

**“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA
PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS
ENERGÉTICOS”**

UNIVERSIDAD INTERNACIONAL SEK

**FACULTAD DE CIENCIAS NATURALES Y
AMBIENTALES**

Trabajo de Fin de Carrera Titulado:

**“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA
PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS
ENERGÉTICOS”**

Realizado por:

JHONATAN FABRICIO ROSERO GARCÍA

Director del proyecto:

Dr. Jesús López Villada.

Como requisito para la obtención del título de:

**MAGISTER EN ECOEFICIENCIA
INDUSTRIAL MENCIÓN EFICIENCIA
ENERGÉTICA**

Quito, abril de 2021

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

DECLARACIÓN JURAMENTADA

Yo, JHONATAN FABRICIO ROSERO GARCÍA, con cédula de identidad # 040099949-6, declaro bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado a 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.



FIRMA

040099949-6

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

DECLARATORIA

El presente trabajo de investigación titulado:

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Realizado por:

JHONATAN FABRICIO ROSERO GARCÍA

como Requisito para la Obtención del Título de:

**MAGISTER EN ECOEFICIENCIA INDUSTRIAL
MENCIÓN EFICIENCIA ENERGÉTICA**

ha sido dirigido por el profesor

DR. JESÚS LOPEZ VILLADA

quien considera que constituye un trabajo original de
su autor

A handwritten signature in blue ink, appearing to be 'JL', is centered on the page. The signature is stylized and somewhat abstract.

FIRMA

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

LOS PROFESORES INFORMANTES

Los Profesores Informantes:



EDILBERTO LLANES



GUSTAVO MORENO

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

FIRMA

FIRMA

Quito, abril de 2021

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

DEDICATORIA

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

AGRADECIMIENTO

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Para someter a:

To be submitted:

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Jhonatan F. Rosero – García ¹, Edilberto A. LLanes – Cedeño¹, Ricardo P. Arciniega – Rocha ², y Jesús López – Villada¹

¹ Universidad Internacional SEK, Facultad de Ciencias Naturales y Ambientales, Quito,
Ecuador.

² Instituto Superior Tecnológico 17 de Julio, Ibarra
Ecuador.

***AUTOR DE CORRESPONDENCIA:**

Teléfono: 0983965306; email: jhonatanroserogarcia@gmail.com

Título corto o Running title: Análisis de predicción y agrupación de consumo eléctrico.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Resumen.

La planificación eléctrica de un país es una creciente necesidad para su desarrollo económico. Sin embargo, este análisis es complejo de realizar debido a los diferentes hábitos de consumo de los participantes de este mercado. En este sentido, en Imbabura-Ecuador, el consumo eléctrico residencial es mayor que la capacidad de producción de esta provincia. Por esta razón, se presenta un análisis de predicción y agrupamiento por municipalidades mediante algoritmos de aprendizaje automático con el fin de determinar tendencias de consumo y presentar reportes para una adecuada planificación eléctrica. Como resultados relevantes los modelos de máquinas de soporte de decisión y bosques aleatorios demostraron ser los adecuados para esta tarea con error de predicción menor del 10 %. Por su parte, el algoritmo k-means pudo agrupar en 4 tipos de consumo eléctrico con una representación del 98 % de la variabilidad de los datos.

Palabras clave: Consumo Eléctrico, análisis eléctrico, modelos de regresión, agrupación.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Abstract.

The electrical planning of a country is a growing need for its economic development. However, this analysis is complex to carry out due to the different consumption habits of the participants in this market. In this sense, in Imbabura-Ecuador, residential electricity consumption is the one that needs the greatest electricity demand. For this reason, a prediction and grouping analysis by municipalities is presented using machine learning algorithms in order to determine consumer trends and present reports for proper electrical planning. As relevant results, the models of decision support machines and random forests proved to be suitable for this task with a prediction error of less than 10 %. For its part, the k-means algorithm was able to group 4 types of electricity consumption with a representation of 98 % of the data variability.

Keywords: Electrical consumption, electrical analysis, regression models, clusterins.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Introducción.

Uno de los principales componentes del desarrollo económico de un país es su capacidad de producción eléctrica de forma confiable y al menor costo posible. Para esto, es necesario determinar el comportamiento actual y futuro de la demanda eléctrica para una adecuada planificación que se debe considerar variables políticas, económicas, sociales, ambientales y tecnológicas (Sauhats, Varfolomejeva, Lmkevics, Petrecenko, Kunickis y Balodis, 2015). No obstante, estos análisis deben ser asumidos por entes gubernamentales y el sector privado con el objetivo de implementar programas y planes que tengan impactos directos sobre la demanda eléctrica. Esto debe estar muy relacionado con la demografía de cada país en conjunto con los diferentes recursos naturales renovables o no renovables a ser empleados como fuentes de producción de electricidad (Dmitri, María y Anna, 2016).

En este sentido, la previsión del consumo eléctrico permite la cooperación de todos los participantes de este mercado para cooperar entre ellos efectivamente.

De esta forma, el sector industrial puede obtener advertencias sobre su excesivo consumo y prevenir a las empresas de suministro eléctrico que realicen la producción de electricidad de manera óptima, al evitar la sobre carga de los equipos de generación (Ye, Wu y Guo, 2018). Por su parte, el sector residencial, el análisis de demanda eléctrica está directamente relacionada a las ampliaciones de sectores urbanos dentro de una ciudad. Con esto, la planificación sobre la necesidad de la implementación de transformadores e instalación de postera se vuelve una tarea tecnificada. No obstante, este proceso de pronóstico eléctrico, es una tarea complicada de realizar por la gran cantidad de datos obtenidos de los usuarios y las diferentes variables que esto implica (Liu, Wang y Ghadimi, 2017). Esto ocasiona que el sistema eléctrico trabaje con cierta incertidumbre que incide en el costo de su producción. Teniendo en consideración que la energía eléctrica producida a gran escala no puede almacenarse (He, Song y Shen, 2010).

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Además, el consumo eléctrico es variable debido a su alta demanda en horarios diferenciados. Otro punto a considerar, es el escaso conocimiento sobre las tendencias de consumo eléctrico en los diferentes sectores de un país. Esto es de gran importancia para evitar un sobre ajuste de políticas y proyectos en el sector eléctrico. Ya que, al establecer agrupaciones de consumo eléctrico, su planificación, no se enfoca en particularidades y busca una generalidad en planes futuros de generación eléctrica. Bajo este criterio, la curva de demanda eléctrica permite representar la demanda real vs la prevista. Este proceso debe ser llevado a cabo mediante algoritmos complejos adaptativos que recolectan información permanentemente (Yildiz, Bilbao, Dore y Sproul, 2017). Sin embargo, este análisis debe ser visto de una manera diferente entre la demografía de cada país y el tipo de consumidor.

Con lo comentado anteriormente, el Ecuador tiene una gran capacidad de generación eléctrica al emplear los diferentes recursos naturales que posee. Ya que su consumo eléctrico tiene un constante crecimiento anualmente. En el 2019, fue del 4.5%, esto deriva en la necesidad de implementar sistemas de soporte de decisión que tenga la capacidad de pronosticar el consumo eléctrico que más se ajuste a los valores reales y, si es el caso recomendar acciones sobre variaciones de hábitos de consumo en la sociedad. Por esta razón se cuenta con grandes repositorios de datos para su análisis. Este proceso solo puede llevarse a cabo bajo criterios robustos de aprendizaje automático alojados en servidores con altos recursos computacionales (Zhang, Grolinger, Capretz y Seewald, 2018). Por lo tanto, se plantea realizar un estudio exhaustivo de la información disponible, para encontrar tendencias de consumo en las diferentes provincias del país, para pronosticar el consumo por cada una de ellas (Zhao y Magoulès, 2012). Con esto, se las puede agrupar según sus principales características y establecer relaciones entre ellas.

El presente trabajo se encuentra enfocado en pronosticar el tipo de consumo por municipalidades dentro de la provincia de Imbabura - Ecuador en relación a las diferentes

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

características de consumo eléctrico recopiladas por las empresas que suministran el servicio eléctrico. Para hacer esto, se recolecta información disponible desde la agencia de regulación y control de energías y recursos naturales no renovables, suscrita al Ministerio de Electricidad del Ecuador. Con esto, se realizará un esquema de análisis de datos con diferentes criterios para determinar la mejor solución a ser implementada. Para esto, se inicia con un proceso de limpieza de datos, posteriormente, se realiza una comparación de algoritmos de regresión lineal multivariante para determinar el modelo que mejor se adapte y tenga una mejor capacidad de predicción de consumo. Además, se realiza una agrupación entre las municipalidades para establecer hábitos de consumo. Todo esto es visto en una interfaz gráfica empleando una herramienta de soporte de decisión por parte de los entes reguladores del país. Como resultados relevantes, los modelos de máquinas de soporte de decisión y bosques aleatorios demostraron ser adecuados para esta tarea con un error de predicción menor al 10 %.

Por su parte, el algoritmo k-means logró agrupar 4 tipos de consumo eléctrico con una representación del 98 % de la variabilidad de los datos.

El resto del documento está estructurado de la siguiente manera: en la sección II se presentan los trabajos relacionados con el tema planteado. Materiales y métodos se muestran en la sección III. Sección IV presenta los resultados. Finalmente, las conclusiones y el trabajo futuro se muestran en la sección V.

2 Trabajos Relacionados.

Este tema ha sido analizado en trabajos tales como (Johnson, Starke, Abdelaziz, Jackson y Tolbert, 2014) que presenta un análisis de datos sobre el consumo eléctrico en domicilios para ser almacenados en un servidor en la nube. En este caso, no presentan una solución a un sector específico y se enfoca a un consumo residencial. Por otra parte, (Rezaei, Sharghi y Motabeli, 2018), realiza un estudio arquitectónico en edificios y su influencia en el consumo

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

eléctrico a partir de tendencias de usuario. (Gong, Han, Li, Tian, 2020) hace uso de modelos matemáticos y simulaciones en diferentes aplicaciones de sistemas eléctricos de potencia. Por su parte, (Çetinkaya, Avci, Bayindir, 2020) utiliza series de tiempo y agrupaciones para predecir el consumo de electricidad residencial. Recientemente, (Parraga-Alava, Moncayo-Nacaza, Revelo-Fuelagán, Rosero-Montalvo, Anaya-Isaza y Peluffo-Ordóñez, 2020) ha publicado con conjunto de datos sobre el consumo eléctrico de Colombia y las diferentes características de los consumidores. En el Ecuador, trabajos como (Parra-Narváez, 2015) realizan estudios sobre la demanda eléctrica en términos generales y su relación a las emisiones de CO_2 que conlleva su producción. (Toapanta-Lema, Gallegos, Rubio-Aguilar, Llanes-Cedeño, Carrascal-García, García-López y Rosero-Montalvo, 2020) presenta una solución de predicción de consumo eléctrico en instituciones públicas. Por otro lado, (Guachimboza-Dávalos, Llanes-Cedeño, Rubio-Aguilar, Peralta-Zurita y Núñez-Barrionuevo, 2021) emplea redes neuronales para la predicción de consumo eléctrico de los diferentes cantones de la provincia de Pichincha-Ecuador. Finalmente (Núñez-Barrionuevo, Llanes-Cedeño, Martínez-Gómez, Guachimboza-Dávalos y López-Villada, 2020), realiza un estudio de agrupamiento por características de consumidor en la citada provincia.

Los trabajos antes mencionados han presentado soluciones adecuados sobre la predicción eléctrica y la agrupación por consumidores. No obstante, se encuentran algunas problemáticas abiertas sobre la temática planteada, como la combinación de varios criterios de aprendizaje automático, la presentación formal de resultados en interfaces interactivas, la relación con los sistemas nacionales de generación eléctrica, entre otros. Por estas razones, el presente trabajo cubre estos criterios mediante la presentación de modelos matemáticos para la extracción del conocimiento intrínseco presente en el conjunto de datos obtenidos sobre los parámetros de consumo eléctrico en la provincia de Imbabura-Ecuador. Con esto, la presentación de los

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

resultados en una herramienta de apoyo a la decisión, busca integrar a los sectores académicos con las entidades públicas para la generación de políticas estatales basadas en información confiable.

3 Materiales y Métodos.

Este capítulo muestra, por un lado, la descripción de la base de datos disponible y sus principales características. Por otro lado, se muestra el esquema de análisis de datos propuesto junto con los criterios de predicción de agrupamiento de información.

3.1 Descripción de la base de datos.

En la provincia de Imbabura, su principal línea de negocio es la producción de artículos de madera y servicios para industrias medianas (Ministerio de Energía y Recursos Naturales no Renovables, 2019). En relación a la producción de energía eléctrica, cuenta con 11 plantas de producción media, que generan una potencia efectiva de 115 MW por año con un total de 130.000 abonados del sector residencial con un consumo anual de 150 GWh. Por ello, Ecuador cuenta con un sistema eléctrico interconectado para abastecer la energía faltante a las provincias a través de sus emblemáticos proyectos como la central hidroeléctrica Coca Codo Sinclair (Urgilés et al., 2015). En este sentido, la información adquirida a través de los repositorios de datos abiertos proporcionados por Ecuador, permitió adquirir los informes de consumo de los clientes residenciales desde 2017 hasta 2020. Este tipo de consumidor se establece por ser el más grande de la provincia. Con esto, tenemos una matriz de datos multivariante de la forma $Y \in R^{m \times n}$, donde m es el número de instancias, y n es el número de variables. Por lo tanto, tenemos $m = 3500$ y $n = 8$ que son las características de: cantón, municipio, tipo de equipo eléctrico (dispositivos electrónicos de 220 voltios), número de clientes por zona geográfica, energía facturada, consumo incremental, consumo residencial y facturación.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

3.2 Propuesta de análisis de datos.

Este esquema de datos se enfoca en tres componentes principales. La transformación, ya que es necesario limpiar los datos y actualizar la información, debido a que el idioma español tiene tildes que en ocasiones se encuentran en las bases de datos, y esto genera un mayor número de atributos erróneos. Además, existen datos vacíos o nulos que generarían errores en los modelos de predicción y agrupación de datos (sección 3.3). Por otro lado, se deben elegir los diferentes modelos de regresión lineal multivariante para determinar el adecuado, estos se muestran en el apartado 3.4. Finalmente, el algoritmo de clustering seleccionado es k-means, debido a su gran efectividad en esta tarea. Sin embargo, es necesario elegir correctamente el valor correcto de clústers dentro de la base de datos propuesta (Núñez-Barrionuevo et al., 2020). Todo este proceso se ve en la figura 1.

3.3 Transformación.

La matriz $Y \in R^{m \times n}$ tiene tres inconvenientes principales. El primero se basa en tener datos nulos o celdas vacías. Para ello, procedimos a eliminar estos datos para evitar inconvenientes. En segundo lugar, tiene variables categóricas con tilde o acento. Esto hace que la misma variable tenga valores diferentes porque tiene una codificación ASCII diferente. Por este motivo, se llevó a cabo un método de búsqueda para eliminar todas estas palabras y

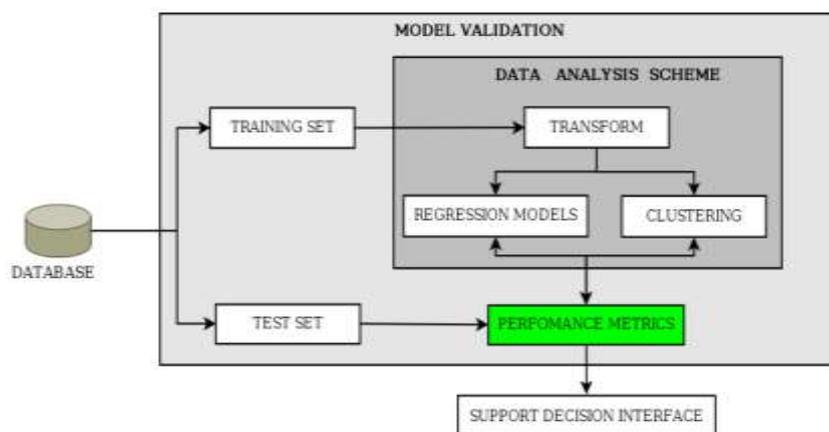


Figura 1. Esquema de análisis de datos

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

estandarizar todas las variables. Este proceso se realizó con una herramienta ETL. Finalmente, es necesario codificar estas variables categóricas para entrenar un modelo de regresión lineal multivariante. Para hacer esto, se ha utilizado el método One-Hot-Encoding de la biblioteca sklearn del entorno de programación de Python. Este método es útil porque cuando codifica las variables, cada valor de categoría se convierte en una nueva columna y se asigna un valor de 1 o 0 (notación para verdadero o falso) a la columna. Esto evita malinterpretar los valores numéricos como si tuvieran algún tipo de jerarquía en ellos. Con el objetivo de no saturar el modelo de regresión lineal con demasiadas variables codificadas. Se ha creado una nueva matriz $X \in R^{p \times s}$. Donde $p = 3450$ al eliminar los datos nulos y $s = 12$ al eliminar las municipalidades como característica aplicable al modelo y al incrementar columnas de los atributos de ciudad que son: IBARRA, ANTONIO ANTE, COTACACHI, OTAVALO, PIMAMPIRO y URCUQUÍ. Además, la variable objetivo se ha eliminado del modelo para crear un vector de características que solo almacena en este caso el valor total facturado.

Por su parte, para realizar el análisis de agrupamiento, se deben eliminar todas las variables categóricas para establecer similitudes en las características de consumo. Por esta razón, se crea una nueva matriz llamada $Z \in R^{p \times t}$. Donde, $p = 3450$ ya que el tamaño de la matriz no ha cambiado. Sin embargo, $t = 6$ ya que se eliminaron las variables categóricas.

3.4 Modelos de regresión.

Para generar un modelo que pronostique el costo total del consumo eléctrico, existen diferentes modelos que utilizan criterios de análisis de regresión multivariante lineal y no lineal. En este sentido, las más importantes en relación a la literatura encontrada se plantean como: (i) matriz de covarianza, (ii) funciones de similitud, (iii) tabla de frecuencias y otras. El primer criterio (matriz de covarianza) es un modelo lineal simple (modelo 1) basado en la ecuación:

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

$$y = b_0 + \sum_{i=0}^s x_i b_i \quad (1)$$

Donde y es la variable dependiente, x_i son las variables independientes hasta el valor de $s = 12$ y b_i son los parámetros de influencia de las variables hacia el modelo.

Para mejorar el ajuste de un modelo lineal, se puede elevar al cuadrado cada uno de sus valores de los atributos de la matriz y encontrar una regresión multivariante (modelo 2) de la ecuación:

$$y = b_0 + \sum_{j=0}^r \sum_{i=0}^s x_i^j b_i \quad (2)$$

Donde el valor de $r = 2$ que representa el grado en que se ha elevado el modelo lineal.

Otro criterio a considerar se basa en distancias a través del algoritmo k-NN. La regresión k-NN usa las mismas funciones de distancia euclidiana que tiene la forma (modelo 3):

$$\sqrt{\sum_{i=0}^s (x_i - y_i)^2} \quad (3)$$

Las máquinas de soporte de decisiones (SVM) se pueden utilizar como métodos de regresión no lineal ya que SVM establece márgenes de tolerancia para minimizar el error, individualizando el hiperplano que maximiza el margen de predicción. Este criterio viene dado por la ecuación 4 que puede usar diferentes métodos de kernel.

$$y = b_0 + \sum_{i=0}^s (x_i - x_i^*) \cdot (kernel) \quad (4)$$

Se puede apreciar que el término x_i^* se convierte en la variable llevada a un hiperplano que ha sido realizada por una función kernel. Por esta razón, el modelo 4 usa una función radial y el modelo 5 usa una función sigmoid.

Finalmente, los algoritmos basados en tablas de frecuencia, como árboles de decisión y

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

bosques aleatorios, permiten un modelo de regresión lineal complejo y altamente no lineal. Está basado en la ecuación 5.

$$y = \frac{1}{B} + \sum_{i=0}^s f_i * x \quad (5)$$

Donde $B = 10$ representa los árboles de decisión utilizados y f_i la frecuencia de pronóstico resultante.

3.5 Agrupamiento.

El método de k-medias tiene como objetivo dividir 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 es el más cercano.

Sin embargo, la aleatoriedad de los valores de k puede provocar diferentes formas de agrupación. Por ello, es necesario definir adecuadamente su valor para agrupar los datos que tienen mayor similitud en sus características. En consecuencia, K-medias ++ permite eliminar este problema mediante el análisis de un conjunto de observaciones (x_1, x_2, \dots, x_n) , donde cada observación es un vector real d - dimensional y la agrupación de K - means tiene como objetivo dividir el n observaciones en k ($\leq n$) conjuntos $S = S_1, S_2, \dots, S_k$ para minimizar la suma de cuadrados dentro del grupo (WCSS) (la varianza). Esto se puede ver en la ecuación 6 (Núñez-Barrionuevo et al., 2020)

$$\arg_S \min \sum_{i=1}^k \sum_{x \in S_i} \|x - u_i\|^2 = \arg_S \min \sum_{i=1}^k |S_i| \text{Var} S_i \quad (6)$$

Donde u_i es el promedio de los puntos S_i .

Gráficamente, la varianza se puede apreciar al utilizar los diferentes valores de k . Para su selección se debe elegir el valor de k que ya no exista una alta variabilidad en su agrupación.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

En este caso $k = 4$. Esto se puede ver en la figura 2.

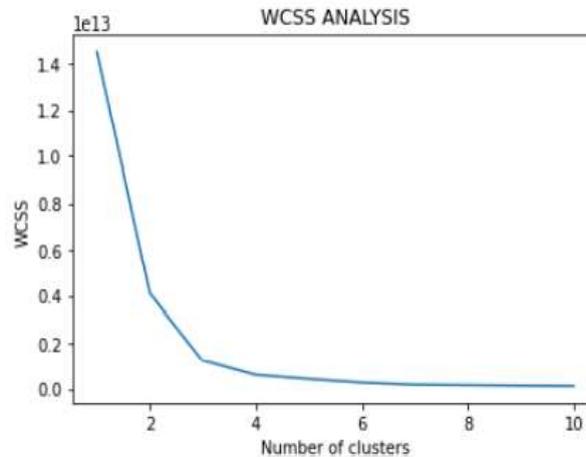


Figura 2. Análisis de k mediante la técnica WCSS

4 Resultados.

Los resultados se muestran con los criterios de: regresión lineal, agrupamiento e interfaz gráfica.

4.1 Modelos de regresión.

Para la selección del modelo de regresión lineal que mejor represente el conjunto de datos de la matriz $X \in R^{p \times s}$. Se realizan pruebas de error de pronóstico. Para hacer esto, la matriz de datos se ha dividido aleatoriamente 10 veces en entrenamiento y prueba. Esto se hace para entrenar el modelo y probar las predicciones contra su respuesta. Para ello se utiliza la suma acumulada de errores de pronóstico (CFE), la desviación media absoluta (MAD), el error cuadrático medio (MSE) y la desviación en términos porcentuales (MAE). La tabla 1 muestra los resultados obtenidos por cada modelo propuesto con un conjunto de prueba de 750 datos. Como se puede observar, la variabilidad (% MSE y % MAE) es muy alta en los modelos 1, 2, 3 y 4. Por su parte, los modelos 5 y 6 tienen menor variabilidad y un error de pronóstico menor al 10 %. Este resultado es muy aceptable considerando la naturaleza de los datos.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Tabla 1. Análisis de tipos de error en la matriz $X \in R^{p \times s}$

Errors type	CFE	%CFE	MAD	MSE	MAE
Model 1	-83007.0	± 7.6	-111.72	152690	14%
Model 2	-82117.9	± 7.52	-110.52	135300	12.43%
Model 3	-83007.0	± 7.6	-111.72	152690	14%
Model 4	-81037.0	± 7.44	-109.06	124017	11.39%
Model 5	-61538.3	± 5.65	-82.82	106848	9.81%
Model 6	-55041.5	± 5.05	-74.08	100914	9.27%

4.2 Agrupamiento

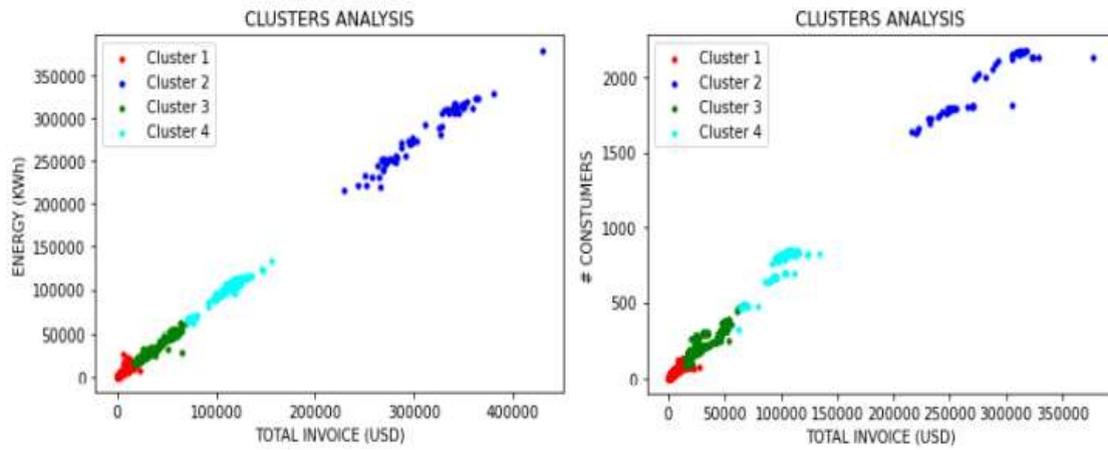
Con el resultado de $k = 4$, se procede a compilar el algoritmo en la base de datos para agrupar por municipio a partir de la matriz $Z \in R^{p \times t}$. Con esto se definen 4 tipos de consumo: EXCESIVO, ALTO, MEDIO y BAJO. Los clústers se presentan gráficamente en la figura 3 la afinidad de los datos y la distancia entre ellos. Esto muestra que el valor de $k = 4$ es adecuado y representa una variabilidad del 98 % de los datos.

4.3 Interface.

La matriz de datos completa se ha cargado en la herramienta de soporte de decisiones junto con los algoritmos de aprendizaje automático para generar informes. Con esto, el personal de las entidades públicas relacionadas con la generación eléctrica puede observar y gestionar los resultados obtenidos de manera amigable. Esto se puede ver en la figura 4, donde las previsiones de los modelos 5 y 6 son muy adecuadas. Además, esta información puede ser consultada por cantones o municipios para un mayor control y seguimiento.

Por otro lado, utilizando sus dos variables categóricas y utilizando un mapa de árbol como organización por clúster, se pueden observar todas las municipalidades con respecto al grupo al que pertenecen y su consumo de predicción por mes.

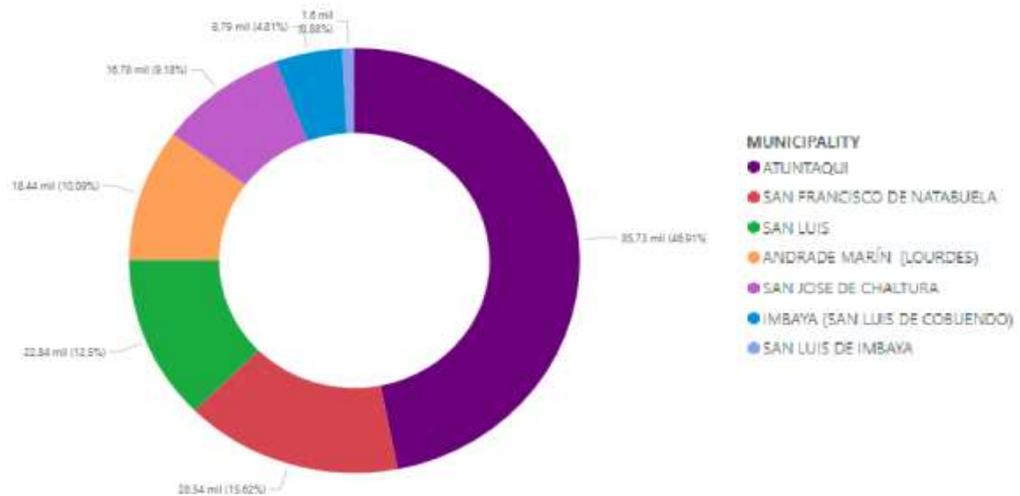
“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”



(a) Análisis de clúster por consumo total de energía

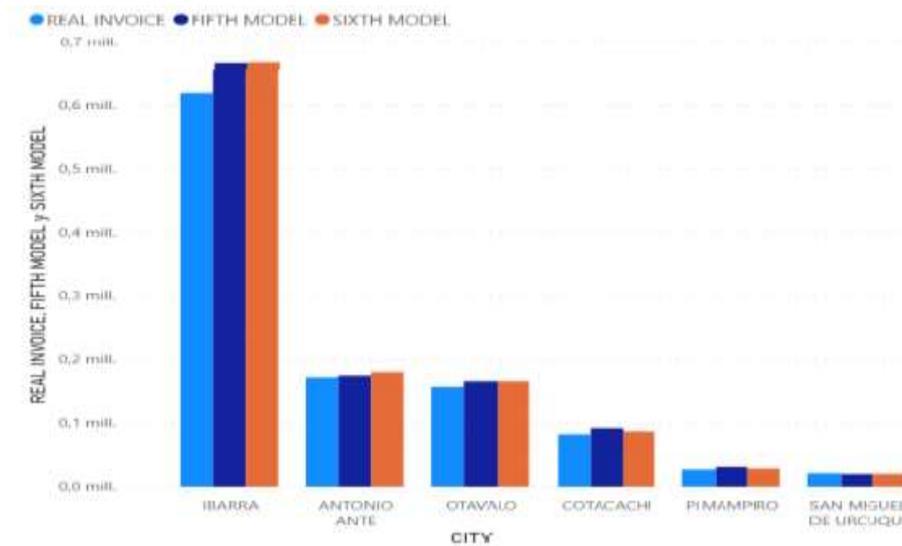
(b) Análisis de clúster por número de clientes

Figura 3. Variabilidad del análisis de clúster



(a) Informe de ejemplo de municipalidades de Antonio Ante

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”



(b) Informes de modelos de predicción de consumo eléctrico de Imbabura

Figura 4. Interfaz de soporte de decisión sobre predicciones de consumo

Esta información es muy importante para la planificación eléctrica en la provincia de Imbabura. Esto se muestra en la figura 6.

Finalmente, utilizando los datos de posición global de cada municipio, cada uno de ellos se puede representar por color con respecto a la agrupación asignada por el algoritmo EXCESIVO, ALTO, MEDIO y BAJO. Con esto, en cada mes que se recompila el algoritmo, se puede observar el cambio en los hábitos de consumo y se pueden presentar alertas. Esto se muestra en la figura 7.

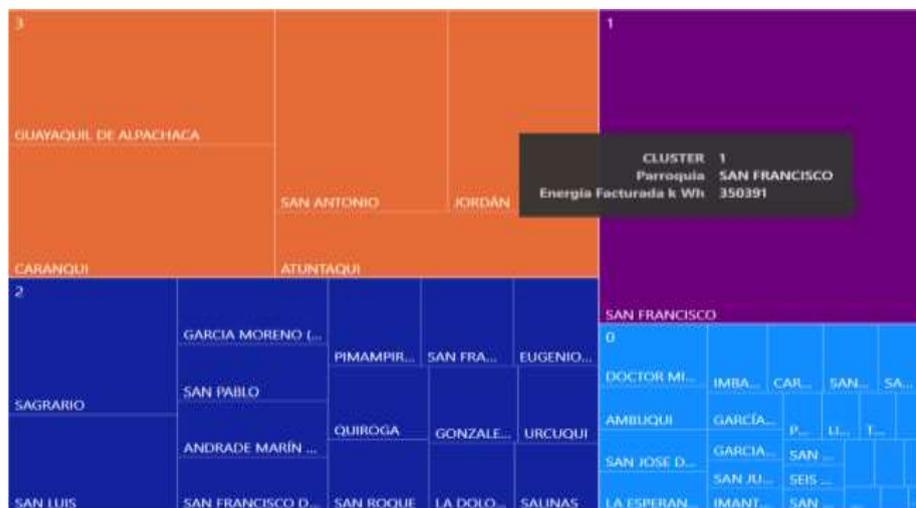


Figura 5. Informes de clúster: * Consumo excesivo, * Consumo alto, * Consumo normal, * Consumo bajo

Figura 6. Informes de clúster por municipalidades

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

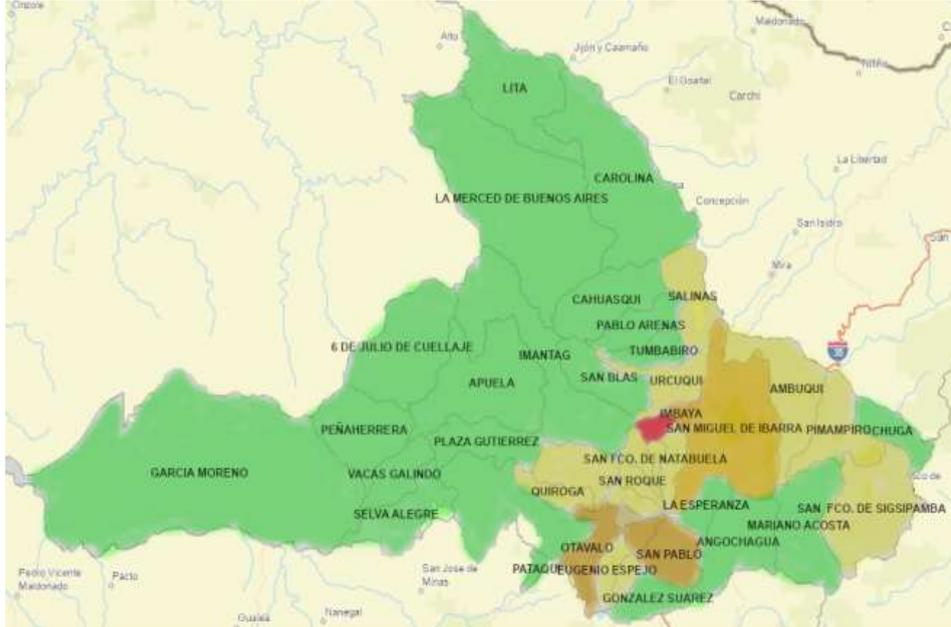


Figura 7. Informes de clúster: * Consumo excesivo, * Consumo alto, * Consumo normal, * Consumo bajo

5 Conclusiones y futuros trabajos.

El análisis del consumo eléctrico mediante algoritmos de aprendizaje automático permite contar con análisis de datos robustos para la adecuada generación de informes que apoyen la planificación y creación de proyectos eléctricos que satisfagan la creciente demanda eléctrica. En este sentido, los modelos de regresión no lineal permiten entrenar modelos con variables categóricas, las cuales son ampliamente utilizadas para definir ciertos factores de las características del usuario. En consecuencia, las máquinas de soporte de decisión y los bosques aleatorios resultaron ser los más adecuados para presentar las tendencias en el consumo eléctrico en la provincia de Imbabura-Ecuador, organizados por cantones y municipalidades.

Por su parte, el algoritmo de k-medias con un valor de $k = 4$ resultó adecuado para determinar similitudes de consumo por cantones y municipios. Con esto, es posible observar cambios mensuales en sus tendencias y alertar sobre ellos.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Finalmente, los informes para una herramienta de apoyo a la decisión, permiten brindar información ágil y eficiente para la planificación eléctrica de un país.

Como trabajo futuro, se propone utilizar estos mismos criterios para que todas las provincias cuenten con una herramienta nacional para predecir y agrupar el consumo eléctrico.

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

REFERENCIAS BIBLIOGRÁFICAS

- Çetinkaya, Ü., Avcı, E., & Bayindir, R. (2020). Time Series Clustering Analysis of Energy Consumption Data. *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*, 409–413. <https://doi.org/10.1109/ICRERA49962.2020.9242763>
- D. Gabriel and T. Urgilés, “UNIVERSIDAD POLITECNICA SALESIANA SEDE CUENCA FACULTAD DE INGENIERIAS CARRERA DE INGENIERIA ELECTRICA,” Tech. Rep.
- Dmitri, K., Maria, A., & Anna, A. (2016). Comparison of regression and neural network approaches to forecast daily power consumption. *2016 11th International Forum on Strategic Technology (IFOST)*, 247–250. <https://doi.org/10.1109/IFOST.2016.7884239>
- Gong, F., Han, N., Li, D., & Tian, S. (2020). *Trend Analysis of Building Power Consumption Based on Prophet Algorithm*. <https://doi.org/10.1109/AEEES48850.2020.9121548>
- Guachimboza-Davalos, J. I., Llanes-Cedeño, E. A., Rubio-Aguilar, R. J., Peralta-Zurita, D. B., & Núñez-Barrionuevo, O. F. (2021). Prediction of Monthly Electricity Consumption by Cantons in Ecuador Through Neural Networks: A Case Study. In M. Botto-Tobar, W. Zamora, J. Larrea Plúa, J. Bazurto Roldan, & A. Santamaría Philco (Eds.), *Systems and Information Sciences* (pp. 28–37). Springer International Publishing.
- He, L., Song, Q., & Shen, J. (2010). k-NN Numeric Prediction Using Bagging and Instance-relevant Combination. *2010 Second International Symposium on Data, Privacy, and E-Commerce*, 3–8. <https://doi.org/10.1109/ISDPE.2010.6>
- Johnson, B. J., Starke, M. R., Abdelaziz, O. A., Jackson, R. K., & Tolbert, L. M. (2014). A method for modeling household occupant behavior to simulate residential energy consumption. *ISGT 2014*, 1–5. <https://doi.org/10.1109/ISGT.2014.6816483>

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

Liu, Y., Wang, W., & Ghadimi, N. (2017). Electricity load forecasting by an improved forecast engine for building level consumers. *Energy*, *139*, 18–30.

<https://doi.org/10.1016/j.energy.2017.07.150>

M. d. E. y. recursos no renovables De, “Plan de mejoramiento de los sistemas de distribución de energía eléctrica (PMD) – Ministerio de Energía y Recursos Naturales no Renovables,” 2019. [Online]. Available: <https://www.rekursyenergia.gob.ec/plan-de-mejoramiento-de-los-sistemas-de-distribucion-de-energia-electrica-pmd/>

Núñez-Barrionuevo, O., Cedeño, E. A., Martínez, J., Guachimboza-Davalos, J., & Lopez-Villada, J. (2020). *Clustering Analysis of Electricity Consumption of Municipalities in the Province of Pichincha-Ecuador Using the K-Means Algorithm* (pp. 187–195).

https://doi.org/10.1007/978-3-030-59194-6_16

Parra Narváez, R. (2015). Factor de emisión de CO₂ debido a la generación de electricidad en el Ecuador durante el periodo 2001-2014. *ACI Avances En Ciencias e Ingenierías*, *7*(2).

<https://doi.org/10.18272/aci.v7i2.269>

Parraga-Alava, J., Moncayo-Nacaza, J. D., Revelo-Fuelagán, J., Rosero-Montalvo, P. D., Anaya-Isaza, A., & Peluffo-Ordóñez, D. H. (2020). A data set for electric power consumption forecasting based on socio-demographic features: Data from an area of southern Colombia. *Data in Brief*, *29*, 105246. <https://doi.org/10.1016/j.dib.2020.105246>

Rezaei, sohrab, Sharghi, A., & motalebi, ghasem. (2018). A framework for analysis affecting behavioral factors of residential buildings’ occupant in energy consumption. *Journal of Sustainable Architecture and Urban Design*, *5*(2), 39–58.

https://jsaud.sru.ac.ir/article_895.html

Sauhats, A., Varfolomejeva, R., Lmkevics, O., Petrecenko, R., Kunickis, M., & Balodis, M. (2015). *Analysis and prediction of electricity consumption using smart meter data*. 17–

22. <https://doi.org/10.1109/PowerEng.2015.7266290>

“ANÁLISIS DE PREDICCIÓN Y AGRUPACIÓN DE CONSUMO ELÉCTRICO EN LA PROVINCIA DE IMBABURA – ECUADOR PARA LA OPTIMIZACIÓN DE RECURSOS ENERGÉTICOS”

- Toapanta-Lema, A., Gallegos, W., Rubio-Aguilar, J., Cedeño, E. A., Carrascal García, J., García-López, L., & Rosero, P. (2020). *Regression Models Comparison for Efficiency in Electricity Consumption in Ecuadorian Schools: A Case of Study* (pp. 363–371). https://doi.org/10.1007/978-3-030-42520-3_29
- Ye, X., Wu, X., & Guo, Y. (2018). *Real-time Quality Prediction of Casting Billet Based on Random Forest Algorithm*. 140–143. <https://doi.org/10.1109/PIC.2018.8706306>
- Yildiz, B., Bilbao, J. I., Dore, J., & Sproul, A. B. (2017). Recent advances in the analysis of residential electricity consumption and applications of smart meter data. In *Applied Energy* (Vol. 208, pp. 402–427). Elsevier Ltd. <https://doi.org/10.1016/j.apenergy.2017.10.014>
- Zhang, X. M., Grolinger, K., Capretz, M. A. M., & Seewald, L. (2018). Forecasting Residential Energy Consumption: Single Household Perspective. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 110–117. <https://doi.org/10.1109/ICMLA.2018.00024>
- Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. In *Renewable and Sustainable Energy Reviews* (Vol. 16, Issue 6, pp. 3586–3592). Pergamon. <https://doi.org/10.1016/j.rser.2012.02.049>