

# UCUENCA

## Universidad de Cuenca

Facultad de Ingeniería

Carrera de Ingeniería Eléctrica

### **Reducción de espacio de búsqueda usando algoritmos de aprendizaje no supervisado aplicado al problema de la expansión del sistema de transmisión de energía eléctrica**


Trabajo de titulación previo a la obtención del título de Ingeniero Eléctrico

**Autor:**

Wilson Patricio Minchala Naula

**Director:**

Santiago Patricio Torres Contreras

ORCID:  0000-0002-8803-6811

**Cuenca, Ecuador**

2023-08-02

## Resumen

La demanda de energía eléctrica actual requiere de adiciones de infraestructura en los sistemas de transmisión que se vuelven más grandes cada vez. Una planificación de la expansión del sistema de transmisión (TEP) se encarga de identificar las adiciones necesarias en el sistema eléctrico. Sin embargo, ese proceso resulta difícil debido a la cantidad de variables a tratar. Estas variables son el resultado del número líneas candidatas que se consideran (espacio de búsqueda) en un modelo de optimización. En este trabajo se presenta una estrategia de agrupación de líneas candidatas con el objetivo de reducir este espacio de búsqueda. Esta estrategia combina herramientas del aprendizaje no supervisado como el clustering y el análisis de la operación del Sistema de Suministro de Energía Eléctrica (SSEE) llamados flujos óptimos de potencia. Esta combinación sirve para clasificar las líneas candidatas bajo 3 criterios: sobrecarga, mínimo esfuerzo y costo-beneficio. Estos criterios se aplican a cada sistema de estudio que se analiza: Garver 6 Barras, IEEE 24 Barras y IEEE 188 Barras, para determinar el más apropiado en cada caso. Esto con el fin de formar e identificar grupos de líneas poco representativos para descartarlos, dando como resultado un nuevo espacio de búsqueda reducido para los sistemas y conseguir una mejora significativa en la eficiencia del proceso de Planificación de la Expansión de Sistemas de Transmisión.

*Palabras clave:* clustering, costo-beneficio, método elbow, planeamiento de la expansión de los sistemas de transmisión



El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Cuenca ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por la propiedad intelectual y los derechos de autor.

**Repositorio Institucional:** <https://dspace.ucuenca.edu.ec/>

### Abstract

The current demand for electrical energy requires infrastructure additions to transmission systems that are becoming larger and larger. A transmission system expansion planning (TEP) is responsible for identifying necessary additions to the power system. However, this process is difficult due to the number of variables to deal with. These variables are the result of the number of candidate lines to consider (search space) in an optimization model. In this paper, a candidate line clustering strategy is presented with the objective of reducing this search space. This strategy combines unsupervised learning tools such as clustering and the analysis of the operation of the Electric Power Supply System (ESS) called optimal power flows. This combination is used to classify the candidate lines under 3 criteria: overload, least effort and cost-benefit. These criteria are applied to each study system under analysis: Garver 6 Buses, IEEE 24 Buses and IEEE 188 Buses, to determine the most appropriate in each case. All this in order to form and identify unrepresentative line groups to discard, resulting in a new reduced search space for the systems and achieving a significant improvement in the efficiency of the Transmission System Expansion Planning process.

*Keywords:* clustering, cost-benefit, elbow method, transmission expansion planning



The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

**Institutional Repository:** <https://dspace.ucuenca.edu.ec/>

## Índice de contenido

Nomenclatura .....	7
Introducción .....	9
Estado del Arte .....	12
Metodología .....	13
Flujo Óptimo de Potencia AC.....	13
Criterios Principales.....	15
<i>Criterio de Sobrecarga (Overload)</i> .....	15
<i>Criterio de Costo-Beneficio</i> .....	15
<i>Criterio de Mínimo Esfuerzo (Least Effort)</i> .....	15
Clustering .....	16
Tipos de clustering.....	16
Medidas de Distancia: .....	16
<i>Distancia Euclidiana</i> .....	16
<i>Distancia Manhattan</i> .....	16
K-means .....	17
Método del Codo (Elbow) .....	18
Procedimiento.....	19
Creación de la base de datos .....	19
Normalización de los datos.....	20
Proceso de clusterización.....	20
Resultados y Discusión.....	21
Presentación de Resultados.....	21
<i>a) Sistema Garver 6 Barras</i> .....	21
<i>b) Sistema IEEE 24 Barras</i> .....	24
<i>c) Sistema IEEE 118 Barras</i> .....	26
Análisis y Discusión.....	27
Conclusiones .....	31
Referencias.....	32
Anexos.....	36
Anexo 1: Datos y espacio de búsqueda del Sistema Garver 6 Barras, IEEE 24 Barras e IEEE 118 Barras.....	36
Anexo 2: Grupos de líneas resultantes de la clusterización con el coeficiente de Costo-Beneficio para el sistema IEEE 118 Barras .....	38

### Índice de figuras

Figura 1. Estado del Sistema Garver previo a una planificación de la expansión de su sistema de transmisión .....	9
Figura 2. Espacio de búsqueda para la expansión del Sistema Garver.....	10
Figura 3. Formación de grupos con las líneas mediante un criterio de clasificación.....	11
Figura 4. Nuevo espacio de búsqueda candidato para el problema TEP. ....	11
Figura 5. Distancia Manhattan vs Distancia Euclidiana .....	17
Figura 6. Método Elbow para un conjunto de datos .....	18
Figura 7. Método Elbow con datos normalizados. ....	19
Figura 8. Método Elbow para el Sistema Garver 6 Barras.....	21
Figura 9. Clusterización utilizando el Coeficiente de sobrecarga en el Sistema Garver 6 Barras .....	22
Figura 10. Combinación de Criterios Sobrecarga vs Mínimo Esfuerzo para el Sistema Garver 6 Barras .....	24
Figura 11. Método Elbow para el Sistema IEEE 24 Barras.....	25
Figura 12. Clusterización usando el criterio de Costo-Beneficio en el Sistema IEEE 24 Barras .....	25
Figura 13. Método Elbow para el Sistema IEEE 118 Barras.....	27
Figura 14. Clusterización usando el criterio de Costo-Beneficio Sistema IEEE 118 Barras. 27	
Figura 15. Clusterización usando el criterio Sobrecarga vs Mínimo Esfuerzo para el Sistema IEEE 24 Barras .....	28

**Índice de tablas**

Tabla 1. Grupos de líneas resultado de la Clusterización con el coeficiente de Sobrecarga	22
Tabla 2. Grupos de líneas resultado de la clusterización con el coeficiente de Costo-Beneficio .....	23
Tabla 3. Grupos de líneas resultado de la clusterización con el coeficiente de Mínimo Esfuerzo.....	23
Tabla 4. Grupos de líneas resultante de la combinación del coeficiente de Mínimo Esfuerzo y Sobrecarga .....	24
Tabla 5. Grupos de líneas resultantes de la clusterización con el coeficiente de Costo-Beneficio para el sistema IEEE 24 Barras.....	26
Tabla 6. Espacio de búsqueda reducido para el Sistema Garver 6 Barras.....	29
Tabla 7. Espacio de búsqueda reducido para el Sistema IEEE 24 Barras.....	29
Tabla 8. Espacio de búsqueda reducido para el Sistema IEEE 118 Barras.....	30
Tabla 9. Datos principales y espacio de búsqueda del Sistema Garver 6 Barras.....	36
Tabla 10. Datos principales y espacio de búsqueda del Sistema IEEE 24 Barras.....	36
Tabla 11. Datos principales y espacio de búsqueda del Sistema IEEE 118 Barras.....	37
Tabla 12. Grupos de líneas resultantes de la clusterización con el coeficiente de costo-beneficio para el sistema ieee 118 Barras.....	38

## Nomenclatura

### Índices y Conjuntos

$k$	Índice de número de agrupaciones.
$i, j$	Índices de las Barras AC.
$l$	Índice de los paths candidatos.

### Funciones

$f(u)$	Costo de generación (Función Objetivo).
$h(u)$	Límites de flujo de potencia en una ruta.
$g(u)$	Ecuaciones de equilibrio de potencia.
$gP$	Función de Ángulo para la potencia activa.
$gQ$	Función de Ángulo para la potencia reactiva.
$f_P^l$	Función de Costo polinomial de inyección de potencia activa.
$f_Q^l$	Función de Costo polinomial de inyección de potencia reactiva.
$h_f$	Función no lineal de magnitudes y ángulos para un bus extremo f.
$h_t$	Función no lineal de magnitudes y ángulos para un bus extremo t.
$\min ()$	Función para encontrar un valor mínimo en un vector.
$\max ()$	Función para encontrar un valor máximo en un vector.

### Variables

$\theta$	Vector de ángulos de fase.
$V_m$	Modulo de Voltaje.
$n_b$	Número de Barras.
$n_l$	Número de líneas del conjunto solución.
$n_r$	Número de líneas para el nuevo espacio de búsqueda.
$O_i^{load}$	Coefficiente de Sobrecarga.
$S_{ij}$	Flujo de Potencia para una línea entre el bus i y bus j.
$S_{ij}^{max}$	Flujo de Potencia máximo en una línea.
$C_{ij}^{max}$	Costo de construcción de una línea entre el bus i y bus j.
$B_i^{cb}$	Coefficiente Costo-Beneficio.
$d_m$	Distancia Manhattan.
$R_{ij}$	Resistencia de una línea candidata entre 2 Barras i-j.

$d_e$	Distancia Manhattan.
$x_i$	Coordenadas para un punto A.
$y_i$	Coordenadas para un punto B.
$\sigma_i^{le}$	Coefficiente de mínimo esfuerzo.
$\gamma_l$	Susceptancia para una línea candidata.
$X_{ij}$	Reactancia de una línea candidata entre 2 Barras i-j.

## Parámetros

$u$	Vector de Optimización de costos de generación.
$P_g$	Vector de potencia activa de generación.
$Q_g$	Vector de potencia reactiva de generación.
$v_m^{l,min}$	Límite inferior para un bus de voltaje
$v_m^{l,max}$	Límite superior para un bus de voltaje
$P_g^{i,min}$	Límite inferior de potencia activa de generación.
$P_g^{i,max}$	Límite superior de potencia activa de generación.
$q_g^{i,min}$	Límite inferior de potencia reactiva de generación.
$q_g^{i,max}$	Límite superior de potencia reactiva de generación.
$A$	Vector columna de dimensión $N \times 1$
$A_+$	Vector columna de dimensión $N \times 1$ completamente positivo.
$A_N$	Vector columna de dimensión $N \times 1$ normalizado.



### Introducción

El aumento de la población junto con la evolución del sector industrial y comercial provocan que la demanda de energía eléctrica crezca de manera considerable. Esta demanda debe estar cubierta en su totalidad con las unidades de generación, pero el Sistema de Suministro de Energía Eléctrica (SSEE) debe contar con la capacidad de llevar esta cantidad de energía desde los puntos de generación hasta los sistemas de subtransmisión y distribución (Palacios, 2020). Esto se logra con una adecuada Planificación del Sistema de Transmisión (por sus siglas en inglés Transmission Expansion Planning, TEP) la misma que tiene como objetivo fortalecer el SSEE para que funcione de una manera estable y confiable; y de esta forma mantener un perfil de operación normal ante una inesperada contingencia (Gallego Pareja et al., 2009).

La Planificación de la Expansión de Sistemas de Transmisión busca determinar la cantidad óptima de líneas necesarias y su ubicación para lograr una expansión eficiente del Sistema Eléctrico de Transmisión, garantizando su correcto funcionamiento y minimizando los costos de inversión. En la Figura 1. se presenta el estado inicial de un sistema que requiere una expansión (Gallego-Pareja et al., 2009).

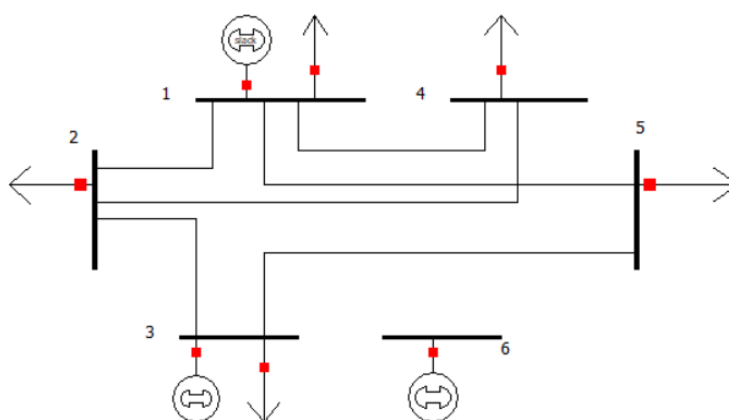


Figura 1. Estado del Sistema Garver previo a una planificación de la expansión de su sistema de transmisión

Entre las técnicas más recientes utilizadas para el TEP se destaca el estudio de los flujos de potencia en corriente alterna (AC) y corriente continua (DC), los cuales permiten analizar los niveles de tensión, la capacidad de carga y las desviaciones angulares del voltaje. Además, están los modelos matemáticos y heurísticos de optimización, estas técnicas consisten en procedimientos que resuelven la formulación matemática del problema de planificación óptima de la expansión de la red eléctrica (TEP). Sin embargo, es importante destacar que

no se puede alcanzar una optimización perfecta debido a que no se tienen en cuenta todas las restricciones y consideraciones que plantea un sistema real, como aspectos financieros y detalles técnicos específicos (Escudero Delgado, 2018).

En la actualidad, la aplicación de una u otra técnica implica un notable esfuerzo computacional debido al incremento en el tamaño de los sistemas, es decir, el espacio de búsqueda para un sistema aumenta exponencialmente con el número de barras que tenga el mismo, a consecuencia de la cantidad de líneas candidatas que pueden considerarse entre las distintas barras. Esto conlleva un aumento significativo en la cantidad de datos, variables y procesos involucrados (Lumbreras & Ramos, 2016). Esta situación se puede apreciar de manera más clara en la Figura 2. donde el espacio de búsqueda de un sistema se resalta de color rojo.

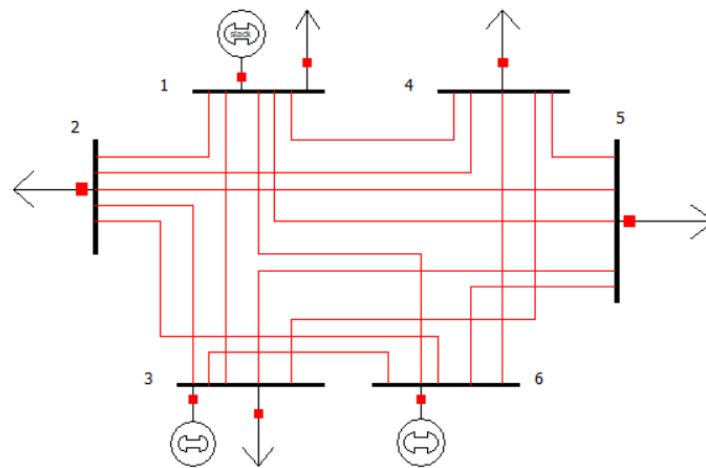


Figura 2. *Espacio de búsqueda para la expansión del Sistema Garver.*

En este trabajo se lleva a cabo un análisis del espacio de búsqueda para una solución al TEP. Consiste en calcular diversos indicadores para las líneas candidatas (rutas), que proporcionen una visión del estado operativo del sistema. Esta información clasifica las líneas según un criterio: Sobrecarga, Costo-beneficio, Mínimo Esfuerzo, para formar grupos como se observa en la Figura 3., donde se forman 3 grupos de líneas (verde, azul y magenta), y seleccionar al que reúna las mejores características para un determinado criterio de clasificación.

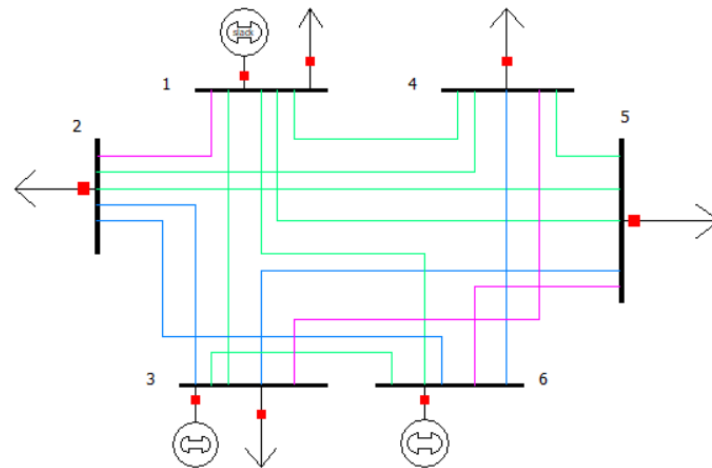


Figura 3. Formación de grupos con las líneas mediante un criterio de clasificación

En la figura 4 se ha descartado 2 grupos y se selecciona uno solo como el candidato para buscar la solución de la expansión del sistema de transmisión. Este grupo de líneas representado en la Figura 4. contiene a la solución final para el TEP que se obtendría con métodos computacionales como las metaheurísticas. Este análisis se realiza con el objetivo de reducir el número de rutas que ingresan como variables a los algoritmos de optimización de los modelos matemáticos, aliviando los procesos y reduciendo sus tiempos de ejecución. Los Sistemas que son analizados en este estudio son: Garver de 6 Barras, IEEE de 24 Barras e IEEE de 118 Barras para determinar el mejor criterio de clasificación para cada uno de ellos (Matute et al., 2020).

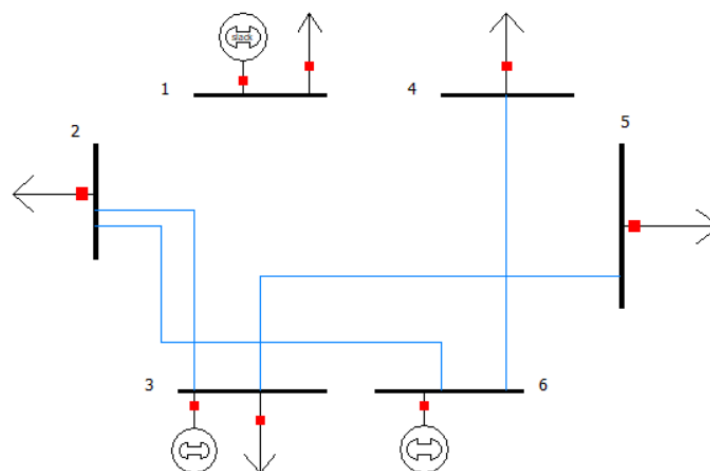


Figura 4. Nuevo espacio de búsqueda candidato para el problema TEP.

### Estado del Arte

En trabajos previos, una estrategia para reducir el espacio de búsqueda se menciona en (Torres R. et al., 2018), donde se aplican dos herramientas para reducir este espacio para el problema TEP. La primera herramienta conocida como “Partición” consiste en dividir al sistema completo en al menos 2 subsistemas, eliminando líneas y dejando barras aisladas. Por esta razón se deben identificar aquellas barras donde el grado de conexión no es muy fuerte eliminando el menor número de líneas posibles para crearlos y evitando que los subsistemas presenten un desbalance entre demanda y generación, lo cual no siempre es posible. Esto se soluciona con la segunda herramienta llamada “Inyección”, la cual suministra potencia activa los subsistemas que lo requieran, este flujo de potencia ingresa por los Barras donde se eliminaron líneas de transmisión. Cuando un subsistema requiera satisfacer su propia demanda, se convierte en una barra de demanda equivalente para el sistema completo, y cuando un subsistema está en condiciones de inyectar potencia a otros, se lo representa como una barra de generación.

El siguiente paso es encontrar y minimizar las pérdidas, para realizar esto se aplica un modelo MILP (Programación Lineal Entera Mixta) para cada subsistema (Alguacil et al., 2003). Dado que algunos subsistemas podrían presentar déficit al momento de satisfacer su propia demanda y las pérdidas, la función objetivo se modifica agregando el deslastre de carga como una variable temporal. El objetivo es reducir a cero el deslastre de carga, lo que conlleva un refuerzo en el número de rutas en cada subsistema. La cantidad de rutas que forman parte de la solución de cada subsistema se convierten en el número máximo de líneas para ese corredor. Al final si una ruta nunca se analizó en un subsistema, simplemente toma el número de adiciones original.

Otra estrategia se plantea en (Dominguez et al., 2017), que reduce el tamaño de la topología inicial de un sistema utilizando un modelo disyuntivo lineal modificado para que se tomen en cuenta algunas contingencias (Duque et al., 2014). Se plantea la búsqueda de todas las soluciones candidatas que podrían tener efectos positivos en el sistema, eliminado así las menos importantes, esta estrategia va enfocada a los sistemas de gran dimensión con multietapa y restricciones de seguridad (da Silva et al., 2015). La cual sigue los siguientes pasos:

1. Buscar todas las soluciones relevantes que podrían causar un efecto significativo para resolver cada TEP de etapa simple.
2. Determinar todos los posibles candidatos que resulten insignificantes para cada etapa del problema TEP.

Para eso se crea un conjunto de soluciones "pool", y los candidatos que nunca aparecen en el pool, se etiquetan como insignificantes.

3. Redefinir el valor de líneas para cada corredor, eliminando aquellas que sean insignificantes y, además, disminuir el número máximo de adiciones en paralelo permitido de la base de datos original del sistema en estudio.

Con esta simple idea se determina una solución total de calidad para el TEP multietapa considerando restricciones de seguridad e incorporando más detalles que mejoran la precisión del modelo.

Como se evidencia, ambas estrategias reducen el espacio de búsqueda al considerar únicamente el número de adiciones en paralelo de las líneas. Sin embargo, hasta el momento no se ha explorado directamente la estrategia de selección de un grupo específico de rutas como reducción del conjunto solución al TEP.

### Metodología

#### Flujo Óptimo de Potencia AC

El problema de flujo óptimo de potencia AC es un modelo de optimización no lineal que mejora el despacho del flujo de potencia en un sistema eléctrico en un solo periodo, dado el estado del sistema en el periodo anterior. El algoritmo se implementa como una extensión del paquete MATPOWER que permite calcular flujos de potencia y flujos de potencia óptimos en MATLAB. A continuación, se describe la versión estándar del problema de optimización AC-OPF que resuelve MATPOWER (Zimmerman et al., 2011).

$$\mathbf{min} v \rightarrow (f(v)) \quad (1)$$

Sujeto a:

$$g(v) = 0 \quad (2)$$

$$h(v) \leq 0 \quad (3)$$

$$v_{min} \leq v \leq v_{max} \quad (4)$$

La función objetivo  $f(v)$  consiste en el costo polinómico de las inyecciones del generador, las restricciones de igualdad  $g(v)$  son las ecuaciones de equilibrio de potencia, las restricciones de desigualdad  $h(v)$  son los límites de flujo en el tramo, y los límites  $v_{min}$  y  $v_{max}$  incluyen ángulos de barra de referencia, magnitudes de voltaje e inyecciones del generador.

El vector de optimización  $v$  para el problema de AC-OPF consiste en los vectores de  $n_b \times 1$  de ángulos de voltaje  $\theta$  y magnitudes  $V_m$ , y los vectores  $n_g \times 1$  de inyecciones de potencia activa y reactiva  $P_g$  y  $Q_g$  respectivamente.

$$v = \begin{bmatrix} \theta \\ V_m \\ P_g \\ Q_g \end{bmatrix} \quad (5)$$

La función objetivo  $f(v)$  en la ecuación 1 es una suma de funciones de costo polinomiales individuales de inyecciones de potencia activa y reactiva, respectivamente, para cada generador:

$$f(P_g, Q_g) = \sum_{l=1}^{n_g} f_p^l(p_g^l) + f_Q^l(q_g^l) \quad (6)$$

Las restricciones de igualdad en la ecuación 2 son el conjunto completo de  $2 \times n_b$  ecuaciones no lineales de balance de potencia activa y reactiva,  $P_{bus}$  y  $Q_{bus}$  respectivamente, expresadas como funciones de los ángulos  $\theta$  y las magnitudes de voltaje  $V_m$ , las inyecciones del generador  $P_g$  y  $Q_g$  y la matriz de conexión  $C_g$ , donde  $P_d$  y  $Q_d$  son la cantidad de potencia activa y reactiva consumida en una barra, respectivamente.

$$gP(\Theta, V_m, P_g) = P_{bus}(\theta, V_m) + P_d - C_g P_g = 0 \quad (7)$$

$$gQ(\Theta, V_m, Q_g) = Q_{bus}(\theta, V_m) + Q_d - C_g Q_g = 0 \quad (8)$$

Las restricciones de desigualdad en la ecuación 3 consiste en dos conjuntos de  $n_l$  límites de flujo del tramo como funciones no lineales de las magnitudes y ángulos de voltaje de la barra, una para el extremo  $f$  y otra para el extremo  $t$  de cada tramo:

$$h_f(\Theta, V_m) = |F_f(\Theta, V_m)| - F_{max} \leq 0 \quad (9)$$

$$h_t(\Theta, V_m) = |F_t(\Theta, V_m)| - F_{max} \leq 0 \quad (10)$$

Los límites variables de la ecuación 4 incluyen una restricción de igualdad en cualquier ángulo de barra de referencia y los límites superior e inferior en todas las magnitudes de voltaje de barra e inyecciones de potencia activa y reactiva del generador.

$$\theta_l^{ref} \leq \theta_l \leq \theta_l^{ref}, l \in I_{ref} \quad (11)$$

$$v_m^{l,min} \leq v_m^l \leq v_m^{l,max}, i = l \dots n_b \quad (12)$$

$$p_g^{l,min} \leq p_g^l \leq p_g^{l,max}, l = 1 \dots n_b \quad (13)$$

$$q_g^{l,min} \leq q_g^l \leq q_g^{l,max}, i = 1 \dots n_b \quad (14)$$

## Criterios Principales

### **Criterio de Sobrecarga (Overload)**

Este criterio se basa en la capacidad de conducción conocida de las líneas de transmisión. Se utiliza la relación entre el flujo de potencia que atraviesa una línea y su capacidad de conducción para clasificarlas y determinar un grupo que se encuentra comprometido con una sobrecarga cuando la demanda de energía aumente en el futuro. Este grupo se considera como candidato para una posible expansión del sistema de transmisión (Matute et al., 2020).

$$O_l^{load} = \frac{S_{ij}}{S_{ij}^{max}} \quad (15)$$

### **Criterio de Costo-Beneficio**

El costo de construcción de una ruta es un dato relevante en un problema de TEP, pero no es el único factor a considerar. Seleccionar únicamente las rutas de menor costo no garantiza encontrar la solución óptima. Además del costo de la línea, es necesario analizar el flujo de potencia que circulará a través de las mismas. Para ello, se utiliza un coeficiente conocido como costo-beneficio, el cual se expresa en [\$/MW] (Matute et al., 2020). Este coeficiente se calcula mediante la ecuación:

$$B_l^{cb} = \frac{C_{ij}}{S_{ij}} \quad (16)$$

### **Criterio de Mínimo Esfuerzo (Least Effort)**

El criterio del mínimo esfuerzo es una estrategia de planificación de transmisión que considera el patrón de distribución de flujos en el SSEE. El camino de menor esfuerzo es aquel que logra una mejor distribución de los flujos de potencia en el sistema. Para seleccionar este camino, se utiliza el mayor coeficiente de sensibilidad de todas las rutas (Monticelli et al., 1982). El coeficiente para el criterio de mínimo esfuerzo se calcula con la siguiente ecuación:

$$\sigma_l^{le} = -\frac{1}{2} \gamma_l \theta_l^2 \quad (17)$$

La susceptancia de todos los derechos de vía (path) se obtiene de esta forma:

$$\gamma_l = -\frac{X_{jk}}{R_{jk}^2 + X_{jk}^2} \quad (18)$$

## Clustering

El término “Clustering” que proviene del inglés, se refiere a la agrupación de datos. Es una de las técnicas fundamentales del aprendizaje no supervisado que busca identificar una estructura en un conjunto de datos no etiquetados. Un clúster se define como un conjunto de datos que comparten un grado de similitud entre sí y que difieren de los datos pertenecientes a otros conjuntos. El análisis se basa en diversas medidas de diferencia entre objetos y utiliza funciones de distancia para llevar a cabo la clasificación de los modelos. Si la clasificación es verdaderamente distintiva o no, depende de la distribución de los vectores de características del patrón (Omran et al., 2007).

### *Tipos de clustering*

Los algoritmos de clustering de datos se pueden clasificar en jerárquicos y particionales. Los algoritmos jerárquicos encuentran clústeres de manera sucesiva utilizando grupos previamente establecidos, es decir, construyen una estructura jerárquica de los mismos. Por otro lado, los algoritmos particionales determinan todos los grupos al mismo tiempo asignando cada dato a un conjunto sin construir una estructura jerárquica (Pascual et al., s. f.).

### **Medidas de Distancia:**

Un paso clave para el método de clustering es la selección de una medida de distancia, entre las principales se tiene:

#### ***Distancia Euclidiana***

Una medida común es la distancia euclidiana, que se calcula hallando el cuadrado de la distancia entre cada variable, sumando los cuadrados y hallando la raíz cuadrada de esa suma. Muchos estudios y resultados de investigación determinan que la distancia euclidiana es la más utilizada en muchos campos entre ellos el área de la salud (González et al., 2015).

$$d_e = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (19)$$

#### ***Distancia Manhattan***

Mide la separación entre elementos utilizando un camino similar a una cuadrícula. La distancia Manhattan entre dos elementos es la suma de las diferencias de sus componentes correspondientes (Rodríguez & Angel, 2021). Este método es una alternativa a la distancia euclidiana, el cual permite menos afectación por datos atípicos, ya que no se eleva al cuadrado las diferencias, sino que se mide el valor absoluto de la siguiente manera:



$$d_m = \sum_{i=1}^n |x_i - y_i| \quad (20)$$

Una comparación gráfica entre la distancia Manhattan y la Euclidiana se puede observar en la Figura 5.

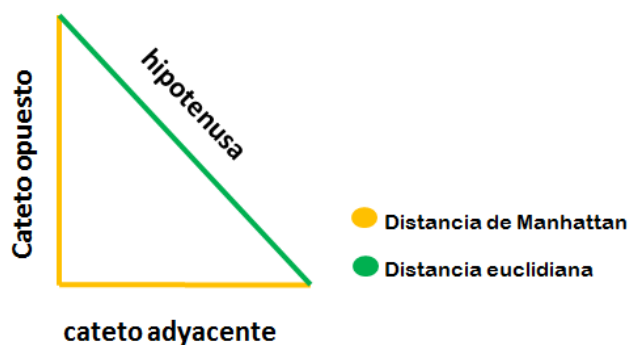


Figura 5. *Distancia Manhattan vs Distancia Euclidiana*

### K-means

El algoritmo K-Means propuesto por J.B. MacQueen es uno de los algoritmos más simples dentro del aprendizaje no supervisado que resuelve el problema conocido como clustering. Este procedimiento sigue una forma sencilla y fácil para clasificar un conjunto de datos dado un número determinado de clústeres predefinidos (Sinaga & Yang, 2020). El algoritmo K-Means se utiliza comúnmente en minería de datos y reconocimiento de patrones, su objetivo es minimizar el índice de rendimiento de los clústeres utilizando el error cuadrático como criterio fundamental. Para obtener el resultado óptimo, el algoritmo busca encontrar k divisiones que cumplan con un determinado criterio (Kodinariya & Makwana, 2013).

#### Procedimiento

1. Definir un número 'k' de centroides para todo el conjunto de datos.
2. Cada centroide toma una posición aleatoria dentro de todo el conjunto.
3. Se calcula la distancia de todos los datos de manera individual a cada centroide con la fórmula:

$$D_c = \sqrt{(X_c - X_i)^2 + (Y_c - Y_i)^2} \quad (21)$$

4. Se asocia cada dato del conjunto a un clúster usando la mínima distancia obtenida en el paso anterior.
5. Una vez terminado se recalcula la posición del centroide en los k grupos.
6. Los centroides en cada grupo se mueven tomando una nueva posición, por lo que es necesario asociar nuevamente todos los datos al centroide más próximo.

7. En este punto se entra en un bucle, el mismo que termina cuando la posición de cada centroide no varía.

### Método del Codo (Elbow)

El método del codo es una regla empírica comúnmente utilizada para determinar el número óptimo de clústeres en métodos de agrupamiento como k-means y la agrupación jerárquica aglomerativa. Este criterio resulta útil al identificar el punto en el cual aumentar más el número de grupos ya no proporciona suficiente información adicional (Syakur et al., 2018).

Este método se basa en el análisis del porcentaje de variación en función del número de clústeres. Se va incrementando gradualmente el valor de 'k' y se registra la Suma Cuadrada de Error (SSE) (Marutho et al., 2018). El SSE representa la suma de las distancias euclidianas promedio de cada punto con respecto al centroide del conjunto al que pertenece. Al observar la disminución drástica del valor del SSE y la formación de un punto de inflexión como se ilustra en la Figura 6. se identifica el número óptimo de grupos, es decir, el valor de 'k'. Los métodos combinados del criterio Elbow y k-means permiten determinar el valor adecuado de k para lograr una clusterización óptima de los datos. Al emplear esta técnica, es posible encontrar la configuración de clústeres que mejor se ajuste a la estructura subyacente de los datos, facilitando así un análisis más preciso y una interpretación significativa de los resultados obtenidos (Kodinariya & Makwana, 2013), (Nainggolan et al., 2019).

En datos normalizados es común representar la curva, tal y como se muestra en la Figura 7.

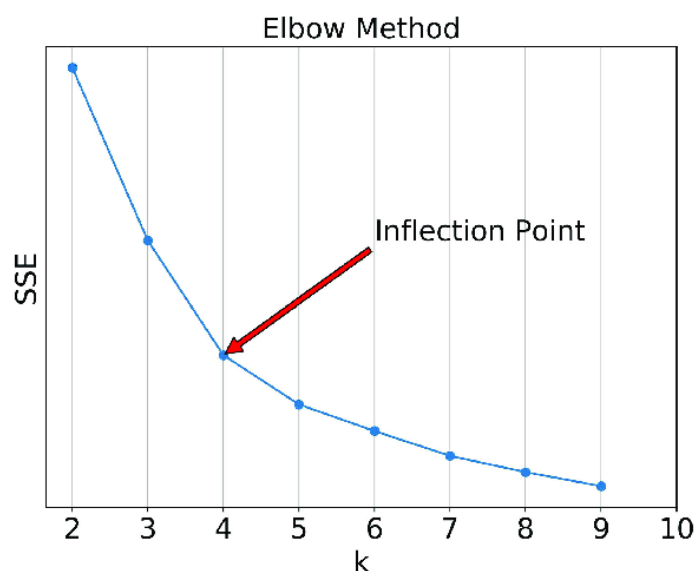


Figura 6. Método Elbow para un conjunto de datos

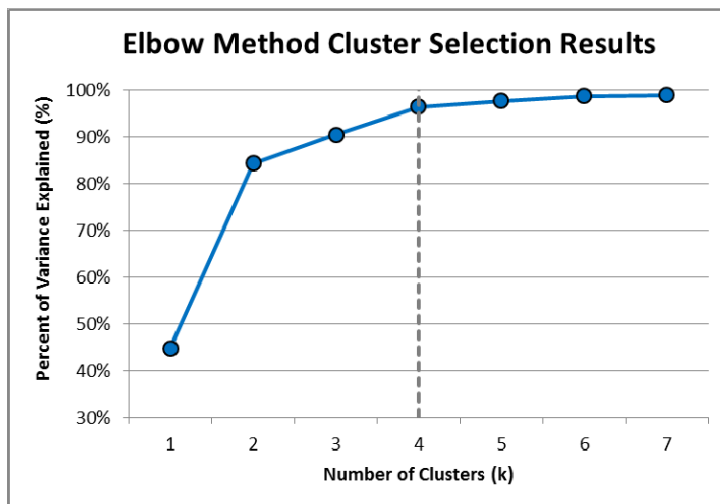


Figura 7. Método Elbow con datos normalizados.

### Procedimiento

Para obtener la base de datos representativo para los sistemas de estudio se procede a generar un conjunto de escenarios, los cuales son procedentes de un mismo estado, pero a un nivel de cargabilidad distinto. Para lo cual se combinan un código propio, herramientas de clustering y funciones adicionales del paquete de MATPOWER, todo esto en el lenguaje de programación MATLAB (Mass, 2012).

### Creación de la base de datos

En total se han realizado 10 ensayos y la demanda de cada sistema aumenta en pasos del 7%, iniciando en un 50% hasta llegar a 120%.

#### Procedimiento

1. Seleccionar un sistema de estudio que deberá estar estructurado en formato 'case.m' para que MATPOWER pueda acceder a toda la información del mismo.
2. Comprobar que el sistema pueda converger para un algoritmo de flujos de potencia. Si no ocurre se debe reforzar el sistema con más rutas entre barras y evitar así barras aisladas.
3. Establecer un nivel inicial del 50% de carga en todas las barras del sistema.
4. Correr un algoritmo de flujos de potencia ('Runopf') implementado en MATPOWER.
5. Con los flujos de potencia y el costo de líneas que se obtiene en el paso anterior, calcular el coeficiente costo beneficio con la ecuación (16) para todas las rutas.
6. De los resultados del paso 4 tomar el flujo máximo de las líneas y los flujos de potencia para calcular el coeficiente de sobrecarga a todas estas con (15).

7. Para calcular el coeficiente de mínimo esfuerzo se toma los ángulos de tensión de los Barras y la impedancia de todas las líneas que se obtuvo en el paso 4 y estimar el coeficiente con la fórmula (17)
8. Normalizar los datos de cada vector de manera individual.
9. Almacenar los valores normalizados de cada vector en una matriz propia para cada coeficiente.
10. Incrementar el nivel de carga en 7% y regresar al paso 4 hasta llegar al número de ensayos establecido.

### Normalización de los datos

Es frecuente encontrarse con escenarios en los que un método de aprendizaje automático no logra comprender completamente el contexto de los datos que recibe. Esto se debe a las diferencias numéricas existentes entre los datos, lo que provoca que el algoritmo otorgue más importancia a aquellos que tienen valores más grandes. Este problema conlleva a la introducción de sesgos en el proceso de aprendizaje, por lo tanto, es necesario homogeneizar los datos. Para abordar esta problemática se emplea un proceso de escalamiento, estandarización o normalización de los datos con las siguientes fórmulas:

$$A_+ = A - \min(A) \quad (22)$$

$$A_N = \frac{A_+}{\max(A_+)} \quad (23)$$

De esta manera, se logra ajustar los valores de los datos de manera adecuada, permitiendo al algoritmo comprender mejor el contexto y realizar un análisis más preciso.

### Proceso de clusterización

Como se menciona anteriormente en la teoría, uno de los parámetros de entrada del algoritmo de k-means es el número de grupos denominado 'k'. Para determinar este valor de manera óptima se utiliza el método Elbow en la base de datos de cada sistema. Una vez obtenido el número óptimo de agrupaciones se procede al análisis de cada base de datos considerando cada criterio y sistema, con el objetivo de obtener las agrupaciones correspondientes. A continuación, se presentan los resultados de este análisis, detallando las agrupaciones obtenidas para cada sistema y criterio.

## Resultados y Discusión

Los sistemas objeto de este estudio son: Garver 6 Barras, IEEE 24 Barras y IEEE 118 Barras, la descripción de los mismos se adjunta en la Tabla 9., Tabla 10. y Tabla 11. del Anexo 1.

### Presentación de Resultados

#### a) Sistema Garver 6 Barras

En este primer caso el número de corredores es reducido, se cuenta únicamente con 15. Usando en el Método Elbow, se ha determinado que el valor adecuado de clúster ( $k = 3$ ), como se muestra en la Figura 8.

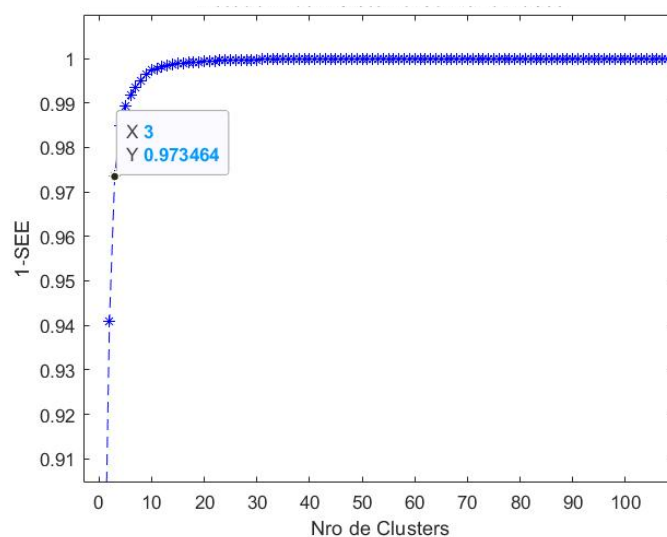


Figura 8. Método Elbow para el Sistema Garver 6 Barras.

El criterio de sobrecarga es el primero en analizarse. En la Figura 9. se muestra la gráfica de la agrupación correspondiente. De acuerdo con la teoría previamente mencionada, el objetivo es identificar el grupo que presenta los valores más altos según la Figura 9. Por lo tanto, se selecciona clúster 1 identificado en color azul, el cual cumple con esta premisa. El número de ensayos indicado en la Tabla 1. y en las siguientes tablas indica cuantas veces una ruta está contenida en un grupo específico, además, se ha numerado a cada ruta para su identificación dentro del espacio de búsqueda. Esta información es útil para resaltar líneas que representen un impacto significativo dentro del grupo, lo que permite centrar el análisis en aquellas rutas que son más relevantes.

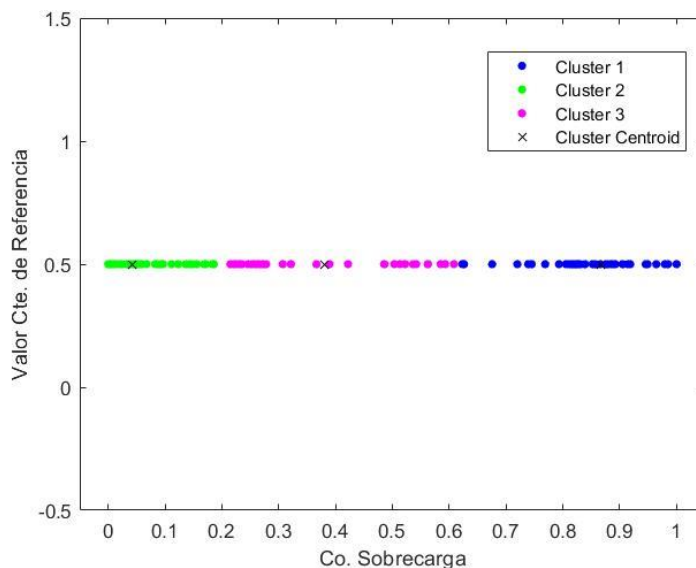


Figura 9. Clusterización utilizando el Coeficiente de sobrecarga en el Sistema Garver 6 Barras

Tabla 1. Grupos de líneas resultado de la Clusterización con el coeficiente de Sobrecarga

Sistema Garver 6 Barras					
Clúster 1		Clúster 2		Clúster 3	
Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos
6	10	7	10	9	10
11	10	8	10	10	10
14	9	12	10	15	10
4	4	13	10	1	6
2	1	2	5	3	5
5	1	3	5	4	5
1	0	1	4	5	5
3	0	5	4	12	4
7	0	4	1	14	1

Continuando con el análisis, se procede a utilizar los criterios de costo-beneficio y mínimo esfuerzo, cuyas graficas de clusterización presentan similitudes a la Figura 9., debido a que los datos se analizan utilizando solo un coeficiente a la vez, es decir, son unidimensionales. En este contexto, se ha seleccionado un grupo que contiene los valores mínimos en términos del coeficiente de costo-beneficio, mientras que se buscan los valores máximos para conformar otro grupo en el caso de utilizar el criterio de mínimo esfuerzo. Esta estrategia de selección de grupos basada en criterios específicos permite identificar soluciones optimas en función de diferentes aspectos y contribuye a la diversificación de resultados en el proceso de análisis de clustering. Por ende, solo se muestran los resultados en la Tabla 2. y Tabla 3. donde los mejores grupos están en el clúster 1 para ambos criterios.

Tabla 2. Grupos de líneas resultado de la clusterización con el coeficiente de Costo-Beneficio

Sistema Garver 6 Barras					
Clúster 1		Clúster 2		Clúster 3	
Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos
1	10	13	10	8	6
2	10	7	4	7	5
4	10	1	0	3	3
6	10	2	0	5	2
9	10	3	0	1	0
10	10	4	0	2	0
11	10	5	0	4	0
12	10	6	0	6	0
14	10	8	0	9	0
15	10	9	0	10	0
5	8	10	0	11	0
3	7	11	0	12	0

Tabla 3. Grupos de líneas resultado de la clusterización con el coeficiente de Mínimo Esfuerzo

Sistema Garver 6 Barras					
Clúster 1		Clúster 2		Clúster 3	
Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos
14	9	15	10	7	10
6	8	1	4	8	10
9	8	3	3	12	10
11	8	2	2	13	10
4	3	4	2	10	8
2	2	5	2	3	7
5	2	6	2	1	6
1	0	9	2	2	6
3	0	10	2	5	6
7	0	11	2	4	5

Con el objetivo de mejorar la reducción del espacio de búsqueda se utiliza simultáneamente dos criterios durante el análisis de clustering. Esta estrategia permite combinar múltiples aspectos relevantes en la agrupación de datos, de esta manera contribuye a una mayor eficiencia en la selección de soluciones óptimas. Al considerar dos criterios a la vez, se busca maximizar la precisión y la capacidad de discernimiento del proceso de clustering, lo que conduce a una exploración más efectiva del espacio de soluciones disponibles.

Como resultado, la mejor combinación de criterios se muestra en la Figura 10., donde el clúster 3 (en color magenta) se ajusta a los criterios seleccionados. Los grupos resultantes de esta combinación se presentan en la Tabla 4.

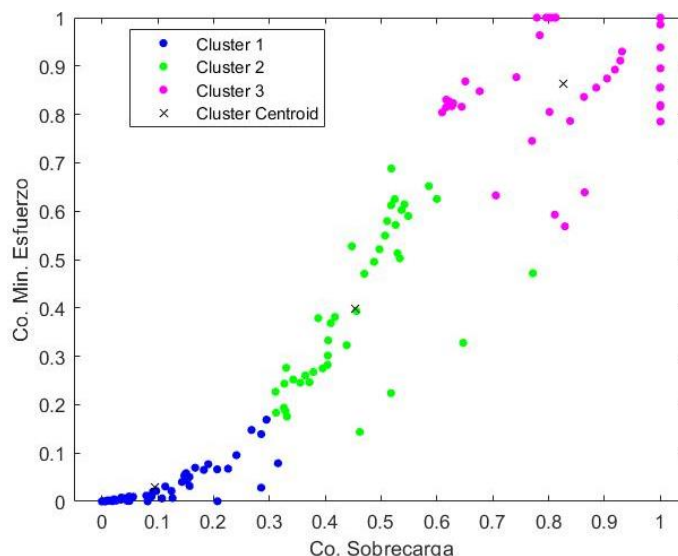


Figura 10. *Combinación de Criterios Sobrecarga vs Mínimo Esfuerzo para el Sistema Garver 6 Barras*

Tabla 4. *Grupos de líneas resultante de la combinación del coeficiente de Mínimo Esfuerzo y Sobrecarga*

Sistema Garver 6 Barras					
Clúster 1		Clúster 2		Clúster 3	
Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos
7	10	15	10	6	10
8	10	10	9	11	9
12	10	1	6	14	9
13	10	3	4	9	7
2	6	4	3	4	4
3	6	5	3	2	2
5	5	9	3	5	2

### b) Sistema IEEE 24 Barras

El sistema IEEE 24 Barras contiene un total de 41 vías en su espacio de búsqueda, el Método Elbow sugiere un número de agrupaciones ( $k = 6$ ), como se observa en la Figura 11. El criterio que reúne el mejor grupo es costo beneficio y la clusterización respectiva con este



criterio se observa en la Figura 12., donde el clúster 1 (en color azul) presenta los valores mínimos. La Tabla 5. muestra las rutas que contiene cada clúster.

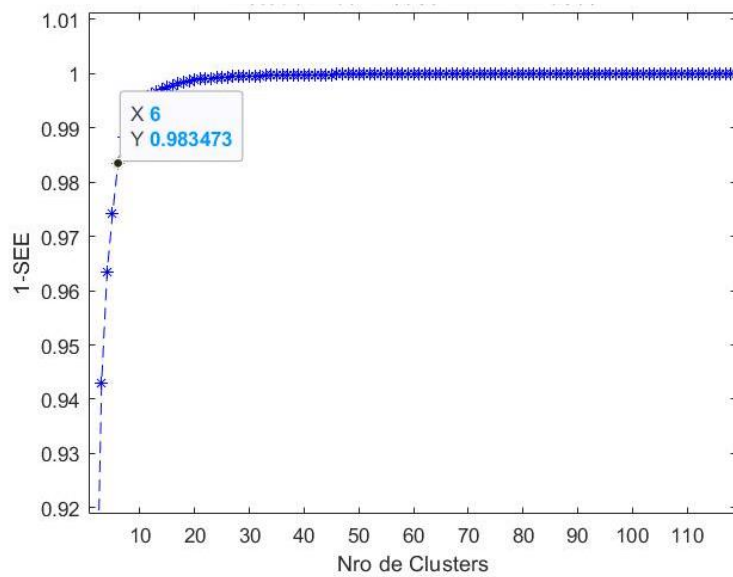


Figura 11. Método Elbow para el Sistema IEEE 24 Barras

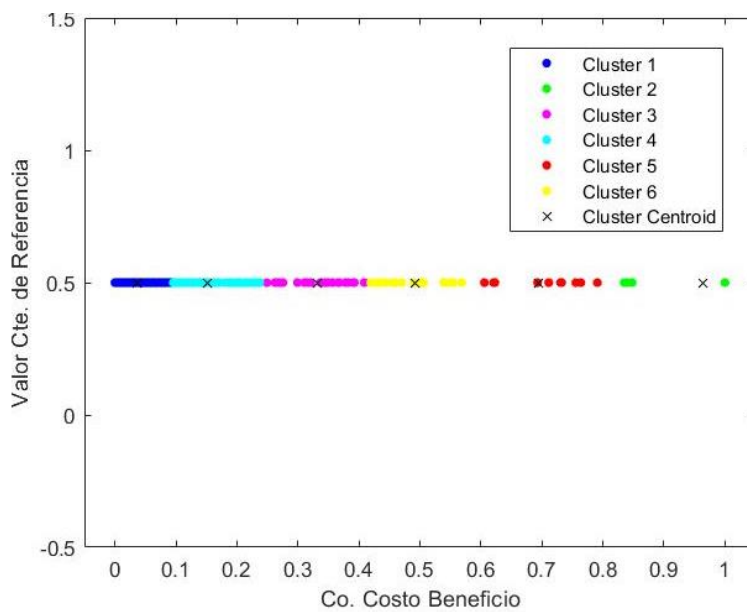


Figura 12. Clusterización usando el criterio de Costo-Beneficio en el Sistema IEEE 24 Barras

Tabla 5. Grupos de líneas resultantes de la clusterización con el coeficiente de Costo-Beneficio para el sistema IEEE 24 Barras

Sistema IEEE 24 Barras											
Clúster 1		Clúster 2		Clúster 3		Clúster 4		Clúster 5		Clúster 6	
Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos
1	10	22	4	32	5	2	7	22	3	13	3
3	10	12	3	9	3	21	6	40	3	2	2
10	10	9	2	21	3	41	6	13	2	8	2
11	10	13	2	30	3	6	5	6	1	9	2
16	10	40	2	40	3	32	5	9	1	12	2
17	10	1	0	6	2	5	4	31	1	22	2
23	10	2	0	8	2	8	4	1	0	6	1
24	10	3	0	12	2	30	4	2	0	30	1
25	10	4	0	34	2	37	4	3	0	40	1
27	10	5	0	13	1	39	4	4	0	1	0
28	10	6	0	1	0	19	3	5	0	3	0
33	10	7	0	2	0	35	3	7	0	4	0
18	9	8	0	3	0	4	2	8	0	5	0
20	9	10	0	4	0	7	2	10	0	7	0
29	9	11	0	5	0	14	2	11	0	10	0
4	8	14	0	7	0	15	2	12	0	11	0
7	8	15	0	10	0	26	2	14	0	14	0
14	8	16	0	11	0	31	2	15	0	15	0
15	8	17	0	14	0	34	2	16	0	16	0
26	8	18	0	15	0	36	2	17	0	17	0
36	8	19	0	16	0	38	2	18	0	18	0
38	8	20	0	17	0	9	1	19	0	19	0

**c) Sistema IEEE 118 Barras**

Por último, se realiza el mismo procedimiento al sistema más grande IEEE 118, primero se obtiene el número de clústeres con el método Elbow como muestra la Figura 13.

El criterio que se utiliza para este sistema es el de costo-beneficio, la Figura 14. muestra la clusterización de los datos y al clúster 5 identificado en color rojo que concentra a los valores mínimos. El Anexo 2: Grupos de líneas resultantes de la clusterización con el coeficiente de Costo-Beneficio para el sistema IEEE 118 Barras

Tabla 12. muestra los resultados de la clusterización para el sistema.

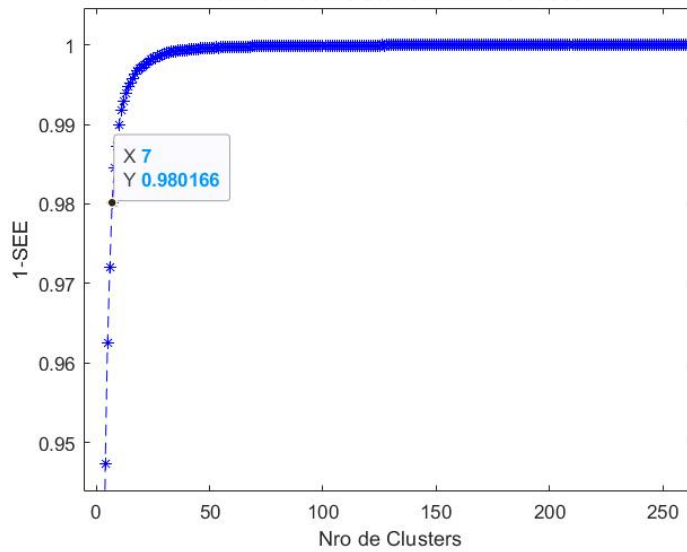


Figura 13. Método Elbow para el Sistema IEEE 118 Barras

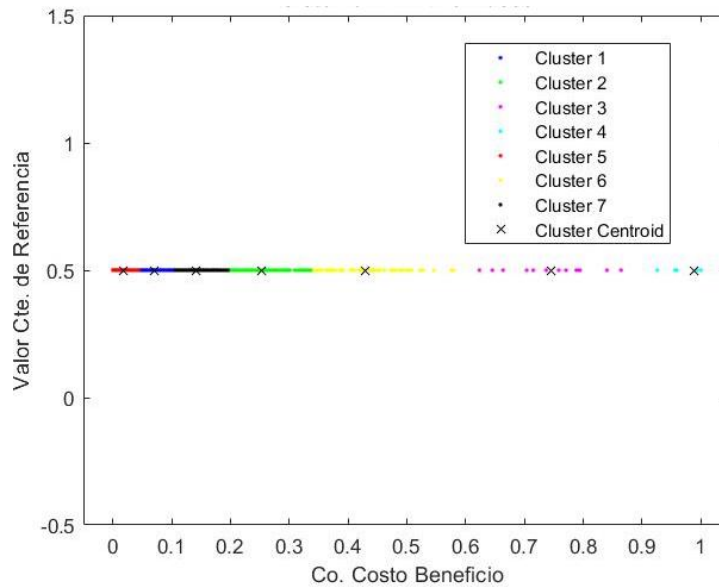


Figura 14. Clusterización usando el criterio de Costo-Beneficio Sistema IEEE 118 Barras

**Análisis y Discusión**

Para la selección de las rutas que forman parte del grupo candidato para la solución del TEP, se consideran aquellas rutas que cuentan con al menos 5 ensayos en el caso del Sistema Garver, y al menos 8 ensayos en los Sistemas del IEEE 24 Barras e IEEE 118 Barras. Esta elección se basa en la premisa de que, al tener un conjunto más amplio de líneas, se requiere obtener una reducción de mayor calidad para lograr una solución óptima en el problema del TEP.

En los sistemas de mayor dimensión no fue posible realizar una combinación de variables debido a la dificultad de determinar un grupo específico, dada la complejidad y la cantidad de

líneas involucradas en estos sistemas, esto se muestra en la Figura 15. En donde se observa una diferencia notable con respecto a la Figura 10. que pertenece al Sistema Garver, donde el clúster 3 (representado en color magenta) como único grupo cumple simultáneamente ambos criterios, alcanzando los valores máximos para cada uno de ellos.

Sin embargo, en la Figura 15. del sistema IEEE 24 se puede apreciar dos grupos: el clúster 3 (magenta) y 6 (amarillo) los cuales se aproximan a la misma posición que se muestra en la Figura 10.

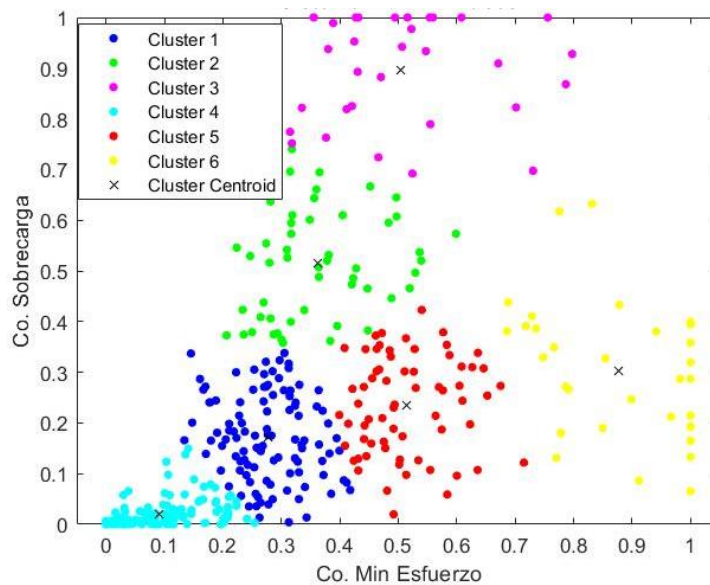


Figura 15. Clusterización usando el criterio Sobrecarga vs Mínimo Esfuerzo para el Sistema IEEE 24 Barras

Esta situación genera una incertidumbre en la selección de un solo grupo como la solución óptima, ya que ambos grupos muestran características relevantes. Se requiere una evaluación más detallada para determinar el más adecuado en función de otros factores y criterios de análisis. Por lo tanto, el nuevo espacio de búsqueda para cada sistema se muestra en la Tabla 6., Tabla 7. y Tabla 8. Las mismas que contienen a la solución final obtenida con metaheurísticas presentadas en (Matute et al., 2020) y (Morquecho et al., 2021).

Tabla 6. *Espacio de búsqueda reducido para el Sistema Garver 6 Barras*

<b>Sistema Garver 6 Barras</b>	
Espacio de Búsqueda Original Nro. de Líneas $n_l = 15$	
<b>Solución Final obtenida por metaheurísticas y enfoques matemáticos</b>	
<b>Escenario con generación despachable</b>	Rutas selectas $l_{2-6}, l_{3-5}, l_{4-6}$
<b>Escenario sin generación despachable</b>	Rutas selectas $l_{2-6}, l_{3-5}, l_{4-6}$
<b>Espacio Reducido obtenido por Clustering</b>	
<b>Criterio de Costo-Beneficio</b>	
<b>Rutas Propuestas</b>	$l_{2-3}^6, l_{3-5}^{11}, l_{4-6}^{14}, l_{2-6}^9$ $l_{1-2}^1, l_{1-3}^2, l_{1-5}^4, l_{1-6}^5$ $l_{3-4}^{10}, l_{5-6}^{15}, l_{3-6}^{12}, l_{1-4}^3$
<b>Porcentaje de Reducción</b>	20 %
<b>Criterio Mínimo Esfuerzo y Sobrecarga</b>	
<b>Rutas Propuestas</b>	$l_{3-5}^{11}, l_{4-6}^{14}, l_{2-6}^9, l_{2-3}^6$
<b>Porcentaje de Reducción</b>	73.23 %

Tabla 7. *Espacio de búsqueda reducido para el Sistema IEEE 24 Barras*

<b>Sistema IEEE 24 Barras</b>			
Espacio de Búsqueda Original Nro. de Líneas $n_l = 41$			
<b>Solución Final obtenida por metaheurísticas y enfoques matemáticos</b>			
<b>Escenario con generación Despachable y considerando contingencias</b>			
$l_{6-10}$	$l_{7-8}$	$l_{14-16}$	$l_{16-17}$
<b>Escenario sin generación Despachable</b>			
$l_{6-10}$	$l_{7-8}$	$l_{10-12}$	

Espacio Reducido obtenido por Clustering					
Criterio Utilizado Costo-Beneficio					
Rutas Propuestas $n_r = 22$					
$l_{1-2}^1$	$l_{20-23}^{33}$	$l_{1-5}^3$	$l_{2-8}^{36}$	$l_{6-10}^{10}$	$l_{2-4}^4$
$l_{7-8}^{11}$	$l_{3-24}^7$	$l_{10-11}^{16}$	$l_{16-19}^{28}$	$l_{15-24}^{26}$	$l_{10-12}^{17}$
$l_{14-16}^{23}$	$l_{15-16}^{24}$	$l_{15-21}^{25}$	$l_{9-12}^{15}$	$l_{13-14}^{38}$	$l_{16-17}^{27}$
$l_{11-13}^{18}$	$l_{12-13}^{20}$	$l_{9-11}^{14}$	$l_{17-18}^{29}$		
<b>Porcentaje de Reducción</b>				46.34 %	

Tabla 8. Espacio de búsqueda reducido para el Sistema IEEE 118 Barras

Sistema IEEE 118 Barras							
Espacio de Búsqueda Original Nro. de Líneas $n_l = 186$							
Solución Final obtenida por metaheurísticas y enfoques matemáticos							
Escenario con compensación reactiva en derivación							
$l_{8-9}$	$l_{8-5}$	$l_{23-25}$	$l_{25-27}$	$l_{9-10}$	$l_{15-17}$	$l_{94-100}$	$l_{17-113}$
Escenario sin compensación reactiva en derivación							
$l_{1-3}$	$l_{3-5}$	$l_{8-9}$	$l_{8-5}$	$l_{9-10}$	$l_{15-17}$	$l_{23-25}$	$l_{25-27}$
	$l_{63-59}$	$l_{49-66}$	$l_{77-78}$	$l_{94-100}$	$l_{17-113}$	$l_{68-116}$	
Espacio Reducido obtenido por Clustering							
Rutas Propuestas $n_r = 69$							
$l_{1-3}^2$	$l_{4-5}^3$	$l_{5-6}^5$	$l_{8-9}^7$	$l_{8-5}^8$	$l_{75-118}^{185}$	$l_{68-116}^{183}$	$l_{100-103}^{163}$
$l_{9-10}^9$	$l_{11-12}^{12}$	$l_{15-17}^{21}$	$l_{17-18}^{23}$	$l_{18-19}^{24}$	$l_{3-5}^4$	$l_{4-11}^{10}$	$l_{5-11}^{11}$
$l_{21-28}^{28}$	$l_{23-25}^{31}$	$l_{26-25}^{32}$	$l_{25-27}^{33}$	$l_{30-17}^{36}$	$l_{26-30}^{38}$	$l_{23-24}^{30}$	$l_{8-30}^{37}$
$l_{35-36}^{46}$	$l_{34-36}^{49}$	$l_{34-37}^{50}$	$l_{38-37}^{51}$	$l_{30-38}^{54}$	$l_{54-56}^{78}$	$l_{64-61}^{95}$	$l_{89-92}^{141}$
$l_{60-61}^{90}$	$l_{63-59}^{93}$	$l_{63-64}^{94}$	$l_{38-65}^{96}$	$l_{64-65}^{97}$	$l_{49-66}^{98}$	$l_{70-71}^{110}$	$l_{69-77}^{119}$
$l_{49-66}^{99}$	$l_{62-66}^{100}$	$l_{65-66}^{102}$	$l_{65-68}^{104}$	$l_{68-69}^{107}$	$l_{74-75}^{117}$	$l_{96-96}^{156}$	$l_{105-106}^{169}$
$l_{77-78}^{121}$	$l_{77-80}^{123}$	$l_{79-80}^{125}$	$l_{68-81}^{126}$	$l_{81-80}^{127}$	$l_{82-83}^{129}$	$l_{6-7}^6$	$l_{23-32}^{41}$
$l_{85-86}^{133}$	$l_{86-87}^{134}$	$l_{94-95}^{147}$	$l_{94-100}^{155}$	$l_{99-100}^{159}$	$l_{17-113}^{178}$	$l_{47-49}^{65}$	$l_{104-105}^{168}$
$l_{2-12}^{13}$	$l_{69-70}^{108}$	$l_{69-75}^{116}$	$l_{77-32}^{128}$	$l_{66-67}^{103}$			
<b>Porcentaje de Reducción</b>				62.90 %			

### Conclusiones

En los grupos que se forman existen rutas que no resultan relevantes, estas se han identificado gracias al número de ensayos que se estableció al inicio. Las rutas que se descartan son aquellas que no cumplen con una cantidad de ensayos adecuado.

El uso del método Elbow es de suma importancia en un análisis de clustering porque proporciona una idea precisa sobre el factor de agrupamiento, puesto que K no incrementa de manera excesiva a medida que los sistemas lo hacen, es decir, su crecimiento no es lineal.

En el sistema Garver 6 Barras clasificar bajo el criterio de costo-beneficio no es adecuado debido a que el costo de las líneas en un sistema tan reducido no es significativo. Por lo tanto, una clusterización bajo criterios de sobrecarga o mínimo esfuerzo es más conveniente como se mostró en la tabla VI donde el porcentaje de reducción es mayor para estos criterios.

En sistemas de gran dimensión como el IEEE 118 Barras, un análisis técnico-operativo resultó inadecuado debido a la gran cantidad de líneas candidatas. Estas líneas distribuyen de manera más eficiente los flujos de potencia. Lo que hace que el coeficiente de mínimo esfuerzo sea poco útil, al igual que el coeficiente de sobrecarga.

Para los sistemas IEEE 24 y 118 Barras el criterio de costo-beneficio se presenta como la mejor opción para la clusterización, ya que el costo de construcción de una ruta puede variar significativamente entre ellas. De esta manera, se pueden seleccionar las líneas que proporcionen un espacio de búsqueda óptimo operativamente y a un costo de inversión reducido.

### Referencias

- Alguacil, N., Motto, A. L., & Conejo, A. J. (2003). Transmission expansion planning: A mixed-integer LP approach. *IEEE Transactions on Power Systems*, 18(3), 1070-1077. <https://doi.org/10.1109/TPWRS.2003.814891>
- da Silva, E. F., Rahmani, M., & Rider, M. J. (2015). A Search Space Reduction Strategy and a Mathematical Model for Multistage Transmission Expansion Planning with \$\$\$N-1\$\$\$ Security Constrains. *Journal of Control, Automation and Electrical Systems*, 26(1), 57-67. <https://doi.org/10.1007/s40313-014-0154-2>
- Dominguez, A. H., Macedo, L. H., Escobar, A. H., & Romero, R. (2017). Multistage Security-Constrained HVAC/HVDC Transmission Expansion Planning With a Reduced Search Space. *IEEE Transactions on Power Systems*, 32(6), 4805-4817. <https://doi.org/10.1109/TPWRS.2017.2669323>
- Duque, A., Escobar, A. H., & Gallego, R. A. (2014). Multistage transmission expansion planning via network partitioning and principal variables identification. *2014 IEEE PES Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA)*, 1-6. <https://doi.org/10.1109/TDC-LA.2014.6955221>
- Escudero Delgado, P. V. (2018). *Modelo de expansión de un sistema de transmisión basado en linealización de flujos óptimos de potencia AC*. [BachelorThesis]. <http://dspace.ups.edu.ec/handle/123456789/15590>
- Gallego Pareja, L. A., Romero, R. A., & López Lezama, J. M. (2009). Planeamiento de la expansión de sistemas de transmisión considerando contingencias y demanda incierta. *Revista Facultad de Ingeniería Universidad de Antioquia*, 48, 188-200.
- Gallego-Pareja, L. A., Romero, R. A., & López-Lezama, J. M. (2009). Transmission network expansion planning considering contingencies and uncertain demand. *Revista Facultad de Ingeniería Universidad de Antioquia*, 48, Article 48. <https://doi.org/10.17533/udea.redin.16529>



- González, R. C., Berrio, A. R., & Sáenz, M. C. (2015). Caracterización de fincas productoras de palto (*Persea americana* Mill.) y mandarina (*Citrus* spp.) en Cañete, Lima, Perú. *Aporte Santiaguino*, ág. 33-44. <https://doi.org/10.32911/as.2015.v8.n1.241>
- Kodinariya, T., & Makwana, P. (2013). Review on Determining of Cluster in K-means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1, 90-95.
- Lumbreras, S., & Ramos, A. (2016). The new challenges to transmission expansion planning. Survey of recent practice and literature review. *Electric Power Systems Research*, 134, 19-29. <https://doi.org/10.1016/j.epsr.2015.10.013>
- Marutho, D., Hendra Handaka, S., Wijaya, E., & Muljono. (2018). The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. *2018 International Seminar on Application for Technology of Information and Communication*, 533-538. <https://doi.org/10.1109/ISEMANTIC.2018.8549751>
- Mass, N. (2012). *MathWorks Announces Release 2012a of the MATLAB and Simulink Product Families*. <https://www.mathworks.com/company/newsroom/mathworks-announces-release-2012a-of-the-matlab-and-simulink-product-families.html>
- Matute, N. E., Torres, S. P., Morquecho, E. G., Astudillo-Salinas, F., Lopez, J. C., & Flores, W. C. (2020). Improving the AC Transmission Expansion Planning by Using Initial Solutions Algorithms. *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 494-498. <https://doi.org/10.1109/ISGT-Europe47291.2020.9248778>
- Monticelli, A., Santos, A., F. Pereira, M. V., Cunha, S. H., Parker, B. J., & G. Praca, J. C. (1982). Interactive Transmission Network Planning Using a Least-Effort Criterion. *IEEE Transactions on Power Apparatus and Systems*, PAS-101(10), 3919-3925. <https://doi.org/10.1109/TPAS.1982.317043>
- Morquecho, E. G., Torres, S. P., & Castro, C. A. (2021). An efficient hybrid metaheuristics optimization technique applied to the AC electric transmission network expansion

- planning. *Swarm and Evolutionary Computation*, 61, 100830.  
<https://doi.org/10.1016/j.swevo.2020.100830>
- Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics: Conference Series*, 1361(1), 012015.  
<https://doi.org/10.1088/1742-6596/1361/1/012015>
- Omran, M., Engelbrecht, A., & Salman, A. (2007). An overview of clustering methods. *Intell. Data Anal.*, 11, 583-605. <https://doi.org/10.3233/IDA-2007-11602>
- Palacios, J. (2020). Estado del arte de la planeación de expansión de sistemas de transmisión. *I+D Tecnológico*, 16(2), Article 2. <https://doi.org/10.33412/idt.v16.2.2835>
- Pascual, D., Pla, F., & Sánchez, S. (s. f.). *Algoritmos de agrupamiento*.
- Rodríguez, A., & Angel, M. (2021). Comparación de métricas de distancia en el algoritmo K-Vecinos Más Cercanos para el problema de reconocimiento automático de dígitos manuscritos. [http://opac.pucv.cl/pucv\\_txt/txt-3000/UCD3128\\_01.pdf](http://opac.pucv.cl/pucv_txt/txt-3000/UCD3128_01.pdf).  
<http://repositorio.ucv.cl/handle/10.4151/92531>
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8, 80716-80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1), 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- Torres R., M. A., Castro, C. A., & Rider, M. J. (2018). A Search Space Reduction Strategy within the MILP Branch Flow Model for Concurrent AC Multistage Transmission Expansion and Reactive Power Planning. *2018 North American Power Symposium (NAPS)*, 1-6. <https://doi.org/10.1109/NAPS.2018.8600545>
- Zimmerman, R. D., Murillo-Sanchez, C. E., & Thomas, R. J. (2011). MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and

Education. *IEEE Transactions on Power Systems*, 26(1), 12-19.

<https://doi.org/10.1109/TPWRS.2010.2051168>

Anexos

Anexo 1: Datos y espacio de búsqueda del Sistema Garver 6 Barras, IEEE 24 Barras e IEEE 118 Barras

Tabla 9. Datos principales y espacio de búsqueda del Sistema Garver 6 Barras

<b>Sistema Garver 6 Barras</b>	
Número de Barras $n_b = 6$	
$P_D = 760 MW$	$Q_D = 152 MWAr$
Líneas Iniciales del Sistema	
$l_{1-2}$	$l_{1-4}$
$l_{2-3}$	$l_{2-4}$
$l_{1-5}$	$l_{3-5}$
<b>Espacio de Búsqueda Total</b>	$n_l = 15$

Tabla 10. Datos principales y espacio de búsqueda del Sistema IEEE 24 Barras

<b>Sistema IEEE 24 Barras</b>					
Número de Barras $n_b = 24$					
$P_D = 8550 MW$			$Q_D = 1740 MWAr$		
Líneas Iniciales del Sistema					
$l_{1-2}$	$l_{1-3}$	$l_{1-5}$	$l_{2-4}$	$l_{2-6}$	$l_{3-9}$
$l_{19-23}$	$l_{3-24}$	$l_{4-9}$	$l_{5-10}$	$l_{6-10}$	$l_{7-8}$
$l_{8-9}$	$l_{8-10}$	$l_{9-11}$	$l_{9-12}$	$l_{10-11}$	$l_{10-12}$
$l_{11-13}$	$l_{11-14}$	$l_{12-13}$	$l_{12-23}$	$l_{13-23}$	$l_{14-16}$
$l_{15-16}$	$l_{15-21}$	$l_{15-24}$	$l_{16-17}$	$l_{16-19}$	$l_{17-18}$
$l_{17-22}$	$l_{18-21}$	$l_{19-20}$	$l_{20-23}$	$l_{21-22}$	$l_{1-8}$
$l_{2-8}$	$l_{6-7}$	$l_{13-14}$	$l_{14-23}$	$l_{16-23}$	
<b>Espacio de Búsqueda Total</b>	$n_l = 41$				

Tabla 11. Datos principales y espacio de búsqueda del Sistema IEEE 118 Barras

Sistema IEEE 118 Barras											
Número de Barras $n_b = 118$											
$P_D = 6240 MW$						$Q_D = 2470 MVar$					
Líneas Iniciales del Sistema											
$l_{1-2}$	$l_{1-3}$	$l_{4-5}$	$l_{3-5}$	$l_{5-6}$	$l_{6-7}$	$l_{8-9}$	$l_{8-5}$	$l_{9-10}$	$l_{4-11}$	$l_{5-11}$	$l_{11-12}$
$l_{2-12}$	$l_{3-12}$	$l_{7-12}$	$l_{11-13}$	$l_{12-14}$	$l_{13-15}$	$l_{12-16}$	$l_{15-17}$	$l_{16-17}$	$l_{17-18}$	$l_{18-19}$	$l_{19-20}$
$l_{15-19}$	$l_{20-21}$	$l_{21-22}$	$l_{22-23}$	$l_{23-24}$	$l_{23-25}$	$l_{26-25}$	$l_{25-27}$	$l_{27-28}$	$l_{28-29}$	$l_{30-17}$	$l_{8-30}$
$l_{26-30}$	$l_{17-31}$	$l_{29-31}$	$l_{23-32}$	$l_{31-32}$	$l_{27-32}$	$l_{15-33}$	$l_{19-34}$	$l_{35-36}$	$l_{35-37}$	$l_{33-37}$	$l_{34-36}$
$l_{34-37}$	$l_{38-37}$	$l_{37-39}$	$l_{37-40}$	$l_{30-38}$	$l_{39-40}$	$l_{40-41}$	$l_{40-42}$	$l_{41-42}$	$l_{43-44}$	$l_{34-43}$	$l_{44-45}$
$l_{45-46}$	$l_{46-47}$	$l_{46-48}$	$l_{47-49}$	$l_{42-49}$	$l_{42-49}$	$l_{45-49}$	$l_{48-49}$	$l_{49-50}$	$l_{49-51}$	$l_{51-52}$	$l_{52-53}$
$l_{53-54}$	$l_{49-54}$	$l_{54-55}$	$l_{54-56}$	$l_{55-56}$	$l_{56-57}$	$l_{50-57}$	$l_{56-58}$	$l_{51-58}$	$l_{54-59}$	$l_{56-59}$	$l_{56-59}$
$l_{55-59}$	$l_{59-60}$	$l_{59-61}$	$l_{60-61}$	$l_{60-52}$	$l_{61-62}$	$l_{63-59}$	$l_{63-64}$	$l_{64-61}$	$l_{38-65}$	$l_{64-65}$	$l_{49-66}$
$l_{49-66}$	$l_{62-66}$	$l_{62-67}$	$l_{65-66}$	$l_{66-67}$	$l_{65-68}$	$l_{47-69}$	$l_{49-69}$	$l_{68-69}$	$l_{69-70}$	$l_{24-70}$	$l_{70-71}$
$l_{24-72}$	$l_{71-72}$	$l_{71-73}$	$l_{70-74}$	$l_{70-75}$	$l_{69-75}$	$l_{74-75}$	$l_{76-77}$	$l_{69-77}$	$l_{75-77}$	$l_{77-78}$	$l_{78-79}$
$l_{77-80}$	$l_{77-80}$	$l_{79-80}$	$l_{68-81}$	$l_{81-80}$	$l_{77-82}$	$l_{82-83}$	$l_{83-84}$	$l_{83-85}$	$l_{84-85}$	$l_{85-86}$	$l_{86-87}$
$l_{85-88}$	$l_{85-89}$	$l_{88-89}$	$l_{89-90}$	$l_{89-90}$	$l_{90-91}$	$l_{89-92}$	$l_{98-92}$	$l_{91-92}$	$l_{92-93}$	$l_{92-94}$	$l_{93-94}$
$l_{94-95}$	$l_{80-96}$	$l_{82-96}$	$l_{94-96}$	$l_{80-97}$	$l_{80-98}$	$l_{80-99}$	$l_{92-100}$	$l_{94-100}$	$l_{95-96}$	$l_{96-97}$	$l_{98-100}$
$l_{99-100}$	$l_{100-101}$	$l_{92-102}$	$l_{101-102}$	$l_{100-103}$	$l_{100-104}$	$l_{103-104}$	$l_{103-105}$	$l_{100-106}$	$l_{104-105}$	$l_{105-106}$	$l_{105-107}$
$l_{105-108}$	$l_{106-107}$	$l_{108-109}$	$l_{103-110}$	$l_{109-110}$	$l_{110-111}$	$l_{110-112}$	$l_{17-113}$	$l_{32-113}$	$l_{32-114}$	$l_{27-115}$	$l_{114-115}$
$l_{68-116}$	$l_{12-117}$	$l_{75-118}$	$l_{76-118}$	$l_{49-54}$	$l_{14-15}$						
Espacio de Búsqueda Total								$n_l = 186$			

**Anexo 2:** Grupos de líneas resultantes de la clusterización con el coeficiente de Costo-Beneficio para el sistema IEEE 118 Barras

Tabla 12. Grupos de líneas resultantes de la clusterización con el coeficiente de costo-beneficio para el sistema IEEE 118 Barras

Sistema IEEE 118 Barras													
Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6		Cluster 7	
Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos	Nro. Ruta	Nro de Ensayos
118	7	39	5	64	3	77	4	2	10	64	4	15	6
140	7	162	5	154	3	18	2	3	10	73	4	84	5
1	6	53	4	18	1	59	2	5	10	86	4	87	5
52	6	115	4	19	1	19	1	7	10	154	4	170	5
74	6	136	4	25	1	25	1	8	10	84	3	175	5
101	6	142	4	59	1	64	1	9	10	85	3	14	4
114	6	143	4	73	1	73	1	12	10	142	3	35	4
124	6	179	4	85	1	109	1	21	10	172	3	75	4
144	6	43	3	106	1	128	1	23	10	43	2	76	4
161	6	45	3	1	0	1	0	24	10	57	2	101	4
173	6	55	3	2	0	2	0	28	10	66	2	113	4
176	6	73	3	3	0	3	0	31	10	67	2	120	4
181	6	75	3	4	0	4	0	32	10	77	2	138	4
14	5	76	3	5	0	5	0	33	10	87	2	143	4
16	5	77	3	6	0	6	0	36	10	109	2	148	4
27	5	80	3	7	0	7	0	38	10	146	2	160	4
29	5	82	3	8	0	8	0	46	10	166	2	165	4
35	5	85	3	9	0	9	0	49	10	177	2	167	4
47	5	86	3	10	0	10	0	50	10	39	1	171	4
56	5	87	3	11	0	11	0	51	10	44	1	172	4
60	5	154	3	12	0	12	0	54	10	45	1	174	4
63	5	164	3	13	0	13	0	78	10	61	1	179	4
68	5	172	3	14	0	14	0	90	10	75	1	184	4
81	5	174	3	15	0	15	0	93	10	76	1	1	3
88	5	17	2	16	0	16	0	94	10	82	1	17	3
92	5	18	2	17	0	17	0	96	10	83	1	18	3
111	5	19	2	20	0	20	0	97	10	105	1	19	3
113	5	20	2	21	0	21	0	98	10	112	1	22	3
131	5	25	2	22	0	22	0	99	10	136	1	45	3

148	5	44	2	23	0	23	0	100	10	143	1	57	3
160	5	59	2	24	0	24	0	102	10	158	1	66	3
165	5	60	2	26	0	26	0	104	10	162	1	67	3
171	5	64	2	27	0	27	0	107	10	167	1	74	3
180	5	83	2	28	0	28	0	117	10	1	0	81	3
184	5	84	2	29	0	29	0	121	10	2	0	83	3
17	4	111	2	30	0	30	0	123	10	3	0	85	3
22	4	135	2	31	0	31	0	125	10	4	0	86	3
25	4	138	2	32	0	32	0	126	10	5	0	106	3
34	4	145	2	33	0	33	0	127	10	6	0	130	3
40	4	166	2	34	0	34	0	129	10	7	0	136	3
42	4	177	2	35	0	35	0	133	10	8	0	164	3
44	4	14	1	36	0	36	0	134	10	9	0	166	3
57	4	15	1	37	0	37	0	147	10	10	0	181	3
58	4	48	1	38	0	38	0	155	10	11	0	20	2
61	4	57	1	39	0	39	0	159	10	12	0	25	2
62	4	58	1	40	0	40	0	168	10	13	0	27	2
66	4	61	1	41	0	41	0	178	10	14	0	39	2
67	4	63	1	42	0	42	0	183	10	15	0	43	2
71	4	66	1	43	0	43	0	185	10	16	0	44	2
72	4	67	1	44	0	44	0	4	9	17	0	48	2
79	4	81	1	45	0	45	0	10	9	18	0	53	2
80	4	109	1	46	0	46	0	11	9	19	0	55	2
89	4	112	1	47	0	47	0	13	9	20	0	58	2
91	4	120	1	48	0	48	0	30	9	21	0	59	2
105	4	130	1	49	0	49	0	37	9	22	0	60	2
112	4	150	1	50	0	50	0	95	9	23	0	63	2
115	4	158	1	51	0	51	0	110	9	24	0	68	2
120	4	165	1	52	0	52	0	119	9	25	0	80	2
132	4	167	1	53	0	53	0	156	9	26	0	82	2
135	4	170	1	54	0	54	0	169	9	27	0	88	2
138	4	1	0	55	0	55	0	6	8	28	0	92	2
139	4	2	0	56	0	56	0	41	8	29	0	105	2
149	4	3	0	57	0	57	0	65	8	30	0	109	2
151	4	4	0	58	0	58	0	103	8	31	0	114	2
153	4	5	0	60	0	60	0	108	8	32	0	115	2
157	4	6	0	61	0	61	0	116	8	33	0	142	2

---

164	4	7	0	62	0	62	0	128	8	34	0	145	2
167	4	8	0	63	0	63	0	141	8	35	0	146	2
170	4	9	0	65	0	65	0	163	8	36	0	150	2

---