

PROYECTO DE INVESTIGACIÓN CUANTITATIVO

TITULO

TÉCNICAS DE MINERÍA DE DATOS, APLICADAS AL ANÁLISIS PREDICTIVO EN EL APRENDIZAJE ESTUDIANTIL DE LA I.E. N ° 64103 TENIENTE DIEGO FERRE SOSA, AÑO 2022.

RESUMEN

La presente investigación tiene como objetivo Aplicar técnicas de minería de datos, aplicadas al análisis predictivo para determinar el aprendizaje estudiantil de la I.E. N° 64103 Teniente Diego Ferre Sosa. El estudio es de enfoque cuantitativo, de tipo aplicada, de nivel descriptiva - explicativa, cumple con un diseño pre experimental. En cuanto a los resultados, se pretende predecir el aprendizaje estudiantil aplicando técnicas de minería de datos.

Esta investigación cuenta con los siguientes puntos importantes tales como: justificación, antecedentes, hipótesis, objetivos y metodología correspondiente, de tal manera se especifica el procesamiento de los datos a obtener, el impacto esperado, los recursos necesarios, la localización en la que se desarrollara el estudio, cronograma detallado de las actividades y el respectivo presupuesto.

Palabras claves

Minería de datos, aprendizaje estudiantil, análisis predictivo, base de datos.

Abstract

The objective of this research is to apply data mining techniques applied to predictive analysis to determine student learning at I.E. N° 64103 Teniente Diego Ferre Sosa. The study is of quantitative approach, applied type, descriptive - explanatory level, complies with a pre-experimental design. As for the results, it is intended to predict student learning by applying data mining techniques.

This research has the following important points such as: justification, background, hypothesis, objectives and corresponding methodology, in such a way that the processing of the data to be obtained, the expected impact, the necessary resources, the location where the study will be developed, detailed schedule of activities and the respective budget are specified.

Keywords

Data Mining, student learning, predictive analytics, database.

I. PLANTEAMIENTO DEL PROBLEMA.

Durante años, el mundo ha tenido que afrontar el problema que se encuentra evidenciada de gran magnitud en países en desarrollo respecto al nivel de aprendizaje de los estudiantes, especificando que estos países suelen frecuentemente considerar a manera de ejemplo las estrategias preventivas de carácter emergentes de los países del primer mundo, siendo una decisión riesgosa, pues las condiciones de cada pueblo o ciudad no son las mismas de las que se encuentran en los países del primer mundo (Moreno y Aguilar, 2019).

Debido a ello, es que en Colombia se hace necesario buscar la innovación educativa, siendo este enfocado en la transformación digital de los tradicional, pues gracias al uso de las tecnologías es que se puede integrar y mejorar el desempeño de cada una de las áreas de estudio, a fin de obtener resultados mejorados en base al nivel de aprendizaje que obtenían anteriormente los estudiantes en los modelos tradicionales (Banoy, 2019).

En otro aspecto, se presenta en una I.E de Ecuador, en el que el cambio repentino de la gestión administrativa afectó considerablemente la toma de decisiones, pues no se adquiere un apoyo tecnológico para absolver dichas dificultades, además de poder estimar como es que los estudiantes se encuentran desempeñándose académicamente, por lo que se hizo necesario proponer la implementación de una técnica de predicción, el cual atribuye netamente en la del estado académico de los estudiantes en cuanto a su nivel de aprendizaje, siendo este otorgado por los mismos docentes quienes han seguido brindando sesiones de aprendizaje bajo una gestión administrativa deficiente (Villao, 2021).

Por otro lado, en el Perú, se manifestó algo similar en una I.E. en Huacho, siendo el principal problema el de poder aproximar una estimación de cómo es que se encuentran los estudiantes desempeñándose académicamente en base al nivel de aprendizaje según sus asignaturas, siendo este aspecto crucial para poder emplear estrategias de aprendizaje que permitan mejorar el rendimiento de estos estudiantes, haciendo así factible la inclusión de las tecnologías, siendo en este caso la aplicación de minería de datos, en el que se pudo predecir con exactitud y precisión el rendimiento de estos estudiantes, por medio del programa WEKA, siendo esta técnica específicamente el árbol

de decisiones (Díaz et al., 2021).

En este contexto, se demostró en la I.E. que se utilizó como caso de estudio que en todo su proceso educativo colaboraron con el Ministerio de Educación, y cada año elaboraban exámenes que medían las competencias de los estudiantes de la institución educativa con el fin de obtener información sobre el avance del nivel educativo. Sin embargo, debido a que las instituciones educativas no monitorean los factores que conducen al bajo rendimiento de los estudiantes, desconocen los desarrollos tecnológicos. Las técnicas de minería de datos son extremadamente útiles porque analizan cantidades masivas de datos, lo que les permite revelar un patrón previsto de todos los datos que se han recopilado. Cabe resaltar que la información adquirida será fundamental para identificar a los estudiantes que tienen dificultades con su aprendizaje y evaluar si se deben implementar o no medidas preventivas. Según la UNESCO, los hábitos alimentarios son un aspecto muy importante en el desarrollo de las capacidades cognitivas, y la mayoría de estos niños llegan a clase sin haber ingerido su alimento particular (desayuno), lo que se suma a un desequilibrio en su rendimiento académico. Por otro lado, está la intervención de los padres, que juegan un papel fundamental en la educación de sus hijos, y la comunicación entre ellos es fundamental. Los padres tienen un gran impacto en las experiencias educativas de sus hijos. En consecuencia, podemos afirmar que inciden en el grado de dedicación, desempeño y responsabilidad que demuestran los alumnos en sus clases. Una de las dificultades más desafiantes que enfrentan los instructores es la atención inadecuada de los padres tanto a sus hijos como al trabajo que se realiza en el aula. Esto complica el rendimiento académico de los niños ya que se les anima a creer que sus padres no están interesados en su educación, lo que afecta el estado emocional de los estudiantes y por ende su rendimiento académico.

Por lo tanto, se precisa que las herramientas tecnológicas son muy útiles y eficaces hoy en día, y que su correcta aplicación mejorará indiscutiblemente nuestra vida cotidiana; en consecuencia, el I.E. N° 64103 TENIENTE DIEGO desconoce la correcta aplicación que se hará de la minería de datos, por lo que no puede predecir cómo progresarán los alumnos según su nivel educativo.

II. JUSTIFICACIÓN DEL PROYECTO

La presente investigación se justifica considerando los siguientes criterios: Justificación teórica, porque las técnicas de minería de datos son importantes en cualquier rubro ya que mediante estas técnicas es posible hallar anomalías, patrones y correlaciones en conjuntos de datos extensos o grandes con el fin de poder predecir resultados. Por ende, al aplicar técnicas de minería de datos en el ámbito educativo, se hará mucho más óptimo el procesamiento de la información que se tiene almacenada; ya que permitirá conocer que alumnos presentan problemas relacionados a su aprendizaje y cuales son estos factores que influyen en su bajo rendimiento.

Justificación práctica, porque al aplicar técnicas de minería de datos en el ámbito educativo, coadyuvará a conocer de manera rápida, que factores son los que influyen negativamente en el aprendizaje del estudiante y de esta manera plantear estrategias para mejorar estas deficiencias, siendo soporte para que los estudiantes refuercen esos aspectos en los cuales manifiestan dificultad.

Justificación metodológica, la presente investigación utilizará la metodología KDD para el proceso de minería de datos, este proceso deberá ser asistido por un conocedor en la materia, con el fin de cumplir adecuadamente cada una de las fases de la metodología, asegurándose de aplicarlo al análisis predictivo, obteniendo de esto datos relevantes.

En la actualidad, residimos en un mundo globalizado y abundante en innovaciones tecnológicas; sin embargo, la mayoría de las veces no se le da el uso previsto a esta riqueza tecnológica, sino que se priorizan otras herramientas por encima de las que están diseñadas para solucionar los problemas que surgen en el día a día. Como consecuencia de este panorama, podemos comprender la dificultad de la región con su nivel educativo, que es negativo e insatisfactorio. Esto se debe a que la situación fue provocada por las circunstancias.

Este proyecto llevará a cabo una investigación de las muchas técnicas de minería de datos disponibles en la actualidad para extraer patrones específicos que se reflejan en nuestros cuatro tipos distintos de situaciones para explorar y, como resultado, proporcionar un resultado de predicción más

preciso sobre el rendimiento académico. De esta manera, el objetivo de esta investigación es realizar un examen de los diversos enfoques de minería de datos accesibles en la actualidad.

III. HIPOTESIS

3.1. Hipótesis General

La aplicación de técnicas de minería de datos, aplicadas al análisis predictivo determina el aprendizaje estudiantil de la I.E. N° 64103 Teniente Diego Ferre Sosa.

3.2. Hipótesis Específicos

- Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación institucional de la I.E. N° 64103 Teniente Diego Ferre Sosa.
- Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación social de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.
- Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación personal de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.
- Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación académica de los alumnos de la I.E. N° 64103 Teniente Diego Ferre Sosa.

IV. OBJETIVOS

4.1. Objetivo General

Aplicar técnicas de minería de datos, aplicadas al análisis predictivo para determinar el aprendizaje estudiantil de la I.E. N ° 64103 Teniente Diego Ferre Sosa.

4.2. Objetivos Específicos

- Estimar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación institucional de la I.E. N ° 64103 Teniente Diego Ferre Sosa.
- Definir la contribución de las técnicas de minería de datos, aplicadas al

análisis predictivo a la situación social de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.

- Precisar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación personal de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.
- Definir la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación académica de los alumnos de la I.E. N° 64103 Teniente Diego Ferre Sosa.

V. ANTECEDENTES

Para los antecedentes de estudio se tiene en primera instancia en el aspecto **internacional**, donde se consideró a Orozco et al., (2021) quienes han mejorado el nivel académico de los alumnos de la institución educativa considerada como caso de estudio. A lo que emplearon una metodología observacional – exploratorio, además, se ha tenido que considerar el uso de la inteligencia de negocios para la toma de decisiones, habiendo recolectado información históricamente de los estudiantes. Obteniendo como resultado que entre las técnicas de minería de datos se tiene al árbol de decisiones, pues es el que mejor precisión ha obtenido de la data procesada, siendo las precisiones las siguientes: MAE (0.41), MSE (0.38) y RMSE (0.89). Concluyendo así que, los datos obtenidos han permitido identificar los factores que inciden directamente en la aprobación final de los estudiantes en su periodo académico, siendo estos factores el de la educación de la madre y faltas graves. Calva et al., (2021) tuvieron como objetivo desarrollar un modelo de aprendizaje automático supervisado para predecir las posibilidades de que un estudiante de la Escuela Politécnica Nacional complete con éxito el curso obligatorio de nivelación. Se examinan elementos relacionados con las cualidades sociales, demográficas, familiares, institucionales y de rendimiento académico presentados en la solicitud, así como el curso de nivelación en el que se encuentra matriculado el estudiante, para explicar el potencial de admisión. El modelo de árbol de decisiones logró un nivel de precisión del 96 % y un área bajo la curva ROC de 89,1 en el conjunto de datos de entrenamiento, los cuales se consideran niveles de precisión generalmente aceptables. Además, una tasa de precisión del 96 % es similar al nivel aceptable típico de precisión. Los

resultados de la regresión logística, por su parte, muestran que factores como la nota ponderada del estudiante en los dos primeros meses, la nota con la que se presentó, su horario de estudio y su ubicación geográfica de origen, entre otros, están relacionados con la probabilidad de que el estudiante apruebe de alguna manera el curso de nivelación.

Hasan et al (2020), tuvieron como objetivo predecir el rendimiento general de los estudiantes al final del semestre utilizando técnicas de análisis de aprendizaje por vídeo y de minería de datos. Este estudio se llevó a cabo con 772 estudiantes matriculados en módulos de comercio electrónico y tecnologías de comercio electrónico en un IES. Se analizaron los datos del sistema de información de estudiantes, del sistema de gestión del aprendizaje y de las aplicaciones móviles utilizando ocho algoritmos de clasificación diferentes. Además, se llevaron a cabo técnicas de transformación y preprocesamiento de datos para reducir las características. Además, se llevaron a cabo la búsqueda genética y el análisis de componentes principales para reducir aún más las características. Además, el inductor de reglas CN2 y la proyección multivariante pueden utilizarse para ayudar al profesorado a interpretar las reglas para obtener información sobre las interacciones de los estudiantes. Los resultados mostraron que Random Forest predijo con precisión a los estudiantes que habían tenido éxito al final de la clase con una exactitud del 88,3% con una relación de amplitud y ganancia de información iguales.

Namoun & Alshanqiti (2020), tuvieron como objetivo examinar una década de trabajos de investigación realizados entre 2010 y noviembre de 2020 para presentar un conocimiento fundamental de las técnicas inteligentes utilizadas para la predicción del rendimiento de los estudiantes. Sintetizaron y analizaron un total de 62 artículos relevantes centrándose en tres perspectivas, (1) las formas en que se predicen los resultados del aprendizaje, (2) los modelos de análisis predictivo, desarrollados para predecir el aprendizaje de los estudiantes, y (3) los factores dominantes que impactan en los resultados de los estudiantes. Se aplicaron las prácticas PICO y PRISMA, para sintetizar e informar de los principales resultados. El logro de los resultados del aprendizaje se midió principalmente como posiciones de la clase de rendimiento (es decir, clasificaciones) y puntuaciones de rendimiento (es decir, calificaciones). Se emplearon con frecuencia modelos de regresión y de aprendizaje automático

supervisado para clasificar el rendimiento de los estudiantes. Por último, las actividades de aprendizaje en línea de los estudiantes, las calificaciones de las evaluaciones trimestrales y las emociones académicas de los estudiantes fueron los predictores más evidentes de los resultados del aprendizaje. Concluyen el estudio destacando algunos de los principales retos de la investigación y sugiriendo un resumen de recomendaciones significativas para motivar futuros trabajos en este campo.

Doleck et al (2020), tuvieron como objetivo evaluar el grado en que los marcos y bibliotecas de aprendizaje profundo (deep learning) conocidos, como Keras, Theano, Tensorflow, fast.ai y Pytorch, son predictores precisos de los resultados futuros. Según los resultados de los experimentos, el nivel de rendimiento, medido por la precisión predicha, difiere según el optimizador que se haya utilizado. Además, los resultados de las pruebas posteriores realizadas modificando la configuración de la red producen los mismos resultados. Llegaron a la conclusión de que el rendimiento del aprendizaje profundo está a la par con el de otras técnicas de aprendizaje automático, incluidas las máquinas de vectores de apoyo, los vecinos más cercanos a k, el clasificador Bayes ingenuo y la regresión logística.

Khan & Ghosh (2020), tuvieron como objetivo realizar un análisis exhaustivo de la investigación de EDM sobre los efectos del aprendizaje en el aula en el rendimiento de los alumnos. Primeramente, identificaron algunas cuestiones de investigación importantes y en curso relacionadas con este campo. De acuerdo con las conclusiones del meta-análisis de los elementos que los investigadores identificaron como que tienen un efecto sobre el rendimiento. Además, se examinan las investigaciones anteriores sobre la previsión del rendimiento de los alumnos, prestando especial atención al objetivo de cada estudio, así como a la oportunidad de sus predicciones. Existe una gran cantidad de investigaciones disponibles sobre la predicción del éxito de los estudiantes después del inicio del curso. Sin embargo, la predicción temprana antes del inicio de la carrera sigue siendo un difícil problema abierto. Además, se ha creado una taxonomía de posibles vías de investigación. No cabe duda de que este estudio en profundidad será de gran ayuda para los investigadores de la EDM a la hora de avanzar en el área de la predicción del grado en los próximos años. Esta predicción avanzada también puede facilitar el desarrollo de

diversos sistemas educativos inteligentes basados en ciencias informáticas.

Fernandes et al (2019), tuvieron como objetivo realizar un estudio predictivo del rendimiento académico de los niños que asisten a las escuelas públicas del Distrito Federal de Brasil a lo largo de los períodos escolares de 2015 y 2016 con el fin de evaluar el rendimiento académico. Se realizó un análisis estadístico descriptivo de los datos para obtener información. El primer conjunto de datos comprendía variables que se recogieron antes del inicio del año académico, y el segundo conjunto de datos incluía factores académicos que se recogieron dos meses después del inicio del semestre. Para cada conjunto de datos, se desarrollaron modelos de clasificación basados en el algoritmo Gradient Boosting Machine (GBM) con el fin de proporcionar predicciones sobre los resultados académicos del rendimiento de los estudiantes al finalizar el año escolar. Según los resultados, las características demográficas de "barrio", "escuela" y "edad" también son indicadores potenciales del éxito o el fracaso académico de un estudiante. Aunque los atributos "calificaciones" y "ausencias" fueron los más relevantes para predecir los resultados académicos del rendimiento de los alumnos al final del año, los resultados mostraron que las "ausencias" y las "calificaciones" fueron las más relevantes para predecir los resultados académicos del rendimiento de los alumnos al principio del año siguiente.

Aldowah et al (2019), tuvieron como objetivo realizar una revisión exhaustiva de la minería de datos educativos (EDM) y la analítica del aprendizaje (LA) en la educación superior. Esta revisión abarcó los estudios más relevantes relacionados con cuatro dimensiones principales: análisis de aprendizaje asistido por ordenador (CSLA), análisis predictivo asistido por ordenador (CSPA), análisis de comportamiento asistido por ordenador (CSBA) y análisis de visualización asistido por ordenador (CSVA) desde el año 2000 hasta 2017. Se identificaron y compararon las técnicas pertinentes de GED y AL en estas dimensiones. Sobre la base de los resultados de 402 estudios, se encontró que las técnicas específicas de EDM y LA podrían ofrecer los mejores medios para resolver ciertos problemas de aprendizaje. La aplicación de la GED y la AL en la educación superior puede ser útil para desarrollar una estrategia centrada en el estudiante y proporcionar las herramientas necesarias que las instituciones podrán utilizar con fines de mejora continua.

Kiu (2018), tuvo como objetivo realizar un análisis con el fin de determinar la relevancia e influencia de los antecedentes de los estudiantes, las actividades sociales de los estudiantes y el rendimiento académico de los estudiantes para predecir el éxito académico. Con el fin de predecir el rendimiento de los estudiantes de secundaria en matemáticas, se utilizaron métodos de minería de datos educativos supervisados como el bayesiano ingenuo, el perceptrón multicapa, el árbol de decisión J48 y el Random Forest. Se utilizó una clasificación de 2 niveles y una clasificación de 5 niveles sobre la nota final para hacer la previsión. Los resultados de este experimento indican que los antecedentes de los estudiantes y las actividades en las que participan socialmente son factores importantes para predecir el rendimiento de los estudiantes en una clasificación de dos niveles. El modelo puede utilizarse para hacer predicciones tempranas sobre el rendimiento de los estudiantes en un tema y también puede ayudar a los estudiantes a mejorar su rendimiento en esa asignatura.

Campos et al., (2018) quienes han analizado la comprensión y predicción acerca del rendimiento estudiantil que poseen los alumnos en un área en específico. Para ello, recolectaron información del periodo 2014-2015 y 2015-2016 para poder determinar a la población, siendo este conformado por un total de 124 estudiantes, a quienes posteriormente se les aplicó un cuestionario conformado por un total de 30 preguntas. Teniendo como resultados que, se puede realizar una predicción eficiente siempre y cuando se obtenga información disponible, pues se precisó que se pudo obtener información mucho más precisa con un 70% de acierto, dicha información puede aportar en brindar las motivaciones pertinentes a cada uno de los estudiantes con el fin de que mejoren su rendimiento en sus asignaturas. Concluyendo así que, la minería de datos sirve efectivamente para predecir el estado en el que se encuentran los estudiantes respecto a su rendimiento académico, en el que se identificarán los factores deficientes para subsanarlos.

En cuanto al aspecto **nacional**, se consideró a Ramírez et al., (2022) quienes tuvieron como finalidad integrar técnicas de minería de datos para caracterizar estudiantes de un programa educativo de la Universidad Autónoma del Estado de Hidalgo que pueden afrontar dificultades originadas por la rápida adaptación a clases virtuales, entre otras medidas que se adoptaron para evitar propagar

la COVID-19. Desarrollando una metodología de enfoque cuantitativo y cualitativa, para aplicar las técnicas de MD se efectuó un sondeo a los estudiantes para obtener datos respecto a las condiciones sociodemográficas, económicas, salud mental y académicas que posibilitan hallar patrones que repercuten en el desempeño académico de los estudiantes. En los resultados que se obtuvieron se puede constatar que, existen 2 grupos (clústeres), uno de ellos con condiciones mejores que otro, además una lista de ítems que influyen de manera negativa como positiva a que el estudiante se vea afectado en su rendimiento escolar.

Tapia (2021), tuvo como objetivo construir un modelo predictivo basado en modelos de conjunto con el objetivo de descubrir y categorizar a los estudiantes universitarios que probablemente abandonarán la escuela a la que asisten actualmente. Se adoptó el enfoque CRISP-DM, específicamente para tratar con técnicas de procesamiento de datos desequilibradas como Re muestreo y variables ficticias, entre otras. Esto se hizo empleando la estrategia CRISP-DM. Se utilizaron algoritmos basados en enfoques supervisados durante el procedimiento de clasificación de datos. Estos algoritmos comprendían clasificadores bayesianos, árboles de decisión y métodos de ensamblaje más complejos como Random Forest. Como parte del procedimiento de validación, los resultados de las clasificaciones producidas a partir de los datos de la prueba se compararon con los grupos reales en los que se ubicaron los estudiantes. Esta revisión incluyó metodologías como la validación cruzada y consideró métricas como el área bajo la curva ROC, el puntaje F1, exactitud, sensibilidad, especificidad y precisión, entre otras. Se determinó que, al usar el conjunto de datos actual y aplicar el aprendizaje automático, el clasificador basado en bosques aleatorios superó a los demás presentados. Además, dado el conjunto de datos actual, fue posible generar pronósticos precisos sobre los niños que probablemente abandonarían la escuela. Esto permitió pronósticos confiables sobre los alumnos que probablemente abandonarían la escuela.

Quiñones y Carrasco (2020) en su investigación tuvieron como objeto predecir el rendimiento académico de los alumnos de la profesión de Ingeniera de Industrias Alimentarias, por lo cual se optó por emplear técnicas de minería de datos. Considerando 142 estudiantes de la profesión de Ingeniería de Industrias Alimentarias del II y X ciclo hasta el periodo de noviembre del año 2019,

realizándose una base de datos la cual fue conformada por 142 instancias y 27 variables; con la finalidad de generar y validar modelos basados en minería de datos que hicieron posible determinar el rendimiento académico. Asimismo, la metodología que se utilizó fue CRISP-DM. Obteniendo como resultado que, mediante la metodología utilizada se pudo hallar que los pronósticos de 3 algoritmos de clasificación realizados en el programa Weka pudieron obtener una confiabilidad superior al 83%.

Garcia (2021), tuvo como objetivo desarrollar un método de predicción del rendimiento académico que se basara en diferentes elementos que influyen en el rendimiento académico y que ofreciera información sobre el rendimiento académico de los estudiantes. En primer lugar, se utilizó el método CRISP-DM de minería de datos para extraer información de los datos de los alumnos y crear el algoritmo fundamental del sistema de predicción. Para ello se utilizó el método CRISP-DM (Random Forest o árbol de decisiones). Luego se integró en un sistema que cubre el proceso de autorización de los apoderados de los estudiantes y el muestreo de los informes para las predicciones que se hicieron. Este sistema se llevó a cabo utilizando la metodología RUP. A continuación, se automatizó el modelo para que los datos de los alumnos se introdujeran a través de una interfaz y devolvieran una predicción en tiempo real. Se determinó que Random Forest era el algoritmo que producía la medida más ideal de asertividad con una puntuación de 0,95; este algoritmo fue elegido para servir de base para la automatización del modelo. También se determinó el grado de importancia de cada variable (factor de impacto) en el modelo, y se evaluó la asociación que se encontró entre cada componente psicológico y el rendimiento académico mediante la escala de Pearson.

Rondo (2020), tuvo como objetivo proponer el uso de la minería de datos en el Colegio Nacional de la Libertad de la ciudad de Huaraz. Se presentó un enfoque de investigación cuantitativo descriptivo con un diseño transversal que no incluyó ningún experimento para los fines de este análisis. El número de estudiantes de quinto año de secundaria incluidos en la muestra se redujo a 43. El siguiente es un resumen abreviado de lo que descubrimos: Para la primera dimensión, que se centró en el nivel de satisfacción con el procedimiento existente de reconocimiento de posibles casos de riesgo, el 81% de los encuestados estaba satisfecho con la forma en que se llevaron a cabo las

actividades. Para evaluar esto se utilizó el nivel de satisfacción con el método existente. cosas; en cuanto a la segunda dimensión, se observó que el cien por ciento de los encuestados apoya el uso de la minería de datos para identificar posibles instancias de peligro. Según las conclusiones de estos dos elementos, el 81% de los encuestados está satisfecho con la forma en que van las cosas y el 100% siente que se está realizando la minería de datos. Concluye que la minería de datos propuesta es necesaria si se quieren descubrir los patrones de comportamiento asociados con la deserción escolar.

Holgado (2018), tuvo como objetivo determinar los factores que contribuyen al mediocre desempeño de los alumnos respecto al aprendizaje que pertenecen a una casa de estudio. Por lo que, se desarrolló un modelo CRISP-DM, correspondiente a una técnica de minería de datos. Se pudo determinar, con la ayuda del script Random Forest, a fin de establecer la incidencia de las variables en la predicción del desempeño estudiantil. Este fue el caso de todos estos factores. El algoritmo de clasificación C5.0 fue el que arrojó mayor resultado para el modelo de estudiantes que presentaban bajo rendimiento académico. Se encontró que la medida de precisión de clasificación para C5.0 tenía una precisión del 77,8 % y su coeficiente kappa era de 0,56. Este método particular logró los mejores resultados cuando se aplicó al modelo de clasificación. La precisión del C5.0 fue de 77.8%, y su valor kappa fue Después de aplicar los scripts CART y C5.0, se estipularon un perfil estudiantil con deficiente desempeño académico, siendo esto referido a alumnos que pasaban satisfactoriamente más de seis cursos, pero pasaban un monto menor a los 62, siendo, además, los que presentan deuda con la casa de estudios a la que pertenecen.

Yamao (2018) tuvo como objetivo pronosticar el rendimiento académico de los alumnos de nuevo ingreso en la Escuela Profesional de Ingeniería Informática y de Sistemas. Para ello, se diseñó metodológicamente una investigación de enfoque cuantitativo, consistente en un diseño correlacional, causal y explicativo, siendo la población considerada 1304 estudiantes. Teniendo en cuenta diversos factores (como la media, la situación socioeconómica y familiar, el modo de ingreso, el lugar de origen, etc.), se realizan predicciones mediante técnicas de regresión lineal, técnicas de árboles de decisión y técnicas de máquinas de vectores de apoyo. La nota del examen, el sexo, la edad, el modo

de ingreso y la distancia del domicilio al centro de estudios fueron los criterios que demostraron tener un impacto más significativo en el rendimiento académico, según estos resultados. La técnica del árbol de decisión fue la que obtuvo los mejores resultados, con una tasa de precisión del 82,87%. Se pudo concluir que la aplicación de la minería de datos permitió crear predicciones sobre el rendimiento académico de los estudiantes que ingresan. Esto permitió identificar a los estudiantes que tienen más probabilidades de tener problemas con sus estudios.

Alania (2018), tuvo como objetivo utilizar métodos de minería de datos para hacer proyecciones sobre el número de estudiantes que abandonarían el programa académico. Con la ayuda de esta predicción, las autoridades podrán determinar la proyección del tamaño de las secciones y llevar a cabo otras acciones apropiadas, así como identificar a los estudiantes que abandonan la escuela. El método del árbol de decisión C4.5 se utiliza para analizar los datos de las calificaciones del último semestre de un estudiante con el fin de determinar si éste se retirará o no de la escuela en el siguiente semestre (J48). Con el uso del resultado del árbol de decisión, pudimos hacer una conjetura sobre el número de alumnos que tienen más probabilidades de abandonar la Universidad. El resultado puede ser utilizado por las autoridades para que puedan tomar medidas para mejorar la toma de decisiones. Una vez finalizado el análisis con el conjunto de datos primarios, se introduce en el sistema un conjunto de datos de prueba para poder examinar los resultados. Se muestra que el mayor número de alumnos que abandonan los estudios se debe a las bajas calificaciones. Según las conclusiones de la investigación, la predicción también ha contribuido a determinar mejor la asociación entre las dos cualidades, "Cursos aprobados" y "Abandono", demuestra que la mayor proporción de abandonos es la de aquellos que han recibido menos de 24 notas de aprobación a lo largo de su carrera académica.

Mollo (2018), tuvo como objetivo utilizar el Data Warehouse y la minería de datos para realizar un análisis predictivo de las tasas de deserción estudiantil en la Universidad Nacional Jorge Basadre Grohmann desde el año 2012 al 2018. Para lograr este objetivo, se ha construido un Data Warehouse para poder realizar un análisis de los datos referentes a la deserción estudiantil utilizando los métodos creados por Ralph Kimball. Los resultados de esta

investigación arrojan luz sobre los indicios de abandono estudiantil basadrina que se producen con mayor frecuencia. Para aplicar las técnicas de minería de datos, se escogieron tres de ellas, y la estrategia que se eligió fue el árbol de decisión por su precisión en la predicción de datos y la sencillez con la que se pueden entender sus hallazgos. Se logró crear una investigación predictiva de la deserción estudiantil utilizando Data Warehouse y minería de datos bajo metodologías que permiten el registro de información histórica y la interpretación de sus resultados predichos. Esta es la conclusión general que se puede extraer de los hallazgos del estudio. Esto fue preciso en la mayoría de las situaciones.

Menacho (2017) tuvo como objetivo utilizar la información académica de los estudiantes matriculados en el curso de Estadística General de la UNALM. Específicamente, para ello se utilizó la TMD de regresión logística, redes neuronales y bayesianas, y árboles de decisión. Para estimar el estado final (aprobado o reprobado) de los estudiantes que se inscribirían en el curso en cuestión, se tomaron en cuenta 914 estudiantes que se habían inscrito entre los periodos 2013 II y 2014 I. Se aplicaron las métricas de la matriz de confusión para validar y seleccionar el modelo ideal para categorizar a los nuevos estudiantes; esto ayudará a los profesores a identificar a los estudiantes que tienen desafíos académicos para que puedan darles una asesoría oportuna y efectiva. Cuando se utilizó la red Naive Bayes para clasificar los datos, el resultado final fue una tasa de clasificación del 71,0%. Esta tasa se calculó utilizando cada una de las cuatro técnicas en relación con la precisión de cada clase, con un mayor porcentaje de clasificación correcta para la clase de Aprobado y un menor porcentaje para la clase de Reprobado.

VI. MARCO TEÓRICO

Técnicas de minería de datos

El proceso de extraer nueva información de una gran base de datos existente se conoce como "minería de datos". La minería de datos se define como la técnica de revisar una enorme base de datos existente. Una tecnología poderosa tiene la capacidad de ayudar a las empresas a concentrar su atención en los datos en sus almacenes de datos almacenados que son de suma importancia. Al ser más proactivo y tomar decisiones más basadas en el

conocimiento, la empresa podrá prever tendencias futuras utilizando las herramientas y técnicas de minería de datos. El uso de técnicas de minería de datos podría brindar respuestas a problemas relacionados con la empresa que, en el pasado, eran difíciles y requerían mucho tiempo para abordar. El propósito de este estudio es proporcionar una visión general de los procedimientos de minería de datos. La agrupación en clústeres, la decisión, la predicción de árboles y las redes neuronales son algunos de los enfoques que serán el énfasis principal de esta discusión (León et al., 2019).

Es una serie de técnicas específicas para identificar patrones implícitos que pertenecen a grandes grupos de datos, con la finalidad de deducir sus mecanismos de comportamiento, interrelación y potencial evolución a futuro (Datahack, 2021).

Dimensiones de técnicas de minería de datos

Selección de datos:

Tiene como objetivo elegir los datos que deben almacenarse durante la recopilación de datos o que deben compartirse – archivarse una vez finalizado el proyecto.

Preprocesamiento:

El preprocesamiento de datos es un paso importante en el descubrimiento de información. Esta etapa está a cargo de la limpieza, integración, transformación y reducción de datos en preparación para la siguiente etapa de minería de datos (García et al., 2016).

Data Mining:

Es una técnica automatizada o semiautomatizada para analizar y sintetizar enormes cantidades de datos dispersos en conocimiento útil. Se encarga de buscar patrones, anomalías y correlaciones entre millones de registros, con el fin de pronosticar resultados (Choez et al., 2021).

Interpretación y evaluación:

El proceso de revisión, limpieza y cambio de datos para obtener información que luego se puede utilizar para sacar conclusiones particulares que pueden ayudar en la toma de decisiones y proporcionar información se conoce como interpretación de datos. Esto se hace con el fin de extraer información de los datos que luego se pueden utilizar para crear conclusiones claras que ayuden con la toma de decisiones y la claridad. El objetivo es mejorar tanto la lucidez

como la capacidad de decisión (EUROINNOVA, 2022). Por otro lado, la evaluación es el proceso de identificación de datos relativos a un aspecto concreto que permita establecer criterios para evaluar en qué medida estos elementos alcanzan los fines y objetivos definidos (Yanes et al., 2021).

Aprendizaje estudiantil

Es el conocimiento que se adopta o adquiere mediante los hechos que se manifiestan en la vida y también mediante los conocimientos impartidos por un docente especialista de una determinada área; de tal modo la persona desarrolla nuevas habilidades y capacidades que enriquecen su conocimiento, también permite amoldarse al medio y responder frente a las modificaciones que se consigan dar a su alrededor (Carrillo & Garcia, 2020).

Dimensiones de aprendizaje estudiantil

Situación institucional:

Se refiere a conocer la capacidad y recursos con los que cuenta la institución para complementarlos con la calidad de enseñanza que se debe impartir en él.

Situación social:

Está relacionado con la posición de una persona en la sociedad en la que vive. En consecuencia, la idea se refiere a la posición de un tema con respecto al entorno o al contexto en el que existe.

Situación personal:

Son situaciones o acontecimientos en los que está expuesta la persona, tales como lidiar con un problema ya sea personal, familiar, salud, problemas económicos, entre otros factores.

Situación académica:

Se refiere al listado de las diversas situaciones académicas que presenta el alumno desde su primera etapa escolar y en el trayecto de las diferentes instancias académicas (Delgado, 2017).

VII. METODOLOGÍA

7.1. Lugar de estudio

El lugar de estudio se realizará en la Institución Educativa Nacional N° 64103 Teniente Diego Ferre Sosa, perteneciente a la ciudad Pucallpa, de

la provincia Coronel Portillo, Distrito Yarinacocha, del departamento de Ucayali.

7.2. Población y tamaño de muestra

Población.

Se refiere al conjunto conformado por individuos u objetos de la misma naturaleza las cuales están en un lugar y tiempo específico (Sánchez et al., 2018). Por ende, estará conformado por los alumnos del nivel primario de la I.E. N° 64103 Teniente Diego Ferre Sosa, quienes poseen registros de notas dentro del periodo 2018 – 2022.

Muestra

Hace referencia a un subgrupo el cual es seleccionado o extraído de la población de importancia de estudio del mismo que se pretende recopilar datos relevantes para su análisis y posterior solución (Baena, 2017). Por lo tanto, se consideró todo el nivel de 6°, conformado por 5 secciones, cada una de ellas con 30 alumnos, siendo un total de (150 alumnos).

7.3. Descripción detallada de los métodos, uso de materiales, equipos o insumos.

a) Diseño de muestreo

Corresponderá a un muestreo probabilístico, Otzen y Manterola (2017) refieren que la muestra será extraída de una población específica, de tal forma que cualquier individuo tendrá la probabilidad de ser partícipe de la muestra

b) Descripción detallada del uso de materiales, equipos, insumos, entre otros.

EQUIPOS Y BIENES DURADEROS
Laptop Acer Corei5
Impresora EPSON L355

ELEMENTOS DE OFICINA
Papel Bond A4 Alpha
Tinta para impresora EPSON L355
Libreta
Lapiceros

SERVICIOS DE ENERGÍA ELÉCTRICA
Energía eléctrica
SERVICIOS DE TELEFONÍA E INTERNET
Internet claro 80 Mbps
Telefonía móvil
SERVICIOS DE IMPRESIONES, ENCUADERNACIÓN Y EMPASTADO
Empastado

c) Descripción de variables a ser analizados en el objetivo específico

Para continuar con el procedimiento respecto al desarrollo del estudio, se dará a conocer a los directivos de la I.E. respecto al procedimiento que se empleara, se dará a conocer mediante el Consentimiento informado, el cual manifestara su autorización de manera voluntaria.

Estimar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación institucional de la I.E. N ° 64103 Teniente Diego Ferre Sosa.

Para el primer objetivo específico, se aplicará una encuesta mediante la cual se podrá recopilar información respecto a la situación institucional según el punto de vista de los estudiantes, de acuerdo con la extracción de predicciones a partir de datos históricos, eso conllevará al análisis predictivo, que se consigue mediante el uso de técnicas de modelado estadístico, grandes cantidades de datos y aprendizaje automático. Puesto que, para poder asegurar el correcto manejo y proyección del estado situacional institucional del centro educativo; seguidamente los datos serán procesados utilizando el software Weka; con el fin de determinar el grado en el que contribuye las técnicas de minería de datos una vez aplicado el análisis predictivo.

Definir la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación social de los padres

de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.

Para el segundo objetivo específico, se aplicará una encuesta mediante la cual se podrá recopilar información respecto a la situación social en la que viven los estudiantes; de la misma forma, para poder analizar de manera correcta los datos de acuerdo al estado situacional social, se deberán realizar los procesos por medio de la herramienta y Software Weka, y a través de un algoritmo adecuado, podremos determinar y dar a conocer como contribuye las técnicas de minería de datos.

Precisar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación personal de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.

Para obtener información sobre las circunstancias personales de los padres de los alumnos que asisten a la institución en cuestión, también se realizará una encuesta como parte del tercer objetivo específico que se cumplirá. de la misma manera, se deberá aplicar el análisis correspondiente a los datos obtenidos, de esta manera, a través del Software Weka, y con la ayuda de un algoritmo adecuado, se podrá evaluar la contribución de las técnicas de minería de datos aplicadas al análisis predictivo a las numerosas partes de la situación personal en el que se ven envueltos los alumnos diariamente.

Definir la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación académica de los alumnos de la I.E. N° 64103 Teniente Diego Ferre Sosa.

Para lograr el cuarto objetivo específico, se aplicó una encuesta para extraer información sobre las circunstancias académicas de los estudiantes. Para el manejo correcto de la información, se deberá tomar en cuenta que, para procesar la información, se hará uso de la herramienta y software Weka para así poder determinar la importancia de la contribución de las técnicas de minería de datos.

d) Aplicación de prueba estadística inferencial.

No se aplicará la estadística inferencial, puesto que se utilizará el programa Weka para el procesamiento de los datos.

673
674

7.4. Tabla de recolección de datos por objetivos específicos.

OBJETIVOS ESPECIFICOS	TECNICA DE RECOLECCIÓN DE DATOS
Estimar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación institucional de la I.E. N ° 64103 Teniente Diego Ferre Sosa.	Encuesta – Software Weka
Definir la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación social de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa.	
Precisar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación personal de los padres de familia de la I.E. N° 64103 Teniente Diego Ferre Sosa	
Definir la contribución de las técnicas de minería de datos aplicadas al análisis predictivo a la situación académica de los alumnos de la I.E. N° 64103 Teniente Diego Ferre Sosa	

675
676
677

VIII. CRONOGRAMA DE ACTIVIDADES

Actividad	2022							
	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
Generalidades del proyecto								
Realidad problemática								
Formulación del problema								
Antecedentes y justificación								
Marco teórico y conceptual								
Diseño, hipótesis y operacionalización de variables								
Población y muestra								

Técnicas e instrumentos de recolección de datos								
Revisión y ajustes finales del asesor de la investigación								
Presentación del proyecto a la FISyIC.								

VII. PRESUPUESTO

Descripción	Unidad de medida	Costo Unitario (S/.)	Cantidad	Costo total (S/.)
Laptop Acer Corei5	Unidad	2500.00	1	2500.00
Impresora EPSON L355	Unidad	700.00	1	700.00
Papel Bond A4 Alpha	Unidad	12.00	2	24.00
Tinta para impresora EPSON L355	Unidad	45.00	4	180.00
Libreta	Unidad	20.00	1	20.00
Lapiceros	Unidad	2.00	5	10.00
Energía eléctrica	Mensual	50.00	9	450.00
Internet claro 80 Mbps	Mensual	90.00	9	810.00
Telefonía móvil	Mensual	30.00	9	270.00
Empastado	Unidad	30.00	3	90.00
PRESUPUESTO TOTAL				5,984.00

VIII. BIBLIOGRAFÍA

- Alania, P. (2018). *Aplicación de técnicas de Minería de Datos para predecir la deserción estudiantil de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión* [Universidad Nacional Daniel Alcides Carrión]. http://repositorio.undac.edu.pe/bitstream/undac/829/1/T026_40573846_M.pdf
- Aldowah, H., Al-Samarraie, H., & Fauzy, W. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/https://doi.org/10.1016/j.tele.2019.01.007>
- Baena, G. (2017). *Metodología de la investigación* (Tercera ed). Grupo Editorial Patria. <https://cutt.ly/gKbErXN>
- Banoy, W. (2019). El uso pedagógico de las tecnologías de la información y la comunicación (TIC) y su influencia en el aprendizaje significativo de estudiantes de media técnica en Zipaquirá, Colombia. *Revista Academia y Virtualidad*, 12(2), 23–46. <https://doi.org/https://doi.org/10.18359/ravi.4007>
- Calva, K., Flores, M., Porras, H., & Cabezas, A. (2021). Academic performance prediction model for the propedeutic course of the Escuela Politécnica Nacional and the implementation of an automated supervised learning model. *Research Articles for the Regular Issue*, 8(2), 58–71. <https://doi.org/https://doi.org/10.5281/zenodo.5770905>
- Campos, J., Ramos, G., Morales, R., & Baena, M. (2018). Minería de datos educativos para la predicción personalizada del rendimiento académico. *Conferencia Internacional de Procesamiento de La Información*. <https://riuma.uma.es/xmlui/handle/10630/15477>
- Carrillo, S., & Garcia, D. (2020). *Clima Escolar y Logro de Aprendizaje de la Institución Educativa Mariscar Orbegoso, Provincia la Unión de Cotahuasi, Arequipa 2019*. [Tesis de pregrado, Universidad Nacional de San Agustín]. <http://hdl.handle.net/20.500.12773/11991>
- Choez, J., Figueroa, V., Barreto, J., Villacreses, C., & Marcillo, M. (2021). Data Mining como soporte para la Toma de Decisiones en el Entorno Empresarial. *Revista Científica Multidisciplinaria*, 5(4), 2602–8166. <https://revistas.unesum.edu.ec/index.php/unesumciencias/article/view/392>
- Datahack. (2021). *Las 7 técnicas de minería de datos más utilizadas en Big Data*. <https://www.datahack.es/tecnicas-mineria-datos/>

- 720 Delgado, J. (2017). *Situación académica desde la perspectiva de los estudiantes del*
721 *quinto grado de secundaria de las Instituciones Emblemáticas de la Ciudad de*
722 *Puno, año 2016* [Tesis de pregrado, Universidad Nacional del Altiplano].
723 <http://repositorio.unap.edu.pe/handle/UNAP/4463>
- 724 Díaz, B., Meleán, R., & Marín, W. (2021). Rendimiento académico de estudiantes en
725 educación superior: Predicciones de factores influyentes a partir de árboles de
726 decisión. *Revista de Estudios Interdisciplinarios En Ciencias Sociales*, 23(3), 616–
727 639. <https://dialnet.unirioja.es/servlet/articulo?codigo=8133982>
- 728 Doleck, T., Lemay, J., Basnet, B., & Bazelais, P. (2020). Predictive analytics in
729 education: a comparison of deep learning frameworks. *Educ Inf Technol*, 25,
730 1951–1963. <https://doi.org/https://doi.org/10.1007/s10639-019-10068-4>
- 731 EUROINNOVA. (2022). *Qué es la interpretación de datos*.
732 <https://www.euroinnova.pe/blog/que-es-la-interpretacion-de-datos>
- 733 Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G.
734 (2019). Educational data mining: Predictive analysis of academic performance of
735 public school students in the capital of Brazil. *Journal of Business Research*, 94,
736 335–343. <https://doi.org/https://doi.org/10.1016/j.jbusres.2018.02.012>
- 737 Garcia, L. (2021). *Sistema predictivo de rendimiento académico en base a factores*
738 *influyentes en estudiantes del 1° Secundaria en un colegio de Lambayeque*
739 [Universidad Católica Santo Toribio de Mogrovejo].
740 [https://tesis.usat.edu.pe/bitstream/20.500.12423/4723/1/TL_GarciaPeredoLuis.p](https://tesis.usat.edu.pe/bitstream/20.500.12423/4723/1/TL_GarciaPeredoLuis.pdf)
741 [df](https://tesis.usat.edu.pe/bitstream/20.500.12423/4723/1/TL_GarciaPeredoLuis.pdf)
- 742 García, S., Ramírez, S., & Fancisco, J. (2016). Big Data: Preprocesamiento y calidad
743 de datos. *Novática*, 237.
744 [https://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133_Nv237-Digital-](https://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133_Nv237-Digital-sramirez.pdf)
745 [sramirez.pdf](https://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133_Nv237-Digital-sramirez.pdf)
- 746 Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, U., & Sattar, U. (2020).
747 Predicting student performance in higher educational institutions using video
748 learning analytics and data mining techniques. *Applied Sciences*, 10(11), 38–94.
749 <https://doi.org/https://doi.org/10.3390/app10113894>
- 750 Holgado, L. (2018). *Detección de patrones de bajo rendimiento académico mediante*
751 *Técnicas de Minería de Datos de los estudiantes de la Universidad Nacional*
752 *Amazónica de Madre de Dios* [Universidad Nacional del Altiplano].
753 <http://repositorio.unap.edu.pe/handle/UNAP/9815>


- 754 Khan, A., & Ghosh, S. (2020). Student performance analysis and prediction in classroom
755 learning: A review of educational data mining studies. *Education and Information*
756 *Technologies*, 26, 205–240. <https://doi.org/10.1007/s10639-020-10230-3>
- 757 Kiu, C. (2018). Data mining analysis on student's academic performance through
758 exploration of student's background and social activities. *Fourth International*
759 *Conference on Advances in Computing, Communication & Automation (ICACCA)*,
760 1–5. <https://doi.org/10.1109/ICACCAF.2018.8776809>
- 761 León, E., Proaño, E., Muirragui, V., & Cajamarca, J. (2019). Minería de datos en el
762 análisis de tendencias políticas en redes sociales. *Ciencia Digital*, 3(4), 91–103.
763 <https://cienciadigital.org/revistacienciadigital2/index.php/CienciaDigital/article/view/837/2023>
- 764
- 765 Menacho, C. (2017). Predicción del rendimiento académico aplicando técnicas de
766 minería de datos. *Anales Científicos*, 7(1), 26–33.
767 <https://revistas.lamolina.edu.pe/index.php/acu/article/view/811>
- 768 Mollo, N. (2018). *Análisis predictivo de la deserción estudiantil utilizando data*
769 *warehouse y minería de datos en la Universidad Nacional Jorge Basadre*
770 *Grohmann – Tacna, 2012-2018* [Universidad Nacional Jorge Basadre Grohmann-
771 Tacna].
772 [http://tesis.unjbg.edu.pe/bitstream/handle/UNJBG/3506/185_2018_mollo_condor](http://tesis.unjbg.edu.pe/bitstream/handle/UNJBG/3506/185_2018_mollo_condor_i_nap_espg_ingenieria_sistemas.pdf?sequence=1&isAllowed=y)
773 [i_nap_espg_ingenieria_sistemas.pdf?sequence=1&isAllowed=y](http://tesis.unjbg.edu.pe/bitstream/handle/UNJBG/3506/185_2018_mollo_condor_i_nap_espg_ingenieria_sistemas.pdf?sequence=1&isAllowed=y)
- 774 Moreno, J., & Aguilar, F. (2019). Fundamentos ontológicos del sistema educativo
775 finlandés como referente para superar problemáticas en contextos emergentes.
776 *Sophia, Colección de Filosofía de La Educación*, 27.
777 <https://doi.org/https://doi.org/10.17163/soph.n27.2019.08>
- 778 Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining
779 and learning analytics techniques: A systematic literature review. *Applied*
780 *Sciences*, 11(1), 237. <https://doi.org/https://doi.org/10.3390/app11010237>
- 781 Orozco, W., Villao, A., Orozco, J., & Villarroel, M. (2021). Aplicación de técnicas de
782 minería de datos para predecir el desempeño académico de los estudiantes de la
783 escuela 'Lic. Angélica Villón L. *Revista Científica y Tecnología UPSE*, 8(2), 68–
784 75. <https://doi.org/https://doi.org/10.26423/rctu.v8i2.637>
- 785 Otzen, T., & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a
786 Estudio. *Int. J. Morphol.*, 35(1), 227–232.
787 <https://scielo.conicyt.cl/pdf/ijmorphol/v35n1/art37.pdf>

- Quiñones, L., & Carrasco, Y. (2020). Rendimiento académico empleando minería de datos. *Revista Espacios*, 41(44). <https://unj.edu.pe/wp-content/uploads/2021/09/a20v41n44p17.pdf>
- Ramírez, L., Delgado, E., & Montúfar, M. (2022). Aplicación de técnicas de minería de datos para la caracterización de estudiantes bajo el efecto de la COVID-19. *Pädi Boletín Científico De Ciencias Básicas e Ingenierías Del ICBI*, 10(2), 75–81. <https://doi.org/https://doi.org/10.29057/icbi.v10iEspecial2.8669>
- Rondo, E. (2020). *Propuesta de implementación de minería de datos para el colegio nacional de La Libertad-Huaraz* [Universidad Católica los Ángeles de Chimbote]. <http://repositorio.uladech.edu.pe/handle/20.500.13032/18452>
- Sánchez, H., Reyes, C., & Mejía, K. (2018). *Manual de términos en investigación, tecnológica y humanística* [Universidad Ricardo Palma]. <https://cutt.ly/gLdZJAY>
- Tapia, J. (2021). *Modelo predictivo de clasificación basado en aprendizaje automatizado para la detección temprana de posibles estudiantes universitarios desertantes* [Universidad Nacional de San Agustín de Arequipa]. <http://repositorio.unsa.edu.pe/handle/20.500.12773/13409>
- Villao, A. (2021). *Aplicación de técnicas de minería de datos para predecir el desempeño académico de los estudiantes de la escuela Lic. Angélica Villón L* [Tesis de pregrado, Universidad Estatal Península de Santa Elena]. <https://repositorio.upse.edu.ec/handle/46000/6493>
- Yamamoto, E. (2018). *Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de las Escuela Profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres, Lima-Perú* [Tesis de posgrado, Universidad de San Martín de Porres]. <https://hdl.handle.net/20.500.12727/3555>
- Yanes, J., Sepúlveda, R., & Díaz, H. (2021). La evaluación de la calidad de datos: una aproximación criptográfica. *Computación y Sistemas*, 23(2), 557–568. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-55462019000200557



818
819

IX. ANEXO.

FORMULACIÓN DEL PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES, DIMENSIONES E INDICADORES	METODOLOGÍA DE LA INVESTIGACIÓN
¿De qué manera las técnicas de minería de datos, aplicadas al análisis predictivo determina el aprendizaje estudiantil de la I.E. N° 64103 teniente Diego Ferre Sosa?	Aplicar técnicas de minería de datos, aplicadas al análisis predictivo para determinar el aprendizaje estudiantil de la I.E. N° 64103 teniente Diego Ferre Sosa.	La aplicación de técnicas de minería de datos, aplicadas al análisis predictivo determina el aprendizaje estudiantil de la I.E. N°64103 teniente Diego Ferre Sosa	VARIABLE 1: TÉCNICAS DE MINERÍA DE DATOS APLICADAS AL ANÁLISIS PREDICTIVO DIMENSIÓN: SELECCIÓN DE DATOS <ul style="list-style-type: none">Analizar y Extraer fuentes de datos. DIMENSIÓN: PREPROCESAMIENTO <ul style="list-style-type: none">Preparación y limpieza de los datos extraídos. DIMENSIÓN: DATA MINING <ul style="list-style-type: none">Extracción de patrones. DIMENSIÓN: INTERPRETACIÓN Y EVALUACIÓN <ul style="list-style-type: none">Identificar y evaluar los patrones. VARIABLE 2: APRENDIZAJE ESTUDIANTIL DIMENSIÓN: SITUACIÓN INSTITUCIONAL <ul style="list-style-type: none">Logros de aprendizaje.Acciones para mejorar el rendimiento.Capacitación a los docentes.Docentes de la I.E.Invitación a participar. DIMENSIÓN: SITUACIÓN SOCIAL <ul style="list-style-type: none">Normas de convivencia.Fomenta valores.Afición a la lectura.Servicio de internet.Número de hijos.Tipo de vivienda. DIMENSIÓN: SITUACIÓN PERSONAL <ul style="list-style-type: none">Hábito a la lectura.Motivación para cumplir las tareas.Preparación de los alimentos.Servicio de Salud.Comunicación entre sus hijos. DIMENSIÓN: SITUACIÓN ACADÉMICA <ul style="list-style-type: none">Notas de Matemática, Comunicación y Ciencia y ambiente.	ENFOQUE DE INVESTIGACIÓN: Cuantitativo TIPO DE INVESTIGACIÓN: Aplicada NIVEL DE INVESTIGACIÓN: Descriptiva – Explicativo DISEÑO DE LA INVESTIGACIÓN: Pre experimental ESQUEMA  Donde: <ul style="list-style-type: none">X = Técnicas de minería de datos, aplicadas al análisis predictivo.O = Observación de la variable independiente (Predicción del aprendizaje estudiantil en la I.E. N° 64103 teniente Diego Ferre Sosa). POBLACIÓN: Está conformado por los alumnos del nivel primario de la I.E. N° 64103 teniente Diego Ferre Sosa MUESTRA: Se encuentra conformado por todo el nivel de 6 TO grado, conformado por 5 secciones, cada uno de ellas con 30 alumnos, siendo un total de (150 alumnos).
<ul style="list-style-type: none">¿De qué manera contribuye las técnicas de minería de datos, aplicadas al análisis predictivo a la situación institucional de la I.E. N° 64103 teniente Diego Ferre Sosa?¿De qué manera contribuye las técnicas de minería de datos, aplicadas al análisis predictivo a la situación social de los padres de familia de la I.E. N° 64103 teniente Diego Ferre Sosa?¿De qué manera contribuye las técnicas de minería de datos, aplicadas al análisis predictivo a la situación personal de los padres de familia de la I.E. N° 64103 teniente Diego Ferre Sosa?¿De qué manera contribuye las técnicas de minería de datos, aplicadas al análisis predictivo a la situación académica de los alumnos de la I.E. N° 64103 Teniente Diego Ferre Sosa?	<ul style="list-style-type: none">Estimar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación institucional de la I.E. N ° 64103 teniente Diego Ferre Sosa.Definir la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación social de los padres de familia de la I.E. N° 64103 teniente Diego Ferre Sosa.Precisar la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación personal de los padres de familia de la I.E. N° 64103 teniente Diego Ferre Sosa.Definir la contribución de las técnicas de minería de datos, aplicadas al análisis predictivo a la situación académica de los alumnos de la I.E. N° 64103 Teniente Diego Ferre Sosa.	<ul style="list-style-type: none">Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación institucional de la I.E. N° 64103 teniente Diego Ferre Sosa.Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación social de los padres de familia de la I.E. N° 64103 teniente Diego Ferre Sosa.Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación personal de los padres de familia de la I.E. N ° 64103 teniente Diego Ferre Sosa.Las técnicas de minería de datos, aplicadas al análisis predictivo contribuyen significativamente a la situación académica de los alumnos de la I.E. N° 64103 teniente Diego Ferre Sosa.		

820