



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

**NOMBRE:** WALTER RAUL PEREZ MACHINENA

**MATRICULA:** 1603647

**GRUPO:** 003

**NOMBRE ACTIVIDAD:** MODELO DE AGRUPAMIENTO

**NUMERO DE ACTIVIDAD:** 4

**MATERIA:** APRENDIZAJE AUTOMATICO

**DOCENTE:** JOSE ANASTACIO HERNANDEZ SALDAÑA



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



# FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

## Contenido

Introducción.....	3
Hallazgos.....	4
Grafica 1 .....	4
Grafica 2 .....	5
Conclusiones.....	6
Referencias .....	6



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



## Introducción

En esta actividad se considero una base de datos sobre eficiencia energética en los 12 diferentes formas de construcción de edificios, que considera 8 variables descriptivas, donde se tienen dos variables objetivo, siendo una para la calefacción y otra para el enfriamiento.

En este caso nuestra variable objetivo es una discreta que es la X6, por referirse a la orientación del edificio, y se eliminan del contexto las variables objetivo-identificadas como Y1 y Y2 para no viciar la información.

Se busca crear un agrupamiento con k-means para el modelo de clasificación, donde se compare con el modelo de clasificación. A si mismo generar un diagrama de codo o de (elbow-method), el cual nos ayudara a identificar el valor de k de forma gráfica.

### Tabla de variables

Nombre de la variable	Role	Tipo	Descripción
X1	Característica	Continuo	Compacidad relativa
X2	Característica	Continuo	Área de superficie
X3	Característica	Continuo	Área de la pared
X4	Característica	Continuo	Área del techo
X5	Característica	Continuo	Altura total
X6	Característica	Entero	Orientación
X7	Característica	Continuo	Área de acristalamiento
X8	Característica	Entero	Distribución del área de acristalamiento
Y1	Objetivo	Continuo	Carga de calefacción
Y2	Objetivo	Continuo	Carga de enfriamiento



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



## Hallazgos

Modelo	Cross validation accuracy
KNN	0.04 (+/- 0.04)
Arbol de Decision	0.01 (+/- 0.03)
SVN	0.18 (+/- 0.07)
Logistic Regression	0.22 (+/- 0.07)

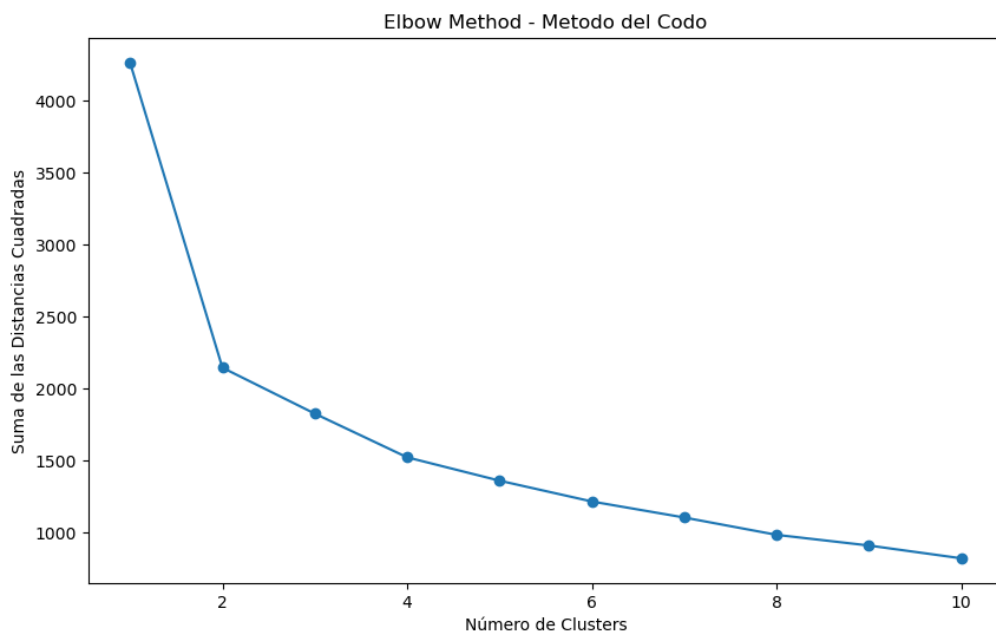
Con el fin de no ser repetitivo, en el Reporte 3 se mostro que el mejor modelo era Logistic Regression, ahora el interés principal es determinar si es mejor utilizar la clasificación o el agrupamiento.

Siguiendo las indicaciones realice lo siguiente:

Realizar k-means con la intención que por medio de la grafica de codo (elbow method), identificar cual sería el valor optimo.

En este caso se determino que era el 2 como se muestra en la gráfica a continuación.

Grafica 1





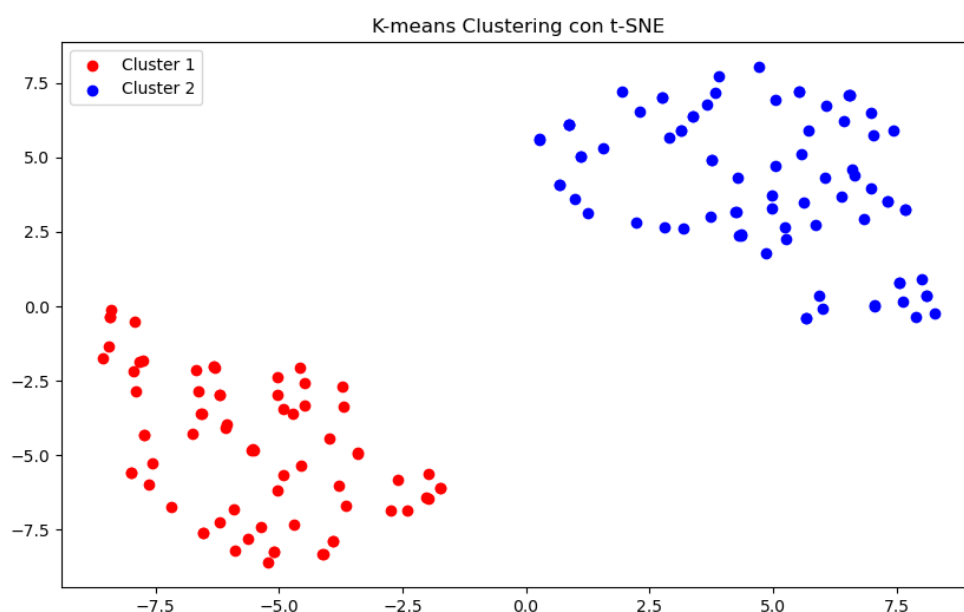
# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



Una vez identificado que el valor correcto de  $k$  sería el 2 decidí utilizar TSNE, el cual permite mostrar los datos de Kmeans de forma gráfica, y ayuda a reducir la dimensionalidad, generando la siguiente grafica.

Grafica 2



A si mismo para realizar una correcta comparación entre nuestro modelo de clasificación y el modelo de agrupamiento como se trata de modelos diferentes no se pueden comparar de una forma tan sencilla, es por esto por lo que utilice algo llamado silhouette\_score el cual va desde el -1, 0 y 1. Esto ayuda a calcular la calidad del agrupamiento realizado por K-means.

Es decir, se utilizó precisión para el modelo de clasificación y coeficiente de silueta para el modelo de agrupamiento, de esta manera poder realizar la comparación.

Donde se identifico que **el mejor modelo, es el de agrupación.**



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



## Conclusiones

En la presente actividad se realizaron acciones de agrupamiento, donde se identificó el valor óptimo, siendo este el 2, esto se realizó de forma visual (gráfica de codo – Gráfica 1) de esta manera se realizó un reentrenamiento ya con el valor óptimo identificado. También logramos visualizar la forma en que se realizó el agrupamiento de forma visual con la Gráfica 2.

Sin embargo, nos encontramos con el problema de que al momento de comparar ambos modelos era complicado, ya que cada uno fue creado para obtener diferentes resultados, considerando lo anterior, se optó por delimitar unos umbrales o parámetros aceptables para cada modelo y de esta manera identificar cual era el **mejor siendo este el Modelo de Agrupamiento**.

## Referencias

Guido, A. C. (2016). *Introduction to Learning with Python*. California: O'REILLY.

Scikit-Learn. (2024, 07 21). *Scikit-Learn*. Retrieved from <https://scikit-learn.org/stable/>

Xifara, A. T. (2024, 07 21). *archive.ics.uci*. Retrieved from <https://archive.ics.uci.edu/dataset/242/energy+efficiency>

Material de Clase.