

## MONITOREO Y CARACTERIZACION DE LOS PERFILES DE DEMANDA HORARIA DE ENERGIA ELECTRICA DE USUARIOS DE LA ANDE

Hans Mersch<sup>1</sup>, Ing. PhD. Diego Stalder<sup>3</sup>, Ing. MSc. Carlos Sauer<sup>3</sup>, Ing. Gustavo Verón<sup>3</sup>

<sup>1</sup>Alumno, hmersch@fiuna.edu.py, Facultad de Ingeniería. Universidad Nacional de Asunción. San Lorenzo, Paraguay

<sup>3</sup>Orientador, Facultad de Ingeniería. Universidad Nacional de Asunción. San Lorenzo, Paraguay

### Palabras clave:

- Agrupamiento
- Kmeans
- Perfil de consumo de energía
- Agrupamiento spectral

### Resumen

Este trabajo final de grado se centra en la identificación de perfiles de consumo horario de energía eléctrica, crucial para mejorar la planificación y eficiencia en su suministro. Detalla una comparativa exhaustiva de métodos no supervisados para definir estos perfiles. Tras varios análisis computacionales, se confirma que la combinación de algoritmos PCA (Análisis de Componentes Principales) y Kmeans resultó la más efectiva. Posteriormente se aplicó a datos clientes residenciales como industriales y en áreas geográficas específicas del país. Este enfoque permitió reconocer patrones de consumo y define perfiles específicos. Además, se introduce un sistema de medición no intrusivo que alimenta una base de datos y un panel interactivo para analizar el consumo diario de electricidad en los hogares.

### Keywords

- Clustering
- Kmeans
- Electricity consumption profile
- Spectral Clustering

### Abstract

This final degree project focuses on identifying hourly electricity consumption profiles, a critical aspect in enhancing planning and efficiency in its supply. It provides a comprehensive comparison of unsupervised methods to define these profiles. Following various computational analyses, it confirms that the combination of PCA (Principal Component Analysis) and Kmeans proved most effective. Subsequently applied to both residential and industrial customer data in specific geographical areas of the country, this approach facilitated the recognition of consumption patterns and the establishment of specific profiles. Additionally, it introduces a non-intrusive measurement system that feeds data into a database and an interactive panel for analyzing daily electricity consumption in households.

## 1. INTRODUCCIÓN

Las compañías de provisión de energía eléctrica necesitan predecir de la demanda energética en diferentes escalas de tiempo, a corto (horas), mediano (meses) y largo plazo (años). Una previsión precisa permite planificar mejor la generación y satisfacer la demanda, así como también en la optimización de la transmisión de energía en las redes de distribución y el logro de una operación energéticamente eficiente. Por ejemplo, la identificación de periodos de demanda energética pico permitiría implementar mejores estrategias operativas para la transmisión de la energía.

Una buena planificación del sistema eléctrico comienza directamente con análisis de la demanda del consumidor considerando, por ejemplo, la demanda mensual acumulada o la demanda máxima de potencia. Sin embargo, si se tienen acceso a las series temporales de consumo de energía eléctrica de los usuarios, se pueden aplicar algoritmos de agrupamiento para realizar un análisis mas detallado de las curvas de demanda en función del tiempo para cada tipo de usuario. Estudios recientes, mostraron que este tipo de algoritmos permitirían agrupar varios consumidores identificando características comunes[1, 2, 3]. Identificar grupos de usuarios similares posibilitaría

la generación de perfiles de demanda característicos que posteriormente podrían combinarse para explicar la demanda de un determinado subsistema e inclusive la demanda global[4, 5, 6, 7, 8, 9, 10, 11].

### 1.1. Objetivos Específicos

- Implementar un sistema de monitoreo del consumo de energía por hora en una muestra de clientes residenciales.
- Pre-procesamiento de los datos de demanda de energía, recolectados previamente por la ANDE.
- Agrupar los perfiles de demanda de usuarios residenciales y no residenciales.
- Identificar y buscar características relevantes para agrupar las series temporales de demanda.

### 1.2. Alcance y Limitaciones

En los últimos años, la Administración Nacional de Energía Eléctrica (ANDE) declaro en su compilación estadística[12] que las pérdidas por distribución de energía eléctrica aumentaron anualmente. En el 2019, estas pérdidas alcanzaron los 3.506.467[MWh], lo cual es un salto significativo en comparación a años anteriores.

Para llevar un control más estricto de las redes de distribución, es necesario establecer una métrica para identificar cualquier comportamiento anómalo en la línea. Sin embargo, si no existe una métrica adecuada, es necesario diseñarla. La detección temprana de posibles fallas en la red de distribución nos permite evitar pérdidas de energía de distribución significativas. Además, al detectar el área de distribución que posee el comportamiento anómalo, solo es necesario mandar al personal a dicha área en vez de mandar a varias cuadrillas por la zona para encontrar el problema. Por lo tanto, la implementación de este sistema de detección permitiría a la ANDE optimizar sus recursos humanos y reducir sus pérdidas energéticas.

El análisis de este trabajo se enfocara en la creación de perfiles de demanda energética de usuarios de la ANDE y su posterior agrupación, a partir de los datos obtenidos en un muestreo de datos de los usuarios, tanto en el área metropolitana como en el interior del país.

Además del análisis de los datos, se contempla el diseño y la integración de un sistema de monitoreo de bajo costo no invasivo del consumo de energía de una vivienda. Se plantea la utilización de un sensor inductivo en conjunto con un microcontrolador que posea la capacidad de conectarse a internet para monitorear el consumo de energía en una vivienda.

## 2. MARCO TEÓRICO

### 2.1. Series temporales

Las series de tiempo representan conjuntos de datos que se presentan en una secuencia a lo largo de intervalos temporales específicos. Estos datos pueden ser cuantitativos, categóricos o incluso de tipo booleano. A través del análisis de estas series, se puede descubrir información valiosa aplicable a numerosos campos de investigación.

Históricamente, se percibía que las series temporales se subdividían en tendencia, oscilaciones estacionales e irregularidades, sin embargo, este enfoque no siempre ha sido óptimo [13]. Actualmente, se considera que las series temporales pueden categorizarse en cuatro componentes esenciales [14]:

- **Tendencia:** Es la manifestación de una dirección clara y sostenida en el transcurso de la serie temporal, reflejando cambios a lo largo del tiempo.
- **Variaciones estacionales:** Refieren a alteraciones que se presentan con regularidad en determinados intervalos. Estas oscilaciones pueden estar influenciadas por factores climáticos, tradiciones sociales o prácticas culturales.
- **Fluctuaciones cíclicas:** Se relacionan con cambios no estacionales que se identifican en intervalos establecidos. Estas fluctuaciones pueden compararse con etapas del ciclo económico, incluyendo fases como el auge, declive, baja y crecimiento.
- **Variaciones Irregulares:** Son las perturbaciones que persisten tras descontar los otros tres componentes de la serie. Estas pueden originarse por eventos aleatorios o no determinísticos.

### 2.2. Agrupamiento de series temporales

El agrupamiento de series temporales representa un área significativa en el ámbito del aprendizaje no supervisado, dedicada a la identificación de patrones y estructuras inherentes en conjuntos de datos temporales. En los métodos de aprendizaje sin supervisión, los datos se presentan sin etiquetas previas y únicamente se cuenta con sus valores intrínsecos como variables autónomas. El propósito principal de estos métodos es descubrir patrones inherentes en los datos sin establecer predicciones concretas, lo que significa que no se opera con una variable de respuesta. Al trabajar con series temporales, es crucial considerar dos aspectos: la técnica de representación y el criterio de distancia o similitud.

La elección de una técnica de representación adecuada es fundamental para conseguir algoritmos que sean tanto eficientes como escalables[15]. Las dos representaciones más comúnmente adoptadas son la orientada a la forma y la basada en atributos.

- **Representación orientada a la forma:** Se emplean los valores iniciales ordenados temporalmente, utilizando la secuencia natural de observaciones a lo largo del tiempo.
- **Representación basada en atributos:** Se genera a partir de la extracción de atributos distintivos de la serie temporal, lo que conduce a una disminución en la dimensionalidad de la serie, pero manteniendo la esencia informativa de la serie original.

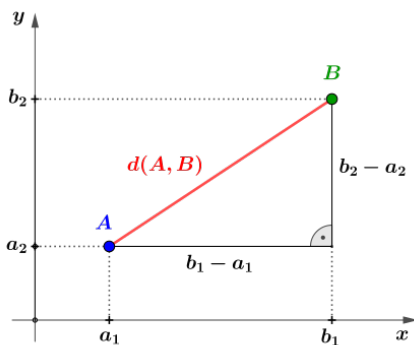
Para determinar cuán similares son dos series temporales,  $T_1$  y  $T_2$ , se han propuesto diversas métricas de similitud. Estas métricas pueden ser representaciones funcionales de la serie temporal en un espacio específico o evaluaciones basadas en la cercanía de los datos entre las series [16].

Se ha observado que simplemente calcular la distancia entre dos valores de manera directa no siempre es la estrategia óptima para las series temporales. Por ello, se han ideado numerosas funciones de distancia para preservar la integridad de la información. Al evaluar la similitud, es posible tomar en cuenta la cercanía entre puntos o, alternativamente, la relación basada en la dinámica de la serie, como por ejemplo, su tasa de crecimiento.

La medida de proximidad escogida para este trabajo final de grado es la distancia euclidiana. Dadas dos series temporales,  $T_1 = (u_1, u_2, \dots, u_n)$  y  $T_2 = (v_1, v_2, \dots, v_n)$ , observadas en los momentos  $t_1, t_2, \dots, t_n$ , la distancia euclidiana entre  $T_1$  y  $T_2$  se expresa como:

$$DE(T_1, T_2) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2} \quad (1)$$

La Figura 1 ilustra que la función  $DE(T_1, T_2)$  se centra únicamente en la cercanía de los valores registrados en sus respectivos momentos. Debido a que la distancia se mantiene con transformaciones que alteran la secuencia temporal de observaciones, es posible que la distancia euclidiana no capture ciertas relaciones entre los datos [17].



**Figura 1:** Visualización de la distancia euclidiana.  
**Fuente:** [18]

## 2.3. Algoritmos de agrupamiento particional

El agrupamiento particional busca dividir el conjunto de datos en grupos a través de métodos iterativos. Dado un conjunto  $X = \{x_1, x_2, \dots, x_n\}$ , este tipo de agrupamiento busca dividir  $X$  en  $k$ -particiones, orientándose comúnmente por una función de costo.

El algoritmo K-means, por ejemplo, inicia seleccionando aleatoriamente  $K$  centroides. Luego, asigna cada dato al centroide más cercano. Tras esto, recalcula los centroides y repite el proceso hasta que la posición de los centroides ya no cambie.

A continuación, se describirán con mas detalle los algoritmos a utilizar:

### 2.3.1. Algoritmo de agrupamiento K means

El algoritmo K-Means es un método de agrupamiento particional que tiene como objetivo dividir un conjunto de  $n$  observaciones en  $k$  clústeres en los que cada observación pertenece al clúster con la media más cercana [19, 20]. Este procedimiento sigue un enfoque simple y eficaz para particionar el conjunto de datos en clústeres disjuntos, minimizando la varianza dentro de cada agrupación.

La formulación del algoritmo se basa en encontrar los centroides de los clústeres,  $c_i$ , que minimizan la suma de las distancias cuadradas entre cada punto de datos  $x_j$  y el centroide de su respectivo clúster. La función objetivo, conocida como la inercia o la suma de cuadrados dentro del clúster (WCSS), se expresa como:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - c_i\|^2, \quad (2)$$

donde  $C_i$  es el conjunto de puntos asignados al clúster  $i$  y  $c_i$  es el centroide de  $C_i$ . La norma usualmente es la distancia euclidiana, aunque K-Means puede ser adaptado a otras métricas de distancia.

### 2.3.2. Algoritmo de agrupamiento espectral

El agrupamiento espectral es un método de agrupamiento particional. Es una técnica que aplica teoría espectral de grafos para realizar la partición de un conjunto de datos en grupos que pueden no ser linealmente separables. Es ampliamente conocido por su versatilidad y capacidad para detectar la estructura de los datos basándose en las propiedades de los autovectores de las matrices derivadas del grafo de similitudes de los datos [21].

El paso crítico en este proceso es la selección de los autovectores apropiados, lo que se puede expresar como la minimización del siguiente criterio:

$$\min_H \text{Tr}(H^T L H) \quad \text{con la condición} \quad H^T H = I, \quad (3)$$

donde  $H$  es la matriz que contiene los vectores propios seleccionados y  $I$  es la matriz identidad. Este enfoque garantiza que los puntos de datos que son similares queden mapeados a puntos cercanos en el espacio de características de baja dimensión formado por los vectores propios, lo que facilita su posterior agrupación [22].

#### 2.4. Algoritmos de reducción de dimensionalidad

La multicolinealidad se refiere a la interrelación entre factores de riesgo y presenta problemas porque: usualmente es difícil de detectar y; cuando esta presente, podría afectar a la eficacia del análisis estadístico adversamente. Si esta presente, una severa multicolinealidad dificulta significativamente la cuantificación y clasificación de las contribuciones de los factores individuales que se consideren importantes[23]. Una forma de mitigar la multicolinealidad es empleando algoritmos de reducción de dimensionalidad. Estos algoritmos se encargan de reducir la dimensión de los datos de entrada, conservando exclusivamente los datos mas significativos para el modelo.

- **Principal Component Analysis(PCA):** Supongamos que tenemos  $n$  mediciones en un vector  $x$  de  $p$  variables aleatorias y deseamos reducir la dimensión de  $p$  a  $q$ . El PCA hace esto al hallar las combinaciones lineales  $a'_1x, a'_2x, \dots, a'_qx$ , denominado *componentes principales*, que sucesivamente tienen la varianza máxima de datos, sujetos a ser no correlacionados con los anteriores  $a'_kx$ .

La resolución de este problema de maximización nos permite obtener los vectores  $a_1, a_2, \dots, a_q$ , los cuales son los *eigenvectores* de la *matriz de covarianza*  $S$ , de los datos, correspondientes a los  $q$  valores propios mas grandes. Estos valores propios dan las varianzas de sus respectivos componentes principales, y la relación entre la suma de los primeros  $q$  valores propios y la suma de las varianzas de todas las  $p$  variables originales representa la proporción de la varianza total en el conjunto de datos original, explicada por el primer  $q$  componentes principales[24].

- **Feature Agglomeration(FA):** La aglomeración de características (*feature agglomeration*) es una técnica de reducción de dimensionalidad que agrupa variables similares en un conjunto único, similar al clustering jerárquico aplicado a las características en lugar de a las muestras. Esta técnica reduce el número de variables basándose en la similitud [25].

El criterio para la aglomeración puede basarse en la distancia entre características. Si se utiliza la distancia euclidiana, la similitud entre dos conjun-

tos de características  $X_i$  y  $X_j$  se puede calcular como:

$$d(X_i, X_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (4)$$

donde  $x_{ik}$  y  $x_{jk}$  son los valores de las características  $i$  y  $j$  para la muestra  $k$ , respectivamente, y  $n$  es el número total de muestras.

La aglomeración se realiza iterativamente de la siguiente manera:

1. Calcular la matriz de proximidad  $D$  basada en la distancia escogida entre todas las características.
2. Fusionar las características con la menor distancia en  $D$ .
3. Actualizar la matriz de proximidad  $D$  para reflejar la fusión, recalculando las distancias.
4. Repetir los pasos 2 y 3 hasta alcanzar el número deseado de características.

El resultado final es un conjunto de características reducido que puede utilizarse para análisis posterior o la construcción de modelos [26].

#### 2.5. Selección del numero de agrupaciones

Tanto en los algoritmos jerárquicos como en los particionales, un aspecto crucial es determinar el número de clústeres. La determinación de este número no es una tarea trivial y suele depender en gran medida del contexto y naturaleza de los datos en estudio[27].

Aunque no hay una técnica universalmente aceptada para determinar el número óptimo de clústeres, el método del codo es ampliamente reconocido y utilizado en la comunidad científica. Esta técnica se basa en una evaluación iterativa de diferentes valores de  $K$  (número de clústeres). Para cada valor de  $K$ , se calcula una métrica, como la distorsión o la inercia, que representa la cohesión interna de los clústeres. Idealmente, el valor óptimo de  $K$  es aquel donde la métrica comienza a mostrar ganancias marginales, es decir, cuando la curva comienza a aplanarse.

La métrica de distorsión, comúnmente citada en la literatura, es una medida que cuantifica cuán dispersos están los puntos en un clúster en relación con su centroide. A menudo se define como el promedio de las distancias cuadradas de los puntos a su centroide.

##### 2.5.1. Estadística de gap

Al implementar los algoritmos de reducción de dimensiones mencionados en la sección 2.4, se puede implementar el estadística de gap[28] para determinar el numero óptimo de agrupamientos para el algoritmo de agrupación seleccionado.

La estadística de gap es una metodología para la estimación del número de aglomeraciones a utilizar para un set de datos. La técnica utiliza el resultado de cualquier algoritmo de agrupamiento, comparando el cambio en la dispersión dentro del grupo con el esperado bajo una distribución nula de referencia apropiada. La estadística de gap para un número dado de aglomeraciones  $k$  se define como:

$$\text{Gap}_n(k) = E_n^*\{\log(W_k)\} - \log(W_k) \quad (5)$$

donde  $W_k$  es la suma de las distancias al cuadrado de cada punto al centroide de su clúster, y  $E_n^*\{\log(W_k)\}$  es el valor esperado de  $\log(W_k)$  bajo una muestra de referencia de la distribución nula. El valor óptimo de  $k$  es entonces aquel que maximiza la estadística de gap.

### 2.6. Clustering Bootstrap

El clustering bootstrap es una técnica utilizada para evaluar la estabilidad y la validez de los agrupamientos identificados en un conjunto de datos. Esta técnica fue explorada en detalle por [29] y [30].

La metodología general del clustering bootstrap implica la generación de múltiples conjuntos de datos a través del muestreo con reemplazo (bootstrap) del conjunto de datos original. Luego, se aplica un algoritmo de agrupamiento a cada uno de estos conjuntos de datos generados. La estabilidad de los agrupamientos se evalúa observando la consistencia de los resultados del agrupamiento a través de estos múltiples conjuntos de datos bootstrap.

## 3. AGRUPAMIENTO DE LOS DATOS

La ANDE realizó un proyecto piloto para obtener los datos del consumo eléctrico de diferentes categorías de clientes a lo largo del país. Sin embargo, aun no cuenta con modelos que les permita definir perfiles de consumo. En esta sección, se realizará el agrupamiento del perfil de consumo de los clientes de ANDE.

### 3.1. Datos

En 2021, la Administración Nacional de Energía (ANDE) llevó a cabo un experimento piloto para recopilar datos de varias categorías de consumidores, incluidas 141 (social), 142 (residencial), 245 (comercial), 343 (industrial), 410 (otros), 546 (general) y 846 (gubernamental). Cuadrillas de funcionarios rastrearon las zonas que contaban con medidores inteligentes, que tenían la capacidad de almacenar datos del consumo horario de los clientes durante un determinado periodo de tiempo, y subieron los datos a una base de datos. En la Tabla 1 se puede visualizar la distribución de los clientes según su región y categoría.

Como los datos de todos los medidores no se extrajeron al mismo tiempo, fue necesario establecer

una franja temporal en la cual se realizara el análisis. La franja seleccionada fue desde el 8 de enero a las 00:00:00 horas hasta el 30 de enero a las 23:00:00 horas.

**Tabla 1:** Cantidad de clientes por categoría según la región

Region	Categorías					
	141	142	245	343	546	846
Metropolitano	50	122	118	72	93	45
Este	0	76	52	0	27	12
Sur	13	37	52	0	29	21

**Fuente:** Elaboración propia.

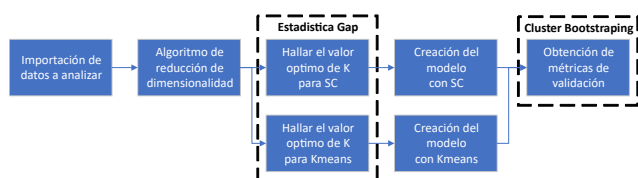
### 3.2. Metodología

En los siguientes experimentos se sigue la misma metodología descrita en la Figura 2. A continuación, se describirá detalladamente los pasos a realizar:

- **Importación de datos a analizar:** Se seleccionan los datos de la zona de interés y la categoría a analizar.
- **Algoritmo de reducción de dimensionalidad:** En todos los experimentos se aplican los algoritmos de PCA y FA a los datos antes de obtener el valor óptimo de  $K$ , de tal forma a trabajar con los datos más significativos.
- **Hallar el valor óptimo de  $K$  para SC/Kmeans:** Se aplica la estadística de gap para así obtener el valor óptimo de  $K$  con el cual crearemos nuestros modelos de agrupación espectral y Kmeans.
- **Creación del modelo con SC/Kmeans:** A partir del modelo óptimo de  $K$  obtenido en el paso anterior, se crean los modelos de agrupación espectral y Kmeans.
- **Obtención de métricas de validación:** Se aplica el clustering bootstrap para evaluar la estabilidad y la validez de los agrupamientos identificados en el conjunto de datos. También se compara los resultados entre las 4 posibles combinaciones de algoritmos de reducción de dimensionalidad y algoritmos de agrupamiento, de tal forma a seleccionar a la combinación que genere los resultados más óptimos.

#### 3.2.1. Primer Experimento

En el trabajo de referencia[31] se crearon modelos utilizando 4 combinaciones de algoritmos: FA+SC, FA+KMC, PCA+SC y PCA+KMC. En la Tabla 2 se presentan los resultados de dicho trabajo, en donde



**Figura 2:** Flujograma de los experimentos. **Fuente:** Elaboración propia.

podemos verificar que los modelos con el mejor rendimiento se obtuvieron con la combinación de algoritmos PCA+KMC.

El primer experimento consiste en realizar la aglomeración de clientes para corroborar que los resultados presentados son o no replicables. Para descartar la posibilidad de que esta combinación esta ligada exclusivamente a un tipo de consumidor, por ejemplo, el consumidor residencial. En este experimento se realiza la agrupación de clientes de distintas categorías del área metropolitana. Los clientes a agrupar pertenecen a las categorías 142(residencial) e industrial(343).

Si en ambas categorías el modelo con mayor rendimiento es el que combina los algoritmos PCA+KMC entonces, podemos afirmar que los resultados del material de referencia son replicables, y no están ligados a solo un tipo de clientes. Se utilizara la estadística de gap para determinar el numero óptimo de aglomeraciones a utilizar.

**Tabla 2:** Resultados de los experimentos realizados en el material de referencia

p	Framework	% Aciertos	% Fallos
2	FA+SC	22.67	77.33
2	FA+KMC	29.07	70.93
2	PCA+SC	18.78	81.22
2	PCA+KMC	<b>76.28</b>	23.72
3	FA+SC	21.34	78.66
3	FA+KMC	24.98	75.02
3	PCA+SC	17.60	82.40
3	PCA+KMC	<b>67.15</b>	32.85

**Fuente:** [31]

### 3.2.2. Segundo Experimento

Una vez validado los resultados del material de referencia, se procede a realizar el agrupamiento de clientes de una misma categoría, pero de zonas distintas. Es decir, realizar el agrupamiento de los clientes de la categoría 142(residencial) del área metropolitana, este y sur del país.

Estos agrupamientos nos permiten determinar si los consumidores de diversas zonas del país poseen el mismo patrón de consumo. De ser así, sería necesario realizar una agrupación de una categoría a ni-

vel país para identificar los patrones de consumo del consumidor residencial paraguayo. Caso contrario, el resultado del experimento nos proveerá un panorama sobre el consumo del cliente residencial en cada zona del país.

## 3.3. Resultados

### 3.3.1. Primer experimento

Comparando los resultados entre las Tablas 2, 3 y 4, podemos apreciar que el mejor modelo para todos los casos se obtiene al combinar PCA con KMC. Por lo tanto, hemos confirmado la hipótesis de que los resultados del material de referencia son replicables sin importar el tipo de cliente que estamos analizando.

Asimismo, también podemos definir el perfil de los clientes de la categoría 142 y los clientes de la categoría 343 del área metropolitana. A continuación, se realizara la descripción de dichos perfiles.

**Tabla 3:** Resultados del primer experimento para clientes de la categoría residencial(142)

p	Framework	% Aciertos	% Fallos
2	FA+SC	16.02	83.98
2	FA+KMC	45.73	54.27
2	PCA+SC	18.37	81.63
2	PCA+KMC	<b>55.45</b>	44.55
3	FA+SC	15.02	84.98
3	FA+KMC	44.83	55.17
3	PCA+SC	16.95	83.05
3	PCA+KMC	<b>49.25</b>	50.75

**Fuente:** Elaboración propia

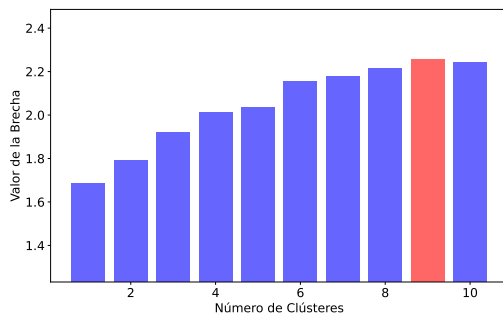
**Tabla 4:** Resultados del primer experimento para clientes de la categoría industrial(343)

p	Framework	% Aciertos	% Fallos
2	FA+SC	20.68	79.32
2	FA+KMC	54.59	45.40
2	PCA+SC	14.01	85.99
2	PCA+KMC	<b>88.67</b>	11.33
3	FA+SC	19.98	80.02
3	FA+KMC	53.65	46.35
3	PCA+SC	14.88	85.12
3	PCA+KMC	<b>85.56</b>	14.44

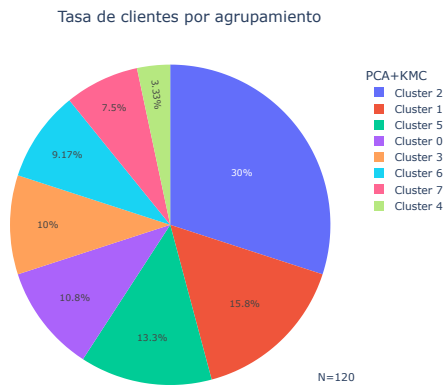
**Fuente:** Elaboración propia



### 3.3.2. Categoría metropolitana 142



**Figura 3:** Resultados de la estadística de gap para la categoría metropolitana 142. **Fuente:** Elaboración propia.

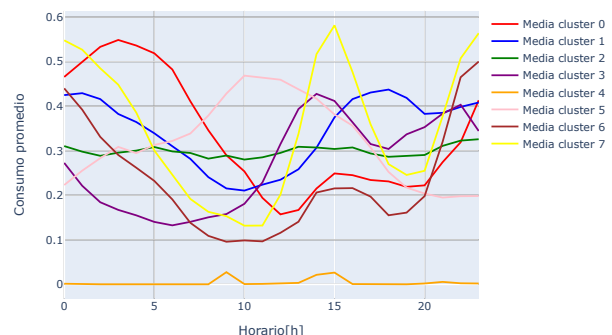


**Figura 4:** Tasa de clientes por agrupación de la categoría metropolitana 142. **Fuente:** Elaboración propia.

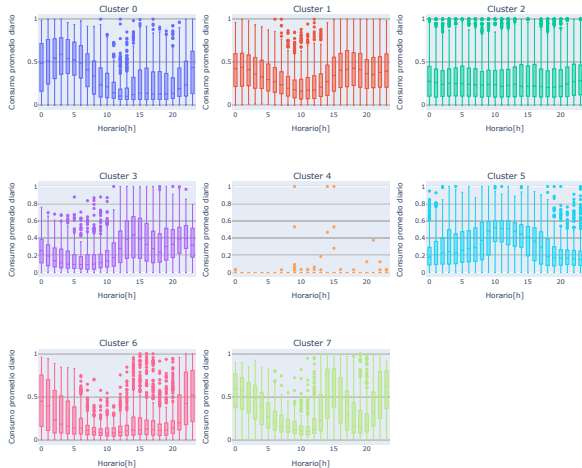
La cantidad de agrupaciones recomendada por la estadística de gap es nueve. Las agrupaciones están distribuidas como se muestra en la figura ???. Sin embargo, como el cluster 8 contiene solo a un cliente, esa agrupación no posee un tamaño representativo, por lo que fue descartado en el perfilamiento final de clientes. Quedándonos finalmente con la distribución presentada en la figura 4. A continuación, se describirán los perfiles de consumo de cada agrupación.

- **Cluster 0:** Presenta un pico de consumo en la noche, lo que podría ser indicativo de uso residencial cuando las personas están en casa después del trabajo, utilizando electrodomésticos, calefacción, aire acondicionado, termocalefón, entre otros.
- **Cluster 1:** Con picos en la mañana y la noche, este patrón sugiere actividad doméstica típica al comienzo y al final del día, posiblemente debido a la preparación de comidas y actividades de relajación nocturna.

- **Cluster 2:** Exhibe un consumo más uniforme con un leve aumento en las horas de la tarde. Esto podría ser representativo de hogares con ocupantes presentes durante todo el día, como aquellos con trabajadores a distancia o personas jubiladas.
- **Cluster 3:** La curva tiene un aumento sostenido desde la tarde hasta la noche. Este perfil puede ser común en hogares donde los ocupantes pasan el día fuera y regresan en la tarde.
- **Cluster 4:** Exhibe un consumo relativamente bajo y estable a lo largo del día, lo que podría indicar un uso eficiente de la energía o que la vivienda está desocupada durante gran parte del día.
- **Cluster 5:** Con un pico temprano en la mañana y un consumo más bajo durante el resto del día, este perfil podría corresponder a hogares donde los ocupantes se preparan para salir temprano, como ir a la escuela o al trabajo.
- **Cluster 6:** Presenta un patrón de dos picos, uno en la mañana y otro en la noche, típico de los horarios de inicio y fin de día en hogares con rutinas tradicionales de trabajo y escuela.
- **Cluster 7:** Muestra picos de consumo en la tarde y una disminución hacia la noche, lo que puede reflejar actividades específicas de tarde, como la cocción de comidas o el uso de entretenimiento en el hogar.

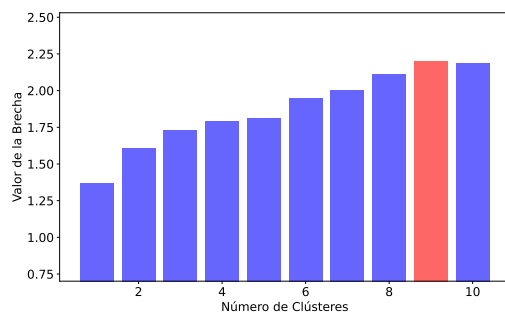


**Figura 5:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Metropolitano 142. **Fuente:** Elaboración propia.

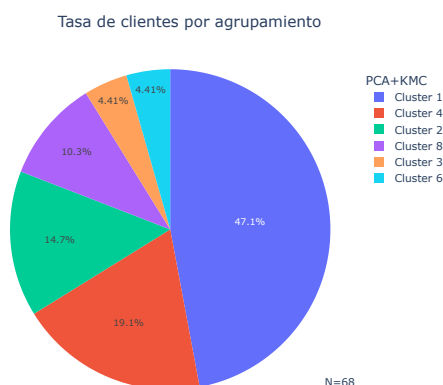


**Figura 6:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Metropolitano 142. **Fuente:** Elaboración propia.

### 3.3.3. Categoría metropolitana 343



**Figura 7:** Resultados de la estadística de gap para la categoría metropolitana 343. **Fuente:** Elaboración propia.



**Figura 8:** Tasa de clientes por agrupación de la categoría metropolitana 343. **Fuente:** Elaboración propia.

La cantidad de agrupaciones recomendada por la estadística de gap es nueve. Las agrupaciones están distribuidas como se muestra en la figura ???. En este caso, las agrupaciones 0, 5 y 7 no poseen un tamaño representativo, por lo cual fueron descartadas. La distribución final de etiquetas se muestra en la figura 8. A continuación, se describirán los perfiles de consumo de cada agrupación.

- **Cluster 1:** esta curva muestra un consumo relativamente estable con una ligera disminución durante la media mañana. Esto podría indicar una industria que opera constantemente, como una planta de tratamiento de agua o una instalación de servidores que requiere un consumo de energía consistente.

- **Cluster 2:** Esta curva tiene un patrón irregular con varios picos y valles. Podría indicar una operación con ciclos de trabajo discontinuos o con máquinas que se encienden y apagan frecuentemente, como una planta de fabricación con diferentes líneas de producción.

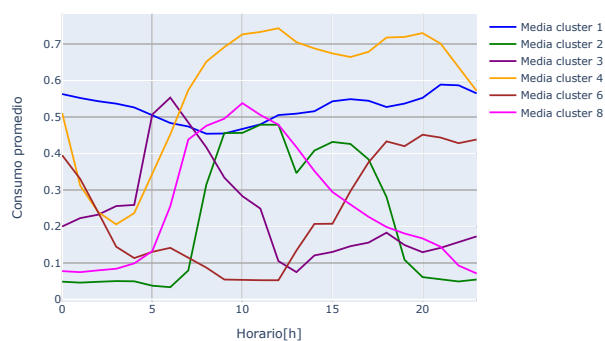
- **Cluster 3:** Esta curva tiene un pico pronunciado en las primeras horas y luego disminuye gradualmente. Esto puede corresponder a una industria que realiza procesos intensivos en energía en la mañana, como una panadería industrial o una fábrica de metales que calienta materiales al inicio de su jornada.

- **Cluster 4:** Esta curva tiene un pico en la mañana y otro en la tarde con una caída significativa entre ellos. Podría ser representativo de una industria con dos turnos de trabajo intensos, como una manufactura de bienes de consumo que opera con mayor intensidad al comenzar el día y antes de finalizar la jornada laboral.

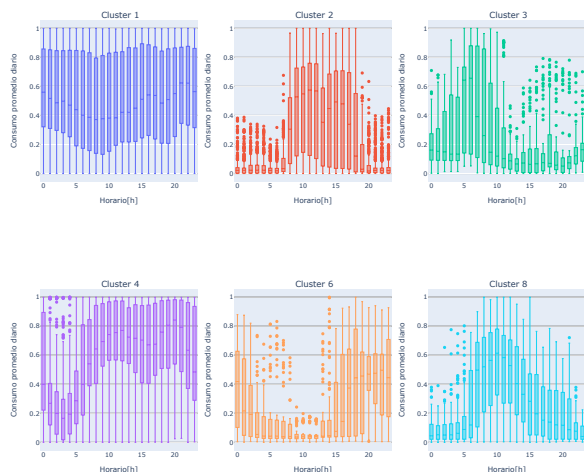
- **Cluster 6:** La curva tiene una caída en el consumo al principio, seguida de un pico en la tarde. Esto puede sugerir una operación que se intensifica en la tarde, como una fábrica que aumenta su producción en un turno específico.

- **Cluster 8:** La curva muestra un aumento gradual durante el día con un pico en la tarde, lo que puede reflejar una industria que aumenta su carga de trabajo a lo largo del día hasta alcanzar un máximo en la tarde, como puede ser una empresa de logística y envíos que tiene más actividad antes de la noche.





**Figura 9:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Metropolitano 343. **Fuente:** Elaboración propia.



**Figura 10:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Metropolitano 343. **Fuente:** Elaboración propia.

### 3.4. Segundo experimento

Para reducir la cantidad de gráficos de barra presentados en esta sección, se recurrió a elaborar la Tabla 5, en la cual podemos visualizar los resultados de la estadística de gap, junto con la K que finalmente fue utilizada.

**Tabla 5:** Resultados del experimento 2

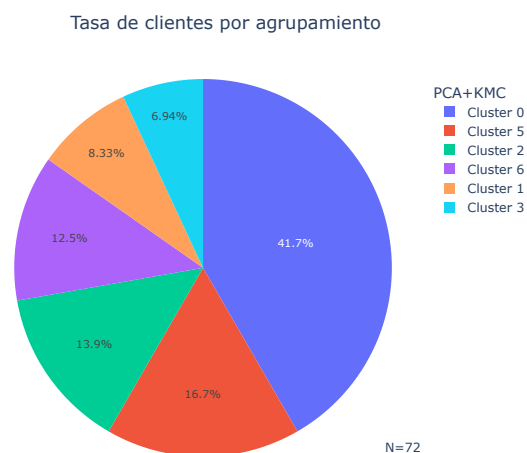
Area	K(Gap Statistic)	K(utilizada)
Metropolitana	9	8
Este	9	6
Sur	8	5

**Fuente:** Elaboración propia

#### 3.4.1. Categoría metropolitana 142

Los agrupamientos son exactamente los mismos que los agrupamientos presentados en la sección 3.3.2.

#### 3.4.2. Categoría este 142



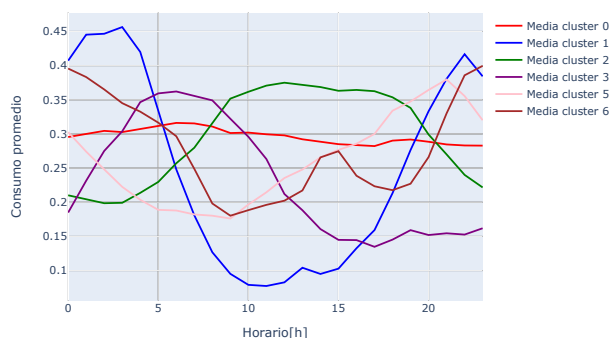
**Figura 11:** Tasa de clientes por agrupación de la categoría este 142. **Fuente:** Elaboración propia.

La cantidad de agrupaciones recomendada por la estadística de gap es nueve. En este caso, las agrupaciones 4, 7 y 8 no poseen un tamaño representativo, por lo cual fueron descartadas. La distribución final de etiquetas se muestra en la figura 11. A continuación, se describirán los perfiles de consumo de cada agrupación.

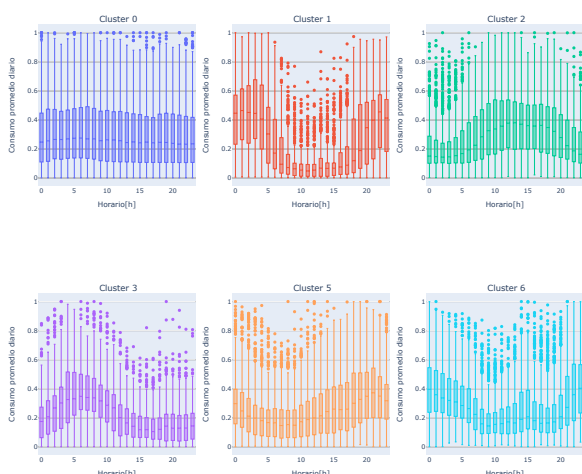
- **Cluster 0:** Presenta un consumo bastante constante durante el día con una leve disminución en la noche. Esto puede sugerir presencia continua en el hogar con menos actividad nocturna, posiblemente indicando ocupantes como personas que trabajan desde casa o están retiradas.
- **Cluster 1:** Exhibe un patrón bimodal, con picos de consumo en las horas de la mañana y de la noche, lo cual es común en residencias donde las actividades principales se llevan a cabo antes de ir al trabajo o la escuela y después de regresar a casa.
- **Cluster 2:** Con un pico pronunciado en la mitad del día y un consumo reducido por la noche, podría representar hogares con actividades centradas en la tarde, como el uso de aparatos de cocina o sistemas de climatización en horas pico.
- **Cluster 3:** Tiene un pico de consumo por la mañana y cae por la tarde. Este consumo puede ser

de personas que tienen un turno tarde a nocturno y solo están en sus hogares por la mañana.

- **Cluster 5:** Muestra un aumento del consumo en la noche. Esto puede indicar hogares donde la actividad principal ocurre en este periodo, tal vez debido al uso de electrodomésticos, sistemas de calefacción o enfriamiento después del regreso del trabajo.
- **Cluster 6:** Esta curva tiene un consumo relativamente estable con un pequeño aumento en la tarde, lo que podría reflejar una rutina diaria constante con un ligero incremento en la actividad al final del día laboral o escolar.

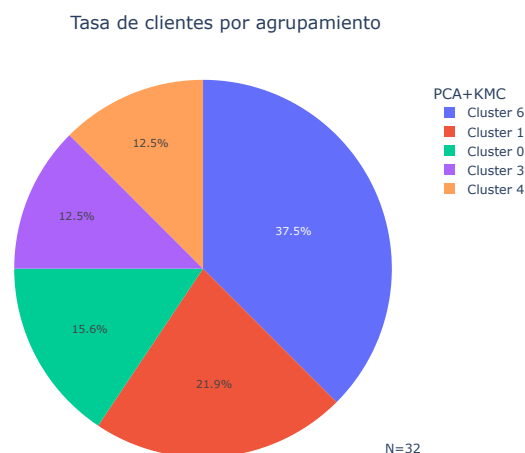


**Figura 12:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Este 142. **Fuente:** Elaboración propia.



**Figura 13:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Este 142. **Fuente:** Elaboración propia.

### 3.4.3. Categoría sur 142



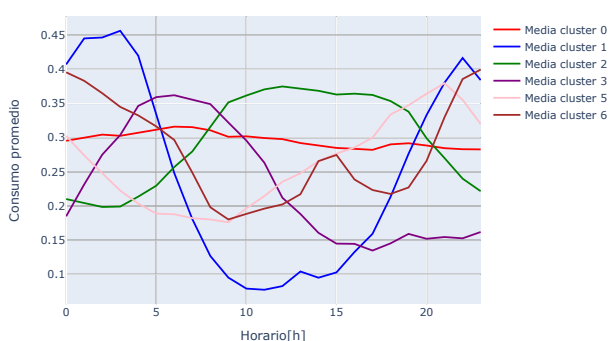
**Figura 14:** Tasa de clientes por agrupación de la categoría sur 142. **Fuente:** Elaboración propia.

La cantidad de agrupaciones recomendada por la estadística de gap es ocho. Las agrupaciones están distribuidas como se muestra en la figura ???. En este caso, las agrupaciones 4, 7 y 8 no poseen un tamaño representativo, por lo cual fueron descartadas. La distribución final de etiquetas se muestra en la figura 17. A continuación, se describirán los perfiles de consumo de cada agrupación.

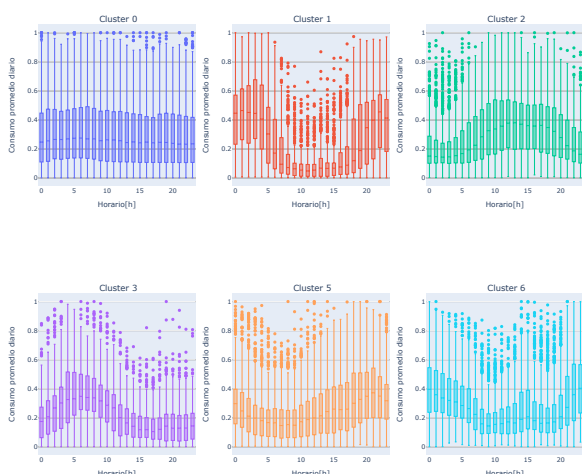
- **Cluster 0:** Esta curva muestra un consumo decreciente desde la madrugada hasta las primeras horas de la mañana y luego se mantiene estable. Esto podría indicar hogares que reducen su actividad después de la noche, tal vez debido a que la mayoría de los electrodomésticos y luces se apagan cuando las personas se van a dormir.
- **Cluster 1:** Muestra un consumo elevado por la tarde, alcanzando el pico cerca de medianoche, presumiblemente por el uso de acondicionadores de aire junto con termocalefones. Este perfil de consumo pertenece a las personas que vuelven de su trabajo a la siesta, y permanecen en sus hogares el resto del día.
- **Cluster 3:** Con un perfil de consumo en forma de U, esta curva indica un mayor uso de energía por la mañana y la noche, lo que es típico de hogares donde las personas están presentes solo antes de salir y después de regresar a casa.
- **Cluster 4:** Exhibe un pico de consumo alrededor de las 15 horas, lo que puede sugerir una tarde de actividad intensa en el hogar, posiblemente

debido a la calefacción o refrigeración, la preparación de comidas, o el uso de electrodomésticos y entretenimiento.

- **Cluster 6:** Presenta un consumo bastante constante con un pico notable en la tarde. Este perfil puede ser común en hogares con ocupantes presentes durante todo el día, como aquellos con trabajadores a distancia o personas que no salen para ir a trabajar o a la escuela.



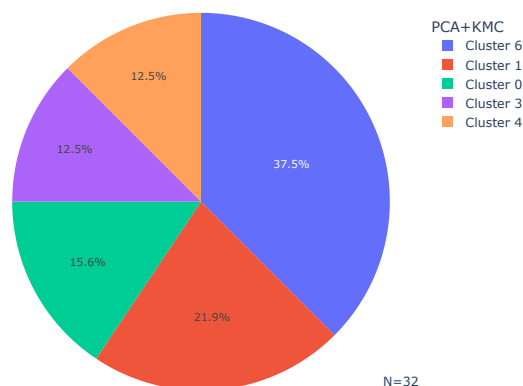
**Figura 15:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Este 142. **Fuente:** Elaboración propia.



**Figura 16:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Este 142. **Fuente:** Elaboración propia.

## 3.4.4. Categoría sur 142

Tasa de clientes por agrupamiento



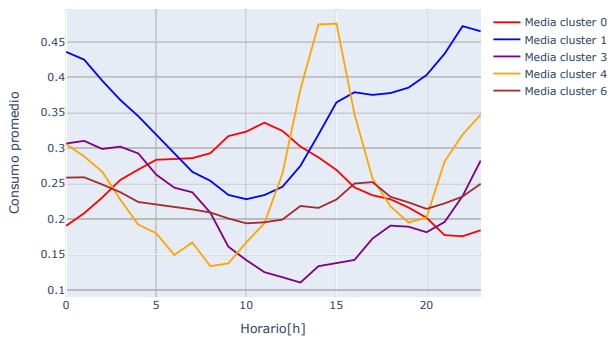
**Figura 17:** Tasa de clientes por agrupación de la categoría sur 142. **Fuente:** Elaboración propia.

La cantidad de agrupaciones recomendada por la estadística de gap es ocho. Las agrupaciones están distribuidas como se muestra en la figura ???. En este caso, las agrupaciones 4, 7 y 8 no poseen un tamaño representativo, por lo cual fueron descartadas. La distribución final de etiquetas se muestra en la figura 17. A continuación, se describirán los perfiles de consumo de cada agrupación.

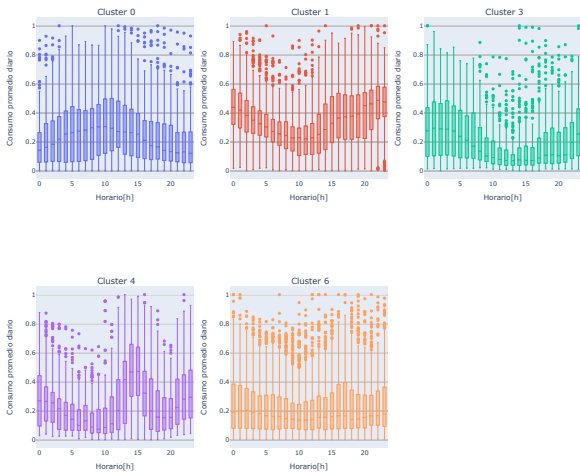
- **Cluster 0:** Esta curva muestra un consumo decreciente desde la madrugada hasta las primeras horas de la mañana y luego se mantiene estable. Esto podría indicar hogares que reducen su actividad después de la noche, tal vez debido a que la mayoría de los electrodomésticos y luces se apagan cuando las personas se van a dormir.
- **Cluster 1:** Muestra un consumo elevado por la tarde, alcanzando el pico cerca de medianoche, presumiblemente por el uso de acondicionadores de aire junto con termocalefones. Este perfil de consumo pertenece a las personas que vuelven de su trabajo a la siesta, y permanecen en sus hogares el resto del día.
- **Cluster 3:** Con un perfil de consumo en forma de U, esta curva indica un mayor uso de energía por la mañana y la noche, lo que es típico de hogares donde las personas están presentes solo antes de salir y después de regresar a casa.
- **Cluster 4:** Exhibe un pico de consumo alrededor de las 15 horas, lo que puede sugerir una tarde de actividad intensa en el hogar, posiblemente

debido a la calefacción o refrigeración, la preparación de comidas, o el uso de electrodomésticos y entretenimiento.

- **Cluster 6:** Presenta un consumo bastante constante con un pico notable en la tarde. Este perfil puede ser común en hogares con ocupantes presentes durante todo el día, como aquellos con trabajadores a distancia o personas que no salen para ir a trabajar o a la escuela.



**Figura 18:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Sur 142. **Fuente:** Elaboración propia.



**Figura 19:** Curvas de consumo diario promedio normalizado de las agrupaciones para la categoría Sur 142. **Fuente:** Elaboración propia.

#### 4. Diseño e implementación de hardware

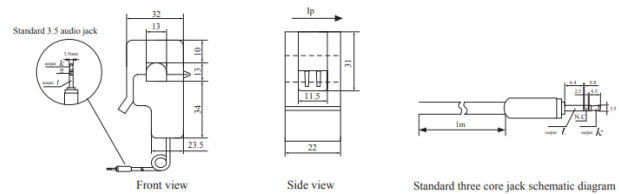
En esta sección se describirá el diseño e implementación de la placa que se ha diseñado para realizar muestreos de consumo eléctrico.

A continuación, se explicará con mayor detalle cada parte del circuito.

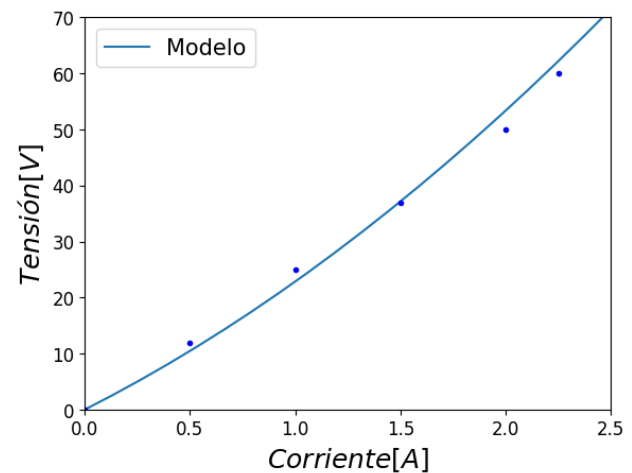
##### 4.0.1. Sensor inductivo

El sensor que nos permitirá realizar el muestreo de forma no invasiva en la acometida de ANDE es el SCT013-50 de YHDC. Este sensor tiene la capacidad de realizar mediciones de circuitos de 220 Voltios hasta 50 Ampere. Esta capacidad lo hace ideal para mediciones de hogares trifásicos, donde las llaves son de 32 Ampere para arriba. La salida del sensor es un jack de audio, lo cual nos facilita la conexión y desconexión de los sensores en caso de una avería.

Cuando el sensor envuelve a un cable, se genera una tensión en la salida del sensor. Esta tensión es medida por la placa embebida y, tras un preprocesamiento de datos, se envían las mediciones a las bases de datos.



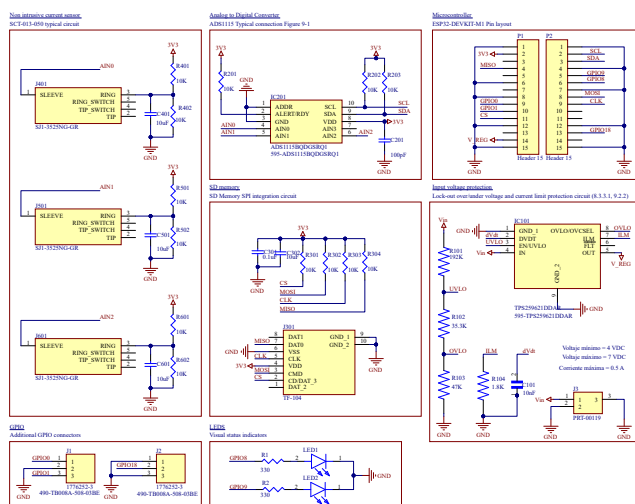
**Figura 20:** Dimensiones del sensor inductivo. **Fuente:** [32].



**Figura 21:** Curva característica del sensor inductivo. **Fuente:** Elaboración propia.

##### 4.0.2. Esquemático

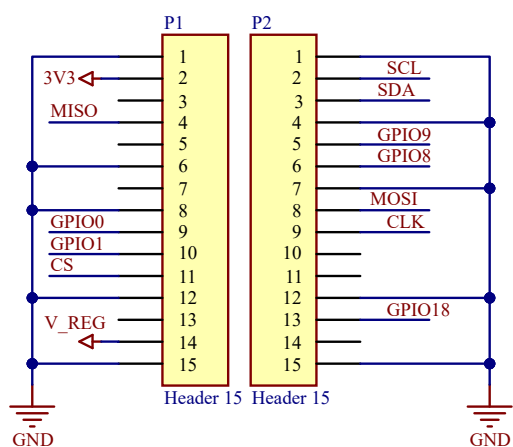
El circuito de la placa está compuesto por los siguientes bloques:



**Figura 22:** Esquemático general del circuito. **Fuente:** Elaboración propia.

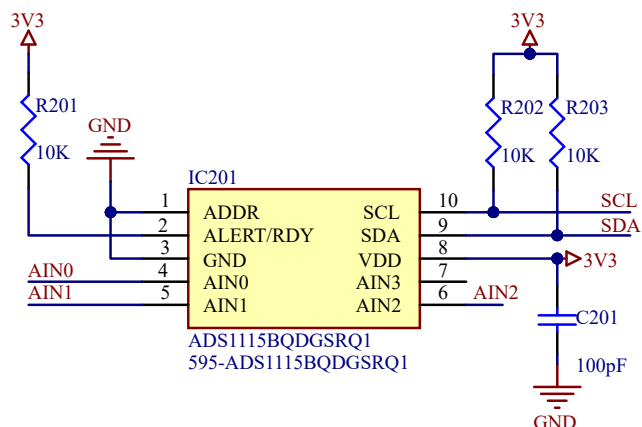
A continuación, se realizara una descripción de la funcionalidad de cada bloque del esquemático.

**El microcontrolador** Como estoy utilizando equipos prestados del Laboratorio de Mecánica y Energía, fue necesario diseñar la placa priorizando la extracción simplificada de los equipos prestados. Entonces, utilizo headers para realizar la conexión entre la placa y el microcontrolador con el ESP-32.



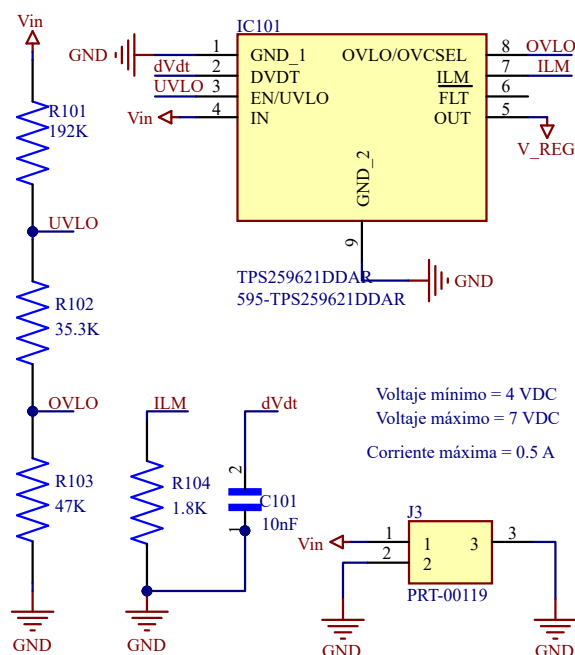
**Figura 23:** Esquemático del header para el microcontrolador. **Fuente:** Elaboración propia.

**Convertor analógico digital(ADC)** El ADS1115 es un convertor analógico digital de 16 bits de alta precisión y alta potencia, compatibles con I2C. Suele emplearse en módulos de cámara sin procesamiento, displays de información de centros automotrices como también en el monitoreo de la tensión y corriente para uso general. El esquemático de la figura 24 es el conexionado básico para la conexión I2C con el microcontrolador. Este integrado recibirá la señal proporcionada por el sensor inductivo, y realizara la conversión analógica digital de la señal.



**Figura 24:** Esquemático del ADC. **Fuente:** Elaboración propia.

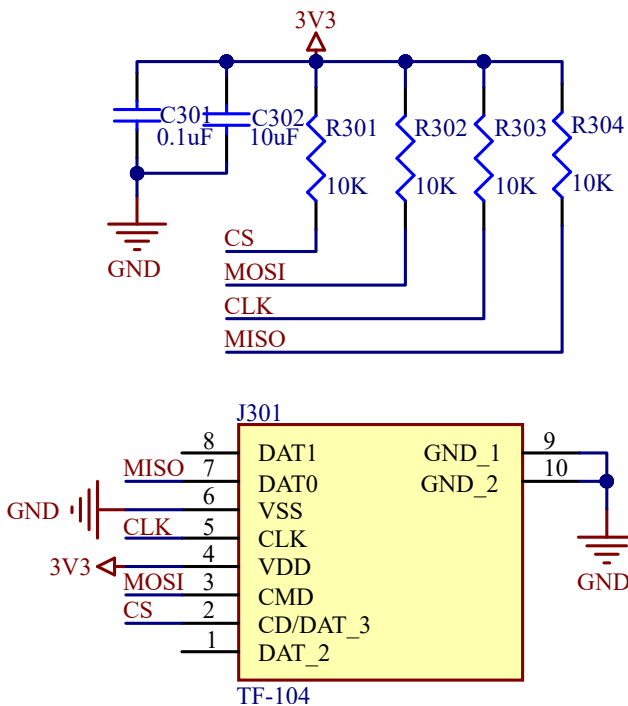
**Protección contra tensión de entrada** El TPS2596 es un circuito integrado con la capacidad de proteger el circuito contra la sobre tensión de entrada, opera en un amplio rango de voltaje (de 2.7 a 19 Volts), cuenta con un límite de corriente ajustable con monitor de corriente de carga, es inmune a transitorios eléctricos rápidos, posee una protección contra sobre temperatura, y su huella es pequeña, lo que nos permite integrarlo con facilidad en una placa. El esquema mostrado en la figura 25 es la conexión recomendada por el fabricante para proteger el circuito.



**Figura 25:** Esquemático de la protección contra sobre tensión. **Fuente:** Elaboración propia.

**Tarjeta SD** En caso de que exista alguna falla al conectar la placa a internet, es necesario prever un medio alternativo en donde almacenar los datos, al menos, temporalmente. Una opción de bajo costo y am-

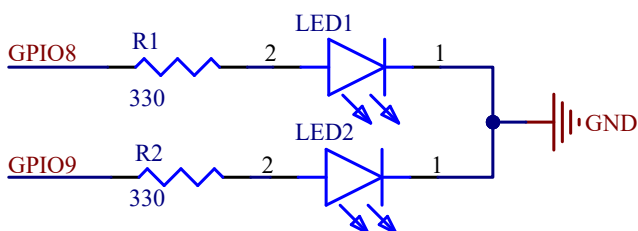
plamente disponible en todo el mundo es la tarjeta SD.



**Figura 26:** Esquemático de la tarjeta SD. **Fuente:** Elaboración propia.

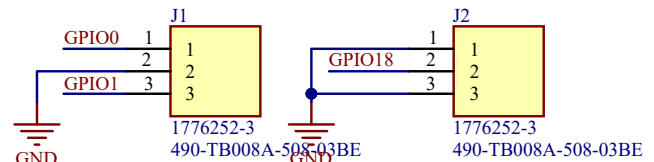
**Leds de estado** Como se priorizo la reducción de costos de producción en esta placa, la misma no cuenta con una interfaz gráfica (una pantalla) para identificar el estado en el que se encuentra. Como alternativa, añadí dos leds que nos permitirán identificar los siguientes estados:

- **Transmisión de datos:** Cuando la placa envíe datos a las bases, uno de los leds comenzará a titilar para indicar que se está realizando la transmisión de datos.
- **Conexión a la red:** Cuando la placa se haya conectado exitosamente a la red wifi para la cual fue configurada, el segundo led se encenderá para indicar dicho estado



**Figura 27:** Esquemático de los leds de estado. **Fuente:** Elaboración propia.

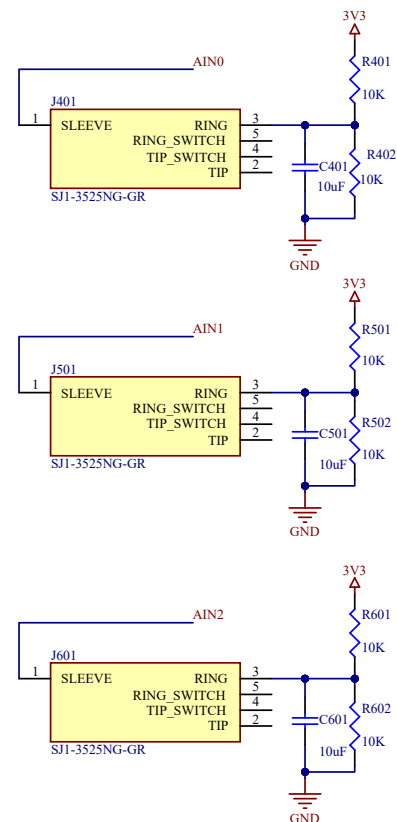
**Puertos GPIO adicionales** Se añadieron dos puertos GPIO adicionales para realizar posibles mejoras en el futuro. Una de las posibles mejoras es la adición de un sensor de voltaje en el sistema, para así asegurarnos de que la medición realizada por el sensor inductivo sea apropiada.



**Figura 28:** Esquemático de puertos GPIO adicionales. **Fuente:** Elaboración propia.

### Entradas de los sensores de corriente no intrusivos

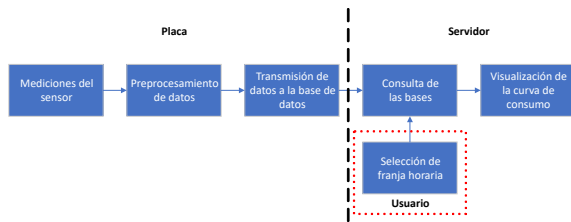
Los sensores a utilizar para este proyecto cuentan con una salida de jack para conectarse a cualquier dispositivo. Si bien también era una opción tajar dicha entrada para utilizar conectores tradicionales, opte por conservar dicha entrada para facilitar su extracción en caso de averías. La placa cuenta con 3 entradas de jack, una para cada fase de la acometida.



**Figura 29:** Esquemático de entradas de los sensores de corriente no intrusivos. **Fuente:** Elaboración propia.



#### 4.0.3. Software



**Figura 30:** Metodología implementada para la transmisión y visualización de los datos adquiridos. **Fuente:** Elaboración propia.

En la figura 30 se describe el sistema implementado desde el muestreo de datos hasta la visualización de los mismos en un dashboard. El procedimiento a seguir es el siguiente:

- **Mediciones del sensor:** El sensor realiza mediciones cada 5 milisegundos. Para ello, el mismo genera una tensión a partir del flujo electromagnético en su bobina. La figura 21 muestra la curva característica del tensión del sensor.
- **Preprocesamiento de datos:** La señal de corriente recibida por la placa es procesada por software. Esta corriente es la corriente pico  $I_p$ , la misma es introducida en la siguiente formula para obtener la potencia consumida:

$$P = \frac{24,84353434 * \frac{2 * I_p}{1000}}{1,41}$$

Para obtener la corriente pico a pico, se multiplica el dato por 2. Es necesario dividir el producto por 1000 para obtener el resultado en Amperes. Posteriormente se multiplica el resultado por 24,84353434 para ajustar el valor de la curva según la Figura 21 y, para trabajar con valores eficaces, se divide el resultado por 1,41 para obtener la corriente eficaz. Una vez obtenida la corriente eficaz, suponiendo una tensión eficaz constante de 220 Vrms, se realiza el producto entre la tensión y la corriente para así obtener la potencia eficaz consumida.

- **Transmisión de datos a la base de datos:** Los datos son transmitidos cada 60 segundos a la base de datos por medio de Wifi.
- **Selección de la franja horaria:** Por defecto, el dashboard mostrara los datos recibidos del día en el que se accede al mismo. Sin embargo, el usuario tiene la posibilidad de visualizar los datos en la franja horaria que desee. Inicialmente, los datos visualizados serán de un solo día.

- **Consulta a las bases:** A partir de los datos recibidos, se realiza una consulta a las bases de datos para obtener el consumo horario de las fases donde el equipo fue instalado.

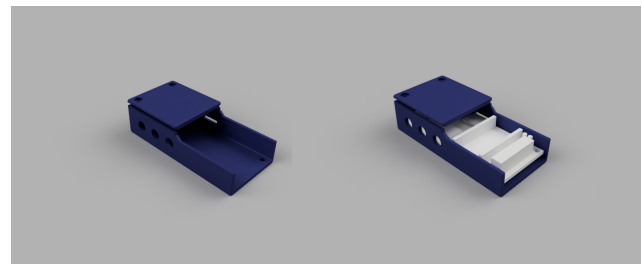
- **Visualización de la curva de consumo:** Es la visualización del consumo en la franja horaria por defecto, o la fijada por el usuario.

#### 4.0.4. Encapsulado

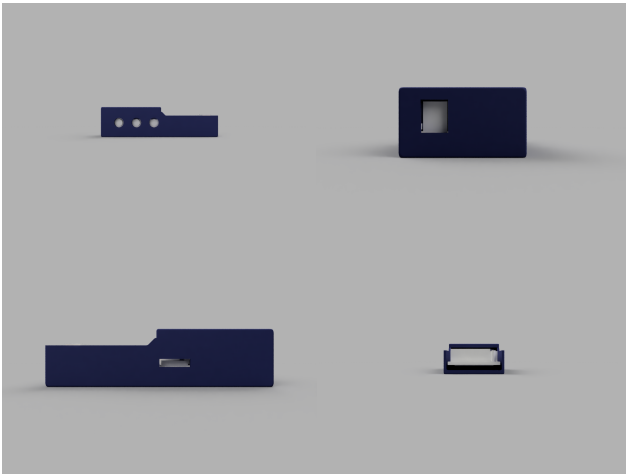
Para proteger al circuito de elementos externos, por ejemplo, el agua; es necesario elaborar un encapsulado a medida para el mismo. Teniendo en cuenta los conocimientos adquiridos en las materias de diseño dentro de la malla curricular, y la experiencia adquirida en los proyectos de diversas cátedras, opte por el diseño del encapsulado con impresión 3D. La impresión 3D presenta los siguientes beneficios:

- La impresión 3D te permite realizar diseños que serian imposibles de realizar con otros métodos de manufactura.
- El costo es relativamente bajo, lo cual hace que sea un material ideal para elaboración de prototipos.
- Los diseños son hechos a medida para su aplicación.
- El tiempo de impresión de encapsulados es relativamente bajo (entre dos a tres horas) por lo que se podría elaborar todos los encapsulados necesarios en un corto lapso de tiempo.

El modelo final a imprimir es visualizado en las figuras 31 y 32, donde se visualizan las vistas isométricas, con o sin la placa; y las vistas laterales del encapsulado.



**Figura 31:** Vistas isométricas renderizadas. **Fuente:** Elaboración propia.



**Figura 32:** Vistas laterales renderizadas. **Fuente:** Elaboración propia.

## 4.1. Resultados

### 4.1.1. Encapsulado

La distribución de los conectores y la forma que posee el microcontrolador con el ESP-32 presento ser todo un desafío al momento de realizar el diseño del encapsulado. Se priorizo cubrir los componentes SMD principales de la placa, y solo cubrir parcialmente las borneras, para facilitar el acceso a las mismas.

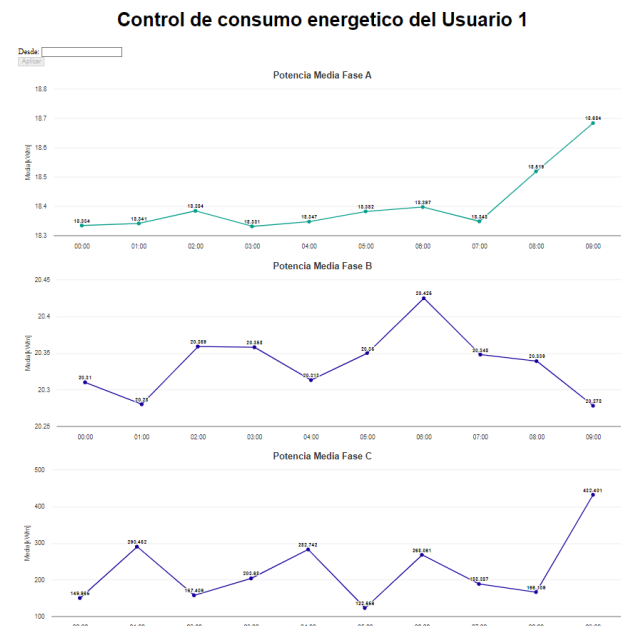


**Figura 33:** Tomas realizadas al encapsulado impreso. **Fuente:** Elaboración propia.

### 4.1.2. Dashboard

En la figura 34 se puede visualizar la versión final del dashboard de monitoreo. El mismo muestra los datos con saltos de 1 hora. Esta visualización fue obtenida al ingresar al dashboard a las 9:34 hs, por lo que solo muestra los datos hasta las 9:00 hs de ese día. Las curvas de consumo visualizadas representan el consumo de cada fase del tablero. Cada usuario que posea una de estas placas también posee un dashboard

para visualizar su consumo energético



**Figura 34:** Dashboard de consumo energético del usuario01. **Fuente:** Elaboración propia.

## 5. CONCLUSIÓN

En este trabajo se presento un estudio comparativo del método no supervisado de agrupamiento de series temporales para definir perfiles de consumo eléctrico.

Inicialmente se generaron modelos realizando las cuatro combinaciones posibles de PCA, SC, SC y KMC. Primeramente se aplicaron los algoritmos de reducción de dimensionalidad (PCA y SC), se halló el valor óptimo de K para SC y KMC por medio de la estadística de gap. Y finalmente se aplicó el cluster bootstrapping para determinar que combinación presenta los mejores resultados.

Para demostrar que la metodología es aplicable a cualquier sector, se realizaron las agrupaciones de las categorías residenciales e industriales del área metropolitana. Tras las validaciones se ha determinado que la mejor combinación de algoritmos es PCA+KMC para ambas categorías, validando así los resultados presentados en el material de referencia[31].

Una vez realizada la validación, se procedió a elaborar el agrupamiento de perfiles de consumo de categorías residenciales del área metropolitana, este y sur del país. Luego de descargar agrupaciones con cantidades no representativas, el área metropolitana quedó definida por 8 perfiles de consumo; el área este quedó definida por 6 perfiles de consumo; y el área sur quedó definida por 5 perfiles de consumo.

También se ha diseñado un sistema compuesto por una placa y una base de datos. La placa es capaz de

realizar mediciones de potencia directamente desde el tablero residencial de forma no intrusiva, enviando las mediciones realizadas cada minuto al servidor donde se encuentra la base de datos.

A partir de los datos disponibles en la base de datos, se elaboró un dashboard interactivo en el cual uno puede fijar una fecha para verificar el consumo eléctrico diario en el hogar.

## Referencias

- [1] A. A. Francisquini, "Estimação de curvas de carga em pontos de consumo e em transformadores de distribuição," 2006.
- [2] E. L. Bravo, "Metodología de la sumatoria de las curvas de cargas," *Ingeniería Energética*, vol. 24, no. 3, pp. 11–a, 2003.
- [3] Y. Wang, Q. Chen, C. Kang, Q. Xia, Y. Tan, Z. Zeng, and M. Luo, "Residential smart meter data compression and pattern extraction via non-negative k-svd," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.
- [4] J. Jardini, C. Tahan, M. Gouvea, S. Ahn, and F. Figueiredo, "Curvas de carga típicas e agregação dos consumidores residenciais, comerciais e industriais de baixa tensão," *4to Encontro Luso-Afro-Brasileiro de Planejamento e Exploração de Redes de Energia-ELAB*, pp. 391–397, 1999.
- [5] J. F. M. Pessanha and L. da Costa Laurencel, "Clustering electric load curves: The brazilian experience," in *Workshop Franco-Brésilien sur la Fouille des Données, Recife*, 2009.
- [6] M. M. F. Gemignani, "Proposição e análise comparativa de métodos alternativos de seleção e classificação de curvas de carga para a definição de tipologias para estudos tarifários." Ph.D. dissertation, Universidade de São Paulo, 2009.
- [7] F. F. Andrade *et al.*, "Formação de agrupamentos de consumidores de energia elétrica com base nos perfis de carga individuais," 2003.
- [8] L. A. G. Pareja, O. G. Carmona, A. Francisquini, and A. Padilha-Feltrin, "Estimación de curvas de demanda de consumidores, transformadores de distribución y alimentadores primarios en sistemas de distribución." *Scientia et Technica*, vol. 3, no. 35, pp. 107–112, 2007.
- [9] S. C. Cerchiari, "Determinação de curvas típicas de demanda de consumidores de baixa tensão utilizando mapas auto-organizáveis (som) para agrupamentos e conjuntos aproximados para classificação de consumidores," 2006.
- [10] R. Platon, V. R. Dehkordi, and J. Martel, "Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis," *Energy and Buildings*, vol. 92, pp. 10–18, 2015.
- [11] F. Morales, M. García-Torres, G. Velázquez, F. Dumas-Ladouce, P. E. Gardel-Sotomayor, F. Gómez-Vela, F. Divina, J. L. Vázquez Noguera, C. Sauer Ayala, D. P. Pinto-Roa *et al.*, "Analysis of electric energy consumption profiles using a machine learning approach: A paraguayan case study," *Electronics*, vol. 11, no. 2, p. 267, 2022.
- [12] D. d. E. E. Dirección de Planificación y Estudios, División de Estudios y Gestión de Inversiones, "Compilacion estadística 1999-2019," *Documentos Contables de la ANDE*, 2019.
- [13] M. G. Kendall, A. Stuart, and J. K. Ord, *Kendall's Advanced Theory of Statistics*. USA: Oxford University Press, Inc., 1987.
- [14] C. Chatfield, *Time-Series Forecasting*, 1st ed. Chapman and Hall/CRC - Estados Unidos, 2000.
- [15] C. Ratanamahatana, E. Keogh, A. Bagnall, and S. Lonardi, "A novel bit level time series representation with implication of similarity search and clustering," 05 2005, pp. 771–777.
- [16] L. A. Rodríguez-Astrain, "Medidas de disimilitud en series temporales," Ph.D. dissertation, 2017.
- [17] L. A. Rodríguez Astrain, "Medidas de disimilitud en series temporales," Ph.D. dissertation, 2017.
- [18] "Glosario de tecnología de medición." [Online]. Available: <https://www.elovis.com/es/empresa/glosario-de-tecnologia-de-medicion/>
- [19] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.
- [20] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [21] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, pp. 395–416, 2007.
- [22] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 14, 2001.

- [23] S. Lafi and J. Kaneene, "An explanation of the use of principal-components analysis to detect and correct for multicollinearity," *Preventive Veterinary Medicine*, vol. 13, no. 4, pp. 261–275, 1992.
- [24] I. Jolliffe, "Principal component analysis," *Encyclopedia of statistics in behavioral science*, 2005.
- [25] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.
- [26] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.
- [27] J. V. Bogado, D. H. Stalder, C. E. Schaerer, and S. Gómez-Guerrero, "Time series clustering to improve dengue cases forecasting with deep learning," *2021 XLVII Latin American Computing Conference (CLEI)*, pp. 1–10, 2021.
- [28] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.
- [29] A. K. Jain and J. Moreau, "Bootstrap technique in cluster analysis," *Pattern Recognition*, vol. 20, no. 5, pp. 547–568, 1987.
- [30] C. A. Field and A. H. Welsh, "Bootstrapping clustered data," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 69, no. 3, pp. 369–390, 2007.
- [31] M. Jain, T. AlSkaif, and S. Dev, "A clustering framework for residential electric demand profiles," *CoRR*, vol. abs/2105.08537, 2021. [Online]. Available: <https://arxiv.org/abs/2105.08537>
- [32] *Split core current transformer*, YHDC.