

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

OBTENCIÓN DE CURVAS HORARIAS DE CARGA CARACTERÍSTICAS DESAGREGADAS POR TIPO DE CONSUMIDOR MEDIANTE LA TÉCNICA DE AGRUPAMIENTO ESPACIAL BASADO EN DENSIDAD

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO ELÉCTRICO**

CARLOS ANDRÉS PEÑALOZA ARAUJO

DIRECTOR: OTERO VALLADARES PATRICIA ELIZABETH

CODIRECTOR: VÁSQUEZ MIRANDA PAÚL FABRICIO

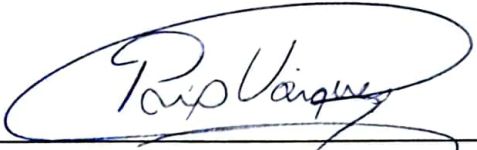
Quito, abril 2023

AVAL

Certificamos que el presente trabajo fue desarrollado por Carlos Andrés Peñaloza Araujo, bajo nuestra supervisión.



**OTERO VALLADARES PATRICIA ELIZABETH
DIRECTOR DEL TRABAJO DE TITULACIÓN**

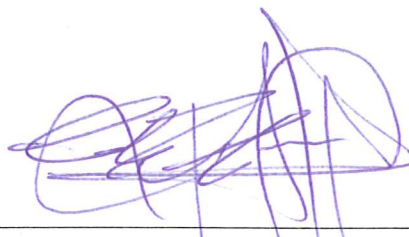


**VÁSQUEZ MIRANDA PAÚL FABRICIO
CODIRECTOR DEL TRABAJO DE TITULACIÓN**

DECLARACIÓN DE AUTORÍA

Yo Carlos Andrés Peñaloza Araujo, declaro bajo estricto juramento que el trabajo aquí descrito es de mi total autoría; que no ha sido presentado con anterioridad para ninguna calificación profesional o grado; y, que he investigado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración, hago constancia de que la Escuela Politécnica Nacional podrá hacer uso del presente trabajo según los términos estipulados en la Ley, Reglamentos y Normas vigentes.

A handwritten signature in purple ink, consisting of several overlapping loops and vertical strokes, positioned above a horizontal line.

CARLOS ANDRES PEÑALOZA ARAUJO

AGRADECIMIENTO

Le agradezco a mi familia, en especial a mi papá, mamá, hermana y abuelita por el apoyo, consejería y buen ejemplo recibido en esta experiencia universitaria.

También es grato reconocer la buena voluntad, inteligencia y excelencia en los profesores que fueron parte de mi vida universitaria. En especial al Dr. Jesús Játiva y MSc. Patricia Otero.

A mi grupo de estudio Christian, Fausto, Felipe y Andrés que desde el inicio demostraron la capacidad de ayuda y de compañerismo que permanece intacta hasta el día de hoy.

ÍNDICE DE CONTENIDO

AVAL	I
DECLARACIÓN DE AUTORÍA.....	II
AGRADECIMIENTO.....	III
ÍNDICE DE CONTENIDO.....	IV
ÍNDICE DE FIGURAS	VI
INDICE DE TABLAS	VIII
RESUMEN	IX
ABSTRACT	X
1. INTRODUCCIÓN.....	1
1.1 OBJETIVOS	2
1.2 ALCANCE	2
1.3 MARCO TEÓRICO.....	3
1.3.1 SISTEMAS ELÉCTRICOS DE TRANSMISIÓN Y DISTRIBUCIÓN.....	3
1.3.2 CARACTERIZACIÓN DE LA DEMANDA, DEFINICIONES Y SU INCREMENTO	4
1.3.2.1 Curva de demanda de potencia	6
1.3.2.2 Factor de carga	7
1.3.3 TÉCNICAS DE MINERÍA DE DATOS.....	7
1.3.3.1 Métodos de agrupamiento.....	8
1.3.3.2 Algoritmo basado en densidad DBSCAN	8
1.3.3.3 Validación de agrupamientos mediante el índice Silhouette (IS)	9
1.3.4 ALGORITMOS COMPLEMENTARIOS.....	11
1.3.4.1 Algoritmo KNN.....	11
1.3.4.2 Knee Definición	12
1.3.5 SOFTWARE EMPLEADOS	13
1.3.5.1 Excel MS	13
1.3.5.2 Matlab.....	14
2. METODOLOGÍA.....	15
2.1 TRATAMIENTO DE INFORMACIÓN REMITIDA POR LA EEA.....	15
2.1.1 INFORMACIÓN DE MEDIDORES.....	15
2.1.2 MACRO CON LENGUAJE VISUAL BASIC	16
2.1.3 ORGANIZACIÓN DE MEDICIONES.....	16
2.2 CREACIÓN DE BASE DE DATOS EN EXCEL MS.....	17
2.2.1 ETIQUETAS EN MEDICIONES	17

2.2.2	TRATAMIENTO DE VARIABLES EN POR UNIDAD	18
2.3	ALGORITMO DBSCAN DESARROLLADO EN MATLAB	18
2.3.1	KNEESEARCH	19
2.3.2	KNEEPOINT	20
2.4	VALIDACIÓN DE GRUPOS MEDIANTE INDICE SILHOUETTE	21
2.5	ORDENAMIENTO Y CLUSTERING.....	21
2.5.1	FLUJO DE ORDENAMIENTO	22
2.5.2	CURVA CARACTERÍSTICA POR MEDIDOR	23
2.5.3	CLUSTERING, FILTROS Y VALIDACIÓN.....	24
2.5.4	DIAGRAMA DE FLUJO TOTAL EMPLEADO EN MATLAB.....	26
3.	RESULTADOS Y DISCUSIÓN	28
3.1	ANÁLISIS ENTRE SEMANA.....	28
3.2	ANÁLISIS FIN DE SEMANA	37
3.2.1	SÁBADO	38
3.2.2	DOMINGO.....	42
3.3	RESULTADOS FINALES DEL SISTEMA	47
3.4	COMPARACIÓN DE MÉTODOS	50
4.4.1	ENTRE SEMANA.....	51
4.4.2	FIN DE SEMANA	54
4.	CONCLUSIONES Y RECOMENDACIONES.....	58
4.1	CONCLUSIONES.....	58
4.2	RECOMENDACIONES	59
5.	REFERENCIAS BIBLIOGRÁFICAS	60
ANEXOS		62
ANEXO A.....		62
ANEXO B.....		63
ANEXO C.....		68

ÍNDICE DE FIGURAS

Figura 1.1. Evolución de la demanda de energía eléctrica en el SNI	5
Figura 1.2. Representación horaria de curva de demanda para 3 tipos de consumidor	6
Figura 1.3. Representación de puntos borde y núcleo, b. Puntos del Clúster directamente accesible o no a la densidad.	9
Figura 1.4. (a) Obtención de distancia métrica con K=1. (b) Obtención de distancias métricas con K=4	12
Figura 1.5. Gráfico de distancias tipo K-dist y rodilla de función	13
Figura 2.1. Diagrama de flujo del ordenamiento en MATLAB	23
Figura 2.2. Clustering para la obtención de curvas características por medidor	24
Figura 2.3. Diagrama de Flujo para la validación de resultados finales	25
Figura 2.4. Diagrama de flujo total del ordenamiento y clustering	27
Figura 3.1. Codo de la función obtenido utilizando algoritmos complementarios	29
Figura 3.2. Clúster 1 de entresemana residencial en zona rural.	30
Figura 3.3. Clúster 3 de entresemana residencial urbana	30
Figura 3.4. Clúster 6 de entresemana otros rural.	31
Figura 3.5. Clúster 7 de entresemana otros urbano	32
Figura 3.6. Clúster 8 de entresemana RUIDO.	32
Figura 3.7. Índice de validación SILHOUETTE, -1 no pertenece al clúster, 1 pertenece.	33
Figura 3.8. Codo de la función encontrado utilizando algoritmos complementarios.....	33
Figura 3.9. Filtro 1 Ruido, Clúster 2 de entresemana Residencial Rural	34
Figura 3.10. Filtro 1 Ruido, Clúster 5 de entresemana Industrial Urbana	35
Figura 3.11. Filtro 1 Ruido, Clúster 6 de entresemana Comercial Urbana	35
Figura 3.12. Validación SILHOUETTE “R”, -1 no pertenece al clúster, 1 pertenece	36
Figura 3.13. Filtro 2 Ruido, Clúster 2 de entresemana Residencial Rural	37
Figura 3.14. Valor de codo de la función utilizando los algoritmos complementarios	38
Figura 3.15. Clúster 4 de sábado Residencial Urbano	38
Figura 3.16. Clúster 8 de sábado Otros Rural	39
Figura 3.17. Clúster 15 de sábado considerado RUIDO.	40
Figura 3.18. Validación con SILHOUETTE, -1 no pertenece al clúster, 1 pertenece	40
Figura 3.19. Codo numérico de la función obtenido con algoritmos complementarios	41
Figura 3.20. Filtro 1 Ruido, Clúster 1 de sábado Otros Urbano	41
Figura 3.21. Valor de codo de la función obtenido con los algoritmos complementarios ..	42
Figura 3.22. Clúster 9 de domingo Industrial Urbano	42
Figura 3.23. Clúster 1 de domingo Residencial Urbano	43
Figura 3.24. Clúster 14 de domingo considerado RUIDO.	44
Figura 3.25. Índice de SILHOUETTE, -1 no pertenece al clúster, 1 pertenece	44
Figura 3.26. Valor codo de la función obtenido con algoritmos complementarios.....	45
Figura 3.27. Filtro 1 Ruido, Clúster 1 de domingo Residencial Rural	45
Figura 3.28. Filtro 1 Ruido, Clúster 13 de domingo Comercial Urbana	46
Figura 3.29. Filtro 1 Ruido, Clúster 8 de domingo Industrial Urbana	46
Figura 3.30. Gráficas en por unidad PU de curvas típicas de la EEA. a) Curvas de días entresemana. b) Curvas de fin de semana.....	47
Figura 3.31. Curvas de demanda obtenidas de grupos de estudio de carga del capítulo 2 “Load Characteristics” de Turan Gönen [18].	50
Figura 3.32. Curva final entre semana Residencial urbana a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN.....	52
Figura 3.33. Curva final entre semana Residencial rural a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN.....	52

Figura 3.34. Curva final entre semana Comercial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	53
Figura 3.35. Curva final entre semana Industrial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	53
Figura 3.36. Curva final entre semana Otros a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	54
Figura 3.37. Curva resultante fin de semana Residencial urbana a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	55
Figura 3.38. Curva resultante fin de semana Residencial rural a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	55
Figura 3.39. Curva resultante fin de semana Comercial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	56
Figura 3.40. Curva resultante fin de semana Industrial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	56
Figura 3.41. Curva resultante fin de semana Otros a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN	57

INDICE DE TABLAS

Tabla 1.1. Estadísticas de equipamiento y capacidades en diferentes niveles del sistema eléctrico [4].	4
Tabla 2.1. Información de medidores por tipo de consumidor	16
Tabla 3.1. Descripción del tipo de consumidor con numeración.....	28
Tabla 3.2. Resultados en por unidad PU de curvas tipo entre semana.	48
Tabla 3.3. Resultados en por unidad PU de curvas tipo fin de semana	49

RESUMEN

En el presente trabajo se aborda la obtención de curvas características por tipo de consumidor utilizando un algoritmo de agrupamiento en el programa Matlab. Las curvas son logradas usando mediciones de los equipos de calidad de la Empresa Eléctrica Azogues (EEA). Se ingresan los datos de las mediciones en el programa MS Excel, para preparar la información y generar una base de datos ordenada con un formato homogéneo para el análisis. De forma complementaria se utiliza el sistema de cómputo numérico Matlab, que importa la base de datos y mediante un algoritmo iterativo de agrupamiento basado en densidad, genera curvas de demanda de potencia diaria para cada medición, con una desagregación de días laborales y fin de semana; identificándolos por tipo de consumidor. Los argumentos de entrada de la función Density Based Spatial Clustering Algorithm with Noise (DBSCAN) de Matlab requieren de distancias euclidianas, así como de su forma para agruparlos considerando la similitud de las curvas estudiadas.

El algoritmo DBSCAN permite identificar clúster entre grupos, tomando en cuenta su densidad y descubriendo grupos con formas arbitrarias. Los argumentos de entrada de esta función son: Épsilon o radio de la vecindad (Eps) y Mínimo de puntos en la vecindad (Mpts). La selección de los argumentos se realizó en base a la identificación del codo de la función en un gráfico de distancias tipo K-dist para encontrar Eps y el otro al considerar ciertos puntos núcleo y de borde en un mismo agrupamiento estableciendo el Mpts.

Para finalizar, se validan los grupos obtenidos mediante un índice de comprobación Silhouette (IS) y se identifican malos agrupamientos para luego determinar cuáles son los resultados finales que son presentados mediante tablas con valores horarios en días laborales y fin de semana. Se añade, también, la comparación gráfica de las curvas resultantes con la metodología anterior de mediciones reales; ya que presentaron picos diferentes en ciertos valores horarios en las curvas encontradas.

PALABRAS CLAVE: Curvas características, tipo de consumidor, algoritmo, Matlab, DBSCAN, comprobación.

ABSTRACT

In the present work the obtaining of characteristic curves by type of consumer is approached using a clustering algorithm in the Matlab program. The curves are obtained using measurements of quality equipment of the Electric Company under study. Measurement data is entered into the MS Excel program to prepare the information and generate an ordered database with a homogeneous format for analysis. In addition, the Matlab numerical computing system is used, which imports the database and through an iterative grouping algorithm based on density, generates daily load curves for each measurement, with a breakdown of weekdays and weekends, identifying them by type of consumer. The input arguments of the Matlab Density Based Spatial Clustering Algorithm with Noise (DBSCAN) function require Euclidean distances, as well as their shape to group them considering the similarity of the studied curves.

The DBSCAN algorithm allows to identify clusters among groups, considering their density and discovering groups with arbitrary shapes. The input arguments to this function are Epsilon or neighborhood radius (Eps) and Minimum points in the neighborhood (Mpts). The selection of the arguments was made based on the identification of the elbow of the function in a distance graph of the K-dist type to find Eps and the other when considering certain core and edge points in the same grouping, establishing the Mpts.

Finally, the groups obtained are validated by means of a Silhouette verification index and bad groupings are identified to later determine the results that are presented by means of tables with hourly values on weekdays and weekends. The graphic comparison of the resulting curves with the previous methodology of real measurements is also added; since they presented different peaks at certain hourly values in the curves found.

KEYWORDS: Characteristic curves, algorithm, type of consumer, Matlab, DBSCAN, verification.

1. INTRODUCCIÓN

Tomando como punto de partida la institucionalización del sector eléctrico y la expedición de la Ley Orgánica del Servicio Público de Energía Eléctrica (LOSPEE), las empresas eléctricas de distribución han aumentado su gestión orientando sus actividades al reforzamiento, renovación y modernización de infraestructura (operativa, administrativa, eléctrica) sujetándose a los mejores estándares de tecnología de la información, comunicación e infraestructura [1].

En este contexto, las empresas eléctricas realizan estudios en el sistema de distribución de forma periódica, para identificar y ejecutar las inversiones necesarias en infraestructura, con el objetivo de reducir pérdidas técnicas y mantener los estándares de calidad [1]. Estos estudios involucran el manejo de diversos datos de mediciones en bajo voltaje y mediciones en cabecera de alimentador para desarrollar un estudio correcto de acuerdo con la necesidad del operador del sistema y de los usuarios finales. Ciertos datos pueden presentar gran similitud u homogeneidad, es decir, que son datos fácilmente tratables en una herramienta de agrupamiento o de inteligencia artificial. Por tanto, los estudios no se realizan por cada elemento de un universo, sino se utilizan herramientas matemáticas ya desarrolladas que permiten realizar una distinción global del problema y presentar resultados correctos utilizando un esfuerzo computacional menor y menos exhaustivo para el operador del sistema eléctrico así como beneficios técnicos para el usuario final.

Se conocen diferentes alternativas metodológicas para la determinación de parámetros típicos o estándares que no son únicos, sino cambian según el modelo en estudio. En este caso, los métodos de agrupamiento o clustering son los indicados para el manejo de un gran volumen de información para su evaluación. Para esto se ha incursionado en análisis de minería de datos con un algoritmo de clustering basado en densidades que ofrece una alternativa rápida para el manejo de bases de datos y la obtención de curvas de demanda de potencia horarias.

La minería de datos son técnicas matemáticas para estimar comportamientos futuros en base al análisis de datos obtenidos. Un algoritmo en esta gran herramienta es el clustering o agrupamiento ya que realiza la identificación de patrones importantes y distribuciones semejantes en grandes bases de información. Existen algunos métodos en la obtención de grupos o clústeres que utilizan promedios, dendogramas, cuadrículas, pero el elegido en el presente trabajo de titulación es el método basado en densidades DBSCAN que identifica grupos con formas arbitrarias en presencia de datos atípicos o ruido [2].

1.1 OBJETIVOS

El objetivo general de este Proyecto Técnico es:

Obtener las curvas horarias de carga características desagregadas por tipo de consumidor mediante la técnica de agrupamiento espacial basado en densidad.

Los objetivos específicos del Proyecto Técnico son:

- Realizar una base teórica con mediciones reales para la obtención de curvas de demanda características según el tipo de consumidor en cabeceras de alimentador, circuitos secundarios y transformadores con información remitida por la empresa de distribución a utilizar como caso de estudio.
- Desarrollar un algoritmo en Matlab para analizar y validar los datos usando minería de datos con el algoritmo de agrupamiento basado en densidades para aplicaciones con ruido (DBSCAN).
- Validar los grupos obtenidos mediante el algoritmo de agrupamiento utilizando el índice silhouette.
- Presentar los resultados de las curvas de carga y evaluar su desempeño al compararlo con mediciones reales

1.2 ALCANCE

El presente trabajo de titulación ofrece una alternativa a la obtención de las curvas de demanda de potencia en las empresas eléctricas estudiando el algoritmo de agrupamiento basado en densidades DBSCAN y su potencial en la obtención de curvas de carga características por tipo de consumidor. Esto se consigue analizando la información remitida por la Empresa Eléctrica Azogues (EEA) respecto a las mediciones de los equipos de calidad instalados en cabeceras de alimentadores, clientes conectados en medio voltaje y transformadores de distribución para luego realizar la correcta agrupación utilizando un criterio técnico de minería de datos y finalizar con la validación mediante el índice de comprobación silhouette.

Una vez seleccionado el algoritmo de clustering o agrupamiento y el índice de desempeño, se desarrollará una herramienta computacional en Matlab que permita la identificación de curvas de carga desagregadas por tipo de consumidor utilizando mediciones en bajo y medio voltaje de clientes residencial rural y urbano, comercial, industrial y otros. Para iniciar, la herramienta extraerá los valores de mediciones procesados en una base de Datos

Excel y luego utilizará el algoritmo de agrupamiento basado en densidades para obtener las curvas de carga que más se asemejen al comportamiento de los diferentes usuarios. Luego, se tendrá un índice de validación de silhouette que realiza una comprobación de agrupamientos para filtrar el ruido o mediciones atípicas.

En base a la información analizada, se determinan las curvas de carga de los tipos de consumidores pertenecientes a la empresa distribuidora utilizando la técnica de minería de datos que consiste en el algoritmo de agrupamiento espacial basado en densidad de aplicaciones con ruido o Density-based Spatial clustering of applications with Noise (DBSCAN). Las curvas se presentarán de forma gráfica con los datos correspondientes, permitiendo la identificación de su comportamiento.

Finalmente, se realiza una comparación entre la metodología propuesta y la utilizada anteriormente por la distribuidora para evidenciar la mejora en la identificación de las curvas.

1.3 MARCO TEÓRICO

En esta sección se informará de los elementos y unidades encontradas en un sistema eléctrico así como su caracterización. También, se hará énfasis en las técnicas de minería de datos así como de todos los algoritmos complementarios utilizados para encontrar las curvas de demanda de potencia por tipo de consumidor.

1.3.1 SISTEMAS ELÉCTRICOS DE TRANSMISIÓN Y DISTRIBUCIÓN

En los países industrializados, los sistemas de distribución están encargados de entregar la energía eléctrica a todos los sectores concesionados por las empresas eléctricas. La energía se toma de las centrales de generación para distribuirlo a los usuarios finales mediante los sistemas de transmisión y distribución. De estos dos sistemas, el de menor atención o inversión es el de distribución; sin embargo, muchas veces es el componente más crítico en términos de afectaciones a la confiabilidad, la calidad del servicio, y el impacto estético (mayormente visual) que se tiene en la urbe o en el sector rural [3].

El sistema de distribución es el encargado del acondicionamiento de la energía. Es decir, proporciona electricidad con altos índice de confiabilidad para no interrumpir el servicio hacia el usuario final. Por tanto, es preciso desarrollar el párrafo anterior, en términos de equipamiento y capacidad porque según varían los niveles de voltaje, también cambiará el número de unidades y las capacidades promedio mostrándose estos datos de referencia en la Tabla 1.1. En un sistema eléctrico primero están los equipos de generación en las grandes centrales, comúnmente ubicados en sectores alejados; seguidamente el sistema

de transmisión y distribución (T&D) mueve la energía de las mencionadas centrales a través de líneas de T&D hacia subestaciones de transmisión/subtransmisión, transformadores de potencia y otro conjunto de equipos de seccionamiento/protección que permiten suministrar la energía eléctrica a la compañía distribuidora [4]. Es aquí, que la empresa distribuidora comienza a manejar el equipamiento en niveles de medios voltajes encargándose de los transformadores de distribución, alimentadores y circuitos secundarios. La Tabla 1.1. esquematiza lo antes dicho y proporciona estadísticas para cada una de las etapas del sistema eléctrico.

Tabla 1.1. Estadísticas de equipamiento y capacidades en diferentes niveles del sistema eléctrico [4].

Nivel del Sistema	Voltaje [kV]	Número de unidades	Capacidad promedio [MVA]	Capacidad Total [MVA]
Transmisión	345-138	12	150	1400
Sub-Transmisión	138-69	25	65	1525
Subestación	138/23,9 - 69/13,8	45	44	1980
Primario o alimentador	23,9 - 13,8	227	11	2497
Transformador de distribución	0,12 - 0,24	60000	0,05	3000
Secundario o acometida	0,12 - 0,24	250000	0,014	3500
Cliente	0,12	250000	0,005	1250

Revisando la Tabla 1.1. se identifica que mientras más cerca se encuentre a los usuarios finales el número de unidades se incrementa, así como la capacidad total de los mismos. Esto es debido a la no coincidencia de la demanda máxima en los usuarios finales dado por su diferencia en los hábitos de consumo y a intereses de confiabilidad en el sistema. Por lo que, a nivel de transformador de distribución siempre existirá una mayor capacidad total instalada, superando la capacidad del alimentador primario. Se entiende entonces que es fundamental realizar una caracterización de la carga para conocer los diferentes tipos de consumidores que se encuentran inmersos en el área de concesión de las empresas eléctricas y así entender las estadísticas mostradas previamente.

1.3.2 CARACTERIZACIÓN DE LA DEMANDA, DEFINICIONES Y SU INCREMENTO

Dado que la conducta de consumo de los usuarios varía por sector, tiempo y condiciones ambientales, se empezará caracterizando la demanda de acuerdo con 2 términos que ayudarán a comprender mejor las variables eléctricas principales del consumo energético para el usuario final, uno de ellos: la carga; y, así mismo, el comportamiento de la demanda

que termina ejemplificándose en curvas diarias de demanda de potencia. A continuación, se mostrarán las definiciones junto a sus unidades de medida.

- Carga: Hace referencia a lo estático, a la potencia eléctrica requerida de cierta instalación o equipo y sus unidades pueden ser en kVA o kW [5].
- Demanda: Es necesario incluir dos consideraciones más: el tiempo y la potencia promedio. En otras palabras, es la potencia requerida en kVA O kW promediado durante un intervalo de tiempo (15 min, 30min, 1 hora). Comúnmente se da en amperios a un nivel de voltaje específico [5].

Así mismo, la demanda en un sistema de distribución puede variar y su crecimiento se da por tres razones o por la combinación de ellas [1], la Figura 1.1. ilustra la evolución de la demanda en el Sistema Nacional Interconectado (SNI) del Ecuador.

1. Consideración del crecimiento tendencial del consumo y la población: Toma en cuenta el aumento del número de usuarios, es decir, nuevos usuarios utilizarán la energía eléctrica para su consumo aumentando así la energía demandada por esa zona.
2. La integración de importantes cargas como: sistema petrolero, proyectos mineros, proyectos camaroneros. Esta razón ejemplifica el aumento de demanda como incorporación de nuevas cargas que corresponden a proyectos especiales de minería, refinería, transporte, petroquímica, metalurgia, etc.
3. El cambio en la matriz productiva. (todo lo anterior, pero sumando las industrias básicas como Alumineras). Se debería ir apalancando esa demanda con la participación de la generación hidroeléctrica.

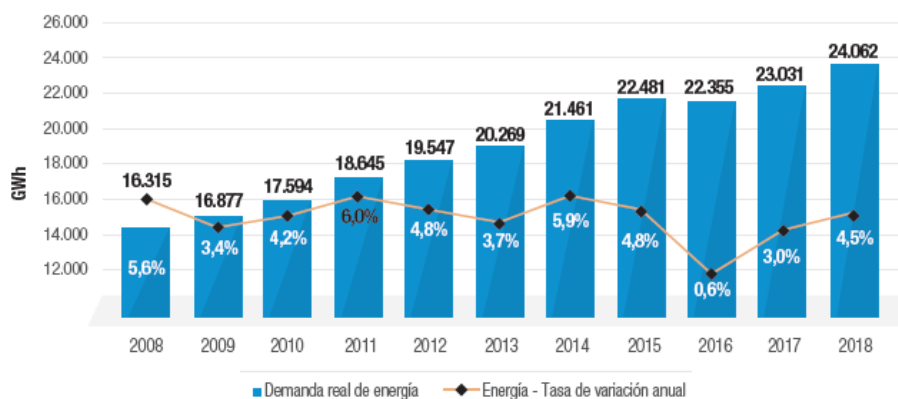


Figura 1.1. Evolución de la demanda de energía eléctrica en el SNI

1.3.2.1 Curva de demanda de potencia

El consumo de potencia eléctrica en el tiempo o energía dado por la carga eléctrica varía en cada instante del día, semana, temporada, mes o año. Por ejemplo, en verano se utilizan ventiladores, aires acondicionados, refrigeradoras, etc. En cambio, en época fría se utilizan otros equipos como calefacción o bombas de calor. [5]. El plan maestro de electrificación (PME) contempla la evolución de estos consumos y diferencia dos términos eléctricos: oferta y demanda [1]. En la demanda se tienen dos tipos: de potencia y de energía. La curva de demanda de potencia es conocida como curva de carga y así se optó en el título, los objetivos y el alcance. Sin embargo, desde este punto en adelante se utilizará el término curva de demanda porque los datos muestran la demanda promedio horaria mas no la carga instalada.

La demanda o requerimiento de potencia del usuario final denota la potencia requerida o consumida en cierto periodo de tiempo escogido por el analista. De modo que, el conocimiento de la variación de la carga y su comportamiento tipo debe ser representado por la curva de demanda con valores medidos para los diferentes usuarios [5]. Se representa en un eje cartesiano donde las ordenadas corresponden a la unidad de kWh/h y en las abscisas se tiene el tiempo normalmente representado en valores horarios. La Figura 1.2. ilustra esta relación denotando 3 diferentes tipos de consumidores y la discrepancia existente entre curvas.

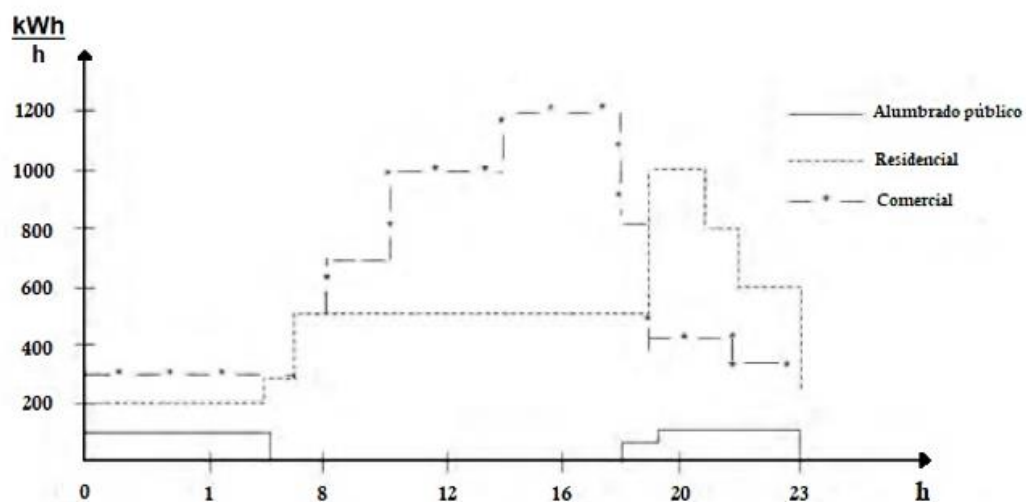


Figura 1.2. Representación horaria de curva de demanda para 3 tipos de consumidor

1.3.2.2 Factor de carga

Se define como la división entre la demanda promedio y la demanda máxima dentro de un periodo de tiempo. Se sabe que la potencia en un intervalo de tiempo es energía, por tanto, es necesario tener una relación con este término también. A continuación, se muestra la fórmula (1.1) ejemplificando este factor:

$$F_c = \frac{\bar{D}}{D_{MAX}} = \frac{E}{t * D_{MAX}} \quad (1.1)$$

Dónde:

\bar{D} : Demanda promedio en kW

D_{MAX} : Demanda máxima coincidente en kW

E : Energía consumida en kWh (día, semana, mes, año)

t : Número de horas durante el intervalo de tiempo (24, 96, 730, 8760)

El factor de Carga F_c es un valor numérico entre 0 y 1 e indica el grado de duración de la demanda máxima en el intervalo de tiempo estudiado. Este factor también permite tener una idea de la forma de la curva de demanda y del tipo de consumidor asociado a esa curva [6]. Por ejemplo, el tipo Residencial tiene un factor de carga que varía entre 0,5 y 0,6 debido a las costumbres variables de los usuarios. Después, el tipo comercial tiene factores que oscilan entre 0,6 y 0,7. Seguidamente, para industrias pesadas se tienen factores entre 0,7 y 0,8 notándose en la curva muy pocas variaciones. Finalmente, el alumbrado público al sufrir variaciones con picos, tanto en el horario matutino como vespertino, tiene factores de carga bajos que cambian entre 0,3 y 0,4 [5].

1.3.3 TÉCNICAS DE MINERÍA DE DATOS

La minería de datos es el proceso de extraer patrones o características importantes de bases de datos muy extensas, teniendo aplicación en la toma de decisiones y predicciones de comportamientos futuros mediante algoritmos matemáticos. Las técnicas de agrupamiento son una herramienta muy utilizada en la minería de datos y son algoritmos actualizados con métodos de análisis efectivos para hacer el uso correcto de la información. El mencionado clustering posee un conjunto de datos que se divide en diferentes grupos e identifica patrones importantes y distribuciones con formas interesantes en grandes bases de datos. La técnica de agrupamiento estudiada en este trabajo puede tener otras categorías como de particiones, jerárquico, en cuadrícula [7]. Pero el seleccionado será el algoritmo basado en densidades.

1.3.3.1 Métodos de agrupamiento

Los métodos de agrupamiento se basan en funciones de distancia y encuentran patrones o semejanzas para los datos en estudio con algoritmos computacionales [7]. Los algoritmos de partición buscan determinar un número k de particiones que optimizan una cierta función objetivo donde cada dato se asigna al grupo más cercano. Este algoritmo matemático fue propuesto en 1990 por Kaufman & Rousseeau [8], siendo así muy antigua para su uso actual. En adición, Kaufman explicó que los métodos jerárquicos trabajan de forma ordenada creando una descomposición jerárquica del conjunto dado de datos formando dendogramas con características de acumulación o de división. Lo negativo en este método es que no identifica clústeres con datos de formas arbitrarias.

Por otra parte, existen los algoritmos de agrupamiento basados en cuadrículas que utilizan una estructura de datos de cuadrícula donde se dividen los datos de entrada en celdas hiperrectangulares, se descartan las celdas con menor densidad y los clústeres son formados agrupando celdas adyacentes de alta densidad. Tal procedimiento exige un alto esfuerzo computacional y no es eficiente cuando se trata de largas bases de datos.

Los algoritmos basados en densidades superan la dificultad de detectar clústeres de formas arbitrarias al extraer regiones densas definiéndolas como un grupo y comprobando si estos puntos son accesibles a la densidad o no [7].

1.3.3.2 Algoritmo basado en densidad DBSCAN

Es un algoritmo de agrupamiento de datos o data clustering que matemáticamente encuentra un número de grupos comenzando por una estimación de la distribución de densidad de los nodos correspondientes. Identifica el vecino k más cercano de cada punto, en términos de distancia Euclidiana, luego calcula el promedio de todas estas distancias. Para cada punto del conjunto de datos, el algoritmo identifica los puntos vecinos que sean directamente accesibles a la densidad, esto es; que cumpla el umbral de radio de la vecindad (Eps) y el mínimo número de puntos ($Mpts$) proporcionado por el usuario, y los clasifica en puntos centrales, de borde y de ruido. Luego recorre todos los puntos del conjunto y para los puntos centrales construye un nuevo clúster con la ayuda de la función `RegionQuery` que comprueba la distancia a todos los puntos cercanos en los diferentes clústeres encontrados. La idea clave del algoritmo DBSCAN es que, para cada punto de un clúster, el vecindario delimitado por un radio dado contiene al menos un número mínimo de puntos, es decir, la densidad en el vecindario tiene que superar un umbral predefinido [8].

Explicando lo anterior, parte de un dato u objeto arbitrario. Luego, si el vecindario a su alrededor está dentro de Eps y satisface el valor Mpts este se convertiría en un punto núcleo y la búsqueda continuaría con su vecindad y se detiene en los puntos de borde teniendo que todos los puntos dentro del clúster están en la cercanía de uno de los puntos núcleo. Seguidamente, otro conjunto de datos es analizado y seleccionado para repetir el proceso hasta que todos los puntos de datos en el conjunto analizado hayan sido colocados en agrupamientos o clústeres. Existen puntos que no son núcleo ni están asociados a un punto núcleo de su vecindad, estos son etiquetados como puntos ruido o Noise. DBSCAN no necesita un número final de clústeres finales que se especifique previamente, más bien detecta automáticamente el número de clústeres de acuerdo con las densidades en las regiones [7].

La elección del umbral predefinido se la realiza con ayudas computacionales o funciones matemáticas en software de simulaciones, pero en esencia si se tienen 100 puntos en un clúster, cada uno de estos no cumplirá el Mpts cuando se elige un valor alto para este argumento, es decir que habría varios puntos de borde que no cumplirían la condición, pero que si pertenecen al clúster como se muestra en la Figura 1.3. en esa forma arbitraria de puntos. Tampoco se recomienda elegir un valor de 1 debido a que todos los puntos serían un punto núcleo y, por tanto, se tendrá un grupo diferente por cada dato [9].

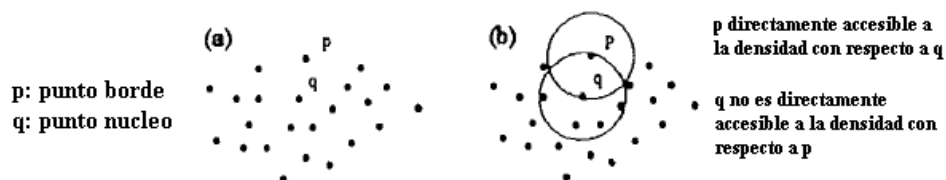


Figura 1.3. Representación de puntos borde y núcleo, b. Puntos del Clúster directamente accesible o no a la densidad

1.3.3.3 Validación de agrupamientos mediante el índice Silhouette (IS)

Este método se utiliza con bastante frecuencia para evaluar los resultados del agrupamiento obtenido analizando la cohesión y la separación de los valores [10]. En adición, no se necesita un conjunto inicial de condiciones para evaluar la veracidad del grupo formado, esto hace que sea más apropiado para las tareas de agrupación. Es así como, en este trabajo se usa el índice de Silhouette para evaluar la eficiencia del clúster [11].

El ancho de silueta $s(x_i)$ para el punto x_i comienza calculando las distancias promedio representados por a_i y b_i de la siguiente manera [10]:

$$a_i = \frac{1}{m_j - 1} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(x_i^j, x_r^j), \quad i = 1, 2, 3, \dots, m_j \quad (1.2)$$

Donde:

a_i : Distancia promedio del i-ésimo dato con todos los datos en un clúster

m_j : Cantidad de datos en el j-ésimo clúster

$d(x_i^j, x_r^j)$: Distancia del i-ésimo con el r-ésimo dato de un clúster j

$$b_i = \min_{\substack{n=1, \dots, k \\ n \neq j}} \left\{ \frac{1}{m_n} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(x_i^j, x_r^n) \right\}, \quad i = 1, 2, 3, \dots, m_j \quad (1.3)$$

Siendo:

b_i : Distancia mínima promedio del i-ésimo dato con todos los datos en los k clúster

k : Numero de clúster encontrados

m_n : Cantidad de datos en el k-ésimo clúster

$d(x_i^j, x_r^n)$: Distancia del i-ésimo dato con el r-ésimo dato de todos los datos

Obteniendo el IS para el i-ésimo dato con (1.4):

$$IS_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (1.4)$$

Calculando el valor medio de los anchos de silueta IS usando (1.5):

$$IS_j = \frac{1}{m_j} \sum_{i=1}^{m_j} IS_i \quad (1.5)$$

Finalmente, se calcula el índice de silueta global como el valor medio de los índices de los grupos totales con (1.6):

$$IS = \frac{1}{k} \sum_{j=1}^k IS_j \quad (1.6)$$

1.3.4 ALGORITMOS COMPLEMENTARIOS

Luego de haber comprendido la parte teórica y matemática de los algoritmos principales es necesario explicar los algoritmos complementarios que aportan al correcto desarrollo y resultados de DBSCAN. Los criterios corresponden a reconocimiento y clasificación de patrones pero esta vez utilizando las distancias métricas entre los datos en estudio.

Con el aumento de la tecnología, los ordenadores son más potentes y su menor costo, sumado a la necesidad de analizar enormes conjuntos de datos con millones de filas, han permitido que se desarrollen técnicas computacionales para la realización de gráficos tipo K-dist con comandos de K-nearest neighbor (KNN) y también se suma la Knee definición para el análisis del mencionado gráfico resultando en un valor numérico de codo de la función de distancias euclidianas [12].

1.3.4.1 Algoritmo KNN

Para el algoritmo KNN se pueden tener 2 tipos de predicción utilizando la minería de datos. El primero, usado en la lógica de programación de este trabajo, es el de clasificación siendo este solo una predicción de a cuál categoría de datos va a pertenecer.

El segundo tipo de predicción es la regresión, donde se predice que valor numérico tendrá una variable en base a valores existentes. En la tarea de reconocer patrones, el algoritmo KNN o los K-vecinos cercanos es el método más simple para la clasificación de objetos basado en ejemplos de entrenamiento o de bases de datos. Es correcto afirmar que KNN es un tipo de minería de datos que tiene un aprendizaje basado en instancias o aprendizaje rápido donde la función solo se aproxima localmente y todo el cálculo computacional se realiza desde la clasificación y determinación de distancias euclidianas hasta una posible agrupación de datos como se muestra en la Figura 1.4.a [12].

KNN es una técnica de clasificación fundamental cuando hay poco o ningún conocimiento previo sobre la distribución de los datos. Esta regla procede a retener todo el conjunto de entrenamiento durante el aprendizaje y asigna a cada grupo una clase representada por la etiqueta de sus k-vecinos más cercanos en el conjunto de entrenamiento. En otras palabras, K es el número de vecinos cercanos en los que se desea el cálculo de distancia métrica, esto luego se traduce a un término "n" representando el número de mediciones para cada conjunto de datos. En la Figura 1.4.b se ejemplifica el cálculo de forma espacial teniendo dos tipos de agrupaciones de acuerdo con el valor que se asigna a K.

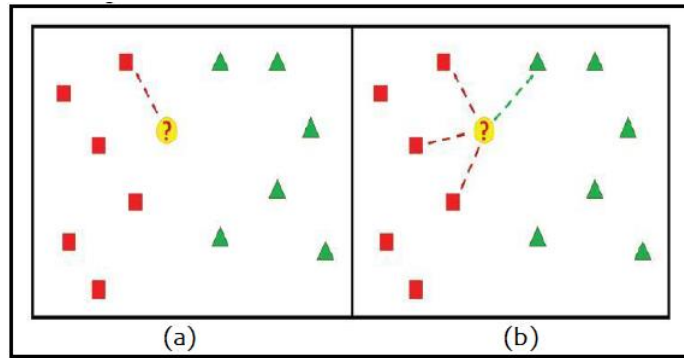


Figura 1.4. (a) Obtención de distancia métrica con K=1. (b) Obtención de distancias métricas con K=4

Como se mencionó anteriormente, KNN hace predicciones basadas en el resultado de los vecinos K más cercanos a ese punto. Pero para lograr predicciones correctas con KNN, se necesita definir una métrica para medir la distancia entre el dato estudiado y los casos de la muestra disponibles para el cálculo. Una de las opciones más utilizadas para medir esta distancia se conoce como euclidiana. Otras medidas incluyen euclidiana al cuadrado, bloque de ciudad y Chebychev.

1.3.4.2 Knee Definición

El termino Knee hace referencia a una rodilla pero si se entiende de forma matemática establece un cambio brusco representado como un pico donde fácilmente se puede asemejar a encontrar también el codo de la función. La dificultad de definir formalmente una rodilla es que los resultados “muy buenos” en un sistema pueden no ser “suficientemente buenos” en otro. Dado que las rodillas solo sirven como una aproximación, el analista del sistema las interpreta de manera diferente en diferentes situaciones. Por lo tanto, la detección de la rodilla es un proceso inherentemente heurístico. Sin embargo, para programar un algoritmo de detección de rodilla independiente de la aplicación, se requiere una definición consistente aplicable a cualquier sistema. En este trabajo, como en [13], se define de forma matemática a la rodilla para una función continua como la curvatura encontrada en su gráfica. Para cualquier función continua $f(x)$, existe una forma cerrada estándar $K_f(x)$, que define la curvatura de $f(x)$ en cualquier punto como una función de su primera y segunda derivada como se muestra en (1.7):

$$K_f(x) = \frac{f''(x)}{(1 + f'(x)^2)^{1.5}} \quad (1.7)$$

El punto de máxima curvatura se adapta bien al método operacional AD hoc usado para seleccionar una rodilla, debido a que la curvatura es una medida matemática de cuánto

difiere una función de una línea recta. Como resultado, la curvatura máxima captura los puntos anteriores y posteriores utilizados para identificar el valor de la rodilla. Es muy importante notar que la curvatura es independiente de la aplicación y no depende de la relación entre los parámetros del sistema y el rendimiento. La detección de rodilla depende de la selección adecuada y ajustable de parámetros del sistema y métricas de rendimiento, valores que luego se adaptan como valor de entrada en funciones de Matlab preestablecidas. Es decir, los ejes son ajustables a la aplicación deseada. Para este caso, los datos de las distancias métricas obtenidas con cierto número n de mediciones horarias son provistas por el algoritmo KNN, y su gráfica se asemeja a la Figura 1.5. donde se observa un gráfico tipo K-dist con su respectiva inflexión o curvatura denotada en rojo [13].

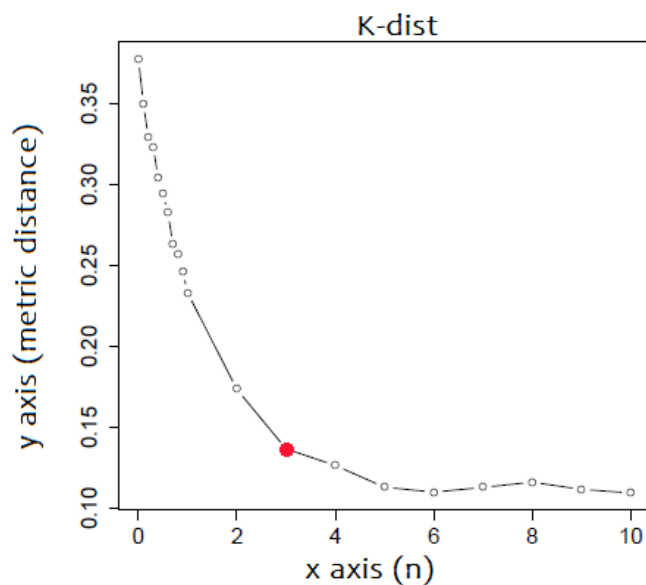


Figura 1.5. Gráfico de distancias tipo K-dist y rodilla de función

1.3.5 SOFTWARE EMPLEADOS

Para la implementación del algoritmo DBSCAN es necesario trabajar con 2 programas computacionales. Uno para generar la base de datos y otro programa capaz de tomar información de la base de datos y ejecutar el algoritmo de clustering basado en densidades. Los programas que se ajustan a esta necesidad son: Excel MS y Matlab.

1.3.5.1 Excel MS

Microsoft Excel es un programa del paquete Office con hojas cálculo de última generación. Estas hojas son herramientas muy convenientes para los cálculos numéricos y, de hecho, en la parte computacional esta herramienta equivale a muchos softwares basados en

lenguajes de programación más elaborados para la computación numérica [15]. Los programas con hojas de cálculo son únicos por las siguientes propiedades importantes:

- Creación de fórmulas con una interfaz de usuario basado en apuntar y hacer clic.
- Referencias de celdas relativas y absolutas en lugar de variables con nombre.
- Iteración y cálculos múltiples copiando fórmulas.
- Recálculo automático al cambiar los valores de entrada.

Se utilizaron comandos de condición como: SI, MAX.SI.CONJUNTO ayudando a la obtención de datos siguiendo cierta especificación. También, se usaron comandos de ubicación como BUSCARV para satisfacer la necesidad de encontrar elementos en una tabla y migrar a otra tabla o solo un rango por fila especificada.

1.3.5.2 Matlab

Matlab (abreviación de MATrix LABoratory) es un software muy utilizado en los campos de ciencia e ingeniería ya que integra la matemática computacional, herramientas visuales y un poderoso lenguaje que provee un ambiente flexible para la utilización de algoritmos computacionales.

Es de conocimiento profesional que Matlab es preferido por sus cálculos matriciales y vectoriales altamente optimizados y también ofrece un lenguaje intuitivo para expresar problemas y sus soluciones de forma matemática en scripts o de forma visual mediante interfaz gráfica [16]. Los usos típicos se denotan a continuación:

- Desarrollo de algoritmos y cómputo numérico
- Cálculo simbólico (con las funciones matemáticas).
- Modelado, simulación y creación de prototipos.
- Procesamiento de señales y análisis de datos.
- Visualización en Interfaces gráficos de ingeniería.

2. METODOLOGÍA

En este capítulo se presenta la descripción detallada de los diferentes procesos realizados para la obtención de curvas de demanda utilizando los dos softwares mencionados previamente. Primero, se comenzará con una explicación de las funciones y fórmulas aplicadas en Excel para generar la base de datos y así utilizar esa base de datos como elemento de entrada en el algoritmo de clustering basado en densidades DBSCAN desarrollado en MATLAB.

2.1 TRATAMIENTO DE INFORMACIÓN REMITIDA POR LA EEA

Las empresas eléctricas con el afán de cumplir una de las políticas expuestas en el capítulo 6 del plan maestro de electricidad ecuatoriano que especifica: “Atender la demanda del servicio de energía eléctrica, propendiendo al incremento de cobertura, mejorando los niveles de calidad, confiabilidad, continuidad, minimizando las pérdidas e incluyendo además el aprovechamiento de los recursos de energía distribuida [1]. Han comenzado con campañas de mediciones en medio y bajo voltaje para proyectos de Balances energéticos, pérdidas, análisis de armónicos, etc.

Estas mediciones, en intervalos de tiempo variables ayudan también al análisis de los diferentes tipos de consumidor al ejemplificar los datos en curvas de demanda horarias. Estos datos de mediciones migradas a un formato específico puede dar una idea del comportamiento del usuario, así como al grupo de consumo al que pertenece. Aprovechando esta identificación, se solicitó las mediciones en bruto de 236 medidores en medio y bajo voltaje colocados en diferentes puntos de la red en estudio.

Para el manejo de estos datos se utilizaron los programas que vienen con el manual de usuario del equipo de medición teniendo más facilidad en generar un archivo que se pueda entender y que sea modificado por el software Excel Ms. Este archivo se seleccionó con formato “.txt” para que así sea exportado a un archivo Excel mediante la utilización de Visual Basic.

2.1.1 INFORMACIÓN DE MEDIDORES

Los 236 medidores con datos de mediciones son obtenidos de los alimentadores principales del subsistema de distribución de la EEA, así como de usuarios conectados a estos circuitos primarios. Estos medidores se encuentran clasificados por tipo de consumidor en el catastro de usuarios para la EEA. Por tanto, la Tabla 2.1 muestra lo antes dicho, identificando el nivel de voltaje (NV) de las mediciones junto al año en que estos

datos fueron provistos para el análisis. Los valores totales para cada tipo de consumidor son mostrados en negrilla.

Tabla 2.1. Información de medidores por tipo de consumidor

TIPO CONSUMIDOR	NV - BAJO			TOTAL BAJO	NV-MEDIO		Total MEDIO	Total general
	2017	2018	2019		2018	2019		
Comercial		15	10	25	11	9	20	45
Industrial	1	10	7	18	9	5	14	32
Otros		6	3	9	8	1	9	18
Residencial		88	43	131				131
Sin Estrato		3	2	5	5		5	10
Total general	1	122	65	188	33	15	48	236

La tabla anterior muestra una peculiaridad denotando un tipo de consumidor Sin Estrato, esto se atribuye a diferentes usuarios que tienen convenios de pago, hospitales, consumos gubernamentales o educación en general.

2.1.2 MACRO CON LENGUAJE VISUAL BASIC

Excel tiene un complemento especial para programadores. Todas las versiones actuales de los programas de Microsoft Office tienen un lenguaje de programación incorporado “VBA” y un entorno de desarrollo integrado “IDE” para este lenguaje. El lenguaje es dinámico y relativamente completo. También permite acceder a otras bibliotecas instaladas ya sea en el mismo lugar de trabajo o incluso en otra computadora accesible por una conexión de red [15]. Este mecanismo combinado permite a los programadores definir nuevas funciones de hoja de trabajo perfectamente integradas en Excel, siendo de mucha ayuda para este trabajo de titulación ya que integrará también documentos tipo “.txt” de forma masiva en las diferentes pestañas que puede tener la hoja de cálculo Excel y así se aprovechará el orden correcto a la hora de tener las mediciones por cada equipo.

2.1.3 ORGANIZACIÓN DE MEDICIONES

Una vez terminada la migración de archivos “.txt” a Excel utilizando una macro de organización. Se tienen todos los medidores en las pestañas del archivo Excel y se procede a realizar gráficas horarias del consumo energético en kWh. Para esto se deben organizar las columnas y elegir la Potencia promedio consumida versus las 24 horas del día representado de 0-23h, obteniendo así las Curvas de Demanda diaria para cada medidor. Pero en este paso, se observan mediciones atípicas que no coinciden con el comportamiento diario de los usuarios asociados dando lugar a la creación de una base de

datos más elaborada donde se espera determinar y eliminar esas mediciones que no deberían estar consideradas para el análisis.

2.2 CREACIÓN DE BASE DE DATOS EN EXCEL MS

Para la base de datos se pasaron los 236 medidores junto al consumo energético total registrado de forma horaria por estos equipos de medición y se habilitaron nuevas columnas que muestran los datos con más información para la presentación posterior de resultados como la hora sola, la fecha, la Potencia máxima por día, el día de la semana, la clave con el grupo de consumo y, finalmente, la clave junto a la parroquia donde se encuentra este medidor para una mejor ejemplificación de los resultados.

2.2.1 ETIQUETAS EN MEDICIONES

Se han destinado acrónimos para identificar las diferentes columnas de la base de datos a utilizar para el análisis de clustering. Esta base es uno de los objetivos específicos del trabajo de titulación y es necesario que se entiendan los valores tanto reales como en por unidad (pu) utilizados como elemento de entrada en el algoritmo de agrupamiento de Matlab. Las etiquetas mostradas en la primera fila de la base de datos son asignadas a columnas específicas de la base de mediciones y se utilizan para el manejo correcto de datos en el algoritmo de agrupamiento en Matlab. DBSCAN presenta los resultados finales de forma horaria desagregado por tipo de consumidor, periodo de la semana y forma de la curva característica. Es así como, toma gran importancia y es necesaria la explicación de las etiquetas junto a sus definiciones.

- Medidor: Representa el equipo de medición al que le pertenecen todos los datos.
- Fecha: Fecha en formato largo obtenido desde el software de los equipos de medición.
- Hora: Hora en formato largo obtenida del software de los equipos de medición.
- Promedio P: Valores promedios de potencia en kW obtenidos de los equipos de medición.
- Promedio Q: Valores promedios de potencia kVAr obtenidos de los equipos de medición.
- FechaSola: Fecha de medición en formato corta obtenida de la columna "Fecha".
- HoraSola: Hora de medición en valores del 0-23 para la ejemplificación de resultados.

- PmaxPorDia: Potencia Máxima en kW obtenida por día de cada medidor.
- PuPotProm: Valor en pu de las mediciones por día por cada medidor.
- DiaSem: Identificador de día de la semana para la presentación de resultados.
- ClaveGCon: Etiqueta de tipo de consumidor para los medidores 10, 20, 30 o 40.
- GrupodeConsum: Identifica etiqueta ClaveGCon atribuyendo el tipo de consumidor.

2.2.2 TRATAMIENTO DE VARIABLES EN POR UNIDAD

El tratamiento de variables en pu permite el trabajo de variables reales de la ingeniería eléctrica en un sistema que simplifica las operaciones matemáticas ofreciendo respuestas acertadas y con un tiempo de resolución más rápido. Tomando en cuenta lo anterior, se normalizaron los valores reales de “Promedio P” en cada día obtenido del medidor tomando como potencia base la máxima para el mismo periodo diario. En otras palabras, se tendrán curvas de días entre semana y fines de semana en cada medidor para normalizar los datos por día utilizando la ecuación (2.1):

$$P_{pu} = \frac{\text{Promedio } P}{P_{\text{maxPorDia}}} \quad (2.1)$$

Donde:

P_{pu} : Valor adimensional en pu que representa el consumo de potencia medida.

Promedio P: Valor promedio de potencia activa en kW de la medición.

$P_{\text{maxPorDia}}$: Valor máximo de potencia activa en kW de la medición por día.

2.3 ALGORITMO DBSCAN DESARROLLADO EN MATLAB

Una vez que se dispone de una base de datos organizada, se importa cada medición utilizando ciclos iterativos en MATLAB, se agrupan dos tipos de curvas de demanda características de acuerdo con el día de la semana, es decir, un grupo de lunes a viernes y otro de sábados con domingos. También, se identifica el grupo de consumo siendo residencial, comercial, industrial y otros. En este proceso se utiliza la función *Knnsearch* para encontrar un gráfico de k-distancias de las curvas, luego la función *knee_pt* para encontrar el codo de esta gráfica, es decir el valor de *Eps* y su zona típica, que es un argumento necesario para establecer los clústeres en la función DBSCAN. El argumento *Mpts* se establece a un valor relativamente bajo, de manera que se incluyan todos los puntos pertenecientes al mismo grupo.

Seguidamente, la función DBSCAN realiza el clúster entre las mediciones y se guardan las curvas de demanda obtenidas en arreglos matriciales. El algoritmo final permite identificar curvas atípicas o Ruido para luego validar los agrupamientos usando el índice de comprobación de clúster o índice de silueta. Este índice indica cuales son los grupos que no reflejan similitud en sus curvas, para reasignarlos de acuerdo con filtros de tratamiento del ruido ingresados en el mencionado algoritmo.

Para las curvas de demanda de lunes a viernes, se aplicó el índice de silueta y se comprobaron los grupos formados, así como los agrupamientos de las mediciones ingresadas de forma errónea como ruido. Para las curvas de fin de semana, se aplicó un solo filtro de comprobación de los grupos. Al final se obtuvieron curvas de demanda típicas: residenciales urbanas y rurales, comerciales, industriales y otros. Este algoritmo requiere de tres parámetros de entrada y retorna un vector idx con el agrupamiento o clúster resultante. Seguidamente, la formula en Matlab:

$$idx = DBSCAN(X, Eps, Mpts) \quad (2.2)$$

Donde:

X : Tamaño de la lista de vecinos o matriz de datos;

Eps : Radio que delimita el área del vecindario de un punto (vecindario-Eps);

$Mpts$: Número mínimo de datos u objetos en la vecindad de Eps

Una posible estrategia para seleccionar un valor de épsilon es generar un gráfico de K distancias con la función `Knnsearch` del punto 2.3.1; luego, para cada punto se encuentra la distancia al punto más cercano y se trazan los valores ordenados de distancias. Generalmente, el gráfico contiene un punto de inflexión “rodilla” y la coordenada en x que corresponde a dicho punto, se encuentra con la función `knee_pt` del punto 2.3.2. Típicamente este valor o valores inferiores representan una buena opción para establecer el umbral adecuado de Eps, porque fuera de esta región sus valores se ubicarían en territorio atípico o también llamado ruido [8].

2.3.1 KNEESEARCH

Esta función busca los vecinos más cercanos, de acuerdo con distancias euclidianas, y devuelve sus índices en un vector de columna d y también sus respectivas distancias kD .

Utiliza datos de entrada o una matriz de datos. En la ecuación (2.3) se muestra su estructura en MATLAB para el correcto uso de esta.

$$[d, kD] = knnsearch(X, Y, Name, Value) \quad (2.3)$$

Donde:

X: Valores en por unidad PU de datos 0-23 horas para los días de medición.

Y: Valores en por unidad PU de datos 0-23 horas para los días de medición.

Name: Se escribe "K" y calcule las distancias de vecinos más cercanos.

Value: Tamaño de matriz de cuantos vecinos cercanos en la métrica de distancia.

El uso de las distancias en un gráfico de K-dist permite un cálculo eficiente del vecino k más cercano a un punto, especialmente para base de datos grandes. Por lo tanto, es necesario realizar el gráfico de la primera columna de la matriz D, ya que esta contendría el vector de distancias más cercanos en los datos estudiados.

2.3.2 KNEEPOINT

La función kneepoint avanza a lo largo de la gráfica tipo K-dist de distancias, un punto de bisección a la vez y ajustando dos líneas siendo estas la primera derivada y la segunda derivada. Es decir, se utiliza el principio estudiado en el punto 2.4.2 pero Matlab mejora esta función introduciendo una función objetivo que en lugar de calcular una línea en cada lado se calculan líneas hacia todos los puntos a la izquierda del punto de bisección y otra a todos los puntos a la derecha de la bisección. Se considera que la rodilla está en un punto de bisección que minimiza la suma de errores para los dos ajustes. Esta se considera como si fuera un punto de partida (threshold) que corresponde a un umbral, donde un cambio agudo de gradiente ocurre a lo largo de la curva de k-distancias. Cualquier valor menor a este umbral de densidad Eps, puede agrupar eficientemente patrones debido a que estos se encontrarían en un territorio típico del gráfico k-dist, lo que implica patrones o puntos pertenecientes a cierta densidad dentro de los datos estudiados [13].

$$x = knee_pt(distancias) \quad (2.4)$$

Donde:

distancias: Valores resultantes de distancias kD euclidianas del grafico tipo K-dist.

x : Valor en x del codo del grafico tipo K-dist.

2.4 VALIDACIÓN DE GRUPOS MEDIANTE INDICE SILHOUETTE

El procedimiento matemático explicado en el punto 2.3.3 es representado mediante una gráfica en Matlab que valida las técnicas de partición o agrupamiento. Cada grupo está representado por una silueta, que se basa en la comparación de su cercanía y separación. Esta silueta muestra los objetos que están bien clasificados dentro de su grupo, y cuáles simplemente están infiltrados entre los grupos. La agrupación se muestra combinando las siluetas en una sola gráfica, lo que permite apreciar la calidad relativa de los grupos y una descripción general de la configuración de datos. El ancho promedio de la silueta proporciona una evaluación de la validez de agrupamiento, y podría usarse para seleccionar un número "apropiado" de clústeres [17]. El formato de la función se lo muestra en la ecuación (2.5), donde se observa que tiene dos elementos de entrada y un vector de retorno.

$$s = \text{silhouette}(X, idx) \quad (2.5)$$

Donde:

X : es la recopilación de los datos entre objetos

idx : es la partición obtenida (por la aplicación de alguna técnica de agrupamiento o clúster)

s : son los valores de retorno de la función, se encuentran entre -1 y 1, denotando a 1 como perteneciente al clúster y a -1 como la no pertenencia al clúster.

$$-1 < s < 1 \quad (2.6)$$

2.5 ORDENAMIENTO Y CLUSTERING

El ordenamiento y clustering se realiza en el ambiente de programación de MATLAB utilizando como elemento de entrada la base de datos elaborada para el presente estudio.

Se puede observar en la Figura 2.4. el diagrama de flujo total del algoritmo y se comienza con el ordenamiento de datos y usando un algoritmo de clustering obtener una curva característica para los días entre semana, para sábado y otra para domingo. Después, mediante un procesamiento de filtros se obtienen resultados parciales que posteriormente se validan con el índice de Silhouette para mostrar los finales.

En los diagramas de los puntos siguientes se notan conectores circulares con letras y números, estos resaltan la relación existente entre los diferentes flujogramas. Basta con analizar entre la Figura 2.1., Figura 2.2., y la Figura 2.3. para establecer los caminos de conexión que resultan en el diagrama total expuesta en Figura 2.4.

2.5.1 FLUJO DE ORDENAMIENTO

Así como se menciona en el punto 1.3.5.2 uno de los beneficios del uso de MATLAB es el análisis de datos. En este caso, la herramienta utilizada para dividir el análisis de clustering de forma iterativa para las curvas de demanda horarias por medidor es el lazo FOR. En la Figura 2.1. se pueden observar los tres lazos del flujo de ordenamiento representados por las variables i , j y k donde ya se trabajan en valores pu obtenidos de la base de datos. Esto resulta en una matriz de algunas curvas diarias por medidor que son clasificadas en etiquetas de días entre semana y otra de fin de semana para realizar el clustering explicado en el diagrama de flujo del siguiente punto 2.5.2 y así obtener curvas características de datos entre semana junto a los datos de fin de semana para cada medidor.

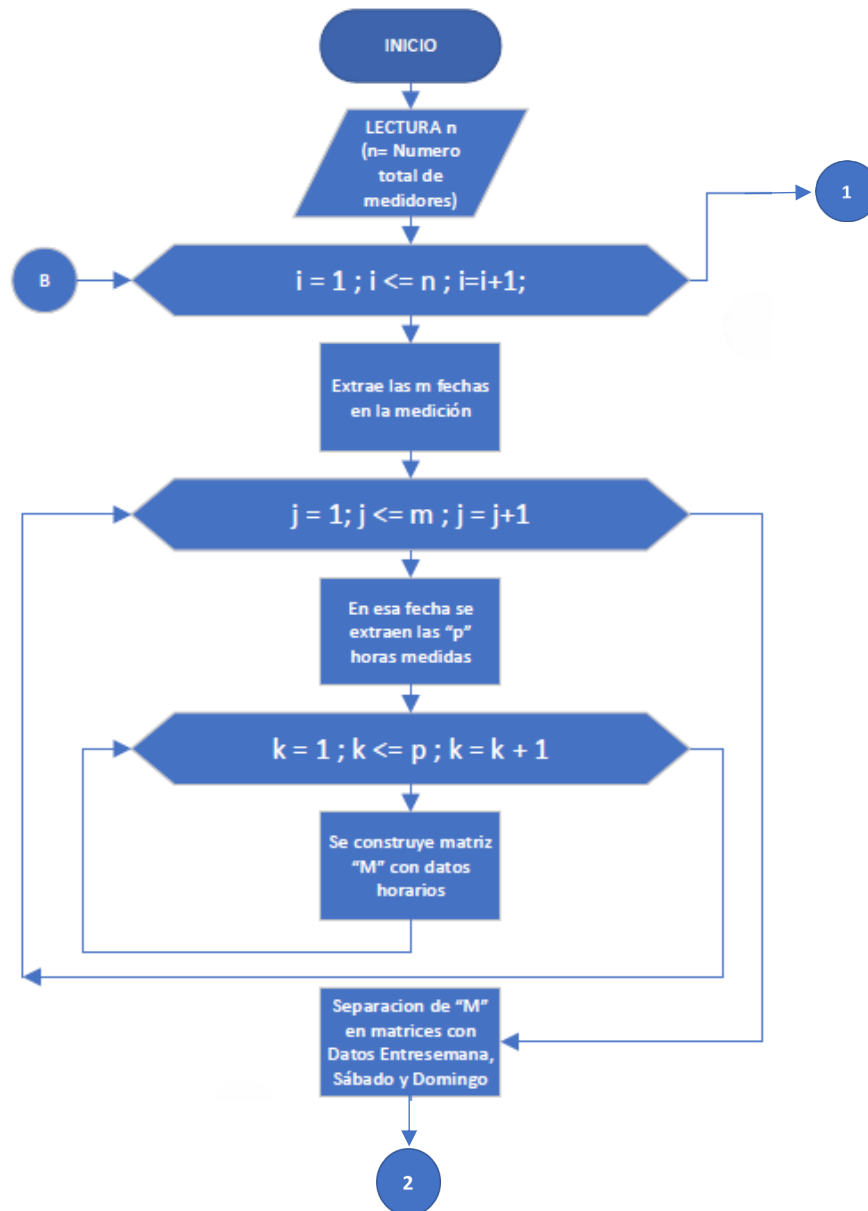


Figura 2.1. Diagrama de flujo del ordenamiento en MATLAB

2.5.2 CURVA CARACTERÍSTICA POR MEDIDOR

Una vez obtenidas algunas curvas diarias de un medidor se procede a realizar el clustering de los datos entresemana y fines de semana para obtener la curva característica por cada medidor. El rango de medición del equipo de calidad es de 10-15 días; por tanto, existe un esfuerzo computacional mayor para encontrar la curva característica entre semana que la de fin de semana. En la Figura 2.2. se puede observar el procedimiento establecido para el clustering de las mediciones tanto para datos entre semana como los de fin de semana. En el clustering entresemana se tienen 3 pasos. Primero, se encuentran los datos de

entrada para el DBSCAN; estos son: Eps y Mpts. Segundo, se realiza el clustering con la función DBSCAN. Tercero, se obtiene el promedio se alisa y se almacena en una matriz auxiliar que almacena las curvas características de los 236 medidores.



Figura 2.2. Clustering para la obtención de curvas características por medidor

2.5.3 CLUSTERING, FILTROS Y VALIDACIÓN

Luego de culminar las curvas características para los 236 medidores se procede con las tres etapas, la primera, el clustering global con la información de curvas entre semana y fin de semana para obtener la curva característica total del sistema. Segundo, el uso de filtros para obtener curvas que no se encontraron en un agrupamiento inicial. Y, tercero, validar los grupos obtenidos con el índice de Silhouette. La Figura 2.3. muestra el diagrama de flujo que incluye los 3 pasos mencionados anteriormente y que se ayuda de las etapas de

filtrado para eliminar ruido indeseable de curvas atípicas, así como del índice de Silhouette para comprobación del clustering y luego almacenar los resultados finales.

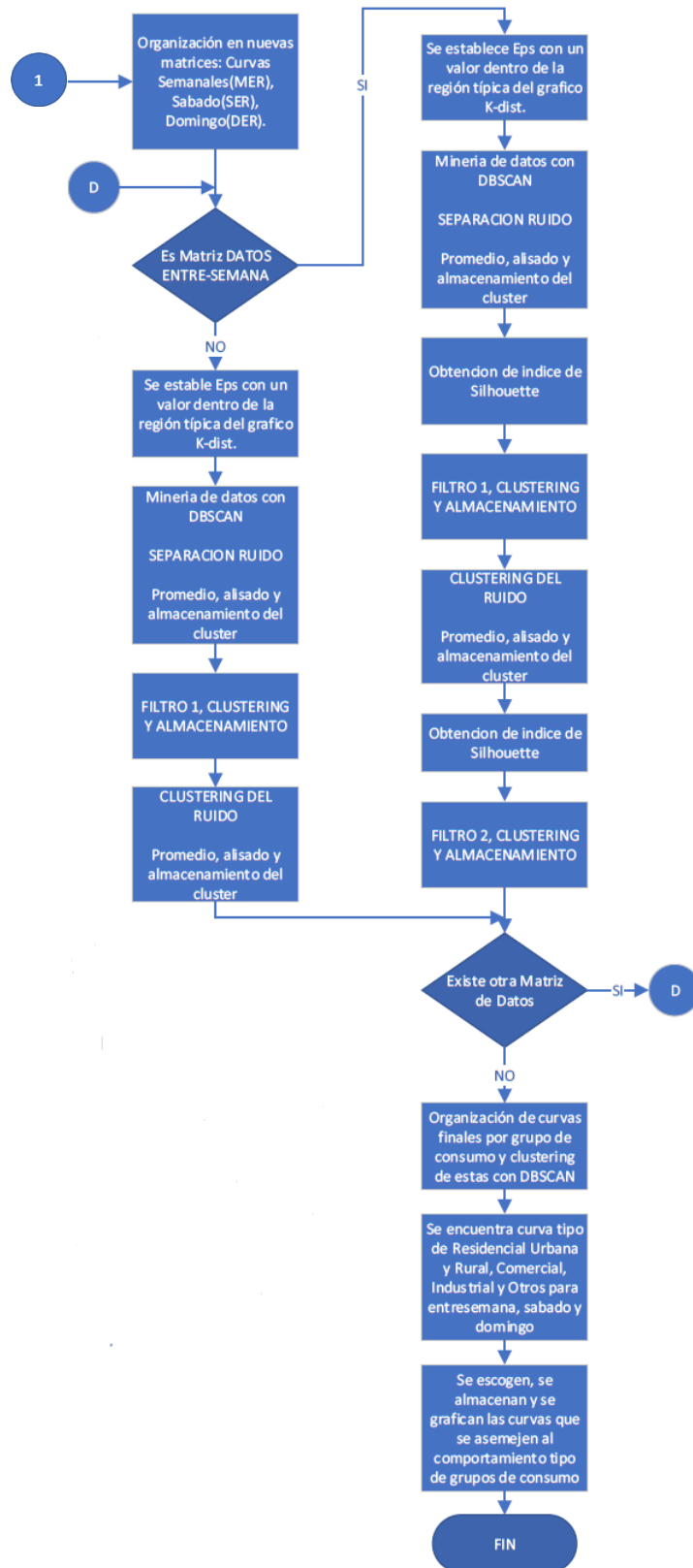


Figura 2.3. Diagrama de Flujo para la validación de resultados finales

2.5.4 DIAGRAMA DE FLUJO TOTAL EMPLEADO EN MATLAB

La Figura 2.4. se analiza de forma vertical y es la unión de los 3 diagramas de flujo explicados en los puntos anteriores. Se comienza por el lado izquierdo del diagrama que es flujo de ordenamiento resultando en la obtención de curvas características de entresemana y fines de semana por medidor para luego desarrollar la parte de clustering global del algoritmo observado hacia el lado derecho, aplicando filtros de ruido y así obtener más agrupamientos que finalmente se validan con el índice de comprobación. Este diagrama contiene todos los flujogramas estudiados y termina en una matriz de resultados en Matlab que luego son presentados en este trabajo de titulación con un formato de tabla adecuado.

Se hace hincapié en los conectores ya que ayudan visualmente al entendimiento del diagrama de flujo total. Comúnmente, se parte desde el bloque condicional o de iteración y con la ayuda de estos círculos se logra el salto a otra sección del flujograma total.

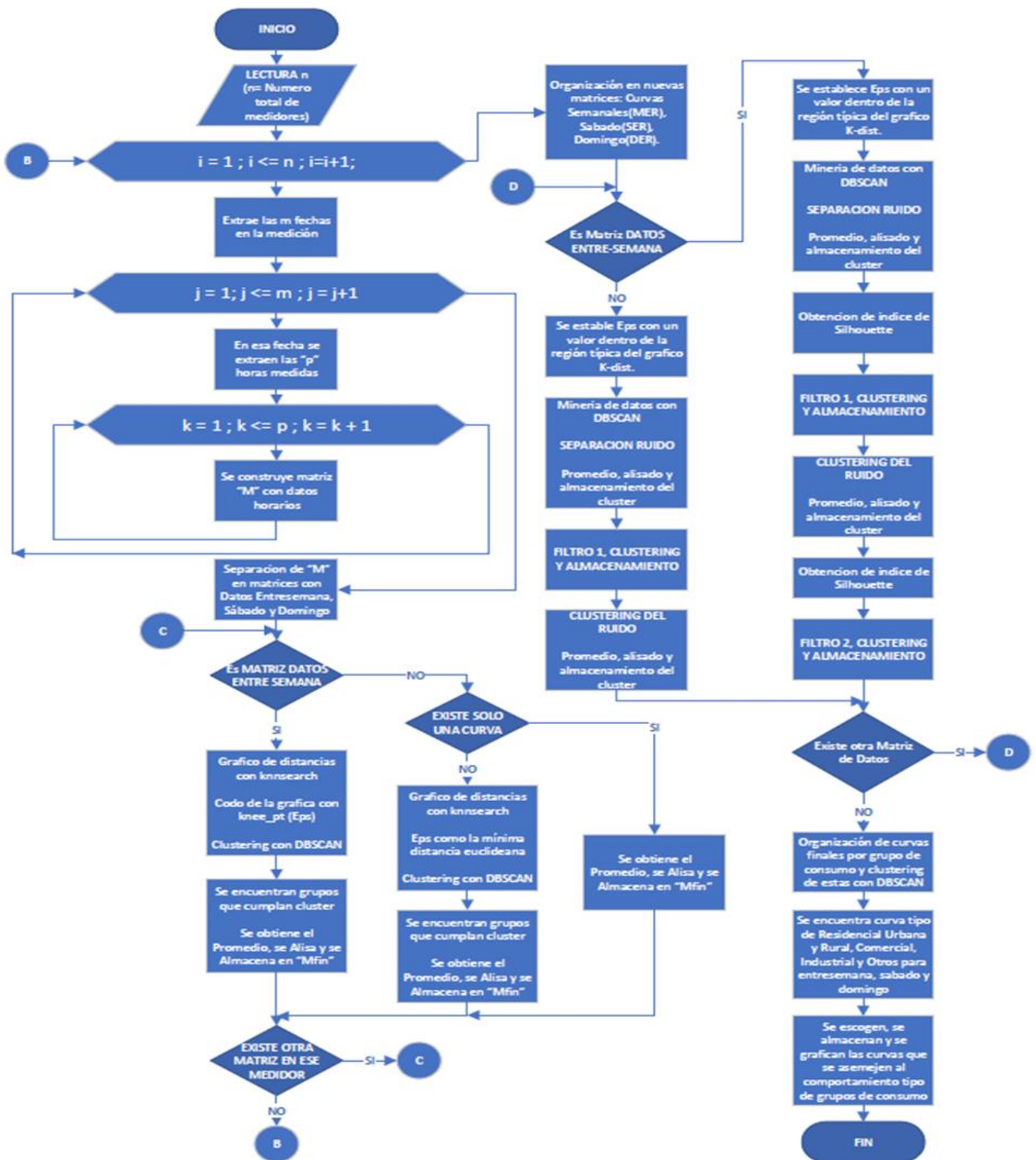


Figura 2.4. Diagrama de flujo total del ordenamiento y clustering

3. RESULTADOS Y DISCUSIÓN

En el punto 2.2.1 se habló de las etiquetas utilizadas en la base de mediciones, estas son importantes también para mostrar los resultados del análisis del algoritmo DBSCAN junto con la clasificación por tipo de consumidor que se requiere para días entre semana y fines de semana. Una etiqueta es llamada “ClaveGCon” y es la utilizada para diferenciar el usuario residencial urbano y rural, comercial, industrial y otros el momento de presentar los resultados gráficos y numéricos. En la Tabla 3.1 se muestra el significado de la nomenclatura utilizada para la diferenciación de estos usuarios.

Tabla 3.1. Descripción del tipo de consumidor con numeración

TIPO CONSUMIDOR	NUMERACIÓN
RESIDENCIAL	10
COMERCIAL	20
INDUSTRIAL	30
OTROS	40

3.1 ANÁLISIS ENTRE SEMANA

Es necesario hacer una relación de los flujogramas del punto 2.5 con la presentación de los resultados gráficos para días entre semana. Es decir, se presentan las figuras que describen cada parte del diagrama de flujo total de ordenamiento y clustering aplicado en Matlab y que utiliza el algoritmo de agrupamiento DBSCAN junto a otras funciones complementarias que determinan los elementos de entrada para este algoritmo matemático.

Los elementos de entrada para la función son: Eps y Mpts. Para los Mpts se optó en fijar el valor en un número entero mayor que uno ya que eso es lo que recomienda la ayuda en Matlab y es mencionado también en el punto 1.3.3.2. Para encontrar Eps se realiza un proceso más elaborado. Se utilizan los algoritmos complementarios explicados anteriormente que son: Knnsearch y Knee_pt.

El fundamento matemático ya es conocido y Matlab aplica las definiciones para ofrecer una mayor facilidad en la obtención de los datos de entrada al algoritmo. Para el análisis de entre semana, se debe encontrar la épsilon “Eps” aplicando las recomendaciones ya provistas. Es decir, se realiza un gráfico tipo K-dist como el observado en la Figura 3.1. con el respectivo codo de la función denotado en color rojo y teniendo en cuenta que cualquier valor hasta ese codo representa un valor correcto para su uso como elemento de entrada en DBSCAN.

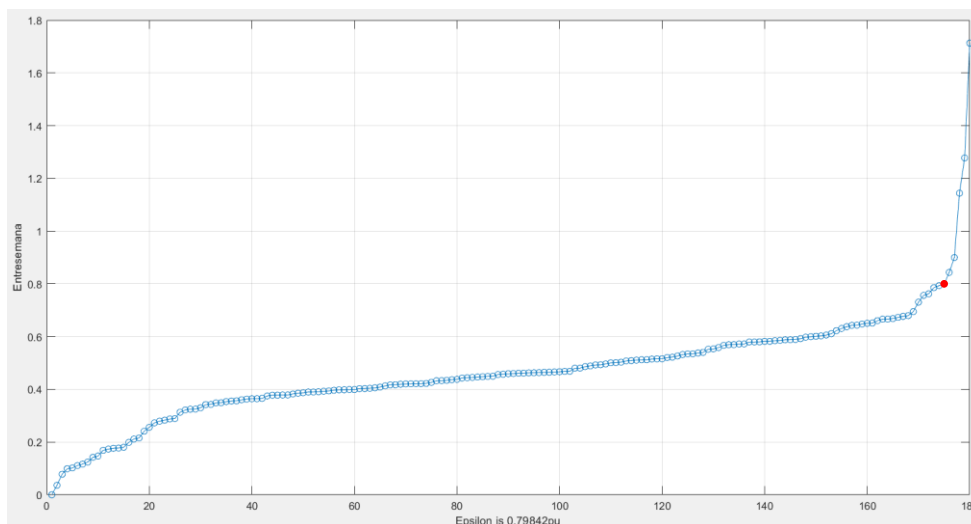


Figura 3.1. Codo de la función obtenido utilizando algoritmos complementarios

Una vez obtenido el codo de la función para establecer Eps y fijado el Mpts a un valor mayor a 1 se comienzan a obtener las gráficas como la de la Figura 3.2. que contienen descripciones en el lado inferior, superior y en los lados laterales. Esta explicación del formato de la imagen será la misma para todos los grupos resultantes cuando se aplica el DBSCAN en Matlab.

Del lado derecho se puede observar el número de datos de medidores con los que se obtuvo la curva característica, es decir 13 datos rojos, en valores por unidad y la curva resultante en amarillo para este tipo de consumidor. El texto en la parte inferior de la imagen hace referencia a la Tabla 3.1. denominando en este caso el grupo de consumo residencial “10” seguido de la información complementaria de zona rural o urbana que se obtuvo de un catastro de usuarios de la EEA. Para el lado izquierdo, se tendrá escrita la denominación entre semana, sábado o domingo.

Esto se realizó con el afán de realizar la separación de días entre semana y fines de semana para los resultados totales. Finalmente, en la parte superior se tiene el número de clúster que para este caso de resultados entre semana se tienen 8 agrupamientos.

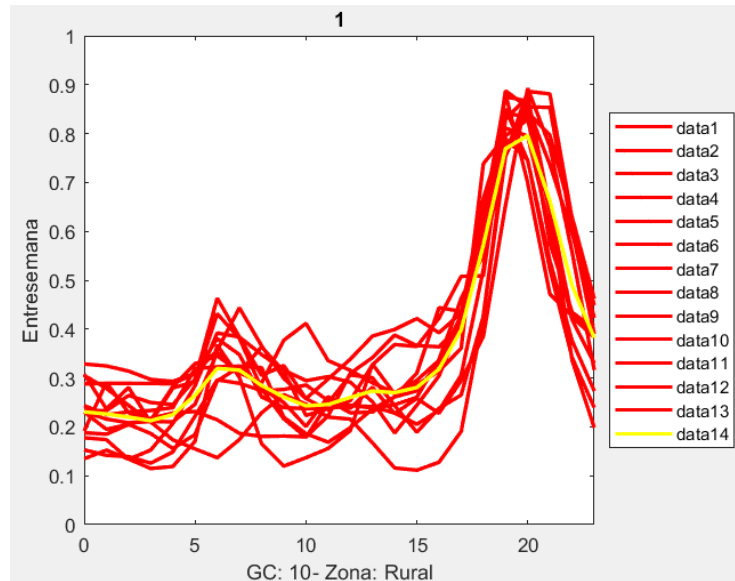


Figura 3.2. Clúster 1 de entresemana residencial en zona rural

Luego de observar la curva característica del agrupamiento 1 se analiza ahora el agrupamiento 3 mostrado en la Figura 3.3. donde solamente se ha hecho el agrupamiento con 7 curvas parecidas denotándolas con color verde y muestra la resultante en color amarillo. Sigue siendo un tipo de consumidor residencial “10” pero la zona es urbana en este caso.

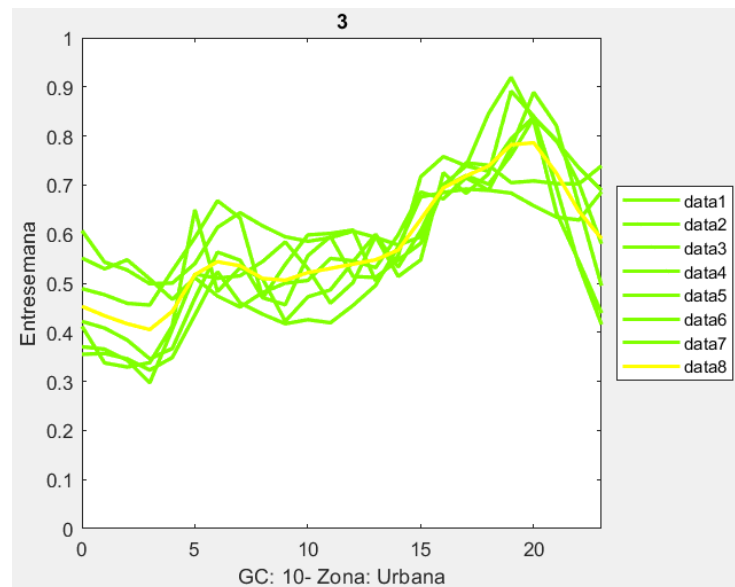


Figura 3.3. Clúster 3 de entresemana residencial urbana

Ahora, se analiza el agrupamiento 6 en la Figura 3.4. que tiene una peculiar forma de curva donde se observa un consumo mayor en horas de la noche y madrugada mientras que en la tarde se tiene un decrecimiento significativo en el valor por unidad de su potencia activa. Para este caso, se realizó el agrupamiento de 3 medidores similares mostrados en azul y su curva resultante de amarillo. Es un tipo de consumidor catalogado en otros y en una zona rural.

De forma empírica y por experiencia en campañas de medición en bajo y medio voltaje esto se asimila a un tipo Industrial con consumos nocturnos para aprovechar la tarifa especial por kWh o también a un consumo de alumbrado público donde se tienen lámparas defectuosas prendidas también en horarios donde se tiene buena iluminación.

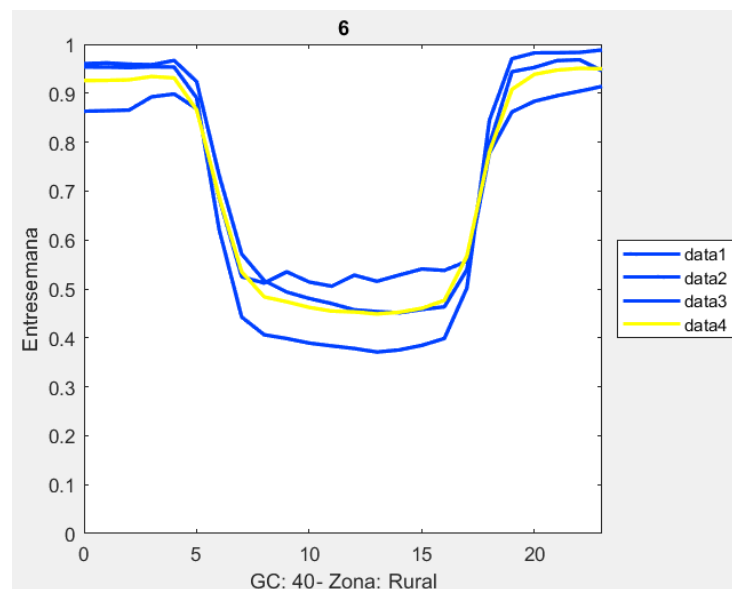


Figura 3.4. Clúster 6 de entresemana otros rural

Seguidamente, se muestra la curva resultante del agrupamiento 7 en la Figura 3.5. esta denota un comportamiento donde se tienen valores bajos en por unidad para horarios nocturnos mientras que en horarios matutinos y vespertinos se obtienen valores mayores que se asemejan a un consumo comercial, pero, revisando el catastro se encuentra en un estrato de Otros. Es importante mencionar que esta empresa eléctrica tiene catalogado a otros según normativa ecuatoriana que podrían ser Comercial con demanda horaria, Comercial Sin demanda, antenas, alumbrado público, etc. Es así como se busca el clustering de estas curvas para mostrar gráficas con un comportamiento parecido a las curvas observadas en [18].

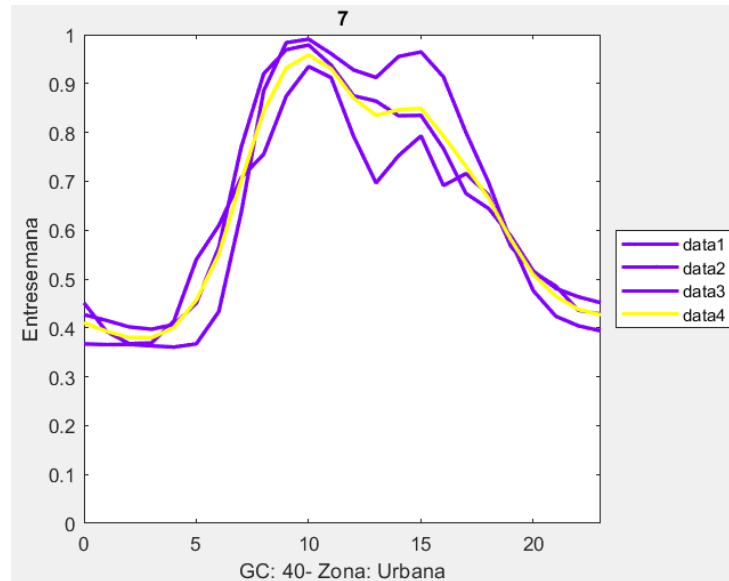


Figura 3.5. Clúster 7 de entresemana otros urbano

Es así como se llega a la Figura 3.6. que es el agrupamiento de medidores considerados Ruido. Dada la cantidad de datos en esta figura, no se distingue un comportamiento parecido en las curvas características y se observa también una curva resultante de color amarillo que no es del todo veraz para este agrupamiento.

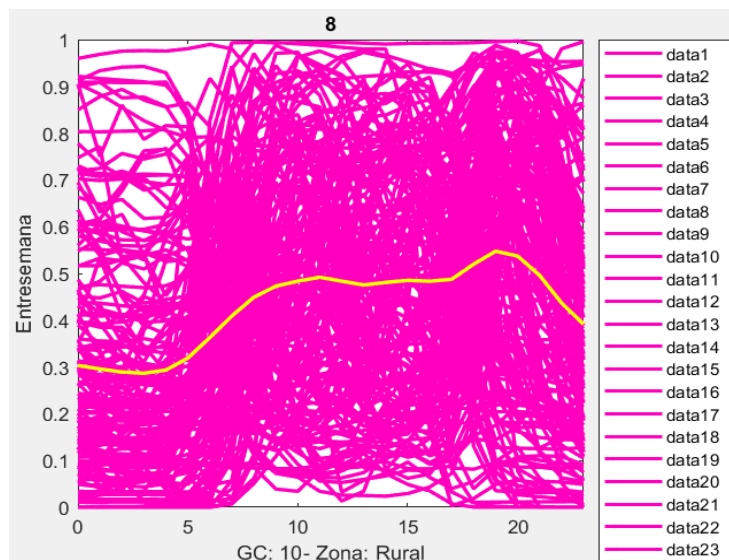


Figura 3.6. Clúster 8 de entresemana RUIDO

La detección del clúster que tiene Ruido se la puede hacer visualmente como se hizo con la figura anterior; pero, es necesario establecer una forma matemática que valide el agrupamiento realizado. Para esta tarea, se utiliza el IS mostrado en la Figura 3.7. donde, matemáticamente, se detecta cuál clúster esta agrupado incorrectamente utilizando los criterios explicados en el punto 2.4 de pertenencia o no al clúster y así aplicar el primer filtro de ruido entresemana.

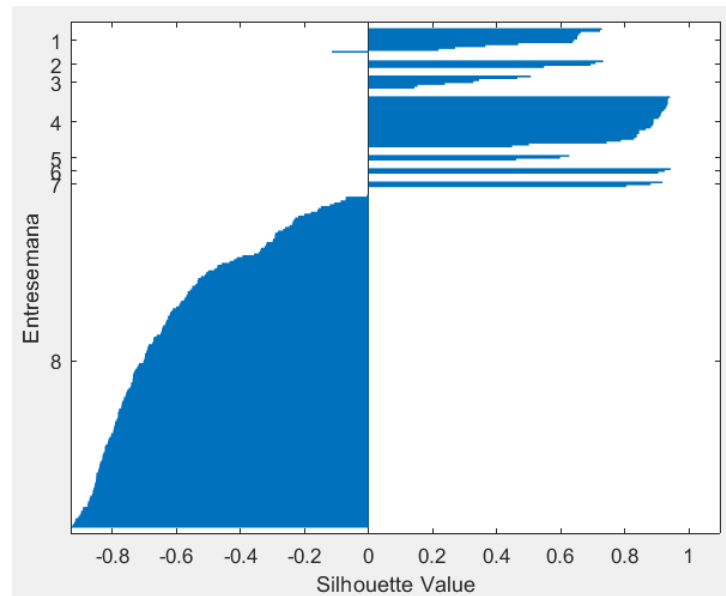


Figura 3.7. Índice de validación SILHOUETTE, -1 no pertenece al clúster, 1 pertenece

Obteniendo el codo de la función en la Figura 3.8. para el primer filtro de ruido “R”:

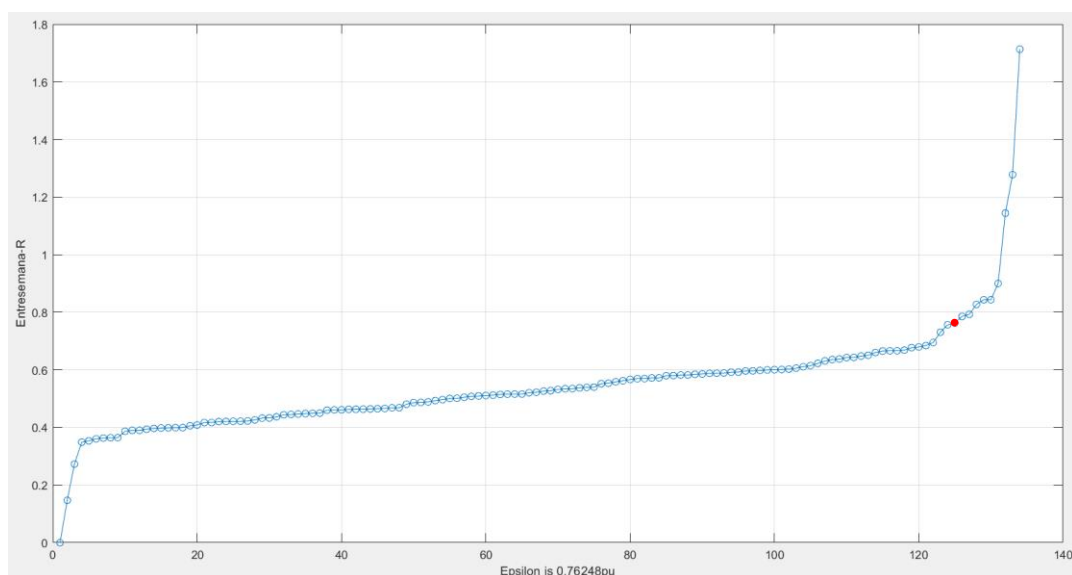


Figura 3.8. Codo de la función encontrado utilizando algoritmos complementarios

Ya obtenido el argumento de la función Eps con el gráfico tipo K-dist y establecido el valor de Mpts se encuentran las curvas características por tipo de consumidor para el primer filtro de ruido. El resultado son 8 grupos que luego de la validación con el índice de silhouette se observa una incorrecta agrupación para el grupo 0 dando lugar a una segunda etapa de Ruido del Filtro "RR" y así obtener más curvas características.

En la Figura 3.9. se observa una curva resultante obtenido del agrupamiento de 7 gráficas de medidores. Esta curva es el resultado del primer filtro de ruido "R" para la obtención de más curvas por tipo de consumidor.

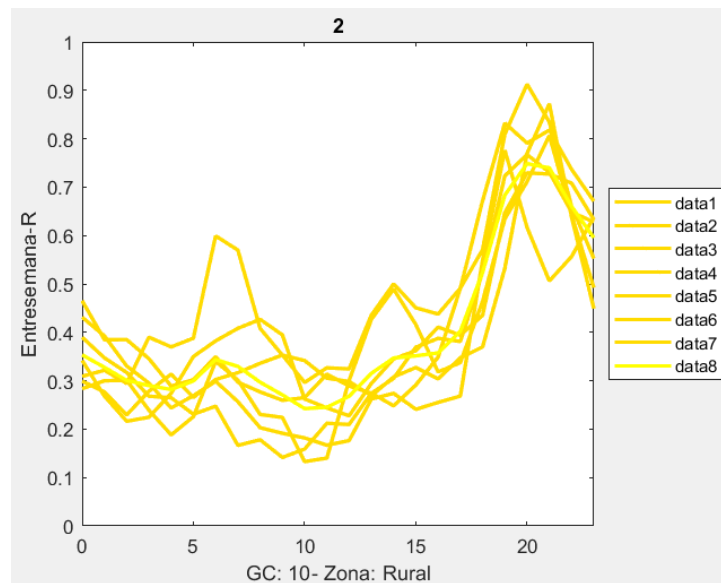


Figura 3.9. Filtro 1 Ruido, Clúster 2 de entresemana Residencial Rural

La Figura 3.10. muestra otro agrupamiento y su curva resultante en amarillo. El tipo de consumidor es Industrial con valores en por unidad PU pequeños para el horario nocturno mientras que desde las 6 am comienza a haber un incremento en el valor por unidad de la potencia activa. Luego, a las 13 horas existe una disminución en este valor de potencia en PU mostrando la típica situación de hora de almuerzo para los trabajadores de las industrias o comercios que son considerados como tipo industrial debido a su alto consumo energético.

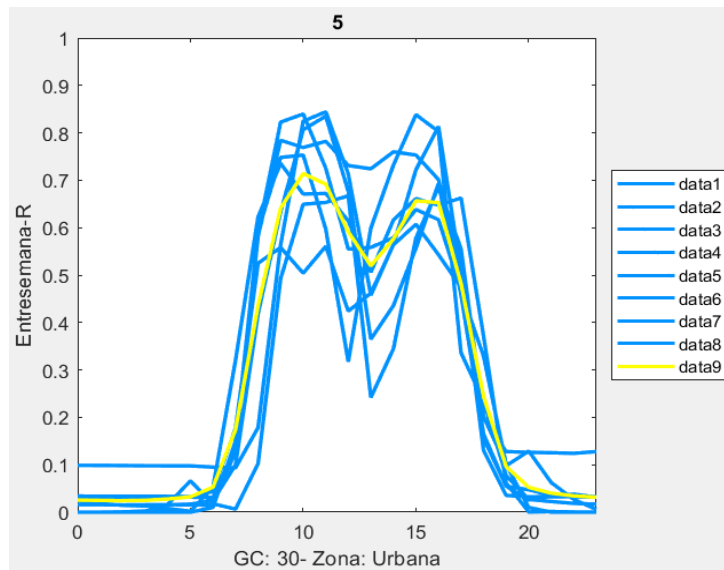


Figura 3.10. Filtro 1 Ruido, Clúster 5 de entresemana Industrial Urbana

Se viene observando que el primer filtro de ruido “R” obtuvo curvas resultantes para el tipo de consumidor residencial e industrial. Ahora se muestra, también, una curva resultante para un tipo Comercial agrupando 4 medidores y obteniendo la curva amarilla de la Figura 3.11.

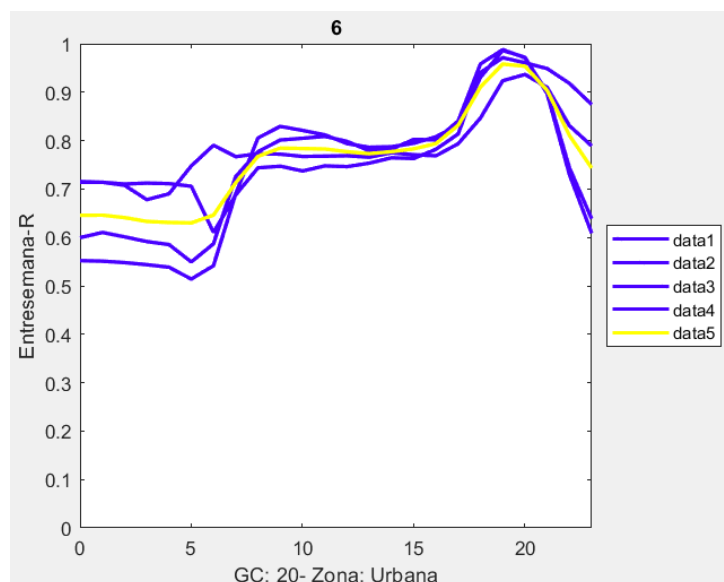


Figura 3.11. Filtro 1 Ruido, Clúster 6 de entresemana Comercial Urbana

En la Figura 3.12. se observa la validación de los agrupamientos utilizando el índice de Silhouette. Este índice muestra la pertenencia o no pertenencia al clúster, para el primer filtro de ruido se obtuvieron 8 grupos de los cuales el grupo 0 denota la no pertenencia al clúster porque sus valores están hacia el -1.

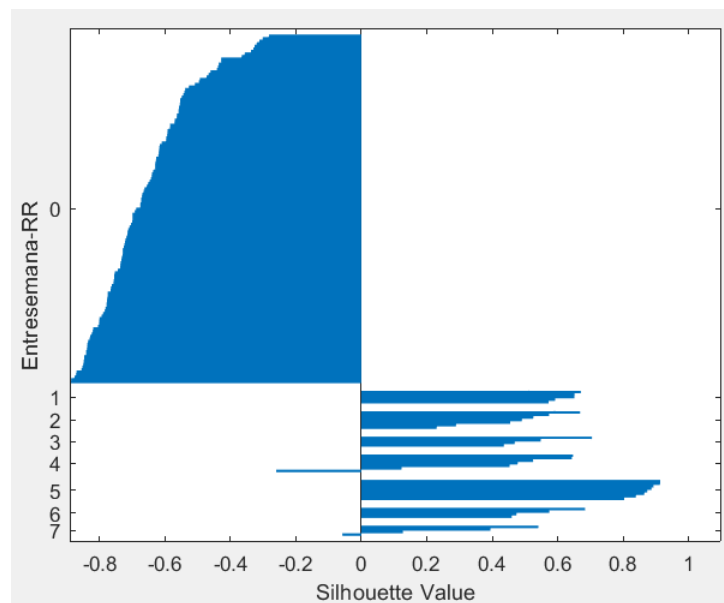


Figura 3.12. Validación SILHOUETTE “R”, -1 no pertenece al clúster, 1 pertenece

Una vez validado los agrupamientos se incluye un segundo filtro del ruido “RR” con la misma finalidad, que es, obtener más curvas tipo para que el clúster final pueda mostrar valores resultantes que no sean atípicos y que muestren cierta relación con su tipo de consumidor y con las gráficas de la bibliografía utilizada.

La Figura 3.13. muestra las curvas del agrupamiento 2 obtenido del segundo filtro de ruido. Son dos medidores que hacen un clúster, se obtiene la curva resultante en amarillo y es un tipo de consumidor residencial en zona rural. Este segundo filtro de ruido es explicado en el punto 2.5.3 donde se tiene una sección del diagrama de flujo de ordenamiento y clustering del algoritmo implementado en MATLAB que muestra en la parte lateral derecha los pasos del clustering, validación y las dos etapas de filtros para datos entresemana.

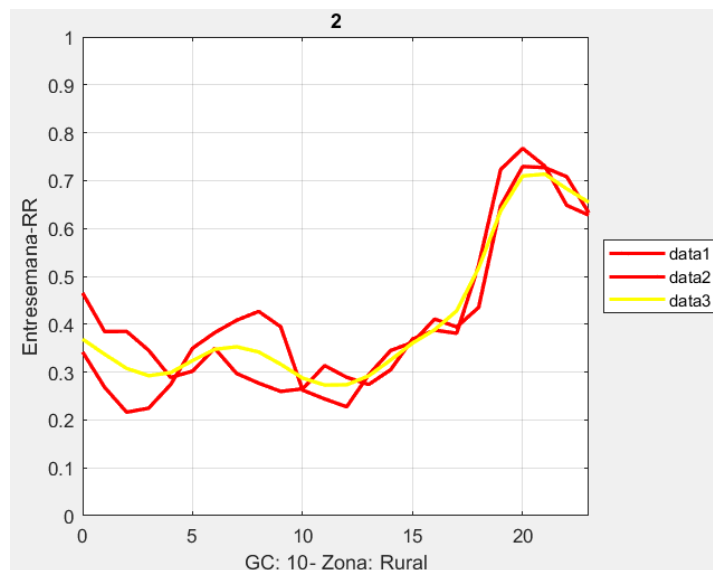


Figura 3.13. Filtro 2 Ruido, Clúster 2 de entresemana Residencial Rural

3.2 ANÁLISIS FIN DE SEMANA

En este análisis se tendrá la diferenciación de curvas características de sábado y domingo. Se realiza así previamente para descartar mediciones atípicas o de ruido y solo tener curvas con la forma correcta de acuerdo con el tipo de consumidor y al comportamiento de fin de semana. Luego, se hace el agrupamiento de los resultados parciales de sábado y domingo para obtener la curva característica de fin de semana.

Dado que existen pocas mediciones para sábados y domingos se realiza el clustering obteniendo diferentes agrupamientos que luego de validar con IS el clustering se determina el uso de un filtro de ruido solamente para obtener más curvas características de fin de semana. Estos pasos se muestran resumidos en el punto 2.5.2, esto debe analizarse porque se encontraron usuarios con solo una medición de fin de semana, resultando esa única como la curva característica.

Utilizando el mismo criterio para la obtención de curvas tipo de entresemana se comienza con el gráfico tipo K-dist de distancias euclidianas como el de la Figura 3.14. para obtener el valor de Eps. Se fija el valor de Mpts y una vez listos los argumentos de entrada para el DBSCAN se procede con la obtención de los diferentes agrupamientos para sábados y domingos.

3.2.1 SÁBADO

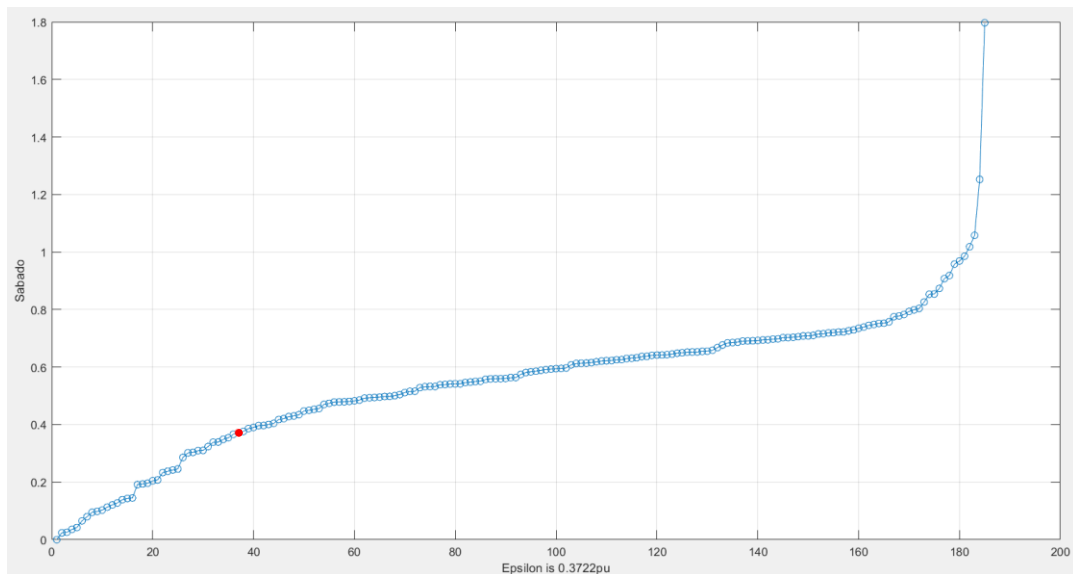


Figura 3.14. Valor de codo de la función utilizando los algoritmos complementarios

En la Figura 3.15. se observa la curva resultante en amarillo obtenido de 3 datos de medidores, esta gráfica es para un usuario residencial zona urbana pero que tiene un comportamiento diferente a los picos residenciales característicos obtenidos en las anteriores figuras de entresemana.

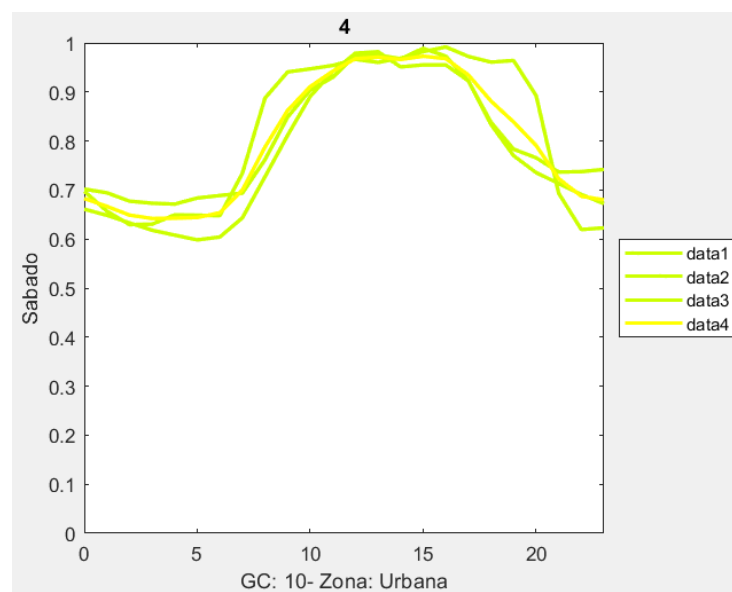


Figura 3.15. Clúster 4 de sábado Residencial Urbano

La curva de la Figura 3.16. se observó anteriormente, teniendo un comportamiento industrial o de alumbrado público con luminarias en mal estado. Esta gráfica resultante en amarillo representa un tipo de consumidor Otros, el sábado en zona Rural.

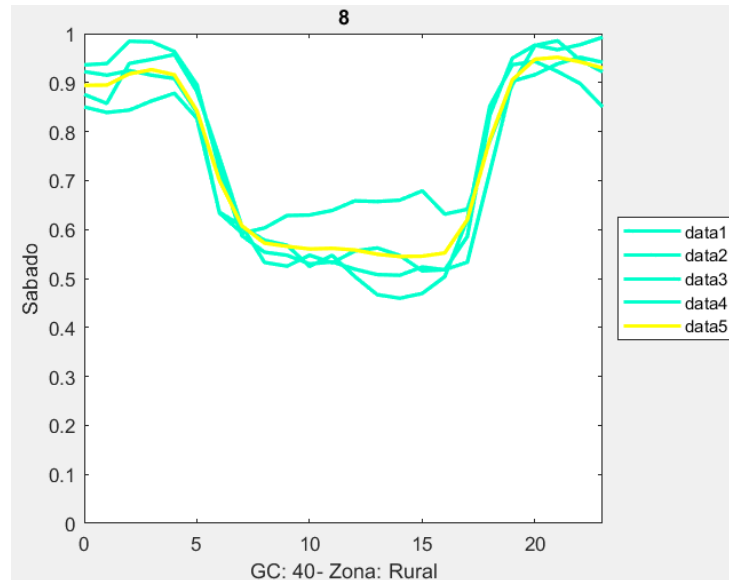


Figura 3.16. Clúster 8 de sábado Otros Rural

El agrupamiento 15 mostrado en la Figura 3.17. tiene un conjunto de curvas en rojo donde visualmente se puede identificar un mal agrupamiento de datos tal que la curva resultante en amarillo podría tener valores incorrectos. Lo interesante es que esta gráfica se asemeja a la final para la zona residencial fin de semana, teniendo un comportamiento de picos en las horas características para horas de usuarios residenciales: 7 am y 8 pm. Pero al no mostrar una relación como otros agrupamientos se realiza la comprobación de estos grupos con el índice de Silhouette mostrado en la Figura 3.18. notando que el agrupamiento 15 no es el adecuado porque todas sus curvas muestran una comprobación matemática hacia -1.

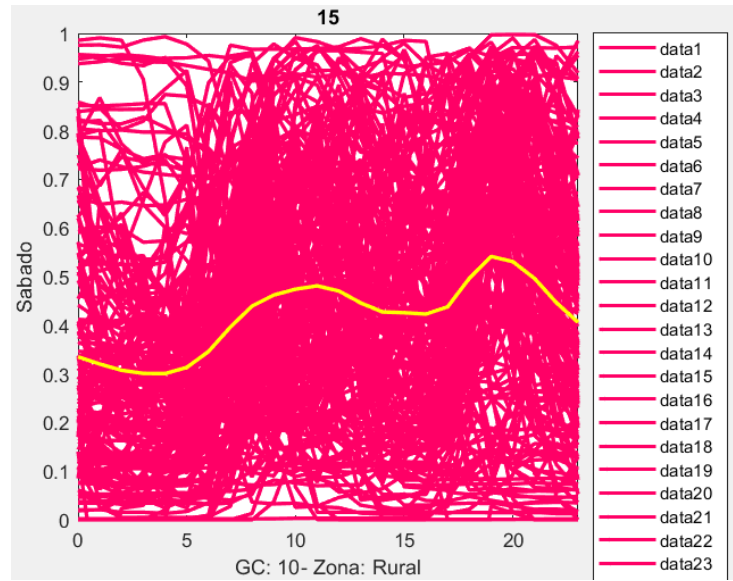


Figura 3.17. Clúster 15 de sábado considerado RUIDO

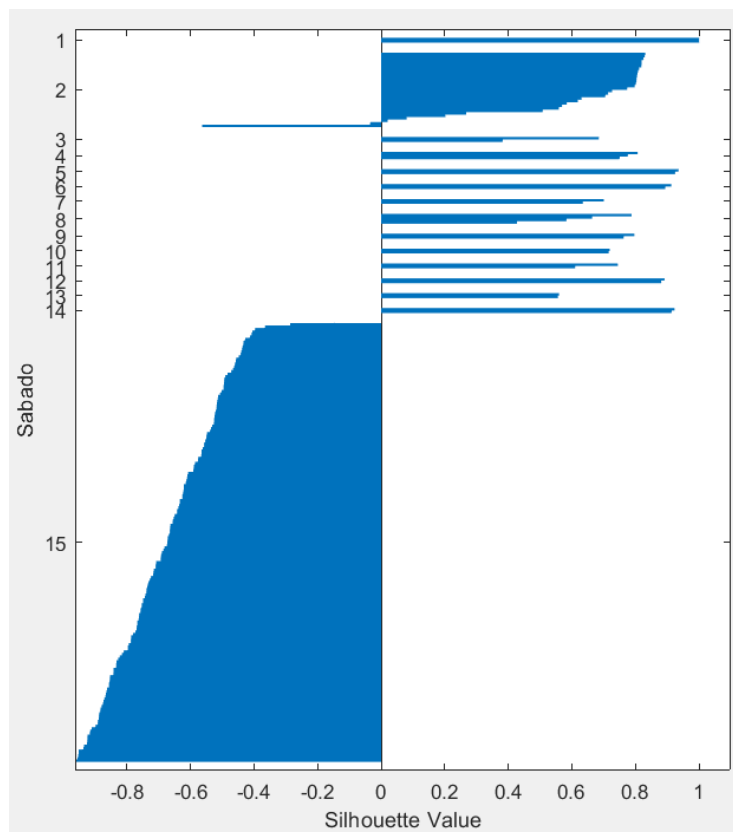


Figura 3.18. Validación con SILHOUETTE, -1 no pertenece al clúster, 1 pertenece

Se encuentra el grafico tipo K-dist de distancias euclidianas como el de la Figura 3.19. y se obtiene el codo de la función para sábado primer filtro del ruido "R".

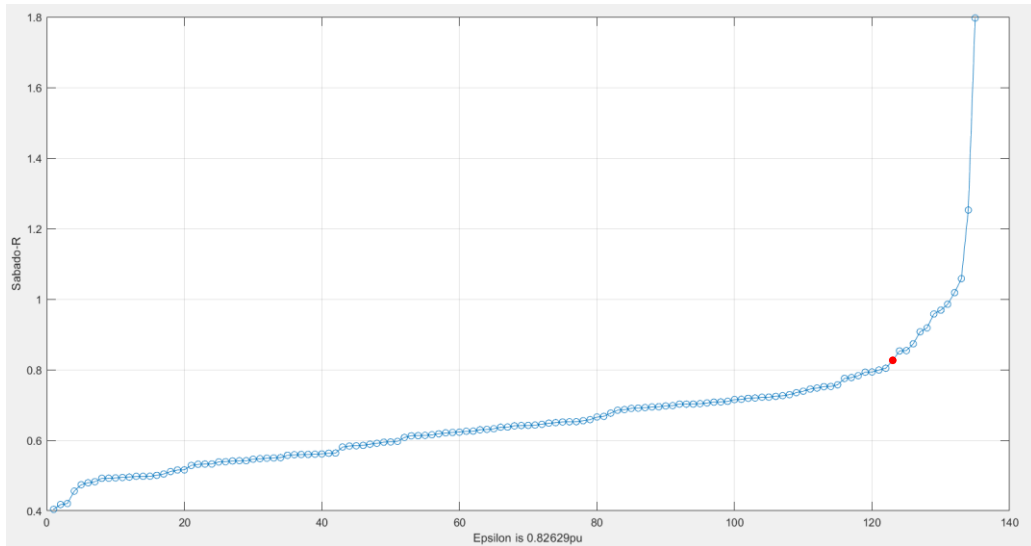


Figura 3.19. Codo numérico de la función obtenido con algoritmos complementarios

En la Figura 3.20. se encuentra la curva resultante en amarillo para el primer filtro del ruido "R" en el algoritmo DBSCAN en sábado otros, zona Urbana.

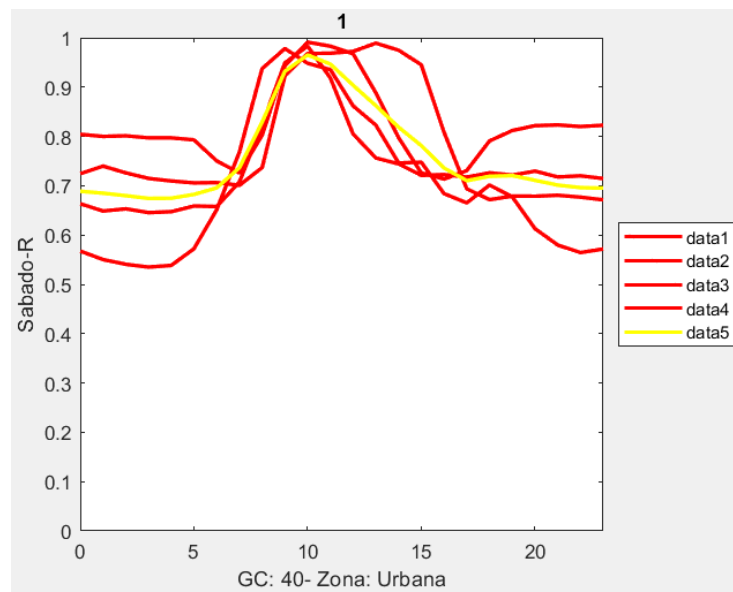


Figura 3.20. Filtro 1 Ruido, Clúster 1 de sábado Otros Urbano

3.2.2 DOMINGO

El análisis de domingo es semejante al aplicado para el sábado, se comienza con un gráfico tipo K-dist como el de la Figura 3.21. y se encuentra el codo de la función. Este valor sería el Eps de elemento de entrada para el algoritmo basado en densidades DBSCAN, luego se valida el clúster para finalmente obtener más agrupamientos con la ayuda del filtro de ruido "R".

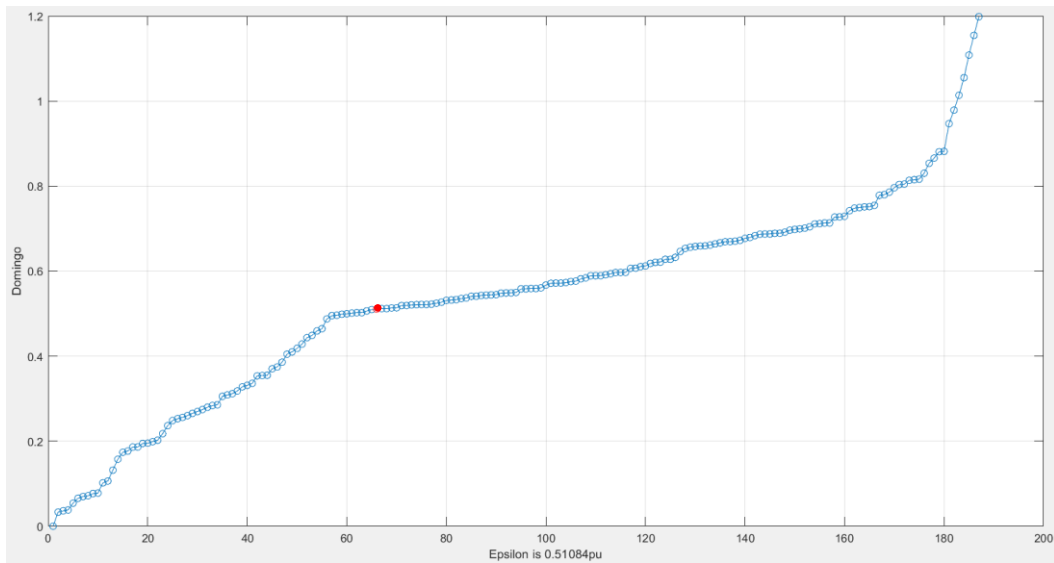


Figura 3.21. Valor de codo de la función obtenido con los algoritmos complementarios

La Figura 3.22. muestra la curva resultante en amarillo del clúster de 5 datos diferentes.

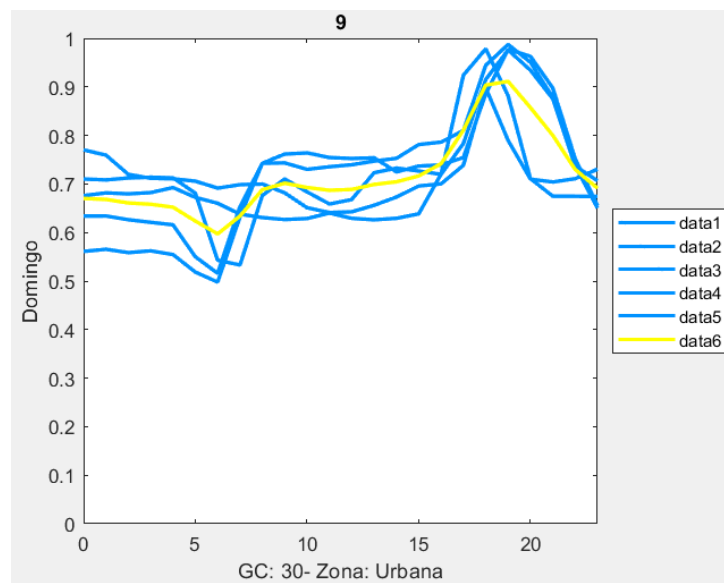


Figura 3.22. Clúster 9 de domingo Industrial Urbano

Se conoce que la curva de demanda muestra el comportamiento de consumo energético de los usuarios, y para un domingo el comportamiento religioso introduce curvas para este día donde se tienen los valores más altos de potencia en por unidad en el horario de 7:00-11:00. Este periodo normalmente es utilizado por devotos de la iglesia para asistir a las respectivas ceremonias y también el comercio alrededor se activa paralelamente, es así como la curva resultante viene de 7 datos de medidores y se aprecia en la Figura 3.23.

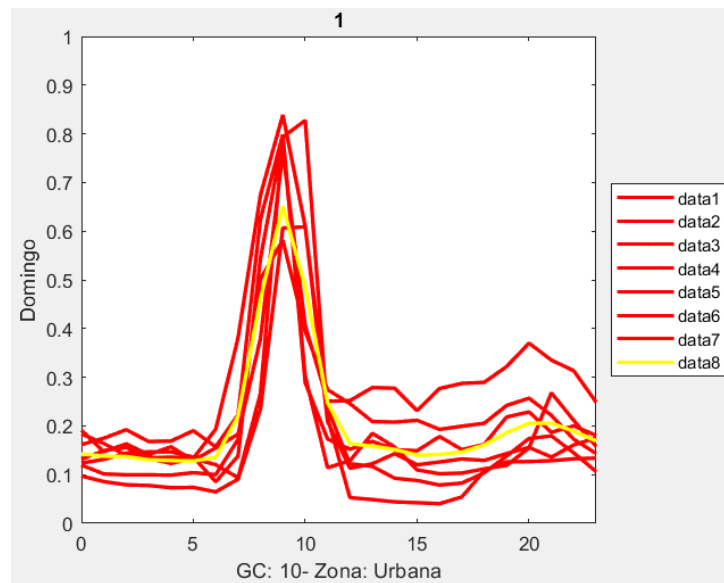


Figura 3.23. Clúster 1 de domingo Residencial Urbano

En la Figura 3.24. se tiene nuevamente el agrupamiento del ruido, este se lo identifica visualmente por la cantidad de curvas que no muestran un comportamiento parecido sino más bien se agrupan incorrectamente debido a que son mediciones atípicas o mediciones incorrectas por mala manipulación del equipo.

La curva resultante en amarillo vuelve a mostrar un comportamiento residencial donde se tienen picos en las horas características de 7:0- y 20:00. Como ya se sabe, el índice de Silhouette es el apropiado para la identificación matemática de un clúster mal realizado, la Figura 3.25. muestra este resultado y nos informa el mal agrupamiento del clúster 14 para así proceder a la etapa de filtrado y obtención de la mayor cantidad de curvas por tipo de consumidor.

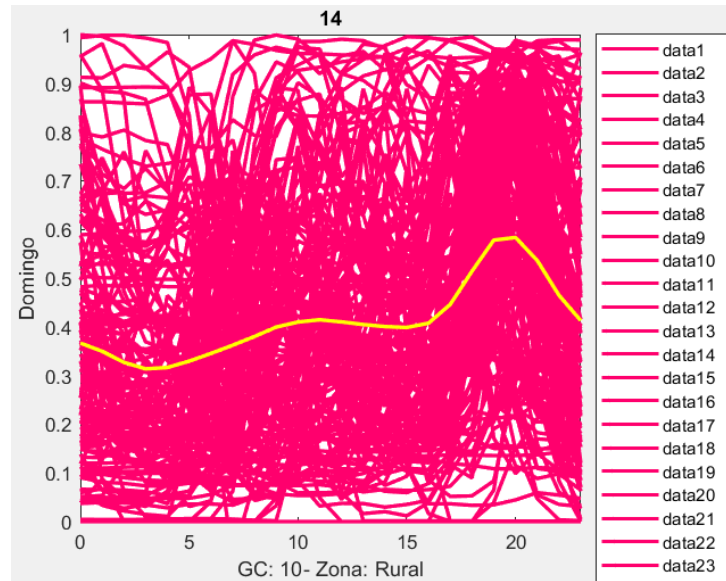


Figura 3.24. Clúster 14 de domingo considerado RUIDO

Índice de comprobación de Silhouette para domingo en la Figura 3.25. que muestra un incorrecto agrupamiento del clúster 14.

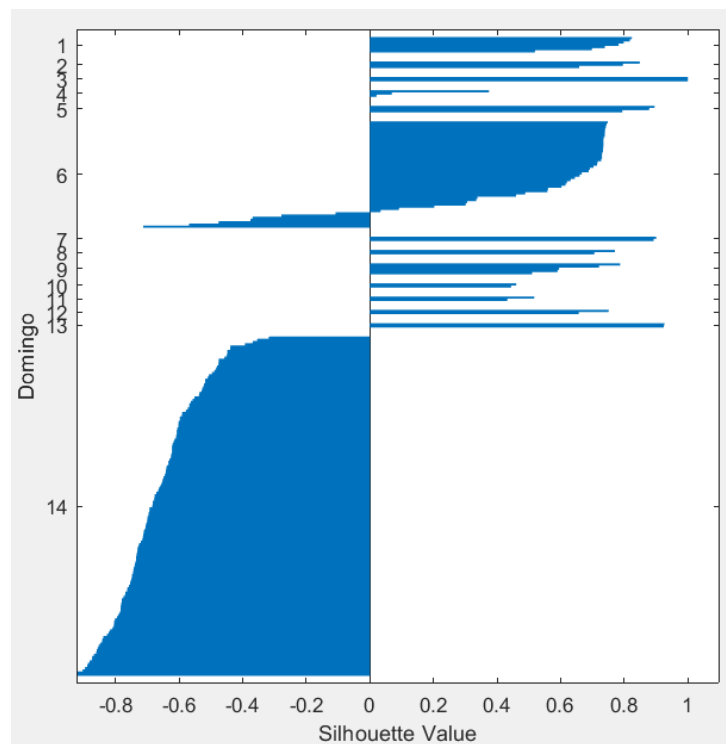


Figura 3.25. Índice de SILHOUETTE, -1 no pertenece al clúster, 1 pertenece

Se aplica el filtro de ruido "R" para domingo, en esta se vuelve a utilizar el algoritmo de agrupamiento DBSCAN, por tanto, se encuentran los valores de entradas de la función. La Figura 3.26. muestra el grafico tipo K-dist para encontrar Eps.

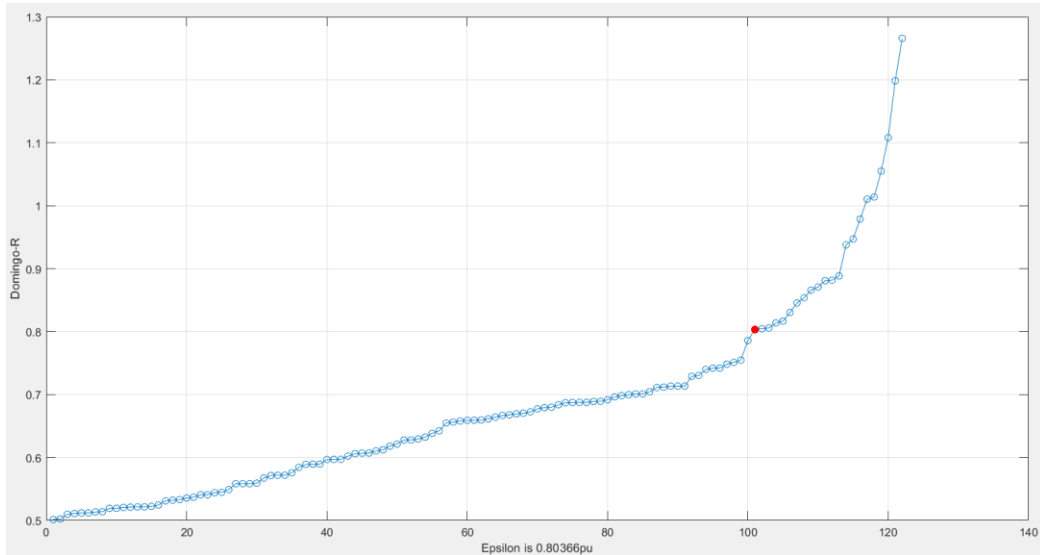


Figura 3.26. Valor codo de la función obtenido con algoritmos complementarios

En la Figura 3.27. se muestra el agrupamiento 1 obtenido de un filtro de ruido con una curva resultante amarilla que viene del agrupamiento de 22 datos de medidores.

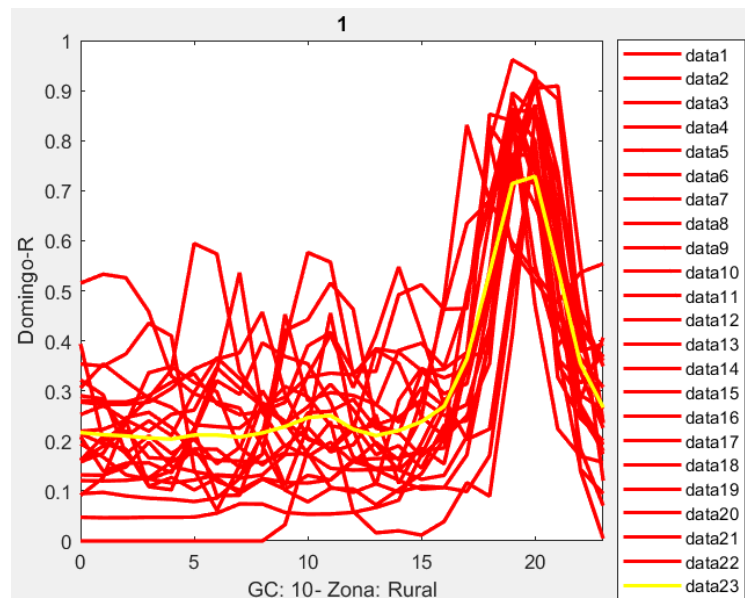


Figura 3.27. Filtro 1 Ruido, Clúster 1 de domingo Residencial Rural

La Figura 3.28. contiene una curva resultante en color amarillo de la agrupación de 3 datos para un grupo de consumo comercial en zona urbana para el filtro de ruido en domingo.

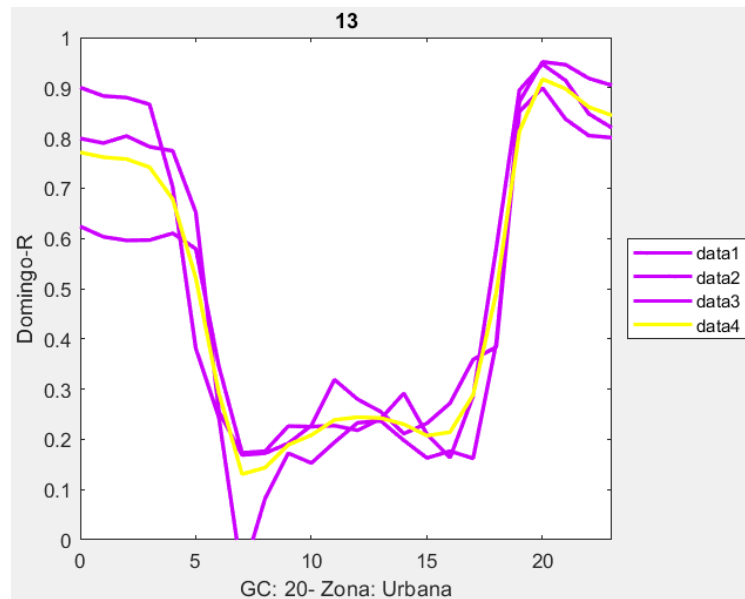


Figura 3.28. Filtro 1 Ruido, Clúster 13 de domingo Comercial Urbana

El agrupamiento número 8 sólo tuvo 2 datos parecidos y la curva resultante se muestra en la Figura 3.29. para un tipo de consumidor Industrial en zona urbana obtenido en un primer filtro de domingo.

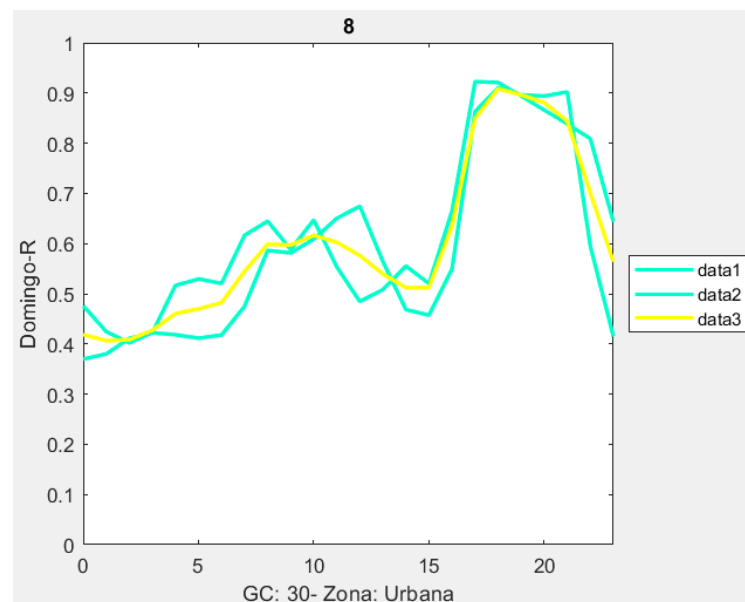


Figura 3.29. Filtro 1 Ruido, Clúster 8 de domingo Industrial Urbana

3.3 RESULTADOS FINALES DEL SISTEMA

Luego de someter la base de datos al algoritmo propuesto en Matlab se encontraron curvas tipo de entresemana y fin de semana. Un gran repertorio de curvas se puede observar al ejecutar el programa creado para resolver este problema de obtención de curvas de demanda por tipo de consumidor en pu. Todas ellas pasaron la fase de validación pero ejemplifican comportamientos que no son característicos del sector o del tipo de consumidor. Esto se debe a que no hubo una desagregación por alimentador o considerando la georreferenciación.

El procedimiento gráfico ya se analizó en los puntos 3.1 y 3.2., pero es necesario establecer solo una curva por tipo de consumidor. Para esto, se utiliza la referencia [18]. En la Figura 3.31 se encuentran curvas características de los diferentes grupos de consumo y una curva total del sistema. Esto se consideró en el estudio de curvas de demanda del presente trabajo de titulación y los resultados gráficos se pueden observar en la Figura 3.30.

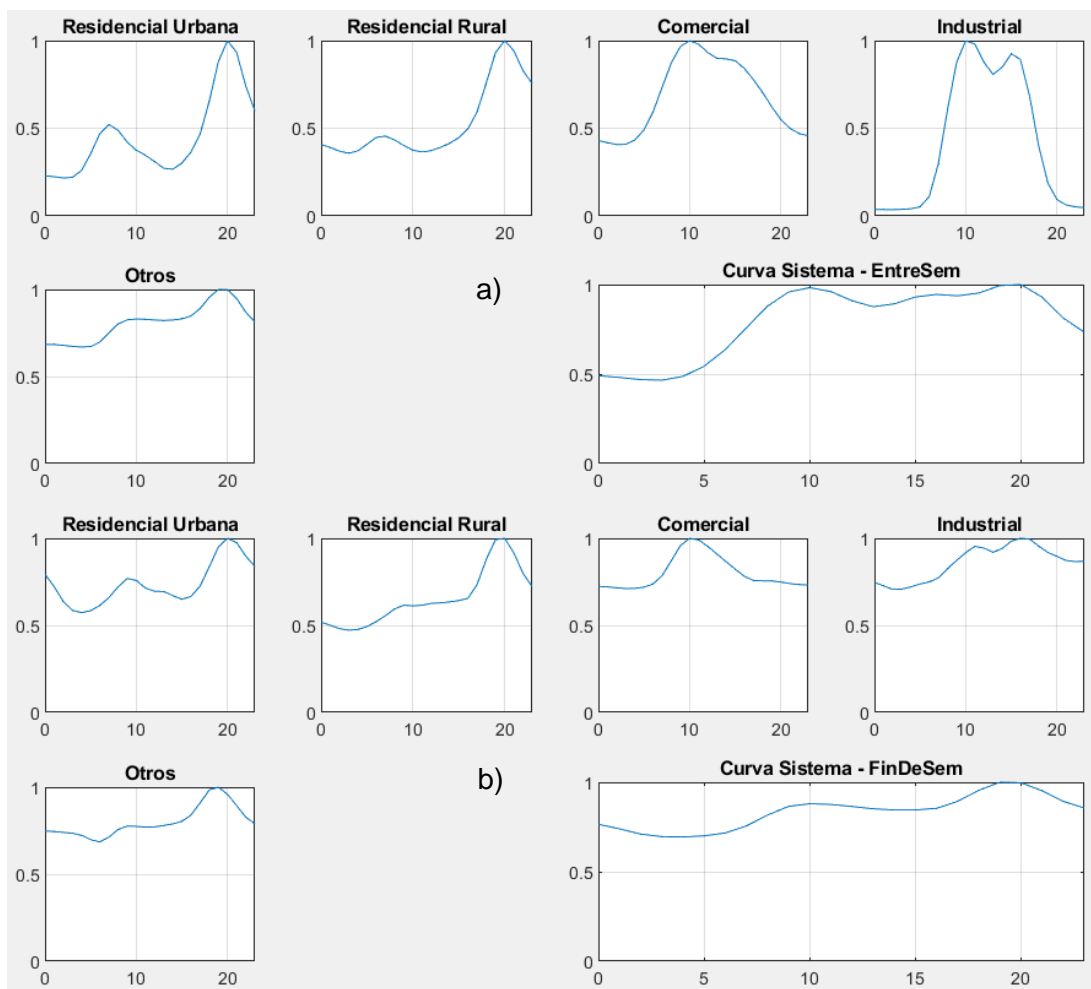


Figura 3.30. Gráficas en por unidad PU de curvas típicas de la EEA. a) Curvas de días entresemana. b) Curvas de fin de semana

Los resultados del sistema eléctrico de la EEA muestran comportamientos característicos de los 4 tipos de consumidores: Residencial Urbano y Rural, Comercial, Industrial y otros para días entresemana. Por tanto, es necesario presentar los resultados numéricos de las curvas. Se seleccionó un formato de cuatro decimales para la Tabla 3.2. y se obtuvieron valores de 1 por unidad para representar la hora en que se tiene la demanda máxima por tipo de consumidor y también la hora de la demanda máxima en la curva del sistema.

Tabla 3.2. Resultados en por unidad PU de curvas tipo entre semana.

Hora	Residencial U	Residencial R	Comercial	Industrial	Otros	Total Sistema - EntreSem
0	0.2295	0.4077	0.4293	0.0369	0.6848	0.4910
1	0.2234	0.3907	0.4179	0.0363	0.6836	0.4810
2	0.2166	0.3690	0.4070	0.0360	0.6790	0.4689
3	0.2203	0.3577	0.4083	0.0372	0.6730	0.4658
4	0.2592	0.3706	0.4330	0.0411	0.6695	0.4869
5	0.3544	0.4089	0.4916	0.0524	0.6723	0.5435
6	0.4687	0.4478	0.5940	0.1105	0.6973	0.6365
7	0.5211	0.4553	0.7355	0.2970	0.7510	0.7578
8	0.4886	0.4338	0.8741	0.6037	0.8016	0.8791
9	0.4213	0.4020	0.9673	0.8748	0.8257	0.9586
10	0.3743	0.3749	1.0000	1.0000	0.8305	0.9829
11	0.3463	0.3649	0.9785	0.9806	0.8287	0.9607
12	0.3098	0.3726	0.9308	0.8761	0.8245	0.9099
13	0.2718	0.3909	0.8990	0.8071	0.8219	0.8761
14	0.2661	0.4143	0.8957	0.8488	0.8245	0.8922
15	0.3017	0.4465	0.8853	0.9244	0.8312	0.9305
16	0.3647	0.4969	0.8404	0.8927	0.8480	0.9453
17	0.4668	0.5931	0.7768	0.6857	0.8909	0.9372
18	0.6531	0.7578	0.7039	0.3999	0.9550	0.9526
19	0.8800	0.9303	0.6224	0.1884	1.0000	0.9942
20	1.0000	1.0000	0.5499	0.0934	0.9988	1.0000
21	0.9326	0.9422	0.4998	0.0625	0.9487	0.9296
22	0.7421	0.8288	0.4698	0.0517	0.8692	0.8131
23	0.5992	0.7542	0.4571	0.0476	0.8148	0.7339

La Tabla 3.3. muestra los resultados numéricos de fin de semana para los 4 tipos de consumidores analizados en este trabajo de titulación. Estos son: Residencial Urbano y Rural, Comercial, Industrial y Otros. Se hizo la distinción de días entre semana y fin de semana para obtener diferentes comportamientos de consumo en los mismos medidores. Gráficamente la Figura 3.30 muestra muchas similitudes entre los dos análisis, pero se observan picos menos pronunciados en las curvas aunque en el área residencial nuevamente obtenemos el valor de potencia máxima en por unidad en las 20:00.

Tabla 3.3. Resultados en por unidad PU de curvas tipo fin de semana

Hora	Residencial U	Residencial R	Comercial	Industrial	Otros	Total Sistema - FinDeSem
0	0.7927	0.5184	0.7238	0.7468	0.7484	0.7648
1	0.7219	0.5000	0.7206	0.7263	0.7454	0.7397
2	0.6370	0.4811	0.7157	0.7064	0.7401	0.7107
3	0.5861	0.4721	0.7115	0.7083	0.7351	0.6961
4	0.5732	0.4762	0.7121	0.7213	0.7233	0.6945
5	0.5842	0.4936	0.7195	0.7379	0.6984	0.7005
6	0.6141	0.5212	0.7385	0.7494	0.6853	0.7168
7	0.6598	0.5554	0.7867	0.7726	0.7129	0.7555
8	0.7203	0.5941	0.8722	0.8246	0.7563	0.8162
9	0.7689	0.6157	0.9593	0.8731	0.7767	0.8652
10	0.7579	0.6118	1.0000	0.9177	0.7750	0.8801
11	0.7140	0.6155	0.9901	0.9538	0.7702	0.8760
12	0.6959	0.6265	0.9531	0.9438	0.7722	0.8647
13	0.6944	0.6290	0.9091	0.9199	0.7798	0.8519
14	0.6704	0.6347	0.8653	0.9420	0.7888	0.8452
15	0.6497	0.6425	0.8223	0.9834	0.8039	0.8453
16	0.6661	0.6558	0.7814	1.0000	0.8397	0.8542
17	0.7262	0.7342	0.7578	0.9952	0.9103	0.8934
18	0.8355	0.8788	0.7555	0.9549	0.9839	0.9551
19	0.9500	0.9916	0.7555	0.9188	1.0000	1.0000
20	1.0000	1.0000	0.7485	0.8959	0.9585	0.9972
21	0.9745	0.9164	0.7398	0.8728	0.8937	0.9526
22	0.8981	0.7987	0.7341	0.8659	0.8272	0.8935
23	0.8380	0.7223	0.7322	0.8689	0.7893	0.8559

Las curvas características por tipo de consumidor de [18] son encontradas en un sistema de distribución perteneciente a la Illinois Power Company, USA y no se observan datos numéricos que den una idea del recurso matemático utilizado para hallar las gráficas en pu, pero muestran el comportamiento característico de los diferentes tipos de consumidores, indicando picos y formas de curvas en horarios que se asemejan a las gráficas de resultados finales de este trabajo de titulación.

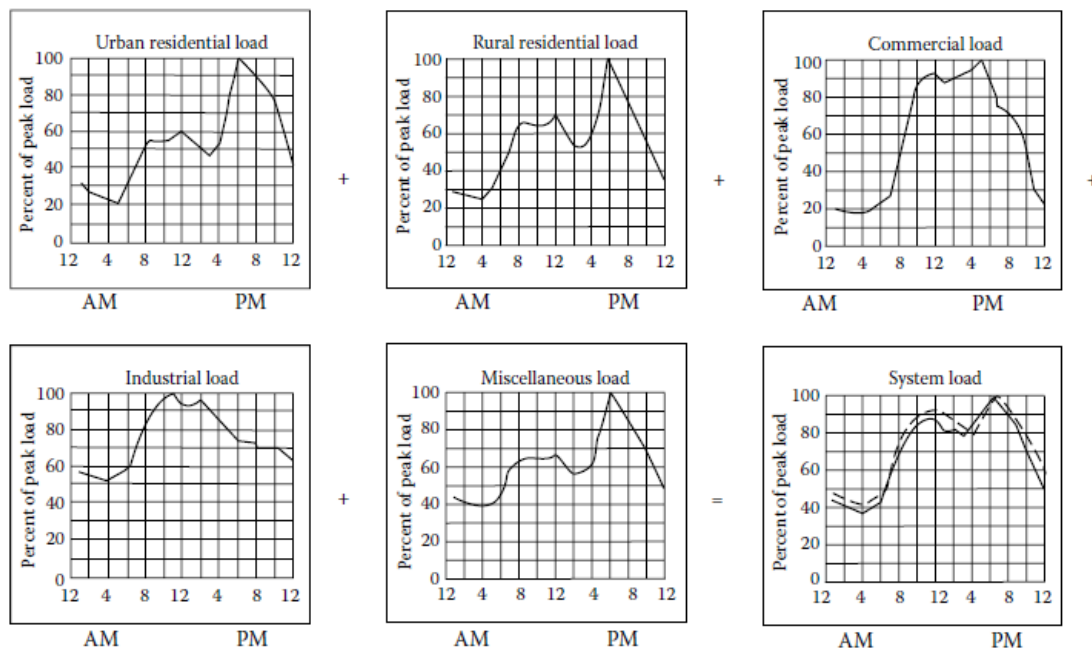


Figura 3.31. Curvas de demanda obtenidas de grupos de estudio de carga del capítulo 2 “Load Characteristics” de Turan Gönen [18].

3.4 COMPARACIÓN DE MÉTODOS

El método de obtener curvas de demanda utilizado por la EEA se enfoca en utilizar las mediciones reales exportándolas a un archivo Excel y luego con una serie de filtros tanto por georreferenciación, nivel de voltaje o alimentadores obtener curvas semejantes en valores reales de Potencia Activa kWh. Luego, empíricamente se van eliminando curvas que no representan un comportamiento característico de ese grupo de estudio para finalmente agrupar y obtener el promedio de las curvas parecidas.

Este método se emplea con índices altos de eficiencia ya que los encargados de planificación tienen la experiencia necesaria en determinar ubicaciones y filtros en catastros de usuarios de su empresa eléctrica. Estas curvas se utilizan posteriormente en los programas de simulación de redes de distribución y deben estar en pu para obtener resultados de análisis eléctricos, flujos de potencia, pérdidas. Pero si desde el inicio el método no encuentra valores horarios de la curva de demanda característica en pu, el algoritmo DBSCAN es una excelente propuesta, ya que solo necesita la base de datos en el formato especificado y poseer la herramienta de programación Matlab instalada en el computador.

Para recrear el método empírico de la EEA en la obtención de curvas de demanda se armó un archivo Excel con los mismos datos de medidores clasificándolos por tipo de consumidor y se aplicaron filtros de niveles de voltaje y formas de curva para así obtener la curva resultante final en kWh. Cabe recalcar que este procedimiento de agrupamiento es referencial ya que las empresas eléctricas manejan un volumen mayor de datos e incluyen condiciones técnicas de acuerdo con su zona de concesión.

Las curvas encontradas de esta manera muestran picos en diferentes valores horarios a los obtenidos de la Figura 3.30. que no son característicos del comportamiento del usuario y resultan de agrupamientos que no se validan con herramientas matemáticas. Las mediciones con valores reales no permiten el correcto agrupamiento por la no homogeneidad de sus gráficos en Excel. Por eso, la base de datos para el algoritmo en Matlab contiene los valores de potencia promedio ya normalizados en pu.

Para finalizar, se comparan las gráficas obtenidas en Excel mediante filtrado de datos con las gráficas resultantes obtenidas del algoritmo DBSCAN en Matlab. Se observa claramente que los cálculos en pu muestran un beneficio tanto gráfico como numérico en la ejemplificación de resultados mejorando la metodología para el agrupamiento de curvas de demanda y así encontrar las curvas que son características por tipo de consumidor.

4.4.1 ENTRE SEMANA

Se realiza la comparación gráfica para las curvas características por tipo de consumidor para días entre semana. La Figura 3.32. muestra la curva final entre semana Residencial urbana donde el lado izquierdo es el resultado con la metodología anterior de la empresa eléctrica en Excel mientras que el lado derecho muestra la curva resultante obtenida con el algoritmo DBSCAN en Matlab.

Es notoria la diferencia de los picos de las curvas en ciertas horas del día, la curva de la izquierda muestra 4 picos mientras que la gráfica de la derecha solo dos. Este se debe a que con el método anterior se necesitan más datos tanto de ubicación como de alimentador para encontrar comportamientos parecidos entre los usuarios. Esto no pasa con DBSCAN ya que trabajando en pu se logra que todas las gráficas tengan el máximo en 1 pu siendo así posible el análisis de agrupamiento con mejores resultados.

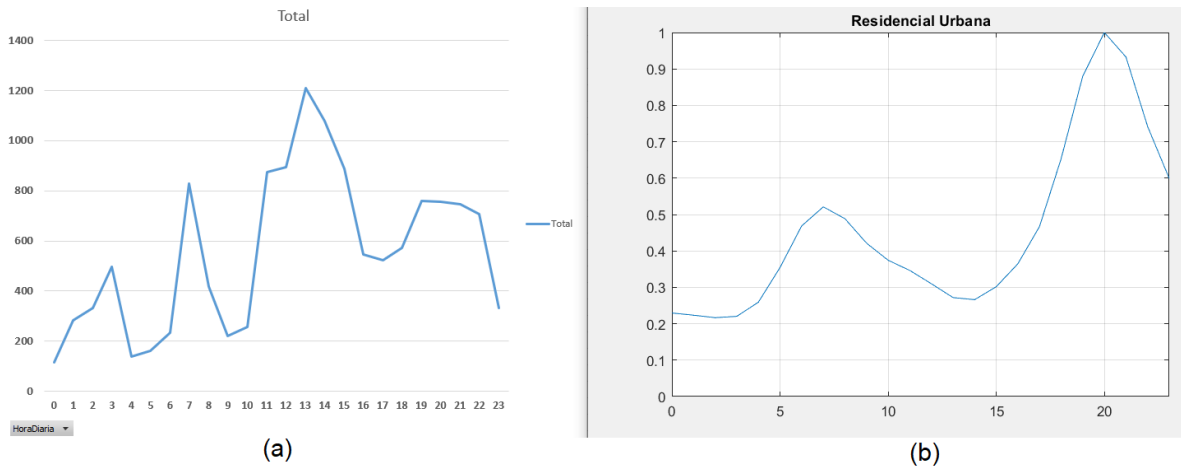


Figura 3.32. Curva final entre semana Residencial urbana a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

La Figura 3.33. señala la curva final entre semana Residencial rural donde el lado izquierdo es el resultado de la metodología anterior en Excel mientras que el lado derecho muestra la curva resultante obtenida con el algoritmo DBSCAN en Matlab. Nuevamente se observan picos que no se asemejan a las curvas encontradas en la bibliografía de [18] para el consumidor Residencial, es decir, estos pocos usuarios introducen picos que no son representativos cuando se analiza una muestra grande del sistema eléctrico de la EEA. También, se requieren una mayor cantidad de datos o condiciones que empíricamente conoce el personal por su experiencia y lograr resultados óptimos solamente usando MS Excel.

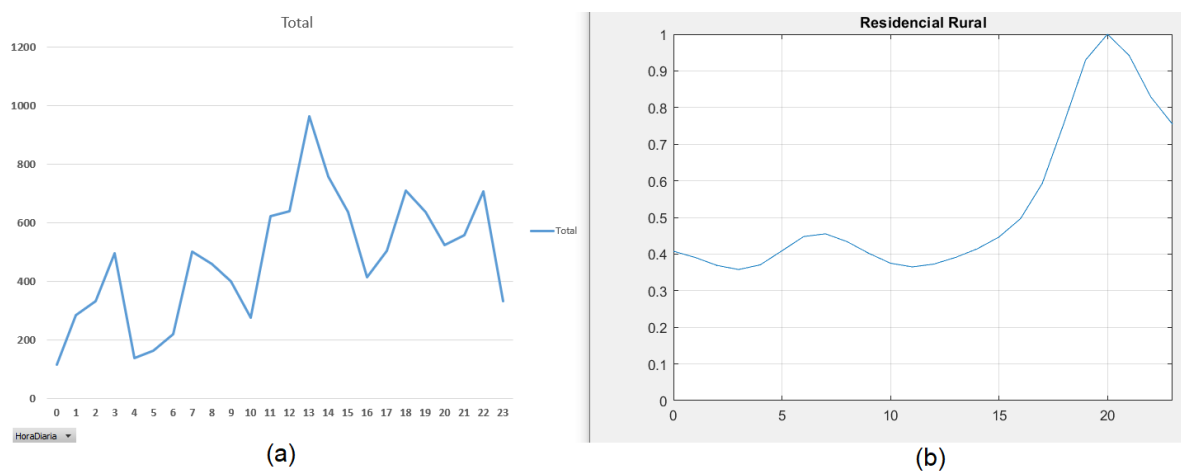


Figura 3.33. Curva final entre semana Residencial rural a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

En la Figura 3.34. se observa la curva final entre semana Comercial donde el lado izquierdo es el resultado de filtros en Excel mientras que el lado derecho, muestra la curva resultante obtenida con el algoritmo DBSCAN en Matlab. Su comportamiento es parecido pero los picos son dados en otros valores horarios. Esto se debe a que en la antigua metodología faltan datos para la realización de filtros mientras que la nueva provee el agrupamiento directo de las curvas parecidas.

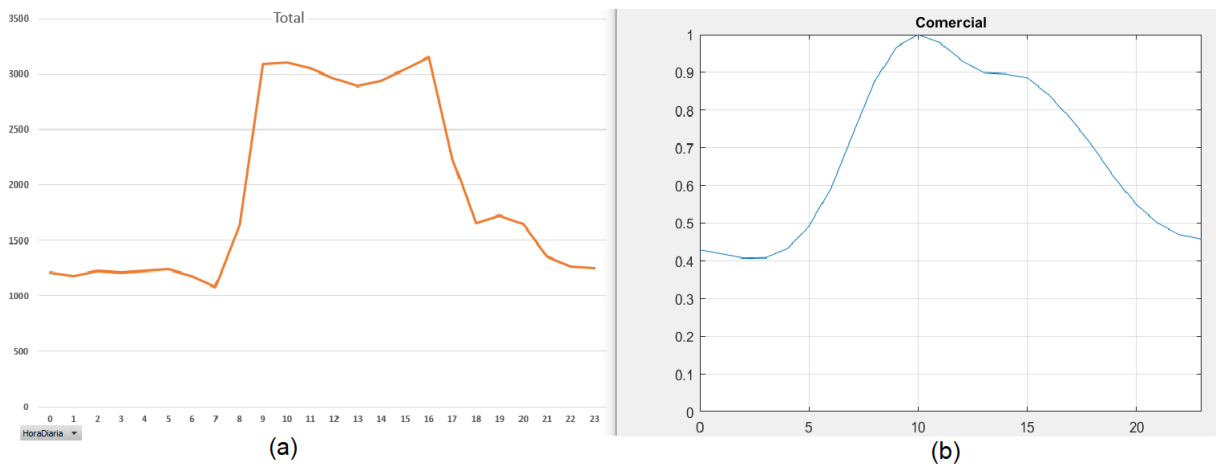


Figura 3.34. Curva final entre semana Comercial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

La Figura 3.35. muestra la curva final entre semana Industrial donde el lado izquierdo es el resultado por filtros en Excel mientras que el lado derecho se observa la curva resultante obtenida con el algoritmo DBSCAN en Matlab.

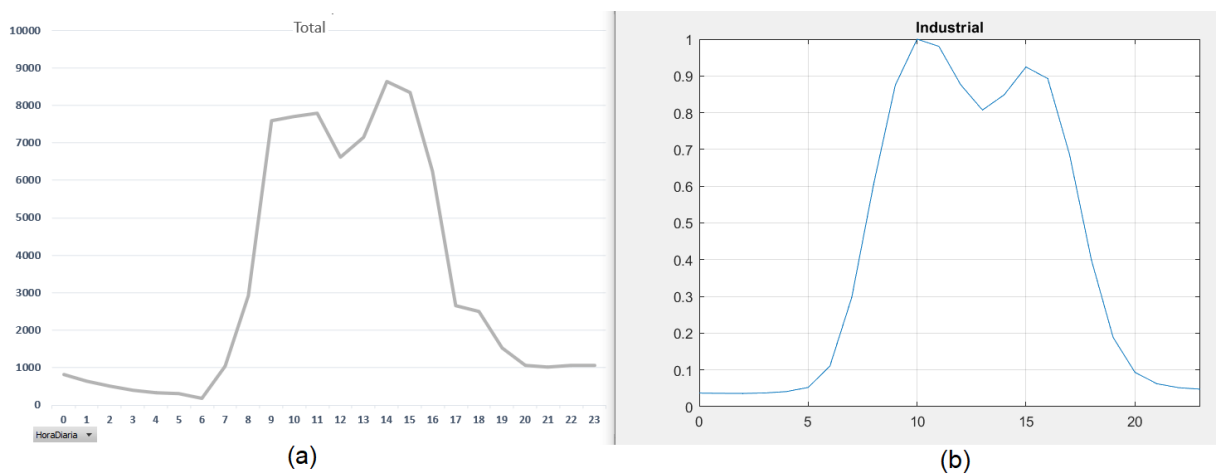


Figura 3.35. Curva final entre semana Industrial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

La Figura 3.36. muestra la curva final entre semana Otros donde el lado izquierdo es el resultado por filtros en Excel mientras que en el lado derecho se tiene la curva resultante obtenida con el algoritmo DBSCAN en Matlab. Las curvas tienen un gran parecido hasta en los picos de consumo. Esta vez los datos en Excel incluyeron muchas curvas parecidas para la denominación del consumidor Otros y su etapa de agrupación fue exitosa. De igual forma, los valores en pu convienen para una mejor identificación de la curva resultante.

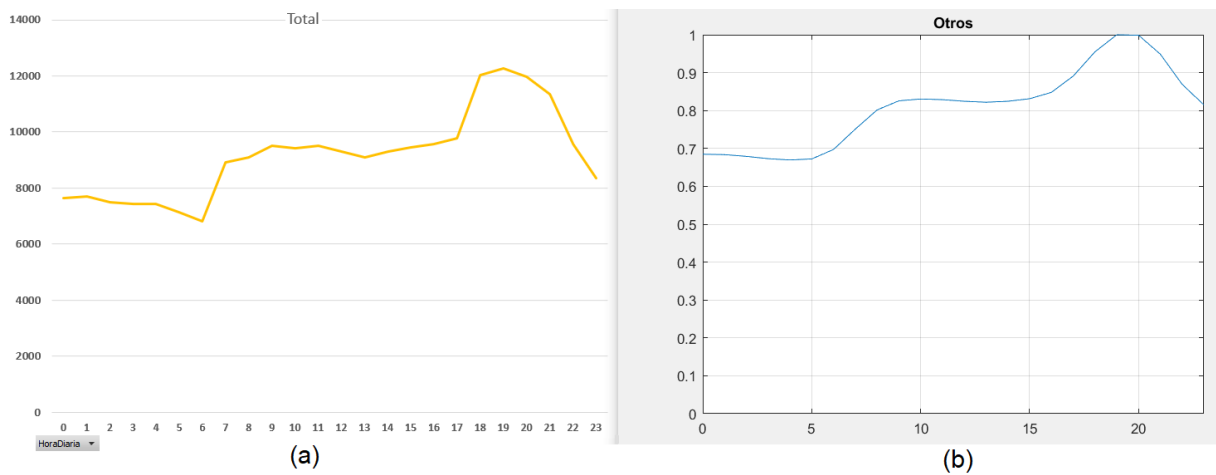


Figura 3.36. Curva final entre semana Otros a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

4.4.2 FIN DE SEMANA

Se realiza la comparación gráfica para las curvas características por tipo de consumidor para días de fin de semana. La Figura 3.37. muestra la Curva resultante fin de semana Residencial urbana. Al lado izquierdo con la antigua metodología en Excel que muestra una mayor cantidad de picos a las gráficas entre semana. Esto representaría el comportamiento de uno, dos hasta tres usuarios pero si se quiere analizar el sistema más bien estos interfieren en el resultado final ya que no tiene parecido a la bibliografía utilizada para este trabajo de titulación. En el lado derecho ya se observa una mejora en la gráfica con la nueva metodología en MATLAB.

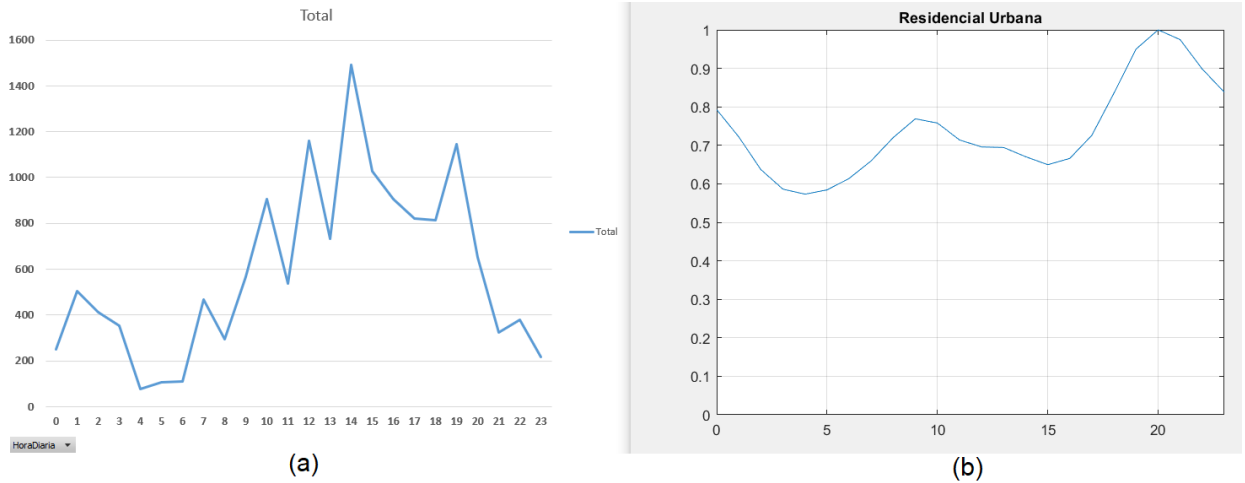


Figura 3.37. Curva resultante fin de semana Residencial urbana a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

La Figura 3.38. muestra la Curva resultante fin de semana Residencial rural donde el lado izquierdo es el resultado por filtros en Excel mientras que en el lado derecho se tiene la curva resultante obtenida con el algoritmo DBSCAN en Matlab.

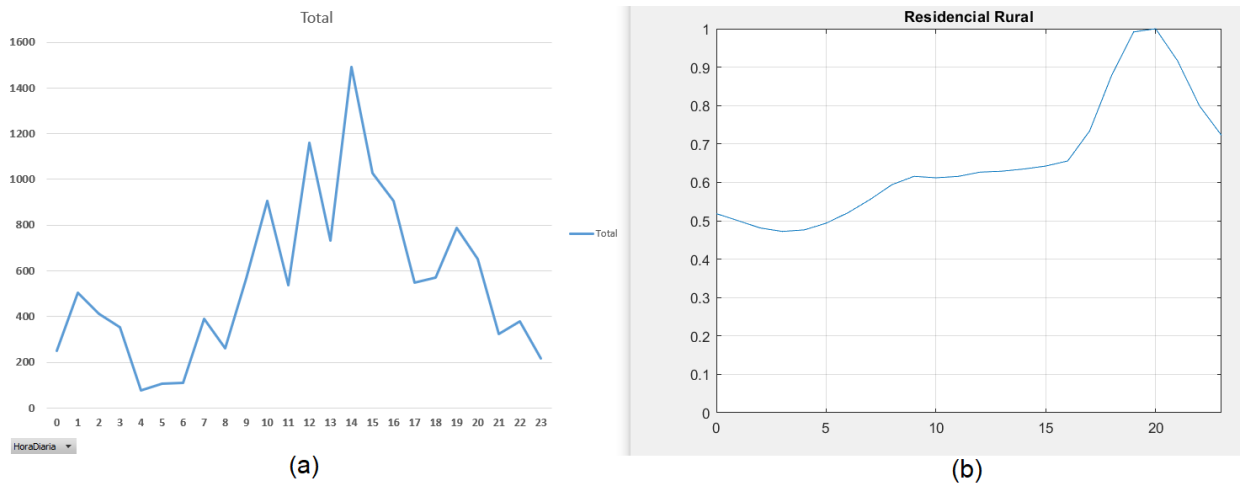


Figura 3.38. Curva resultante fin de semana Residencial rural a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

En la Figura 3.39. se observa la Curva resultante fin de semana Comercial donde el lado izquierdo es el resultado por filtros en Excel mientras que en el lado derecho se tiene la curva resultante obtenida con el algoritmo DBSCAN en Matlab.

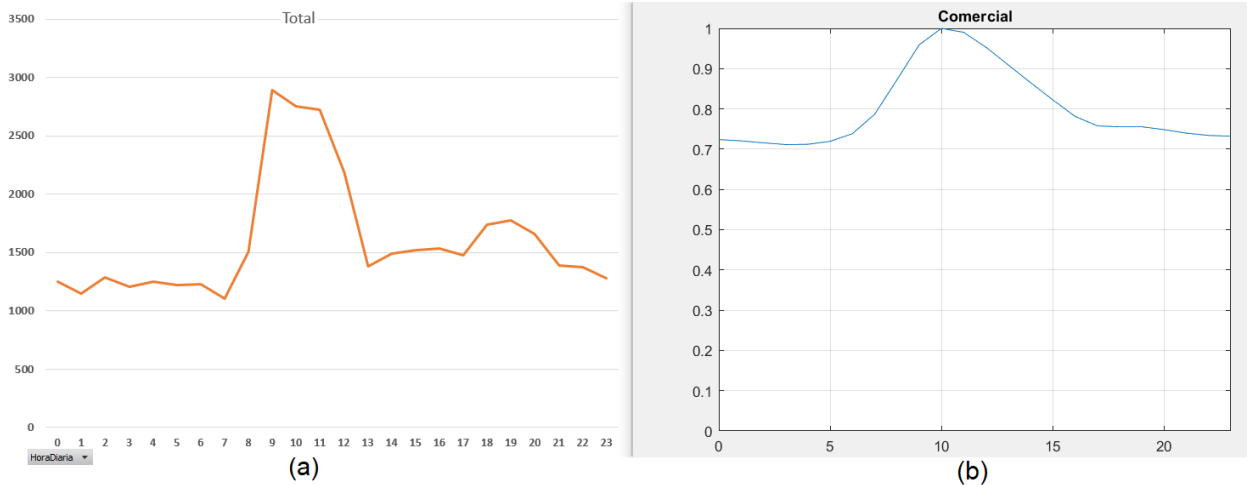


Figura 3.39. Curva resultante fin de semana Comercial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

Las curvas resultantes de la Figura 3.40. muestran del lado izquierdo la figura con valores en Excel mientras que del lado derecho se tiene la curva obtenida del algoritmo de clustering DBSCAN. Esta figura muestra un comportamiento de fin de semana Industrial.

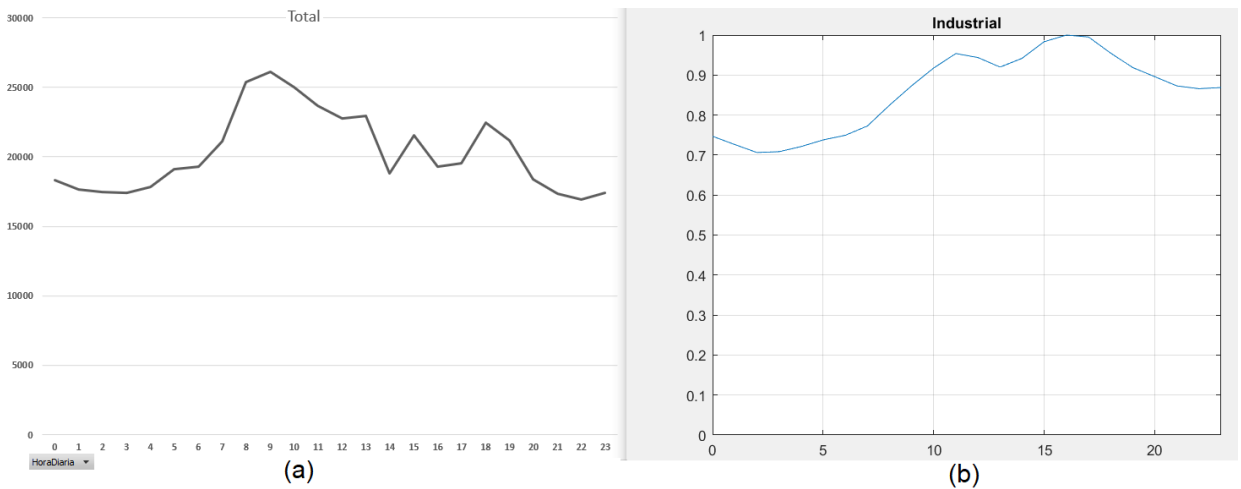


Figura 3.40. Curva resultante fin de semana Industrial a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

Las curvas resultantes de la Figura 3.41. muestran del lado izquierdo una figura con valores en Excel mientras que del lado derecho se tiene la curva final obtenida del algoritmo de clustering DBSCAN. Esta figura muestra un comportamiento de fin de semana Industrial.

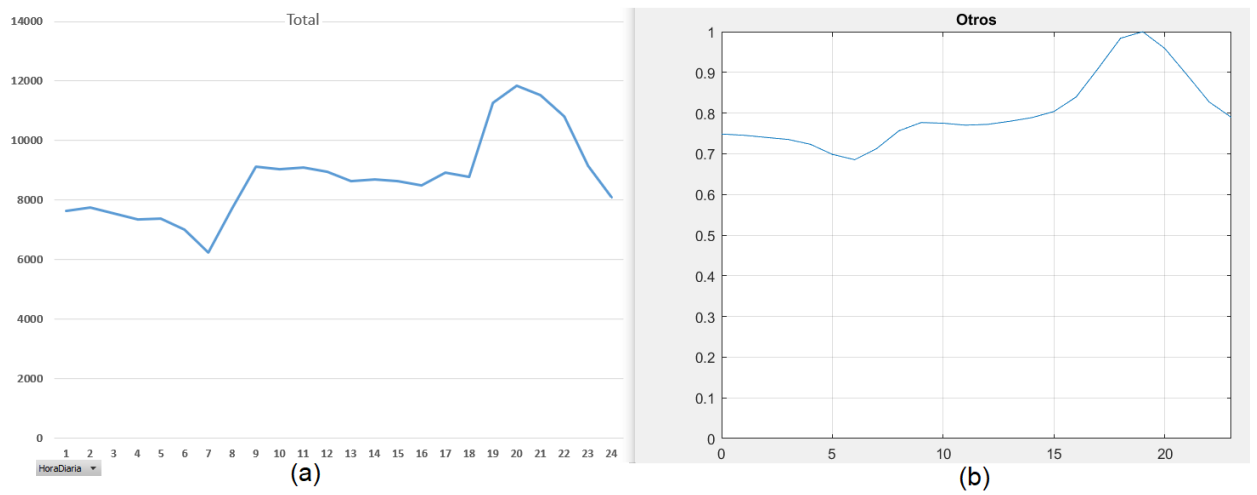


Figura 3.41. Curva resultante fin de semana Otros a) Obtenida con metodología anterior en Excel, b) Lograda con algoritmo DBSCAN

4. CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES

Se realizó la base de datos con mediciones reales y valores pu para la obtención de curvas de demanda características según el tipo de consumidor. Con esta base se estableció la metodología del trabajo de titulación y se incluyó la técnica de agrupamiento DBSCAN para la obtención de resultados.

La forma horaria de la curva y los datos atípicos hacían que el análisis mediante distancias euclidianas sea más complejo y que su esquematización proporcione formas arbitrarias. El método que se ajusta a este análisis se basa en densidades de vecinos cercanos o DBSCAN proporcionando así una excelente agrupación para cada tipo de consumidor.

Debido a la poca cantidad de curvas por medidor y a su similitud, el método de obtener ϵ utilizando Knnsearch y knee_pt fue un éxito para encontrar la curva característica de cada usuario en estudio. El momento de establecer el clúster entre medidores y obtener ya la curva definitiva por tipo de consumidor este método de encontrar argumentos con algoritmos complementarios solo proporcionaba el punto de partida hacia territorio atípico. Por tanto, se utilizaron valores dentro de la zona típica de distancias.

Las curvas para cada grupo de consumo muestran un comportamiento parecido al observado en "Load Characteristics" de Turan Gönen. Existieron mínimas variaciones en todas las curvas; esto es debido al déficit de industrialización en esta ciudad, al bajo contenido comercial y al cambio de densidad de demanda en zonas urbanas y rurales. También se pudo observar que las mediciones atípicas debido al mal uso del equipo afectaron en una pequeña cantidad a las curvas resultantes.

En los resultados durante el procedimiento del clustering se pueden observar curvas que tienen comportamiento de alumbrado público y de industrias que solo operan en horas de la madrugada, esta característica de consumo no entró en la obtención de curvas finales debido a su poca intervención en el sistema eléctrico de distribución.

La ventaja del método presentado en este trabajo escrito es el uso de un algoritmo de agrupamiento para la obtención de curvas de demanda en pu y la herramienta de validación para descartar mediciones atípicas. La metodología anterior de la EEA no tiene etapa de validación solo utiliza filtros en MS Excel y experiencia del analista.

4.2 RECOMENDACIONES

Para lograr la optimización en los argumentos de entrada con el algoritmo basado en densidades DBSCAN, se recomienda instalar el equipo de medición de calidad un rango mayor de días incluyendo más datos de fines de semana.

En trabajos futuros, se recomienda incluir etiquetas de georreferenciación, alimentador o estrato al que pertenece cada medidor para que el estudio dirija las curvas características hacia sectores o grupos en específico.

Utilizando información complementaria del catastro y aplicando el método de agrupamiento de este trabajo de titulación se lograría un nuevo procedimiento de estratificación para los usuarios finales. La forma de curva identificaría si pertenece o no al estrato asignado y la correcta estratificación ayudaría a reducir las pérdidas en el sistema de distribución al planificar nueva infraestructura, reconfiguración de alimentadores o repotenciación de transformadores.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] M. d. E. y. E. N. N. Renovables, «Plan Maestro de Electricidad 2018-2027,» 1 Febrero 2018. [En línea]. Available: <https://www.rekursosyenergia.gob.ec/plan-maestro-de-electricidad/>. [Último acceso: 22 Junio 2022].
- [2] S. Fong y S. U. Rehman, «DBSCAN: Past, Present and future,» de *The fifth International Conference on the Applications of Digital Information and Web Technologies, IEEE*, Chennai, India, 2014.
- [3] T. Short, *Electric Power Distribution Handbook*, Boca Raton, USA: CRC Press, 2004.
- [4] H. Willis, *Power Distribution Planning Reference Book*, Raleigh, USA: Marcel Decker, Inc., 2002.
- [5] V. Kamaraju, *Electrical Power Distribution Systems*, Kakinada, India: Tata McGraw Hill, 2009.
- [6] S. R. Castaño, *Redes de Distribución de Energía*, Bogota, Colombia: Universidad Nacional de Colombia, 2004.
- [7] . Y. El-Sonbaty, M. A. Ismail y M. Farouk, «An efficient Density Based Clustering Algorithm for Large Databases,» de *ICTAI 16th, IEEE*, Boca Raton, USA, 2004.
- [8] M. Ester, H.-P. Kriegel y J. Sander, «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,» de *Second International Conference on Knowledge Discovery and Data Mining*, Protland, USA, 1998.
- [9] K. Sawant y M. Gaonkar, «DBSCAN with Eps Automatic for Large Dataset,» *International Journal on Advanced Computer Theory and Engineering* , vol. 2, nº 2, pp. 2319-2526, 2013.
- [10] Poerwanto, B;, «Evaluating the K-Means Analysis in Clustering Area Based on Estates Productivity in Tana Luwu using Silhouette Index,» de *3rd International Conference on Statistics*, Makassar, Indonesia, 2019.

- [11] N. N. Kachouie y M. Shutawi, «Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering,» *Entropy MDPI*, vol. 23, nº 759, pp. 1-17, 2021.
- [12] M. Bolandraftar, «Application of K-nearest neighbor (KNN) approach for predicting economic events: Theoretical background,» *Int. Journal of Engineering Research and Applications*, vol. III, nº 5, pp. 605-610, 2013.
- [13] D. Irwin, J. Albrecht y V. Satopa, «Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior,» de *31st International Conference on Distributed Computing Systems Workshops*, Minneapolis, USA, 2011.
- [14] L. Tylee, «User's Guide Learn Visual Basic 6.0,» KIDware, Washington, USA, 2002.
- [15] E. Neuwirth y T. Baie, «Excel :: COM :: R,» *Computational Statistics*, vol. XXII, pp. 91-108, 2007.
- [16] R. V. Dukkupati, «Matlab Basics» in *MATLAB An introduction with Applications*, Daryaganj, India: New Age International Ltd. Publishers, 2010.
- [17] P. J. Rousseeuw, «Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,» *Journal of Computational and Applied mathematics*, vol. XX, pp. 53-65, 1999.
- [18] T. Gönen, *Electric Power Distribution Engineering*, Boca Raton, Usa: CRC Press, 2014.

ANEXOS

Este trabajo de titulación incluye 3 anexos que incorporan información relevante a la obtención de curvas de demanda por tipo de consumidor, pero que, por su extensión, no fueron ser incorporadas directamente en ninguna de las secciones anteriores.

ANEXO A.

Archivo .m llamado GenerarCurvasDeCarga que contiene el algoritmo de Clustering basado en densidades DBSCAN.

ANEXO B.

Información de medidores empleados en el algoritmo DBSCAN:

Medidores	Nivel de Voltaje	Grupo de Consumo
1000055237	BAJA	Residencial
1000055290	BAJA	Residencial
1000055999	BAJA	Residencial
1000056718	BAJA	Residencial
1000056799	BAJA	Residencial
1000057625	BAJA	Residencial
1000345353	BAJA	Residencial
1000345363	BAJA	Residencial
1000345673	BAJA	Residencial
1000346047	BAJA	Residencial
1000346067	BAJA	Residencial
1000346072	BAJA	Residencial
1000347008	BAJA	Residencial
1000347343	BAJA	Residencial
1000347374	BAJA	Residencial
1000347469	BAJA	Residencial
1000348072	BAJA	Residencial
1000348743	BAJA	Residencial
1000348817	BAJA	Residencial
1000349096	BAJA	Residencial
1000349571	BAJA	Residencial
1000349743	BAJA	Residencial
1000349893	BAJA	Residencial
1000350009	BAJA	Residencial
1000350448	BAJA	Residencial
1000350800	BAJA	Residencial
1000351113	BAJA	Residencial
1000351393	BAJA	Residencial
1000351640	BAJA	Residencial
1000351702	BAJA	Residencial
1000352284	BAJA	Residencial
1000352328	BAJA	Residencial
1000352486	BAJA	Residencial
1000352488	BAJA	Residencial
1000352511	BAJA	Residencial
1000352634	BAJA	Residencial
1000352831	BAJA	Residencial
1000352871	BAJA	Residencial
1000353992	BAJA	Residencial
1000354180	BAJA	Residencial
1000354308	BAJA	Residencial

Medidores	Nivel de Voltaje	Grupo de Consumo
1000354341	BAJA	Residencial
1000354591	BAJA	Residencial
100055358	NoExiste	NoExiste
1000570997	BAJA	Residencial
1000571140	BAJA	Residencial
1000572687	BAJA	Residencial
1000572958	BAJA	Residencial
1000573169	BAJA	Residencial
1000579823	BAJA	Residencial
1000579931	BAJA	Residencial
1000579946	BAJA	Residencial
1000580029	BAJA	Residencial
1000585954	BAJA	Residencial
1000596883	BAJA	Industrial
1000596888	BAJA	Residencial
1000597278	MEDIA	Comercial
1000597334	BAJA	Comercial
100774	BAJA	Residencial
101034	BAJA	Comercial
101099	BAJA	Residencial
101193	BAJA	Residencial
101248	BAJA	Otros
101278	BAJA	Residencial
101285	BAJA	Otros
101304	BAJA	Residencial
101315	BAJA	Residencial
102301	BAJA	Otros
110113	MEDIA	Comercial
110151	BAJA	Comercial
110303	BAJA	Residencial
1107001180	BAJA	Residencial
1107001549	BAJA	Otros
1107701024	BAJA	Otros
110804	BAJA	Residencial
110882	BAJA	Residencial
11212312	MEDIA	Industrial
11212319	MEDIA	Comercial
11212322	MEDIA	Comercial
11212324	BAJA	Industrial
113746	BAJA	Residencial
113757	BAJA	Residencial
114418	BAJA	Residencial
117453	BAJA	Residencial
117591	BAJA	Residencial
118402	BAJA	Residencial

Medidores	Nivel de Voltaje	Grupo de Consumo
118887	BAJA	Residencial
119850	BAJA	Residencial
121027	BAJA	Residencial
123392	BAJA	Residencial
124360	BAJA	Otros
124370	BAJA	Comercial
124924	BAJA	Residencial
125470	BAJA	Residencial
126390	BAJA	Residencial
127018	BAJA	Residencial
128471	BAJA	Residencial
129257	BAJA	Comercial
130118	BAJA	Residencial
1408701010	BAJA	Otros
1408701101	BAJA	Residencial
1408701474	BAJA	Residencial
1408701608	BAJA	Residencial
1408701645	BAJA	Comercial
1408701723	BAJA	Comercial
1408701746	BAJA	Residencial
1408701948	BAJA	Residencial
1408702106	BAJA	Residencial
1601701182	BAJA	Comercial
1601701188	BAJA	Industrial
1601701239	BAJA	Residencial
1601701242	BAJA	Industrial
1601701244	MEDIA	Industrial
16220837	MEDIA	Otros
16606806	BAJA	Comercial
1717119	BAJA	Comercial
19695392	MEDIA	Industrial
19695399	MEDIA	Otros
19695400	BAJA	Industrial
19695401	MEDIA	Industrial
19695641	MEDIA	Comercial
19814045	BAJA	Comercial
2010110198	BAJA	Residencial
2010110364	BAJA	Residencial
2010111024	BAJA	Residencial
2010111339	BAJA	Residencial
201105	BAJA	Residencial
201299	BAJA	Industrial
201311056	BAJA	Residencial
201340113	BAJA	Residencial
201373	BAJA	Residencial

Medidores	Nivel de Voltaje	Grupo de Consumo
2014207826	BAJA	Residencial
2014208283	BAJA	Residencial
2014208323	BAJA	Residencial
2014208644	BAJA	Residencial
2014208795	BAJA	Residencial
2014209727	BAJA	Residencial
2014400122	MEDIA	Otros
2014400125	BAJA	Industrial
201632	BAJA	Residencial
201636	BAJA	Industrial
2017101348	BAJA	Residencial
201770	BAJA	Industrial
201819	MEDIA	Industrial
2018202002	BAJA	Comercial
201918	BAJA	Comercial
202109	BAJA	Residencial
202140	BAJA	Residencial
202283	BAJA	Residencial
202327	MEDIA	Otros
202374	BAJA	Residencial
202379	BAJA	Residencial
202476	BAJA	Residencial
202506	BAJA	Residencial
202507	BAJA	Comercial
2798954	BAJA	Otros
2831842	BAJA	Comercial
2838559	BAJA	Otros
300496	BAJA	Comercial
300599	MEDIA	Comercial
300609	BAJA	Industrial
300612	BAJA	Industrial
300631	BAJA	Comercial
300640	BAJA	Industrial
300658	MEDIA	Comercial
300674	BAJA	Comercial
300695	BAJA	Comercial
300709	MEDIA	Comercial
300726	BAJA	Comercial
300749	BAJA	Comercial
300786	BAJA	Residencial
300803	BAJA	Comercial
300808	BAJA	Residencial
300828	MEDIA	Otros
300849	BAJA	Comercial
3313204	NoExiste	NoExiste

Medidores	Nivel de Voltaje	Grupo de Consumo
400002	MEDIA	Otros
400012	MEDIA	Comercial
400017	MEDIA	Otros
400023	MEDIA	Comercial
400027	MEDIA	Comercial
400032	BAJA	Industrial
400036	MEDIA	Comercial
400041	BAJA	Comercial
400042	MEDIA	Comercial
400045	MEDIA	Industrial
400046	MEDIA	Industrial
400049	BAJA	Industrial
400055	MEDIA	Comercial
400057	BAJA	Industrial
400064	MEDIA	Industrial
400069	BAJA	Comercial
400071	BAJA	Industrial
400072	MEDIA	Industrial
400073	MEDIA	Otros
400074	MEDIA	Otros
400077	MEDIA	Industrial
400089	MEDIA	Industrial
400091	MEDIA	Industrial
400096	MEDIA	Industrial
400097	MEDIA	Comercial
400098	MEDIA	Comercial
400115	NoExiste	NoExiste
4093104	NoExiste	NoExiste
4093114	NoExiste	NoExiste
4093159	MEDIA	Comercial
4384167	NoExiste	NoExiste
4604290	NoExiste	NoExiste
5204288	BAJA	Residencial
5204292	BAJA	Residencial
5204365	BAJA	Residencial
5204452	BAJA	Residencial
5204613	BAJA	Residencial
5204644	BAJA	Residencial
5204686	BAJA	Residencial
5204760	BAJA	Residencial
5205102	BAJA	Residencial
5205208	BAJA	Residencial
5205247	BAJA	Residencial
5205293	BAJA	Residencial
5205434	BAJA	Residencial

Medidores	Nivel de Voltaje	Grupo de Consumo
5206020	BAJA	Residencial
5206295	BAJA	Residencial
5206299	BAJA	Residencial
5206705	BAJA	Residencial
5207184	BAJA	Residencial
55835	BAJA	Industrial
61564654	MEDIA	Comercial
73772	MEDIA	Comercial
7398837	BAJA	Residencial
8063068	NoExiste	NoExiste
8063074	MEDIA	Industrial
8063091	MEDIA	Comercial
86965	BAJA	Industrial
T2496- UNAE	NoExiste	NoExiste
T2497- UNAE	NoExiste	NoExiste

ANEXO C.

Base de Datos adjuntada como archivo Excel y usada en el algoritmo DBSCAN

ORDEN DE EMPASTADO