# Uso de algoritmo K-means para clasificar perfiles de clientes con datos de medidores inteligentes de consumo eléctrico: Un caso de estudio

Using K-means algorithm to classify customer profiles with data from smart energy consumption meters: A case study

Lester Marrero<sup>1</sup> Dante Carrizo<sup>2</sup> Luis García-Santander<sup>1</sup>\* Fernando Ulloa-Vásquez<sup>3</sup>

Recibido 6 de enero de 2021, aceptado 28 de agosto de 2021 Received: January 6, 2021 Accepted: August 28, 2021

#### RESUMEN

La eficiencia energética es parte de los objetivos establecidos por los gobiernos de todo el mundo para reducir la huella energética y proporcionar un desarrollo sostenible para todos. El advenimiento de nuevas tecnologías que permiten el monitoreo y autocontrol del consumo eléctrico dentro de los hogares, como los medidores inteligentes, permiten a los usuarios finales integrarse en los sistemas de gestión inteligente de la red eléctrica, al proporcionar información sobre los perfiles de consumo de energía eléctrica y los precios de ésta. El presente trabajo realiza una clasificación de clientes residenciales a partir de los datos de consumo obtenidos desde medidores inteligentes. Para esto, se utiliza una metodología basada en el algoritmo K-means para identificar patrones de comportamiento en el consumo asociado a 1179 clientes conectados a redes eléctricas de distribución de baja tensión del sur de Chile equipados con medidores inteligentes. Los resultados fueron validados a través del uso de medidas de la teoría de los conjuntos aproximados. En particular, se caracterizan grupos conformados por medio de sus centroides, posibilitando la conversión de los grandes volúmenes de datos coleccionados en conocimientos útiles, lo que resulta en beneficio tanto para el propio cliente residencial como para la empresa distribuidora de energía eléctrica. Los resultados muestran que dos clústeres son los que mejor representan el conjunto de clientes. La oportunidad de recopilar datos de consumo en tiempo real mediante estos dispositivos, ofrece perspectivas para optimizar las políticas públicas y privadas sobre distribución de energía eléctrica.

Palabras clave: Medidores inteligentes, K-means, agrupamiento, perfiles de consumo, clientes residenciales.

## **ABSTRACT**

Energy efficiency is part of the goals set by governments around the world to reduce the energy footprint and provide sustainable development for all. The arrival of new technologies that allow the monitoring and self-control of electricity consumption within homes, such as smart meters, allow end users to integrate into the intelligent management systems of the electricity grid, by providing information on the flow of energy and prices. This work performs a classification of residential customers from the consumption data obtained from smart meters. For this, a methodology based on the simple K-means algorithm is used to identify patterns of behavior in the consumption of 1179 homes connected to a real low-voltage electricity distribution network in southern Chile equipped with smart meters, and the validation and

<sup>&</sup>lt;sup>1</sup> Univ. de Concepción. Depto. de Ing. Eléctrica. Concepción, Chile. E-mail: lmarrero@udec.cl; luis.garcia@udec.cl

Univ. de Atacama. Depto. de Ing. Informática y Cs. de la Computación. Copiapó, Chile. E-mail: dante.carrizo@uda.cl

Univ. Tec. Metropolitana. Depto. de Electricidad. Santiago, Chile. E-mail: fulloa@utem.cl

<sup>\*</sup> Autor de correspondencia: luis.garcia@udec.cl

refinement of the results using numerical measures belonging to the rough sets theory. Final groups are characterized based on their centroids, making possible to convert the large volumes of data collected into useful knowledge, which is beneficial for both the residential customer and the electric power distribution company. The results show that two clusters are the ones that best represent the set of clients. The opportunity to collect consumption data in real time through these devices offers perspectives to optimize public and private policies on electricity distribution.

Keywords: Smart meters, K-means, clustering, load profiles, residential clients.

# INTRODUCCIÓN

Los medidores inteligentes (*smart meters*) son los dispositivos electrónicos de medición que permiten la comunicación bidireccional entre el consumidor y la empresa distribuidora. Estos equipos de medición permiten registrar datos de voltaje, corriente, potencia y energía los cuales quedan disponibles tanto para los propios usuarios como la empresa distribuidora de energía, y transmitirlos de manera remota [1, 2].

Estos medidores tienen la capacidad de llevar dos transmisiones con la finalidad de proporcionar formas diferentes de comunicación de la información. La primera de ellas es para la lectura automática remota (AMR, acrónimo en inglés de *Automatic Meter Reading*), así como para el monitoreo en tiempo real de las variables eléctricas en el consumidor y transmitir la información a un punto de acceso con la empresa distribuidora [3]. La segunda permite un mayor control del consumo y uso de la energía eléctrica en los equipos eléctricos, lo que permite a los consumidores habilitar una red de área domiciliaria (HAN, acrónimo en inglés de *Home Área Network*) [4] y gestionar sus hábitos [5].

La transmisión de datos de consumo permite conocer hábitos de consumo de los clientes, apoyando con ello la toma de decisiones en políticas públicas de eficiencia energética y en el desarrollo de ciudades inteligentes [6]. Conocer los hábitos de consumo de los clientes es útil para las distribuidoras eléctricas ya que les permite adecuar sus programas de planificación y operación, así como también establecer políticas de tarifas flexibles [7].

Para poder sistematizar y dirigir estas decisiones es necesario conocer los perfiles de los consumidores con el fin de encontrar similitudes y poder agruparlos apropiadamente. En esta dirección, este trabajo pretende caracterizar grupos de clientes residenciales a partir de los datos de medidores inteligentes de consumo eléctrico. Para esto, se presenta un estudio que utiliza una metodología basada en el análisis clúster para identificar similitudes en el comportamiento de los perfiles de carga en clientes residenciales a pequeña escala y la caracterización de los grupos conformados por medio de sus centroides. El análisis clúster resulta un método efectivo para comprender el comportamiento de parámetros medibles en el sistema de potencia, centrándose sus principales aplicaciones dentro del mismo en la evaluación de la seguridad, la optimización de la operación, el pronóstico de carga y la detección de eventos [8].

La teoría de los conjuntos aproximados (RST, siglas en inglés de *Rough Sets Theory*) es utilizada para validar dichas agrupaciones [9]. Mediante el empleo de la RST para valorar los resultados del agrupamiento, es posible realizar una validación no supervisada, evaluando la calidad y precisión de los resultados local y globalmente. La metodología es implementada en MATLAB.

El estudio consideró lecturas de consumos de energía obtenidas de 1179 medidores inteligentes, instalados en igual número de clientes residenciales pertenecientes a empresas de distribución del sur de Chile, como parte de la política de cambio e implementación de esta nueva tecnología establecida a partir del año 2017. Dichas mediciones se efectúan con intervalos de 15 minutos y corresponden al periodo comprendido entre el 4 de marzo y el 4 de abril de 2019.

El trabajo está organizado de la siguiente manera: La Sección 2 presenta los trabajos relacionados con este estudio; la Sección 3 describe las técnicas analíticas de la metodología presentada; en la Sección 4, la metodología es aplicada a las mediciones reales y se exponen los resultados más relevantes y su análisis.

Finalmente, la Sección 5 exhibe las conclusiones del trabajo.

## TRABAJOS RELACIONADOS

La creciente penetración de medidores inteligentes brinda una excelente oportunidad para monitorear y analizar el consumo de energía de clientes residenciales. Sin embargo, el manejo de grandes cantidades de datos entregados por sistemas de medición inteligentes requiere tecnologías avanzadas de análisis de datos y algoritmos de detección de patrones. Existen varios trabajos en esta dirección.

Fahim y Sillitti [10] proponen un marco para procesar los perfiles observados de consumo de energía para inferir las características del hogar en edificios residenciales. Su enfoque se basa en la extracción automatizada de características de datos de series de tiempo univariantes y el desarrollo de un modelo a través de una variante de la técnica de árboles de decisión.

Liang y otros [11] se centran en reducir el consumo innecesario de energía doméstica y reemplazar los refrigeradores y congeladores de baja eficiencia mediante el uso de datos de medidores inteligentes y datos de temperatura diaria. Hay trabajos como el de Rajabi y otros [12] que realizan un estudio comparativo de diferentes técnicas para el agrupamiento de patrones de carga. Pero la mayoría aplican alguna estrategia para analizar los datos.

Ullah y otros [13] proponen un análisis basado en el agrupamiento del consumo de energía para clasificar el uso de electricidad de los consumidores en diferentes niveles, entrenando un *autoencoder* profundo que transfiere los datos de consumo de energía de baja dimensión a representaciones de alto nivel y luego un algoritmo de agrupamiento adaptativo de mapa autoorganizado.

Sala, Li, y Christensen [14] aplicaron métodos de aprendizaje automático de uso común en los datos de consumo de calefacción de dos hogares individuales en un edificio de apartamentos y la subestación de calefacción de distrito del edificio de apartamentos que incluye 72 hogares, para identificar cómo las características de los datos afectan el resultado del análisis de datos. Se aplicaron dos enfoques de

agrupamiento utilizando el algoritmo *K-means* para agrupar perfiles diarios de calentamiento similares.

Zhu y otros [15] desarrollaron un proceso sistemático de uso de datos de medición inteligente para cuantificar la construcción de perfiles de carga diaria (es decir, patrones de consumo de energía) con un conjunto de estadísticas: Carga base, carga máxima, tiempo de subida, etc.

Yilmaz, Chambers, y Patel [16] presentan una evaluación rigurosa del agregado de muestras y los enfoques de agrupamiento que usan cinco características para crear perfiles representativos de demanda de electricidad de medidores inteligentes de 656 hogares en Suiza. También Markovic y otros [17] proponen una técnica novedosa para el agregado de datos inspirada en los principios de la ciencia de redes. Para ello utilizan un conjunto de datos de consumo medidos por hora de un año de 2201 usuarios.

Finalmente, Cerquitelli, Chicco, y Di Corso [21] presentan una metodología innovadora y escalable llamada CONDUCTS para descubrir comportamientos de consumo de electricidad residencial a lo largo del tiempo. CONDUCTS explota el procesamiento de flujo de datos en ventanas de tiempo juntamente con el aprendizaje automático no supervisado en datos independientes del tiempo.

Los trabajos previos evidencian como el uso de las recientes tecnologías de comunicación y análisis de datos, y los sistemas de medición inteligente, permiten una mayor flexibilidad en la gestión energética dentro de la red. Los grandes volúmenes de estos datos de consumo revelan información importante acerca de los perfiles de carga de los usuarios, y el usarlos para mejorar los servicios, promover la eficiencia energética y apoyar el desarrollo del sistema de distribución es un punto en común del presente estudio con respecto a dichas investigaciones.

## METODOLOGÍA PROPUESTA

El estudio utiliza un algoritmo que permite realizar una clasificación de los perfiles de clientes en base a los datos de consumo eléctrico provenientes de los medidores inteligentes instalados. Para esto, se diseñó una metodología que es presentada en la Figura 1.

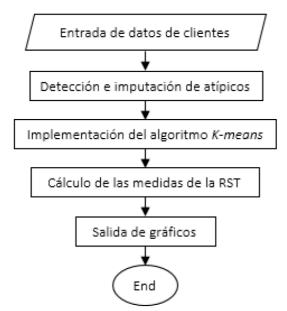


Figura 1. Metodología de investigación.

## Detección e imputación de atípicos

En la búsqueda de su estructura, el análisis clúster desarrollado es muy sensible a la inclusión de variables irrelevantes. Los datos atípicos pueden representar tanto observaciones verdaderas, que no son representativas en general, como una muestra reducida del grupo que provoca una mala representación. En ambos casos, los atípicos distorsionan la verdadera estructura y hacen que los conglomerados deducidos no sean representativos de esta [19].

Para la detección de estos datos anómalos, se utiliza en el estudio la perspectiva univariante que examina la distribución de las observaciones y selecciona como atípicos aquellos casos fuera de un rango de distribución con valor umbral de tres desviaciones estándar. En el caso de lecturas atípicas, el método de imputación seleccionado es el valor medio (valor más probable).

#### Algoritmo *K-means* simple

El principal objetivo del análisis clúster se enfoca en agrupar objetos basándose en las características que poseen. Los conglomerados de objetos resultantes muestran altos grados de homogeneidad interna y de heterogeneidad externa. Al conformar grupos homogéneos, resulta factible la descripción de taxonomías, la simplificación de datos y la identificación de relaciones.

En el presente estudio, se aplica el método no jerárquico *K-means* que particiona los individuos, en este caso clientes, en un número específico de grupos. Este algoritmo ha mantenido su popularidad en aplicaciones de agrupamiento debido a su buen rendimiento y competitividad con enfoques sugeridos más recientemente [20].

Partiendo de un conjunto de N observaciones de una variable aleatoria X d-dimensional  $\{x_1, x_2, ..., x_N\}$ , este algoritmo divide el conjunto de datos en un número K de clústeres conocido. Considerando a  $\mu k$ , con k = 1, ..., K, como el conjunto de vectores d-dimensionales que representan los centros de los clústeres, los datos son asignados de manera tal que para cada uno la distancia a su respectivo centro, definida en (1), resulta mínima en comparación con las distancias al resto de los centros.

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \left\| x_n - \mu_k \right\|^2 \tag{1}$$

Aquí  $r_{nk}$  es un indicador binario que refiere a cuál de los K clústeres el objeto  $x_n$  es asignado. Suponiendo su asignación al clúster k, entonces  $r_{nk} = 1$  y  $r_{nj} = 0$  con  $j \neq k$ . El procedimiento es iterativo, y cada iteración implica dos pasos sucesivos correspondientes a minimizaciones sucesivas de J, en la primera fase con respecto a  $r_{nk}$  manteniendo  $\mu_k$  fijo (se requiere una partición inicial), y en la segunda fase con respecto a  $\mu_k$  manteniendo  $r_{nk}$  fijo. Esta optimización de dos etapas se repite hasta la convergencia [21], donde se obtienen los clústeres con sus respectivos clientes, los que representan los valores  $r_{nk}$  así como sus centroides, es decir, los valores  $\mu_k$ , cuyo comportamiento puede ser caracterizado para la variable bajo análisis.

Diferentes particiones iniciales pueden conducir a diferentes óptimos locales, aunque con datos bien estructurados es razonable esperar una convergencia al mismo óptimo, con suerte el global, en la mayoría de las configuraciones iniciales. Por lo tanto, es recomendable ejecutar el algoritmo de optimización varias veces con diferentes particiones iniciales.

Finalmente, y según se expone en la literatura especializada, las reglas para seleccionar el número de grupos son muy subjetivas porque hacen suposiciones sobre la estructura del grupo y solo funcionarían bien cuando se cumplan estos supuestos [21].

## Teoría de los conjuntos aproximados

La RST consiste básicamente en aproximar cualquier concepto, un subconjunto duro del dominio (ejemplo, un grupo resultante del proceso de agrupamiento), por un par de conjuntos exactos: La aproximación inferior y la superior. Al par ordenado U, A, conformado por el conjunto de atributos  $\{A=a_1, a_2, ..., a_m\}$  y el conjunto U llamado universo y descrito por los atributos, se le denomina sistema de información.

Los objetos que tienen la misma descripción son inseparables (similares). Esta relación de inseparabilidad induce una partición de U en bloques con estos objetos. Cualquier concepto k de U se puede expresar en términos de estos bloques de forma exacta o aproximada [22].

La extensión de la RST clásica acepta que objetos que no son inseparables, pero si suficientemente cercanos o similares puedan pertenecer a la misma clase. El objetivo es construir relaciones de similitud R' a partir de relaciones de inseparabilidad, relajando las condiciones iniciales de esta, siendo  $R'_{xn}$  la clase de similitud de  $x_n$ , es decir,  $R'(x_n) = \{x_1, x_2, ..., x_M\}$  $\in U$ :  $x_m$  está relacionado con  $x_n$  si y sólo si  $s(x_n)$  $(x_m) > \xi$ . Así, es posible calcular, para cada objeto agrupado, el conjunto de objetos relacionados con él, donde  $\xi$  es el umbral de similitud considerado y  $s(x_n, x_m)$  retorna un valor de similitud entre los objetos  $x_n$  e  $x_m$ . Para el cálculo de  $\xi$  existen varias expresiones [23], en el trabajo se seleccionó la media de las similitudes entre todos los posibles pares de objetos, descrita por (2):

$$\xi = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d(x_i, x_j)$$
 (2)

Donde  $d(x_i, x_j)$  es el valor de la distancia entre los vectores  $x_i$  y  $x_j$ , siendo la Euclidiana la elegida en este caso.

A partir de las relaciones de inseparabilidad, se definen los conceptos de aproximación inferior en (3):

$$R_I(k) = \left\{ x \in k : R'(x) \subseteq k \right\} \tag{3}$$

y de aproximación superior en (4):

$$R_s'(k) \bigcup_{x \in k} R'(x)$$
 (4)

Con el cálculo de las aproximaciones inferiores y superiores por grupos, se valida el agrupamiento en general y cada grupo en particular mediante la aplicación de medidas ofrecidas por la RST para evaluar los conceptos definidos sobre sistemas de información. Una medida que permite evaluar cada concepto es la precisión de la aproximación, definida por (5):

$$\alpha(k) = \frac{\left| \dot{R_I}(k) \right|}{\left| \dot{R_s}(k) \right|} \tag{5}$$

Donde · denota la cardinalidad del conjunto, finito y no vacío. Obviamente,  $0 \le \alpha(k) \le 1$ . Si  $\alpha(k) = 1$ , k es duro o exacto, si  $\alpha(k) < 1$ , k es aproximado, vago o inexacto.

La calidad de la aproximación, expuesta en (6), es otra medida que permite evaluar conceptos:

$$\gamma(k) = \frac{\left| R_I(k) \right|}{|k|} \tag{6}$$

Esta expresa el porcentaje de objetos que pueden ser correctamente asignados a k. Además  $0 \le \gamma(k) \le 1$ , y  $\gamma(k) = 0$  si  $\alpha(k) = 0$ .

La medición de la precisión y calidad de la aproximación considerando el sistema de información y los conceptos definidos sobre él, también puede ser ejecutada.

La calidad del agrupamiento, representada en (7), describe la inexactitud de los conceptos, expresando la proporción de los objetos que pueden estar correctamente asignados a los grupos en el sistema. Si es uno, el sistema de información según los conceptos definidos es consistente.

$$\Gamma = \frac{\sum_{k=1}^{K} \left| R_I(k) \right|}{\left| U \right|} \tag{7}$$

La precisión del agrupamiento, ver (8), expresa las posibles asignaciones correctas a grupos y señala la proporción entre la cantidad de objetos que pudieran estar bien agrupados y la cantidad de objetos que pudieran o no pertenecer a los grupos del sistema de información [23].

$$A = \frac{\sum_{k=1}^{K} |\vec{R_I}(k)|}{\sum_{k=1}^{K} |\vec{R_S}(k)|}$$
(8)

## Datos e implementación

Los datos iniciales corresponden a las lecturas de consumo registradas por los 1179 medidores inteligentes durante el período comprendido entre el lunes 4 de marzo y el jueves 4 de abril de 2019, con intervalos de registros de energía cada 15 minutos.

Previo al empleo del algoritmo de agrupamiento, resultó necesario un preprocesamiento de estos datos puesto que muchas lecturas no estaban registradas (valores ausentes). MATLAB® en estos casos las declara como NaN, por lo que estas lecturas fueron completadas por el valor medio del resto de las lecturas válidas para cada uno de los medidores, al igual que en el caso de los valores atípicos.

Se encontró que 39 consumidores tenían menos de 10 kWh en el mes, por lo que no se consideraron en la etapa de análisis, resultando el resto 1140 clientes.

Atendiendo a la subjetividad en la selección del número adecuado de grupos, en el trabajo se consideró en la implementación del algoritmo un rango de interés desde dos hasta seis clústeres, y, de acuerdo con los resultados de la RST, se eligió el número definitivo de grupos.

Por otro lado, siguiendo el enfoque que establece la ejecución del algoritmo con diferentes particiones iniciales, el procedimiento realizado en la investigación contempló 100 distintas particiones iniciales aleatorias (valor tomado arbitrariamente) y la selección de la solución con el valor más bajo para las sumas dentro del clúster de las distancias de cada objeto (cliente) a su correspondiente centroide.

La evaluación de los resultados del algoritmo se realizó empleando las medidas de calidad y precisión, tanto para cada agrupación ( $\gamma k$  y  $\alpha k$  respectivamente), como para el resultado global del agrupamiento ( $\Gamma$  y A respectivamente), definidas como parte de la RST.

# ANÁLISIS DE RESULTADOS

La Tabla 1 muestra los resultados para los cinco casos analizados: Desde seis hasta dos clústeres. El

umbral obtenido según los datos fue de 25,2573%. Valores mayores implican una mayor precisión y calidad, tanto local como globalmente, en los resultados del agrupamiento, mientras que la presencia de valores NaN se debe a la ausencia de aproximaciones inferiores y superiores dentro de un grupo en particular, lo que es símbolo de una distribución no óptima.

Tras estos resultados se concluye que dos clústeres son los que mejor representan el conjunto de datos. La distribución de los clientes en las dos agrupaciones resulta en 984 clientes conformando el clúster 1 y 156 el clúster 2.

Los gráficos resultantes para cada uno de los clústeres durante el período de medición se exponen en la Figura 2 y la Figura 3 respectivamente. La curva de color negro en ambos gráficos señala el centroide. En general, el clúster 1 presenta un menor consumo

Tabla 1. Clústeres de clientes equipados con medidores inteligentes.

mediaores intengentes.				
	$\alpha(k)$	$\gamma(k)$	A	Γ
	Seis clústeres			
1	0,2609	1	0,3892	0,9711
2	0,5385	0,3043		
3	0,711	0,9986		
4	NaN	0		
5	0,1643	0,8846		
6	0,1519	0,9277		
	Cinco clústeres			
1	NaN	0	0,4662	0,9728
2	0,3008	0,6607		
3	0,7632	0,9987		
4	0,2012	0,8293		
5	0,2505	1		
	Cuatro clústeres			
1	0,2677	0,9138	0,4497	0,9719
2	0,6	0,2222		
3	0,2575	1		
4	0,7283	0,9986		
	Tres clústeres			
1	0,5385	0,25	0,5251	0,9711
2	0,8534	0,9977		
3	0,2032	0,9556		
	Dos clústeres			
1	0,9265	0,999	0,7463	0,9728
2	0,2965	0,8077		

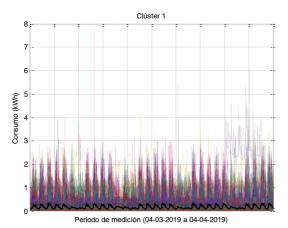


Figura 2. Centroide clúster 1 para el período de medición.

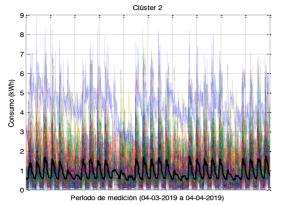


Figura 3. Centroide clúster 2 para el período de medición.

de energía con respecto al clúster 2 para el período de prueba.

En la Figura 2 se aprecia que existen algunos registros de consumo de energía que son extremadamente altos y que parecen escapar de los valores medios del conjunto de datos, así se puede identificar por ejemplo un registro cercano a los 8 kWh, otro de 6 kWh y dos registros de consumo con aproximadamente 5.5 kWh. Estos registros requieren ser analizados con mayor énfasis en trabajos futuros para lograr identificar su origen.

Por su parte, en la Figura 3 se percibe que los consumos del clúster 2 son mayores a los del clúster 1. Al igual que en la clasificación de los registros del clúster 1, se visualizan registros puntuales superiores y que varían entre 7 kWh y 9 kWh.

En la Figura 4 se muestran los perfiles de ambos centroides para observar en mayor detalle sus patrones de comportamiento.

La Figura 4 superior (curva de color rojo) resume el comportamiento del primer conjunto de registros cuyo centroide corresponde al clúster 1 y donde se puede observar que se tiene claramente identificable el comportamiento de un día hábil (lunes a viernes), un día feriado (viernes 15 de marzo) y los fines de semana. Se destaca que, tanto en el día feriado como en los fines de semana, los clientes residenciales agrupados en este clúster 1 reducen significativamente

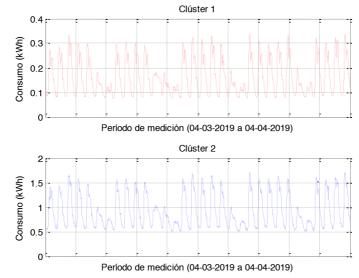


Figura 4. Gráfico de ambos centroides para el período de medición.

su consumo de energía en comparación al consumo energético de un día hábil. Los valores de consumo de energía para este centroide varían desde un mínimo aproximado de 0,09 kWh a un máximo de 0,34 kWh.

De manera similar, se puede notar en la Figura 4 inferior (curva de color azul) que el centroide del segundo grupo sigue un comportamiento similar al del clúster 1, identificando los días hábiles, el feriado y los fines de semana. Los valores de consumo mínimo y máximo varían significativamente, alcanzando en este caso un mínimo de 0,5 kWh y un máximo de 1,7 kWh.

Por tanto, de Figura 4 se puede concluir que los comportamientos de ambos clústeres son similares difiriendo sólo en las magnitudes de los consumos agrupados. Prueba de ello resulta el cálculo del coeficiente de correlación entre ambos centroides, cuyo valor es de 0,9876. La diferencia en magnitud admite su etiquetamiento bajo el concepto de pequeños y grandes consumidores. El diagrama de caja y bigote de la Figura 5 resume ambos grupos de consumidores estadísticamente, donde se comprueban las lecturas máximas y mínimas y se distingue el valor de la mediana de estos, así como la existencia de una mayor concentración de las lecturas en el caso del clúster 1 y un mayor

grado de dispersión entre el 50% y el 75% dentro del rango intercuartílico para ambos casos.

Este trabajo sólo consideró como dato de entrada la energía consumida en intervalos de 15 minutos para cada uno de los 1179 clientes residenciales. En estudios siguientes se pretende incorporar las variables de georreferenciación, datos meteorológicos y datos aportados por censo de población de modo de que la clasificación e identificación de clústeres aporte una mayor información del comportamiento de la demanda estableciendo así condiciones de entrada para políticas públicas del país.

## **CONCLUSIONES**

El presente estudio consideró sobre un millar de consumidores de energía eléctrica. Los datos provenientes de los medidores inteligentes fueron tratados con el algoritmo *K-means* implementado en MATLAB® para encontrar comportamientos similares de consumo.

La aplicación del algoritmo reveló que dos clústeres bien definidos logran la mejor representación de los clientes analizados. Este resultado es confiable basado en los valores de los indicadores de precisión y calidad proporcionados por la RST.

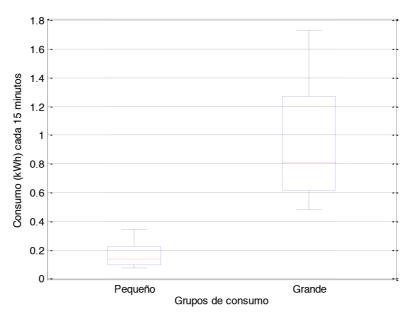


Figura 5. Diagrama de caja y bigote de los clústeres 1 y 2 etiquetados como pequeños y grandes consumidores respectivamente.

Los centroides presentan comportamientos similares para los días hábiles, el día feriado y los fines de semana, con una alta correlación de 98,76%, siendo diferentes solo en las magnitudes de los consumos energéticos. Este estudio de caso permitiría considerar decisiones de tarifas para incentivar cambios en los hábitos de consumo eléctrico.

El trabajo permitió además comprobar que la aplicación de tecnologías de análisis de datos permite generar información relevante para la toma de decisiones.

Como trabajo futuro se proyecta continuar con la clasificación de perfiles incorporando las variables de georreferenciación, datos meteorológicos y datos de censo de población para caracterizar con mayor precisión, mejorando la gestión de la demanda y estableciendo recomendaciones en políticas públicas y probadas en el campo energético.

### **AGRADECIMIENTOS**

Esta investigación ha sido apoyada por los proyectos internacionales *An ICT platform for sustainable energy ecosystem in smart cities ITCity* (ELAC2015/T10-0643) y Red CYTED Ciudades Inteligentes Totalmente Integrales, Eficientes y Sostenibles (CITIES).

# REFERENCIAS

- [1] B.Q. Yang, S.L. Liu, M. Gaterell and Y. Wang. "Smart metering and systems for low-energy households: Challenges, issues and benefits". Advances in Building Energy Research. Vol. 13 N° 1, pp. 80-100. 2017.
- [2] F. Ulloa, L. García-Santander, D. Carrizo and C. Hurtado. "Towards a home energy management model through a coordinator of smart sockets". Latvian Journal of Physics and Technical Sciences. Vol. 55 N° 4, pp. 35-43. 2018. DOI: 10.2478/lpts-2018-0027.
- [3] A. Kumar, S. Thakur and P. Bhattacharjee. "Real time monitoring of AMR enabled energy meter for AMI in smart city - an IoT application". IEEE 4th International Symposium on Smart Electronic Systems (ISES 2018). Hyderabad, India. pp. 219-222. 2018.
- [4] D. Yeolekar, H.H. Kulkarni, D.G. Bhardwaj. "Real time AMR & control of household

- energy meter with ZIGBEE communication". International Conference on Computing Communication Control and Automation (ICCUBEA). Pune, India. pp. 172-176. 2015.
- [5] T. Schwartz, G. Stevens, T. Jakobi, S. Denef, L. Ramirez, V. Wulf and D. Randall. "What people do with consumption feedback: A long-term living lab study of a home energy management system". Interacting with Computers. Vol. 27 Issue 6, pp. 551-576. 2015.
- [6] K. Ahuja and A. Khosla. "Importance of Smart Metering in Diverse Fields of Smart Cities". International Journal of Grid and Distributed Computing. Vol. 11 N° 3, pp. 33-49. 2018.
- [7] K. Hopf, M. Sodenkamp and T. Staake. "Enhancing energy efficiency in the residential sector with smart meter data analytics". Electronic Markets. Vol. 28 N° 4, pp. 453-473. 2018.
- [8] J. Yang, Y. Yang, J. Chen, L. Fu and Z. He. "Determining the harmonic contributions of multiply harmonic sources using data clustering analysis". IEEE Int. Symposium on Smart Electric Distribution Systems and Technologies (EDST). Vienna, Austria. pp. 405-411. 2015.
- [9] M.D. Khan and V.S. Patel. "Knowledge and approximations: A formal study under the perspective of information systems and rough set theory". Information Sciences. Vol. 524, pp. 97-115. 2020. DOI: 10.1016/j. ins.2020.03.017.
- [10] M. Fahim and A. Sillitti. "Analyzing load profiles of energy consumption to infer household characteristics using smart meters". Energies. Vol. 12 N° 5. 2019.DOI: 10.3390/ en12050773.
- [11] H.S. Liang, J. Ma, R.F. Sun and Y.L. Du. "A data-driven approach for targeting residential customers for energy efficiency programs". IEEE Transactions on Smart Grid. Vol. 11 N° 2, pp. 1229-1238. 2020. DOI: 10.1109/TSG.2019.2933704.
- [12] A. Rajabi, M. Eskandari, M.J. Ghadi, L. Li, J.F. Zhang and P. Siano. "A comparative study of clustering techniques for electrical load pattern segmentation". Renewable & Sustainable Energy Reviews. Vol. 120. 2020. DOI: 10.1016/j.rser.2019.109628.
- [13] A. Ullah, K. Haydarov and H. Ui. "Deep Learning Assisted Buildings Energy

- Consumption Profiling Using Smart Meter Data". Sensors. Vol. 20 N° 3. 873. 2020.
- [14] J. Sala, R. Li and M. Christensen. "Clustering and classification of energy meter data: A comparison analysis of data from individual homes and the aggregated data from multiple homes". Building Simulation. 2019. DOI: 10.1007/s12273-019-0587-4.
- [15] J. Zhu, Y. Shen, S. Yingjun and Z. Song. "Data-driven building load profiling and energy management". Sustainable Cities and Society. Vol. 49. 101587. 2019.
- [16] S. Yilmaz, J. Chambers and M.K. Patel. "Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management". Energy. Vol. 180, pp. 665-677, 2019.
- [17] R. Markovic, M. Gosak, V. Grubelnik, M. Marhl and P. Virtič. "Data-driven classification of residential energy consumption patterns by means of functional connectivity networks". Applied Energy. Vol. 242. pp. 506-515. 2019.

- [18] T. Cerquitelli, G. Chicco and E. Di Corso. "Electricity Consumption Over Time for Residential Consumers Through Cluster Analysis". 14th International Conference on Development and Application Systems (DAS). Suceava, Romania. pp. 164-169. 2018.
- [19] J.F. Hair, W.C. Black, B.J. Babin and R.E. Anderson. "Multivariate Data Analysis". Pearson. 7th edition. 2014.
- [20] B.S. Everitt, S. Landau, M. Leese and D. Stahl. "Cluster analysis". Wiley. 5th edition.
- [21] C.M. Bishop. "Pattern Recognition and Machine Learning". Springer Science + Business Media, LLC. 2006.
- [22] J. Komorowski, Z. Pawlak, and L. Polkowski. "Rough-Fuzzy Hybridization: A New Trend in Decision Making". Springer - Verlag, 1999.
- [23] L. Arco. "Agrupamiento basado en la intermediación diferencial y su valoración utilizando la teoría de los conjuntos aproximados". PhD. Dissertation. Univ. Central de Las Villas, Cuba. 2008.