

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

Tema: “Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar”

Trabajo de Titulación previo a la obtención del Grado Académico de Magister en
Matemática Aplicada

Modalidad de titulación Proyecto de Desarrollo

Autor: Lic. Edgar Patricio Mejía Zamora

Director: Ing. Marco Enrique Benalcázar Palacios, Ph.D.

Ambato-Ecuador

2023

i

APROBACIÓN DEL TRABAJO DE TITULACIÓN

A la Unidad de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica E Industrial.

El Tribunal receptor de la Defensa del Trabajo de Titulación presidido por la Ing. Elsa Pilar Urrutia Urrutia Mg., e integrado por los señores: Ing. Oscar Eduardo Ruiz Robalino, Mg. e Ing. Víctor Santiago Manzano Villafuerte Mg., designados por la Unidad de Titulación de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar”, elaborado y presentado por el señor Licenciado Edgar Patricio Mejía Zamora, para optar por el Grado Académico de Magister en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato

Ing. Elsa Pilar Urrutia Urrutia Mg.
Presidente y Miembro del Tribunal de Defensa

Ing. Víctor Santiago Manzano Villafuerte, Mg
Miembro del Tribunal de Defensa

Ing. Oscar Eduardo Ruiz Robalino, Mg
Miembro del Tribunal de Defensa

AUTORÍA DEL TRABAJO DE TITULACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el Trabajo de Titulación presentado con el tema: “Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar”, le corresponde exclusivamente al Licenciado Edgar Patricio Mejía Zamora, Autor bajo la Dirección de Ingeniero Marco Enrique Benalcázar Palacios, Ph.D. Director del Trabajo de Titulación; y el patrimonio intelectual a la Universidad Técnica de Ambato.

Lic. Edgar Patricio Mejía Zamora.
AUTOR

Ing. Marco Enrique Benalcázar Palacios, Ph.D.
DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el Trabajo de Titulación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos de mi Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

Lic. Edgar Patricio Mejía Zamora
C.C. 1801769058

ÍNDICE GENERAL

PORTADA.....	i
APROBACIÓN DEL TRABAJO DE TITULACIÓN	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN	iii
ÍNDICE GENERAL.....	v
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS.....	viii
AGRADECIMIENTO	ix
DEDICATORIA	x
RESUMEN EJECUTIVO	xi
EXECUTIVE SUMMARY.....	1
CAPÍTULO I.....	3
EL PROBLEMA DE INVESTIGACIÓN.....	3
1.1. Planteamiento.....	3
1.2. Justificación	4
1.3. Objetivos.....	4
1.3.1. General	4
1.3.2. Específicos.....	4
CAPITULO II.....	5
ANTECEDENTES INVESTIGATIVOS	5
CAPITULO III.....	19
MARCO METODOLÓGICO.....	19
3.1. Ubicación.....	19
3.2. Equipos y materiales.....	19
3.3. Tipo de investigación.....	20
3.4. Prueba de hipótesis	20
3.5 Recolección de información.....	20
CAPITULO IV	22
RESULTADOS Y DISCUSIÓN	22
CAPITULO V	30
CONCLUSIONES, RECOMENDACIONES, BIBLIOGRAFÍA Y ANEXOS.....	30
5.1. Conclusiones.....	30

5.2. Recomendaciones	32
5.3. BIBLIOGRAFÍA	33
5.4 ANEXOS	36

ÍNDICE DE TABLAS

Tabla 1. Resultado con datos de testeo en SVM	23
Tabla 2. Resultado con datos de testeo en RNA feed forward	24
Tabla 3. Resultados finales en SVM.....	25
Tabla 4. Resultados finales en RNA feed forward entrenado.....	25
Tabla 5. Resultados finales entrenados en SVM y RNA feed -forward	28

ÍNDICE DE FIGURAS

Figura 1. Estructura de una Red Neuronal	14
Figura 2. Vectores de Soporte	15
Figura 3. Ubicación geográfica del ISTB.....	19
Figura 4. Resultado de Matriz de Confusión datos desbalanceados.....	25
Figura 5. Resultado de Matriz de Confusión datos balanceados	25
Figura 6. Resultado de Matriz de Confusión con RNA datos desbalanceados	27
Figura 7. Resultado de Matriz de Confusión con RNA datos balanceados	27

Otros Anexos

Anexo 1. Resultado de la institución del encuestado.	36
Anexo 2. Resultado del estado civil	36
Anexo 3. Resultado de la edad del encuestado.	37
Anexo 4. Resultado de gastos para la educación.....	37
Anexo 5. Resultado de cargas familiares del encuestado.....	38
Anexo 6. Resultado del semestre actual del encuestado.	38
Anexo 7. Resultado de la provincia de residencia actual del encuestado.....	39
Anexo 8. Resultado de la dificultad del proceso de aprendizaje.....	39
Anexo 9. Resultado del desequilibrio económico del encuestado.	40
Anexo 10. Resultado del apoyo del docente.	40
Anexo 11. Resultado del cálculo de la muestra poblacional	40
Anexo 12. Preguntas elaboradas a estudiantes No desertores.....	41
Anexo 13. Encuesta Elaborada y enviada a los estudiantes desertores	41
Anexo 14. Resultado de la encuesta aplicada a los estudiantes No desertores	42
Anexo 15. Resultado de la encuesta aplicada a los estudiantes Desertores	42
Anexo 16. Ponderado a aplicarse entre 0 y 1	43
Anexo 17. Resultados con datos ponderados	43
Anexo 18. Resultado de prueba con 41 ensayos en RNA feed-forward	44
Anexo 19. Resultado de prueba con 41 ensayos en RNA	44
Anexo 20. Resultado de prueba con 41 ensayos en SVM.....	44
Anexo 21. Resultado de prueba con 41 ensayos en SVM.....	45
Anexo 22. Programa en Matlab del entrenamiento para obtener los resultados en Matriz de confusión con SVM	45
Anexo 23. Programa en Matlab del entrenamiento para obtener los resultados en Matriz de confusión con RNA feed forward	46

AGRADECIMIENTO

A Dios por darme la vida, a mis padres por ser el pilar para el progreso tanto profesional y humano.

A mi esposa Mónica a mis hijos Andrea y Erick por el gran apoyo dado en este camino largo y duro de la maestría

Al director de trabajo de titulación Ing. Marco Enrique Benalcázar Palacios, Ph. D por compartir sus excelentes conocimientos y su don de gente.

A todos los docentes que a su tiempo compartieron los módulos en la maestría A los directivos de la Universidad Técnica de Ambato por organizado la presente maestría

DEDICATORIA

El presente trabajo de investigación está dedicado de forma muy especial a mi familia que me apoyó en todo momento para alcanzar el objetivo propuesto que con la bendición de Dios se hace realidad.

También a mis padres que siendo dos son un solo ejemplo quienes me inculcaron principios y valores

UNIVERSIDAD TÉCNICA DE AMBATO

FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

TEMA:

“Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar”

AUTOR: Lic. Edgar Patricio Mejía Zamora

DIRECTOR: Ing. Marco Enrique Benalcázar Palacios, Ph.D.

LÍNEA DE INVESTIGACIÓN: Tecnología de la Información y sistemas de control

FECHA: 4 de enero de 2023

RESUMEN EJECUTIVO

La deserción estudiantil es un indicador de retraso en la educación de cualquier país del mundo y esto ocurre con mayor impacto en el Ecuador. Por esta razón es importante tomar en cuenta el trabajo de investigación que se presenta a través de la siguiente tesis. La presente investigación tiene como finalidad predecir la deserción de los estudiantes del Instituto Superior Tecnológico Bolívar de la ciudad de Ambato. En los archivos del Instituto constan que se matricularon 1036 estudiantes en el semestre mayo – octubre 2022 en cinco carreras.

El tamaño de la muestra calculado es de 280 encuestas, usando un nivel de confianza del 95% y un margen de error del 5%. Las encuestas se aplicaron a dos grupos: el primero a 115 estudiantes que por cualquier motivo abandonaron sus estudios en cualquier semestre, pero regresaron a continuar la carrera. Un segundo grupo de 165 estudiantes que no desertaron en ningún momento, para la encuesta se usó de Google forms a través de preguntas de opción múltiple. Por tener un dataset asimétrico se requiere balancear los datos para que los modelos a utilizarse funcionen correctamente.

En el presente trabajo se hace uso de dos algoritmos, Máquinas de Vectores de Soporte (SVM) y Redes Neuronales Artificiales Feed-forward. Al comparar los resultados obtenidos en función del F1 score se determinó que el mejor algoritmo fue Redes Neuronales Artificiales Feed-forward con un valor de F1score de 82,2% mientras que en Máquinas de Vectores de Soporte arrojó un valor de 72,5 %.

DESCRIPTORES: Deserción, Balanceo de datos, validación, regresión lineal binaria, modelos matemáticos, variables dependientes, validación cruzada.

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL
MAESTRÍA EN MATEMÁTICA APLICADA

THEME:

Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Bolívar

AUTHOR: Lic. Edgar Patricio Mejía Zamora.

DIRECTED BY: Ing. Marco Enrique Benalcázar Palacios, Ph.D.

LINE OF RESEARCH: Technology of Information and Control System.

DATE : January 4, 2023

EXECUTIVE SUMMARY

Student desertion is an indicator of delay in education in any country in the world and this occurs with a greater impact in Ecuador. For this reason, it is important to take into account the research work that is presented through the following thesis. The purpose of this research is to predict the desertion of students from the Bolívar Higher Technological Institute in the city of Ambato. The Institute's archives show that 1,036 students were enrolled in the May-October 2022 semester in five majors.

The calculated sample size is 280 surveys using a confidence level of 95% and a margin of error of 5%. The surveys were applied to two groups, the first to: 115 students who for any reason abandoned their studies in any semester but returned to continue their studies. A second group of 165 students who did not drop out at any time, used Google forms for the survey through multiple choice questions . Due to having an asymmetric dataset, it is necessary to balance the data so that the models to be used work correctly.

In the present work, two algorithms are used: Support Vector Machines (SVM) and Feed-forward Artificial Neural Networks. When comparing the results obtained based on the F1 score, it was determined that the best algorithm was Feed-forward Artificial Neural Networks with an F-score value of 82.2% while in Support Vector Machines it yielded a value of 72.5 %.

DESCRIPTORS: Attrition, Data balancing, validation, binary linear regression, mathematical models, dependent variables, cross-validation

CAPÍTULO I

EL PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento

La deserción estudiantil en todos los niveles y sistemas educativos de Latinoamérica constituye una problemática en el proceso de obtención de un título superior. En nuestro país la deserción estudiantil no es la excepción, pues constituye un fenómeno muy notorio y su alcance e incidencia es alto en el proceso educativo [1]. Este problema ocasiona un gran costo social y económico en la educación pública igual en la privada y afecta de manera directa en el crecimiento de la economía [2]. A un estudiante se le considera desertor cuando se descarta la posibilidad de que retome los estudios en un mismo período. En la Educación Superior es más notorio el abandono de los estudios por lo que en la actual investigación se pretende investigar los motivos de la deserción y plantear un modelo de predicción [3].

1.2. Justificación

En los sistemas educativos de Latinoamérica una problemática significativa es la deserción estudiantil, en todos los niveles de educación. Por esta razón los gobiernos implementan medidas políticas públicas para combatir la deserción estudiantil [4]. La ministra de Educación del Ecuador en el 2021 manifestó que debido a la pandemia más de 90.000 estudiantes no asistieron a clases, con lo cual los índices de deserción estudiantil ha ido en aumento [5].

En el Ecuador la educación superior es gratuita, según lo dispuesto en el artículo 356 de la constitución política, hasta el tercer nivel, los gobiernos garantizarán los fondos necesarios para su cumplimiento [6]. El derecho a una educación superior establece la semejanza de oportunidades en función de los méritos tanto académica como profesional. Cuando los estudiantes no han alcanzado ciertos niveles en la educación, afecta de manera directa en el crecimiento en las economías del estado [7].

En el Instituto Superior Bolívar en el periodo académico mayo - octubre 2022 un gran número de estudiantes no terminó el ciclo de estudios, es decir, son considerados como desertores; los mismos que por diferente causa o motivo tomaron la decisión de no seguir estudiando para dedicarse a otras actividades. La presente investigación, ayudará a las autoridades para una posible toma de decisiones para prevenir el índice de desertores [8]. En la actualidad, quien no posea por lo menos un título de tercer nivel le es muy difícil aplicar a una oferta laboral en instituciones públicas o privadas

1.3. Objetivos

1.3.1. General

Determinar el grado de deserción mediante un modelo matemático para los estudiantes del Instituto Superior Tecnológico Bolívar.

1.3.2. Específicos

- Analizar las variables socioeconómicas mediante el modelo de Redes Neuronales Feed-forward y Máquinas de Vectores de Soporte para los estudiantes del instituto Superior Tecnológico Bolívar
- Definir el modelo de mejor ajuste a las necesidades del ISTB para predecir un posible caso de deserción
- Establecer el grado de deserción estudiantil para el Instituto Superior Tecnológico Bolívar mediante el modelo matemático con el uso de software de programación.

CAPITULO II

ANTECEDENTES INVESTIGATIVOS

2.1 Estado del arte

González y Arismendi, en su trabajo “Deserción Estudiantil en la Educación Superior Técnico Profesional: Analiza los factores que influyen en los alumnos de primer año” [9], lograron identificar diferentes variables que explican la deserción de los estudiantes utilizando datos obtenidos de la cohorte 2014-2016. La muestra que se toma en consideración fue de 1876 estudiantes de primer año del instituto técnico-profesional ubicado en la región de Valparaíso, los datos se procesaron con el software R. Los resultados arrojados fueron que, los factores con mayor relevancia estadística corresponden al ‘género’, ‘año de egreso de enseñanza media’ y ‘jornada de estudio’. Al final, es importante señalar que todo esto depende de la escuela de formación, se dice que el modelo cambia de acuerdo con las características de cada campo de formación.

En la investigación de Pérez, Escobar, Toledo, Gutiérrez y Reyes con título “Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O’ Higgins” [10]. El objetivo fue diseñar un modelo predictivo para la captación de los estudiantes de primer año de la Universidad Bernardo O’ Higgins (Santiago de Chile). Se identifican las variables consideradas de entrada, sean estas académicas, sociales o familiares, identificadas como significativas para el análisis. A partir de la automatización de los historiales de las cohortes de los años 2014 y 2015 se elabora una matriz comprendida en tres partes con los datos relacionados a las variables que sale del análisis de los estudiantes consultados. Posteriormente, el análisis de dos en dos variables permite determinar diecisiete variables, notablemente ligadas con la deserción estudiantil. Con esto se puede descubrir las relaciones de dependencia con la decisión de abandono de estudios. Hicieron uso del modelo multivariado que predice en un 86,4 % la conducta de abandono indicando siete variables categóricas independientes que, al final se muestran como factores sobresalientes del modelo de predicción.

Los autores Esteban, Bernardo, Tuero, Cervero y Casanova [11]. En la investigación indica que primero es necesario indicar que, a pesar de que la muestra analizada en la investigación llegaba a 1.311 participantes, se excluye varios grupos. Luego, los análisis estadísticos son aplicados sobre una muestra de 677 alumnos (198 abandonan y 479 permanecen). Al inicio, se implementan en un análisis descriptivo y correlacional, cuyos resultados indican el predominio de un amplio grupo de variables.

El modelo terminado, incluyendo 11 variables y los predictores estadísticamente valiosos, $X^2 = (11, N = 677) = 410.974$, con un $p < 0.01$, señala que el modelo es capaz de diferenciar, acertadamente en un 70.5% de los casos, a los estudiantes que permanecen en la carrera de los que abandonan. Por otro lado, el modelo ajustado muestra que entre el 45.4% (R^2 Cox & Snell) y el 64.6% (R^2 Nagelkerke) de la varianza, clasifica acertadamente el 87.3% de los casos. Se determina que solo 3 de las 11 variables contribuyen de manera sustancialmente al modelo “elección de la titulación por vocación, porcentaje de créditos aprobados sobre el total de la titulación y porcentaje de créditos suspensos sobre el total de la titulación”. Los resultados muestran que el predictor más fuerte es la estabilidad del estudiante en la institución superior es su vocación, con una odds ratio de 2.099. Esto indica que aquellos alumnos que tienen desde su ingreso la vocación obtienen el doble de posibilidades de permanecer en la carrera hasta completarla, con respecto a aquellos alumnos que no escogieron a esta como principal motivación. Por otro lado, el rendimiento académico tiene un rol significativo menor, y es de baja magnitud.

El trabajo de Tapasco, Ruiz, Osorio y Ramírez, con su tema “Deserción estudiantil: incidencia de factores institucionales relacionados con los procesos de admisión” [12], estudian la incidencia que distintas variables instauradas en el proceso de admisión para determinar el riesgo de deserción estudiantil. La población del estudio es los estudiantes legalmente matriculados en la Universidad de Caldas para las cohortes 2010-2012. Los modelos analizados arrojan ciertas variables en el ingreso o admisión dispuestas a ser analizadas, como el ‘ingreso por segunda ocasión’, el ‘puntaje de admisión’, el ‘ingreso por segunda ocasión’, y el ‘valor económico de matrícula’ tienen un suceso significativo en el riesgo de desertar. En conclusión, para mermar la deserción estudiantil e incrementar el porcentaje de predecir el desempeño futuro del estudiante, se sugiere contemplar acciones como el repensar la admisión por segunda ocasión, se debe revisar periódicamente las políticas de ingreso por méritos académicos, de acción afirmativa también se debe modificar las ponderaciones actualmente asignadas a las áreas valoradas en el proceso de admisión.

Por otro lado, Albarrán en su investigación “La deserción estudiantil en la Universidad de Los Andes (Venezuela)” [13]. El autor indicó que consultó varias fuentes bibliográficas y, por no tener datos oficiales del gobierno que midan el fenómeno, utilizó información sacada de estimaciones de organismos no gubernamentales. Los resultados revelan que 50% de los estudiantes pueden dejar la institución educativa superior. Esto se debe a la falta de plazas laborales, aumento de sus egresos tanto personales como académicos, también influye la suspensión de las labores académicas por las regulares protestas sociales, falta de muchos recursos económicos, mínima formación escolar en la secundaria y desmotivación. El estudio indica que la deserción en la ULA es un fenómeno que va en constante aumento a medida que la crisis económica que sufre el país se siga agravando. Por lo tanto,

se expuso que se debe dar más apoyo en lo motivacional y por ende en lo académico a todos los estudiantes con el propósito de que continúen en la institución.

De igual manera Torres, A., Acuña, J., & Moguel, J. Los términos rezago, reprobación, y deserción, van tomando fuerza como indicadores que influye en la calidad educativa. La estadística internacional ubica a México como un país con mayores deserciones. Dado que la reprobación es sinónimo del fracaso escolar, pasó a ser un tema que preocupa y mantiene ocupados a todos los implicados en el sistema educativo.

Al inicio del trabajo el propósito fue determinar el índice existente de reprobados en los estudiantes de los varios programas educativos ofertados en la Facultad de Ciencias de la Salud de la Universidad Autónoma del Carmen. La investigación tiene un enfoque “cuantitativo, de alcance descriptivo, con diseño no experimental y delineación transversal retrospectivo”. La muestra recogida es de tipo no probabilístico esto por convenio. El índice de reprobación da como resultado 43.22 ± 7.65 , sobre lo recomendado por los organismos evaluadores de la ciudad de México [14].

De igual forma Zambrano, Rodríguez y Guevara en su revisión bibliográfica “Deserción estudiantil en las universidades del Ecuador y América” [15]. Los estudios arrojan resultados que, en los establecimientos de educación superior, muchos estudiantes tienen dificultades durante el tiempo que dura sus estudios, llegan a reprobado varias asignaturas y en el peor de los casos deciden retirarse de la universidad. Identificar esta realidad puede servir como guía a las autoridades a la toma de decisiones académicas para enfrentar esta problemática y mejorar la eficiencia final, que sirve como un indicador valioso para la acreditación de cada institución. El objetivo de estudio de este tema es determinar las tasas de deserción y reprobación de los alumnos de cualquier facultad, y en qué instante de los estudios estas tasas son más altas.

La investigación realizada a 1900 estudiantes de la Universidad de Cuenca, iniciado la cohorte en marzo del 2013, y se realiza un seguimiento de tipo longitudinal terminando en julio del 2018. Aquí se hizo uso de la técnica de investigación documental, para esto se procesa la base de datos académica que reposa en los archivos de la universidad. Se identifica que, tanto las tasas de deserción como la de repetición son más altas en los primeros períodos de estudio en todas las facultades. También se evidencia que en las carreras de Ciencias Médicas y Arquitectura gozan bajos índices de deserción y reprobación de asignaturas del alumnado, mientras que en otras facultades del área de Ciencias Sociales poseen elevados índices de deserción y reprobación. Es de vital importancia conocer estos indicadores para que cada una de las facultades realice un seguimiento sobre el historial académico de los estudiantes, para que las autoridades sepan tomar decisiones de mejora para la oferta académica.

Según Camacho, Montalvo y Galezo en una investigación efectuada “Determinantes de la deserción estudiantil en estudiantes universitarios” [16], la deserción de estudiantes es un problema muy grave para el sistema educativo. El trabajo se centraliza en un estudio de índole descriptivo, correlacional que considera a los 16.927 alumnos de pregrado legalmente matriculados en diferentes cohortes en la Universidad de Cartagena. Se establecen los posibles ítems con el riesgo de desertar y/o patrones transitorios de deserción. Para los datos, se pide autorización previa en la plataforma de SPADIES. Se toma en cuenta una muestra de 1547 estudiantes que desertaron anteriormente se evidencia que es influyente a cada programa y se desarrolló un instrumento con veinte y dos preguntas, ese instrumento es plasmado por varios investigadores. Se encuentra que, en los últimos períodos, la tasa de deserción en la Universidad tiende a disminuir, y está en correspondencia con la tasa promedio de deserción, considerando el nivel nacional. De acuerdo con el estudio, los alumnos de los programas que eligen la modalidad a distancia presentan mucho más riesgo de abandonar los estudios. Para quienes ingresan con una calificación baja en las pruebas con nombre Saber presentan un mayor rango de deserción al inicio de los primeros ciclos de estudio.

En el estudio de los escritores Vizúete y Chasiluisa con el tema “Deserción escolar de los estudiantes del colegio técnico Vicente Rocafuerte de la ciudad de Quito de la especialidad de mecánica automotriz y la repercusión en el ámbito laboral (año 2017-2018)” [17]. Los investigadores estudiaron el abandono escolar y la influencia en los estudiantes del colegio Técnico Vicente Rocafuerte en el ciclo escolar 2017. Esto lo hicieron con el anhelo de determinar los factores influyentes en la deserción educativa y sus consecuencias en el campo laboral. Para esto se considera efectuar una investigación científica fundamentada en cierta bibliografía orientada en la socio-pedagogía, además se utiliza la encuesta a estudiantes del colegio. Obtuvieron como conclusión que la deserción de los estudiantes repercute en la gran mayoría de los casos negativamente y que afecta principalmente a aquellos estudiantes de segundo y tercer año de bachillerato de la institución en estudio.

Los autores Machado y Castillo en su proyecto de titulación “Modelado del tiempo de estancia estudiantil en la FICFM de la Universidad Central del Ecuador” [18]. Los autores utilizan un modelo no paramétrico bi-dimensional para determinar la probabilidad de desertar de los estudiantes que pertenecen a la Facultad de Ingeniería, Ciencias Físicas y Matemática de la Universidad Central del Ecuador. Se fundamentan en las siguientes variables aleatorias, tiempo de estancia al igual que el porcentaje de créditos aprobados, se acudió a los registros existentes en la Facultad. Se aplica un modelo para proyectar la disminución en la población de alumnos de la facultad FICFM de la Universidad Central del Ecuador. En base a los vectores recomendados y haciendo uso del software libre R, se logra determinar la función de densidad de las variables” tiempo de estancia” y “porcentaje

de créditos aprobados”, valiéndose del método de interpolación de densidad de kernel.

Los autores proceden a encontrar el valor de probabilidad que ocurra la deserción de estudiantes, mediante el uso de métodos matemáticos. Estos métodos facilitan hallar el volumen por debajo de la superficie por la densidad conjunta y la fracción de área donde se define la mencionada densidad. Luego los autores con esos datos que obtienen manifiestan que aplicaron los modelos de supervivencia y demográficos, con los que analizaron el incremento de estudiantes de la FICFM de la UCE mediante tablas de permanencia.

De igual forma Reyes, Rosero y Medina con tema “Deserción escolar en estudiantes de 5 a 17 años en las Provincias de Pichincha y Guayas en el período 2013-2017” [19]. Identifican los factores que influyen en la deserción en alumnos del período indicado, en las dos provincias para ello se utilizan la base de datos existentes de la ENEMDU. Luego se determina la probabilidad de deserción estudiantil utilizando el modelo conocido como de probabilidad de efectos fijos, lo que permite identificar en que área y en qué grado de educación existe más deserción. Se obtiene como resultado significativo que la población estudiantil que cursa los últimos años de Bachillerato General Unificado tiende a desertar. También señalan que mientras exista mayor oferta de estudiantes en la especial en las provincias de Guayas y Pichincha mayor es el grado de deserción estudiantil especialmente de mujeres. Esto se debe especialmente a factores como falta de mínimos recursos económicos y la necesidad de buscar empleo a temprana edad. En la provincia del Guayas existe el mayor número de desertores y el área que predomina es del sector rural ya que no cuentan con los recursos económicos necesarios y suficientes para terminar los estudios.

La autora Rodríguez en su investigación “análisis de los factores que inciden en la deserción estudiantil de los niños, niñas y adolescentes ecuatorianos en el periodo 2009 - 2019” [20]. Manifiesta que el abandono del sistema educativo generado por los estudiantes influye directamente en el progreso económico, social, político, y cultural de la sociedad. Para enfrentar este problema recomienda analizar las diversas causas que lo producen. Para este fin la investigadora utiliza un modelo de respuesta binaria Probit con la fuente de datos de la Encuesta Nacional de Empleo, Desempeño y Subempleo (ENEMDU). En el archivo Maestro de Instituciones Educativas (AMIE) correspondiente al período 2009-2019. En la investigación se verifica que ciertas características de índole individual como, por ejemplo, ser mujer y vivir en un área rural incrementan la probabilidad de deserción. Por otro lado, la probabilidad de deserción se incrementa si los estudiantes se desenvuelven en un a jurisdicción bilingüe. De igual forma concluye que si existe un aumento de estudiantes en centros de educación en los cantones esto incrementa la posibilidad de que los estudiantes opten por la decisión de abandono de los estudios.

El escritor Campoverde manifiesta en su proyecto titulado “Deserción Estudiantil en el Instituto Tecnológico Superior de Fútbol en Guayaquil, 2019” [21]. El autor se enfoca en estudiar el alto número de deserción estudiantil que actualmente existe en los alumnos del Instituto Tecnológico Superior de Fútbol de Guayaquil. El Universo objeto de estudio estuvo compuesto por 300 estudiantes y se toma como muestra a 60 de ellos, existiendo un muestreo no probabilístico intencional. Se utiliza un cuestionario conformado por 20 preguntas, con un nivel de confiabilidad aproximado de 0.835 y cuya validez sobre el contenido se analizó a través del juicio de validación de tres expertos del tema. Se concluye que el 61.70% de estudiantes se encuentran en el nivel regular de la escolaridad académica que corresponde a la variable ‘deserción estudiantil’.

Para las autoras Segura y Loza en su trabajo “Uso de árboles de decisión para predecir el rendimiento académico en función de factores socioeconómicos” [22]. Se propusieron determinar si los factores socioeconómicos influyen en la educación y logros alcanzados por parte de los estudiantes de una Escuela Politécnica. Para cumplir con esto, consideraron los datos de los registros institucionales desde diez años atrás los relacionados a socioeconómicos y académicos. Posterior utilizaron técnicas de aprendizaje automático y algoritmos de clasificación para determinar qué factores tienen más influencia en el rendimiento académico de los estudiantes. Se efectuó un estudio comparativo de los varios métodos de árboles de decisiones. Para concluir las autoras manifestaron que las variables ‘título de escuela secundaria’, ‘edad’, ‘beca académica’ y ‘provincia’ influyen de forma directa en el porcentaje de rendimiento académico de los alumnos. El resultado del trabajo aporta información importante para los que dirigen el aspecto académico y para los trabajadores sociales involucrados en el único objetivo de mejorar las condiciones de los aspirantes a un título académico.

Así mismo J. Sopalo, Guevara y Burbano en su investigación “Análisis de los factores que inciden en la deserción estudiantil de los niños, niñas y adolescentes ecuatorianos en el periodo 2009-2019” [20]. Efectuada en la Escuela Politécnica Nacional. Aquí los autores indican que la deserción de los estudiantes crea un problema que infiere negativamente al progreso social, económico, político, y cultural de una sociedad. El análisis cuyo objetivo es estudiar cómo afectan las características tanto individuales y contextuales de las niñas, niños y adolescentes ecuatorianos en la deserción estudiantil. Para alcanzar este propósito, se utiliza un modelo de respuesta binaria Probit y como fuente de datos se consultó la ENEMDU (Encuesta Nacional de Empleo, Desempleo y Subempleo) y el Archivo Maestro de Instituciones Educativas en el periodo del 2009 - 2019.

Al concluir, se verifica que ciertas particularidades tanto individuales como ‘ser mujer’ y ‘vivir en el área rural’ incrementan en porcentaje en la probabilidad de retiro de sus estudios. Así mismo, se

encuentra que un mayor ingreso económico familiar garantiza la permanencia del estudiante en el sistema educativo. Por otra parte, las características contextuales delataron que la probabilidad de deserción incrementa si los alumnos se encuentran en una ‘jurisdicción bilingüe’. Por último, se evidencia que un aumento exagerado de estudiantes por unidad educativa amplía la posibilidad de abandono estudiantil.

Según Frutos y Buenaño en su investigación “Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato” [23]. Señalan que la deserción estudiantil es problema exclusivamente social que no es extraño a ninguna institución de educación superior, incluso incluido los institutos superiores a nivel nacional. Los autores en su investigación analizan el problema valiéndose de la técnicas y algoritmos de Minería de Datos que sirve para reconocer patrones de interés. Luego aplicaron la metodología KDD que toma un proceso recomendado de selección, preprocesamiento, transformación, así como minería de datos e interpretación de resultados. Su conclusión es que el análisis sobre la tasa de retención se debe tomar muy en cuenta ya que es muy importante la detección temprana de patrones de deserción estudiantil.

En la investigación de MESA y BARRERA realizan una comparación de las técnicas de clasificación, las mismas que sirven para determinar la deserción estudiantil en la Universidad de Antioquía de Colombia en la facultad de ingeniería [24]. En el artículo se busca determinar perfiles de estudiantes que han desertado de la facultad. Se utiliza un dataset de 2761 registros y con un total de 58 variables las mismas que facilitan información de alumnos desertores del período 2010 a 2018. La información clasifica en cuatro categorías a saber individuales, institucionales, socio económicas, y académicas, la investigación, en factores como el nivel de pregrado que concierne al semestre de carrera que tiene que ver con la probabilidad de desertar y el número de créditos académicos. Las máquinas de soporte vectorial desarrollado por Vladimir Vapnik se aplica para este tipo de trabajo este método ya que se relacionan de forma directa con problemas de clasificación el mismo que etiqueta clases de una muestra.

Los investigadores ACOSTA y RAMIREZ. Realizan un “estudio comparativo de técnicas de analítica del aprendizaje para predecir el rendimiento académico de los estudiantes de educación superior”[25]. Aquí mencionan que la deserción educativa concierne a diversos factores, entre ellos, se puede mencionar el compromiso del alumno, a través del cual se logre predecir el éxito o fracaso en los estudios. Este compromiso tiene varios frentes tales como, conductual, emocional y también el cognitivo.

El objetivo del estudio es analizar la eficiencia de la regresión lineal comparado con dos técnicas de minería de datos que servirán como parámetro para comparar y predecir el rendimiento en el sector

académico de los alumnos en la educación superior. Los autores realizaron una encuesta a 222 estudiantes de la institución pública ubicada en la ciudad de México en donde aplicaron la regresión lineal jerárquico. Al mismo tiempo aplican la técnica de análisis de aprendizaje por medio de redes neuronales y máquinas de vectores de soporte haciendo el análisis comparativo de regresión lineal y de máquinas de vectores de soporte. Como conclusión el pronóstico sobre el desempeño del estudiante menciona que el algoritmo de máquinas de vectores de soporte para este tipo de pronósticos es el más adecuado, ya que no requiere la comprobación de ningún supuesto estadístico.

En los trabajos de investigación detallados anteriormente, los autores predicen la deserción escolar en diferentes niveles educativos, utilizan variables predictoras y de la misma manera aplicaron diferentes modelos, la problemática es algo parecido a la investigación que he planteado, lo que motivó para tomar la decisión de utilizar el algoritmo de Redes Neuronales Artificiales y Máquinas de Vectores de Soporte..

2.2 MÉTODOS DE INVESTIGACIÓN

2.2.1 Estudio de caso

La presente investigación es de estudio de caso, es una herramienta poderosa para la investigación mediante la cual se analiza la conducta de las personas involucradas ya que se analiza a un estudiante o una realidad de carácter individual por medio de la investigación de campo. También es una investigación descriptiva, debido a que describen las variables socioeconómicas de las estudiantes llenadas en un formulario al inicio de cada periodo de matrícula que tiene el nombre de ficha socio económica la misma que determinará la influencias de la probabilidad de desertar [27].

2.2.2 Técnica de encuesta

La encuesta es una técnica en la mayoría de los casos es utilizada cuando se aplica un método de investigación, ya que con ello permite acceder a datos de manera rápida y eficaz, por medio de la recolección de respuestas a todas las preguntas que se planteen. También se define a la encuesta como aquella técnica que se vale de varios procedimientos investigativos mediante los que se recolecta y analiza una serie de datos obtenidos de una muestra de casos que son muy representativos de una determinada población de la que se pretende explorar, describir, predecir explicar una serie de caracteres [28]. La fuente de información primaria recibe de forma directa la información es decir es original sin que sea manipulada por segundos o terceros filtros antes de que sea analizada o interpretada por cualquier otra persona [29].

2.3 Marco Teórico

Procesamiento de información.

El procesamiento e información según define una serie de actividades que sirven para ordenar, almacenar y por sobre todo preparar los archivos con la información recibida. Esto con el propósito de explotar dicha información y presentar los resultados estadísticos mediante regresión lineal múltiple que es una técnica que permite identificar las relaciones existentes entre la variable dependiente y una o varias variables independientes. El propósito es llegar a fórmulas o modelos matemáticos con la finalidad de predecir algún resultado [26]. El cuestionario se compone de diez atributos los mismos que forman parte de variables independientes que mediante la aplicación de los modelos se determinará que se acepta o rechaza la hipótesis nula.

Maching Learning

Maching Learning es una definición que reúne varias estrategias analíticas cuyo fin es desarrollar algoritmos para datos ya sea para clasificación o para regresión. Machine Learning es sinónimo de Inteligencia Artificial y sirve como un recurso para intentar, usar medios computacionales, tomar decisiones. Este algoritmo tiene la capacidad de identificar y al mismo tiempo aprender patrones en datos para efectuar predicciones.

Se puede clasificar en tres grandes grupos según sea el paradigma de aprendizaje que se aplique y estos son: supervisados y no supervisados, reforzado, la diferencia principal de los algoritmos mencionados es la presencia o ausencia de la variable que da como resultado, llamada variable dependiente a la cual se le conoce como etiqueta, una de la técnica que ha dado mayor resultado en los últimos tiempos es Redes Neuronales Artificiales [30].

Inteligencia Artificial

La tecnología de Inteligencia Artificial es capaz de aprender en forma jerarquizada eso quiere decir, que la información se aprende por niveles en dónde. la primera capa aprende conceptos muy concretos en las otras capas, se utiliza lo aprendido previamente para aprender conceptos más abstractos, mientras más capas se añada seguirá aprendiendo conceptos cada vez más abstractos y así se puede añadir capas sin límite [31].

Redes Neuronales

Son capaces de aprender en forma ordenada eso quiere decir, que la información se aprende por niveles en la cual, la primera capa aprende conceptos muy concretos. En las otras capas se utiliza lo aprendido previamente para aprender conceptos más abstractos. En la redes neuronales mientras más capas se añada seguirá aprendiendo conceptos cada vez más abstractos y así se puede añadir capas sin límite [32].

Las Redes Neuronales es una familia de modelos bastante grande y de esta familia en la presente investigación utilizará la topología Feed-forward. Las mismas se distinguen por ejecutar el procesamiento de datos en una y solo una dirección de izquierda hacia la derecha, son muy fáciles de programar por ende son las más usadas en el contexto científico. Las Redes Neuronales Artificiales feed-forward permite obtener una función compuesta que es de uso universal porque se aplica a cualquier problema sea de regresión o clasificación [33]. Redes neuronales artificiales tienen la capacidad de ejercitar acciones automáticas por medio del reconocimiento de la voz, por lo que se considera como exponente de la inteligencia artificial IA bajo el dominio del aprendizaje automático o machine learning.

Lo que hace posible que las máquinas logren ejecutar automáticamente tareas que antes solo se podía realizar por personas y no se podía realizar automatizar con los sistemas convencionales. Es por esto por lo que la Inteligencia Artificial posibilita una interrupción de algo tecnológico sin precedentes que día a día está en constante cambio de una manera vertiginosa. Las redes neuronales artificiales representan un modelo computacional que facilita simular el comportamiento de nuestro cerebro humano, es decir otorgar a las máquinas capacidades de aprender de una forma similar a como aprende nuestro cerebro. Una red neuronal está compuesta por neuronas artificiales, las mismas que son unidades o nodos que reciben información de la parte externa [32].

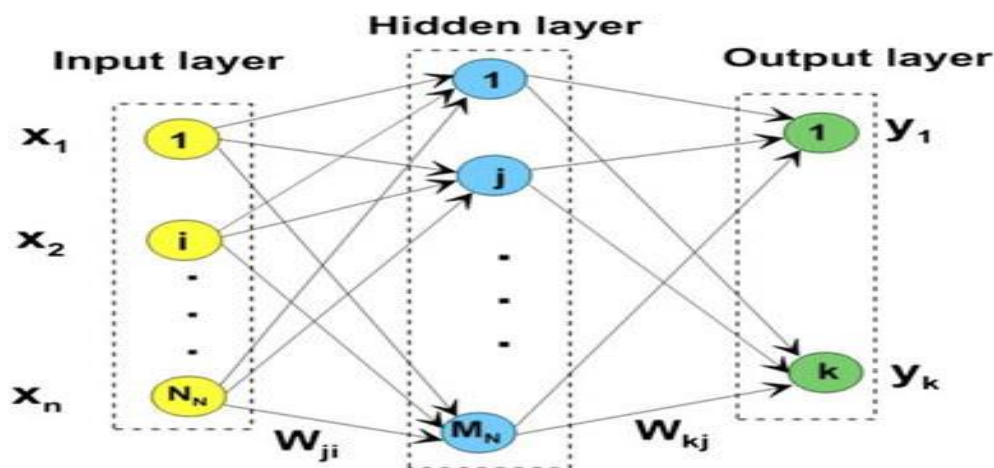


Figura 1. Estructura de una Red Neuronal
Fuente. Revista Espacios [34].

Máquinas de Vectores de Soporte

Al inicio por la década de los 90 se desarrolla en la ciencia computacional y se implementa en un método de clasificación binaria, luego la información se expande a problemas de clasificación múltiple y regresión. Dentro de Maching learning se considera el mejor clasificador dentro del aprendizaje estadístico. Las Máquinas de Vectores de Soporte es el algoritmo que trata de buscar un hiperplano que separe lo mejor posible las clases, que es un espacio p-dimensional definido como aquel

subespacio plano y afín con dimensiones $p-1$, se dice afín porque el subespacio no tiene que cruzar por el origen. En un espacio de dos dimensiones, la idea principal es de mapear todos los datos a una dimensión superior y encontrar el límite para hacer la clasificación con la ayuda de kernel que es el encargado de transformar el espacio de datos de entrada en la forma que más convenga [35].

La función de Máquinas de Vectores de Soporte es elaborar un modelo que señale los puntos que se muestra en el espacio e intentar separar esos dos espacios de manera más amplia posible. Esos espacios se conoce como un hiperplano de separación, señalado como el vector entre los dos puntos de las dos clases que estén más cerca el uno del otro, conocido como el vector de soporte que determina a cuál de los dos lados del plano pertenece un dato y se clasifica en una rama u otra de la características es decir clasificar los datos en dos grupos y buscar la recta que mejor divida los grupos de datos [33].

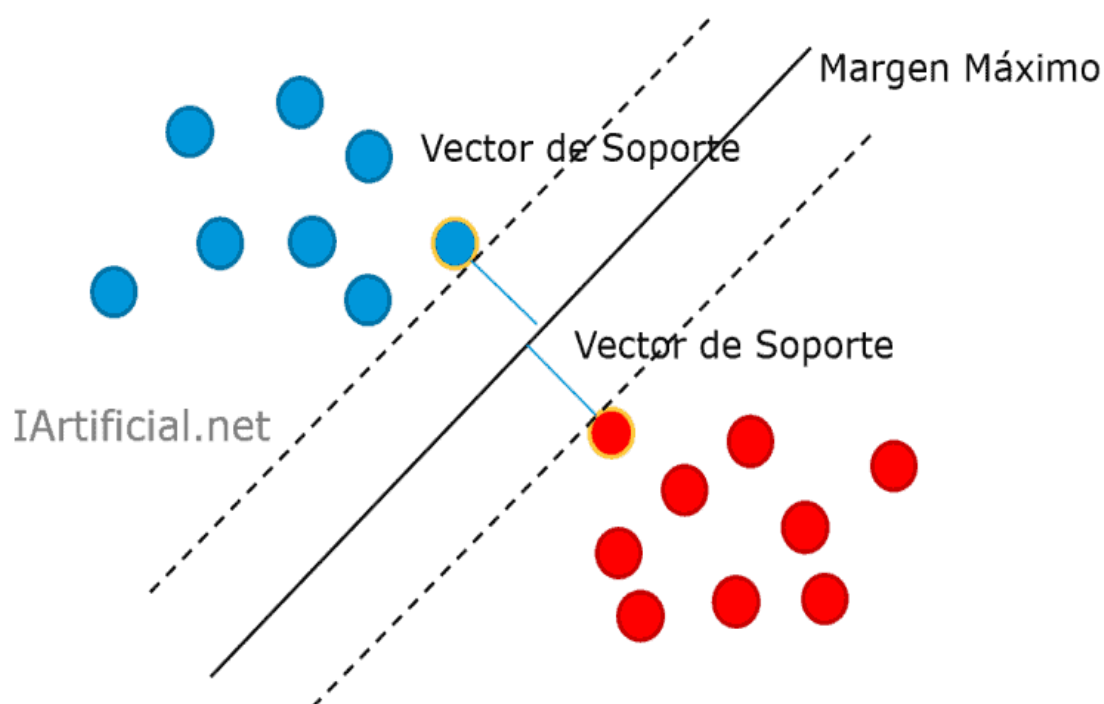


Figura 2. Vectores de Soporte

Fuente. E.E. García Díaz y F. Lozano Martínez [36].

Matriz de confusión

Es una definición importante en la inteligencia artificial y por ende en el aprendizaje automático, es una poderosa herramienta que analiza los resultados de la forma de cómo trabaja un algoritmo de aprendizaje supervisado, la matriz por lo general se muestra en forma de tabla en donde cada columna señala el número de predicciones de la clase, mientras que en las filas se señala la cantidad de instancias de cada clase. La matriz de confusión relaciona las predicciones realizadas por un algoritmo de aprendizaje y los resultados de forma correcta que se debería haber logrado, es decir determinar el número de aciertos y errores en el momento de pasar por un proceso de aprendizaje. En conclusión permite medir lo bueno que es un modelo de clasificación sobre dicho sistema de aprendizaje [37].

Dataset

Es el total de conjunto datos que contiene la información sobre lo que se quiere desarrollar algún algoritmo de Maching Learning con la finalidad de conseguir un modelo que mejor lo represente. Este conjunto de datos se conforma por variables dependientes (es la variable predicha en el modelo por la variable independiente) e independientes (son aquellas variables que en el modelo van a predecir lo mejor posible a la variable dependiente) [38].

Balanceo de datos

Es utilizado en Maching learning para tratar de equilibrar todas las variables en un modelo, es el caso en que el número de observaciones no varíe para todas las clases de un dataset usado en la clasificación. Existe varios tipos de balanceo de datos basados en el remuestreo, es una técnica para tratar un conjunto de datos altamente desequilibrado. Se tiene un sobremuestreo cuando se trata de eliminar muestras de la clase mayoritaria, en cambio submuestreo consiste en agregar muestras a la clase minoritaria para equilibrar en número a la clase mayoritaria [39]. Lo recomendable es desbalancear los datos para ello se requiere de un clasificador que genere una alta precisión para la clase minoritaria, pero sin descuidar el peligro en la precisión de la clase dominante [40].

Validación del modelo

Consiste en realizar pruebas en modelos de aprendizaje para verificar la capacidad de predecir nuevos casos que no se hayan utilizado en el momento de ser entrenado, con esto se verifica si el modelo se generaliza. El procedimiento consiste en dividir el dataset de entrenamiento en dos partes, la una parte para entrenamiento lo que conlleva a entrenar el modelo de clasificación y la otra parte minoritaria en porcentaje, sirve para validar el modelo que no ha sido utilizado en la fase de entrenamiento [39].

Para medir la efectividad del modelo se recomienda hacer uso de la validación cruzada, normalmente se usa para comparar y seleccionar un modelo para aplicar en un problema de modelado predictivo por su facilidad de entender, es fácil de implementar y arroja como resultado valoraciones de habilidades. En nuestro caso nos centraremos en el tipo específico llamado F-folds,

Estas valoraciones normalmente tienen sesgo más bajo al compararlos con otros métodos, aquí se utiliza todos los datos para entrenar y al mismo tiempo todos los datos para evaluar. El procedimiento contiene un único parámetro con nombre K que se refiere al número de división de grupos que se puede dividir una muestra de datos previamente determinado, cuando se selecciona por ejemplo un valor específico para $k = 5$ convirtiéndose en 5 veces o grupos la validación cruzada [41].

El aprendizaje costo sensitivo que es el costo que afecta a la toma de decisión, proporciona una alternativa posible a los métodos de muestreo para mandos de clasificación de datos desbalanceados. Se basa en realizar modificaciones en los umbrales de decisión o asignando pesos a nuevas situaciones remuestradas según la matriz de decisión [36].

Métricas de evaluación del modelo

Varios clasificadores pretenden aprender otras características de la clase minoritaria creando un alto costo si se comete errores en la clasificación de las muestras, siempre las medidas de rendimiento con más frecuencia que son usadas son la efectividad o exactitud (accuracy), la precisión, la sensibilidad, especificidad, tasa de error, puntuación F1-score, motivando un problema primordial de clasificación binaria teniendo clases positivas y negativas logrando una representación del rendimiento de clasificación el mismo que puede ser ilustrado en una matriz de confusión tomando en cuenta a la clase minoritaria como la clase positiva.

Exactitud

La accuracy del modelo relaciona al número total de predicciones correctas dividido por el número total de predicciones, que no es más que determinar qué tan confiable es un modelo en el momento de responder si un punto pertenece a esa clase [42].

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

Precisión

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (2)$$

Tiene que ver sobre la clase de interés que tan bien funciona el modelo, es la cercanía de las mediciones entre si [42].

Sensibilidad

$$\text{(Recall)Sensibilidad} = \frac{VP}{\text{Valores positivos} + \text{Falsos Negativos}} \quad (3)$$

cuando la clase es positiva, que porcentaje logra clasificar bien el modelo [42].

$$\text{Especificidad} = \frac{VN}{\text{Valores negativos} + \text{Falsos positivos}} \quad (4)$$

cuando la clase es negativa, que porcentaje logra clasificar el modelo [42].

$$\text{Tasa de error} = \frac{F P + F N}{\text{TOTAL}} \quad (5)$$

qué porcentaje de la data clasifica incorrectamente [42].

$$\text{La clase de error} = 1 - \text{Exactitud} \quad (6)$$

Puntuación F1 Score Es conocida como una medida de precisión y al mismo tiempo de robustez del modelo, se utiliza cuando los datos son desbalanceados y combina las métricas de precisión y la sensibilidad

$$\text{F1 Score} = 2 \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (7)$$

Población y Muestra

El Universo de estudio se consideró a todos los estudiantes legalmente matriculados en el período académico Mayo - Octubre 2022 del Instituto Superior Tecnológico Bolívar el mismo que está ubicado en la ciudad de Ambato que oferta las carreras de: Administración de empresas, Contabilidad, Desarrollo de Software, Marketing, Redes y Telecomunicaciones, se considerará la siguiente fórmula para obtener una muestra significativa.

$$n = \frac{Z^2 P Q N}{Z^2 P Q + (N-1)e^2} \quad (8)$$

Z= Nivel de confiabilidad 95%

P= Probabilidad de ocurrencia 50%

Q= Probabilidad de no ocurrencia 50%

e= Error de muestreo

N= Población

CAPITULO III

MARCO METODOLÓGICO

3.1. Ubicación

Los datos se obtienen del Instituto Superior Tecnológico Bolívar que está ubicado en la zona central del centro del Ecuador, específicamente en la ciudad de Ambato en las calles sucre entre las calles Lalama y Martínez como se muestra en la figura 3.

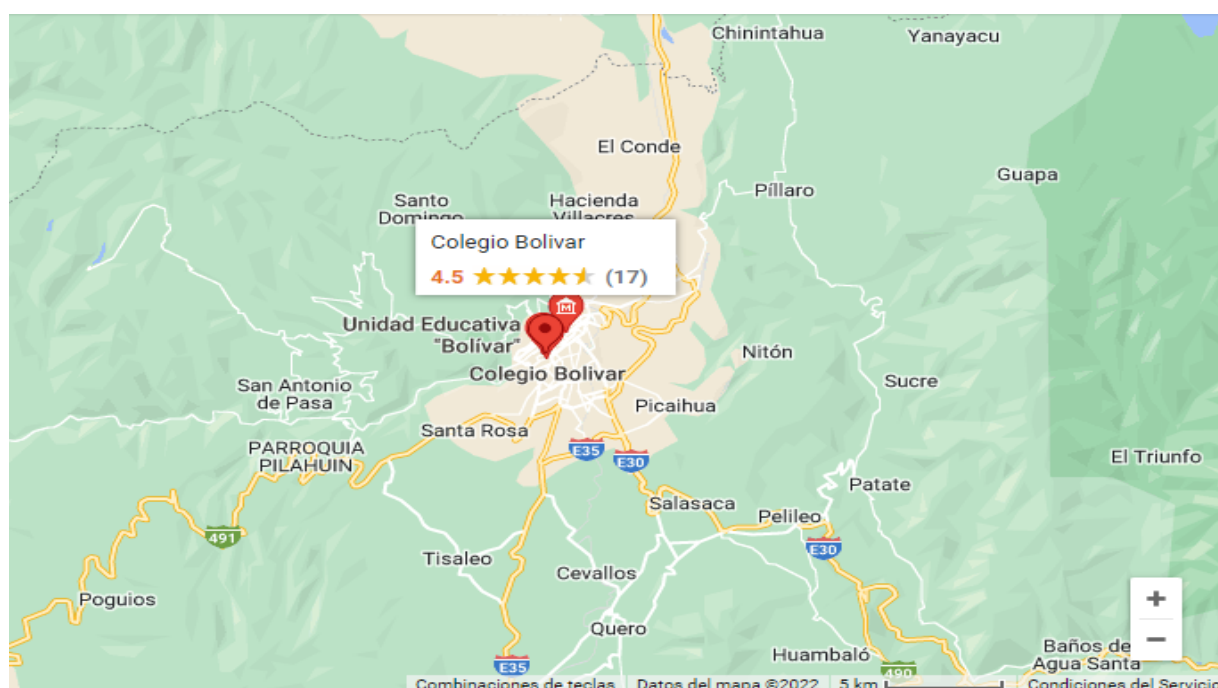


Figura 3. Ubicación geográfica del ISTB

Fuente: Google Maps.

3.2. Equipos y materiales

Para la presente investigación se necesitó de los siguientes materiales:

Hardware

Nombre del dispositivo: DESKTOP-G5UHPGV

Procesador: AMD A6-9225 RADEON R4, 5 COMPUTE CORES 2C+3G 2.60 GHz

RAM: 4,00 GB (3,88 GB utilizable)

Sistema: Sistema operativo de 64 bits, procesador x64

Edición: Windows 10 Home

Software

- Base de datos de estudiantes
- Sistema RLM

- Programa Matlab
- Paquete Microsoft Office 2016

3.3. Tipo de investigación

En la presente investigación se empleó tres tipos de investigación, la bibliográfica porque se utilizó varios libros como fuente de información. La Investigación de campo porque se aplica las encuestas a los estudiantes de forma presencial en el Instituto. La investigación cuantitativa porque se recolecta y procesó datos mediante procedimientos matemáticos apoyándose en la estadística [43].

3.4. Prueba de hipótesis

En la presente investigación, se consideraron las siguientes hipótesis:

3.4.1 Hipótesis nula (H_0):

Las Redes Neuronales Artificiales no permite construir un modelo para predecir la deserción estudiantil en los estudiantes del Instituto Superior Tecnológico Bolívar.

Las Máquinas de Vectores de Soporte no permite implementar un modelo para predecir la deserción estudiantil en los estudiantes del Instituto Superior Tecnológico Bolívar.

3.4.2 Hipótesis alternativa (H_a):

Las Redes Neuronales Artificiales permite construir un modelo para predecir la deserción estudiantil en los estudiantes del Instituto Superior Tecnológico Bolívar.

Las Máquinas de Vectores de Soporte permite implementar un modelo para predecir la deserción estudiantil en los estudiantes del Instituto Superior Tecnológico Bolívar

3.5 Recolección de información

Los datos recolectados para el presente trabajo de investigación pertenecen a los estudiantes del Instituto Superior Bolívar. La información se obtiene en base a entrevistas personales y encuestas realizadas a estudiantes de la carrera de contabilidad, Marketing, Desarrollo de Software, Administración financiera, Redes y Telecomunicaciones utilizando la herramienta google forms, también se recopiló a información de la ficha socio -económico, es una matriz que reposa en archivos de la institución. La información se guarda en una matriz en Microsoft Excel. El universo de estudio se considera de 1036 estudiantes se envía una encuesta por medio de google forms y respondieron la encuesta 280 estudiantes, cantidad requerida aplicando la fórmula (7) de muestra poblacional [27].

Para el análisis de las variables se utiliza el sistema de programación MatLab en el cual se analizará con los algoritmos Redes Neuronales Artificiales feed-forward (RNA) y Máquinas de Vectores de

Soporte (SVM) que se encuentran disponibles en google y no tiene costo, se procede a importar los datos desde el Excel. Como los datos son con varias opciones cualitativas es decir son asimétricas por lo que se convierte esas respuestas a números, condición necesaria que se requiere para trabajar con los algoritmos. (ver anexos del 12 al 16).

CAPITULO IV

RESULTADOS Y DISCUSIÓN

Para el avance de la presente investigación se presenta por oficio al señor rector del Instituto Superior Tecnológico Bolívar, con la petición para que brinde la facilidad para obtener la información necesaria sobre los datos de los estudiantes. Esa información se obtendrá mediante una encuesta, como también de los archivos de la institución, la solicitud es respondida satisfactoriamente, luego de consultar en la secretaría el número de estudiantes matriculados en el período 2022 (1036) y por cada carrera (Contabilidad 262), Marketing (193), Desarrollo de Software (148), Redes y comunicaciones (158) y Administración financiera (275), para determinar una muestra de una población se aplica la fórmula generalizada (ver anexo 11).

Con un margen de error del 5% y nivel de confianza del 95% que en la tabla de distribución normal se obtiene un valor de 1,96 dando como resultado 280 encuestas que se debe aplicar como muestra. Las variables son elaboradas con base a una revisión de la bibliografía de estudios afines, sobre las variables socioeconómicas se escogió algunas preguntas que pueden influir en los estudiantes en su posible deserción. Las preguntas son de índole general que se relacionan con preguntas de datos personales, socio afectivos académicos y también se incluyó preguntas socio económico del núcleo familia. Algunos de los datos requeridos se encuentran en las fichas socio económicas que los estudiantes llenan al momento de realizar la matrícula para cada periodo lectivo, se elabora la encuesta con las variables que pueden servir para lo que se quiere predecir (ver Anexo 12).

La encuesta se envía a los estudiantes mediante el uso de la herramienta web por medio del aplicativo google forms (ver Anexo 13). El resultado numérico de la encuesta solicitada se encuentra en el Anexo (14,15). Y el resultado gráfico de la encuesta y el análisis de cada pregunta se puede observar en los anexos del (1 al 10). La información recopilada es de tipo primaria porque se obtuvo de la fuente, luego se procede a tabular las encuestas en formato Excel. Los resultados de la encuesta son asimétricos, lo que se los categorizan para que los datos sean cuantitativos, luego se los pondera debido a que el resultado obtenido no es perfecto el conjunto de datos está muy desbalanceado (ver Anexo 16 y 17).

Para evaluar el modelo se utiliza los algoritmos de Redes Neuronales Artificiales feed-forward y

Máquinas de Vectores de Soporte, se divide el dataset en dos sectores dejando el uno con el 60 % de los datos como entrenamiento y el otro sector de datos que es el 40 % como datos de prueba, posteriormente se entrena el modelo de nuevo pero con los datos de entrenamiento, en el siguiente paso, se aplica a los datos que se quedaron como reserva y se evalúa el rendimiento del modelo obteniendo como resultado la métricas de exactitud (Accuracy), la métrica de precisión (Precisión), especificidad, Sensibilidad y F1-Score en Redes Neuronales Artificiales feed-forward y Máquinas de Vectores de Soporte, con datos balanceados y desbalanceados. Para lograr lo manifestado, se realiza 41 ensayos con cada algoritmo (ver anexo 18-19), de los cuales se selecciona en Máquinas de Vectores de Soporte, los mejores resultados. Los experimentos realizados en Máquinas de Vectores de Soporte, fue cambiar el Kernels y el número de repeticiones hasta obtener los mejores resultados como se muestra en la tabla 1. Por otra parte, en Redes Neuronales Artificiales Feed-forward se experimenta cambiando el número de las capas ocultas, la cantidad de neuronas en cada capa, la función de activación en cuanto al número y a la clase, el número de iteraciones, el número de repeticiones hasta obtener los mejores resultados como se muestra en la tabla 2.

Tabla 1. Resultado con datos de testeo en SVM

SVM				Resultados					
fractionForTrainingAndValidation	numFolds	kernel	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)
0,6	20	polynomial	10	0,85	0,93	0,74	0,84	0,93	0,74
0,6	50	polynomial	10	0,86	0,93	0,79	0,54	0,57	0,89

Autor: Patricio Mejía

SVM

Después de realizar varias corridas se obtiene los mejores resultados utilizando el 60% de los datos para entrenamiento y validación con número de folds 66, con la función de activación linear, con 66 repeticiones se obtuvo como resultado con datos balanceados la especificidad 60%, sensibilidad 93% y el F1-score 80%.

Después de realizar varias corridas se obtiene los mejores resultados utilizando el 60% de los datos para entrenamiento y validación con número de folds 59, con la función de activación gauss, con 20 repeticiones se obtuvo como resultado con datos balanceados la especificidad 74 %, sensibilidad 87% y el F1-score 80%.

Después de realizar varias corridas se obtiene los mejores resultados utilizando el 60% de los datos para entrenamiento y validación con número de folds 20 con la función de activación polynomial, con 10 repeticiones se obtuvo como resultado con datos balanceados la especificidad 74 %, sensibilidad

93% y el F1 Score 85 % con esto se puede determinar que es la mejor opción (ver anexo 20). De lo anterior se detecta que con un kernel polynomial se obtiene mejores resultados.

RNA

Después de realizar varias experiencias se obtiene los mejores resultados utilizando el 60% de los datos para entrenamiento y validación con número de 13 folds con 15 iteraciones y con 5 repeticiones se obtuvo como resultado con datos balanceados, la especificidad 76%, sensibilidad 92%, y el F1-score 81% (ver anexo 19).

Con Redes Neuronales Artificiales feed-forward se realiza con los 15 mejores resultados nuevamente se efectúa un cambio en el número de neuronas y el número de capas ya que el F1-score con datos desbalanceados es mayor que el resultado con datos balanceados y selecciona los 3 mejores resultados, como se muestra en la tabla 2.

Tabla 2. Resultado con datos de testeo en RNA feed forward

RNA feed-forward							Resultados					
fractionForTrainingAndValidation	capas ocultas	funcion de activacion	lambda	numFolds	iteraciones	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)
0,6	80,75,22,1	tanh,tanh,tanh,logsig	0	20	30	10	0,84	0,92	0,76	0,75	0,8	0,84
0,6	20,60,50,1	relu,softplus,tanh,logsig	0	20	50	10	0,83	0,91	0,73	0,7	0,75	0,81
0,6	100,75,20,1	tanh,tanh,tanh,logsig	0	20	40	5	0,83	0,9	0,81	0,81	0,87	0,84

Autor: Patricio Mejía

Después de realizar varias corridas en el algoritmo Redes Neuronales Artificiales feed-forward se obtiene los mejores resultados utilizando el 60% de los datos para entrenamiento y validación con número de 20 folds con 30 iteraciones y con 10 repeticiones utilizando una capa de entrada, 3 capas ocultas y 3 funciones de activación tanh, tanh, tanh, y la capa de salida logsig se obtuvo como resultado con datos balanceados la especificidad 76%, sensibilidad 92%, el F1 Score 84%. El siguiente paso consiste en entrenar el modelo con el total de los datos, es decir con todo el 60%. Obteniendo los siguientes resultados con datos desbalanceados y balanceados ver tabla 3.

Tabla 3. Resultados finales en SVM

fractionForTrainingAndValidation	numFolds	kernel	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)
				BALANCEADOS			DESBALANCEADOS		
0,6	20	polynomial	10	0,83	0,929	0,758	0,834	0,929	0,758
RESULTADO FINAL-MODELO ENTRENADO									
				BALANCEADOS			DESBALANCEADOS		
0,6	20	polynomial	10	0,756	0,913	0,651	0,756	0,913	0,65

Autor: Patricio Mejía

Con el resultado obtenido en Máquinas de Vectores de Soporte se observa que el modelo antes de entrenar de obtiene un F1-score balanceado es de 83,4 % mientras que después que el modelo fue entrenado se logra un F1-score de 75,6 % balanceado aquí se visualiza que de acuerdo al F1-score disminuye cuando el modelo es entrenado.

Tabla 4. Resultados finales en RNA feed forward entrenado

fractionForTrainingAndValidation	capas ocultas	funcion de activacion	lambda	numFolds	iteraciones	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)
							BALANCEADOS			DESBALANCEADOS		
0,6	80,75,22,1	tanh,tanh,tanh,logsig	0	20	30	10	0,725	0,823	0,7	0,654	0,585	0,94
RESULTADO FINAL-MODELO ENTRENADO												
							BALANCEADOS			DESBALANCEADOS		
0,6	80,75,22,1	tanh,tanh,tanh,logsig	0	20	30	10	0,822	0,956	0,744	0,72	0,587	0,97

Autor: Patricio Mejía

Los resultados logrados en Redes Neuronales Artificiales feed-forward se pueden observar que el modelo antes de ser entrenado tiene un F1-score balanceado de 72,5 % luego de ser entrenado el modelo se logra un F1 score balanceado del 82,2 % esto es bueno porque se logró mejores resultados.

Matriz de Confusión

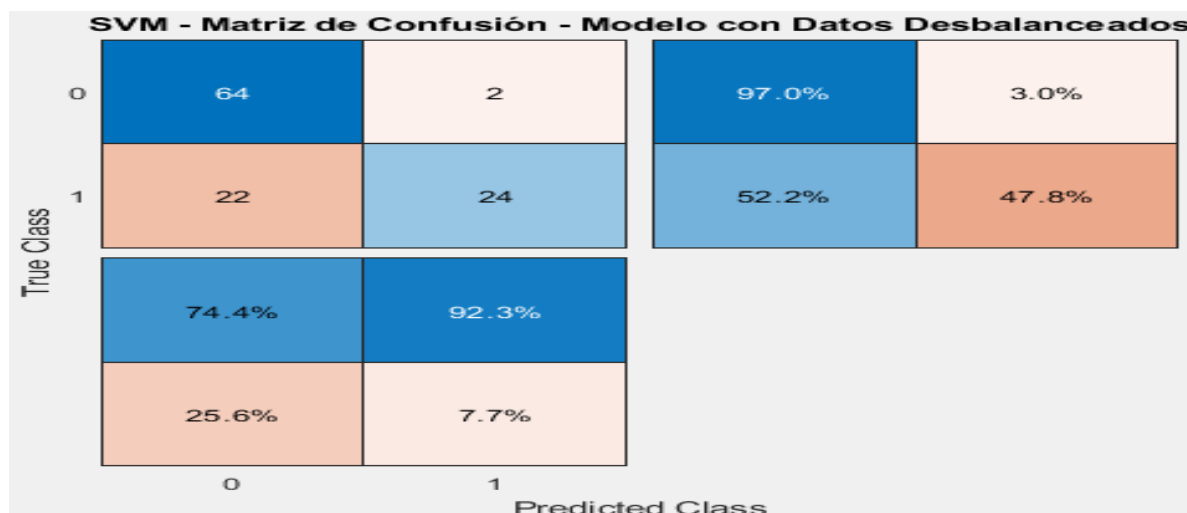


Figura 4. Resultado de Matriz de Confusión datos desbalanceados

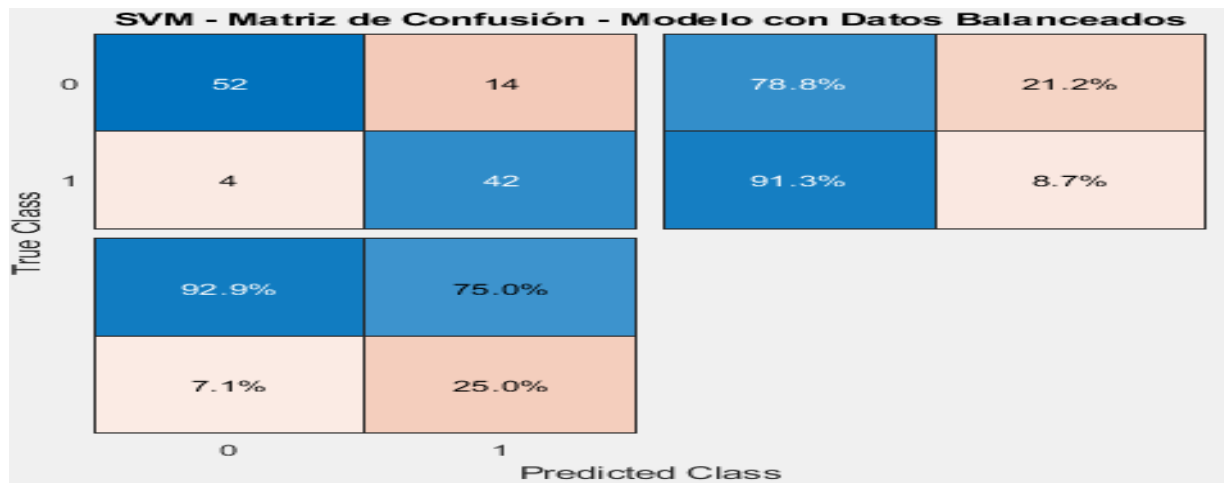


Figura 5. Resultado de Matriz de Confusión datos balanceados

Resultado Final de Matriz de Confusión

SVM

Los datos que se muestra en la diagonal principal $a_{11}= 52$ y $d_{22}= 42$ se corresponden con los valores que se han estimado de forma correcta por el modelo, tanto los valores verdaderos positivos VP (a), así como los valores negativos VN (d), la otra diagonal representa todos los casos en los que el modelo de ha equivocado ($b_{12}= 14$ falsos positivos FP, $c_{21}= 4$ falsos negativos FN)

$$\text{Especificidad} = \frac{VN}{VN+FP} = \frac{42}{42+14} = 0,75 \approx 75 \% \quad (9)$$

$$\text{Sensibilidad} = \frac{VP}{VP+FN} = \frac{52}{52+4} = 0,93 \approx 93\% \quad (10)$$

En nuestro ejemplo, la sensibilidad = 0,93 y a la especificidad = 0,75 por lo tanto este modelo es mucho más sensible que específico. Esto es lo que nos interesa cuando el objetivo es evitar falsos positivos [44].

$$\text{Precisión} = \frac{VP}{TOTAL POSITIVOS} = \frac{52}{52+14} = 0,78 \approx 78 \% \quad (11)$$

$$F1 \text{-Score} = 2 \frac{\text{precisión} \times \text{sensibilidad}}{\text{precisión} + \text{sensibilidad}} = 2 \frac{0,78 \times 0,93}{0,78 + 0,93} = 0,85 \approx 85 \% \quad (12)$$

RNA feed- forward

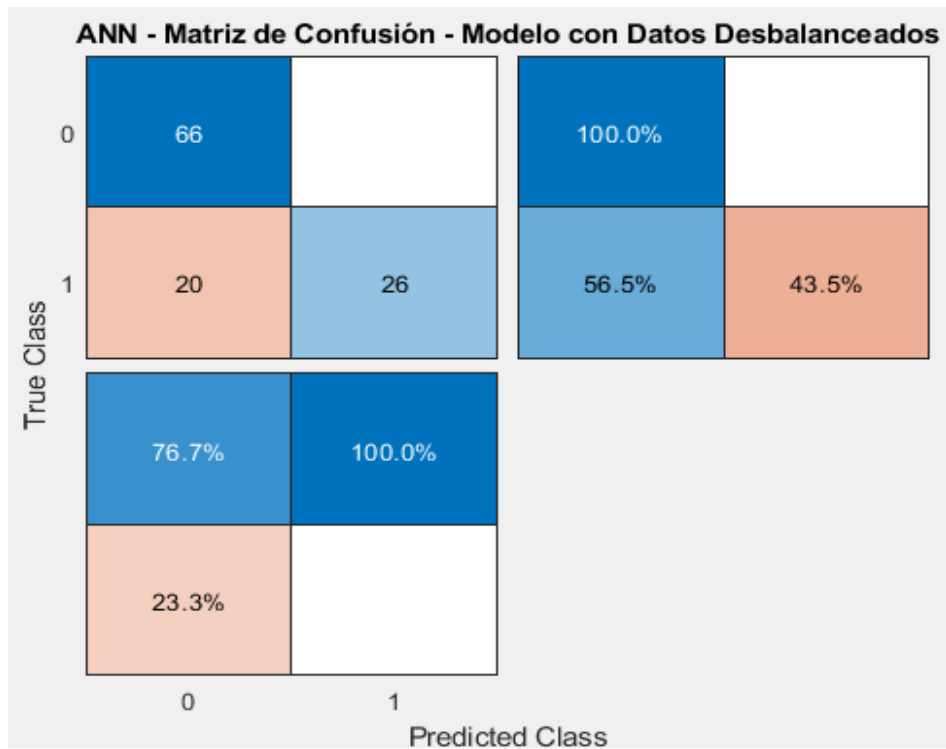


Figura 6. Resultado de Matriz de Confusión con RNA datos desbalanceados

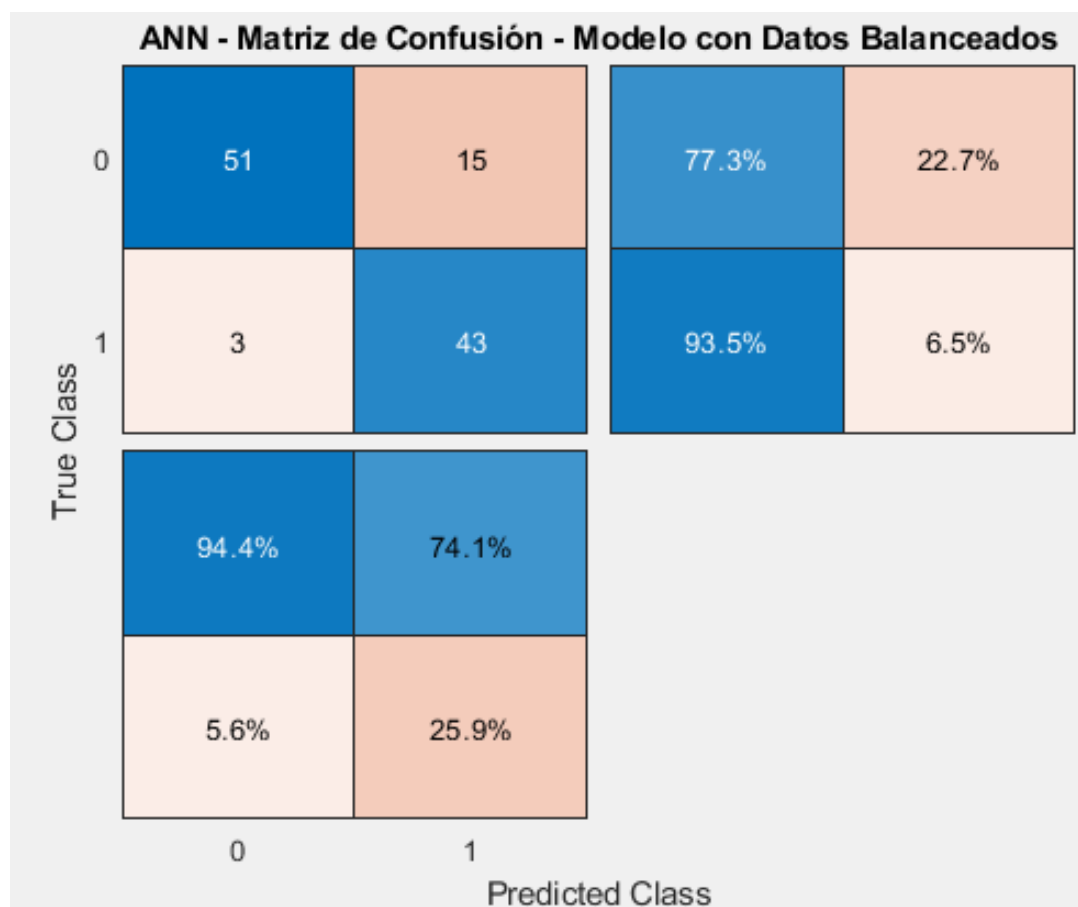


Figura 7. Resultado de Matriz de Confusión con RNA datos balanceados

RNA feed-forward

Los datos que se muestra en la diagonal principal $a= 51$ y $d= 43$ se corresponden con los valores que se han estimado de forma correcta por el modelo, tanto los valores verdaderos positivos VP (a), así como los valores negativos VN (d), la otra diagonal representa todos los casos en los que el modelo se equivoca ($b= 15$ falsos positivos FP, $c= 3$ falsos negativos FN)

$$\text{Especificidad} = \frac{VN}{VN+FP} = \frac{43}{43 + 15} = 0,74 \approx 74 \% \quad (13)$$

$$\text{Sensibilidad} = \frac{VP}{VP+FN} = \frac{51}{51 + 3} = 0,94 \approx 94 \% \quad (14)$$

En nuestro ejemplo, la sensibilidad = 0,94 y a la especificidad = 0,74 por lo tanto este modelo es mucho más sensible que específico lo que indica que el porcentaje de estudiantes desertores es correctamente identificado por el modelo.

$$\text{Precisión} = \frac{VP}{TOTAL POSITIVOS} = \frac{51}{51 + 15} = 0,77 \approx 77 \% \quad (15)$$

$$F1 \text{-Score} = 2 \frac{\text{precisión} \times \text{sensibilidad}}{\text{precisión} + \text{sensibilidad}} = 2 \frac{0,77 \times 0,94}{0,77 + 0,94} = 0,84 \approx 84 \% \quad (16)$$

Con Redes Neuronales Artificiales feed-forward los falsos negativos disminuyen, eso es bueno, porque el programa detecta menor cantidad de estudiantes que siendo desertores los identifica como no desertores.

Tabla 5. Resultados finales entrenados en SVM y RNA feed -forward

RESULTADO FINAL-MODELO ENTRENADO SVM									
fractionForTrainingAndValidation	numFolds	kernel	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)
				BALANCEADOS			DESBALANCEADOS		
0,6	20	polynomial	10	0,756	0,913	0,651	0,756	0,913	0,65

RESULTADO FINAL-MODELO ENTRENADO RNA												
fractionForTrainingAndValidation	capas ocultas	funcion de activacion	lambda	numFolds	iteraciones	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)
							BALANCEADOS			DESBALANCEADOS		
0,6	80,75,22,1	tanh,tanh,tanh	0	20	30	10	0,822	0,956	0,744	0,72	0,587	0,97

Autor: Patricio Mejía

Al final, comparando los resultados de los dos modelos, se observa que el modelo con el algoritmo de Redes Neuronales Artificiales feed-forward se logra mejores resultados con los datos balanceados.

La especificidad con un 74,4 %, esto representa la cantidad de la clase negativa y de ella que porcentaje logra clasificar como negativa, esto indica la capacidad de la prueba de detectar a los no desertores.

La sensibilidad del 95,6 %, es un valor alto indica el porcentaje de estudiantes desertores que fueron correctamente identificados por el modelo esto indica la capacidad de la prueba de detectar a los desertores.

F1-score nos señala que, de 100 casos posibles, 82 son predichos correctamente; el modelo indica y combina las métricas de precisión y sensibilidad.

Al analizar los datos resultantes de visualiza que el F1-score mejora con datos entrenados usando Redes Neuronales Artificiales feed-forward, por lo tanto, vale la pena realizar los cambios para obtener mejores resultados en la matriz de confusión. Con esto, a pesar de que se obtiene poca diferencia en los resultados en lo referente al F1- score y siendo los dos modelos de poca dificultad, se recomienda para este tipo de problema utilizar el modelo de Redes Neuronales Artificiales feed-forward. Los modelos utilizados no permiten determinar las variables que influyen más o menos, lo que permite predecir los modelos es que si los estudiantes con esa información van a desertar o no. Este sería el modelo matemático para predecir la deserción de los estudiantes del Instituto Superior Tecnológico Bolívar.

CAPITULO V

CONCLUSIONES, RECOMENDACIONES, BIBLIOGRAFÍA Y ANEXOS

5.1. Conclusiones

- Para predecir la deserción estudiantil de los estudiantes del Instituto Superior Bolívar de la ciudad de Ambato del periodo mayo octubre 2022 se analizó las variables con el modelo de Máquinas de vectores de Soporte y Redes Neuronales feed-forward se utilizó las siguientes variables predictoras al mismo tiempo se indica según los resultados el rango de aporte más significativo: Tipo de bachillerato (fiscal 82%), estado civil (solteros 52%), edad (18 a 20 años 29%) , gastos económicos (padres 53%), cargas familiares (no 54 %), semestre de estudio (primero 45%), lugar de residencia(en la provincia 78%), la pandemia ocasionó dificultad en el proceso de aprendizaje (si 53%), la pandemia ocasionó desequilibrio económico (81%), falta de apoyo del docente (si 72%), y la variable dependiente es si desertó o no.
- Después de experimentar varios cambios con los kernels: polynomial, gaussiano, linear, y el número de folds, el mejor modelo obtenido con Máquinas de Vectores de Soporte fue con un kernel polynomial de segundo orden obteniendo un F1-score de 75,6 % después de balancear los datos vs un F1-score de 65,4 % antes del balanceo de datos.

En la arquitectura de Redes Neuronales Artificiales feed-forward se inserta cinco capas, una capa de entrada tres ocultas y una capa de salida. Después de efectuar varios cambios se logra los mejores resultados con el número de variables predictoras en la primera capa oculta y en las tres capas ocultas 80, 75, 22 neuronas y una única neurona de salida. Cuando se entrenó las Redes Neuronales Artificiales Feed-forward lo que produjo más impacto en el resultado fue cuando se experimentó con cinco capas y cuando se cambió el número de neuronas en cada una de ellas. También influyó cuando se experimenta varios ensayos con diferentes funciones de activación como elu, relu, softplus, tanh y con las que se logró mejor resultado es tanh, tanh, tanh, y la función logsig como capa de salida, con esto se logró un F1- score de 82,2% después de aplicar el balanceo de datos vs el 72% antes del balanceo de datos, con esto se determina que el modelo de mejor ajuste es el de Redes Neuronales Artificiales feed-forward lo que permite determinar los posibles casos de deserción.

- Para la evaluación de los algoritmos implementados en este trabajo se utilizó validación cruzada a través de 20 folds y comparando los F1-score de los dos modelos implementados el mejor

modelo es Redes Neuronales Artificiales feed-forward con lo que se obtiene que el grado de deserción estudiantil en el Instituto Superior Bolívar en el período 2022 es menor al 18%.

5.2. Recomendaciones

- Luego del análisis de los resultados logrados, se recomienda que se elija otro tipo de variables predictoras y con la utilización de la misma arquitectura de los algoritmos de Máquinas de Vectores de Soporte y Redes Neuronales Artificiales utilizados en esta investigación para comparar si mejora el resultado del F1-score.
- Se recomienda para futuras investigaciones la utilización del algoritmo de máquinas de vectores de soporte con un Kernel polynomial con el que se logra un mejor f1-score.
- Después de aplicar el modelo predictivo del algoritmo de Redes neuronales Artificiales feed-forward se recomienda utilizar cinco capas, la función de activación tanh en las tres capas ocultas con eso se logró mejores resultados los mismos que ayudarán a los directivos de las instituciones superiores a una toma de decisiones preventiva en la deserción estudiantil.

5.3. BIBLIOGRAFÍA

- [1] M. Castillo-Sánchez, R. Gamboa-Araya, y R. Hidalgo-Mora, «Factores que influyen en la deserción y reprobación de estudiantes de un curso universitario de matemáticas», *Uniciencia*, vol. 34, n.º 1, pp. 219-245, 2020, doi: <https://doi.org/10.15359/ru.34-1.13>.
- [2] D. M. Barragán-Díaz y D. M. Barragán-Díaz, «La experiencia estudiantil universitaria en América Latina: una revisión de literatura», *Rev. Colomb. Educ.*, n.º 78, pp. 147-172, abr. 2020, doi: 10.17227/rce.num78-6708.
- [3] Arias, Bastidas, Salazar, «Estudio sobre la deserción estudiantil universitaria y sus implicaciones académicas, económicas y sociales | Bolentín de Coyuntura». <https://revistas.uta.edu.ec/erevista/index.php/bcoyu/article/view/677> (accedido 16 de agosto de 2022).
- [4] A. King-Domínguez, P. Backhouse Erazo, L. Améstica-Rivas, A. King-Domínguez, P. Backhouse Erazo, y L. Améstica-Rivas, «Deserción y graduación. Midiendo la eficiencia de las universidades estatales en Chile», *Mendive Rev. Educ.*, vol. 18, n.º 2, pp. 326-335, jun. 2020.
- [5] diario el universo, «La pandemia en Ecuador provoca más de 90 mil deserciones escolares», <https://www.eluniverso.com/noticias/ecuador/en-ecuador-90-mil-estudiantes-dejaron-de-asistir-a-clase-durante-la-pandemia-nota/>, 16 de junio de 2021. <https://www.eluniverso.com/noticias/ecuador/en-ecuador-90-mil-estudiantes-dejaron-de-asistir-a-clase-durante-la-pandemia-nota/> (accedido 2 de mayo de 2022).
- [6] Asamblea Nacional del Ecuador, «CONSTITUCIÓN DE LA REPÚBLICA DEL ECUADOR». 2008. [En línea]. Disponible en: https://www.educacionsuperior.gob.ec/wp-content/uploads/downloads/2012/07/Normas_constitucionales.pdf
- [7] Asamblea Nacional del Ecuador, «LEY ORGANICA DE EDUCACION SUPERIOR, LOES». 2010. [En línea]. Disponible en: https://www.educacionsuperior.gob.ec/wp-content/uploads/downloads/2014/03/LEY_ORGANICA_DE_EDUCACION_SUPERIOR_LOES.pdf
- [8] Definicion.de, «Definición de deserción — Definicion.de», *Definición.de*. <https://definicion.de/desercion/> (accedido 2 de mayo de 2022).
- [9] F. I. González Catalán y K. J. Arismendi Vera, «Deserción Estudiantil en la Educación Superior Técnico-Profesional: Explorando los factores que inciden en alumnos de primer año», *Rev. Educ. Super.*, vol. 47, n.º 188, pp. 109-137, dic. 2018.
- [10] A. M. Pérez, C. R. Escobar, M. R. Toledo, L. B. Gutierrez, y G. M. Reyes, «Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins», *Educ. E Pesqui.*, vol. 44, jun. 2018, doi: 10.1590/S1678-4634201844172094.
- [11] M. Esteban, A. Bernardo, E. Tuero, A. Cervero, y J. Casanova, «Variables influyentes en progreso académico y permanencia en la universidad», *Eur. J. Educ. Psychol.*, vol. 10, n.º 2, Art. n.º 2, 2017, doi: 10.30552/ejep.v10i2.125.
- [12] O. A. Tapasco Alzate *et al.*, «Deserción estudiantil: incidencia de factores institucionales relacionados con los procesos de admisión», *Educ. Educ.*, vol. 22, n.º 1, pp. 81-100, abr. 2019, doi: 10.5294/edu.2019.22.1.5.
- [13] S. V. Muñoz-Camacho, T. Gallardo, M. Muñoz-Bravo, y C. A. Muñoz-Bravo, «Probabilidad de Deserción Estudiantil en Cursos de Matemáticas Básicas en Programas Profesionales de la Universidad de Los Andes-Venezuela», *Form. Univ.*, vol. 11, n.º 4, pp. 33-42, ago. 2018, doi: 10.4067/S0718-50062018000400033.
- [14] Á. E. Torres-Zapata *et al.*, «Reprobación, síntoma de deserción escolar en licenciatura en Nutrición de la Universidad Autónoma del Carmen», *RIDE Rev. Iberoam. Para Investig. El Desarro. Educ.*, vol. 10, n.º 20, jun. 2020, doi: 10.23913/ride.v10i20.602.
- [15] F. E. B. Guerrero, M. P. P. Ortega, y L. E. I. Peña, «Reprobación y abandono de estudiantes universitario», *UNIANDES EPISTEME*, vol. 7, n.º 4, Art. n.º 4, oct. 2020.
- [16] J. A. Gallegos *et al.*, «Factores Determinantes en la Deserción Universitaria. Caso

Facultad de Ciencias Económicas y Administrativas de la Universidad Católica de la Santísima Concepción (Chile)», *Form. Univ.*, vol. 11, n.º 3, pp. 11-18, jun. 2018, doi: 10.4067/S0718-50062018000300011.

[17] R. S. Vizuete Córdova, «Deserción escolar de los estudiantes del colegio técnico Vicente Rocafuerte de la ciudad de Quito de la especialidad de mecánica automotriz y la repercusión en el ámbito laboral (año 2017-2017)», 2017, Accedido: 15 de agosto de 2022. [En línea]. Disponible en: <http://www.dspace.uce.edu.ec/handle/25000/14152>

[18] J. L. Machado Vallejo, «Modelado del tiempo de estancia estudiantil en la FICFM de la Universidad Central del Ecuador», 2019, Accedido: 15 de agosto de 2022. [En línea]. Disponible en: <http://www.dspace.uce.edu.ec/handle/25000/18608>

[19] J. P. Reyes Velásquez y S. E. Rosero Chango, «Deserción escolar en estudiantes de 5 a 17 años en las provincias de Pichincha y Guayas en el período 2013-2017», 2019, Accedido: 15 de agosto de 2022. [En línea]. Disponible en: <http://www.dspace.uce.edu.ec/handle/25000/18085>

[20] S. M. Armas Sopalo, «Análisis de los factores que inciden en la deserción estudiantil de los niños, niñas y adolescentes ecuatorianos en el periodo 2009-2019», sep. 2020, Accedido: 14 de agosto de 2022. [En línea]. Disponible en: <http://bibdigital.epn.edu.ec/handle/15000/21111>

[21] S. V. Campoverde Aguilar, «Deserción estudiantil en el instituto tecnológico superior de fútbol, Guayaquil, 2019», *Repos. Inst. - UCV*, 2020, Accedido: 15 de agosto de 2022. [En línea]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/46493>

[22] R. Timarán-Pereira, J. Caicedo-Zambrano, y A. Hidalgo-Troya, «Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11º», *Rev. Investig. Desarro. E Innov.*, vol. 9, n.º 2, Art. n.º 2, feb. 2019, doi: 10.19053/20278306.v9.n2.2019.9184.

[23] C. A. Frutos Ortega, «Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato», 2017, Accedido: 14 de agosto de 2022. [En línea]. Disponible en: <https://repositorio.uta.edu.ec:8443/jspui/handle/123456789/25797>

[24] Fundación Universitaria Internacional del Trópico Americano, J. E. Chaparro, J. Cuatindoy, Universidad de Medellín. Colombia, N. Barrera, y Universidad Pedagógica y Tecnológica de Colombia, «Análisis comparativo de técnicas de clasificación para determinar la deserción estudiantil de la facultad de ingeniería de la Universidad de Antioquia, Colombia», *ESPACIOS*, vol. 42, n.º 07, pp. 63-81, abr. 2021, doi: 10.48082/espacios-a21v42n07p05.

[25] E. Acosta-Gonzaga y A. Ramirez-Arellano, «Estudio comparativo de técnicas de analítica del aprendizaje para predecir el rendimiento académico de los estudiantes de educación superior», *CienciaUAT*, vol. 15, n.º 1, pp. 63-74, dic. 2020, doi: 10.29059/cienciauat.v15i1.1392.

[26] A. Carrasquilla-Batista *et al.*, «Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal», *Rev. Tecnol. En Marcha*, vol. 29, pp. 33-45, dic. 2016, doi: 10.18845/tm.v29i8.2983.

[27] Piedad Cristina Martínez Carazo, «(PDF) El método de estudio de caso: Estrategia metodológica de la investigación científica», marzo de 2006. https://www.researchgate.net/publication/28195979_El_metodo_de_estudio_de_caso_Estrategia_metodologica_de_la_investigacion_cientifica (accedido 15 de agosto de 2022).

[28] G. Montes, «METODOLOGÍA Y TÉCNICAS DE DISEÑO Y REALIZACIÓN DE ENCUESTAS EN EL ÁREA RURAL», *Temas Soc.*, n.º 21, pp. 39-50, 2000.

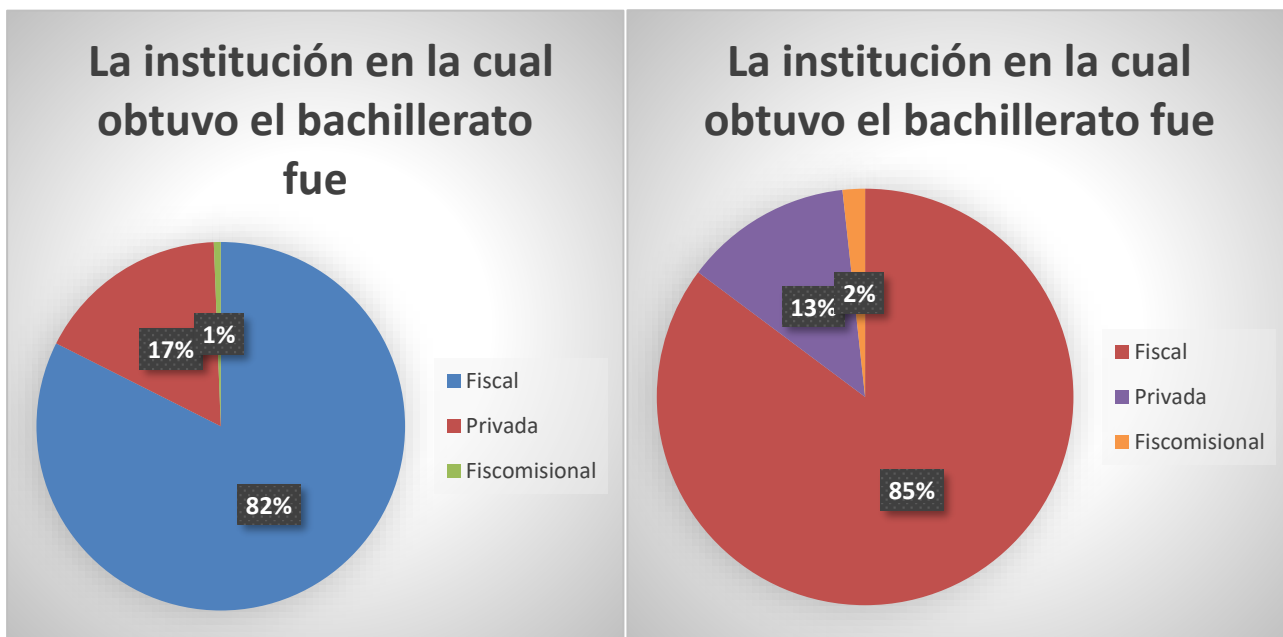
[29] «Fuentes primarias y secundarias. ¿Por qué no abusar de las últimas?», *Comunicar. Escuela de Autores*. <https://www.revistacomunicar.com/wp/escuela-de-autores/fuentes-primarias-y-secundarias-por-que-no-abusar-de-las-ultimas/> (accedido 17 de septiembre de 2022).

[30] V. Pedrero *et al.*, «Generalidades del Machine Learning y su aplicación en la gestión

- sanitaria en Servicios de Urgencia», *Rev. Médica Chile*, vol. 149, n.º 2, pp. 248-254, feb. 2021, doi: 10.4067/s0034-98872021000200248.
- [31] «Inteligencia artificial y condición humana: ¿Entidades contrapuestas o fuerzas complementarias?», *Rev. Cienc. Soc.*, 2021, doi: 10.31876/rsc.v27i2.35937.
- [32] «Redes neuronales artificiales: qué son y cuáles son sus usos», *UNIR*. <https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/> (accedido 1 de septiembre de 2022).
- [33] «RPods - Máquinas de Vector Soporte». https://rpubs.com/Cristina_Gil/SVM (accedido 4 de septiembre de 2022).
- [34] E. G. Muñoz, F. O. Cedeño, S. M. Ruiz, y J. C. Cruz, «Aplicación de redes neuronales para predecir el éxito de la compra de deuda a una entidad financiera», *Rev. Espac.*, vol. 40, n.º 20, jun. 2019, Accedido: 16 de noviembre de 2022. [En línea]. Disponible en: <https://www.revistaespacios.com/a19v40n20/19402006.html>
- [35] Joaquin Amat Rodrigo, «Máquinas de Vector Soporte (Support Vector Machines, SVMs)». https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines (accedido 8 de octubre de 2022).
- [36] E. E. García Díaz y F. Lozano Martínez, «Máquinas de vectores de soporte», *Rev. Ing.*, n.º 24, pp. 62-70, nov. 2006.
- [37] S. L. González-Ruiz, I. Gómez-Gallego, J. L. Pastrana-Brincones, y A. Hernández-Mendo, «Algoritmos de clasificación y redes neuronales en la observación automatizada de registros», *Cuad. Psicol. Deporte*, vol. 15, n.º 1, pp. 31-40, ene. 2015, doi: 10.4321/S1578-84232015000100003.
- [38] R. KeepCoding, «¿Qué son los Datasets? [4 sitios donde encontrarlos]», 27 de febrero de 2020. <https://keepcoding.io/blog/que-son-datasets/> (accedido 8 de octubre de 2022).
- [39] «Clasificación con datos desbalanceados | Aprende Machine Learning». <https://www.aprendemachinelearning.com/clasificacion-con-datos-desbalanceados/> (accedido 8 de octubre de 2022).
- [40] J. K. H. Osorio y G. D. Santacoloma, «METODOLOGÍA DE CLASIFICACIÓN DE DATOS DESBALANCEADOS BASADO EN MÉTODOS DE SUBMUESTREO.», p. 67.
- [41] J. Brownlee, «A Gentle Introduction to k-fold Cross-Validation», *Machine Learning Mastery*, 22 de mayo de 2018. <https://machinelearningmastery.com/k-fold-cross-validation/> (accedido 1 de septiembre de 2022).
- [42] «La matriz de confusión y sus métricas – Inteligencia Artificial –», *Juan Barrios*, 26 de julio de 2019. <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/> (accedido 8 de octubre de 2022).
- [43] M. de los A. Cienfuegos Velasco, A. Cienfuegos Velasco, M. de los A. Cienfuegos Velasco, y A. Cienfuegos Velasco, «Lo cuantitativo y cualitativo en la investigación. Un apoyo a su enseñanza», *RIDE Rev. Iberoam. Para Investig. El Desarro. Educ.*, vol. 7, n.º 13, pp. 15-36, dic. 2016.
- [44] «Cómo interpretar la matriz de confusión: ejemplo práctico», *Think Big*, 13 de diciembre de 2021. <https://empresas.blogthinkbig.com/como-interpretar-la-matriz-de-confusion-ejemplo-practico/> (accedido 29 de septiembre de 2022).

5.4 ANEXOS

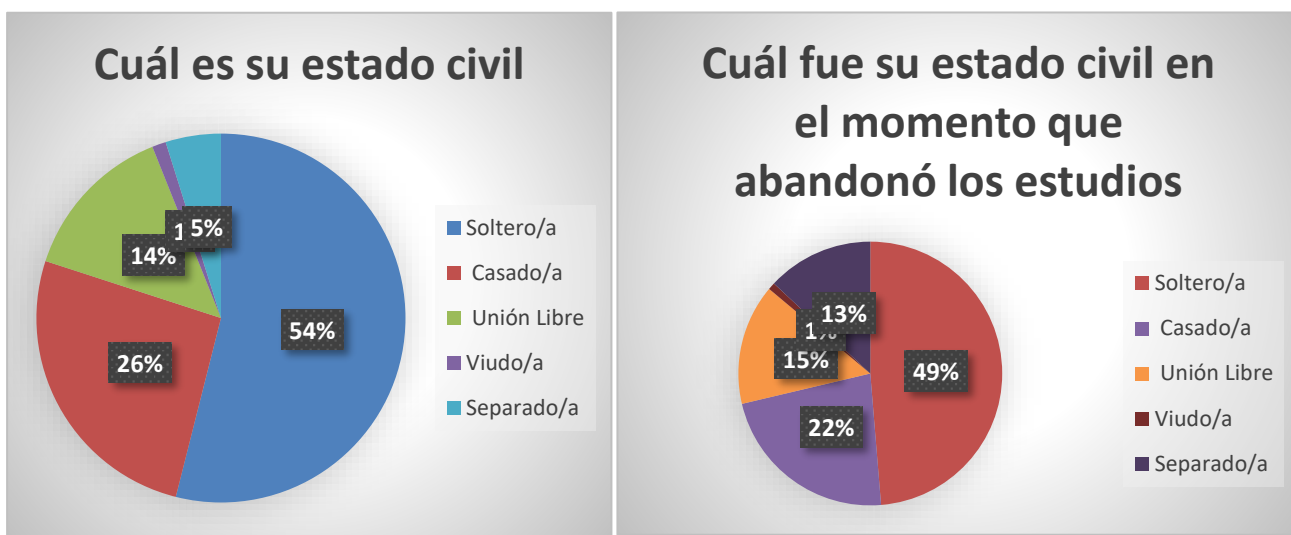
Anexo 1. Resultado de la institución del encuestado.



(a) (b)
Figura 1A. Resultado institución que obtuvo el bachillerato. (a) No desertores. (b) desertores

Al observar el resultado se nota que existe una similitud marcada entre los diferentes factores encontrados indistintamente de la procedencia de las instituciones objeto de estudio

Anexo 2. Resultado del estado civil



(a) (b)
Figura 2A. Resultado del estado civil. (a) No desertores. (b) desertores

Analizando el factor estado civil de los encuestados se observa que tanto el segmento de los estudiantes regulares tanto como los desertores atienden a un conglomerado de solteros

Anexo 3. Resultado de la edad del encuestado.

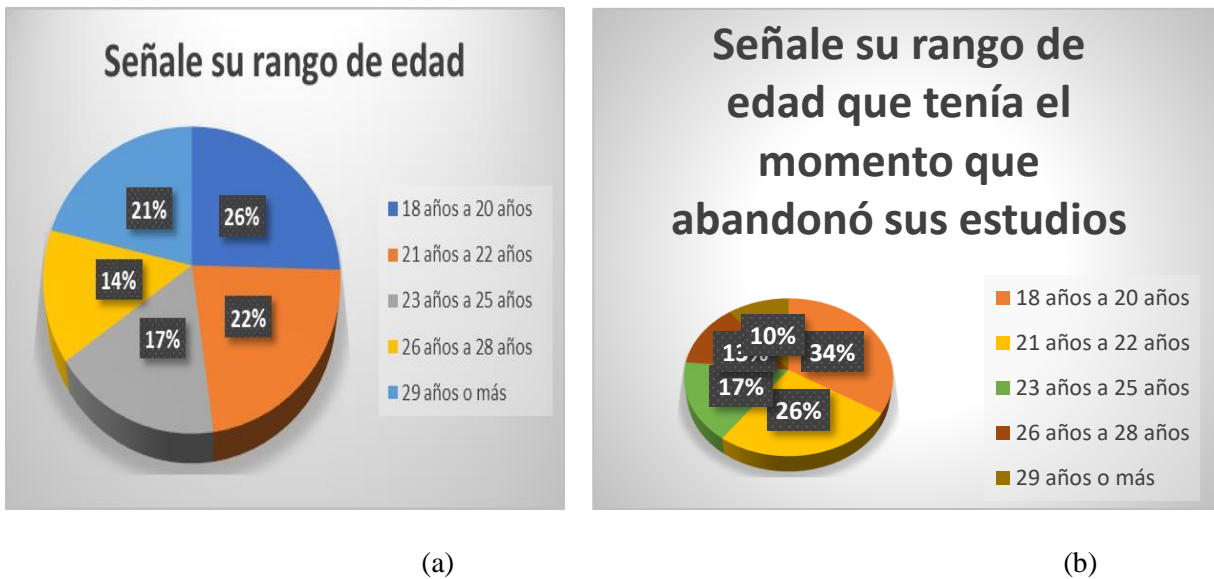


Figura 3A. Edad del encuestado. (a) No desertores. (b) desertores

De acuerdo a los datos obtenidos se visualiza que los estudiantes regulares ingresan con margen de edad más temprana a diferencia de quienes desertan.

Anexo 4. Resultado de gastos para la educación.

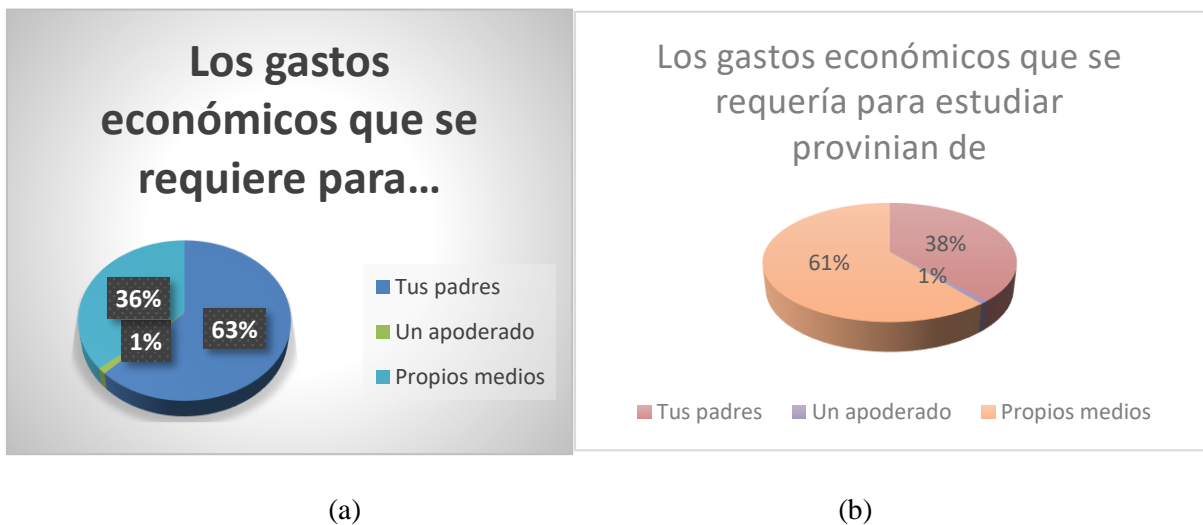
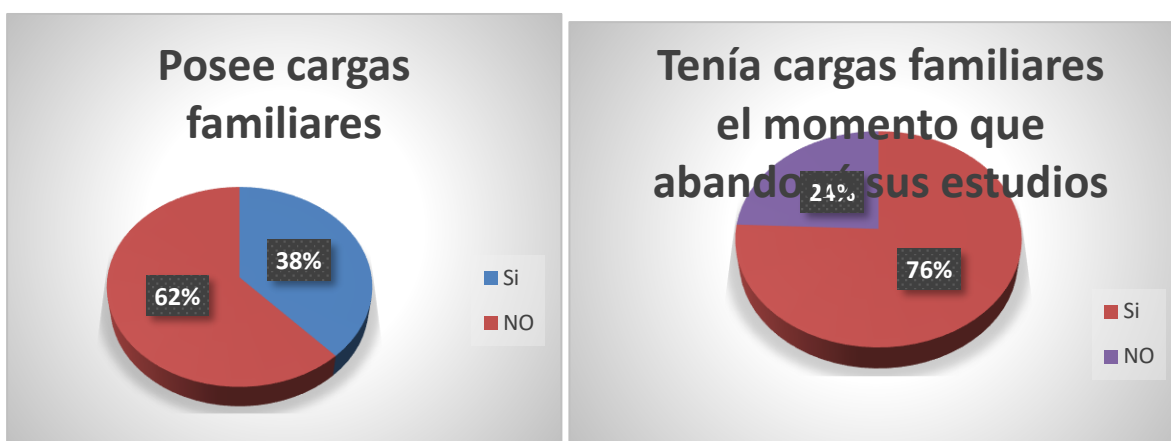


Figura 4A. Gastos para la educación. (a) No desertores. (b) desertores

De acuerdo a los gráficos presentados se observa que un importante porcentaje de estudiantes desertores toman la decisión de abandonar sus estudios por cuestiones de trabajo

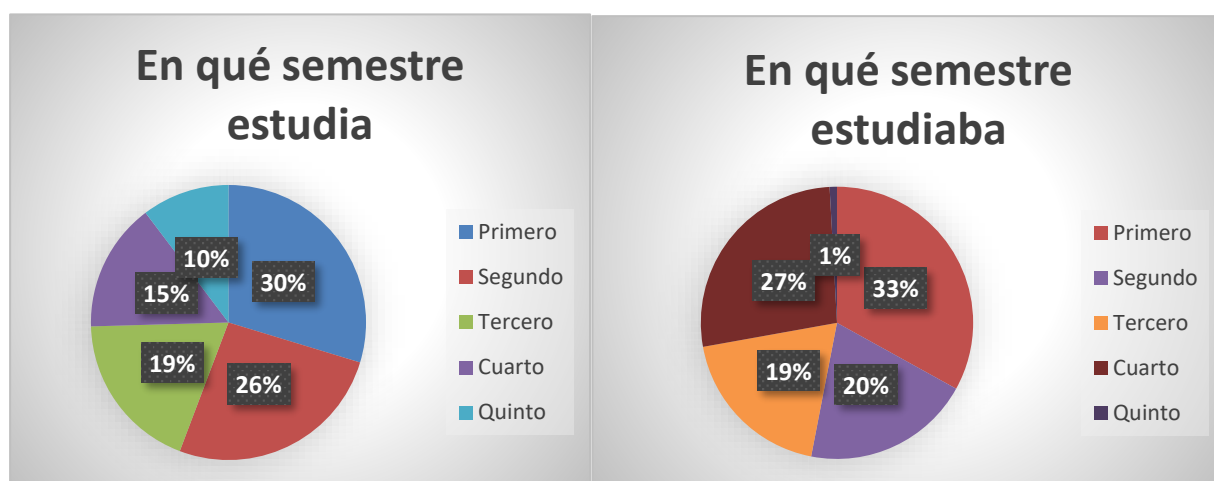
Anexo 5. Resultado de cargas familiares del encuestado.



(a) (b)
Figura 5A. Cargas familiares. (a) No desertores. (b) desertores

Según los resultados se visualiza que los estudiantes que toman la decisión de desertar son los que tienen responsabilidades familiares en un porcentaje mayor que los estudiantes que son regulares.

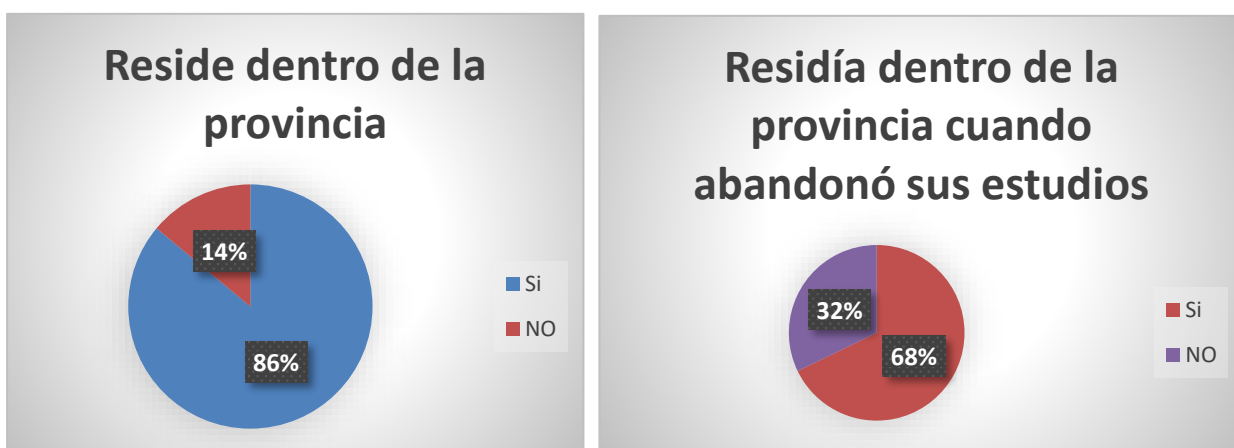
Anexo 6. Resultado del semestre actual del encuestado.



(a) (b)
Figura 6A. Semestre actual de estudio. (a) No desertores. (b) desertores

De acuerdo con los resultados obtenidos en la pregunta se evidencia que los estudiantes regulares llegan al último semestre en un mayor porcentaje que los estudiantes desertores

Anexo 7. Resultado de la provincia de residencia actual del encuestado.



(a) (b)
Figura 7A. Provincia de residencia. (a) No desertores. (b) desertores

Según los datos obtenidos en la pregunta es notorio que son más propicios a desertar los estudiantes que viven fuera de la provincia.

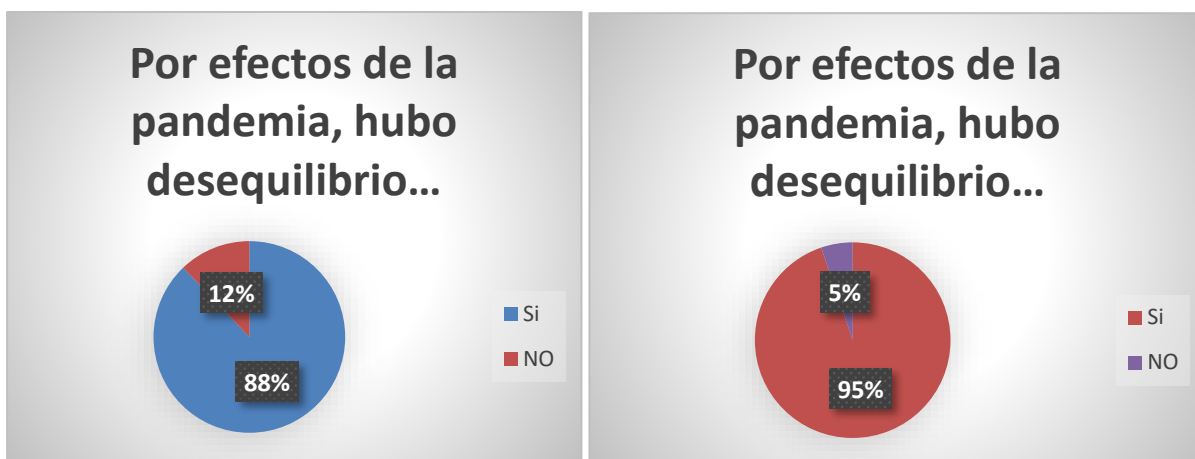
Anexo 8. Resultado de la dificultad del proceso de aprendizaje.



(a) (b)
Figura 8A. Dificultad de aprendizaje. (a) No desertores. (b) desertores

Según los datos obtenidos en la pregunta es notorio que los estudiantes que tuvieron más dificultad en la emergencia sanitaria de manera especial los que desertaron definitivamente.

Anexo 9. Resultado del desequilibrio económico del encuestado.



(a) (b)
Figura 9A. Desequilibrio económico. (a) No desertores. (b) desertores

De acuerdo a los resultados obtenidos se puede evidenciar que hubo más desequilibrio económico por efectos de la pandemia en los estudiantes desertores.

Anexo 10. Resultado del apoyo del docente.



(a) (b)
Figura 10A. Apoyo del docente. (a) No desertores. (b) desertores

De acuerdo a los resultados obtenidos se puede evidenciar que los encuestados manifiestan que faltó apoyo por parte de los docentes para los desertores como para los no desertores.

Anexo 11.. Resultado del cálculo de la muestra poblacional

Cuadro de Muestra de Acuerdo a la Población (N)												
	1%	2,0%	2,5%	3,0%	3,5%	4,0%	5,0%	6,0%	7,0%	8,0%	9,0%	10,0%
2000	1.655	1.091	869	696	563	462	322	235	179	140	112	92
4000	2.824	1.501	1.110	843	656	522	351	250	187	145	115	94
6000	3.693	1.715	1.223	906	693	546	361	255	190	146	116	95
8000	4.366	1.847	1.289	942	714	558	367	258	193	147	117	95
10000	4.899	1.936	1.332	964	727	566	370	260	192	148	117	95
12000	5.335	2.001	1.362	980	736	572	372	261	193	148	117	95
14000	5.696	2.050	1.385	992	742	576	374	262	193	148	118	95
16000	6.002	2.088	1.402	1.000	747	579	375	262	194	149	118	95
18000	6.263	2.118	1.416	1.007	751	581	376	263	194	149	118	96
20000	6.495	2.141	1.427	1.012	754	582	376	263	194	149	118	96
22000	6.688	2.165	1.436	1.018	757	584	378	264	194	149	118	96
24000	6.859	2.183	1.444	1.022	759	586	378	264	194	149	118	96
26000	7.013	2.198	1.451	1.025	761	587	379	264	195	149	118	96
28000	7.151	2.211	1.457	1.028	763	588	379	264	195	149	118	96
30000	7.275	2.223	1.462	1.030	764	588	379	264	195	149	118	96
32000	7.387	2.233	1.466	1.033	765	589	380	265	195	149	118	96
34000	7.489	2.243	1.470	1.035	766	590	380	265	195	149	118	96
36000	7.581	2.251	1.474	1.036	767	590	380	265	195	149	118	96

MARGEN DE ERROR (común en auditoría)	5,0%
TAMAÑO POBLACIÓN	1.036 *
NIVEL DE CONFIANZA (común en auditoría)	95% **
Valores Z (valor del nivel de confianza)	90% 95% 97% 98% 99%
Varianza (valor para reemplazar en la fórmula)	1,645 1,960 2,170 2,326 2,576

Nota:
 * Ingresar Tamaño de la Población - Universo
 ** Valor fijo para auditoría
 *** Ingresar los datos de la escala de acuerdo al tamaño de la población (universo)

$$\text{TAMAÑO DE LA MUESTRA} = \frac{N * (\alpha_e * 0,5)^2}{1 + (e^2 * (N - 1))} = 280$$

Donde:
 α_e

Fuente: ISTB
Autor: Patricio Mejía

Anexo 12. Preguntas elaboradas a estudiantes No desertores

Estimados señores y señoritas estudiantes del Instituto Superior Tecnológico Bolívar, por medio del presente formulario, solicito a usted se digne llenar el mismo, con la finalidad de poder desarrollar un proyecto de investigación. La encuesta, está dirigida a aquellos estudiantes que NO tuvieron dificultades en aprobar el semestre. La presente encuesta es de carácter anónimo Gracias por su colaboración
1. La institución en la cual obtuvo el bachillerato fue A. Fiscal b. Particular c. Fiscomisional
2. ¿Cuál es su estado civil? a. Solter@ b. Casad@ c. Unión libre d. Separad@
3. Señale su rango de edad. a. 18 años a 20 años b. 21 años a 22 años c. 23 años a 25 años d. 26 años a 28 años e. 29 años o más
4. ¿Los gastos económicos que se requiere para estudiar proviene de? a. Tus padres b. Un apoderado c. Propios medios
5. ¿Posee cargas familiares? a. SI b. NO
6. ¿En qué semestre estudia? a. Primero b. Segundo c. Tercero d. Cuarto e. Quinto
7. ¿Reside dentro de la provincia? a. SI b. NO
8. ¿La emergencia sanitaria a través de la pandemia Covid 19 ocasionó dificultad en su proceso de aprendizaje? a. SI b. NO
9. ¿Por efectos de la pandemia, hubo desequilibrio económico en su entorno familiar? a. SI b. NO
10. ¿Cree usted que le falta apoyo por parte de sus docentes en el proceso de aprendizaje? a. SI b. NO

Fuente: ISTB

Anexo 13. Encuesta Elaborada y enviada a los estudiantes desertores

Estimados señores y señoritas estudiantes del Instituto Superior Tecnológico Bolívar, por medio del presente formulario, solicito a usted se digne llenar el mismo, con la finalidad de poder desarrollar un proyecto de investigación. La encuesta, está dirigida a aquellos estudiantes que por diferentes situaciones tuvieron que abandonar sus estudios. La presente encuesta es de carácter anónimo. Gracias por su colaboración
1. La institución en la cual obtuvo el bachillerato fue A. Fiscal b. Particular c. Fiscomisional
2. ¿Cuál es su estado civil? a. Solter@ b. Casad@ c. Unión libre d. Separad@
3. Señale su rango de edad. a. 18 años a 20 años b. 21 años a 22 años c. 23 años a 25 años d. 26 años a 28 años e. 29 años 0 más
4. ¿Los gastos económicos que se requiere para estudiar provenían de? a. Tus padres b. Un apoderado c. Propios medios
5. ¿Tenía cargas familiares el momento que abandonó sus estudios? a. SI b. NO
6. ¿En qué semestre estudiaba? a. Primero b. Segundo c. Tercero d. Cuarto e. Quinto
7. ¿Residía dentro de la provincia cuando abandonó sus estudios? a. SI b. NO
8. ¿Cuándo abandonó los estudios la pandemia Covid 19 ocasionó dificultad en su proceso de aprendizaje? a. SI b. NO

9. ¿Por efectos de la pandemia, hubo desequilibrio económico en su entorno familiar cuando abandonó sus estudios? a. SI b. NO
10. ¿Cree usted que le faltó apoyo por parte de sus docentes en el proceso de aprendizaje? a. SI b. NO

Fuente: ISTB

Anexo 14. Resultado de la encuesta aplicada a los estudiantes No desertores

RESULTADO DE LA ENCUESTA	
1. La institución en la cual obtuvo el bachillerato fue A. Fiscal 136 b. Privada 28 c. Fiscomisional 1	
2. ¿Cuál es su estado civil? a. Solter@ 89 b. Casad@ 43 c. Unión libre 23 d. Viud@ 2 e. Separad@ 8	
3. Señale su rango de edad. a. 18 años a 20 años 42 b. 21 años a 22 años 37 c. 23 años a 25 años 28 d. 26 años a 28 años 33 e. 29 años o más 35	
4. ¿Los gastos económicos que se requiere para estudiar proviene de? a. Tus padres 103 b. Un apoderado 2 c. Propios medios 60	
5. ¿Posee cargas familiares? a. SI 63 b. NO 102	
6. ¿En qué semestre estudia? a. Primero 49 b. Segundo 43 c. Tercero 31 d. Cuarto 25 e. Quinto 17	
7. ¿Reside dentro de la provincia? a. SI 142 b. NO 23	
8. ¿La emergencia sanitaria a través de la pandemia Covid 19 ocasionó dificultad en su proceso de aprendizaje? a. SI 144 b. NO 21	
9. ¿Por efectos de la pandemia, hubo desequilibrio económico en su entorno familiar? a. SI 145 b. NO 20	
10. ¿Cree usted que le falta apoyo por parte de sus docentes en el proceso de aprendizaje? a. SI 96 b. NO 69	

Fuente: ISTB

Anexo 15. Resultado de la encuesta aplicada a los estudiantes Desertores

RESULTADO DE LA ENCUESTA	
1. La institución en la cual obtuvo el bachillerato fue A. Fiscal 98 b. Particular 15 c. Fiscomisional 2	
2. ¿Cuál es su estado civil? a. Solter@ 56 b. Casad@ 26 c. Unión libre 17 d. Viud@ 1 e. Separad@ 15	
3. Señale su rango de edad. a. 18 años a 20 años 39 b. 21 años a 22 años 30 c. 23 años a 25 años 19 d. 26 años a 28 años 15 e. 29 años 0 más 12	
4. ¿Los gastos económicos que se requiere para estudiar provenían de? a. Tus padres 44 b. Un apoderado 1 c. Propios medios 70	
5. ¿Tenía cargas familiares el momento que abandonó sus estudios? a. SI 87 b. NO 28	
6. ¿En qué semestre estudiaba? a. Primero 38 b. Segundo 23 c. Tercero 22 d. Cuarto 31 e. Quinto 1	
7. ¿Residía dentro de la provincia cuando abandonó sus estudios? a. SI 78 b. NO 37	
8. ¿Cuándo abandonó los estudios la pandemia Covid 19 ocasionó dificultad en su proceso de aprendizaje? a. SI 104 b. NO 11	
9. ¿Por efectos de la pandemia, hubo desequilibrio económico en su entorno familiar cuando abandonó sus estudios? a. SI 109 b. NO 6	
10. Cree usted que le faltó apoyo por parte de sus docentes en el proceso de aprendizaje? a. SI 65 b. NO 50	

Fuente: ISTB

Anexo 16. Ponderado a aplicarse entre 0 y 1

PONDERACIÓN DE LA ENCUESTA	
1. La institución en la cual obtuvo el bachillerato fue	A. Fiscal 1 b. Privada 0,05 c. Fiscomisional 0,5
2. ¿Cuál es su estado civil?	a. Solter@ 0,05 b. Casad@ 1 c. Unión libre 0,5 d. Viud@ 0,25 e. Separad@ 0,75
3. Señale su rango de edad.	a. 18 años a 20 años 0,05 b. 21 años a 22 años 0,25 c. 23 años a 25 años 0,5 d. 26 años a 28 años 0,75 e. 29 años o más 1
4. ¿Los gastos económicos que se requiere para estudiar proviene de?	a. Tus padres 0,05 b. Un apoderado 0,5 c. Propios medios 1
5. ¿Posee cargas familiares?	a. SI 1 b. NO 0,05
6. ¿En qué semestre estudia?	a. Primero 1 b. Segundo 0,75 c. Tercero 0,5 d. Cuarto 0,25 e. Quinto 0,05
7. ¿Reside dentro de la provincia?	a. SI 0,05 b. NO 1
8. ¿La emergencia sanitaria a través de la pandemia Covid 19 ocasionó dificultad en su proceso de aprendizaje?	a. SI 1 b. NO 0,05
9. ¿Por efectos de la pandemia, hubo desequilibrio económico en su entorno familiar?	a. SI 1 b. NO 0,05
10. ¿Cree usted que le falta apoyo por parte de sus docentes en el proceso de aprendizaje?	a. SI 1 b. NO 0,05

Fuente: ISTB

Anexo 17. Resultados con datos ponderados

	A	B	C	D	E	F	G	H	I	J	K
100	1	1	0,25	0	1	1	0	1	1	1	0
101	1	1	0,25	0	1	1	0	1	1	1	0
102	1	1	0,25	0	1	1	0	1	1	1	0
103	1	1	0,25	0	1	0,75	0	1	1	1	0
104	1	1	0,25	0	1	0,75	0	1	1	1	0
105	1	1	0,25	0	0	0,75	0	1	1	1	0
106	1	1	0,25	0	0	0,75	0	1	1	0	0
107	1	1	0,25	0	0	0,75	0	1	1	0	0
108	1	1	0,5	0	0	0,75	0	1	1	0	0
109	1	1	0,5	0	0	0,75	0	1	1	0	0
110	1	1	0,5	0	0	0,75	0	1	1	0	0
111	1	1	0,5	0	0	0,75	0	1	1	0	0
112	1	1	0,5	0	0	0,75	0	1	1	0	0
113	1	1	0,5	0	0	0,75	0	1	1	0	0
114	1	1	0,5	0	0	0,75	0	1	1	0	0
115	1	1	0,5	0	0	0,75	0	1	1	0	0
116	1	1	0,5	0	0	0,75	0	1	1	0	0
117	1	1	0,5	0	0	0,75	0	1	1	0	0
118	1	1	0,5	0	0	0,75	0	1	1	0	0
119	1	1	0,5	0	0	0,75	0	1	1	0	0

Fuente: Autor

Anexo 18. Resultado de prueba con 41 ensayos en RNA feed-forward

		ANN							Resultados						
nForTraining	capas ocultas	funcion de activacion	neuronas	lambda	numFolds	iteraciones	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalanceados)	Sensibilidad (desbalanceados)	Especificidad (desbalanceados)		
2															
3	0,6	10,5,1	relu, tanh, logsig	3	0,01	5	10	5	0,75	0,88	0,68	0,67	0,67	0,81	
4	0,6	15,13,10,1	tanh,relu,softplus,logsig	4	0,000001	10	8	4	0,71	0,73	0,81	0,72	0,76	0,9	
5	0,6	10, 5, 1	tanh, relu, logsig	3	0	2	10	10	0,69	0,84	0,58	0,37	0,4	0,81	
6	0,6	20,60,50,1	tanh,elu,elu,logsig	4	0	7	8	8	0,69	0,81	0,63	0,5	0,56	0,75	
7	0,6	15,25,10,1	tanh,relu,elu,logsig	4	0	5	8	4	0,66	0,87	0,47	0,39	0,47	0,75	
8	0,6	20,60,50,1	relu,softplus,tanh,logsig	4	0	5	8	8	0,65	0,72	0,66	0,13	0,14	0,93	
9	0,6	80,75,20,1	tanh,tanh,tanh,logsig	4	0	11	5	5	0,65	0,82	0,54	0,16	0,17	0,91	
10	0,6	8,6,1	tanh,softplus,logsig	3	0,001	5	5	4	0,64	0,97	0,23	0	0	1	
11	0,6	20,60,50,1	tanh,elu,elu,logsig	4	0	11	5	5	0,64	0,89	0,37	0,12	0,13	0,95	
12	0,6	10,30,15,1	elu,tanh,tanh,logsig	4	0	11	5	5	0,64	0,86	0,44	0,2	0,23	0,88	
13	0,6	10,5,1	relu, relu, logsig	3	0,01	3	10	5	0,63	0,78	0,52	0,55	0,69	0,58	
14	0,6	15,15,20,1	tanh,tanh,elu,logsig	4	1E-11	10	8	6	0,63	0,72	0,64	0,31	0,36	0,78	
15	0,6	20,25,30,1	relu,relu,tanh,logsig	4	1E-11	10	8	6	0,63	0,79	0,56	0,28	0,32	0,84	
16	0,6	50,70,30,7	relu,elu,tanh,tanh,logsig	5	0	12	5	3	0,61	0,89	0,35	0,087	0,1	0,93	
17	0,6	9,5,1	tanh,elu,logsig	3	0,001	5	5	5	0,6	0,81	0,41	0,21	0,21	0,87	
18	0,6	9,5,1	elu,elu,logsig	3	0,001	5	5	5	0,59	0,73	0,53	0,35	0,47	0,71	
19	0,6	30,20,35,1	tanh,logsig,elu,logsig	4	1E-11	5	8	6	0,58	0,83	0,4	0,21	0,24	0,87	
20	0,6	10,5,1	tanh,relu,logsig	3	0	5	10	5	0,57	0,82	0,26	0,14	0,14	0,82	

Anexo 19. Resultado de prueba con 41 ensayos en RNA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
22	0,6	15,13,10,1	tanh,relu,relu,logsig	4	0,000001	10	8	6	0,57	0,75	0,5	0,22	0,26	0,87
23	0,6	15,25,10,1	relu,relu,elu,logsig	4	0	5	8	4	0,57	0,71	0,56	0,19	0,2	0,88
24	0,6	9,6,1	relu, tanh, logsig	3	0	6	5	10	0,56	0,72	0,54	0,17	0,2	0,9
25	0,6	8,6,1	softplus,elu,logsig	3	0,001	5	5	4	0,56	0,84	0,3	0,11	0,13	0,93
26	0,6	20,25,30,1	softplus,relu,tanh,logsig	4	1E-11	10	8	6	0,55	0,72	0,54	0,25	0,29	0,85
27	0,6	15,25,10,1	tanh,tanh,relu,logsig	4	0	5	8	4	0,54	0,66	0,57	0,23	0,27	0,61
28	0,6	9,5,1	logsig,softplus,logsig	3	0,001	5	5	3	0,53	0,85	0,25	0,04	0,03	1
29	0,6	10,12,8,1	softplus,elu,relu,logsig	4	0,000001	10	8	4	0,53	0,72	0,48	0,19	0,23	0,9
30	0,6	10,25,15,9	tanh,relu,relu,tanh,logsig	5	1E-09	10	10	5	0,53	0,71	0,49	0,097	0,1	0,96
31	0,6	50,30,20,3	relu,elu,tanh,tanh,logsig	5	0	12	5	3	0,53	0,8	0,34	0,14	0,15	0,92
32	0,6	9,5,1	tanh,relu,logsig	3	0,001	5	5	5	0,47	0,71	0,38	0,18	0,24	0,84
33	0,6	30,40,15,0	tanh,elu,elu,relu,logsig	5	0	10	5	6	0,47	0,77	0,25	0,03	0,03	0,99
34	0,6	9,5,1	relu,relu,logsig	3	0,001	5	5	5	0,46	0,61	0,6	0,13	0,14	0,94
35	0,6	30,15,10,2	elu,tanh,stanh,logsig	5	1E-09	12	3	5	0,42	0,72	0,27	0	0	1
36	0,6	30,15,10,2	tanh,relu,tanh,relu,logsig	5	0	12	3	5	0,41	0,7	0,3	0	0	1
37	0,6	20,10,20,3	tanh,elu,tanh,relu,logsig	5	0	12	2	5	0,4	0,68	0,34	0	0	1
38	0,6	50,70,80,7	tanh,elu,elu,relu,logsig	5	0	13	2	3	0,39	0,6	0,47	0	0	1
39	0,6	90,80,85,9	tanh,relu,relu,relu,logsig	5	0	10	5	4	0,39	0,59	0,48	0	0	1
40	0,6	30,15,10,2	elu,tanh,softplus,logsig	5	1E-09	12	3	5	0,29	0,52	0,47	0	0	1
41	0,6	10,25,15,9	tanh,elu,relu,logsig,logsig	5	0,000001	10	10	5	0,266	0,46	0,54	0	0	1
42														

Anexo 20 Resultado de prueba con 41 ensayos en SVM

Resultados SVM										
	fractionForTrainingAndValida	numFolds	kernel	numRepetitions	F1Score (balanceados)	sensibilidad (balanceados)	Especificidad (balanceados)	F1Score (desbalance)	Sensibilidad (desbalanceado)	Especificidad (desbalanceados)
2	0,7	50	linear	20	0,69	0,9184	0,638	0,5877	0,59	0,99
3	0,7	80	linear	30	0,76	0,88	0,64	0,54	0,56	0,98
4	0,6	67	linear	30	0,79	0,91	0,62	0,51	0,52	0,98
5	0,6	69	linear	20	0,55	0,58	0,91	0,54	0,55	0,98
6	0,8	79	linear	40	0,7	0,8	0,67	0,56	0,58	0,96
7	0,6	59	linear	50	0,74	0,87	0,57	0,5	0,5	0,97
8	0,6	69	linear	5	0,58	0,6	0,91	0,57	0,57	0,98
9	0,6	68	linear	50	0,74	0,86	0,64	0,52	0,54	0,97
10	0,6	69	linear	60	0,75	0,87	0,65	0,53	0,55	0,97
11	0,6	69	linear	70	0,8	0,92	0,67	0,53	0,55	0,98
12	0,6	69	linear	65	0,71	0,83	0,63	0,48	0,5	0,97
13	0,6	65	linear	66	0,8	0,93	0,6	0,44	0,46	0,96
14	0,7	69	linear	10	0,73	0,86	0,61	0,57	0,57	0,96
15	0,7	79	linear	20	0,74	0,85	0,67	0,58	0,58	0,98
16	0,7	80	linear	25	0,65	0,74	0,7	0,53	0,53	0,98
17	0,7	80	linear	10	0,77	0,89	0,62	0,52	0,53	0,95
18	0,7	80	linear	30	0,73	0,84	0,66	0,56	0,57	0,96
19	0,6	69	polynomial	60	0,66	0,74	0,78	0,62	0,65	0,95

Anexo 21. Resultado de prueba con 41 ensayos en SVM

22	0,7	69	polynomial	15	0,84	0,92	0,76	0,64	0,65	0,98
23	0,7	69	polynomial	50	0,82	0,92	0,67	0,47	0,55	0,76
24	0,7	69	polynomial	100	0,77	0,85	0,72	0,63	0,63	0,99
25	0,6	59	polynomial	20	0,84	0,92	0,69	0,56	0,57	0,96
26	0,6	50	polynomial	10	0,86	0,93	0,79	0,54	0,57	0,89
27	0,6	45	polynomial	20	0,74	0,9	0,68	0,56	0,56	0,96
28	0,6	40	polynomial	30	0,79	0,94	0,71	0,59	0,58	1
29	0,6	40	polynomial	40	0,71	0,86	0,72	0,6	0,61	0,97
30	0,6	40	gauss	10	0,64	0,85	0,6	0,46	0,46	0,97
31	0,6	59	gauss	20	0,798	0,87	0,74	0,53	0,54	0,99
32	0,6	50	gauss	10	0,73	0,83	0,7	0,54	0,55	0,97
33	0,6	60	gauss	20	0,62	0,64	0,87	0,57	0,57	0,98
34	0,7	69	gauss	10	0,72	0,81	0,72	0,54	0,56	0,96
35	0,7	70	gauss	20	0,67	0,75	0,71	0,57	0,57	0,98
36	0,6	50	gauss	20	0,63	0,67	0,85	0,61	0,61	0,97
37	0,6	55	gauss	30	0,79	0,91	0,61	0,57	0,58	0,96
38	0,6	56	gauss	30	0,69	0,78	0,73	0,52	0,53	0,97
39	0,6	59	gauss	60	0,72	0,83	0,59	0,49	0,5	0,96
40	0,6	30	gauss	60	0,72	0,76	0,82	0,67	0,6	0,98
41	0,6	20	gauss	60	0,71	0,8	0,7	0,61	0,5	0,96
42	0,6	10	gauss	60	0,76	0,9	0,68	0,65	0,56	0,93

Anexo 22. Programa en Matlab del entrenamiento para obtener los resultados en Matriz de confusión con SVM

Configuración para la validación cruzada

```

numFolds = 20;
kernelFunction = 'polynomial'
numRepetitions = 10;
minNumExamples = min(numExamples_withLabel0_TrainVal,
numExamples_withLabel1_TrainVal);
if numFolds > minNumExamples
    error(['El número de grupos para la validación cruzada no puede ser
mayor que ' num2str(minNumExamples)]);
end
examplesPerGroup_withLabel0 =
floor(numExamples_withLabel0_TrainVal/numFolds);
examplesPerGroup_withLabel1 =
floor(numExamples_withLabel1_TrainVal/numFolds);
accuracyCV_WithBalancedData = zeros(numRepetitions, numFolds);
sensitivityCV_WithBalancedData = zeros(numRepetitions, numFolds);
specificityCV_WithBalancedData = zeros(numRepetitions, numFolds);
F1ScoreCV_WithBalancedData = zeros(numRepetitions, numFolds);

```

Anexo 23. Programa en Matlab del entrenamiento para obtener los resultados en Matriz de confusión con RNA feed forward

Configuración para la validación cruzada

```
numFolds = 20;
n = size(dataX_withLabel0_TrainVal,2);
numNeuronsPerLayer = [n 80 75 22 1];
transferFunctions{1} = 'none';
transferFunctions{2} = 'tanh';
transferFunctions{3} = 'tanh';
transferFunctions{4} = 'tanh';
transferFunctions{5} = 'logsig';
options.lambda = 0;
options.numIterations = 30;
options.reluThresh = 0;
numRepetitions = 10;
minNumExamples = min(numExamples_withLabel0_TrainVal,
numExamples_withLabel1_TrainVal);
if numFolds > minNumExample
```