

**DETECCIÓN DE PATRONES DE VALOR AGREGADO DE LOS
PROGRAMAS ACADÉMICOS DE LA UNIVERSIDAD DE NARIÑO CON
TÉCNICAS DE BIG DATA**

**CUASPUD ESTRADA DAVID ALEXANDER
HUERTAS REVELO JHON CAMILO**

**PROGRAMA DE INGENIERÍA DE SISTEMAS
DEPARTAMENTO DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE NARIÑO
ENERO, 2023**

**DETECCIÓN DE PATRONES DE VALOR AGREGADO DE LOS
PROGRAMAS ACADÉMICOS DE LA UNIVERSIDAD DE NARIÑO CON
TÉCNICAS DE BIG DATA**

Autores:

CUASPUD ESTRADA DAVID ALEXANDER, dav-id449@hotmail.com
HUERTAS REVELO JHON CAMILO, camilo.199613@gmail.com

Informe final de trabajo de grado presentado como requisito para optar al título de Ingeniero de
Sistemas, en modalidad Investigación.

Director

ING. GUERRERO MATEO

Codirector

ING. PATIÑO JAIRO

**PROGRAMA DE INGENIERÍA DE SISTEMAS
DEPARTAMENTO DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE NARIÑO
ENERO, 2023**

Nota exclusión de responsabilidad intelectual

“Las ideas y conclusiones aportadas en este Trabajo de Grado son responsabilidad de los autores”.

Artículo 1° del Acuerdo No. 324 de octubre 11 de 1966, emanado del Honorable Consejo Directivo de la Universidad de Nariño.

Nota de Aceptación

Firma del director

Firma del jurado evaluador

Firma del jurado evaluador

San Juan de Pasto, enero de 2023

AGRADECIMIENTOS

“A Dios, por darme siempre un día más de vida, darme fortaleza y salud para poder salir adelante y cumplir mis metas. A mis padres Jhon y Patricia por ser mi inspiración día a día, apoyarme siempre para triunfar en la vida y dejarme el mejor legado que es ser un profesional. Al profesor Mateo Guerrero que gracias a su experiencia, conocimiento y confianza para poder sacar el proyecto adelante. Y por último a los profesores que a lo largo de nuestra carrera que siempre fomentaron responsabilidad, buena ética y su conocimiento para ser buenos profesionales y desempeñarnos de la mejor manera en el campo laboral.”

Jhon Camilo Huertas Revelo

“En primer lugar quiero agradecer al profesor Mateo Guerrero quien, con sus conocimientos y apoyo, me guió a través de cada una de las etapas de este proyecto para alcanzar los resultados que buscaba.

Por último, quiero agradecer a todos mis compañeros y a mi familia, por apoyarme aun cuando mis ánimos decaían. En especial, quiero hacer mención de mis padres Richard y Rosa, que siempre estuvieron ahí para darme palabras de apoyo y un abrazo reconfortante para renovar energías.

Muchas gracias a todos.”

David Alexander Cuaspud Estrada

DEDICATORIA

“A mis padres, quienes siempre me han apoyado para salir adelante todos los días, que me han dado su confianza incondicional, en los momentos difíciles me han dado sus consejos, oraciones y han sido mi fortaleza para poder continuar y superarlos.

Los quiero mucho”

Jhon Camilo Huertas Revelo

“A mis padres, por su amor, trabajo y sacrificio en todos estos años, gracias a ustedes he logrado llegar hasta aquí y convertirme en lo que soy. Ha sido un orgullo y privilegio de ser su hijo, son los mejores padres. A mi hermano por estar siempre presente, acompañándome y por el apoyo moral, que me brindo a lo largo de esta etapa de mi vida”

David Alexander Cuaspud Estrada

RESUMEN

En el siguiente documento se presentan los resultados del proyecto de investigación que tuvo como objetivo la detección de patrones de valor agregado de los programas académicos de la Universidad de Nariño que presentaron pruebas Saber 11 y pruebas Saber Pro entre los años 2008 a 2018 con técnicas de Big Data. Para la obtención de patrones se utilizó la herramienta Apache Spark donde se implementó algoritmo a priori para la generación de reglas de asociación y como paso siguiente la creación de árboles de decisión, los cuales brindan información limpia y de calidad que ayudará a directivos de los programas académicos de la universidad de Nariño, a la toma de decisiones para mejorar la calidad educativa de los estudiantes que egresan de la institución.

Palabras Clave: Árbol de decisión, Apache Spark, Big Data, Patrones, Valor agregado, Pruebas Saber 11, Pruebas Saber Pro.

ABSTRACT

The following document presents the results of the research Project that aimed to detect patterns of added value of the academic programs of the University of Nariño that presented Saber 11 tests and Saber Pro tests between the years 2008 to 2018 with techniques of Big Data. To obtain patterns, the Apache Spark tool was used, where an a priori algorithm was implemented for the generation of association rules and as a next step, the creation of decision trees provides clean and quality information that will help as a manager of the programs academic of the University of Nariño, to decision-making to improve the educational quality of students who graduate from the Institution.

Keywords: Decision tree, Apache Spark, Big Data, Patterns, Added value, Saber 11 tests, Saber Pro tests.

TABLA DE CONTENIDO

	Pág.
GLOSARIO.....	15
I. CONTEXTUALIZACIÓN	17
A. GRUPO Y LÍNEA DE INVESTIGACIÓN	17
B. PLANTEAMIENTO DEL PROBLEMA	17
C. JUSTIFICACIÓN	18
D. OBJETIVOS.....	19
E. DELIMITACIÓN	19
II. MARCO TEÓRICO	20
A. MARCO DE ANTECEDENTES	20
B. SUPUESTOS TEÓRICOS	21
III. METODOLOGÍA	28
A. COMPRENSIÓN DEL NEGOCIO.....	28
B. COMPRENSIÓN DE LOS DATOS	28
C. PREPARACIÓN DE LOS DATOS	29
D. MODELADO	33
E. EVALUACIÓN.....	36
F. DESPLIEGUE.....	37
IV. RESULTADOS DE LA INVESTIGACIÓN.....	38
A. COMPRENSIÓN DEL NEGOCIO.....	38
B. COMPRENSIÓN DE LOS DATOS	38
C. PREPARACIÓN DE LOS DATOS	39
D. MODELADO	44
E. EVALUACIÓN.....	79
F. DESPLIEGUE.....	101
V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	108
CONCLUSIONES	109
RECOMENDACIONES	110
REFERENCIAS	111

LISTA DE FIGURAS

	Pág.
Fig. 1. Componentes de la metodología spark	23
Fig. 2. Ciclo de la metodología CRISP-DM	24
Fig. 3. Ciclo del algoritmo A Priori	26
Fig. 4. Variable con su respectiva categoría Fuente: esta investigación.....	28
Fig. 5. Selección de tablas y filtro de estudiantes de la Universidad de Nariño.	29
Fig. 6. Selección de variables socioeconómicas	29
Fig. 7. Limpieza del repositorio inicial	30
Fig. 8. Variable cuantitativa estrato vivienda.....	30
Fig. 9. Repositorio Final Fuente: esta investigación.	30
Fig. 10. Matriz de correlaciones de los estudiantes de la Universidad de Nariño	31
Fig. 11. Estandarización de datos.....	31
Fig. 12. Matriz Estandarizada	32
Fig. 13. Código Componentes principales dataframe	32
Fig. 14. Código generación de biplots.....	33
Fig. 15. Eliminación de variables nulas	34
Fig. 16. Import mlxtend con algoritmo apriori.....	34
Fig. 17. Algoritmo de reglas asociativas	34
Fig. 18. Selección de variables con la vista clase_va.....	35
Fig. 19. Algoritmo train y test	35
Fig. 20. Algoritmo arboles de decisión	36
Fig. 21. Algoritmo árbol de decisión con los mejores parámetros.....	36
Fig. 22. Algoritmo Confusión matrix.....	37
Fig. 23. Valores máximos pruebas Saber 11	38
Fig. 24. Valores máximos pruebas Saber Pro	38
Fig. 25. Calculo valor agregado de estudiantes de la Universidad de Nariño.....	39
Fig. 26. Tabla valor agregado de estudiantes de la Universidad de Nariño	39
Fig. 27. Componentes razonamiento cuantitativo.....	40
Fig. 28. Biplot razonamiento cuantitativo.....	40
Fig. 29. Componentes lectura crítica.....	41
Fig. 30. Biplot lectura critica.....	41
Fig. 31. Componentes Competencias ciudadanas	42
Fig. 32. Biplot Competencias ciudadanas	42
Fig. 33. Componentes Inglés.....	43
Fig. 34. Biplot inglés	43
Fig. 35. Conteo de reglas a priori	44
Fig. 36. Árbol 01 - Competencias Ciudadanas.....	49
Fig. 37. Árbol 02 - Competencias Ciudadanas.....	50
Fig. 38. Árbol 03 - Competencias Ciudadanas.....	51
Fig. 39. Árbol 04 - Competencias Ciudadanas.....	52
Fig. 40. Árbol 05 - Competencias Ciudadanas.....	53
Fig. 41. Árbol 06 - Competencias Ciudadanas.....	54
Fig. 42. Árbol 07 - Competencias Ciudadanas.....	55

Fig. 43. Árbol 08 - Competencias Ciudadanas.....	56
Fig. 44. Árbol 09 - Competencias Ciudadanas.....	56
Fig. 45. Árbol 10 - Competencias Ciudadanas.....	57
Fig. 46. Árbol 01 – inglés.....	58
Fig. 47. Árbol 02 – inglés.....	59
Fig. 48. Árbol 03 – inglés.....	60
Fig. 49. Árbol 04 – inglés.....	61
Fig. 50. Árbol 05 – inglés.....	62
Fig. 51. Árbol 06 – inglés.....	63
Fig. 52. Árbol 07 – inglés.....	64
Fig. 53. Árbol 08 – inglés.....	65
Fig. 54. Árbol 09 – inglés.....	66
Fig. 55. Árbol 10 – inglés.....	66
Fig. 56. Árbol 01 – Lectura critica	67
Fig. 57. Árbol 02 – Lectura critica	68
Fig. 58. Árbol 03 – Lectura critica	68
Fig. 59. Árbol 04 – Lectura critica	69
Fig. 60. Árbol 05 – Lectura critica	69
Fig. 61. Árbol 06 – Lectura critica	70
Fig. 62. Árbol 07 – Lectura critica	70
Fig. 63. Árbol 08 – Lectura critica	71
Fig. 64. Árbol 09 – Lectura critica	71
Fig. 65. Árbol 10 – Lectura critica	72
Fig. 66. Árbol 01 – Razonamiento cuantitativo	73
Fig. 67. Árbol 02 – Razonamiento cuantitativo	74
Fig. 68. Árbol 03 – Razonamiento cuantitativo	74
Fig. 69. Árbol 04 – Razonamiento cuantitativo	75
Fig. 70. Árbol 05 – Razonamiento cuantitativo	75
Fig. 71. Árbol 06 – Razonamiento cuantitativo	76
Fig. 72. Árbol 07 – Razonamiento cuantitativo	76
Fig. 73. Árbol 08 – Razonamiento cuantitativo	77
Fig. 74. Árbol 09 – Razonamiento cuantitativo	77
Fig. 75. Árbol 10 – Razonamiento cuantitativo	78
Fig. 76. Métricas árbol 01 - Competencias ciudadanas.....	79
Fig. 77. Métricas árbol 02 - Competencias ciudadanas.....	80
Fig. 78. Métricas árbol 03 - Competencias ciudadanas.....	80
Fig. 79. Métricas árbol 04 - Competencias ciudadanas.....	81
Fig. 80. Métricas árbol 05 - Competencias ciudadanas.....	81
Fig. 81. Métricas árbol 06 - Competencias ciudadanas.....	82
Fig. 82. Métricas árbol 07 - Competencias ciudadanas.....	82
Fig. 83. Métricas árbol 08 - Competencias ciudadanas.....	83
Fig. 84. Métricas árbol 09 - Competencias ciudadanas.....	83
Fig. 85. Métricas árbol 10 - Competencias ciudadanas.....	84
Fig. 86. Métricas árbol 01 – inglés.....	85
Fig. 87. Métricas árbol 02 – inglés.....	85
Fig. 88. Métricas árbol 03 – inglés.....	86
Fig. 89. Métricas árbol 04 – inglés.....	86

Fig. 90. Métricas árbol 05 – inglés.....	87
Fig. 91. Métricas árbol 06 – inglés.....	87
Fig. 92. Métricas árbol 07 – inglés.....	88
Fig. 93. Métricas árbol 08 – inglés.....	88
Fig. 94. Métricas árbol 09 – inglés.....	89
Fig. 95. Métricas árbol 10 – inglés.....	89
Fig. 96. Métricas árbol 01 - Lectura critica.....	90
Fig. 97. Métricas árbol 02 - Lectura critica.....	91
Fig. 98. Métricas árbol 03 - Lectura critica.....	91
Fig. 99. Métricas árbol 04 - Lectura critica.....	92
Fig. 100. Métricas árbol 05 - Lectura critica.....	92
Fig. 101. Métricas árbol 06 - Lectura critica.....	93
Fig. 102. Métricas árbol 07 - Lectura critica.....	93
Fig. 103. Métricas árbol 08 - Lectura critica.....	94
Fig. 104. Métricas árbol 09 - Lectura critica.....	94
Fig. 105. Métricas árbol 10 - Lectura critica.....	95
Fig. 106. Métricas árbol 01 - Razonamiento cuantitativo.....	96
Fig. 107. Métricas árbol 02 - Razonamiento cuantitativo.....	96
Fig. 108. Métricas árbol 03 - Razonamiento cuantitativo.....	97
Fig. 109. Métricas árbol 04 - Razonamiento cuantitativo.....	97
Fig. 110. Métricas árbol 05 - Razonamiento cuantitativo.....	98
Fig. 111. Métricas árbol 06 - Razonamiento cuantitativo.....	98
Fig. 112. Métricas árbol 07 - Razonamiento cuantitativo.....	99
Fig. 113. Métricas árbol 08 - Razonamiento cuantitativo.....	99
Fig. 114. Métricas árbol 09 - Razonamiento cuantitativo.....	100
Fig. 115. Métricas árbol 10 - Razonamiento cuantitativo.....	101

LISTA DE TABLAS

Pág.

TABLA I REGLA 1 - RAZONAMIENTO CUANTITATIVO.....	44
TABLA II REGLA 2 - RAZONAMIENTO CUANTITATIVO	45
TABLA III REGLA 3 - RAZONAMIENTO CUANTITATIVO	45
TABLA IV REGLA 1 - COMPETENCIAS CIUDADANAS.....	46
TABLA V REGLA 2 - COMPETENCIAS CIUDADANAS	46
TABLA VI REGLA 3 - COMPETENCIAS CIUDADANAS.....	47
TABLA VII REGLA 1 - INGLÉS.....	47
TABLA VIII REGLA 2 - INGLÉS	48
TABLA IX REGLA 3 - INGLÉS.....	48
TABLA X MATRIZ DE CONFUSIÓN ÁRBOL 01 - COMPETENCIAS CIUDADANAS.....	79
TABLA XI MATRIZ DE CONFUSIÓN ÁRBOL 02 - COMPETENCIAS CIUDADANAS	79
TABLA XII MATRIZ DE CONFUSIÓN ÁRBOL 03 - COMPETENCIAS CIUDADANAS.....	80
TABLA XIII MATRIZ DE CONFUSIÓN ÁRBOL 04 - COMPETENCIAS CIUDADANAS.....	81
TABLA XIV MATRIZ DE CONFUSIÓN ÁRBOL 05 - COMPETENCIAS CIUDADANAS	81
TABLA XV MATRIZ DE CONFUSIÓN ÁRBOL 06 - COMPETENCIAS CIUDADANAS.....	82
TABLA XVI MATRIZ DE CONFUSIÓN ÁRBOL 07 - COMPETENCIAS CIUDADANAS	82
TABLA XVII MATRIZ DE CONFUSIÓN ÁRBOL 08 - COMPETENCIAS CIUDADANAS	83
TABLA XVIII MATRIZ DE CONFUSIÓN ÁRBOL 09 - COMPETENCIAS CIUDADANAS.....	83
TABLA XIX MATRIZ DE CONFUSIÓN ÁRBOL 10 - COMPETENCIAS CIUDADANAS	84
TABLA XX MATRIZ DE CONFUSIÓN ÁRBOL 01 - INGLÉS	84
TABLA XXI MATRIZ DE CONFUSIÓN ÁRBOL 02 - INGLÉS	85
TABLA XXII MATRIZ DE CONFUSIÓN ÁRBOL 03 - INGLÉS.....	85
TABLA XXIII MATRIZ DE CONFUSIÓN ÁRBOL 04 - INGLÉS	86
TABLA XXIV MATRIZ DE CONFUSIÓN ÁRBOL 05 - INGLÉS	86
TABLA XXV MATRIZ DE CONFUSIÓN ÁRBOL 06 - INGLÉS.....	87
TABLA XXVI MATRIZ DE CONFUSIÓN ÁRBOL 07 - INGLÉS	87
TABLA XXVII MATRIZ DE CONFUSIÓN ÁRBOL 08 - INGLÉS.....	88
TABLA XXVIII MATRIZ DE CONFUSIÓN ÁRBOL 09 - INGLÉS.....	89
TABLA XXIX MATRIZ DE CONFUSIÓN ÁRBOL 10 - INGLÉS	89
TABLA XXX MATRIZ DE CONFUSIÓN ÁRBOL 01 - LECTURA CRITICA	90
TABLA XXXI MATRIZ DE CONFUSIÓN ÁRBOL 02 - LECTURA CRITICA	90
TABLA XXXII MATRIZ DE CONFUSIÓN ÁRBOL 03 - LECTURA CRITICA.....	91
TABLA XXXIII MATRIZ DE CONFUSIÓN ÁRBOL 04 - LECTURA CRITICA	91
TABLA XXXIV MATRIZ DE CONFUSIÓN ÁRBOL 05 - LECTURA CRITICA	92
TABLA XXXV MATRIZ DE CONFUSIÓN ÁRBOL 06 - LECTURA CRITICA.....	93
TABLA XXXVI MATRIZ DE CONFUSIÓN ÁRBOL 07 - LECTURA CRITICA	93
TABLA XXXVII MATRIZ DE CONFUSIÓN ÁRBOL 08 - LECTURA CRITICA.....	94
TABLA XXXVIII MATRIZ DE CONFUSIÓN ÁRBOL 09 - LECTURA CRITICA.....	94
TABLA XXXIX MATRIZ DE CONFUSIÓN ÁRBOL 10 - LECTURA CRITICA	95
TABLA XL MATRIZ DE CONFUSIÓN ÁRBOL 01 - RAZONAMIENTO CUANTITATIVO	95
TABLA XLI MATRIZ DE CONFUSIÓN ÁRBOL 02 - RAZONAMIENTO CUANTITATIVO.....	96
TABLA XLII MATRIZ DE CONFUSIÓN ÁRBOL 03 - RAZONAMIENTO CUANTITATIVO	96
TABLA XLIII MATRIZ DE CONFUSIÓN ÁRBOL 04 - RAZONAMIENTO CUANTITATIVO	97
TABLA XLIV MATRIZ DE CONFUSIÓN ÁRBOL 05 - RAZONAMIENTO CUANTITATIVO.....	97
TABLA XLV MATRIZ DE CONFUSIÓN ÁRBOL 06 - RAZONAMIENTO CUANTITATIVO	98
TABLA XLVI MATRIZ DE CONFUSIÓN ÁRBOL 07 - RAZONAMIENTO CUANTITATIVO.....	99

TABLA XLVII	MATRIZ DE CONFUSIÓN ÁRBOL 08 - RAZONAMIENTO CUANTITATIVO.....	99
TABLA XLVIII	MATRIZ DE CONFUSIÓN ÁRBOL 09 - RAZONAMIENTO CUANTITATIVO ...	100
TABLA XLIX	MATRIZ DE CONFUSIÓN ÁRBOL 10 - RAZONAMIENTO CUANTITATIVO.....	100

GLOSARIO

APACHE SPARK: es un motor de código abierto para el almacenamiento, procesamiento y análisis de grandes volúmenes de datos. Grandes compañías como CISCO, IBM, NASA lo usan para aplicaciones con minería de datos.

ÁRBOL DE DECISIÓN: es un mapa de posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una entidad de cualquier naturaleza pueda comparar acciones posibles entre si según el criterio establecido y tomar la mejor decisión posible.

BIG DATA: es un término que describe grandes volúmenes de un conjunto de datos o combinaciones de conjuntos de datos, para su procesamiento se requiere herramientas especializadas debido a la alta complejidad y tamaño de los datos.

PATRONES: es una serie de variables constantes, identificables dentro de un conjunto mayor de datos. En minería de datos, los patrones explican el comportamiento de los datos en un contexto específico.

VALOR AGREGADO: el concepto de valor agregado en el sistema educativo se refiere al logro de los estudiantes, expresado como la progresión del conocimiento, habilidades y aptitudes que estos han obtenido en su paso por la institución de educación.

INTRODUCCIÓN

Con la creciente complejidad de los sistemas de información utilizados por la educación media y superior en el país, es importante analizar la gran cantidad de datos que generan estos sistemas, por lo tanto, el procesamiento de estos volúmenes de información utilizando minería de datos facilitaría el planteamiento de acciones eficaces que permitan a las entidades educativas del país tomar decisiones efectivas y significativas para mejorar la calidad de la educación. Dicho procesamiento consiste en la medición de eficiencia en el proceso de aprendizaje que recibe el estudiante a lo largo de su carrera, esto se conoce como valor agregado de la institución. Tomando la definición de valor agregado propuesto por ICFES “El valor agregado se entiende como la contribución al progreso neto de los estudiantes, hacia objetivos de aprendizaje establecidos, una vez eliminada la influencia de otros factores ajenos a la institución que pueden contribuir a ese progreso” [1]. En contextos internacionales, según el artículo del Observatorio de Políticas de Evaluación Educativa [2] se informó acerca del uso de medidas de valor agregado para estimar el trabajo de instituciones de educación y docentes. Dicho uso es creciente en los Estados Unidos y en Reino Unido, países que cuentan con grandes bases de información sobre el desempeño de sus estudiantes y sobre diversidad de variables relacionadas con su entorno, sus instituciones y sus docentes. Actualmente en Colombia existen proyectos enfocados a la medición del valor agregado de las competencias ciudadanas en la educación superior [3], sin embargo, no se utiliza técnicas de minería de datos para el procesamiento de la información.

El enfoque del proyecto es el análisis de datos de los estudiantes de la Universidad de Nariño, para identificar patrones de valor agregado de la educación superior, lo anterior es aplicado con herramientas de minería de datos, teniendo como objetivo, generar nuevos conocimientos a partir del análisis de la información y posteriormente realizar planes de acción que permitan mejorar la calidad de la educación prestada en la Universidad de Nariño.

En el transcurso del documento se evidenciarán los procesos realizados que comprenden la obtención de patrones de valor agregado como la construcción del repositorio limpio de datos, análisis de variables y generación de árboles de decisión que permitirán dar una idea al desarrollo del proyecto.

Es necesario indicar que para la obtención de datos e información relevante sobre el problema planteado es obtenida mediante bases de datos Saber 11 de 2008 a 2018 y Saber Pro desde 2012 hasta 2018, las cuales tienen información actualizada con respecto al tema del proyecto.

Este documento está estructurado de la siguiente manera. En primer lugar, se tiene la Contextualización la cual contiene información relacionada con el grupo y línea de investigación, planteamiento del problema, justificación y objetivos, sección con la cual se introduce al lector a entender el problema, por qué se realizó la investigación, y el aporte que genera a la Universidad de Nariño. Posteriormente, se encuentra el Marco teórico donde se describen los antecedentes que son estudios similares que se han hecho sobre el tema en cuestión, y supuestos teóricos que se muestran como definiciones clave de herramientas y técnicas que fueron utilizadas para realizar la investigación. A continuación, en la Metodología se muestra una descripción detallada de las actividades y tareas que se realizaron para la ejecución del proyecto. En la siguiente sección se tienen los resultados de la investigación donde se detallan los modelos que se obtuvieron a partir de todo el proceso de investigación. Finalmente se encuentran las conclusiones.

I. CONTEXTUALIZACIÓN

A. GRUPO Y LÍNEA DE INVESTIGACIÓN

La línea de investigación en la que se inscribe este proyecto es: Línea Software y manejo de Información. [4]

B. PLANTEAMIENTO DEL PROBLEMA

Para los entes de control de la educación del país es de vital importancia determinar la calidad de aprendizaje de la comunidad estudiantil, es por eso que se aplica el Examen de Estado Saber Pro realizado por el ICFES cada año, que permite evaluar a los estudiantes que están en proceso de formación universitaria mediante cuatro componentes principales las cuales son: Competencias Ciudadanas, Inglés, Lectura Crítica y Razonamiento Cuantitativo, con el fin de determinar lo que han logrado durante su preparación. Con la obtención de estos datos y con el uso de técnicas de minería de datos se pretende descubrir patrones de valor agregado para medir la calidad y congruencia de los conocimientos adquiridos.

Las instituciones de educación superior tienen como fin construir el capital humano de los estudiantes, evaluando sus competencias genéricas o competencias específicas propias de cada profesión. Para la medición de la calidad educativa, las instituciones deben tomar en cuenta todos los aspectos que influyen en los resultados de los estudiantes, el conocimiento y condiciones tanto de la etapa de ingreso como de culminación. Si las competencias resultan en ganancias, se afirma que la institución genera valor agregado en la formación de sus estudiantes.

Dentro de los programas académicos de la Universidad de Nariño, no se ha efectuado un estudio para detectar patrones de valor agregado que obtienen los estudiantes a lo largo de su carrera, con lo cual, no se ha logrado tener una idea clara de cómo los programas académicos han afectado positivamente o negativamente las habilidades y conocimientos del estudiante desde que inició el programa de pregrado hasta que presentó la prueba Saber Pro.

Por lo anterior, se plantean algunos aspectos importantes dentro de la investigación: integrar y transformar un conjunto de datos a partir de las bases de datos de Saber 11 y Saber Pro que permita la aplicación de técnicas de Big Data para la detección de patrones de valor agregado; evaluar la calidad de patrones de valor agregado de los programas académicos de la Universidad de Nariño; y fomentar la continuación del estudio mediante esta técnica para los estudiantes quienes presenten la prueba Saber Pro a partir del año 2019 en adelante.

C. JUSTIFICACIÓN

En la actualidad, la calidad de la Educación Superior en Colombia debe estar presente en las instituciones educativas tanto de media académica como de nivel superior con el objetivo de evaluar el conocimiento que se brinda a los estudiantes a lo largo de su formación en las diferentes áreas del aprendizaje, que les permita desarrollar habilidades y competencias para que puedan desenvolverse en la sociedad.

El saber interpretar datos es de gran importancia para llevar al éxito y a diferenciarse de otras instituciones. Esto permitirá que se pueda evaluar la forma de enseñanza aplicada en los estudiantes, llevando a tomar decisiones de tal manera que se pueda mejorar el proceso académico en los estudiantes de la Universidad de Nariño. Además, que asegure que la institución brinde un valor agregado para el capital humano pues aportan de manera didáctica al ámbito de la investigación y la formación de los futuros profesionales.

De lo anterior, surge esta investigación que tiene como objetivo la detección de patrones de valor agregado de los programas académicos de la Universidad de Nariño, obteniendo información valiosa que contribuirá a mejorar la formación académica y los métodos de enseñanza que se aplican a los estudiantes.

La investigación es interesante ya que para la obtención de patrones de valor agregado se generan modelos de predicción los cuales se conocen como árboles de decisión que se infieren a partir del repositorio de datos y utilizando inteligencia artificial se crean los diagramas de construcciones lógicas, que, en otros términos, representan una serie de condiciones socioeconómicas de los estudiantes que se presentan de manera sucesiva en forma de reglas. [5]

La investigación es útil porque si se aplica la metodología descrita permitirá el manejo de grandes volúmenes de datos de manera sencilla y aplicando los algoritmos se puede continuar con el estudio en años posteriores.

La investigación es novedosa porque no se ha realizado un estudio de generación de patrones de valor agregado utilizando técnicas de Big Data para los estudiantes de todos los programas de la Universidad de Nariño, estas técnicas permiten comprender de manera eficiente los resultados obtenidos a partir del repositorio de datos transformado, producto de las bases de datos Saber 11 y Saber Pro.

D. OBJETIVOS

1) Objetivo general

Detectar patrones de valor agregado en los programas académicos de la Universidad de Nariño utilizando técnicas de Big Data que genere información de calidad para la toma de decisiones en cuanto al mejoramiento de la calidad de la educación.

2) Objetivos específicos

- Apropiar el conocimiento sobre técnicas de Big Data.
- Limpiar, normalizar y estructurar la base de datos para la inferencia del valor agregado.
- Definir los algoritmos necesarios para la obtención del valor agregado mediante la aplicación de técnicas de Big Data.
- Interpretar los patrones obtenidos de valor agregado.

E. DELIMITACIÓN

La presente investigación tiene el propósito de detectar patrones de valor agregado en la formación de los estudiantes de los programas académicos de la Universidad de Nariño con el uso de base de datos de las Pruebas Saber Pro de los años 2012 - 2018, para ello se creará un repositorio datos limpio y transformado haciendo uso de la herramienta Apache Spark de Big Data

II. MARCO TEÓRICO

A. MARCO DE ANTECEDENTES

FORMULACIÓN DE INDICADORES PARA DETERMINAR EL VALOR AGREGADO EN LA FORMACIÓN DE ESTUDIANTES DEL PROGRAMA DE INGENIERÍA DE SISTEMAS SEDE PASTO DE LA UNIVERSIDAD DE NARIÑO.

Este proyecto se enfocó en el estudio de los indicadores que determinan el valor agregado de la educación recibida por los estudiantes a lo largo de su formación del programa de ingeniería de sistemas sede Pasto, para establecer que tanto aportó el programa de estudios y la enseñanza de los docentes. [6]

Este proyecto de investigación docente es un referente fundamental para el proyecto de grado por su enfoque a la obtención de indicadores que contribuyen a determinar valor agregado a la educación del estudiante, se utilizó la base de datos de las Pruebas Saber 11 y Saber Pro de los años 2006 a 2018.

MODELO DE VALOR AGREGADO PARA LA CARRERA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN A PARTIR DE SABER PRO 2015 Y 2016

Este proyecto fue construido para automatizar el cálculo del valor agregado desarrollando un software que utilice el modelo lineal jerárquico de la Universidad Católica de Colombia aplicado a las competencias genéricas Saber Pro 2016-2017 y Saber 11, para medir el efecto aula que las IES tienen sobre sus estudiantes. [7]

Para el desarrollo de este proyecto se utilizó el lenguaje de programación Python junto con bibliotecas como pandas, sqlalchemy y statsmodel. Para el desarrollo de la interfaz gráfica se utilizó el framework Web2Py, éste framework utiliza JQuery con un template Bootstrap. Como DBMS se utilizó MariaDB.

De este proyecto se tomó en cuenta la integración del lenguaje Python con las bases de datos SQL, esto fue utilizado en los books de Apache Spark para la generación de algoritmos.

DESARROLLO DE UNA HERRAMIENTA DE ANÁLISIS, PROCESADO Y VISUALIZACIÓN DE FUENTES DE DATOS EMPLEANDO TÉCNICAS DE BIG DATA UTILIZANDO APACHE SPARK

En el presente trabajo de grado se abarca el estudio del conjunto de técnicas Big Data, que permiten el análisis, procesado y visualización de grandes cantidades de datos. El objetivo es una familiarización con el concepto de Big Data y todas las herramientas que lo engloban, para después proceder al montaje de una arquitectura en completo y finalizar con un ejemplo de uso con la herramienta Apache Spark como motor de procesamiento. [8]

De estos estudios, se toma como referencia el uso de la tecnología Apache Spark que se emplea como herramienta junto con las técnicas de Big Data para validación del repositorio limpio de datos.

CONSTRUCCIÓN DE UN REPOSITORIO LIMPIO DE DATOS DE LOS ESTUDIANTES DE LOS PROGRAMAS DE LA UNIVERSIDAD DE NARIÑO PARA ANÁLISIS DE DESERCIÓN CON MINERÍA DE DATOS

En la presente investigación se construyó un repositorio de datos limpio y adecuado al objeto de estudio, transformado con los factores académicos y socioeconómicos de los estudiantes de los programas de pregrado de la Universidad de Nariño. La recolección de datos se hizo a partir de la información histórica que reposa en la Oficina de Registro y Control Académico (OCARA), se utilizó la metodología CRISP-DM donde se aplicaron las tres primeras fases y están relacionadas con la construcción del repositorio. [9]

De estos estudios se toma como referencia la herramienta de PostgreSQL para la construcción del repositorio, y la metodología CRISP-DM la cual la aplicamos en el desarrollo de nuestra investigación.

CALCULO DE VALOR AGREGADO GENERADO POR LA UNIVERSIDAD DE NARIÑO EN RELACION A LAS PRUEBAS SABER 11 – SABER PRO 2010-2014.

En Colombia se han realizado varios estudios como por ejemplo el artículo realizado por Burgos, M., Rúales, K., Bastidas, Y., Ortiz, E., que surge como resultado del proyecto de investigación con el mismo nombre realizado con financiación otorgada por la Universidad de Nariño a través del Sistema de Investigaciones en la Convocatoria Docente, cuyo objetivo es la medición y análisis de los efectos generados por la Universidad de Nariño en los estudiantes que culminaron los estudios de pregrado en el periodo 2010-2014, mediante la estimación de valor agregado su citado por la Universidad de Nariño a partir del cálculo de la diferencia entre los resultados de las Pruebas Saber 11 y Saber Pro; Por ende, este cálculo surge de la necesidad de evaluar el papel de la Universidad de Nariño en la instrucción de sus estudiantes en las competencias básicas y ciudadanas. [10]

De esta investigación se toma como referencia el estudio que se hace por competencias genéricas y ciudadanas de los estudiantes y pautas para la redacción de resultados.

B. SUPUESTOS TEÓRICOS

GESTIÓN DEL CONOCIMIENTO

La gestión del conocimiento es un método para la identificación, recopilación, depuración, distribución y evaluación del conocimiento organizacional. En entidades ésta disciplina se considera el activo más importante, el conjunto de habilidades y capacidades de la propia organización y de las personas que la componen son el elemento principal que permite a las organizaciones una precisa toma de decisiones.

Según The Gartner Group: “La Gestión del Conocimiento es la disciplina que promueve un enfoque integral a la identificación, captura, evaluación, recuperación y el compartir todos los activos de información de una empresa, para generar valor y nuevas oportunidades” [11]

La presente investigación busca a partir de la recolección de datos relacionados con el valor agregado, generar conocimiento mediante el empleo de técnicas de análisis definidas en Big Data.

BIG DATA

Es un término que describe un conjunto de datos estructurados o no estructurados cuyo volumen, variabilidad y velocidad de crecimiento dificulta su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales como las bases de datos estadísticas convencionales.

La gran utilidad de Big Data es el hecho de que proporciona respuestas a muchas preguntas gracias a la cantidad tan grande de información que maneja, moldeando los datos o probando de cualquier manera que alguna entidad considere adecuada. Al hacerlo, las organizaciones son capaces de identificar los problemas de una forma más comprensible. [12]

La presente investigación cuenta con una gran base de datos no estructurada que será ordenada y procesada mediante el uso de herramientas de Big Data.

MINERÍA DE DATOS

Es un proceso para clasificar grandes volúmenes de datos, con el fin de encontrar información importante y útil para una finalidad específica, el principal enfoque de la minería de datos es la detección de patrones.

Su base comprende tres disciplinas científicas que funcionan conjuntamente:

- Estadística que comprende el estudio numérico de relaciones de datos
 - Inteligencia artificial
 - Machine Learning que comprende aprendizaje de datos para hacer predicciones.
- [13]

VALOR AGREGADO

Tomando la definición de “Taylor R.S” [14] se explica el valor agregado como:

El sistema agrega valor a toda la información que transmite o maneja en el proceso de generación, selección, visualización, recuperación, distribución y uso, permitiendo a la persona obtener un flujo de información relevante que le puede ayudar a tomar una conclusión o a la toma de una decisión.

VALOR AGREGADO EN LA EDUCACIÓN

Los estudios de valor agregado varían en diferentes formas. En primer lugar, los estudios pueden diferir en la unidad de análisis, centrándose algunos en las instituciones y otros en los programas o en los profesores. En segundo lugar, los estudios pueden concentrarse en la asociación de distintos factores con el valor agregado o en las variables que deben ser consideradas como parte de las condiciones iniciales del individuo.

Finalmente, los estudios también difieren en la medida en que logren identificar qué explica la asociación de los factores estudiados con el valor agregado estimado, encontrándose en esta literatura tanto estudios descriptivos y de análisis de determinantes como estudios que realizan inferencia causal. [15]

SPARK

Spark es una plataforma de código libre para procesamiento en clústeres que está orientada a manejar grandes volúmenes de datos y ejecutar procesamiento computarizado intensivo sobre

ellos; debido a la eficiencia en Big Data se dice que es la evolución de las técnicas de procesamiento de datos, que nos ofrece varias ventajas y reduce significativamente los tiempos de ejecución. [16]

Las ventajas más notables de Spark son:

- 100% Open Source
- Soporte de múltiples lenguajes
- Hasta 100 veces más rápida que MapReduce
- Tiene módulos para implementar Machine Learning, Streaming, acceso a datos, grafos.

Spark se compone de los siguientes elementos:

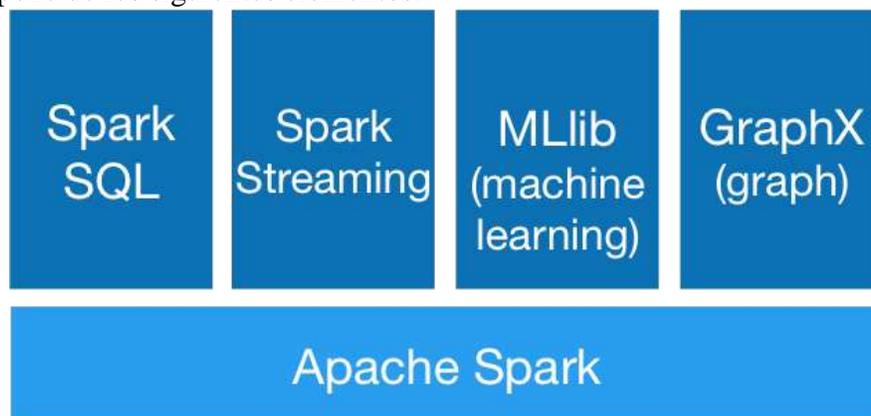


Fig. 1. Componentes de la metodología spark
Fuente: [17]

Spark SQL: Nos permite acceder a los datos de forma estructurada e integrar Spark con diferentes herramientas como Hive, ODBC, JDBC y herramientas BI

Spark Streaming: Ofrece soporte para procesamiento en casi tiempo real a través de un sistema de empaquetamiento por lotes.

MLlib: Incluye una biblioteca de algoritmos de machine learning, compatibles para ejecutar con Spark

GraphX: Es un API para computación paralela de gráficos.

Para esta investigación es de vital importancia el uso de esta herramienta, que nos brinda un entorno de programación para el desarrollo de todo el procesamiento de la información y obtención de los resultados.

DBEAVER:

Es una herramienta de software de gran utilidad que permite la gestión de bases de datos, multiplataforma y de código abierto, la característica principal es la incorporación de Bases de Datos SQL y NoSQL.

Permite la conexión con múltiples gestores de bases de datos SQL como lo son: PostgreSQL, MySQL, MariaDB, Oracle, SQLite y los gestores NoSQL como MongoDB, Cassandra. [18]

La segunda es la discriminación de datos (Data Discrimination), que es una comparación entre las características generales de los objetos de una clase respecto a las de otro conjunto contrastante. Finalmente, también se puede aplicar una combinación de ambas.

- **Análisis de asociación:** Es el descubrimiento de reglas de asociación que muestran condiciones del tipo atributo-valor que ocurre con frecuencia dentro de un conjunto de datos.

La minería mediante reglas de asociación es el proceso de búsqueda interesante de correlaciones entre un conjunto grande de datos. El descubrimiento de reglas de asociación en grandes volúmenes de transacciones de negocios, puede facilitar el proceso de toma de decisiones.

En la presente investigación, se busca a partir de las técnicas descriptivas de datos la limpieza y estructuración de la información, y mediante algoritmos encontrar las reglas de asociación más coherentes para la investigación.

TÉCNICAS PREDICTIVAS DE DATOS

CLASIFICACIÓN Y PREDICCIÓN

Son dos tipos de análisis de datos, aquellos que pueden ser usados para clasificar datos y los que se usan para predecir tendencias. La clasificación de datos predice clases de etiquetas mientras la predicción de datos predice funciones de valores continuos. Aplicaciones típicas incluyen análisis de riesgo para préstamos y predicciones de crecimiento. Algunas técnicas para clasificación de datos incluyen: clasificaciones bayesianas. K-NearestNeighbor, algoritmos en éticos, entre otros. [22]

En la presente investigación, se busca a partir de las técnicas predictivas de datos, la clasificación de reglas de asociación y mediante algoritmos obtener predicciones para lograr los objetivos propuestos.

ARBOLES DE DECISIÓN

Definen un conjunto de clases, asignando a cada dato de entrada una clase y determina la probabilidad de que ese registro pertenezca a la clase.

Podemos distinguir dos tipos de árboles, el primero es el árbol de decisión de clasificación, donde cada registro a clasificar fluye por una rama del árbol. La rama a seguir es determinada por una serie de preguntas definidas por los nodos de la rama. Cuando el registro llega a un nodo hoja, se le asigna a la clase del nodo hoja.

El segundo es el árbol de decisión de regresión, cuando el registro llega a un nodo hoja, a la variable de salida de ese nodo, se le asigna el promedio de los valores de la variable de salida de los registros que cayeron en ese nodo hoja durante el proceso de entrenamiento.

En este proyecto se requiere encontrar árboles de decisión de clasificación, para un análisis y obtención de patrones de valor agregado. [23]

ALGORITMO A PRIORI [24]:

Se trata de una serie de pasos a seguir para el descubrimiento de elementos frecuentes dentro de una base de datos. Es una técnica de minería de datos que realiza dos funciones principales: Unión y Poda. En la Unión éste genera elementos a partir de los conjuntos de datos para posteriormente unir los resultados. En la poda se analiza el conteo de cada elemento en la base de datos.

A continuación, se presenta un gráfico descriptivo del proceso que realiza el algoritmo A Priori:

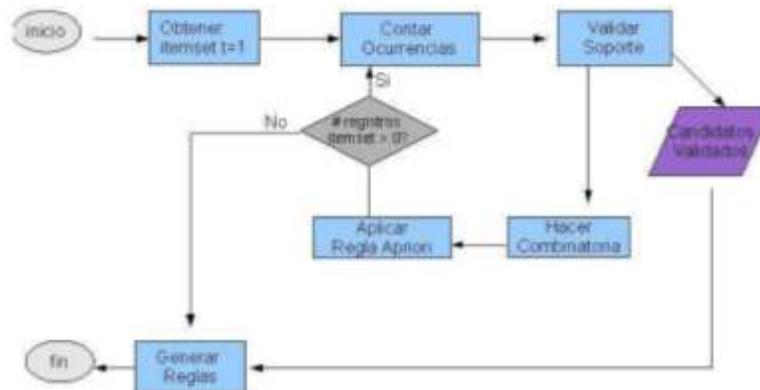


Fig. 3. Ciclo del algoritmo A Priori

Fuente: [25]

REGLAS DE ASOCIACIÓN:

Las reglas de asociación [26] son condiciones que ayudan a mostrar la probabilidad de relación entre los subconjuntos de datos, dentro de grandes volúmenes de datos en diversas clases de bases de datos.

Se utilizan modelos de Machine Learning para analizar los datos en busca de patrones, o concurrencias, en un conjunto de datos. Una regla de asociación se divide en dos partes:

- **ANTECEDENTE:** Se describe como un elemento que se encuentra dentro del dataset.
- **CONSECUENTE:** Se describe como el elemento que se encuentra en combinación con el antecedente.

En la presente investigación, se descubrieron reglas de asociación interesantes y con buen soporte, que nos permitieron generar arboles de decisión.

La medición de la efectividad de las reglas de asociación se mide bajo los siguientes ítems:

SOPORTE Y CONFIANZA:

Para la evaluación de las reglas de asociación se toman en cuenta estos dos ítems importantes. El soporte es el número de transacciones que contiene una regla en la base de datos, mientras que la confianza define el porcentaje de fiabilidad de los datos obtenidos a partir de la base de datos.

LIFT:

Éste ítem estadístico compara la frecuencia observada de una regla con la frecuencia esperada. Si la regla aparece con mayor frecuencia sobre el conjunto de datos, ésta tendrá más confianza y soporte para la validación de datos. [27]

MÉTRICAS DE CLASIFICACIÓN [28]:

Se trata de un conjunto de valores resultantes del análisis de un conjunto de datos, que nos ayudan a medir la efectividad de un modelo en cualquier campo de estudio. A continuación, se describirán las siguientes métricas:

- **MATRIZ DE CONFUSIÓN:** Describe el rendimiento de un modelo en los datos de prueba, donde se desconocen los verdaderos valores. Mediante el uso de esta métrica le facilita al sistema la detección de los valores aportantes y los valores errados.
- **PRECISIÓN:** Es el porcentaje de elementos o valores clasificados correctamente en un modelo.

PRUEBAS SABER 11:

El examen de estado de la educación media del país, Saber 11, es una herramienta de evaluación estandarizada que mide oficialmente la calidad de la educación formal de las personas que han concluido la educación secundaria.

Saber 11 está compuesto por cinco pruebas: Ciencias Naturales, Matemáticas, Lectura Crítica, Sociales y Ciudadanas e Inglés. [29]

PRUEBAS SABER PRO:

El examen de estado de la calidad de la educación superior, Saber Pro, que permite medir externamente la calidad de la educación superior que evalúa las competencias de los estudiantes que están próximos a terminar los diversos programas profesionales universitarios.

Saber Pro está compuesto por competencias de dos tipos: genéricas y específicas. El primero conjunto evalúa cinco módulos: Razonamiento Cuantitativo, Lectura Crítica, Competencias Ciudadanas, Comunicación escrita e inglés. Y el segundo conjunto se trata de una serie de pruebas específicas de acuerdo al programa de formación del estudiante. [30]

III. METODOLOGÍA

El tipo de esta investigación es cuantitativa porque se emplearon métodos para obtención de patrones del valor agregado y su medición mediante métricas, la población de estudio comprende estudiantes que cursaron sus programas pregrado entre los periodos 2012 a 2018, se utilizaron herramientas como PostgreSQL para la importación inicial de la base de datos que comprende Pruebas Saber 11 de 2008 a 2018 y Saber Pro de 2012 a 2018, Anaconda que fue integrada con Apache Spark para la generación de la interfaz de trabajo en la cual se realizó la limpieza de los datos, creación del repositorio, clasificación de variables, generación de reglas de asociación, construcción de árboles de decisión y métricas de evaluación, se utilizó DBeaver como herramienta de administración de bases de datos, en la cual se hizo los cruces con las bases de Saber 11 y Saber Pro para sacar las variables más determinantes que se trabajaron en el proyecto.

En este capítulo se describen los procesos de construcción, limpieza y transformación de datos para obtener un repositorio limpio. Luego se analizan una a una las variables socioeconómicas que permitan identificar correlaciones; con el objetivo de descubrir patrones interesantes que ayuden a encontrar valor agregado en los estudiantes de la Universidad de Nariño utilizando la metodología CRISP-DM

A. *COMPRESIÓN DEL NEGOCIO*

CONSTRUCCIÓN MARCO TEÓRICO

En esta fase, se realizaron las actividades que permitieron profundizar y apropiar de una manera completa el problema objeto de estudio, los objetivos y los requisitos de esta investigación, que posibilitaron analizar las diferentes variables que componen la base de datos y así poder consolidar e interpretar adecuadamente los resultados. En esta fase, descubrir patrones interesantes que generen valor agregado para la Universidad de Nariño, se convirtió en un problema a resolver con Big data. Como resultado de la fase de comprensión se construyó un marco teórico de diferentes técnicas de Big Data para abordar el problema; contemplado en el capítulo anterior

B. *COMPRESIÓN DE LOS DATOS*

Una vez construido el marco teórico, se procedió a analizar la base de datos de las pruebas SABER PRO tomando como población a los estudiantes de la Universidad de Nariño. Construyendo una tabla donde se calcula el valor agregado de cada estudiante en cada competencia que evalúa el ICFES. Luego se analizaron las variables socioeconómicas de las pruebas SABER PRO de los años 2012 a 2018 con sus respectivas categorías con la herramienta apache Spark.

```
In [86]: df.groupby("fami_hogaractual").count().show()

+-----+-----+
|  fami_hogaractual| count|
+-----+-----+
|Es habitual o per...|102843|
|Es temporal por r...| 22937|
|                   |  25024|
+-----+-----+
```

Fig. 4. Variable con su respectiva categoría
Fuente: esta investigación.

Dando como resultado que muchas variables se encontraban en todos los años y otras que se podrían descartar lo que sirvió de base para las siguientes fases.

C. PREPARACIÓN DE LOS DATOS

Para la construcción del repositorio se recopilarán, limpiaran y transformaran los datos a partir de las bases de datos de las pruebas SABER 11 Y SABER PRO, donde se clasificarán teniendo en cuenta las relaciones entre las variables socioeconómicas y las competencias específicas que evalúa el ICFES, para posteriormente aplicar técnicas de Big data.

CREAR EL REPOSITORIO LIMPIO DE DATOS

En esta fase se construyó el primer el repositorio de datos va_repositorio_inicial con datos desde el año 2012 a 2018, con todas las variables socioeconómicas analizadas y filtrando a los estudiantes de la Universidad de Nariño, con la herramienta apache Spark.

```
tablas=["archivos.ftp_sbpro_gen_20121","archivos.ftp_sbpro_gen_20122","archivos.ftp_sbpro_gen_20131",
"archivos.ftp_sbpro_gen_20132","archivos.ftp_sbpro_gen_20141","archivos.ftp_sbpro_gen_20142",
"archivos.ftp_sbpro_gen_20151","archivos.ftp_sbpro_gen_20152","archivos.ftp_sbpro_gen_2016",
"archivos.ftp_sbpro_gen_2017","archivos.ftp_sbpro_gen_2018"]
datos=[]
for t in tablas:
df = spark.read \
.format("jdbc") \
.option("url", "jdbc:postgresql://localhost:5432/va") \
.option("dbtable", t) \
.option("user", "postgres") \
.option("password", "david") \
.option("driver", "org.postgresql.Driver") \
.load()
df=df.where("inst_nombre_institucion = 'UNIVERSIDAD DE NARIÑO-PASTO' or inst_nombre_institucion='UNIVERSIDAD DE NARIÑO'")
datos.append(df)
```

Fig. 5. Selección de tablas y filtro de estudiantes de la Universidad de Nariño.

Fuente: esta investigación.

```
In [3]: variables=["periodo","estu_expio_reside","estu_genero","estu_estadocivil","estu_etnia","estu_limita_notrir","estu_limita_invidente",
"estu_limita_sordocominterprete","estu_limita_sordosininterprete","estu_limita_sordoceguera",
"estu_porcentajecreditosaprob","estu_valormatriculauiversidad","estu_pagomatriculabeca",
"estu_pagomatriculacredito","estu_pagomatriculapadres","estu_pagomatriculapropio","estu_tomo_cursopreparacion",
"estu_cursodocentesles","estu_cursosleapoyosoterno","estu_cursosleexterna","estu_simulacrotipolcfes",
"estu_actividadrefuerzooareas","estu_actividadrefuerzogenetic","estu_semestrecursa","fami_hogaractual",
"fami_cabezafamilia","fami_numpersonasacargo","fami_educacionpadre","fami_educacionmadre","fami_ocupacionpadre",
"fami_ocupacionmadre","fami_estratovivienda","fami_nivel_sisben","fami_personahogar","fami_cuartoshogar",
"fami_pisos_hogar","fami_tieneinternet","fami_tiene_serviciotv","fami_tienecomputador","fami_tienelavadora",
"fami_tiene_microondas","fami_tiene_horno","fami_tieneautomovil","fami_tiene_dvd","fami_tiene nevera",
"fami_tiene celular","fami_telefono","fami_tienebormicroondas","fami_tienemotocicleta","fami_numlibros",
"estu_dedicacionlecturadiaria","estu_dedicacioninternet","fami_tieneconsolavideojuegos","fami_cuantoscompartebano",
"fami_ingreso_familiar_sensual","estu_trabaja_actualmente","estu_horassasanatrabaja","estu_tiporegeneracion",
"estu_pripn_academico","grupopreferencia","inst_origen","estu_privado_libertad","estu_nse_individual","estu_nse_ies"]
```

Fig. 6. Selección de variables socioeconómicas

Fuente: esta investigación.

Seguidamente se analizaron las columnas creadas en el repositorio inicial para mirar el tipo de datos y los niveles que contenían para posteriormente, hacer una limpieza a cada una de ellas, encontrando que los datos tenían espacios en blanco, estaban las letras en mayúsculas, minúsculas, contenían caracteres especiales para lo cual se les dio un tratamiento con la herramienta apache Spark y así poder realizar una limpieza a los datos del repositorio como lo muestra la siguiente figura.

```

In [8]: from pyspark.sql.functions import regexp_replace
        from pyspark.sql.functions import lower, col
        import re
        for c in df.columns:
            df=(df
                .withColumn(c, lower(col(c)))
                .withColumn(c, regexp_replace(c, ' ', '_'))
                .withColumn(c, regexp_replace(c, ',', ''))
                .withColumn(c, regexp_replace(c, 'á', 'a'))
                .withColumn(c, regexp_replace(c, 'é', 'e'))
                .withColumn(c, regexp_replace(c, 'í', 'i'))
                .withColumn(c, regexp_replace(c, 'ó', 'o'))
                .withColumn(c, regexp_replace(c, 'ú', 'u'))
                .withColumn(c, regexp_replace(c, 'ñ', 'n')))

In [32]: df=(df
        .withColumn("estu_acpio_reside", regexp_replace("estu_acpio_reside", 'ü', 'u'))
        .withColumn("estu_estadocivil", regexp_replace("estu_estadocivil", '/', '_'))
        .withColumn("estu_trabaja_actualmente", regexp_replace("estu_trabaja_actualmente", '/', '_'))
        .withColumn("estu_porcentajecreditosaprob", regexp_replace("estu_porcentajecreditosaprob", '%', 'por_ccentaje')))

```

Fig. 7. Limpieza del repositorio inicial
Fuente: esta investigación.

Posteriormente, se creó una nueva tabla de datos **va_repositorio_v2**, con todos los datos limpios, en seguida se realizó otro análisis donde se encontró variables que se podían transformar en variables cuantitativas algunas de ellas: estrato de la vivienda, cuántas personas conforman el hogar, numero de libros que ha leído etc. para las cuales se utilizaron sentencias SQL para poder actualizar los valores antes mencionados. Como lo muestra la siguiente figura.

```

--fami_estratovivienda
update proyecto.va_repositorio_v3 set fami_estratovivienda='no aplica' where trim(fami_estratovivienda)='';
update proyecto.va_repositorio_v3 set fami_estratovivienda='0' where trim(fami_estratovivienda)='estrato 0';
update proyecto.va_repositorio_v3 set fami_estratovivienda='1' where trim(fami_estratovivienda)='estrato 1';
update proyecto.va_repositorio_v3 set fami_estratovivienda='2' where trim(fami_estratovivienda)='estrato 2';
update proyecto.va_repositorio_v3 set fami_estratovivienda='3' where trim(fami_estratovivienda)='estrato 3';
update proyecto.va_repositorio_v3 set fami_estratovivienda='4' where trim(fami_estratovivienda)='estrato 4';
update proyecto.va_repositorio_v3 set fami_estratovivienda='5' where trim(fami_estratovivienda)='estrato 5';
update proyecto.va_repositorio_v3 set fami_estratovivienda='0' where trim(fami_estratovivienda)='estrato 6';
update proyecto.va_repositorio_v3 set fami_estratovivienda='0' where trim(fami_estratovivienda)='vive en una zona rural donde no hay estratificación socioeconómica';
update proyecto.va_repositorio_v3 set fami_estratovivienda='99' where fami_estratovivienda in('sin estrato', 'no aplica');

```

fami_estratovivienda	count
0	270
1	3
2	1,591
99	124
5	3,540
6	31
3	43
4	2,997

Fig. 8. Variable cuantitativa estrato vivienda
Fuente: esta investigación.

Por último, se creó un repositorio final, donde se unieron las tablas **va_repositorio_v3** y **valor agregado** agregando que uno (1) son los estudiantes que generan valor agregado y cero (0) a los estudiantes que no generan valor agregado. Quedando como resultado un repositorio limpio y listo para el desarrollo del trabajo.

```

create or replace view proyecto.view_clase_va
as select r.*, case when va razonamiento_cuantitativo > 0 then 1 else 0 end clase_va_razonamiento_cuantitativo,
case when va lectura_critica > 0 then 1 else 0 end clase_va_lectura_critica,
case when va competencias_ciudadanas > 0 then 1 else 0 end clase_va_competencias_ciudadanas,
case when va ingles > 0 then 1 else 0 end clase_va_ingles,
cast(anio_pro as int)-cast(anio_s11 as int) anio_preparacion
from proyecto.va_repositorio_v3 r join proyecto.valor_agregado va on upper(r.estu_consecutivo) = va.estu_consecutivo

```

Fig. 9. Repositorio Final
Fuente: esta investigación.

ANÁLISIS EXPLORATORIO DEL REPOSITORIO DE DATOS DE VALOR AGREGADO

En esta fase se analizaron las diferentes correlaciones que tienen las variables socioeconómicas de los estudiantes de la Universidad de Nariño con las variables de valor agregado (competencias ciudadanas, inglés, lectura crítica y razonamiento cuantitativo). Para lo cual se utilizó la herramienta Apache Spark con el método `corr()`, donde se trataron los datos para obtener una matriz de correlaciones.

```
#df_va_dummies.corr(), to_excel("correlaciones.xlsx")
df_va_dummies.corr()
```

	período	estu_porcentajecreditosaprob	estu_cursodocentesies	estu_cursoiesapoyoexterno	estu_cursoiesexterna
período	1.000000	-0.860947	-0.757901	-0.760264	-0.759013
estu_porcentajecreditosaprob	-0.860947	1.000000	0.912133	0.913896	0.912220
estu_cursodocentesies	-0.757901	0.912133	1.000000	0.997384	0.995426
estu_cursoiesapoyoexterno	-0.760264	0.913896	0.997384	1.000000	0.997665
estu_cursoiesexterna	-0.759013	0.912220	0.995426	0.997665	1.000000
estu_nse_individual_4	0.147779	-0.106394	-0.101612	-0.101818	-0.101670
estu_nse_individual_nse1	0.322566	-0.467068	-0.441141	-0.441910	-0.441381
estu_nse_individual_nse2	0.201496	-0.329872	-0.305033	-0.305958	-0.305194
estu_nse_individual_nse3	0.088754	-0.188631	-0.172331	-0.172678	-0.172165
estu_nse_individual_nse4	0.098606	-0.133894	-0.127875	-0.128134	-0.127948

229 rows x 229 columns

Fig. 10. Matriz de correlaciones de los estudiantes de la Universidad de Nariño
Fuente: esta investigación.

Después se procedió a exportar esta tabla, y analizar, las posibles correlaciones entre las variables antes mencionadas, con un porcentaje mayor al 15% de incidencia en los resultados.

SELECCIONAR VARIABLES SOCIOECONÓMICAS A EVALUAR

Una vez terminada la fase de análisis exploratorio, se comprobó con gráficos exploratorios (biplots), las interrelaciones entre las clases y variables que se encontraron en la anterior fase, para así poder seleccionar variables para posteriormente realizar los respectivos modelados y encontrar patrones interesantes. Se utilizó la librería **StandardScaler**, que estandariza los datos eliminando la media y escalando los datos de forma que su varianza sea igual a 1.

```
#estandarizar datos
from sklearn.preprocessing import StandardScaler
#
scaler = StandardScaler()
scaler.fit(df_va_dummies)
scaled_matrix = scaler.transform(df_va_dummies)

df_scaled_matrix=pd.DataFrame(scaled_matrix,columns=df_va_dummies.columns)
df_scaled_matrix.head()
```

Fig. 11. Estandarización de datos
Fuente: esta investigación.

```
df_scaled_matrix.corr()
```

	periodo	estu_porcentajecreditosaprob	estu_cursodocentesies	estu_cursoiesapoyoexterno
periodo	1.000000	-0.860947	-0.757901	-0.760264
estu_porcentajecreditosaprob	-0.860947	1.000000	0.912133	0.913696
estu_cursodocentesies	-0.757901	0.912133	1.000000	0.997384
estu_cursoiesapoyoexterno	-0.760264	0.913696	0.997384	1.000000
estu_cursoiesexterna	-0.759013	0.912220	0.995426	0.997665
...
estu_nse_individual_4	0.147779	-0.106394	-0.101612	-0.101818
estu_nse_individual_nse1	0.322566	-0.467068	-0.441141	-0.441910
estu_nse_individual_nse2	0.201496	-0.329872	-0.305033	-0.305958
estu_nse_individual_nse3	0.088754	-0.188631	-0.172331	-0.172678
estu_nse_individual_nse4	0.098606	-0.133894	-0.127875	-0.128134

229 rows x 229 columns

Fig. 12. Matriz Estandarizada
Fuente: esta investigación.

Después se utilizó el análisis de componentes principales o **PCA**, es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información, y la librería visual para generar los componentes principales respecto a las 4 clases y sus respectivas variables.

```
from sklearn.decomposition import PCA
import visuals as vs

#pca = PCA(n_components=len(df_scaled_matrix.keys())).fit(df_scaled_matrix)

#Numero de componentes
pca = PCA(n_components=3).fit(df_scaled_matrix)

pca_samples = pca.transform(df_scaled_matrix)

# Generate PCA results plot
pca_results = vs.pca_results(df_scaled_matrix, pca)
```

Fig. 13. Código Componentes principales dataframe
Fuente: esta investigación.

Por consiguiente, se utilizó la librería **NumPy** es una librería especializada en el cálculo numérico y el análisis de datos, especialmente para un gran volumen de datos, y la librería **matplotlib** especializada en la creación de gráficos en dos dimensiones. Para generar los biplots. Los cuales se muestran en el capítulo resultados.

```

import numpy as np
n = len(df_scaled_matrix.columns)
pca_model = PCA(n_components=n)
pca = pca_model.fit_transform(df_scaled_matrix)
pca.shape

from matplotlib import pyplot as plt

def myplot(score, cp1, cp2, coeff, labels=None):
    fig, ax = plt.subplots(figsize = (14,8))
    xs = score[:,cp1]
    ys = score[:,cp2]
    n = coeff.shape[0]
    scalex = 1.0/(xs.max() - xs.min())
    scaley = 1.0/(ys.max() - ys.min())
    plt.scatter(xs * scalex, ys * scaley, s=5)
    for i in range(n):
        plt.arrow(0, 0, coeff[i,0], coeff[i,1], alpha = 0.5,
                 head_width=0.02, head_length=0.02, linewidth=1, color='red')
        if labels is None:
            plt.text(coeff[i,0]* 1.15, coeff[i,1] * 1.15, "Var"+str(i+1), color = 'green', ha = 'center', va = 'center')
        else:
            plt.text(coeff[i,0]* 1.15, coeff[i,1] * 1.15, labels[i], color = 'green', ha = 'center', va = 'center')

    plt.xlabel("PC{}".format(cp1+1))
    plt.ylabel("PC{}".format(cp2+1))

    plt.grid()

cp1=0
cp2=1

myplot(pca, cp1, cp2, np.transpose(pca_model.components_)[[cp1, cp2]. : ], list(df_scaled_matrix.columns))
plt.show()

```

Fig. 14. Código generación de biplots
Fuente: esta investigación.

D. MODELADO

Es esta fase se pretenden construir modelos con técnicas de Big Data a partir del repositorio limpio de datos, que muestren que variables socioeconómicas influyen en la generación de patrones de valor agregado de los estudiantes de la Universidad de Nariño.

TÉCNICAS DE BIG DATA

ALGORITMO A PRIORI

Una vez finalizada la etapa de preparación de los datos que dio como resultado una matriz precisa de correlaciones entre las variables socioeconómicas de los estudiantes y las competencias ciudadanas a evaluar, se encontraron reglas de asociación, mediante el algoritmo a priori, para lo cual se creó una nueva vista llamada **proyecto.view_class_va**, donde se almacenaron las variables previamente analizadas, y además se dio la transformación de algunas variables (cuantitativas, cualitativas, discretas, continuas, etc.), para un mejor resultado e interpretación en las reglas.

También se descartaron las variables que tienen un valor igual a **Null**, o bien denominadas **no_aplica** para evitar asociaciones erróneas en el método a priori.

```

#CONVERTIR VARIABLES A DUMMIES
import pandas as pd
df_va_dummies= pd.get_dummies(df_pan.drop_first=True)
df_va_dummies.drop(columns=['fami_tienehornodirroogas_no_aplica',
                             'fami_tienemotocicleta_no_aplica',
                             'estu_nse_individual_no_aplica',
                             'estu_horassemestrabaaja_no_aplica',
                             'fami_tieneconsolavideojuegos_no_aplica',
                             'estu_pagowatriculabeca_no_aplica',
                             'estu_pagowatriculacredito_no_aplica',
                             'estu_pagowatriculapropio_no_aplica',
                             'estu_tiposmuneracion_no_aplica',
                             'fami_tieneinternet_no_aplica',
                             'estu_pagowatriculapadres_no_aplica',
                             'estu_etnia_no_aplica',
                             'fami_hogarectual_no_aplica',
                             'fami_caberafamilia_no_aplica',
                             'estu_valorwtriculainiversidad_no_aplica',
                             'fami_educacionmadre_no_aplica',
                             'fami_tienecomputador_no_aplica',
                             'fami_educacionpadre_no_aplica',
                             'fami_tienelavadora_no_aplica',
                             'fami_tieneautomovil_no_aplica',
                             'fami_tiene_servisiotv_no_aplica',
                             'fami_ocupacionpadre_no_aplica',
                             'fami_ocupacionmadre_no_aplica',
                             'estu_estadocivil_no_aplica',
                             'estu_cuidocrotipoicfes_no_aplica',
                             'estu_actividadrefuerzorees_no_aplica',
                             'estu_actividadrefuerzogenetic_no_aplica',
                             'estu_tomo_cursoreparacion_no_aplica',
                             'estu_trabajo_actualmente_no_aplica',
                             'fami_pisos_hogar_no_aplica',
                             'fami_tiene_microondas_no_aplica',
                             'fami_tiene_horno_no_aplica',
                             'fami_tiene_dvd_no_aplica',
                             'fami_telefono_no_aplica',
                             'fami_ingreso_familiar_sensual_no_aplica',
                             'fami_tiene_nevera_no_aplica',
                             'fami_tiene_celular_no_aplica'],inplace=True)

```

Fig. 15. Eliminación de variables nulas

Fuente: esta investigación.

Posteriormente se procedió a importar la librería **mlxtend** que permite identificar los patrones frecuentes del itemset limpio que se ha construido, para ello se emplea el algoritmo a priori y los parámetros como el soporte mínimo de 0.1 para mayor efectividad en la obtención de reglas.

```

In [11]: import mlxtend
         from mlxtend.frequent_patterns import apriori
         itemset_freq=apriori(df_va_dummies, min_support=0.1, use_colnames=True)

```

Fig. 16. Import mlxtend con algoritmo apriori.

Fuente: esta investigación.

```

In [13]: from mlxtend.frequent_patterns import association_rules
         reglas=association_rules(itemset_freq, metric="confidence", min_threshold=0.7)

```

Fig. 17. Algoritmo de reglas asociativas

Fuente: esta investigación.

Más adelante se importó la librería **association_rules** que utiliza como medida la métrica confidence la cual es la confianza o fiabilidad de cada regla a priori del modelo.

El cual se importó como una tabla de datos llamada **reglas** para su respectivo análisis que se muestra en el capítulo resultados.

ARBOLES DE DECISIÓN

Para aplicar este algoritmo se utilizó la vista **proyecto.view_clase_va**, antes mencionada, con la cual se aplicó un modelo donde se selecciona la clase (Y) y las variables del dataframe (X), para poder aplicar el algoritmo de árboles de decisión.

```
In [185]: Y=df_pan['clase_va_lectura_critica']
X=df_pan.drop(columns=['estu_consecutivo','estu_nse_ies','estu_limita_motriz','periodo',
'estu_limita_invidente','estu_limita_sordoconinterprete','estu_limita_sordosininterprete',
'estu_limita_sordoceguera','inst_origen','estu_privado_libertad','clase_va_lectura_critica',
'clase_va_razonamiento_cuantitativo','clase_va_competencias_ciudadanas','clase_va_ingles'])
```

Fig. 18. Selección de variables con la vista clase_va
Fuente: esta investigación.

Además, se eliminaron las variables de menor importancia debido a que estas afectaban el resultado en los árboles de igual manera se eliminan las clases debido a que el algoritmo debe comparar solo con las variables socioeconómicas. Y también se descartan las variables Null (**no aplica**) como en el anterior algoritmo.

Se importó la librería **train_test_split** que permite dividir un dataset en dos bloques, típicamente bloques destinados al entrenamiento y validación del modelo train (bloque de entrenamiento) y test (bloque de pruebas), se asignaron los valores de train 0.7 y el test 0.3 para aplicar el algoritmo.

```
In [109]: from sklearn.model_selection import train_test_split

# Split the 'features' and 'income' data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_dummies,
                                                    Y,
                                                    test_size = 0.3,
                                                    random_state = 0)

# Show the results of the split
print("Total ejemplos train {}".format(X_train.shape[0]))
print("Total ejemplos test {}".format(X_test.shape[0]))

print("Valor Agregado train:",y_train[y_train==1].count())
print("No Valor Agregado train:",y_train[y_train==0].count())

print("Valor Agregado test:",y_test[y_test==1].count())
print("No Valor Agregado test",y_test[y_test==0].count())
```

```
Total ejemplos train 3595.
Total ejemplos test 1542.
Valor Agregado train: 2263
No Valor Agregado train: 1332
Valor Agregado test: 977
No Valor Agregado test 565
```

Fig. 19. Algoritmo train y test
Fuente: esta investigación.

Esta librería permite analizar los datos que van a hacer procesados a la hora de aplicar el algoritmo de árboles de decisión para la cual se necesita una librería **tree** la cual permite entrenar árboles de decisión teniendo en cuenta diferentes parámetros.

```

In [110]: from sklearn import tree

# Crear Arbol de decision con profundidad = 4
modelo1 = tree.DecisionTreeClassifier(criterion='entropy',
                                     min_samples_leaf=10,
                                     max_depth = 4
                                     )

modelo1.fit(X_train, y_train)

Out[110]: DecisionTreeClassifier(criterion='entropy', max_depth=4, min_samples_leaf=10)

In [111]: print(modelo1.get_depth())
print(modelo1.get_n_leaves())

4
15

```

Fig. 20. Algoritmo arboles de decisión
Fuente: esta investigación.

Con la aplicación de este modelo se creó un árbol de decisión utilizando el criterio de entropy que mide la impureza del conjunto de datos, max_depth (profundidad máxima que puede alcanzar el árbol) en este caso tiene una profundidad de 4 y min_samples_leaf (número mínimo de observaciones que debe de tener cada uno de los nodos hijos para que se produzca la división) en este modelo es de 10, generando un árbol de profundidad de 4 y el nivel de hojas que en este caso es 15.

Más adelante, se modelo un algoritmo que permite analizar la clase (Y) y las variables del dataframe (X), con las diferentes métricas que utiliza el algoritmo de árboles de decisión y que elija los mejores valores para generar arboles de decisión precisos.

```

In [117]:
%%time
gb = Class_Fit(clf = tree.DecisionTreeClassifier)
param_grid = {
    'criterion':['entropy'],
    'ccp_alpha':[0,0.0025,0.0050,0.0075,0.008,0.0085,0.01,0.05],
    'max_depth':[3,4,5,6,7,8,9,10,15,20],
    'min_samples_split':[20,30,40,50,60,70,80,90,100],
    'max_leaf_nodes':[3,4,5,6,7,8,9,10,20]
}

gb.grid_search(parameters = param_grid, kfold = 10)
gb.grid_fit(X = X_train, Y = y_train)

print("Parámetros árbol:",gb.grid.best_params_)
mejor_arbol=gb.grid.best_estimator_

Parámetros árbol: {'ccp_alpha': 0, 'criterion': 'entropy', 'max_depth': 4, 'max_leaf_nodes': 5, 'min_samples_split': 10}
Wall time: 40min 25s

```

Fig. 21. Algoritmo árbol de decisión con los mejores parámetros
Fuente: esta investigación.

E. EVALUACIÓN

Para esta fase se seleccionarán los mejores parámetros para la generación de árboles de decisión, las métricas de evaluación de cada modelo donde mostrarán la precisión de cada resultado obtenido.

MÉTRICAS DE EVALUACIÓN

Se modeló un algoritmo con la librería **confusion_matrix**, que permite visualizar el desempeño de un algoritmo, se utilizaron las diferentes métricas de la librería. La cuales se explican a continuación.

- Precisión del entrenamiento (Accuracy train)
- Precisión de las pruebas (Accuracy test)
- Precisión de entrenamiento equilibrada (Balanced Accuracy train)
- Precisión de pruebas equilibrada (Balanced Accuracy test)

```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score, precision_score, balanced_accuracy_score, f1_score

def _score_func(estimator, X, y):
    y_pred_test = estimator.predict(X)
    return balanced_accuracy_score(y, y_pred_test)

class Class_Fit(object):
    def __init__(self, clf, params=None):
        if params:
            self.clf = clf(**params)
        else:
            self.clf = clf()

    def train(self, x_train, y_train):
        self.clf.fit(x_train, y_train)

    def predict(self, x):
        return self.clf.predict(x)

    def grid_search(self, parameters, Kfold):
        self.grid = GridSearchCV(estimator = self.clf, param_grid = parameters, cv = Kfold, scoring=_score_func)

    def grid_fit(self, X, Y):
        self.grid.fit(X, Y)

    def grid_predict(self, X, Y):
        self.predictions = self.grid.predict(X)
        print("Precision: {:.2f} % ".format(100*metrics.balanced_accuracy_score(Y, self.predictions)))

from sklearn.model_selection import GridSearchCV
```

Fig. 22. Algoritmo Confusión matrix

F. DESPLIEGUE

Se detectaron patrones interesantes de valor agregado los cuales se describen mejor en el capítulo de resultados.

IV. RESULTADOS DE LA INVESTIGACIÓN

A. *COMPRENSIÓN DEL NEGOCIO*

En la fase de comprensión del problema se obtuvo un marco teórico referente a todas las técnicas utilizadas en esta investigación para tratar los resultados de los estudiantes de las pruebas Saber Pro. En la fase de preparación de los datos se obtuvo un repositorio limpio de datos obteniendo como resultado que desde el año 2012 a 2018, se presentaron a las pruebas Saber Pro 8.599 estudiantes de la Universidad de Nariño.

B. *COMPRENSIÓN DE LOS DATOS*

En esta fase, se idéntico, recopiló y familiarizó con la base de datos de las pruebas SABER 11 y SABER PRO. Se inició analizando una tabla **va_cruces** donde se almacenan los consecutivos de todos los estudiantes de Colombia que presentaron las pruebas SABER 11 y SABER PRO, obteniendo como resultado que 1.201.464 de estudiantes fueron a las universidades de Colombia.

Después se construyó dos tablas de datos **va_maximos_11**, **va_maximos_pro**, donde se calcularon los valores máximos de puntajes obtenidos, en las diferentes pruebas que evalúa el ICFES de los periodos de 2006 a 2018, filtrando por el departamento de Nariño y los estudiantes de la Universidad de Nariño.

abc año	123 max_punt_matematicas	123 max_punt_lectura_critica_s11	123 max_punt_sociales	123 max_punt_ingles_s11
2006	86,9	70,875	66,39	85,62
2007	90,85	68,995	82,77	76,45
2008	95,12	71,845	78,11	83,41
2009	95,4	69,79	75,89	87,56
2010	113,46	82,625	80,92	116,95
2011	86	80	107	89
2012	97	77,5	79	92
2013	88	74	78	94
2014	83	79	75	83
2017	66	58	59	74

Fig. 23. Valores máximos pruebas Saber 11

Fuente: esta investigación.

abc año	123 max_punt razonamiento_cuantitativo	123 max_punt_lectura_critica_pro	123 max_punt_competencias_ciudadanas	123 max_punt_ingles_pro
2012	14,5	13,5	13,9	15
2013	16,2	14,3	13,7	15
2014	15,4	14,6	13,9	14,9
2015	13,5	13,8	14,6	14,95
2016	300	300	300	300
2017	300	300	227	231
2018	300	300	230	227

Fig. 24. Valores máximos pruebas Saber Pro

Fuente: esta investigación.

Posteriormente se creó una tabla de datos inicial, **valor_agregado** donde se calculó el valor agregado en cada una de las pruebas presentadas de los estudiantes ante el ICFES, con el respectivo año de presentación en cada una de las pruebas, identificados por el consecutivo de las pruebas Saber Pro utilizando el siguiente **SELECT**.

```

select pro_estu_consecutivo, substring(pro_periodes,0,5) anio_pro, substring(s11_periodes,0,5) anio_s11,
(pro_punt_razonamiento_cuantitativo/mpro_max_punt_razonamiento_cuantitativo)-(s11_punt_matematicas/m11_max_punt_matematicas) va_razonamiento_cuantitativo,
(pro_punt_lectura_critica/mpro_max_punt_lectura_critica)-(s11_punt_lectura_critica/m11_max_punt_lectura_critica_s11) va_lectura_critica,
(pro_punt_competencias_ciudadanas/mpro_max_punt_competencias_ciudadanas)-(s11_punt_sociales/m11_max_punt_sociales) va_competencias_ciudadanas,
(pro_punt_ingles/mpro_max_punt_ingles)-(s11_punt_ingles/m11_max_punt_ingles_s11) va_ingles:
FROM proyecto_va_saber_pro_pro
JOIN proyecto_va_crszas t ON pro_estu_consecutivo = c_estu_consecutivo_pro
JOIN proyecto_va_saber_ii s11 ON s11_estu_consecutivo = c_estu_consecutivo_ii
JOIN va_maximos_pro mpro ON substring(pro_periodes,0,5)=mpro_anio
JOIN va_maximos_s11 m11 ON substring(s11_periodes,0,5)=m11_anio
WHERE pro_departamento_programa = 'NARIÑO':text
AND pro_nombre_institucion = 'UNIVERSIDAD DE NARIÑO-PASTO':text

```

Fig. 25. Calculo valor agregado de estudiantes de la Universidad de Nariño

Fuente: esta investigación.

	estu_consecutivo	anio_pro	anio_s11	va_razonamiento_cuantitativo	va_lectura_critica	va_competencias_ciudadanas	va_ingles
1	EK20130020411	2015	2006	0,1115520249	0,0996593218	0,0908578821	0,2258062576
2	EK20131000934	2013	2006	-0,0251544886	-0,0721864189	0,0286253646	0,1203270264
3	EK201330164707	2013	2006	0,0102217676	0,0406151873	0,0340148844	0,1093983294
4	EK201430220569	2014	2006	-0,0126417386	-0,0093377787	0,0230822357	0,3261994234
5	EK201730144342	2017	2006	-0,1904909858	-0,3317636684	-0,0722210831	0,001381823
6	EK201210078710	2012	2006	0,071505099	-0,0043738977	0,0243254109	0,1867647746
7	EK201630184347	2016	2006	-0,0296979407	-0,0065879645	-0,255927097	-0,070660099
8	EK201330161388	2013	2006	0,1259728063	-0,0022964423	-0,0328683086	0,4250820246
9	EK201530043541	2015	2006	0,0065889602	-0,024897879	0,0399830256	0,1188663606
10	EK201430225279	2014	2006	0,1459148446	0,0851102897	-0,0574132517	0,2300796522
11	EK201230112323	2012	2006	-0,0388198881	-0,1758024691	-0,0794329442	0,0979976675
12	EK201230112542	2012	2006	0,0911273362	0,2316049383	0,1094231700	0,0628731605
13	EK201630184311	2016	2006	0,0383369808	-0,0018719088	-0,2016169188	-0,046194814

Fig. 26. Tabla valor agregado de estudiantes de la Universidad de Nariño

Fuente: esta investigación

C. PREPARACIÓN DE LOS DATOS

VARIABLES SOCIOECONÓMICAS A EVALUAR

Una vez terminada la fase de análisis exploratorio, se comprobó con gráficos exploratorios (biplots), las interrelaciones entre las clases y variables que se encontraron en la anterior fase, para así poder seleccionar variables para posteriormente realizar los respectivos modelados y encontrar patrones interesantes.

A continuación, se muestran los biplots generados con respecto a las 4 clases y sus respectivas variables.

CLASE VA RAZONAMIENTO CUANTITATIVO

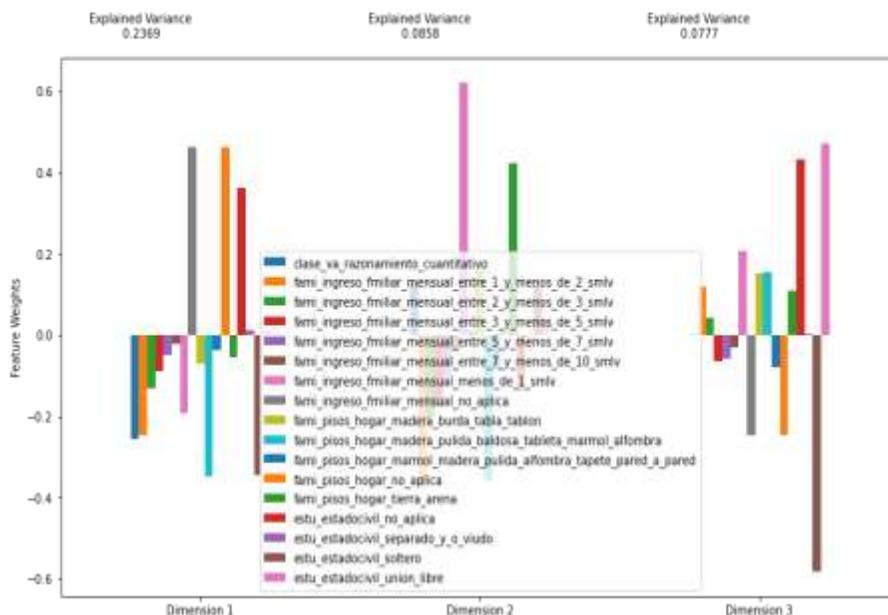


Fig. 27. Componentes razonamiento cuantitativo
Fuente: esta investigación.

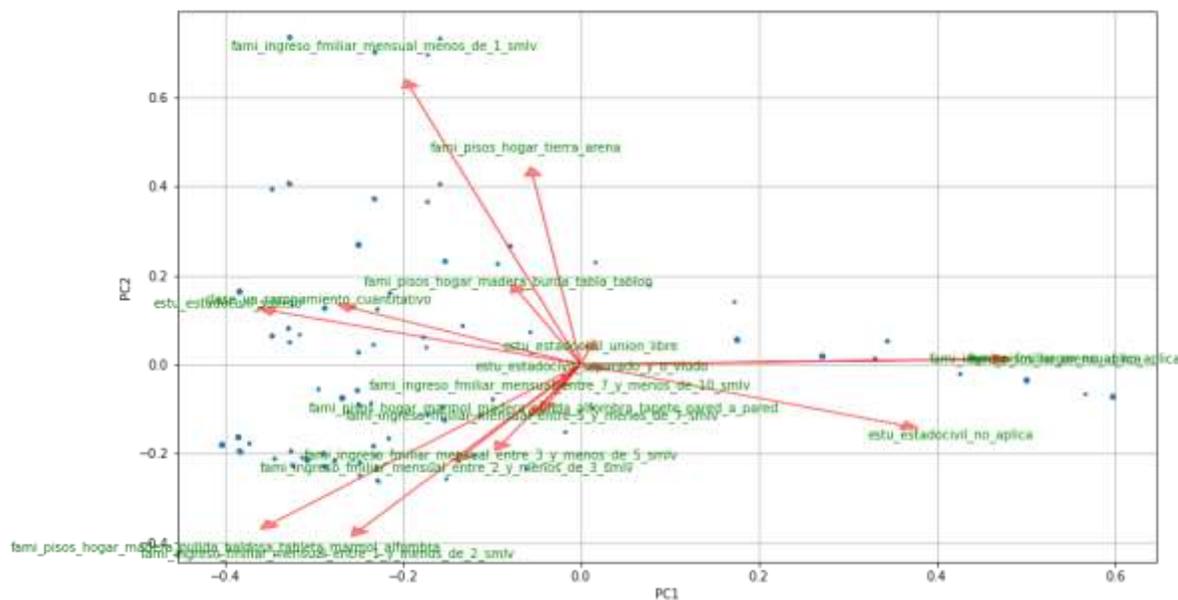


Fig. 28. Biplot razonamiento cuantitativo
Fuente: esta investigación.

Variables correlacionadas con va razonamiento cuantitativo:

- fami_ingreso_familiar_mensual
- fami_pisos_hogar
- estu_estadocivil

CLASE VA LECTURA CRITICA

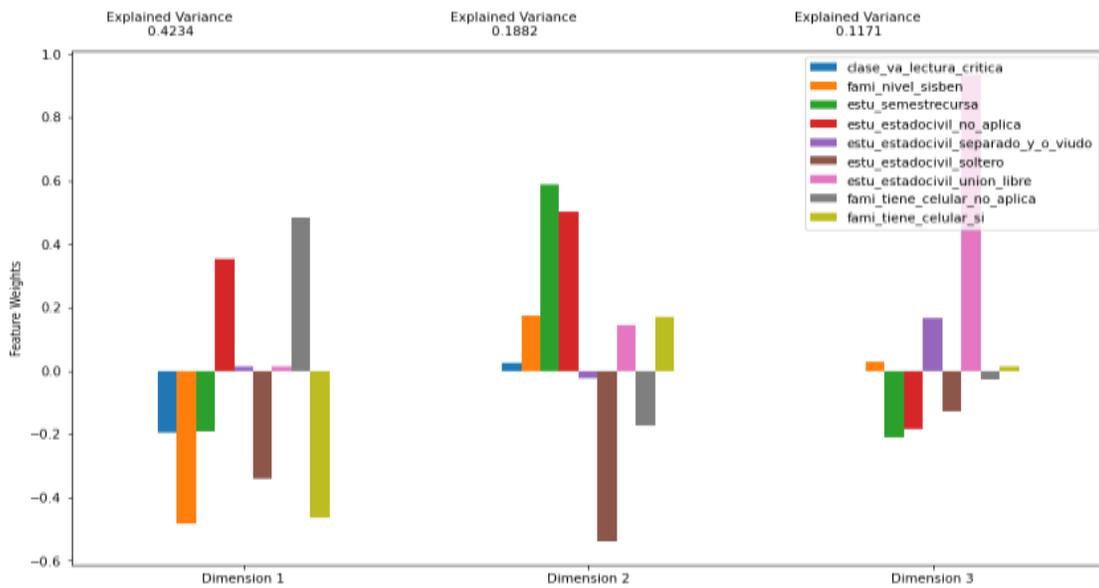


Fig. 29. Componentes lectura crítica
Fuente: esta investigación.

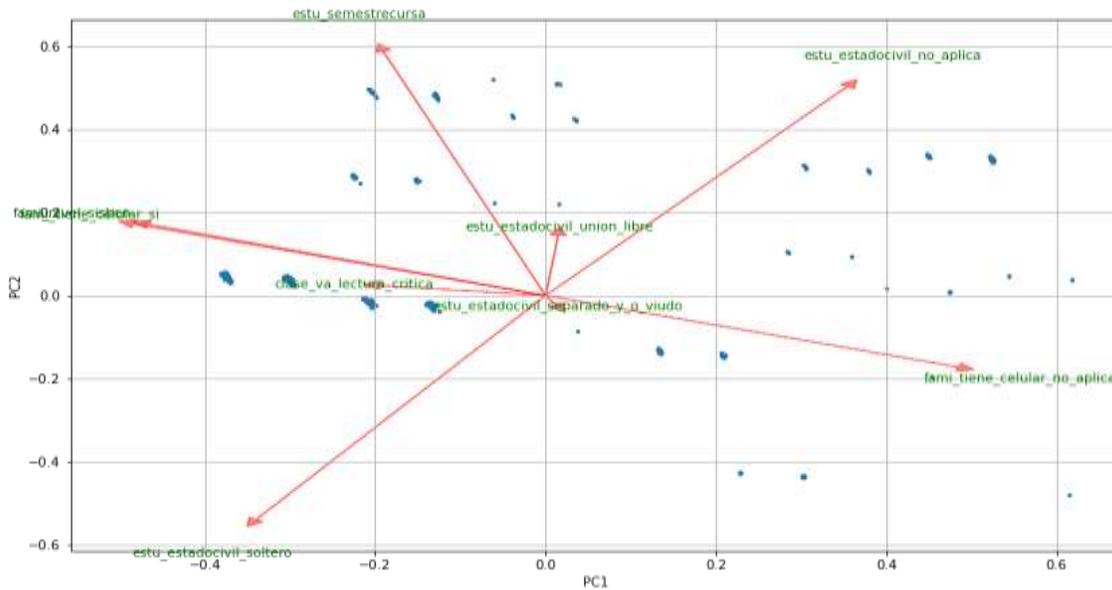


Fig. 30. Biplot lectura crítica
Fuente: esta investigación.

Variables correlacionadas con va_lectura_crítica:

- estu_estadocivil
- fami_nivel_sisben
- fami_tiene_celular
- estu_semestrecursa

CLASE VA COMPETENCIAS CIUDADANAS

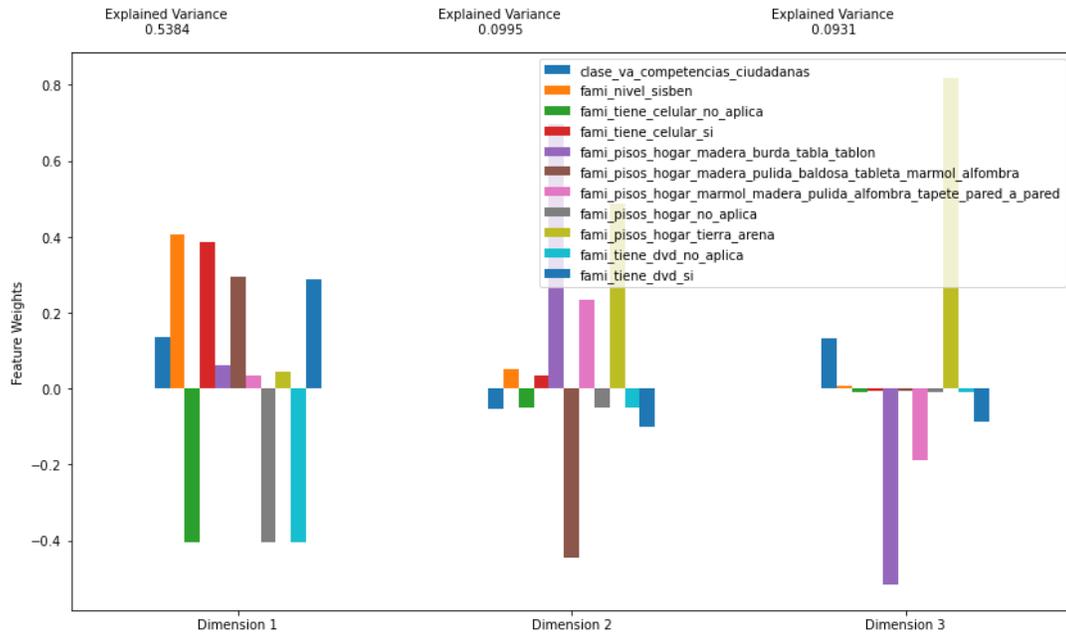


Fig. 31. Componentes Competencias ciudadanas

Fuente: esta investigación.

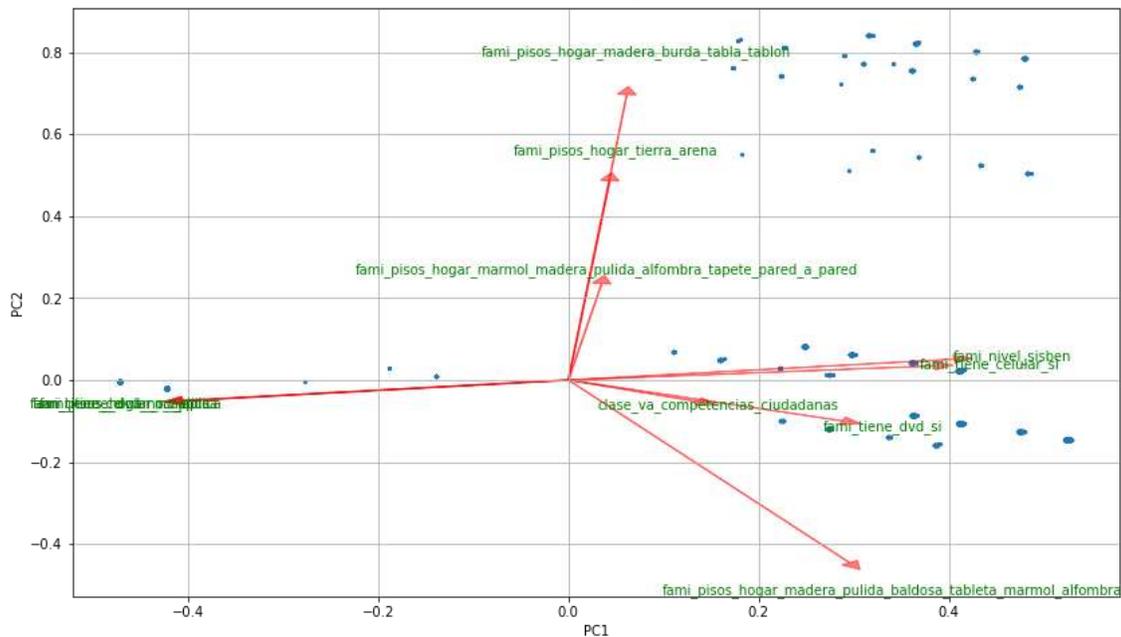


Fig. 32. Biplot Competencias ciudadanas

Fuente: esta investigación.

Variables correlacionadas con va competencias ciudadanas:

- fami_ocupacionmadre
- fami_nivel_sisben
- fami_tiene_celular
- fami_pisos_hogar
- fami_tiene_dvd

D. MODELADO

En esta fase se obtuvieron reglas de asociación y árboles de decisión para obtener patrones interesantes de valor en los estudiantes de Universidad de Nariño.

ALGORITMO A PRIORI

Con la aplicación del algoritmo se obtuvieron **6.021.962** reglas de asociación como resultado.

```
In [14]: reglas.count()
Out[14]: antecedents      6021962
         consequents      6021962
         antecedent support 6021962
         consequent support 6021962
         support          6021962
         confidence       6021962
         lift             6021962
         leverage         6021962
         conviction       6021962
         dtype: int64
```

Fig. 35. Conteo de reglas a priori
Fuente: esta investigación.

A continuación, se muestran reglas interesantes que se encontró tomando en cuenta las métricas de Confianza y lift:

Reglas clase va razonamiento cuantitativo

TABLA I
REGLA 1 - RAZONAMIENTO CUANTITATIVO

Antecedentes	Soporte	Confianza	Lift
estu_cursoiesapoyoexterno_0, estu_estadocivil_soltero, estu_cursoiesexterna_0, estu_cursodocentesies_0	0.103173055	0.8998273	1.4266707

- Estado civil soltero(a).
- No se preparó para el examen saber pro en una institución externa.
- No se preparó para el examen saber pro con docentes de la institución.
- No se preparó para el examen saber pro en la institución.

El 90% de los estudiantes que no tomaron curso de apoyo externo, con estado civil soltero/a, no tomaron curso en una institución externa y no tomaron curso impartido por los docentes de la institución, generan valor agregado a la clase de razonamiento cuantitativo, el 10% de todos los estudiantes, cumplen este patrón, y la métrica lift es mayor a uno, por lo tanto, los antecedentes tienen correlación positiva con la consecuente.

TABLA II
REGLA 2 - RAZONAMIENTO CUANTITATIVO

Antecedentes	Soporte	Confianza	Lift
estu_cursoiesapoyoexterno_0, estu_valormatriculauniversidad_menos_de_500_mil, fami_tiene_celular_si, anio_preparacion_5, estu_mcpio_reside_pasto	0.12263967	0.88359046	1.4009272

- No se preparó para el saber pro en un curso de la institución con apoyo de entidades externas.
- El costo de la matrícula en el último semestre fue menor a \$500 mil pesos.
- Cuenta con un teléfono móvil.
- Presento el examen cursando el quinto año de carrera.
- Reside en Pasto.

El 88% de los estudiantes que no tomaron curso de apoyo externo, pagaron menos de \$500.000 de matrícula, tienen celular, están en quinto año de la carrera y residen en pasto generan valor agregado a la clase de razonamiento cuantitativo, el 12% de todos los estudiantes cumplen este patrón y la métrica lift es mayor a uno, por lo tanto, los antecedentes tienen correlación positiva con el consecuente.

TABLA III
REGLA 3 - RAZONAMIENTO CUANTITATIVO

Antecedentes	Soporte	Confianza	Lift
estu_cursoiesapoyoexterno_0, estu_valormatriculauniversidad_menos_de_500_mil, fami_tiene_celular_si, estu_estadocivil_soltero, anio_preparacion_5, fami_numpersonasacargo_0, estu_cursoiesexterna_0, estu_mcpio_reside_pasto, estu_cursodocentesies_0	0.10589838	0.87741935	1.391143

- No se preparó para el saber pro en un curso de la institución con apoyo de entidades externas.
- El costo de la matrícula en el último semestre fue menor a \$500 mil pesos
- Cuenta con un teléfono móvil.
- Es soltero(a).
- Presento el examen cursando el quinto año de carrera.
- No tiene personas a cargo.
- No se preparó en un instituto externo.
- Reside en pasto.
- No se preparó para el examen saber pro con docentes de la institución.

El 87% de estudiantes que no tomaron curso de apoyo externo, pagaron menos de \$500.000 de matrícula académica, tienen celular, son solteros/as, están en quinto año de la carrera, no tienen personas a cargo, no tomaron curso con una institución externa, residen en pasto y no tomaron

curso de preparación con docentes de la institución educativa superior, generan valor agregado a la clase de razonamiento cuantitativo. El 10% de todos los estudiantes cumplen este patrón y la métrica lift es mayor a 1, por lo tanto, los antecedentes tienen correlación positiva con los consecuentes.

Reglas clase va competencias ciudadanas

TABLA IV
REGLA 1 - COMPETENCIAS CIUDADANAS

Antecedentes	Soporte	Confianza	Lift
fami_tieneinternet_si, fami_tiene_celular_si, estu_pagomatriculapadres_si, fami_tienecomputador_si	0.10239439	0.821875	1.406855

- Cuenta con conexión a internet en el hogar.
- Tiene dispositivo móvil.
- Los padres del estudiante pagan la matrícula de la Universidad.
- Cuenta con laptop o pc de sobremesa.

El 82% de estudiantes que tienen internet, tienen celular, los padres pagan sus matriculas académicas y tienen computador generan valor agregado en la clase de competencias ciudadanas, el 10% de todos los estudiantes cumplen con este patrón, la métrica lift es mayor a 1, por lo tanto, los antecedentes se correlacionan positivamente con los consecuentes.

TABLA V
REGLA 2 - COMPETENCIAS CIUDADANAS

Antecedentes	Soporte	Confianza	Lift
estu_valormatriculauniversidad_menos_de_500_mil, estu_estadocivil_soltero, estu_pagomatriculapadres_si, fami_numpersonasacargo_0, estu_cursoiesexterna_0	0.10492505	0.8093093	1.3853456

- El costo de la matrícula en el último semestre fue menor a \$500 mil pesos.
- Es soltero(a).
- Los padres del estudiante pagan la matrícula de la Universidad.
- No tiene personas a cargo
- No tomo cursos de preparación para el examen saber pro en un instituto externo.

El 81% de los estudiantes que pagaron menos de \$500.000 de matrícula financiera, son solteros/as, los padres pagaron su matrícula, no tienen personas a cargo y no tomaron curso de preparación con una institución externa generan valor agregado en la clase de competencias ciudadanas, el 10% de todos los estudiantes cumplen este patrón, la métrica lift es mayor a 1, por lo tanto, los antecedentes se correlacionan positivamente con los consecuentes.

TABLA VI
REGLA 3 - COMPETENCIAS CIUDADANAS

Antecedente	Soporte	Confianza	Lift
fami_tieneinternet_si, fami_tiene_celular_si, estu_estadocivil_soltero, fami_numpersonasacargo_0, estu_mcpio_reside_pasto, fami_pisos_hogar_madera_pulida_baldosa_tableta_mar mol_alfombra, fami_tienecomputador_si	0.10044772	0.80751175	1.3822685

- El estudiante cuenta con conexión a internet en el hogar.
- Tiene teléfono móvil.
- Es soltero(a).
- No tiene personas a cargo.
- Reside en Pasto.
- El lugar de residencia cuenta con piso de tipo madera, baldosa, mármol o alfombra.
- Tiene laptop o PC de sobremesa

El 80% de los estudiantes que tiene internet, tienen celular, son solteros/as, no tienen personas a cargo, residen en Pasto, tienen un tipo de piso en su hogar y tienen computador generan valor agregado a la clase de competencias ciudadanas, el 10% de todos los estudiantes cumplen con este patrón, la métrica lift es mayor a uno, por lo tanto, los antecedentes se correlacionan positivamente con los consecuentes.

Reglas clase va inglés

TABLA VII
REGLA 1 - INGLÉS

Antecedentes	Soporte	Confianza	Lift
estu_valormatriculauniversidad_menos_de_500_mil, anio_preparacion_5, fami_numpersonasacargo_0, estu_cursoiesexterna_0,estu_mcpio_reside_pasto, fami_tienecomputador_si	0.10083706	0.92007107	1.1350636

- El costo de la matrícula en el último semestre fue menor a \$500 mil pesos.
- Presento el examen en el último año de la carrera.
- No tiene personas a cargo.
- No tomo curso de preparación para el examen saber pro en un instituto externo.
- Reside en Pasto.
- Tiene laptop o PC de sobremesa.

El 92% de los estudiantes que pagaron menos de \$500.000 de matrícula financiera, tomaron 5 años de carrera, no tienen personas a cargo, no tomaron curso de preparación en una institución externa, residen en Pasto y tienen computador generan valor agregado a la clase de inglés, el 10% de todos los estudiantes cumplen este patrón, la métrica lift es mayor a 1, por lo tanto, los antecedentes se correlacionan positivamente con los consecuentes.

TABLA VIII
REGLA 2 - INGLÉS

Antecedente	Soporte	Confianza	Lift
estu_cursoiesapoyoexterno_0, estu_estadocivil_soltero, anio_preparacion_5, fami_numpersonasacargo_0, estu_cursoiesexterna_0, estu_mcpio_reside_pasto, fami_tienecomputador_si, estu_cursodocentesies_0	0.11485302	0.9133127	1.126726

- No tomo un curso de preparación de la institución de educación superior con apoyo de un instituto externo.
- Es soltero(a).
- Presento el examen en el último año de la carrera.
- No tiene personas a cargo.
- No tomo curso de preparación en un instituto externo.
- Reside en pasto.
- Cuenta con laptop o PC de sobremesa.
- No tomo curso de preparación con docentes de la institución.

El 91% de los estudiantes que no tomaron curso de preparación externo, son solteros/as, tomaron 5 años de preparación, no tienen personas a cargo, no tomaron curso de preparación en una institución externa, residen en Pasto, tienen computador y no tomaron curso de preparación con docentes de la institución generan valor agregado a la clase inglés, el 11% de todos los estudiantes cumplen este patrón, la métrica lift es mayor a 1, por lo tanto, los antecedentes se correlacionan positivamente con los consecuentes.

TABLA IX
REGLA 3 - INGLÉS

Antecedente	Soporte	Confianza	Lift
estu_porcentajecreditosaprob_81, estu_cursoiesapoyoexterno_0, fami_tiene_celular_si, estu_mcpio_reside_pasto, fami_tienecomputador_si, estu_cursodocentesies_0	0.1047303	0.9118644	1.1249393

- El estudiante aprobó entre el 81% y 90% de créditos para optar por el título.
- No tomo un curso de preparación de la institución de educación superior con apoyo de un instituto externo.
- Cuenta con teléfono móvil.
- Reside en pasto.
- Cuenta con laptop o PC de sobremesa.
- No tomo curso de preparación con docentes de la institución.

El 91% de los estudiantes que aprobaron entre el 81% y el 90% de los créditos para poder realizar trabajo de grado, no tomaron curso de preparación externo, tienen celular, residen en pasto, tienen computador y no tomaron curso de preparación con docentes de la institución generan valor agregado a la clase inglés, la métrica lift es mayor a 1, por lo tanto, los antecedentes tienen correlación positiva con los consecuentes.

ARBOLES DE DECISIÓN

En esta investigación, para la generación de patrones interesantes se realizó el análisis de información con un total de 3595 estudiantes de la Universidad de Nariño, y en cada nodo raíz de los árboles, las variables socioeconómicas difieren, obteniendo los siguientes resultados:

CLASE COMPETENCIAS CIUDADANAS

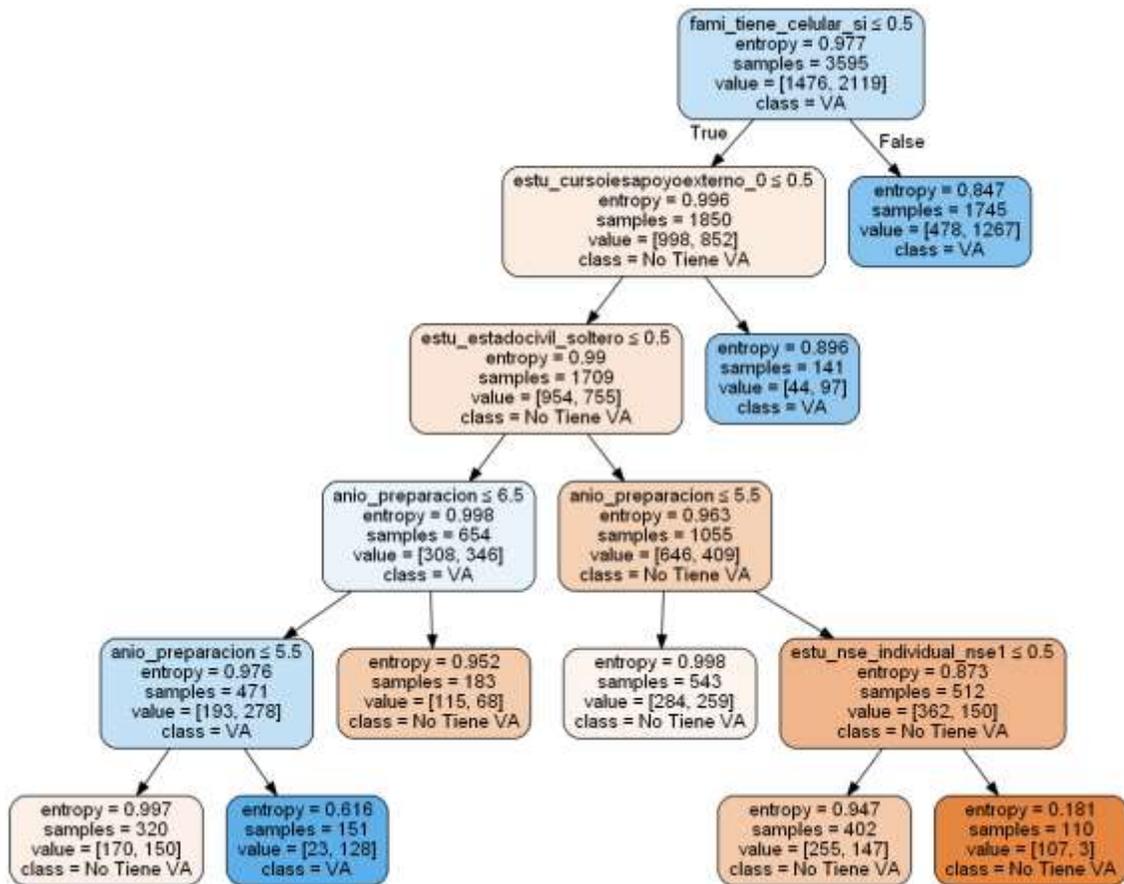


Fig. 36. Árbol 01 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa que 1745 estudiantes cumplieron la condición de que tienen dispositivos móviles, los cuales 1267 aportan valor agregado y 478 no aportan valor agregado, 1850 estudiantes no cumplieron con la condición y se pasa al segundo nivel del árbol donde 141 estudiantes tomaron un curso de apoyo externo impartido por la Universidad para las pruebas saber pro, de los cuales 97 aportaron valor agregado y 44 no aportaron, 1709 estudiantes no cumplieron con la condición y pasamos al tercer nivel del árbol donde se evaluó el estado civil de los estudiantes, los cuales 1055 no están solteros y no generan valor agregado, mientras que 654 estudiantes solteros generan

valor agregado, pasamos al nivel 4 donde tenemos una medición que consiste en los años que le tomo al estudiante acabar la carrera universitaria, 183 estudiantes tomaron más de 6.5 años en terminar la carrera y no generan valor agregado mientras que 471 generan valor agregado ya que cumplen la condición, y 151 estudiantes que tomaron menos de 5.5 años de carrera generan valor agregado.

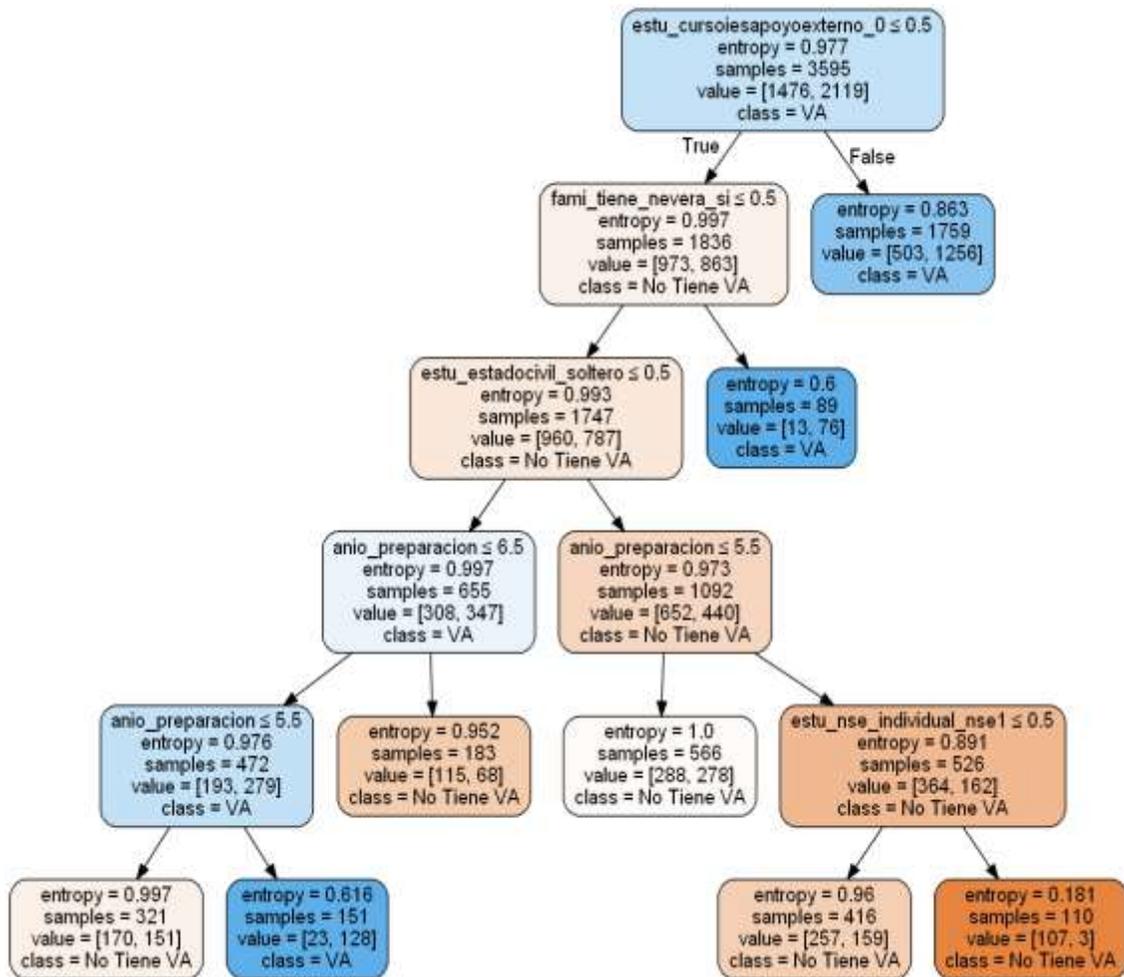


Fig. 37. Árbol 02 - Competencias Ciudadanas

Fuente: esta investigación.

Se evalúa que 1759 tomaron un curso de apoyo externo con docentes de la institución, de lo cual se obtiene que 1256 aportaron valor agregado y 503 no aportaron. 1836 estudiantes no tomaron el curso y no aportan valor agregado, entonces se evalúa la condición de que, si tienen nevera en casa, 89 estudiantes no tienen nevera de los cuales 76 aportaron valor agregado y 13 no aportaron. 1747 estudiantes tienen nevera y no aportan valor agregado, se evalúa la condición de su estado civil, 1092 estudiantes tienen un estado civil diferente a soltero por lo tanto no tienen valor agregado, 655 estudiantes solteros tienen estado civil soltero, de los cuales 347 aportaron valor agregado y 308 no aportaron. Se evalúa la condición de los años de preparación que tomo el estudiante, 183 estudiantes que tomaron más de 6.5 años en terminar pregrado no aportaron valor agregado, 151 estudiantes tomaron menos de 5.5 años en terminar pregrado, de los cuales 128 aportaron valor

agregado y 23 no aportaron, 321 estudiantes que terminaron pregrado entre 5.5 y 6.5 años no aportaron valor agregado.

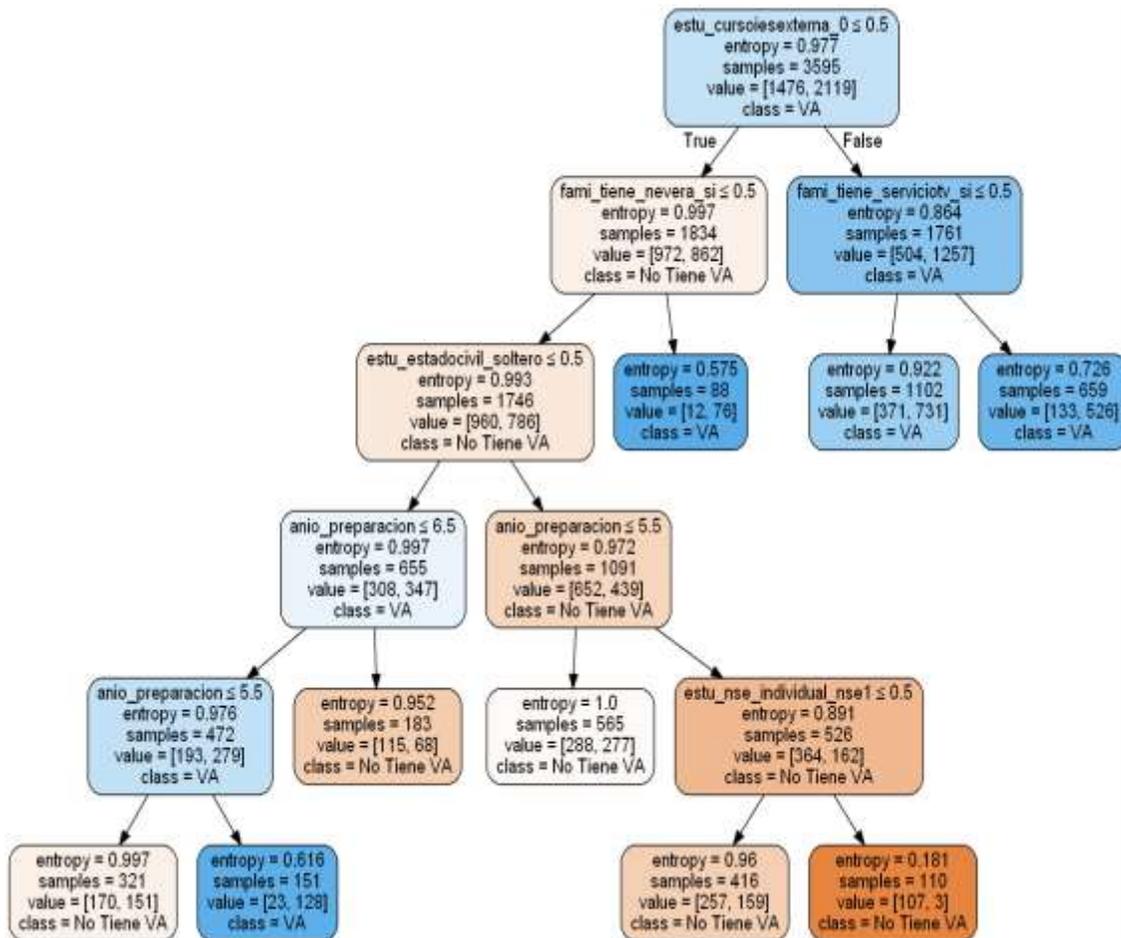


Fig. 38. Árbol 03 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa que 1761 no tomaron el curso, entonces se evalúa si tienen servicio de televisión, de 659 estudiantes que no tienen servicio de televisión, 526 aportaron valor agregado y 133 no aportaron, y de 1102 estudiantes que tienen servicio de televisión, 731 aportaron valor agregado y 371 no aportaron valor agregado. 1834 estudiantes tomaron el curso de apoyo externo, entonces se evalúa si en el hogar cuentan con nevera, 88 estudiantes no cuentan con nevera de los cuales 76 aportaron valor agregado y 12 no aportaron, 1746 estudiantes cuentan con nevera, se evalúa si son solteros, 1091 estudiantes no son solteros, por lo tanto, no aportan valor agregado, 655 estudiantes están solteros de los cuales 347 aportaron valor agregado y 308 no aportaron. Se evalúa la condición de los años de preparación que tomaron los estudiantes para terminar el pregrado, los cuales 183 estudiantes no aportaron valor agregado porque tomaron más de 6.5 años en terminar, 151 estudiantes tomaron menos de 5.5 años en terminar la carrera, los cuales 128 aportaron valor agregado y 23 no aportaron. 321 estudiantes tomaron entre 5.5 años y 6.5 años terminar el pregrado, por lo tanto, no aportaron valor agregado.

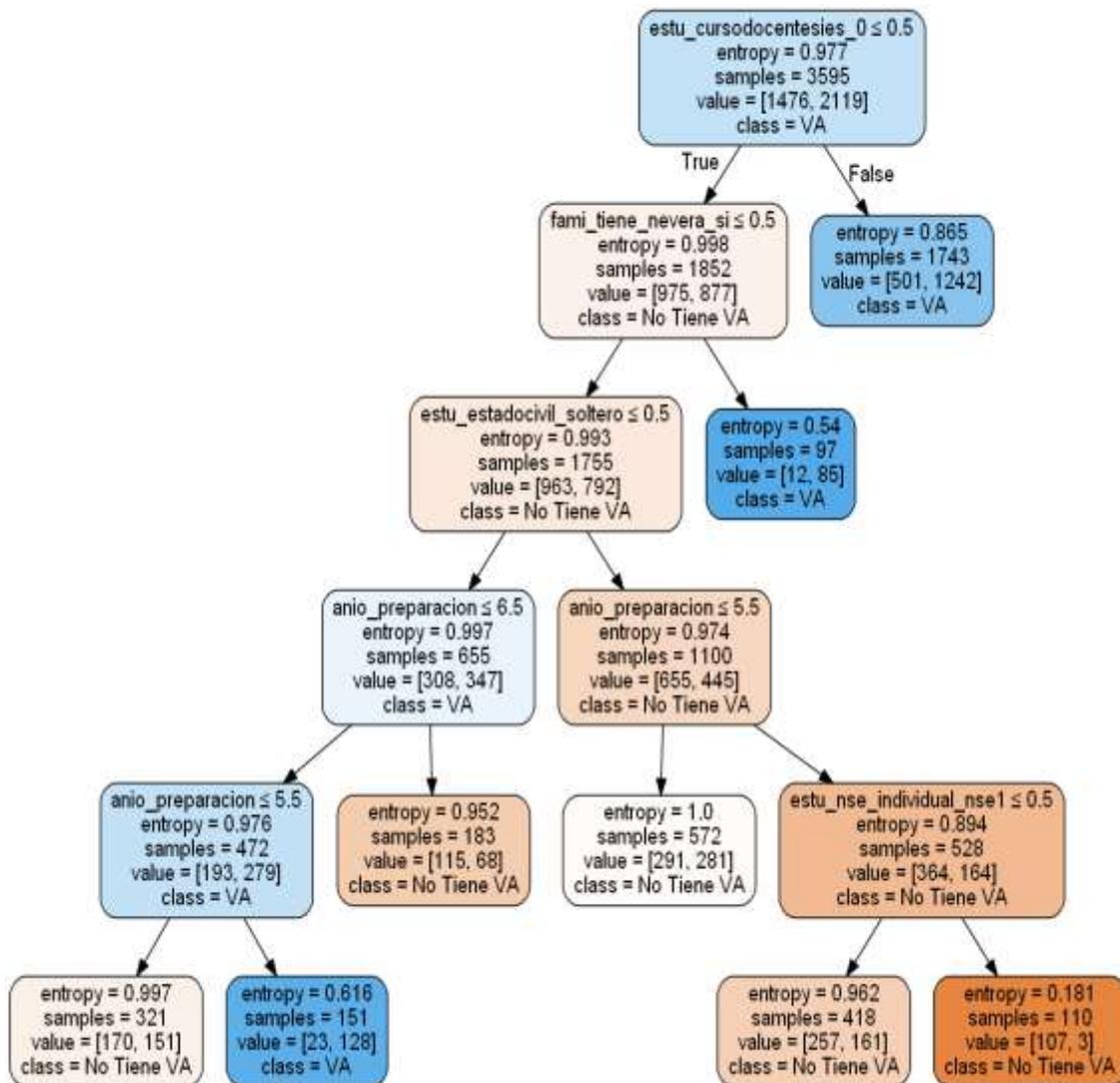


Fig. 39. Árbol 04 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa que 1743 tomaron el curso con docentes, de los cuales 1242 aportaron valor agregado y 501 no aportaron. 1852 no tomaron el curso, entonces se evalúa la condición si tienen nevera en el hogar, 97 estudiantes no tienen nevera de los cuales 85 aportaron valor agregado y 12 no aportaron, 1755 estudiantes tienen nevera, entonces se evalúa la condición de los años de preparación para terminar pregrado, 151 estudiantes terminaron pregrado entre 5.5 años y 6.5 años, de los cuales 128 aportaron valor agregado y 23 no aportaron, 321 estudiantes que terminaron de 5.5 años o menos no aportaron valor agregado.

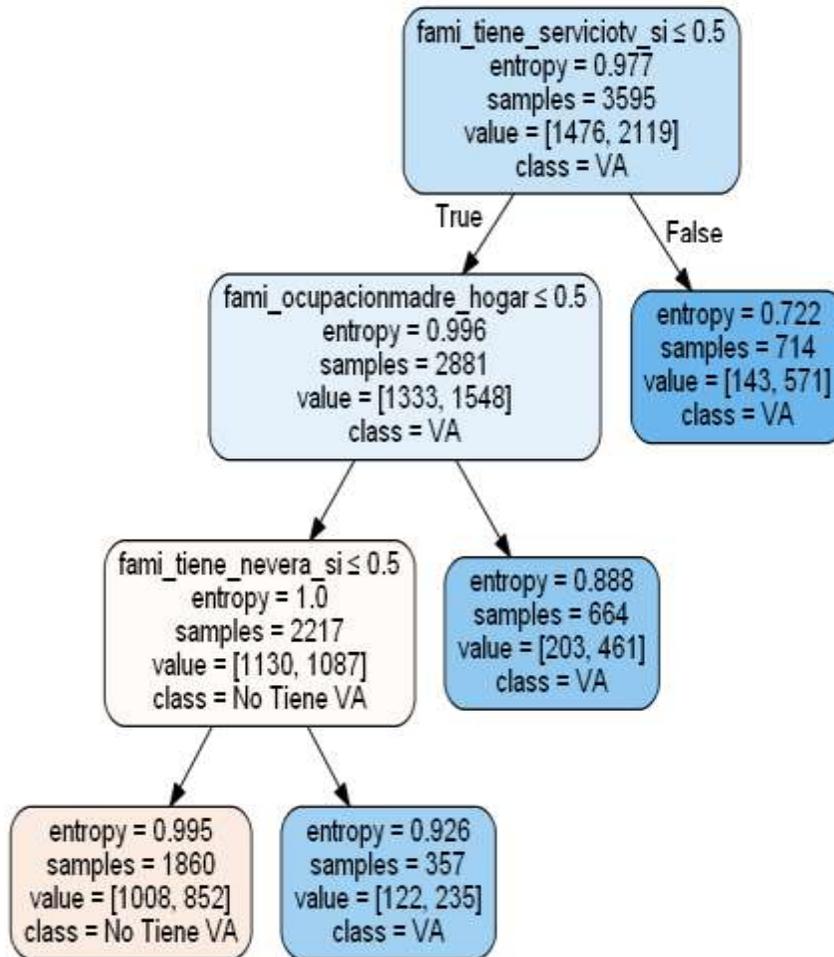


Fig. 40. Árbol 05 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa que 714 no tiene servicio de televisión de los cuales 571 aportaron valor agregado y 143 no aportaron valor agregado, si tienen servicio de televisión, se evalúa la ocupación de la madre de familia, 664 estudiantes están bajo la condición de que la madre no es ama de casa, los cuales 461 aportaron valor agregado y 203 no aportaron valor agregado, si la madre se ocupa del hogar, se evalúa la condición de que si tienen nevera, 357 estudiantes no tienen nevera de los cuales 235 aportaron valor agregado y 122 no aportaron valor agregado.

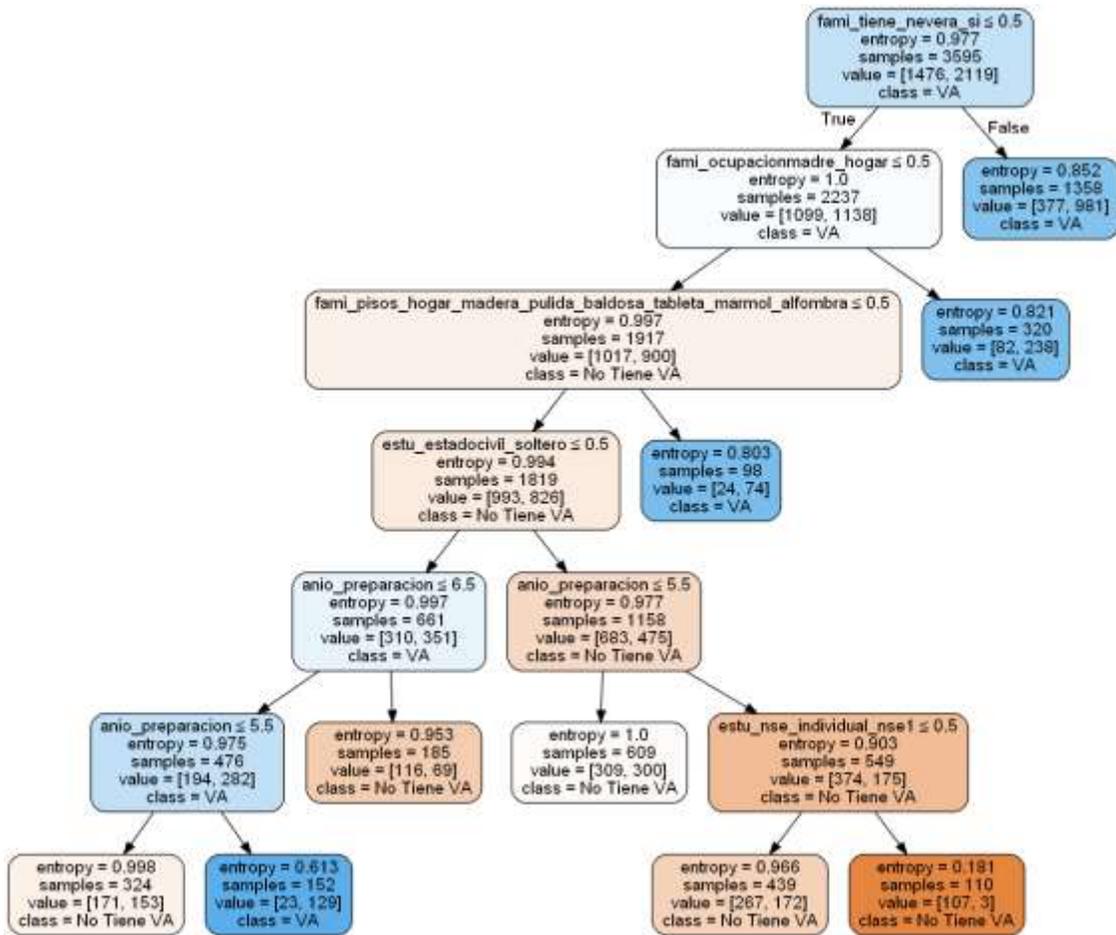


Fig. 41. Árbol 06 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa que 1358 estudiantes no tienen nevera de los cuales 981 generan valor agregado y 377 no generan valor agregado, 2237 estudiantes tienen nevera, entonces se evalúa la ocupación de la madre, 320 estudiantes están bajo la condición de que la madre no es ama de casa, los cuales 238 aportaron valor agregado y 82 no aportaron valor agregado, 1917 estudiantes están bajo la condición de que la madre tiene otra ocupación diferente al hogar, entonces se evalúa la condición del piso de la vivienda donde reside el estudiante, 98 estudiantes no cuentan con ningún tipo de piso comercial, lo cual se debe a condiciones de mucha pobreza, de los cuales 74 aportaron valor agregado y 24 no aportaron, 1819 cuentan con piso en casa, se evalúa la condición de estado civil soltero, 1158 estudiantes no son solteros, por lo tanto no aportan valor agregado, 661 estudiantes son solteros, de los cuales 351 aportaron valor agregado y 310 no aportaron, se evalúa la condición de los años de preparación para acabar el pregrado, 185 estudiantes tomaron más de 6.5 años en terminar, por lo tanto, no aportaron valor agregado, 152 estudiantes tomaron 5.5 años o menos en terminar, de los cuales 129 aportaron valor agregado y 23 no aportaron, 324 estudiantes que tomaron más de 5.5 años y menos de 6.5 años no aportaron valor agregado.

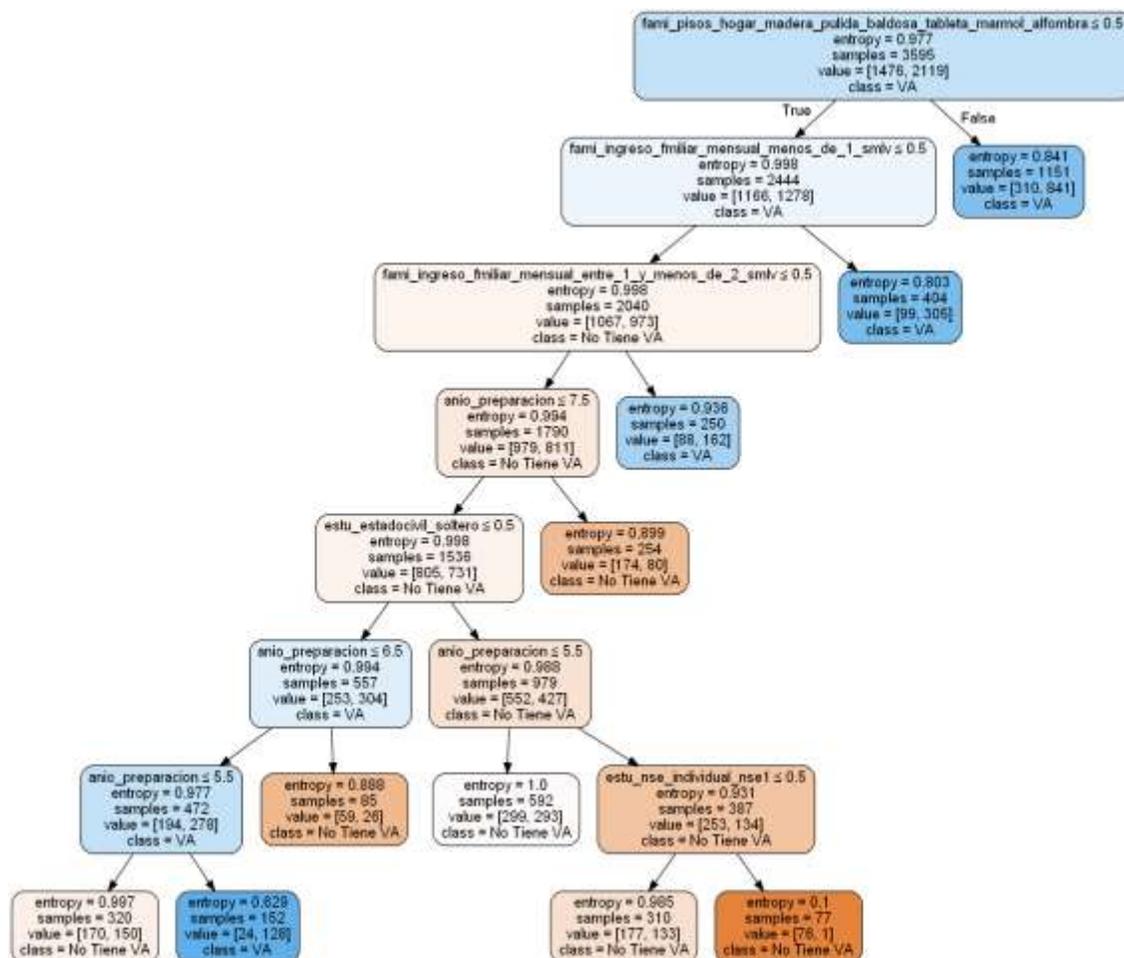


Fig. 42. Árbol 07 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa que 1151 estudiantes no cuentan con un tipo de piso comercial, de los cuales 841 aportaron valor agregado y 310 no aportaron, 2444 estudiantes cuentan con un tipo de piso comercial en sus viviendas, por lo tanto, se evalúa la condición del ingreso económico mensual familiar 404 estudiantes registraron ingreso económico menor a 1 salario mínimo, de los cuales 305 aportaron valor agregado y 99 no aportaron. 250 estudiantes registraron ingresos entre 1 y 2 salarios mínimos, de los cuales 162 aportaron valor agregado y 88 no aportaron, 1790 registraron ingresos mayores a 2 salarios mínimos entonces se evalúa la condición de estado civil soltero, 979 estudiantes no son solteros, por lo tanto, no aportaron valor agregado, 557 estudiantes se registraron como solteros, de los cuales 304 aportaron valor agregado y 253 no aportaron, se evalúa la condición de los años de preparación que tomaron los estudiantes para acabar el pregrado, 85 estudiantes tomaron más de 6.5 años en finalizar estudios, por lo tanto, no aportaron valor agregado, 152 estudiantes tomaron de 5.5 años o menos en terminar la carrera, de los cuales 128 aportaron valor agregado y 24 no aportaron.

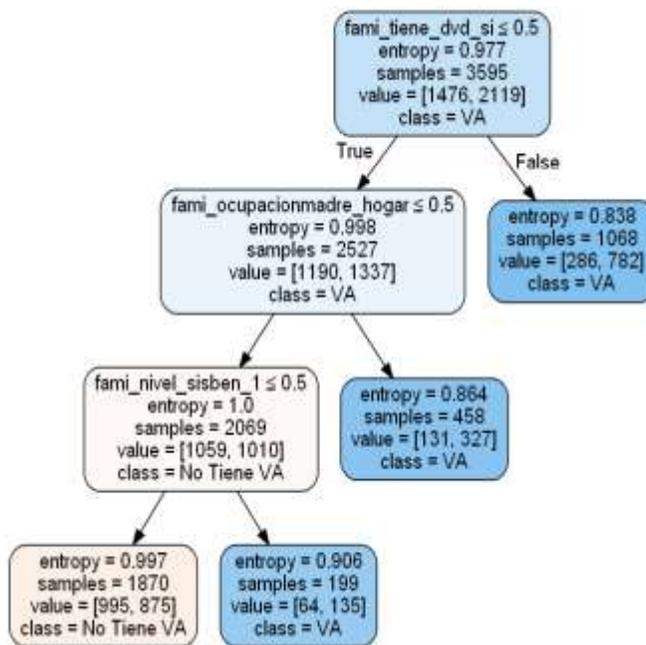


Fig. 43. Árbol 08 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa la condición de que si en el hogar se cuenta con dispositivo reproductor DVD, 1068 estudiantes no cuenta con reproductor de DVD, de los cuales 782 aportaron valor agregado y 286 no aportaron valor agregado, se evalúa la condición que registra si la madre de familia es ama de casa, 458 estudiantes cumplen con la condición de que la madre de familia no es ama de casa, entonces 327 estudiantes aportaron valor agregado y 131 no aportaron, 2069 estudiantes cumple la condición de que la madre de familia es ama de casa, por lo tanto, se evalúa la condición del nivel de Sisbén de la familia, 199 estudiantes tienen nivel de Sisbén 1, de los cuales 135 aportaron valor agregado y 64 no aportaron.

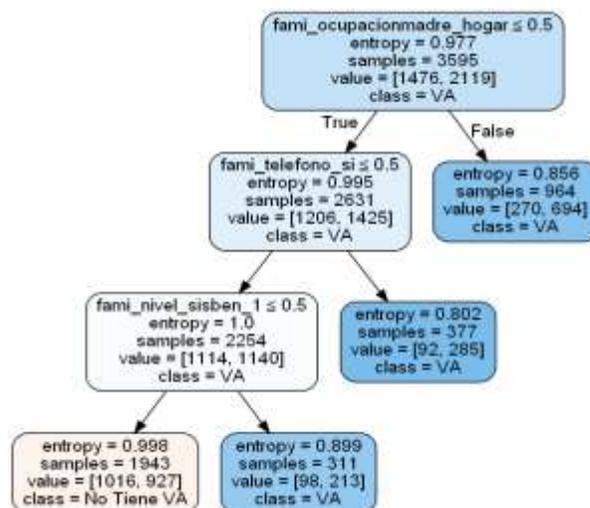


Fig. 44. Árbol 09 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa la condición que registra si la madre es ama de casa, 964 estudiantes viven bajo la condición de que la madre de familia no es ama de casa, los cuales 694 aportaron valor agregado y 270 no aportaron, 2631 viven bajo la condición de que la madre de familia es ama de casa, entonces se evalúa la condición de que si hay teléfono en el hogar, 377 estudiantes no cuentan con teléfono, 285 aportaron valor agregado y 92 no aportaron valor agregado, 2254 cuentan con teléfono entonces se evalúa la condición del nivel de Sisbén 1, 311 estudiantes tienen nivel de Sisbén 1, 213 aportaron valor agregado y 98 no aportaron valor agregado.

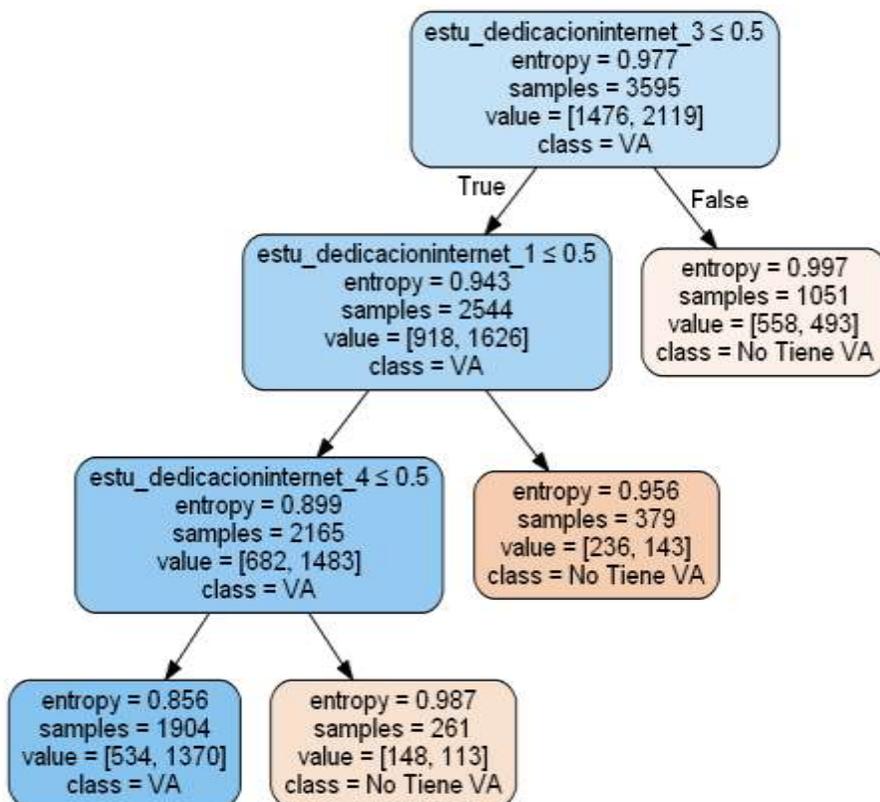


Fig. 45. Árbol 10 - Competencias Ciudadanas
Fuente: esta investigación.

Se evalúa la condición de la dedicación entre 1 y 3 horas diarias a internet fuera de actividades académicas, 1051 estudiantes que dedicaron entre 1 y 3 horas a internet no aportaron valor agregado, 2544 dedicaron menos de 1 hora diaria a internet, de los cuales 1626 aportaron valor agregado y 918 no aportaron, se evalúa la condición de los estudiantes que dedicaron más de 4 horas a internet, 261 que dedicaron más de 4 horas no aportaron valor agregado, 1904 se dedicaron menos de 4 horas diarias, de los cuales 1370 aportaron valor agregado y 534 no aportaron valor agregado.

CLASE INGLÉS

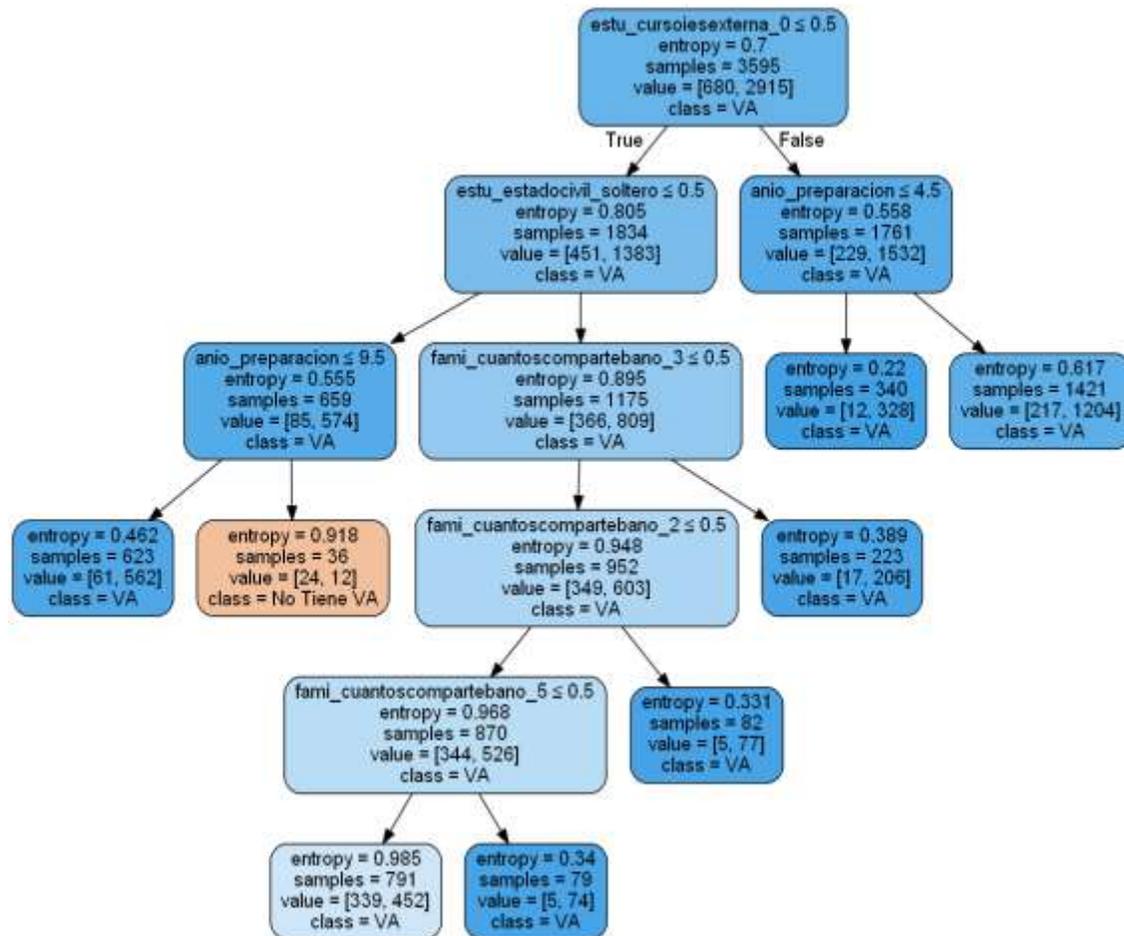


Fig. 46. Árbol 01 – inglés

Fuente: esta investigación.

Se evalúa la condición si tomaron curso para preparación de saber pro fuera de la institución, 1761 no tomaron el curso de los cuales 1532 aportaron valor agregado y 229 no aportaron, se evalúa la condición de que si tomaron 4.5 años o menos en terminar el pregrado, 1421 tomaron más de 4.5 años de los cuales 1204 aportaron valor agregado y 217 no aportaron valor agregado, y 340 tomaron menos de 4.5 años en terminar pregrado, de los cuales 328 aportaron valor agregado y 12 no aportaron. 1834 estudiantes tomaron curso de preparación fuera de la institución de los cuales 1383 aportaron valor agregado y 451 no aportaron, entonces se evalúa la condición si el estado civil de los estudiantes es soltero, 1175 estudiantes no son solteros, de los cuales 809 aportaron valor agregado y 386 no aportaron valor agregado, se evalúa la condición de las personas que comparten baño en un hogar, 223 comparten baño con 3 personas, de los cuales 206 aportaron valor agregado y 17 no aportaron, 82 estudiantes comparten baño con 2 personas, de los cuales 77 estudiantes aportaron valor agregado y 5 no aportaron, 791 estudiantes comparten baño con 5 personas, de los cuales 452 aportaron valor agregado y 339 no aportaron valor agregado. 659 estudiantes son solteros entonces se evalúa la condición del tiempo de preparación para pregrado, 623 tomaron menos de 9.5 años, de los cuales 562 aportaron valor agregado y 61 no aportaron.

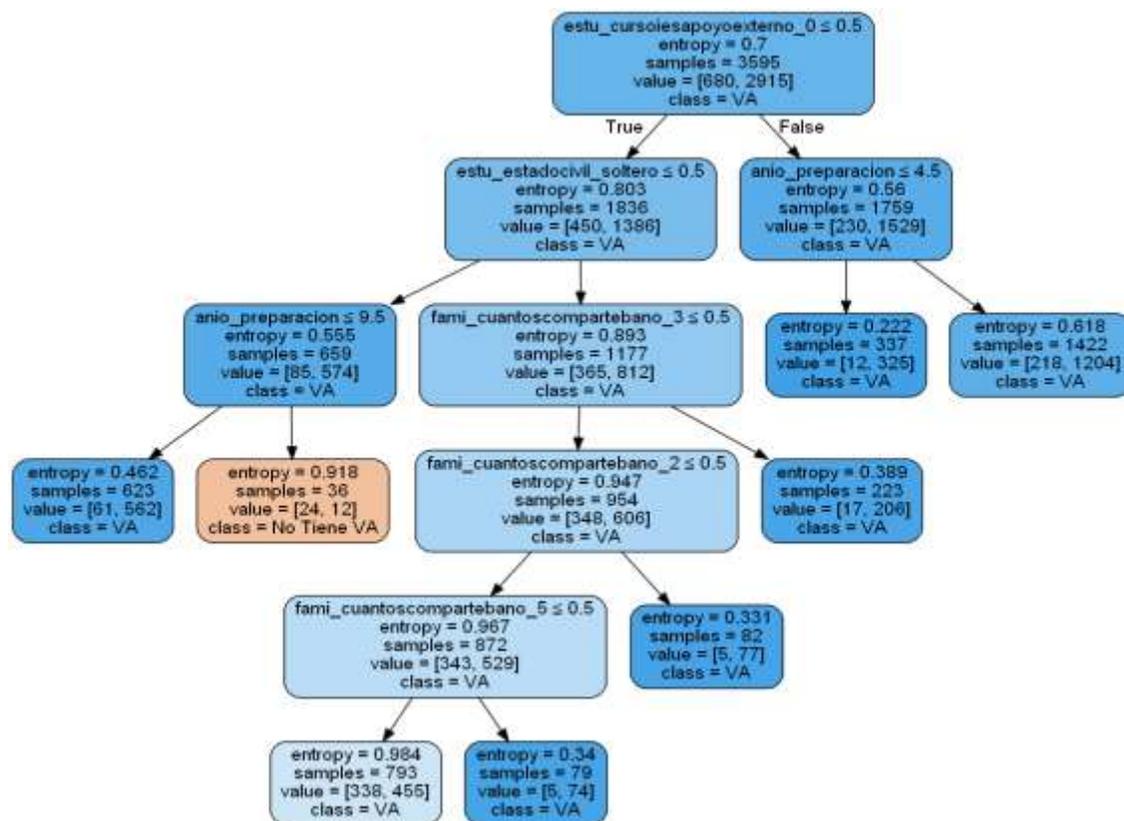


Fig. 47. Árbol 02 – inglés
Fuente: esta investigación.

Se evalúa la condición de que si el estudiantes tomo curso de preparación externo con apoyo de docentes internos, 1759 estudiantes no tomaron curso de preparación, de los cuales 1529 aportaron valor agregado y 230 no aportaron valor agregado, se evalúa la condición de los años de preparación para finalizar el pregrado, 337 tomaron menos de 4.5 años, de los cuales 325 aportaron valor agregado y 12 no aportaron, 1422 tomaron más de 4.5 años de los cuales 1240 aportaron valor agregado y 218 no aportaron. 1836 estudiantes tomaron curso de preparación, de los cuales 1386 aportaron valor agregado y 450 no aportaron. Se evalúa la condición del estado civil de los estudiantes, 1177 estudiantes no son solteros, 812 aportaron valor agregado y 365 no aportaron, se evalúa la condición del número de personas con las que el estudiantes comparte baño, 223 comparten baño con 3 personas, los cuales 206 aportaron valor agregado y 17 no aportaron, 82 personas comparten baño con 2 personas los cuales 77 aportaron valor agregado y 5 no aportaron, 793 compartieron baño con 5 personas los cuales 455 aportaron valor agregado y 338 no aportaron. 659 estudiantes son solteros de los cuales 574 aportaron valor agregado y 85 no aportaron, se evalúa la condición de los años de preparación para terminar pregrado, 623 tomaron menos de 9.5 años, de los cuales 562 aportaron valor agregado y 61 no aportaron.

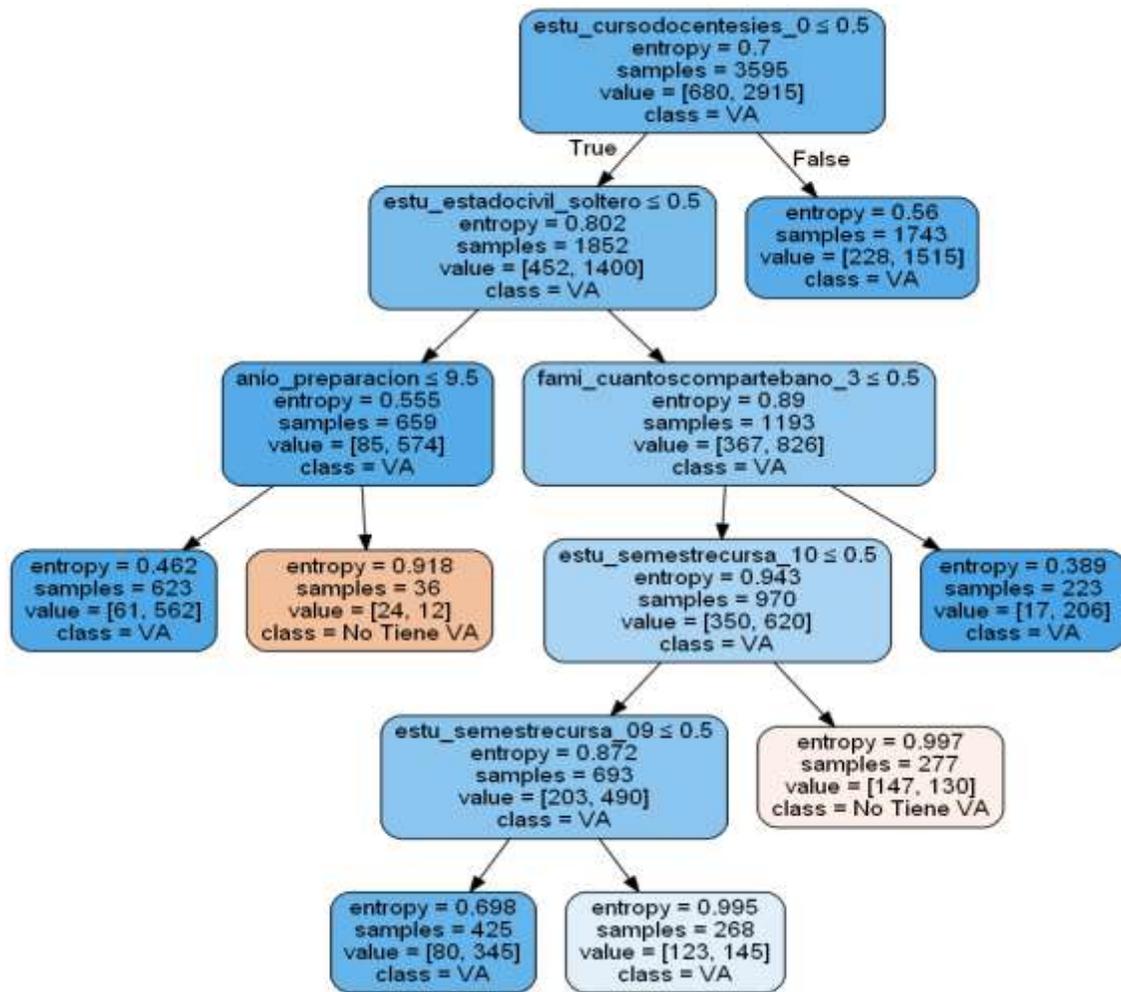


Fig. 48. Árbol 03 – inglés
Fuente: esta investigación.

Se evalúa la condición si tomo curso de preparación con docentes de la institución, 1743 no tomaron el curso, de los cuales 1515 aportaron valor agregado y 228 no aportaron, 1852 tomaron el curso de preparación de los cuales, 1400 aportaron valor agregado y 452 no aportaron. Se evalúa la condición del estado civil del estudiantes, 1193 no son solteros, de los cuales 826 aportaron valor agregado y 367 no aportaron, se evalúa la condición de cuántas personas comparten baño con el estudiantes, 223 comparten baño con 3 personas, las cuales 206 aportaron valor agregado y 17 no aportaron, se evalúa la condición del semestre que cursó el estudiantes en ese entonces, 693 estaban en decimo semestre, de los cuales 490 aportaron valor agregado y 203 no aportaron, 425 estudiantes estaban en noveno semestre, de los cuales 345 aportaron valor agregado y 80 no aportaron.

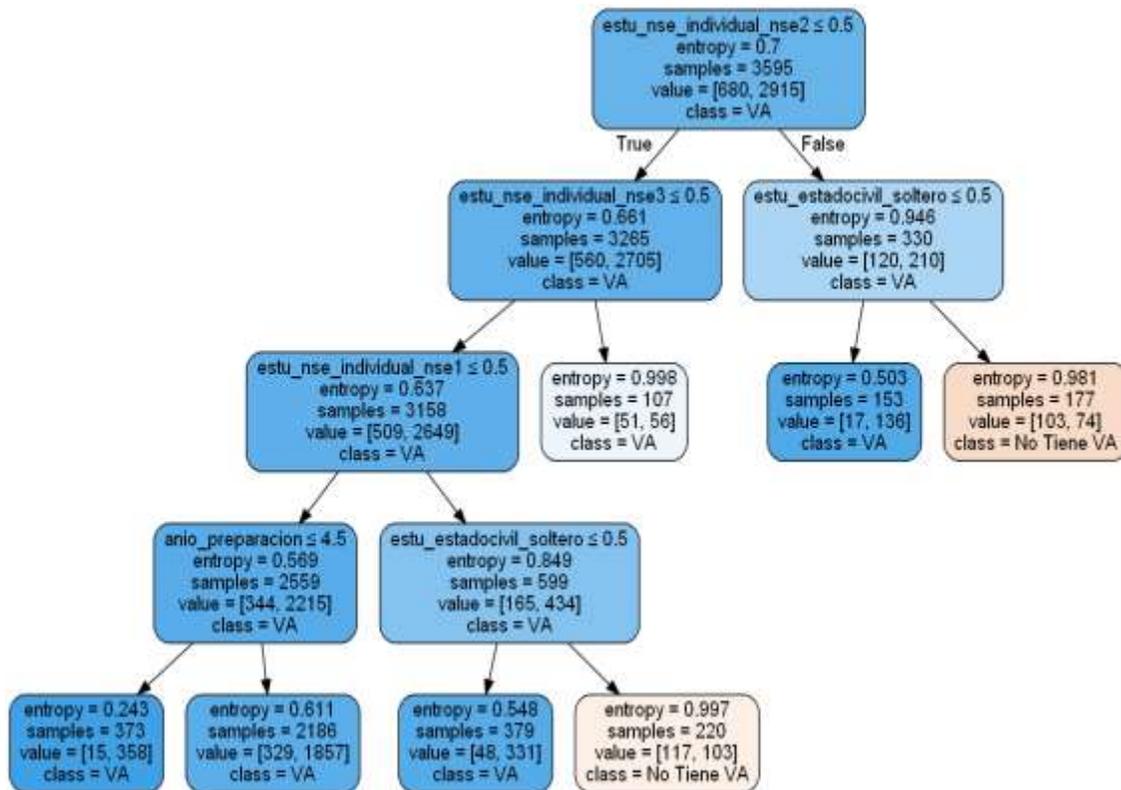


Fig. 49. Árbol 04 – inglés
Fuente: esta investigación.

Se evalúa la condición del nivel socioeconómico individual 2, 330 estudiantes no cumplen con la condición, de los cuales 210 aportaron valor agregado y 120 no aportaron, entonces, se evalúa la condición del estado civil de los estudiantes, 153 estudiantes son solteros, de los cuales 136 aportaron valor agregado y 17 no aportaron. 3565 estudiantes cumplen con la condición los cuales 2705 aportaron valor agregado y 560 no aportaron. Se evalúa la condición del nivel socioeconómico individual 3, 107 no cumplen con la condición de los cuales 56 aportaron valor agregado y 51 no aportaron, 3158 cumplen con la condición, de los cuales 2649 aportaron valor agregado y 509 no aportaron, entonces, se evalúa la condición del nivel socioeconómico individual 1, 379 estudiantes, no tienen nivel socioeconómico 1 y no son solteros, de los cuales 331 aportaron valor agregado y 48 no aportaron. 2186 estudiantes tomaron 4.5 años o menos en terminar el pregrado y tienen nivel socioeconómico 1, de los cuales 1857 aportaron valor agregado y 329 no aportaron.

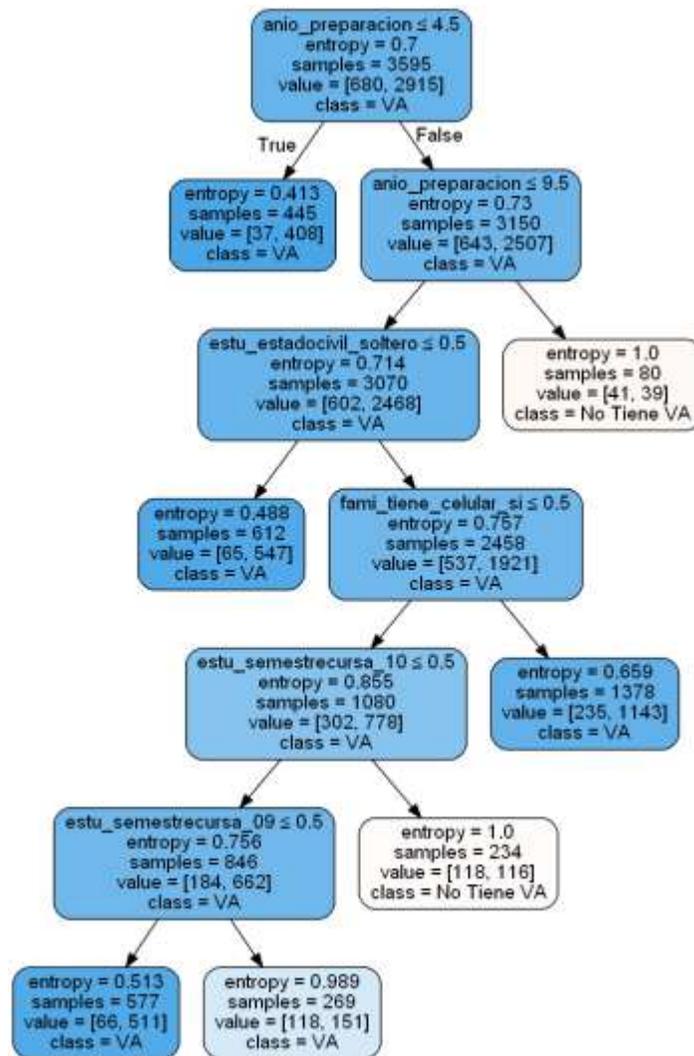


Fig. 50. Árbol 05 – inglés
Fuente: esta investigación.

Se evalúa la condición de los años que tomo los estudiantes en terminar pregrado, 445 estudiantes tomaron de 4.5 años o menos en terminar pregrado, de los cuales 408 aportaron valor agregado y 37 no aportaron, 3150 tomaron más de 4.5 años de preparación.

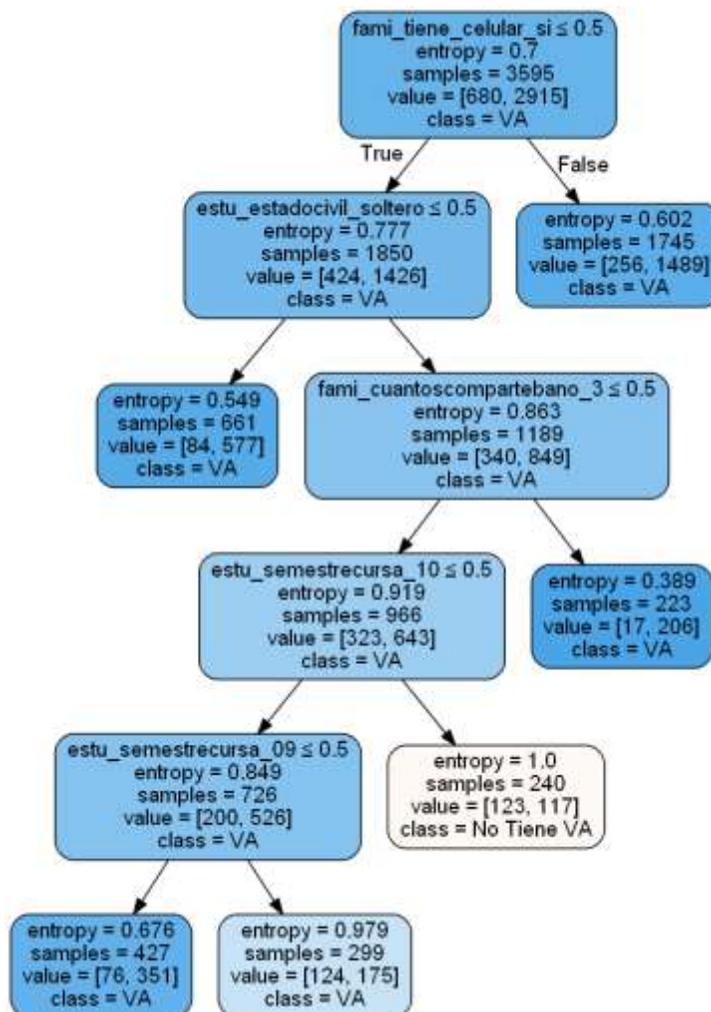


Fig. 51. Árbol 06 – inglés
Fuente: esta investigación.

Se evalúa la condición de si el estudiantes cuenta con teléfono celular, 1745 estudiantes no cuentan con teléfono celular, de los cuales 1489 aportaron valor agregado y 256 no aportaron valor agregado, 1850 cuentan con teléfono celular, de los cuales 1426 aportaron valor agregado y 424 no aportaron valor agregado, entonces, se evalúa el estado civil del estudiantes, 661 estudiantes son solteros, de los cuales 577 aportaron valor agregado y 84 no aportaron, 1189 estudiantes no son solteros, de los cuales 849 aportaron valor agregado y 340 no aportaron valor agregado, se evalúa la condición del número de personas con las que comparte baño el estudiantes, 223 estudiantes comparten baño con 3 o 4 estudiantes, de los cuales 206 aportaron valor agregado y 17 no aportaron, 966 estudiantes no comparten baño con 3 o 4 estudiantes, de los cuales 643 aportaron valor agregado y 323 no aportaron. Se evalúa la condición del semestre que estaban cursando en ese entonces los estudiantes, 427 estudiantes estaban en noveno semestre de los cuales 351 aportaron valor agregado y 76 no aportaron y 299 estudiantes estaban en semestres inferiores de los cuales 175 aportaron valor agregado y 124 no aportaron.

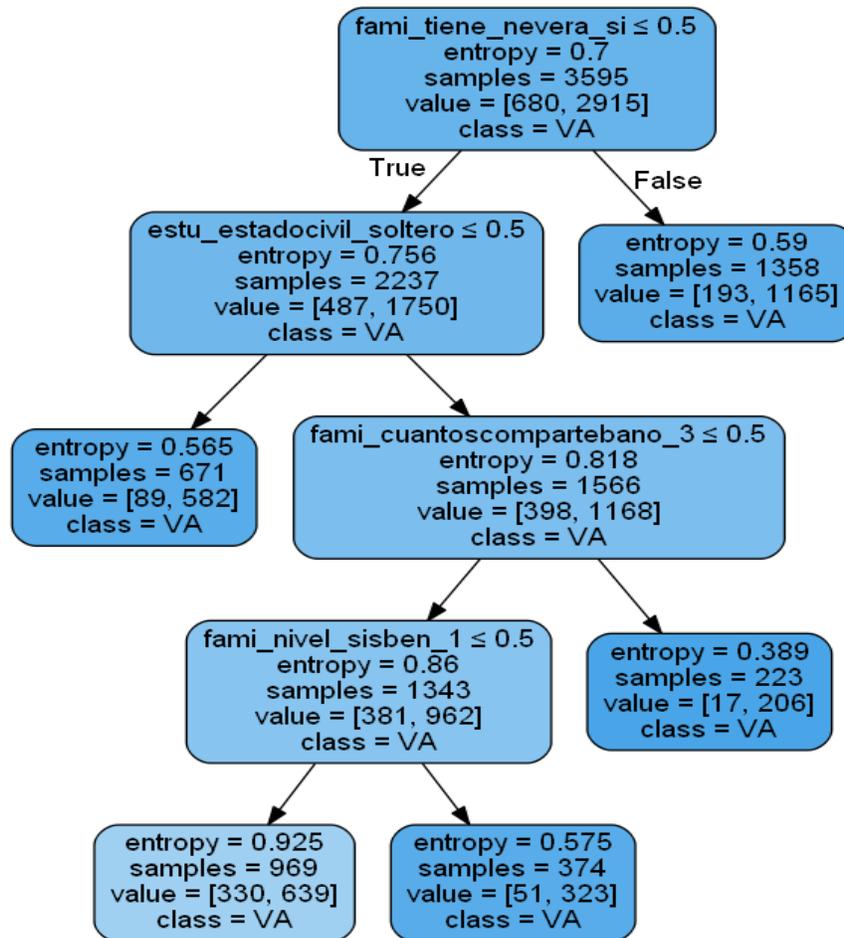


Fig. 52. Árbol 07 – inglés
Fuente: esta investigación.

Se evalúa la condición de si en el hogar hay nevera, 1358 estudiantes registraron que no tienen nevera, de los cuales 1165 aportaron valor agregado y 193 no aportaron, 2237 estudiantes tienen nevera, de los cuales 1750 aportaron valor agregado y 487 no aportaron, se evalúa la condición del estado civil de los estudiantes, 671 estudiantes son solteros de los cuales 582 aportaron valor agregado y 89 no aportaron, 233 estudiantes no son solteros y comparten baño con 3 o 4 personas, de los cuales 206 aportaron valor agregado y 17 no aportaron, 969 estudiantes tienen nivel de Sisbén 1 y no comparten baño con 3 o 4 personas, de los cuales 639 aportaron valor agregado y 330 no aportaron, 374 estudiantes tienen nivel de Sisbén diferente a 1, de los cuales 323 aportaron valor agregado y 51 no aportaron.

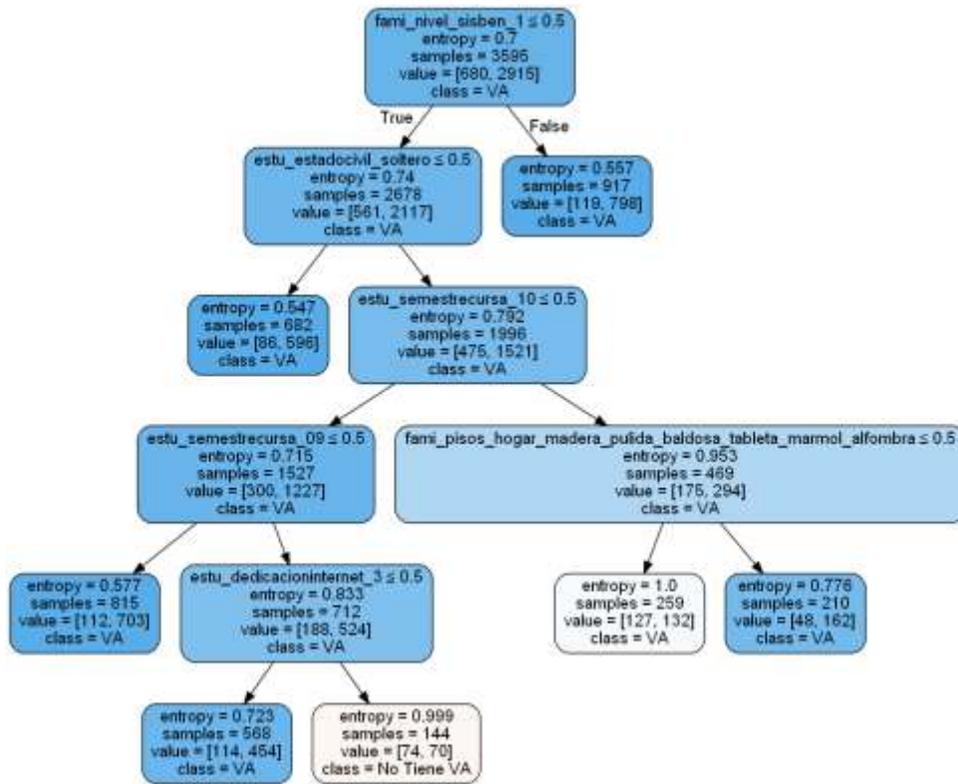


Fig. 53. Árbol 08 – inglés

Fuente: esta investigación.

Se evalúa la condición primaria del nivel de Sisbén del hogar, 917 estudiantes tienen nivel de Sisbén diferente a 1, de los cuales 798 aportaron valor agregado y 119 no aportaron, se evalúa la condición del estado civil del estudiantes, 682 estudiantes son solteros y tienen nivel de Sisbén 1, de los cuales 596 aportaron valor agregado y 86 no aportaron, se evalúa la condición del semestre que el estudiantes cursaba en ese entonces, y la condición del tipo de piso que tiene el lugar de vivienda del estudiantes, 210 estudiantes no cursan decimo semestre y no tienen un tipo de piso conocido, de los cuales 162 aportaron valor agregado y 48 no aportaron. 815 estudiantes están en noveno semestre de los cuales 703 aportaron valor agregado y 112 no aportaron. Se evalúa la condición del número de horas de dedicación a internet fuera de actividades académicas por parte de los estudiantes, 568 estudiantes se dedican de 1 a 3 horas a internet y cursan noveno semestre, de los cuales 454 aportaron valor agregado y 114 no aportaron.

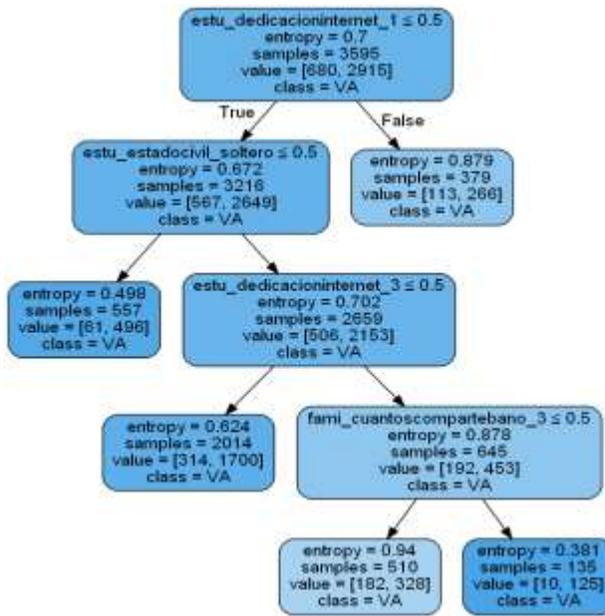


Fig. 54. Árbol 09 – inglés
Fuente: esta investigación.

Se evalúa la condición del número de horas de dedicación a internet fuera de actividades académicas por parte del estudiantes, 379 se dedicaron más de una hora, de los cuales 266 aportaron valor agregado y 113 no aportaron, se evalúa la condición del estado civil del estudiantes, 557 estudiantes son solteros y se dedican 1 hora a internet, de los cuales 496 aportaron valor agregado y 61 no aportaron, se evalúa la condición del número de personas con las que comparte baño el estudiantes, 135 estudiantes comparten baño con menos de tres personas, dedican menos de 3 horas a internet y no son solteros de los cuales 125 aportaron valor agregado y 10 no aportaron.

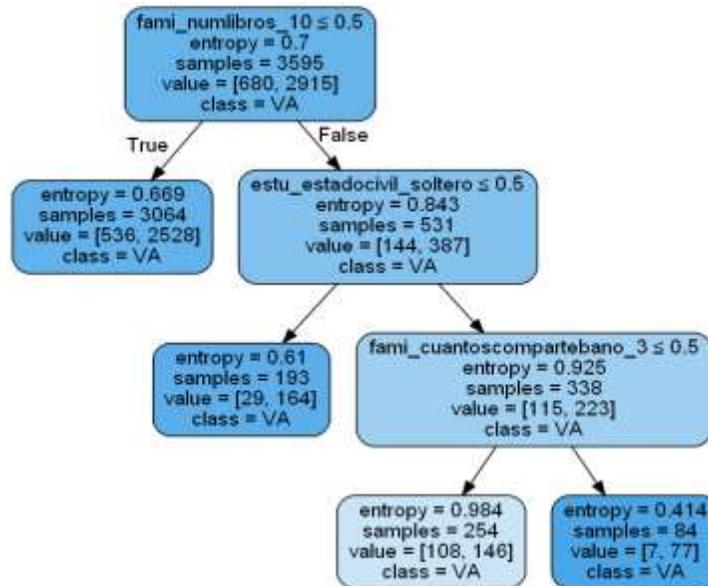


Fig. 55. Árbol 10 – inglés
Fuente: esta investigación.

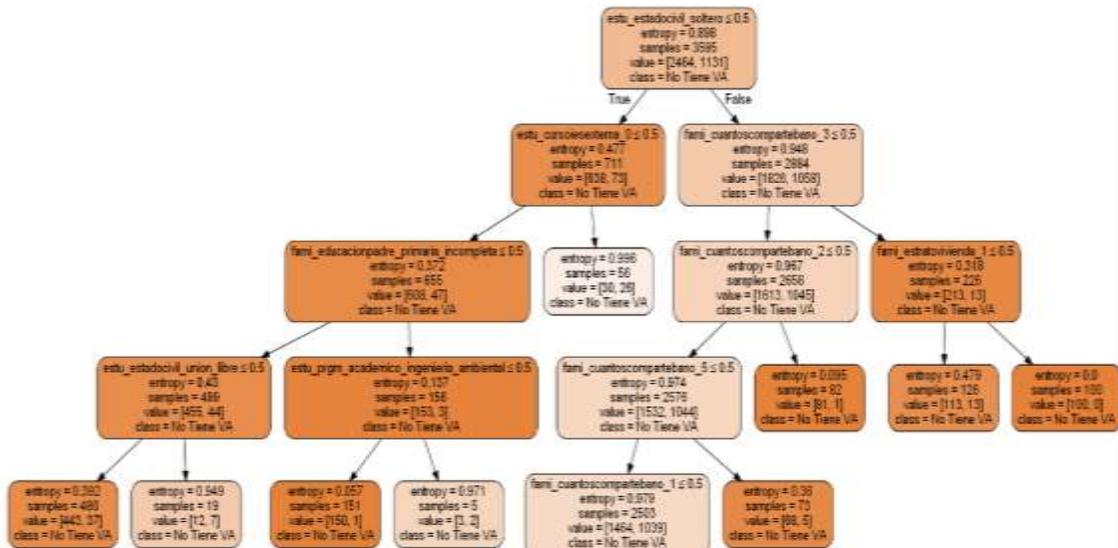


Fig. 57. Árbol 02 – Lectura critica
Fuente: esta investigación.

La primera condición, si el estado civil es soltero se tiene que 2464 estudiantes no generan valor agregado con respecto a la clase lectura crítica, además comparten baño con 3 personas y el estrato de la vivienda es 1 no generan valor agregado 100 estudiantes.

Si el estado civil es diferente de soltero y no realizo un curso en un instituto de preparación externo para las pruebas Saber Pro, además la educación del padre es primaria incompleta y él estudia ingeniería ambiental no generan valor agregado 150 estudiantes.

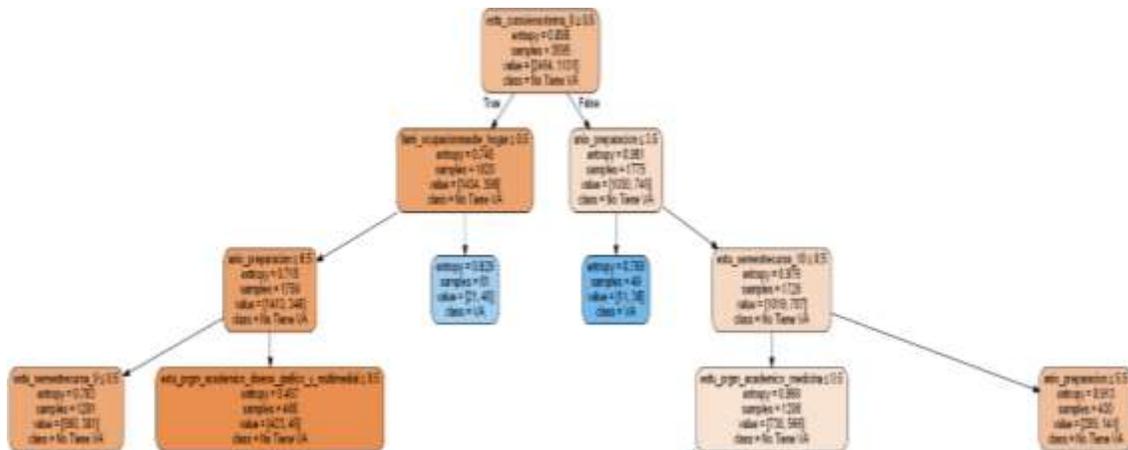


Fig. 58. Árbol 03 – Lectura critica
Fuente: esta investigación.

No tomo curso de preparación externo y el año de preparación es menor a 3.5 años 38 estudiantes generan valor agregado, si los estudiantes no cursan decimo semestre y estudian medicina 730 estudiantes no generan valor agregado.

Si los estudiantes no tomaron un curso en un instituto de preparación externo y la ocupación de la madre es hogar 40 estudiantes generan valor agregado. Si es la ocupación es diferente a hogar y el

año de preparación es mayor a 6.5 años y estudian diseño gráfico y multimedia 423 estudiantes no generan valor agregado.

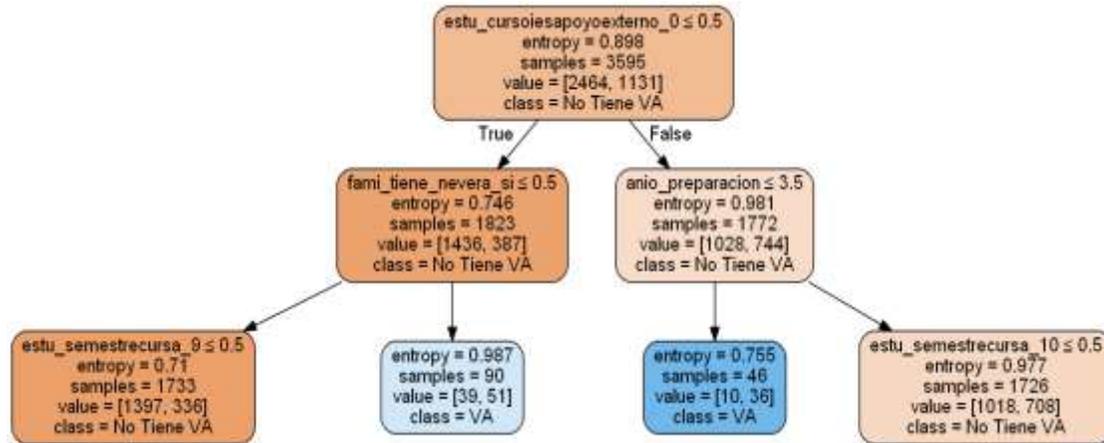


Fig. 59. Árbol 04 – Lectura crítica
Fuente: esta investigación.

La primera condición no tomo curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos y el año de preparación es menor o igual a 3.5 años de los cuales 36 estudiantes generan valor agregado. y si es mayor a 3.5 años de preparación y está en un semestre menor a 10 no generan valor agregado 1018 estudiantes.

Si no cumple que tomo el curso de preparación y tiene nevera 51 estudiantes generan valor agregado, de lo contrario si no tiene nevera y están en semestres menores a 9 1397 estudiantes no generan valor agregado.

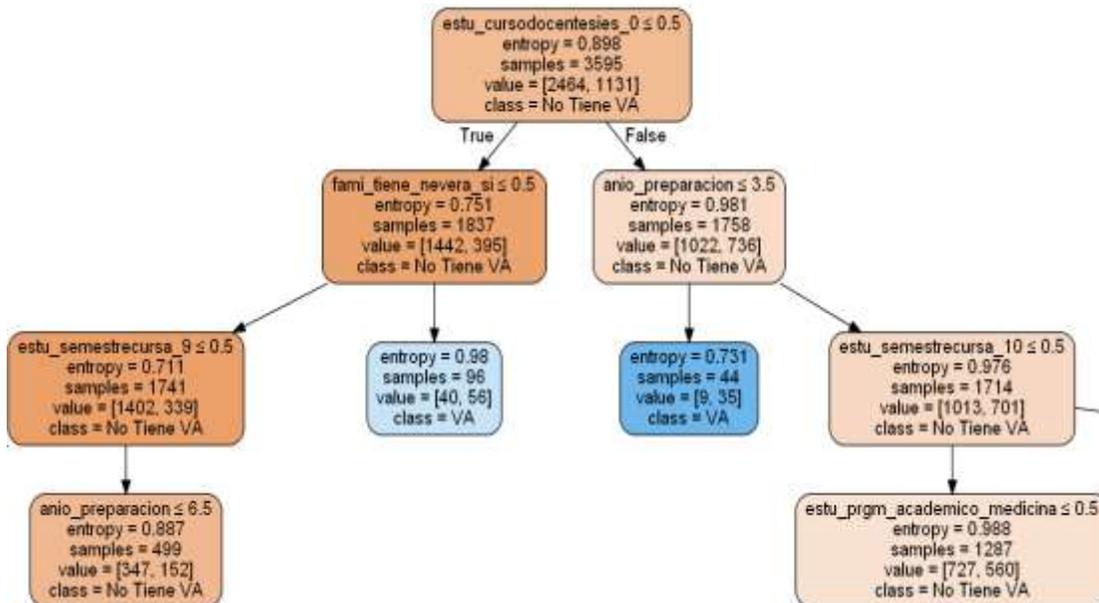


Fig. 60. Árbol 05 – Lectura crítica
Fuente: esta investigación.

La primera condición no se preparó para el examen con docentes de la institución y el año de preparación es menor o igual a 3.5 años 35 estudiantes generan valor agregado, si es mayor a 3.5 años de preparación y cursa semestres inferiores a 10 semestre y son de programas diferentes a medicina 727 estudiantes no generan valor agregado.

Si no se preparó para el examen con docente de la institución y tiene nevera 56 estudiantes generan valor agregado para la clase lectura crítica.

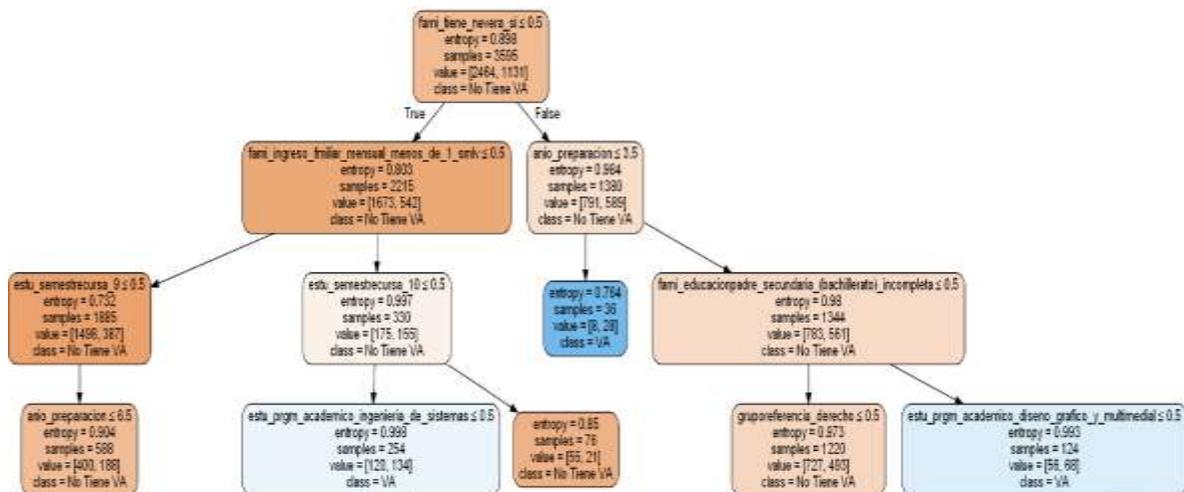


Fig. 61. Árbol 06 – Lectura crítica

Fuente: esta investigación.

La condición si tiene nevera y los años de preparación son menores o iguales 3.5 años generan valor agregado 28 estudiantes, si es mayor a 3.5 años, la educación del padre secundaria incompleta y el programa académico es diseño gráfico y multimedia 68 estudiantes tienen valor agregado. Si la educación del padre es diferente a secundaria incompleta y el grupo de referencia es derecho 727 estudiantes no tienen valor agregado.

Si no tiene nevera, el ingreso familiar mensual menos de 1 SMLV y cursa decimo semestre 55 estudiantes no tienen valor agregado. sí cursa semestre diferente a decimo semestre y el programa académico es ingeniería de sistemas 134 estudiantes tienen valor agregado.

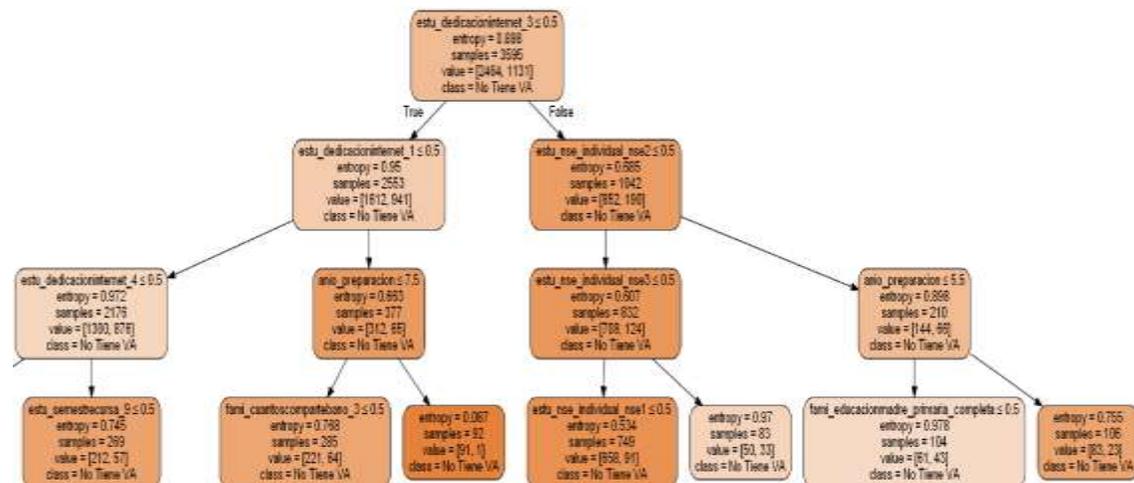


Fig. 62. Árbol 07 – Lectura crítica

Fuente: esta investigación.

La primera condición si los estudiantes no comparten baño con menos de tres personas, el nivel Sisbén es 1 y los años de preparación son menores o iguales a 2.5 años 11 estudiantes generan valor agregado. si el nivel de Sisbén es diferente a 1, los pisos del hogar son madera pulida, baldosa, tableta, mármol, alfombra y el programa académico es tecnología en promoción en salud existen 12 estudiantes que generan valor agregado.

Si comparten baño 3 personas y el estrato de vivienda es 1, no generan valor agregado 102 estudiantes.

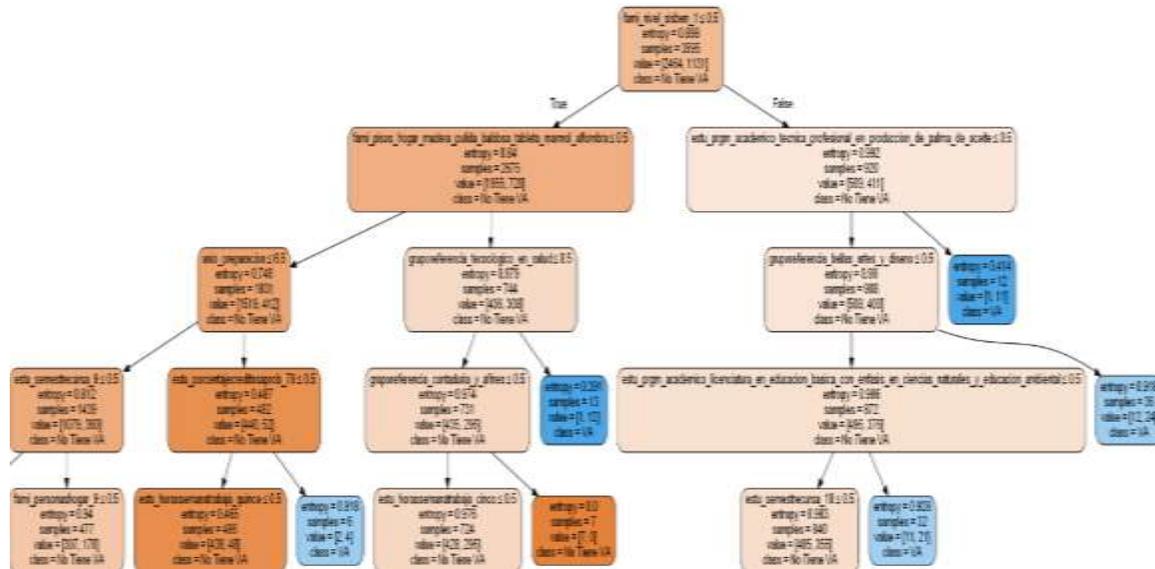


Fig. 65. Árbol 10 – Lectura crítica
Fuente: esta investigación.

La primera condición nivel de Sisbén 1 y el programa académico técnico profesional en producción de palma de aceite 11 estudiantes generan valor agregado, si no es técnico profesional en producción de palma de aceite, pero el grupo de referencia es artes y diseño 24 estudiantes generan valor agregado, si no es grupo de referencia es artes y diseño, pero el programa académico es licenciatura en educación básica con énfasis en ciencias naturales y educación ambiental 21 estudiantes generan valor agregado.

Si no es nivel de Sisbén 1, tiene pisos en el hogar madera pulida, baldosa tableta, mármol, alfombra y el grupo de referencia es tecnológico en salud 12 estudiantes generan valor agregado. si no tiene pisos en el hogar madera pulida, baldosa tableta, mármol, alfombra, los años de preparación son 6.5 y el porcentaje de crédito aprobado entre el 76% y 80% generan valor agregado 4 estudiantes.

Si el porcentaje crédito aprobado no está entre el 76% y 80% y las horas de trabajo son 15 generan valor agregado 2 estudiantes.

CLASE RAZONAMIENTO CUANTITATIVO

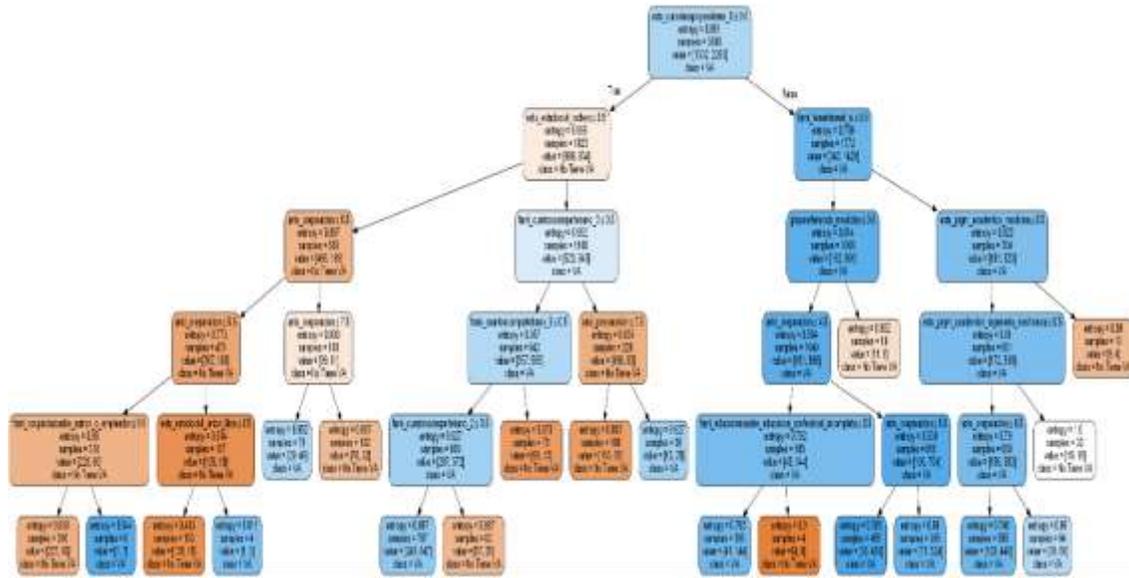


Fig. 66. Árbol 01 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición se preparó para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos de menos de 20 horas, tiene internet y estudia en el programa académico de medicina 9 estudiantes no generan valor agregado, si estudiaron el programa académico de ingeniería electrónica y el año de preparación es mayor a 6.5 años 58 estudiantes generaron valor agregado y si el año de preparación es menor o igual a 6.5 años 445 estudiantes generan valor agregado.

Si no se preparó para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos menos de 20 horas, el estado civil es soltero, comparte baño con 3 personas y el año de preparación es mayor a 7.5 años 25 estudiantes generan valor agregado. Si comparten baño menos de dos personas 547 estudiantes generan valor agregado

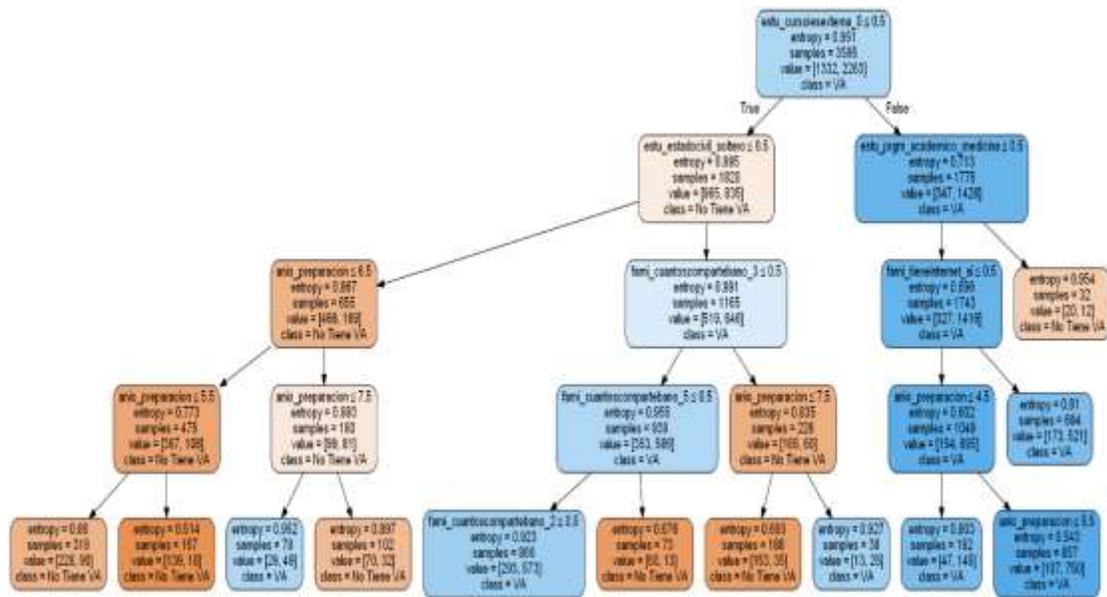


Fig. 67. Árbol 02 – Razonamiento cuantitativo

Fuente: esta investigación.

La primera condición no se preparó para el examen Saber Pro en un instituto de preparación de exámenes menos de 20 horas, pertenece al programa académico de medicina 20 estudiantes no generan valor agregado, si no pertenecen a medicina, tienen internet 521 estudiantes generan valor agregado.

Si no se preparó para el examen Saber Pro en un instituto de preparación de exámenes menos de 20 horas, el estado civil es soltero, comparte baño con 3 personas y el año de preparación es mayor a 7.5 años 25 estudiantes generan valor agregado.

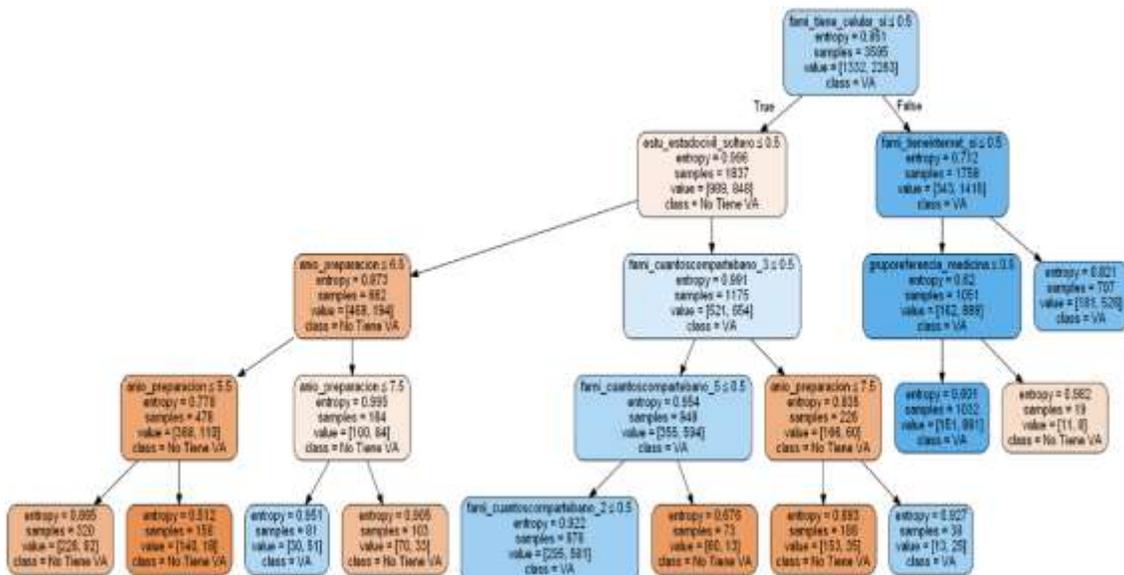


Fig. 68. Árbol 03 – Razonamiento cuantitativo

Fuente: esta investigación.

La primera condición si tiene celular y tiene internet 526 estudiantes generan valor agregado, si no tienen internet el grupo de referencia no es medicina 881 estudiantes generan valor agregado.

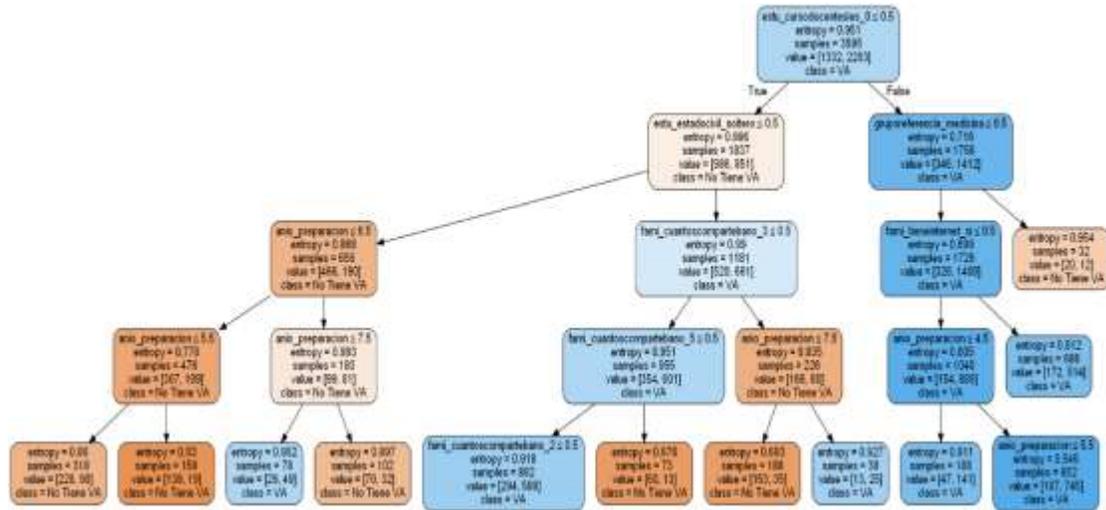


Fig. 69. Árbol 04 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición no se preparó para el examen Saber Pro en su IES con docentes de la institución menos de 20 horas, el grupo de referencia no es medicina y tienen internet 514 estudiantes generan valor agregado, si no tienen internet y los años de preparación son menores o iguales a 4.5 años 141 estudiantes generan valor agregado

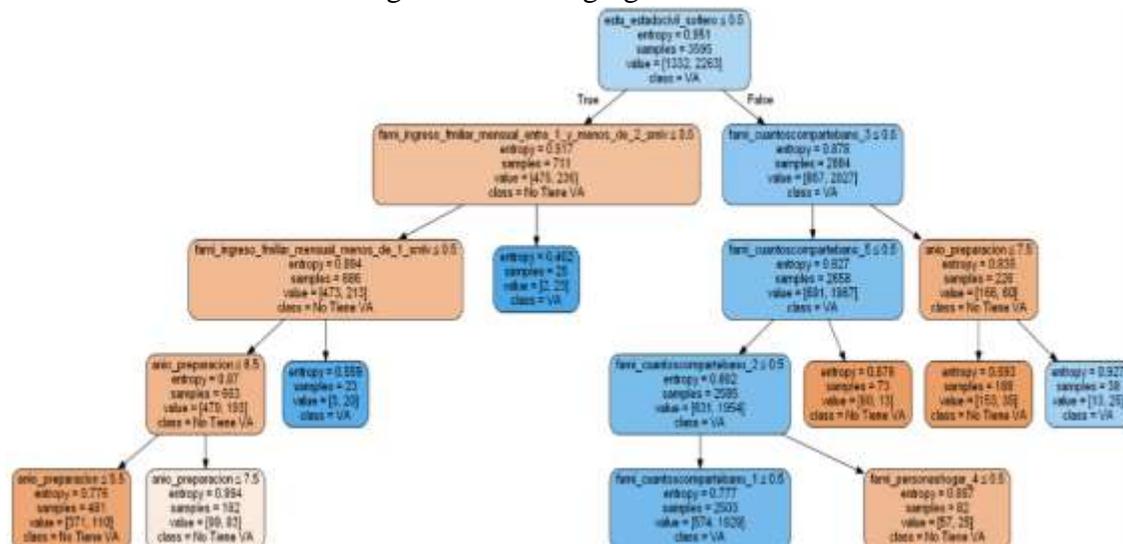


Fig. 70. Árbol 05 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición si es soltero, comparte baño con 3 personas y el año de preparación es mayor a 7.5 años 25 estudiantes generan valor agregado. Si no es soltero y el ingreso familiar mensual entre 1 y menos de 2 SMLV 23 estudiantes generan valor agregado, Si el ingreso mensual es menos de 1 SMLV 20 estudiantes generan valor agregado.

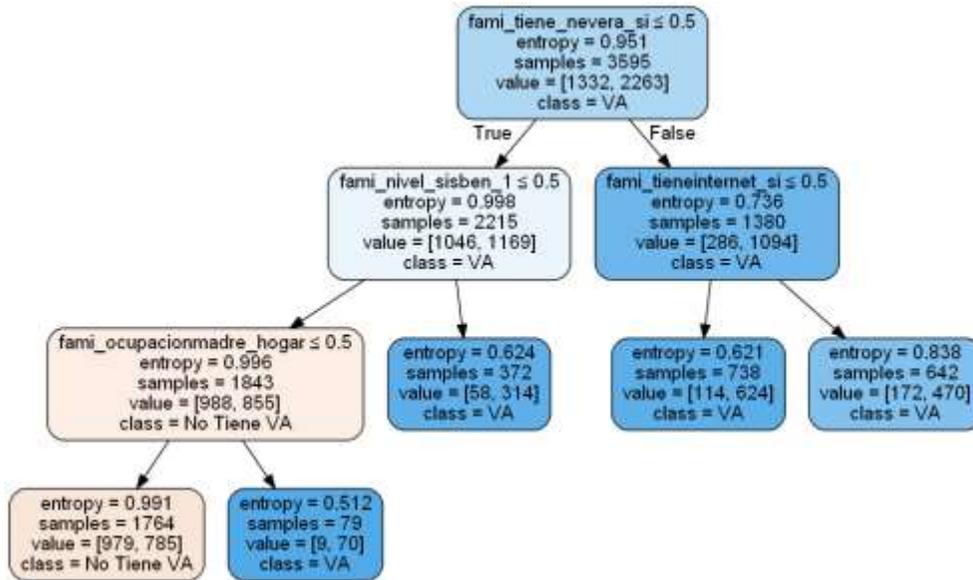


Fig. 71. Árbol 06 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición si tiene nevera y tiene internet 470 estudiantes generan valor agregado, si no tienen internet 624 estudiantes generan valor agregado.

Si no tiene nevera y el nivel de Sisbén es 1 generan valor agregado 314 estudiantes. sí es nivel de Sisbén no es 1 y la ocupación de la madre es hogar 70 estudiantes generan valor agregado.

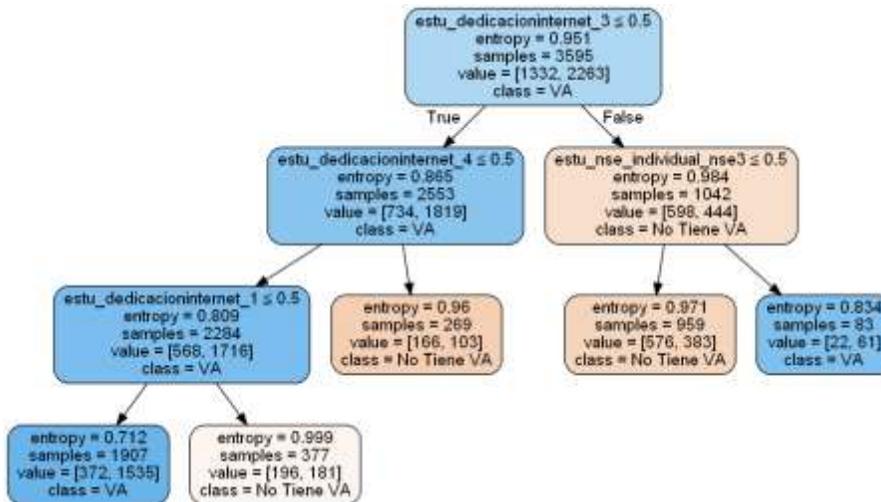


Fig. 72. Árbol 07 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición si se dedica 3 horas a internet y el nivel socioeconómico a nivel de estudiantes es 3 tienen valor agregado 61 estudiantes. si la dedicación de internet es menor a 3, 4, 1 horas, 1535 estudiantes generan valor agregado.

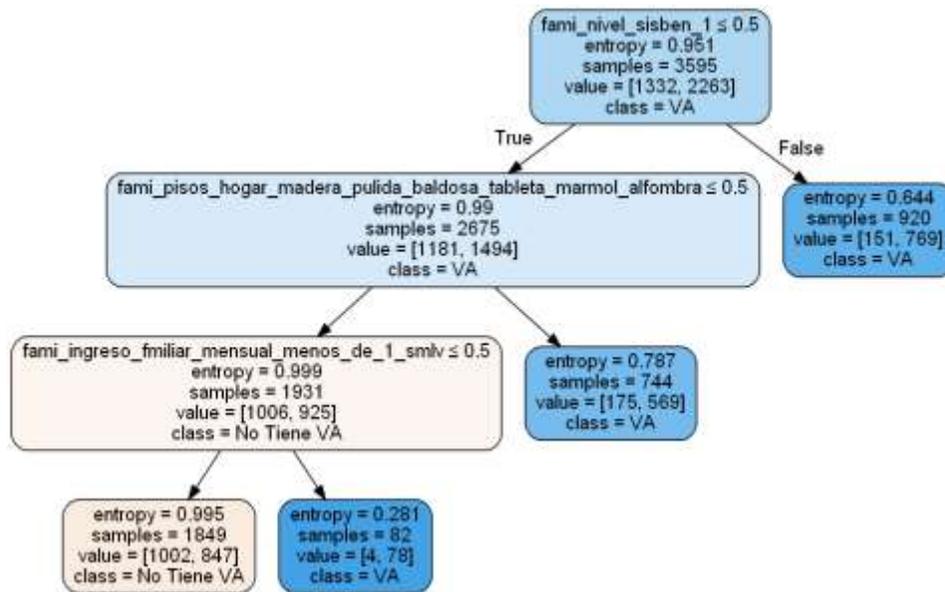


Fig. 73. Árbol 08 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición si el nivel de Sisbén es 1 769 estudiantes generan valor agregado. Si el nivel de Sisbén no es 1 y los pisos del hogar son madera pulida, baldosa, tableta, mármol, alfombra 569 estudiantes generan valor agregado.

Si los pisos del hogar no son madera pulida, baldosa, tableta, mármol y el ingreso familiar mensual es menos de 1 SMLV 78 estudiantes generan valor agregado.

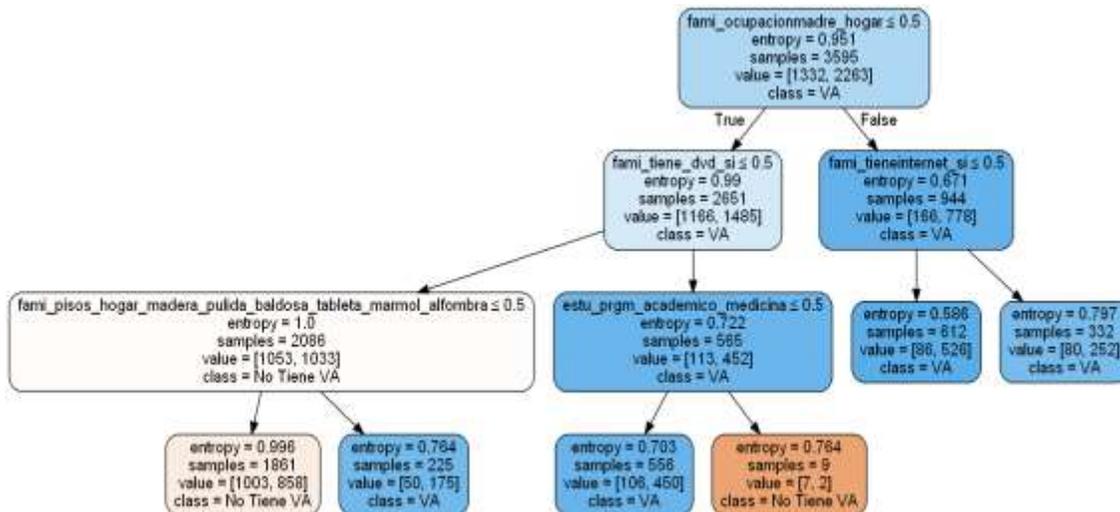


Fig. 74. Árbol 09 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición si la ocupación de la madre es hogar y tiene internet 252 estudiantes generan valor agregado. Si no tienen internet 526 estudiantes generan valor agregado.

Si la ocupación de la madre no es hogar, tiene DVD y el programa académico no es medicina 450 estudiantes generan valor agregado.

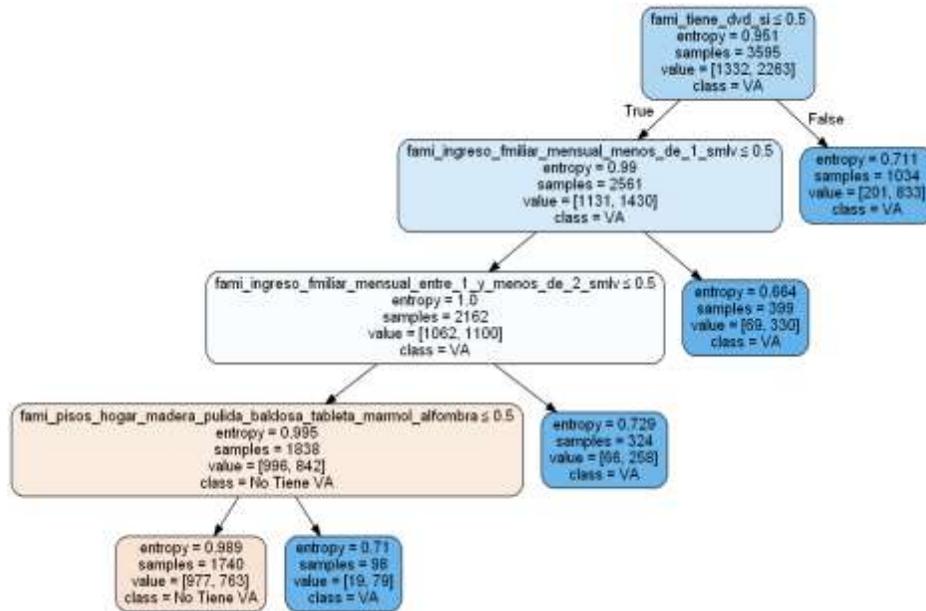


Fig. 75. Árbol 10 – Razonamiento cuantitativo
Fuente: esta investigación.

La primera condición si tiene DVD 833 estudiantes generan valor agregado. Si no tienen DVD y el ingreso familiar mensual es menor de 1 SMLV 330 estudiantes generan valor agregado.

Si el ingreso familiar mensual es entre de 1 y menos de 2 SMLV 258 estudiantes generan valor agregado.

E. EVALUACIÓN

MÉTRICAS DE EVALUACIÓN

Para la prueba de rendimiento de los datos obtenidos, se obtuvo las métricas y balanceos correspondientes, las cuales se muestran a continuación:

COMPETENCIAS CIUDADANAS

TABLA X
MATRIZ DE CONFUSIÓN ÁRBOL 01 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	416	244	660
1	256	626	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 72 %, acertando que 626 estudiantes generan valor agregado frente a 244 estudiantes que se toman como desaciertos en este modelo.

```
Modelo 4
accuracy test: 0.6763942931258107
accuracy train: 0.6739916550764952
Balanced_accuracy test: 0.670021645021645
Balanced_accuracy train: 0.6674322589143777
precision recall f1-score support
0 0.62 0.63 0.62 660
1 0.72 0.71 0.72 882
```

Fig. 76. Métricas árbol 01 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 67% de los estudiantes generan valor agregado y un 33% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XI
MATRIZ DE CONFUSIÓN ÁRBOL 02 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	416	244	660
1	265	617	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 72 %, acertando que 617 estudiantes generan valor agregado frente a 244 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6712062256809338
accuracy train: 0.666759388038943
Balanced_accuracy test: 0.6656771799628942
Balanced_accuracy train: 0.6619140477624692
      precision    recall  f1-score   support

     0       0.61      0.63      0.62       660
     1       0.72      0.70      0.71       882

```

Fig. 77. Métricas árbol 02 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 67% de los estudiantes generan valor agregado y un 33% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XII
MATRIZ DE CONFUSIÓN ÁRBOL 03 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	416	244	660
1	265	617	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 72 %, acertando que 617 estudiantes generan valor agregado frente a 244 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6712062256809338
accuracy train: 0.666759388038943
Balanced_accuracy test: 0.6656771799628942
Balanced_accuracy train: 0.6619140477624692
      precision    recall  f1-score   support

     0       0.61      0.63      0.62       660
     1       0.72      0.70      0.71       882

```

Fig. 78. Métricas árbol 03 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 67% de los estudiantes generan valor agregado y un 33% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XIII
MATRIZ DE CONFUSIÓN ÁRBOL 04 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	416	244	660
1	265	617	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 72 %, acertando que 617 estudiantes generan valor agregado frente a 244 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6699092088197146
accuracy train: 0.6662030598052852
Balanced_accuracy test: 0.6645433931148217
Balanced_accuracy train: 0.6617505061317721
      precision  recall  f1-score  support
0         0.61     0.63     0.62     660
1         0.72     0.70     0.71     882
    
```

Fig. 79. Métricas árbol 04 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 66% de los estudiantes generan valor agregado y un 34% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XIV
MATRIZ DE CONFUSIÓN ÁRBOL 05 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	449	211	660
1	362	520	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 71 %, acertando que 520 estudiantes generan valor agregado frente a 211 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6251621271076524
accuracy train: 0.6328233657858137
Balanced_accuracy test: 0.631720263863121
Balanced_accuracy train: 0.6404251890560435
      precision  recall  f1-score  support
0         0.55     0.68     0.61     660
1         0.71     0.59     0.64     882
    
```

Fig. 80. Métricas árbol 05 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 62% de los estudiantes generan valor agregado y un 38% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XV
MATRIZ DE CONFUSIÓN ÁRBOL 06 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	356	304	660
1	229	653	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 68 %, acertando que 653 estudiantes generan valor agregado frente a 304 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6517509727626459
accuracy train: 0.6603616133518776
Balanced_accuracy test: 0.6372294372294373
Balanced_accuracy train: 0.6423015215286649
precision recall f1-score support
0 0.61 0.54 0.57 660
1 0.68 0.74 0.71 882
    
```

Fig. 81. Métricas árbol 06 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 65% de los estudiantes generan valor agregado y un 35% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XVI
MATRIZ DE CONFUSIÓN ÁRBOL 07 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	343	317	660
1	229	653	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 67 %, acertando que 653 estudiantes generan valor agregado frente a 317 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6459143968871596
accuracy train: 0.6545201668984701
Balanced_accuracy test: 0.6296485260770975
Balanced_accuracy train: 0.6351877003904536
precision recall f1-score support
0 0.60 0.52 0.56 660
1 0.67 0.74 0.71 882
    
```

Fig. 82. Métricas árbol 07 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 64% de los estudiantes generan valor agregado y un 36% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XVII
MATRIZ DE CONFUSIÓN ÁRBOL 08 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	455	205	660
1	388	494	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 71 %, acertando que 494 estudiantes generan valor agregado frente a 205 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6147859922178989
accuracy train: 0.6228094575799722
Balanced_accuracy test: 0.6237940630797774
Balanced_accuracy train: 0.630594306768929
precision    recall  f1-score   support

0           0.54    0.69    0.60     660
1           0.71    0.56    0.62     882
  
```

Fig. 83. Métricas árbol 08 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 61% de los estudiantes generan valor agregado y un 39% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XVIII
MATRIZ DE CONFUSIÓN ÁRBOL 09 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	455	205	660
1	406	476	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 70 %, acertando que 476 estudiantes generan valor agregado frente a 205 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6044098573281452
accuracy train: 0.6141863699582754
Balanced_accuracy test: 0.6156771799628943
Balanced_accuracy train: 0.6254381892568336
precision    recall  f1-score   support

0           0.53    0.69    0.60     660
1           0.70    0.54    0.61     882
  
```

Fig. 84. Métricas árbol 09 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 60% de los estudiantes generan valor agregado y un 40% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XIX
MATRIZ DE CONFUSIÓN ÁRBOL 10 - COMPETENCIAS CIUDADANAS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	422	238	660
1	326	556	882
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 70 %, acertando que 556 estudiantes generan valor agregado frente a 238 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6381322957198443
accuracy train: 0.643115438108484
Balanced_accuracy test: 0.6386724386724387
Balanced_accuracy train: 0.6423713824207614
precision    recall  f1-score   support
0           0.57    0.64    0.60     660
1           0.70    0.63    0.67     882

```

Fig. 85. Métricas árbol 10 - Competencias ciudadanas
Fuente: esta investigación.

Obteniendo así que un 63% de los estudiantes generan valor agregado y un 37% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

INGLÉS

TABLA XX
MATRIZ DE CONFUSIÓN ÁRBOL 01 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	6	287	293
1	13	1236	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 81 %, acertando que 1236 estudiantes generan valor agregado frente a 287 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8099870298313878
accuracy train: 0.8141863699582754
Balanced_accuracy test: 0.5091431506980328
Balanced_accuracy train: 0.5155887397840783
precision    recall  f1-score   support

0           0.50     0.02     0.05     293
1           0.81     0.99     0.89    1249

```

Fig. 86. Métricas árbol 01 – inglés

Fuente: esta investigación.

Obteniendo así que un 80% de los estudiantes generan valor agregado y un 20% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXI
MATRIZ DE CONFUSIÓN ÁRBOL 02 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	6	287	293
1	13	1236	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 81 %, acertando que 1236 estudiantes generan valor agregado frente a 287 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8099870298313878
accuracy train: 0.8141863699582754
Balanced_accuracy test: 0.5091431506980328
Balanced_accuracy train: 0.5155887397840783
precision    recall  f1-score   support

0           0.50     0.02     0.05     293
1           0.81     0.99     0.89    1249

```

Fig. 87. Métricas árbol 02 – inglés

Fuente: esta investigación.

Obteniendo así que un 80% de los estudiantes generan valor agregado y un 20% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXII
MATRIZ DE CONFUSIÓN ÁRBOL 03 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	88	205	293
1	75	1174	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 85 %, acertando que 1174 estudiantes generan valor agregado frente a 205 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8190661478599222
accuracy train: 0.8189151599443671
Balanced_accuracy test: 0.6205469495049964
Balanced_accuracy train: 0.6013785188174755
precision    recall  f1-score   support
0           0.54    0.30    0.39     293
1           0.85    0.94    0.89    1249

```

Fig. 88. Métricas árbol 03 – inglés
Fuente: esta investigación.

Obteniendo así que un 81% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXIII
MATRIZ DE CONFUSIÓN ÁRBOL 04 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	105	188	293
1	88	1161	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 86 %, acertando que 1161 estudiantes generan valor agregado frente a 188 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8184176394293126
accuracy train: 0.8228094575799721
Balanced_accuracy test: 0.642351423801157
Balanced_accuracy train: 0.631404500050449
precision    recall  f1-score   support
0           0.53    0.36    0.43     293
1           0.86    0.93    0.89    1249

```

Fig. 89. Métricas árbol 04 – inglés
Fuente: esta investigación.

Obteniendo así que un 81% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXIV
MATRIZ DE CONFUSIÓN ÁRBOL 05 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	85	208	293
1	75	1174	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 85 %, acertando que 1174 estudiantes generan valor agregado frente a 208 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8119325551232166
accuracy train: 0.811961057023644
Balanced_accuracy test: 0.6109187691450089
Balanced_accuracy train: 0.5903251437796387
precision recall f1-score support
0 0.51 0.29 0.37 293
1 0.85 0.94 0.89 1249

```

Fig. 90. Métricas árbol 05 – inglés
Fuente: esta investigación.

Obteniendo así que un 81% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXV
MATRIZ DE CONFUSIÓN ÁRBOL 06 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	73	220	293
1	63	1186	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 84 %, acertando que 1186 estudiantes generan valor agregado frente a 220 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8190661478599222
accuracy train: 0.8125173852573018
Balanced_accuracy test: 0.5996483193380643
Balanced_accuracy train: 0.5703725658359399
precision recall f1-score support
0 0.55 0.25 0.34 293
1 0.84 0.95 0.90 1249

```

Fig. 91. Métricas árbol 06 – inglés
Fuente: esta investigación.

Obteniendo así que un 81% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXVI
MATRIZ DE CONFUSIÓN ÁRBOL 07 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	293	0	293
1	0	1249	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 81 %, acertando que 1249 estudiantes generan valor agregado

```

Modelo 4
accuracy test: 0.8099870298313878
accuracy train: 0.8108484005563282
Balanced_accuracy test: 0.5
Balanced_accuracy train: 0.5
precision recall f1-score support
0 0.00 0.00 0.00 293
1 0.81 1.00 0.90 1249

```

Fig. 92. Métricas árbol 07 – inglés
Fuente: esta investigación.

Obteniendo así que un 80% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXVII
MATRIZ DE CONFUSIÓN ÁRBOL 08 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	23	270	293
1	25	1224	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 82 %, acertando que 1224 estudiantes generan valor agregado frente a 270 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.8086900129701686
accuracy train: 0.811961057023644
Balanced_accuracy test: 0.5305473047379883
Balanced_accuracy train: 0.5424049036424176
precision recall f1-score support
0 0.48 0.08 0.14 293
1 0.82 0.98 0.89 1249

```

Fig. 93. Métricas árbol 08 – inglés
Fuente: esta investigación.

Obteniendo así que un 80% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXVIII
MATRIZ DE CONFUSIÓN ÁRBOL 09 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	293	0	293
1	0	1249	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 82 %, acertando que 1249 estudiantes generan valor agregado

```

Modelo 4
accuracy test: 0.8099870298313878
accuracy train: 0.8108484005563282
Balanced_accuracy test: 0.5
Balanced_accuracy train: 0.5
precision recall f1-score support
0 0.00 0.00 0.00 293
1 0.81 1.00 0.90 1249

```

Fig. 94. Métricas árbol 09 – inglés
Fuente: esta investigación.

Obteniendo así que un 80% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXIX
MATRIZ DE CONFUSIÓN ÁRBOL 10 - INGLÉS

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	293	0	293
1	0	1249	1249
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 81 %, acertando que 1249 estudiantes generan valor agregado.

```

Modelo 4
accuracy test: 0.8099870298313878
accuracy train: 0.8108484005563282
Balanced_accuracy test: 0.5
Balanced_accuracy train: 0.5
precision recall f1-score support
0 0.00 0.00 0.00 293
1 0.81 1.00 0.90 1249

```

Fig. 95. Métricas árbol 10 – inglés
Fuente: esta investigación.

Obteniendo así que un 80% de los estudiantes generan valor agregado y un 19% no genera valor agregado, este modelo no presenta sobreajuste y se encuentra equilibrado.

LECTURA CRITICA

TABLA XXX
MATRIZ DE CONFUSIÓN ÁRBOL 01 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	977	74	1051
1	383	108	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 61%, acertando que 108 estudiantes generan valor agregado frente a 74 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.7068741893644618
accuracy train: 0.7123783031988874
Balanced_accuracy test: 0.577153753287045
Balanced_accuracy train: 0.5763654448999277
precision  recall  f1-score  support
0          0.72    0.93    0.81    1051
1          0.61    0.22    0.32    491
    
```

Fig. 96. Métricas árbol 01 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 70% de los estudiantes generan valor agregado y un 30% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXI
MATRIZ DE CONFUSIÓN ÁRBOL 02 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	998	53	1051
1	447	44	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 48%, acertando que 44 estudiantes generan valor agregado frente a 53 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6789883268482491
accuracy train: 0.7004172461752434
Balanced_accuracy test: 0.523056307541455
Balanced_accuracy train: 0.5504199464328775
precision recall f1-score support
0 0.69 0.95 0.80 1051
1 0.48 0.09 0.16 491

```

Fig. 97. Métricas árbol 02 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 67% de los estudiantes generan valor agregado y un 33% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXII
MATRIZ DE CONFUSIÓN ÁRBOL 03 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	1030	21	1051
1	447	44	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 63%, acertando que 44 estudiantes generan valor agregado frente a 21 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6932555123216602
accuracy train: 0.6998609179415856
Balanced_accuracy test: 0.5324373450946728
Balanced_accuracy train: 0.5339900760159381
precision recall f1-score support
0 0.70 0.98 0.81 1051
1 0.63 0.09 0.16 491

```

Fig. 98. Métricas árbol 03 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 69% de los estudiantes generan valor agregado y un 31% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXIII
MATRIZ DE CONFUSIÓN ÁRBOL 04 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	1019	32	1051
1	447	44	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 63%, acertando que 44 estudiantes generan valor agregado frente a 32 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6932555123216602
accuracy train: 0.6987482614742698
Balanced_accuracy test: 0.5329799376406138
Balanced_accuracy train: 0.535570033414861
precision    recall  f1-score   support

0           0.70     0.97     0.81     1051
1           0.62     0.09     0.16     491
    
```

Fig. 99. Métricas árbol 04 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 69% de los estudiantes generan valor agregado y un 31% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXIV
MATRIZ DE CONFUSIÓN ÁRBOL 05 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	1009	42	1051
1	442	49	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 56%, acertando que 49 estudiantes generan valor agregado frente a 42 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6880674448767834
accuracy train: 0.7012517385257302
Balanced_accuracy test: 0.5302592235888234
Balanced_accuracy train: 0.5397879778267709
precision    recall  f1-score   support

0           0.70     0.96     0.81     1051
1           0.56     0.10     0.16     491
    
```

Fig. 100. Métricas árbol 05 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXV
MATRIZ DE CONFUSIÓN ÁRBOL 06 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	977	74	1051
1	432	59	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 45%, acertando que 59 estudiantes generan valor agregado frente a 74 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6738002594033723
accuracy train: 0.7087621696801113
Balanced_accuracy test: 0.5252189264031346
Balanced_accuracy train: 0.5636825459023735
      precision  recall  f1-score  support
0         0.69     0.93     0.80     1051
1         0.45     0.12     0.18     491
  
```

Fig. 101. Métricas árbol 06 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 67% de los estudiantes generan valor agregado y un 33% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXVI
MATRIZ DE CONFUSIÓN ÁRBOL 07 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	1030	21	1051
1	466	25	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 57%, acertando que 25 estudiantes generan valor agregado frente a 21 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.685473411154345
accuracy train: 0.6995827538247567
Balanced_accuracy test: 0.5169618305522236
Balanced_accuracy train: 0.5306780145142214
      precision  recall  f1-score  support
0         0.69     0.98     0.81     1051
1         0.57     0.05     0.10     491
  
```

Fig. 102. Métricas árbol 07 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXVII
MATRIZ DE CONFUSIÓN ÁRBOL 08 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	980	71	1051
1	442	49	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 50%, acertando que 49 estudiantes generan valor agregado frente a 71 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6815823605706874
accuracy train: 0.7004172461752434
Balanced_accuracy test: 0.5255018496592325
Balanced_accuracy train: 0.5480283007222663
precision recall f1-score support
0 0.69 0.96 0.80 1051
1 0.50 0.10 0.16 491
  
```

Fig. 103. Métricas árbol 08 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXVIII
MATRIZ DE CONFUSIÓN ÁRBOL 09 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	1040	11	1051
1	481	10	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 52%, acertando que 10 estudiantes generan valor agregado frente a 11 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.682230869001297
accuracy train: 0.6951321279554937
Balanced_accuracy test: 0.5069868479442525
Balanced_accuracy train: 0.51929966585139
precision recall f1-score support
0 0.68 0.99 0.81 1051
1 0.52 0.02 0.05 491
  
```

Fig. 104. Métricas árbol 09 - Lectura critica
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XXXIX
MATRIZ DE CONFUSIÓN ÁRBOL 10 - LECTURA CRITICA

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	1030	21	1051
1	471	20	491
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 50%, acertando que 20 estudiantes generan valor agregado frente a 21 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6815823605706874
accuracy train: 0.7001390820584145
Balanced_accuracy test: 0.5103092583728812
Balanced_accuracy train: 0.5303663649568822
precision recall f1-score support
0 0.69 0.98 0.81 1051
1 0.50 0.04 0.07 491
    
```

Fig. 105. Métricas árbol 10 - Lectura crítica
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

RAZONAMIENTO CUANTITATIVO

TABLA XL
MATRIZ DE CONFUSIÓN ÁRBOL 01 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	305	260	565
1	117	860	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 77%, acertando que 860 estudiantes generan valor agregado frente a 260 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.7594033722438391
accuracy train: 0.772461752433936
Balanced_accuracy test: 0.7138513238104727
Balanced_accuracy train: 0.7286162101120122
precision recall f1-score support
0 0.73 0.54 0.62 565
1 0.77 0.88 0.82 977

```

Fig. 106. Métricas árbol 01 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 75% de los estudiantes generan valor agregado y un 25% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLI
MATRIZ DE CONFUSIÓN ÁRBOL 02 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	333	232	565
1	107	870	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 79%, acertando que 870 estudiantes generan valor agregado frente a 232 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.7782101167315175
accuracy train: 0.7824756606397775
Balanced_accuracy test: 0.7376491154971423
Balanced_accuracy train: 0.7421297236255257
precision recall f1-score support
0 0.75 0.59 0.66 565
1 0.79 0.89 0.84 977

```

Fig. 107. Métricas árbol 02 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 77% de los estudiantes generan valor agregado y un 23% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLII
MATRIZ DE CONFUSIÓN ÁRBOL 03 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	333	232	565
1	107	870	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 79%, acertando que 870 estudiantes generan valor agregado frente a 232 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.7788586251621271
accuracy train: 0.7821974965229486
Balanced_accuracy test: 0.7385340712493547
Balanced_accuracy train: 0.7422176374341642
precision recall f1-score support
0 0.75 0.59 0.66 565
1 0.79 0.89 0.84 977
    
```

Fig. 108. Métricas árbol 03 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 77% de los estudiantes generan valor agregado y un 23% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLIII
MATRIZ DE CONFUSIÓN ÁRBOL 04 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	339	226	565
1	117	860	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 79%, acertando que 860 estudiantes generan valor agregado frente a 226 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.7788586251621271
accuracy train: 0.7830319888734353
Balanced_accuracy test: 0.7407731814023424
Balanced_accuracy train: 0.7438070527443041
precision recall f1-score support
0 0.75 0.60 0.66 565
1 0.79 0.88 0.84 977
    
```

Fig. 109. Métricas árbol 04 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 77% de los estudiantes generan valor agregado y un 23% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLIV
MATRIZ DE CONFUSIÓN ÁRBOL 05 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	333	232	565
1	107	870	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 79%, acertando que 870 estudiantes generan valor agregado frente a 232 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.7782101167315175
accuracy train: 0.7835883171070932
Balanced_accuracy test: 0.7380223005226403
Balanced_accuracy train: 0.7423957873029902
precision recall f1-score support
0 0.75 0.59 0.66 565
1 0.79 0.89 0.84 977
    
```

Fig. 110. Métricas árbol 05 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 77% de los estudiantes generan valor agregado y un 23% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLV
MATRIZ DE CONFUSIÓN ÁRBOL 06 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	429	136	565
1	352	625	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 82%, acertando que 625 estudiantes generan valor agregado frente a 136 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6802853437094682
accuracy train: 0.6834492350486787
Balanced_accuracy test: 0.6961974982110669
Balanced_accuracy train: 0.6940501593064563
precision recall f1-score support
0 0.55 0.76 0.63 565
1 0.82 0.64 0.72 977
    
```

Fig. 111. Métricas árbol 06 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLVI
MATRIZ DE CONFUSIÓN ÁRBOL 07 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	429	136	565
1	352	625	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 82%, acertando que 625 estudiantes generan valor agregado frente a 136 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6802853437094682
accuracy train: 0.6834492350486787
Balanced_accuracy test: 0.6961974982110669
Balanced_accuracy train: 0.6940501593064563
precision recall f1-score support
0 0.55 0.76 0.63 565
1 0.82 0.64 0.72 977
    
```

Fig. 112. Métricas árbol 07 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLVII
MATRIZ DE CONFUSIÓN ÁRBOL 08 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	429	136	565
1	352	625	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 82%, acertando que 625 estudiantes generan valor agregado frente a 136 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6802853437094682
accuracy train: 0.6834492350486787
Balanced_accuracy test: 0.6961974982110669
Balanced_accuracy train: 0.6940501593064563
precision recall f1-score support
0 0.55 0.76 0.63 565
1 0.82 0.64 0.72 977
    
```

Fig. 113. Métricas árbol 08 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 68% de los estudiantes generan valor agregado y un 32% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLVIII
MATRIZ DE CONFUSIÓN ÁRBOL 09 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	429	136	565
1	391	586	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 81%, acertando que 586 estudiantes generan valor agregado frente a 136 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6562905317769131
accuracy train: 0.6712100139082059
Balanced_accuracy test: 0.6783815363991268
Balanced_accuracy train: 0.6891158723902868
      precision  recall  f1-score  support
0          0.52    0.76    0.62     565
1          0.81    0.60    0.69     977
  
```

Fig. 114. Métricas árbol 09 - Razonamiento cuantitativo
Fuente: esta investigación.

Obteniendo así que un 65% de los estudiantes generan valor agregado y un 35% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

TABLA XLIX
MATRIZ DE CONFUSIÓN ÁRBOL 10 - RAZONAMIENTO CUANTITATIVO

	Verdaderos Negativos	Falsos Positivos	
Clase	0	1	Soporte
0	424	141	565
1	332	645	977
	Falsos Negativos	Verdaderos Positivos	1542

Este modelo para la clase 1 obtuvo una precisión del 82%, acertando que 645 estudiantes generan valor agregado frente a 141 estudiantes que se toman como desaciertos en este modelo.

```

Modelo 4
accuracy test: 0.6906614785992218
accuracy train: 0.6890125173852573
Balanced_accuracy test: 0.703266274762004
Balanced_accuracy train: 0.6981602127978619
      precision    recall  f1-score   support

0         0.56      0.75      0.64       565
1         0.82      0.66      0.73       977

```

Fig. 115. Métricas árbol 10 - Razonamiento cuantitativo

Fuente: esta investigación.

Obteniendo así que un 69% de los estudiantes generan valor agregado y un 31% no genera valor agregado y este modelo no presenta sobreajuste y se encuentra equilibrado.

F. DESPLIEGUE

PATRONES INTERESANTES DE VALOR AGREGADO.

COMPETENCIAS CIUDADANAS

P1: Si no tiene celular entonces genera valor agregado.

P2: Si tiene celular y tomó curso de apoyo externo entonces genera valor agregado.

P3: Si tiene celular y no tomó curso de apoyo externo y tiene estado civil soltero y los años de preparación están entre 5.5 y 6.5 años entonces genera valor agregado.

P4: Si no tomo curso de apoyo externo genera valor agregado.

P5: Si tomó curso de apoyo externo y no tiene nevera, genera valor agregado.

P6: Si tomó curso de apoyo externo y tiene nevera y tiene estado civil soltero/a y el año de preparación esta entre 5.5 años y 6.5 años genera valor agregado.

P6: Si no tomó curso de preparación externo y no tiene servicio de televisión entonces genera valor agregado.

P7: Si tomó curso de preparación externo y no tiene nevera entonces genera valor agregado.

P8: Si tomó curso de preparación externo, tiene nevera, tiene estado civil soltero/a y el año de preparación esta entre 5.5 años y 6.5 años entonces genera valor agregado

P9: Si no tomó curso de preparación con docentes de la institución entonces genera valor agregado

P10: Si tomó curso de preparación con docentes de la institución y no tiene nevera entonces genera valor agregado

P11: Si tomó curso de preparación con docentes de la institución y tiene nevera y el estado civil es soltero/a y el año de preparación esta entre 5.5 y 6.5 años entonces genera valor agregado.

P12: Si no tiene servicio de televisión entonces genera valor agregado

P13: Si tiene servicio de televisión y la madre tiene una ocupación diferente a la del hogar entonces genera valor agregado

P14: Si tiene servicio de televisión y la madre se ocupa del hogar y no tiene nevera en el hogar genera valor agregado

P15: Si no tiene nevera entonces genera valor agregado

P16: Si tiene nevera y la madre no se ocupa del hogar entonces genera valor agregado

P17: Si tiene nevera y la madre se ocupa del hogar y la casa no cuenta con un tipo de piso conocido entonces generan valor agregado.

P18: Si tiene nevera y la madre se ocupa del hogar y la casa cuenta con un tipo de piso conocido y el estado civil es soltero/a y el año de preparación esta entre 5.5 y 6.5 años entonces genera valor agregado.

P19: Si la casa no cuenta con un tipo de piso entonces genera valor agregado

P20: Si la casa cuenta con un tipo de piso y tiene un ingreso mensual menor a un salario mínimo entonces generan valor agregado.

P21: Si la casa cuenta con un tipo de piso y tiene un ingreso mensual entre 1 y menos de 2 salarios mínimos entonces generan valor agregado

P22: Si la casa cuenta con un tipo de piso y tiene un ingreso mayor a 2 salarios mínimos y tiene estado civil soltero y el año de preparación esta entre 5.5 y 6.5 años entonces genera valor agregado.

P23: Si no cuenta con dispositivo reproductor de DVD entonces genera valor agregado

P24: Si cuenta con dispositivo reproductor de DVD y la madre tiene otra ocupación diferente a la del hogar entonces genera valor agregado

P25: Si cuenta con dispositivo reproductor de DVD y la ocupación de la madre es el hogar y tiene nivel de Sisbén 1 entonces genera valor agregado.

P26: Si la madre no tiene la ocupación del hogar entonces genera valor agregado

P27: Si la ocupación de la madre es el hogar y no tienen servicio de teléfono entonces genera valor agregado

P28: Si la ocupación de la madre es el hogar y tienen servicio de teléfono y el nivel de Sisbén es 1 entonces genera valor agregado.

P29: Si se dedicó entre 1 y menos de 4 horas diarias a internet entonces genera valor agregado

Para competencias ciudadanas los estudiantes que generan valor agregado a la Universidad de Nariño son de estratos bajos, se debería apoyar e incentivar estos grupos de individuos y hacer una investigación a los grupos de estudiantes con niveles socioeconómicos altos.

INGLÉS

P1: Si no tomo curso de apoyo con una institución externa y el año de preparación es mayor a 4.5 años entonces genera valor agregado

P2: Si tomo curso de apoyo con una institución externa y tiene estado civil soltero y el año de preparación es menor a 9.5 años entonces genera valor agregado

P3: Si tomo curso de apoyo con una institución externa y no tiene estado civil soltero y comparte baño con 3 o 4 personas entonces genera valor agregado

P4: Si tomo curso de apoyo con una institución externa y no tiene estado civil soltero y no comparte baño con 3 o 4 personas y comparte baño con 2 o 3 personas entonces genera valor agregado.

P5: Si tomo curso de apoyo con una institución externa y no tiene estado civil soltero y no comparte baño con 3 o 4 personas y no comparte baño con 2 o 3 personas y comparte baño con 5 o más personas entonces genera valor agregado.

P6: Si no tomo curso de apoyo externo con docentes de la institución y el año de preparación es mayor a 4.5 años entonces genera valor agregado

P7: Si tomo curso de apoyo con una institución externa y tiene estado civil soltero y el año de preparación es menor a 9.5 años entonces genera valor agregado

P8: Si tomo curso de apoyo externo con docentes de la institución y no tiene estado civil soltero y comparte baño con 3 o 4 personas entonces genera valor agregado

P9: Si tomo curso de apoyo externo con docentes de la institución y no tiene estado civil soltero y no comparte baño con 3 o 4 personas y comparte baño con 2 o 3 personas entonces genera valor agregado.

P10: Si tomo curso de apoyo externo con docentes de la institución y no tiene estado civil soltero y no comparte baño con 3 o 4 personas y no comparte baño con 2 o 3 personas y comparte baño con 5 o más personas entonces genera valor agregado.

P11: Si no tomó curso de apoyo entonces genera valor agregado

P12: Si tomó curso y tiene estado civil diferente a soltero y comparte baño con menos de 3 personas entonces genera valor agregado.

P13: Si tomó curso de apoyo y tiene estado civil soltero/a y el año de preparación es menor o igual a 9.5 años entonces genera valor agregado

P14: Si tomó curso y tiene estado civil diferente a soltero/a y comparte baño con 3 personas y cursa entre noveno y décimo semestre genera valor agregado.

P15: Si tiene un nivel socioeconómico diferente a 2 y está soltero/a entonces genera valor agregado

P16: Si tiene un nivel socioeconómico 2 entonces genera valor agregado

P17: Si tiene un nivel socioeconómico entre 1 y 3 y el año de preparación es mayor a 4.5 años entonces genera valor agregado

P18: Si tiene un nivel socioeconómico entre 2 y 3 y es soltero/a entonces genera valor agregado.

P19: Si el año de preparación es menor o igual a 4.5 años entonces genera valor agregado

P20: Si el año de preparación es mayor a 4.5 años y menor a 9.5 años y está soltero entonces genera valor agregado

P21: Si el año de preparación es menor o igual a 4.5 años y menor a 9.5 años y no tiene estado civil soltero y si no tiene celular entonces genera valor agregado.

P22: Si el año de preparación es menor o igual a 4.5 años y menor a 9.5 años y no tiene estado civil soltero y tiene celular y está en cursando noveno semestre entonces genera valor agregado

P23: Si no tiene celular entonces genera valor agregado

P24: Si tiene celular y tiene estado civil soltero entonces genera valor agregado

P25: Si tiene celular y no tiene estado civil soltero y no comparten baño con 3 o 4 personas entonces genera valor agregado

P26: Si tiene celular y no tiene estado civil soltero y comparte baño con 3 o 4 personas y cursa entre noveno y décimo semestre entonces genera valor agregado

P27: Si no tiene nevera entonces genera valor agregado

P28: Si tiene nevera y tiene estado civil soltero entonces genera valor agregado

P29: Si tiene nevera y tiene estado civil diferente a soltero y no comparte baño con 3 o 4 personas entonces genera valor agregado

P30: Si tiene nevera y tiene estado civil diferente a soltero y comparte baño con 3 o 4 personas y tiene nivel de Sisbén 1 entonces genera valor agregado

P31: Si no tiene nivel de Sisbén 1 entonces genera valor agregado

P32: Si tiene nivel de Sisbén 1 y tiene estado civil soltero entonces genera valor agregado.

P33: Si tiene nivel de Sisbén 1 y no tiene estado civil soltero y no cursa decimo semestre y en la vivienda no cuenta con un tipo de piso comercial entonces genera valor agregado.

P34: Si tiene nivel de Sisbén 1 y no tiene estado civil soltero y está en décimo semestre entonces genera valor agregado.

P35: Si tiene nivel de Sisbén 1 y no tiene estado civil soltero y está en noveno semestre y dedican entre 1 y 3 horas a internet entonces generan valor agregado.

P36: Si se dedica menos de una hora a internet entonces genera valor agregado

P37: Si no se dedica menos de una hora a internet y tiene estado civil soltero entonces genera valor agregado

P38: Si no se dedica menos de una hora a internet y no tiene estado civil soltero y se dedica entre 1 y 3 horas a internet entonces genera valor agregado

P39: Si no se dedica menos de una hora a internet y no tiene estado civil soltero y no se dedica entre 1 y 3 horas a internet y comparte baño con menos de 3 personas entonces genera valor agregado

P40: Si tiene entre 0 y 10 libras entonces genera valor agregado

P41: Si tiene más de 10 libras y tiene estado civil soltero entonces genera valor agregado

P42: Si tiene más de 10 libras y no tiene estado civil soltero y comparte baño con 3 o 4 personas entonces genera valor agregado.

En la clase ingles se evidencia que el uso de la tecnología son factores que influyen en la generación de patrones de valor agregado, se deberían tomar en cuenta y crear proyectos que ayuden y fortalezcan estos grupos de individuos para que tengan acceso a estas herramientas.

LECTURA CRITICA

P1: Tiene celular y el año de preparación es menor o igual a 3.5 años entonces genera valor agregado.

P2: Tiene celular, el año de preparación es mayor a 3.5, no tomo curso para prueba con docentes de la institución y tiene horno entonces genera valor agregado.

P3: No tiene celular, es soltero, cursa semestre 9, el año de preparación es menor o igual a 6.5 años, la educación del padre no es técnica o tecnológica incompleta entonces genera valor agregado.

P4: No tomo curso de preparación externo y año de preparación menor o igual a 3.5 años genera valor agregado.

P5: Tomo curso de preparación externo y ocupación de la madre es el hogar entonces genera valor agregado.

P6: No tomo curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos y año de preparación menor o igual a 3.5 años entonces genera valor agregado.

P7: Tomo curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos y tiene nevera entonces genera valor agregado.

P8: No se preparó para el examen con docentes de la institución y año de preparación menor o igual a 3.5 años entonces genera valor agregado.

P9: Se preparó para el examen con docentes de la institución y tiene nevera entonces genera valor agregado.

P10: Tiene nevera y año de preparación menor o igual a 3.5 años entonces genera valor agregado.

P11: Tiene nevera, año de preparación mayor a 3.5 años, educación del padre bachillerato incompleto y programa académico diseño gráfico y multimedia entonces genera valor agregado.

P12: No tiene dedicación de lectura diaria de 60, 30, 1 hora y año de preparación es menor o igual a 3.5 años genera valor agregado.

P13: Tiene dedicación de lectura diaria de 60 horas, año de preparación es menor o igual a 6.5 años, el nivel socioeconómico de estudiantes no es 3 y el número de personas a cargo es cero tiene valor agregado.

P14: Si no comparten 3 personas el baño, el nivel de Sisbén es 1 y año de preparación es menor o igual a 2.5 años generan valor agregado

P15: Si no comparten 3 personas el baño, el nivel de Sisbén no es 1, el piso del hogar es madera pulida, baldosa, tableta, mármol, alfombra y el programa académico es tecnología en promoción de la salud entonces genera valor agregado.

P16: Nivel de Sisbén es 1 y programa académico técnico profesional en producción de palma de aceite entonces genera valor agregado.

P17: Nivel de Sisbén es 1 y grupo de referencia bellas artes y diseño entonces genera valor agregado.

P18: Nivel de Sisbén es 1 y programa académico es licenciatura en educación básica con énfasis en ciencias naturales y educación ambiental entonces genera valor agregado.

P19: Nivel de Sisbén no es 1, pisos del hogar pisos en el hogar madera pulida, baldosa tableta, mármol, alfombra y el grupo de referencia es tecnológico en salud entonces genera valor agregado.

La generación de patrones de valor agregado para la clase lectura crítica tiene un déficit; en la investigación se refleja que los estudiantes de la Universidad de Nariño generan valor agregado en un mínimo porcentaje, por cual se deberían realizar proyectos de investigación para encontrar la problemática que afecta estos resultados y generar soluciones para dichos problemas.

RAZONAMIENTO CUANTITATIVO

P1: Toma curso para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos de menos de 20 horas de intensidad entonces genera valor agregado.

P2: Toma curso para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos de menos de 20 horas de intensidad, tiene internet, programa académico de medicina entonces genera valor agregado.

P3: Toma curso para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos de menos de 20 horas de intensidad, tiene internet, programa académico de ingeniería electrónica entonces genera valor agregado.

P4: Toma curso para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos de menos de 20 horas de intensidad y grupo de referencia medicina entonces genera valor agregado.

P5: No toma curso para el examen Saber Pro en un curso organizado por la institución con apoyo de un instituto de preparación de exámenes externos de menos de 20 horas de intensidad, estado civil soltero, comparten baño 3 personas y año de preparación mayor a 7.5 años entonces generan valor agregado.

P6: Tiene celular, tiene internet entonces genera valor agregado.

P7: Estado civil no soltero e ingreso familiar mensual entre 1 y menos de 2 SMLV entonces genera valor agregado.

P8: Estado civil no soltero e ingreso familiar mensual menos de 1 SMLV entonces genera valor agregado.

P9: Tiene nevera, tiene internet entonces genera valor agregado.

P10: No tiene nevera y nivel de Sisbén 1 entonces genera valor agregado.

P11: No tiene nevera, nivel de Sisbén diferente a 1 y ocupación de madre es hogar entonces genera valor agregado.

P12: Dedicación a internet de 3 horas y nivel socioeconómico de estudiantes es de 3 entonces genera valor agregado.

P13: Nivel de Sisbén es 1 entonces genera valor agregado.

P14: Nivel de Sisbén diferente a 1 y pisos del hogar son madera pulida, baldosa, tableta, mármol, alfombra entonces genera valor agregado.

P15: Nivel de Sisbén diferente a 1 y pisos del hogar diferente a madera pulida, baldosa, tableta, mármol, alfombra e ingreso familiar mensual menos de 1 SMLV entonces genera valor agregado.

P16: Ocupación de madre hogar y tiene internet entonces genera valor agregado

P17: Ocupación de madre hogar y no tiene internet entonces genera valor agregado

P18: Ocupación de madre diferente a hogar, tiene DVD, programa académico diferente a medicina entonces genera valor agregado.

P19: Tiene DVD entonces genera valor agregado

P20: No tiene DVD y el ingreso familiar mensual menos de 1 SMLV entonces genera valor agregado.

En la clase razonamiento cuantitativo tomar un curso de preparación para las pruebas SABER PRO es una de las variables que influye en la obtención de valor agregado para la Universidad de Nariño, por lo cual se deberían crear alternativas de estudio ya sean gratis o económicamente accesibles para así aportar conocimientos y preparación para presentar dichas pruebas.

V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

Se analizaron los resultados construidos con el conjunto de datos limpio de los estudiantes de la Universidad de Nariño que presentaron las pruebas Saber Pro en los años 2012 – 2018. Los árboles de decisión mostraron que los cuatro componentes principales de esta investigación, Competencias ciudadanas, Inglés, Lectura Crítica y Razonamiento Cuantitativo; están ligadas a las diferentes variables socioeconómicas tales como: el estrato social, el tiempo de preparación para presentar dichas pruebas, ocupación de los padres, las herramientas tecnológicas etc.

Con este estudio se logró identificar que los estudiantes generan valor agregado, son de estratos bajos, viven en condiciones normales, tienen ingresos de menos de 2 SMLV, logran obtener resultados que destacan la calidad de la educación en la institución.

En el desarrollo de la investigación se evidenció que la lectura tiene un papel fundamental en el resultado de la prueba Saber Pro. Los estudiantes que dedican mayor tiempo diario a la lectura obtienen mejores resultados en la prueba, por lo que se puede concluir que la lectura diaria influye en el puntaje.

Por otro lado, se debería buscar alternativas para preparar a los estudiantes para presentar este examen, debido a que los resultados de los patrones muestran que a medida que estos mismos toman cursos de preparación ya sea con instituciones o con la misma institución logran obtener resultados destacados.

CONCLUSIONES

Durante la investigación se logró afianzar los conocimientos, en cuanto a los procesos de selección, transformación y limpieza de datos logrando así aplicar estos mismos, para la construcción del repositorio de datos que fue la base para aplicar técnicas de Big Data con el objetivo de generar árboles de decisión los cuales arrojaron patrones. Con lo anterior se encontraron modelos y parámetros que facilitaron el trabajo de los equipos de cómputo para el procesamiento de los datos.

Con la construcción, limpieza y transformación del repositorio de datos de las pruebas Saber Pro; se logró estructurar usando diferentes métodos de transformación de variables para eliminar datos incorrectos que interfirieran en la búsqueda de patrones interesantes de valor agregado. El desarrollo de este proceso que abarca los aspectos relacionados a la preparación de datos, quedó demostrado que en la selección del subconjunto de datos y la selección de las variables socioeconómicas para aplicar técnicas de Big data fue la correcta.

Con la generación de algoritmos para la construcción de reglas asociativas (a priori) y árboles de decisión, se encontró que los parámetros correctos generaban patrones de valor agregado, mediante el análisis descendente de los árboles de decisión, es decir, desde el nodo principal hasta los nodos hoja, teniendo en cuenta las clases, para competencias ciudadanas se obtuvo una media de precisión del 70%, para Inglés se obtuvo una media de precisión del 82%, para lectura crítica una media de precisión del 52% y para razonamiento cuantitativo una media de precisión del 80%.

RECOMENDACIONES

Para próximas investigaciones, dando continuidad a este proyecto, se recomienda hacer profundización en el estudio de la competencia de lectura crítica, estructurando la información a partir del año 2017 en adelante, para aplicar técnicas de minería de datos y obtener resultados que permitan definir acciones encaminadas a la mejora de los resultados obtenidos por parte de los estudiantes.

Se recomienda continuar el estudio con bases de datos SABER 11 como SABER PRO a partir del año 2019 en adelante, para que se realicen comparaciones con los resultados de este proyecto y se haga un análisis en base a las variables socioeconómicas y si estas varían en cuanto a la generación de valor agregado en la Universidad de Nariño.

REFERENCIAS

- [1] ICFES, «Medición de los efectos de la educación superior en Colombia sobre el aprendizaje estudiantil.,» ICFES, Bogota D.C, Agosto, 2014.
- [2] Grupo de Trabajo sobre Estandares y Evaluación, «Usaid.gov,» 01 04 2011. [En línea]. Available: https://pdf.usaid.gov/pdf_docs/PBAAB628.pdf. [Último acceso: 20 Agosto 2022].
- [3] E. Muñoz, «Repositorio Universidad Autonoma de Bucaramanga,» 26 10 2019. [En línea]. Available: <https://repository.unab.edu.co/handle/20.500.12749/12316>. [Último acceso: 05 09 2022].
- [4] F. d. I. d. Sistemas, «Universidad de Nariño,» 01 01 2019. [En línea]. Available: <http://facultades.udenar.edu.co/wp-content/uploads/2019/05/Anexo-2-Guia-elaboracion-trabajos-de-grado-pasantias.pdf>. [Último acceso: 20 02 2021].
- [5] Netec, «MINERÍA DE DATOS: Qué es, importancia y técnicas de su implementación,» Netec, 27 Mayo 2019. [En línea]. Available: <https://www.netec.com/post/mineria-de-datos-que-es-importancia-y-tecnicas-de-su-implementacion>. [Último acceso: 19 12 2022].
- [6] M. Guerrero y J. Patiño, «Formulación de indicadores para determinar el valor agregado en la formación de estudiantes del programa de Ingeniería de Sistemas sede Pasto de la Universidad de Nariño,» Pasto, 2022.
- [7] S. Robertson y L. Camilo, «Software para el cálculo del valor agregado aportado a los estudiantes por las universidades a partir de las pruebas saber pro y saber 11 para los años 2016 y 2017.,» ANP, Bogotá, 2019.
- [8] R. Cacho, «Desarrollo de una herramienta de análisis, procesado y visualización de fuentes de datos empleando técnicas de Big Data utilizando Apache Spark. Trabajo de grado (Ingeniería de tecnologías y servicios de telecomunicación).,» ANP, Madrid, 2017.
- [9] D. Martines y A. Torres, «Construcción de un repositorio limpio de datos de los estudiantes de los programas de la Universidad de Nariño para análisis de deserción con minería de datos,» Pasto, 2016.
- [10] B. F. M. Antonio, R. S. K. Julieth, B. G. Y. Estivel y O. B. Edinson, «Calculo de valor agregado generado por la Universidad de Nariño en las pruebas saber 11 – saber pro 2010-2014,» *VIDA UNIVERSITARIA*, vol. XX, n° 1, pp. 203-226, 2019.
- [11] T. G. Group, «Ministerio de Educación,» 15 02 2015. [En línea]. Available: https://www.mineduccion.gov.co/1759/articles-324587_archivo_pdf_4_Gestion_Conocimiento_MEN.pdf. [Último acceso: 10 11 2020].
- [12] J. Molina y J. García, «Universidad Carlos III de Madrid,» 01 01 2006. [En línea]. Available: http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf. [Último acceso: 05 02 2021].

- [13 R. Lopez, «Iebschool,» Comunidad, 13 05 2015. [En línea]. Available:
] <https://comunidad.iebschool.com/bigdata/2015/05/13/que-es-la-mineria-de-datos/>. [Último
acceso: 06 02 2021].
- [14 G. Vega, «Repensando el concepto de “Valor Agregado a la Información”,» Hypotheses, 21
] 08 2019. [En línea]. Available: <https://bdcv.hypotheses.org/2672>. [Último acceso: 17 02
2021].
- [15 ICFES, «Ministerio de Educación Nacional,» 15 06 2017. [En línea]. Available:
] <https://www.icfes.gov.co/valoragregado>. [Último acceso: 18 02 2021].
- [16 M. Zaforas, «Spark: un destello en el universo Big Data.,» 03 03 2016. [En línea]. Available:
] <https://www.paradigmadigital.com/dev/spark-un-destello-en-el-universo-big-data/>. [Último
acceso: 19 02 2021].
- [17 S. Hernandez, «Spark,» Datademia, 30 06 2022. [En línea]. Available:
] <https://datademia.es/blog/que-es-apache-spark>. [Último acceso: 19 12 2022].
- [18 M. Febrer, «DBeaver: software para la gestión de bases de datos,» 20 10 2022. [En línea].
] Available: <https://geoinnova.org/blog-territorio/dbeaver-gestion-de-bases-de-datos/>. [Último
acceso: 12 01 2023].
- [19 D. L. Delgado y R. P. Navarro, «Ciencia de datos y estudios globales: aportaciones y desafíos
] metodológicos.,» *Articulo de Scielo.org*, vol. 1, n° 102, pp. 41-62, 2020.
- [20 I. Corporation, «IBM,» IBM SPSS Modeler, 17 08 2020. [En línea]. Available:
] <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>. [Último
acceso: 21 02 2021].
- [21 IBM, «Minería de Datos,» IBM, 30 07 2021. [En línea]. Available:
] <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>. [Último
acceso: 10 11 2021].
- [22 Desconocido, «<http://mineriadatos1.blogspot.com/2013/06/tecnicas-mineria-de-datos.html>,»
] Desconocido, 10 06 2013. [En línea]. Available:
<http://mineriadatos1.blogspot.com/2013/06/tecnicas-mineria-de-datos.html>. [Último acceso:
21 02 2021].
- [23 R. Amat, «Árboles de decisión, random forest, gradient boosting y C5.0,» 05 11 2020. [En
] línea]. Available:
[https://www.cienciadedatos.net/documentos/33_arboles_decision_random_forest_gradient_
boosting_C50.html](https://www.cienciadedatos.net/documentos/33_arboles_decision_random_forest_gradient_boosting_C50.html). [Último acceso: 23 02 2021].
- [24 ServName, «Apriori Algorithm Data Mining,» 2021. [En línea]. Available:
] [https://spa.myservername.com/apriori-algorithm-data-
mining#Apriori_Algorithm_8211_Frequent_Pattern_Algorithms](https://spa.myservername.com/apriori-algorithm-data-mining#Apriori_Algorithm_8211_Frequent_Pattern_Algorithms). [Último acceso: 11 01
2023].

- [25 D. A. R. Sanchez, «Repositorio UniAndes,» 2007. [En línea]. Available:
] <https://repositorio.uniandes.edu.co/bitstream/handle/1992/25528/u336448.pdf?sequence=1>.
- [26 Aprende IA, «AprendeIA,» 30 06 2022. [En línea]. Available: <https://aprendeia.com/reglas-de-asociacion/>. [Último acceso: 11 01 2023].
- [27 R. Amat, «Reglas de Asociación y Algoritmo Apriori,» 01 06 2018. [En línea]. Available:
] https://www.cienciadedatos.net/documentos/43_reglas_de_asociacion. [Último acceso: 11 01 2023].
- [28 BIGDATA, «Sitio BigData,» 19 01 2021. [En línea]. Available:
] <https://sitiobigdata.com/2019/01/19/machine-learning-metrica-clasificacion-parte-3/>.
[Último acceso: 11 01 2023].
- [29 ICFES, «Pruebas Saber 11,» ICFES, 10 10 2022. [En línea]. Available:
] <https://www.icfes.gov.co/acerca-del-examen-saber-11%C2%B0>. [Último acceso: 12 01 2023].
- [30 ICFES, «Pruebas Saber Pro,» ICFES, 10 10 2022. [En línea]. Available:
] <https://www.icfes.gov.co/acerca-del-examen-saber-pro>. [Último acceso: 12 01 2023].