

Identificación del nivel de aprendizaje a nivel superior durante la pandemia por COVID-19 con minería de datos.

Luz Angela Medina Muñoz, Javier Romero Torres, Doricela Gutiérrez Cruz.

^a Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N Estado de México, Méx., acuario1402luz@hotmail.com, Estado de México, México.

^b Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N Estado de México, Méx., jromero@uaemex.mx, Estado de México, México.

^c Centro Universitario UAEM Nezahualcóyotl, Av. Bordo de Xochiaca S/N Estado de México, Méx., dgutierrezcr@uaemex.mx, Estado de México, México.

Resumen

Se analizó el nivel de aprendizaje académico a nivel superior en época de pandemia por COVID-19, mediante la minería de datos. Se analizaron 1600 datos de 64 estudiantes de la carrera de Ingeniería en Sistemas Inteligentes del Centro Universitario UAEM Nezahualcóyotl, las variables que se evaluaron fueron edad, situación laboral, dispositivos tecnológicos, el dispositivo principal es de uso compartido o propio, la cantidad de hermanos que tienen y la acreditación de materias de ambos ciclos escolares (2020 A-2020B y 2021A-2021B). Para la agrupación se utilizó el algoritmo SimpleKmeans y el algoritmo de clasificación J48 empleando el software WEKA. Se logró identificar que el nivel de aprendizaje en los estudiantes se vio afectado por el COVID-19 en ambos ciclos escolares (2020A-2020B y 2021A-2021B), mostrando que un 70.3% de estudiantes no acreditaron todas las materias en ciclo 2020A-2020B y un 73.4% en el ciclo 2021A-2021B, lo anterior debido a que no contaban con equipo propio para seguir con la modalidad a distancia y las plataformas empleadas para realizar las actividades no fueron lo suficientemente bien manejadas en el ciclo escolar 2020A-2020B.

Palabras clave—COVID-19, Nivel Académico, Minería de Datos, KDD, Kmeans.

Abstract

The academic learning level at a higher level in times of the COVID-19 pandemic was analyzed through data mining. 1600 data from 64 students of the Intelligent Systems Engineering career from to Centro Universitario UAEM Nezahualcoyotl were analyzed, variables like age, employment situation, technological devices, the main device is for shared or own use, the number of siblings they have and the accreditation of subjects of both school cycles (2020 A-2020B and 2021A-2021B) were evaluated. To analyze the data, SimpleKmeans algorithm and the J48 classification algorithm were using. In the result was possible to identify that the level of student learning was affected by COVID-19 in both school cycles (2020A-2020B and 2021A-2021B), it showing that 70.3% of students did not pass all subjects in the 2020A-2020B cycle. and 73.4% in the 2021A-2021B cycle, the above due to the fact that they did not have their own equipment to continue with the distance modality and the

platforms used to carry out the activities were not sufficiently well managed in the 2020A-2020B school cycle.

Keywords— Covid-19, Academic level, Data Mining, KDD, K.means.

1. INTRODUCCIÓN

A principios del año 2020, México se vio afectado por una pandemia ocasionada por una enfermedad infecciosa provocada por el virus de nombre SARS-CoV-2. (COVID-19), medir el impacto de dicha pandemia en distintos ámbitos ha sido un reto que enfrentan los países en el mundo, en específico en el sector educativo[1], representando un rezago importante en los distintos niveles educativos, transformando usos y estilos de vida para luchar contra la enfermedad en un intento por contener la propagación de la COVID-19 las autoridades educativas se vieron obligadas a implementar medidas para evitar el contagio en estudiantes y docentes, estableciendo el cierre temporal de instituciones a nivel nacional.[2][3].

Lo anterior condujo a la implementación de clases en línea a través del uso de diversos formatos y plataformas, en donde tanto alumnos como docentes desafiaron su capacidad de adaptación a escenarios completamente virtuales, iniciando una transformación digital repentina y dramática en la sociedad. [4]

La modalidad de educación a distancia tuvo un efecto negativo en el aprendizaje de estudiantes, afectando principalmente a los de entornos menos favorecidos económicamente o con la poca capacidad de habilidades digitales tanto de alumnos como de maestros. [2]

Al ser el COVID-19 un acontecimiento tan reciente, varios autores e instituciones han analizado el impacto que ha tenido en diferentes ámbitos de la vida, precisando las posibles repercusiones que pudo tener a nivel educativo. Dichos artículos afirman un rezago académico. [2]

En junio de 2022 L. Ramírez publicó un artículo integrando técnicas de minería de datos (MD), como el análisis de clúster jerárquico y la regresión logística, realizando un sondeo a los alumnos para obtener datos acerca de las condiciones sociodemográficas, económicas, técnicas, de salud mental y académicas que permitieron encontrar patrones que incidieran en el desempeño académico del alumnado. Obteniendo como resultado la existencia de 2 grupos, uno con mejores condiciones que otro y también un listado de ítems que influyen tanto negativamente como positivamente a que un alumno se vea afectado en su rendimiento escolar.[5]

Otros autores como Adriana Olvera utilizaron una metodología cuantitativa descriptiva con unidades de análisis asociadas al abandono escolar, en el contexto de la pandemia del COVID-19, obteniendo la información a través de una encuesta en línea aplicada a 840 docentes de todos los niveles y modalidades educativas, teniendo como resultado una disminución en la matrícula escolar. [6]

En la actualidad el uso de técnicas de minería de datos a nivel sector salud en pandemia han sido de utilidad para estimar con antelación, el número de casos y muertes por COVID-19, la demanda de recursos hospitalarios, el número de vacunas y medicamentos necesarios, permitiendo tomar decisiones estratégicas y planificadas para luchar de manera eficiente contra el COVID-19.

Por lo anterior queda destacado que el Data Mining o Minería de datos es una herramienta que integra los datos de diferentes fuentes para, posteriormente, extraer un importante conocimiento, es decir, identificar información trascendente, valiosa y útil, a partir de lo cual las instituciones van a poder tomar alguna significativa decisión.[7]

Combinando técnicas de la estadística, inteligencia artificial y otras áreas, las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Facilitando la búsqueda automática de grandes almacenes de datos para descubrir patrones y tendencias que van más allá del análisis simple, utilizando sofisticados algoritmos matemáticos para segmentar los datos y predecir la probabilidad de eventos futuros basados en eventos pasados. [8] [9] [7].

Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos. [9].

El objetivo del presente trabajo de estudio es aplicar técnicas de minería de datos, para identificar el rezago de aprendizaje como consecuencia de la modalidad de estudio en los estudiantes de la carrera de Ingeniería en Sistemas Inteligentes en el periodo 2020-2022 a causa de la pandemia del COVID-19, además si existe una relación en el abandono de escolar. Para que de acuerdo a los resultados obtenidos en este trabajo, la universidad al ya trabajar en un sistema híbrido se logre mejorar la forma de impartir el contenido de las materias y a su vez incorporar asesorías.

2. MATERIALES Y MÉTODOS.

Se analizaron 1600 datos correspondientes a 64 estudiantes de la carrera de Ingeniería en Sistemas Inteligentes que se obtuvieron a partir de un formulario realizado en Google Drive. El diseño se realizó a partir de preguntas de opciones múltiples y abiertas, para identificar cuáles fueron los retos a los que se enfrentaron con la educación a distancia derivada de la pandemia del COVID-19.

El formulario está compuesto por 25 preguntas el cual fue contestado voluntariamente en línea, alcanzando un número de 64 personas encuestadas, abordando aspectos socioeconómicos, tecnológicos, etc. Los resultados del formulario de Google drive se exportaron a una hoja de

cálculo. Para validar la encuesta se utilizó la técnica de Alfa de Cronbach que es el promedio de las correlaciones entre los ítems que son parte de un instrumento produciendo valores que oscilan entre 0 y 1[10], con ayuda del software IBM SPSS Statistics Base 22 [11], se obtuvo un coeficiente de confiabilidad de 0.76 que de acuerdo a los rangos de Alfa de Cronbach la encuesta es aceptable.

La investigación se desarrolló bajo el enfoque cuantitativo de tipo descriptivo. Utilizando el proceso general del KDD (Knowledge Discovery in Databases) para los datos recabados, donde de acuerdo con [12], se contempla: entendimiento del problema, selección de datos, preprocesamiento, limpieza y transformación de los datos.

2. 1. Entendimiento del problema

En esta fase se requiere comprender con exactitud el problema al que se le va a dar solución por medio de la minería de datos. Permitiendo recolectar la información necesaria para interpretar con asertividad los resultados encontrados. [12]

2. 2. Selección y preprocesamiento de datos

La selección y preprocesamiento de los datos se obtuvo mediante la identificación de variables (edad, situación laboral, acreditación de materias, inscripciones a ciclo escolar, obstáculos que presento el estudiante para tomar clases en línea) la limpieza de datos consistió en seleccionar los datos más relevantes eliminando registros duplicados o eliminando anomalías; por el tipo de rigor, al momento de recabar la información y procesarla, se tuvieron inconsistencias, por lo que fue necesario reparar datos.[13] En la tabla 1 se muestra la descripción de cada uno de los atributos seleccionados como edad, situación laboral, situación del hogar, situación escolar, ventaja y desventajas de la pandemia por COVID-19, para el análisis y modelo de minería de datos.

Tabla 1. Descripción de atributos basados en las preguntas del cuestionario.

Atributo	Definición	Descripción
Edad	Edad	Edad de la persona encuestada.
Vives con	Vives con	Indica si vive solo o con sus padres.
Vives en	Vives en	Indica el tipo de vivienda.
AA	¿Cuántos hermanos tienes?	Indica la cantidad de hermanos que tiene.
AB	¿Quiénes trabajan en tu hogar?	Indican las personas que trabaja en el hogar.
AC	¿Trabajas?	Indica si la persona encuestada trabaja.
AD	¿Mientras tomabas clases en línea tenías un espacio propio para realizar dicha actividad?	Indica si tenían un espacio designado para su estudio.

AE	¿Cuentas con las siguientes herramientas tecnológicas?	Indica si contaba con diferentes herramientas tecnológicas.
AF	¿Estuviste inscrito en el ciclo escolar 2020A - 2020B?	Indica sí estuvo inscrito ese ciclo escolar.
AG	¿Acreditaste todas tus materias?(ciclo 2020A - 2020B)	Indica si acredito las materias.
AH	¿Cuáles materias debes? (ciclo 2020A - 2020B)	Indica las materias que debe en ese ciclo.
AI	¿Concluiste el ciclo 2020A - 2020B?	Indica si concluyo el ciclo.
BA	¿Estuviste inscrito en el ciclo escolar 2021A - 2021B?	Indica sí estuvo inscrito ese ciclo escolar.
BB	¿Acreditaste todas tus materias? (ciclo 2021A - 2021B)	Indica si acredito las materias.
BC	¿Cuáles materias debes? (ciclo 2021A - 2021B)	Indica las materias que debe en ese ciclo.
BD	¿Concluiste el ciclo 2021A - 2021B?	Indica si concluyo el ciclo.
BE	¿Consideras que hubo un posible rezago académico durante pandemia?	Indica si cree que existe un posible rezago académico.
BF	¿Consideras que los siguientes sucesos fueron motivos por los cuales hubo un rezago académico en época COVID-19?	Indica si el posible rezago de debe a diferentes sucesos.
BG	¿Con que frecuencia utilizaste los siguientes dispositivos electrónicos para realizar tu actividades durante el ciclo escolar en pandemia?	Indica la frecuencia en la que utilizo diferentes dispositivos electrónicos.
BH	¿El dispositivo PRINCIPAL que utilizó para sus actividades escolares...?	Indica cual fue el dispositivo principal que utilizo para tomar clases.
CA	¿Consideras que el dispositivo que utilizo tuvo influencia en tu aprendizaje?	Indica si cree que el dispositivo tuvo que ver en su aprendizaje.
CB	¿Qué medio(s) utilizaba tu institución para	Indica los medios que utilizo la escuela para dejar actividades.

	informar sobre las actividades escolares durante el periodo de pandemia anterior año escolar?	
CC	Para tomar las clases a distancia por COVID-19 ¿Se realizó algún gasto que no se tuviera contemplado para...?	Indica si se realizó un gasto no contemplado.
CD	¿Cuáles han sido las desventajas de las clases a distancia?	Indica las desventajas de estudiar a distancia.
CE	¿Cuáles han sido las ventajas de las clases a distancia?	Indica las ventajas de estudiar a distancia.

Fuente: elaboración propia.

2.3. Análisis de datos

Se realizó el análisis de datos de las variables evaluadas (edad, situación laboral, dispositivos tecnológicos, si era de uso compartido el dispositivo con el que realizaba sus actividades, la cantidad de hermanos que tiene y la acreditación de materias), mediante el software WEKA. [14] Se hizo uso del análisis descriptivo, ya que lo que se quiere obtener es, identificar los patrones que permitan explorar las propiedades de los datos y sus relaciones. [15]

2.4. Clasificación

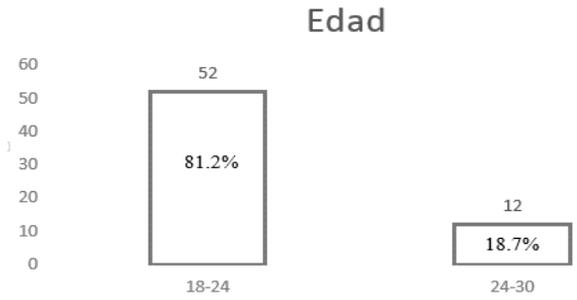
Los datos incluidos en el conjunto en su mayoría eran categóricos, sin embargo al estar la base compuesta por algunas preguntas de más de dos opciones se optó por utilizar clúster y classify, utilizando el simple k-means y el algoritmo J48.

3. RESULTADOS

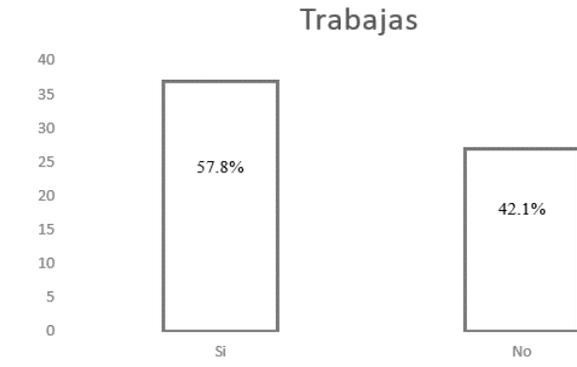
De acuerdo con los resultados obtenidos en las diferentes pruebas realizadas en la etapa de minería de datos con las tareas de clasificación, el análisis de frecuencia de las variables estudiadas (edad, situación laboral, dispositivos tecnológicos, si tuvo que compartir el dispositivo con el que realizaba sus actividades, la cantidad de hermanos que tiene y la acreditación de materias) para las 64 personas corresponden con un rango de edad de 18 a 24 años que equivale al 81.2% y de 24 a 30 años 18.7% (grafica 1 a), de los cuales el 57.8% si trabajan y el 42.1% no (grafica 1 b), un 39% tienen un solo hermano y el 53.1% tiene 2 o más hermanos (grafica 1 c), en cuanto a los dispositivos tecnológicos el 34.3% cuentan con computadora de escritorio, el 64% tienen laptop, 14% Tablet, el 100% celular inteligente y 95% internet (grafica 1 d), dichos dispositivos el 43.7% eran de uso propio y el 56.2% lo compartían, lo tuvieron que compartir o lo pidieron prestado a algún familiar (grafica 1e), en lo referente a si acreditaron todas las materias en el ciclo escolar 2020A-2020B(AG) el 70.3% y en el

2021A-2021B(BB) el 73.4% no las acreditaron todas (grafica 1 f y 1 g).

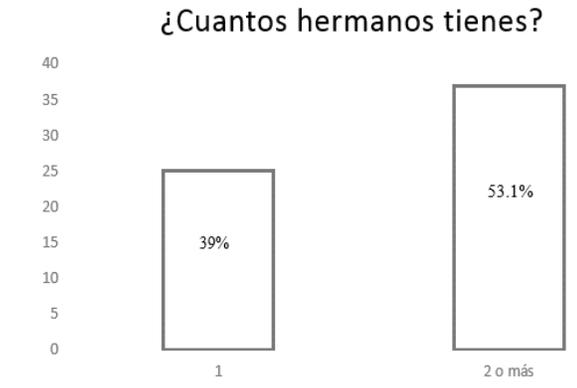
Gráfica 1. Resultados de las diferentes pruebas realizadas.



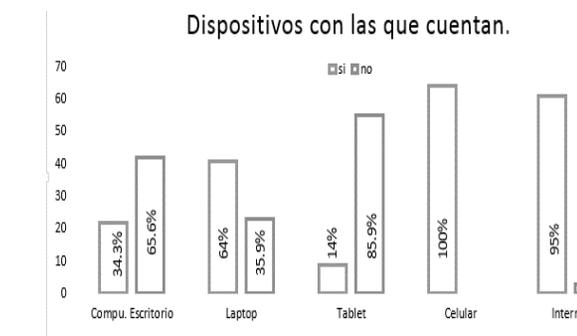
(a)



(b)

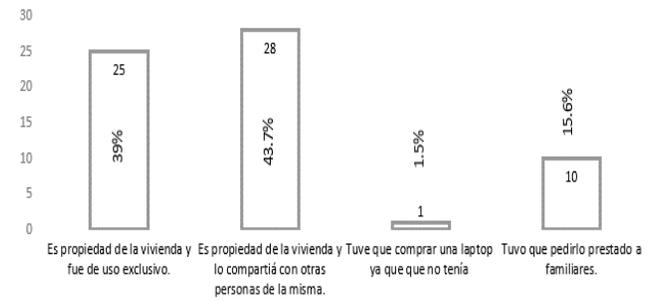


(c)



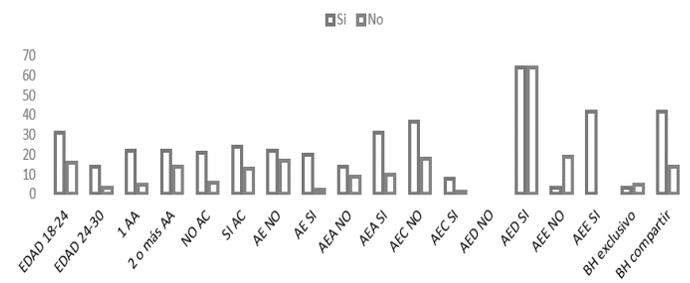
(d)

El dispositivo PRINCIPAL que utilizó para sus actividades escolares..



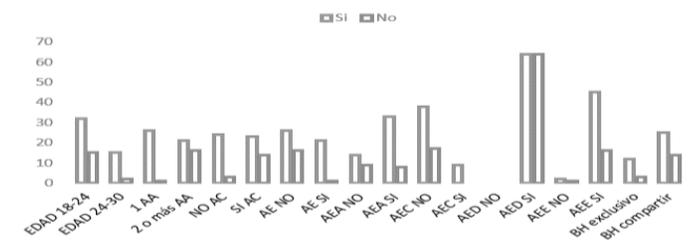
(e)

¿Acreditaste todas tus materias? (ciclo 2020A - 2020B)



(f)

¿Acreditaste todas tus materias? (ciclo 2021A - 2021B)



(g)

Fuente: elaboración propia.

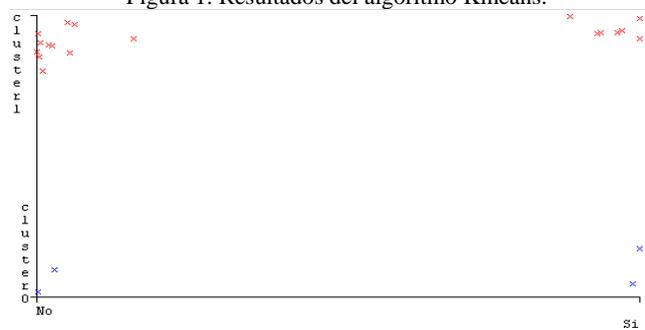
Nota: a) Rangos de edad, b) Trabajan, c) Hermanos que tiene, d) Dispositivos que cuentan, e) Si el dispositivo era de uso propio, f) Acredito todas las materias ciclo 2020A-2020B, g) Acredito todas las materias ciclo 2021A-2021B.

Lo anterior promueve el descubrimiento de patrones inesperados que conllevan al surgimiento de la etapa de minería de datos, en donde pueden ser aplicados algoritmos de clasificación, clustering y asociaciones. Por la estructura de la base se optó por utilizar el algoritmo SimpleKmeans que es un método de agrupamiento (clustering), que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano.[16]

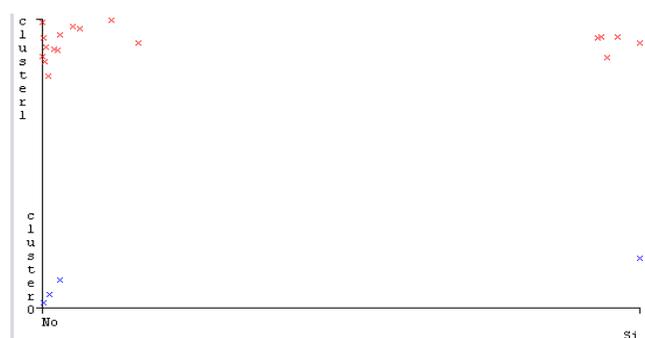
Se dividió la base en 2 grupos (Cluster# 0 y Cluster# 1), realizando un numero de 5 iteraciones. Teniendo para el Cluster# 0 28 datos y el Cluster# 1 36 datos, con un margen de error de 30.8. En el resultado obtenido por el algoritmo de

clasificación SimpleKmeans se presentan las respuestas con mayor concentración en los dos grupos. En la figura 1 se aprecia una visualización más concreta de la aplicación del SimpleKmeans donde se observa que la mayor concentración de datos para la variable AG (¿Acreditaste todas tus materias? (ciclo 2020A - 2020B)) gran parte de los datos no acreditaron todas las materias (Figura 1 a), y en BB (¿Acreditaste todas tus materias? (ciclo 2021A - 2021B)) la mayoría de los datos no acreditaron todas las materias (figura 1 b)

Figura 1. Resultados del algoritmo Kmeans.



(a)



(b)

Fuente. Elaboración Propia.

Nota: a) ¿Acreditaste todas tus materias? (ciclo 2020A - 2020B)), b) ¿Acreditaste todas tus materias? (ciclo 2021A - 2021B))

Para las variables, edad (Edad) están en un rango de 18 a 24 años, en la variable AA (¿Cuántos hermanos tienes?) la mayoría tienen 2 hermanos, en AC (¿Trabajas?) la cantidad de personas que si trabajaron predomina más, para AE (Contaban con computadora de escritorio) la mayoría no tenían computadora de escritorio, en AEA (Contaban con computadora portátil) gran parte de los datos si tenían computadora portátil, para AEC (Contaban con Tablet) casi todos no tenían Tablet, para AED (Contaban con Celular inteligente) todos si contaban con dicho dispositivo y para BH (¿El dispositivo PRINCIPAL que utilizó para sus actividades escolares..) la mayoría lo compartían el dispositivo.

Dentro de las variables analizadas, la cantidad de hermanos que tienen los estudiantes, el hecho de que trabajaran y los dispositivos con los que cuentan eran de uso compartido tuvieron cierta influencia en el acto de que la mayoría de estudiantes no acreditaran todas las materias, si se toma en cuenta que los estudiantes tienen 2 o más hermanos, contaban con dispositivos para realizar sus actividades pero estos eran

compartidos y la expectativa de acreditar las materias es baja, se puede deducir que el rendimiento en el nivel de educativo bajo.

El Kmeans al ser un algoritmo no supervisado es decir que basa su proceso de entrenamiento en un juego de datos sin etiquetas donde no se puede reflejar una relación de variables entre sí. Recurriremos a la aplicación de un algoritmo supervisado como es el J48, en el cual se tiene una variable dependiente, y el objetivo del clasificador es determinar el valor de dicha variable, comenzando el proceso de la creación de un árbol seleccionando la variable a partir de la cual se va a dividir el nodo raíz.[17] Aplicaremos el J48 a las variables variable AG (¿Acreditaste todas tus materias? (ciclo 2020A - 2020B)) y BB (¿Acreditaste todas tus materias? (ciclo 2021A - 2021B)), la media de las instancias correctamente clasificadas es de 80% y de instancias incorrectamente clasificadas es de 12%

Tabla 1. Resultados del algoritmo J48.

AI = Si: No (49.0/6.0)
AI = No
CB = Plataformas virtuales(Classroom, Zoom, Microsoft Teams, etc.): Si (12.0)
CB = Redes sociales (grupos de Whatsapp, Facebook, etc.): No (3.0/1.0)
CB = Correo electrónico.: Si (0.0)
CB = Todos los anteriores : Si (0.0)
CB = Pocas veces lo supe debido a la falta de conexión de mi parte: Si (0.0)

(c)

BD = Si: No (54.0/7.0)
BD = No: Si (10.0)

(d)

Fuente. Elaboración propia.

Nota: c) ¿Acreditaste todas tus materias? (ciclo 2020A - 2020B), d) ¿Acreditaste todas tus materias? (ciclo 2021A - 2021B)

Interpretando los resultados obtenidos por el algoritmo J48 para AG (¿Acreditaste todas tus materias? (ciclo 2020A - 2020B)) se tiene como nodo raíz a la variable AI (¿Concluiste el ciclo 2020A - 2020B?), los que si concluyeron no acreditaron todas las materias, de los que no concluyeron nos lleva a la variable CB (¿Qué medio(s) utilizaba tu institución para informar sobre las actividades escolares durante el periodo de pandemia anterior año escolar?) y utilizaban Plataformas virtuales(Classroom, Zoom, Microsoft Teams, etc.) si acreditaron, si utilizaban Redes sociales (grupos de Whatsapp, Facebook, etc.) no acreditaron. Teniendo un total de 52 datos que no acreditaron todas las materias en el ciclo 2020A - 2020B

Para la variable BB (¿Acreditaste todas tus materias? (ciclo 2021A - 2021B)) el árbol es más sencillo, teniendo como nodo raíz la variable BD (¿Concluiste el ciclo 2021A - 2021B?) si concluíste no acreditaste todas las materias de ese ciclo y si no lo concluíste si acreditaste todas las materias.

Con un total de 54 datos que no acreditaron todas las materias en el ciclo 2021A - 2021B.

Para este algoritmo se puede obtener más información dependiendo de cuál sea nuestra variable independiente.

4. CONCLUSIONES

Con base a los resultado obtenidos en ambos algoritmos, se concluye que en el periodo de pandemia debido al COVID-19 que abarco los ciclos escolares 2020A, 2020B, 2021A y 2021B, la mayoría de los estudiantes no acreditaron todas las materias esto puede ser consecuencia de ciertos factores y acontecimientos que se presentaron, entre los que destaca el hecho de la falta de uso exclusivo de los dispositivos electrónicos y a su vez la falta de dispositivos electrónicos y al hecho de que la mayoría de estudiantes trabajaban, las plataformas que la institución utilizo para las actividades y si concluyeron los ciclos escolares, generando que el nivel de aprendizaje disminuyera, lo cual coincide con los resultados obtenidos en el artículo Aplicación de técnicas de minería para la caracterización de estudiantes bajo el efecto de la COVID-19.[5]

Al comparar el uso de ambos algoritmos se puede destacar el hecho de que el algoritmo SimpleKmeans resultó ser eficiente con la investigación, ya que se logró obtener resultados que pueden ser de utilidad a futuro, sin embargo el algoritmo J48 nos proporcionó más información que nos permitió llegar a un resultado más concreto, ya que al ser un algoritmo supervisado nos proporciona mayor información al momento de interactuar con la variable independiente. Es necesario puntualizar y enfatizar lo importante que ha sido la minería de datos para analizar datos de esta índole ya que gracias ella se pudo llegar a determinar este fenómeno relacionado con el nivel de aprendizaje en la época de COVID-19.

Los resultados de este artículo proporcionan elementos que serán de gran ayuda para proponer estrategias para mejorar la forma de impartir el contenido de las materias al trabajar en un sistema híbrido, así como incorporar asesorías a alumnos que lo necesiten.

5. TRABAJOS FUTUROS

En lo subsecuente con base a los resultados obtenidos se pretende realizar una nueva encuesta para saber si los alumnos que recibieron asesoría presencial lograron mejorar su aprendizaje, así mismo se está trabajando con otros docentes en la parte didáctica que involucre el desarrollo de material, con ello realizar algún instrumento que permita conocer la aceptación y aprendizaje del alumno respecto a este tema. Así mismo, se dará el seguimiento a los datos recopilados y se realizara lo conducente con técnicas de minería.

6. REFERENCIAS

[1] L. Medina, C. Chao Rebolledo, E. Garduño, M. C. González, M. Baptista, L. Montes, L. Medina Velázquez, M. Navarro, C. Covarrubias, L. Rojas, J. Ojeda, C. Monereo Font, A. García, A. Siqueiros, W. Rojas, A. Jiménez y H. García, “

Educación en contingencia durante la covid-19 en México”, Fundación SM, 2021.

[2] CEPAL-UNESCO, “La educación en tiempos de la pandemia de COVID-19”, 2020.

[3] INEGI, “Encuesta para la Medición del Impacto COVID-19 en la Educación (ECOVID-ED), 2020.

[4] B. Flores, J. Pérez, “Los retos de la educación a distancia en las prácticas educativas durante la pandemia de COVID-19.”, 2021.

[5] L. Ramírez, “Aplicación de técnicas de minería de datos para la caracterización de estudiantes bajo el efecto de la COVID-19.”, Publicación Semestral Padi Vol. 10 No. Especial 2 (75-81), 2022.

[6] A. Carro Olvera, J. Lima Gutiérrez, “Pandemia, rezago y abandono escolar: Sus factores asociados.”, Revista Andina de Educación 5(2) 005210, 2022.

[7] M.C B. Martínez, “Minería De Datos”, 2016.

[8] J. Riquelme, “Minería de Datos: Conceptos y Tendencias”, Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, pp. 11-18, 2006.

[9] R. Suárez, Y. Amador, “Herramientas de Minería de Datos”, Revista Cubana de Ciencias Informáticas, vol. 3, núm. 3-4, pp. 73- 80, 2009.

[10] J. Vinicio, M. Angel Duque, A. Reinoso, “Alfa de Cronbach para validar un cuestionario de uso de tic en docentes universitarios”, Revista mktDescubre, ISSN-1390-7352, N° 3232-12,2017

[11] IBM SPSS Statistics Base 22, Versión 29.0, 2022, Licencia de prueba.

[12] M. Aular, T. Pereira, “Minería de Datos como soporte a la toma de decisiones empresariales”, Opción, vol. 23, núm. 52, pp. 104-118, 2007.

[13] F. Herrera, J. Riquelme, “Preproceso de Datos”, Reunión Red Nacional DM y ML, 2004.

[14] Weka, Versión 3.8.4 (1999-2019).

[15] H. Marisol Marin, “Minería de Datos”, Universidad Politécnica de Victoria, 2007.

[16] C. García, I. Gómez, “Algoritmos de aprendizaje: KNN y KMEANS, Universidad Carlos III de Madrid 2012.

[17] C. Hernán, N. Gelvez, J. Palacios, “Análisis de datos mediante el algoritmo de clasificación J48, sobre un clúster en la nube de AWS”, Universidad Distrital Francisco José de Caldas, Volumen especial E-ISSN: 2248-762X, 2016.