

MODELOS DE PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL DURANTE  
LOS TRES PRIMEROS SEMESTRES DEL PROGRAMA DE INGENIERÍA DE  
SISTEMAS MEDIANTE TÉCNICAS DE MINERÍA DE DATOS UTILIZANDO UN  
SITIO WEB.

LUIS GABRIEL CARLOSAMA DAZA  
ESTEBAN ANDRES URBINA BRAVO

UNIVERSIDAD CESMAG  
FACULTAD DE INGENIERIA PROGRAMA DE  
INGENIERIA DE SISTEMAS  
JUNIO DEL 2022

MODELOS DE PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL DURANTE  
LOS TRES PRIMEROS SEMESTRES DEL PROGRAMA DE INGENIERÍA DE  
SISTEMAS MEDIANTE TÉCNICAS DE MINERÍA DE DATOS UTILIZANDO UN  
SITIO WEB.

LUIS GABRIEL CARLOSAMA DAZA  
ESTEBAN ANDRES URBINA BRAVO

Asesor: MG CARLOS FERNANDO GONZÁLEZ GUZMÁN

Trabajo de grado para optar el título de ingeniero de sistemas

UNIVERSIDAD CESMAG  
FACULTAD DE INGENIERIA PROGRAMA DE  
INGENIERIA DE SISTEMAS  
JUNIO DEL 2022

**NOTA DE ACEPTACIÓN**

**NOTA DE ACEPTACIÓN**

---

---

---

---

---

---

---

---

**Mg. JURADO 1**

---

**Mg. JURADO 2**

## NOTA DE EXCLUSIÓN

“El pensamiento que se expresa en esta obra es exclusiva responsabilidad del autor y no compromete la ideología de la Universidad CESMAG”.

## **DEDICATORIA**

Primero que todo a Dios quien me ayudo en tiempos difíciles de mi carrera y fue de su mano que seguí a delante, pese a todo obstáculo que la vida me coloco para poder culminar mis estudios universitarios, toda la gloria y honra a Dios por darme la capacidad de afrontar el camino que el escogió para mí y gracias a él es que termino con satisfacción mi carrera de ingeniería de sistemas.

A mis padres, Luis Eduardo Carlosama quien es un guerreo de la vida y me ha enseñado a afrontar la vida de la mejor manera, igualmente a mi madre María del rosario daza, quien siempre ha estado pendiente de mí y de mi carrera, jamás me dejo solo, de una u otra manera siempre me brindo su ayuda y sobre todo por su amor incondicional.

A mis hermanos Christian Carlosama y Diana Carlosama quienes siempre me motivaron a seguir luchando por mi objetivo, quienes con su fe hacia Dios me enseñaron que todo lo que uno se proponga lo puede lograr.

A mis tías, Ana Daza, quien me enseñó a ser sencillo ante todas las cosas de la vida. A Rosa Daza quien es un eje fundamental en el inicio y culminación de mi carrera, quien hizo hasta lo imposible para seguir apoyándome en todo este proceso.

Al resto de mi familia quienes de una u otra manera me brindaron su ayuda para culminar mis estudios universitarios.

Y como no a mi hijo Mathias Carlosama, quien es la fuente de inspiración de mi carrera, por él y con el hasta que Dios me lo permita estar con mi hijo ya que en todo momento me ha brindado su gran amor.

A mi persona, ya que me creo capaz de realizar todos los sueños que me proponga y hare todo lo necesario para dar lo mejor de sí mismo, sin importar el obstáculo que la vida me coloque, todo de la mano de Dios.

**Luis Gabriel Carlosama Daza**

## DEDICATORIA

Quiero expresar mi profundo agradecimiento y reconocimiento a las personas que me han apoyado en los momentos más difíciles de mi formación.

En primer lugar, quiero dedicar este logro a mi padre, Fabián Urbina †, quien, desde el más allá, me ha brindado su apoyo incondicional y me ha inculcado valores irreprochables que han forjado mi carácter. Gracias, padre mío, por ser una guía en mi vida.

A mi madre, Mónica Bravo, le debo un agradecimiento especial. Ha sido fundamental en este proceso, siendo mi inspiración constante y brindándome su apoyo incondicional en la consecución de mis metas. Este triunfo es dedicado a estas personas que amo profundamente y que me han dado la vida y una formación basada en valores inquebrantables.

A mi hermano, Fabián, le agradezco por estar a mi lado y motivarme en los momentos difíciles, enseñándome que los triunfos se logran con perseverancia y dedicación.

También quiero dedicar este logro a mis dos sobrinas, Juliana y Helen. Ellas son mi fuente de inspiración y espero ser un modelo a seguir para ellas.

Por último, quiero expresar mi gratitud a mis familiares: abuelos, tías, tíos y primos. Siempre han creído en mí y han sido mi mayor apoyo para alcanzar mis objetivos.

A todos ellos, les agradezco de todo corazón por estar a mi lado y ser parte fundamental en mi camino hacia el éxito

“Nunca consideres el estudio como una obligación, si no como una oportunidad De penetrar en el bello y maravilloso Mundo del saber”

Albert Einstein

**Esteban Andres Urbina Bravo**

## **AGRADECIMIENTOS**

En primer lugar, queremos expresar nuestro profundo agradecimiento a Dios por iluminar nuestro camino de aprendizaje y brindarnos la fortaleza para completar nuestro proyecto de formación como ingenieros de sistemas.

En segundo lugar, extendemos nuestro sincero agradecimiento al profesor y asesor de investigación, Mg Carlos Fernando Gonzáles Guzmán, por su generosa contribución y orientación invaluable en el desarrollo de este proyecto investigativo.

Además, deseamos expresar nuestra gratitud a todos los docentes de la Universidad CESMAG, quienes nos han brindado un valioso apoyo y han dejado una huella significativa en nuestra formación académica. Apreciamos su calidad humana y su dedicación.

Finalmente, queremos agradecer a la Universidad CESMAG y a su distinguido cuerpo directivo por brindarnos la oportunidad de crecer y desarrollarnos profesionalmente bajo una filosofía humanizadora. Estamos agradecidos por su respaldo y confianza en nuestro proceso de formación.

## RESUMEN ANALÍTICO DE ESTUDIO R.A.E

<b>FACULTAD</b>	INGENIERÍA
<b>PROGRAMA</b>	INGENIERÍA DE SISTEMAS
<b>FECHA DE ELABORACIÓN DEL RAE</b>	ABRIL DE 2023
<b>AUTORES DE LA INVESTIGACIÓN</b>	LUIS GABRIEL CARLOSAMA DAZA  ESTEBAN ANDRÉS URBINA BRAVO
<b>DIRECTOR DE LA INVESTIGACIÓN</b>	MG CARLOS FERNANDO GONZÁLES GUZMÁN
<b>TÍTULO DE LA INVESTIGACIÓN</b>	MODELOS DE PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL DURANTE LOS TRES PRIMEROS SEMESTRES DEL PROGRAMA DE INGENIERÍA DE SISTEMAS MEDIANTE TÉCNICAS DE MINERÍA DE DATOS UTILIZANDO UN SITIO WEB

### **PALABRAS O FRASES CLAVES**

Deserción estudiantil, Minería de datos, Modelos predictivos, Dashboard, Desarrollo web, Bases de datos

### **DESCRIPCION**

El objetivo de este proyecto de investigación fue analizar la deserción estudiantil en los primeros tres semestres del programa de Ingeniería de Sistemas mediante el uso de técnicas de minería de datos. El propósito principal fue desarrollar un modelo predictivo basado en el rendimiento académico de los estudiantes, con el fin de presentar y visualizar los resultados a través de un tablero de control en una aplicación web

## ABSTRACT

The objective of this research project was to analyze student dropout in the first three semesters of the Systems Engineering program through the use of data mining techniques. The main purpose was to develop a predictive model based on the academic performance of the students, in order to present and visualize the results through a dashboard in a web application.

## CONTENIDO

**Capítulo 1:** El proyecto comienza con la identificación del problema de investigación, donde se exponen las principales dificultades abordadas en este estudio. Además, se definen el objeto de investigación, el área y la línea de investigación correspondientes. Posteriormente, se plantea la formulación del problema y se establecen el objetivo general, los objetivos específicos, la justificación y la delimitación del proyecto.

**Capítulo 2:** En el marco teórico, se realiza una exhaustiva revisión de la literatura pertinente. Se abordan antecedentes a nivel internacional, nacional y regional que se encuentran vinculados a esta investigación. Además, se definen las variables de estudio y se plantean las hipótesis de investigación correspondientes.

**Capítulo 3:** En la sección de metodología, se presenta de manera objetiva y precisa los diferentes aspectos considerados en el proceso de desarrollo de esta investigación. Se describen el paradigma, enfoque, método, tipo de investigación, diseño de investigación, población de estudio, técnicas de recolección de información, validación y confiabilidad de las técnicas utilizadas, así como los instrumentos empleados para recopilar los datos.

**Capítulo 4:** En este capítulo, se aplicó de manera rigurosa la metodología CRISP-DM para guiar el desarrollo de la investigación. Se llevó a cabo un análisis estadístico exhaustivo de todas las variables de estudio con el objetivo de construir un modelo de regresión lineal múltiple. Además, se exploraron diversas alternativas de modelos, incluyendo un modelo de clasificación, para evaluar su pertinencia y aplicabilidad. Como parte del proceso, se implementó la metodología Scrum en todas las fases de desarrollo del proyecto, asegurando una gestión eficiente y ágil. Se presenta en detalle el proceso de evaluación y análisis, se

destacan los resultados obtenidos y se describe minuciosamente el diseño y desarrollo de la aplicación web correspondiente

**Capítulo 5:** En este capítulo, se llevó a cabo el análisis de los resultados de la investigación mediante la realización de una post-prueba a un grupo de estudiantes de la Universidad CESMAG. Además, se procedió a realizar simulaciones en la aplicación web con el objetivo de analizar los resultados obtenidos antes, durante y después de aplicar los modelos predictivos. Por último, se destacan las gráficas generadas a partir de una encuesta de satisfacción del aplicativo, proporcionando una visión más detallada de la percepción de los usuarios.

**Recomendaciones:** Aquí se presentan las sugerencias sobre el proyecto investigativo

## **METODOLOGIA**

El proyecto investigativo se fundamentó en el enfoque del positivismo, utilizando una metodología cuantitativa y siguiendo el método científico de manera rigurosa y objetiva. Durante la fase de desarrollo del proyecto, se optó por emplear la metodología scrum siendo muy utilizada en proyectos de corto plazo, reconocida por su enfoque flexible y colaborativo en la planificación y ejecución de procesos iterativos e incrementales en el desarrollo del software. Esta elección permitió una adaptabilidad efectiva a los cambios y una mayor eficiencia en la entrega de resultados.

## **AREA DE INVESTIGACIÓN**

Deserción estudiantil

## **LÍNEA DE INVESTIGACION**

Minería de datos: Es el estudio de técnicas para extraer conocimiento y patrones de grandes bases de datos. Es utilizada en diferentes ámbitos para obtener información valiosa y tomar decisiones estratégicas

## **CONCLUSIONES**

La investigación concluye con resultados positivos obtenidos durante el desarrollo de la herramienta web, al aplicar cada uno de los procesos inherentes a la investigación. Mediante la realización de un cuestionario dirigido a un grupo

específico de estudiantes, se validaron diversos aspectos relevantes de los modelos de predicción. De esta manera, logramos cumplir con el objetivo principal de la investigación proporcionar una herramienta web que brinde altos niveles de predicción al analizar la deserción estudiantil en el programa de ingeniería de sistemas de la Universidad CESMAG.

## FUENTES

CESMAG, I.U. Áreas y líneas de investigación programa de Ingeniería de Sistemas. Colombia,2015.

QUIJANO, Armando. Guía de investigación cuantitativa. Pasto, Colombia. Primera edición,2009.

ICONTEC, Instituto Colombiano de Normas Técnicas y Certificación. Norma técnica NTC-ISO/IEC colombiana 27001. Sistema de Gestión de la Seguridad de la Información. Disponible en: <http://intranet.bogotaturismo.gov.co/sites/intranet.bogotaturismo.gov.co/files/file/Norma.%20NTC-ISO-IEC%2027001.pdf>

TIMARÁN, R. y JIMÉNEZ, J. Extracción de perfiles de deserción estudiantil en la Institución Universitaria CESMAG, Colombia. IRE: Ciencias Sociales y Humanas, VI (1), 30-44. 2015. doi:

TIMARÁN-PEREIRA, R., CAICEDO-ZAMBRANO, J. y HIDALGO-TROYA, A. Árboles de decisiones para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas saber 11°, 2019, Rev.investig. desarro. innova.

TREVIÑO-REYES, Rosalba; RIVERA-RODRÍGUEZ, F.; GARZA-ALONSO, J. La analítica de datos como ventaja competitiva en las organizaciones. VINCULATEGICA EFAN, 2020, vol. 6, no 2 Ortiz, M. D., Joyanes, L., y Giraldo, L.M. Los desafíos del marketing en la era del big data. E Ciencias de la Información, 6(1), 2016,16-45

## CONTENIDO

	Pág.
INTRODUCCIÓN	19
1. PROBLEMA DE INVESTIGACIÓN	21
1.1 OBJETO O TEMA DE ESTUDIO	21
1.2 LÍNEA DE INVESTIGACIÓN	21
1.3 PLANTEAMIENTO DEL PROBLEMA	21
1.4 formulación del problema	23
1.5 Objetivo de la investigación	23
1.5.1 Objetivo general	23
1.5.2 Objetivos específicos	23
1.6 JUSTIFICACIÓN	24
1.7 DELIMITACIÓN	25
2. MARCO TEÓRICO	26
2.1 ANTECEDENTES	26
2.1.1 Internacionales	26
2.1.2 Nacionales	31
2.1.3 Regionales	36
2.2 Supuesto teóricos de la investigación	39
2.2.1 Ámbito educativo	39
2.2.2 Deserción estudiantil	40
2.2.3 Documentos institucionales y de programa	40
2.2.4 Ingeniería de software	42
2.2.5 Metodología Scrum	42
2.2.6 Bases de datos	44
2.2.7 Web	45
2.2.7 Framework	45
2.2.8 Lenguaje de programación	45
2.2.9 Servidor web	47
2.2.10 Minería de datos	48
2.2.11 Metodología CRISP-DM	49

2.2.12 Herramientas para aplicación de minería de datos	51
2.2.14 Estadística.	52
2.2.15 Spss	52
2.2.17 Regresión lineal múltiple	55
2.3 Variables de estudio	56
2.3.1 variable dependiente	56
2.3.2 variables independientes	56
2.4 Definición nominal de las variables	56
2.5 Definición operativa de las variables	58
2.6 Formulación de hipótesis	58
2.6.1 hipótesis de investigación	58
2.6.2 hipótesis nula	59
2.6.3 hipótesis alterna	59
3.METODOLOGIA DE LA INVESTIGACION	60
3.1 Paradigma	60
3.2 Enfoqué	60
3.3 Método	61
3.4 Tipo de investigación	61
3.5 Diseño investigación	62
3.6. Población	62
3.7 Muestra	62
3.8 Técnicas de recolección de información	63
3.9 Validez de las técnicas de recolección de la información	63
3.10 Confiabilidad de las técnicas de recolección	64
3.11 Instrumentos de recolección de información	64
4. RESULTADOS DE INVESTIGACIÓN.	66
4.1 ANALIZAR LA INFORMACIÓN DE LOS REGISTROS DE NOTAS	66
4.2 PROCESOS DE MINERÍA DE DATOS Y METODOLOGÍA CRISP-DM	67
4.3 DESARROLLO DEL SISTEMA DE INFORMACION	80
4.3.1 Metodología SCRUM	80

4.3.2 Fases metodología scrum	80
4.3.2.1 Levantamiento de requerimientos	80
4.3.2.2 Atributos de calidad	84
4.3.2.3 Escenarios de calidad priorizados	85
4.3.2.4 Product Backlog	88
4.3.2.5 Historias de usuario	89
4.3.2.6 Sprint plannig	91
4.3.2.7 Workitem	91
4.4 VALIDACION DEL APLICATIVO CON LA MÉTRICA DE CALIDAD ISO 25010	98
5. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	100
5.1 CARACTERIZACIÓN DE LA POBLACIÓN	100
5.2 APLICACIÓN DE LA PREPRUEBA	101
5.3 APLICACIÓN DE LA POSTPRUEBA	103
5.5 RESULTADOS DE ENCUESTA DE SATISFACION	108
5.6 DISCUSION DE RESULTADOS	112
5.6.1 Resultados de la investigación referentes al marco teórico	113
CONCLUSIONES	117
RECOMENDACIONES	118
BIBLIOGRAFIA	119

## Lista de tablas

	Pág.
Tabla 1. Área de formación profesional	57
Tabla 2 Métricas de calidad ISO 25010	58
Tabla 3 Grado de correlación	69
Tabla 4 correlación de Pearson ciencias básicas	69
Tabla 5 correlación de Pearson comunicación oral y escrita	70
Tabla 6: Resumen de modelos ciencias básicas	72
Tabla 7: Resumen de modelos comunicación oral y escrita	73
Tabla 8: Análisis de varianza ciencias básicas	74
Tabla 9: análisis de varianza comunicación oral y escrita	75
Tabla 10: Coeficientes ciencias básicas	76
Tabla 11: coeficientes comunicación oral y escrita	77
Tabla 12 Matriz de confusión	79
Tabla 13 Tabla roles Scrum	80
Tabla14. RF1-Registro	81
Tabla15. RF2-Insertar datos	81
Tabla16. RF3-Visualizacion	82
Tabla17. RF4-Visualizacion de gráficos	82
Tabla 18. RNF1-Compatibilidad	82
tabla 19. RNF2-Eficiencia	83
Tabla 20. RNF3-Usabilidad	83
Tabla 21. RNF4-Seguridad	83
Tabla 22. Atributo de calidad-Compatibilidad	84
Tabla 23. Atributo de calidad-Eficiencia	84
Tabla 24. Atributo de calidad-Interacción	85
Tabla 25. Atributo de calidad-Seguridad	85
Tabla 26. Escenario de calidad 1	85
Tabla 27. Escenario de calidad 2	86
Tabla 28. Escenario de calidad 3	87

Tabla 29. Escenario de calidad 4	87
Tabla 30 Product Backlog	88
Tabla 31. HU-001 Gestión registro	89
Tabla 32. HU-002 Inicio de sesión	89
Tabla 33. HU-00 Modelos de predicción ciencias básicas	90
Tabla 34. HU-004 Modelos de predicción comunicación oral y escrita	90
Tabla 35. HU-005 Reportes de graficas	90
Tabla 36 Sprint plannig	91
Tabla 37 Mockus – 1	92
Tabla 38. MockUp – 2	93
Tabla 39. MockUp – 3	94
Tabla 40. MockUp – 4	95
Tabla 41. MockUp – 5	96
Tabla 4.2 Lista de chequeo métricas de calidad	99
Tabla 43 Población	100
Tabla 44. Resultados de preprueba	102
Tabla 45. Total, resultados preprueba	103
Tabla 46 Total resultados preprueba	103
Tabla 47 Resultados de post-prueba modelo de regresión lineal	104
Tabla 48 Resultados de post-prueba modelo de clasificación	105
Tabla 49. Total resultados post-prueba	107
Tabla 50 Resultados prueba satisfacción	108

## Lista de figuras

	Pág.
Figura 1 ¿Fue fácil ingresar a la plataforma?	109
Figura 2 ¿Tuvo dificultad al manejar la aplicación?	109
Figura 3 ¿Las funciones de la aplicación fueron claras y fáciles de entender?	110
Figura 4 ¿Recomendaría el software a los demás?	111
Figura 5 ¿Estas satisfecho con el rendimiento de nuestro software?	111
Figura 6 ¿Cree usted que la herramienta cumple con los requerimientos de seguridad?	112

## Tabla de anexos

	Pág.
Anexo 1 manual de usuario	126

## INTRODUCCIÓN

La formación es fundamental para el desarrollo de una nación, por ello es necesario conocer la deserción de estudiantes, que viene siendo un tema complejo y crucial dentro del proceso de aprendizaje<sup>1</sup>. La cual se caracteriza por presentar diferentes causales, y por ello diversas clasificaciones de los factores involucrados, en general, de carácter personal e institucional<sup>2</sup>, y donde se tiene en cuenta el determinante: académico, social y cultural; la intención de abandonar los estudios universitarios tiene una amplia gama de consecuencias que no solo involucran al desertor, también afecta a la familia, docentes y comunidad<sup>3</sup>.

Para la Universidad CESMAG al igual que para otras universidades el problema de la deserción es un tema de mayor relevancia y preocupación, ya que, según el Ministerio de Educación Nacional, el Sistema Nacional de Información de la Educación Superior (SNIES) y Observatorio de la Universidad de Colombia<sup>4</sup>, el 42% de las personas acceden a educación superior, que anualmente serian cerca de 2,2 millones de estudiantes, de los cuales solo el 18% se gradúa. También se señala que cuatro de cada diez personas no tienen recursos para matricularse en una universidad, o no alcanza para clasificar e ingresar a la educación pública; de igual forma se reporta que hoy solo el 49% de los adolescentes, entre 17 y 21 años, tiene la oportunidad de estudiar una carrera profesional.

El objetivo de este trabajo consiste en desarrollar un modelo predictivo para evaluar el rendimiento académico de los estudiantes en los primeros semestres del programa de ingeniería de sistemas en la Universidad CESMAG. Para lograrlo, se

---

<sup>1</sup> ABREU ALVES, Sara, SINVAL, Jorge, NETO, Lía Lucas, MAROCO, João, GONÇALVES FERREIRA, António y OLIVEIRA, Pedro. Burnout e intención de abandono en estudiantes de medicina: el papel protector del compromiso académico. BMC Med Educ. 2022. p.3. En: <https://doi.org/10.1186/s12909-021-03094-9>

<sup>2</sup> FREIXA NIELLA, M., LLANES ORDÓNEZ, J. y VENCESLAO PUEYO, M. El abandono en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona. Revista de Investigación Educativa, 2018. p.185-202. doi: <http://dx.doi.org/10.6018/rie.36.1.278971>

<sup>3</sup> PREET SINGH, Harman y NAFIL ALHULAIL, Hilal. Predicción del riesgo de abandono escolar de los estudiantes-docentes e identificación temprana: un enfoque de regresión logística de cuatro pasos. India. En: IEEE Access. Vol.10, p. 6470-6482, 2022, doi: 10.1109/ACCESS.2022.3141992.

<sup>4</sup> MINISTERIO DE EDUCACIÓN NACIONAL, SISTEMA DE INFORMACIÓN DE EDUCACIÓN SUPERIOR y OBSERVATORIO DE LA UNIVERSIDAD DE COLOMBIA. Colombia. Bogotá. 2020. Disponible en:

[[https://caracol.com.co/programa/2021/11/04/sanamente/1636049871\\_761122.html#:~:text=por%20la%20educaci%C3%B3n%20superior%20en,cifras%20hablan%20por%20s%C3%AD%20solas.P.1.](https://caracol.com.co/programa/2021/11/04/sanamente/1636049871_761122.html#:~:text=por%20la%20educaci%C3%B3n%20superior%20en,cifras%20hablan%20por%20s%C3%AD%20solas.P.1.)]

empleará el análisis de datos y la utilización de las tecnologías de la información y comunicación (TIC) en el procesamiento y tratamiento de la información.

Con el propósito de comprender a fondo el factor académico, se llevó a cabo un análisis de los registros académicos de los estudiantes del tercer semestre, investigando las variables involucradas. A partir de estas variables, se generó un conjunto de datos (dataset) que fue sometido a la metodología CRISP DM para obtener bases de datos limpios y, posteriormente, desarrollar un modelo de predicción eficaz.

La importancia de la minería de datos dentro de sus dependencias de la universidad<sup>5</sup>, siendo una técnica que ha dado buenos resultados en diferentes campos del conocimiento, y de la cual se puede llegar a obtener datos válidos, útiles y comprensibles, puede radicar en la fortaleza de buscar patrones de datos válidos, novedosos, útiles y lo más importante comprensibles

Después de completar el procesamiento de la información, se inició el levantamiento de requisitos y se procedió al desarrollo de la herramienta. Para asegurar la eficiencia del proceso, se optó por implementar la metodología ágil Scrum. Gracias a esta metodología, se logró crear una página web que cumple con los requisitos establecidos y se ajusta a los estándares de calidad de software exigidos.

---

<sup>5</sup> CARNEIRO, Roberto, TOSCANO, Juan Carlos y DÍAZ, Tamara. Los desafíos de las TIC para el cambio educativo. 2021. p.127. En: <https://www.oei.es/uploads/files/microsites/28/140/lastic2.pdf>

## 1. PROBLEMA DE INVESTIGACIÓN

### 1.1 OBJETO O TEMA DE ESTUDIO

Modelos de deserción estudiantil y minería de datos.

### 1.2 LÍNEA DE INVESTIGACIÓN

“Minería de datos: Fomenta actividades investigativas para el estudio y desarrollo de métodos para la extracción automática de conocimiento y patrones de grandes bases de datos estructuradas y no estructuradas, Implementar la utilización de técnicas, algoritmos y métodos para la explotación, los sistemas de información, se ha puesto en valor el conocimiento como recurso estratégico y las bases de datos con información transaccional han pasado a convertirse en valiosos repositorios para las técnicas minería de datos que pasan a ser herramientas fundamentales en cualquier ámbito científico o empresarial<sup>6</sup> “.

### 1.3 PLANTEAMIENTO DEL PROBLEMA

De acuerdo con el <sup>7</sup>Banco Mundial en América Latina, Colombia es el segundo país con mayor deserción universitaria con un 42% de estudiantes en los primeros años entre las edades de 17 a 24 años. El Ministerio de Educación Nacional (MEN)<sup>8</sup>, en la nación como en América Latina los sistemas de educación superior tienen un alto nivel de excavación. Para el año 2018, El Sistema de Prevención y Análisis de la deserción en las instituciones de Educación superior (SPADIES) <sup>9</sup>, Colombia presenta una tasa con valor cercano a 37%. Según Timarán y Jiménez<sup>10</sup>, los

---

<sup>6</sup> Universidad Cesmag [unicesmag.edu/ingeniería-en-sistemas]. Pasto Nariño; [Consultado: 20 de febrero del 2022]. disponible en: <https://ingenieria.unicesmag.edu.co/index.php/ingenieria-en-sistemas>

<sup>7</sup> GUZMAN, Carolina; DURÁN, Diana y FRANCO, Jorge. Deserción estudiantil En: la educación superior colombiana: *Metodología de seguimiento, diagnóstico y elementos para su prevención*, 2009. 13-17 p.

<sup>9</sup> MINISTERIO DE EDUCACION NACIONAL. Estadísticas de deserción. 2018. p.1. En: [https://www.mineduacion.gov.co/sistemasdeinformacion/1735/w3-article357549.html?\\_noredirect=1](https://www.mineduacion.gov.co/sistemasdeinformacion/1735/w3-article357549.html?_noredirect=1)

<sup>10</sup> TIMARÁN, R. y JIMÉNEZ, J. Extracción de perfiles de deserción estudiantil en la Institución Universitaria CESMAG, Colombia. IRE: Ciencias Sociales y Humanas.

estudiantes que se registran en la Universidad CESMAG renuncian a aprender en los primeros semestres de la carrera lo cual ha sido evidenciado mediante la verificación de base de datos. También es importante resaltar que a pesar de que un porcentaje de estudiantes terminan sus estudios no obtienen su título profesional; Todo ello, debido a diferentes problemas, los cuales pueden ser: económicos, socio económicos, académicos y personales

El problema económico es lo más frecuente en los estratos bajos<sup>11</sup>, decidir si subsistir o educarse profesionalmente. El entorno socio económico del estudiante incide sobre los resultados de su desempeño académico. Por lo tanto, los estudiantes universitarios con mayores recursos tienen menor probabilidad de desertar, a diferencia de las personas becadas y que reciben ayudas financieras o créditos universitarios<sup>12</sup>. Los problemas de carácter académicos influyen al desconocimiento de la carrera, por lo tanto, no se genera una visión profesional y terminan abandonando sus estudios. No obstante, la mala preparación o el bajo rendimiento en su educación media hace que se genere dificultad para la comprensión de materias que son vitales para aprobar el semestre. En lo personal la falta de oportunidades de empleo para los estudiantes universitarios, incide en el abandono de sus carreras, por causa que muchos de ellos solventan su semestre con su trabajo.

El bajo nivel de rendimiento académico como una de las causas de la deserción<sup>13</sup>, pueden darse por la falta de hábitos de estudio, desconocimiento de las materias a desarrollarse durante la carrera<sup>14</sup> lo que se ve reflejado en los bajos resultados en las pruebas de estado (Saber 11), donde se demuestra el bajo conocimiento en las áreas básicas que son evaluadas en diferentes carreras universitarias y donde se evidencia que no hubo una buena preparación, ni importancia desde la etapa del colegio, donde probablemente persistía el bajo nivel educativo, la deficiencia en espacios y material didáctico, el bajo apoyo económico y poco aprovechamiento por

---

<sup>11</sup> TIMARAN PEREIRA, Silvio Ricardo. Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos En: Memorias de la 8ª Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI. 2009.

<sup>12</sup> LARROUCAU DE MAGALHAES-CALVET, Tomas. Estudio de los factores determinantes de la deserción, en el sistema universitario chileno, 2013. p.1.

<sup>13</sup> TIMARAN P., Silvio Ricardo, JIMÉNEZ, Javier Alejandro y ROMERO C., Andrés. Detención de patrones de deserción estudiantil con minería de datos. En: Editorial universitaria-Universidad de Nariño, 2017, vol. 7.

<sup>14</sup> BLANCO VILLAFÁÑE, Vanessa Paola. (2016). Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos, 2016. p.5.

parte de los estudiantes en las clases<sup>15</sup>, lo que trajo en consecuencia el fracaso universitario.

Por lo que se puede decir que el abandono por el aprendizaje tiene también que ver con situaciones de carácter emocional e institucional. La deserción estudiantil con causas emocionales también puede darse a consecuencia de la falta de identidad con la carrera<sup>16</sup>, donde los estudiantes universitarios no están completamente relacionados con la profesión a cursar, se puede afirmar que existe un porcentaje de estudiantes que desertan al no cumplirse las expectativas que exige el programa de ingeniería de sistemas<sup>17</sup>. Cabe destacar que las instituciones universitarias consideran como una pérdida de talento e ingresos monetarios el realizar este tipo de acciones por parte de los estudiantes.<sup>18</sup>

#### 1.4 formulación del problema

¿Cómo se pueden utilizar los datos históricos de los estudiantes para desarrollar un modelo predictivo de deserción estudiantil?

#### 1.5 Objetivo de la investigación

##### 1.5.1 Objetivo general

Desarrollar un modelo predictivo para determinar la deserción estudiantil del programa de Ingeniería de Sistemas de la Universidad CESMAG durante los tres primeros semestres en un aplicativo web.

##### 1.5.2 Objetivos específicos

---

<sup>15</sup> BAZANTES Z., P., Ruiz Carpio, M. L. y ÁLVAREZ Gutiérrez, M. L. Deserción estudiantil universitaria en Ecuador y su influencia en la calidad del egresado. Magazine De Las Ciencias: Revista De Investigación E Innovación, 2016, Vol. 1 núm. p.65-70

<sup>16</sup> FREIXA NIELLA, Montserrat, LLANES, Juan y VENCESLAO, Marta. El abandono en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona. En: Revista de Investigación Educativa, 2018, vol. 1.

<sup>17</sup> SIBRIÁN ESCOBAR, Luisa Amelia. Expectativas de la formación universitaria desde la perspectiva de los estudiantes. Editorial Universitaria Don Bosco. No.19. 2017. p.25-37. En: <https://www.lamjol.info/index.php/DIALOGOS/article/view/5481/5149>

<sup>18</sup> GALVIS, Carlos Angulo. Efectos de la deserción. Colombia, Cali. 2017. p.19.

- Analizar la información de los registros de notas de los estudiantes del programa de ingeniería de sistemas de la Universidad CESMAG hasta tercer semestre
- Aplicar el proceso de minería de datos con la información referente a los resultados académicos en las áreas y componentes de la malla curricular del programa de ingeniería de sistemas aplicando metodología CRISP-DM
- Elaborar un sitio web que mida y visualice por un modelo predictivo, deserción estudiantil en el programa de ingeniería de sistemas de la Universidad CESMAG
- Validar el funcionamiento del sitio web utilizando la norma de calidad ISO 25010 para producto software.

## 1.6 JUSTIFICACIÓN

La deserción universitaria tiene un papel fundamental en nuestra sociedad, por lo tanto, es importante moderar esta problemática. Por esta razón es de suma importancia culminar la educación superior por la cual se encontrará mejores ofertas laborales e incrementar la solvencia económica individual<sup>19</sup>, teniendo en cuenta que una persona sin trabajo representa gastos mayoritarios por parte del estado en financiación de programas sociales para estos sectores que no logran generar recursos propios. Por ello este fenómeno, en la actualidad, es motivo de interés tanto para el estado como para las directivas de las instituciones<sup>20</sup>, entre ellas la Universidad CESMAG.

Además, la deserción universitaria se da por cambios en la comunidad el ámbito económico como problema más frecuente en los estratos bajos, como también lo académico, el bajo nivel de rendimiento y falta de motivación de los estudiantes que ven como obligación aprender y no como una oportunidad para penetrar el bello y

---

<sup>19</sup> LOZANO TREVIÑO, David Fernando y MALDONADO MALDONADO, Lauro. Asociación entre factores económicos y sociales con la propensión de deserción escolar en colegios militarizados. revista de Estudios y Experiencias en Educación, vol. 19, núm. 40, pp. 35-52. 2020. En: <https://doi.org/10.21703/rexe.20201940lozano2>

<sup>20</sup> ERAZO ORTIZ, Eduar Alirio y MORA POTILLA, Sergio Aldemar. Aplicación de la minería de datos para la detección de factores asociados a la deserción estudiantil en los programas profesionales de las ciencias naturales y ciencias sociales de la universidad de Nariño sede Pasto. Colombia, Pasto. 2018. p.1-92.

maravilloso mundo del saber.<sup>21</sup> de igual manera los recursos de los centros educativos que pueden influir en la comodidad de los estudiantes como en docentes, también en materiales de estudio, herramientas de trabajo y conectividad de internet.

La tecnología ocupa un rol importante, la minería de datos como objetivo principal busca estudiar y analizar los altos niveles de deserción que existen en el programa de ingeniería de sistemas de la Universidad CESMAG en los primeros tres semestres, recolectando inicialmente datos de estudiantes para después clasificarlos de acuerdo a los factores involucrados, y llegar a realizar un modelo de predicción para obtener resultados que le permitan a la universidad tener información exacta de deserción estudiantil.

De desarrollarse con éxito, sería de total importancia para las dependencias interesadas dentro de la Universidad CESMAG, y basándose en los resultados obtenidos se encontrarían posibilidades para contrarrestar la problemática principal; fortaleciendo la planificación, implementando enfoques innovadores, entre otros, que permitan aumentar el número de estudiantes graduados del programa de ingeniería de sistemas dentro del programa, ya que dicho aumento traería consigo mayor reconocimiento de la carrera y universidad en la comunidad.

## 1.7 DELIMITACIÓN

La presente investigación contará con un tiempo investigativo y de desarrollo de 8 meses iniciando desde el mes de febrero del 2022 hasta enero del 2023 contando con el periodo de vacaciones de 2 meses por cada semestre, en el cual se implementará el desarrollo del producto software. De igual forma, se tomarán datos de los estudiantes de la Universidad CESMAG - programa de ingeniería de sistemas desde el primer semestre hasta el tercero, los datos serán concretamente sobre espacios académicos en materias como las ciencias básicas y socio humanísticas del plan 3 para la elaboración y desarrollo del producto software. Los datos fundamentales de los aspirantes que ingresen al programa serán también importantes para el desarrollo de la investigación.

---

<sup>21</sup> TIMARAN P., Silvio Ricardo, JIMÉNEZ, Javier Alejandro y ROMERO C., Andrés. Detención de patrones de deserción estudiantil con minería de datos. En: Editorial universitaria-Universidad de Nariño, 2017, vol. 7.

## 2. MARCO TEÓRICO

### 2.1 ANTECEDENTES

#### 2.1.1 Internacionales

La investigación **El abandono en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona**, realizada en el año 2018, en Barcelona – España, Montserrat Freixa Niella, Juan Llanes y Marta Venceslao<sup>22</sup>. plantea como objetivos analizar el momento de abandono a lo largo del periodo de estudio, identificar las diferentes causas y describir los motivos de dicho acontecimiento. La metodología usada en este caso fue de carácter cuantitativo descriptivo, donde el instrumento para la recolección de información fue un cuestionario de preguntas abiertas y cerradas, cuyas dimensiones utilizadas fueron (estatus actual, procesos y causas de abandono, valoración académica y la situación académica actual).

Los resultados obtenidos confirmaron que el abandono se produce por desmotivación, expectativas no cumplidas entre el trabajo y el estudio, fracaso académico y la falta de identidad con la carrera elegida, también se logró establecer que la deserción no está ligada únicamente como el abandono por el sistema educativo o el no querer adquirir conocimientos, ya que varios estudiantes que han desertado inicialmente, regresan a reintegrar otras carreras educativas. Finalmente se logró concluir que es importante construir y confrontar una identidad profesional siendo este un aspecto que resulta especialmente sensible para los estudiantes que acceden a ciclos de formación; para los cursos que resultan difíciles para el estudiante y que es otro motivo de abandono académico, se sugirió: ayudar a los responsables a crear planes específicos de tutoría y asesoramiento que prevengan dichas acciones negativas, de igual forma buscar que los itinerarios formativos sean replanteados a fin de disminuir la deserción a causa de la relación entre trabajo y estudio.<sup>23</sup>

---

<sup>22</sup> FREIXA NIELLA, Montserrat, LLANES, Juan y VENCESLAO, Marta. El abandono en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona. En: Revista de Investigación Educativa, 2018, vol. 1.

<sup>23</sup> FREIXA NIELLA, Montserrat, LLANES, Juan y VENCESLAO, Marta. El abandono en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona. En: Revista de Investigación Educativa, 2018, vol. 1.

El anterior antecedente resulta importante para la presente investigación porque se enfoca en los motivos de deserción estudiantil, abarcando factores de carácter económico, académico y personal.

El estudio **Análisis de la deserción estudiantil en la Universidad de Chile, con énfasis en las ayudas financieras** escrito por Rafael Díaz Escobar<sup>24</sup>, en el año 2021, en Santiago de Chile. Establece como objetivo de investigación analizar la contribución de los distintos tipos de ayuda financiera, tanto del estado como de la institución, para ello se propuso identificar la probabilidad de deserción entre los estudiantes que reciben y los que no reciben las ayudas financieras, como segundo parámetro se estableció identificar la probabilidad de deserción académica y no académica, tales como el quintil de ingreso del estudiante o de la de sus padres. La metodología empleada para el desarrollo de la investigación es el modelo logít de riesgo en tiempo discreto que permite vincular sistemáticamente el riesgo de deserción con los factores financieros, socioeconómicos, académicos y no económicos.

En general, los resultados indicaron que los estudiantes que cuentan con ayuda financiera presentan menor riesgo de deserción en comparación con quienes no tienen ese apoyo. Esto se reafirmó cuando se analizó el tipo de ayuda financiera, siendo importantes las de arancel y manutención, ya que los beneficiados por becas estatales presentan niveles bajos de deserción tanto en los modelos Logít, como al utilizar el Propensity Score Matching (PSM). Este análisis se realizó con datos de admisión estudiantil, ayuda financiera, y Prueba de Selección Universitaria (PSU). En conclusión, el estudio afirma que las mujeres tienen mayor probabilidad de permanecer en el sistema; de acuerdo con las evidencias los estudiantes procedentes de regiones de bajos recursos y con padres cuyo nivel de educación no sobrepasa la secundaria, son quienes requieren soporte o ayuda estatal e institucional al presentar mayores riesgos de deserción; la variable que más contribuyen a reducir el riesgo de deserción son las denominadas becas de manutención donde se distribuyen ayudas de transferencia monetaria de libre disposición; finalmente se determinó que la deserción es un problema a nivel mundial donde se puede implementar métodos

---

<sup>24</sup> DÍAZ ESCOBAR, Rafael. Análisis de la deserción estudiantil en la Universidad de Chile, con énfasis en las ayudas financieras, 2021.p. II.

cualitativos y modelos matemáticos para llegar a encontrar resultados certeros al momento de estudiar las variables.<sup>25</sup>

La importancia de esta investigación radica en qué implementando modelos matemáticos se tiene mayor eficiencia en el estudio de variables, lo cual llevará a tener una investigación mucho más eficiente, donde se logrará tener resultados idóneos como los que presenta el autor.

El estudio **Deserción estudiantil universitaria en Ecuador y su influencia en la calidad del egresado**, llevado a cabo en el 2017, en Babahoyo – Ecuador, por Zoila Piedad Basantes, María Lucía Ruiz Carpio y Marcela Leonor Álvarez Gutiérrez<sup>26</sup>. Presenta como objetivo estudiar el abandono de los estudios universitarios, como inconveniente social que se presenta con frecuencia en los diferentes centros de educación superior a nivel mundial.

Como resultados dentro del estudio se demostró el desinterés en las clases por parte de los estudiantes quienes se dedican cumplir solo por compromiso social y en muchos casos con o sin familia sus propios proyectos de vida, la persistencia del bajo nivel educativo, la deficiente infraestructura y material didáctico, el bajo apoyo económico y poco aprovechamiento de las clases, entre otros son los causantes de la deserción.<sup>27</sup>

---

<sup>25</sup> DÍAZ ESCOBAR, Rafael. Análisis de la deserción estudiantil en la Universidad de Chile, con énfasis en las ayudas financieras, 2021.p. II.

<sup>26</sup> BAZANTES Z., P., RUIZ CARPIO, M. L. y ÁLVAREZ GUTIÉRREZ, M. L. Deserción estudiantil universitaria en Ecuador y su influencia en la calidad del egresado. Magazine De Las Ciencias: Revista De Investigación E Innovación, 2016, Vol. 1 núm. p.65-70.

<sup>27</sup> BAZANTES Z., P., Ruiz Carpio, M. L. y ÁLVAREZ Gutiérrez, M. L. Deserción estudiantil universitaria en Ecuador y su influencia en la calidad del egresado. Magazine De Las Ciencias: Revista De Investigación E Innovación, 2016, Vol. 1 núm. p.65-70

Por ende, resulta fundamental estudiar esta clase de variables para generar un análisis más apto y poder dar solución social a los estudiantes que deciden desistir de la educación.

La investigación **Una aproximación teórica a la deserción estudiantil universitaria**, realizada en el año 2014, en Lima, Perú, por Héctor Tudela Viela<sup>28</sup>. Presenta como objetivo demostrar la relación entre los modelos existentes con la deserción universitaria. Como metodología se utilizó el modelo Fishbein y Ajzen: un modelo general de las relaciones entre actitudes, convicciones, presión social; también el Modelo de Spady que indica que la deserción se presenta por falta de integración por parte del estudiante en el entorno educativo.

Dentro de sus resultados se establecieron diferentes modalidades de deserción, una de ellas es denominada “voluntaria” la cual hace que por méritos propios el estudiante tome la decisión de dejar su carrera por motivos personales, la siguiente modalidad es por violación al reglamento interno de la universidad y cuando el alumno no ha realizado trámites académico administrativos durante siete o más trimestres consecutivos. Finalmente se concluyó que las causas de deserción son provocadas por emociones y temas psicológicos.<sup>29</sup>

La anterior referencia representa una serie de modelos que resultan importantes porque se pueden utilizar para clasificar el tipo de deserción en la presente investigación.

---

<sup>28</sup> TUDELA VIELA, Héctor. una aproximación teórica a la deserción estudiantil universitaria Revista digital de Investigación en Docencia universitaria, 2014, p. 59-76.

<sup>29</sup> TUDELA VIELA, Héctor. una aproximación teórica a la deserción estudiantil universitaria Revista digital de Investigación en Docencia universitaria, 2014, p. 59-76.

De igual manera la tesis de maestría titulada **Estudio de los factores determinantes de la deserción, en el sistema universitario chileno**, realizado en Santiago de Chile, en el año 2013, Tomas Larroucau de Magalhaes-Calvet<sup>30</sup>, Propuso un modelo econométrico de tres ecuaciones que tiene como objetivo seleccionar la muestra de ayudas estudiantiles clasificando la deserción temporal y definitiva, analizando la heterogeneidad de carreras y tipos de universidades. El estudio de este modelo se hace a partir de pruebas internas de país denominadas SIMCE elaborada en el 2003 y prueba de selección universitaria (PSU) en los años 2005 a 2006, para ello se realizó la clasificación de los diferentes tipos de deserción de estudio.

resultado se obtuvo que la deserción universitaria está influida por factores de habilidades individuales, socioeconómicos y demográficos, pero también factores relacionados a enseñanza media. Las características previas al ingreso de los universitarios y el ranking de calidad académica en educación media son relevantes a las habilidades por medio de la PSU, además se constató que la deserción es una problemática financiera donde estudiantes con mayores recursos tienen menos probabilidad de desertar, a diferencia de estudiantes becados o con ayudas financieras y créditos universitarios. Finalmente, en función de los resultados obtenidos, se propuso medidas de políticas públicas para reducir las elevadas tasas de deserción del sistema universitario.<sup>31</sup>

Es significativo implementar modelos basados en datos de diferentes pruebas para obtener mejores resultados al momento de implementar el análisis de las variables.

---

<sup>30</sup> LARROUCAU DE MAGALHAES-CALVET, Tomas. Estudio de los factores determinantes de la deserción, en el sistema universitario chileno, 2013. p.1.

<sup>31</sup> LARROUCAU DE MAGALHAES-CALVET, Tomas. Estudio de los factores determinantes de la deserción, en el sistema universitario chileno, 2013. p.1.

### 2.1.2 Nacionales

En la tesis de maestría titulada **Análisis de la deserción y permanencia académica en la educación superior aplicando minería de datos**, desarrollada en el año 2016, en Bogotá, Colombia, por Álvaro Agustín Oñate Bowen<sup>32</sup>. Establece como objetivos: construir un conjunto de datos de estudiantes que contemplen información de tipo socioeconómico, historial académico y desempeño en pruebas de estado, evaluar mediante un modelo descriptivo que estudiantes admitidos y los matriculados por cuatro periodos académicos, y evaluar un modelo predictivo para predecir la pérdida de la condición académica de un estudiante en los primeros cuatro semestres. Donde la metodología empleada incluyó la minería de datos en la educación, el cual fue realizado para modelar el comportamiento de los estudiantes inscritos en los programas de Ingeniería Electrónica e Ingeniería de Sistemas por la Universidad Popular de César.

Se implementó dos tareas de minería de datos: Primero, basado en tareas descriptivas y algoritmo K-means para la selección de grupos de alumnos, en segundo lugar, clasificación impulsada por dos técnicas llamadas árboles de decisiones y Naive Bayes para predecir la pérdida de estatus académico por los estudiantes que tienen malas calificaciones en sus primeros cuatro semestres. Para el entrenamiento y prueba de los modelos, se utilizaron los expedientes académicos y los datos recogidos durante el proceso de admisión de los estudiantes y se evaluaron utilizando la técnica de validación cruzada. Los resultados experimentales demostraron que la predicción de la pérdida académica mejora cuando se añaden los datos de la matrícula académica anterior.<sup>33</sup>

---

<sup>32</sup> OÑATE BOWEN, Álvaro Agustín. Análisis de la deserción y permanencia académica en la educación superior aplicando minería de datos, 2016. p.5.

<sup>33</sup> OÑATE BOWEN, Álvaro Agustín. Análisis de la deserción y permanencia académica en la educación superior aplicando minería de datos, 2016. p.5.

Para la investigación a desarrollar, lo anterior permite estudiar la posibilidad de implementar los modelos de minería de datos en base a árboles de decisiones, ya que ayuda clasificar los datos de una manera más eficiente y llegar a resultados infalibles.

La investigación **Relación y uso de las TIC con la deserción en estudiantes pertenecientes a la Facultad de Ingeniería de la Universidad Nacional de Colombia**. Por Héctor Julián García Otálora<sup>34</sup>, en Bogotá – Colombia, en el año 2017. Presentó como objetivo de tesis determinar si existe una relación entre el uso de tecnologías de la información y la comunicación y la deserción en estudiantes de la Facultad de Ingeniería de la Universidad, la construcción de la metodología se basó principalmente en la relación de la deserción educativa con las TIC, la cual consistió

En la modificación del actual modelo descriptivo de deserción de Vincent Tinto, incluyendo el impacto de las TIC, la elección del modelo subyacente comenzó con la búsqueda de definiciones de conceptos relevantes, la exploración de diferentes mecanismos que explican el fenómeno de la deserción escolar y la búsqueda en la literatura de herramientas que puedan proporcionar una transición de la educación tradicional al mundo de la educación, tecnología de la información y la comunicación; una vez se validó un entorno real se clasificó la población objetivo por herramientas estadísticas y se elaboró encuestas para los estudiantes que no pertenecen a la facultad de ingeniería por motivos de deserción o por culminación de estudios. Al alimentar el modelo propuesto se nota una concordancia con la población que abandonó sus estudios, y se pudo determinar que existe evidencia estadística que las TIC tienen relación con la deserción educativa.<sup>35</sup>

---

<sup>34</sup> GARCÍA OTÁLORA, Héctor Julián. Relación y uso de las TIC con la deserción en estudiantes pertenecientes a la Facultad de Ingeniería de la Universidad Nacional de Colombia, 2017. p.1.

<sup>35</sup> GARCÍA OTÁLORA, Héctor Julián. Relación y uso de las TIC con la deserción en estudiantes pertenecientes a la Facultad de Ingeniería de la Universidad Nacional de Colombia, 2017. p.1. <sup>44</sup>

El anterior antecedente es importante dentro de la investigación porque brinda una posible metodología para la presente investigación, y una relación evidente entre las TIC y la deserción educativa.

El estudio titulado **Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos**, llevado a cabo en el año 2015, en Bogotá, Colombia, desarrollado por Vanessa Paola Blanco Villafañe.<sup>36</sup> Donde se estableció como objetivos: preparar un conjunto de datos apropiados para la aplicación de las diferentes técnicas de minería de datos, proponer e implementar un modelo descriptivo para el agrupamiento de datos y realizar la evaluación sistemática del modelo propuesto y un análisis de los resultados obtenidos. Presenta detalles de la metodología aplicada basada en CRISP-DM, la cual describe un modelo de proceso jerárquico de un conjunto de tareas, en este caso consiste de 6 fases dinámicas: comprensión del negocio, comprensión de los datos, preparación de datos, modelado, evaluación y desarrollo.

Finalmente se logró obtener como resultados que el entorno socioeconómico del estudiante incide en los resultados de su desempeño académico, encontrando que a mayor nivel socioeconómico es mayor el puntaje en la prueba, también que las zonas rurales tienden a presentar bajo desempeño educativo, de igual forma se identificó que los niveles de rendimiento más bajos se presentan en el área de filosofía. En conclusión, el nivel socioeconómico juega un papel importante dentro de la educación, las zonas rurales presentan menor calidad educativa, los niveles de educación de los padres inciden de forma indirecta sobre el desempeño de los estudiantes.<sup>37</sup>

Es fundamental contar con este tipo de información puesto que la deserción estudiantil tiene muchos factores de decisión, tanto para lograr identificar en qué

---

<sup>36</sup> BLANCO VILLAFañE, Vanessa Paola. (2016). Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos, 2016. p.5.

<sup>37</sup> BLANCO VILLAFañE, Vanessa Paola. (2016). Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos, 2016. p.5. <sup>46</sup> Palmira, 2017. p.9.

nivel de aprendizaje terminan los estudiantes bachilleres como para identificar sus habilidades y destrezas en el campo formativo.

La investigación **Estudio de factores que inciden en el retiro universitario y el rezago en los planes de estudio en los programas de pregrado de la Universidad Nacional de Colombia sede Palmira**, realizada en el año 2017, por Eivar Alejandro Mora Medina<sup>38</sup>, establece como objetivo analizar los factores que conllevan al desistimiento universitario y rezago en los planes de estudio de pregrado en la universidad nacional con sede en Palmira. La metodología de la investigación tuvo un enfoque exploratorio, descriptivo y explicativo, y se llevó a cabo en dos etapas, la primera se basó en la recolección de información secundaria que permitió conocer los aportes acerca del tema, después se recogió la información primaria y se realizó el análisis de resultados; aquí se abordó variables de carácter cualitativo y cuantitativo relacionadas con estadísticas y comportamientos personales de los estudiantes.

Los resultados Indicaron los factores de deserción y a su vez razones o causas que conllevaron a que el alumno no complete su plan de estudio en un tiempo esperado resultaron multicausales, destacándose un alto porcentaje de estudiantes que abandonan o no logran graduarse por problemas económicos, también se logró evidenciar que los aspirantes no reciben orientación profesional antes de ingresar a una institución educativa, y como conclusión se obtuvo que el retiro y deserción estudiantil es un problema que surge a consecuencia de procesos socioeconómicos, motivacionales, académicos, políticos y personales.<sup>39</sup>

Considerando los factores que se presenta en el anterior estudio es de fundamental importancia en esta investigación debido a que podemos llegar a estudiar qué tipo

---

<sup>38</sup> MORA BRAVO, Eivar Alejandro. Estudio de factores que inciden en el retiro universitario y el rezago en los planes de estudio en los programas de pregrado de la Universidad Nacional de Colombia sede

<sup>39</sup> MORA BRAVO, Eivar Alejandro. Estudio de factores que inciden en el retiro universitario y el rezago en los planes de estudio en los programas de pregrado de la Universidad Nacional de Colombia sede Palmira, 2017. p.9.

de factores pueda afectar en la Universidad CESMAG para evitar el retiro de los estudiantes ya sea por lo académico, socioeconómico o falta de motivación para así disminuir es desistimiento en la universidad.

La investigación **Método para la detección de estudiantes en riesgo de deserción, basado en un diseño de métricas y una técnica de minería de datos.** Elaborada en el 2021, por Daniel Zapata Medina<sup>40</sup>, en Medellín, Colombia. Plantea como intención principal elaborar métodos basados en métricas para cambiar las causas principales, aportando el procesamiento y la comprensión profunda de los datos, de esa forma seleccionar la representación adecuada de las características para llevarlas como entradas de clasificadores expertos en el tipo de características. Donde se empleó una mezcla en la clasificación para mejores respuestas, puesto que hubo errores en algunas muestras, con esta implementación se mejoró el rendimiento del modelo de clasificación para optar por buenos resultados y analizar la interpretación de una manera más fácil.

Como resultados se permitió comprobar la capacidad de detección de la deserción utilizando la transformación de características a partir de métricas, logrando un 82% de precisión y 64% de recall, estos resultados demostraron el aumento significativo con respecto al 71% de precisión y 57% de recall alcanzado con las características iniciales sin el uso de métricas. En resumen, se concluyó que de acuerdo a los estudios realizados en esta investigación resulta factible aplicarlo en la analítica de datos para determinar la precisión temprana, para así, presidir al estudiante de permanecer en las instituciones educativas.<sup>41</sup>

La investigación anterior brinda información relevante para el desarrollo de la presente, puesto que plantea y establece estrategias combinadas con métricas y

---

<sup>40</sup> ZAPATA MEDINA, Daniel. Método para la detección de estudiantes en riesgo de deserción, basado en un diseño de métricas y una técnica de minería de datos. Universidad Nacional de Colombia,2021. p.9.

<sup>41</sup> ZAPATA MEDINA, Daniel. Método para la detección de estudiantes en riesgo de deserción, basado en un diseño de métricas y una técnica de minería de datos. Universidad Nacional de Colombia,2021. p.9.

técnicas de minería de datos dentro de la deserción estudiantil, convirtiéndose en un campo de poca exploración.

### 2.1.3 Regionales

El estudio **Detención de patrones de deserción estudiantil con minería de datos**, por Silvio Ricardo Timaran Pereira y otros<sup>42</sup>, elaborado en el año 2017, en Pasto, Colombia, planteo como objetivo estudiar patrones de deserción a partir de variables socioeconómicos, académicos y disciplinares para los programas de pregrado para la universidad de Nariño y la institución universitaria CESMAG, aplicando técnicas de minería de datos para la detención temprana. La metodología de desarrollo inició seleccionando bases de datos de los dos centros educativos, después filtrando datos socioeconómicos, académicos y disciplinarios a partir de unos años predeterminados aplicados por los autores, con los datos se elaboró repositorios limpios desarrollados en postgre SQL, con la finalidad de aplicar etapas de procesamientos y transformación para obtener datos limpios y poder aplicar las técnicas de minería de datos. Las técnicas que se llevaron a cabo fueron arboles de decisión, como también de asociación de clustering. Se usó una herramienta libre para el procesamiento de minería de datos denominada weka.

Los resultados mostraron un patrón general al tener materias pérdidas en los primeros semestres de la carrera y un puntaje bajo en las pruebas ICFES en ambas instituciones, además se pudo apreciar que en la universidad de Nariño incide el valor de matrícula a pesar que es un valor muy bajo, para estudiantes oriundos de la zona sur de la ciudad, es complejo solventar ese valor por la situación económica.<sup>43</sup>

---

<sup>42</sup> TIMARAN P., Silvio Ricardo, JIMÉNEZ, Javier Alejandro y ROMERO C., Andrés. Detención de patrones de deserción estudiantil con minería de datos. En: Editorial universitaria-Universidad de nariño,2017, vol. 7.

<sup>43</sup> TIMARAN P., Silvio Ricardo, JIMÉNEZ, Javier Alejandro y ROMERO C., Andrés. Detención de patrones de deserción estudiantil con minería de datos. En: Editorial universitaria-Universidad de nariño,2017, vol. 7.

Para esta investigación es grato contar con este tipo de estudios realizados en la región, puesto que se puede aplicar un estudio para comparar resultados con los de autor y las propuestas en esta investigación.

La investigación **Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11**, de los autores: Silvio Ricardo Timaran Pereira y otros<sup>44</sup>, realizado en el año 2018, en Pasto - Colombia. Plantea como objetivo analizar los factores de desempeño académico de estudiantes colombianos de grado once que presentaron sus pruebas saber 11 entre años 2015 y 2016 implementando modelos de clasificación en arboles de decisión. El estudio implementó la metodología CRISP-DM donde se estudió las bases de datos del ICFES, filtrando la información socioeconómica, académica e institucional de los estudiantes. Con la herramienta WEKA se elaboró los árboles de decisión, que en ello identifico desempeños académicos, para clasificar en un rendimiento alto o bajo en las pruebas saber 11.

La información expuesta en el estudio anterior es importante para el presente estudio porque estableció una de las problemáticas de la deserción estudiantil, en este caso la mala preparación académica de los estudiantes para las pruebas ICFES, y dónde el puntaje de las pruebas influye a la hora de elegir de forma acertada las carreras a estudiar y evitar una deserción.

Finalmente, el estudio **Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos**, en el año 2009, en la ciudad de Pasto, Colombia, por Silvio Ricardo Timaran<sup>45</sup>. El objetivo planteado

---

<sup>44</sup> TIMARÁN-PEREIRA, R., CAICEDO-ZAMBRANO, J. y HIDALGO-TROYA, A. Árboles de decisiones para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas saber 11°, 2019, Rev.investig. desarro. innova.

<sup>45</sup> TIMARAN PEREIRA, Silvio Ricardo. Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos En: Memorias de la 8ª Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI. 2009.

fue determinar deserción y rendimiento estudiantil en la población universitaria UDENAR, a partir de datos almacenados durante los últimos 15 años, para el procesamiento de los datos se implementó una herramienta desarrollada por el departamento de ingeniería de UDENAR denominada TariyKDD, la cual hizo el proceso de minería de datos para llegar a los resultados propuestos. La etapa de selección de datos se clasificó en externa e interna, donde la externa utilizó una base de datos de la universidad de Nariño y la interna la información de colegios del país por medio del Ministerio de Educación Nacional de Colombia, como resultados se obtuvo que gran parte de estudiantes de primer semestre, oriundos de la zona sur de Nariño, de estratos bajos que pertenecen a facultades de ciencias básicas, humanas y naturales presentan un bajo rendimiento. Como recomendación se sugirió implementar estrategias de seguimiento a estudiantes con estos perfiles con el fin de prevenir que caigan en bajo rendimiento y disminuir el grado de deserción que se presenta en dicha institución.

La investigación bibliográfica anterior, delimitada a deserción estudiantil y minería de datos, expuso diferentes maneras de implementar minería de datos, cada una de ellas con ventajas y desventajas al usarlas, y donde lo más relevante fue trabajar con árboles de decisiones que serán de gran ayuda dentro del presente estudio.

## 2.2 Supuesto teóricos de la investigación

### 2.2.1 Ámbito educativo

En el ámbito educativo es necesario la integración de las TIC, como dice Gargallo Castel.<sup>46</sup> El ámbito educativo ha incorporado gradualmente las oportunidades ofrecidas por las Tecnologías de la Información y Comunicación (TIC). Si bien este dato resulta positivo, desde el punto de vista organizativo, es necesario complementar la implantación de las TIC con una estrategia global e integradora que aporte coherencia entre los aspectos puramente tecnológicos y los modelos educativos, para alcanzar resultados superiores,

Examen Saber 11° : Esta prueba se ha realizado con el fin de evaluar el conocimiento de los estudiantes en nuestro país, Moncayo Cabrera, afirma que “Las diferentes evaluaciones que realiza el Instituto Colombiano para el Fomento de la Educación Superior – ICFES, no miden con exactitud el conocimiento de los estudiantes colombianos, su proceso educativo o sus distintas habilidades, pero siguen siendo la principal herramienta de evaluación de la calidad de la educación en el país y del rendimiento académico de los estudiantes. Por ende, la calidad de la educación vista desde el rendimiento académico de los estudiantes, es actualmente un tema primordial dentro de los asuntos educativos, y especialmente dentro de la planeación de las políticas educativas del gobierno”<sup>47</sup>.

Examen Saber Pro: como dice el MEN<sup>48</sup> esta prueba es aplicada a todos los estudiantes de educación superior la cual su presentación a esta prueba es obligatoria desde el año 2010 para poder obtener el título de la carrera la cual se ha estado preparando. Otra de las definiciones que Plazas Estepa y Becerra Rodríguez afirman que “Las Pruebas Saber Pro conforman un papel preponderante en el desempeño y la calidad de la educación superior en Colombia. La ley 1324 de 2009 establece los parámetros que regulan la medición de la calidad en las instituciones universitarias. Los hallazgos permiten indicar el rol fundamental de las pruebas SABER PRO en el tejido de la educación superior en Colombia y su posible comparación con otros países de la región, las pruebas SABER PRO se consolidan como una herramienta muy importante en el proceso integral de

---

<sup>46</sup> GARGALLO CASTEL, Ana Felicitas. La integración de las TIC en los procesos educativos y organizativos. Pdf, junio 2018.

<sup>47</sup> MONCAYO CABRERA, M. A. (2016). Determinantes que influyen en el rendimiento académico: un estudio aplicado para Colombia a partir de las Pruebas Icfes - Saber 11. Retrieved from

<sup>48</sup> MEN, la educación en Colombia [línea], 2016

evaluación de la calidad de la educación superior en Colombia, ya que al ser una valoración externa, contribuye a que las instituciones universitarias como los mismos estudiantes puedan establecer mecanismos para mejorar los programas y el propio desarrollo de la sociedad”<sup>49</sup>.

### 2.2.2 Deserción estudiantil

La deserción universitaria es la causa principal del bajo rendimiento en los programas académicos, <sup>50</sup>a partir de la identificación de las causas que motivan la deserción de los estudiantes que ingresan a las carreras universitarias, para lo cual se tendrá en cuenta los elementos que determinan el abandono de los estudios. Si se acepta que la deserción es la comparación numérica entre la matrícula inicial menos el número de egresados de último año, se tiene un concepto cuantitativo de ella. Estudios recientes han concluido que el tema de la deserción universitaria se presenta principalmente en los primeros años de la carrera. Para tal efecto se han desarrollado investigaciones sobre la deserción y mortalidad estudiantil, bajo análisis cuantitativo y cualitativo. Estos estudios han hecho énfasis en determinar las causas de la deserción universitaria.

### 2.2.3 Documentos institucionales y de programa

Pei: (Proyecto educativo institucional)

Para el ingreso de estudiantes a la Universidad CESMAG se requiere de tener un perfil donde en el PEI nos afirma que “El estudiante de la Universidad CESMAG es una persona que conoce, respeta, comparte y se identifica con la misión y visión institucional y asume un papel activo en su proceso de formación integral, demostrando capacidad y disposición para Aprender a Aprender y para participar en el desarrollo de la vida universitaria”<sup>51</sup> Dicho lo anterior se requiere seguir con

---

<sup>49</sup> PLAZAS ESTEPA, r., & becerra rodríguez, r. (2016). una mirada a las pruebas saber pro en la calidad de la educación superior en Colombia. Pensamiento Republicano,

<sup>50</sup> BURGOS FLOREZ, Marco Antonio, et al. Diagnóstico del fenómeno de deserción estudiantil en los programas de pregrado de la Universidad de Nariño—sede Pasto 2012–2016. 2018.

<sup>51</sup> Universidad Cesmag, PEI-UNICESMAG, [sitio web]. Pasto Nariño; [consultado en febrero de 2022]. Disponible en <https://www.unicesmag.edu.co/wp-content/uploads/2021/01/PEI-UNICESMAG-2020.pdf>

unos pasos en específico el cual está escrito en el reglamento estudiantil donde nos afirma que

“ARTÍCULO 7. REQUISITOS PARA LA INSCRIPCIÓN DE ASPIRANTES A PRIMER SEMESTRE. El aspirante a un programa académico ofrecido por la Universidad podrá solicitar su ingreso, a través del registro en el aplicativo web o el medio establecido por la Universidad, radicando dentro del término establecido en el calendario académico, los siguientes documentos:

- Formulario de inscripción debidamente diligenciado.
- Documento de identidad.
- Comprobante o validación de pago de los derechos de inscripción
- Fotografía digital en el medio y características definidos por la Universidad
- Resultados de las pruebas de Estado realizadas para la culminación de la educación media, en el caso de pregrado.
- Acta de grado y/o diploma que acredite su condición de bachiller o certificado expedido por el establecimiento educativo, donde indique la próxima fecha de graduación con un tiempo de expedición no superior a treinta (30) días, en el caso de pregrado.
- Acta de grado de estudios de pregrado, en caso de posgrado”<sup>52</sup>

Pep: (Proyecto educativo del programa de ingeniería de sistemas)

Como dice el PEP<sup>53</sup>, para poder ingresar al programa de ingeniería de sistemas se requiere seguir un proceso que es otorgado por las directrices de la universidad, el cual el perfil del estudiante de ingeniería de sistemas de la Universidad Cesmag debe ser de un nivel de creatividad, ingenio, pensamiento lógico y curiosidad en la solución de problemas que se presenten en la vida cotidiana del ser humano haciendo uso de la tecnología, siendo esta la base fundamental del programa, debido a esto el estudiante aspirante deberá tener el conocimiento previo a las matemáticas ya que es una de las bases fundamentales del postgrado. Debido a lo anterior y si cumple con lo establecido se podrá continuar con el debido proceso el cual deberá realizar una inscripción para luego pasar por una selección y por último la admisión que estará a cargo el comité de admisiones de la facultada de ingeniería de sistemas, de hecho, esto se encuentra en el reglamento general estudiantil.

---

<sup>52</sup> Universidad Cesmag, La institución-reglamento, [sitio web]. Pasto Nariño; [consultado en febrero de 2022]. Disponible en <https://www.unicesmag.edu.co/reglamentos/>

<sup>53</sup> PEP, proyecto educativo del programa de ingeniería de sistemas, 2021

## 2.2.4 Ingeniería de software

La ingeniería de software según León Socha afirma que es un “Conjunto de conocimientos y técnicas científicas, empíricas y prácticas aplicadas a la invención, el diseño, el desarrollo, la construcción, el mantenimiento y el perfeccionamiento de tecnologías, estructuras, máquinas, herramientas, sistemas, materiales y procesos para la resolución de problemas prácticos”<sup>54</sup>.

## 2.2.5 Metodología Scrum

La metodología Scrum según Roger S<sup>55</sup> es un proceso en el que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto. En Scrum se realizan entregas parciales y regulares del producto final, priorizadas por el beneficio que aportan al receptor del proyecto. Por ello, Scrum está especialmente indicado para proyectos:

- En entornos complejos, donde se necesita obtener resultados pronto
- Donde los requisitos son cambiantes o poco definidos
- Donde la innovación, la competitividad, la flexibilidad y la productividad son fundamentales.

### Roles SCRUM:

- Un proyecto de desarrollo se puede llevar a cabo mediante uno o más equipos Scrum
- Un equipo Scrum está formado por personas que juegan tres tipos de roles: Product Owner, Scrum Master y Development team member
- Un equipo Scrum se auto-organiza y no necesita jefes o gestores, aunque si serán necesarios en el contexto de la organización: contratación, formación, establecimiento y control de objetivos, gestión económica, asignación de personas y tareas, etc.

---

<sup>54</sup> LEÓN SOCHA, Fredy Alonso. Ingeniería de software 1. 1ed, noviembre de 2017, p9

<sup>55</sup> ROGER S, Pressman. Ingeniería de software I. 7ed

## Actores SCRUM. Product Owner.

Un product owner es uno de los futuros usuarios del sistema, alguien que sabe lo que quieren los usuarios del sistema en desarrollo. El product owner es clave en un proyecto ágil. Y si no realiza su trabajo puede poner en riesgo el proyecto.

## Actores SCRUM. SCRUM Máster

Velar por que todos los participantes del proyecto sigan los valores y principios ágiles, las reglas y proceso de Scrum y guiar la colaboración intraequipo y con el cliente de manera que las sinergias sean máximas. Esto implica:

- Asegurar que exista una lista de requisitos priorizada y que esté preparada antes de la siguiente iteración.
- Facilitar las reuniones de Scrum (planificación de la iteración, reuniones diarias de sincronización del equipo, demostración, retrospectiva), de manera que sean productivas y consigan sus objetivos.
- Enseñar al equipo a auto gestionarse. No da respuestas, sino que guía al equipo con preguntas para que descubran por sí mismo una solución.

## Paso de la metodología SCRUM

### Planificación del sprint

Si entendemos el significado del sprint como un mini proyecto dentro del proyecto principal, cada uno de ellos tiene un objetivo en particular. Por ejemplo, el primer intervalo puede ser plantear cuál será el presupuesto general a utilizar, por lo que se necesitará de un equipo de profesionales expertos en el tema económico. En la primera reunión del equipo se definirán aspectos como la funcionalidad, objetivos, riesgos del sprint, plazos de entrega, entre otros. Posteriormente se realiza una junta entre el equipo y el jefe del proyecto para explicar cómo se desarrollará cada punto del intervalo. Aquí se evaluarán cambios, toma de decisiones, mejoras y más factores.

### Etapas de desarrollo

Cuando el trabajo del sprint está en curso, los encargados deben garantizar que no se generen cambios de último momento que puedan afectar los objetivos del mismo. Además, se asegura el cumplimiento de los plazos establecidos para su término.

### Revisión del sprint

Al final del desarrollo del intervalo, es posible analizar y evaluar los resultados. Si es necesario, todo el equipo colaborará para saber qué aspectos necesitan ser cambiados. En esta fase se fomenta la colaboración y retroalimentación entre todos. Como Colaboración entre equipos, supervisores, jefes y dueños de productos y se admite un análisis externo como forma de complementación

### Retroalimentación

Los resultados pueden ser entregados para recibir un feedback no solo por parte de los profesionales dentro del proyecto, sino también de las personas que utilizarán directamente lo que se desea lograr; es decir, los clientes potenciales. Las lecciones aprendidas durante esta etapa permitirán que el siguiente sprint pueda ser mucho más efectivo y ágil. La metodología Scrum no se utiliza en todos los casos. Se emplea cuando la empresa posee los recursos disponibles, la madurez y experiencia del equipo encargado, una estructura organizacional ágil e innovadora, entre otros factores. Contar con un profesional que asegure estos principios será el primer paso<sup>56</sup>

### 2.2.6 Bases de datos

Las bases de datos almacenan gran cantidad de datos, relacionados y estructurados, que pueden ser consultados rápidamente de acuerdo con las características selectivas que los usuarios deseen así lo afirma J.C “Un sistema de bases de datos es básicamente un sistema computarizado para llevar registros. Es posible considerar a la propia base de datos como una especie de armario electrónico para archivar; es decir, es un depósito o contenedor de una colección de archivos de datos computarizados. Los usuarios del sistema pueden realizar una variedad de operaciones sobre dichos archivos. por ejemplo: Agregar nuevos archivos vacíos a la base de datos, Insertar datos dentro de los archivos existentes, recuperar datos de los archivos existentes, modificar datos en archivos existentes,

---

<sup>56</sup> Ken Schwaber & Jeff Sutherland; La Guía Definitiva de Scrum: Las Reglas del Juego: noviembre 2022

eliminar datos de los archivos existentes, eliminar archivos existentes de la base de datos”<sup>57</sup>.

### 2.2.7 Web

<sup>58</sup> Es una colección de documentos web enlazados entre sí por enlaces de hipertexto, disponibles en Internet que pueden comunicarse a través de tecnología digital. "Hipertexto" significa una mezcla de textos, gráficos y archivos de todo tipo, en un mismo documento. La web es el diminutivo de *Word wide web* o *www* cuyas tecnologías para su funcionamiento (HTML, URL, HTTP) fueron desarrolladas en el año 1990 por Tim Berners Lee. Para usar la web es necesario tener acceso a internet y un navegador web, por la cual se solicita una página dinámica llamada también página web. El navegador web como, por ejemplo, Google Chrome, se comunica con el servidor web mediante el protocolo web o HTTP para entregar la petición deseada.

### 2.2.7 Framework

Laravel como dice Tobinson; <sup>59</sup> Este framework fue diseñado para desarrollar aplicaciones rápidamente en php, utilizando la arquitectura MVC, Laravel es uno de los marcos más populares actualmente. Cuenta con muchos paquetes específicos como lo son: el motor de plantillas blade, pruebas de unidades, ORM, un sistema de empaquetado, controladores RESTful, entre muchos otros. Como características principales encontramos desarrollo rápido de aplicaciones, arquitectura MVC y php 7, prueba unitaria HHVM y alta nivel de abstracción

### 2.2.8 Lenguaje de programación

#### Php

Php acrónimo de Hypertext Preprocessor, como dice Mehdi es un lenguaje de secuencias de comandos de propósito general de código abierto diseñado para el desarrollo web y se puede incrustar en páginas HTML. Su sintaxis utiliza C, Java y Perl, lo que facilita su aprendizaje. El objetivo principal de este lenguaje es permitir

---

<sup>57</sup> J.C, Datos. Introducción a los sistemas de base de datos. 7 ed.

<sup>58</sup> LATORRE, Marino. Historia del web, 1.0, 2.0, 3.0 y 4.0. Recuperado de Universidad Marcelino Champagnat. pdf, 2018.

<sup>59</sup> Laravel, un framework de PHP, WHITEPAPERS

a los desarrolladores web escribir páginas web generadas dinámicamente y rápidamente, aunque se puede hacer mucho más con PHP.

Segundo en el documento Análisis comparativo entre ASP.NET y PHP dentro de las ventajas de PHP se mencionarán las siguientes

- “Es un lenguaje sencillo y fácil de estudiar y aprender, una de sus características es la rapidez.
- Lo soportan la mayoría de las plataformas de alojamiento web.
- Tiene ciertas características de los lenguajes orientados a objetos como la utilización de clases y herencias
- Puede manejar ficheros y conectarse a distintas bases de datos (MySQL, Oracle, SQL Server, Informix, PostgreSQL, etcétera).
- El software que permite soportarlo en los servidores de hosting es libre y gratuito
- Está en continuo desarrollo y soporta numerosas funcionalidades.”<sup>60</sup>

## JavaScript

Como indica Pérez Javier.<sup>61</sup> JavaScript es un lenguaje de programación utilizado principalmente para crear páginas web dinámicas. Una página web dinámica es una página web que contiene efectos, como texto y desapariciones, animaciones, acciones activadas al pulsar botones y mensaje de advertencia al usuario. JavaScript es técnicamente un lenguaje de programación interpretado, por lo que no es necesario compilar los programas para ejecutarlos. En otras palabras, el programa escrito en JavaScript se puede probar directamente en cualquier navegador sin que se requiera un proceso intermedio.

Segundo JavaScript. Innovación y Cualificación; 2012. Dentro de las ventajas están las siguientes<sup>62</sup>

- Velocidad: JavaScript tiende a ser muy rápido porque a menudo se ejecuta de inmediato en el navegador. Por lo tanto, no se permite que JavaScript se

---

<sup>60</sup> MINA, Marcos Antonio Espinoza; CEDEÑO, Adriana Yolanda Sierra. Análisis comparativo entre ASP.NET y PHP. INNOVA Research Journal, 2018, vol. 3, no 4, p. 25-43

<sup>61</sup> PÉREZ, Javier Eguíluz. introducción a JavaScript. 2019.

<sup>62</sup> MAZA, Miguel Ángel Sánchez. JavaScript. Innovación Y Cualificación, 2012.

retrase por las llamadas del servidor back-end siempre que no se requieran recursos externos.

- **Simplicidad:** La sintaxis de JavaScript está inspirada por Java y es relativamente sencillo de aprender comparado a otros lenguajes de programación populares como C++.
- **Popularidad:** JavaScript esta por todas partes de la web, y con la llegada de Node.js, se ha incrementado su uso en backend. Hay incontables recursos para aprender JavaScript. Tanto StackOverflow como GitHub muestran un creciente número de proyectos que usan JavaScript, y la popularidad que ha alcanzado en los recientes años se espera que siga creciendo.
- **Server Load:** JavaScript es client-side, entonces esto reduce la demanda de servidores en general, y las aplicaciones sencillas puede que no necesiten el servidor del todo.
- **Interfaces sencillas:** JavaScript puede ser usado para crear características como arrastrar y soltar, y componentes tales como las diapositivas, lo cual mejora enormemente la interfaz de usuario y la experiencia del sitio.
- **Funcionalidad extendida:** Los desarrolladores pueden extender la funcionalidad de las páginas web mediante fragmentos de JavaScript para un tercer grupo de extensiones como Greasemonkey.
- **Versatilidad:** Muchos métodos para usar JavaScript mediante servidores Node.js. Si tú estás en Bootstrap Node.js con Express, puedes usar un documento database como MongoDB, y usar JavaScript en el frontend para clientes. Es posible desarrollar una aplicación completa de JavaScript desde principio a fin solo usando JavaScript.

### 2.2.9 Servidor web

#### Xamp

Es un servidor de plataforma libre, es un <sup>63</sup>software que integra en una sola aplicación, un servidor web Apache, intérpretes de lenguaje de scripts PHP, un

---

<sup>63</sup> BOU, Ramón Carrión. Usando XAMPP con Bootstrap y WordPress. Mercedes Gómez Alcalá, 2019.

servidor de base de datos MySQL, un servidor de FTP FileZilla, el popular administrador de base de datos escrito en PHP, MySQL, entre otros módulos. Está disponible en 11 idiomas y es compatible con diferentes plataformas, como Windows, macOS y Linux. Es una plataforma que proporciona un entorno adecuado para probar y verificar el funcionamiento de proyectos basados en Apache, Perl, base de datos MYSQL y PHP a través del sistema del propio host.

Segundo FRIENDS, Apache. XAMPP las ventajas son las siguientes:

- “Es un software multiplataforma disponible para todo tipo de sistemas operativos como Linux y Windows.
- Posee muchos otros módulos esenciales como phpMyAdmin, OpenSSL, Media Wiki, WordPress, Joomla y más.
- Viene en una versión estándar y una versión completa.
- El usuario puede iniciar y finalizar toda la pila del servidor web más la base de datos con un solo comando”<sup>64</sup>

## 2.2.10 Minería de datos

### Analítica de datos

La analítica de datos tiene como objetivo hacer <sup>65</sup>comprender al consumidor la manera de pensar y analizar. La variación de las variables cualitativas y cuantitativas en datos capaces de procesar ha sido el pilar principal de la evolución digital que se vive en el mundo del negocio en la actualidad.

Ortiz, Joyanes y Giraldo <sup>77</sup>afirman que es responsabilidad de las organizaciones no sólo dedicarse a recopilar los datos de las principales fuentes, sino procesarlos e interpretarlos de forma correcta. La analítica de datos es definida como la

---

<sup>64</sup> FRIENDS, Apache. XAMPP Apache+ MariaDB+ PHP+ Perl. Retrieved april, 2019, vol. 5.

<sup>65</sup> TREVIÑO-REYES, Rosalba; RIVERA-RODRÍGUEZ, F.; GARZA-ALONSO, J. La analítica de datos como ventaja competitiva en las organizaciones. *VINCULATEGICA EFAN*, 2020, vol. 6, no 2 <sup>77</sup> Ortiz, M. D., Joyanes, L., y Giraldo, L.M. Los desafíos del marketing en la era del big data. *E Ciencias de la Información*, 6(1), 2016,16-45

recolección a través de las principales fuentes de información de los datos más importantes para ser procesados, teniendo su impacto directo en la forma en que se toman las decisiones estratégicas dentro de la organización, su forma de operar en el mercado y su productividad; y manteniendo como principal filosofía la generación de valor a través de productos y servicios, así como la entera satisfacción del consumidor final.

Hablar de análisis de datos es incluir en la discusión el Big Data. McKinsey Global Institute citado en Ortiz, Joyanes y Giraldo menciona que el Big Data se refiere a los datos cuyo tamaño está más allá de las capacidades con que cuentan los softwares que hoy utilizan las organizaciones para almacenar, organizar y procesar todos los datos que son recopilados día con día.

Para Laynez el término Big Data <sup>66</sup>hace referencia a un grupo masivo de datos acumulados que son procesados a través de diversas herramientas que ayudan a transformar y hacer más amena la comprensión de todos ellos, extrayendo su información más relevante. Laynez enlista las principales ventajas competitivas:

- La mejora de la gestión debido al mejor entendimiento y funcionamiento a través de datos reales
- La velocidad de la toma de decisiones gracias a la facilidad con que se accede a los datos necesarios para tomar una decisión importante.
- La innovación en los productos y procesos que tiene a través del análisis de datos

### 2.2.11 Metodología CRISP-DM

CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining, es <sup>67</sup>un método probado para orientar sus trabajos de minería de datos. El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases.

proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los modelos de ciclo de vida de desarrollo de software. El modelo CRISP-DM

---

<sup>66</sup> Laynez, M. Estudio del Big Data. Tesis de Licenciatura, Universidad de Sevilla. Recuperado.2017

<sup>67</sup> Manual CRISP-DM de IBM SPSS Modeler, v15

cubre las fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas. En este nivel de descripción no es posible identificar todas las relaciones; las relaciones podrían existir entre cualquier tarea según los objetivos, el contexto, y el interés del usuario sobre los datos.

La metodología CRISP-DM contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en la elaboración de los modelos. Este contexto tiene en cuenta la existencia de un cliente que no es parte del equipo de desarrollo, así como el hecho de que el proyecto no sólo no acaba una vez se halla el modelo idóneo, sino que está relacionado con otros proyectos, y es preciso documentarlo de forma exhaustiva para que otros equipos de desarrollo utilicen el conocimiento adquirido y trabajen a partir de él.

## Fases

### Fase I. Business Understanding. Definición de necesidades del cliente (comprensión del negocio) <sup>68</sup>

El objetivo de esta fase inicial es comprender los objetivos del proyecto. Este conocimiento de datos se transforma luego en una definición de problema de minería de datos y un plan preliminar destinado a lograr el objetivo.

### Fase II. Data Understanding. Estudio y comprensión de los datos

La fase de comprensión de datos comienza con la recopilación inicial de datos, seguida de actividades para familiarizarse con los datos, identificar problemas de calidad, descubrir conocimientos preliminares sobre los datos y/o descubrir subconjuntos interesantes para formular hipótesis sobre información oculta.

### Fase III. Data Preparation. Análisis de los datos y selección de características

La fase de preparación de datos cubre todas las actividades requeridas para construir el conjunto de datos final a partir de los datos sin procesar iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para las herramientas de modelado.

---

<sup>68</sup> FERNÁNDEZ, Diego Buenaño; MORA, Sergio Luján. Uso de la metodología CRISP-DM para guiar el proceso de minería de datos en LMS. En Tecnología, innovación e investigación en los procesos de enseñanza-aprendizaje. Octaedro, 2016. p. 2385-2393.

#### Fase IV. Modeling. Modelado

En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema entre más sean es mucho mejor.

#### Fase V. Evaluation. Evaluación (obtención de resultados)

En esta etapa del proyecto, se han construido uno o más modelos que parecen tener suficiente calidad desde el punto de vista del análisis de datos. Antes de proceder al despliegue final de un modelo, es importante evaluarlo a fondo y revisar los pasos realizados para crearlo, comparando el modelo obtenido con el objetivo

#### Fase VI. Deployment. Despliegue (puesta en producción)

Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización

### 2.2.12 Herramientas para aplicación de minería de datos

#### RapidMiner

Es un programa informático para el análisis y minería de datos. <sup>69</sup>Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. Se usa en investigación y en aplicaciones empresariales.

Es sin duda el sistema de código abierto líder en el mundo para la minería de datos. Está disponible como una aplicación independiente para Análisis e integración de datos como motor de minería de datos en tu propio producto Miles de aplicaciones RapidMiner en más de 40 países Dar a sus usuarios una ventaja competitiva

Segundo RAPIDMINER. 2010. las características principales son las siguientes

- Desarrollado en Java
- Multiplataforma
- Representación interna de los procesos de análisis de datos en ficheros XML

---

<sup>69</sup> BELTRAN diego y POVEDA diego; RAPIDMINER; UNI-FO-13 V 1.0, 2010

- Permite el desarrollo de programas a través de un lenguaje de script
- Puede usarse de diversas maneras: Gui, en línea de comandos, en batch
- Extensible
- Incluye gráficos y herramientas de visualización de datos

#### 2.2.14 Estadística.

La estadística descriptiva es una <sup>70</sup> rama de la estadística que proporciona recomendaciones sobre cómo resumir de forma clara y sencilla los datos de la encuesta en forma de gráficos, tablas, figuras o gráficos. Antes de realizar un análisis descriptivo, es necesario retomar el objetivo de estudio, así como determinar las medidas de las diversas variables que se registraron en el estudio. El objetivo de las tablas o cuadros es proporcionar información puntual de los resultados. Las tendencias de los gráficos aparecen y pueden ser histogramas, representaciones en pastel, cajas con bigotes, gráficos de líneas o puntos de dispersión. Las imágenes sirven para representar ejemplos de conceptos o reforzar el suceso. La selección de un cuadro, gráfico o imagen debe basarse en los objetivos del estudio.

#### 2.2.15 Spss

Desarrollada por SPSS Inc. en 1968, la <sup>71</sup>herramienta existe desde hace más de 50 años. Posteriormente IBM adquirió SPSS (en 2009) y el nombre de la herramienta cambió a IBM SPSS Statistics. Pero nadie se refiere al software con ese nombre; todos lo llaman SPSS. SPSS es una herramienta compleja diseñada originalmente para ayudar en el análisis y la gestión de datos Hoy en día, el software tiene un uso generalizado en el mundo académico, pero los académicos no son los únicos profesionales que lo utilizan.

#### Característica

- Gestor de bases de datos para que se pueden procesar de modo dinámico los datos, permite actualizar los cambios operados y realizar informes personalizados.

---

<sup>70</sup> RENDÓN-MACÍAS, Mario Enrique; VILLASÍS-KEEVER, Miguel Ángel; MIRANDA-NOVALES, María Guadalupe. Estadística descriptiva. Revista Alergia México, 2016, vol. 63, no 4, p. 397-407.

<sup>71</sup> VISAUTA VINACUA, Bienvenido. Análisis estadístico con SPSS para Windows. *Estadística*, 1997.

- Report Builder le permite preparar informes convincentes sobre sus investigaciones al combinar el texto del informe, tablas, gráficos y resultados estadísticos en un solo archivo.
- El análisis de datos puede extraer toda la información de los archivos para realizar procedimientos estadísticos descriptivos, inferenciales y multivariantes.
- Gestor de minerías de datos para llevar a cabo búsquedas inteligentes, extraer información oculta elaborando árboles de decisión, segmentaciones de mercados o diseños de redes neuronales de inteligencia artificial

Segundo IBM SPSS Statistics sus ventajas son:

- “Fácil de utilizar: Analice datos fácilmente a través de una interfaz de usuario intuitiva sin tener que escribir código.
- Integrado con código abierto: Mejore la sintaxis de SPSS con R y Python utilizando una biblioteca de extensiones o creando la suya propia.
- Completo: Ejecute estadísticas avanzadas y descriptivas, análisis de regresión, árboles de decisiones y más con una interfaz integrada
- Flexible: Elija entre suscripción o licencias tradicionales, con múltiples opciones de funciones según lo que necesita”<sup>72</sup>.
- Permite trabajar con grandes cantidades de datos, utilizando muestras mayores e incluyendo más variables.
- Permite trasladar la atención desde las tareas mecánicas de cálculo a las tareas conceptuales: decisiones sobre el proceso, interpretación de resultados, análisis críticos.
- Es compatible con muchos de los programas para manejo de datos. Por lo tanto, permite la transmisión o traspaso de datos de un programa a otro.

### 2.2.16 Excel

Es una aplicación que permite realizar hojas de <sup>73</sup>cálculo y cuenta con un lenguaje de programación macro llamado Visual Basic, que se encuentra integrada en el conjunto ofimático de programas Microsoft Office.

---

<sup>72</sup> IBM Corp. IBM SPSS Statistics para Windows, versión 28.0.1 2020 Armonk, Nueva York: IBM Corp

<sup>73</sup> Manual Microsoft office Excel 2010

Una hoja de cálculo sirve para trabajar con números de forma sencilla e intuitiva. Para ello se utiliza una cuadrícula donde en cada celda de la cuadrícula se pueden introducir números, letras y gráficos.

Excel permite a los usuarios elaborar tablas y formatos que incluyan cálculos matemáticos mediante fórmulas; las cuales pueden usar operadores matemáticos como son: suma, resta, multiplicación, división, potenciación, mayor, menor, porcentaje, mayor o igual que, menor o igual que; además de poder utilizar elementos denominados funciones como, por ejemplo: Suma, Promedio, Buscar, etc.

### Ventajas

- Organización avanzada de datos<sup>74</sup>: Este es uno de los puntos más destacables de Excel. Tiene uno de los sistemas de análisis y organización de datos más completos para el usuario común.
- Fórmulas incorporadas: Excel ha agregado consistentemente más fórmulas integradas a medida que los usuarios lo han solicitado. Estas fórmulas no son de su propiedad, ya que son solo una adaptación de conocimiento matemático
- Soporte: Para aprender sobre funciones avanzadas o solucionar dudas, se puede consultar fácilmente el centro de ayuda o acceder a cualquiera de los foros en línea que pueden ayudarle.
- Funcionalidad de gráficos: Los gráficos de Excel son muy fáciles de usar en comparación con otros sistemas de información, y tienen varias funciones útiles integradas.
- Aparentemente Gratuito: Se incluye como parte de Microsoft Office, una suite que se ha convertido en un estándar para todas las organizaciones y que la mayoría de las personas tienen en su computadora de forma predeterminada

---

<sup>74</sup> ALVARADO claudia victoria, Ventajas y desventajas de usar Excel para reportar resultados de la estrategia. Pensemos. 2021

## 2.2.17 Regresión lineal múltiple

En su documento Montero<sup>75</sup> afirma que la regresión lineal múltiple se centra en ajustar modelos lineales o linealizables que establecen relaciones entre una variable dependiente y varias variables independientes. En este tipo de modelos, es fundamental examinar la presencia de heterocedasticidad, multicolinealidad y la precisión de la especificación del modelo. En el análisis de regresión, se calculan promedios y desviaciones típicas para interpretar conjuntamente los resultados. Se evalúa si los valores son altos o bajos, tanto en términos absolutos como en relación a una distribución normal. Además, la regresión lineal múltiple, trabaja con una variable dependiente y varias variables independientes. Se analiza el fenómeno de forma individualizada, aunque en casos raros puede haber múltiples variables dependientes, lo cual presenta dificultades en su resolución

### VARIABLES CONTINUAS

Las variables continuas son aquellas que toman valores numéricos en un rango amplio y se pueden expresar con números reales, incluso si su rango no cubre todos los números desde menos infinito hasta más infinito. Estas variables suelen ser de naturaleza cuantitativa, como peso, edad, notas académicas, donde los valores pueden variar de manera gradual. Sin embargo, también es posible considerar variables cualitativas continuas cuando se pueden ordenar y tienen un número suficientemente grande de elementos, generalmente más de siete, lo que permite analizar y comparar las diferencias entre ellos

### HIPÓTESIS Y ESTIMACIÓN

Para conocer los resultados de la regresión sean confiables es necesario que:

La relación entre las variables debe ser lineal, lo cual no implica necesariamente que deba ser una línea recta, sino que también puede ser lineal mediante alguna transformación.

---

<sup>75</sup> GRANADOS, R. Montero. Modelos de regresión lineal múltiple. *Granada, España: Departamento de Economía Aplicada, Universidad de Granada*, 2016.

## 2.3 Variables de estudio

### 2.3.1 variable dependiente

- Sitio web

### 2.3.2 variables independientes

- Comunicación oral y escrita
- Matemáticas básicas
- Matemáticas discretas
- Calculo diferencial
- Matemáticas discretas 2
- Algebra lineal
- Calculo integral

## 2.4 Definición nominal de las variables

Sitio web:

Elemento fundamental en el actual proyecto donde se encuentran plasmadas los resultados de la investigación, en donde se agrega información referente a los estudiantes que se encuentran cursando el primero, segundo o tercer semestre del programa de ingeniería de sistemas, donde por medio del modelo de regresión lineal múltiple se puede predecir si el estudiante deserta o no mediante el ingreso de datos en las variables anteriormente mencionadas.

Comunicación oral y escrita:

Dados los resultados de la investigación se da por concluido que una de las materias importantes junto al área de la formación básica de las matemáticas es la materia de la comunicación oral y escrita que según la malla curricular del programa de ingeniería de sistemas se la matricula en el primer semestre del programa y que nos brinda una de las correlaciones más certeras a la hora de realizar el análisis para obtener los resultados de las probabilidades de desertar del programa.

Matemáticas básicas y algebra lineal:

Dentro del área de formación básica se encuentran todas las materias de la matemática que es fundamental dentro de la formación como ingeniero, el cual

dichos elementos se convierten en materias base para obtener el conocimiento adecuado en todas las ramas que tienen las matemáticas dentro del programa de ingeniería de sistemas de la universidad CESMAG plasmado en su malla curricular

Matemáticas Discretas 1 y 2:

Después de realizar los respectivos filtros de las bases de datos que fueron facilitadas por parte de la oficina de software, se escogió el modelo matemático de la regresión lineal, donde de acuerdo con los estudios, la correlación, el error estándar y la parte de la significancia eran aproximada a cero, el cual nos demuestra que se podía trabajar y seguir los estudios con este modelo, donde la materia de matemáticas discretas estaba dentro del modelo que automáticamente arrojó el software con el cual se está trabajando como IBM SPSS

Calculo Diferencial e Integral:

Variable que se encuentra en el área de formación básica, donde están todas las materias de matemáticas que se estudian en el programa de ingeniería de sistemas según la malla curricular que maneja la universidad CESMAG, esta variable también hizo parte del análisis que se realizó y que por su significancia que se aproximó a cero fue arrojada por el software que se está trabajando.

**Tabla 1. Área de formación profesional**

**AREA FORMACION BASICA**

Calculo		Matemáticas básicas	
0-1	Bajo	0-1	Bajo
1-3	Regular	1-3	Regular
3-4	Bueno	3-4	Bueno
4-5	Alto	4-5	Alto

Algebra		Matemáticas discretas	
0-1	Bajo	0-1	Bajo
1-3	Regular	1-3	Regular
3-4	Bueno	3-4	Bueno
4-5	Alto	4-5	Alto

Fuente: investigación propia

Descripción:

Tabla de rendimiento académico en las variables de estudio

## 2.5 Definición operativa de las variables

Se utilizará como <sup>76</sup>“métrica de calidad la norma ISO 25010”, el cual tiene como objetivo medir el producto con los estándares de calidad para su funcionamiento, se va medir en una escala de 0 a 100 según su calificación a continuación se presenta el grafico:

**Tabla 2 Métricas de calidad ISO 25010**

Compatibilidad		Usabilidad	
0 - 20	Baja compatibilidad	0 - 20	Usabilidad Baja
20 - 60	Normal compatibilidad	20 - 60	Usabilidad Normal
60 - 80	Excelente compatibilidad	60 - 80	Usabilidad Excelente
80 - 100	compatibilidad Muy excelente	80 - 100	Usabilidad Muy excelente
Seguridad		Fiabilidad	
0 - 20	Seguridad Baja	0 - 20	Fiabilidad Baja
20 - 60	Seguridad Normal	20 - 60	Fiabilidad Normal
60 - 80	Seguridad Excelente	60 - 80	Fiabilidad Excelente
80 - 100	Seguridad Muy excelente	80 - 100	Fiabilidad Muy excelente
Portabilidad		Mantenibilidad	
0 - 20	Portabilidad Baja	0 - 20	Mantenibilidad Baja
20 - 60	Portabilidad Normal	20 - 60	Mantenibilidad Normal
60 - 80	Portabilidad Excelente	60 - 80	Mantenibilidad Excelente
80 - 100	Portabilidad Muy excelente	80 - 100	Mantenibilidad Muy excelente

Calculamos cada una de las variables independientes con las métricas de calidad de la norma ISO 25010

## 2.6 Formulación de hipótesis

### 2.6.1 hipótesis de investigación

---

<sup>76</sup> BLAS, María Julia; GONNET, Silvio M.; LEONE, Horacio P. Especificación de la Calidad en Software-as-a-Service: Definición de un Esquema de Calidad basado en el Estándar ISO/IEC 25010. En Simposio Argentino de Ingeniería de Software (ASSE 2016)-JAIIO 45 (Tres de Febrero, 2016). 2016.

Los procesos de minería de datos permiten elaborar modelos predictivos para determinar deserción estudiantil del programa de Ingeniería de sistemas de la Universidad CESMAG durante los tres primeros semestres.

#### 2.6.2 hipótesis nula

Los procesos de minería de datos no permiten elaborar modelos predictivos precisos para determinar la deserción estudiantil del programa de Ingeniería de Sistemas de la Universidad CESMAG durante los tres primeros semestres.

#### 2.6.3 hipótesis alterna

Los procesos de minería de datos sí permiten elaborar modelos predictivos precisos para determinar la deserción estudiantil del programa de Ingeniería de Sistemas de la Universidad CESMAG durante los tres primeros semestres.

### 3.METODOLOGIA DE LA INVESTIGACION

#### 3.1 Paradigma

Rico C indica que el “paradigma positivista se califica de cuantitativo, empírico analítico, racionalista, sistemático gerencial y científico tecnológico<sup>77</sup>”. Por lo tanto, el paradigma positivista se destaca por su énfasis en la medición cuantitativa de los fenómenos observables y la recolección de datos empíricos. Además, se enfoca en el análisis detallado de los elementos individuales para comprender su relación y contribución al todo. También se basa en la razón y la lógica como medios para obtener conocimiento válido. El paradigma positivista adopta un enfoque sistemático en la gestión y toma de decisiones, utilizando métodos científicos y tecnológicos para resolver problemas.

Por consiguiente, el paradigma positivista de acuerdo a Hernández, Fernández y Baptista <sup>78</sup> . El paradigma positivista encuentra sus fundamentos en los planteamientos formulados por Augusto Comte, específicamente en su obra centrada en el concepto del espíritu positivo. A partir de estas concepciones filosóficas, el paradigma positivista ha evolucionado y se ha consolidado como una perspectiva dominante en el campo de la investigación científica.

Por lo tanto, este proyecto se enmarca dentro del enfoque positivista debido a su orientación metodológica principalmente cuantitativa. El objetivo es investigar la deserción estudiantil a través del análisis del rendimiento académico. El enfoque cuantitativo permite recopilar y analizar datos numéricos, lo que facilita el estudio y la comprensión de patrones y relaciones estadísticas. En este proyecto, se utilizarán técnicas estadísticas para examinar la relación entre el rendimiento académico y la deserción estudiantil, lo que proporcionará información objetiva y cuantificable.

#### 3.2 Enfoqué

Metodológicamente, el enfoque cuantitativo de investigación se <sup>79</sup>caracteriza por privilegiar la lógica empírico-deductiva, se utiliza para contestar preguntas de

---

<sup>77</sup> Ricoy, C. Contribución sobre los paradigmas de investigación. Revista do Centro de Educación, 31 2006, 11-22

<sup>78</sup> HERNÁNDEZ Roberto, FERNÁNDEZ Carlos, BAPTISTA, Pilar. Metodología de la Investigación. México D.F, Interamericana Editores, S.A. de C.V. 2010, p.56.

<sup>79</sup> Hernández Sampieri y otros. Metodología de la investigación. McGraw-Hill Interamericana.

investigación y probar hipótesis previamente para establecer con exactitud patrones del comportamiento de la población. A partir de procedimientos rigurosos, métodos experimentales y el uso de técnicas de recolección de datos estadísticos.

Además, Bonilla y Rodríguez<sup>80</sup> sostiene que el enfoque cuantitativo en la investigación se apoya en los principios del positivismo. Este enfoque sostiene que la ciencia persigue la búsqueda de consenso, empleando un método unificado similar al utilizado en las ciencias exactas y naturales.

Basándose en lo mencionado anteriormente, la presente investigación se enmarca dentro de un enfoque cuantitativo. Este enfoque se caracteriza por la recopilación y análisis de datos numéricos con el objetivo de obtener conclusiones objetivas y generalizables.

### 3.3 Método

El método utilizado en esta investigación es el método científico, el cual cumple con las características mencionadas de manera conceptual en el libro guía de investigación cuantitativa de Armando José Quijano<sup>81</sup> afirma que el enfoque positivista se caracteriza por considerar los datos o representaciones del objeto de estudio como elementos que pueden ser abordados desde una perspectiva medible y cuantificable

### 3.4 Tipo de investigación

El presente estudio se enmarca dentro de una investigación descriptiva, cuyo propósito principal es medir con precisión las variables de interés para realizar una descripción detallada del objeto de estudio. En este caso, se busca analizar el rendimiento académico de los estudiantes y desarrollar modelos de predicción con el objetivo de presentar los resultados a través de una aplicación web. De esta manera, se cumple con la finalidad del proyecto al proporcionar información precisa y útil relacionada con el rendimiento académico de los estudiantes.

---

México, D.F., 2003

<sup>80</sup> BONILLA Elsy y RODRÍGUEZ Penélope. Más allá del dilema de los métodos. La investigación en Ciencias Sociales. 3ra. Ed. Ediciones Uniandes. Bogotá. Colombia, 1997. p.21.

<sup>81</sup> Quijano Vodniza, Armando José. Guía de Investigación Cuantitativa. [en línea], 2009, pág. 76, p.4

### 3.5 Diseño investigación

El diseño utilizado en este estudio se clasifica como pre-experimental, ya que se trabajó exclusivamente con un grupo específico de estudiantes. Este diseño nos permite realizar una preprueba y una post-prueba, con el objetivo de obtener resultados cuantitativos que posteriormente serán analizados y discutidos de manera minuciosa.

*O1 = pre – prueba grupo experimental*

*O2 = post prueba grupo experimental*

*G1 = grupo experimental*

*X = tratamiento experimental*

### 3.6. Población

Para Suárez <sup>82</sup> la población hace referencia a un conjunto de individuos que desempeñan un papel fundamental en relación a la pregunta de investigación o al objetivo que se busca alcanzar. Específicamente, la población o universo de estudio se refiere a la totalidad de sujetos que serán analizados durante la investigación. Esta población comprende a todos los participantes que serán considerados en el estudio, sin excluir a ningún sujeto de la muestra.

Para la investigación la población son los estudiantes de la Universidad CESMAG del programa de ingeniería de sistemas de los primeros 3 semestres correspondiente al periodo B-2017 hasta el A-2022

### 3.7 Muestra

De acuerdo Hernández, Fernández y Batista <sup>83</sup>, Desde una perspectiva cuantitativa, se plantea que la muestra se define como un subgrupo representativo de la población que tiene como objetivo proporcionar datos e información que reflejen con precisión a toda la población. Se resalta la importancia de establecer

---

<sup>82</sup> SUÁREZ Patricio. Población de estudio y muestra. Curso de Metodología de la Investigación Unidad Docente de MF y C. 2011, p.35.

<sup>83</sup> HERNÁNDEZ, Roberto, FERNÁNDEZ, Carlos y BAPTISTA, Pilar. Metodología de la Investigación. Mac Graw Hill. Bogotá. Colombia, 2004, p.65.

de manera precisa y anticipada los criterios de selección de la muestra. En el contexto específico de la depuración de datos durante la segunda fase de la metodología CRISP-DM, se realizaron eliminaciones de registros inconsistentes, lo cual condujo a la obtención de una muestra final consistente en 354 registros. (Ver capítulo 4)

### 3.8 Técnicas de recolección de información

El análisis detallado de registros históricos es una técnica que se utiliza para examinar y estudiar conjuntos de datos previamente recopilados con el fin de obtener información y extraer conocimientos significativos. Este tipo de análisis implica la aplicación de métodos estadísticos, técnicas de minería de datos o enfoques de visualización de datos para explorar los datos existentes y descubrir información valiosa. Para Useche<sup>84</sup>. Los datos son representaciones simbólicas, numéricas o alfabéticas que contienen información significativa. Estos datos son el resultado de un proceso cuidadoso de registro y recopilación de información relevante. Posteriormente, se empleará un cuestionario con el propósito de recopilar datos e información relevante para el avance de la investigación.

En el contexto de la investigación, se aplica la técnica de análisis detallado de los registros históricos alojados en la oficina de desarrollo de software de la Universidad CESMAG. Mediante esta metodología, se obtendrá información relevante de los estudiantes, como registros de notas y resultados de las pruebas Saber 11.

Además, se utilizará un cuestionario con el objetivo de recopilar información adicional que contribuirá al progreso de la investigación

### 3.9 Validez de las técnicas de recolección de la información

La validez de la técnica de información es viable porque son fuentes de datos escritos que corresponden a la Universidad CESMAG y están depositados en la oficina de desarrollo de software, son datos actualizables en el tiempo y tratables. Lo que genera que toda esta información que nos suministra estos repositorios es confiable y a su vez tiene una viabilidad alta

---

<sup>84</sup> USECHE, María Cristina, et al. Técnicas e instrumentos de recolección de datos cuali-cuantitativos. 2019

### 3.10 Confiabilidad de las técnicas de recolección

Los registros obtenidos en la revisión de los repositorios son altamente confiables debido a su naturaleza como datos escritos provenientes de un repositorio extraído de la base de datos de la oficina de desarrollo de software. Estos registros son actualizables en tiempo real, lo que garantiza que la información se mantenga al día y precisa. Además, presentan una notable estabilidad y son fácilmente manipulables, lo que facilita su tratamiento y análisis. Esta combinación de características proporciona un sólido respaldo a la confiabilidad de los datos obtenidos, generando así una base sólida para el trabajo y la manipulación de las variables principales

### 3.11 Instrumentos de recolección de información

De acuerdo con Hernández <sup>85</sup>. El término instrumento de recolección de información se refiere a las diversas técnicas y herramientas utilizadas por los analistas para obtener datos y desarrollar sistemas de información de manera eficiente. Estas técnicas incluyen la realización de entrevistas personalizadas con los actores involucrados, la aplicación de encuestas y cuestionarios estructurados para recopilar datos cuantitativos, el análisis detallado de registros históricos y documentos relevantes, así como la observación directa de los procesos en estudio.

Sin embargo, dataset<sup>86</sup>, también conocido como conjunto de datos, es una colección estructurada de información que se organiza en forma tabular. En un dataset, cada columna representa una variable específica y cada fila corresponde a una entidad o miembro del conjunto de datos. Estos conjuntos de datos pueden contener valores asociados de un objeto. Además, un dataset puede contener información para uno o varios miembros, y puede haber relaciones entre diferentes tablas que componen el conjunto de datos

No obstante, para Hincapié M<sup>87</sup> las listas de verificación, también conocidas como cheks-list, son un conjunto de preguntas en forma de cuestionario que abordan temas relevantes. Su aplicación proporciona información específica para evaluar

---

<sup>85</sup> HERNÁNDEZ SAMPIERI, R., FERNÁNDEZ COLLADO, C., & BAPTISTA Lucio, P. Metodología de la Investigación (6 ed.). México: Mc Graw Hill, 2014. p. 569

<sup>86</sup> Deusto formación, Blog Programación y TIC [sitio web]. [consultado: abril 15 de 2022]. Disponible en: <https://www.deustoformacion.com/blog/programacion-tic/que-son-datasets-dataframes-big-data>

<sup>87</sup> HINCAPIÉ MORENO, Catalina; WILCHES LÓPEZ, Laura Esther. Diseño de listas de chequeo para la aplicación de las guías de buenas prácticas de seguridad del paciente. 2016.

la usabilidad, confiabilidad y seguridad de un sistema de información. Casal<sup>88</sup> afirma que las listas de verificación consisten en un conjunto de pautas en forma de cuestionario que deben ser tenidas en cuenta desde el inicio de un proyecto. Esta perspectiva resalta que las listas de verificación siguen siendo relevantes tanto en las etapas iniciales como en etapas posteriores, incluso después. Es fundamental destacar que estas herramientas resultan muy útiles para evaluar tanto la eficacia como el cumplimiento

Al utilizar estos instrumentos, los analistas pueden obtener información valiosa que les permita comprender las necesidades, requerimientos y problemas del sistema en desarrollo. La combinación de estas técnicas les brinda una visión completa y precisa de la situación, lo que facilita la toma de decisiones informadas y la creación de sistemas de información efectivos y adaptados a las necesidades del usuario

La investigación se llevará a cabo mediante el análisis detallado de los registros históricos de la Universidad CESMAG como herramienta principal. El enfoque estará centrado en el análisis de variables clave relacionadas con el departamento de ciencias básicas y comunicación oral y escrita durante los primeros tres semestres del programa de Ingeniería de Sistemas.

Además, se examinarán indicadores importantes como el puntaje obtenido en la prueba saber 11 en razonamiento cuantitativo y lectura crítica, los promedios de notas, las calificaciones por cada corte y cualquier otro indicador relevante que proporcione una visión completa del rendimiento académico de los estudiantes. Estos análisis abarcarán los periodos comprendidos desde 2017B hasta 2022A

---

<sup>88</sup> Martín Casal, Miguel "la Técnica de las Chesklisen" – "curso de Técnica Legislativa" Grupo de Estudios de Técnica legislativa", Serie de Técnica Legislativa I, Madrid, Centro de estudios Constitucionales, Madrid , 1989, pag. 255.

## 4. RESULTADOS DE INVESTIGACIÓN.

### 4.1 ANALIZAR LA INFORMACIÓN DE LOS REGISTROS DE NOTAS

Durante el periodo académico 2017B-2022A, se llevó a cabo un minucioso análisis de los repositorios de la Universidad CESMAG, con el objetivo de obtener información valiosa sobre el desempeño académico de los estudiantes. Este análisis se dividió en dos categorías distintas, cada una enfocada en aspectos específicos.

La primera categoría se centra en el repositorio de resultados académicos, el cual se enfoca en las asignaturas de comunicación oral y escrita, así como en las materias del departamento de ciencias básicas hasta el tercer semestre. Estas asignaturas incluyen Matemáticas Básicas, Cálculo Diferencial, Matemáticas Discretas 1 y 2, Álgebra Lineal y Cálculo Integral. Mediante el análisis de este repositorio, se buscaba comprender el rendimiento de los estudiantes en estas asignaturas fundamentales

Por otro lado, el segundo repositorio se ha dedicado a recopilar las notas de las pruebas Saber 11. Esta prueba nacional es ampliamente reconocida y utilizada como una medida del rendimiento académico de los estudiantes a nivel nacional. En este caso, se ha llevado a cabo una clasificación de los puntajes de razonamiento cuantitativo y lectura crítica. El análisis de este repositorio proporciona información valiosa sobre las habilidades de razonamiento y comprensión lectora de los estudiantes, lo que puede ayudar a identificar fortalezas y debilidades en estos aspectos cruciales del aprendizaje.

En conclusión, el exhaustivo análisis de los repositorios de la Universidad CESMAG durante el periodo académico 2017B-2022A ha permitido obtener información detallada sobre el desempeño académico de los estudiantes en asignaturas específicas y en las pruebas Saber 11. Este enfoque de análisis proporciona una base sólida para la identificación de áreas de mejora y la implementación de estrategias educativas efectivas en beneficio de los estudiantes y su desarrollo académico.

## 4.2 PROCESOS DE MINERÍA DE DATOS Y METODOLOGÍA CRISP-DM

### Fase 1: Compresión del negocio

El objetivo principal del proyecto es utilizar la minería de datos para predecir la deserción estudiantil durante los primeros 3 semestres del programa de ingeniería de sistemas en la Universidad CESMAG. Para lograrlo, se desarrollará un modelo predictivo utilizando los datos disponibles. El propósito final es implementar un sistema web que permita identificar cuándo un estudiante corre el riesgo de desertar en función de su historial de calificaciones

### Fase 2: Estudio en compresión de los datos

En este proyecto, se lleva a cabo un análisis exhaustivo del repositorio de notas académicas desde el período 2017-B hasta 2022-A. Los registros son procesados y clasificados según el año y las materias correspondientes. Además, se crea una nueva base de datos denominada BD\_MATE para migrar los registros obtenidos del repositorio original proporcionado por la Universidad.

BD\_MATE incluye un total de 950 registros de notas de estudiantes, distribuidos con variables como identificación, jornada, sexo, código, matemáticas básicas, matemáticas discretas, cálculo diferencial, matemáticas discretas 2, álgebra lineal y cálculo integral. Se migran todas las notas, incluyendo los cortes parciales (corte 1, corte 2, corte 3) y la nota definitiva para cada variable.

Una vez completada la migración, se aplican técnicas de procesamiento de datos con el objetivo de obtener una calidad óptima de los datos para el desarrollo de modelos predictivos. Estas técnicas permiten reducir el número de registros a 354 se excluyendo identificación, sexo y código. Esto garantiza que los datos sean más adecuados y relevantes para el análisis y la construcción de los modelos predictivos.

Se ha desarrollado una segunda base de datos denominada BD\_LEC, que almacena 950 registros de notas de estudiantes con las siguientes variables: identificación, jornada, sexo, código, comunicación oral en los cortes 1, 2 y 3 y la nota definitiva. Al igual que en el caso anterior, se aplica proceso de procesamiento de datos con el objetivo de obtener datos de calidad y lograr los mejores modelos predictivos. Como resultado de este proceso, se extraen 354 registros y se excluye identificación, sexo y código

En la base de datos RES\_SAB11 se encuentran almacenados los registros de las pruebas Saber 11 de los estudiantes, incluyendo los puntajes obtenidos en las áreas de razonamiento cuantitativo y lectura crítica. Para el procesamiento de estos datos, se aplicó una fórmula matemática que permite ajustar los puntajes, los cuales se encuentran en un rango de 0 a 100.

$$x = \frac{\text{puntaje prueba saber } 11 * 5}{100}$$

Con esta fórmula toman valores de 0 a 5 para garantizar datos con coeficientes iguales

#### Fase 3: preparación de datos

En esta etapa del proceso, una vez que se han procesado los datos en las tres bases de datos, se procede a unificar la variable "razonamiento cuantitativo" en los registros de la base de datos BD\_MATE. El objetivo principal es utilizar esta variable como dependiente para simular un proceso de predicción y obtener un dato final.

Por otra parte, en la BD\_LEC se unifica la variable "lectura crítica" como variable dependiente para simular también un proceso de predicción y obtener un dato final.

Sin embargo, es importante destacar que las variables mencionadas presentan una correlación baja entre sí en ambos casos. Para abordar esta situación, se ha optado por crear una nueva variable llamada "resultados" en cada base de datos. Esta variable se obtiene al calcular el valor final a partir de los datos de corte 1, corte 2 y corte 3 de todas las variables de estudio. (Ver Tabla 4 y 5)

Al utilizar esta nueva variable, se logra obtener una correlación significativa y alta. De esta manera, se asegura que el modelo aplicado sea capaz de generar datos objetivos, confiables y más precisos.

#### Fase 4: Modelado

Para esta etapa se desarrolló un modelo de regresión lineal múltiple con respecto al desempeño académico en el aplicativo IBM SPSS Statistics y se obtuvieron los siguientes datos

#### Correlación de Pearson

La correlación en estadística es una medida cuantitativa que permite evaluar la relación entre dos variables, indicando la dirección (positiva o negativa) y la intensidad de esa relación

**Tabla 3 Grado de correlación**

R	Grado de correlación
1	Correlación perfecta
0.80 – 0.99	Correlación muy alta
0.66 – 0.79	Correlación alta
0.40 – 0.59	Correlación moderada
0.20 – 0.39	Correlación baja
0.01 – 0.20	Correlación muy baja
0	Correlación nula

Autor: fuente propia

Descripción de la tabla: La tabla numero 3 representa el grado de correlación que existen en estadística

**Tabla 4 correlación de Pearson ciencias básicas**

\*\* La correlación es significativa en el nivel 0,01 (bilateral)

\* La correlación es significativa en el nivel 0,05 (bilateral).

		Correlaciones																			
	RES ULTA	MBC1_ S1	MBC2_ _S1	MBC3_ _S1	MDC 1_S1	MDC 2_S1	MDC3_ _S1	CDC 1_S2	CDC2_ _S2	CDC 3_S2	MD2C 1_S2	MD2C 2_S2	C3_S 2	AGC1_ _S3	AGC2_ _S3	AGC3_ _S3	CINC 1_S3	CINC 2_S3	CINC 3_S3	T_B1 1	
RESULTADO	Correlación de Pearson	1	,422**	,451**	,455**	,194**	,161**	,140**	,849**	,883**	,849**	,868**	,852**	,809**	,879**	,872**	,861**	,880**	,866**	,862**	-0,007
	Sig. (bilateral)		0,000	0,000	0,000	0,002	0,008	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,898
	N	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354

Autor: fuente propia

**Tabla 5 correlación de Pearson comunicación oral y escrita**

\*\* La correlación es significativa en el nivel 0,01 (bilateral)

\* La correlación es significativa en el nivel 0,05 (bilateral).

		<b>Correlaciones</b>			
		RESULTD	CEC1_S1	CEC2_S1	CEC3_S1
		O	X1	X2	X3
RESULTDO	Correlación de Pearson	1	,831**	,873**	,894**
	Sig. (bilateral)		,000	,000	,000
	N	354	354	354	354
CEC1_S1 X1	Correlación de Pearson	,831**	1	,671**	,591**
	Sig. (bilateral)	,000		,000	,000
	N	354	354	354	354
CEC2_S1 X2	Correlación de Pearson	,873**	,671**	1	,638**
	Sig. (bilateral)	,000	,000		,000
	N	354	354	354	354
CEC3_S1 X3	Correlación de Pearson	,894**	,591**	,638**	1
	Sig. (bilateral)	,000	,000	,000	
	N	354	354	354	354
saber11	Correlación de Pearson	-,018	,029	-,057	-,013
	Sig. (bilateral)	,733	,592	,288	,801
	N	354	354	354	354

\*\* La correlación es significativa en el nivel 0,01 (bilateral).

Autor fuente propia

Descripción de las tablas: Al realizar la correlación de Pearson, se consideran todas las variables para obtener un análisis más preciso. Sin embargo, se realiza una distinción en la variable resultado, que se trata como la variable dependiente, buscando una correlación alta con las demás variables de estudio. Al interpretar la tabla, se toman en cuenta dos aspectos importantes: la correlación de Pearson y el valor de significancia bilateral (p-value). En el gráfico, se observa la variable resultada junto con las variables de estudio, concluyendo que existe una correlación alta entre la variable resultado y las variables de estudio debido a que

los valores son mayores a 0.40. Para que el valor de significancia bilateral sea considerado válido y confiable, debe ser menor a 0.05.

En el caso del modelo de ciencias básicas, se selecciona una variable por materia siguiendo un criterio específico: se elige aquella variable que presenta un valor superior a 0.40, garantizando así una correlación moderada

### **Selección de variables**

#### ***Variables X ciencias básicas:***

MBC3\_S1 = matemáticas básicas corte 3 semestre 1  
MDC1\_S1 = matemáticas discretas corte 1 semestre 1  
CDC2\_S2 = calculo diferencial corte 2 semestre 2  
MD2C1\_S2= matemáticas discretas 2 corte 1 semestre 2  
AGC1\_S3= algebra lineal corte 1 semestre 3  
CINC1\_S3= calculo integral corte 1 semestre 3

#### ***Variable Y ciencias básicas:***

RESULTADO

#### ***Variables X comunicación oral y escrita:***

CEC1\_S1=Comunicación oral y escrita corte 1 semestre 1  
CEC2\_S1=Comunicación oral y escrita corte 2 semestre 1  
CEC3\_S1=Comunicación oral y escrita corte 3 semestre 1

#### ***Variable Y comunicación oral y escrita:***

RESULTADO

Posteriormente, se realizó la construcción del modelo de regresión lineal múltiple utilizando el software IBM SPSS Statistics. Como resultado de dicho proceso, se obtuvieron los siguientes hallazgos:

**Tabla 6: Resumen de modelos ciencias básicas**

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,889 <sup>a</sup>	,790	,789	,5526
2	,963 <sup>b</sup>	,927	,927	,3250
3	,979 <sup>c</sup>	,958	,957	,2480
4	,990 <sup>d</sup>	,981	,981	,1670
5	,995 <sup>e</sup>	,991	,991	,1165
6	1,000 <sup>f</sup>	1,000	1,000	,0000

a. Predictores: (Constante), CINC1\_S3  
b. Predictores: (Constante), CINC1\_S3, CDC2\_S2  
c. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1  
d. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1, MD2C1\_S2  
e. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1, MD2C1\_S2, AGC1\_S3  
f. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1, MD2C1\_S2, AGC1\_S3, MDC1\_S  
g. Variable dependiente: Y

Autor: fuente propia

Descripción de la tabla: En el estudio realizado, se generaron un total de 6 modelos automáticos de regresión lineal múltiple que utilizaban variables independientes ( $x$ ) y una variable dependiente ( $y$ ). Con el objetivo de asegurar la calidad y confiabilidad de los modelos, se tuvo en cuenta que el coeficiente de correlación  $R$  se acercara a 1, lo que indicaría una fuerte relación entre las variables.

En la tabla 6 se presentan los resultados obtenidos por los 6 modelos, los cuales combinan diferentes variables. En el caso del modelo número 6, se observa un coeficiente de correlación  $R$  de 1, lo que indica una relación del 100% entre las variables. Además, el coeficiente de determinación  $r^2$  también es igual a 1, lo que implica que el 100% de la variabilidad de la variable dependiente ( $y$ ) puede ser explicada por la variable independiente ( $x$ ). Asimismo, el coeficiente de determinación múltiple ajustado también es 1, lo que sugiere que la variable dependiente ( $y$ ) puede ser predicha con una exactitud del 100%.

Estos resultados demuestran que la selección de variables (consultar página 71) cumple con una correlación moderada, lo que contribuye a obtener resultados eficientes y confiables en el modelo

**Tabla 7: Resumen de modelos comunicación oral y escrita**

<b>Resumen del modelo</b>				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,894 <sup>a</sup>	,800	,799	,3427
2	,977 <sup>b</sup>	,954	,953	,1653
3	1,000 <sup>c</sup>	1,000	1,000	,0000

a. Predictores: (Constante), CEC3\_S1 X3  
b. Predictores: (Constante), CEC3\_S1 X3, CEC2\_S1 X2  
c. Predictores: (Constante), CEC3\_S1 X3, CEC2\_S1 X2, CEC1\_S1 X1  
d. Variable dependiente: RESULTADOS

Autor: fuente propia

Descripción de las tablas:

En el contexto específico de este estudio, se han desarrollado tres modelos automatizados de regresión lineal múltiple que involucran la combinación de variables X e Y. Para seleccionar el modelo más apropiado para su aplicación, se han tenido en cuenta los mismos criterios y consideraciones utilizados en el modelo de ciencias básicas, los cuales se detallan en la descripción presentada en la tabla 6.

**Tabla 8: Análisis de varianza ciencias básicas**

Modelo		Anova			F	Sig.
		Suma de cuadrados	GI	Media cuadrática		
1	Regresión	403,363	1	403,363	1320,725	,000 <sup>b</sup>
	Residuo	107,504	352	,305		
	Total	510,867	353			
2	Regresión	473,804	2	236,902	2243,500	,000 <sup>c</sup>
	Residuo	37,064	351	,106		
	Total	510,867	353			
3	Regresión	489,340	3	163,113	2651,905	,000 <sup>d</sup>
	Residuo	21,528	350	,062		
	Total	510,867	353			
4	Regresión	501,137	4	125,284	4493,441	,000 <sup>e</sup>
	Residuo	9,731	349	,028		
	Total	510,867	353			
5	Regresión	506,140	5	101,228	7452,341	,000 <sup>f</sup>
	Residuo	4,727	348	,014		
	Total	510,867	353			
6	Regresión	510,867	6	85,145	22526 112730 775616,000	,000 <sup>g</sup>
	Residuo	,000	347	,000		
	Total	510,867	353			

a. Predictores: (Constante), CINC1\_S3  
b. Predictores: (Constante), CINC1\_S3, CDC2\_S2  
c. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1  
d. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1, MD2C1\_S2  
e. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1, MD2C1\_S2, AGC1\_S3  
f. Predictores: (Constante), CINC1\_S3, CDC2\_S2, MBC3\_S1, MD2C1\_S2, AGC1\_S3, MDC1\_S

Autor: Fuente propia

**Tabla 9: análisis de varianza comunicación oral y escrita**

ANOVA <sup>a</sup>						
Modelo		Suma de cuadrados	Gl	Media cuadrática	F	Sig.
1	Regresión	165,415	1	165,415	1408,315	,000 <sup>b</sup>
	Residuo	41,345	352	,117		
	Total	206,760	353			
2	Regresión	197,166	2	98,583	3606,950	,000 <sup>c</sup>
	Residuo	9,593	351	,027		
	Total	206,760	353			
3	Regresión	206,760	3	68,920	.	, <sup>d</sup>
	Residuo	,000	350	,000		
	Total	206,760	353			

a. Variable dependiente: RESULTADOS  
b. Predictores: (Constante), CEC3\_S1 X3  
c. Predictores: (Constante), CEC3\_S1 X3, CEC2\_S1 X2  
d. Predictores: (Constante), CEC3\_S1 X3, CEC2\_S1 X2, CEC1\_S1 X1

Autor: fuente propia

Descripción de las tablas: Según anova el modelo 6 para ciencias básicas y el modelo 3 para comunicación oral y escrita es estadísticamente significativo es decir que alcanza a realizar la redición

**Tabla 10: Coeficientes ciencias básicas**

		Coeficientes				
		Coeficientes no estandarizados		Coeficientes estandarizados		
Modelo		B	Desv. Error	Beta	t	Sig.
1	(Constante)	1,627	,040		40,834	,000
	X6	,588	,016	,889	36,342	,000
2	(Constante)	1,275	,027		47,036	,000
	X6	,351	,013	,531	26,604	,000
	X3	,353	,014	,516	25,828	,000
3	(Constante)	,720	,041		17,745	,000
	X6	,353	,010	,533	35,019	,000
	X3	,305	,011	,445	28,026	,000
	X1	,182	,011	,188	15,893	,000
4	(Constante)	,612	,028		21,985	,000
	X6	,311	,007	,470	43,876	,000
	X3	,181	,009	,265	19,153	,000
	X1	,184	,008	,190	23,883	,000
	X4	,184	,009	,274	20,570	,000
5	(Constante)	,564	,020		28,843	,000
	X6	,171	,009	,258	19,296	,000
	X3	,174	,007	,254	26,318	,000
	X1	,198	,005	,204	36,510	,000
	X4	,163	,006	,242	25,670	,000
	X5	,159	,008	,259	19,193	,000
6	(Constante)	4,441E-16	,000		,000	1,000
	X6	,167	,000	,252	35737030,543	,000
	X3	,167	,000	,243	47681269,753	,000
	X1	,167	,000	,172	55602860,648	,000
	X4	,167	,000	,248	49732577,034	,000
	X5	,167	,000	,271	37981149,753	,000
	X2	,167	,000	,102	35363758,553	,000

a. Variable dependiente: Y

Autor: Fuente propia

**Tabla 11: coeficientes comunicación oral y escrita**

Coeficientes						
Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Desv. Error			
1	(Constante)	1,675	,067		25,071	,000
	CEC3_S1	,613	,016	,894	37,528	,000
	X3					
2	(Constante)	,559	,046		12,177	,000
	CEC3_S1	,391	,010	,570	38,135	,000
	X3					
	CEC2_S1 X2	,472	,014	,509	34,084	,000
3	(Constante)	1,554E-15	,000		,000	1,000
	CEC3_S1	,333	,000	,486	245013291,5	,000
	X3				25	
	CEC2_S1	,333	,000	,359	166561962,4	,000
	X2				41	
	CEC1_S1 X1	,333	,000	,303	147207171,5	,000
				81		

a. Variable dependiente: RESULTADOS

Autor: fuente propia

Descripción de las tablas: En el marco del análisis estadístico, existen diferentes modelos que permiten obtener predicciones a partir de un conjunto de datos. En particular, en el modelo 6 de ciencias básicas y el modelo 3 de comunicación oral y escrita se pueden visualizar los valores necesarios para reemplazar en la fórmula de regresión lineal múltiple. Al introducir estos valores en la fórmula, se pueden obtener predicciones precisas que permiten hacer estimaciones sobre diferentes variables de interés. Es importante destacar que este proceso se basa en la relación entre las variables incluidas en el modelo, y requiere una adecuada selección y tratamiento de los datos para obtener resultados confiables. En consecuencia, la comprensión y aplicación de este modelo puede ser de gran utilidad en diferentes campos, como la investigación científica, la planificación empresarial y la toma de decisiones en general.

Fórmula de regresión lineal múltiple:

$$y = b_0 + (b_1x_1) + (b_2x_2) + (b_3x_3) + (n_n n_n)$$

*Donde:  $B_0$  es igual a coeficientes y  $x_1$  valor de entrada de variables*

Modelo de regresión lineal

Este modelo presenta similitudes con el modelo de regresión múltiple, pero se diferencia en su ecuación al permitir únicamente una variable dependiente y una variable independiente. En el contexto de esta investigación específica, se cuenta con un total de 6 variables independientes, lo cual no resulta apropiado para el estudio e investigación en cuestión.

Modelo de clasificación

El modelo de clasificación utiliza técnicas ampliamente utilizadas en el campo del aprendizaje automático, incluyendo árboles de decisión y un clasificador bayesiano. La razón para seleccionar estos algoritmos es su sencillez y capacidad de ser interpretados fácilmente.

La técnica utilizada para la clasificación de datos en este modelo es el árbol de decisión. Se trata de un algoritmo en el que una instancia es clasificada al seguir una serie de condiciones, desde la raíz del árbol hasta llegar a una hoja. En este caso, se utiliza el algoritmo C4.5, que es uno de los más populares y versátiles para construir árboles de decisión a partir de datos.

En resumen, el árbol de decisión es una técnica eficaz y fácilmente interpretable para la clasificación de datos, y C4.5 es un algoritmo muy popular para su construcción que puede manejar atributos tanto categóricos como continuos.

Para predecir deserción estudiantil se agregó dos variables categóricas, una con respecto al desempeño académico y su jornada estudiantil

Para su evaluación del modelo se implementó la validación cruzada, se divide los registros de la base de datos en un 70% para su entrenamiento y 30% para su evaluación.

**Tabla 12 Matriz de confusión**

<b>Predicción</b>	<b>Positivos</b>	<b>Negativos</b>
<b>Positivos</b>	<b>Verdaderos positivos (VP)</b>	<b>Falsos negativos (FN)</b>
<b>Negativos</b>	<b>Falsos positivos (FP)</b>	<b>Verdaderos negativos (VN)</b>

Autor: fuente propia

Descripción

- **Exactitud:** Es la proporción del número total de predicciones que son correctas.  
**Accuracy** =  $(VP + VN) / N$ ;  $N = FN + FP + VP + VN$   
 Donde N es el total de ejemplos del conjunto de validaciones
- **Exhaustividad:** Son casos positivos que fueron clasificados correctamente  
**Recall** =  $VP / (VP + FN)$
- **Precisión:** Es la predicción de casos positivos que fueron clasificados correctamente.  
**Precisión** =  $VP / (VP + FP)$
- **Falsos Negativos:** son los casos positivos que son clasificados incorrectamente como negativos.  
**FN** =  $FN / (VP + FN)$
- **Falsos Positivos:** son los casos negativos que son clasificados incorrectamente como positivos.  
**FP** =  $FP / (VN + FP)$
- **Specificity:** son los casos negativos que fueron clasificados correctamente.  
 Especificidad =  $VN / (FP + VN)$

## Resume

Verdadero positivo (VP): El modelo predijo correctamente la clase positiva

Verdadero negativo (VN): El modelo predijo correctamente la clase negativa

Falso positivo (FN): El modelo predijo incorrectamente la clase positiva

Falso negativo (FV): El modelo predijo incorrectamente la clase negativa

### 4.3 DESARROLLO DEL SISTEMA DE INFORMACION

#### 4.3.1 Metodología SCRUM

Para el desarrollo de la aplicación, para conocer y visualizar la deserción estudiantil de la Universidad CESMAG en el programa de ingeniería de sistemas durante los 3 semestres se trabajó una metodología ágil que permite trabajar colaborativamente, en equipo. Scrum es una metodología ágil ampliamente utilizada en la gestión de proyectos, que se centra en la colaboración, la adaptabilidad y la entrega iterativa de productos. Se fundamenta en la premisa de equipos autónomos y diversos que trabajan en períodos de tiempo limitados conocidos como sprint. Durante cada sprint, se eligen y priorizan las tareas a abordar, y el equipo las aborda de forma colaborativa, llevando a cabo revisiones y ajustes de forma continua

#### 4.3.2 Fases metodología scrum

##### 4.3.2.1 Levantamiento de requerimientos

**Tabla 13 Tabla roles Scrum**

<b>Roles</b>	<b>Descripción</b>	<b>Persona</b>
<b>Product Owner</b>	Representa a los stakeholders y es responsable de gestionar el backlog del producto	Programa de Ingeniería de Sistemas
<b>Scrum Master</b>	se encarga de asegurar que se sigan los principios y prácticas de Scrum	Esteban Andrés Urbina
<b>Scrum team</b>	Diseñador, desarrollador	Luis Gabriel Carlosama, Esteban Andrés Urbina

Autor: Fuente propia

## Requerimientos funcionales

Los requerimientos funcionales del presente proyecto se describen en las siguientes tablas con sus respectivas características, las cuales están desde la

**Tabla14. RF1-Registro**

RF-0001	GESTIÓN DE AUTENTICACIÓN DE USUARIOS
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema deberá mostrar una interfaz amigable donde el usuario pueda agregar sus datos de registro e inicio de sesión</i>
Estado	Pendiente revisión
Importancia	Vital
Estabilidad	Media

Autor: Fuente propia

**Tabla15. RF2-Insertar datos**

RF-0002	REGISTRO Y GESTION DE VARIABLES
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema deberá permitir agregar datos en las diferentes variables de estudio expuestas en la interfaz de cada uno de los módulos</i>
Estado	Pendiente revisión
Importancia	Importante
Estabilidad	Media

Autor: Fuente propia

**Tabla16. RF3-Visualizacion**

RF-0003	FUNCIONALIDAD DE IMPRESIÓN DE MODELOS
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema debe visualizar los datos de entrada y de salidas dl los modelos en una ventana emergente</i>
Estado	Pendiente revisión
Importancia	Vital
Estabilidad	Alta

Autor: fuente propia

**Tabla17. RF4-Visualizacion de gráficos**

RF-0004	VISUALIZACIÓN DE DATOS
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema debe visualizar gráficos respecto a la deserción estudiantil durante el periodo de investigación</i>
Estado	Pendiente revisión
Importancia	Vital
Estabilidad	Media

Autor: Fuente propia

### Requerimientos no funcionales

A continuación, se muestra de la tabla 18 a la 21 los requerimientos no funcionales que fueron definidos para el sistema donde se analizará la predicción de la deserción o no de los estudiantes de ingeniería de sistemas de la Universidad CESMAG.

**Tabla 18. RNF1-Compatibilidad**

RNF-0001	COMPATIBILIDAD
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema debe ser compatible con los navegadores actuales de internet.</i>

Estado	Pendiente revisión
Importancia	Importante
Estabilidad	Alta

Autor: Fuente propia

**tabla 19. RNF2-Eficiencia**

<b>RNF-0002</b>	<b>EFICIENCIA</b>
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema deberá ser rápido y eficiente para que el resultado de las predicciones sean cien por ciento más seguras.</i>
Estado	Pendiente revisión
Importancia	Importante
Estabilidad	Alta

Autor: Fuente propia

**Tabla 20. RNF3-Usabilidad**

<b>RNF-0003</b>	<b>USABILIDAD</b>
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza
Descripción	<i>El sistema deberá tener una interfaz de usuario sencilla para que sea de fácil manejo para los usuarios del sistema.</i>
Estado	Pendiente revisión
Importancia	Importante
Estabilidad	Alta

Autor: Fuente propia

**Tabla 21. RNF4-Seguridad**

<b>RNF-0004</b>	<b>SEGURIDAD</b>
Versión	1.0
Autores	Estaban Andrés Urbina Bravo, Luis Gabriel Carlosama Daza

Descripción	<i>El sistema deberá resguardar la información suministrada, para brindar seguridad y privacidad de los datos.</i>
Estado	Pendiente revisión
Importancia	Importante
Estabilidad	Alta

Autor: Fuente propia

#### 4.3.2.2 Atributos de calidad

**Tabla 22. Atributo de calidad-Compatibilidad**

Atributo de calidad:		Compatibilidad	
	ID	Descripción	Prioridad
Adaptabilidad	<i>RNF-0001</i>	El sistema debe ser compatible con los navegadores actuales de internet.	<b>100</b>

Autor: Fuente propia

**Tabla 23. Atributo de calidad-Eficiencia**

Atributo de calidad:		Eficiencia	
	ID	Descripción	Prioridad
Operatividad	<i>RNF-0002</i>	El sistema deberá ser rápido y eficiente para que el resultado de las predicciones sean cien por ciento más seguras.	<b>100</b>

Autor: Fuente propia

**Tabla 24. Atributo de calidad-Interacción**

Atributo de calidad:		Interacción	
	ID	Descripción	Prioridad
Usabilidad	<i>RNF-0003</i>	El sistema deberá tener una interfaz de usuario sencilla para que sea de fácil manejo para los usuarios del sistema.	<b>100</b>

Autor: Fuente propia

**Tabla 25. Atributo de calidad-Seguridad**

Atributo de calidad:		Seguridad	
	ID	Descripción	Prioridad
Integridad	<i>RNF-0004</i>	El sistema deberá resguardar la información suministrada, para brindar seguridad y privacidad de los datos.	<b>100</b>

Autor: Fuente propia

#### 4.3.2.3 Escenarios de calidad priorizados

**Tabla 26. Escenario de calidad 1**

Escenario de calidad:	<i>RNF-0001</i>	Prioridad:	Alta
Atributo de Calidad	Compatibilidad (adaptabilidad)		
Justificación	El usuario puede utilizar el sitio web en cualquier navegador de internet ya sea (Chrome, Firefox, opera, safari)		

Fuente	Usuario
Estimulo	Tener un mejor acceso al sistema desde cualquier parte donde este, solamente es necesario tener una conexión a internet
Artefacto	Sitio web
Ambiente	Condiciones normales de funcionamiento del equipo donde se encuentre conectado como también conexión a internet
Respuesta	El sitio web permite acceder a la información solicitada
Medida de la respuesta	El sistema debe de ser compatible con los actuales navegadores de internet

Autor. Fuente propia

**Tabla 27. Escenario de calidad 2**

Escenario de calidad:	<i>RNF-0002</i>	Prioridad:	Alta
Atributo de Calidad	Eficiencia (operabilidad)		
Justificación	El usuario puede utilizar el sitio web de manera rápida y eficaz en cualquier momento		
Fuente	Usuario		
Estimulo	Analizar la información expuesta en las variables que están dentro de cada uno de los módulos		
Artefacto	Sitio web		
Ambiente	Condiciones normales de funcionamiento del equipo donde se encuentre conectado como también conexión a internet		
Respuesta	El sitio web permite analizar la información con dos modelos matemáticos como son: análisis de datos con regresión lineal y análisis de datos con el modelo de clasificación		
Medida de la respuesta	El sistema deberá ser rápido y eficiente para que el resultado de las predicciones sean cien por ciento más seguras.		

Autor: Fuente propia

**Tabla 28. Escenario de calidad 3**

<b>Escenario de calidad:</b>	<i>RNF-0003</i>	Prioridad:	Alta
Atributo de Calidad	Interacción (usabilidad)		
Justificación	El usuario puede utilizar el sitio web con facilidad ya que su interfaz es cómoda y adaptable a el usuario		
Fuente	Usuario		
Estimulo	El fácil manejo a la hora de ingresar las variables de estudios hace mejor la interacción con el usuario		
Artefacto	Sitio web		
Ambiente	Condiciones normales de funcionamiento del equipo donde se encuentre conectado como también conexión a internet		
Respuesta	El sitio web permite una interacción cómoda para el usuario		
Medida de la respuesta	El sistema deberá tener una interfaz de usuario sencilla para que sea de fácil manejo para los usuarios del sistema.		

Autor: Fuente propia

**Tabla 29. Escenario de calidad 4**

<b>Escenario de calidad:</b>	<i>RNF-0004</i>	Prioridad:	Alta
Atributo de Calidad	Seguridad (integridad)		
Justificación	El sistema conserva toda información segura y confiable		
Fuente	Usuario		
Artefacto	Sitio web		
Ambiente	Condiciones normales de funcionamiento del equipo donde se encuentre conectado como también conexión a internet		
Respuesta	La información ingresada en el sistema será resguardada y segura para el usuario		
Medida de la respuesta	La información debe de ser segura y de calidad cuando el usuario intente observarla		

Autor: Fuente propia

#### 4.3.2.4 Product Backlog

**Tabla 30 Product Backlog**

Historias de usuario	Titulo	Descripción
HU-001	Gestión registro	Yo como usuario del sistema espero realizar un registro previo para obtener una cuenta como administrador.
HU-002	Inicio de sesión	Yo como interesado en usar el sitio web, me gustaría una vez registrado, iniciar sesión con el Usuario y contraseña creados anteriormente.
HU-003	Módulo de predicción ciencias básicas	Yo como interesado en utilizar el sitio web, quisiera ingresar los valores de cada una de las variables que se utilizaran para el análisis en el área de las ciencias básicas
HU-004	Módulo de predicción comunicación oral y escrita	Yo como interesado en utilizar el sitio web, quisiera ingresar los valores de cada una de las variables que se utilizaran para el análisis en el área de la comunicación oral y escrita
HU-005	Reportes de graficas	Yo como interesado en utilizar el sitio web, quisiera obtener los resultados de los estudios de deserción en grafica

#### 4.3.2.5 Historias de usuario

Las historias de usuario que se emplearon para el presente proyecto se desarrollaron a partir de los posibles módulos que se crearon, por lo tanto, se especifica detalladamente las funcionalidades del sistema que debe de cumplir para que el proceso de deserción estudiantil se lleve con éxito y con la mayor probabilidad posible para que el software sea cien por ciento efectivo.

A continuación, se expone las historias de usuario que se obtuvieron durante el desarrollo del sitio web, las cuales se encuentran desde la tabla 31 hasta la tabla 36

**Tabla 31. HU-001 Gestión registro**

HU-001 – Gestión registro	
<b>Como:</b>	Usuario
<b>Quiero:</b>	Ingresar datos en formulario de registro en la página de inicio
<b>Para:</b>	Crear cuenta como administrador del sistema
OBSERVACIONES	
Ninguna	

Autor: Fuente propia

**Tabla 32. HU-002 Inicio de sesión**

HU-002 – Inicio de sesión	
<b>Como:</b>	Usuario
<b>Quiero:</b>	Ingresar nombre de usuario y contraseña correcta
<b>Para:</b>	Acceder al sistema e ingresar las variables principales que pueden decidir la deserción o no del estudiante
OBSERVACIONES	
Ninguna	

Autor: Fuente propia

Historias de usuario segundo Sprint

**Tabla 33. HU-00 Modelos de predicción ciencias básicas**

HU-003 – Modelos de predicción ciencias básicas	
<b>Como:</b>	Usuario
<b>Quiero:</b>	Ingresar valores de variables principales en el sistema
<b>Para:</b>	Calcular el porcentaje de predicción de deserción
OBSERVACIONES	
Las variables de resultados saber 11, año de prueba saber 11 y la jornada son variables de información para el modelo de regresión lineal múltiple	

Autor: Fuente propia

Historias de usuario tercer Sprint

**Tabla 34. HU-004 Modelos de predicción comunicación oral y escrita**

HU-004 – Modelos de predicción comunicación y escrita	
<b>Como:</b>	Usuario
<b>Quiero:</b>	Ingresar valores de variables principales en el sistema
<b>Para:</b>	Calcular el porcentaje de predicción de deserción
OBSERVACIONES	
Las variables de resultados saber 11, año de prueba saber 11 y la jornada son variables de información para el modelo de regresión lineal múltiple	

Autor: Fuente propia

Historias de usuario cuarto Sprint

**Tabla 35. HU-005 Reportes de graficas**

HU-005 – Reportes de graficas	
<b>Como:</b>	Usuario
<b>Quiero:</b>	Visualizar la información de deserción gráficamente
<b>Para:</b>	Conocer la deserción que se obtuvo durante el periodo de investigación
OBSERVACIONES	

Autor: Fuente propia

#### 4.3.2.6 Sprint plannig

Para realizar el sprint plannig se tuvo en cuenta cada uno de los workitem del proyecto, el cual se toma como referencia la estimación de puntos de las historias de usuario y las horas trabajas en cada ítem, por lo tanto, se obtiene como resultado la siguiente tabla.

**Tabla 36 Sprint plannig**

Identificador	Titulo	Puntos de historia	Tiempo de desarrollo (Horas)
HU-001	Gestión registro	2	24
HU-002	Inicio de sesión	2	20
HU-003	Modelos de predicción ciencias básicas	5	48
HU-004	Modelos de predicción comunicación oral y escrita	5	48
HU-005	Reporte de graficas	3	30
	<b>TOTAL</b>	17	170

Autor: Fuente propia

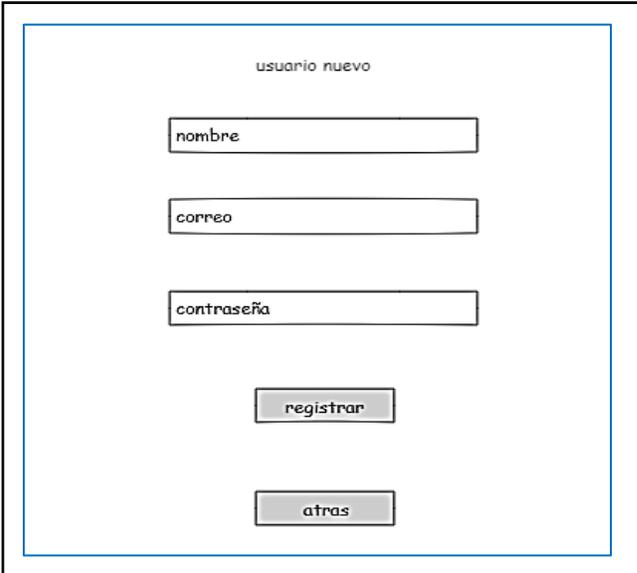
#### 4.3.2.7 Workitem

Inicialmente se establecieron unos diseños iniciales los cuales debían cumplir con los requerimientos que se establecieron por cada uno de las historias de usuario, los cuales se encuentran agrupados por modelos como el de regresión lineal y el modelo de clasificación y como resultado del diseño inicial y final las encontramos en las siguientes tablas.

Sprint 1

**Tabla 37 Mockus – 1**

MOCK-UP	
Identificador	Nombre de historia de usuario
HU-001	Gestión registro
❖ Se presenta el diseño inicial de la página principal con su respectivo mockup y el resultado que se obtiene al realizar la interfaz del sitio web a utilizar, donde se debe de ingresar nombre de usuario y contraseña correcta	

MOCK-UP	RESULTADO
 <p>The mockup shows a registration form titled 'usuario nuevo'. It contains three input fields: 'nombre', 'correo', and 'contraseña'. Below the fields are two buttons: 'registrar' and 'atras'.</p>	 <p>The screenshot shows the implemented registration page with a dark blue background. The form is titled 'Usuario nuevo' and contains three input fields: 'Nombre' (filled with 'esteban1994@hotmail.com'), a password field (filled with '****'), and two buttons: 'Registrar' and 'Atras'.</p>

Autor: Fuente propia  
Sprint Review Sprint1

Entrega:	Esteban Andres Urbina Bravo – Luis Gabriel Carlosama Daza
Actividades cumplidas:	Se cumplió satisfactoriamente con la historia de usuario 001

Sprint 2

**Tabla 38. MockUp – 2**

MOCK-UP	
Identificador	Nombre de historia de usuario
HU-002	Inicio de sesión
❖ Se presenta el diseño inicial y final de la interfaz del inicio de sesión donde el usuario ya debe de estar registrado y debe de ingresar al sistema con las credenciales registradas	

MOCK-UP	RESULTADO
	

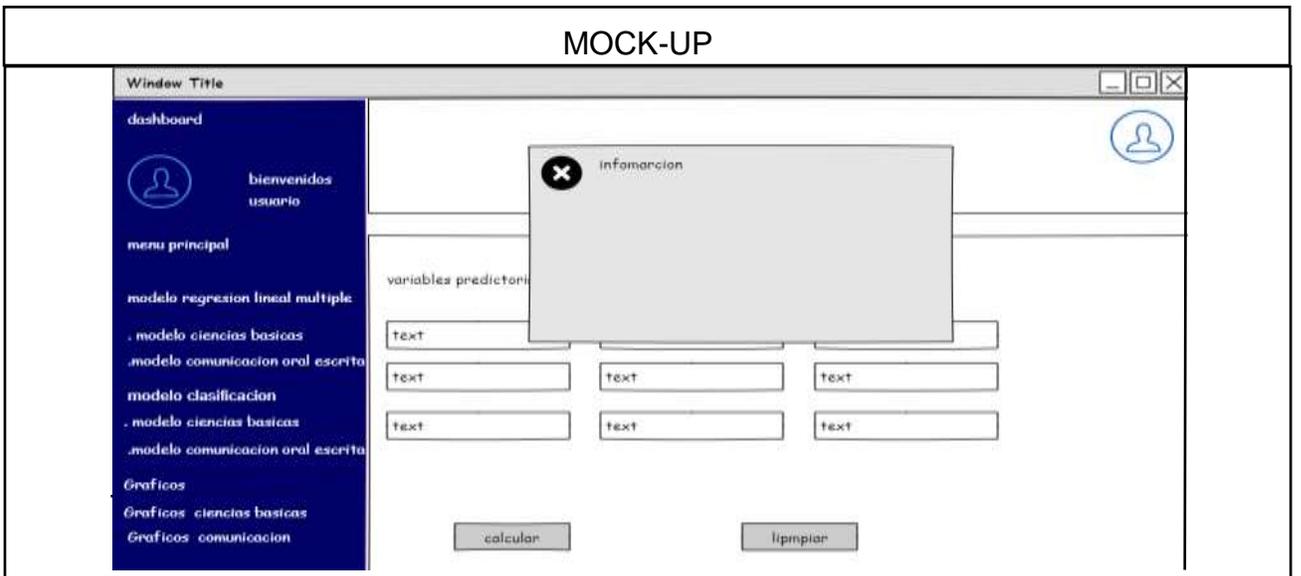
Autor: Fuente propia  
Sprint Review Sprint2

Entrega:	Esteban Andres Urbina Bravo – Luis Gabriel Carlosama Daza
Actividades cumplidas:	Se cumplió satisfactoriamente con la historia de usuario 001
Próximo:	Historia de usuario 3, 4 y 5

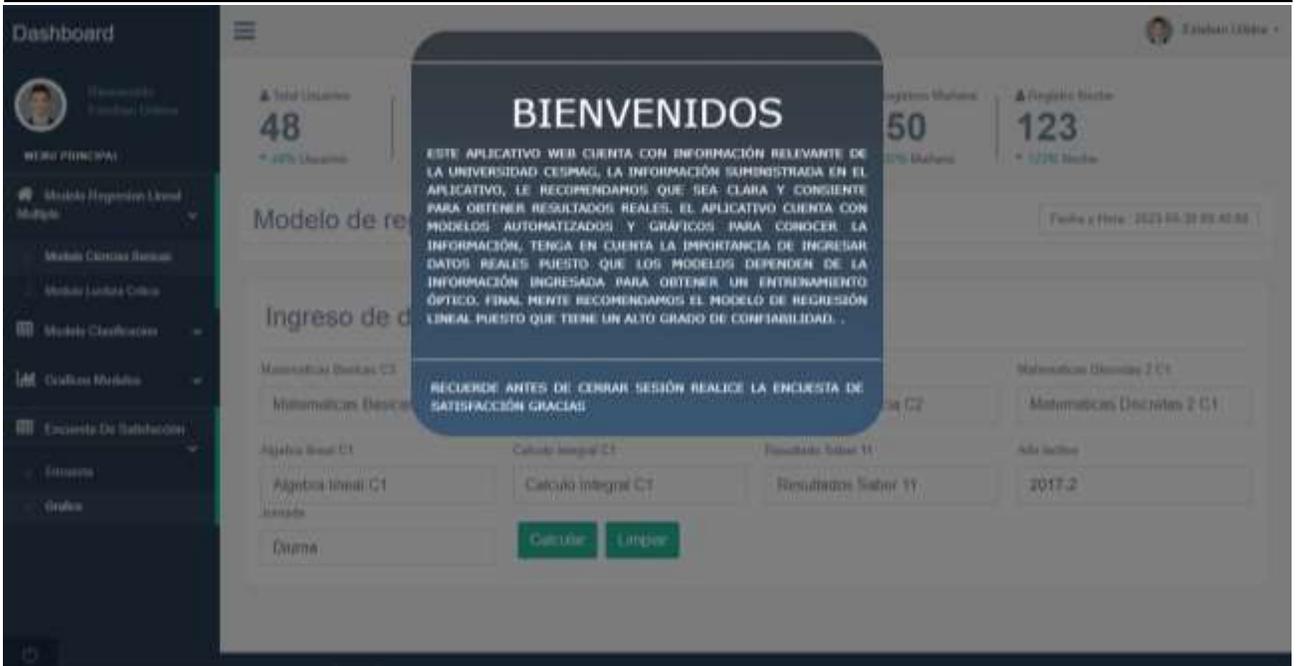
Sprint 3

Tabla 39. MockUp – 3

MOCK-UP	
Identificador	Nombre de historia de usuario
HU-003	Modelo de predicción ciencias básicas
<p>❖ En los diseños siguientes se muestra la interfaz para ingresarlas variables de estudio con las cuales se iniciará el proceso de análisis y su respectiva predicción por módulos, en este caso el módulo de ciencias básicas</p>	



**RESULTADOS**



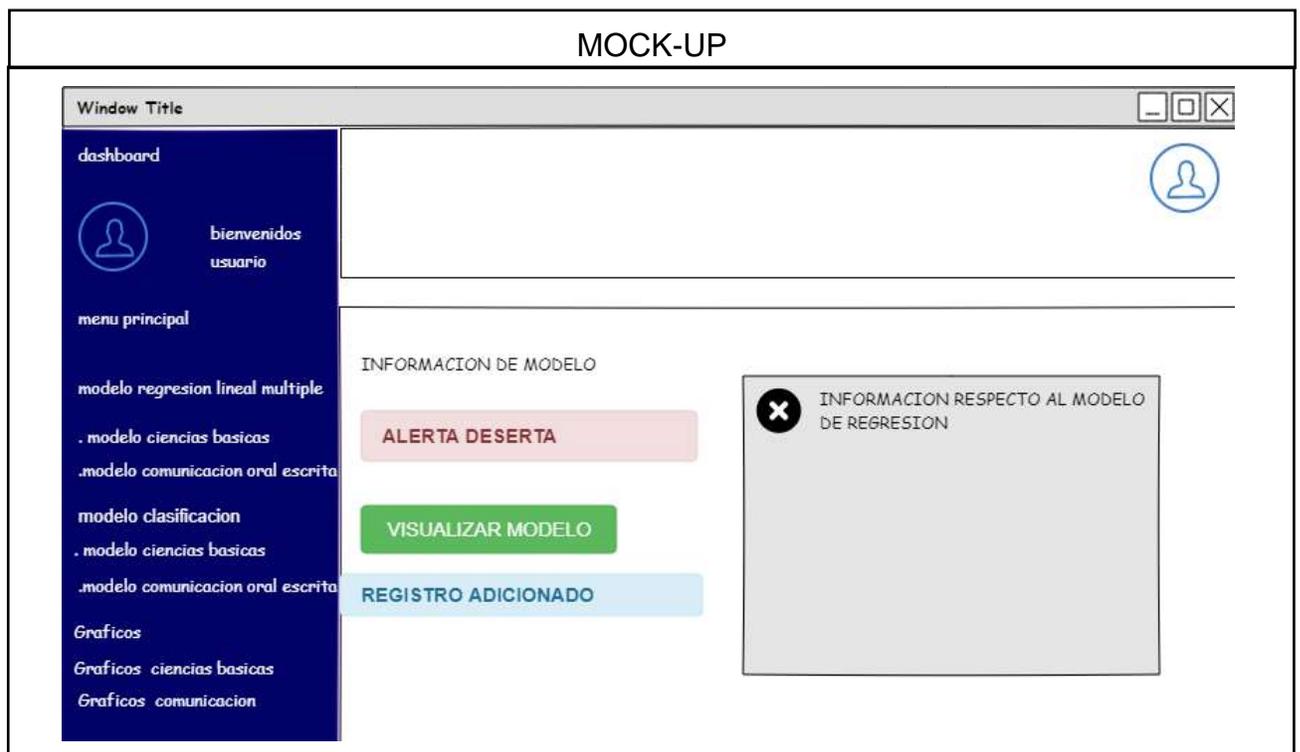
Autor: Fuente propia  
Sprint Review Sprint3

Entrega:	Esteban Andres Urbina Bravo – Luis Gabriel Carlosama Daza
Actividades cumplidas:	Se cumplió satisfactoriamente con la historia de usuario 003
Próximo:	Historia de usuario 4 y 5

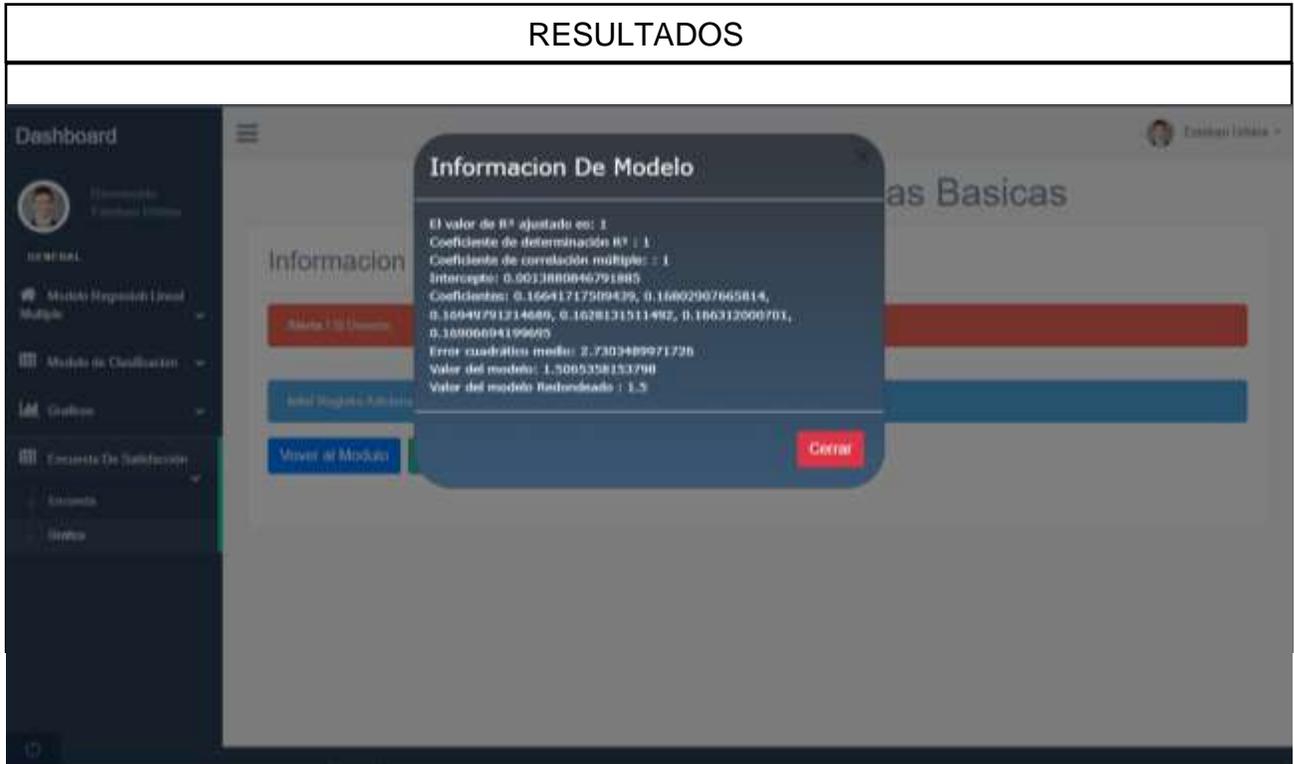
Sprint 4

**Tabla 40. MockUp – 4**

MOCK-UP	
Identificador	Nombre de historia de usuario
HU-004	Modelo de predicción comunicación oral y escrita
<p>❖ En los diseños siguientes se muestra la interfaz para ingresarlas variables de estudio con las cuales se iniciará el proceso de análisis y su respectiva predicción por módulos, en este caso el módulo de comunicación oral y escrita</p>	



## RESULTADOS



Autor: Fuente propia  
Sprint Review Sprint4

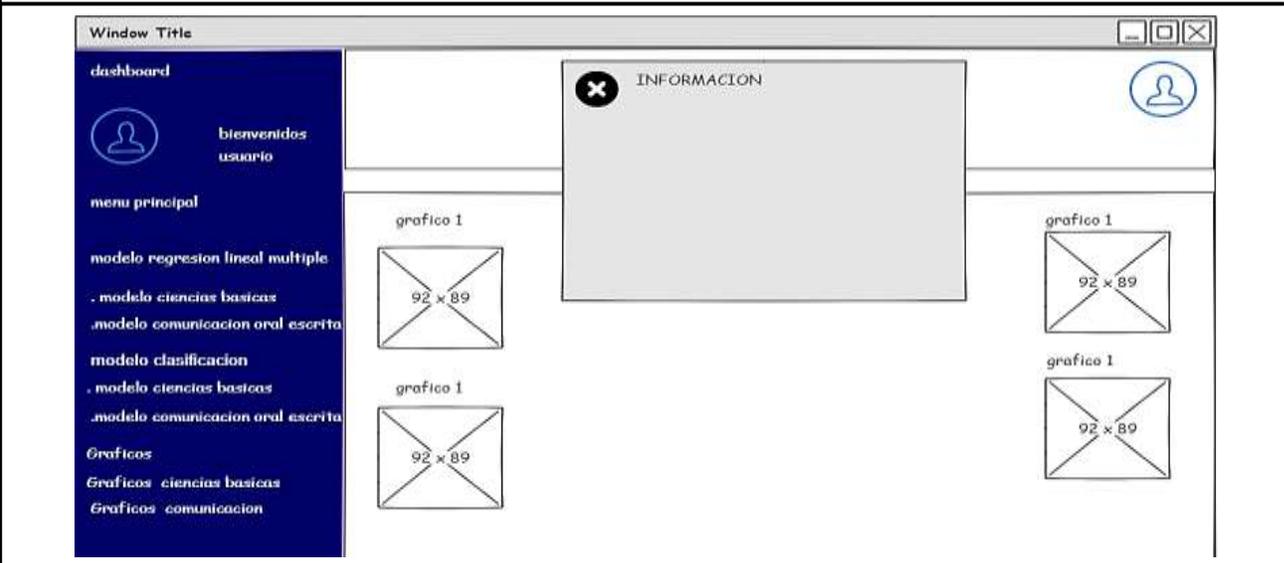
Entrega:	Esteban Andrés Urbina Bravo – Luis Gabriel Carlosama Daza
Actividades cumplidas:	Se cumplió satisfactoriamente con la historia de usuario 004
Próximo:	Historia de usuario 5

Sprint 5

**Tabla 41. MockUp – 5**

MOCK-UP	
Identificador	Nombre de historia de usuario
HU-005	Reporte de graficas
<p>❖ Como resultado de la última historia de usuario y su respectivo requerimiento se presenta el módulo de graficas el cual contiene los resultados del análisis de variables representadas en diferentes gráficos, el cual nos ayudara a identificar con más precisión la posible deserción de estudiantes.</p>	

# MOCK-UP



# RESULTADOS



Autor. Fuente propia  
Sprint Review Sprint5

Entrega:	Esteban Andres Urbina Bravo – Luis Gabriel Carlosama Daza
Actividades cumplidas:	Se cumplió satisfactoriamente con la historia de usuario 005

#### 4.4 VALIDACION DEL APLICATIVO CON LA MÉTRICA DE CALIDAD ISO 25010

ISO 25010 es un estándar internacional que define un conjunto de métricas de calidad para evaluar el software en diferentes aspectos como la compatibilidad, usabilidad, seguridad, fiabilidad, portabilidad y mantenibilidad. Estas métricas están diseñadas para ayudar a los desarrolladores de software a medir la calidad del producto y mejorar su desempeño en estas áreas clave.

El software utilizado en el proyecto actual ha sido evaluado mediante las métricas de calidad ISO 25010. Estas métricas se han seleccionado para medir aspectos clave del producto como la compatibilidad, usabilidad, seguridad, fiabilidad, portabilidad y mantenibilidad, con el objetivo de mejorar su calidad. La evaluación se ha realizado de manera cuantitativa y objetiva, lo que ha permitido obtener resultados precisos sobre la calidad del software utilizado en el proyecto.

Para presentar de forma clara los resultados de esta evaluación, se ha elaborado una tabla que muestra las características y el uso del software en relación a cada una de las métricas de calidad. De esta manera, se puede visualizar de manera fácil y comprensible el rendimiento del software en cada una de las áreas evaluadas.

**Tabla 4.2 Lista de chequeo métricas de calidad**

Lista de chequeo ISO 25010				
Evaluar la calidad del producto software				
Métrica de calidad	submétricas		Aprueba	No aprueba
Usabilidad	Operabilidad	¿el software le permite al usuario operarlo y controlarlo?	SI	
	Atractividad	¿el software cuenta con una interfaz adecuada y fácil de utilizar?	SI	
Seguridad	confidencialidad	¿el software es seguro y confiable?	SI	
	integridad	¿el software permite modificar datos ya guardados en base de datos?	SI	
Fiabilidad	Tolerancia a fallos	¿el producto software cuenta con la capacidad de restablecimiento de servicio y recuperación de datos?	SI	
Portabilidad	Adaptabilidad	¿el producto software tiene la capacidad de adaptarse a cualquier navegador ?	SI	
	Facilidad de instalación	¿el producto software tiene la capacidad de ser instalado en cualquier equipo PC?	SI	
Mantenibilidad	Cambiabilidad	¿el producto software tiene la capacidad de aceptar modificaciones que sean implementadas?	SI	
	Prueba	¿el producto software cuenta con alguna validación de calidad?	SI	

Autor. Fuente propia

Dadas los resultados de las anteriores métricas, demuestra la calidad que se obtiene del producto software ya que aprueba cada una de las métricas de calidad de la norma ISO 25010.

## 5. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

### 5.1 CARACTERIZACIÓN DE LA POBLACIÓN

El despliegue del aplicativo web contó con la participación de un total de 6 estudiantes pertenecientes a las jornadas diurna y nocturna de la universidad.

El objetivo principal fue la simulación, a través del uso de modelos de predicción, del posible riesgo de deserción académica por parte de los estudiantes, en función de su historial académico.

Es importante destacar que la población participante fue seleccionada de manera aleatoria, con el fin de contar con un grupo representativo de los estudiantes de la institución, que permitiera una aproximación a la problemática de la deserción escolar.

**Tabla 43 Caracterización de la Población**

<b>JORNADA</b>	<b>CANTIDAD ESTUDIANTES</b>
Diurna	2
Nocturna	1
Diurna	3
<b>TOTAL</b>	<b>6</b>

Autor fuente propia

Imagen 1 grupo estudiantil



Imagen 2 grupo estudiantil



## 5.2 APLICACIÓN DE LA PREPRUEBA

Antes de iniciar la prueba en sí, se lleva a cabo una preprueba en la que se administra un cuestionario a un grupo de estudiantes. Este cuestionario busca obtener información acerca del rendimiento académico de los estudiantes, en función de las variables predictoras del modelo que se utiliza en el aplicativo web.

El objetivo de esta preprueba es recopilar información real sobre los estudiantes, que luego será utilizada para simular en los modelos y obtener resultados precisos.

Es importante destacar que el éxito del aplicativo web depende en gran medida de la calidad de los datos que se ingresan en el modelo. Por lo tanto, es fundamental contar con información precisa y confiable sobre el rendimiento académico de los estudiantes. De esta manera, se puede garantizar que los resultados obtenidos sean útiles y relevantes para la toma de decisiones

**Tabla 44. Resultados de preprueba**

N°	PREGUNTAS	USUARIOS					
		1	2	3	4	5	6
1	¿Cuál fue tu nota en matematicas basicas corte 3 ?	1,4	2,0	3,5	3,9	4,8	2,5
2	¿Cuál fue tu nota en matematicas discretas corte 1 ?	4,1	4,1	0,2	3,0	4,4	3,5
3	¿Cuál fue tu nota en matematicas en calculo diferencial corte 2 ?	4,2	2,0	1,8	2,0	2,7	5,0
4	¿Cuál fue tu nota en matematicas discretas 2 corte 1 ?	2,8	4,4	4,7	3,0	0,5	3,2
5	¿Cuál fue tu nota en algebra lineal corte 1?	4,5	5,0	0,8	4,0	3,6	3,2
6	¿Cuál fue tu nota en calculo integra corte 1 ?	1,8	1,0	3,1	2,9	4,4	1,4
7	¿Cuál fue tu resultado en la prueba saber 11 en base a razonamiento cuantitativo?	0,8	2,0	0,8	1,8	4,0	1,0
8	¿En que año lectivo ingresaste a la universidad ?	2018-1	2018-1	2018	2018	2017-2	2019-1
9	¿Cuál fue tu nota en comunicación oral y escrita corte 1 ?	3,1	5,0	5,0	2,8	1,7	3,2
10	¿Cuál fue tu nota en comunicación oral y escrita corte 2 ?	3,0	2,0	1,8	3,0	3,1	3,3
11	¿Cuál fue tu nota en comunicación oral y escrita corte 3?	2,9	4,0	4,7	2,7	2,6	2,9
12	¿Cuál fue tu resultado en la prueba saber 11 en base a lectura critica ?	2,9	1,9	0,2	2,3	4,3	2,9
13	¿En que jornada estudias?	DIURNA	DIURNA	DIURNA	DIURNA	DIURNA	NOCTURNA
14	¿Cuál fue tu promedio global en materias de ciencias basicas ?	3,1	3,1	2,3	3,1	3,4	3,1
15	¿Cuál fue tu promedio global en comunicación oral y escrita ?	3,0	3,7	3,8	2,8	2,5	3,1
16	¿Haz repetido alguna materia de las que se menciono anteriormente?	NO	NO	SI	SI	SI	NO
<b>TOTAL</b>		<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
<b>TOTAL TIEMPO EN MINUTOS</b>		<b>15</b>	<b>15</b>	<b>16</b>	<b>20</b>	<b>10</b>	<b>16</b>

Autor fuente propia

Descripción de la tabla: De acuerdo a los resultados de la tabla 44 se obtuvo lo siguiente:

El objetivo de la tabla es recopilar información de estudiantes activos de la Universidad CESMAG, para esta prueba se realizó un cuestionario a estudiantes activos donde se preguntó datos de sus notas cursadas durante el periodo de formación equivalente a las variables predictoras.

**Tabla 45. Total, resultados preprueba**

<b>TOTAL</b>	<b>DESERTADOS</b>	<b>NO DESERTADOS</b>	<b>TOTAL VARIABLES</b>
	3	9	12
	25%	75%	100%

Autor fuente propia

**Tabla 46 Total resultados preprueba**

<b>TOTAL</b>	<b>REPITIO</b>	<b>NO REPITIO</b>	<b>TOTAL USUARIOS</b>
	3	3	6
	50%	50%	100%

Autor fuente propia

Descripción de la tabla: En la tabla 45 y 46 muestra que 25% de los usuarios tuvo un bajo rendimiento en las variables predictoras y el 75% obtuvo un rendimiento considerable, no obstante, en tabla 46 el 50 % de encuestados repitió alguna vez una materia de las variables predictoras

### 5.3 APLICACIÓN DE LA POSTPRUEBA

La post prueba es una fase importante del proceso de evaluación en la que se pide a los estudiantes que ingresen nuevamente los mismos datos al modelo predictivo del aplicativo web. El objetivo de la post prueba es obtener los resultados reales y compararlos con los resultados simulados en la preprueba, con el fin de evaluar la precisión del modelo.

Esta comparación permite determinar la calidad del modelo predictivo y su capacidad para predecir el rendimiento académico de los estudiantes. Los resultados obtenidos en la post prueba proporcionan información valiosa que puede utilizarse para mejorar el modelo y hacer ajustes necesarios.

En conclusión, la post prueba es una fase fundamental del proceso de evaluación, ya que permite evaluar la precisión del modelo predictivo del aplicativo web. La comparación de los resultados obtenidos en el post prueba con los resultados simulados en la preprueba proporciona información valiosa para mejorar el modelo y hacer ajustes necesarios. La participación de los estudiantes es esencial para obtener resultados precisos y confiables.

**Tabla 47 Resultados de post-prueba modelo de regresión lineal**

N°	PREGUNTAS	USUARIOS					
		1	2	3	4	5	6
1	¿Cuál fue tu nota en matematicas basicas corte 3 ?	1,4	2,0	3,5	3,9	4,8	2,5
2	¿Cuál fue tu nota en matematicas discretas corte 1 ?	4,1	4,1	0,2	3,0	4,4	3,5
3	¿Cuál fue tu nota en matematicas en calculo diferencial corte 2 ?	4,2	2,0	1,8	2,0	2,7	5,0
4	¿Cuál fue tu nota en matematicas discretas 2 corte 1 ?	2,8	4,4	4,7	3,0	0,5	3,2
5	¿Cuál fue tu nota en algebra lineal corte 1?	4,5	5,0	0,8	4,0	3,6	3,2
6	¿Cuál fue tu nota en calculo integra corte 1 ?	1,8	1,0	3,1	2,9	4,4	1,4
7	¿Cuál fue tu resultado en la prueba saber 11 en base a razonamiento cuantitativo?	0,8	2,0	0,8	1,8	4,0	1,0
8	¿En que año lectivo ingresaste a la universidad ?	2018-1	2018-1	2018	2018	2017-2	2019-1
9	¿Cuál fue tu nota en comunicación oral y escrita corte 1 ?	3,1	5,0	5,0	2,8	1,7	3,2
10	¿Cuál fue tu nota en comunicación oral y escrita corte 2 ?	3,0	2,0	1,8	3,0	3,1	3,3
11	¿Cuál fue tu nota en comunicación oral y escrita corte 3?	2,9	4,0	4,7	2,7	2,6	2,9
12	¿Cuál fue tu resultado en la prueba saber 11 en base a lectura crítica ?	2,9	1,9	0,2	2,3	4,3	2,9
13	¿En que jornada estudias?	DIURNA	DIURNA	DIURNA	DIURNA	DIURNA	NOCTURNA
14	¿Cuál fue el resultado del modelo en el modulo de ciencias basicas ?	3.14	3.0	2,3	3.13	3.42	3.14
15	¿Cuál fue el resultado del modelo en el modulo de comunicación oral y escrita ?	3.6	3.6	3.8	2.8	2.5	3.13
<b>TOTAL</b>		<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
<b>TOTAL TIEMPO EN MINUTOS</b>		<b>15</b>	<b>15</b>	<b>16</b>	<b>20</b>	<b>10</b>	<b>16</b>

Autor fuente propia

Descripción de la tabla: De acuerdo a la tabla se obtuvo lo siguiente:

Los datos visualizados en la tabla número 44 se simuló en el aplicativo uno a uno.

Los resultados obtenidos en la post prueba fueron muy similares a los resultados simulados en la preprueba, lo que indica que el modelo de regresión lineal múltiple utilizado en el aplicativo web es altamente confiable y preciso en la predicción del rendimiento académico de los estudiantes en los módulos de ciencias básicas y comunicación oral y escrita.

La alta fiabilidad del modelo es una señal de la calidad de la información recopilada en la preprueba y de la capacidad del modelo para hacer predicciones precisas en base a las variables predictoras seleccionadas. Estos resultados son muy alentadores para la utilización del aplicativo web en la evaluación del rendimiento académico de los estudiantes.

En conclusión, los resultados obtenidos en la post prueba indican que el modelo de regresión lineal múltiple utilizado en el aplicativo web es altamente confiable y preciso en la predicción del rendimiento académico de los estudiantes.

#### **Tabla 48 Resultados de post-prueba modelo de clasificación**

N°	PREGUNTAS	USUARIOS					
		1	2	3	4	5	6
1	¿Cuál fue tu nota en matematicas basicas corte 3 ?	1,4	2,0	3,5	3,9	4,8	2,5
2	¿Cuál fue tu nota en matematicas discretas corte 1 ?	4,1	4,1	0,2	3,0	4,4	3,5
3	¿Cuál fue tu nota en matematicas en calculo diferencial corte 2 ?	4,2	2,0	1,8	2,0	2,7	5,0
4	¿Cuál fue tu nota en matematicas discretas 2 corte 1 ?	2,8	4,4	4,7	3,0	0,5	3,2
5	¿Cuál fue tu nota en algebra lineal corte 1?	4,5	5,0	0,8	4,0	3,6	3,2
6	¿Cuál fue tu nota en calculo integra corte 1 ?	1,8	1,0	3,1	2,9	4,4	1,4
7	¿Cuál fue tu resultado en la prueba saber 11 en base a razonamiento	0,8	2,0	0,8	1,8	4,0	1,0
8	¿En que año lectivo ingresaste a la universidad ?	2018-1	2018-1	2018	2018	2017-2	2019-1
9	¿Cuál fue tu nota en comunicación oral y escrita corte 1 ?	3,1	5,0	5,0	2,8	1,7	3,2
10	¿Cuál fue tu nota en comunicación oral y escrita corte 2 ?	3,0	2,0	1,8	3,0	3,1	3,3
11	¿Cuál fue tu nota en comunicación oral y escrita corte 3?	2,9	4,0	4,7	2,7	2,6	2,9
12	¿Cuál fue tu resultado en la prueba saber 11 en base a lectura critica ?	2,9	1,9	0,2	2,3	4,3	2,9
13	¿En que jornada estudias?	DIURNA	DIURNA	DIURNA	DIURNA	DIURNA	NOCTURNA
14	¿Según el modelo de clasificacion deserto en ciencias basicas ?	NO	NO	NO	NO	NO	NO
15	¿Según el modelo de clasificacion deserto en comunicación oral y	NO	SI	NO	NO	NO	NO
<b>TOTAL</b>		<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
<b>TOTAL TIEMPO EN MINUTOS</b>		<b>15</b>	<b>15</b>	<b>16</b>	<b>20</b>	<b>10</b>	<b>16</b>

Autor fuente propia

Descripción de la tabla: De acuerdo a la tabla se obtuvo lo siguiente:

Los datos visualizados en la tabla número 44 se simuló en el aplicativo uno a uno.

Los resultados del modelo de clasificación para la predicción del rendimiento académico fueron desfavorables en comparación con el modelo lineal utilizado. El modelo de clasificación detectó un estudiante basándose en las variables de entrada, y se observó que este estudiante presentaba notas bajas en el módulo de comunicación oral y escrita.

Además, uno de los factores que puede haber contribuido a la baja precisión del modelo de clasificación es la falta de inclusión de factores no predefinidos que también influyen en la deserción de los estudiantes, como la falta de interés, la

desmotivación, la falta de información sobre la materia, entre otros. Estos factores pueden ser difíciles de cuantificar y medir, pero es importante tenerlos en cuenta en el desarrollo de modelos de predicción más precisos y completos.

En conclusión, los resultados del modelo de clasificación para la predicción del rendimiento académico fueron desfavorables en comparación con el modelo lineal utilizado en el estudio.

**Tabla 49. Total, resultados post-prueba**

<b>MODELO LINEAL</b>	<b>DESERTA</b>	<b>NO DESERTA</b>	<b>TOTAL</b>
	3	9	12
PORCENTAJE	25%	75%	100%
<b>MODELO DE CLASIFICACION</b>	<b>DESERTA</b>	<b>NO DESERTA</b>	<b>TOTAL</b>
	1	11	12
PORCENTAJE	8%	92%	100%

Autor fuente propia

Descripción de la tabla: En la tabla 49, se analiza los resultados de los dos modelos. Visualmente el modelo que menos desertan es el de clasificación esto indica que:

Para la investigación se recomienda el modelo de regresión lineal múltiple puesto que el modelo se analizó estadísticamente, obteniendo una correlación entre variables unos R cuadrados, un error cuadrático a favor de las variables, teniendo en cuenta que las predicciones se hacen en base a su rendimiento e historial académico por ende el modelo de regresión es más efectivo a razón que se basa en fórmula matemática.

Sin embargo, el modelo de clasificación puede lograr una predicción precisa, en este caso el entrenamiento del modelo no está predefinidos los factores que si inciden en la deserción como: económico, emocional, falta de interés, falta de información, nivel académico de los padres entre otros. Por causa que las bases de datos que se adquirió desde desarrollo de software no se encontró dicha información en cada registro.

## 5.5 RESULTADOS DE ENCUESTA DE SATISFACION

Después de realizada la encuesta de satisfacción a todos los usuarios que analizaron nuestro software, se obtuvo un resultado favorable, el cual plasma los resultados en la siguiente tabla.

Donde se obtienen los resultados finales en cada una de las preguntas de acuerdo a la cantidad de usuarios que respondieron, en total 6 estudiantes encuestados

**Tabla 50 Resultados prueba satisfacción**

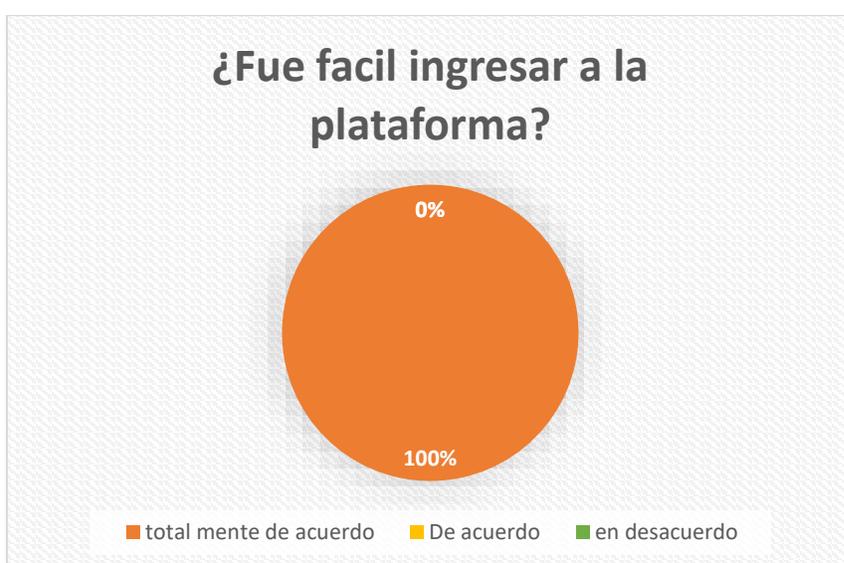
<b>N°</b>	<b>PREGUNTAS</b>	<b>TOTAL, MENTE DE ACUERDO</b>	<b>DE ACUERDO</b>	<b>EN DESACUERDO</b>	<b>TOTAL, PERSONAS ENCUESTADAS</b>
<b>1</b>	¿Fue fácil ingresar a la plataforma?	6	0	0	6
<b>2</b>	¿Tuvo dificultad al manejar la aplicación?	5	1	0	6
<b>3</b>	¿Las funciones de la aplicación fueron claras y fáciles de entender?	6	0	0	6
<b>4</b>	¿Recomendaría el software a los demás?	6	0	0	6
<b>5</b>	¿Está satisfecho con el rendimiento del software?	5	1	0	6
<b>6</b>	¿Cree usted que la herramienta cumple con los requerimientos de seguridad?	5	1	0	6
<b>TOTAL</b>		<b>33</b>	<b>3</b>	<b>0</b>	<b>36</b>

Autor fuente propia

Descripción de la tabla: Como se observar en la tabla de la encuesta de satisfacción, 33 de 36 respuestas están total mente de acuerdo con la aceptación del software

Para concluir con el estudio de la encuesta de satisfacción para usuarios se realiza la tabulación por cada una de las respuestas que genero dicha encuesta a continuación:

Figura 1 ¿Fue fácil ingresar a la plataforma?

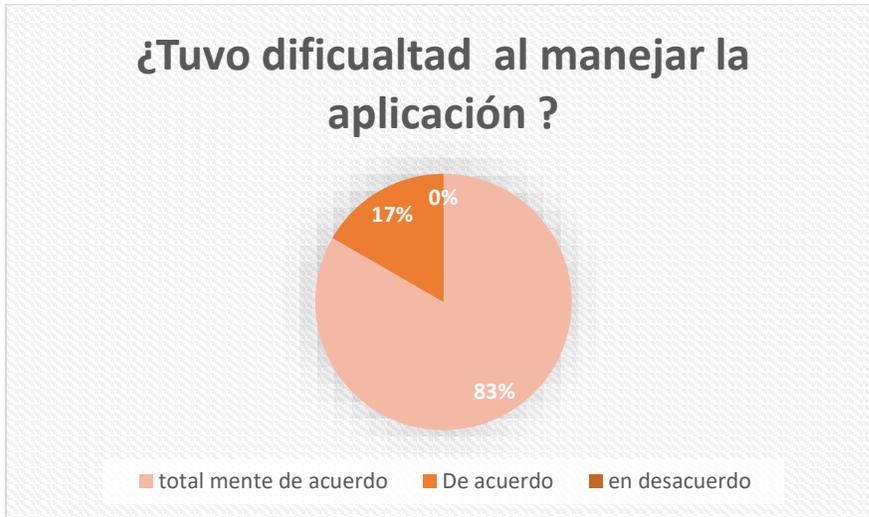


Fuente: Resultado de investigación

Descripción de la figura:

De acuerdo a los resultados de la investigación en su mayoría de encuestados no tuvo problemas a la hora de acceder a la plataforma, por tal motivo es un resultado favorable para el proyecto.

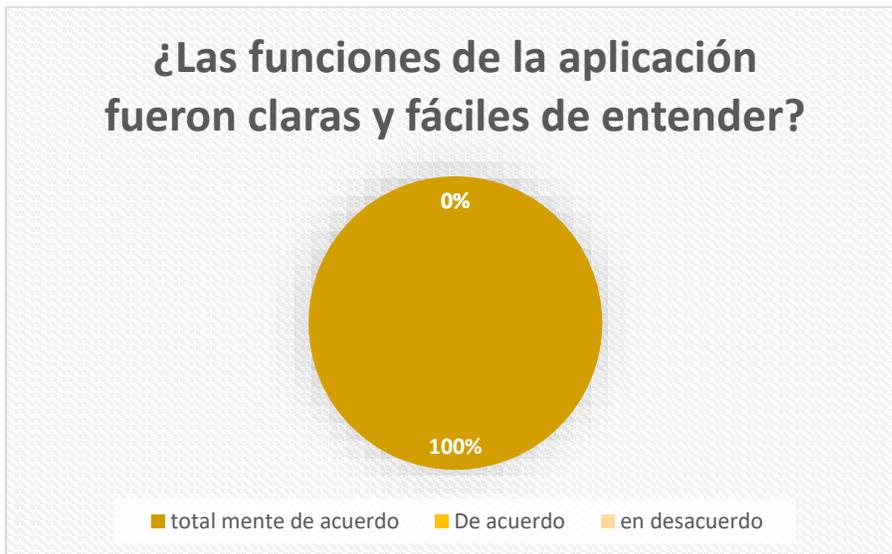
Figura 2 ¿Tuvo dificultad al manejar la aplicación?



Fuente: Resultado de investigación

Como resultado de la operabilidad del sistema favorece en un alto grado de sencillas al utilizar el aplicativo

Figura 3 ¿Las funciones de la aplicación fueron claras y fáciles de entender?



Fuente: Resultado de investigación

Descripción de la figura: La usabilidad de la aplicación fue de un porcentaje positivo y muestra resultados favorables para su funcionamiento.

Figura 4 ¿Recomendaría el software a los demás?



Fuente: Resultado de investigación

Descripción de la figura: Según el resultado se comprende que la mayoría de encuestados recomendaría el aplicativo

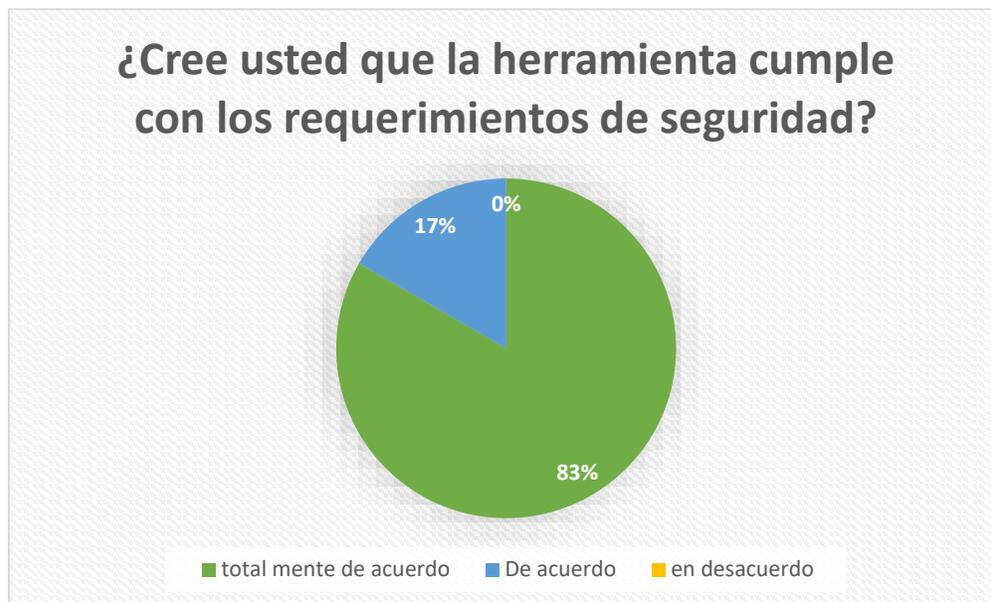
Figura 5 ¿Estas satisfecho con el rendimiento de nuestro software?



Fuente: Resultado de investigación

Descripción del grafico: La satisfacción del cliente es muy importante, ya que muestra su interés de seguir utilizando mucho más el producto, por lo tanto, en el análisis de la encuesta se encontró que dichos usuarios dan un resultado positivo en la funcionalidad del software.

Figura 6 ¿Cree usted que la herramienta cumple con los requerimientos de seguridad?



Fuente: Resultado de investigación

Descripción del grafico: La gran parte de encuestados está en totalmente de acuerdo que el aplicativo cumple con sus estándares de calidad

## 5.6 DISCUSION DE RESULTADOS

Basándose en los resultados obtenidos en la post prueba, se puede concluir que el modelo de regresión lineal utilizado en el aplicativo web ha demostrado ser altamente confiable y preciso en la predicción del rendimiento académico de los estudiantes en los módulos de ciencias básicas y comunicación oral y escrita.

Una de las principales fortalezas de este modelo es su capacidad para hacer predicciones precisas y consistentes, lo que sugiere que puede ser una herramienta útil para la evaluación y el apoyo académico de los estudiantes.

Además, la calidad de la información recopilada en la preprueba ha demostrado ser una contribución importante para la precisión del modelo.

En resumen, los resultados obtenidos en la post prueba sugieren que el modelo de regresión lineal múltiple utilizado en el aplicativo web puede ser una herramienta valiosa para la evaluación del rendimiento académico de los estudiantes y puede ayudar a los educadores a identificar a los estudiantes que necesitan apoyo adicional para mejorar su desempeño académico.

El modelo de clasificación presentó resultados limitados al predecir el rendimiento académico de los estudiantes, con solo un estudiante siendo identificado como propenso a desertar. Se observó que este estudiante tenía notas bajas en el módulo de comunicación oral y escrita. Sin embargo, la baja precisión del modelo puede estar relacionada con la falta de inclusión de factores no predefinidos que también influyen en la deserción de los estudiantes, como la falta de interés o la desmotivación. Es importante tener en cuenta estos factores para mejorar la precisión y completitud del modelo de predicción.

En general, se concluye que el modelo de clasificación requiere una mayor refinación y consideración de los factores relevantes para mejorar su precisión y utilidad en la predicción del rendimiento académico.

#### 5.6.1 Resultados de la investigación referentes al marco teórico

Realizando una comparación de la presente investigación frente a lo que propone el autor:

**Álvaro Agustín Oñate Bowen, en su tesis de maestría titulada "Análisis de la deserción y permanencia académica en la educación superior aplicando minería de datos".**

La principal diferencia entre la investigación mencionada y la otra investigación radica en el enfoque de recopilación de datos. Mientras que la investigación mencionada recopiló información socioeconómica, historial académico y desempeño en pruebas de estado, la otra investigación se basó en variables categóricas en lugar de variables continuas.

La inclusión de variables continuas en el estudio permite un análisis más preciso y detallado de los factores que afectan la deserción y permanencia académica en la educación superior. Al utilizar variables continuas, se pueden capturar matices y

variaciones más sutiles, lo que enriquece la comprensión de los patrones y relaciones presentes en los datos.

Por otro lado, aunque la investigación se centró principalmente en el historial académico para desarrollar modelos predictivos, es importante reconocer la relevancia de considerar otros factores, como la motivación, el interés y la falta de información sobre la materia. Estos aspectos psicológicos y socioemocionales desempeñan un papel fundamental en las decisiones de los estudiantes y, por lo tanto, deben tenerse en cuenta para mejorar la precisión y completitud de los modelos de predicción.

En resumen, la inclusión de variables continuas en la recopilación de datos y la consideración de factores adicionales más allá del historial académico representan mejoras significativas que enriquecen el análisis de la deserción y permanencia académica en la educación superior. Al combinar estos enfoques, es posible obtener una comprensión más profunda y precisa de los factores que influyen en la condición académica de los estudiantes.

### **Realizando una comparación en su artículo detección de Patrones de Bajo Rendimiento académico y deserción estudiantil con técnicas de minería de datos de Silvio Ricardo Timaran.**

Los hallazgos sobre el bajo rendimiento académico en estudiantes de ciencias básicas. Ambos estudios indican que los estudiantes pertenecientes a facultades de ciencias básicas presentan un bajo rendimiento académico. En el artículo se centra en la población universitaria de la Universidad UDENAR, se encontró que una parte considerable de los estudiantes de primer semestre, especialmente aquellos de estratos bajos y de la zona sur de Nariño, exhiben un bajo rendimiento académico en las facultades de ciencias básicas, humanas y naturales.

En contraste, la investigación se centra en el rendimiento académico de las materias del departamento de ciencias básicas durante los primeros tres semestres, los cuales abarcan el contenido específico de dichas asignaturas. De manera consistente con los hallazgos del artículo, se concluye que estas materias desempeñan un papel relevante en la deserción estudiantil debido a las dificultades que los estudiantes enfrentan al aprender dichos contenidos.

Es importante tener en cuenta que estos hallazgos son específicos de las poblaciones y contextos mencionados en cada estudio, y no se pueden generalizar de manera indiscriminada a todas las universidades o programas académicos

**La metodología CRISP-DM** es crucial en la gestión de proyectos de minería de datos, ya que brinda una estructura clara y eficiente. Su enfoque paso a paso permite comprender los objetivos del negocio, recopilar y preparar los datos, explorar y modelarlos, y evaluar los resultados. La colaboración entre equipos es fomentada, lo que permite adaptarse según sea necesario durante el proyecto. Al seguir CRISP-DM, se garantiza una comprensión exhaustiva de los datos y se logran resultados más confiables y efectivos en la minería de datos.

La aplicación de la metodología CRISP-DM en la investigación es eficiente debido a su enfoque claro y comprensible en cada una de sus fases. Esta metodología fomenta el trabajo colaborativo y se caracteriza por ser iterativa, lo que permite a los investigadores realizar ciclos repetidos de análisis y ajustes a lo largo del proyecto. La flexibilidad que proporciona CRISP-DM les permite retroceder a etapas anteriores para realizar mejoras o modificaciones en función de los hallazgos y los cambios en los requisitos del proyecto. Esto garantiza una mayor eficiencia y calidad en el proceso de investigación, al adaptarse y refinar continuamente los resultados obtenidos

### **Modelos de regresión lineal múltiple**

La regresión lineal múltiple es una técnica estadística utilizada para establecer relaciones entre una variable dependiente y múltiples variables independientes. Durante el análisis, se examina la presencia de heterocedasticidad, multicolinealidad y se asegura una especificación precisa del modelo.

La importancia de aplicar el modelo de regresión lineal múltiple en esta investigación radica en su capacidad para establecer relaciones entre variables. El objetivo principal del estudio fue analizar el rendimiento académico utilizando variables continuas y desarrollar un modelo predictivo. Durante el análisis de regresión, se calcularon promedios y desviaciones típicas para interpretar de manera conjunta los resultados, evaluando los valores tanto en términos absolutos como en relación a las variables consideradas. Además, se clasificaron las variables de estudio en independientes y dependientes, lo que respalda la elección del modelo para examinar el fenómeno de la deserción estudiantil. Es relevante destacar que el estudio estadístico realizado demostró una confiabilidad del 100% (ver capítulo 4).

## **Metodología SCRUM**

Scrum es un enfoque que implica la aplicación regular de un conjunto de prácticas efectivas para colaborar en equipo y lograr los mejores resultados posibles en un proyecto. Esta metodología se destaca por hacer entregas parciales y periódicas del producto final, priorizando aquellas que generan mayores beneficios para el receptor del proyecto. En el contexto del desarrollo de aplicativos webs, la metodología ágil, como Scrum, se muestra altamente eficiente. Su enfoque iterativo permite a los desarrolladores llevar a cabo ciclos de entrega en tiempos determinados, promoviendo así la productividad y la planificación del proyecto. La flexibilidad que Scrum proporciona se convierte en un factor crucial, ya que permite realizar cambios y ajustes incluso en etapas avanzadas del proceso, adaptándose a las necesidades cambiantes del cliente. Esta capacidad de efectuar modificaciones de última hora contribuye a un trabajo más eficiente y efectivo para el cliente, ofreciendo un resultado final que cumple con sus expectativas y requerimientos

## CONCLUSIONES

Fueron analizados con éxito cada uno de los datos proporcionados por la oficina de software la cual se obtuvo una base de datos de más de 950 registros de estudiantes del programa de ingeniería de sistemas de los tres primeros semestres en las jornadas diurna y nocturna de la universidad CESMAG, como también los datos de resultados de las pruebas saber 11 de cada uno de los aspirantes de los años 2017- B a 2022-A donde se depuro cada uno de estos datos y se obtuvo como resultado las variables más importantes en las áreas que toman mayor influencia en el programa de ingeniería de sistemas, como son las áreas de ciencias básicas referente a todas la matemáticas estudiadas durante estos tres semestres y la comunicación oral y escrita donde toman un alto porcentaje para ser analizadas durante la aplicación de este proyecto.

Los datos expuestos por la oficina de software de la universidad, se analizaron para predecir la posible deserción del estudiante de ingeniería de sistemas, la cual se aplicó con éxito cada una de las fases que contiene la metodología CRISP DM. Dado, el análisis de gestión de datos exactamente fue realizada por el software IBM SPSS Statistics, el cual fue fundamental para encontrar el modelo matemático de regresión lineal múltiple con una predicción del 100% de los datos estudiados, fue recomendado también el modelo de clasificación, donde al realizar el análisis de los datos se observa un comportamiento menor que el estudiado anteriormente, para corroborar esta teoría se aplicó preprueba y post-prueba a un determinado grupo de estudiantes que obtuvo como resultado varios aspectos relevantes de los modelos de predicción, donde el modelo matemático de regresión lineal fue el que mayor influencia obtuvo durante las pruebas realizadas. En conclusión, los resultados del modelo de clasificación para la predicción del rendimiento académico fueron desfavorables en comparación con el modelo lineal utilizado en el estudio.

Por medio de las métricas de calidad establecidas por el estándar ISO 25010 fueron realizados cada uno de los requerimientos funcionales y no funcionales del software, para dar así una mejor calidad del producto a entregar, esto hace que sea más confiable y seguro a la hora de utilizar el producto. Las métricas de calidad fueron: la confiabilidad, usabilidad, portabilidad y seguridad las cuales evalúan el proyecto, obteniendo un resultado positivo tras los estudios realizados a los estudiantes de los diferentes semestres del programa de ingeniería de sistemas de la universidad CESMAG.

## RECOMENDACIONES

Se recomienda incluir en el análisis otros factores relevantes para la deserción estudiantil, como los aspectos socioeconómicos, familiares y psicológicos. Estos factores pueden tener una mayor influencia y deben ser considerados en la herramienta web utilizada.

Se recomienda incorporar una función que permita exportar la base de datos seleccionada y depurada, de manera que el sistema pueda realizar el análisis correspondiente en cada uno de los modelos presentados

El aplicativo está en la capacidad de actualizar sus datos para que en un futuro cuente con mejoras que sean pertinentes con el objetivo de la investigación

## BIBLIOGRAFIA

- ABREU ALVES, Sara, SINVAL, Jorge, NETO, Lía Lucas, MAROCO, João, GONÇALVES FERREIRA, António y OLIVEIRA, Pedro. Burnout e intención de abandono en estudiantes de medicina: el papel protector del compromiso académico. BMC Med Educ. 2022. p.3. En:
- ALANIA RICARDI, Frank. Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la facultar de ingeniería de la universidad nacional Daniel Alcides Carrión. Perú, Pasco. Universidad Nacional Daniel Alcides Carrión. 2018. p.1-90. En:
- ALVARADO claudia victoria, Ventajas y desventajas de usar Excel para reportar resultados de la estrategia. Pensemos. 2021
- BAZANTES Z., P., Ruiz Carpio, M. L. y ÁLVAREZ Gutiérrez, M. L. Deserción estudiantil universitaria en ecuador y su influencia en la calidad del egresado. Magazine De Las Ciencias:
- BELTRAN diego y POVEDA diego; RAPIDMINER; UNI-FO-13 V 1.0, 2010
- BLANCO VILLAFañE, Vanessa Paola. (2016). Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos, 2016. p.5.
- BLAS, María Julia; GONNET, Silvio M.; LEONE, Horacio P. Especificación de la Calidad en Software-as-a-Service: Definición de un Esquema de Calidad basado en el Estándar ISO/IEC 25010. En Simposio Argentino de Ingeniería de Software (ASSE 2016)-JAIIO 45 (Tres de Febrero, 2016). 2016.
- BLANCO VILLAFañE, Vanessa Paola. (2016). Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos, 2016. p.5.
- BOU, Ramón Carrión. Usando XAMPP con Bootstrap y WordPress. Mercedes Gómez Alcalá, 2019.
- BONILLA Elsy y RODRÍGUEZ Penélope. Más allá del dilema de los métodos. La investigación en Ciencias Sociales. 3ra. Ed. Ediciones Uniandes. Bogotá. Colombia, 1997. p.21.
- BURGOS FLOREZ, Marco Antonio, et al. Diagnóstico del fenómeno de deserción estudiantil en los programas de pregrado de la Universidad de Nariño–sede Pasto 2012–2016. 2018.
- CARNEIRO, Roberto, TOSCANO, Juan Carlos y DÍAZ, Tamara. Los desafíos de las TIC para el cambio educativo. 2021. p.127. En:  
<https://www.oei.es/uploads/files/microsites/28/140/lastic2.pdf>

- COLOMBIA, MINISTERIO DE EDUCACION NACIONAL. Decreto 869(17, marzo,2010).  
 Por el cual se reglamenta el Examen de Estado de la Educación Media,  
 ICFESSABER 11°.Santafe de Bogotá, D.C: El ministerio,2010
- DESERCIÓN Estudiantil con Técnicas de Minería de Datos En: Memorias de la 8ª  
 Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI. 2009.
- DEUSTO FORMACIÓN, Blog Programación y TIC [sitio web]. [consultado: abril 15 de  
 2022]. Disponible en: <https://www.deustoformacion.com/blog/programacion-tic/que-son-datasets-dataframes-big-data>
- DÍAZ ESCOBAR, Rafael. Análisis de la deserción estudiantil en la Universidad de Chile,  
 con énfasis en las ayudas financieras, 2021.p. II.
- Disponible en:  
 [https://caracol.com.co/programa/2021/11/04/sanamente/1636049871\_761122.htm  
 l#:~:text=por%20 la%20educaci%C3%B3n%20superior%20en,  
 cifras%20hablan%20por%20s%C3%AD%20solas.
- Educativa, 2018. p.185-202. doi: <http://dx.doi.org/10.6018/rie.36.1.278971>
- ERAZO ORTIZ, Eduard Alirio y MORA POTILLA, Sergio Aldemar. Aplicación de la minería  
 de datos para la detección de factores asociados a la deserción estudiantil en los  
 programas profesionales de las ciencias naturales y ciencias sociales de la  
 universidad de Nariño sede Pasto. Colombia, Pasto. 2018. p.1-92.
- revista de Estudios y Experiencias en Educación, vol. 19, núm. 40, pp. 35-52. 2020. En:  
 FERNÁNDEZ, Diego Buenaño; MORA, Sergio Luján. Uso de la metodología CRISP-DM  
 para guiar el proceso de minería de datos en LMS. En Tecnología, innovación e  
 investigación en los procesos de enseñanza-aprendizaje. Octaedro, 2016. p. 2385-  
 2393.
- FERNÁNDEZ, S. Estadística descriptiva. Edición. ESIC Editorial,2edi.2002
- FREIXA NIELLA, M., LLANES ORDÓNEZ, J. y VENCESLAO PUEYO, M. El abandono  
 en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona.  
 Revista de Investigación  
 Universidad de Barcelona. En: Revista de Investigación Educativa, 2018, vol. 1.
- FRIENDS, Apache. XAMPP Apache+ MariaDB+ PHP+ Perl. Retrieved april, 2019, vol.5.
- GALVIS, Carlos Angulo. Efectos de la deserción. Colombia, Cali. 2017. p.19.
- GARCÍA OTÁLORA, Héctor Julián. Relación y uso de las TIC con la deserción en  
 estudiantes pertenecientes a la Facultad de Ingeniería de la Universidad Nacional  
 de Colombia, 2017. p.1.
- GARCÍA, Beatriz. Una mirada al método científico. Hojitas de Conocimiento. Ciencia; nº  
 40, 2020.
- GARGALLO CASTEL, Ana Felicitas. La integración de las TIC en los procesos  
 educativos y organizativos. Pdf, junio 2018.
- GUZMAN, Carolina; DURÁN, Diana y FRANCO, Jorge. Deserción estudiantil En: la  
 educación superior colombiana: *Metodología de seguimiento, diagnóstico y  
 elementos para su prevención*, 2009. 13-17 p

GRANADOS, R. Montero. Modelos de regresión lineal múltiple. Granada, España: Departamento de Economía Aplicada, Universidad de Granada, 2016.

HINCAPIÉ MORENO, Catalina; WILCHES LÓPEZ, Laura Esther. Diseño de listas de chequeo para la aplicación de las guías de buenas prácticas de seguridad del paciente. 2016.

HERNÁNDEZ Sampieri y otros. Metodología de la investigación. McGraw-Hill Interamericana.

HERNÁNDEZ Roberto, FERNÁNDEZ Carlos, BAPTISTA, Pilar. Metodología de la Investigación. México D.F, Interamericana Editores, S.A. de C.V. 2010, p.56

HERNÁNDEZ SAMPIERI, R., FERNÁNDEZ COLLADO, C., & BAPTISTA Lucio, P. Metodología

de la Investigación (6 ed.). México: Mc Graw Hill, 2014. p. 569.

IBM Corp. IBM SPSS Statistics para Windows, versión 28.0.1 2020 Armonk, Nueva York: IBM Corp.

J.C, Datos. Introducción a los sistemas de base de datos. 7 ed.

Ken Schwaber & Jeff Sutherland; La Guía Definitiva de Scrum: Las Reglas del Juego: noviembre 2022

Laravel, un framework de PHP, WHITEPAPERS

LARROUCAU DE MAGALHAES-CALVET, Tomas. Estudio de los factores determinantes de la deserción, en el sistema universitario chileno, 2013. p.1.

LATORRE, Marino. Historia del web, 1.0, 2.0, 3.0 y 4.0. Recuperado de Universidad Marcelino Champagnat. pdf, 2018.

Layne, M. Estudio del Big Data. Tesis de Licenciatura, Universidad de Sevilla. Recuperado.2017

LEÓN SOCHA, Fredy Alonso. Ingeniería de software 1. 1ed, noviembre de 2017, p9

LOZANO TREVIÑO, David Fernando y MALDONADO MALDONADO, Lauro. Asociación entre factores económicos y sociales con la propensión de deserción escolar en colegios militarizados.

Manual CRISP-DM de IBM SPSS Modeler, v15

Manual Microsoft office Excel 2010

María Guadalupe. Estadística descriptiva. Revista Alergia México, 2016, vol. 63, no 4, p. 397-407.

MARTÍNEZ MONCALEANO, Carlos Javier y FAJARDO, Ofelia Palencia. 2020. p.23-30. En: <http://www.scielo.org.co/pdf/sdn/v12n26/2027-5692-sdn-12-26-23.pdf>

- MARTÍNEZ, Rocío Erandi Barrientos, et al. Árboles de decisión como herramienta en el diagnóstico médico. Revista médica de la Universidad Veracruzana, 2009, vol. 9, no 2, p. 19-24.
- MAZA, Miguel Ángel Sánchez. JavaScript. Innovación Y Cualificación, 2012.
- MEN, la educación en Colombia [línea], 2016  
México, D.F.,2003
- MINA, Marcos Antonio Espinoza; CEDEÑO, Adriana Yolanda Sierra. Análisis comparativo entre ASP. NETy PHP. INNOVA Research Journal, 2018, vol. 3, no 4, p. 25-43
- MINISTERIO DE EDUCACIÓN NACIONAL, SISTEMA DE INFORMACIÓN DE EDUCACIÓN SUPERIOR y OBSERVATORIO DE LA UNIVERSIDAD DE COLOMBIA. Colombia. Bogotá. 2020.
- MINISTERIO DE EDUCACION NACIONAL. Estadísticas de deserción. 2018. p.1. En: [https://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article357549.html?\\_noredirect=1](https://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article357549.html?_noredirect=1)
- MONCAYO CABRERA, M. A. (2016). Determinantes que influyen en el rendimiento académico:
- MORA BRAVO, Eivar Alejandro. Estudio de factores que inciden en el retiro universitario y el rezago en los planes de estudio en los programas de pregrado de la Universidad Nacional de Colombia sede Palmira, 2017. p.9.
- OÑATE BOWEN, Álvaro Agustín. Análisis de la deserción y permanencia académica en la educación superior aplicando minería de datos, 2016. p.5.
- PEP, proyecto educativo del programa de ingeniería de sistemas, 2021
- PÉREZ, Javier Eguíluz. introducción a JavaScript. 2019.
- PLAZAS ESTEPA, r., & becerra rodríguez, r. (2016). una mirada a las pruebas saber pro en la calidad de la educación superior en Colombia. Pensamiento Republicano
- Quijano Vodniza, Armando José. Guía de Investigación Cuantitativa. [en línea], 2009, pág. 76, p.4
- PREET SINGH, Harman y NAFIL ALHULAIL, Hilal. Predicción del riesgo de abandono escolar de los estudiantes-docentes e identificación temprana: un enfoque de regresión logística de cuatro pasos. India. En: IEEE Access. Vol.10, p. 6470-6482, 2022, doi: 10.1109/ACCESS.2022.3141992
- RENDÓN-MACÍAS, Mario Enrique; VILLASÍS-KEEVER, Miguel Ángel; MIRANDA-NOVALES,  
Revista De Investigación E Innovación, 2016, Vol. 1 núm. p.65-70
- Ricoy, C. Contribución sobre los paradigmas de investigación. Revista do Centro de Educación, 31 2006, 11-22

- ROGER S, Pressman. Ingeniería de software I. 7ed
- SUÁREZ Patricio. Población de estudio y muestra. Curso de Metodología de la Investigación  
Unidad Docente de MF y C. 2011, p.35.
- SIBRIÁN ESCOBAR, Luisa Amelia. Expectativas de la formación universitaria desde la perspectiva de los estudiantes. Editorial Universitaria Don Bosco. No.19. 2017. p.25-37. En: <https://www.lamjol.info/index.php/DIALOGOS/article/view/5481/5149>
- TIMARAN P., Silvio Ricardo, JIMÉNEZ, Javier Alejandro y ROMERO C., Andrés. Detención de patrones de deserción estudiantil con minería de datos. En: Editorial universitaria-Universidad de Nariño, 2017, vol. 7
- TIMARÁN, R. y JIMÉNEZ, J. Extracción de perfiles de deserción estudiantil en la Institución Universitaria CESMAG, Colombia. IRE: Ciencias Sociales y Humanas, VI (1), 30-44. 2015. doi:
- TIMARÁN-PEREIRA, R., CAICEDO-ZAMBRANO, J. y HIDALGO-TROYA, A. Árboles de decisiones para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas saber 11°, 2019, Rev.investig. desarro. innova.
- TREVIÑO-REYES, Rosalba; RIVERA-RODRÍGUEZ, F.; GARZA-ALONSO, J. La analítica de datos como ventaja competitiva en las organizaciones. *VINCULATEGICA EFAN*, 2020, vol. 6, no 2 Ortiz, M. D., Joyanes, L., y Giraldo, L.M. Los desafíos del marketing en la era del big data. E Ciencias de la Información, 6(1), 2016,16-45
- TUDELA VIELA, Héctor. una aproximación teórica a la deserción estudiantil universitaria Revista digital de Investigación en Docencia universitaria, 2014, p. 59-76.
- UNIVERSIDAD Cesmag [[unicesmag.edu/ingeniería-en-sistemas](https://unicesmag.edu/ingeniería-en-sistemas)]. Pasto Nariño; [Consultado: 20 de febrero del 2022]. disponible en: <https://ingenieria.unicesmag.edu.co/index.php/ingenieria-en-sistemas>
- UNIVERSIDAD Cesmag [[unicesmag.edu/ingeniería-en-sistemas](https://unicesmag.edu/ingeniería-en-sistemas)]. Pasto Nariño; [Consultado: 20 de febrero del 2022]. disponible en: <https://ingenieria.unicesmag.edu.co/index.php/ingenieria-en-istemas>
- UNIVERSIDAD Cesmag, La institución-reglamento, [sitio web]. Pasto Nariño; [consultado en febrero de 2022]. Disponible en <https://www.unicesmag.edu.co/reglamentos/>
- UNIVERSIDAD Cesmag, PEI-UNICESMAG, [sitio web]. Pasto Nariño; [consultado en febrero de 2022]. Disponible en <https://www.unicesmag.edu.co/wp-content/uploads/2021/01/PEI-UNICESMAG-2020.pdf>
- USECHE, María Cristina, et al. Técnicas e instrumentos de recolección de datos cuali-cuantitativos. 2019
- VISAUTA VINACUA, Bienvenido. Análisis estadístico con SPSS para Windows. *Estadística*, 1997.

ZAPATA MEDINA, Daniel. Método para la detección de estudiantes en riesgo de deserción, basado en un diseño de métricas y una técnica de minería de datos. Universidad Nacional de Colombia,2021. p.9.

# ANEXOS

<b>CONTENIDO</b>		Pag
Descripción del software		127
Interfaz de ingreso		127
1.1. Descripción de la función de interfaz		127
1.2. Descripción de elementos de interfaz		127
2. Interfaz registro de usuario		128
2.1 Descripción de la función de interfaz		128
2.2. Descripción de elementos de interfaz		128
3. Interfaz modelos analíticos		129
3.1 Descripción de la función de interfaz		129
3.2. Descripción de elementos de interfaz		129
4. Interfaz modelo de predicción regresión lineal		130
4.1 Descripción de la función de interfaz		130
4.2. Descripción de elementos de interfaz		131
4.3. Descripción de elementos de interfaz		131
5. Interfaz modelo de predicción de clasificación		132
5.1. Descripción de elementos de interfaz		132
5.2. Descripción de elementos de interfaz		133
6. Interfaz de graficas		133
6.1. Descripción de elementos de interfaz		134
6.2. Descripción de elementos de interfaz		134
6.3. Descripción de elementos de interfaz		135
6.4. Descripción de elementos de interfaz		135
7. Interfaz de encuesta de satisfacción.		135
7.1. Descripción de elementos de interfaz		136
7.2. Descripción de elementos de interfaz		136

## Descripción del software

El sitio web tiene como objetivo principal buscar por medio de modelos de predicción la posible deserción o no del estudiante del programa de ingeniería de sistemas de la universidad CESMAG, en las siguientes secciones se describen las interfaces de usuario del sitio web.

### Interfaz de ingreso

En la siguiente figura se describe la interfaz de ingreso, la cual se debe acceder una vez registrado.

*Figura 1. Ingreso al sistema*



#### 1.1. Descripción de la función de interfaz

Dentro del sistema se debe ingresar un correo y contraseña una vez registrado dentro de la base de datos del sitio web, la cual dará el respectivo ingreso.

#### 1.2. Descripción de elementos de interfaz

A. Ingreso de usuario válido en el sistema

B. contraseña de nivel alto con letras mayúsculas, minúsculas y números

C. si el registro es satisfactorio se redirecciona a la interfaz donde se analizará la deserción, en caso contrario se tendrá que verificar los datos ingresados.

D. en caso de ingresar datos incorrectos de selecciona la opción limpiara para borrar todos los datos ingresados

E. redirecciona a la interfaz de registro de usuarios.

## 2. Interfaz registro de usuario

A continuación, se describe la funcionalidad que tiene el registro de nuevos usuarios dentro del sistema.

*Figura 2. Registro de usuarios*



The image shows a user registration form titled "Usuario nuevo". It features three input fields: "Nombre", "Correo electronico", and "Combina numeros y mayusculas". Below these fields are two buttons: "Registrar" (labeled B) and "Volver" (labeled C). A bracket labeled A groups the three input fields.

### 2.1 Descripción de la función de interfaz

En la siguiente interfaz se debe ingresar datos como: nombre, email y contraseña para la validación de un nuevo usuario.

### 2.2. Descripción de elementos de interfaz

A. se ingresa los datos básicos como nombre, email y contraseña para validación de usuario.

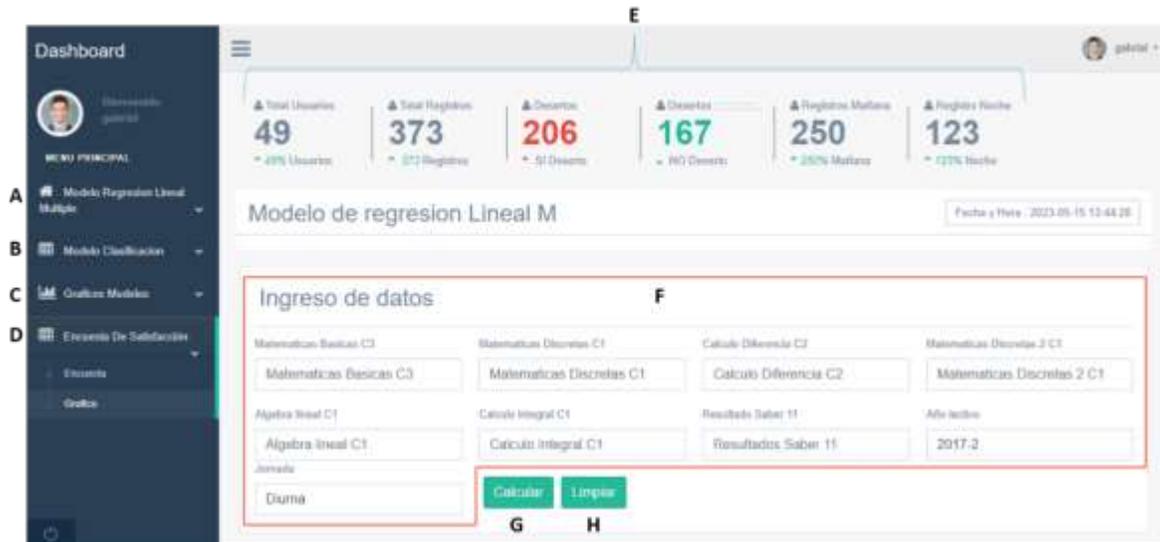
B. Si el registro es exitoso, lo redireccionara a la interfaz de inicio, como lo muestra la figura 1

C. lo redirecciona a la interfaz principal de ingreso al sistema

### 3. Interfaz modelos analíticos

A continuación, se describe la interfaz de los modelos matemáticos con los cuales se realizó el respectivo análisis en cada uno de ellos.

Figura 3. Interfaz modelos



#### 3.1 Descripción de la función de interfaz

En la siguiente interfaz se describe el funcionamiento de los modelos matemáticos como: regresión lineal y modelo de clasificación para el análisis de la deserción en estudiantes.

#### 3.2. Descripción de elementos de interfaz

A. En la pestaña del modelo de regresión lineal se visualizarán dos módulos de los cuales se hablará más adelante como: módulo de ciencias básicas y módulo de comunicación oral y escrita.

B. En la pestaña del modelo de clasificación se visualizarán dos módulos de los cuales se hablará más adelante como: módulo de ciencias básicas y módulo de comunicación oral y escrita.

C. En la pestaña gráficos módulos se visualizarán las gráficas respectivas de cada uno de los módulos que se mencionó anteriormente.

D. Al presionar la pestaña de encuesta de satisfacción lo redirigirá a una serie de preguntas sobre la calificación del software.

E. En la siguiente sección se visualizarán el total de los datos que se han ingresado en el sistema de acuerdo a la información que arrojo la base de datos.

F. En esta sección se encuentran alojadas todas las variables de estudio que se analizó en el módulo de ciencias básicas.

G. Al presionar el botón calcular se observarán los resultados del análisis en el módulo para determinar la deserción.

H. al presionar el botón limpiar se eliminarán todos los datos ingresados en las variables de estudio.

#### 4. Interfaz modelo de predicción regresión lineal

A continuación, se describe la interfaz del modelo matemático de regresión lineal con dos módulos: ciencias básicas y comunicación oral y escrita.

Figura 4. Interfaz modelo regresión lineal- modulo ciencias básicas



#### 4.1 Descripción de la función de interfaz

El modelo matemático de regresión lineal muestra dos módulos con los cuales se realiza los respectivos análisis.

#### 4.2. Descripción de elementos de interfaz

A. Pestaña donde se visualizarán los modelos con los cuales se realizarán el análisis.

B. Al presionar la pestaña del módulo de ciencias básicas se visualizarán las variables a ingresar dentro del sistema.

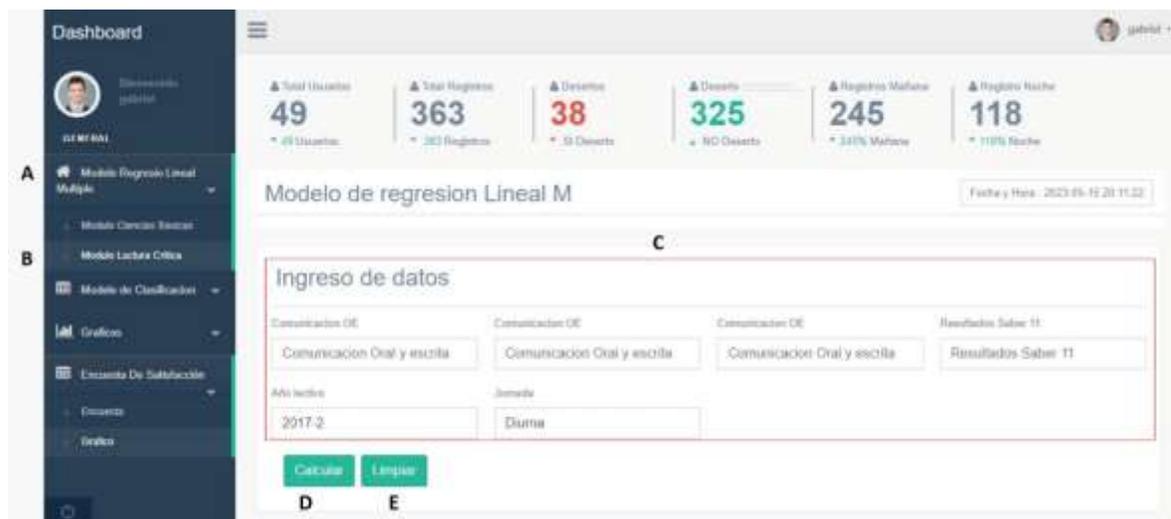
C. Pestaña donde se visualizarán las variables a ingresar dentro del módulo de comunicación oral y escrita.

D. Al presionar el botón calcular se observarán los resultados del análisis en el módulo para determinar la deserción.

E. Al presionar el botón limpiar se eliminarán todos los datos ingresados en las variables de estudio.

F. Al presionar el botón de apagado saldrá de la interfaz de los módulos y lo redirigirá a la interfaz de ingreso al sistema.

Figura 5. Interfaz modelo regresión lineal- modulo comunicación oral y escrita



#### 4.3. Descripción de elementos de interfaz

A. Pestaña donde se visualizarán los modelos con los cuales se realizarán el análisis.

B. Al presionar la pestaña del módulo de lectura crítica se visualizarán las variables a ingresar dentro del sistema.

C. Variables de estudio del módulo comunicación oral y escrita.

D. Al presionar el botón calcular se observarán los resultados del análisis en el módulo para determinar la deserción.

E. Al presionar el botón limpiar se eliminarán todos los datos ingresados en las variables de estudio.

## 5. Interfaz modelo de predicción de clasificación

A continuación, se describe la interfaz del modelo matemático de clasificación con dos módulos: ciencias básicas y comunicación oral y escrita.

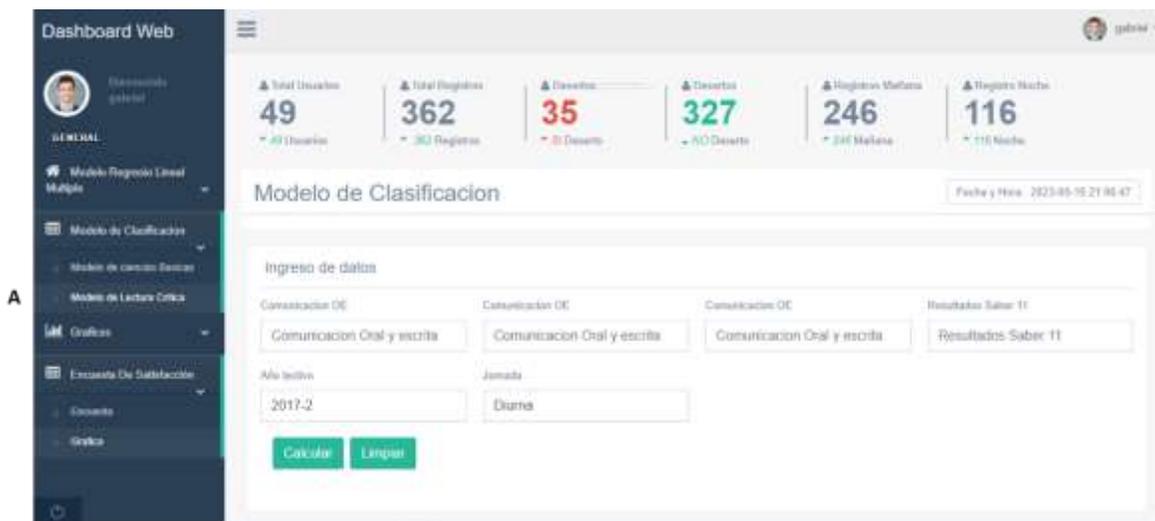
Figura 6. Interfaz modelo de clasificación - modulo ciencias básicas



### 5.1. Descripción de elementos de interfaz

Para el modelo de clasificación se obtienen las mismas variables de estudio que el modelo de regresión lineal, pero con predicciones no tan precisas, es por eso que en su interfaz muestra las mismas funcionalidades.

Figura 7. Interfaz modelo de clasificación - modulo comunicación oral y escrita



## 5.2. Descripción de elementos de interfaz

Sección que se encuentra alojada en la interfaz con las mismas variables de estudio que en el modelo de regresión lineal en el módulo de lectura crítica, el cual cumple con el mismo funcionamiento, pero con predicciones no tan precisas.

## 6. Interfaz de graficas

A continuación, se describe la interfaz de cada una de las gráficas que se realizaron dentro de la investigación, en los módulos de ciencias básicas y comunicación oral y escrita.

Figura 7. interfaz gráfica – modulo ciencias básicas



### 6.1. Descripción de elementos de interfaz

Esta funcionalidad permite visualizar graficas de los datos ingresados en cada uno de los módulos de estudio, los cuales se van alimentando y actualizando con datos que anteriormente se ingresó.

### 6.2. Descripción de elementos de interfaz

A. Al desplegar la pestaña gráficos ciencias básicas, se obtiene los resultados gráficamente de cada uno de los datos ingresados anteriormente en el módulo.

Figura 8. interfaz gráfica – módulo de comunicación oral y escrita



### 6.3. Descripción de elementos de interfaz

Esta funcionalidad permite visualizar graficas de los datos ingresados en cada uno de los módulos de estudio, los cuales se van alimentando y actualizando con datos que anteriormente se ingresó.

### 6.4. Descripción de elementos de interfaz

A. Al desplegar la pestaña gráficos lectura crítica, se obtiene los resultados gráficamente de cada uno de los datos ingresados anteriormente en el módulo.

## 7. Interfaz de encuesta de satisfacción.

A continuación, se describe la funcionalidad de la interfaz de encuesta de satisfacción dentro del sistema.

Figura 9. Interfaz encuesta



### 7.1. Descripción de elementos de interfaz

Esta funcionalidad permite acceder a la encuesta de satisfacción del cliente después de utilizar el sitio web.

### 7.2. Descripción de elementos de interfaz

A. Al acceder en el botón de encuesta se desplegará una ventana donde se visualizará preguntas acordes al manejo del sitio web y su funcionamiento, el cual se lo calificará por medio de estrellas.

 <p>UNIVERSIDAD <b>CESMAG</b> NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	<b>CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)</b>	<b>CÓDIGO:</b> AAC-BL-FR-032
		<b>VERSIÓN:</b> 1
		<b>FECHA:</b> 09/JUN/2022

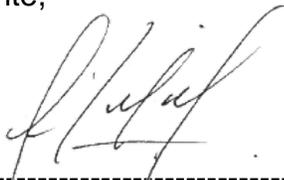
San Juan de Pasto, 15 De Julio del 2023

Biblioteca  
**REMIGIO FIORE FORTEZZA OFM. CAP.**  
Universidad CESMAG  
Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominad: **Modelos de predicción de la deserción estudiantil durante los tres primeros semestres del programa de ingeniería de sistemas mediante técnicas de minería de datos utilizando un sitio web**, presentado por los autores Esteban Andrés Urbina bravo y Luis Gabriel Carlosama daza del Programa Académico ingeniería al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a), que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,



-----  
**Mg. CARLOS FERNANDO GONZÁLEZ GUZMÁN**  
Director Programa de Ingeniería de Sistemas  
UNICESMAG  
Calle 20 No. 14 – 54 Pasto, Nariño - Sede C  
[www.unicesmag.edu.co](http://www.unicesmag.edu.co)

 <b>UNIVERSIDAD CESMAG</b> <small>NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</small>	<b>AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL</b>	<b>CÓDIGO:</b> AAC-BL-FR-031
		<b>VERSIÓN:</b> 1
		<b>FECHA:</b> 09/JUN/2022

<b>INFORMACIÓN DEL (LOS) AUTOR(ES)</b>	
<b>Nombres y apellidos del autor:</b> <b>LUIS GABRIEL CARLOSAMA DAZA</b>	<b>Documento de identidad:</b> <b>1085303785</b>
<b>Correo electrónico:</b> <b>lcarlosama32@gmail.com</b>	<b>Número de contacto:</b> <b>311 3551656</b>
<b>Nombres y apellidos del autor:</b> <b>ESTEBAN ANDRES URBINA BRAVO</b>	<b>Documento de identidad:</b> 1085312696
<b>Correo electrónico:</b> <b>e.andres2012@hotmail.com</b>	<b>Número de contacto:</b> 3187170062
<b>Nombres y apellidos del asesor:</b> <b>CARLOS FERNANDO GONZÁLEZ GUZMÁN</b>	<b>Documento de identidad:</b>
<b>Correo electrónico:</b> <b>cfgonzalez@unicesmag.edu.co</b>	<b>Número de contacto:</b>
<b>Título del trabajo de grado:</b> <b>MODELOS DE PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL DURANTE LOS TRES PRIMEROS SEMESTRES DEL PROGRAMA DE INGENIERÍA DE SISTEMAS MEDIANTE TÉCNICAS DE MINERÍA DE DATOS UTILIZANDO UN SITIO WEB</b>	
<b>Facultad y Programa Académico:</b> <b>Faculta de ingeniería: Programa ingeniería de sistemas</b>	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- a) La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.
- b) Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
- c) Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación,

 <p>UNIVERSIDAD <b>CESMAG</b> NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	<b>AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL</b>	<b>CÓDIGO:</b> AAC-BL-FR-031
		<b>VERSIÓN:</b> 1
		<b>FECHA:</b> 09/JUN/2022

distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.

- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

**NOTA:** En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG, por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 15 días del mes de junio del año 2023

<i>Esteban Urbina</i> Firma del autor	 Firma del autor
Nombre del autor: <b>ESTEBAN URBINA</b>	Nombre del autor: <b>LUIS CARLOSAMA</b>
Firma del asesor _____ Nombre del asesor:	