

PREDICCIÓN DEL RESULTADO DE LAS COMPETENCIAS GENÉRICAS EN
LAS PRUEBAS SABER PRO PARA INGENIERÍA INDUSTRIAL Y CIVIL A TRAVÉS
DEL PERFIL SOCIOECONOMICO Y SOCIODEMOGRAFICO DE LOS
ESTUDIANTES

OLFER EFRAIN OLAYA SUAREZ

UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA INDUSTRIAL
BUCARAMANGA
2019

PREDICCIÓN DEL RESULTADO DE LAS COMPETENCIAS GENÉRICAS EN LAS PRUEBAS SABER PRO PARA INGENIERÍA INDUSTRIAL Y CIVIL A TRAVÉS DEL PERFIL SOCIOECONOMICO Y SOCIODEMOGRAFICO DE LOS ESTUDIANTES

OLFER EFRAIN OLAYA SUAREZ ID:231721

Estudiante director del proyecto:
OLFER EFRAIN OLAYA SUAREZ

Asesor metodológico:
MANUEL ARTURO JIMENEZ RAMIREZ

UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERIA
FACULTAD DE INGENIERÍA INDUSTRIAL
BUCARAMANGA
2019

Copyright © 2015 por Olfer Efraín Olaya Suarez. Todos los derechos reservados.

Dedicatoria

Este proyecto va con especial dedicatoria para mi padre Fernando Olaya Castro y mi madre Olga Suarez Valencia por cada grano de confianza y apoyo que me dieron durante todo este proceso. También a Dios que es el que nos da salud y bendiciones todos los días de la vida.

Agradecimientos

Nuevamente infinitas gracias doy a mis padres que aparte de la confianza y el apoyo, me brindaron la motivación en los momentos donde los necesitaba y por todo el esfuerzo económico para lograr este logro y a Dios por brindarme unos excelentes padres.

A mi director de tesis Manuel Arturo Jiménez Ramírez quien, con gran paciencia, dedico su tiempo, conocimiento para la realización de este proyecto.

A mi novia Diana quien fue un pilar emocional para mí y de gran apoyo para todo lo que necesité.

Tabla de contenido

1. Introducción	12
2. Delimitación del problema	14
3. Antecedentes	17
4. JUSTIFICACIÓN	21
5. Objetivos	23
5.1 Objetivo general.....	23
5.2 Objetivos específicos.	23
6. Marco referencial	25
6.1 Marco teórico	25
6.1.1. Pruebas Saber Pro.....	25
6.1.2. Análisis de correspondencia (AC).....	28
6.1.3. Árbol de decisión.....	30
7. Diseño metodológico	35
7.1 Alcance metodológico	35
7.2 Definición de la población.....	35
7.2.1. Unidad de investigación.	35
7.2.2. Tamaño de la población.....	35
7.3 Metodología.....	36
7.3.1. Procedimiento.	37
8. Resultados	39
8.1 Preparación y construcción de la base de datos	39
8.2 Perfil Estudiantes por Facultad.....	43
8.2.1. Análisis de correspondencia múltiple Variables Sociodemográficas y socioeconómicas.....	45
8.2.2. Resumen Modelo.....	45

8.3 Pronostico y validación de los árboles de clasificación de los programas académicos.....	51
8.3.1. Validación del modelo de predicción.	51
8.3.2. Validación de la predicción de Ingeniería industrial 2018.....	53
8.3.3. Validación de la predicción de ingeniería civil 2018.	62
8.3.4. Resultados ingeniería industrial.....	73
8.3.5. Resultados Ingeniería Civil.	82
9. Conclusiones y recomendaciones	92
10. Bibliografía.....	97
11. Anexos.....	104
12. Apéndice	109

Lista de tablas

Tabla 1. <i>Escala de puntajes a partir del 2016.</i>	27
Tabla 2. <i>Nivel de desempeño de los puntajes</i>	27
Tabla 3. <i>Comparación modelos CART y CHAID.</i>	34
Tabla 4. <i>Muestra de estudiantes</i>	36
Tabla 5. <i>Resumen del modelo ACM Variables sociodemográficas y económicas.</i>	46
Tabla 6. <i>Definición de los niveles de desempeño</i>	51
Tabla 7. <i>VARIABLES DE ENTRADA Árboles de decisión.</i>	52
Tabla 8. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de razonamiento cuantitativo.</i>	54
Tabla 9. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de lectura crítica.</i>	56
Tabla 10. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de competencia ciudadana.</i>	58
Tabla 11. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de inglés.</i>	60
Tabla 12. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia ciudadana.</i>	62
Tabla 13. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de razonamiento cuantitativo.</i>	64
Tabla 14. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de lectura crítica.</i>	66
Tabla 15. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia ciudadana.</i>	69
Tabla 16. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de inglés.</i>	71
Tabla 17. <i>Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de comunicación escrita.</i>	73

Lista de figuras

<i>Figura 1</i> Descripción de los módulos de competencias genéricas. ICFES (2017)	26
<i>Figura 2.</i> Variables no tomadas en cuenta por no tener trascendencia en los periodos 2015 a 2018	41
<i>Figura 3.</i> Frecuencias de las variables por concentración en respuestas.	42
<i>Figura 4.</i> Comparación de tenencia de bienes y servicios según profesión.....	44
<i>Figura 5.</i> ACM variables sociodemográficas y económicas	47
<i>Figura 6.</i> Mapa Perceptual dimensión educación-estrato.	49
<i>Figura 7.</i> Mapa perceptual dimensión competencias-rendimiento	50
<i>Figura 8.</i> Árbol Razonamiento cuantitativo de validación Ingeniería industrial.	53
<i>Figura 9.</i> Árbol lectura crítica de validación Ingeniería industrial.	55
<i>Figura 10.</i> Árbol competencia ciudadana de validación Ingeniería industrial.	57
<i>Figura 11.</i> Árbol Razonamiento Inglés de validación Ingeniería industrial.	59
<i>Figura 12.</i> Árbol comunicación escrita de validación Ingeniería industrial.	61
<i>Figura 13.</i> Árbol Razonamiento cuantitativo de validación Ingeniería civil.	63
<i>Figura 14.</i> Árbol lectura crítica de validación ingeniería civil.	65
<i>Figura 15.</i> Árbol competencias ciudadanas de validación ingeniería civil.	68
<i>Figura 16.</i> Árbol inglés de validación ingeniería civil.	70
<i>Figura 17.</i> Árbol comunicación escrita de validación ingeniería civil.	72
<i>Figura 18.</i> Árbol razonamiento cuantitativo ingeniería industrial.....	75
<i>Figura 19.</i> Árbol Lectura crítica ingeniería industrial.	77
<i>Figura 20.</i> Árbol Competencia ciudadana ingeniería industrial.	78
<i>Figura 21.</i> Árbol Desempeño inglés ingeniería industrial.	80
<i>Figura 22.</i> Árbol Comunicación escrita ingeniería industrial.....	81
<i>Figura 23.</i> Árbol rendimiento cuantitativo ingeniería civil.	83
<i>Figura 24.</i> Árbol lectura crítica ingeniería civil.....	85
<i>Figura 25.</i> Árbol competencia ciudadana ingeniería civil.	87
<i>Figura 26.</i> Árbol rendimiento inglés ingeniería civil.....	89
<i>Figura 27.</i> Árbol comunicación escrita ingeniería civil.....	91

RESUMEN GENERAL DE TRABAJO DE GRADO

TITULO: PREDICCIÓN DEL RESULTADO DE LAS COMPETENCIAS GENÉRICAS EN LAS PRUEBAS SABER POR PARA INGENIERÍA INDUSTRIAL Y CIVIL A TRAVÉS DEL PERFIL SOCIOECONÓMICO Y SOCIODEMOGRÁFICO DE LOS ESTUDIANTES

AUTOR(ES): Olfier Efraín Olaya Suárez

PROGRAMA: Facultad de Ingeniería Industrial

DIRECTOR(A): Manuel Arturo Jiménez Ramírez

RESUMEN

El siguiente proyecto es de tipo transversal y correlacional ya que permite conocer relación entre las variables en los periodos 2016 a 2018 y tiene como propósito predecir los resultados de los exámenes de la educación superior Saber Pro 2019 de las competencias genéricas, tomando como variables de entrada las preguntas de tipo personal, sociodemográficas y socioeconómicas de los estudiantes de Ingeniería industrial y civil de la Universidad Pontificia Bolivariana de Bucaramanga en los periodos 2016 a 2018, se utilizó técnicas de estadística multivariada de Análisis de Correspondencias Múltiple y Árboles de decisión CHAID con el programa estadístico SPSS STATISTICS. Inicialmente se obtuvo las bases de datos y luego la eliminación de aquellos que no fueron óptimos para el análisis de datos, luego se realizó el Análisis de Correspondencias Múltiples que permitió generar el perfil de los estudiantes de ingeniería civil e industrial, con variables que generaron impacto en su rendimiento. Seguido de ello se realizó arboles de decisión para cada una de las cinco competencias genéricas y también por facultad, esto permitió identificar que el género, edad y tipo de colegio Privado o Público en los estudiantes de ingeniería industrial y civil afecten su desempeño en las competencias, paralelo a esto se hizo validación del modelo donde se tomó los periodos 2016 y 2017 para predecir el año 2018 y luego comparar con los resultados obtenidos por los estudiantes en el 2018. Finalmente se plantea conclusiones y recomendaciones, para que se presenten resultados con mayor asertividad en futuros estudios.

PALABRAS CLAVE:

Árboles de decisión, CHAID, Análisis correspondencia múltiple, Minería de datos

V° B° DIRECTOR DE TRABAJO DE GRADO

GENERAL SUMMARY OF WORK OF GRADE

TITLE: PREDICTION OF THE RESULT OF GENERIC COMPETITIONS IN THE TESTS SABER PRO FOR INDUSTRIAL AND CIVIL ENGINEERING THROUGH THE SOCIOECONOMIC AND SOCIODEMOCRATIC PROFILE OF STUDENTS

AUTHOR(S): Oifer Efraín Olaya Suárez

FACULTY: Facultad de Ingeniería Industrial

DIRECTOR: Manuel Arturo Jiménez Ramírez

ABSTRACT

The purpose of this project is to predict the results of the Saber Pro 2019 higher education examinations of generic competences, taking personal questions as input variables, In the period 2016 to 2018, multivariate statistical techniques of Multiple Correspondence Analysis and CHAID Decision Trees were used with the statistical program SPSS STATISTICS. Initially the databases were obtained and then the elimination of those that were not optimal for the data analysis, then the Multiple Correspondence Analysis was carried out that allowed generating the profile of the students of civil and industrial engineering, with variables that generated impact in their performance. This was followed by decision trees for each one of the five generic competencies and also by faculty, this allowed to identify that the gender, age and type of Private or Public school in the students of industrial and civil engineering affect their performance in the competencies, parallel to this was validated the model where the periods 2016 and 2017 were taken to predict the year 2018 and then compare with the results obtained by the students in 2018. Finally, conclusions and recommendations are proposed so that results can be presented with greater assertiveness in future studies.

KEYWORDS:

Decision tree, CHAID, Multiple correspondence analysis, Data mining

V° B° DIRECTOR OF GRADUATE WORK

1. Introducción

Las pruebas de estado en Colombia son manejadas por el Ministerio de Educación Nacional (MEN), que a su vez son regidas por la Institución Colombiana del Fomento de la Educación Superior (ICFES); en la educación superior, los Exámenes de Estado de Calidad de la Educación Superior, ECAES antiguamente llamado así por sus siglas y que en el año 2009 la Ley 1324 estableció que dichos exámenes debían ser de obligatoriedad para obtener el título de pregrado. Para el segundo semestre de 2010 dichas pruebas pasaron a llamarse Prueba Saber Pro, que según (Decreto N° 1781 de 2003) afirma que los resultados de estas son de utilidad al gobierno nacional para evaluar la calidad del servicio de educación superior, medir las falencias y determinar las competencias del estudiante en las instituciones del país. De esta manera yace una condición de interés para esta investigación, ya que si bien las pruebas son resultados de metodologías, pensum, programas de estudio, entre otras de las universidades, se hace importante conocer el nivel incidencia de condiciones sociodemográficas (o de otro tipo de condiciones) en los resultados, y en especial en las que enmarca las competencias genéricas, que están definidas por el ICFES (2018) como “las capacidades de entender, interpretar y evaluar textos que pueden encontrarse tanto en la vida cotidiana, como en ámbitos académicos no especializados” (p,16)

Estos resultados se podrían ver favorecidos por las condiciones de calidad de vida en que se desarrolla el estudiante, son diversos estudios los que han demostrado que la calidad de vida son determinantes en el rendimiento estudiantil , así como las condiciones

psicológicas, biológicas y socioeconómicas (Duran, Prieto & García, 2017; Ruiz de miguel, 2001); Por otra parte, el provenir de un colegio mixto, la educación de los padres, el poco habito a la lectura afectan negativamente al desempeño académico (Vélez & Roa, 2005). De esta manera se precisa descubrir mediante el análisis de resultados de competencias genéricas de las pruebas Saber Pro-2016 a 2018, las condiciones o variables sociodemográficas y socioeconómicas que afecten el rendimiento en dichas pruebas, es por esta razón que este proyecto pretende predecir los resultados de las pruebas genéricas del Saber Pro en los periodos 2016 a 2018 a los estudiantes de ingeniería industrial y civil. Para esto se planeó las siguientes actividades a realizar:

- Construcción y preparación de las bases de datos Saber Pro de 2016 a 2018 con los datos luego a analizar.
- Caracterización de los estudiantes de ingeniería industrial y civil según los perfiles sociodemográficos y socioeconómicos.
- Asociación entre las características sociodemográficas y socioeconómicas de los estudiantes de ingeniería industrial y civil de la Universidad Pontificia Bolivariana Bucaramanga y los resultados de los módulos genéricos.
- Validación de cada uno de los modelos de predicción CHAID.

Todo lo nombrado anteriormente se realizó mediante técnicas estadísticas de Análisis de Correspondencias Múltiple (ACM) y Árboles de decisión tipo CHAID.

2. Delimitación del problema

La educación es considerada como factor influyente y fundamental en el desarrollo económico, tecnológico y social de un país. Al respecto, son amplios los estudios en el que han encontrado relación entre el nivel de alfabetización de sus habitantes con el crecimiento económico del mismo y por tanto, la calidad de vida e ingresos individuales (Bonilla, 2013; Briceño, 2011; Flores, 2016; Moratilla, 2018; Rajimon, 2010; Ramos et al., 2017;), específicamente, Hanushek y Wöbmann (2007) investigaron en más de 50 países que la calidad de educación evaluada por pruebas estandarizadas influye en el nivel económico de un país. Encontraron que por cada año de escolaridad aumenta el crecimiento económico a largo plazo y por ende se confirma que, sí hay una asociación de las tasas de crecimiento anual del PIB per cápita y el nivel de educación de cada país, dando un estatus diferenciador a la educación en términos países desarrollados de los no desarrollados (Márquez, 2017).

Para poder comparar a Colombia frente al mundo en cuanto a educación, el Programa para la Evaluación Internacional de Alumnos (PISA por sus siglas en inglés) es una prueba que se realiza cada tres años para medir la calidad de la educación a países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y también a países invitados. La metodología utilizada es aplicar dichas pruebas a estudiantes de 15 años de edad sin determinar el grado de escolaridad que se encuentren, se evalúa el desempeño en las áreas de lectura, matemáticas y ciencias (ICFES, 2015). En el año 2016 la OCDE (2016) publicó un informe de los resultados de las pruebas PISA aplicadas en el año 2015, en el que Singapur ocupó el primer puesto de 70 países, para ese año Colombia en el área

de Ciencias ocupó el puesto 57, con una media de 416; en el módulo de comprensión lectora se situó en el puesto 54, con una media de 425 y en el área de matemáticas se ubicó en el puesto 61 cuya media fue de 390, lo cual está dentro de promedio.

En cuanto a Colombia el Instituto Colombiano para el Fomento de la Educación Superior (ICFES) es el organismo encargado de obtener la medición de la educación media y la educación superior en Colombia, por medio de las pruebas de estado que se aplican a los grados 3, 5, 9 y 11, en el que en el último es de obligatoriedad para obtener grado de bachiller, en cuanto la educación superior se aplican las pruebas Saber Pro, las cuales son requisito para obtener el título de pregrado desde el 2009 (ICFES, 2018a, 2018c). De esta manera, las pruebas evalúan cinco módulos las competencias genéricas que son: lectura crítica, razonamiento cuantitativo, competencias ciudadanas, comunicación escrita e inglés, además de otros 40 módulos de temáticas y contenidos específicos de acuerdo al área de formación profesional; los resultados son obtenidos a nivel individual y se puede filtrar por entidad educativa, ciudad, departamento, pregrado, entre otras (ICFES, 2018a, 2018b; 2018 d).

La revista dinero (2017; 2018) en el 2017 comenzó a generar un ranking de las universidades, teniendo en cuenta los resultados de las pruebas Saber Pro, tomando como referencia las pruebas de competencias específicas y genéricas y también por programas de pregrado; para el 2017 la Universidad Pontificia Bolivariana (UPB) Bucaramanga en los programas de ingeniería se ubicó en el puesto 36; en Santander la Universidad Industrial de Santander (UIS) lidera seguida por la UPB Bucaramanga que comparte el segundo con la Universidad Autónoma de Bucaramanga (UNAB); ambas con un puntaje promedio de 164. Para el año 2018 la UPB Bucaramanga ocupó el puesto 28 siendo así la tercera en el

departamento de Santander por debajo de la UNAB y UIS; finalmente, según la resolución 018735 [Ministerio de Educación Nacional] la UPB Bucaramanga se ubicó en el puesto 147 de 151 universidades a nivel nacional y puntuó en último lugar a nivel de Santander

Teniendo en cuenta la relevancia de la educación en la economía de los países y los puntajes mostrados con anterioridad, es necesario profundizar qué factores afectan el rendimiento académico y por tanto los resultados en las pruebas de estado, se ha encontrado que algunos de ellos son el tipo de aprendizaje, el origen sociocultural, nivel educativo de la madre y padre, dinámicas familiares, variables sociodemográficas, género, edad, estado civil, financiación de estudios, pertenecía a grupos minoritarios, clima educativo, grado de responsabilidad, técnicas de estudios, automotivación, ambiente educativo, entre otras (Avendaño et al., 2016; Atundanga, 2008; Barrios & Frios, 2016; Monroy & Hernández, 2014; Olave et al., 2016; Ruiz, 2001). Adicionalmente, en Colombia, Sánchez y Otero (2012) evidenciaron que el mayor puntaje de las pruebas de estados ICFES 11 la obtuvieron en la mayoría los estudiantes que pertenecen a estratos altos, mientras los estudiantes de estratos bajos obtuvieron puntajes bajos en dichas pruebas.

En otro sentido, cabe mencionar los autores (Ruiz. C, 2001; Duran et al. (2017); Cortes y Acadio (2015); Cañon et al. (2009); Ramírez. C (2014) en los artículos de investigación dedicados a los diversos factores, motivos y causas que determinan el rendimiento académico en los estudiantes, basados en los resultados de las pruebas de estado (Saber 11 y Saber Pro), Concuerdan con que las variables sociodemográficas y socioeconómicas, tales como: edad, genero, estrato socioeconómico, apoyo económico familiar, nivel de educación de los padres, pertenece a algún grupo minoritario étnico y entre otras más, son

variables que tienen afectación positiva para un bajo desempeño académico. Es por esto que para las instituciones de educación superior se hace importante tener resultados sobresalientes en las pruebas Saber Pro, ya que estos exámenes reflejan una buena calidad en la educación de las universidades. Actualmente la universidad Pontificia Bolivariana seccional Bucaramanga los programas académicos de ingeniería Industrial y civil se encuentran acreditados con educación de alta calidad por parte del Ministerio de Educación Nacional. A partir de ello y dado al bajo rendimiento en los resultados que se han obtenido en las últimas pruebas la UPB Bucaramanga, se hace necesario determinar cuáles variables socioeconómicas y demográficas tienen impacto como predictores del rendimiento de los estudiantes en las pruebas Saber Pro; es por ello, que aparece la necesidad de preguntarse ¿cómo implementar un modelo de predicción que permita predecir la influencia de variables socioeconómicas y demográficas en los resultados del módulo de competencias genéricas de las pruebas Saber Pro, de los estudiantes de la ingeniería industrial e ingeniería civil de la Universidad Pontificia Bolivariana seccional Bucaramanga?

3. Antecedentes

Para las Instituciones de Educación Superior (IES), el poder saber cuáles son las razones que afectan el rendimiento académico de sus estudiantes es de importancia, ya que, identificando los factores de afectación, se podría estudiar, diseñar y aplicar algún plan de mejoramiento. En el campo de la estadística y de la minería de datos educativa, existen

técnicas de predicción que sirven de ayuda para saber qué características son las que afectan o simplemente darle otro tema de investigación.

Se ha estudiado el impacto de variables sociodemográficas en diferentes etapas escolares, Leiva y Valdés (2016), desarrollaron un modelo para predecir el puntaje de desarrollo psicomotor de los niños de 4 y 5 años en función del nivel socioeconómico, para ello recolectaron una muestra de 228 niños a quienes se aplicó una encuesta sociodemográfica, se le realizó un análisis correlacional bivariado no paramétrico y pruebas de ANOVA y al final un análisis de regresión múltiple, los resultados del artículo arrojaron 3 modelos de predicción: el primer modelo; permitió predecir un 4,2% del puntaje en función de la variable ingreso per cápita, el segundo predijo un 11% del puntaje de lenguaje junto con la variable ingreso per cápita y el último modelo que predijo 2,2% del puntaje coordinación con la variable nivel de educación del jefe de hogar.

Por otra parte, Ramaswami y Bhaskaran (2010) realizaron una investigación, donde llevaron a cabo la técnica de árboles de clasificación con el modelo de algoritmo CHAID, para identificar la relación entre varias sociodemográficas y el rendimiento académico, en este sentido escogieron una población de dos escuelas en la India, una privada y la otra pública, para un total de una muestra de 224 estudiantes de secundaria. El modelo de predicción obtuvo una precisión de 44.69%, en la que también evaluó correlaciones positivas entre factores como escuela, educación parental, condición económica, compromiso de los padres, hábitos alimenticios y tutoría privada con el alto o bajo rendimiento académico.

Una de las técnicas de predicción usadas es los árboles de decisión, es así como, Márquez, Romero y Ventura (2012), llevaron a cabo una investigación, para reconocer los

factores que motivan a los estudiantes al abandono de sus estudios, así como para predecir o detectar a los estudiantes que presenten dichos factores, para reducir la deserción escolar. Ante la creación de varios modelos de algoritmos de clasificación, los factores más comunes entre los modelos son: las notas de asignaturas del semestre y la inasistencia a las clases de: física, humanidades, matemáticas e inglés; en otros modelos aparecieron la edad (mayores de 15), tener hermanos (uno), el grupo al que asistió, vivir en una ciudad de gran población, y considerar que la materia más difícil es matemáticas.

Sin embargo, no sólo las condiciones sociodemográficas influyen en el desempeño académico, Flores, Sánchez y Martínez (2016), desarrollaron un estudio descriptivo, predictivo y transversal a estudiantes que cursaron materias de primero y segundo años en la carrera de medicina, según el desempeño del docente. Primero, realizaron un análisis factorial exploratorio y confirmatorio, para así obtener una estructura unidimensional. Y por último realizaron un análisis de regresión entre las dimensiones obtenida de las variables independientes, sobre el rendimiento académico del desempeño docente, y así saber su impacto predictivo en el rendimiento académico de los estudiantes. El desarrollo de los modelos predijo el 41% y los factores que tuvieron mayor peso de predicción: Psicopedagógica (18%), comunicación y evaluación (14%), intervención (10%) y humanística (9%).

Un estudio más a fondo en el que se emplearon varias técnicas de predicción, y que luego pudieron clasificar cuales fueron la más “predecibles” lo realizaron Visbal, Mendoza y Orejuela (2017) el principal objetivo del artículo era realizar un análisis de la eficiencia técnica de las universidades públicas de Colombia, mediante un análisis envolvente de datos y posteriormente con técnicas de minería de datos (Árbol de clasificación, Vecino

más próximo, Random forest y máquina de soporte vertical (SVM)), y así, determinar las variables que mejor clasifiquen a las universidades, de este modo, la validación del poder predictivo de cada modelo estudiado. Así mismo, los resultados permitieron resaltar cuál de los modelos tiene un mayor poder predictivo, tal es el caso, que los ordenaron de mayor poder a menor predictivo: SVM clasificó correctamente el 75%, árbol de decisión ordenó correctamente de 15 a 24 observaciones (62,5%) y vecino más próximo clasificó 4 a 7 (57,14%) observaciones correctamente (57,14%). Ya por otro lado la técnica de random forest permitió ver variables de mayor importancia al momento de predecir la eficiencia de una IES que son: Números de alumnos matriculados en pregrado, número de alumnos matriculados en postgrados y número de estudiantes con resultados en las pruebas Saber Pro.

Respecto a las pruebas de estado de la calidad de la educación, Pereira, Hernández y Gómez (2011), desarrollaron un artículo en el que planteaban si las pruebas de estado SABER 11 pueden predecir el rendimiento académico del estudiante al ingresar a un programa de educación superior. Con una muestra de 2888 estudiantes, se analizó para cada uno de ellos las variables estudio y, por consiguiente, realizaron la técnica de correlación y pruebas de casualidad, como también una metodología de modelos multivariados lineales y no lineales. El análisis del artículo concluyó que las universidades no se deben basar únicamente en las pruebas de estado como método de ingreso a los candidatos, ya que, este tiene un bajo valor predictivo ante el desempeño de los estudiantes en la universidad. Con respecto a la prueba SABER11, obtuvieron un valor predictivo de las materias de biología, geografía, física, historia, química y lenguaje, con valores significativos bajos. Determinaron que las universidades que utilizan la pruebas SABER 11

como requisito de inscripción, no es confiable decir que un buen puntaje en las pruebas define un perfil que obtenga buenas notas en la carrera universitaria.

En cuanto a los modelos predictivos aplicados a la educación superior y más aún a las pruebas Saber Pro. García y Valbuena (2016), efectuaron un artículo en el que trabajaron las pruebas Saber Pro-2011-2 a 2013-2 de la facultad de psicología en todas las IES del país. Preliminarmente iniciaron con caracterizar a 17.733 estudiantes de los programas de psicología del país y los estudiantes en cuestión de variables sociodemográficas, económicas y de calidad del programa. Luego un análisis factorial de correspondencias que les permitió la identificación de perfiles de estudiantes e instituciones que se relacionan con el desempeño de los estudiantes, en los módulos genéricos de las pruebas, en cuanto a las variables de caracterización de los estudiantes y las instituciones. Y por último un modelo predictivo donde tres variables, han sido explicadas entre el 10% y el 34% de la varianza de instituciones, los autores concluyen que, con el ánimo de aumentar la varianza explicada en los dos niveles, se debe de probar distintas combinaciones y agregar más variables.

4. JUSTIFICACIÓN

La UPB, realiza para sus estudiantes de últimos semestres próximos a presentar las pruebas de estado como requisito obligatorio el programa reto al saber, es un programa para el fortalecimiento académico que ayuda a desarrollar las capacidades humanas y competencias de cada estudiante. Dicho programa se divide en 3 etapas: básico,

acompañamiento para estudiantes durante los dos primeros semestres; ciclos, son evaluaciones de ciclo al lograr cursado el 70% de los ciclos de la carrera (ciclo básico, formación humanística y profesional) y la última etapa Pro, se enfatiza en los estudiantes que presentan las pruebas Saber Pro (UPB, 2016). Por otro lado, la universidad brinda incentivos, beneficios como descuentos a carreras de posgrado, exención derechos de grado de pregrado, exención de preparatorios derecho y reconocimientos académicos, para aquellos estudiantes que logren un buen puntaje en las pruebas.

Esta iniciativa de fortalecer competencias y brindar incentivos que realiza la UPB hacia sus estudiantes, nace como estrategia para hacer frente a la problemática del bajo rendimiento que obtuvieron los estudiantes de la UPB Bucaramanga durante los últimos años en las pruebas de estado para la calidad de la educación superior (Resolución Número 018735 del 10 de diciembre de 2018. [Ministerio de Educación Nacional]). Por consiguiente, una vez realizada la revisión de antecedentes enfocada a los proyectos de modelos de predicción hacia la educación, en el que se puede observar distintas técnicas para predecir y en el que también se observaron factores de mayor común denominador, como son sociodemográficos, socioeconómicos, rendimiento docente y psicosociales. Surge la necesidad de revisar los factores de influencia en la UPB en mira del rendimiento académico evaluado en la prueba de estado, mediante la realización de todo un modelo de predicción que permita saber cuáles son los posibles resultados de las pruebas, determinar diferentes perfiles de estudiantes dado por ciertas variables de afectación hacia su rendimiento académico durante la duración de la carrera. Del mismo modo que ayude a la UPB y a las demás universidades, a tener el conocimiento de los factores que afecten a su estudiantado, y así, poder focalizar esfuerzo en pro de fortalecer y mejorar el rendimiento

académico del estudiante, y posicionar a la Universidad como institución representativa en la educativa superior.

El alcance del proyecto es de manera correlacional-explicativo, ya que no solamente permitirá predecir los posibles resultados que obtengan los estudiantes de ingeniería industrial y civil de la UPB Bucaramanga en el periodo de tiempo 2016 -2018 de acuerdo con variables de afectación. Y que permitirá en el mismo sentido brindar información sobre los factores que pueden afectar a los estudiantes e incentivar a futuras investigaciones, con más modelos y técnicas de predicción, para que finalmente se pueda confiar en el modelo de mayor poder de predicción.

5. Objetivos

5.1 Objetivo general.

Implementar el modelo de predicción CHAID a partir de los resultados de las competencias genéricas de las pruebas Saber Pro en el periodo 2016 a 2018 que ayude a predecir los posibles resultados de los estudiantes de ingeniería industrial y civil de la Universidad Pontificia Bolivariana seccional Bucaramanga.

5.2 Objetivos específicos.

- Construir y preparar la base de datos Saber Pro de 2016 a 2018, con las variables a analizar.
- Caracterizar los perfiles sociodemográficos y socioeconómicos de los estudiantes de ingeniería industrial e ingeniería civil de la Universidad Pontificia Bolivariana seccional Bucaramanga
- Establecer la asociación entre las características sociodemográficos y socioeconómicos de los estudiantes de ingeniería industrial e ingeniería civil de la UPB Bucaramanga y los resultados de los módulos genéricos.
- Validar el modelo predicción CHAID utilizando resultados en las pruebas genéricas del año 2018 como datos de prueba.

6. Marco referencial

6.1 Marco teórico

6.1.1. Pruebas Saber Pro. La organización encargada de velar por la evaluación de la educación y en la investigación de factores que inciden en la calidad de educativa es el Instituto Colombiano para la Evaluación de la Educación (ICFES), fue creada en 1968 y comenzó con la evaluación del servicio nacional de pruebas (SNP) con esta se validó la admisión a la educación superior (ICFES, 2012). En 2003 el ICFES comenzó a evaluar a la educación superior con el Examen de Estado de Calidad de la Educación Superior (ECAES), en que se evaluó 22 programas de educación superior y luego, en el 2007 aumentó a 55 programas; durante este periodo la prueba media conocimientos únicamente las competencias específicas por programa y en ese periodo las pruebas no eran un requisito obligatorio para graduarse en la educación superior (ICFES, 2016); para el 2009 entró en vigencia la evaluación de las competencias genéricas que estaban conformadas por pensamiento crítico, solución de problemas, entendimiento interpersonal y escritura. Luego al año siguiente, el ICFES empezó con un nuevo diseño y aplicación de un modelo de examen conocido actualmente como Saber Pro, en el que, en este nuevo módulo de competencias genéricas busca evaluar el desarrollo de las competencias comunes, dicho modulo está conformado: lectura crítica, razonamiento cuantitativo, comunicación escrita, inglés y competencias ciudadanas, y cuya descripción de lo que evalúa cada módulo se refleja en la figura 1.

Competencias genéricas	Descripción
Competencias Ciudadanas	Evalúa los conocimientos y habilidades que posibilitan la construcción de marcos de comprensión del entorno, que promueven el ejercicio de la ciudadanía y la coexistencia inclusiva según la Constitución Política de Colombia. Un ciudadano competente es aquel que conoce su entorno social y político; tiene presentes sus derechos y deberes; reflexiona sobre problemáticas sociales; busca el bienestar de su comunidad; entre otros.
Comunicación Escrita	Evalúa la competencia para comunicar ideas por escrito referidas a un tema dado. Los temas son de dominio público y no requieren conocimientos especializados.
Inglés	Evalúa la competencia para comunicarse efectivamente en inglés. Esta competencia, alineada con el Marco Común Europeo, permite clasificar a los examinados según su nivel de desempeño A-, A1, A2, B1, B2.
Lectura Crítica	Evalúa las capacidades de entender, interpretar y evaluar textos que pueden encontrarse tanto en la vida cotidiana, como en ámbitos académicos no especializados. Pretende indagar sobre la comprensión lectora del evaluado que le permita interpretar, aprender y tomar posturas críticas frente a un texto, aunque no cuente con un conocimiento previo del tema tratado.
Razonamiento Cuantitativo	Evalúa las habilidades para: comprender y transformar información cuantitativa presentada en distintos formatos; formular y ejecutar planes para dar solución a problemas que involucran información cuantitativa u objetos matemáticos; y justificar, argumentar o dar validez a procedimientos y estrategias matemáticas utilizadas para dar solución a problemas planteados.

Figura 1 Descripción de los módulos de competencias genéricas. ICFES (2017)

Las personas que presentan la evaluación externa de la calidad de la educación superior Saber Pro, son estudiantes correspondientes a un programa académico vinculados a una IES y por otro lado los egresados, que deciden volver a presentar la prueba de

manera independiente. (ICFES 2017). Así mismo quienes presenten las pruebas serán evaluados con los siguientes puntajes como se ve en el cuadro tabla 1 y que según el ICFES clasifica por desempeño cada una de las pruebas como se aprecia en la tabla 2.

Tabla 1. *Escala de puntajes a partir del 2016.*

Prueba	Media	Desviación		
		estándar	Mínimo	Máximo
Razonamiento				
Cuantitativo	150	30	0	300
Lectura crítica	150	30	0	300
Comunicación				
escrita	150	30	0	300
Inglés	150	30	0	300
Competencias				
ciudadanas	150	30	0	300
Puntaje global	150	30	0	300

Nota. Fuente: ICFES (2017)

Tabla 2. *Nivel de desempeño de los puntajes*

Módulo	Nivel Desempeño 0	Nivel de Desempeño 1	Nivel de Desempeño 2	Nivel de Desempeño 3	Nivel de Desempeño 4
Comunicación escrita	No aplica	0 - 125	126 - 155	156 - 190	191 - 300

Razonamiento cuantitativo	No aplica	0 - 125	126 - 155	156 - 190	191 - 300
Lectura crítica	No aplica	0 - 125	126 - 155	156 - 190	191 - 300
Competencias ciudadanas	No aplica	0 - 125	126 - 155	156 - 190	191 - 300
Inglés	(-A1)	A1	A2	B1	B2

Nota. Fuente: ICFES (2018)

6.1.2. Análisis de correspondencia (AC). Es un método eficaz para analizar las tablas de contingencias con datos de frecuencias numérica, porque representa de manera gráfica la interpretación de los datos de una manera simple. Este análisis fue ideado en los años 60 por el matemático Jean Paule Benzecri junto con su equipo de trabajo, en una universidad de Paris (Greenacre, 2008). El AC como explica Peña (2013) se interpreta de dos formas, la primera, es una forma de representar variables en una dimensión menor, a componentes principales permitiendo así definir la distancia entre el punto de manera clara, para interpretar los datos con la distancia de Ji-cuadrado; la segunda manera es asignar valores numéricos a variables cualitativas.

El análisis de correspondencia es un método multivariante que permite aplicarse a variables cualitativas de la Fuente (2011) y tiene como objetivo resumir una gran cantidad de datos, por medio, de la reducción de dimensiones y permitiendo que no exista gran cantidad de datos perdidos, también, permite estudiar las posibles relaciones existentes entre variables permitiendo caracterizar un perfil con los factores comunes del análisis (Lebart, Morineau, & Piron, 1995)

6.1.2.1 Análisis de correspondencia simple (ACS). Es un método en el que explica la relación entre dos variables cualitativas, por medio de representaciones, a partir de su tabla de contingencia, en otras palabras, la distribución de sus frecuencias contenidas por categorías (Saavedra 2012). Este análisis permite analizar la relación de las variables de manera gráfica en un espacio dimensional, en el que las categorías de cada variable, esta gráficamente más próximas o más alejadas en las todas las dimensiones en función del grado de similitud o diferencias (Visauta & Martori, 2003).

6.1.2.2 Análisis de Correspondencia múltiple (ACM). El ACM al igual que ACS, se utilizan para la reducción de datos y así mismo poder caracterizar variables con ciertas categorías, mediante la representación de dichas variables. La diferencia de estos dos métodos es que el ACM se desarrolla con dos o más variables cualitativas de tipo ordinal y nominal (Aránzazu et al,2007), otra diferencia se encuentra en que según Lebart, Morineau y Piron (1995, p. 41):

El análisis de correspondencia aplicados a tres tipos de tablas (tabla de contingencia, matriz disyuntiva y matriz Burt) dan los mismo ejes factoriales, pero con diferentes valores propios y por tanto diferentes tasas de inercia, la diferencia es que la tabla de contingencia dará la tasa de inercia más alta.

En decir, el ACM resume un espacio de propiedades que generan nuevas variables-resumen que son denominadas factores que revelan la diferencias entre los sujetos de estudio, lo cual transforma los datos en gráficos para que permitan ver las distancias entre

modalidades y sujetos de estudio en los espacios originales. Es entonces, que las semejanzas entre los sujetos de estudio aparecen en la siguiente dimensión y como consecuencia las modalidades en el espacio de los individuos (Algañaraz, 2016). El análisis de correspondencia se lleva por medio de **Matriz de Burt**, que es una estructura alternativa para los datos de ACM, “es una matriz compuesta por todas las tablas resultantes de cruzar todas las variables de interés dos a dos” (Greenacre, 2008, p.189). las tablas diagonales corresponden a los cruces de cada una de las variables analizadas que contienen frecuencias marginales y fuera de ella todas las combinaciones de 2 a 2 de las frecuencias cruzadas (de la Fuente, 2011), el modelo de Burt se basa en:

$$B = Z'Z$$

Donde B: la tabla de contingencia Burt y la Z: es la representación de una tabla disyuntiva completa. (Lebart, Morineau. & Piron, 1995)

6.1.3. Árbol de decisión. Los árboles de decisión o de clasificación, es una técnica de clasificación, para desarrollar algoritmos de predicción para una variable explicativa (Song. & Lu, 2015). Y en el que consiste en “buscar una variable independiente x que mejor explique a la variable dependiente y define una primera división de la muestra en dos subconjuntos, el procedimiento repite para los dos segmentos buscando una segunda mejor variable y así sucesivamente”. (Pérez, 2015, p.285-286). La estructura de un árbol está conformada por (Song, & Lu, 2015):

- **Nodos:** el árbol tiene tres nodos nodo raíz, nodo interno y nodo de hoja, el primero también conocido como nodo de decisión representa una opción que viene de la

subdivisión de los registros de dos o más subconjuntos; el segundo también llamado nodo aleatorio representa una de las opciones disponible; y el último, también conocido como nodo final, es en el que reposa la decisión final.

- Ramas: representan resultados aleatorios o las apariciones que resultan de los nodos
- Reglas de parada: Detiene la creación de nuevas divisiones, evitando que el modelo se vuelva complejo. Algunos parámetros de ayuda para evitar que suceda son: número mínimo de registros en una hoja, el número mínimo de dividir un nodo, el número de pasos o profundidad.
- Poda: antes de realizar una poda de árbol, es recomendable que se realice en dos fases, la primera es dejar desarrollar el árbol que contenga cientos de ramas y la segunda fase es realizar la poda eliminando los nodos o ramas innecesarias

Las características de los árboles de decisión es que las variables de entrada del son variables independientes (predictivas) que pueden ser categóricas o numéricas, mientras que la variable de salida (objetivo) es dependiente; los árboles se clasifican en binarios y categóricos, los binarios significan que únicamente tendrán como respuesta dos ramas, mientras el categórico tendrá como ramas finales dos o más ramas (Milanovic & Stamenkovic, 2016).

6.1.3.1 Árboles CHAID. El método Chi-square Automatic Interaction Detector (CHAID), desarrollado por G. V. Kass, es un método de tipo no binario, útil para la identificación de variables importantes e interacciones enfocada en la segmentación y análisis descriptivos (Pérez, 2015). El criterio estadístico para la ramificación es la prueba de chi- cuadrado, y se usa variables cualitativas para determinar la independencia de la tabla dinámica, en otras palabras, la combinación de categorías de una variable

dependiente y un predictor para caso de variable (Pawliczek, Kozen, & Villanova, 2018). Las variables cuantitativas, se recurre al método de análisis de varianza, evaluando el nodo que tenga el par de categorías con el menor valor de chi-cuadrado, si el nivel de significancia es menor que el de otro nivel crítico se unen ambas variables y si es mayor se convierten en candidatas para la división de la variable, evaluando así en cada par de categorías (Pérez, 2015).

Los árboles de decisión tipo CHAID procede en pasos, primero se encuentra la mejor partición para cada variable predictora. Luego las variables predictoras son comparadas y es escogido el mejor de ellos, los datos son subdivididos de acuerdo con los predictores elegidos y cada uno de estos subgrupos son reanalizados independiente para producir otras subdivisiones parra el análisis Kass (1980).

El tipo de cada predictor determina la permisividad de las agrupaciones de esas categorías, para construir la tabla de contingencia con los niveles de significancia más altos de acuerdo con la prueba de chi-cuadrado. Esto implica que estos no son suficientes observaciones para asegurar la validez de la prueba. Si esto no es el caso luego algún otro criterio puede ser usado, tal como test de Fisher Ouyang. J, Patel. N & Sethi. I (2011).

La manera cómo funciona el algoritmo CHAID está dado por la siguiente ecuación:

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}}$$

Donde n_{ij} : Valor que asume el predictor y el criterio ij

\widehat{m}_{ij} es la frecuencia esperada bajo la hipótesis nula

El valor de significancia P esta dado por $p = \Pr (N_d^2 > X^2)$ donde N_d^2 sigue la distribución Chi-cuadrado con grados de libertad $d = (J - 1) (I - 1)$ y Pr es el valor de probabilidad observado entre el predictor y la variable dependiente.

Kass, (1980) asume que el algoritmo se puede realizar en 5 pasos:

1. Realice una tabla cruzada con cada predictor y la variable dependiente.
2. Encuentra el par de categorías del predictor donde $2 \times d$ (d es el número de clases) es menor significativamente diferente.
3. Para cada categoría combinada compuesta de tres o más categorías originales, encuentre la división binaria más significativa en el que se puede resolver la combinación.
4. Calcule la significancia de chi-cuadrado de cada combinación de los predictores y se toma el más significativo. Si esta significancia es mayor al valor permitido, subdivida los datos de acuerdo con las categorías (fusionadas del predictor elegido).
5. Para cada partición de los datos que no han sido analizados aun regrese al paso 1 (p,121).

6.1.3.2 Arboles CART. Classification and regression trees (CART), fue ideado por Leo Breiman y otros en 1984, el modelo CART, este modelo nació para solventar las debilidades del modelo CHAID y posee una estadística más robusta razón por la cual es utilizado en campos de la investigación de medicina y el marketing (Pérez, 2015). es un

método en el que puede manejar variables numéricas o categóricas y se caracteriza por la invarianza en la estructura de los árboles de clasificación (Díaz, 2012). Este método se desarrolla al dividir los subconjuntos para evaluar los predictores y así encontrar las mejores agrupaciones de categorías, seleccionándose el predictor y la división que produzca mayor bondad de ajuste. Por otro lado, para los predictores cuantitativos se utilizan la minimización del error cuadrático o de la desviación media absoluta respecto a la mediana y en los predictores cualitativos se utiliza el coeficiente Gini. (Pérez, 2015).

Por último, Song y Lu (2015), llevaron a cabo la comparación entre los tipos de árboles de decisión. Esta comparación refleja (tabla 3) cada una de las características de cada modelo.

Tabla 3. *Comparación modelos CART y CHAID.*

MÉTODOS	CART	CHAID
Medida utilizada para seleccionar variable entrada	Índice de Gini; criterios de dosificación.	Chi-cuadrado.
Poda	Prepoda usando un algoritmo de un solo paso.	Pre-poda usando la prueba de Chi-cuadrado para independencia.
Variable dependiente	Categorico/ continuo.	Categorica
Variable de entrada	Categorico/ continuo.	Categorico/ continuo.
Dividir en cada nodo	Binario; Dividir en combinaciones lineales	Múltiple

Nota: Fuente: Song. Y., Lu. Y (2015)

7. Diseño metodológico

7.1 Alcance metodológico

El alcance del proyecto es de manera correlacional, ya que no solamente permitirá predecir los posibles resultados que obtengan los estudiantes de ingeniería industrial de la UPB Bucaramanga en la prueba Saber Pro que se presentará en el año 2019; en el modelo de predicción a implementar se utilizará las bases datos del ICFES de los periodos comprendidos entre el año 2016 a 2018 de acuerdo con variables de afectación y que permitirá en el mismo sentido brindar información sobre los factores que pueden afectar a los estudiantes e incentivar a futuras investigaciones, con más modelos y técnicas de predicción. Finalmente, el modelo implementado se validará teniendo como base de comparación los resultados de las pruebas Saber Pro del 2018 y de esta forma conocer la brecha entre los resultados reales y los del modelo de predicción.

7.2 Definición de la población.

7.2.1. Unidad de investigación. La población objeto para el desarrollo del proyecto son los estudiantes de ingeniería industrial que presentaron las pruebas Saber Pro en el periodo 2012 a 2018 de la Universidad Pontificia Bolivariana seccional Bucaramanga.

7.2.2. Tamaño de la población. La población, la cual se obtuvo en los repositorios de las bases de datos de los resultados de las pruebas Saber Pro, se logró tener una muestra de 1119 estudiantes de ingeniería industrial de la Universidad Pontificia Bolivariana seccional Bucaramanga que presentaron las pruebas en el periodo 2016-2018, los cuales están distribuidos en la tabla 4.

Tabla 4. *Muestra de estudiantes*

Año	Cantidad estudiantes ingeniería industrial	Cantidad estudiantes ingeniería civil
2016	160	211
2017	157	231
2018	141	187
Total	458	429

Nota. Cantidad de estudiantes de ingeniería industrial y civil de la UPB seccional Bucaramanga que presentaron las pruebas en los años 2016, 2017 y 2018

7.3 Metodología.

El presente proyecto es de enfoque cuantitativo por cuanto se hará uso de variables categóricas, de tipo transversal puesto que se utilizará resultados obtenidos en un rango de tres años y de alcance correlacional ya que permitirá conocer si existe relación entre las variables (Hernández, Fernandez, & Baptista, 2014).

7.3.1. Procedimiento. Para dar cumplimiento al objetivo general de implementar un modelo de predicción CHAID a partir de los resultados de las competencias genéricas de las pruebas Saber Pro en el periodo 2016 a 2018 que ayude a predecir los posibles resultados de los estudiantes de ingeniería industrial de la universidad pontificia seccional Bucaramanga; se planteó un procedimiento que consta de tres fases, los cuales a partir de ciertas actividades responden a cada objetivo específico.

Fase I. Preparación y construcción de la base de datos. La fase I da cumplimiento al objetivo uno de construir y preparar la base de datos Saber Pro de 2016 a 2018, con las variables a analizar y al segundo objetivo con el que se pretende caracterizar los perfiles sociodemográficos y socioeconómicos de los estudiantes de ingeniería industrial de la universidad pontificia bolivariana seccional Bucaramanga, para ello se realizará las actividades siguientes actividades:

- Las bases de datos se encuentran en el repositorio para investigadores de la página de ICFES, dichas variables al descargar se encuentran en lectura de texto tipo código, para que se pueda realizar en cualquier software de análisis de datos. Por consiguiente, en primer lugar, se procederá a trasladar las bases de datos en formato texto, al programa de Excel para la disposición de futuros análisis.
- En segundo lugar, se revisará la información relacionada con los posibles factores sociodemográficos, socioeconómicos y personales que influyan en el rendimiento académico estudiantil.

- Por último, se llevará a cabo una tabla de frecuencias para tener conocimiento de aquellas variables que carecen de información y así seleccionar las variables que sirvan para el estudio.

En respuesta al objetivo tres de establecer la asociación entre las características sociodemográficas y socioeconómicas de los estudiantes de ingeniería industrial de la UPB Bucaramanga y los resultados de los módulos genéricos, se planteó la fase II y III

Fase II. Caracterización de los perfiles. En esta fase se llevará a cabo el análisis de correspondencias múltiples a las variables sociodemográficas, socioeconómicas y de información personal, de manera que permita generar un perfil de los estudiantes que presentaron las pruebas Saber Pro en el periodo de tiempo 2016 a 2018 y a su vez entregando las variables de entrada para la siguiente fase.

Fase III. Implementación de un modelo predicción por medio la técnica arboles de decisión tipo CHAID. La III fase de la investigación va ligada con la creación de un modelo de predicción, que se desarrollará por medio de la técnica de árboles de decisión tipo CHAID, ya que permite trabajar con variables de entrada categoricas que permitirá pronosticar los resultados del módulo de competencias genéricas que está conformado por: Competencias ciudadanas, Comunicación escrita, inglés, lectura crítica y razonamiento cuantitativo, de las pruebas Saber Pro.

Fase IV. Validación del modelo de predicción CHAID. La última fase de la investigación va enlazada con la validación del modelo del árbol de decisión tipo CHAID, para esto se realizará una predicción tomando los periodos 2016-2017 para predecir los resultados de las competencias genéricas de las pruebas Saber Pro-2018. Luego, se

realizará el perfil dado por la predicción con los datos del año 2018. Y se comparará si los perfiles acertaron los resultados reales de ingeniería civil e industrial.

8. Resultados

8.1 Preparación y construcción de la base de datos

En el repositorio del ICFES donde se encuentra la base de datos de los resultados de las pruebas Saber Pro, cada periodo de presentación de las pruebas contiene aproximadamente 105 o más variables de información entre las que el ICFES las clasifica en: información personal, información de contacto, datos académicos SABER PRO, información socioeconómica, datos citación del examen, resultados y variables adicionales. Como primera instancia para la preparación y construcción de la base de datos se tuvo como criterio que en todos los periodos de presentación debían tener el mismo número de variables y que las variables fueran igual para todos los periodos, de esta manera el periodo de años se redujo de 2012-2018 a 2015-2018 porque antiguamente de 2012-2014 las pruebas SABER PRO se presentaban de manera semestral y no anualmente como se lleva a cabo en la actualidad. En 2015 la variable valor de la matrícula del último semestre, las opciones de respuestas estaban acotadas de manera distinta a la de los años 2016, 2017 y 2018 y no hubo manera de relacionarla con otra variable que dijera un valor de impacto económico que tuviera el estudiante; de tal manera que la brecha de tiempo quedo de 2016 a 2018.

En segundo paso se procedió a verificar que las variables que son en común en los tres periodos (2016-2017-2018) y a la eliminación de las variables que no estuvieran en común, como se observa (figura 2) fueron eliminadas en su totalidad variables como: información de la institución educativa la cual pertenecía, si el estudiante sufría o tenía alguna

discapacidad física, motriz u otra, así como también el tipo documento de identificación con el que presentó las pruebas Saber 11, semestre que cursa actualmente y área donde reside, por último las variables socioeconómicas como: si el estudiante tiene videojuegos, con cuantas personas comparte el baño de la casa, la ocupación de los padres no se tomó en cuenta ya que para el último año esta pregunta cambio por la labor de la madre o el padre de familia y dio paso a obtener una respuesta más precisa sobre la ocupación de los padres.

INFORMACIÓN PERSONAL			
	2018	2017	2016
ESTU_LIMITA_MOTRIZ		x	x
ESTU_LIMITA_INVIDENTE		x	x
ESTU_LIMITA_CONDICIONESPECIAL		x	x
ESTU_LIMITA_SORDO		x	x
ESTU_LIMITA_AUTISMO		x	x
ESTU_LIMITA_SORDOCONINTERPRETE			
ESTU_LIMITA_SORDOSININTERPRETE			
ESTU_LIMITA_SORDOCEGUERA			
INFORMACIÓN DE CONTACTO			
ESTU_AREARESIDE	x		x
ESTU_PAGOMATRICULAEXT		x	
ESTU_PAISDOCUMENTOSB11		x	
ESTU_SEMESTRECURSA		x	x
INFORMACIÓN SOCIOECONÓMICA			

FAMI_OCUPACIONPADRE		x	x
FAMI_OCUPACIONMADRE		x	x
FAMI_TRABAJOLABORPADRE	x		
FAMI_TRABAJOLABOR_MADRE	x		
FAMI_TIENECONSOLAVIDEOJUEGOS	x	x	
FAMI_CUANTOSCOMPARTEBAÑO	x	x	

Figura 2. Variables no tomadas en cuenta por no tener trascendencia en los periodos 2015 a 2018

El tercer paso fue verificar que las respuestas de cada una de las variables no estuviesen concentradas en una sola categoría de respuesta, para ello se realizó una tabla de frecuencias de todas las variables y en la figura 3 se muestra aquellas que sus respuestas tenían más del 90 % de concentración en una sola categoría. Las que se eliminaron han sido variables como: Nacionalidad, si pertenece algún grupo étnico, tipo de documento de identificación con el que presentó la prueba, el tipo de preparación que tuvo para las pruebas Saber Pro, información de la IES, datos de citación del examen.

	INGENIERÍA INDUSTRIAL	INGENIERÍA CIVIL
VARIABLES ELIMINADAS POR CONCENTRACION EN LA CATEGORIA NS/NR		
ESTU_CURSODOCENTESIES	92.80%	92.80%
ESTU_CURSOIESAPOYOEXTERNO	92.80%	92.80%
ESTU_CURSOIESEXTERNA	93.10%	93%

ESTU_SIMULACROTIPOICFES	92.60%	92.60%
ESTU_ACTIVIDADREFUERZOAREAS	92.60%	92.60%
ESTU_ACTIVIDADREFUERZOGENERIC	92.60%	92.60%
VARIABLES ELIMINADAS POR CONCENTRACIÓN EN UNA SOLA		
CATEGORIA		
ESTU_TIPODOCUMENTO	Cedula Ciudadanía (100%)	Cedula Ciudadanía (100%)
ESTU_NACIONALIDAD	Colombiana (100%)	Colombiana (99.8%)
ESTU_TIENEETNIA	No (100%)	No (100%)
ESTU_PRIVADO LIBERTAD	Si (100%)	Si (100%)
ESTU_COD_MCPIO_PRESENTACIÓN	Bucaramanga (96.3%)	Bucaramanga (95.7%)
ESTU_COD_DPTO_PRESENTACIÓN	Santander (98.9%)	Santander (97.8%)

Figura 3. Frecuencias de las variables por concentración en respuestas.

Teniendo en cuenta que para el programa SPSS STATISTICS los modelos estadísticos que se plantean trabajar como ACM y Arboles de decisión tipo CHAID se recomienda que los datos de entrada deben estar codificados con números y el análisis que se realice sea preciso. De esta manera lo que se realizó fue recodificar de manera binaria (1:si y 0:no) aquellas variables de tipo sí o no y para el resto de las variables se codificaron con el número 1 en adelante para cada una de sus categorías de respuesta. Por ultimo las respuestas en donde el estudiante no respondió o no sabía la respuesta se codifico con el código 99 indicando (NS/NR) (Anexo1).

Dentro de las variables se tuvo que reestructurar dos de ellas:

EDAD_PRESENTACION y ESTU_COD_RESIDE_MCPIO, la primera reestructuración

consistió en cambiar la variable “Fecha- Mes- Año de nacimiento” a las categorías 20 años o menos, 21 años, 22 años, 23 años, 24 años y 25 años o más, para la segunda variable se hizo un procedimiento similar y se reorganizó priorizando cada uno de los municipios del Área Metropolitana de Bucaramanga (AMB) y en una sola opción aquellos que no lo eran dado que eran menores que los anteriores. Por otro lado, la variable ESTU_CODDANE_COLE_TERMINO, no pudo ser tomada por razón de que tenía más de 100 respuestas y el estudio tendría una dispersión no deseada; en este mismo sentido para no dejar a un lado esta variable se creó una variable similar en donde se utilizó el código DANE de la institución educativa para obtener información de los colegios que permitiera satisfacer esta variable, de esta manera se creó una nueva variable que brinda información del tipo de entidad educativa de donde se graduó el estudiante (si es pública o privada) nombrada: ESTU_COL_PRIV_PUBL.

8.2 Perfil Estudiantes por Facultad.

Con el fin de identificar el perfil de los estudiantes con respecto a variables sociodemográficas y de tenencia de artículos, primero se realizó una graficas de las frecuencias de la tenencia de bienes y servicios en los estudiantes de ingeniería civil e industrial de la Universidad Pontificia Bolivariana de Bucaramanga, con el propósito de identificar un perfil de acuerdo con las variables mencionadas. Luego, se realizó las técnicas de análisis de correspondencias múltiples con las variables sociodemográficas y socioeconómicas.

En la Figura 4 se aprecia la comparación de bienes y servicios que tienen los estudiantes de ingeniería industrial y civil de la UPB Bucaramanga en los periodos 2016, 2017 y 2018. Estas variables no presentan diferencias entre facultades y demuestran las

comodidades con las que cuentan los estudiantes en dicho periodo. En los estudiantes de los periodos estudiados es común contar con lavadora, tv, computador y servicio de internet ya que sus porcentajes de tenencia superan el 90%. Estas variables no han sido ingresadas en el análisis de correspondencias, ya que genera bifurcación de datos para el árbol de decisión.

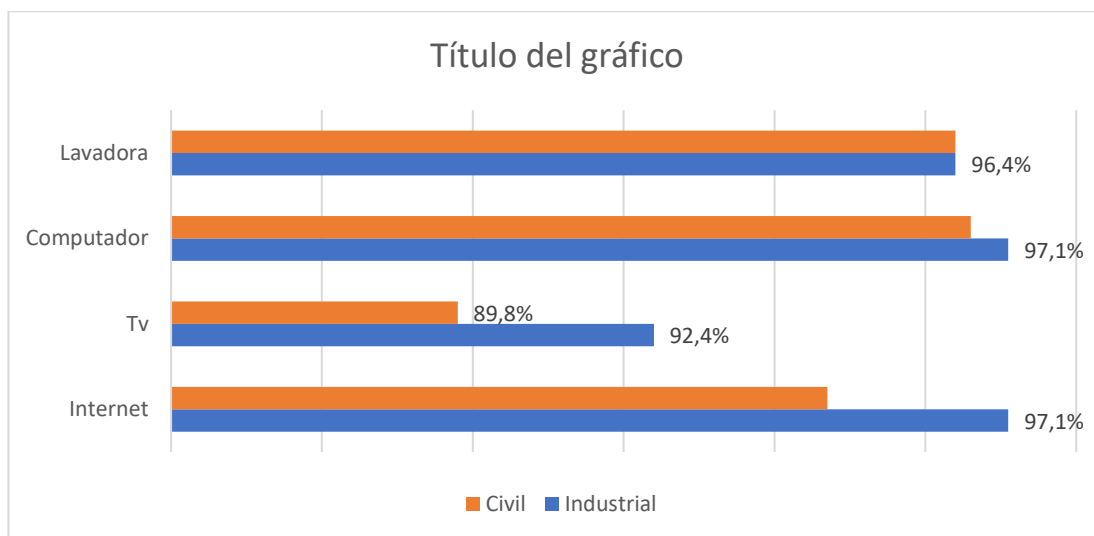


Figura 4. Comparación de tenencia de bienes y servicios según profesión.

8.2.1. Análisis de correspondencia múltiple Variables Sociodemográficas y

socioeconómicas. El propósito del análisis de correspondencia múltiple (ACM), también conocido como análisis de homogeneidad, es encontrar cuantificaciones que sean óptimas en el sentido de que las categorías están separadas entre sí tanto como sea posible. Esto implica que los objetos en la misma categoría se trazan cerca uno del otro y los objetos en diferentes categorías se trazan lo más separados posible (Algarañaz,2016). El término homogeneidad también se refiere al hecho de que el análisis será más exitoso cuando las variables sean homogéneas; es decir, cuando dividen los objetos en grupos con las mismas categorías o similares.

Para el presente estudio, se introdujeron las siguientes variables para el ACM: Género, Edad, Periodo, municipio, título bachiller, tipo de colegio, tipo de hogar, Número de personas a cargo, educación padre y madre, estrato, personas en el hogar y número de cuartos. Además de facultad y los diferentes tipos de desempeño: razonamiento cuantitativo, lectura crítica, competencia ciudadana, aprendizaje inglés y comunicación escrita.

8.2.2. Resumen Modelo. El análisis de homogeneidad puede calcular una solución para varias dimensiones, para nuestro caso dos.

Según Oviedo & Arias (2005), establecen que valores de alfa Cronbach entre 0,7 y 0,9 son valores buenos, por lo que quiere decir que para este modelo se encontraron valores de alfa Cronbach aceptables en la primera dimensión (80,9%) y en la segunda dimensión (74,6%), indicadores que estas dos dimensiones explican de buena manera las variables consideradas en el modelo. Ambas dimensiones explican de manera similar la variación en los datos, el (55,8%) por la primera dimensión y (44,1%) por la segunda (tabla 5)

Tabla 5. *Resumen del modelo ACM Variables sociodemográficas y económicas.*

Dimensión	Alfa de Cronbach	Total (autovalor)	Varianza explicada
1	,809	4,361	0,5582
2	,746	3,451	0,4418
Total		7,812	1,000
Media	,781 ^a	3,906	0,500

a. La media de alfa de Cronbach se basa en la media de autovalor.

Para determinar las medidas de discriminación se requiere calcular para cada variable, una carga de componentes al cuadrado, para cada dimensión. Esta medida es también la varianza de la variable cuantificada en esa dimensión, tiene un valor máximo de 1, que se logra si las puntuaciones de los objetos caen en grupos mutuamente excluyentes y todas las puntuaciones de los objetos dentro de una categoría son idénticas (de la Fuente, 2011). Las medidas de discriminación grandes corresponden a un gran aporte a la dimensión correspondiente, cuando las variables están cerca del origen, dichas variables no aportan.

La gráfica de medidas de discriminación (figura 5) muestra que la dimensión uno (eje y) está influenciada en primer momento por el desempeño del inglés, además, de educación, tanto de la madre como del padre y el estrato, esto debido a que son las que tienen mayores cargas de ahí que se colocó el nombre de educación estando bien alejadas del origen. En la dimensión 2 (eje X), sobresalen las variables de tenencia como son: tener horno microondas o gas y auto, además, que aportan también el estrato y la educación de la madre.

De otro lado, las variables cercanas al origen no discriminan en nada o muy poco, entre ellas, sobre salen: el tipo de ingeniería identificada en la variable facultad, el género, la edad y el número de personas a cargo, no aportan al desempeño de las competencias ya mencionadas no a la dimensión capacitación de los padres. El título de bachiller, municipio y periodo aportan un poco a la capacitación de padres y prácticamente nada al rendimiento de las competencias.

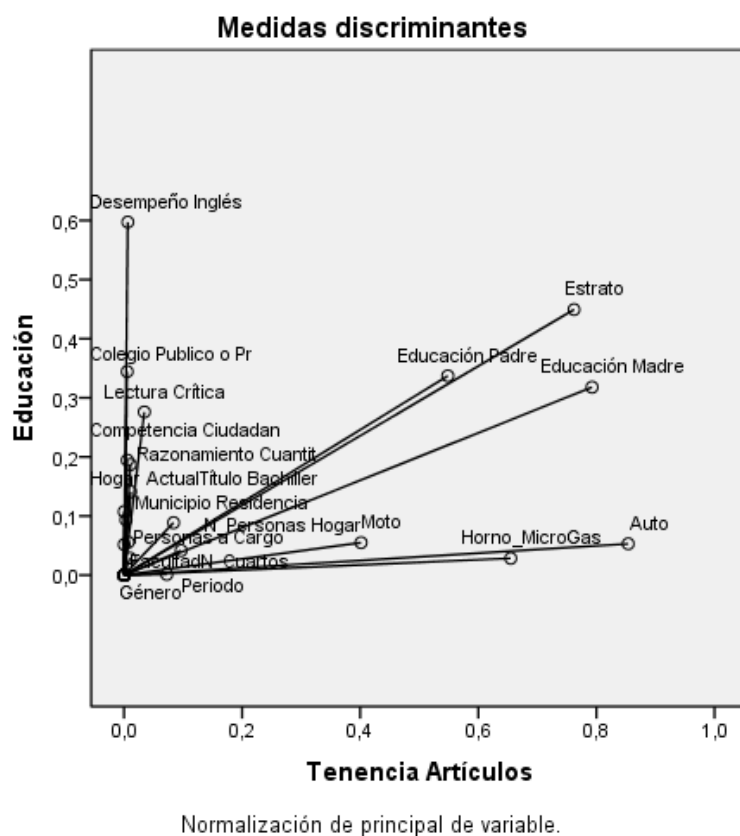


Figura 5. ACM variables sociodemográficas y económicas

A continuación, se presentan las cuantificaciones de las categorías que mostraron influencia en las dos dimensiones en la figura 6 de medidas discriminantes, en primer lugar, se analizan las tres variables relevantes en la dimensión educación padres.

Las gráficas de cuantificación de categoría proporcionan un método alternativo para mostrar la discriminación de variables que pueden identificar relaciones de categoría. En este gráfico, se muestran las coordenadas de cada categoría en cada dimensión. Por lo tanto, puede determinar qué categorías son similares para cada variable.

A continuación (Figura 6), se presentan las cuantificaciones de las categorías que mostraron influencia en las dos dimensiones en la figura de medidas discriminantes, en primer lugar, se analizan las tres variables relevantes en la dimensión educación padres. Se aprecia como los estudiantes que provienen de colegios privados en su mayoría provienen de los estratos más altos (4-5-6) y los padres, tanto madres como papás tienen estudios superiores altos, específicamente postgrados. En contra parte, los que vienen de colegios públicos (elipse 1), en la elipse 2 se aprecia que la formación de los padres posee estudios bajos, donde sobresalen primaria y secundaria incompleta, se destaca, además, que con estas características pertenecen en su mayoría a estratos bajos (1-2-3).

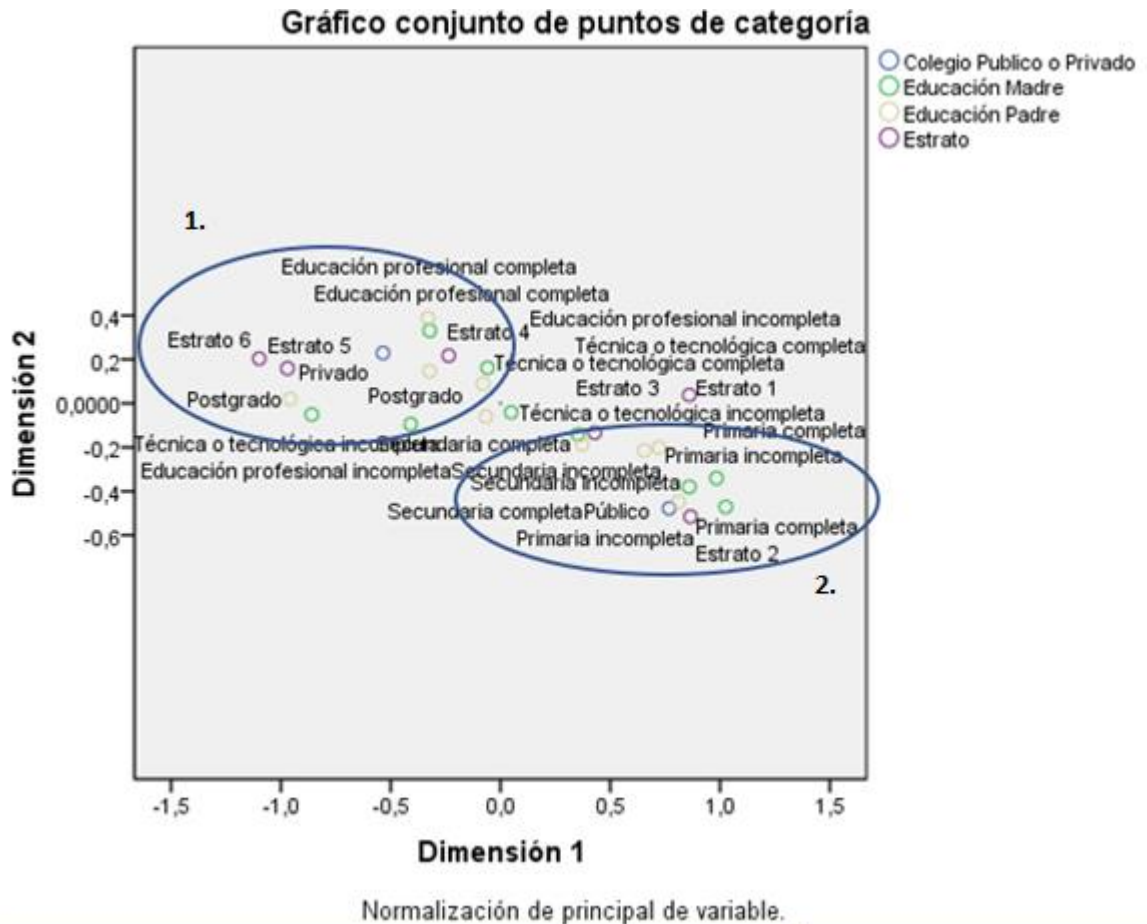


Figura 6. Mapa Perceptual dimensión educación-estrato.

Se aprecia (Figura 7) como se pueden configurar dos grupos bien identificados en la explicación de la dimensión rendimiento, las competencias que mostraron discriminación fueron: desempeño del inglés, comunicación escrita y competencia ciudadana. En general, se observa como los estudiantes de ingeniería industrial y civil provienen más de colegios privados, de estratos altos (4-5-6), con un manejo del inglés alto, lectura crítica, competencia ciudadana y competencia ciudadana entre IV y III, este perfil se aprecia en la elipse 1 dibujada en el mapa perceptual previo. En el mismo gráfico, se dibujó una segunda elipse donde se determina la cercanía de las siguientes categorías que identifican un segundo perfil. Los estudiantes con procedencia de colegios públicos, estratos bajos, tienden a tener bajos rendimientos fundamentalmente en el desempeño del inglés,

competencia ciudadana y lectura crítica. Los estudiantes de ingeniería están cerca de valoraciones medias y altas en lectura crítica y competencia ciudadana, los periodos de tiempo no parecen ser discriminantes en sus categorías.

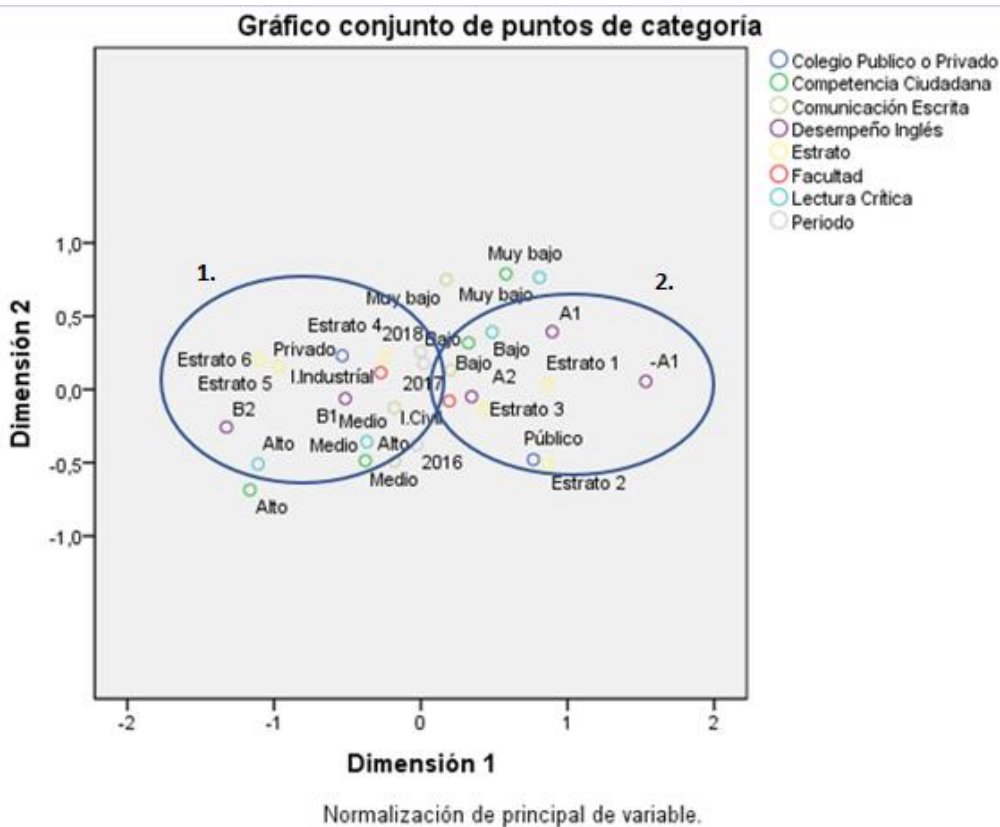


Figura 7. Mapa perceptual dimensión competencias-rendimiento

8.3 Pronostico y validación de los árboles de clasificación de los programas académicos.

8.3.1. Validación del modelo de predicción. Con el fin de identificar que variables están relacionadas con el rendimiento académico en los programas de ingenierías civil e Industrial. A continuación (tabla 6), se presentan las variables dependientes que se recodificaron en cuatro categorías, I, II, III y IV, donde IV es el mayor desempeño y I el menor desempeño. Para cada una de ellas, se utilizaron las mismas variables independientes.

Tabla 6. *Definición de los niveles de desempeño*

Nivel de desempeño	Definición
IV	Muestra un desempeño sobresaliente en las competencias esperadas en cada módulo del examen.
III	Muestra un desempeño adecuado en las competencias exigibles para los módulos del examen. Este es el nivel esperado que todos o la gran mayoría de los estudiantes deberían alcanzar.
II	Supera las preguntas de menor complejidad de cada módulo del examen.
I	No supera las preguntas de menor complejidad de los módulos del examen

Nota. Tomada de ICFES (2018)

De acuerdo con el previo análisis de correspondencias, se establece que las variables de la Tabla 7 se encuentran aquellas que más influencia tienen con las competencias genéricas como: estrato, Educación de los padres, Tenencia de moto, carro, horno microondas o gas, número de personas con las que vive, título bachiller, hogar actual, tipo de colegio y aquellas que no tuvieron gran influencia pero son importantes para el siguiente análisis como: genero, Número de cuartos en casa, municipio de residencia y personas que tiene a cargo.

VARIABLES DEPENDIENTES: Razonamiento Cuantitativo, Lectura Crítica, Competencia Ciudadana, Desempeño Inglés. Comunicación Escrita.

VARIABLES INDEPENDIENTES:

Tabla 7. Variables de entrada Árboles de decisión.

Información personal	Variables de información académica	Variables Socioeconómicas	Variables de familia
Genero	Título de bachiller	Valor matricula	Tipo de residencia hogar actual
Edad	Colegio (público o privado)	Horas trabajadas	Cabeza familia
Municipio de residencia	Educación padre	Estrato socioeconómico	Personas a cargo
	Educación madre	Tenencia de moto	Número de personas en el hogar
		Tenencia de carro	
		Tiene en casa	
		horno microondas o gas	Numero de cuartos en el hogar
			Dedicación internet
			Numero de libros en casa

La idea de la validación de los modelos es corroborar si la predicción que se realizó para el año 2019 clasifica de manera correcta o similar los resultados para las pruebas genéricas. Para ello, se planteó que se realizaría nuevos árboles de decisión tomando las mismas variables a utilizar en la predicción de los resultados del año 2019, pero con los periodos 2016 y 2017 para así predecir los resultados del año 2018.

8.3.2. Validación Ingeniería industrial. Para la validación de los árboles de decisión de los estudiantes de ingeniería industrial, se tomaron los diferentes perfiles arrojados por la predicción de estos; posteriormente se realizaron los perfiles de la predicción con los datos del año 2018 para luego ser comparados con los resultados reales que tuvieron los estudiantes y así considerar si el árbol realizado clasificó de manera acertada o no.

Razonamiento cuantitativo. De acuerdo con la predicción del árbol de la figura 18 para la competencia genérica de razonamiento cuantitativo se espera que los estudiantes de ingeniería industrial menores o igual a 21 años tengan mayor probabilidad de tener un resultado III, mientras que para los mayores a 21 es ligeramente menor estar en un desempeño III.

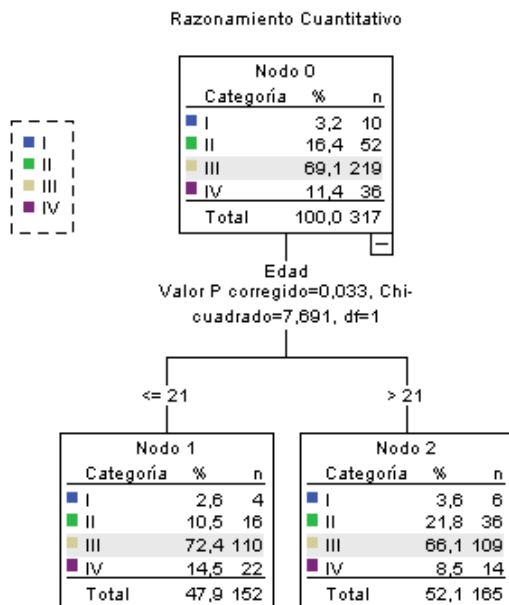


Figura 8. Árbol Razonamiento cuantitativo de validación Ingeniería industrial.

Los resultados que tuvieron los estudiantes de ingeniería industrial para las competencias de razonamiento cuantitativo en el año 2018 la mayoría de los estudiantes (70%) tuvieron un desempeño III (Ver apéndice A. Resultados de los estudiantes ingeniería industrial en las pruebas razonamiento cuantitativo 2018.)

En la Tabla 8 se reflejan los resultados obtenidos en el 2018 según los perfiles que se desarrollaron con el árbol de decisión de la figura 18, en el que un estudiante de ingeniería de industrial de menor o igual a 21 años según los resultados reales el (75%) tuvo un desempeño III y de igual manera la mayoría de los estudiantes (67,9%) que tenían más de 21 años tuvieron un desempeño III para este módulo. En conclusión, el modelo es válido para la competencia de razonamiento cuantitativo del 2018, ya que ambos perfiles que se generaron según el Árbol de decisión CHAID arrojaron porcentajes cercanos a los resultados reales de las pruebas.

Tabla 8. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de razonamiento cuantitativo.

Estudiantes menores o igual a 21 años			Estudiantes mayores a 21 años		
Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%
I	2	3,1%	I	1	1,3%
II	2	3,1%	II	18	23,4%
III	49	76,6%	III	50	64,9%
IV	11	17,2%	IV	8	10,4%

Lectura crítica. La Figura 10 muestra que la predicción para las pruebas de lectura crítica la mayoría de estudiantes de ingeniería industrial tendrán un resultado III, esto dependiendo de la edad ya que los menores o igual a 21 tendrán más probabilidad (61,8%)

que los mayores a 21 años (50,3%). De esta forma se dice que el segundo resultado con más probabilidad de tener es un desempeño Bajo.

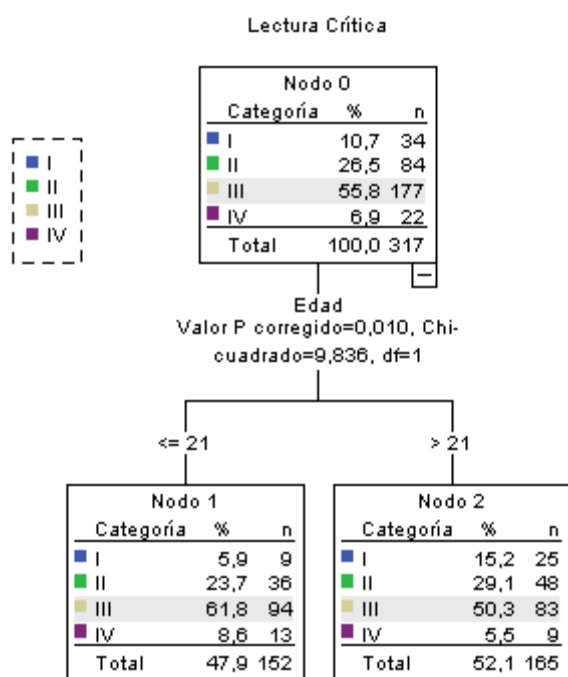


Figura 9.Árbol lectura crítica de validación Ingeniería industrial.

De acuerdo los resultados que tuvieron los estudiantes la mayoría fue de desempeño III de tal manera que el 43% de los estudiantes ocuparon dicho lugar (ver Apéndice B. Resultados de los estudiantes ingeniería industrial en las pruebas lectura crítica 2018). La Tabla 9 refleja los resultados obtenidos de acuerdo con los posibles perfiles que brinda la predicción, la mayoría de estudiantes menores o igual a 21 años tuvieron un desempeño III (51,4%) y la mayoría de estudiantes mayores a 21 años obtuvieron resultado II (43,4%).

La predicción para el módulo de lectura crítica es válida, ya que uno de los perfiles predictores que clasificó el árbol acertó de acuerdo con el resultado de la prueba Saber Pro-2018. Es de resaltar que la predicción clasificó a los estudiantes menores o igual a 21 años con mayor probabilidad de obtener desempeño III a comparación del resto de edades.

Tabla 9. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de lectura crítica.

Estudiantes menores o igual a 21 años			Estudiantes mayores a 21 años		
Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%
I	7	10,9%	I	10	13,0%
II	17	26,6%	II	34	44,2%
III	34	53,1%	III	27	35,1%
IV	6	9,4%	IV	6	7,8%

Competencia ciudadana. Para la predicción de competencias ciudadanas (Figura 11) presenta que la mayoría de estudiantes menores o igual a 21 años tendrán posibilidad de estar en el desempeño III (57,2%), la misma manera sucede con los mayores a 21 años, pero en una probabilidad menor (44,8%).

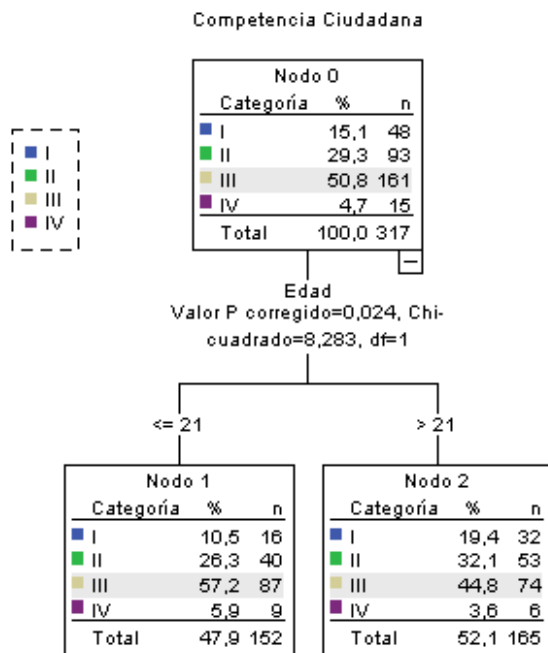


Figura 10. Árbol competencia ciudadana de validación Ingeniería industrial.

Los resultados obtenidos para el módulo de competencias ciudadanas de los estudiantes que presentaron las pruebas Saber Pro-2018, la tabla refleja que la mayoría de estudiantes tuvieron un desempeño III (43%), mientras que la segunda con más estudiantes fue II (35%) (ver Apéndice B. Resultados de los estudiantes ingeniería industrial en las pruebas competencias ciudadanas 2018).

Para validar el árbol de decisión se tomó el perfil de la predicción que acertó a los resultados y se genera con los datos de las pruebas de 2018 para corroborar dicho perfil. El perfil que cumple con los resultados del 2018 son los estudiantes menores o igual a 21 años en el que la mayoría logró un desempeño III (53,1%) (Figura10), De esta manera el árbol para el módulo de competencia ciudadana clasificó de manera similar a los resultados reales de las pruebas.

Tabla 10. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de competencia ciudadana.

Estudiantes menores o igual a 21 años			Estudiantes mayores a 21 años		
Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%
I	13	20,3%	I	17	22,1%
II	17	26,6%	II	32	41,6%
III	34	53,1%	III	27	35,1%
IV	0	0,0%	IV	1	1,3%

Inglés. La predicción (figura 12) para las pruebas de inglés dio 3 posibles perfiles de los estudiantes, el primero es si un estudiante tiene en casa de 0 a 25 libros y se graduó de un colegio privado tendrá como resultado un desempeño B1 (35,7%); el segundo son las mismas condiciones que el primero pero si se graduó de un colegio público esperará un resultado A2 (34,6%); mientras que el ultimo perfil, si el estudiante tiene en casa más de 26 libros su desempeño será B1 (43,2%).

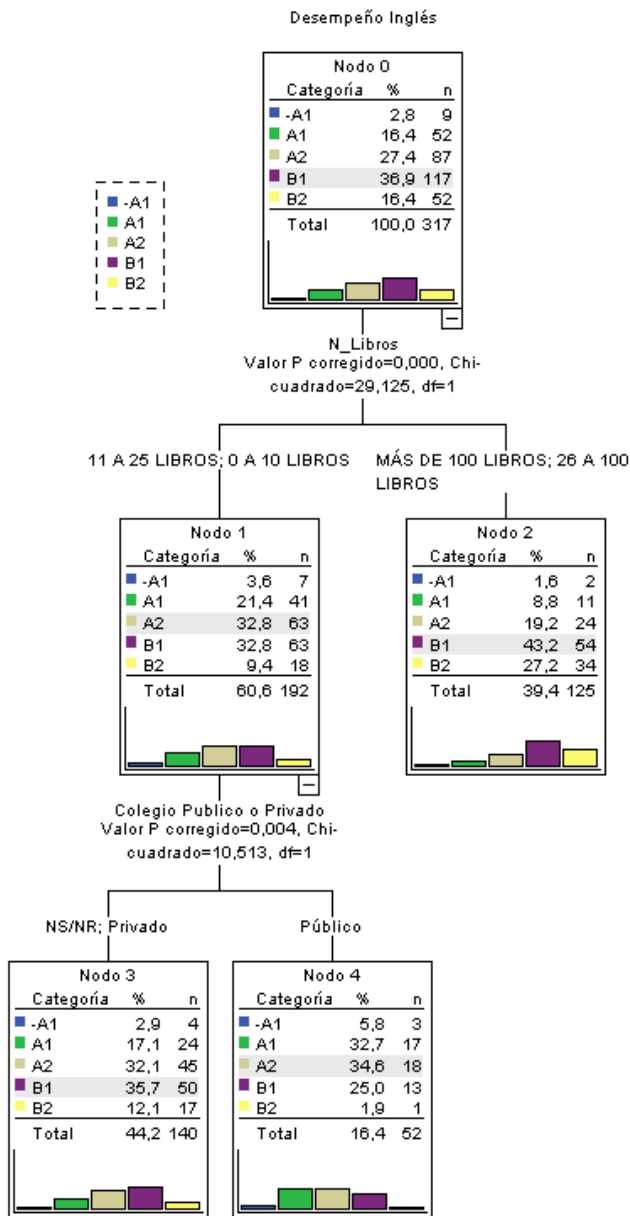


Figura 11. Árbol Razonamiento Inglés de validación Ingeniería industrial.

Los resultados que tuvieron los estudiantes en dicha competencia muestran que la mayoría de estudiantes tuvieron un desempeño de B1 (36%) y la segunda de mayor concentración es el desempeño A2 (31%) (ver Apéndice D. Resultados de los estudiantes ingeniería industrial en las pruebas de inglés 2018).

De acuerdo, con los perfiles dados por el árbol de decisión el perfil que cumple con la condición de los resultados reales (Tabla 11) es: los estudiantes que en casa tengan de 0 a 25 libros y se hayan graduado de un colegio privado lograron un desempeño B1 (34,7%). Al compararse el resultado real junto con el resultado obtenido por el perfil de los estudiantes del periodo 2018 que clasificó el árbol, la predicción es acertada para dicho módulo.

Tabla 11. *Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de inglés.*

Estudiantes que tienen en casa de 0 a 25 libros y se graduaron de un colegio privado		
Nivel de desempeño	Número de estudiantes	%
(-A1)	2	4,1%
A1	6	12,2%
A2	13	26,5%
B1	17	34,7%
B2	11	22,4%

Comunicación escrita. La predicción para la prueba de comunicación escrita (figura 13) está dada por tres perfiles: tanto el primero como el segundo perfil dan con mayor probabilidad al desempeño III (49,2% y 50%), mientras que el tercer perfil lo clasifica con desempeño II.

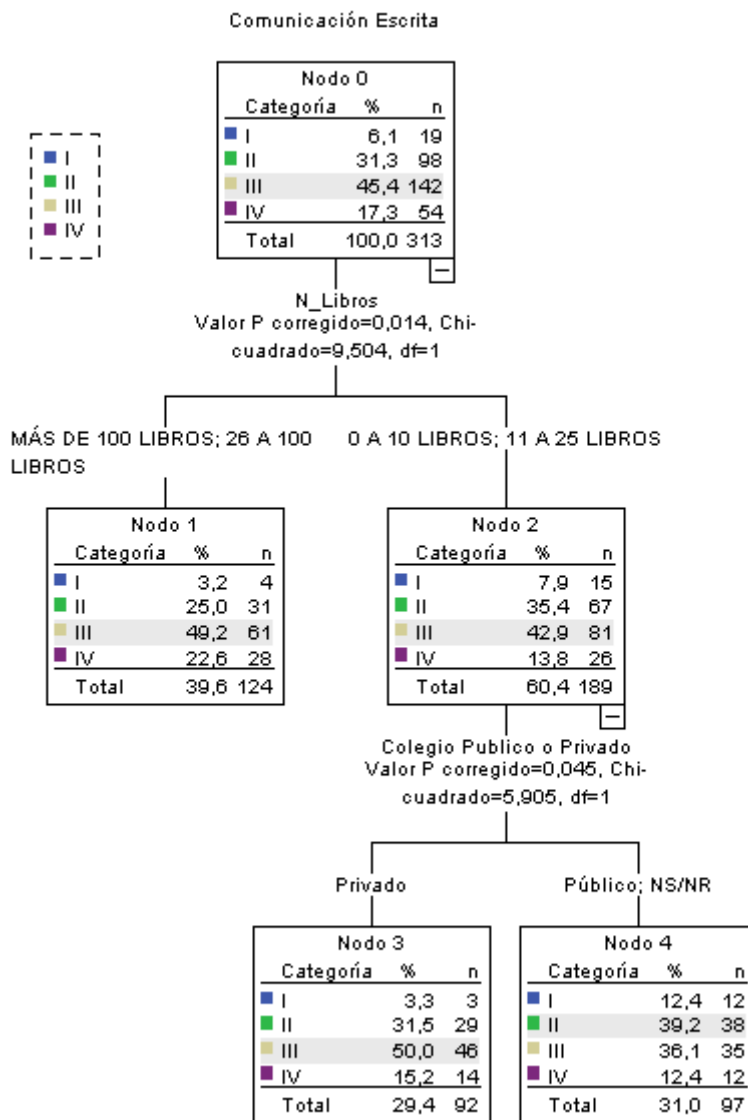


Figura 12. Árbol comunicación escrita de validación Ingeniería industrial.

Los resultados para la competencia de comunicación escrita, refleja que la mayoría de los estudiantes lograron un desempeño II (43%). El perfil que predijo que los estudiantes iban a tener un desempeño II fue el que tuviera en casa de 0 a 25 libros y se haya graduado de un colegio público, la Tabla 12 muestra los resultados del perfil con los datos del 2018 y se evidencia que para esta competencia la predicción no fue acertada ya que la mayoría de

estudiantes obtuvieron un desempeño III (43,5%), sin embargo el porcentaje del desempeño II (42,5%) no es muy lejano al de desempeño III.

Tabla 12. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia ciudadana.

Estudiantes que tienen en casa de 0 a 25 libros y se graduaron de un colegio público			Estudiantes que tienen en casa de 0 a 25 libros y se graduaron de un colegio privado			Estudiantes que tienen en casa más de 26 libros en casa		
Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%
I	2	10%	I	3	7,5%	I	5	10%
II	9	45%	II	15	37,5%	II	20	39%
III	10	50%	III	17	42,5%	III	17	33%
IV	2	10%	IV	5	12,5%	IV	9	18%

8.3.3. Validación de la predicción de ingeniería civil 2018. Para la validación de los árboles de decisión de los estudiantes de ingeniería civil, se tomaron los diferentes perfiles arrojados por la predicción de estos; posteriormente se realizaron los perfiles de la predicción con los datos del año 2018 para luego ser comparados con los resultados reales que tuvieron los estudiantes y así considerar si el árbol realizado clasificó de manera acertada o no.

Razonamiento cuantitativo. La predicción para esta competencia (Figura 14) es generada por tres perfiles: el primero son las mujeres cuya mayor probabilidad (68,1%) es obtener un desempeño III, el segundo son los estudiantes hombres que en casa tengan más de 26 libros tendrán mayor probabilidad de lograr desempeño (67%) y por último los estudiantes masculinos con 0 a 25 libros en casa la mayor probabilidad será de desempeño III (74,7%).

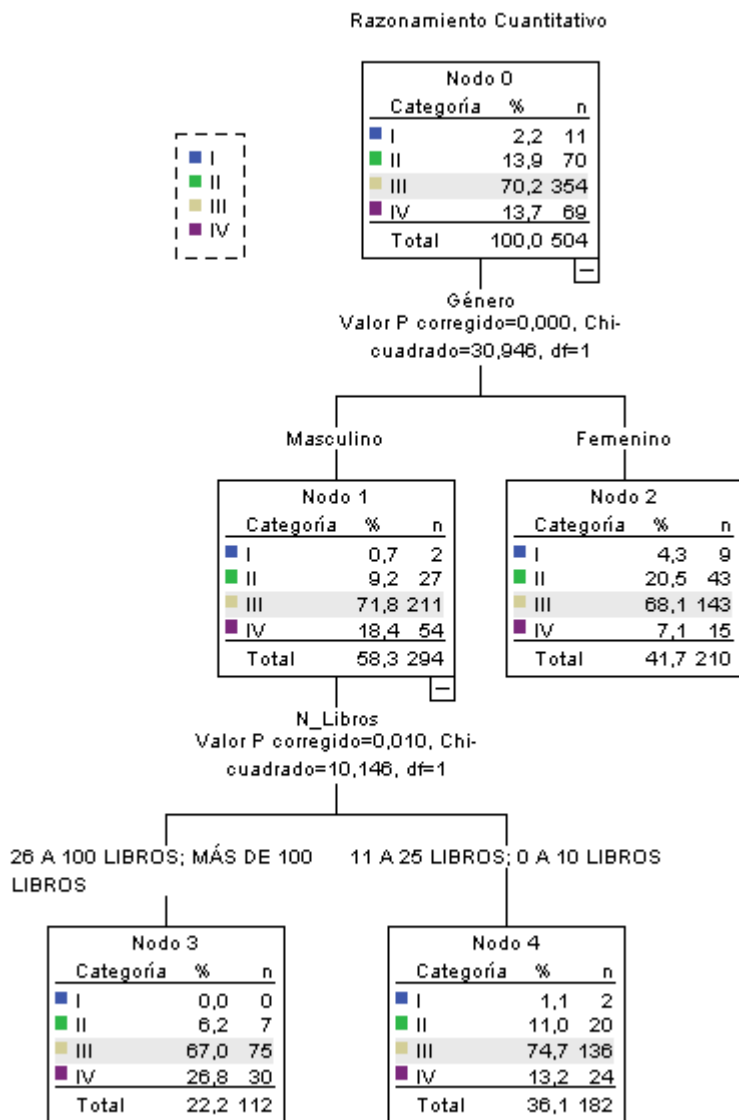


Figura 13.Árbol Razonamiento cuantitativo de validación Ingeniería civil.

La mayoría de estudiantes para el año 2018 tuvieron un desempeño III (69%) (ver Apéndice F. Resultados de los estudiantes ingeniería civil en las pruebas Razonamiento cuantitativo 2018), de esta manera, se verificó los perfiles dados por la predicción que cumplen con los resultados obtenidos por los estudiantes en el 2018. La tabla 13 refleja que

los tres perfiles dados por la predicción son acertados, ya que en los tres la mayoría de estudiantes lograron un desempeño III, el árbol de decisión para este módulo es válido ya que los tres perfiles fueron acertados.

Tabla 13. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de razonamiento cuantitativo.

Estudiante de género femenino			Estudiante de género masculino y tiene en casa más de 26 libros			Estudiante masculino y tiene en casa de 0 a 25 libros.		
Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%
I	0	0,0%	I	0	0,0%	I	0	0,0%
II	4	13,8%	II	0	0,0%	II	2	8,7%
III	22	75,9%	III	7	70,0%	III	18	78,3%
IV	3	10,3%	IV	3	30,0%	IV	3	13,0%

Lectura crítica. La predicción para las competencias de lectura crítica (Figura 15) es que un estudiante que viva actualmente en un hogar temporal y trabaje entre 11 a 30 horas semanales o no trabaje y tenga menos o igual de 21 años tendrá un desempeño III (62,1%) en dichas competencias; el segundo perfil con mayor probabilidad (59,6%) de tener un desempeño III es el estudiantes que viva en un hogar temporal trabaje entre 11 a 30 horas o no trabaje y tenga más de 21 años y el tercer perfil con mayor probabilidad (54,4%) de desempeño III es aquel estudiante que viva en un hogar habitual o permanente.

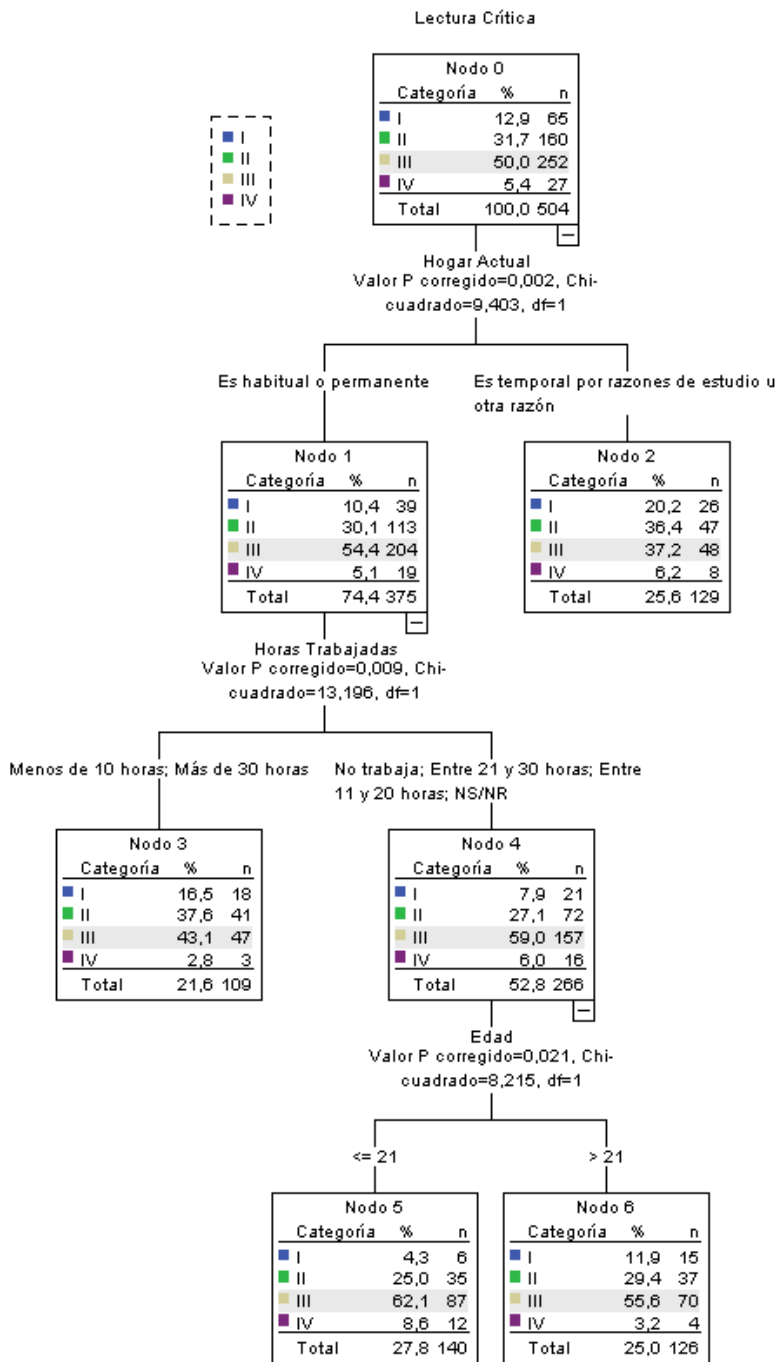


Figura 14. Árbol lectura crítica de validación ingeniería civil.

Los resultados de las pruebas de lectura crítica en el 2018 por los estudiantes de ingeniería civil muestran que la mayoría de estudiantes tuvieron un desempeño III (48%) seguido de desempeño II (36%) (ver Apéndice G. Resultados de los estudiantes ingeniería civil en las pruebas lectura crítica 2018), La Tabla 14 muestra los perfiles que fueron generados con los resultados 2018 a partir de la predicción, de esta manera, los perfiles que cumplen con que la mayoría de estudiantes tuvieron desempeño III fueron : los estudiantes que vivieron actualmente en un hogar temporal, trabajaron entre 11 a 30 horas semanales o no trabajaron y tuviesen menos o igual de 21 años (55%); el segundo perfil presenta las mismas condiciones que el primero pero siendo este mayor a 21 años (70%) y el tercer perfil es aquel estudiante que vivió en un hogar habitual o permanente (54,3%). Por consiguiente, el árbol de decisión para el módulo de lectura crítica es válido, ya que los perfiles dados por la predicción clasificaron acertadamente de acuerdo con los resultados reales.

Tabla 14. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de lectura crítica.

Hogar habitual, no trabaja o trabaja entre 21 y 30 horas y es menor o igual a 21 años			Hogar habitual, no trabaja o trabaja entre 11 a 20 horas y es mayor a 21 años			Hogar habitual, trabaja menos de 10 horas o más de 30 horas.		
Número			Número			Número		
Nivel de desempeño	de estudiantes	%	Nivel de desempeño	de estudiantes	%	Nivel de desempeño	de estudiantes	%
I	0	0,0%	I	2	20%	I	1	6%
II	7	35,0%	II	0	0%	II	8	50%
III	11	55,0%	III	7	70%	III	7	44%
IV	2	10,0%	IV	1	10%	IV	0	0%

Competencia ciudadana. El perfil con mayor probabilidad dado por la predicción de las pruebas de competencia ciudadana (Figura 16) es que el estudiante que presentó las pruebas en el 2018 viviera en un hogar habitual o permanente, se haya graduado de bachiller académico y de un colegio privado; es quien tendrá la mayor probabilidad de tener un desempeño III (49,6%), seguido del estudiante que tenga un hogar habitual y sea bachiller técnico o normalista (63,9%).

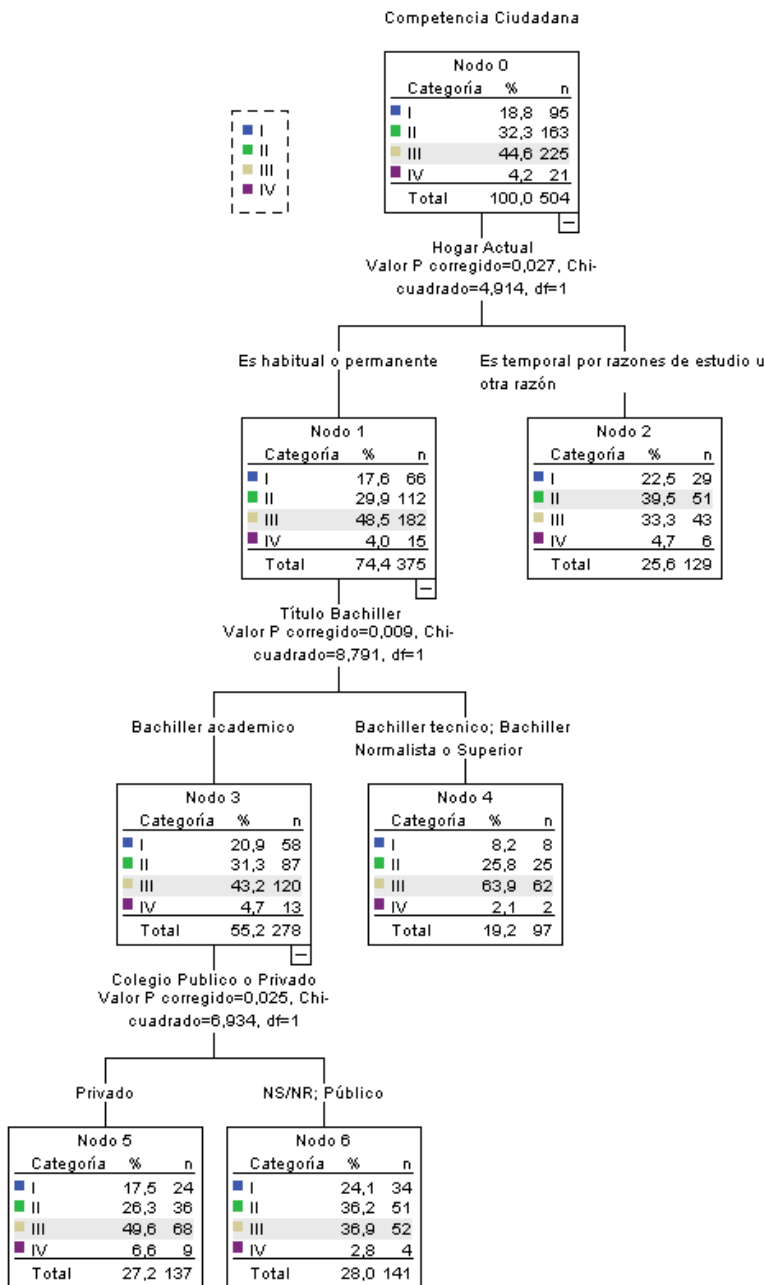


Figura 15. Árbol competencias ciudadanas de validación ingeniería civil.

Los resultados que obtuvieron los estudiantes para las pruebas de 2018, en este el desempeño que la mayoría obtuvo fue III (37%) mientras que el segundo fue II (32%) (ver Apéndice H. Resultados de los estudiantes ingeniería civil en las pruebas de competencias ciudadanas 2018). Para este modelo la Tabla 15 muestra que los perfiles que cumplen con

que la mayoría de estudiantes hayan obtenido un resultado III es el estudiante que vivió en un hogar habitual y sea bachiller técnico o normalista (63%) y el estudiante que haya vivido en un hogar temporal (37,5%). Así mismo, el árbol es válido ya que la predicción de los dos perfiles fue acertada de acuerdo con los resultados de 2018.

Tabla 15. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia ciudadana.

Hogar temporal			Viva en un hogar habitual y sea bachiller técnico o normalista		
Nivel de desempeño	Número de estudiantes	%	Nivel de desempeño	Número de estudiantes	%
I	4	25,0%	I	0	0%
II	5	31,3%	II	3	38%
III	6	37,5%	III	5	63%
IV	1	6,3%	IV	0	0%

Inglés. La predicción de la prueba de inglés (Figura 17) arrojó que los estudiantes que se hayan graduado de un colegio privado y que los padres tuvieran educación profesional completa, postgrado o ninguna, tendrían una mayor probabilidad de tener un resultado B1 (56%). Sin embargo, aquellos que lo hayan hecho de un colegio público tendrán la mayor probabilidad de obtener un desempeño A2 (39,8%).

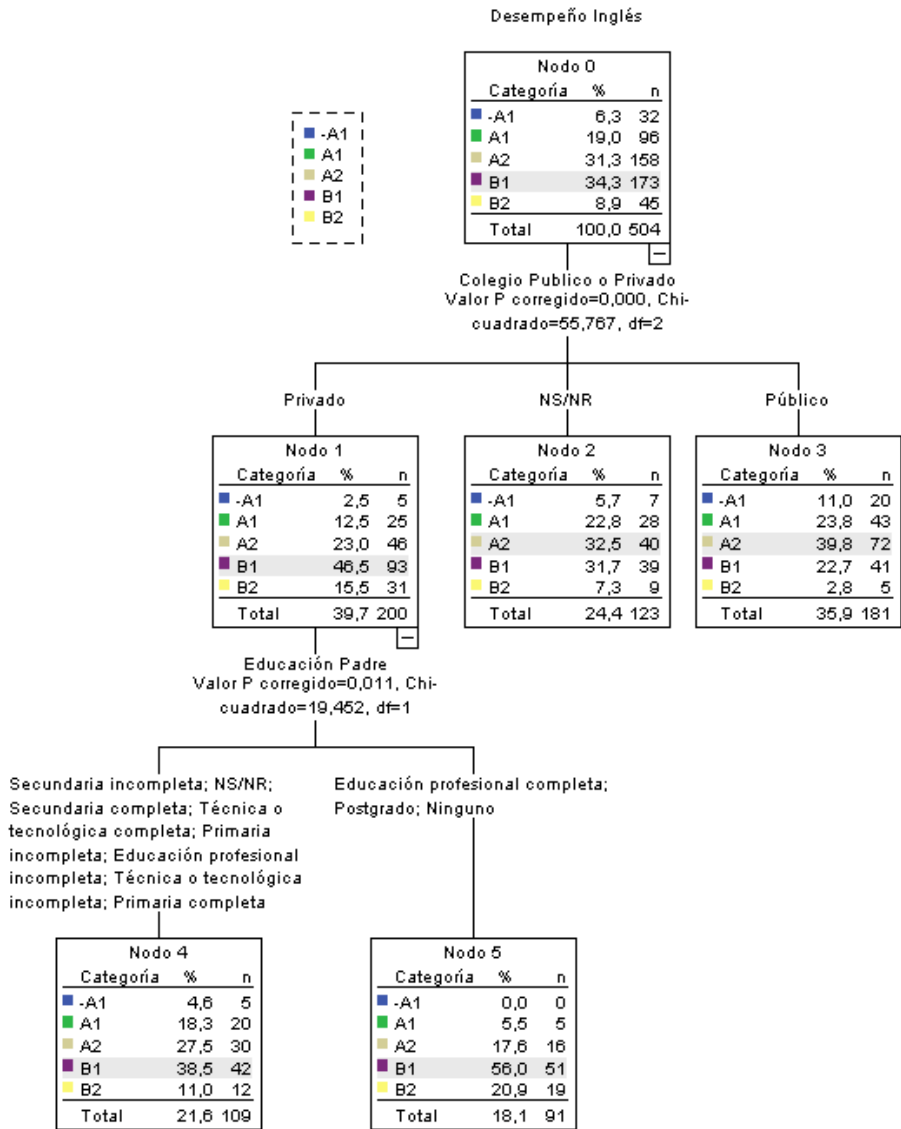


Figura 16. Árbol inglés de validación ingeniería civil.

La mayoría de estudiantes de ingeniería civil que presentaron las pruebas de inglés en el 2018 ocuparon el desempeño de A2 (38%), seguido del desempeño B1 (30%). (ver Apéndice I. Resultados de los estudiantes ingeniería civil en las pruebas de inglés 2018). La predicción que cumple con los resultados A2 en el 2018 está dado por los estudiantes que se haya graduado de un colegio público (37,8%) (Tabla16) este perfil tiene valores

cercanos a la predicción (39,8%) (figura 17), por ende, el árbol realizado es válido para los resultados del módulo de inglés en el año 2018.

Tabla 16. *Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de inglés.*

Estudiante graduado en colegio público		
Nivel de desempeño	Número de estudiantes	%
(-A1)	3	8,1%
A1	8	21,6%
A2	14	37,8%
B1	9	24,3%
B2	3	8,1%

Comunicación escrita. La Figura 18 muestra la predicción para la prueba de comunicación escrita, en el que presenta que los estudiantes que tienen más de 11 libros en casa tendrán la probabilidad de estar en el desempeño III (42,1%), mientras los que tienen de 0 a 10 su desempeño será de II (46,6%).

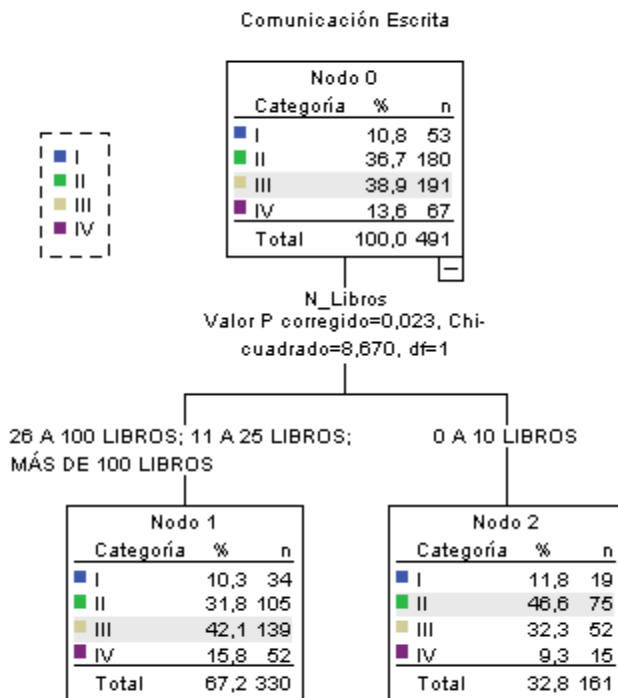


Figura 17. Árbol comunicación escrita de validación ingeniería civil.

Los resultados de Saber Pro-2018 en la prueba de comunicación escrita muestran que el desempeño II (44%) es la calificación que más estudiantes tiene, seguido del desempeño III (34%) (ver Apéndice J. Resultados de los estudiantes ingeniería civil en las pruebas comunicación escrita 2018). En el perfil generado a partir de la predicción con los resultados reales de 2018 (Tabla 17), muestra que el estudiante que tenga de 0 a 10 libros en casa tuvo desempeño II (55%). De esta manera, el modelo de predicción es válido ya que uno de los perfiles acertó generados por el árbol fue cumple con los resultados reales de las competencias de comunicación escrita 2018.

Tabla 17. Resultados de 2018 validados con perfiles de árboles de decisión en la competencia de comunicación escrita.

Estudiante tiene en casa de 0 a 10 libros		
Nivel de desempeño	Número de estudiantes	%
I	3	15,0%
II	11	55,0%
III	6	30,0%
IV	0	0,0%

8.3.4. Resultados ingeniería industrial. Considerando las características de las variables dependientes se usó el procedimiento CHAID, ya descrito previamente, además, de utilizar la estrategia de construir dos árboles en cada caso una para la facultad de ingeniería industrial y uno de ingeniería civil, con el fin de lograr predecir los resultados de las pruebas Saber Pro en las competencias genéricas para el año 2019.

Las variables de entrada que se añadieron para la predicción de las competencias genéricas 2019 son las que se reflejan en la tabla 7, ya que fueron las mismas que se trabajaron para la validación.

Razonamiento cuantitativo. Para este modelo (Figura 19) el estrato es la primera variable predictora, como primer detalle los estudiantes que vivan en estratos 1-5-6 (nodo 2) tienen mayor probabilidad de obtener mejores resultados a comparación de los estratos 2-3-4 (nodo 1), en el que se evidencia que para los estratos del nodo 2 tienen mayor porcentaje en el resultado IV (18,6%) opuesto a (9,7%) de los del nodo 1. Por contrario el nodo 1 muestra que los estudiantes tendrán mayor posibilidad de un desempeño I (3,6%) mientras que el nodo 2 no existe esa posibilidad (0%).

En segunda instancia la edad es predictora para este modelo, los estudiantes de menos o igual a 22 años pueden obtener un desempeño III (71,3%) a comparación de los de

mayor de 22 años cuya probabilidad es menor (61%). De la misma manera como se comporta la primera variable, en la segunda, los de menor o igual a 22 años tienen una mayor posibilidad de lograr el resultado IV (10,9%), mientras que para los mayores a 22 años lograrán con mayor probabilidad resultados bajos (desempeño II 23,2%) (Desempeño I 9,8%).

Las variables que presenta el modelo son significativas para el resultado ya que sus valores P son menores a 0,05. En cuanto a el porcentaje de asertividad es un valor alto (69,4%). Prediciendo que, los estudiantes de ingeniería industrial que presentaron las pruebas Saber Pro-2019 del módulo de razonamiento cuantitativo de estrato 2-3-4 y que tengan menos o igual a 22 años tendrán un desempeño III.

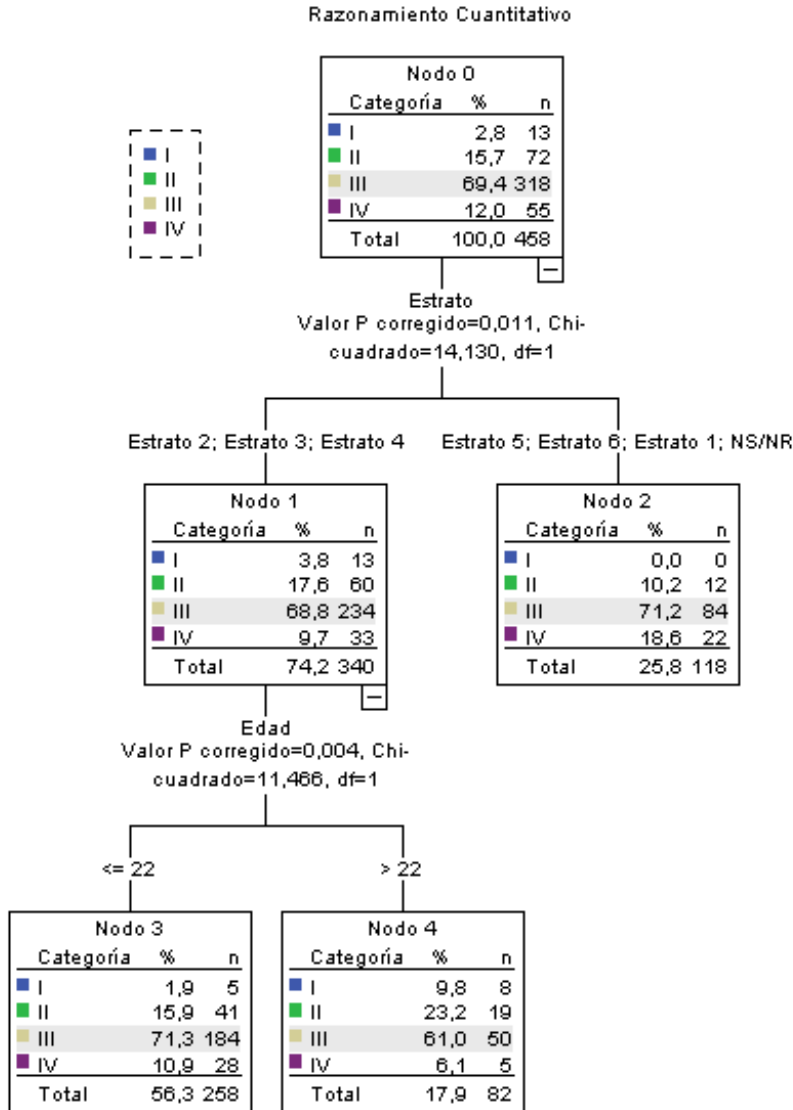


Figura 18. Árbol razonamiento cuantitativo ingeniería industrial

Lectura crítica. La figura 20 muestra que la edad tiene influencia en los resultados de dicha prueba, aquellos que tengan menos o igual a 21 años obtendrán mayor probabilidad de un resultado III (58,3%) a comparación de los mayores de 21 años (46,3%), también los mayores a 21 años tendrán mayor posibilidad de tener un resultado I (15,3%).

La segunda variable predictora es el tipo de colegio (público o privado) que se graduaron los estudiantes, en el que se evidencia que a los de colegio privado tienen una mayor probabilidad de alcanzar un desempeño III (52,1%) en contra (40,5%) de colegio público, también los que hayan terminado su bachillerato de un colegio privado tienen mayor posibilidad de tener un resultado IV (8,3%) mientras que los de un colegio público esta posibilidad disminuye (3,3%).

Las variables de edad y de tipo de colegio son significativas siendo menores a 0,05, mientras que el valor de asertividad para el modelo es aceptable (52, %). De esta forma el modelo permite predecir que los estudiantes de ingeniería industrial mayores a 21 años y se hayan graduado de un colegio privado tendrá un desempeño III en las pruebas de lectura crítica del presente año.

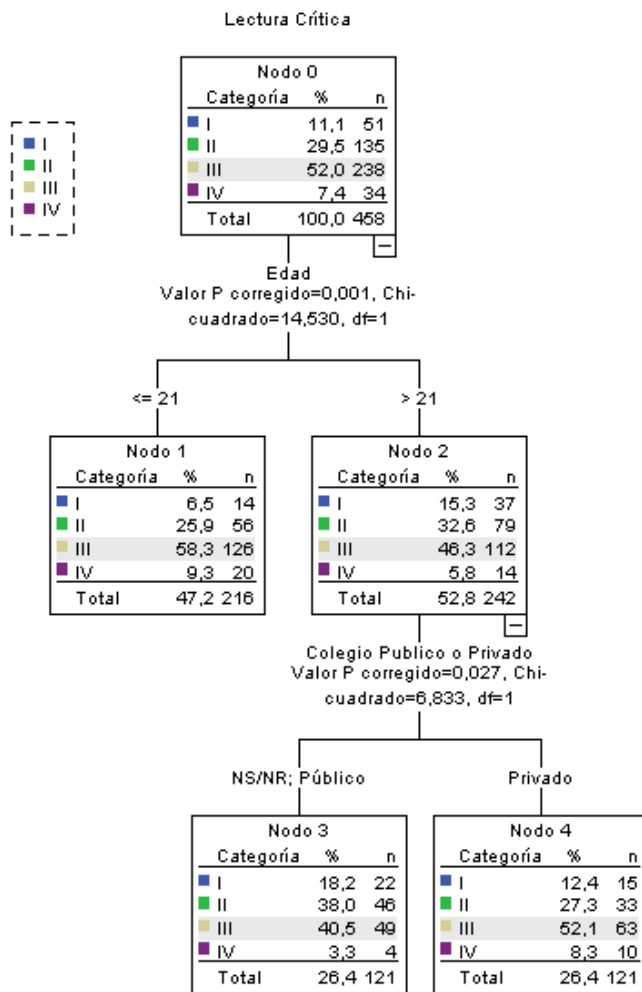


Figura 19. Árbol Lectura crítica ingeniería industrial.

Competencias ciudadanas. Para las pruebas de competencia ciudadana (figura 21) se presentó solo una variable predictora que es la edad, en este caso se evidencia que los menores o igual a 21 años tendrán mayor probabilidad de obtener un resultado III (56,9%) en comparación a los mayores de 21 años cuya probabilidad es menor (40,5%), otro detalle

a resaltar es que los estudiantes que tengan más de 21 años tendrán mayor posibilidad de tener un resultado II (20,7%) mientras los estudiantes menores a 21 es menos (13%).

Este modelo presenta un porcentaje de asertividad aceptable (48,3%) y también su única variable predictora un valor P de 0,003 indicando que es significativa para el módulo de competencia ciudadana. Por consiguiente, el estudiante de ingeniería industrial que tenga menos o igual a 21 años tendrá un desempeño III en las pruebas de competencia ciudadana.

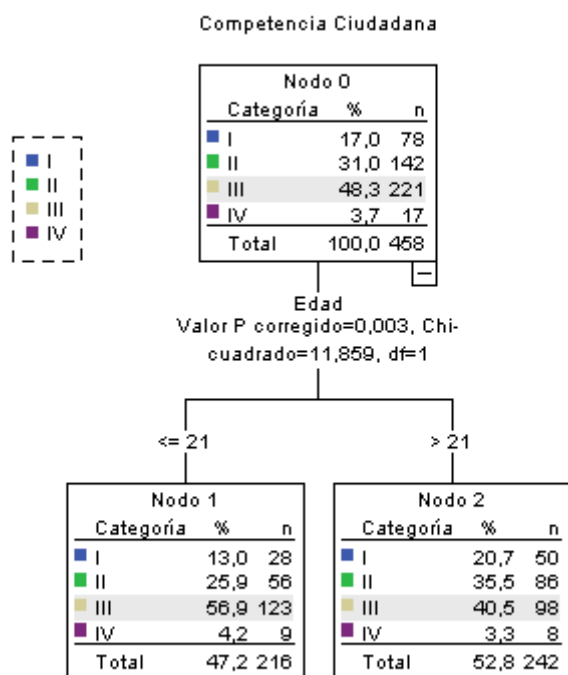


Figura 20. Árbol Competencia ciudadana ingeniería industrial.

Inglés. Para la prueba de inglés (figura 11) la primera variable predictora es el estrato, se evidencia que los estudiantes que pertenezcan a estratos 4 (43,6%) y estratos 5-6

(48,1%) obtendrán un resultado alto B1, mientras que para los estratos 1-2-3 tendrán mayor probabilidad de lograr resultados bajos A2 (39,9%), A1 (22,5%) y -A1 (4,5%).

De segundo se encuentra la variable de tipo de colegio (público o privado) del que se graduó y los estudiantes que se graduaron de un colegio privado tendrán mayor posibilidad de lograr resultados altos como B1 (27,3%) y B2 (12,7%), por el otro lado los de colegio público tendrán mayor probabilidad de tener resultados bajos como A2 (44,1%), A1 (26,5%).

Por último, la variable Educación de la madre muestra que para aquellos estudiantes que sus madres tengan Educación profesional completa o incompleta; técnica completa; ninguno (nodo 6) son las condiciones que se deben presentar para que los estudiantes tengan desempeños altos en esta prueba, tal es el caso que estas condiciones permitirá al estudiante estar dentro de desempeño B1 (37,7%) y B2 (20,8%). Mientras la otra condición es que si la madre del estudiante tiene educación Técnica incompleta; secundaria completa o incompleta; primaria completa o incompleta la mayor posibilidad es que tenga un desempeño A2 (43,9%).

En resumen, un estudiante de ingeniería industrial que presente las pruebas Saber Pro de 2018 y que pertenezca a estratos 1, 2 y 3, proveniente de un colegio privado y la madre tenga educación profesional completa e incompleta; postgrado; técnica completa; ninguno, obtendrá un resultado B1 en las pruebas de inglés. Mientras que los estudiantes de estrato 4,5 y 6 tendrán un desempeño B1.

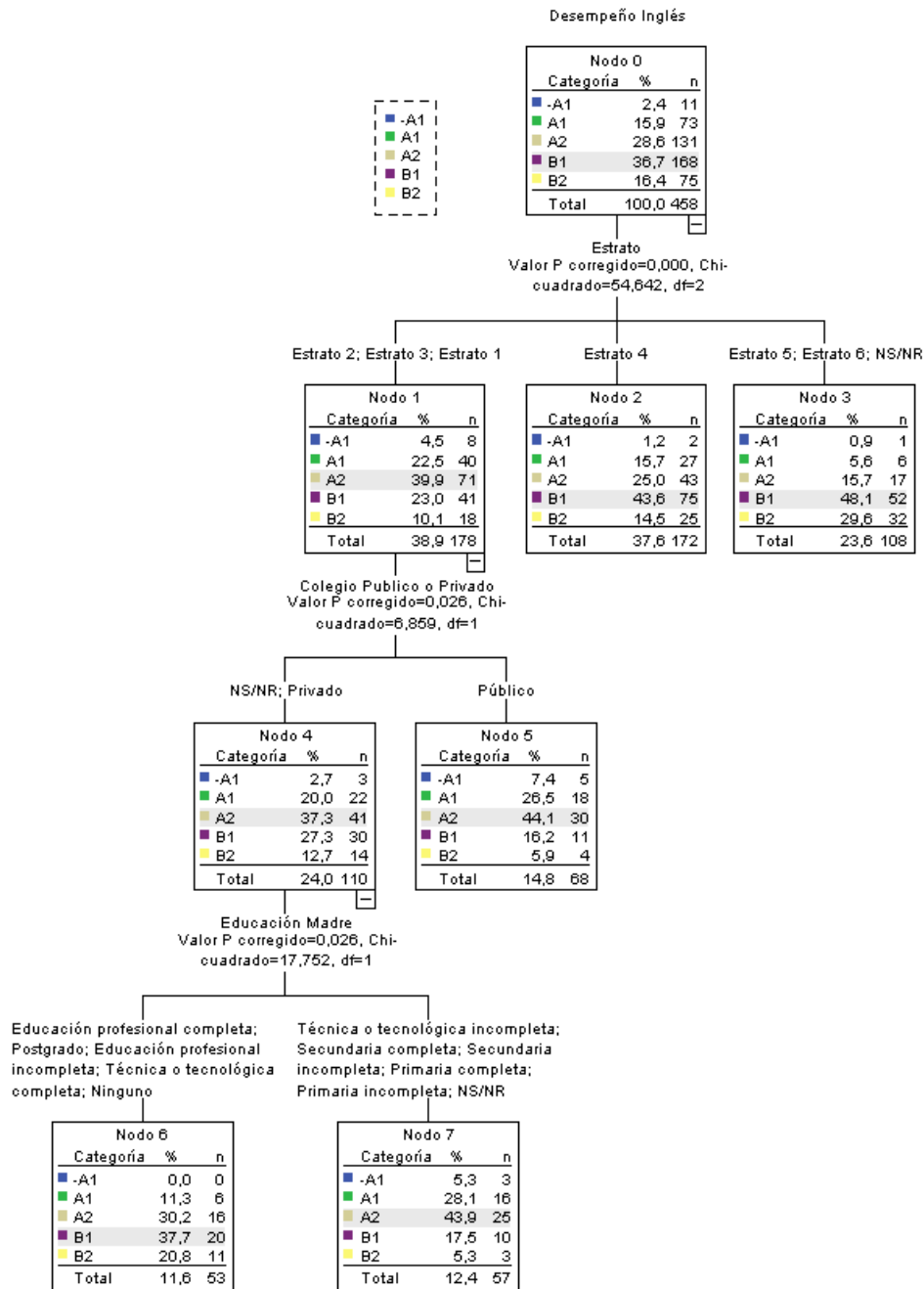


Figura 21. Árbol Desempeño inglés ingeniería industrial.

Comunicación escrita. Para este módulo (figura 23) solo se presentó una variable predictora y es el género del estudiante, las mujeres tienen mayor posibilidad de tener un desempeño III de (44,7%) mientras que para los hombres es de (35,9%), también se evidencia que los hombres tienen mayor probabilidad de lograr desempeño II (38,3%).

El modelo presenta un porcentaje de asertividad ligeramente aceptable de (42,4%) y su única variable predictora tiene un valor P de (0,011) lo que indica que es significativa para el modelo. Para este módulo los estudiantes de ingeniería industrial de 2019 de género femenino tendrán un resultado III mientras que para los de masculino desempeño II.

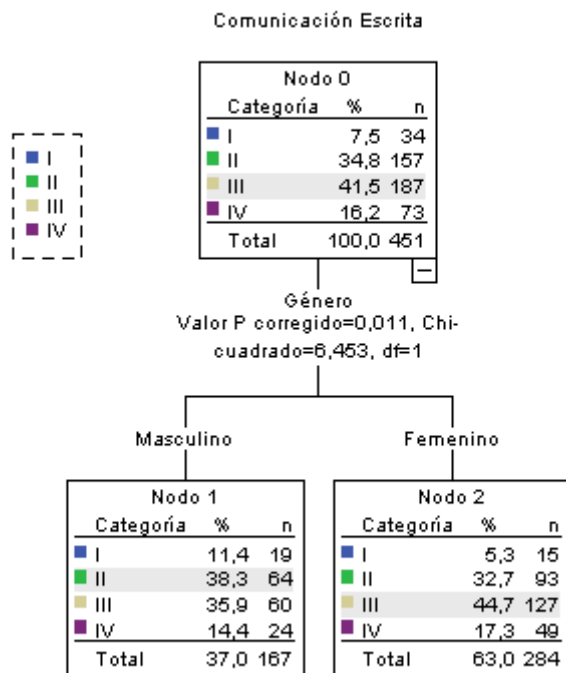


Figura 22. Árbol Comunicación escrita ingeniería industrial.

8.3.5. Resultados Ingeniería Civil.

Razonamiento cuantitativo. La figura 24 presenta el árbol para dicho modulo, en él se refleja que ambos géneros (masculino y femenino) tienen una proporción similar para un desempeño III (69,2% y 69,5%), pero, los hombres son los que tienen mayor probabilidad de estar en un resultado IV (20,6%) frente a las mujeres (7%).

Estos resultados se ven afectados por el tipo de colegio que se graduó el estudiante (público o privado), siendo los de colegio público los que tienen mayor proporción de desempeño III (70,1%) pero no es un muy lejano a los de colegio Privado (67,9%), por contrario, los provenientes de colegio privado tienen mayor posibilidad de un resultado IV (25,8%) frente a (16,8%) de colegio público.

Por último, el (73,8%) es la probabilidad de que un estudiante tenga desempeño III siendo un bachiller académico frente a un (64,5%) de puntaje III que pertenece a un bachiller técnico o normalista, no obstante, los bachilleres técnicos o normalista tienen más porcentaje de estar en el desempeño IV (26,2%) en comparación a los académicos que es menos de la mitad (10,8%).

El árbol presenta un valor de asertividad bueno (69,3%) y sus variables se presentaron por debajo del valor P (0,05) indicando que el estudiante de género masculino, que haya salido de un colegio público y sea bachiller académico tendrá un desempeño III. Mientras de género femenino sin cumplir las condiciones de los hombres obtendrá el mismo desempeño.

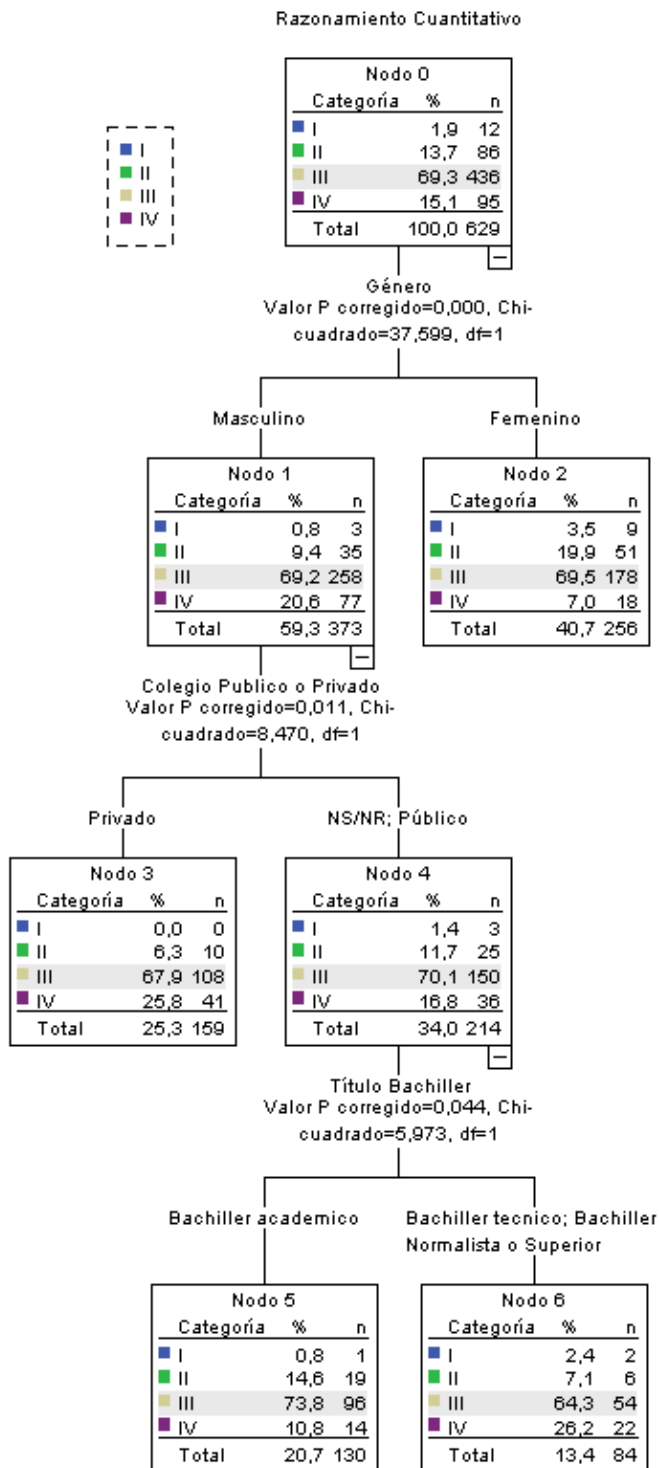


Figura 23.Árbol rendimiento cuantitativo ingeniería civil.

Lectura Crítica. Como primera variable se encuentra la edad del estudiante (figura 25), en la predicción se observa que los estudiantes de menor o igual a 21 años tienen una mayor concentración en el desempeño III (53,2%) en comparación a los mayores de 21 años (44,8%), de tal manera que, se puede afirmar que los menores de 21 años tendrán un mejor rendimiento en estas pruebas, ya que en los desempeños I (9,6%) y II (29,9%) tienen menor concentración que los mayores a 21 años (15,5% I) y (35,7% II) ; también son los menores de 21 años los que tendrán mayor posibilidad de lograr un desempeño IV (7,3%).

Por último, el tipo de hogar en el que vive actualmente el estudiante refleja que los estudiantes que residan en un hogar habitual (Nodo 3) obtendrán desempeño III (48,1%), caso contrario a los que residan en un hogar temporal (Nodo 4) cuya mayor probabilidad es de tener un resultado II (37,9%). Al comparar el desempeño I se muestra que el nodo 4 tiene mayor posibilidad de estar en dicho desempeño (23%) y aquellos de un hogar habitual casi la mitad (12,9%).

El indicador de asertividad se muestra relativamente aceptable (49,1%), sin embargo, las variables presentadas tienen valor P inferior a (0,05) indicando que tienen significancia en los resultados. De esta manera, si un estudiante de ingeniería civil que tenga más de 21 años y resida en un hogar habitual o permanente tendrá un desempeño III en las competencias de lectura crítica; y el estudiante que resida en un hogar temporal por razones de estudio u otra razón estará en la calificación de III.

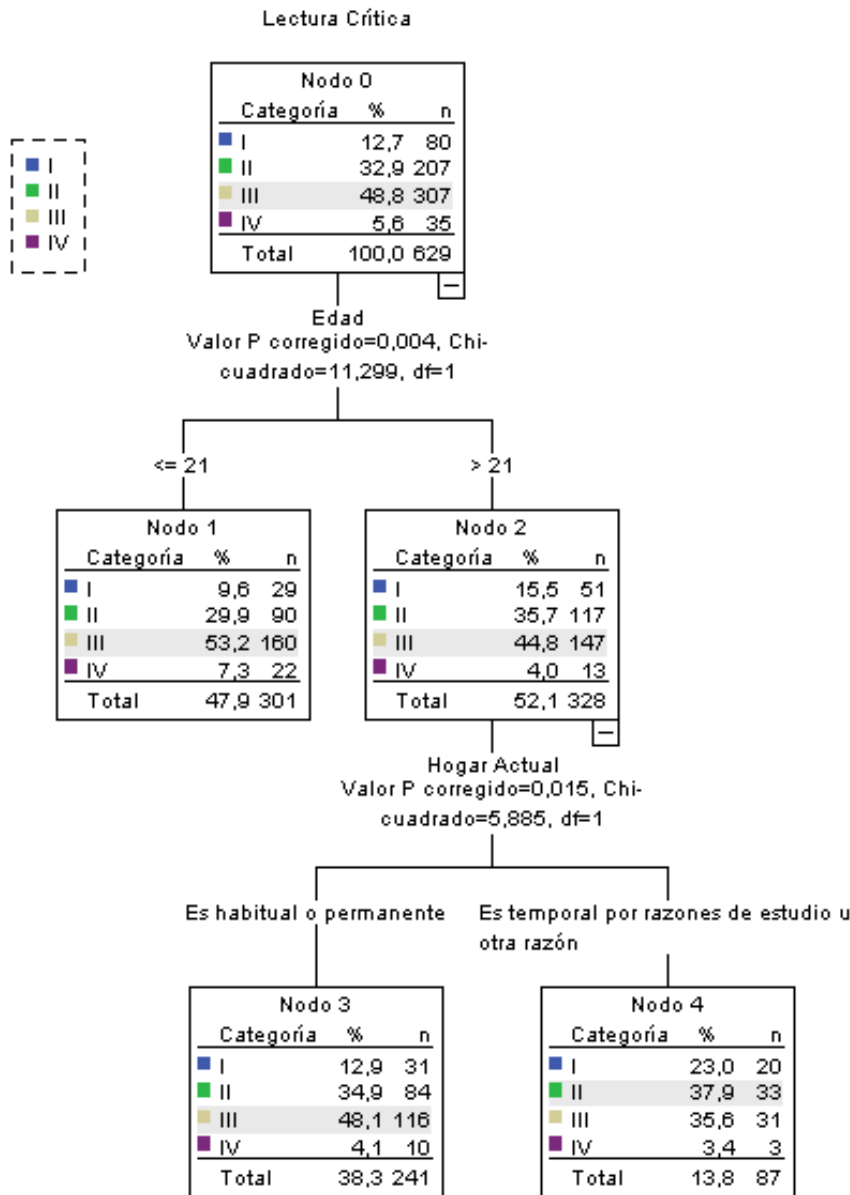


Figura 24. Árbol lectura crítica ingeniería civil

Competencia Ciudadana. En el módulo de competencia ciudadana (figura 26) el título de bachiller resultó como primera variable, en ella el desempeño III tiene mayor concentración por parte de los estudiantes que son bachilleres técnicos o normalista

(52,4%) a comparación de los bachilleres académicos (39,9%), también es pertinente que el desempeño I es más alcanzable por los estudiantes académicos (22,2%) que los técnicos o normalistas (14,1%).

El tipo de colegio del estudiante (privado o público) es influyente en los resultados de esta prueba, aquellos estudiantes que hayan terminado sus estudios en un colegio privado tendrán mayores posibilidades de estar en el desempeño III (44,8%) frente a los de colegio público que su mayor concentración es una calificación de II (37%), de esta manera, los estudiantes que culminaron en una entidad educativa no oficial (privada) tendrán mayor posibilidad de estar en el desempeño IV (5,4%) por lo que para las entidades oficiales será la mitad (2,5%).

El porcentaje de asertividad para el modelo se encuentra relativamente débil (43,9%), pero sus variables predictoras se encuentran correctamente significativas al estar inferior de 0,05 del valor P. En otras palabras, el modelo predice que los estudiantes de ingeniería civil que sean bachilleres académicos y se hayan graduado de un colegio Privado tendrán

como resultado un desempeño III; mientras que los que se hayan graduado de un colegio Público un desempeño II para las pruebas de competencia ciudadana.

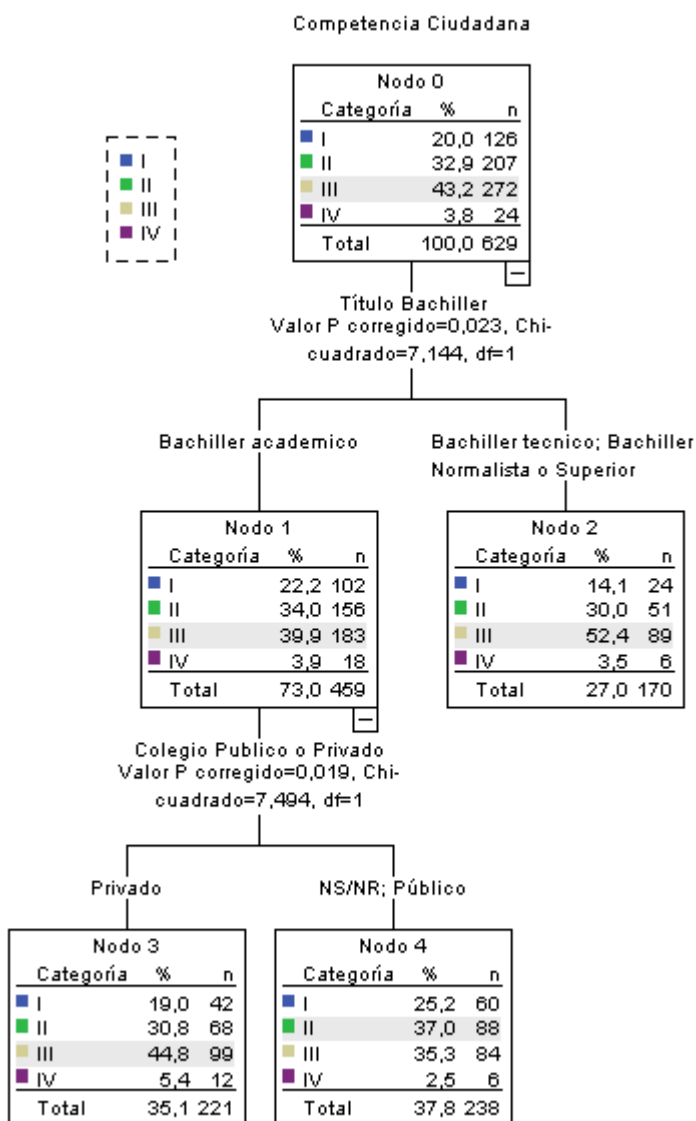


Figura 25. Árbol competencia ciudadana ingeniería civil.

Inglés. La figura 27 muestra que el tipo de colegio (público o privado) que se graduó el estudiante es una variable influyente en dicha prueba, de tal manera que los estudiantes que son de un colegio privado obtendrán calificaciones altas como B1 (45,8%) y B2 (16,5%) y que siendo de un colegio público serán de menor probabilidad dichas calificaciones B1 (22,1%) y B2 (2,2%), también se evidencia que es lo contrario en los estudiantes que se graduaron de un colegio público cuyo desempeño serán bajos A1 (23%) y -A1 (10,2%) y aquellos que no lo sean disminuirá la probabilidad de tener dichos desempeños A1 (12%) y -A1 (2%).

Sin embargo, el modelo relacionó el estrato en el que vive el estudiante como última variable predictora, para todos los estratos el desempeño de mayor probabilidad es la calificación B1, sin embargo este porcentaje es más de la mitad para estratos 5 y 6 (55,9%) mientras que para el resto de estratos es de (42,6%), de igual forma, los de estratos altos tienen mayor posibilidad de alcanzar un desempeño B2 (27,1%) si se compara con los estratos bajos (13,2%).

La asertividad del modelo es un valor relativamente débil (41,8%), no obstante, las variables predictoras se muestran fuertemente relacionadas con la prueba de inglés y son significativas siendo menores al valor P. Siendo así, un estudiante de ingeniería civil que se haya graduado de un colegio privado y viva en estratos 5 o 6 tendrá mayor posibilidad de estar en el desempeño B1 y B2 de la prueba de inglés.

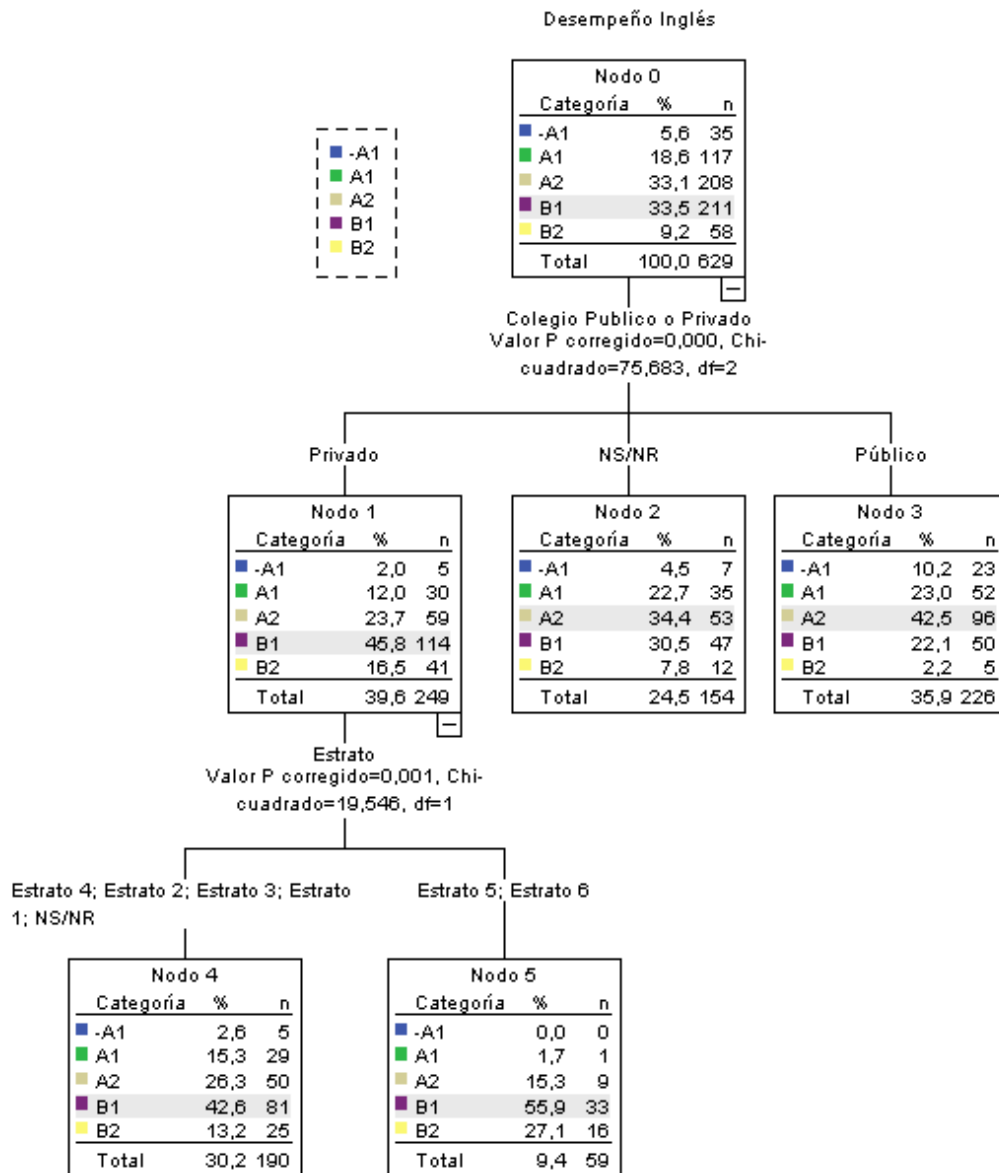


Figura 26. Árbol rendimiento inglés ingeniería civil.

Comunicación Escrita. El siguiente modelo (figura 28) es algo particular, ya que en la primera variable que es la de género del estudiante se evidencia que para ambos géneros los desempeños de II y III se encuentran en una concentración similar, es decir que para los

hombres la mayor concentración se encuentra en II (39,8%) y no muy lejano a este valor el desempeño III (38,2%), en cambio las mujeres la de mayor concentración es III (37,5%) y muy cerca de esa proporción II (36,7%). Sin embargo, se puede afirmar que los estudiantes de género masculino tendrán más calificaciones bajas, ya que son estos los que tienen mayor probabilidad de tener una calificación de I (12,4%) comparado con las mujeres que esperarían el (9,2%), otra manera de afirmar esto es que las mujeres tienen mayor posibilidad de tener desempeño IV (16,7%) y los hombres de (9,6%).

La variable de tipo de hogar en el que vive actualmente el estudiante muestra que aquellos que vivan en un hogar temporal tendrán mayor probabilidad de encontrarse en desempeño II (42,2%). Por otro lado, aquellos que estén en un hogar habitual tienen mayor posibilidad de alcanzar un resultado III (39,6%).

Este modelo cuenta con un asertividad de 40% siendo relativamente débil, aun así, las variables que se relacionaron con la predicción de esta prueba son significativas siendo menores al valor P (0.05). La predicción del modelo es si un estudiante de ingeniería civil siendo mujer y que viva en un hogar temporal tendrá un desempeño II; mientras si vive en un hogar habitual tendrá un desempeño III. Si el estudiante es un hombre su desempeño en las pruebas de comunicación escrita será II.

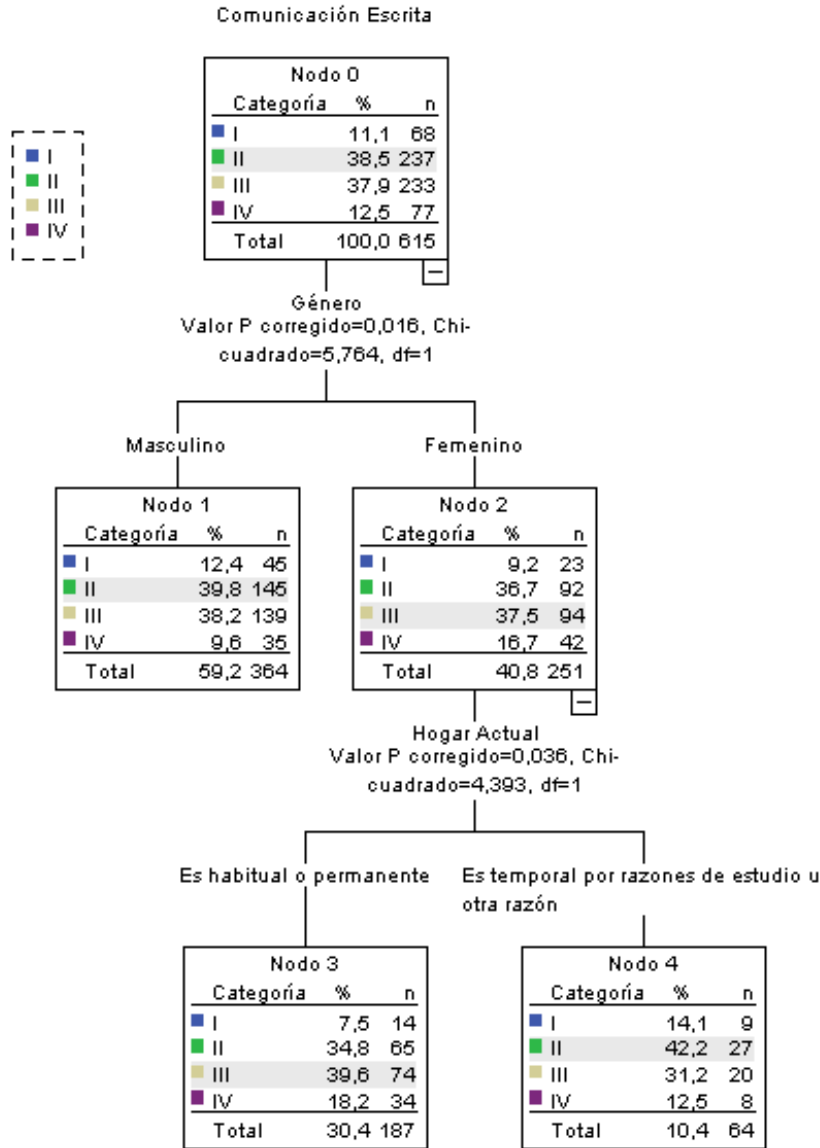


Figura 27. Árbol comunicación escrita ingeniería civil.

9. Conclusiones y recomendaciones

Las variables que dan información socioeconómica del estudiante como Tiene moto, carro, horno, lavadora, computador, servicio de TV, son preguntas que ya no se deben tener en cuenta para el análisis, ya que, en su mayoría los estudiantes poseen dichos bienes. Es decir, se concluye que estos bienes ya hacen parte dentro de la economía de un hogar y que no tienen algún impacto en el desempeño de las pruebas.

Las preguntas referidas a la preparación que tuvo el estudiante para las pruebas no se tomaron, ya que son variables que en su mayoría o su totalidad no tenían respuesta por parte de los estudiantes, así como si pertenecía a algún grupo étnico o tenía alguna discapacidad física o motriz.

Perfil Estudiantes por Facultad.

No se encontraron diferencias en la tenencia de bienes y servicios según profesión, los bienes o servicios muy comunes son el tener acceso a la Internet, la TV, el computador, medianamente poseen autos y lo que menos se posee por parte de los estudiantes en las dos carreras son las motos, poco menos del 25%.

Con el uso del ACM, se determinaron dos dimensiones, una con variables de formación de los padres como discriminadoras y otra como el rendimiento en tres componentes, ellas son: desempeño del inglés, lectura crítica y competencia ciudadana y también aporta a esa dimensión en el tipo de colegio.

Los estudiantes que provienen de colegios privados en su mayoría provienen de estratos altos y los padres, tienen estudios superiores altos, específicamente postgrados.

Los estudiantes que vienen de colegios públicos, la formación de los padres posee estudios bajos, donde sobresalen primaria y secundarias incompletas, además, pertenecen en su mayoría a estratos bajos (1-2-3).

Los estudiantes de ingeniería industrial provienen más de colegios privados, de estratos altos (4-5-6), con un manejo del inglés alto, lectura crítica, competencia ciudadana y competencia ciudadana entre alto y medio.

Los estudiantes con procedencia de colegios públicos, estratos bajos, tienden a tener bajos rendimientos fundamentalmente en el desempeño del inglés, competencia ciudadana y lectura crítica. Los estudiantes de ingeniería están cerca de valoraciones medias y altas en lectura crítica y competencia ciudadana, los periodos de tiempo no parecen ser discriminantes en sus categorías.

Ingeniería Industrial.

La prueba de Razonamiento Cuantitativo es influenciada por el estrato en el que vive y la edad del estudiante, teniendo como mayor probabilidad un resultado III. El desempeño Alto, la mayor probabilidad de obtenerlo es un estudiante que viva en estratos altos.

En lectura crítica, la edad y el tipo de colegio (público o privado) del que se graduó el estudiante son variables predictoras para los resultados de esta prueba, en el que la mayor probabilidad en todos los casos es un desempeño III. Los estudiantes menores o igual a 21 años son los que tienen mayor probabilidad de tener resultados Altos, mientras los del peor resultado son los estudiantes mayores a 21 y se hayan graduado de un colegio público.

En cuanto a competencias ciudadanas solo una variable fue influyente para los resultados y es la edad, donde los que tendrán mejor rendimiento son los estudiantes menores o igual a 21 años y los mayores a 21 tendrán mayor probabilidad de tener un desempeño I.

En el desempeño de Inglés, los estudiantes que obtienen un mejor rendimiento o un puntaje alto son aquellos que pertenecen a estratos altos (4 en adelante). Sin embargo, aquellos estudiantes de estratos bajos, que se hayan graduado de un colegio Privado y la educación de la madre fuese mayor a secundaria también tendrán desempeño B1. Como detalle a resaltar, es evidente que siendo de estratos bajos y de un colegio público su rendimiento será A2.

Finalmente, la prueba de comunicación escrita se vio influenciada por el género del estudiante, en el que el género Femenino tiene una mayor oportunidad de obtenga unos resultados más altos que los de género Masculino.

Ingeniería Civil

Los resultados sobre razonamiento cuantitativo se vieron influenciados principalmente por el género donde los hombres obtienen mayoritariamente mejores valores que las mujeres y dentro de los hombres que se graduaron de un colegio público obtuvieron mejores resultados aquellos que son bachilleres académicos.

En cuanto Lectura Crítica, la edad y el tipo de hogar en el que vive el estudiante influyó en los resultados, los menores a 21 años son los que mayormente tienen mejores resultados. Dentro de los estudiantes mayores a 21 años que vivan en un hogar habitual o permanente tendrán mejores resultados que aquellos que vivan en un hogar temporal.

En lo que respecta a la Competencia Ciudadana, influyen características del colegio del cual se graduó (título de bachiller y tipo de colegio), donde los de mejor rendimiento son los bachilleres técnicos o normalistas y los de bajos rendimientos son los bachilleres académicos de colegio público.

El Desempeño Inglés estuvo influenciado principalmente por el tipo de colegio (público o privado) donde provienen los estudiantes de ingeniería civil, siendo los de mejores rendimientos los de colegios privados. Otro detalle que se percibió fue que en los colegios privados si influye negativamente el pertenecer a estratos medios y bajos (1,2,3 y 4)

Por último, la Comunicación Escrita sólo se vio principalmente influenciada por el género de los estudiantes, donde los hombres tienen desempeños bajos. Al contrario, las mujeres que vivan en un hogar permanente tendrán mejores resultados que aquellas que vivan en un hogar temporal.

Las variables de Tipo de colegio bien sean público o privado, Genero y Edad son las variables con mayor influencia en el desempeño de las pruebas de competencias genéricas del Saber Pro, ya que estas se encuentran predictoras en ambas facultades.

El resultado de la competencia genérica de inglés se ve afectada por la variables estrato, tipo de colegio público o privado en ambas facultades, esto es explicable, ya que la mayoría de los colegios privados tienen mejor énfasis en la enseñanza del inglés a comparación de los colegios públicos; en este orden de ideas son los de estratos medio-altos los que pueden costear un colegio privado y por esta razón son estos los que logran

obtener un alto desempeño, comparado con los de estratos bajo o de un colegio público cuyo desempeño se espera que sea II.

Para los estudiantes de ingeniería industrial las variables que más afectan fueron las de la formación secundaria adquirida, ya que se evidencia que si se gradúa de un colegio privado o es bachiller académico tendrá mayor oportunidad de tener resultados buenos en las competencias genéricas. Por otra parte, si el estudiante es mujer tendrá mejor desempeño en las competencias genéricas.

En la facultad de ingeniería civil también se presenta que los estudiantes que se graduaron de un colegio privado y que vivan en estratos altos se verán mayormente favorecidos a un buen rendimiento en las pruebas.

Es notable que la edad verdaderamente influye en estas pruebas, ya que para ambas facultades los estudiantes que presentaron las pruebas y tenían menos de 22 años tendrían mejores resultados que los mayores a esta edad.

Los modelos encontrados para las diferentes dimensiones de rendimiento académico han sido validados, dado que la predicción que se obtuvo del año 2018 y los resultados reales que tuvieron los estudiantes del 2018 siempre entregaron estructuras similares y resultados de asertividad parecidos para todas las competencias genéricas de las pruebas Saber Pro.

Recomendaciones

Los árboles de decisión en especial CHAID, son técnicas descriptivas que permiten predecir una situación de acuerdo con las variables de entrada que se le suministren, es

recomendable que para estos modelos las variables de entrada no contengan muchas categorías, de esta manera el algoritmo distribuirá o imprimirá un resultado con mayor poder de asertividad para cada una de las variables.

Para futuros estudios las variables que preguntan de cual colegio se graduó el estudiante no permiten manipular de manera correcta, ya que es una pregunta abierta y genera demasiadas opciones de respuesta, se recomienda que se cree nuevas variables de acuerdo con esta, como el tipo de colegio Público o Privado, si es mixto o no, de tal manera que sean derivadas de la variable original.

Se debe replantear las variables de tipo tenencia de bienes como horno, servicio TV, servicio de internet, computador, moto, carro; debido que la mayoría de estudiantes en la actualidad ya posee dichos servicios o más aun algunos de los beneficios mencionados ya están dentro de la canasta familiar. De esta manera se recomienda a las entidades encargadas de generar preguntas que vayan más allá de la tenencia y sean más las preguntas del uso y frecuencia de la utilización a dichos bienes, de esta manera se puedan tener en cuenta para futuros análisis.

10. Bibliografía

- Algañaraz Soria, V. H. (2016). El “Análisis de Correspondencias Múltiples” como herramienta metodológica de síntesis teórica y empírica. Su aporte al estudio del locus universitario privado argentino (1955-1983). *Revista Latinoamericana de Metodología de las Ciencias Sociales*, 6(1), 1-16
- Aranzazu, D., Rodríguez, B., Zapata, M., Bustamante, J., & Restrepo, L. (2007). Aplicación del análisis de factor de correspondencia múltiple en un estudio

de válvulas cardiacas en porcinos. *Revista colombiana de pecuaria*, 20, 129-1

Artunduaga, M. (2008). Variables que influyen en el rendimiento académico en la universidad. Madrid: Depto MIDE (Métodos de Investigación y Diagnostico en Educación)

Avendaño, C.; Gutiérrez, K.; Salgado, C., & Do-antos, M. (2016). Rendimiento académico en estudiantes de ingeniería comercial: modelo por competencias y factores de influencia. *Formación Universitario*, 9(3), 3-10. Doi: 10.4067/S0718-50062016000300002

Barrios, M., & Frías, M. (2016). Factores que influyen en el desarrollo y rendimiento escolar de los jóvenes de Bachillerato. *Revista Colombiana de Psicología*, 25(1), 63-82.

Bonilla, S. (2013). Educación y crecimiento económico entre 1975 y 2005: una aplicación del modelo de datos del panel dinámico de Arellano y Bond. *Sociedad y economía*, 25, 255-274.

Briceño, A. (2011). La educación y su efecto en la formación de capital humano y en el desarrollo de capital humano y en el desarrollo económico de los países. *Apuntes del CENES*, 30(51), 45-59.

Cañon, H., Cortes, R., Lopez, M., Castellanos, F. (2009). Revista investigación en enfermería: imagen y desarrollo. *Rendimiento académico y desempeño en el Examen de Calidad de la Educación Superior (ECAES) de enfermería, año 2003*. 11(1), 67-80 recuperado de:
<https://revistas.javeriana.edu.co/index.php/imagenydesarrollo/article/view/1612>

Celina Oviedo, Heidi, & Campo Arias, Adalberto (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, XXXIV (4), undefined-undefined. [fecha de Consulta 25 de septiembre de 2019]. ISSN: 0034-7450. Disponible en:
<http://www.redalyc.org/articulo.oa?id=806/80634409>

- Cortes, S., Arcadio, C. (2015). Revista Repertorio de medicina y cirugía. *Examen de calidad de la educación superior Saber Pro: factores socodemograficos en el desempeño académico en instrumentación quirúrgica*. 24(3), 206-211.
Recuperado a partir de
<https://revistas.fucsalud.edu.co/index.php/repertorio/article/view/618>
- De la Fuente, S. (2011). Análisis correspondencias múltiples y simples. Universidad Autónoma de Madrid. Recuperado de
<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>
- Díaz Sepúlveda, J. F. *Comparación entre árboles de regresión CART y regresión Lineal* (Doctoral dissertation, Universidad Nacional de Colombia, Medellín).
- Duran, S., Prieto, R & Garcia, J. (2017) Calidad de vida, inclusión social y bienestar humano. *influencia de la calidad de vida en el rendimiento del estudiante universitario*, 1, 89-110. Recuperado de
https://www.researchgate.net/profile/Ronald_Prieto_Pulido/publication/317098519_Influencia_de_la_Calidad_de_Vida_en_el_rendimiento_-_Duran_Prieto_Garcia/data/5ab11c82458515e3ebecbb85/06-Influencia-de-la-Calidad-de-Vida-en-el-rendimiento-Duran-Prieto-Garcia.pdf.
- Flores, F., Sánchez, M., & Martínez, A. (2016). Modelo de predicción del rendimiento académico de los estudiantes del ciclo básico de la carrera de medicina a partir de la evaluación del desempeño docente. *Revista Mexicana de Investigación Educativa*, 21(70), 975-991
- Flores, J. (2016). Simultaneidad y complementariedad entre economía y educación. Economía y educación. *Economía y administración*, 7(1), 65-76.:
<https://doi.org/10.5377/eya.v7i1.4294>
- Greenacre, M. (2008). *La práctica del análisis de correspondencia*. España, Madrid: Fundación BBVA.

- Hanushek, E., & WöBmann, L. (2007). Calidad de la educación y crecimiento económico. *Programa de protección de la reforma educativa en América Latina y El caribe*, 1-30, 4-37.
- Hernández, R., Fernández, C., & Baptista, M. (2014). *Metodología de la investigación*. Ciudad de México, México: McGraw Hill Education.
- Instituto Colombiano de Educación Superior (ICFES). (2018d). Instituciones de educación superior, reporte de resultados, guía de interpretación y uso de resultados del examen Saber Pro. Recuperado de:
<http://www2.icfesinteractivo.gov.co/aporteRelativoInstituciones/pdfs/guia.pdf>
- Instituto Colombiano de Formación y Educación Superior (ICFES). (2012). Recuperado de: <https://studylib.es/doc/7248647/presentaci%C3%B3n-de-powerpoint>
- Instituto Colombiano de Formación y Educación Superior (ICFES). (2015). Informe nacional de resultados Colombia en Pisa 2015. Recuperado de <https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf>
- Instituto Colombiano de Formación y Educación Superior (ICFES). (2016). Documentación del examen Saber Pro. Recuperado de:
<http://www2.icfes.gov.co/docman/investigadores-y-estudiantes-de-posgrado/acceso-a-bases-de-datos/documentacion-y-diccionarios-saber-pro/5515-documentacion-saber-pro/file?force-download=1>
- Instituto Colombiano de Formación y Educación Superior (ICFES). (2017). Saber Pro, informe nacional de resultados examen Saber Pro-2016. Recuperado de:
<http://www.icfes.gov.co/en/docman/investigadores-y-estudiantes-de-posgrado/informes-de-resultados-evaluaciones-nacionales/informes-de-saber-pro/3816-informe-nacional-de-saber-pro-2017/file>
- Instituto Colombiano de Formación y Educación Superior (ICFES). (2018a). Guía de diseño, producción, aplicación y calificación del examen saber 11. Recuperado de <http://www2.icfes.gov.co/docman/atencion-al->

ciudadano/5295-guia-de-diseno-produccion-aplicacion-y-calificacion-del-examen-saber-11

Instituto Colombiano de Formación y Educación Superior (ICFES). (2018b). Icfes 2014-2018, la evaluación como herramienta para mejorar la calidad de la educación. Recuperado de <http://www2.icfes.gov.co/prensa/publicaciones>

Instituto Colombiano de Formación y Educación Superior (ICFES). (2018c). Icfes interactivo 2018, todo sobre las pruebas Icfes Saber, conozca los resultados y encuentra tipo icfes. Recuperado de: <http://icfesinteractivo.info/pruebas-colombia/>

Instituto Colombiano de Formación y Educación Superior (ICFES). (2018). Recuperado de:

<http://www.icfes.gov.co/documents/20143/496194/Guia%20de%20orientacion%20modulos%20de%20competencias%20genericas-saber-pro-2018.pdf>

Kass, G. (1980). An exploratory technique for investing large quantities of categorical Data. *Applied statistics*, vol (29), pp. 119-127- Recuperado de: <http://links.jstor.org/sici?sici=0035-9254%281980%2929%3A2%3C119%3AAETFIL%3E2.0.CO%3B2-N>

Lebart, C., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Francia, Paris: Dunod.

Leiva, M., & Valdés, M. (2016). Modelo estadístico para predecir el puntaje de desarrollo psicomotor de niños de 4 a 5 años de edad en función del nivel socioeconómico. *Archivos de pediatría de Uruguay*, 87(1), 23-27.

Márquez, A. (2017). Educación y desarrollo en la sociedad del conocimiento. *Perfiles Educativos*, XXXIX (158), 3-17.

Márquez, C., Romero, C., & Ventura, S. (2012). Predicción del fracaso escolar mediante técnicas de minería de datos. *Revista Iberoamericana de tecnología del aprendizaje*, 7(3), 109-117

- Milanović, M., Stamenković, M. (2016). Chaid decisión tree: methodological frame an application. *Economic themes*, 54 (4), 563-586. DOI 10.1515/ethemes-2016-0029
- Monroy, F., & Hernández, F. (2014). Factores que influyen en los enfoques de aprendizaje universitario. Una revisión sistemática. *Educación XXI*, 17(2), 105-114. Doi: 10.5944/educxx1.17.2.11481
- Morales, J. (2004). Aplicación e interpretación de técnicas de reducción de datos según escalamiento óptimo. (Tesis de grado). Universidad de Chile. Chile.
- Moratilla, R. (2018). Educación y economía en la España reciente. *ENSAYOS, Revista de la facultad de Educación d Albacete*, 33(1), 189-204.
- Olave, G., Pérez, E., Ortiz, M., Batía, N., Márquez, C., Parra, P., & Ibáñez, P. (2016). Factores que afectan al ambiente educativo en la formación preclínica de medicina según sus docentes. *Revista Médica Chilena*, 144, 1343-1350. Doi: 10.4067/S0034-98872016001000015
- Organización para la Cooperación y el Desarrollo Económico. Pisa 2015, resultados claves. Recuperando de: <https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf>
- Ouyang, J., Patel, N & Sethi, I. (2011). From centralized to distributed decision tree induction using CHAID and fisher's linear discriminant function algorithm. *Intelligent decisión technologies*, (5). 133-149. DOI 10.3233/IDT-2011-0102
- Pawliczek, A., Kozel, R., & Vilamová, Š. (2018). STRATEGIC MANAGEMENT AND PERFORMANCE OF ENTERPRISES EVALUATED USING CHAID DECISION TREE ANALYSIS. *Scientific Papers of the University of Pardubice. Series D, Faculty of Economics & Administration*, 26(44).
- Pereira, C., Hernández, G., & Gómez, I. (2011). El valor predictivo de los exámenes de estado frente al rendimiento académico universitario. *Revista educación y educadores*, 14(1), 51-64
- Pérez, M. (2015). *Minería de datos a través de ejemplos*. 1ª ed. México: Alfaomega

- Rajimon, J. (2010). La economía y la función de producción en educación. *Visión de futuro*, 13(1), pp. 1-7.
- Ramaswami, M & Bhaskaran, R. (2010). IJCSI International Journal of Computer Science Issues. *A CHAID based performance prediction model in educational data mining*, 7 (1). 10-18
- Ramirez, C. (2014). Revista Colombiana de educación. *Factores asociados al desempeño académico según nivel de formación pregrado y género de los estudiantes de educación superior Colombia*. 66, 203-224. Recuperado de <http://www.scielo.org.co/pdf/rcde/n66/n66a09.pdf>
- Ramos, M., Solís, C., Hernández, I., Redondo, M., & Hernández. *Economía y administración*, 7(1), 41-64. Doi: <https://doi.org/10.5377/eya.v7i1.4294>
- Resolución Número 018735 del 10 de diciembre de 2018. [Ministerio de Educación Nacional]. Número 018735 del 10 de diciembre de 2018. "Por la cual se otorga reconocimiento a las instituciones de educación superior que obtuvieron los mejores resultados en el Examen de Estado de Calidad de la Educación Superior en competencias genéricas y específicas, aplicados en el año 2017". Diciembre 10 de 2018
- Revista Dinero. (2017). Las mejores universidades de Colombia en Ingeniería. Recuperado de: <https://www.dinero.com/edicion-impresa/caratula/articulo/mejores-universidades-de-colombia-en-ingenieria-2017/245856>
- Revista Dinero. (2018). Las mejores universidades en ingeniería en Colombia 2018. Recuperado de <https://www.dinero.com/edicion-impresa/caratula/articulo/ranking-de-mejores-universidades-en-ingenieria-de-colombia-2018/258763>
- Ruiz, C. (2001) Revista Complutense de Educación. *Factores familiares vinculados al bajo rendimiento*, 12(1), 81-113. Recuperado de <http://revistas.ucm.es/index.php/RCED/article/view/RCED0101120081A>
- Ruiz, C. (2001). Factores familiares vinculados al bajo rendimiento. *Revista Complutense de Educación*, 12(1), 81-113.
- Saavedra, R. (2012). El análisis de correspondencia conjunto y múltiple ajustado. (Tesis de grado). Pontificia Universidad Católica de Perú, Perú, Lima.

- Sanchez, A., Otero, A. Reportes del emisor investigación e información económica. *Educación y reproducción de la desigualdad en Colombia*, 154, [1-4]. Recuperado de <https://publicaciones.banrepcultural.org/index.php/emisor/article/view/7863/8243>
- Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130-5.
- Valbuena, C., & García, G. (2016). Factores asociados al desempeño de los estudiantes de psicología en Colombia en las Pruebas Saber Pro-2011-2 a 2013-2. *Asociación Colombiana de Facultades de Psicología*, 1-77.
- Vélez, A., y Roa, C. (2005). Factores asociados al rendimiento académico en estudiantes de medicina. *Educación médica*, 8 (2): 77-82. Recuperado de: <http://scielo.isciii.es/pdf/edu/v8n2/original1.pdf>
- Visbal, D., Mendoza, A. & Orejuela, S. Predicción de la eficiencia de las instituciones de educación superior colombianas con análisis envolvente de datos y minería de datos. *Pensamiento y gestión*, 42. 140-161. <http://dx.doi.org/10.14482/pege.41.9704>

11. Anexos

Anexo1. Diccionario de variables

CAMPO	DESCRIPCIÓN DEL CAMPO	OPCIONES DE RESPUESTA
INFORMACIÓN PERSONAL		

ESTU_GENERO	Género	1- Femenino 2- Masculino
EDAD_RECODIFICADA	Fecha de Nacimiento	1- 20 o menos 2- 21 años 3- 22 años 4- 23 años 5- 24 años 6- 25 años 7- Más de 25 años
PERIODO	Periodo de aplicación de la prueba	1- 2016 2- 2017 3- 2018
DATOS ACADÉMICOS SABER PRO		
FACULTAD	Facultad a la que pertenece el estudiante	1- Ingeniería Industrial 2- Ingeniería Civil
ESTU_COD_RESIDE_MCPPIO	Municipio donde reside el estudiante	1- Bucaramanga 2- Floridablanca 3- Girón 4- Piedecuesta 5- Fuera del Área Metropolitana
ESTU_TITULOBTENIDOBACHILLER	Título de bachiller obtenido	1- Bachiller académico 2- Bachiller técnico Bachiller pedagógico o normalista 99- NS/NR
ESTU_COL_PRIV_PUBL	Se graduó de colegio público o privado	1- Público 2- Privado 99- NS/NR
ESTU_VALORMATRICULAUNIVERSIDAD	¿Cuál es el valor de la matrícula del último semestre cursado (sin considerar descuentos o becas)?	0- No pagó matrícula 1- Menos de 500 mil 2- Entre 500 mil y menos de 1 millón 3- Entre 1 millón y menos de 2.5 millones 4- Entre 2.5 millones y menos de 4 millones 5- Entre 4 millones y menos de 5.5 millones 6- Entre 5.5 millones y menos de 7 millones 7- Más de 7 millones 99- NS/NR
ESTU_PAGOMATRICULABECA	Variable que define si el pago de matrícula es por beca	0- No 1- Si 99- NS/NR
ESTU_PAGOMATRICULACREDITO	Variable que define si el pago de matrícula es mediante crédito	0- No 1- Si 99- NS/NR
ESTU_PAGOMATRICULAPADRES	Variable que define si el pago de matrícula lo realizan los padres del estudiante	0- No 1- Si 99- NS/NR
ESTU_PAGOMATRICULAPROPIO	Variable que define si el pago de matrícula es por recursos propios	0- No 1- Si 99- NS/NR

INFORMACIÓN SOCIOECONÓMICA		
FAMI_HOGARACTUAL	Variable que indica si el hogar actual donde vive es permanente o temporal	1- Es habitual o permanente 2- Es temporal por razones de estudio u otra razón 99- NS/NR
FAMI_CABEZAFAMILIA	¿Es usted jefe de hogar o cabeza de familia?	0- No 1- Si 99- NS/NR
FAMI_NUMPERSONASACARGO	¿Cuántas personas dependen económicamente de usted? (Incluya parientes, no parientes y servicio doméstico que viven permanentemente en su hogar)	0- Ninguna 1- Una 2- Dos 3- Tres 4- Cuatro 5- Cinco 6- Seis 7- Siete 8- Ocho 9- Nueve 10- Diez 11- Once 12- Doce o más 99- NS/NR
FAMI_EDUCACIONPADRE	Nivel educativo más alto alcanzado por el padre	0- Ninguno 1- Primaria incompleta 2- Primaria completa 3- Secundaria (Bachillerato) incompleta 4- Secundaria (Bachillerato) completa 5- Técnica o tecnológica incompleta 6- Técnica o tecnológica completa 7- Educación profesional incompleta 8- Educación profesional completa 9- Postgrado 10- No sabe 99- NS/NR
FAMI_EDUCACIONMADRE	Nivel educativo más alto alcanzado por la madre	0- Ninguno 1- Primaria incompleta 2- Primaria completa 3- Secundaria (Bachillerato) incompleta 4- Secundaria (Bachillerato) completa 5- Técnica o tecnológica incompleta 6- Técnica o tecnológica completa 7- Educación profesional incompleta 8- Educación profesional completa 9- Postgrado 10- No sabe 99- NS/NR

FAMI ESTRATOVIVIENDA	Estrato socioeconómico de su vivienda según recibo de energía eléctrica	1- Estrato 1 2- Estrato 2 3- Estrato 3 4- Estrato 4 5- Estrato 5 6- Estrato 6 7- Vive en una zona rural donde no hay estratificación socioeconómica 8- Sin estrato 99- NS/NR
FAMI PERSONASHOGAR	¿Cuántas personas conforman el hogar donde vive actualmente, incluido usted?	1- Una 2- Dos 3- Tres 4- Cuatro 5- Cinco 6- Seis 7- Siete 8- Ocho 9- Nueve 10- Diez 11- Once 12- Doce o más 99- NS/NR
FAMI CUARTOSHOGAR	En total, ¿en cuántos cuartos duermen las personas de su hogar?	1- Una 2- Dos 3- Tres 4- Cuatro 5- Cinco 6- Seis 7- Siete 8- Ocho 9- Nueve 10- Diez o más 99- NS/NR
FAMI TIENEINTERNET	¿Su hogar cuenta con servicio o conexión a internet?	0- No 1- Si 99- NS/NR
FAMI TIENESERVICIOTV	¿Su hogar cuenta con servicio cerrado de televisión?	0- No 1- Si 99- NS/NR
FAMI TIENECOMPUTADOR	¿Cuáles de los siguientes bienes posee su hogar?: Computador	0- No 1- Si 99- NS/NR
FAMI TIENELAVADORA	¿Cuáles de los siguientes bienes posee su hogar?: Máquina lavadora de ropa	0- No 1- Si 99- NS/NR
FAMI TIENEHORNOMICROOGAS	¿Cuáles de los siguientes bienes posee su hogar?: Horno Microondas u Horno eléctrico o a gas	0- No 1- Si 99- NS/NR
FAMI TIENEAUTOMOVIL	¿Cuáles de los siguientes bienes posee su hogar?: Automóvil particular	0- No 1- Si 99- NS/NR
FAMI TIENEMOTOCICLETA	¿Cuáles de los siguientes bienes posee su hogar?: Motocicleta	0- No 1- Si 99- NS/NR

FAMI_NUMLIBROS	¿Cuántos libros físicos o electrónicos hay en su hogar excluyendo periódicos, revistas, directorios telefónicos y libros del colegio?	1- 0 A 10 LIBROS 2- 11 A 25 LIBROS 3- 26 A 100 LIBROS 4- MÁS DE 100 LIBROS 99- NS/NR
ESTU_DEDICACIONINTERNET	Usualmente, ¿cuánto tiempo al día dedica a navegar en internet? Excluya actividades académicas	1- Menos de una hora 2- Entre 1 y 3 horas 3- Más de 4 horas 99- NS/NR
ESTU_HORASSEMANTRABAJA	¿Cuántas horas trabajó usted durante la semana pasada?	0- 0 1- Menos de 10 horas Entre 11 y 20 horas 2- Entre 21 y 30 horas Más de 30 horas 99- NS/NR
RESULTADOS		
MOD_RAZONA_CUANTITAT_DESEM	Nivel de desempeño razonamiento cuantitativo	1- I 2- II 3- III 4- IV
MOD_LECTURA_CRITICA_DESEM	Nivel de desempeño lectura crítica	1- I 2- II 3- III 4- IV
MOD_COMPETEN_CIUADADA_DESEM	Nivel de desempeño competencias ciudadanas	1- I 2- II 3- III 4- IV
MOD_INGLES_DESEM	Nivel de desempeño inglés	1- (-A1) 2- A1 3- A2 4- B1 5- B2
MOD_COMUNI_ESCRITA_DESEM	Nivel de desempeño comunicación escrita	1- I 2- II 3- III 4- IV

12. Apéndice

Apéndice A. Resultados de los estudiantes ingeniería industrial en las pruebas razonamiento cuantitativo 2018.

Periodo	Razonamiento cuantitativo				Total
	I	II	III	IV	
2018	2	20	99	19	141
%	1%	14%	70%	13%	100%

Apéndice B. Resultados de los estudiantes ingeniería industrial en las pruebas lectura crítica 2018.

Periodo	Lectura critica				Total
	I	II	III	IV	
2018	17	51	61	12	141
%	12%	36%	43%	9%	100%

Apéndice C. Resultados de los estudiantes ingeniería industrial en las pruebas Competencias ciudadanas 2018.

Periodo	Competencia ciudadana				Total
	I	II	III	IV	
2018	30	49	60	2	141
%	21%	35%	43%	1%	100%

Apéndice D. Resultados de los estudiantes ingeniería industrial en las pruebas de inglés 2018.

Periodo	Inglés					Total
	(-A1)	A1	A2	B1	B2	
2018	2	21	44	51	23	141

%	1%	15%	31%	36%	16%	100%
---	----	-----	-----	-----	-----	------

Apéndice E. Resultados de los estudiantes ingeniería industrial en las pruebas comunicación escrita 2018.

Comunicación escrita					
Periodo	I	II	III	IV	Total
2018	15	59	45	19	138
%	11%	43%	33%	14%	100%

Apéndice F. Resultados de los estudiantes ingeniería civil en las pruebas razonamiento cuantitativo 2018)

Razonamiento cuantitativo					
Periodo	I	II	III	IV	Total
2018	1	22	129	35	187
%	1%	12%	69%	19%	100%

Apéndice G. Resultados de los estudiantes ingeniería civil en las pruebas lectura crítica 2018.

Lectura crítica					
Periodo	I	II	III	IV	Total
2018	19	67	89	12	187
%	10%	36%	48%	6%	100%

Apéndice H. Resultados de los estudiantes ingeniería civil en las pruebas de competencias ciudadanas 2018.

Competencia ciudadana					
Periodo	I	II	III	IV	Total
2018	51	59	70	7	187
%	27%	32%	37%	4%	100%

Apéndice I. Resultados de los estudiantes ingeniería civil en las pruebas de inglés 2018.

Periodo	Inglés					Total
	(-A1)	A1	A2	B1	B2	
2018	7	34	71	57	18	187
%	4%	18%	38%	30%	10%	100%

Apéndice J. Resultados de los estudiantes ingeniería civil en las pruebas comunicación escrita 2018.

Periodo	Comunicación escrita				Total
	I	II	III	IV	
2018	23	80	62	15	180
%	13%	44%	34%	8%	100%

