

## Aplicación del algoritmo k-means como técnica de minería de datos para determinar el nivel de autoestima en los alumnos universitarios mediante la escala de rosenberg

Application of the k-means algorithm as a data mining technique to determine the level of self-esteem in university students using the rosenberg scale

*Felipe J Núñez-Cárdenas<sup>a</sup>, Juan Fernando Hernández-Reyes<sup>b</sup>, Ana M. Felipe-Redondo<sup>c</sup>, Victor T Tomas-Mariano<sup>d</sup>*

---

### Abstract:

In the present work, a typification is applied applying data mining techniques on the subject of self-esteem in university students of the Computational Sciences career. As an evaluation tool, the Rosenberg scale was applied, as well as through the K-Means algorithm. I make the grouping of the data and as a result, we seek to obtain three clusters, since the evaluation tool diagnoses three types of self-esteem.

### Keywords:

Data Mining Techniques, K-Means Algorithm, Rosenberg self-esteem Scale.

---

### Resumen:

En el presente trabajo se realiza una tipificación aplicando técnicas de minería de datos sobre el tema de la autoestima en alumnos universitarios de la carrera de Ciencias Computacionales, como herramienta de evaluación se aplico la escala de Rosenberg, así también mediante el algoritmo K-Means se realizo la agrupación de los datos y como resultado se busca obtener tres clusters, ya que la herramienta de evaluación diagnostica tres tipos de autoestima.

### Palabras Clave:

Técnicas de Minería de datos, Algoritmo K-Means, Escala de Autoestima de Rosenberg.

---

## Introducción

Uno de los factores que incide en el aprovechamiento académico, así como la deserción escolar que se da en los alumnos que cursan programas de computación es la autoestima que ellos tienen para desarrollarse dentro de estos programas.

La escala de autoestima de Rosenberg, es un cuestionario, el cual se usa para explorar la autoestima personal entendida como los sentimientos de valía personal y de respeto así mismo, esta consta de diez ítems y cuenta con tres posibles diagnósticos, el objetivo

de este test es evaluar el sentimiento de satisfacción que la persona tiene de sí misma.

El proyecto busca a través de herramientas de minería de datos como lo es la aplicación del algoritmo K-Means para la agrupación de datos, tipificar el nivel de autoestima en los alumnos que cursan la carrera de Ciencias Computacionales en la Escuela Superior de Huejutla perteneciente a la Universidad Autónoma del Estado de Hidalgo, usando como herramienta de evaluación, la escala de autoestima de Rosenberg.

---

<sup>a</sup> Autor de correspondencia, Universidad Autónoma del Estado de Hidalgo, Escuela superior de Huejutla, ORCID: 0000-0002-2462-3654, email: felipe\_nunez@uaeh.edu.mx

<sup>b</sup> Universidad Autónoma del Estado de Hidalgo, Escuela Superior de Huejutla, email: juanferdinand17@gmail.com

<sup>c</sup> Universidad Tecnológica de la Huasteca Hidalguense, email: ana.felipe@uthh.edu.mx

<sup>d</sup> Universidad Autónoma del Estado de Hidalgo, Escuela superior de Huejutla, ORCID: 0000-0001-6623-860X, email: victor\_tomas@uaeh.edu.mx

## Estado del Arte

En el año 2017 Mauricio A. Miranda realizó un estudio titulado análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos, en el cual determina cuáles son y cuál es la importancia de las variables que llevan a un estudiante a abandonar sus estudios universitarios, usando técnicas de minería de datos [1].

Así también durante el mismo año en Colombia Cindy Nayid Vega Santamaría escribe un artículo sobre análisis del diagnóstico de enfermedades mentales en la ciudad de Bogotá, mediante técnicas de minería de datos, propone un análisis de la aplicación de una técnica de minería de datos para clasificar las localidades de la ciudad de Bogotá, donde se presenta con mayor incidencia los trastornos mentales en sus habitantes [2].

En el año 2016 Daniel Pazcual, de Cuba en su trabajo que lleva por nombre Algoritmos de agrupamiento, en el cual hace uso del algoritmo de agrupamiento K-Means. Donde muestra los resultados de diferentes técnicas de agrupamiento dentro del campo de estudio de reconocimiento de patrones mostrando el estudio comparativo de 3 algoritmos empleando bases de datos reales y artificiales [3].

En el año 2018 en España Paula Ruiz Gonzales en su trabajo que lleva por nombre, Relación entre la autoestima y la satisfacción con la vida en una muestra de estudiantes universitarios, en el cual propone analizar la capacidad predictiva de la autoestima sobre la satisfacción con la vida, donde la muestra estuvo conformada por 150 estudiantes pertenecientes a diferentes grados de la universidad de Cadiz y se administró la escala de autoestima de Rosenberg y la escala de satisfacción con la vida [4].

Durante ese mismo año 2018 en España también José Enrique Moral García en su trabajo titulado, Influencia de la actitud física en la autoestima y riesgo de dependencia en personas mayores activas y sedentarias, en el cual su objetivo del estudio fue conocer cómo influye la práctica de actividad física en la autoestima y en el nivel de riesgo de dependencia de las personas mayores, diferenciados entre en grupo de control (sedentarios) y el grupo experimental (activos), donde participaron 18 personas mayores de 65 años, de las cuales 84 personas fueron catalogadas como activas y los 84 restantes fueron completamente sedentarias. Para analizar la autoestima se ha utilizado la escala de autoestima de Rosenberg y

para evaluar el riesgo de dependencia se utilizó el test de Barber [5].

En el 2017 Estefanía Ruiz Palomino de España en su proyecto que tiene por nombre El papel de la autoestima en la prevención del VIH se evaluaron a 467 jóvenes de la comunidad de Valencia, se utilizó la Escala de autoestima de Rosenberg y el cuestionario de prevención del Sida de Ballester, los resultados muestran relaciones positivas estadísticamente significativas de la autoestima con autoestima con auto eficiencia en el uso del preservativo [6].

Durante el mismo año Bruno Williams Chacaliza López en su trabajo titulado Relación entre la capacidad funcional y la autoestima en los adultos mayores con diabetes mellitus de los centros de salud de morales y banda de shilcayo del ministerio de salud San Martin 2017, nos habla sobre evaluar la relación entre la capacidad funcional y la autoestima en los adultos mayores con diabetes mellitus atendidos en los centros de salud de Morales y Banda de Shilcayo, en donde el material y método empleado fue una muestra de 48 adultos mayores diabéticos obtenidos por muestreo probabilístico, así como también los instrumentos utilizados fueron la escala de Lawton y Brody para la capacidad funcional y la escala de Rosenberg para la autoestima [7].

## Fundamento Teórico

### Minería de Datos

La minería de datos surgió con la intención o el objetivo de ayudar a comprender una enorme cantidad de datos, y que estos, pudieran ser utilizados para extraer conclusiones para contribuir en la mejora y crecimiento de las empresas, sobre todo, por lo que hace a las ventas o fidelización de clientes.

Su principal finalidad es explorar, mediante la utilización de distintas técnicas y tecnologías, bases de datos enormes de manera automática con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos que se han ido recopilando con el tiempo. Estos patrones pueden encontrarse utilizando estadísticas o algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales.

Por tanto, los datos son el medio o la base para llegar a conclusiones y transformar estos datos en información relevante, para que las empresas puedan abarcar mejoras y soluciones que les ayuden a conseguir sus objetivos.

Las personas que se dedican al análisis de datos a través de este sistema son conocidos como mineros o exploradores de datos, estos intentan descubrir patrones en medio de enormes cantidades de datos. Su intención es la de aportar información valiosa a las empresas para así, ayudarlas en la toma de decisiones futuras. Pero debemos tener claro que la elección del mejor algoritmo para una tarea analítica específica es un gran desafío, ya que podemos encontrar muchos patrones distintos, y además, dependerá de los problemas a resolver. Estos pueden ser la clasificación, regresión, segmentación, asociación y análisis de secuencias.

Los mineros o exploradores de datos a la hora de llevar a cabo un análisis de Data Mining, deberán realizar cuatro pasos distintos:

**Determinación de los objetivos:** El cliente determina qué objetivos quiere conseguir gracias al uso del Data Mining.

**Procesamiento de los datos:** Selección, limpieza, enriquecimiento, reducción y transformación de la base de datos.

**Determinación del modelo:** Primero se debe hacer un análisis estadístico de los datos y después visualización gráfica de los mismos.

**Análisis de los resultados:** En este paso se deberán verificar si los resultados obtenidos son coherentes.

Actualmente este tipo de trabajos se están realizando en seguridad de datos, finanzas, salud, marketing, detección de fraude, búsquedas online, procesamiento del lenguaje natural, coches inteligentes, entre otros. Es por este motivo, que la minería de datos se está convirtiendo en uno de los trabajos con mayor proyección para el futuro [8].

## Algoritmo K-Means

K-means o K-medias es uno de los algoritmos más utilizados para realizar agrupamiento, técnica implementada en Minería de Datos. La idea del k-medias es colocar todos los objetos en un espacio determinado y dadas sus características formar grupos de objetos con rasgos similares pero diferentes a los demás que integran otros grupos.

Sin embargo, el algoritmo presenta algunos inconvenientes:

- El agrupamiento final depende de los centroides iniciales.
- La convergencia en el óptimo global no está garantizada, y para problemas con muchos ejemplares,

requiere de un gran número de iteraciones para converger.

Descripción del algoritmo k-means;

**Paso 1. Inicialización:** Se definen un conjunto de objetos a los cuales se les aplica el proceso de clustering que consiste en la división de los datos en grupos y un centroide (centro geométrico del clusters) para cada uno. Los centroides iniciales se pueden determinar aleatoriamente, mientras que en otros casos procesan los datos y se determinan los centroides mediante cálculos.

**Paso 2. Clasificación:** Para cada dato se calcula la distancia (euclidiana cuadrada) con respecto a los centroides, se determina el centroide más cercano a cada uno de los datos, y el objeto se anexa al clusters del centroide que fue seleccionado.

**Paso 3. Calculo de centroides:** Para cada uno de los clusters se vuelve a recalcular los centroides.

**Paso 4. Verificación de convergencia:** En este paso se comprueba si una de las condiciones del algoritmo se ha cumplido y que este debe parar, a esto se le llama condición de convergencia o paro. A continuación, se mencionan algunas de las condiciones de convergencia:

- El número de iteraciones.
- Cuando los centroides obtenidos en dos iteraciones sucesivas no cambian su valor.
- Cuando la diferencia entre los centroides de dos iteraciones sucesivas no supera cierto umbral.
- Cuando no hay transferencia de objetos entre grupos en dos iteraciones sucesivas.
- Si algunas de las condiciones de convergencia no cumplen se repiten los pasos dos, tres y cuatro del algoritmo [9].

## Escala de Autoestima de Rosenberg

La prueba consta de 10 preguntas, puntuables entre 1 y 4 puntos, lo que permite obtener una puntuación mínima de 10 y máxima de 40. Las frases están enunciadas una mitad en forma positiva y la otra mitad en forma negativa, al objeto de evitar el llamado efecto de "aquiescencia autoadministrada".

No se establecieron inicialmente puntos de corte que permitan clasificar el tipo de autoestima según la puntuación obtenida, aunque sí es posible establecer un rango normal de puntuación en función de la muestra que se tome como referencia [10].

Esta escala es breve, rápida, fiable y goza de gran validez, tanto que es uno de los instrumentos a los que

más se recurre para evaluar el autoconcepto que alguien tiene por parte de los psicólogos. También se utiliza con frecuencia cuando se quiere medir esta variable en el contexto de alguna investigación.

La autoestima es un concepto ampliamente recta relación con el bienestar general de la estudiado y discutido, observándose su di- persona, lo que sugiere que podría ser un buen indicador de la salud mental. Diversas investigaciones han confirmado la relación inversa entre autoestima y síntomas depresivos, obsesivos compulsivos, ansiosos, entre otros. Por tanto, su conocimiento es una valiosa información que posibilitaría la predicción de posibles trastornos psicopatológicos y su prevención

Rosenberg entiende la autoestima como un sentimiento hacia uno mismo, que puede ser positivo o negativo, el cual se construye por medio de una evaluación de las propias características.

30 a 40 puntos	Autoestima Elevada
26 a 29 puntos	Autoestima Media
Menos de 25 puntos	Autoestima Baja

Tabla 1. Diagnostico de Escala de Autoestima de Rosenberg

### Herramienta Weka

Weka es una herramienta de tipo software para el aprendizaje automático y minería de datos diseñado a base de Java y desarrollado en la universidad de Waikato en Nueva Zelanda en el año 1993, esta herramienta por su nombre en inglés (Waikato Environment for Knowledge Analysis) además es una herramienta de distribución de licencia GNU-GLP o software libre.

Weka contiene una colección de algoritmos para realizar análisis de datos y modelado predictivo, también tiene herramientas para la visualización de estos datos, además provee una interfaz gráfica que unifica las herramientas para que estén a una mejor disposición.

Esta es una herramienta muy versátil que soporta muchas tareas estándar de la minería de datos en especial tareas de procesamiento de datos, regresión, clasificación, clusterin entre otras, así mismo permite la visualización y la selección de los datos.

Todas las técnicas en WEKA están basadas en la sunción de datos que están disponibles en un fichero plano o una relación, en donde cada registro de datos esta descrito por un número fijo de atributos nominales o numéricos.

Permite el acceso a otras instancias de bases de datos por medio de SQL, gracias al JDBC, además puede procesar un resultado generado a base de una consulta hecha a una base de datos.

Weka posee un grupo de técnicas que se pueden aplicar con éxito a múltiples capos, tales como el Marketing, manufactura, salud, energía, finanzas, medicina, entre otros, para su aplicación se debe de tener en cuenta el tipo de tarea que se desea realizar. Con base a la tarea se puede implementar en para dar la solución necesaria en cada uno de estos campos, a continuación, se muestran una serie de tareas incluidas en Weka.

- Métodos de clasificación basados en redes neuronales
  - Manipulación de métodos numéricos sobre datos (Funciones estadísticas)
  - Métodos de clasificación basados en máquinas de soporte vectorial
  - Metaclasificadores
  - Métodos de implementación de árboles de decisión
  - Implementación de clustering
  - Métodos de estimación de probabilidades
  - Clases para implementar clasificadores bayes.
- [11].

### Metodología CRISP-DM

La metodología CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica, en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases, como se muestran a continuación [12]:

La sucesión de fases no es necesariamente rígida. Cada fase es estructurada en varias tareas generales de segundo nivel como lo muestra la siguiente Figura 1 .

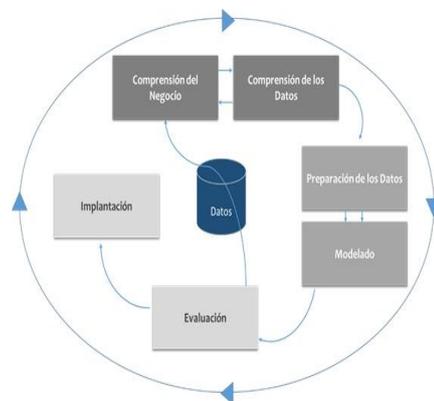


Figura 1 Metodologia CRISP-DM

### *Comprensión del Negocio*

En la fase de Entendimiento del Negocio se deben entender los objetivos del proyecto y los requerimientos desde una perspectiva del negocio y luego convertir este conocimiento en una definición de un problema de explotación de información y diseñar un plan preliminar para lograr dichos objetivos.

### *Entendimiento de los Datos*

El Entendimiento de los Datos comienza con la recolección inicial de datos y procede con las acciones para familiarizarse con ellos, identificar problemas de calidad, identificar primeras pautas en los datos o detectar subconjuntos interesantes de las hipótesis de información oculta.

### *Preparación de los Datos*

La fase de Preparación de los Datos cubre todas las actividades para construir el conjunto de datos final desde los datos iniciales, las tareas de esta fase pueden ser realizadas muchas veces y sin un orden preestablecido, incluye tanto la selección de tablas, registros y atributos como transformación y limpieza de datos para herramientas de modelado.

### *Construcción del Modelo*

El Modelado incluye la selección de técnicas de modelado y la calibración de sus parámetros a los valores óptimos, suelen existir distintas técnicas para un mismo problema de explotación de información y cada una de ellas suele tener ciertos requisitos sobre los datos, muchas veces es necesario volver a la fase de preparación de los datos.

### *Evaluación*

La Evaluación requiere la construcción de uno o varios modelos que aparentan tener la mayor calidad desde una perspectiva de análisis, requiere la evaluación del modelo y revisión de los pasos ejecutados para la construcción del modelo para asegurarnos de lograr los objetivos de negocio, al final de esta fase se debería poder tomar una decisión respecto de la utilización de los resultados.

### *Despliegue*

La fase de despliegue puede ser tan simple como generar un reporte o tan compleja como implementar un

proceso de explotación de información repetible a través de toda la empresa.

## **Desarrollo**

En este apartado se realiza el desarrollo de la metodología CRISP-DM aplicada al presente proyecto en cada una de sus etapas utilizando técnicas de minería de datos, la herramienta weka y como instrumento la escala de autoestima de Rosenberg.

### *Comprensión del Negocio*

El objetivo del proyecto es el de realizar una tipificación detallada a partir de datos recabados en la aplicación de la escala de autoestima de Rosenberg en alumnos universitarios, específicamente a los alumnos de la licenciatura de Ciencias Computacionales de la escuela superior de Huejutla de Reyes, perteneciente a la Universidad Autónoma del Estado de Hidalgo, mediante la técnica de minería de datos algoritmo K-Means.

Esta tipificación resulta útil tanto para docentes y padres de familia, donde se podrá ver en qué porcentaje los alumnos de universidad sufren de autoestima alta o baja y poder ayudarlos.

### *Comprensión de los Datos*

En esta segunda etapa de la metodología CRISP-DM se realiza una recolección de datos y así establecer el primer contacto con el problema, verificando la calidad de los mismos.

Los datos que se utilizaron en este proyecto son referentes a 50 alumnos de la licenciatura en Ciencias Computacionales que estudian en la Escuela Superior de Huejutla de Reyes, los cuales se recabaron a partir de la aplicación de la escala de autoestima de Rosenberg, instrumento seleccionado, a continuación, se presenta La escala de autoestima de Rosenberg Figura 2.

ESCALA DE AUTOESTIMA DE ROSEMBERG

Indicación: Cuestionario para explorar la autoestima personal entendida como los sentimientos de valía personal y de respeto a sí mismo.

Este test tiene por objeto evaluar el sentimiento de satisfacción que la persona tiene de sí misma. Por favor, conteste las siguientes frases con la respuesta que considere más apropiada.

- A. Muy de acuerdo
- B. De acuerdo
- C. En desacuerdo
- D. Muy en desacuerdo

	A	B	C	D
1. Siento que soy una persona digna de aprecio, al menos en igual medida que los demás				
2. Estoy convencido de que tengo cualidades buenas				
3. Soy capaz de hacer las cosas tan bien como la mayoría de la gente				
4. Tengo una actitud positiva hacia mí mismo/a				
5. En general estoy satisfecho/a de mí mismo/a				
6. Siento que no tengo mucho de lo que estar orgulloso/a				
7. En General, me inclino a pensar que soy un fracasado/a				
8. Me gustaría poder sentir más respeto por mí mismo				
9. Hay veces que realmente pienso que soy inútil				
10. A veces creo que no soy buena persona				

Figura 2. Escala de Autoestima de Rosenberg

Preparación de los Datos

Una vez que se han descrito los datos, se procede a preparar los datos iniciales recabados de la aplicación de la escala de autoestima de Rosenberg ( anexo Figura 13) para poder ser usados en el proceso de modelado, se procederá a crear un archivo con extensión (.arff) ya que es el formato requerido para poder manipular los datos en la herramienta WEKA Figura 3, seleccionaremos las variables que queremos analizar y que son apropiadas.

```

1 @relation rosenberg
2 @attribute var1 numeric
3 @attribute var2 numeric
4 @attribute var3 numeric
5 @attribute var4 numeric
6 @attribute var5 numeric
7 @attribute var6 numeric
8 @attribute var7 numeric
9 @attribute var8 numeric
10 @attribute var9 numeric
11 @attribute var10 numeric
12 @data
13 3,3,3,4,2,3,3,3,3,3
14 4,4,4,4,4,4,4,4,4,1
15 3,4,4,4,3,3,4,4,4,4
16 4,4,4,4,4,4,4,4,4,4
17 3,3,3,3,3,3,3,3,3,3
18 3,3,3,3,3,4,4,3,4,3
19 3,3,3,3,2,3,3,3,4,2
20 4,3,2,3,3,1,3,2,2,2
21 2,3,4,3,3,1,1,2,3,2
22 3,3,3,3,3,2,2,3,3,3
23 4,3,3,3,3,2,3,4,4,4
24 4,3,3,3,3,2,2,2,3,4
25 4,3,2,1,2,2,3,4,4,3
    
```

Figura 3. Preparación de los datos en herramienta WEKA

Modelado

En esta fase de la metodología se selecciona la técnica para la agrupación de una serie de vectores según la distancia, vectores que estén más cercanos, aquellos que tengan características comunes. A continuación, se procederá a aplicar dicha técnica sobre los datos para generar el modelo y por último se tendrá que evaluar si dicho modelo ha cumplido los criterios de éxito o no.

Como técnica de modelado para este proyecto se eligió el algoritmo K-Means para realizar el agrupamiento, el modelado usa la herramienta WEKA.

- Abrimos la herramienta WEKA Figura 4.



Figura 4. Herramienta WEKA

- Seleccionamos la opción Explorer Figura 5.

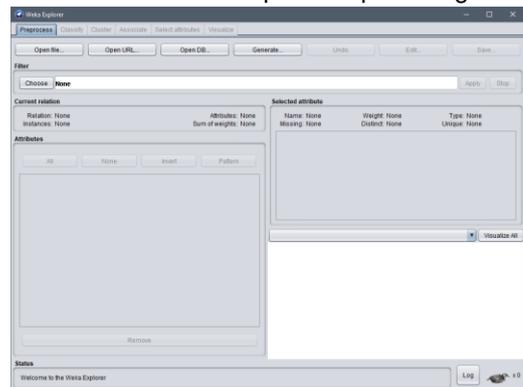


Figura 5. Explorer en la herramienta WEKA

- Cargamos el archivo con extensión (.arff) que se realizó en la etapa de preparación de los datos Figura 6.

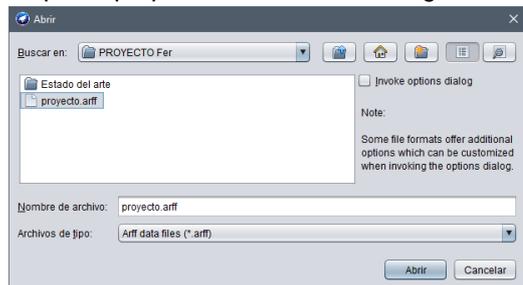


Figura 6. Carga de Datos

- Seleccionamos la opción de Cluster Figura 7.

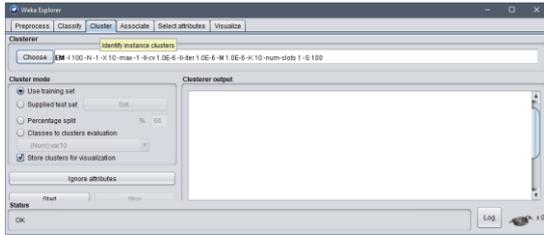


Figura 7. Cluster en la herramienta WEKA

•Seleccionamos el algoritmo K-Means como se muestra la siguiente Figura 8.

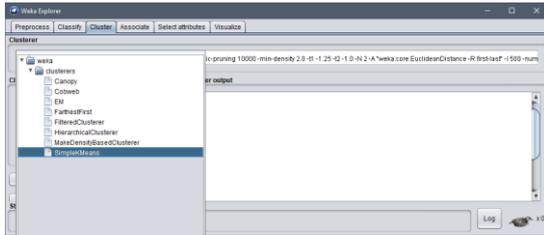


Figura 8. Selección del algoritmo Kmeans

•Colocamos el número de clusters queremos obtener, en este caso de acuerdo con los posibles diagnósticos de autoestima tres Figura 9.

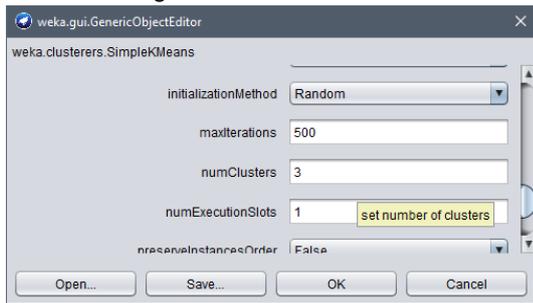


Figura 9. Números de Cluster

•Pulsamos el botón “Start” para iniciar las iteraciones del algoritmo K-Means Figura 10.

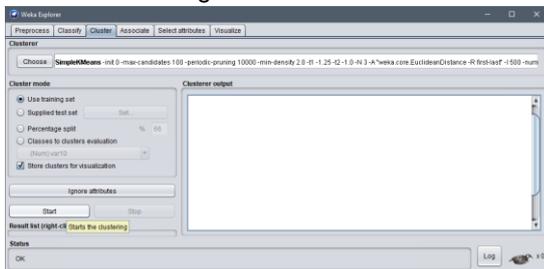


Figura 10. Inicio de Ejecucion del Algoritmo K means

•Una vez terminado se muestran los resultados como muestra la Figura 11.

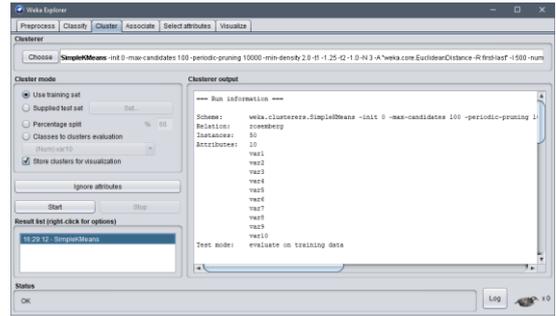


Figura 11. Resultados

•los Clusters finales se conformn como aparecen en la Figura 12.

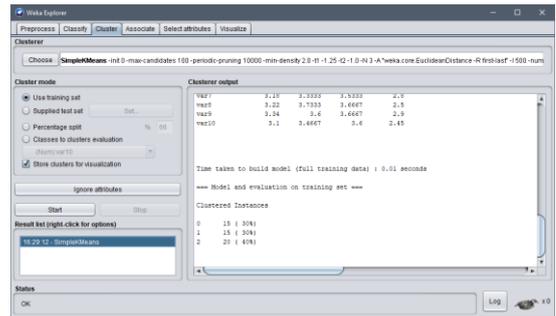


Figura 12. Cluster Finales

### Evaluación

los resultados obtenidos se describen en la siguiente tabla:

Cluster 1	3,2,3,2,3,2,4,4,4,4
Cluster 2	3,3,3,3,3,3,3,4,4,4
Cluster 3	4,3,3,3,2,2,3,2,4,3

Tabla 2. Clusters

En el cluster 1 de acuerdo con los datos arrojados aleatoriamente corresponden a una población de alumnos que de acuerdo con la escala de Rosemberg tienen una autoestima elevada.

El cluster 2, los datos corresponden a una población con características que indican una autoestima elevada.

Por último, en el cluster 3, los datos corresponden a una población descrita según la escala de Rosemberg con autoestima media.

En la siguiente tabla se describirán los clusters y la población que esta asignada a cada cluster.

Cluster	Población	%
1	15	30%
2	15	30%
3	20	40%
<b>Total:</b>	<b>50</b>	<b>100%</b>

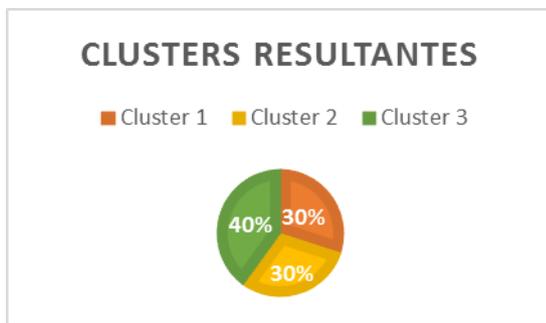
Tabla 3. Descripción de la población en los clusters

- En el cluster 1, quedaron 15 estudiantes universitarios que corresponden al 30% de la población total.
- En el cluster 2, quedaron 15 estudiantes universitarios que corresponden al 30% de la población total.
- En el cluster 2, quedaron 20 estudiantes universitarios que corresponden al 40% de la población total.

Se puede decir que la escala de autoestima de Rosenberg, del total de la población el 60% cuenta con una autoestima elevada, mientras que el 40% su autoestima es media, por lo cual se encuentra en buen nivel para los estudiantes.

**Despliegue**

Como resultado de la aplicación de la escala de Rosenberg a alumnos universitarios, específicamente a los de la licenciatura en Ciencias Computacionales, se aplicó una técnica de agrupamiento, se optó por el algoritmo K-Means, se eligieron tres clusters debido a los posibles diagnósticos de autoestima según la escala de autoestima de Rosenberg, a continuación, se presenta la Gráfica 1 que representa el porcentaje de los clusters finales.



Gráfica 1. Cluster finales

A continuación, se muestra Gráfica 2 que muestra cómo se conforman los datos en cada clúster.

	Cluster 1	Cluster 2	Cluster 3
10	4	4	3
9	4	4	4
8	4	4	2
7	4	3	3
6	2	3	2
5	3	3	2
4	2	3	3
3	3	3	3
2	2	3	3
1	3	3	4

Gráfica 2. Conformacion de Datos en Cluster

Estos resultados obtenidos ayudan a la institución educativa para tomar acciones dentro del departamento de psicopedagogía de tal manera que se puedan

prevenir situaciones como deserción escolar, adicciones, etc.

**Conclusiones y Trabajo Futuro**

En el desarrollo del presente trabajo, se aprendió sobre la importancia que tiene la minería de datos en la actualidad, así como las técnicas que esta utiliza tanto como para descripción como para predicción.

También se usó el Algoritmo K-Means, para realizar una tipificación de la población y describir su comportamiento, aplicando la escala de autoestima de Rosenberg.

Se aplico la metodología CRISP-DM, para poder desarrollar el proyecto, la cual está enfocada o destinada a proyectos directamente relacionados con la minería de datos.

La autoestima forma un papel muy importante en los estudiantes de universidad y suele ser un factor al momento de tomar decisiones importantes, debido a eso se realizó este proyecto apoyándose de la minería de datos y sus técnicas, los resultados obtenidos sirven para encausar las acciones para aumentar el autoestima en aquellos alumnos que la poseen en termino medio a baja, y mas aun acrecentar los diferentes factores que impactan en cada rubro de las diez variables utilizadas para el diagnostico.

Se puede dar continuidad al trabajo, pero ahora aplicando una técnica de minería de datos distinta como puede ser predictiva y así determinar las características en alumnos universitarios que causen baja escolar y que se ven influenciados por algún parámetro o factor que incide en el instrumento usado para este proyecto.

**Anexo**

ESCALA DE AUTOESTIMA DE ROSENBERG

Indicación: Cuestionario para explorar la autoestima personal entendida como los sentimientos de valía personal y de respeto a sí misma.

Este test tiene por objeto evaluar el sentimiento de satisfacción que la persona tiene de sí misma. Por favor, conteste las siguientes frases con la respuesta que considere más apropiada.

A. Muy de acuerdo  
B. De acuerdo  
C. En desacuerdo  
D. Muy en desacuerdo

	A	B	C	D
1. Siento que soy una persona digna de aprecio, al menos en igual medida que los demás		X		
2. Estoy contento/a de que tengo cualidades buenas		X		
3. Soy capaz de hacer las cosas tan bien como la mayoría de la gente		X		
4. Tengo una actitud positiva hacia mí mismo/a	X			
5. En general estoy satisfecho/a de mí mismo/a			X	
6. Siento que no tengo mucho de lo que estar orgulloso/a			X	
7. En General, me inclino a pensar que soy un fracasado/a			X	
8. Me gustaría poder sentir más respeto por mí mismo			X	
9. Hay veces que realmente pienso que soy inútil			X	
10. A veces creo que no soy buena persona			X	

Figura 13. Escala de Autoestima Resuelta

## Referencias

- [1] Miranda, M. A. (2017). análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos. Chile
- [2] Santamaría, C. N. V., Angulo, A. S., & Mariño, P. A. (2017). Análisis del diagnóstico de enfermedades mentales en la ciudad de Bogotá, mediante técnicas de minería de datos. Mundo FESC, (13), 35-47.
- [3] González, D. P. (2010). Algoritmos de agrupamiento basados en densidad y variación de clusters (Doctoral dissertation, Universitat Jaume I, Departament de Llenguatges i Sistemes Informàtics).
- [4] Ruiz-González, P., Medina-Mesa, Y., Zayas, A., & Gómez-Molinero, R. (2018). Relación entre la autoestima y la satisfacción con la vida en una muestra de estudiantes universitarios. International Journal of Developmental and Educational Psychology. Revista INFAD de Psicología., 2(1), 67-76.
- [5] Moral-García, J. E., García, D. O., García, S. L., Jiménez, M. A., & Dios, R. M. (2018). Influencia de la actividad física en la autoestima y riesgo de dependencia en personas mayores activas y sedentarias. Anales de Psicología/Annals of Psychology, 34(1), 162-166.
- [6] Ruiz-Palomino, E., Ballester-Arnal, R., Gil-Llarío, M. D., & Giménez-García, C. (2017). El papel de la autoestima en la prevención del VIH de jóvenes españoles. International Journal of Developmental and Educational Psychology, 2(1), 15-21.
- [7] López, C., & Williams, B. (2018). Relación entre la capacidad funcional y la autoestima en los adultos mayores con diabetes mellitus de los centros de salud de Morales y Banda de Shilcayo del Ministerio de Salud-SanMartín 2017.
- [8]. Hernández Orallo, J., FERRI RAMIREZ, C., & RAMIREZ QUINTANA, M. J. (2004). Introducción a la Minería de Datos.
- [9]. Pascual, D., Pla, F., & Sánchez, S. (2007). Algoritmos de agrupamiento. Método Informáticos Avanzados, 164-174.
- [10] psicopedia. (s.f.). Obtenido de <http://psicopedia.org/1723/medir-la-autoestima-con-la-escala-de-rosenberg/> al 16 de febrero del 2018.
- [11] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
- [12] smartbasegroup. (2016). metodología crisp-dm. obtenido de <http://smartbasegroup.com/metodologia-crisp-dm-final/> al 22 de febrero del 2018.