

Minería de datos educativa para determinar la influencia de las redes sociales en el rendimiento académico de estudiantes universitarios

Nayo Salinas, Luis Chamba-Eras
 Carrera de Ingeniería en Sistemas, Grupo de Investigación en Tecnologías de la Información y Computación
 Universidad Nacional de Loja
 Loja, Ecuador
 {nfsalinasm,lachamba}@unl.edu.ec

Abstract—El objetivo de este trabajo es identificar (patrones) por medio de la minería de datos educativa (MDE), la influencia de las redes sociales (RS) en el rendimiento académico de 438 estudiantes universitarios, para ello, se utilizó la metodología CRISP-DM, en conjunto con el algoritmo “A priori”, para las reglas de asociación, y se realizó el agrupamiento mediante el algoritmo “K-means”, obteniendo como resultados que los estudiantes universitarios tienen cierto grado de influencia en el uso de las RS, debido que utilizan para ocio, distracción y para publicar memes o chistes, además, siempre se conectan a las RS desde la universidad y hogar, finalmente, se identificó que el uso no adecuado del teléfono celular es otro factor que afecta el rendimiento académico.

Keywords— educational, data mining, artificial intelligence

I. INTRODUCCIÓN

En la actualidad la tecnología se ha venido introduciendo en la vida cotidiana de los estudiantes, por ello las universidades demandan un nuevo tipo de enseñanza. Las nuevas herramientas tecnológicas (redes sociales, blogs, plataformas de vídeo), les han dado el poder de compartir, crear, informar y comunicarse, convirtiéndose en un elemento esencial en sus vidas [1].

Las Redes Sociales (RS) se han universalizado, los jóvenes las han incorporado plenamente en sus vidas. Se han convertido en un espacio idóneo para intercambiar información y conocimiento de una forma rápida, sencilla y cómoda. En la actualidad, los estudiantes utilizan constantemente las RS, ocasionando que capten más su atención ante cualquier otra actividad que estén realizando. Los estudios relacionados al uso de las redes sociales en la academia han sido escasos o bien se han hecho, no se han apoyado en las nuevas tecnologías para análisis de datos para su desarrollo [2].

La aplicación de la IA en la educación permite cambiar la forma de enseñar; por un lado se puede automatizar las tareas administrativas del docente, dándole un nuevo significado al proceso de enseñanza aprendizaje; por otra parte para los estudiantes, la producción de software que brinde educación personalizada, detectando los temas que necesitan más trabajo en clases [3]; así mismo la IA puede realizar cambios en la búsqueda e interacción del estudiante con la información, mediante el uso de sistemas de manera más inteligente, que permitan enseñar y apoyar al estudiante. En la década de los 80 aparece la Minería de Datos o Data Mining (DM), es considerada como la ciencia que se encarga del conocimiento mediante el análisis inteligente de datos, guiándose en un conjunto de modelos matemáticos y algoritmos que trabajan conjuntamente con la estadística y la IA [4]; permitiendo así brindar ayuda a la toma de decisiones en diversas áreas.

II. DISEÑO EXPERIMENTAL

A. Minería de datos educativa

La minería de datos educativa (EDM) ha surgido como un dominio independiente de investigación. La EDM se refiere al “desarrollo, investigación, y aplicación de métodos computarizados para detectar patrones en grandes colecciones de datos educacionales que de otro modo sería muy difícil o imposible analizar debido al enorme volumen de datos existentes” [5].

B. Técnicas de minería de datos educativa

En referencia, a los estudios [3], [5]–[8] indican algunas técnicas para aplicar a problemas educativos. En la TABLA I, se muestra el análisis de varios problemas en el contexto educativo que son utilizadas a través de técnicas de minería de datos.

TABLE I. TÉCNICAS APLICADAS EN LA EDM

Categoría de estudio de EDM	Descripción	Técnicas empleadas
Predecir el rendimiento estudiantil	Descubrir características de los estudiantes.	Clustering
Modelamiento de estudiantes	Crear modelos cognitivos de los estudiantes.	Regresión
Detección de comportamientos indeseables de los estudiantes	Detección de estudiantes que poseen problemas o comportamiento inusual.	Clasificación
Agrupamiento de estudiantes	Clasificación en grupos de alumnos de acuerdo a sus características personales.	Reglas de asociación
Análisis de redes sociales.	Analizar las relaciones entre estudiantes.	Filtrado colaborativo
Recomendaciones para estudiantes	Dar recomendaciones directamente a los estudiantes, actividades a realizar, enlaces a visitar, etc.	Reglas de asociación. Clustering. Patrones secuenciales.
Retroalimentación para apoyo a docentes	Información para toma de decisiones y retroalimentación a los docentes.	Estadística, Técnicas de visualización de información, Reglas de asociación, Clustering, Clasificación.

En definitiva, se puede decir que las técnicas de Clustering y reglas de asociación son las que mejor se adaptan a la investigación debido a que esas técnicas permiten identificar el rendimiento estudiantil y comportamiento de los

estudiantes, además, porque se adecúa a los tipos de datos (nominal) que tiene el dataset.

En la TABLA II, existe diferentes técnicas de Clustering y reglas de asociación, se describe que el algoritmo k-means [9], [10], en la actualidad es el más utilizado en la aplicación de minería de datos, según los estudios realizados[9]–[13], el algoritmo k-means ofrecen mejores resultados que los diferentes algoritmos de clustering, además porque mejor se adapta al tipo de datos que tiene al dataset, al igual que el algoritmo A priori.

TABLE II. ALGORITMOS DE EDM

Algoritmos	Técnicas de EDM	
	Clustering	Asociación
A priori		X
K- Means	X	
CN2 Rules		X
Cure	X	
Arboles de decisión (J48)	X	
Mapas de Kohonen	X	
EM	X	
Dbscan	X	
Optics	X	
Knnclust	X	
Svm Fuzy	X	
SNN	X	
DIANA	X	

C. Influencia de las redes sociales

Según los estudios [2], [14]–[19]muestran la influencia positiva o negativa de las redes sociales en la educación, en la TABLA III se detalla.

TABLE III. INFLUENCIA DE LAS REDES SOCIALES

Influencia Positiva	Influencia Negativa
YouTube para mirar tutoriales con fines educativos.	Es la distracción que causa, pero no hasta el punto de arriesgar la influencia negativa de las Redes sociales.
Las RS (facebook) ayudan a conectar a personas con limitaciones para interactuar directamente (timidez, limitaciones económicas, limitaciones físicas, distancias muy largas entre otro).	Influye de manera negativamente el rendimiento académico debido a los estudiantes consume mucho tiempo en las redes sociales.
Permite interactuar con diferentes expertos y personas con conocimientos sobre una temática en específica.	Falta de formación del profesado en modelos pedagógicos en la aplicación didáctica de las redes sociales.
Permite a los estudiantes a acceder a la información fácilmente y compartirlos en cualquier momento y en cualquier lugar.	Se publica cualquier cosa sin previa verificación del contenido que se está publicado o compartiendo.
Permite conocer información sobre cursos, capacitaciones o eventos.	
Las redes sociales (Facebook, Twitter, YouTube, WhatsApp) forman parte de sus vidas y de sus tareas cotidianas; con ellas, está prácticamente conectada durante todo el día.	
Las redes sociales en su proceso educativo es una realidad, debido a que al día dedican parte de su tiempo conectados a alguna red social, permitiendo resolver inquietud expuestos en el aula, o para obtener información.	

D. Algoritmos

La aplicación de algoritmos de minería ha permitido detectar patrones en los datos y, por ende, crear modelos que sustenten la toma de decisiones, y así contribuir al mejoramiento de un problema en particular. Estos algoritmos son operaciones codificadas como un conglomerado finito de reglas que toman los datos como entradas y sus salidas como modelos o patrones [20].

1) *K-means*: El algoritmo de K-means fue propuesto por MacQueen en el año 1968 [21]. Agrupa un conjunto de datos en un número predefinido de clústeres. Comienza con un conjunto aleatorio de centroides de cada uno de los clústeres y continúa reasignando los datos del conjunto de datos a los centroides, basándose en la similitud entre el dato y el centroide. El proceso de reasignación no se detiene hasta que se converge al criterio de parada. El algoritmo k-means se pueden identificar cuatro pasos [21], [22]:

a) *Inicialización*: Se define un conjunto de objetos a particionar, el número de grupos y un centroide por cada grupo. Algunas implementaciones del algoritmo estándar determinan los centroides iniciales de forma aleatoria; mientras que algunos otros procesan los datos y determinan los centroides mediante de cálculos.

b) *Clasificación*: Se define Para cada objeto de la base de datos, se calcula su distancia a cada centroide, se determina el centroide más cercano, y el objeto es incorporado al grupo relacionado con ese centroide.

c) *Cálculo de centroides*: Para cada grupo generado en el paso anterior se vuelve a calcular su centroide.

d) *Condición de convergencia*: Se han usado varias condiciones de convergencia, de las cuales las más utilizadas son las siguientes: converger cuando alcanza un número de iteraciones dado, converger cuando no existe un intercambio de objetos entre los grupos, o converger cuando la diferencia entre los centroides de dos iteraciones consecutivas es más pequeño que un umbral dado. Si la condición de convergencia no se satisface, se repiten los pasos dos, tres y cuatro del algoritmo.

2) *Regla de asociación Apriori*: El algoritmo Apriori es el algoritmo de reglas de asociación más utilizado, busca posibles relaciones entre elementos para descubrir hechos que ocurren dentro de un conjunto de datos [3], [23] y para ello identifica conjuntos de elementos frecuentes de donde luego se derivan las reglas de asociación Este algoritmo considera cada posible combinación de pares atributo-valor y se denomina cada par como ítem y a su vez el conjunto de ítems se denomina ítemsets. Para obtener resultados óptimos se debe establecer la condición de confianza para eliminar ítem-sets con un nivel por debajo del valor establecido. Otro aspecto a tener en cuenta, es que un ítem-set puede dar más de una regla de asociación, así como también ninguna. Las reglas de asociación ligan cualquier atributo y no solamente las clases de un conjunto de datos [3] por lo que presentan las siguientes medidas de fiabilidad: soporte, soporte absoluto, confianza, elevación, soporte multiplicado por confianza y grupo de reglas [24]. Este algoritmo proporciono relaciones entre las palabras basándose en reglas, por lo cual también se obtuvo un porcentaje de confianza por cada regla obtenida.

Las reglas de Asociación nos permiten:

- Encontrar las combinaciones de artículos o ítems que ocurren con mayor frecuencia en una base de datos transaccional.
- Medir la fuerza e importancia de estas combinaciones.

Las podemos representar como:

a) *Lista*: Cada fila representa una transacción, cada fila lista los ítems comprados por el consumidor, las filas pueden tener un número diferentes de columnas.

b) *Representación vertical*: Representa dos columnas, primera indica el numero o identificador y la segunda el ítem presente, es una forma más eficiente de guardar datos.

c) *Representación Horizontal*: Representa cada fila una transacción, cada columna representa un artículo o ítem, si un ítem está presente en una transacción lo marcamos con un 1 caso contrario lo marcamos con un cero.

E. Participantes

La investigación se desarrolló en la Universidad Nacional de Loja, con una muestra de estudiantes de la Facultad de la Energía (Ing. en Sistemas / Computación, Ing. en Geología Ambiental y Ordenamiento Territorial / Minas, Ing. en Electrónica y Telecomunicaciones / Telecomunicaciones, Ing. en Electromecánica, y Ing. en Mecánica Automotriz), matriculados en el periodo octubre 2018 - marzo 2019 (población de 1300 estudiantes). Se aplicó un muestro probabilístico (Ecuación (1)), obteniendo una muestra de 438 estudiantes, a los cuales se les aplicó una encuesta que permitió recolectar datos y con ello elaborar el dataset.

III. RESULTADOS

1) *Fase 1. Compresión de los datos*: Los datos o información fueron recolectados en la Universidad Nacional de Loja matriculados en el periodo octubre 2018 - marzo 2019. Se usó un muestreo probabilístico donde se aplicó el instrumentó de cuestionario de las redes sociales en la Facultad la Energía.

Para la obtención del tamaño de la muestra, se utilizó un análisis probabilístico [2] debido a que la población es finita con un nivel de confianza del 95% y un margen de error del 5%. Utilizando la siguiente fórmula.

$$n = \frac{Z^2 p.q.N}{Ne^2 + Z^2.p.q} \quad (1)$$

Donde:

Z= Nivel de confianza

n= Tamaño de la muestra

N= Tamaño de la población o universo

p=Población a favor

q= Población a contra

e= Error de muestra

a) *Recolección de datos*: En el levantamiento de la información se obtuvo 442 muestras de forma aleatoria a los estudiantes de la Facultad, distribuidos en la TABLA IV, la encuesta fue elaborada y aplicada manualmente basándose en

los estudios [25]–[31] y conformado por las siguientes variables: datos demográficos, frecuencia de uso de las redes sociales, motivación del uso de las redes sociales, uso educativo, afectaciones por el mal uso de las redes sociales, lugar de uso, problemas de las redes sociales y dispositivo de conexión a Internet.

TABLE IV. NÚMERO DE ENCUESTAS APLICADAS POR CARRERA

Carrera	Nº de Encuestas
Ing. en Sistemas / Computación.	115
Ing. en Geología Ambiental y Ordenamiento Territorial / Minas.	98
Ing. en Electrónica y Telecomunicaciones / Telecomunicaciones.	104
Ing. en Electromecánica	101
Ing. en Mecánica Automotriz	24
Total	442

b) *Descripción de los datos*: A continuación, detallamos cada una de las variables que se utilizaron en la investigación.

- Datos demográficos: Permitted obtener la edad, género, carrera que pertenecen los estudiantes.
- Uso de las redes sociales: Permitted determinar las redes sociales que utilizan, además se indagó para que utilizan las RS y tiempo que están diariamente conectados.
- Motivo de uso de las redes sociales: Se indagó las actividades que realizan en las RS como para estar contacto con amigo y familiares, estar al día en eventos, noticias o novedades, además para conocer si utilizan para distracción o para publicar contenido.
- Uso educativo: En esta variable, para el uso educativo se dividen en cuatro componentes que se detallan en la TABLA V.

TABLE V. DESCRIPCIÓN DE LA VARIABLE USO EDUCATIVO

Campo	Nº de Test
Motivador	Participar en grupos o páginas en RS o para promover el uso de activo RS para fines educativos
Informativo	Utilizan para el intercambio de documentos entre estudiantes
Comunicativo	Para una comunicación fluida entre compañeros y docente
Formativo	Para trabajar en equipo o como herramienta de aula virtual

- Afectaciones por el mal uso redes sociales: Dentro de la variable afectaciones por el mal uso de las redes sociales se dividen en dos bloques, la primera para averiguar si existe afectaciones en actividades recreativas como hacer ejercicios, caminar, hacer deporte, etc. La segunda para actividades académicas.
- Lugar de uso: Conocer el lugar de uso de las redes sociales (Universidad, Hogar, Sitios públicos, Cyber, Megas).
- Dispositivos en que se conectan: Celular, Portátil, Pc escritorio, Tablet.

2) *Fase 2. Preparación de los datos:* En esta fase consistió en analizar los datos erróneos, faltantes o en blanco para obtener el dataset final así también uniformar las variables para su posterior modelado en la siguiente fase.

Se procedió a transformar las variables cualitativas a variables cuantitativas, para la herramienta Weka se realizó el filtro (StringToNominal) a todo el dataset, para ello se usó el atributo no supervisado en Weka, configurando 4 rangos (bins) y el useEqualFrequency en True para obtener rangos con el mismo número de frecuencias. En la herramienta rapidminer se aplicó el atributo normalize para que las variables de diferentes unidades o escalas operen al mismo rango sin importar el tipo de datos:

a) *Selección de los datos:* Los datos seleccionados para proceso de minería de datos fueron la información recolectada en la aplicación manual del instrumento de las redes sociales y fueron tabulados en archivo plano en formato .csv, dando como resultado 24 variables que se especifican en la TABLA VI.

TABLE VI. VARIABLE QUE CONFORMA EL DATASET

Nº	Variable	Tipo de dato
1	Género	Nominal
2	Edad	Nominal
3	Carrera	Nominal
4	Frecuencia de uso de las redes sociales	Nominal
5	Tipo de redes sociales	Nominal
6	Tiempo en RS	Nominal
7	Contexto de uso de redes sociales	Nominal
8	Contacto amigo/ Familiares	Nominal
9	Estar al día en eventos o noticias	Nominal
10	Para publicación y distracción	Nominal
11	Motivador	Nominal
12	Informativo	Nominal
13	Comunicativo	Nominal
14	Formativo	Nominal
15	Afectaciones a actividades recreativas	Nominal
16	Afectaciones a actividades educativas	Nominal
17	Usa las RS en la Universidad	Nominal
18	Usa las RS en la hogar	Nominal
19	Usa las RS en la sitios públicos	Nominal
20	Usa las RS en un cyber	Nominal
21	Usa las RS mediante un plan de datos	Nominal
22	Adicción por las RS	Nominal
23	Dispositivo celular	Nominal
24	Dispositivo computador portátil	Nominal

b) *Limpieza datos:* Para la limpieza de los datos se realizó en base al estudio [32], donde es caso existir inconsistencia de los datos que pueden ser por: valores no

existentes, datos incompletos, caracteres inválidos, etc., menciona que se puede aplicar la siguiente limpieza de los datos: ignorar los datos, eliminar el campo o reemplazado el dato.

La Fig.1, se observa que el encuestado a seleccionada sola la primera opción en toda la encuesta por lo cual se procedió a eliminar el instrumento de la recolección, mientras que en la Fig.2 existen campos vacíos donde se realizó el mismo proceso de la Fig.1.

	5	6	7	8	9	10	1
Rara vez	Casi Siempre	Siempre	Rara vez	Algunas veces	Rara vez	Rara vez	
Nunca	Rara vez	Rara vez	Rara vez	Siempre	Casi siempre	Rara vez	
Siempre	Siempre	Siempre	Siempre	Siempre	Siempre	Siempre	
Rara vez	Rara vez	Casi siempre	Algunas veces	Siempre	Casi siempre	Siempre	
Nunca	Rara vez	Casi siempre	Casi siempre	Siempre	Siempre	Siempre	

Fig. 1. Datos basura

Siempre	Casi siempre	Rara vez	Algunas veces	Casi siempre
Siempre	Rara vez	Casi Siempre	Siempre	Siempre
Algunas veces	Nunca		Casi siempre	Rara vez
Rara vez	Rara vez	Algunas veces	Siempre	Nunca
Casi siempre	Algunas veces	Casi Siempre		Casi siempre
Algunas veces	Nunca	Rara vez	Algunas veces	
Algunas veces	Rara vez	Algunas veces	Algunas veces	Rara vez
	Algunas veces	Algunas veces	Siempre	Casi siempre

Fig. 2. Datos incompletos o vacíos

Al finalizar el proceso de limpieza de los datos, las 442 encuestas aplicadas se redujeron a 438

c) *Construcción e integración de los datos:* Se elaboró el dataset y se procedió a elaborar dos hipótesis que serán utilizados en la siguiente fase.

- Obtener los patrones que influyen por el uso de las redes sociales en el rendimiento académico (hipótesis 1).
- Determinar las actividades (Motivador, Informativo, Comunicativo, Formativo) que realizan para el uso educativo de las redes sociales (hipótesis 2).

3) *Fase 3. Modelado:* En esta fase es primordial dentro del proceso CRISP-DM ya que aquí se genera el conocimiento a través de la aplicación de técnicas y algoritmo de minería de datos.

Una vez obtenidos el dataset, y cargado los datos en la herramienta

a) *Generación del plan de prueba:* Para validar el modelo de clustering se utilizó los parámetros de configuración (ver TABLA VII) teniendo en cuenta el número de cluter y semilla, se aplicó en cada uno de los experimentos, para posterior evaluar el experimento de mayor relevancia para la investigación. Para las reglas de asociación se consideraron las 20 primeras reglas de asociación siempre y cuando la precisión sea superior al 0,95 y que sean relevantes para el objeto de estudio.

TABLE VII. CONFIGURACIÓN DEL ALGORITMO CLUSTERING

Parámetro	Descripción	Valor
DisplayStdDevs	Permite mostrar los valores de la división estándar de los valores numéricos	False
DistanceFuntion	Calcula las distancias entre los valores que tiene por defecto.	Euclidean distance funtion
MaxIterations	El máximo número de interacciones del algoritmo.	500
NumClusters	Número de cluster.	NumClusters (2,3 4)
Seed (semilla)	Número de semilla (Para elegir el número de semilla se basa en los experimentos, tomando en cuenta el menor error cuadrático).	Seed 10, Seed 15, Seed 20
Split	Se define un porcentaje de entrenamiento el modelo, y el porcentaje restantes se evalúa el modelo.	El 66% para entrenamiento y el resto para prueba (34%)

b) Estadísticos descriptivos

Redes sociales más utilizadas

Facebook es la RS que más utilizan (ver Fig. 3), seguido de YouTube y WhatsApp y la que menos usan es Twitter.

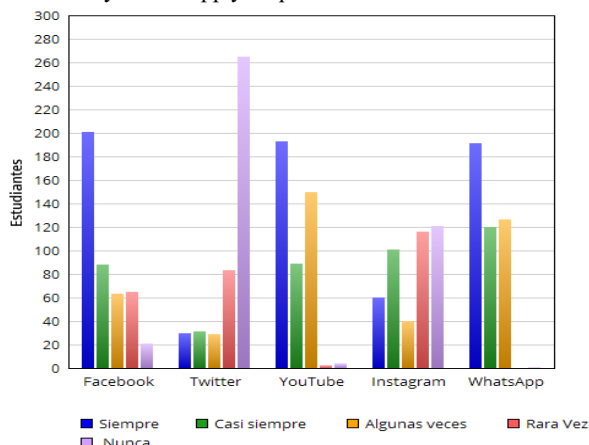


Fig. 3. Redes sociales más utilizadas en la Facultad de la Energía

Tiempo de uso de las redes sociales

El tiempo promedio diariamente que utilizan Facebook es de 2 a 4 horas, mientras Twitter lo usan menos de una hora, YouTube más de 4 horas (Ver Fig. 4).

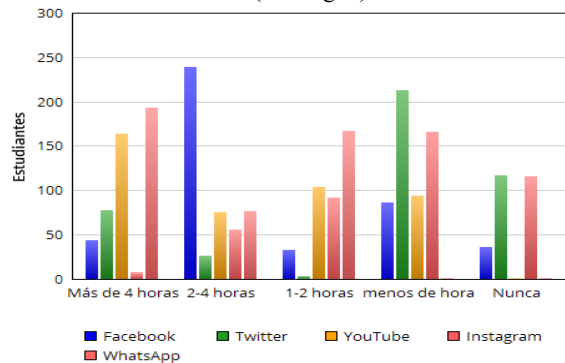


Fig. 4. Tiempo que usan las RS

Lugar de uso de las redes sociales

En el campus universitario es el lugar donde más utilizan las RS, seguido por el hogar mientras que sitios públicos y Cyber son lugares que menos utilizan (ver Fig. 5).

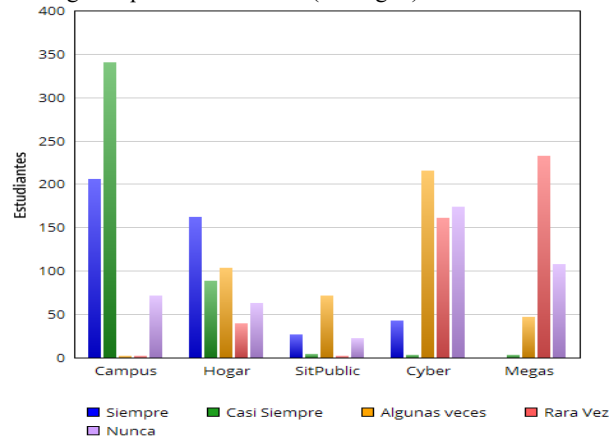


Fig. 5. Lugares de uso de las redes sociales

Relación tipo de redes sociales - para que lo utilizan

En la Fig. 6, observamos que Facebook, YouTube, Instagram y WhatsApp son las redes sociales que utilizan más para Social y ocio mientras que para el uso educativo además de las menciona anteriormente utilizan también Twitter, pero con menor proporción de uso.

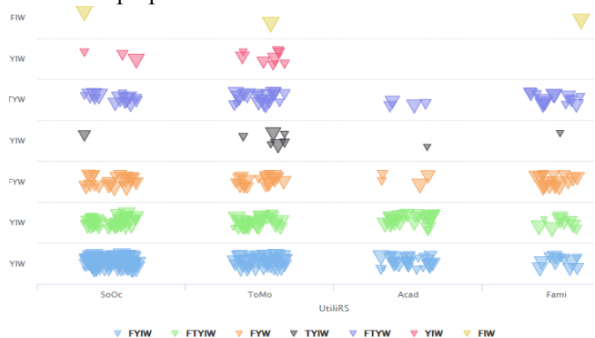


Fig. 6. Relación tipo de redes sociales - para que lo utilizan

Relación tiempo en RS- para que utilizan RS

En la Fig. 7, observamos que Facebook, YouTube, Instagram y WhatsApp son las RS que más utiliza con un promedio de más de 4 horas a diario, especialmente para Social y ocio mientras YouTube y WhatsApp disfrutan el uso RS en el ámbito educativo.

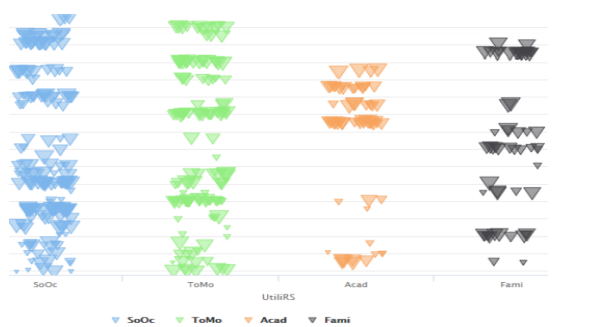


Fig. 7. Relación tiempo en RS- Para que utilizán RS

Relación dispositivo - uso educativo

La Fig. 8, 9, 10 muestra que algunas veces (informativo, comunicativa, formativa), usan las RS como herramienta educativa y siempre se conectan desde el celular o computadora portátil, por lo cual se puede concluir que las RS en el ámbito educativo lo utilizan poco uso.



Fig. 8. Relación dispositivo –Uso educativo (Informativo)



Fig. 9. Relación dispositivo –Uso educativo (Comunicativo)

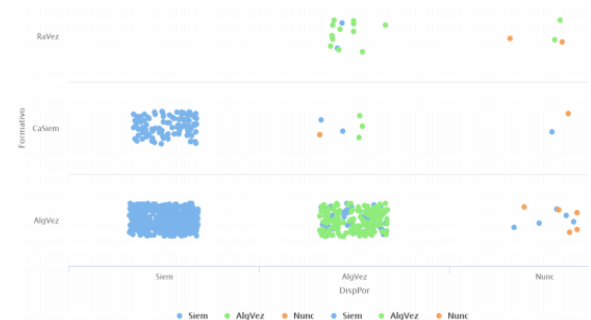


Fig. 10. Relación dispositivo –Uso educativo (Formativo)

c) *Resultados de la minería de datos:* Los experimentos se realizaron en la herramienta Weka y Rapidminer, a continuación, se detalle los resultados obtenidos en cada una de las hipótesis.

Hipótesis 1

Algoritmo K-means

En la TABLA VIII, se muestra los patrones obtenidos en cada uno de los clústeres, a continuación, se detalla.

Clúster 0 (27.40%): se observa que existe una afectación por el uso de las RS (casi simple), debido a que siempre pasa conectado a las RS siendo WhatsApp y YouTube son las RS

que más utiliza con promedio de más de 4 horas, seguido de Facebook con un promedio de 1 a 2 horas principalmente utilizan para estar el día en novedades, noticias, novedades y para estar en contacto entre amigos y familiares, en la universidad y hogar, el lugar donde más se conectan a las RS y siempre se conecta a través del celular.

Clúster 1 (35.84%): algunas veces se ven afectado el rendimiento académico debido que utilizan Facebook, y WhatsApp aproximadamente de dos a cuatro horas diariamente a las RS, seguido de YouTube con un promedio de 1 a dos horas, principalmente emplean el tiempo para ocio, donde utilizan siempre para estar al día en novedades, noticias y novedades, además algunas veces la usan para estar en contacto con amigos y familiares y para publicar y distracción, en la Universidad, hogar y sitios públicos es el lugar donde siempre se conecta principalmente a través de los dispositivos como el celular y computadora.

Clúster 2 (36.76%): se examina que siempre pasan conectados a las RS donde Instagram es la red que más utiliza con un tiempo de dos a cuatro horas, seguido por Facebook, YouTube y WhatsApp con un promedio de uno a dos horas diariamente, por lo cual se ven afectados las actividades académico y recreativas utilizan principalmente para chatear con amigos y familiares, siempre se conectan desde la universidad y hogar a través de dispositivo celular, así como la computadora.

TABLE VIII. EJECUCIÓN DEL ALGORITMO K-MEANS HIPÓTESIS 1

Atributo	Cluster 0	Cluster 1	Cluster 2
Utiliza RS	Siempre		
Tipo de RS	Facebook (F), YouTube (Y), Instagram (I), WhatsApp (W)	Facebook, YouTube, Instagram, WhatsApp	Facebook, YouTube, Instagram, WhatsApp
Tiempo en RS	F: 1-2 h Y: 2-4 h I: menos h W: más 4 h	F: 2-4 h Y: más 4 h I: menos h W: 2-4 h	F: 1-2 h Y: 1-2 h I 2-4 h W: 1-2 h
Actividades	Social y ocio	Social y ocio	Todo Motivo
Utiliza RS para contacto con amigos o familia	Algunas veces	Algunas veces	Algunas veces
Utiliza RS para estar al día en eventos o noticias	Siempre	Siempre	Rara vez
Utiliza RS para publicar y distracción	Rara vez	Algunas veces	Rara vez
Afecta RS actividades recreativas	Rara vez	Rara vez	Algunas veces
Afecta RS actividades educativas	Casi siempre	Algunas veces	Algunas veces
Univ. lugar de uso RS	Siempre	Siempre	Siempre
Hogar lugar de uso RS	Siempre	Siempre	Siempre
Sitios públicos lugar de uso RS	Algunas veces	Siempre	Algunas veces
Cyber lugar de uso RS	Nunca	Algunas veces	Nunca
Mediante plan de datos	Algunas veces	Algunas veces	Nunca
Celular a RS	Siempre	Siempre	Siempre

Utiliza computadora para conectarse a RS	Algunas veces	Siempre	Siempre
--	---------------	---------	---------

En la Fig. 11, se muestra como está compuesta cada uno de los clústeres, el primero se diferencia de los demás es que poco utilizan las RS, mientras que el segundo son los que más tiempo están las RS, a diferencia del tercero grupo son los que menos utilizan las RS.

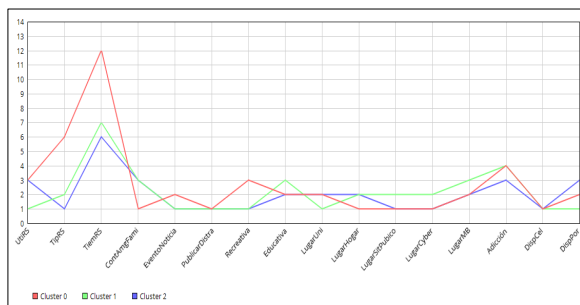


Fig. 11. Modelo de agrupamiento del algoritmo K-means hipótesis 1

Algoritmo A priori

En la TABLA IX, muestra los resultados del algoritmo, tomando en cuenta el criterio del plan de prueba, cumplen con el criterio las reglas.

TABLE IX. EJECUCIÓN DEL ALGORITMO APRIORI HIPÓTESIS 1

Reglas de asociación (A priori)	
Adicción=RaraVez Dispositivo celular=siempre 224=>> Lugar hogar=Siempre Dispositivo Celular= Siempre conf:(1) lift:(1.18) lev:(0.08) [34] conv:(34.26)	
Dispositivo portátil= Siempre 316=>> Lugar hogar=Siempre conf:(1) lift:(1.02) lev:(0.01) [6] conv:(3.61)	
Lugar universidad=Siempre Dispositivo celular=siempre 273=>> Lugar hogar=Siempre 272 conf:(1) lift:(1.02) lev:(0.01) [4] conv:(3.12)	
Educativa= Alguna vez 341=>> Lugar hogar=Siempre 240 conf:(1) lift:(1.02) lev:(0.01) [4] conv:(2.75)	

Los estudiantes que siempre utilizan el celular y computadora portátil, además lo usan en la casa y universidad tienen cierta adicción a las RS o Internet.

Los estudiantes que siempre utilizan el celular, además desde el hogar algunas veces la actividad que más realiza es para estar en contacto con amigos y familiares.

Hipótesis 2

Algoritmo K-means

En la TABLA X, se presenta los resultados de la ejecución del algoritmo, a continuación, se detalla cada uno de los clústeres. **Clúster 0**: utilizan las RS en el ámbito educativo (algunas veces) como informativo, comunicativo, y formativo principalmente utilizan en el campus universitario y hogar a través del celular y computadora. En el **Clúster 1** utilizan como informativo y comunicativo, siempre se conectan en la universidad, hogar y sitios públicos, a través del celular o computadora. Mientras que en el **Clúster 2** utilizan más como motivador, (casi siempre) siempre desde la universidad y hogar a través del celular y portátil. **Clúster**

3 lo utilizan (siempre) como informativo y comunicativo desde el hogar a través de del dispositivo celular y computadora.

TABLE X. EJECUCIÓN DEL ALGORITMO K-MEANS HIPÓTESIS 2

Atributo	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Tipo de RS	FTYIW	FYIW	FYIW	FYIW
Motivador	Rara vez	Rara vez	Casi siempre	Rara vez
Informativo	Algunas veces	Siempre	Algunas veces	Siempre
Comunicativo	Algunas veces	Siempre	Algunas veces	Siempre
Formativo	Algunas veces	Algunas veces	Algunas veces	Algunas veces
Universidad	Siempre	Siempre	Siempre	Algunas veces
Hogar	Siempre	Siempre	Siempre	Siempre
Sitios públicos	Algunas veces	Siempre	Algunas veces	Algunas veces
Cyber	Nunca	Algunas veces	Nunca	Algunas veces
Plan de datos	Algunas veces	Algunas veces	Nunca	Siempre
Celular	Algunas veces	Siempre	Siempre	Siempre
Computadora	Algunas veces	Siempre	Siempre	Siempre

En la Fig. 12, se muestra como está compuesta cada uno del clúster, el primero se diferencia de los demás es que utilizan las RS como herramienta de comunicación en el ámbito educativo, mientras que el segundo, tercero y cuarto clúster lo usan como informativo.

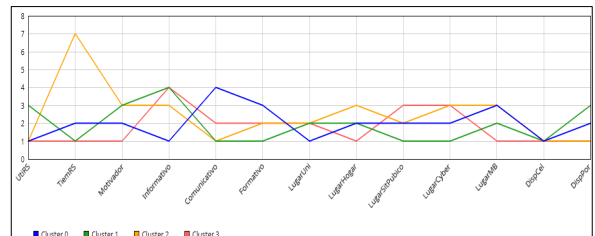


Fig. 12. Modelo de clustering del algoritmo K-means hipótesis 2

Algoritmo A priori

De las reglas generadas en el algoritmo Apriori en la hipótesis 2, por el cual la regla que está en la Fig. 13, es la que cumple con los requerimientos planteados. Donde sí utiliza en el hogar “Siempre” las RS mediante la computadora portátil o celular entonces también utilizan las RS “Algunas veces” en el ámbito educativo lo utiliza para trabajo en equipo o como herramienta de aula virtual.

Reglas de asociación(Apriori)
formativo=AlgVez disppor=Siem 264 ==> lugarhog=Siem dispcel=Siem 264 <conf:(1)> lift:(1.18) lev:(0.09) [40] conv:(40.38)

Fig. 13. Ejecución del algoritmo Apriori hipótesis 2

Graph de las reglas con mayor Lift

En la Fig. 14, se observa la relación entre variables que se asocian con la variable uso del dispositivo celular “Siempre”, quiere decir que también utilizan las RS en el hogar y universidad y “Rara vez”, tiene adicción a las redes sociales. En la Fig. 15, muestra si utilizas como formativo “Algunas veces”, desde el dispositivo celular y computadora portátil también se relaciona con la variable hogar es el lugar donde “Siempre” también utilizan las RS.

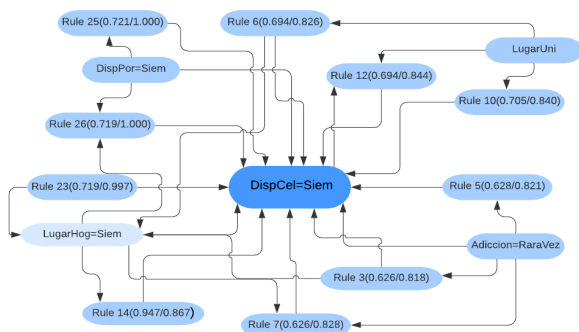


Fig. 14. Relaciones existentes entre las variables hipótesis 1

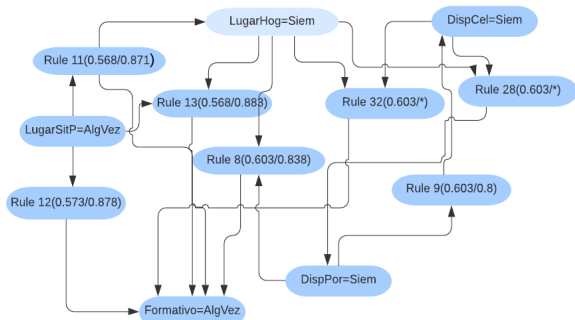


Fig. 15. Relaciones existentes entre las variables hipótesis 2

IV. DISCUSIÓN

Dentro de la información demográfica, el rango de edades de los estudiantes esta entre 22 a 24 años. El género masculino representa el 74.43% y el femenino el 25.57% de los estudiantes de la Facultad que participaron del estudio. Tomando como base de referencia el estudio [33], menciona que Facebook es la más utilizada seguida de 78% Twitter, Instagram y YouTube, mientras los datos obtenidos Facultad de la Energía, Facebook también es la RS más utilizada que representa el 47.50%, seguido por WhatsApp y YouTube, siendo Facebook utilizada en actividades como sociales, académicas, familiares u ocio con un tiempo entre 2-4 horas diariamente, a su vez a YouTube lo dedican más de 4 horas, por lo tanto, los estudiantes de la Facultad tienen un exceso uso de las redes sociales.

Mediante la ejecución del algoritmo Apriori en la herramienta RapidMiner, se presentó un inconveniente, debido que existía alteración de datos puesto que el dataset es de tipo nominal, se tuvo que aplicar dos filtros, el primero “Nominal to Numerical”, y luego el “Numerical Binomial”

porque es quien genera las reglas de asociación ya que trabajan con datos binomiales [34], mientras en la herramienta Weka se trabajó con el filtro “StringToNominal”, el dataset está compuesto de 438X29 en formato CSV para ser ejecutado en la herramienta RapidMiner, mientras para Weka los datos están en formato ARFF.

La afectación por el uso de las RS en la educación se presentó que el 27.40% tiene una afectación (Casi siempre), por el uso RS (WhatsApp y YouTube) principalmente con un promedio de más de cuatro horas, y el 35.84% tiene una afectación (algunas veces) por estar en las RS (Facebook y WhatsApp) con un tiempo promedio de dos a 4 horas, basado en la TABLA III, evidencia que si el estudiante este mucho tiempo en RS influye negativamente el rendimiento académico. Los estudiantes que siempre utilizan el celular o computadora portátil tienen cierto grado de adicción las RS o Internet, el dispositivo celular “Siempre usan”, dese el hogar o universidad tiene cierta adicción a las redes sociales.

En la Fig. 12, se muestra como está compuesta cada uno del clúster, el primero se diferencia de los demás es que poco utilizan las RS, el segundo son los que más tiempo están las RS, a diferencia del tercero grupo son los que menos usa, donde se ve reflejado el nivel de adicción, según [35], menciona a mayor tiempo en las RS mayor se ve reflejada adicción por lo cual se ve afectando el rendimiento académico. Los tres clústeres tienen en común que las redes sociales utilizan para estar en contacto con amigos y familiares, además siempre utilizan desde la universidad y hogar a través del celular.

Las RS para el uso educativo, el 29% utilizan para el intercambio de documentos y como medio de comunicación entre estudiantes y docentes, esto ayuda a mejorar comunicación en caso de existir limitaciones [14]: timidez, física, económica, distancias muy largas entre otros (ver TABLA V); además el dispositivo como celular y computadora son los aparatos tecnológicos que más utilizan en la vida diaria los estudiantes; poco vez es utilizada en la Facultad de la Energía como herramienta de aula virtual las RS.

Finalmente, en la Facultad de la Energía existe cierto grado de influencia de las RS en el rendimiento académico, además se comparten criterio con los estudios [16], [33], [36] donde se menciona que existen gran influencia negativa de las redes sociales en el rendimiento académico de los alumnos donde dedican mayoritariamente para el ocio y entretenimiento, además porque consumen mucho tiempo en estar en las RS.

V. CONCLUSIONES

Al finalizar la investigación se concluye que, por el uso de las redes sociales, principalmente para actividades de ocio, como estar día en novedades y noticias, así como para estar chateando con amigo o familiares, el 27% de la muestra de estudio casi siempre se ve afectado las actividades académicas.

Uno de los hallazgos en el uso de las redes sociales en el ámbito educativo, es que el 29% de la muestra de estudio lo utilizan como informativo (intercambio de información entre estudiantes), comunicativo (comunicación entre compañero y docentes), motivador (promover el uso activo de la RS para fines educativo o la participación en

grupo o páginas de RS) y formativo (trabajar en equipo o como herramienta virtual) poco lo utilizan.

REFERENCIAS

- [1] M. Gómez, S. Roses Pedro Farias Málaga, D. Marisol Gómez Aguilar, and P. Farias Batlle, "El uso académico de las redes sociales en universitarios," *Comunicar*, vol. 19, no. 38, pp. 131–138, 2012, doi: 10.3916/C38-2012-03-04.
- [2] A. Gordon, S. Esteban, N. Vásconez, and D. Frans, "Estudio de los comportamientos y expectativas del uso de las redes sociales en las principales universidades del país y planteamiento de una posible solución para un mejor impacto en la actividad académica," 2018.
- [3] A. Ballesteros Román, D. Sánchez Guzmán, and R. García, "Minería de datos educativa: Una herramienta para la investigación de patroness de aprendizaje sobre un contexto educativo," *Latin-American J. Phys. Educ.*, vol. 7, no. 4, pp. 662–668, 2013, doi: 10.1145/2330601.2330661.
- [4] O. Urgilés and M. Xavier, "Identificación de los factores de deserción de una carrera universitaria en una institución de educación superior, usando técnicas de minería de datos," 2018.
- [5] D. E. Gallardo Corrales, "Análisis de patrones de deserción estudiantil de la unidad educativa Lenin School aplicando minería de datos," 2017.
- [6] C. P. Gustavo Eduardo and S. G. Hugo Victor, "Proceso para el descubrimiento de conocimiento en las bases de datos de la Universidad de Cuenca mediante técnicas de Data Mining," 2017.
- [7] C. Morales Romero, S. Soto Ventura, and C. Martínez Hervás, "Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web," pp. 49–56, 2005.
- [8] F. X. Cuesta Coronel, "Diseño e implementación de un módulo de Moodle para la aplicación de Minería de Datos Educativos para la Universidad Católica Santiago de Guayaquil," 2018.
- [9] H. John, "Clustering Algorithms," pp. 354–359, 2011, doi: 10.1002/9781118165485.scard.
- [10] J. K, "A K-Means Clustering Algorithm," *Bonfring Int. J. Data Min.*, vol. 2, no. 1, pp. 13–16, 2012, doi: 10.9756/bijdm.1106.
- [11] M. Garre, J. J. Cuadrado, M. Sicilia, D. Rodríguez, and R. Rejas, "Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software," *REICIS. Rev. Española Innovación, Calid. e Ing. del Softw.*, vol. 3, no. 1, pp. 6–22, 2007.
- [12] G. Sánchez and J. Alberto, "Análisis comparativo de diferentes métodos de agrupación para el tratamiento de datos de expresión genética," 2018.
- [13] J. F. Torres Carrión, "Análisis de los usos de la tecnología en los estudiantes de la Universidad PUCE - Sede Ibarra," p. 211, 2016.
- [14] R. L. C. Campos, "Dynamics of social interaction on Facebook among university students. An approach from data mining," *Rev. Electrónica sobre Technol. Educ. y Soc.*, vol. 4, no. 7, Jan. 2017.
- [15] A. J. Meza Castillo, "Redes sociales y usos educativos," 2015.
- [16] E. Abu-Shanab and H. Al-Tarawneh, "The influence of social networks on high school students' performance," *Int. J. Web-Based Learn. Teach. Technol.*, vol. 10, no. 2, pp. 49–59, Apr. 2015, doi: 10.4018/IJWLTT.2015040104.
- [17] E. M. Mukhaini, W. S. Qayoudhi, and A. H. Badi, "Adoption of social networking in education: a study of the use of social networks by higher education students in Oman," *J. Int. Educ. Res.*, vol. 10, no. 2, p. 143, Mar. 2014, doi: 10.19030/jier.v10i2.8516.
- [18] A. Infante Moro and M. Muñoz Vázquez, "Las redes sociales en el ámbito universitario social," *La Educ. en la Soc. del Conoc.*, vol. 1, no. 1, pp. 1–12, 2013, doi: 10.1525/sop.2011.54.4.547.This.
- [19] A. E. Hoyos Zavala, "Influencia de las redes sociales en la educación superior," 2016.
- [20] J. Carlos, H. Criado, I. Juan, D. Herrera, E. Luis, and A. Jácome, "Data mining application in the virtual education," 2017.
- [21] J. Pérez *et al.*, "Mejora al algoritmo de agrupamiento K-means mediante un nuevo criterio de convergencia y su aplicación a bases de datos poblacionales de cáncer," pp. 1–7.
- [22] A. Villagra, A. Guzm, D. Pandolfi, C. Olivia, and G. Leguizam, "Análisis de medidas no-supervisadas de calidad en clusters obtenidos por K-means y Particle Swarm Optimization,"
- [23] M. Leyva Vázquez, R. Escobar Jara, C. Espín Riofrio, and K. Pérez Teruel, "Facebook como herramienta para el aprendizaje colaborativo de la inteligencia artificial," *Didascalía Didáctica y Educ.*, 2018.
- [24] C. A. Romero Romero, "Estudio comparativo de algoritmos de inteligencia artificial y minería de datos enfocados a la toma de decisiones empresariales de selección de personal." p. 153, 2018.
- [25] G. L. Campoverde, R. K. Toledo, and G. Pena, "Las redes sociales y su influencia en el comportamiento de los adolescentes," 2014.
- [26] J. González Martínez, "Las redes sociales y la educación superior: Las actitudes de los estudiantes universitarios hacia el uso educativo de las redes sociales, de nuevo a examen," *Educ. Knowl. Soc.*, vol. 17, no. 2, p. 21, Jul. 2016, doi: 10.14201/eks20161722138.
- [27] M. Escurra Mayaute and E. Salas Blas, "Construcción y validación del cuestionario de adicción a redes sociales," *Liberabit*, vol. 20, no. 1, pp. 73–91, 2014, Accessed: Oct. 23, 2019.
- [28] F. G. Paco Cocarico, "La red social Facebook y su incidencia con la trata y tráfico," 2017.
- [29] J. Rodríguez Sánchez and J. Palmero Ruiz, "Uso académico de las redes sociales: análisis comparativo entre estudiantes y profesorado de la Universidad Católica Santo Toribio de Mogrovejo," 2016.
- [30] P. Farias Batlle, M. Gómez Aguilar, and S. Roses Campos, "The academic use of social networks among University Students," *Comunicar*, vol. 19, no. 38, pp. 131–138, 2012, doi: 10.3916/C38-2011-03-04.
- [31] R. Flores, "Ventajas y desventajas del uso de las redes sociales en el estudio universitario," 2014.
- [32] A. Jaramilloa and H. Paz-ariasb, "Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje," vol. 28, no. Agosto, pp. 64–90, 2015.
- [33] R. J. Martelo, I. Jimenez-Pitre, and P. M. Martelo, "Incidencia de las redes sociales en el rendimiento académico de los estudiantes de la universidad de La Guajira (Colombia)," *Espacios*, vol. 38, no. 45, 2017.
- [34] J. C. Blandón Andrade and C. M. Zapata Jaramillo, "Una revisión de la literatura sobre población de ontologías," *Ing. y Desarro.*, vol. 36, no. 1, p. 26, 2018, doi: 10.14482/inde.36.1.10949.
- [35] Á. Ruiz Ramírez, "Minería de datos en redes sociales para pymes," Jan. 2018.
- [36] E. Gordon-salcedo, F. Noguera-vásconez, and M. Morales-morales, "Social networks in universities of the country: Descriptive analysis and approach to a possible solution to improve the impact on academic activity," vol. 2, pp. 16–38, 2019.

