

Efecto que tuvo el aprendizaje a distancia durante la pandemia en estudiantes de nivel básico utilizando minería de datos.

Ana Miriam Garcia Cruz, Dr. Ricardo Rico Molina, Dra. Carmen Liliana Rodríguez Páez, Dr. Israel Gutiérrez Gonzales, Dra. Doricela Gutiérrez Cruz.

^a Centro Universitario UAEM Nezahualcóyotl, anacruzuaem@gmail.com, ricom@uaemex.mx, clrodriguezp@uaemex.mx, Estado de México, Méx.

Resumen

En el presente trabajo se tiene como objetivo primordial analizar el efecto que tuvo la introducción del aprendizaje a distancia o desde casa, para los estudiantes de nivel básico, mediante el apoyo de la minería de datos. Los datos utilizados en este trabajo fueron recabados por el autor y proporcionados por una muestra de 277 alumnos de nivel básico, a través de la aplicación de un cuestionario de 54 preguntas, las cuales abarcan un periodo dividido en tres tiempos: antes, durante y al término de la pandemia, obteniendo la recolección, limpieza y tratamiento de 14,958 datos para la muestra de dicha población estudiantil.

El tratamiento de estos datos se realizó con la orientación de la metodología KDD (Knowledge Discovery in Databases), llevando a cabo cada una de sus fases: recopilación de datos, selección, limpieza, transformación, minería de datos, interpretación y evaluación de los modelos obtenidos. Se desarrolló en la herramienta Weka, utilizando para clasificación el algoritmo J48, en agrupamiento el algoritmo SimpleKmeans y en asociación el algoritmo Apriori; comparando entre ellos los resultados y eligiendo el óptimo para con ello concluir si la implementación del sistema de aprendizaje a distancia afectó de manera positiva o negativa en los alumnos.

Palabras clave—Aprendizaje, Estudiantes, Minería de datos, Pandemia, WEKA.

Abstract

The primary objective of this work is to analyze the effect that the introduction of distance or home learning had for basic level students, through the support of data mining. The data used in this work were collected by the author and provided by a sample of 277 basic level students, through the application of a questionnaire of 54 questions, which cover a period divided into three times: before, during and at the end of the pandemic, obtaining the collection, cleaning and processing of 14,958 data for the sample of said student population.

The treatment of these data was carried out with the guidance of the KDD (Knowledge Discovery in Databases) methodology, carrying out each of its phases: data collection, selection, cleaning, transformation, data mining, interpretation and evaluation of the models obtained. It was developed in the Weka tool, using the J48 algorithm for classification, the SimpleKmeans algorithm for clustering and the A priori algorithm in association; comparing the results

between them and choosing the optimal one to conclude whether the implementation of the distance learning system affected the students positively or negatively.

Keywords—Learning, Students, Data mining, Pandemic, WEKA.

1. INTRODUCCIÓN

A principios del año 2020 los medios de comunicación asiáticos reportaban la primera víctima a causa del coronavirus semanas después la organización mundial de la salud (OMS) anunciaba al COVID-19 como una nueva enfermedad y poco tiempo después se le considero como pandemia, paralizando todas las actividades sociales a nivel internacional [8]. En el caso de México, el 27 de febrero de 2020 la secretaria de salud detecto el primer caso de Covid19, lo que llevo a remplazar las herramientas de enseñanza convencionales durante la pandemia por lo cual las tecnologías en la educación se han transformo en una herramienta fundamental para brindar educación a distancia [1].

Los salones de clase y los métodos de enseñanza han cambiado en todo el mundo, y la virtualidad y diversas metodologías se han apoderado del proceso de enseñanza y aprendizaje [2]. La Organización Iberoamericana de Educación estima que 14 millones de infantes no asisten a la escuela durante la infancia, incluidos 46 millones en la escuela primaria, en los 23 países miembros de la Comunidad Iberoamericana de la OEI, 44 millones de estudiantes de secundaria y más de 18 millones de universitarios, lo que nos arroja que más de 122 millones de estudiantes en América Latina se vean perjudicados. El cierre de escuelas requirió impostergable la experiencia de la enseñanza virtual hasta el momento, porque en muchas latitudes geográficas esta manera de enseñar con tecnologías de la información y la comunicación (TIC) aproximó al punto importante de saber enfrentarla para continuar con la educación a distancia, especialmente cuando en la educación la clausura de instituciones agudizó, la enorme disparidad entre los estudiantes que tienen teléfonos móviles y computadoras y buena conexión a internet, y los estudiantes que con menos posibilidades que no tienen teléfonos móviles ni computadoras y, peor aún, tienen conexiones de internet que no les permiten continuar en las clases virtuales a pesar de las alternativas que hubieron como Moodle o Google Hangouts, que resaltaron como las preferidas entre profesores y escuelas. [3].

El cierre de puertas de las escuelas en todo el mundo planteó una situación sin precedentes en la historia de la educación y trajo diversos desafíos a la calidad de la misma, por lo que los países se propusieron impulsar el proceso educativo mediante el uso de las tecnologías de la información y las comunicaciones como forma estratégica. La reubicación forzada y temporal dio paso a la llamada enseñanza remota de emergencia, ya que la crisis permitió diseñar un plan de enseñanza por lo cual afectó negativamente el aprendizaje de los estudiantes y, por ende, la calidad de la educación [4].

Debido a lo anterior se realizaron estudios que aportan datos mostrando que las pérdidas de aprendizaje ocurren principalmente en las áreas de matemáticas, lengua y literatura. La herramienta de simulación Learning Adjusted Years of Schooling (LAYS) del Banco Mundial, evaluó el impacto del cierre de escuelas en la pérdida de aprendizaje ideal y años de escolaridad, dando como resultado que en promedio se tiene un retraso en años de escolaridad en América Latina y el Caribe de 1,3 años [4].

Debido a lo anterior en el sistema educativo surgen necesidades en los alumnos, que pueden provocar estrés e incomodidad debido a que no encuentran un camino para resolver sus problemas. Por lo tanto, los docentes deben planificar y estructurar actividades participativas e interactivas en función de los temas presentados en los contenidos de tal manera que se mejore el proceso de aprendizaje de la enseñanza de habilidades docentes. [6].

En este sentido se requiere tener análisis que permitan ver los diferentes tiempos educativos que se vivieron antes, durante y después ya que esto permitirá tener la información útil que explique que cambio tuvieron los estudiantes. Para esto se puede usar una técnica como la minería de datos, que en su propia definición es la extracción de datos para detectar patrones de comportamiento útiles para la toma de decisiones. Está estrechamente relacionado con la estadística, utilizando métodos de muestreo, visualización de datos y organización. [5].

Utilizando los métodos y herramientas adecuados, esta información puede resultar útil en su proceso de toma de decisiones. Una de las ventajas de la minería de datos (MD) es que es fácil de usar y aplica el conocimiento adecuado de los diferentes tipos de algoritmos utilizados [7].

El objetivo principal es analizar el efecto que provocó la introducción del aprendizaje a distancia para los estudiantes de nivel básico, mediante el apoyo de la minería de datos utilizando la metodología KDD.

2. MATERIALES Y MÉTODOS

Para esta investigación se utilizó la metodología KDD (Knowledge Discovery in Databases) la cual es un proceso reiterado que involucra el conocimiento del problema con diferentes métodos de análisis de datos tradicionales y avanzados con el objetivo de encontrar relaciones y patrones que se puedan usar para tomar decisiones útiles en el contexto que se usen [9].

Etapas del proceso KDD:

- Etapa de selección: Utilizando técnicas de filtrado de atributos e historial, los datos irrelevantes se eliminan para su posterior análisis [10].
- Limpieza de Datos: Esto es importante porque los datos sucios y desordenados pueden afectar negativamente la calidad de los resultados de la minería de datos [11].
- Transformación de Datos: Podemos asumir que los datos están limpios e imparciales y, después de observar los aspectos cubiertos en nuestro trabajo anterior, ahora podemos usarlos para entrenar

nuestro modelo. Esto casi siempre implica la transformación reversible de los datos originales para producir una nueva versión más adecuada para el análisis KDD. A grandes rasgos se adaptan los datos para que puedan usarse dependiendo el algoritmo con el que serán procesados [12].

- Minería de Datos: Dependiendo del propósito establecido para el estudio si es predictivo; el primer paso para elegir una técnica de minería de datos apropiada, como clasificación, regresión o agrupamiento, implica encontrar un modelo que pueda usarse en el futuro y condiciones desconocidas. Si es descriptivo, únicamente observa su comportamiento [13].
- Interpretación y evaluación: Luego de obtener el conocimiento y el modelo de la fase minera, el siguiente paso es valorar cómo esta información se relaciona con la realidad en términos de los objetivos definidos en la primera fase. [14].

Muchos investigadores han identificado el proceso de obtención de información a partir de grandes conjuntos de datos como un proyecto de investigación importante en el campo de los sistemas, de la privacidad y muchas empresas industriales lo consideran un lugar importante y una oportunidad para obtener enormes ganancias debido a la información que se puede aprovechar y no es tan trivial encontrar. [15].

Para la presente investigación se tomó una muestra de 277 alumnos de nivel básico, a través de la generación y aplicación de un cuestionario de 54 preguntas, las cuales abarcan un periodo dividido en tres tiempos: antes, durante y al término de la pandemia, obteniendo la recolección, limpieza y tratamiento de 14,958 datos para la muestra de dicha población estudiantil.

2.1 Selección y preprocesamiento de Datos

En esta primera fase se optó por seleccionar las principales variables que conformarían el banco de datos final (sexo, edad, ayuda_casa y promedio), en esta etapa se empezó a considerar las variables que no eran necesarias para la integración del conjunto de datos final.

En la figura 1, se muestra el cuestionario que se aplicó a los alumnos para la recopilación de datos.

Fig. 1. Cuestionario.

UAEM Universidad Autónoma del Estado de México

Matemáticas Física Química Historia Inglés

10D¿Qué materia le costaba menos trabajo entender?
Matemáticas Física Química Historia Inglés

11D¿Preguntaba a su profesor cuando tenía dudas de algún tema?
Siempre Casi siempre Ocasionalmente Casi nunca Nunca

12D¿Resolvía sus dudas?
Siempre Casi siempre Ocasionalmente Casi nunca Nunca

13D¿Recibía ayuda de compañeros al hacer sus tareas?
Siempre Casi siempre Ocasionalmente Casi nunca Nunca

14D¿Recibía ayuda en casa al hacer sus tareas?
Siempre Casi siempre Ocasionalmente Casi nunca Nunca

15D¿Qué medios utilizaba para hacer sus tareas?
Libros Enciclopedias Revistas Apuntes Internet

16D¿Cómo entregaba sus tareas?
A mano Impresas USB Correo Plataforma

17D¿Qué plataformas aprendió a usar?
Classroom EdModo Meet Zoom Otra

18D¿Recibió orientación de su escuela para el uso de plataformas?
Siempre Casi siempre Ocasionalmente Casi nunca Nunca

19D¿Cómo le resultaba realizar trabajos en equipo?
Muy Fácil Fácil Difícil Muy difícil No participé

20D¿Qué tan seguro estaba al momento de realizar sus evaluaciones?
Muy seguro=100% Seguro=75% Regular=50% Poco seguro=25% Nada=0%

21D¿Se sentía conforme con los resultados obtenidos en sus evaluaciones?
Siempre Casi siempre Ocasionalmente Casi nunca Nunca

22D¿Qué promedio tenía?
Muy alto=9.5-10 Alto=8.5-9.4 Regular=7.5-8.4 Bajo=6.5-7.4 Muy bajo=6-6.4

Fuente: Elaboración propia.

Fuente: Elaboración propia.

2.2 Limpieza de Datos

Para la siguiente etapa se realizaron pruebas unitarias y tomando en cuenta los resultados obtenidos, se descartaron todos aquellos datos o variables que no daban aportación significativa al análisis, así como también descartar cualquier inconsistencia de datos como por ejemplo los registros incompletos.

2.2 Transformación de Datos

Dentro de esta fase se buscaron las características útiles para representar los datos que nos ayudarán al análisis. En la tabla 1 se presentan las variables finales a evaluar. Mismas que dieron mejores resultados al ser evaluadas de forma unitaria.

Tabla 1. Datos finales.

	Nombre de la variable	Descripción
1	Sexo	Indica el género del alumno.
2	Edad	Indica cuantos años tiene el alumno.
3	Ayuda_casa	Indica si recibía apoyo para realizar sus tareas en casa.
4	Promedio	Indica la calificación general del alumno.

2.3 Minería de Datos

En esta fase se eligió la herramienta Weka (Waikato Environment for Knowledge Analysis) la cual es un Software de aprendizaje automático y minería de datos (data mining) escritos en Java en la Universidad de Waikato [16].

Utilizando para la clasificación el algoritmo J48; por ser de tipo supervisado y categórico, para agrupamiento el algoritmo SimpleKmeans; que nos ayuda a crear agrupación de objetos según sus características y para asociación el algoritmo A priori; que analiza patrones arrojando la relación que tienen. Comparando entre ellos los resultados y eligiendo el óptimo para con ello concluir si la implementación del sistema de aprendizaje a distancia afecto de manera positiva o negativa en los alumnos.

2.3 Resultados

En la figura 2 al aplicar el algoritmo J48 para antes de la pandemia se observa que el promedio Regular en su mayoría lo tenía en sexo femenino quienes indican que sus materias más fáciles eran Historia, Química e Inglés, mientras que el sexo masculino preguntaban sus dudas ocasionalmente y sus materias más fáciles eran Matemáticas y Física, también nos indica que la edad promedio mayor es de 14 años, comprendían mucho y ocasionalmente se reunían con sus compañeros.

Fig. 2. Resultados J48 antes de pandemia.

```
J48 pruned tree
-----
Promedio = REGULAR: F (11.0/3.0)
Promedio = ALTO
| Preguntaba_dudas = OCASIONALMENTE: M (6.0/2.0)
| Preguntaba_dudas = CASI SIEMPRE
| | Materia_dificil = INGLES
| | | Materia_facil = HISTORIA: F (56.0/25.0)
| | | Materia_facil = QUIMICA: F (12.0/5.0)
| | | Materia_facil = INGLES: F (0.0)
| | | Materia_facil = MATEMATICAS: M (10.0/2.0)
| | | Materia_facil = FISICA: M (6.0/2.0)
| | | Materia_dificil = HISTORIA: F (4.0)
| | | Materia_dificil = MATEMATICAS
| | | EDAD = CATORCE
| | | Comprendia = MUCHO
| | | | Reunia_compañeros = OCASIONALMENTE
| | | | Ayuda_casa = CASI SIEMPRE
| | | | | Materia_facil = HISTORIA
| | | | | | Resolvia_dudas = SIEMPRE: M (16.0/5.0)
| | | | | | Resolvia_dudas = CASI SIEMPRE : F (8.0/2.0)
| | | | | | Resolvia_dudas = OCASIONALMENTE: M (0.0)
| | | | | | Materia_facil = QUIMICA: F (5.0/1.0)
| | | | | | Materia_facil = INGLES
| | | | | | | Ayuda_compañeros = OCASIONALMENTE: M (10.0/3.0)
| | | | | | | Ayuda_compañeros = CASI SIEMPRE: M (1.0)
| | | | | | | Ayuda_compañeros = CASI NUNCA: F (13.0/5.0)
```

Fuente: Elaboración propia.

En la figura 3 al aplicar el algoritmo J48 durante la pandemia se observa que el promedio seguía siendo Regular, quienes preguntaban sus dudas casi siempre era el sexo femenino y se sentían capaces de aprender con las clases en línea, la materia más difícil para el sexo masculino era Física, y en general se les dificultaba mucho realizar trabajos en equipo.

Fig. 3. Resultados J48 durante la pandemia.

```

J48 pruned tree
-----
Promedio = REGULAR
EDAD = CATORCE
| Preguntaba_dudas = CASI SIEMPRE: F (40.0/13.0)
| Preguntaba_dudas = SIEMPRE
| Medios_digitales = OCASIONALMENTE
| Capaz_aprender = CAPAZ: F (2.0)
| Capaz_aprender = REGULAR
| Materia_dificil = FISICA: M (1.0)
| Materia_dificil = MATEMATICAS
| Interacción_compañeros = OCASIONAL
| Seguro_evaluaciones = REGULAR
| Materia_facil = HISTORIA
| Trabajos_equipo = DIFICIL
| Ayuda_casa = CASI SIEMPRE
| Dispositivo_electrónico = SIEMPRE: F (40.0/18.0)
| Dispositivo_electrónico = CASI SIEMPRE: M (4.0/1.0)
| Dispositivo_electrónico = OCASIONALMENTE: F (0.0)
| Ayuda_casa = SIEMPRE: M (18.0/6.0)
| Ayuda_casa = OCASIONALMENTE: F (1.0)
| Trabajos_equipo = MUY DIFICIL: F (7.0/1.0)
| Materia_facil = INGLES
| Acceso_internet = SIEMPRE: M (2.0)
| Acceso_internet = CASI SIEMPRE: F (2.0)
| Acceso_internet = OCASIONALMENTE: M (1.0)
| Materia_facil = QUIMICA
    
```

Fuente: Elaboración propia.

En la figura 4 al aplicar el algoritmo J48 al término de la pandemia, se observó que el aprendizaje fue regular en la modalidad a distancia, su dominio de las plataformas fue bueno, la interacción con el profesor en su mayoría fue directa y el promedio en general subió.

Fig.4. Resultados J48 al termino la pandemia.

```

J48 pruned tree
-----
Aprendido_distancia = POCO
| Dominio_plataformas = BUENO
| Interacción_profesor = DIRECTA: F (19.0/4.0)
| Interacción_profesor = CONSTANTE: M (2.0)
| Interacción_profesor = OCASIONAL: F (0.0)
| Dominio_plataformas = REGULAR
| Interacción_profesor = DIRECTA
| Ayuda_compañeros = OCASIONALMENTE
| Promedio = MUY ALTO: M (1.0)
| Promedio = ALTO
| Comprende_enseñan = MUCHO
| Pregunta_dudas = CASI SIEMPRE
| Materia_dificil = QUIMICA
| Habilidad_dispositivos = BUENA: M (6.0/1.0)
| Habilidad_dispositivos = REGULAR: F (2.0)
| Habilidad_dispositivos = MALA: M (0.0)
| Materia_dificil = MATEMATICAS
| Ayuda_casa = CASI SIEMPRE
| aprende_presencial = DEMASIADO: M (4.0/1.0)
| aprende_presencial = MUCHO
| Entrega_tareas = A MANO
| Interacción_compañeros = CONSTANTE: F (8.0)
| Interacción_compañeros = DIRECTA
| Resuelve_dudas = SIEMPRE: M (11.0/3.0)
| Resuelve_dudas = CASI SIEMPRE
    
```

Fuente: Elaboración propia.

Dentro de la tabla 2 se observa la comparación de los resultados del algoritmo J48 en los tres tiempos (antes, durante y al término de la pandemia).

Tabla 2. Comparación de resultados J48.

Variables	Antes	Durante	Término
Sexo	49.8%	47.6%	54.1%
Edad	76.5%	75.4%	75.8%
Ayuda_casa	58.8%	57.4%	79.7%
Promedio	92.78%	92.05%	90.25%

Fuente: Elaboración propia.

Una vez analizados los datos, se puede observar que el sexo femenino en la edad de 14 años que recibían mayor ayuda en casa para realizar sus tareas y actividades escolares, fueron las que aumentaron su promedio de regular a alto al término de la pandemia.

En la figura 5 se muestran 3 cluster que surgieron al aplicar SimpleKmeans en donde se observa que hay atributos similares en algunos cluster por ejemplo en todos los cluster coincide que antes de la pandemia la materia difícil es matemáticas y a su vez la materia fácil es historia, en contraste, existe en el atributo medios tareas una diferencia entre el cluster 0 y el 1 en donde los estudiantes del cluster 0 se enfocaron más en los libros con respecto a apuntes que usaron los del cluster 1. A su vez la Ayuda_compañeros los del cluster 0 ocasionalmente pedían ayuda a sus compañeros a diferencia con los del cluster 1 que casi nunca lo hacían. Ayuda_casa los del cluster 0 casi siempre la recibían con respecto a los del cluster 1 que solo ocasionalmente la recibían.

Fig. 5. Resultados SimpleKmeans antes de pandemia.

Attribute	Full Data (277.0)	Cluster#		
		0 (157.0)	1 (73.0)	2 (47.0)
EDAD	CATORCE	CATORCE	CATORCE	CATORCE
Sentia_asistiendo	BIEN	BIEN	BIEN	BIEN
Interacción_compañeros	DIRECTA	DIRECTA	DIRECTA	DIRECTA
Interacción_profesor	DIRECTA	DIRECTA	DIRECTA	DIRECTA
Comprendia	MUCHO	MUCHO	MUCHO	MUCHO
Materia_dificil	MATEMATICAS	MATEMATICAS	MATEMATICAS	MATEMATICAS
Materia_facil	HISTORIA	HISTORIA	HISTORIA	HISTORIA
Preguntaba_dudas	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE
Resolvía_dudas	SIEMPRE	SIEMPRE	SIEMPRE	SIEMPRE
Ayuda_compañeros	OCASIONALMENTE	OCASIONALMENTE	CASI NUNCA	CASI NUNCA
Ayuda_casa	CASI SIEMPRE	CASI SIEMPRE	OCASIONALMENTE	CASI SIEMPRE
Medios_tareas	LIBROS	LIBROS	APUNTES	APUNTES
Entregaba_tareas	A MANO	A MANO	A MANO	A MANO
Reunía_compañeros	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE
Promedio	ALTO	ALTO	ALTO	ALTO

Fuente: Elaboración propia.

En la figura 6 se muestran 3 cluster que surgieron al aplicar SimpleKmeans en donde se observa que hay atributos similares en algunos cluster por ejemplo en todos los cluster coincide que durante la pandemia la materia difícil es matemáticas y la entrega de tareas era por medio de plataformas, así mismo en el atributo medio de tareas todos se enfocaron en internet, en contraste, existe en el atributo sentía distancia una diferencia entre el cluster 0 y el 1 en donde los estudiantes del cluster 0 se sentían desmotivados mientras que los del cluster 1 se sentían bien. A su vez la Ayuda_casa los del cluster 0 casi siempre recibían ayuda a diferencia con los del cluster 1 y 2 quienes siempre la recibían.

Fig. 6. Resultados SimpleKmeans durante la pandemia.

Attribute	Full Data (277.0)	Cluster#		
		0 (210.0)	1 (50.0)	2 (17.0)
EDAD	CATORCE	CATORCE	CATORCE	TRECE
Acceso_internet	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE
Dispositivo_electrónico	SIEMPRE	SIEMPRE	SIEMPRE	SIEMPRE
Medios_digitales	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE	CASI SIEMPRE
Capaz_aprender	REGULAR	REGULAR	REGULAR	REGULAR
Sentía_distancia	DESMOTIVADO	DESMOTIVADO	BIEN	DESMOTIVADO
Interacción_compañeros	OCASIONAL	OCASIONAL	OCASIONAL	OCASIONAL
Interacción_profesor	CONSTANTE	CONSTANTE	CONSTANTE	CONSTANTE
Comprendía_enseñanza	A VECES	A VECES	A VECES	A VECES
Materia_dificil	MATEMATICAS	MATEMATICAS	MATEMATICAS	INGLES
Materia_facil	HISTORIA	HISTORIA	HISTORIA	QUIMICA
Preguntaba_dudas	SIEMPRE	SIEMPRE	SIEMPRE	SIEMPRE
Resolvía_dudas	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE
Ayuda_compañeros	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE
Ayuda_casa	CASI SIEMPRE	CASI SIEMPRE	SIEMPRE	SIEMPRE
Medios_tareas	INTERNET	INTERNET	INTERNET	INTERNET
Entregaba_tareas	PLATAFORMA	PLATAFORMA	PLATAFORMA	PLATAFORMA
Plataformas_aprendió	CLASSROOM	CLASSROOM	CLASSROOM	CLASSROOM
Orientación_escuela	CASI NUNCA	CASI NUNCA	CASI NUNCA	CASI NUNCA
Trabajos_equipo	DIFICIL	DIFICIL	DIFICIL	DIFICIL
Seguro_evaluaciones	REGULAR	REGULAR	REGULAR	REGULAR
Conforme_resultados	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE
Promedio	REGULAR	REGULAR	REGULAR	REGULAR

Fuente: Elaboración propia.

En la figura 7 se muestran 3 cluster que surgieron al aplicar SimpleKmeans en donde se observa que hay atributos similares en algunos cluster por ejemplo en todos los cluster coincide que al término de la pandemia la materia difícil siguió siendo matemáticas y la interacción_profesor fue directa, así mismo aprendió a distancia fue poco, mientras que aprendió mucho en forma presencial, en contraste, existe en el atributo medios tareas una diferencia entre el cluster 0 y el 1 en donde los estudiantes del cluster 0 se enfocaron más en internet con respecto a apuntes que usaron los del cluster 1. A su vez para la entrega de tareas los del cluster 0 lo hacían de forma impresa a diferencia con los del cluster 1 que las entregaban a mano.

Fig.7. Resultados SimpleKmeans al término de la pandemia.

Final cluster centroids:

Attribute	Full Data (277.0)	Cluster# 0 (172.0)	1 (81.0)	2 (24.0)
EDAD	CATORCE	CATORCE	CATORCE	TRECE
Siente_regreso	MOTIVADO	MOTIVADO	MOTIVADO	MOTIVADO
Interaccion_compañeros	DIRECTA	DIRECTA	DIRECTA	CONSTANTE
Interaccion_profesor	DIRECTA	DIRECTA	DIRECTA	DIRECTA
Comprende_enseñan	MUCHO	MUCHO	MUCHO	MUCHO
Materia_dificil	MATEMATICAS	MATEMATICAS	MATEMATICAS	MATEMATICAS
Materia_facil	HISTORIA	HISTORIA	HISTORIA	INGLES
Pregunta_dudas	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE
Resuelve_dudas	CASI SIEMPRE	CASI SIEMPRE	SIEMPRE	CASI SIEMPRE
Ayuda_compañeros	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE
Ayuda_casa	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE	CASI SIEMPRE
Medios_taeras	INTERNET	INTERNET	APUNTES	APUNTES
Entrega_tareas	A MANO	IMPRESAS	A MANO	A MANO
Trabajos_equipo	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE	OCASIONALMENTE
Aprendio_distancia	POCO	POCO	POCO	POCO
aprende_presencial	MUCHO	MUCHO	MUCHO	MUCHO
Dominio_plataformas	REGULAR	REGULAR	REGULAR	REGULAR
Habilidad_dispositivos	BUENA	BUENA	BUENA	BUENA
Promedio	ALTO	ALTO	ALTO	ALTO

Fuente: Elaboración propia.

Dentro de la tabla 3 se observa la comparación de los resultados del algoritmo SimpleKmeans en los tres tiempos (antes, durante y al término de la pandemia).

Tabla 3. Comparación de resultados J48.

Variables	Antes	Durante	Término
Sexo	44.6%	51.6%	45.91%
Edad	58.4%	47.6%	45.1%
Ayuda_casa	43.3%	51.8%	47.2%
Promedio	48.7%	58.4%	49.4%

Fuente: Elaboración propia.

En la figura 8 se observa que haciendo uso del algoritmo de asociación A priori para el término de la pandemia se generaron 10 reglas en donde la confiabilidad en estas fue mayor a 0.98. En estas se destaca que:

1. Los que tuvieron un dominio regular de la plataforma, ayudaron ocasionalmente a sus compañeros.
2. Los que tenían aprendieron mucho en presencial, su comprensión fue mucha.
3. Los que tenían un promedio alto, ayudaban ocasionalmente a sus compañeros.

Fig.8. Resultados Apriori antes de pandemia.

```
Apriori
*****
Minimum support: 0.9 (249 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 2

Generated sets of large itemsets:
Size of set of large itemsets L(1): 7
Size of set of large itemsets L(2): 11
Size of set of large itemsets L(3): 2

Best rules found:
1. Pregunta_dudas=CASI SIEMPRE 259 ==> Interaccion_profesor=DIRECTA 265 <conf:(0.98)> lift:(1.01) lev:(0.0) [2] conv:(0.12)
2. Interaccion_compañeros=DIRECTA Pregunta_dudas=CASI SIEMPRE 259 ==> Interaccion_profesor=DIRECTA 249 <conf:(0.98)> lift:(1.01) lev:(0.0) [1] conv:(0.11)
3. Siente_regreso=CASIONALMENTE 247 ==> Interaccion_compañeros=DIRECTA 261 <conf:(0.98)> lift:(1.0) lev:(0.0) [2] conv:(0.98)
4. Interaccion_profesor=DIRECTA Siente_regreso=CASIONALMENTE 247 ==> Interaccion_compañeros=DIRECTA 259 <conf:(0.98)> lift:(1.0) lev:(0.0) [3] conv:(0.94)
5. Promedio=ALTO 257 ==> Interaccion_profesor=DIRECTA 262 <conf:(0.98)> lift:(1.0) lev:(0.0) [2] conv:(0.98)
6. Interaccion_profesor=DIRECTA 271 ==> Interaccion_compañeros=DIRECTA 268 <conf:(0.98)> lift:(1.0) lev:(0.0) [2] conv:(0.94)
7. Interaccion_compañeros=DIRECTA 271 ==> Interaccion_profesor=DIRECTA 265 <conf:(0.98)> lift:(1.0) lev:(0.0) [2] conv:(0.94)
8. Siente_regreso=CASIONALMENTE 247 ==> Interaccion_profesor=DIRECTA 260 <conf:(0.98)> lift:(1.0) lev:(0.0) [2] conv:(0.92)
9. Interaccion_compañeros=DIRECTA Siente_regreso=CASIONALMENTE 247 ==> Interaccion_profesor=DIRECTA 259 <conf:(0.98)> lift:(1.0) lev:(0.0) [3] conv:(0.81)
10. Pregunta_dudas=CASI SIEMPRE 259 ==> Interaccion_compañeros=DIRECTA 265 <conf:(0.98)> lift:(1.0) lev:(0.0) [2] conv:(0.8)
```

Fuente: Elaboración propia.

En la figura 9 se observa que haciendo uso del algoritmo de asociación A priori para durante la pandemia se generaron 8 reglas en donde la confiabilidad en estas fue de 1. En estas se destaca que los que interactuaban de forma constante con el profesor y aprendían a utilizar la plataforma Classroom entregaban sus tareas.

Fig. 9. Resultados Apriori durante la pandemia.

```
Apriori
*****
Minimum support: 0.95 (243 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 1

Generated sets of large itemsets:
Size of set of large itemsets L(1): 6
Size of set of large itemsets L(2): 14
Size of set of large itemsets L(3): 16
Size of set of large itemsets L(4): 9
Size of set of large itemsets L(5): 2

Best rules found:
1. Entrega_tareas=PLATAFORMA 277 ==> Interaccion_profesor=CONSTANTE 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. Interaccion_profesor=CONSTANTE 277 ==> Entrega_tareas=PLATAFORMA 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. Plataforma_aprendi=CLASSROOM 277 ==> Interaccion_profesor=CONSTANTE 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. Interaccion_profesor=CONSTANTE 277 ==> Plataforma_aprendi=CLASSROOM 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. Plataforma_aprendi=CLASSROOM 277 ==> Entrega_tareas=PLATAFORMA 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. Entrega_tareas=PLATAFORMA 277 ==> Plataforma_aprendi=CLASSROOM 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. Entrega_tareas=PLATAFORMA Plataforma_aprendi=CLASSROOM 277 ==> Interaccion_profesor=CONSTANTE 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. Interaccion_profesor=CONSTANTE Plataforma_aprendi=CLASSROOM 277 ==> Entrega_tareas=PLATAFORMA 277 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
```

Fuente: Elaboración propia.

En la figura 10 se observa que haciendo uso del algoritmo de asociación A priori se generaron 10 reglas en donde la confiabilidad en estas estuvo en un rango de (0.92-0.94). En estas se destaca que:

1. Los que aprendieron poco ocasionalmente recibían ayuda de sus compañeros.
2. Los que preguntaban dudas casi siempre interactuaban directamente con el profesor.
3. Los que tenían interacción directa con sus compañeros y preguntaban dudas a su vez interactuaban directamente con el profesor.
4. Los que interactuaban de forma directa con el profesor y se reunían ocasionalmente con sus compañeros tenían una interacción directa con sus compañeros.
5. Los que tenían un promedio alto, tenían una interacción directa con el profesor.

Fig. 10. Resultados Apriori al término de la pandemia.

```
Apriori
*****
Minimum support: 0.9 (222 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 4

Generated sets of large itemsets:
Size of set of large itemsets L(1): 9
Size of set of large itemsets L(2): 17

Best rules found:
1. Aprendio_distancia=POCO 255 ==> Ayuda_compañeros=OCASIONALMENTE 240 <conf:(0.94)> lift:(1) lev:(0) [0] conv:(0.98)
2. Dominio_plataformas=REGULAR 284 ==> Ayuda_compañeros=OCASIONALMENTE 239 <conf:(0.94)> lift:(1) lev:(0) [0] conv:(0.96)
3. Comprende_enseñan=MUCHO 253 ==> Ayuda_compañeros=OCASIONALMENTE 238 <conf:(0.94)> lift:(1) lev:(0) [0] conv:(0.97)
4. Interaccion_profesor=DIRECTA 239 ==> Ayuda_compañeros=OCASIONALMENTE 224 <conf:(0.94)> lift:(1) lev:(0) [0] conv:(0.92)
5. aprende_presencial=MUCHO 252 ==> Ayuda_compañeros=OCASIONALMENTE 239 <conf:(0.93)> lift:(0.99) lev:(-0.0) [-1] conv:(0.96)
6. Promedio=ALTO 250 ==> Ayuda_compañeros=OCASIONALMENTE 235 <conf:(0.93)> lift:(0.99) lev:(-0.0) [-1] conv:(0.95)
7. Interaccion_profesor=DIRECTA 239 ==> Aprendio_distancia=POCO 222 <conf:(0.93)> lift:(1.0) lev:(0.0) [1] conv:(1.05)
8. Ayuda_compañeros=OCASIONALMENTE 260 ==> Aprendio_distancia=POCO 240 <conf:(0.92)> lift:(1) lev:(0) [0] conv:(0.98)
9. Comprende_enseñan=MUCHO 253 ==> Dominio_plataformas=REGULAR 233 <conf:(0.92)> lift:(1) lev:(0) [1] conv:(1)
10. aprende_presencial=MUCHO 252 ==> Comprende_enseñan=MUCHO 232 <conf:(0.92)> lift:(1.0) lev:(0.0) [1] conv:(1.04)
```

Fuente: Elaboración propia.

Dentro de la tabla 4 se observa la comparación de los resultados del algoritmo Apriori en los tres tiempos.

Tabla 4. Comparación de resultados Apriori.

VARIABLES	Antes	Durante	Término
Sexo	90%	95%	80%
Edad	90%	95%	80%
Ayuda_casa	90%	95%	80%
Promedio	90%	95%	80%

Fuente: Elaboración propia.

3. CONCLUSIONES

Conforme a los resultados obtenidos se observa un mayor y mejor aprendizaje alcanzado al regresar a clases presenciales, mostrando un aumento del promedio general en 250 alumnos, arrojando que durante las clases en línea los alumnos aprendieron poco y se reflejó en sus promedios, dado a que, más del 50% de la muestra no contaba con acceso a internet, dispositivos electrónicos o medios digitales para realizar sus tareas constantemente, también nos muestra que en su mayoría el sexo femenino se sentía motivado durante las clases en línea, su comunicación era constante con su profesor y compañeros, de la misma forma el sexo masculino logro una buena habilidad en el dominio de los dispositivos electrónicos, pero su dominio de las plataformas para entregar sus actividades fue regular debido a que no contaron con una orientación para su uso por parte de su institución escolar.

Las evaluaciones realizadas nos muestran que la mayoría de los alumnos durante las clases a modalidad distancia siempre preguntaban las dudas que tenían sobre los temas vistos en clase, siendo Matemáticas la materia clasificada como más difícil de aprender, mientras que por su parte Historia se clasifico como la más fácil.

La plataforma más utilizada fue classroom, en cuanto a la comprensión de los temas, el análisis arrojó un incremento al regresar a clases presenciales.

Durante el análisis realizado se observó lo siguiente; la clasificación mejor ponderada fue en la base de datos del término de la pandemia, la mejor agrupación fue en la base de datos durante la pandemia y la asociación con mayor confiabilidad fue en la base de datos durante la pandemia.

En base a la métrica de datos correctamente analizados, concluimos que el algoritmo de asociación A priori fue el mas optimo para este trabajo.

De manera general y con la muestra analizada se encuentran diferencias en los cluster y reglas arrojadas en las diferentes etapas de análisis, mostrando en la virtual más diferencias, esto posiblemente por todo lo que implicó el cambio de paradigma en la forma de se impartía clase, sin embargo, a dado pie a adaptar y mejorar estas herramientas para la educación virtual o mixta.

4. REFERENCIAS

[1] Quiroz, F. M. B. (2022c, agosto 18). “Consecuencias del desarrollo en la educación en tiempos de post pandemia”. <https://www.cienciayeducacion.com/index.php/journal/article/view/149>

[2] Andrés, S., Rojas, C., Yaneth, D. C., & Becerra, R. (s/f). “La educación en tiempos de pandemia”. Edu.co

https://repository.unab.edu.co/bitstream/handle/20.500.12749/17566/2022_Tesis_Sergio_Andres_Castillo_Rojas.pdf?sequence=1

[3] ALTBACH P. y de WIT H. (marzo 25, 2020). “El impacto del coronavirus en la educación superior”. Mérida, México.: <https://educacion.nexos.com.mx/>

[4] “Vista de impacto de la pandemia en la calidad del sistema educativo colombiano” | Ciencia Latina Revista científica Multidisciplinar.(s. f.). <https://ciencialatina.org/index.php/cienciala/article/view/8178/12348>

[5] Gutiérrez, P. L., Gutiérrez, C. D. (2020) “Descubrimiento de conocimiento en incidencia de tipo de cáncer para pacientes terminales mediante minería de datos”. Ciencia Ergo-Sum. (28) (1-10).

[6] Dávila, E.M.A (2020) “Actividades docentes online y el desarrollo de la empatía en la fase de aislamiento de la pandemia covid 19”. <https://repository.uta.edu.ec/jspui/bitstream/123456789/31870/1/Tesis%20Ana%20Da%CC%81vila%20.pdf>

[7] Timarán Pereira, R. (2009). “La minería de datos en el descubrimiento de perfiles de deserción estudiantil en la Universidad De Nariño”. *Universidad Y Salud*. 1(11). <https://revistas.udenar.edu.co/index.php/usalud/article/view/218>

[8] Angélica, V. S. G. (s. f.). “Educación en tiempos de pandemia: una revisión bibliográfica”. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442021000300166

[9] Gervilla García, E., José Montaña Moreno, J., Sesé Abad, A., Palmer Pol, A., & Palmer, A. (s/f). “Sociedad Científica Española de Estudios sobre el Alcohol, el Alcoholismo y las otras Toxicomanías”. Palma de Mallorca, España. Redalyc.org. Recuperado el 19 de enero de 2024, de <https://www.redalyc.org/pdf/2891/289122882009.pdf>

[10] Digitales, S. (2021, 14 mayo). “Breve explicación del proceso KDD” - Laboratorio de Certificación. Laboratorio de Certificación. <https://www.laboratoriodecertificacion.es/breve-explicacion-del-proceso-kdd/>

[11] Neural Networks Chile. (2023, 15 marzo). “Cómo los procesos KDD pueden ayudarte a descubrir información en sus datos.” - Neural networks. Neural Networks - Aprende, Optimiza, Evolucionar. <https://neuralnetworks.cl/como-los-procesos-kdd-pueden-ayudarlo-a-descubrir-informacion-en-sus-datos/#:~:text=La%20limpieza%20de%20datos%20es,en%20el%20proceso%20de%20KDD>

[12] Chartre, F. (s. f.). “Cómo es el proceso de extraer conocimiento a partir de bases de datos” - campusMVP.es. campusMVP.es. <https://www.campusmvp.es/recursos/post/el-proceso-de-extraccion-de-conocimiento-a-partir-de-bases-de-datos.aspx>

[13] Landa, J. (s. f.). “¿Qué es KDD y minería de datos?” – Javier (Paco) Landa. <https://fcojlanda.me/es/ciencia-de-los-datos/kdd-y-mineria-de-datos-espanol/>

[14] Admin. (2020, 6 abril). “¿Qué es el KDD o proceso de Descubrimiento de Conocimiento?” DiagramasUML.com. https://diagramasuml.com/que-es-el-kdd-o-proceso-de-descubrimiento-de-conocimiento/#google_vignette

[15] Riquelme C. J., Ruiz R., Gilbert Karina (2006) “Minería de Datos: Conceptos y Tendencias” Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial (10).

[16] “¿Qué es Weka y qué tiene que ver con Big Data?” (2020, julio 9). Agenciab12.mx. <https://agenciab12.mx/noticia/que-es-weka-que-tiene-que-ver-big-data>