

Universidad de Costa Rica
Escuela de Ingeniería Eléctrica
IE0435 – Inteligencia Artificial Aplicada a la Ingeniería Eléctrica
II ciclo 2023

Profesor: Ing. Marvin Coto Jiménez

Proyecto 01

Katherine Vargas Castro B88198

Septiembre 2023

Índice

Lista de Figuras	II
Lista de Tablas	III
1 Resumen	1
2 Introducción	1
3 Procedimiento	2
4 Resultados	3
4.1 ANEXO A. Circuitos con frecuencia promedio por abonado y duración promedio de las interrupciones superior a los límites normativo	3
4.2 ANEXO B. Circuitos con frecuencia promedio por abonado o duración promedio de las interrupciones superior a los valores normativos	8
5 Conclusiones	13
Referencias	15
Apendice	16

Lista de Figuras

Figure 1:	Visualización de Datos	4
Figure 2:	Visualización de Datos	4
Figure 3:	Método del codo para el AnexoA	5
Figure 4:	Gráfica en 3D para $k=4$	5
Figure 5:	Gráfica en 3D para $k=5$	6
Figure 6:	Gráfica en 3D para $k=6$	6
Figure 7:	Gráfica en 3D para $k=7$	7
Figure 8:	Gráfica en 3D para $k=8$	7
Figure 9:	Visualización de Datos	8
Figure 10:	Visualización de Datos	8
Figure 11:	Método del codo para el AnexoB	9
Figure 12:	Gráfica en 3D para $k=4$	10
Figure 13:	Gráfica en 3D para $k=5$	10
Figure 14:	Gráfica en 3D para $k=6$	11
Figure 15:	Gráfica en 3D para $k=7$	11
Figure 16:	Gráfica en 3D para $k=8$	12
Figure 17:	Gráfica en 3D para $k=8$	12
Figure 18:	Gráfica en 3D para $k=8$	13

Lista de Tablas

1. Resumen

El objetivo del proyecto es evaluar la calidad del servicio de distribución eléctrica de Costa Rica utilizando métodos de agrupamiento de circuitos eléctricos. El suministro de electricidad en Costa Rica se realiza a través de una serie de empresas distribuidoras, y la calidad del servicio es crucial. El proyecto utiliza datos de abonados, DPI y FPI de varios circuitos y los agrupa para identificar patrones y similitudes que ayuden a caracterizar las fallas y problemas de calidad eléctrica. Para lograrlo, se utilizarán algoritmos de agrupamiento y se analizarán los resultados en el contexto de la distribución eléctrica del país.

2. Introducción

La calidad del suministro de electricidad es un tema de gran importancia en cualquier país, ya que afecta directamente la calidad de vida de las personas.

Respecto a esto, la Autoridad Reguladora de los Servicios Públicos (ARESEP) en Costa Rica, publica anualmente un Informe de Calidad del Suministro de Electricidad. La existencia de esta autoridad reguladora es fundamental para evaluar que las empresas distribuidoras cumplan con los estándares de calidad y confiabilidad necesarios, garantizando así que los usuarios reciban un servicio de calidad.

El presente proyecto se basa en los datos proporcionados por esta entidad específicamente en el reporte publicado en el año 2021, donde se incluye información sobre la región, los abonados, el tiempo promedio de interrupción por abonado (DPI) y la frecuencia promedio de las interrupciones por abonado (FPI), (ARESEP 2021). El objetivo principal es agrupar estos circuitos en función de estas variables para identificar patrones y similitudes que ayuden a comprender mejor la calidad eléctrica en diferentes regiones de Costa Rica.

Cantidad de interrupciones promedio percibidas por los usuarios del servicio eléctrico (SCIJ 2015):

$$F.P.I = \frac{(\sum_{i=1}^n A_i \cdot C_i)}{At}$$

En donde:

- A_i = Número de abonados o usuarios afectados por la interrupción i , de nivel I, II, III y IV.
- C_i = Total de interrupciones, asociados al elemento de protección de nivel I, II, III y IV.
- At = Número total de abonados del sistema eléctrico, subestación, circuito o alimentador, etc.

- n = Número de interrupciones en el semestre de estudio.

Tiempo promedio de interrupción percibido (SCIJ 2015):

$$D.P.I = \frac{(\sum_{i=1}^n A_i \cdot T_i)}{At}$$

En donde:

- A_i = Número de abonados o usuarios afectados por la interrupción i , de nivel I, II, III y IV.
- T_i = Tiempo en horas de la interrupción i .
- At = Número total de abonados del sistema eléctrico, subestación, circuito o alimentador, etc.
- n = Número de interrupciones en el semestre de estudio.

En la literatura podemos encontrar diferentes trabajos donde se utilizan algoritmo de agrupamiento en el sector electro, por ejemplo, en el artículo "Partición de una red eléctrica de distribución aplicando algoritmos de agrupamiento", presentan un caso de estudio donde utilizan algoritmos de agrupamiento como K-means y DBSCAN para la partición de la red eléctrica, y se compararon los resultados obtenidos con los de otros métodos de partición de redes eléctricas. Los resultados mostraron que la metodología propuesta es efectiva para la partición de redes eléctricas de distribución y puede mejorar la eficiencia y la calidad del servicio eléctrico (Soto 2023).

3. Procedimiento

Para llevar a cabo este proyecto, el primer paso fue buscar el informe de ARESEP (ARESEP 2021), en el cual se incluían los anexos que contenían las tablas de datos necesarias para probar nuestro método de agrupamiento.

Para obtener estos datos en un formato de archivo CSV, utilizamos CHATGPT. Inicialmente, intentamos procesar todos los datos de una vez, pero notamos que no era posible debido a su cantidad. Por lo tanto, decidimos dividirlos en tres tramos para su procesamiento. Además, utilizamos el editor de texto de Visual Studio Code para realizar algunas modificaciones. Inicialmente, los datos estaban separados por punto y coma ";", pero al final los dejamos separados por coma ",".

Una vez obtenidos los datos, procedimos a procesarlos utilizando la biblioteca Pandas para su posterior análisis.

Antes de pasar a probar el método, fue necesaria la normalización de los datos, permite comparar y analizar los datos de manera más precisa. Este proceso permite comparar y

analizar los datos de manera más precisa al escalar los valores de las variables, asegurando que tengan una escala comparable. La normalización es fundamental ya que las variables pueden tener diferentes unidades o rangos de valores, lo que podría dificultar la comparación y el análisis de los datos.

Lo siguiente fue probar el método de k-means para diferentes valores de k, para la comparación se utilizó la medida de la inercia, la inercia mide qué tan bien fue agrupado un conjunto de datos por K-Means. Se calcula midiendo la distancia entre cada punto de datos y su centroide, elevando al cuadrado esta distancia y sumando estos cuadrados en un solo clúster, donde N es la cantidad de muestras, X es el valor de cada una de esas muestras y C es el centro del grupo (Dicicco 2022).

$$inertia = \sum_{i=1}^N (x_i - C_k)^2$$

Con ello se gráfico el "Método del codo", donde se visualizo cuales valores de k podrían ser útiles, buscando mantener el compromiso entre un bajo valor de inercia acompañado de un bajo número de clústeres K.

En base a los resultados obtenidos, se procedió a probar valores específicos de k. A partir de estos valores, generamos diversas gráficas con el fin de analizar de manera más detallada los resultados y determinar el valor óptimo de k.

Este mismo proceso se realizó tanto para los datos del AnexoA, como los del AnexoB.

4. Resultados

4.1. ANEXO A. Circuitos con frecuencia promedio por abonado y duración promedio de las interrupciones superior a los límites normativo

Primero se sacaron algunas gráficas iniciales para el Anexo A (1 y 2).

En base a estos datos parece que independientemente de la cantidad de abonados, la mayoría tiende a concentrarse en una esquina, con valores muchos más bajos que los del FPI.

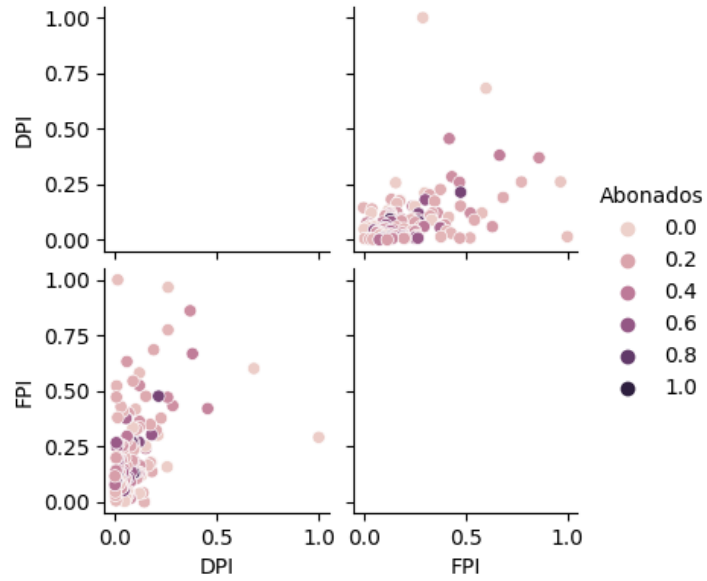


Figura 1: Visualización de Datos

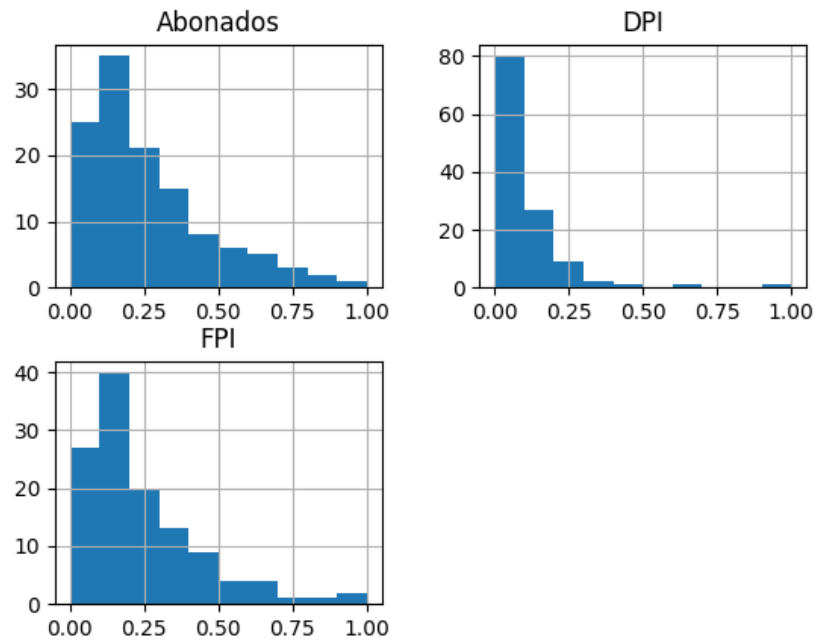


Figura 2: Visualización de Datos

Posteriormente, se procedió a obtener la gráfica del método del codo, la cual nos brindó una idea de los posibles valores de k que podrían resultar útiles para este conjunto de datos.

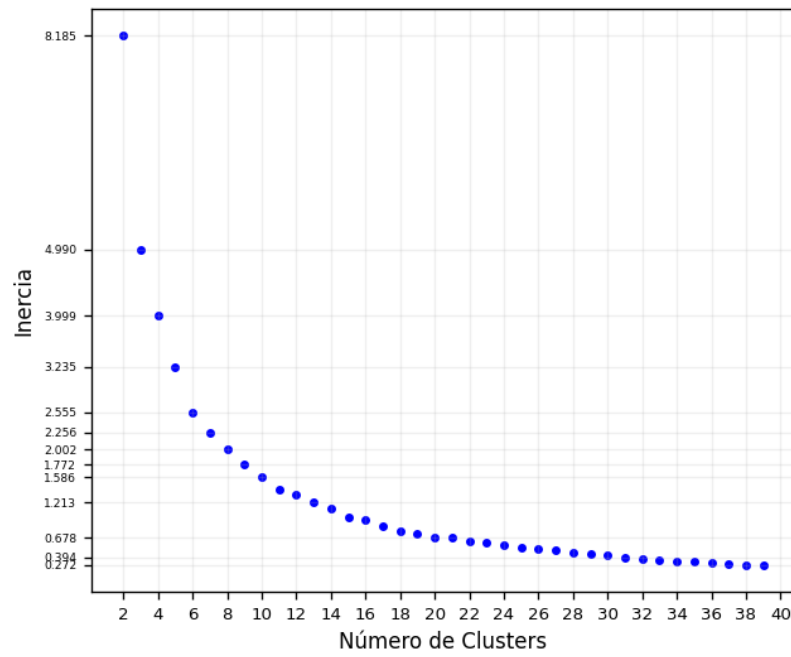


Figura 3: Método del codo para el AnexoA

- Para $k = 4$

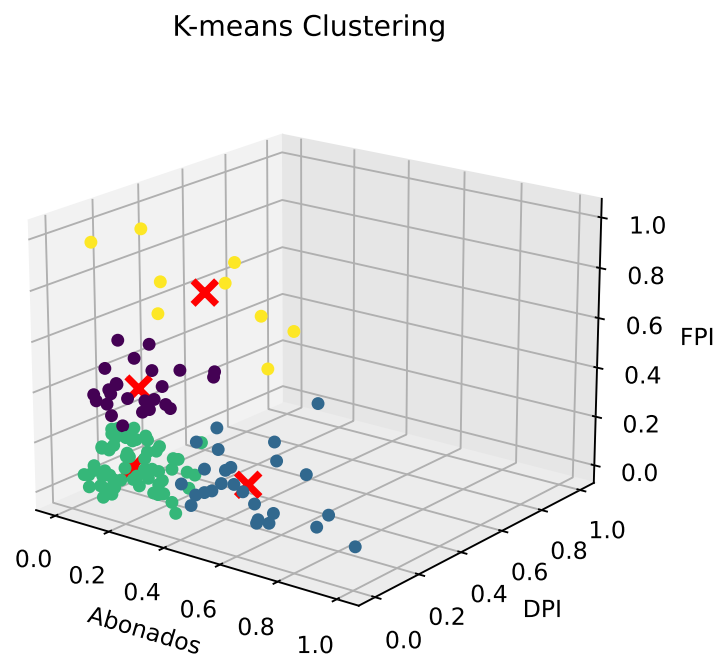


Figura 4: Gráfica en 3D para $k=4$

- Para $k = 5$

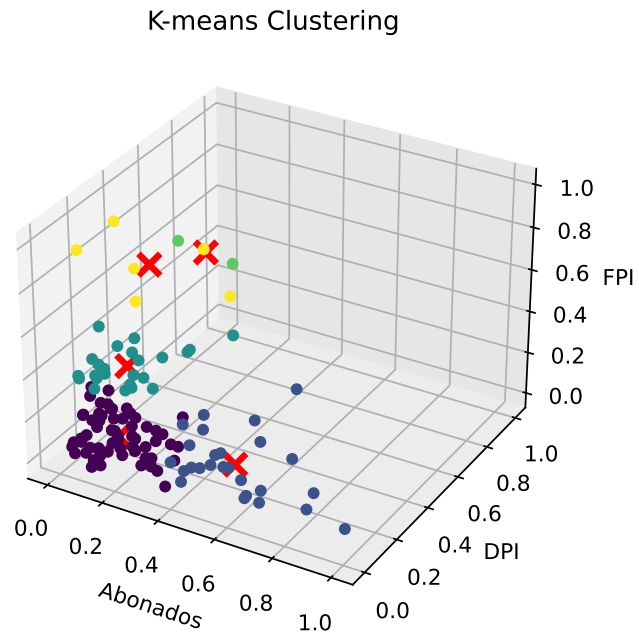


Figura 5: Gráfica en 3D para $k=5$

- Para $k = 6$

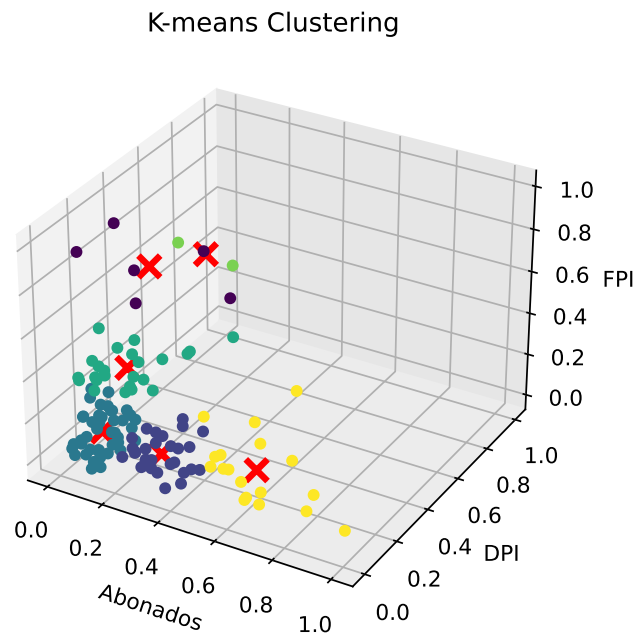


Figura 6: Gráfica en 3D para $k=6$

- Para $k = 7$

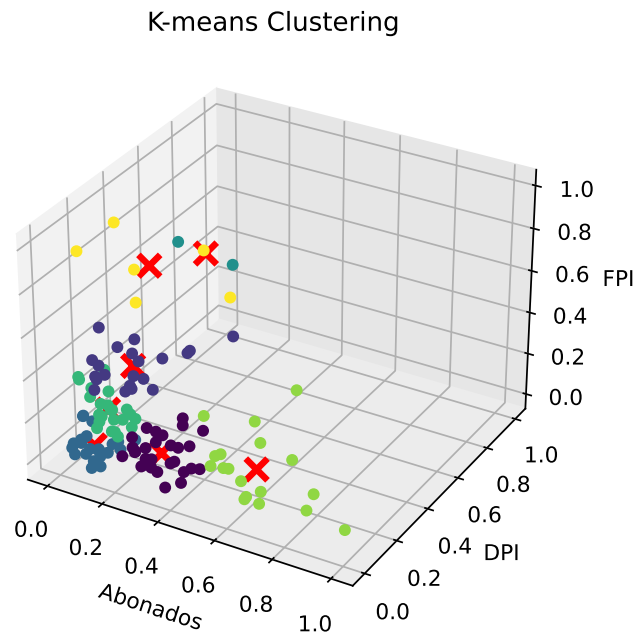


Figura 7: Gráfica en 3D para $k=7$

- Para $k = 8$

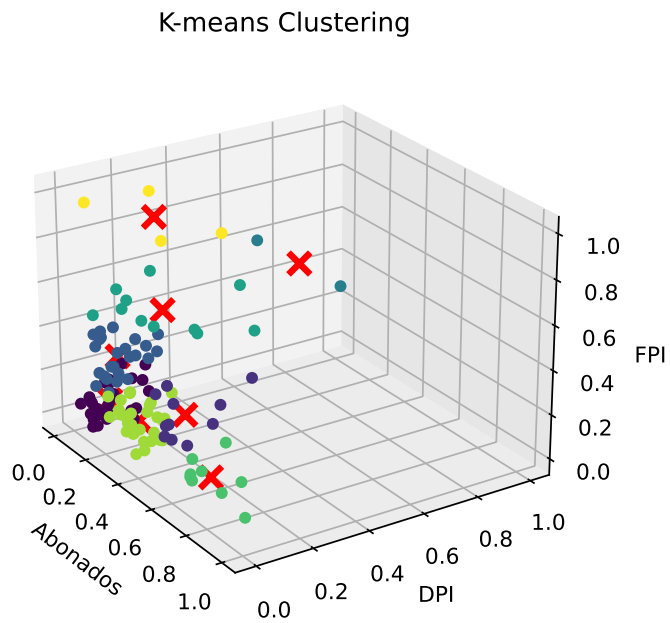


Figura 8: Gráfica en 3D para $k=8$

4.2. ANEXO B. Circuitos con frecuencia promedio por abonado o duración promedio de las interrupciones superior a los valores normativos

Primero se sacaron algunas gráficas iniciales para el Anexo B (9 y 10). En base a estos datos se observa que la mayoría tiende a concentrarse en en el medio, especialmente en el FPI.

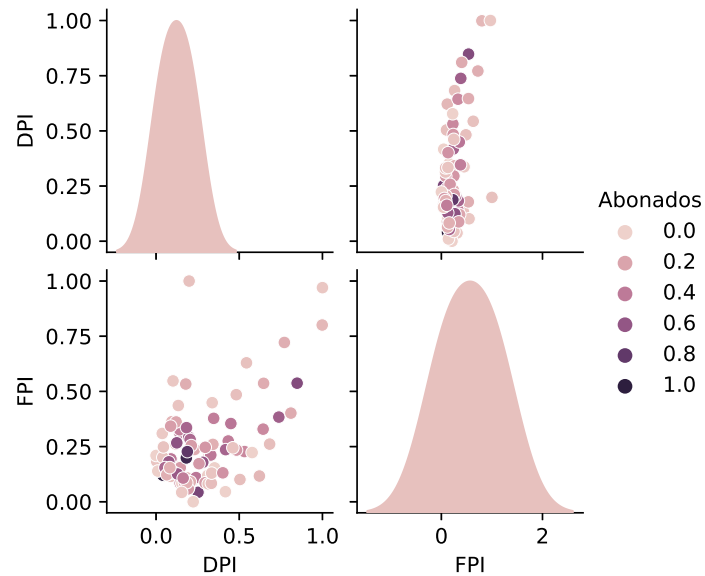


Figura 9: Visualización de Datos

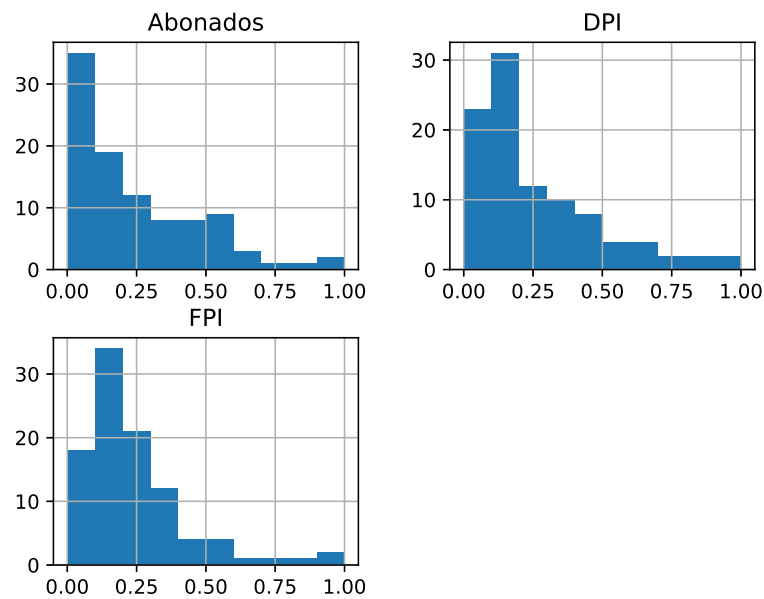


Figura 10: Visualización de Datos

Posteriormente, se procedió a obtener la gráfica del método del codo, la cual nos brindó una idea de los posibles valores de k que podrían resultar útiles para este conjunto de datos del anexo B.

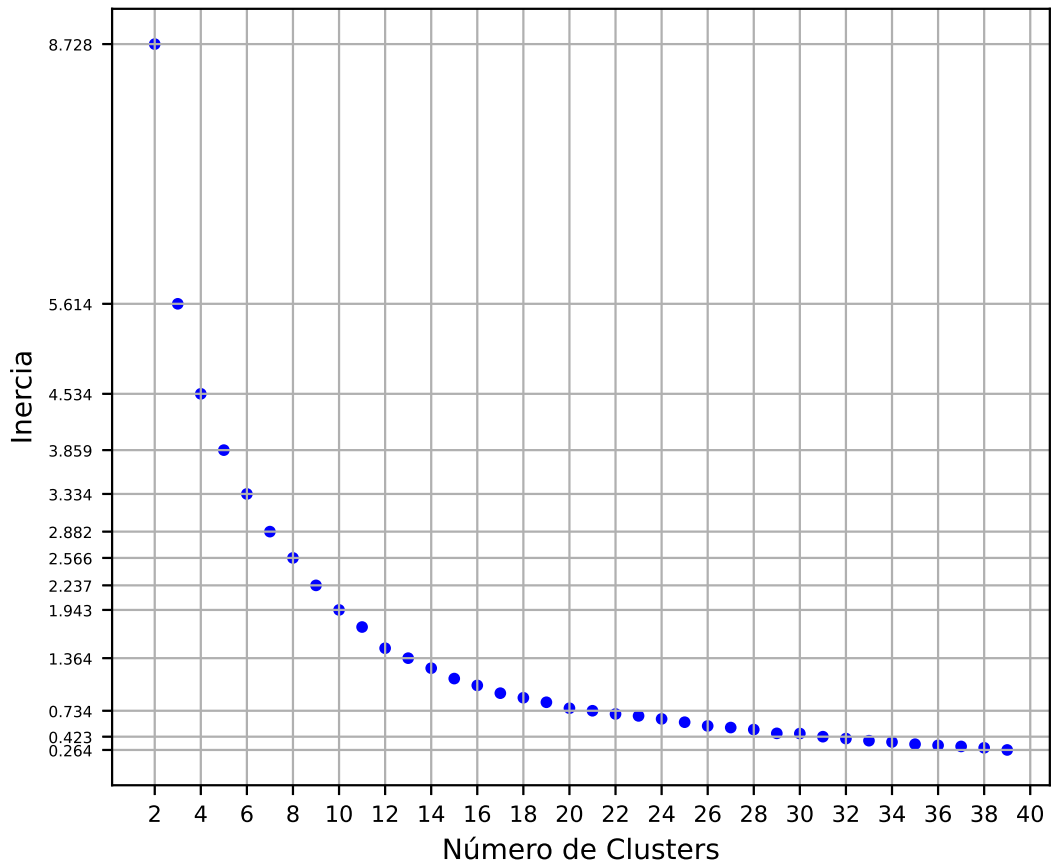


Figura 11: Método del codo para el AnexoB

- Para $k = 4$

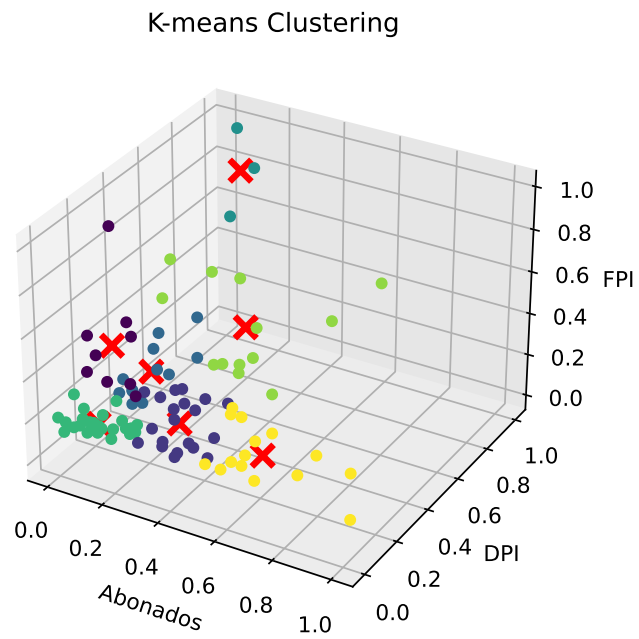


Figura 12: Gráfica en 3D para $k=4$

- Para $k = 5$

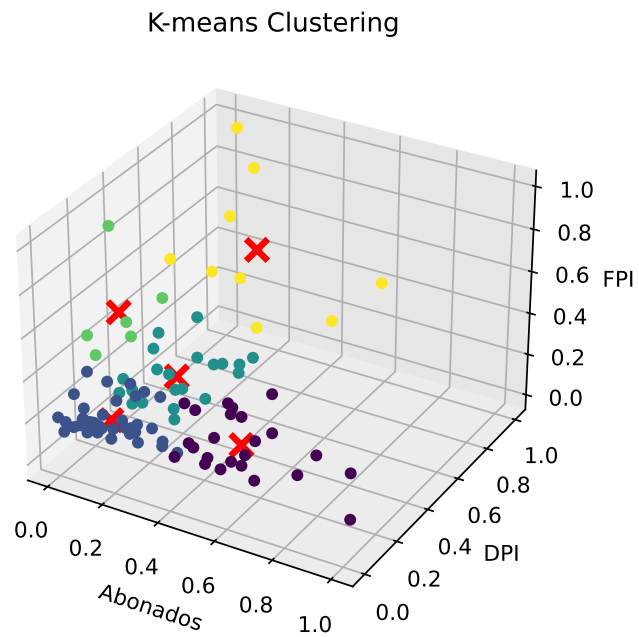


Figura 13: Gráfica en 3D para $k=5$

- Para $k = 6$

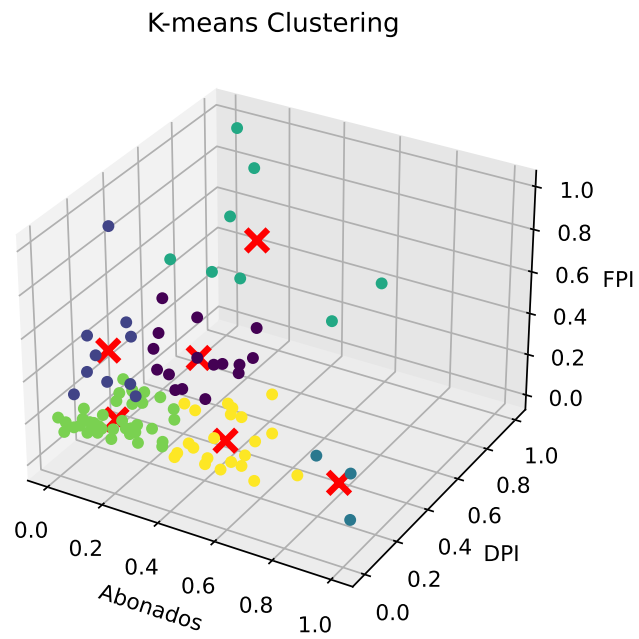


Figura 14: Gráfica en 3D para $k=6$

- Para $k = 7$

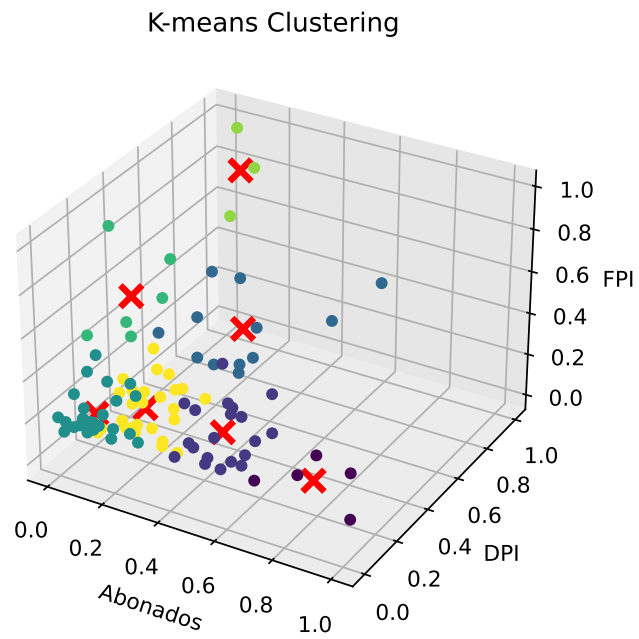


Figura 15: Gráfica en 3D para $k=7$

- Para $k = 8$

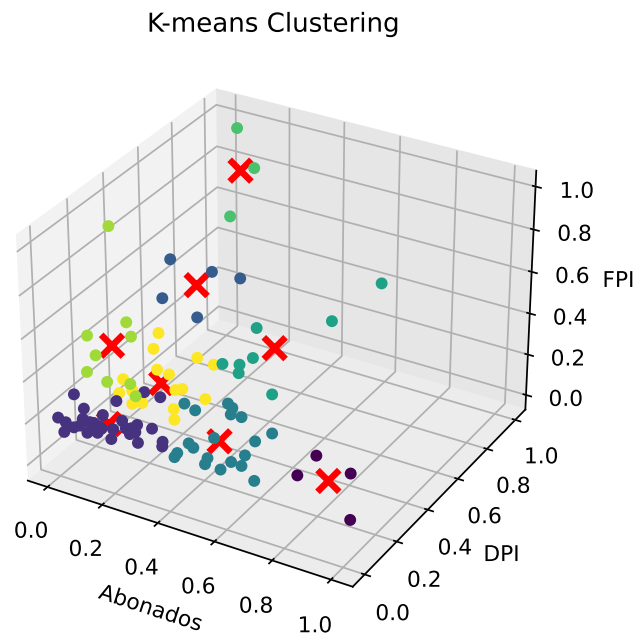


Figura 16: Gráfica en 3D para $k=8$

- Para $k = 9$

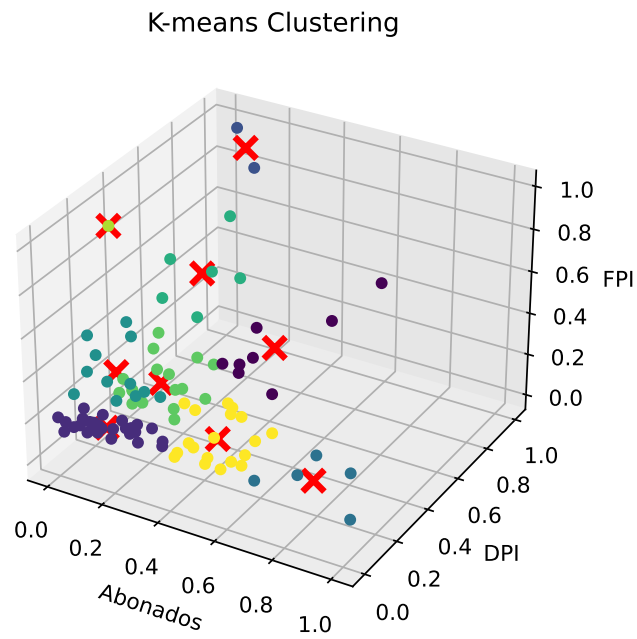


Figura 17: Gráfica en 3D para $k=8$

- Para $k = 10$

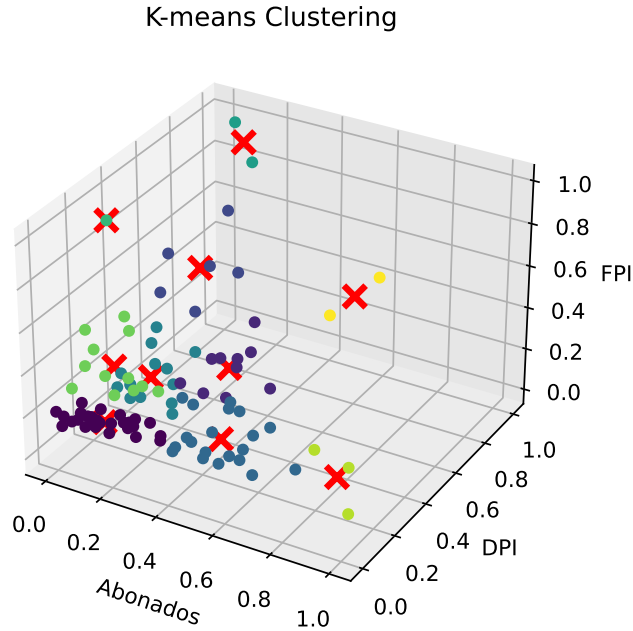


Figura 18: Gráfica en 3D para $k=8$

5. Conclusiones

En el caso del anexo A debido a que los datos tienden agruparse en la esquina donde los valores de DPI y FPI son menores, se observa que no parece útil usar mas de 8 centroides debido a que como se observa quedan algunos centroides demasiado dispersos, al mismo tiempo es importante mantener un bajo valor de k , mientras se posible, visualmente se observa que con 4 se obtiene una separación de grupos de datos necesaria, igualmente lo ideal serian hacer mas pruebas para demostrar que este valor k realmente es el ideal, para ello seria importante evaluar otras métricas.

Con respecto al ".^NEXO B", se observa una tendencia en los datos a agruparse alrededor de la media de los valores de DPI y FPI. Similar al caso anterior, para valores mayores de k , como 8, 9 o 10, parece que no son los más ideales, ya que resultan en centroides con muy pocos datos asignados, lo que podría aportar una cantidad limitada de información. Sin embargo, a diferencia del caso anterior, en este escenario los centroides parecen estar un poco más dispersos.

En este contexto, un valor de k igual a 5 podría ajustarse de manera más adecuada para la agrupación de datos en el ".^NEXO B". No obstante, sería recomendable realizar un análisis más exhaustivo utilizando otras métricas para evaluar de manera más precisa la elección del valor de k óptimo.

Otro detalle importante que debemos observar en los datos del .^ANEXO A.^{es} que estos cumplen con ambas condiciones de falla. Al analizar los grupos principales, notamos que tienden a concentrarse en las regiones con una menor cantidad de abonados. Estas regiones suelen corresponder a zonas rurales, lo que nos indica que tanto la frecuencia promedio por abonado como la duración promedio de las interrupciones son superiores a los límites normativos establecidos para este tipo de áreas.

Por otro lado en el .^ANEXO B", se incluyen los circuitos en los que al menos una de las condiciones se cumple, se observa que los datos se amontonan un poco menos en las áreas de menor cantidad de abonados,

De estos datos también podemos concluir que las áreas con mayor cantidad de abonados tienden a sufrir menos fallas, esta nos da un indicio de que existe una desigualdad en áreas con mayor o menor cantidad de población.

Gracias este trabajo se ha podido comprender la aplicabilidad de los sistemas de agrupamiento en el contexto de la ingeniería eléctrica, reconociendo su utilidad para analizar y clasificar datos relacionados con la calidad del servicio eléctrico.

También se aprendió a como trabajar con datos, desde su extracción hasta su procesamiento y posterior análisis mediante algoritmos de agrupamiento como k means.

Por ultimo, el utilizar diferentes métodos de agrupamiento puede ser útil para obtener una mayor cantidad de información, principalmente debido a que el problema no cuenta con muchos datos. Por lo tanto, aplicar diferentes algoritmos no generaría una gran carga computacional, pero podría ayudar a evaluar el problema desde diversas perspectivas y, de esta manera, lograr extraer información más confiable y precisa.

Referencias

- ARESEP. 2021. *Informe de calidad eléctrica del servicio de distribución 2021*. <https://aresep.go.cr/electricidad-calidad/informe-de-calidad-electrica-del-servicio-de-distribucion-2021/>. [Online; accessed Septiembre-2023].
- Dicicco. 2022. *Inertia For ML Applications*. <https://medium.com/@matthew.dicicco38/inertia-for-ml-applications-8c38de2d10d7>. [Online; accessed Septiembre-2023].
- Learning, Aprende Maching. 2018. *K-Means en Python paso a pasos*. <https://www.aprendemachinelearning.com/k-means-en-python-paso-a-paso/>. [Online; accessed Septiembre-2023].
- SCIJ. 2015. *Supervisión de la calidad del suministro eléctrico en baja y media tensión (norma técnica regulatoria AR-NT-SUCAL)*. http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?nValor1=1&nValor2=79374. [Online; accessed Septiembre-2023].
- Soto. 2023. *Partición de una Red Eléctrica de Distribución Aplicando Algoritmos de Agrupamiento K-means y DBSCAN*. <https://revistaenergia.cenace.gob.ec/index.php/cenace/article/view/572>. [Online; accessed Septiembre-2023].

Apéndice

https://github.com/katyvc29/Proyecto_IA_Sistema_de_agrupamiento.git

Código en python (Learning 2018):

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler # Normalizar
from sklearn.cluster import KMeans
import seaborn as sns
import seaborn as sb
from mpl_toolkits.mplot3d import Axes3D

# Leemos el archivo csv del anexoA y anexoB
dataA= pd.read_csv("anexoA.csv")
dataB = pd.read_csv("anexoB.csv")

# Seleccionamos las columnas 'Abonados', 'DPI' y 'FPI' del DataFrame dataA y
    dataB
datosA = dataA[['Abonados', 'DPI', 'FPI']]
datosB = dataB[['Abonados', 'DPI', 'FPI']]

datosA_i = dataA[['Empresa', 'Circuito']]
datosB_i = dataA[['Empresa', 'Circuito']]
# print(dataA.describe())

# Calcula los valores minimos y maximos para que est n en el rango [0,
    1]
escaladorA = MinMaxScaler().fit(datosA.values)
escaladorB = MinMaxScaler().fit(datosB.values)

# Datos normalizados
# Se aplica la transformaci n de escala a los datos utilizando el
    escalador previamente ajustado
datos_normA = pd.DataFrame(escaladorA.transform(datosA.values), columns = [
    'Abonados', 'DPI', 'FPI'])
datos_normB = pd.DataFrame(escaladorB.transform(datosA.values), columns = [
    'Abonados', 'DPI', 'FPI'])

# Graficas para Anexo A
datosA.drop([],1).hist()
datos_normA.drop([],1).hist()
plt.title('Histogramas_del_Anexo_A')
sb.pairplot(datos_normA.dropna(), hue='Abonados', height=2,vars=[ 'DPI', '
    FPI'], kind='scatter')
plt.show()
```

```

# Graficacion para observar el codo para el anexo A
inerciasA = []
for k in range(2, 40):
    # Generaci n de modelo
    kmeansA = KMeans(n_clusters=k).fit(datos_normA.values)
    inerciasA.append(kmeansA.inertia_)
#print(inerciasA)

# Codigo para imprimir la gr fica de codo del AnexoA
plt.figure(figsize=(6, 5), dpi=120)
plt.scatter(range(2, 40), inerciasA, marker="o", s=10, color="blue")
plt.xlabel("Numero_de_Clusters", fontsize=10) # Etiquetas para los ejes
    x e y
plt.ylabel("Inercia", fontsize=10)
plt.xticks(range(2, 41, 2)) # Definir los ticks (marcas) en el eje x
# Definir los ticks (marcas) en el eje y con valores espec ficos
yticks = inerciasA[0:9] + inerciasA[11:12] + inerciasA[19:20] + inerciasA
    [29:30] + inerciasA[37:38]
plt.yticks(yticks)
plt.yticks(fontsize=5.5) # Ajustar el tama o de las etiquetas en los
    ejes x e y
plt.xticks(fontsize=8)
plt.grid(True, alpha=0.2) # Agregar una cuadr cula al gr fico con
    transparencia
plt.show() # Mostrar el gr fico

# En base a lo obtenido en el metodo del codo se elige probar varios
    valores de k
modeloA_k1 = 4

# Generar modelo para el AnexoA con cierto valor de k
kmeansA1 = KMeans(n_clusters= modeloA_k1).fit(datos_normA.values)
labels = kmeansA1.labels_
labels = kmeansA1.predict(datos_normA.values)

datos_normA1 = datos_normA
datos_normA1["cluster"] = kmeansA1.labels_ # Nueva columna con los grupos
A_k1 = pd.concat([datosA_i, datos_normA1], axis=1) # Uniendo columnas
#print(A_k1) # Imprimiendo tabla completa

print("Imprimiento los centroides para k=", modeloA_k1 )
print("Centroides:\n", kmeansA1.cluster_centers_)
print("Inercia:", kmeansA1.inertia_)

# Plot the data points and their cluster assignments
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

```

```

ax.scatter(datos_normA.values[:, 0], datos_normA.values[:, 1], datos_normA.
    values[:, 2], c=labels, cmap='viridis')
ax.scatter(kmeansA1.cluster_centers_[:, 0], kmeansA1.cluster_centers_[:,
    1], kmeansA1.cluster_centers_[:, 2], marker='x', color='red', s=100,
    linewidths=3)
ax.set_title("K-means Clustering")
ax.set_xlabel('Abonados')
ax.set_ylabel('DPI')
ax.set_zlabel('FPI')
plt.show()

#####
# Graficas para Anexo B
datosB.drop([], 1).hist()
datos_normB.drop([], 1).hist()
plt.title('Histogramas del Anexo B')
sb.pairplot(datos_normB.dropna(), hue='Abonados', height=2, vars=[ 'DPI', '
    FPI'], kind='scatter')
plt.show()

# Graficacion para observar el codo para el anexo B
inerciasB = []
for k in range(2, 40):
    # Generacion de modelo
    kmeansB = KMeans(n_clusters=k).fit(datos_normB.values)
    inerciasB.append(kmeansB.inertia_)
#print(inerciasB)

# Codigo para imprimir la grafica de codo del AnexoA
plt.figure(figsize=(6, 5), dpi=120)
plt.scatter(range(2, 40), inerciasB, marker="o", s=10, color="blue")
plt.xlabel("Numero de Clusters", fontsize=10) # Etiquetas para los ejes
    x e y
plt.ylabel("Inercia", fontsize=10)
plt.xticks(range(2, 41, 2)) # Definir los ticks (marcas) en el eje x
# Definir los ticks (marcas) en el eje y con valores especificos
yticks = inerciasB[0:9] + inerciasB[11:12] + inerciasB[19:20] + inerciasB
    [29:30] + inerciasB[37:38]
plt.yticks(yticks)
plt.yticks(fontsize=5.5) # Ajustar el tamaño de las etiquetas en los
    ejes x e y
plt.xticks(fontsize=8)
plt.grid(True, alpha=0.2) # Agregar una cuadrícula al gráfico con
    transparencia
plt.show() # Mostrar el gráfico

```

```

# En base a lo obtenido en el metodo del codo se elige probar varios
  valores de k
modeloB_k1 = 4

# Generar modelo para el AnexoA con cierto valor de k
kmeansB1 = KMeans (n_clusters= modeloB_k1).fit(datos_normB.values)
labels = kmeansB1.labels_
labels = kmeansB1.predict(datos_normB.values)

datos_normB1 = datos_normB
datos_normB1["cluster"] = kmeansB1.labels_ # Nueva columna con los grupos
B_k1 = pd.concat([datosA_i, datos_normB1], axis=1) # Uniendo columnas
#print(B_k1) # Imprimiendo tabla completa

print("Imprimiendo los centroides para k=", modeloB_k1 )
print("Centroides:\n", kmeansB1.cluster_centers_)
print("Inercia:", kmeansB1.inertia_)

# Plot the data points and their cluster assignments
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(datos_normB.values[:, 0], datos_normB.values[:, 1], datos_normB.
  values[:, 2], c=labels, cmap='viridis')
ax.scatter(kmeansB1.cluster_centers_[:, 0], kmeansB1.cluster_centers_[:,
  1], kmeansB1.cluster_centers_[:, 2], marker='x', color='red', s=100 ,
  linewidths=3)
ax.set_title("K-means Clustering")
ax.set_xlabel('Abonados')
ax.set_ylabel('DPI')
ax.set_zlabel('FPI')
plt.show()

```