia	75.2	
R5. Área ciencias sociales; NO vive con la familia; género femenino		66.7
R6. Facultad de Ciencias de la Salud	89.3	

Fuente: esta investigación.

De los patrones extraídos con las reglas de clasificación del árbol de “Razonamiento Cuantitativo”, en promedio el 78.7% de los estudiantes incluidos en estas las reglas, tienen la posibilidad de obtener puntuaciones sobre la media de la Institución, y el 71.8 de obtener puntuaciones por debajo de este valor. Se observa que los estudiantes que más aportan para subir la media son de la FACIES, de las áreas de ciencias naturales y técnicas, de las ciencias sociales, pero siempre que sean de la FACEA y vivan con la familia. En resumen, se podría decir que los estudiantes que más aportan para la mejora de la media, son de las áreas de ciencias naturales y técnicas y de FACIES. En cambio, los estudiantes que se destacan por el aporte a la disminución de la media, pertenecen a las áreas de ciencias humanas, particularmente, de las facultades de Educación y Artes, y las mujeres de las áreas ciencias sociales que no viven con la familia.

Estos resultados son coherentes con la naturaleza de los planes de estudio, ya que, en general, el mejor desempeño en las competencias de razonamiento cuantitativo se espera en los programas que pertenecen a las ciencias naturales y técnicas, entre los cuales están las ingenierías. En consecuencia, la Institución debería adoptar estrategias que ayuden a los estudiantes de las otras áreas del saber, para que mejoren sus competencias de razonamiento cuantitativo, toda vez que son muy importantes en el desempeño de todas las profesiones, pero además, la Institución debe implementar estrategias que contribuyan a fortalecer dichas competencias en los programas que, en el momento, se destacan positivamente.

Además, con base en la matriz de confusión de la figura 17, se tiene que el porcentaje de estudiantes de la Institución en la “Competencia en Inglés” que presentaron Saber Pro en el período 2012 a 2014, cuya media se encuentra SOBRE la media en Inglés, es del 43.3%; y el porcentaje BAJO la media es del 56.7%.

La Tabla 39 que se presenta en seguida, contiene las reglas más importantes de los patrones que tienen mayor probabilidad de obtener una puntuación SOBRE la media o BAJO la media.

Tabla 39. Patrones destacados en la competencia Inglés

Patrones destacados	% SOBRE la media correctamente clasificados	% BAJO la media correctamente clasificados
R1. Tiene personas a cargo		72.2
R2. NO tiene personas a cargo; Facultad de Ingeniería	71.2	
R3. NO tiene personas a cargo; Facultad de Ciencias Económicas y Administrativas; NO vive con la familia.		73.1
R4. NO tiene personas a cargo; Facultad de Artes; estrato bajo (1 y 2)		67.8
R5. NO tiene personas a cargo; Facultad de Educación		79.4
R6. NO tiene personas a cargo; Facultad de Ciencias Agrícolas		62.9
R7. NO tiene personas a cargo; Facultad de Ciencias Pecuarias; estrato bajo (1 y 2)		67.6
R8. NO tiene personas a cargo; Facultad de Ciencias Económicas y Administrativas; vive con la familia; programa NO acreditado	70.9	
R9. NO tiene personas a cargo; Facultad de Agroindustria		65.4
R10. NO tiene personas a cargo; Facultad de Ciencias Exactas y Naturales; Área Ciencias Naturales y Técnicas	78.9	
R11. NO tiene personas a cargo; Facultad de Derecho; viven con la familia; menor de 23 años	71.1	
R12. NO tiene personas a cargo; Facultad de Derecho; NO viven con la familia		66.7
R13. NO tiene personas a cargo; Facultad de Ciencias de la Salud	81.4	

Fuente: esta investigación.

En la Tabla 39, presentada anteriormente, se puede observar que, en los patrones extraídos sobre el desempeño en competencias en Inglés, en promedio el 74.7% de los estudiantes incluidos en las reglas, tienen probabilidad de obtener puntuaciones SOBRE la media de la Institución, y el 69.4% de obtener puntuaciones por DEBAJO de este valor; este porcentaje se considera muy alto, y requiere, en consecuencia, atención central por parte de la Universidad.

En términos generales, los estudiantes que tienen el 74.7% de probabilidad para obtener puntuaciones SOBRE la media, NO tienen personas a cargo y pertenecen a las facultades de ingeniería, FACEA, FACIEN, Derecho o FACIES. Parece extraño que, entre estos estudiantes, no aparezca la Facultad de Ciencias Humanas, a la cual pertenecen los programas del Departamento de Lingüística e Idiomas, que por su naturaleza deberían aportar significativamente al crecimiento de la media. Por su parte, los estudiantes que tienen una alta probabilidad de que sus puntuaciones en competencias en Inglés estén por debajo de la media (69.4%), tienen personas a cargo; o si no las tienen, son de las Facultades de Artes, Educación, Ciencias Agrarias, Ciencias Pecuarias, Agroindustria; o son de la FACEA o Derecho, que no viven con la familia.

Estos resultados son coherentes con la naturaleza de los programas, puesto que, para los programas que más aportan a mejorar la media, se encuentra mucha bibliografía, en idioma inglés, disponible en la biblioteca de la Institución. Teniendo en cuenta que un altísimo porcentaje de estudiantes de la Universidad, no tiene personas a cargo, las estrategias de la Institución para mejorar el desempeño en inglés, debería orientarse, prácticamente, a toda la población estudiantil.

De los estudiantes que presentaron la Prueba de Estado Saber Pro entre los años 2012 y 2014, el 56.6% obtuvo puntuaciones SOBRE la media en “Competencias Ciudadanas”, y el 43.4% BAJO la media en esta misma competencia. Estos datos se obtienen según los resultados de la matriz de confusión de la figura 19.

La tabla 40 contiene las siete reglas que tienen más probabilidad de ocurrencia, es decir, muestra los patrones que predominan en la obtención de puntuaciones tanto por encima como debajo de la media en la evaluación de “Competencia Ciudadanas”.

De dicha tabla, se tiene que, en promedio, el 74.5% de los estudiantes que pertenecen a los descritos por las reglas que identifican a los estudiantes que más aportan al mejoramiento de la media de la Institución en Competencias Ciudadanas, pertenecen a las facultades de Ciencias Humanas, Derecho, Ciencias de la Salud; también se incluyen los de las facultades de Ingeniería y FACEA que viven con la familia, y los programas acreditados de la FACIEN. En contra parte, los estudiantes de género masculino de la Facultad de Artes, son muy susceptibles de obtener puntuaciones por debajo de la media; es decir, son estudiantes que se destacan por el aporte a la baja de la media de la Institución; hecho que amerita atención institucional importante.

Tabla 40. Patrones destacados en Competencias Ciudadanas

Patrones destacados	% SOBRE la media correctamente clasificados	% BAJO la media correctamente clasificados
R1. Facultad de Ciencias Humanas	67.8	
R2. Facultad de Derecho	88.8	
R3. Facultad de Artes; género masculino; Área Ciencias Humanas		65.3
R4. Facultad de Ciencias Económicas y Administrativas; Vive con la familia	66.9	
R5. Facultad de Ingeniería; vive con la familia	65.8	
R6. Facultad de Ciencias Exactas y Naturales; Programa acreditado	76.5	
R7. Facultad de Ciencias de la Salud	81.3	

Fuente: esta investigación.

Se puede señalar que, para mejorar el desempeño de los estudiantes en las competencias ciudadanas, conviene que la Universidad enfoque sus esfuerzos, principalmente, en los estudiantes de las facultades distintas a Ciencias Humanas, Derecho, Ciencias de la Salud, Ciencias Exactas y Naturales, FACEA e Ingeniería. Vale decir que, las facultades que más aportan al promedio de competencias ciudadanas son, en su orden, Derecho, Ciencias de la Salud, FACIEN con los programas acreditados y la Facultad de Ciencias Humanas. En realidad, estos resultados son coherentes con la naturaleza de los currículos respectivos.

En resumen, entre los atributos que forman parte de los patrones descubiertos en la competencia de lectura crítica, se destacan la edad del estudiante y la facultad a la cual pertenecen, como dos variables importantes asociadas al buen o bajo desempeño académico de los estudiantes de la Universidad de Nariño, en esta competencia.

En la competencia de comunicación escrita, los atributos que forman parte de la mayoría de los patrones descubiertos son el sexo del estudiante, el estado civil y la facultad a la cual pertenecen, que constituyen tres variables importantes asociadas al buen o bajo desempeño académico de los estudiantes en dicha competencia.

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

En la competencia de razonamiento cuantitativo, los atributos que forman parte de la mayoría de los patrones descubiertos hacen referencia al área a la cual pertenece el programa académico y la facultad. 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, formación que, al margen de las Pruebas Saber Pro, es importante. No obstante, lo ideal es que tales cursos sean 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 fomente y 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; todo en la vía de contribuir al fomento y desarrollo de la competencia de razonamiento cuantitativo.

Al parecer, es en este sentido que el MEN (s.f) concibe la alfabetización cuantitativa, entendiéndola como el dominio del lenguaje y los métodos matemáticos, como parte indispensable para la comprensión de problemas y las 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, 1), 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”, afirmación que

hace notar que la competencia del razonamiento cuantitativo es indispensable y, por lo mismo, inaplazable para todo profesional.

A pesar del reconocimiento sobre 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.

Por otra parte, entre los atributos que forman parte de los patrones descubiertos en la competencia de inglés, están tener o no personas a cargo, la facultad a la cual pertenece, el estrato socioeconómico, y si vive o no con su familia, variables que están asociadas al buen o bajo desempeño académico de los estudiantes de la Universidad de Nariño, en las pruebas Saber Pro, en esta competencia.

Estos hechos coinciden con Garbanzo (2007), Seibold (2000) y Montero & Villalobos (2004), en el sentido 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, lo cual 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 donde los estudiantes de estratos bajos tienen alta probabilidad de obtener puntuaciones por debajo de la media de la institución.

En competencias ciudadanas, la facultad a la cual pertenece el estudiante y el hecho de vivir con su familia, forman parte de la mayoría de los patrones descubiertos, y constituyen dos variables importantes asociadas al buen o bajo desempeño académico de los estudiantes de la Universidad de Nariño, en las pruebas Saber Pro, en esta competencia. Esto guarda coherencia con la naturaleza de los programas que más aportan al mejoramiento del promedio en esta competencia, y que pertenecen a las facultades

de Ciencias Humanas, Derecho, Ciencias de la Salud, Ciencias Exactas y Naturales, FACEA e Ingeniería.

Teniendo en cuenta los resultados de esta investigación, se plantea como trabajos futuros, complementar el estudio utilizando otras técnicas de minería de datos que permitan, por ejemplo, relacionar qué atributos se presentan juntos asociados al desempeño académico de los estudiantes de los programas profesionales de la Universidad de Nariño, en las pruebas Saber Pro, y cómo se agrupan los individuos de acuerdo a su rendimiento en dichas pruebas. Además, sería recomendable realizar estudios sobre el análisis relacional entre el rendimiento académico en las pruebas Saber Once, requisito principal para el ingreso a la Universidad de Nariño, y su desempeño académico en la carrera y en las pruebas Saber Pro.

Finalmente, entre las dificultades presentadas en el desarrollo de esta investigación, se cuenta la regular calidad de los datos de las bases de datos del ICFES, por lo cual, se tuvo que descartar varios atributos por la imposibilidad de obtener sus valores, aún recurriendo a otras fuentes. Este hecho pudo, de alguna manera, influir en el descubrimiento de los patrones objeto de este estudio, además de su implicación en el gran consumo de recursos durante los procesos de limpieza y transformación de los datos.

REFERENCIAS BIBLIOGRÁFICAS

- Abu-Saa, A. (2016). Educational Data Mining & Students' Performance Prediction (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5.
- Agrawal, R. & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In: 20th International Conference on Very Large Data Bases, VLDB 1994, (12-15/09/1994). Santiago de Chile (Chile): VLDB. Proceedings. p. 487-499. ISBN: 1-55860-153-8.
- Algarni, A. (2016). Data Mining in Education. International Journal of Advanced Computer Science and Applications IJACSA, vol. 7(6). Disponible en: <https://doi.org/10.14569/IJACSA.2016.070659>.
- 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>.
- Ardila Rodríguez, M., (2011). Calidad de la Educación Superior en Colombia. ¿Problema de Compromiso Colectivo?, Rev. Educ. y Desarrollo. Soc., vol. 5, no 2, pp. 44–55, 2011.
- Azevedo, A., Zantos, M., 2008. KDD, SEMMA and CRISP-DM: A parallel overview. Proceeding of IADIS European Conference Data Mining. ISBN: 978-972-8924-63-8. Amsterdam (Holland).
- Baker, R. (2010). Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier.
- Bodensiek, A. (2010). Estudio sobre los factores que influyen en el rendimiento escolar. Secretaria de Educación Municipal.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de economía*, 60(60), 39–65.
- Castro, F., Vellido, A., Nebot, A. and Mugica, F. (2007). Applying data mining techniques to e-learning problems. Evolution of teaching and learning paradigms in intelligent environment, pp. 183–221, Springer.
- Climent-Bonilla, J. B. (2014). Origen, desarrollo y declive de las competencias individuales en tiempos de incertidumbre. *Educación y Educadores*, 17(1). Disponible en: <http://educacionyeducadores.unisabana.edu.co/index.php/eye/article/view/3499>
- Cohen, J. (1988). Análisis de poder estadístico para las Ciencias del comportamiento (2ª ed.). Nueva Jersey: Lawrence Erlbaum.
- Congreso de Colombia. Ley 30 (1992). Por la cual se organiza el servicio de la educación superior. Disponible en: <http://www.bienestarbogota.unal.edu.co/pdf/egresados/ley30de1992.pdf>
- Congreso de la República de Colombia (2009). Ley 1324 del 13 de Julio del 2009. Colombia: Congreso de la República de Colombia, 2009, p. 9.
- Colombia. Consejo Nacional de Acreditación – CNA. (2013). Lineamientos para la acreditación de programas de pregrado. Bogotá. CNA.
- Colombia. Consejo Nacional de Acreditación – CNA. (s. f.). Acreditación Institucional. Disponible en: <http://www.cna.gov.co/1741/article-186371.html>

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. Disponible en: <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=47530>

Colombia. Ministerio de Educación Nacional MEN. (s. f.). Propuesta de lineamientos para la formación por competencias en educación superior. Disponible en: 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. Disponible en: <http://www.wvu.edu/vpue/documents/issue8.pdf>

Di Gresia, L., & Porto, A. (2004). Características y Rendimiento de Estudiantes Universitarios. El Caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. Recuperado a partir de <http://sedici.unlp.edu.ar/handle/10915/3505>

Erazo, O. (2012). El rendimiento académico, un fenómeno de múltiples relaciones y complejidades. *Revista Vanguardia Psicológica* / Año 2 / Volumen 2 / Numero 2, octubre-marzo / pp. 144-173 / ISSN 2216-0701.

Fayyad, U. M. et al. (1996). From data mining to knowledge discovery: an overview. In Fayyad, U. M. et al (Eds.), *Advances in knowledge discovery and data mining*. AAAI Press / The MIT Press.

Fernández, G. (2009). Extracción de Información de la Web usando Técnicas de Minería de Datos. Recuperado a partir de Disponible en: <http://www.tdg-seville.info/Download.ashx?id=48>

García, G. (2018). Aplicación de la metodología CRISP-DM a la recolección y análisis de datos georreferenciados desde Twitter. Trabajo de grado para optar el título de especialista en Geomática. Universidad Nueva Granada, Bogotá D.C., Colombia. Disponible en: <https://repository.unimilitar.edu.co/bitstream/handle/10654/20099/GarciaVelezGustavoAdolfo2018.pdf>.

Gallardo, J., 2009. Metodología para la Definición de Requisitos en Proyectos de Data Mining (ER-DM). Tesis doctoral. Departamento de Lenguajes y Sistemas, Universidad Politécnica de Madrid. Madrid, España. Disponible en: http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf

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., & Álvarez, A. (2010). Análisis de datos en WEKA—pruebas de selectividad. *línea| disponible en <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>*. Recuperado a partir de <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>

GRIAS (2016). *Yachay Datamining Tool:Manual del usuario*. Grupo de Investigación Aplicada en Sistema, Universidad de Nariño. Pasto, Nariño, Colombia

Gómez, G. Soares, A. (2013). Diferencias de género con relación al desempeño académico en estudiantes de nivel básico. *Alternativas en Psicología*, XVII, 28, 106-118.

González, J., & Wagenaar, R. (2003). "Structures in Europe". Informe Final Proyecto Piloto-Fase 1. Informe realizado por más de 100 universidades, coordinado por la Universidad de Deusto (España) y la Universidad de Groningen (Países Bajos) y apoyado por la Comisión Europea. Editado por Julia González y Robert Wagenaar. Bilbao.

Hall, M., Frank, E., & Witten, I. (2011). *Practical Data Mining: Tutorials*. University of Waikato. Recuperado a partir de <http://www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf>

Han, J., Kamber, M. & Pei, J. (2012). *Data Mining: Concepts and Techniques, Third Edition* (3 edition.). Burlington, MA: Morgan Kaufmann. ISBN: 978-0-12-381479-1.

Hernández, E., & Lorente, R. (2009). *Minería de datos aplicada a la detección de Cáncer de Mama*. Universidad Carlos III de Madrid. Recuperado a partir de <http://tps5to-utn-fre.googlecode.com/svn/trunk/BI/Cancer%20de%20Mama/14.pdf>

Hernández, I., & Ojeda, B. (2009). Competencias: Un reto en la formación de profesionales. *Memorias*, 5(12), 102–119.

Hernández, J., Ramírez, M. J., & Ferri, C. (2005). *Introducción a la Minería de Datos*. Editorial Pearson Educación S.A, Madrid. Recuperado a partir de <http://dspace.ucbscz.edu.bo/dspace/handle/123456789/526>

Huapaya, C., Lizarralde, F., Arona, G., & Massa, S. (2012). Minería de datos educacional en ambientes virtuales de aprendizaje. En *memorias de XIV Workshop de Investigadores en Ciencias de la Computación*, (pp. 996-1000). Entre Ríos, Argentina: RedUNCI. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/18347>.

Icfes (2010). Exámenes de Estado de calidad de la educación superior. Saber Pro (ECAES). Análisis de resultados del período 2004-2008. ICFES. Disponible en: <http://54.208.2.57/datos/Informe%20final%20resultados%20SABER%20PRO%202004%202008.pdf>

Icfes (2012a). Examen Saber Pro: Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior. ICFES. Disponible en: www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias

Icfes (2012b). Saber Pro: Principales resultados en Competencias Genéricas. Santa Marta, Colombia. ICFES. Disponible en: www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias

Icfes (2011a). Lineamientos Saber Pro. Bogotá. ICFES. Disponible en: http://aprendeenlinea.udea.edu.co/lms/moodle/file.php/532/Lineamientos_SABER_PRO_2011_2_30_08_1_.pdf

Icfes (2011b). Módulos de competencias genéricas y específicas disponibles. Evaluación de la calidad de la educación superior. ICFES. Disponible en: <http://acofartes.org.co/docsweb/documento/ICFES%202011,%20M%C3%93DULOS%20COMPETENCIAS%20GEN%C3%89RICAS%20Y%20ESPEC%C3%8DFICAS.pdf>

Icfes (2013). Estructura del Saber Pro. ICFES. Disponible en: http://ead.uis.edu.co/empresarial/images/stories/doc/SABER_PRO_2013_ESTRUCTURA.pdf

Icfes (2016). Informe nacional de resultados Saber Pro 2012-2015. Evaluación de la calidad de la educación superior - ICFES. Santafé de Bogotá D.C., Colombia

Icfes (2017). Guía de orientación módulo de competencias genéricas. Instituto Colombiano para la Evaluación de la Educación - ICFES. Santafé de Bogotá D.C., Colombia.

- Jiménez, A. & Álvarez, H. (2010). Minería de datos en la Educación. ACM Inteligencia en Redes de Comunicación.
- Kurgan, L. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, 21(01), 1-24.
- Larose, D. & Larose, Ch. (2014). *Discovering Knowledge in Data. An Introduction to Data Mining*(2d ed.) . Jhon Willey & Sons. New Jersey (USA).
- Lemos, C., Cuello Daza, E., Roys Romero, N y Rojas Pinzón, L. (2014). Análisis de los resultados de las prueba Saber Pro. Programa de licenciatura en etnoeducación, Riohacha, Guajira, Colombia.
- Lerner, J. (2012). *Rendimiento académico de los estudiantes de pregrado de la Universidad EAFIT*. Perspectiva cuantitativa.
- Lerner, J., & Montes, I. (2012). Rendimiento académico de los estudiantes de pregrado de la Universidad EAFIT*. *Cuadernos de Investigación*, (91). Recuperado a partir de <http://publicaciones.eafit.edu.co/index.php/cuadernos-investigacion/article/view/1254>
- Martínez-Martínez, A., Cegarra Navarro, J., & Rubio Sánchez, J. A. (2012). Aprendizaje basado en competencias, una propuesta para la autoevaluación docente. Recuperado a partir de <http://digibug.ugr.es/handle/10481/23026>.
- Maya-Pérez, N, Aguilar-Cisneros, J., Zamora-Ríos, R. & Barronadame, J. (2016). Propuesta del Diseño de un Modelo Predictivo de alerta temprana en indicadores educativos de Nivel Superior aplicando Minería de Datos. *Revista de Aplicación Científica y Técnica*, 2-4: pp. 29-40.
- Merceron, A. & Yacef, K. (2008). Interestingness measures for associations rules in educational data. *EDM*, vol. 8, pp. 57–66.
- Ministerio de Educación Nacional (2009). Decreto 3963 del 14 de octubre de 2009, por el cual se reglamenta el Examen de Estado de Calidad de la Educación Superior. Ministerio de Educación Nacional. Recuperado a partir de www.mineducacion.gov.co/1621/article-205955.html
- Mizala, A., Romaguera, P., & Reinaga, T. (2002). *Factores que inciden en el rendimiento escolar en Bolivia*. Recuperado a partir de <http://ideas.repec.org/p/edj/ceauch/61.html>
- 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. *RIED*, 14 2, 77-95.
- Montero, E., Villalobos, J. (2004). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico y a la repetición estudiantil en la Universidad de Costa Rica. San José, Costa Rica: Universidad de Costa Rica.
- Orea, S. V., Vargas, A. S., & Alonso, M. G. (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., Restrepo, G., Usuga, O. y Mendoza, R. (2013). 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. Disponible en: <http://www.udea.edu.co/portal/page/portal/bibliotecaSedesDependencias/unidadesAcademicas/FacultadIngenieria/Diseno/Archivos/Tab/Rendimiento%20acad%C3%A9mico%20de%20los%20estudiantes%5B1%5D.pdf>.

- Peña, L. (2008). La competencia oral y escrita en la educación superior. Recuperado a partir de http://mineducacion.gov.co/1621/articles-189357_archivo_pdf_comunicacion.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. R. (1993). *C4. 5: programs for machine learning* (Vol. 1). Morgan Kaufmann. Recuperado a partir 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>
- Ramírez, C. & Teichler, U. (2014). Factores socioeconómicos y educativos asociados con el desempeño académico, según nivel de formación y género de los estudiantes que presentaron la prueba SABER PRO 2009. ICFES, Bogotá, Colombia.
- Robles, N. (2013). Desarrollo de Competencias Genéricas Mediante el Aprendizaje Basado en Problemas. Eleventh LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2013) "Innovation in Engineering, Technology and Education for Competitiveness and Prosperity". August 14-16, 2013. Cancún, México.
- Romero, C. & Ventura, S. (2010). Educational Data Mining: A Review of the State-of-the-Art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 40, no. 6, pp. 601-618.
- Rosado, A. & Vergel, A. (2017). Aplicación de la minería de datos en la educación en línea. Revista Colombiana de Tecnologías de Avanzada. Volumen 1, Número 29. ISSN: 1692-7257.
- Rubio, M. J. (2009). Nuevas Orientaciones y Metodología para la Educación a Distancia Loja - Ecuador: Editorial de la Universidad Técnica Particular de Loja.
- Sánchez-Santillán, M., Paule-Ruiz, M., & Núñez, J. (2016). Predicting Students' Performance: Incremental Interaction Classifiers. Proceedings of the Third ACM Conference on Learning @ Scale. Edinburgh: ACM New York, pp. 217-220.
- Salas, R. (2000). Educación Médica Superior. La calidad en el desarrollo profesional: avances y desafíos. *Educ Med Super*, 14(2). Recuperado de http://scielo.sld.cu/scielo.php?pid=S0864-2141200000200003&script=sci_arttext
- Sattler, K.-U., & 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). ACM. Recuperado a partir de <http://dl.acm.org/citation.cfm?id=502650>
- Sarramona, J. (2002). Evaluación de programas de educación a distancia. Barcelona: Universidad Autónoma de Barcelona.
- 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*. Recuperado de <http://www.rieoei.org/rie23a07.htm>
- Spicker, P., Alvarez, S., & Gordon, D. (s. f.). hH: Hacinamiento Hambruna. Recuperado a partir de <http://biblioteca.clacso.edu.ar/gsd/collect/clacso/index/assoc/D9393.dir/h.pdf>
- Timarán, S.R., Hernández, I., Caicedo, J., Hidalgo, A. & Alvarado, J (2016): Descubrimiento de Patrones de Desempeño Académico con árboles de decisión en las competencias genéricas de la formación profesional. Ediciones Universidad Coope-

rativa de Colombia. Bogotá, Colombia. ISBN:978-958-760-049-0. DOI: <http://dx.doi.org/10.16925/9789587600490>.

Timarán, R., Calderón, A., & Jiménez, J. (2013a). Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil [Application of data mining in extracting student dropout profiles]. *Ventana Informática*, (28). Recuperado a partir de <http://revistasum.umanizales.edu.co/ojs/index.php/ventanainformatica/article/view/181>

Timarán, R., Calderón, A., & 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 a partir de <http://www.acofipapers.org/index.php/acofipapers/2013/paper/view/211>

Timarán, R., & Millán, M. (2006). New algebraic operators and SQL primitives for mining classification rules. En *Computational Intelligence* (pp. 61–65). Recuperado a partir de <http://www.actapress.com/PaperInfo.aspx?PaperID=29048&reason=500>.

Timarán, R., Jiménez, J. & Calderón, A. (2017). Detección de patrones de deserción estudiantil con minería de datos caso de estudio: Universidad de Nariño e Institución Universitaria Cesmag, Editorial Universitaria Universidad de Nariño. Pasto (N), Colombia. 151 p. ISBN: 978-958-8958-38-5.

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. *Revista electrónica de Educación y Desarrollo*, 2(11).

Toro, J. y Villaveces, J. (2008). El pensamiento matemático: Una competencia genérica emergente. Disponible en: http://www.mineducacion.gov.co/1621/articles-189357_archivo_pdf_matematica_1B.pdf,

Universidad Nacional de Colombia - UNAL. (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 - UNAL. Recuperado a partir de www.unal.edu.co/diracad/evaluacion/SaberPro_2012/analisis_de_resultados.pdf

Valera, R. (2010). El proceso de formación del profesional en la educación superior basado en competencias: el desafío de su calidad, en busca de una mayor integralidad de los egresados. *Civilizar*, 117.

Valero Orea, S. (2009). Aplicación de técnicas de minería de datos para predecir deserción. *Aplicación de técnicas de minería de datos para predecir deserción*. México.

Valero, S., Salvador, A., & 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.

Villena, J., 2016. CRISP-DM: La metodología para poner orden en los proyectos de Data Science. Disponible en: <https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science>.

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. Fundación Universitaria Luis Amigó, Facultad de Educación, Licenciatura en Educación Básica con Énfasis en Inglés. Medellín, Colombia.



Editorial
Universidad de **Nariño**

Este libro se terminó de imprimir
en el mes de marzo de 2021,
en Graficolor Pasto sas
Calle 18 No. 29-67
Tels. 7310652 - 7311833
graficolorpasto@hotmail.com

En este libro se presentan los resultados del proyecto de investigación que tuvo como objetivo descubrir patrones de desempeño académico en las competencias genéricas de los estudiantes de programas profesionales de la Universidad de Nariño que presentaron las pruebas Saber Pro en los años 2012, 2013 y 2014.

Se seleccionó, de las bases de datos del ICFES, la información socioeconómica, académica e institucional de estos estudiantes. Se construyó, limpió y transformó un repositorio de datos y a partir de él, se descubrieron patrones asociados al buen o mal desempeño académico de los estudiantes en cada competencia, utilizando un modelo de clasificación basado en árboles de decisión y la metodología CRISP-DM.

El conocimiento descubierto se podrá integrar a los procesos de toma de decisiones de la Universidad de Nariño con el fin de mejorar la calidad de la educación que se imparte en sus programas profesionales.



Universidad de **Nariño**



ISBN: 978-958-5123-43-4



9 789585 1123434