

UNIVERSIDAD INTERNACIONAL SEK

FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

Trabajo de Fin de Carrera Titulado:

**“AGRUPACIÓN POR CARACTERÍSTICAS DE CONSUMO ELÉCTRICO DE
PARROQUIAS DE LA PROVINCIA DE PICHINCHA ECUADOR MEDIANTE EL
ALGORITMO K-MEANS”**

Realizado por:

OSCAR FERNANDO NÚÑEZ BARRIONUEVO

Director del proyecto:

LLANES CEDEÑO EDILBERTO ANTONIO, Ph.D.

Como requisito para la obtención del título de:

**MASTER EN ECOEFICIENCIA INDUSTRIAL MENCIÓN EFICIENCIA
ENERGÉTICA**

Quito, 14 de octubre de 2020

DECLARACIÓN JURAMENTADA

Yo, OSCAR FERNANDO NÚÑEZ BARRIONUEVO, con cédula de identidad # 171685861-6, declaro bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración, cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la UNIVERSIDAD INTERNACIONAL SEK, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.



Oscar Fernando Núñez Barrionuevo
C.I.: 1716858616

DECLARATORIA

El presente trabajo de investigación titulado:

**“AGRUPACIÓN POR CARACTERÍSTICAS DE CONSUMO ELÉCTRICO DE
PARROQUIAS DE LA PROVINCIA DE PICHINCHA ECUADOR MEDIANTE EL
ALGORITMO K-MEANS”**

Realizado por:

OSCAR FERNANDO NÚÑEZ BARRIONUEVO

como Requisito para la Obtención del Título de:

MATER EN ECOEFICIENCIA INDUSTRIAL MENCIÓN EFICIENCIA ENERGÉTICA

ha sido dirigido por el profesor

Ph.D, LLANES CEDEÑO EDILBERTO ANTONIO

quien considera que constituye un trabajo original de su autor



Llanes Cedeño Edilberto Antonio

DIRECTOR

LOS PROFESORES INFORMANTES

Los Profesores Informantes:

JAVIER MARTÍNEZ

JESÚS LÓPEZ

Después de revisar el trabajo presentado,
lo han calificado como apto para su defensa oral ante
el tribunal examinador



Javier Martínez



Jesús López

Quito, 14 de OCTUBRE de 2020

**“Agrupación por características de consumo eléctrico de parroquias de la provincia de
Pichincha Ecuador mediante el algoritmo K-means”**

**“Clustering analysis of electricity consumption of municipalities in the province of
Pichincha-Ecuador using the K-means algorithm”**

Oscar F. Núñez-Barrionuevo, Edilberto A. Llanes-Cedeño,

Javier Martinez-Gomez, Jorge I. Guachimboza-Davalos

Jesús López-Villada

Universidad Internacional SEK, Quito-Ecuador

ofnunez.mee@uisek.edu.ec

Abstract. El presente trabajo muestra un análisis de consumo eléctrico por parroquias de la provincia de Pichincha para definir afinidades de consumo entre ellas. Para ello, se realiza una codificación de variables categóricas para ser usadas con el algoritmo de aprendizaje de máquina no supervisado k-means. Como resultados relevantes se establecen cuatro tipos de consumos y se analizan en relación a la densidad poblacional mientras que el algoritmo representa el 100% de la variabilidad de los datos, y se identifican municipios con diferentes tendencias de consumo.

Keywords: Consumo de energía _ agrupamiento _ k-means

Abstract. This work shows an analysis of electricity consumption by parishes in the province of Pichincha to define consumption affinities between them. To do this, categorical variables are encoded to be used with the unsupervised machine learning algorithm k-means. As relevant results, four types of consumption are established and they are analyzed in relation to population density while algorithm represents 100 % of the variability of the data, and municipalities with different consumption trends are identified.

Keywords: power consumption _ clustering _ k-means

1 Introducción

El calentamiento global se ha convertido en una de las mayores preocupaciones de los gobiernos de todo el mundo, por lo que cada vez se suman esfuerzos para contrarrestar sus efectos y mitigar su impacto en la población. En tal virtud, se han realizado numerosos estudios donde se ha evidenciado que los gases contaminantes generados por las industrias, el transporte y muchos de los hábitos de la población han sido los causantes de este fenómeno. Eventos adversos como incendios, inundaciones y sequías son solo una parte de los efectos del cambio climático en el planeta [1]. Una de las principales causas son las emisiones de gases contaminantes, entre ellos el CO₂. Estas emisiones deben ser controladas de manera inmediata, antes que las consecuencias sean irreversibles para la población, por lo que tanto la emisión de políticas públicas como la búsqueda de nuevas fuentes de energía renovables deben ser propuestas por los gobiernos para su aplicación obligatoria en los ámbitos de la industria [2].

Además, estos efectos del calentamiento global se ven relacionados con la modificación de consumo eléctrico de la población el sufrir temperaturas extremas, ocasionando un aumento en el uso de sistemas de calefacción [3]. Por ello, una de las formas de actuar para limitar e impedir sus consecuencias ambientales, sociales y económicas, relacionadas con el aumento de temperatura, consiste en reducir el consumo energético. Ya que el modelo de generación, transporte y consumo actual, es absolutamente dependiente de los combustibles fósiles. De este modo es insostenible como consecuencia del cambio climático que supone.

En América Latina, se cuenta con un promedio de 0.65 toneladas de CO₂ por cada MWh generado [4][5]. En el Ecuador, este resultado es de 397.5g de CO₂ por KWh, donde el mayor valor corresponde al año 2010, ya que la generación de energía de fuentes no renovables fue del (52.2%) [6]. No obstante, con el cambio de matriz productiva, el 60.85% de producción eléctrica es a base de energía renovable donde el 58.53% está relacionado a fuentes

Hidroeléctricas [7] . Esta producción está orientada en un 30.93% al consumo residencial y a un 26.01% al sector Industrial [6]. Sin embargo, la generación de electricidad por medio de las centrales eléctricas ocasiona cambios en los ecosistemas, desaparición de especies, erosión del suelo, entre otros. Por tal motivo, es necesario realizar un proceso de reconocimiento de parámetros de consumo que permitan descubrir los hábitos de las personas. Este proceso se realiza para contar con información hacia todos los participantes del mercado a cooperar de forma efectiva en el ahorro de consumo eléctrico. Además, brinda información a las centrales de generación de energía para realizar una adecuada planificación y evitar la sobrecarga en sus equipos [8].

Con el constante incremento de la demanda eléctrica domiciliar el análisis del consumo de energía eléctrica (EEC) es cada vez más importante. Por ello, la adquisición y análisis de datos se ha convertido en una herramienta muy usada en estos días ya que permite determinar el comportamiento de EEC de una forma precisa. En este sentido, el uso de etiquetas (identificadores) de usuario y la tecnología adecuada puede aportar una expresión más intuitiva y concisa al análisis de EEC en comparación con el análisis tradicional [9]. Es por ello, que el uso de algoritmos de aprendizaje de máquina permite reconocer patrones más relevantes del fenómeno estudiando a base de cálculos matemáticos complejos que busca emular cierta funcionalidad del cerebro humano como lo es la agrupación en conjuntos a objetos en relación a sus características más evidentes.

En este sentido, al contar con información del consumo eléctrico por cantones y parroquias, el criterio de agrupación permite organizarlos de tal forma que se formen subgrupos con atributos similares entre sí, pero distintos a otros. Este proceso se enmarca dentro del aprendizaje no supervisado, ya que solo se dispone de un conjunto de datos de entrada que debemos obtener información sobre la estructura del dominio de salida, la misma que no se

dispone [10]. Con ello, se pueden establecer un EEC por sectores del país y definir políticas de consumo eléctrico. Sin embargo, el proceso de adquisición de datos puede ser una desafiante tarea. Esto se debe a que la mayoría de bases de datos gubernamentales cuentan información mayormente categórica. En consecuencia, limita significativamente el análisis de datos [11]. Para ello, existen técnicas que permiten codificar variables para convertirlas a un sistema numérico. Sin embargo, es necesario tener cautela en el uso de estos atributos ya que pueden sobredimensionar al algoritmo de aprendizaje de máquina y brindar información errónea al momento de generar el modelo [6].

Por esta razón, el presente trabajo expone un desarrollo de clusters dentro del consumo energético del Ecuador en la provincia de Pichincha para agruparlos a zonas poblaciones con tendencias similares de consumo relacionados a su cantón y parroquia. Para ello, se realiza una etapa de pre-procesamiento de datos para codificar las variables categóricas y eliminar las que no presenten información relevante al algoritmo de aprendizaje automático. Posteriormente, se emplean criterios técnicos para determinar el número de grupos a realizar y finalmente establecer tendencias de consumo eléctrico. Como resultado, este estudio presenta un aceptación del 100% que explica la variabilidad de cada grupo.

El resto del documento se encuentra estructurado de la siguiente manera: La sección 2, se presenta los trabajos relacionados. La sección 3 indica el esquema metodológico de análisis de datos. La sección 4 presenta el análisis clusters para brindar el agrupamiento del consumo eléctrico. La sección 5 se muestra los resultados obtenidos con los diferentes análisis de error. Finalmente, en la sección 6 se presentan los resultados y trabajos futuros.

2 Trabajos Relacionados

Trabajos como [8, 9, 10] han presentado importantes aportes relacionados al agrupamiento en el comportamiento de consumo de electricidad de los clientes residenciales, afín de comprender las demandas personalizadas del usuario y proporcionarles servicios específicos. No obstante, los patrones de consumo pueden ser significativamente diferentes en relación a la ubicación geográfica, densidad poblacional, hábitos de consumo, entre otros. Es por ello que existen problemáticas pendientes, orientado principalmente en la población ecuatoriana, ya que existe una carencia de estos estudios y sobre todo, no usan algoritmos de aprendizaje de máquina.

3 Materiales y Métodos

La presente sección muestra por un lado, la adquisición de datos (sección 3.1), normalización (sección 3.2), el algoritmo de agrupamiento seleccionado (sección 3.3) y el esquema de análisis de datos (sección 3.4).

3.1 Adquisición de datos

La recopilación de datos es mediante la información presentada por el ministerio de Electricidad y Energía renovable en sus rendiciones de cuentas mensuales durante el año 2019. En estos reportes explica detalladamente la empresa que distribuye la electricidad a las provincias, cantones y parroquias. En este caso específico, se toma la información del consumo de los años 2018 y 2019. Todos los datos son almacenados en la matriz $Y \in \mathbb{R}^{m \times n}$ donde m representa el número de muestras y n las atributos de la base de datos. En este caso, $m = 2550$ y $n = 11$. Los atributos hace referencia a: Año de consumo, mes, empresa que factura, provincia, cantón, parroquia, tipo de equipamiento de voltaje a 220, número de clientes, energía facturada,

incremento de consumo, consumo residencial, energía de subsidio generado, facturación del servicio y valor del subsidio.

3.2 Normalización de la base de datos

La normalización de índices significa ajustar los valores medidos en diferentes escalas respecto a una en común. Este proceso se realiza previo a realizar modelos matemáticos [7]. Con ello, se evita que existan variables que según su la naturaleza del dato, aporte en gran cantidad ponderada al modelo al tener valores muy altos en relación al resto. Además, evita que el uso de variables categóricas se vea mermado debido a su baja escala (0 a 1). Es por ello que existen diferentes métodos y formas de normalización. En este caso se usa la normalización estándar dada por la siguiente formula:

$$\frac{X - \mu}{\delta}$$

3.3 Agrupación (Clustering)

Uno de los retos computacionales importantes es tener la capacidad de reconocer características para agrupar a elementos similares. El método de K-means tiene como objetivo la partición de un conjunto de n observaciones en k grupos. Donde cada valor de n pertenece a un grupo de k cuyo valor medio de la distancia sea el más cercano. Para realizar este proceso se establece el siguiente proceso:

Algorithm 1 K-means pseudo-code

Input: Dataset

Output: each m assigned to a group k

1: Choose the number of K fo clusters

2: Select at random K points the centroids

3: Assign each data point to the closest centroid (That forms k clusters)

4: Compute and place the new centroid of each cluster

5: Reassign each data point to the new closest centroid

```

6: if any reassignment took place then
7: go to step 4
8: else
9: go to Fin
10: return: Each n on k cluster

```

No obstante, la aleatoriedad de los valores de k puede ocasionar diferentes formas de agrupación. Por ello, es necesario definir adecuadamente su valor con el objetivo de agrupar a solo los datos que contengan la mayor cantidad de atributos similares [12]. En consecuencia, la versión de K-means ++ permite eliminar este problema mediante el análisis de un conjunto de observaciones $(x_1, x_2, \dots, x_p)$, donde cada observación es un vector real d -dimensional, la agrupación de K-means tiene como objetivo dividir las p observaciones en k ($k \leq p$) conjuntos $S = S_1, S_2, \dots, S_k$ para minimizar la suma de cuadrados dentro del grupo (WCSS) (es decir, la varianza). Esto se puede apreciar en la siguiente ecuación:

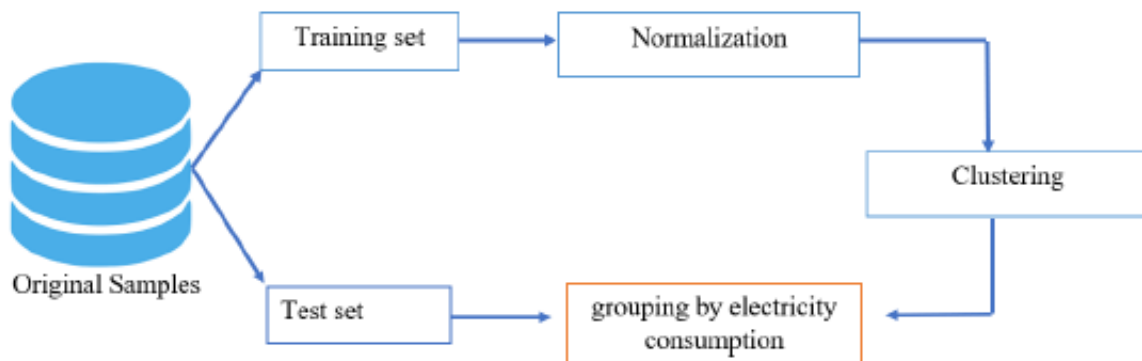
$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mu_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i| \text{Var } S_i$$

Donde μ_i es el promedio de los puntos en S_i .

3.4 Esquema de análisis de datos

El presente trabajo se enfoca en agrupar a las parroquias de la provincia de Pichincha por hábitos de consumo eléctrico. Para ello es necesario un modelo de análisis de datos, el mismo que se muestra en la Fig.1

Fig. 1. Esquema de análisis de datos propuesto



4 Resultados

Se realiza el análisis (WCSS) con un valor de $n = 10$ y determinar el valor adecuado de k . Según la varianza, se determina que $k = 4$ es idóneo en relación a los atributos de la base de datos Y. Esto se puede apreciar en la Fig. 2.

Para conocer el porcentaje de aceptación del modelo en relación a la distribución de los datos. Se realizó el análisis de varianza. Esto junto con la implementación del modelo WCSS. Como resultado, hay una distribución de los datos al modelo con una aceptación del 99,9%. El análisis multidimensional de varianza se puede ver en la Fig.3.

La implementación del algoritmo permite agrupar por cantones en relación al consumo en la provincia de Pichincha-Ecuador. Como principal información, se establecen los valores de los centroides de cada clúster. Donde el consumo eléctrico cuenta con los valores de: Consumo excesivo: 7037116.0 Kwh, alto: 3458488.1, promedio: 1984133.6 y bajo: 295444.3. Con el objetivo de observar gráficamente la organización de los clúster se realiza una reducción de dimensiones por medio del algoritmo Análisis de Componentes Principales (PCA) Fig. 4.

Fig. 2. Análisis WCSS para determinar el valor de k

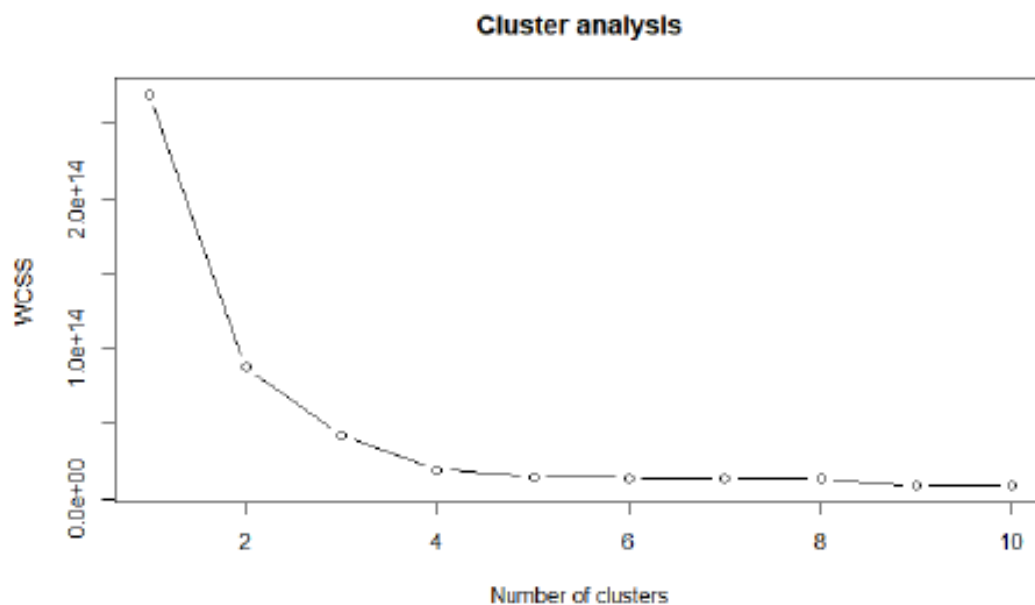


Fig. 3. Análisis de varianza de datos para representar el ajuste del modelo.

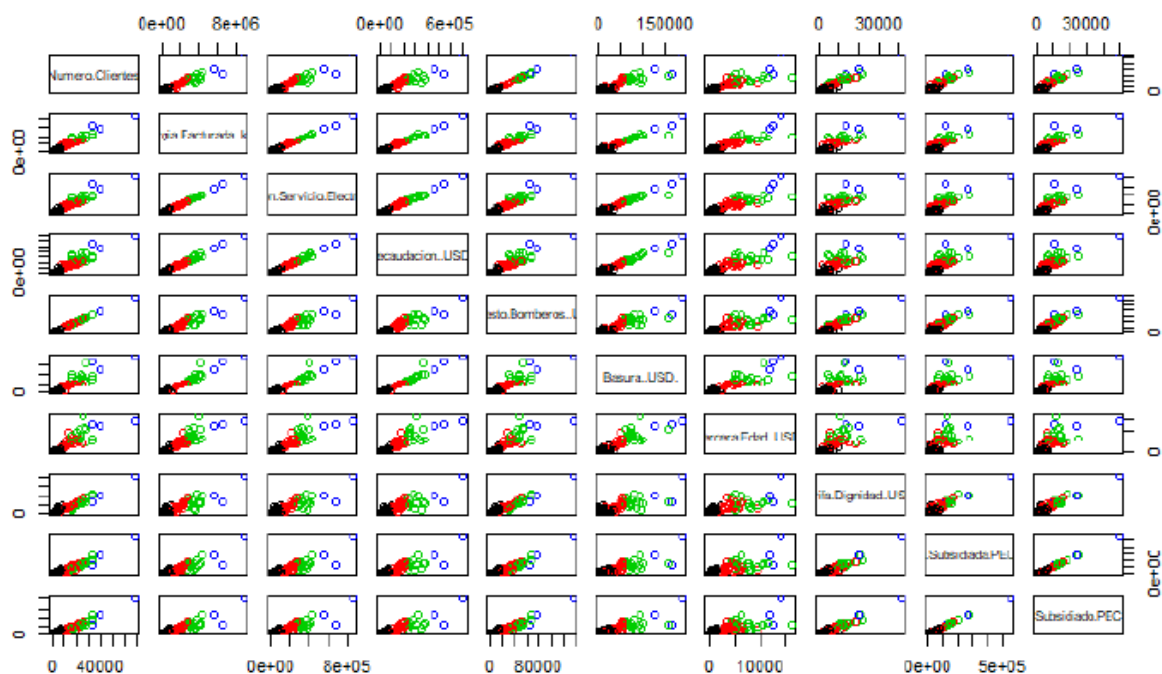
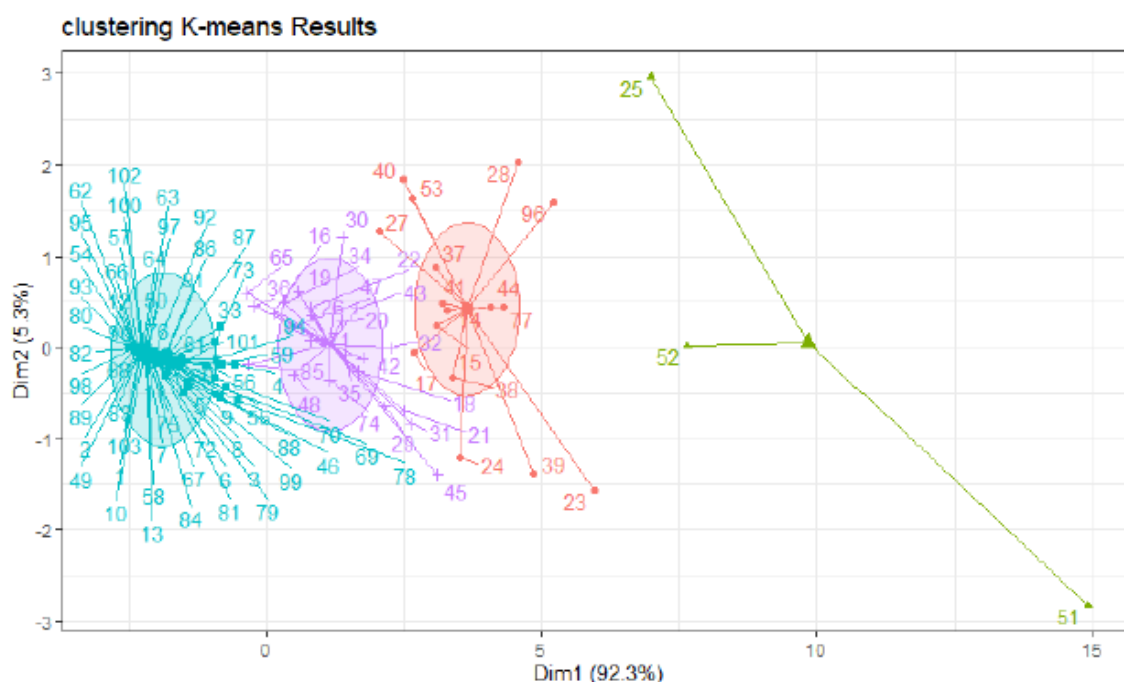


Fig. 4. Resultado de consumo en Pichincha por K means.



Con el algoritmo de aprendizaje automático desarrollado, los municipios de cada cantón se agrupan en relación al tipo de consumo previamente establecido. Se puede hacer un resumen por cantones de la provincia de Pichincha, esto se puede ver en la tabla 1.

El Distrito Metropolitano de Quito cuenta con el mayor consumo eléctrico y la gran variabilidad en relación a sus cantones. Geográficamente se pueden establecer la forma de consumo con colores: Consumo excesivo (rojo), alto (tomate), promedio (amarillo) y bajo (verde). Esto se puede apreciar en la Fig. 5. Se puede observar que las parroquias de Calderón, Conocoto e Iñaquito presentan un mayor consumo en relación al resto, esto por la alta densidad poblacional muy superior al resto. Sin embargo, Iñaquito presenta una densidad poblacional promedio pero con un alto consumo eléctrico. Por otra parte, parroquias como El Condado o San Antonio cuenta una alta densidad poblacional pero no presenta un consumo eléctrico excesivo.

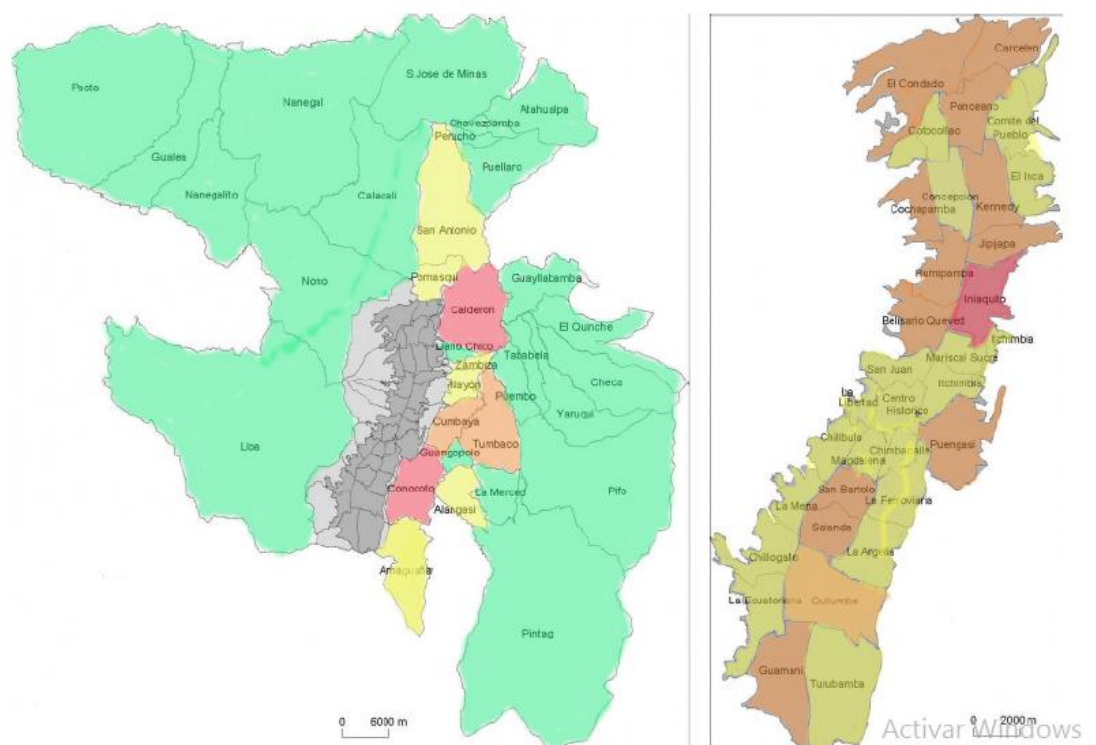
AGRUPACIÓN POR CARACTERÍSTICAS DE CONSUMO ELÉCTRICO DE PARROQUIAS DE LA PROVINCIA DE PICHINCHA ECUADOR MEDIANTE EL ALGORITMO K-MEANS

16

Tabla 1. Resumen de agrupación de municipios por tipo de consumo eléctrico

Power consumption clustering	Municipalities
EXCESSIVE	INAQUITO(25), CALDERON (CARAPUNGO)(51), CONOCOTO(52)
HIGH	BELISARIO QUEVEDO(14), CARCELÉN(15), COCHAPAMBA(17), EL CONDADO(23), GUAMANI(24), JIPIJAPA(27), KENNEDY(28), PONCEANO(37), PUENGASÍ(38), QUITUMBE(39), RUMIPAMBA(40), SAN BARTOLO(41), SOLANDA(44), CUMBAYA(53), TUMBACO(77), SANGOLQUI(96)
MODERATE	CENTRO HISTÓRICO(16), COMITÉ DEL PUEBLO(18), COTOCOLLAO(19), CHILIBULO(20), CHILLOGALLO(21), CHIMBACALLE(22), ITCHIMBIA(26), LA ARGELIA(29), LA CONCEPCIÓN(30), LA ECUATORIANA(31), LA FERROVIARIA(32), LA MAGDALENA(34), LA MENA(35), MARISCAL, SUCRE(36), SAN ISIDRO DEL INCA(42), SAN JUAN(43), TURUBAMBA(45), ALANGASI(47), AMAGUAÑA(48), NAYON(65), POMASQUI(71), SAN ANTONIO(74), MACHACHI(85)
LOW	TANDAPI(1), SAN MIGUEL DE LOS BANCOS(2), PUERTO QUITO(3), CAYAMBE(4), JUAN MONTALVO(5), CANGAHUA(6), OLMEDO (PESILLO)(7), SAN JOSE DE AYORA(8), TABACUNDO(9), LA ESPERANZA(10), MALCHINGUI(11), TOCACHI(12), TUPIGACHI(13), LA LIBERTAD(33), QUITO(46), ATAHUALPA (HABASPAMBA)(49), CALACALI(50), CHAVEZPAMBA(54), CHECA (CHILPA)(55), EL QUINCHE(56), GUALEA(57), GUANGOPOLO(58), GUAYLLABAMBA(59), LA MERCED(60), LLANO CHICO(61), LLOA(62), NANEGAL(63), NANEGALITO(64), NONO(66), PACTO(67), PERUCHO(68), PIFO(69), PINTAG(70), PUELLARO(72), PUEMBO(73), SAN JOSE DE MINAS(75), TABABELA(76), YARUQUI(78), ZAMBIZA(79), CAYAMBE(80), ASCAZUBI(81), CANGAHUA(82), OTON(83), SANTA ROSA DE CUZUBAMBA(84), ALOAG(86), ALOASI(87), CUTUGLAHUA(88), EL CHAUPÍ(89), SAN RAFAEL(95), (TANDAPI)(90), TAMBILLO(91), UYUMBICHO(92), MALCHINGUI(93), SANGOLQUI(94), COTOGCHOA(97), RUMIPAMBA(98), SAN MIGUEL DE LOS BANCOS(99), PEDRO VICENTE MALDONADO(101), PUERTO QUITO(102), MINDO(100)

Fig. 5. Distribución por conglomerados de municipios de la provincia de Pichincha. Exceso de consumo (rojo), alto (tomate), medio (amarillo) y bajo (verde)



5 Conclusiones y trabajos futuros

Este proceso de agrupación por consumo eléctrico en contraste con la densidad poblacional se puede deducir que parroquias con 15 mil habitantes tiene un consumo superior a los 17000 Kwh por mes. No obstante algunas de ellas no presentan esta tendencia similar y es necesario realizar un análisis más afondo sobre las características de consumo eléctrico.

Se puede apreciar en la gráfica que sectores con similares ubicaciones geográficas y densidad poblacional pueden cambiar los hábitos de consumo eléctrico. Esto se puede apreciar específicamente en los conjuntos de alto y medio consumo. Con esto, se pueden presentar mejores políticas gubernamentales para una mejor planificación de producción eléctrica.

Como trabajos futuros, se propone hacer uso de una herramienta de soporte de toma de decisión para contar con una interfaz adecuada sobre los comportamientos de consumo con reportes mensuales y anuales.

Bibliografía

1. X. Huang and S. Wang, "Prediction of bottom-hole flow pressure in coalbed gas wells based on GA optimization SVM," in Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018. Institute of Electrical and Electronics Engineers Inc., dec 2018, pp. 138-141.
2. L. He, Q. Song, and J. Shen, "K-NN numeric prediction using bagging and instance relevant combination," in Proceedings - 2nd International Symposium on Data, Privacy, and E-Commerce, ISDPE 2010, 2010, pp. 3-8.
3. S. Bose, A. Goel, T. Shankar, R. Mageshvaran, and A. Rajesh, "Energy efficient heterogeneous network with daily load variation," in 2017 Innovations in Power and Advanced Computing Technologies, i-PACT 2017, vol. 2017-January. Institute of Electrical and Electronics Engineers Inc., jan 2018, pp. 1-4.
4. X. M. Zhang, K. Grolinger, M. A. Capretz, and L. Seewald, "Forecasting Residential Energy Consumption: Single Household Perspective," in Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018. Institute of Electrical and Electronics Engineers Inc., jan 2019, pp. 110-117.
5. L. Diao, Y. Sun, Z. Chen, and J. Chen, "Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation," Energy and Buildings, vol. 147, pp. 47-66, jul 2017.
6. "Ministerio de Electricidad y Energía Renovable. Ente rector del Sector Eléctrico Ecuatoriano." [Online]. Available: <http://historico.energia.gob.ec/>
7. R. Parra Narváez, "Factor de emisión de CO₂ debido a la generación de electricidad en el Ecuador durante el periodo 2001-2014," Avances en

Ciencias e Ingeniería, vol. 7, no. 2, dec 2015.

8. Y. Wang, Z. Chen, Z. Xu, G. Gang, and J. Lu, "User electricity consumption pattern optimal clustering method for smart grid," in 2018 14th IEEE International Conference on Signal Processing (ICSP), Aug 2018, pp. 567-570.
9. C. Zhong, J. Shao, F. Zheng, K. Zhang, H. Lv, and K. Li, "Research on electricity consumption behavior of electric power users based on tag technology and clustering algorithm," in 2018 5th International Conference on Information Science and Control Engineering (ICISCE), July 2018, pp. 459-462.
10. Li Kangping, Wang Fei, Zhen Zhao, Mi Zengqiang, Sun Hongbin, Liu Chun, Wang Bo, and Lu Jing, "Analysis on residential electricity consumption behavior using improved k-means based on simulated annealing algorithm," in 2016 IEEE Power and Energy Conference at Illinois (PECI), Feb 2016, pp. 1-6.
11. A. Toapanta-Lema, W. Gallegos, J. Rubio-Aguilar, E. Llanes-Cedeño, J. Carrascal-García, L. García-López, and P. D. Rosero-Montalvo, "Regression models comparison for efficiency in electricity consumption in ecuadorian schools: A case of study," in Applied Technologies, M. Botto-Tobar, M. Zambrano Vizuite, P. Torres-Carrion, S. Montes León, G. Pizarro Vásquez, and B. Durakovic, Eds. Cham: Springer International Publishing, 2020, pp. 363-371.
12. Sohrab Rezaei, A. Sharghi, and G. Motalebi, "A framework for analysis affecting behavioral factors of residential buildings' occupant in energy consumption," journal of Sustainable Architecture and Urban Design, vol. 5, no. 2, pp. 39-58, 2018.