ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA EN SISTEMAS

OPTIMIZACIÓN DE SISTEMAS DE INFORMACIÓN EN CONTEXTOS EMPRESARIALES

ANÁLISIS Y SEGMENTACIÓN DE CLIENTES NO REGULADOS DEL SECTOR ELÉCTRICO MEDIANTE ALGORITMOS DE APRENDIZAJE NO SUPERVISADO

TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN CIENCIAS DE LA COMPUTACIÓN

ANDRÉS ANTONIO ZAMBRANO ALQUINGA

andres.zambrano03@epn.edu.ec

DIRECTOR: JOSAFÁ DE JESÚS AGUIAR PONTES

josafa.aguiar@epn.edu.ec

DMQ, julio 2025

Certificaciones

Yo, **Andrés Zambrano**, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

NOMBRE_ESTUDIANTE

Certifico que el presente trabajo de integración curricular fue desarrollado por Andrés Zambrano, bajo mi supervisión.

NOMBRE_DIRECTOR
DIRECTOR

Declaración de autoría

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Andrés Zambrano

Josafá Aguiar

Dedicatoria

A mis padres celestiales, Dios y la santísima Virgen María, en quienes siempre he depositado toda mi fé y confianza a lo largo de toda mi trayectoria académica.

A mis padres, Verito y Marco, quienes a pesar de todas las dificultades que se presentaron a lo largo del camino, nunca dudaron de mí, y en su lugar, siempre supieron alentarme y darme su apoyo incondicional para seguir adelante, sin lugar a dudas, este, y todos mis logros se los dedico a ustedes.

A Edita Vélez, mi segunda mamá, quien me cuidó durante toda mi niñez, llenándome siempre de amor, mimos y mucho cariño.

A mis padrinos, Franklin Vásquez y Silvana Barba, por acogerme con cariño en su hogar durante mis estudios universitarios, de igual manera, a mis primos, Carolina, Dennis y Pamela, quienes más que primos han sido como hermanos para mí.

A Jhonny Sánchez, mi hermano de otra madre, con quien he compartido invaluables momentos durante gran parte de mi niñez. Gracias por ser ese hermano que nunca pude tener, pero que la vida se encargó de darme.

A la memoria de mis abuelitos, Teresa y Manuel, quienes a pesar de ya no estar físicamente conmigo, sigo sintiendo su amor y protección en cada paso que doy.

A mis amigos, compañeros de risas, retos e innumerables experiencias, que siempre han estado presentes, tanto en las buenas como en las malas.

A toda mi familia en general, quienes de manera directa o indirecta han contribuido con su granito de arena para formar la persona que soy hoy en día.

Finalmente, a mis dos peluditos, Rockie y Merlín, especialmente a mi gordo, Merlín, mi más linda compañía durante mi transición por propedéutico, pasó largas noches de vela a mi lado brindándome de su cálida compañía mientras yo estudiaba.

Agradecimientos

Agradezco en primer lugar, a Dios y a la Vírgen María por no desampararme nunca en ninguna etapa de mi vida, por haberme guiado en cada momento, y por empaparme de sabiduría durante toda mi transición por la universidad.

A mis padres, mis dos grandes tesoros, gracias por creer en mí en todo momento, por demostrarme que con esfuerzo y dedicación todo es posible y, sobre todo, por su amor y apoyo incondicional. Gracias por tanto, gracias por ser mis padres.

Quiero agradecer de manera muy especial a mi prima Carolina Vásquez por todo lo que ha hecho por mí. Gracias Carito por ser una guía indispensable y un apoyo incondicional en mi vida, eres como una hermana para mí.

Agradezco de igual manera al ingeniero Boris Astudillo por compartir conmigo sus valiosos consejos durante mi vida universitaria y, por su constante guía y apoyo durante todo el proceso de desarrollo de mi proyecto de titulación.

A mi alma máter, la Escuela Politécnica Nacional y a los docentes que contribuyeron a mi formación académica, por brindarme todos los conocimientos y las herramientas necesarias para desarrollarme como profesional.

Quiero agradecer a todo el equipo de la Empresa Eléctrica Quito, por su apoyo y guía durante el desarrollo de mis prácticas preprofesionales, en especial a los ingenieros e ingenieras Carolina, William, Oscar, Claudia, Isabel y Grace. Agradezco de igual manera al ingeniero Ricardo Dávila por brindarme la confianza y la oportunidad de vivir esta experiencia invaluable para mi desarrollo profesional.

Finalmente, agradezco a mis amigos Carlos, Alexis, Hernán, Galo, Dilan y los que faltan por nombrar, por hacer que la vida universitaria fuera mucho más llevadera. Gracias por todas las experiencias que compartimos, risas, enojos, tristezas, largas charlas, y sobre todo, la remontada del siglo en sexto semestre.

Índice general

1. Resumen

Este Trabajo de Integración Curricular aborda un proyecto de minería de datos enfocado en la implementación de un algoritmo de aprendizaje no supervisado para segmentar clientes en grupos homogéneos a partir de sus curvas características de consumo anual. El objetivo es identificar patrones de consumo energético que permitan una planificación más eficiente y una optimización del uso de la energía en el sector eléctrico.

La metodología aplicada es CRISP-DM, con una modificación en su fase final. Dentro de la misma, se han planteado dos procesos claves a seguir: en primer lugar, se desarrolla un proceso ETL orquestado por Apache Airflow, para la consolidación y transformación de los datos mensuales en una curva característica representativa anual por cada cliente, posteriormente, en el proceso de agrupación, se seleccionan y optimizan varios algoritmos para agrupar a los clientes en base a la similitud de sus curvas de consumo.

Los resultados de cada algoritmo son evaluados mediante diversas métricas, que cuantifican la calidad de las agrupaciones, con el fin de determinar el algoritmo que ofrece las agrupaciones de mejor calidad. Los resultados de agrupación serán presentados de manera visual y cuantitativa.

Palabras clave: minería de datos, segmentación de clientes, curvas de consumo, aprendizaje no supervisado, algoritmos de clustering, planificación energética, proceso ETL, Apache Airflow, CRISP-DM.

2. Abstract

This Curriculum Integration Project focuses on a data mining project aimed at implementing an unsupervised learning algorithm to segment clients into homogeneous groups based on their annual characteristic consumption curves. The goal is to identify energy consumption patterns that allow for more efficient planning and optimization of energy use in the electric sector.

The methodology applied is CRISP-DM, with a modification in its final phase. Within this framework, two key processes are followed: first, an ETL process orchestrated by Apache Airflow is developed to consolidate and transform monthly data into an annual representative characteristic curve for each client. Then, in the grouping process, several algorithms are selected and optimized to group clients based on the similarity of their consumption curves.

The results of each algorithm are evaluated using various metrics that quantify the quality of the groupings, in order to determine which algorithm provides the highest-quality groupings. The grouping results will be presented both visually and quantitatively.

Keywords: data mining, customer segmentation, consumption curves, unsupervised learning, clustering algorithms, energy planning, ETL process, Apache Airflow, CRISP-DM.

3. Introducción

El análisis del consumo energético es un aspecto fundamental para la optimización de recursos en sectores como la distribución eléctrica y la gestión de tarifas, debido a esto, identificar patrones de consumo permite segmentar a los clientes en función de su comportamiento energético, lo cual facilita la toma de decisiones estratégicas, garantizando así una planificación energética eficiente.

El presente trabajo corresponde a un proyecto de minería de datos que propone un enfoque basado en técnicas de aprendizaje no supervisado para la segmentación de clientes en función de la forma de su curva característica anual de consumo energético. El objetivo principal es identificar patrones de consumo que permitan una planificación más eficiente y optimización del uso de la energía en el sector eléctrico.

Bajo este contexto, el desarrollo del componente es realizado bajo la metodología CRISP-DM, con una ligera modificación en su fase final. Mientras que en la metodología original la fase final se centra en la implementación y despliegue del modelo, en este caso, el objetivo final es, entre todas las agrupaciones dadas por los diferentes algoritmos, escoger aquella que tenga la mejor calidad y homogeneidad, basándose en métricas de evaluación. Esta modificación de la fase final es posible debido a que CRISP-DM es sumamente flexible, y permite personalizar sus fases en función de los objetivos del proyecto.

Dentro del flujo de trabajo estructurado que propone la metodología CRISP-DM, se han definido dos procesos claves: en primer lugar, se lleva a cabo un proceso de Extracción, Transformación y Carga (ETL), orquestado por Apache Airflow, para consolidar los datos de consumo mensual de cada cliente en una curva representativa anual. Este proceso asegura la correcta integración de los datos y su transformación para que sean comparables entre sí. A continuación, se aplican técnicas de normalización para garantizar que las curvas puedan ser comparadas de manera justa.

Posteriormente, de desarrolla el proceso de agrupación, donde se determina el número óptimo de grupos de clientes a través de un análisis conjunto con las partes interesadas y el uso de métodos como el del codo. Se implementan y optimizan diferentes algoritmos de clustering, como KMeans, GaussianMixture, Birch y Spectral Clustering, para segmentar a los clientes en base a la similitud de sus curvas de consumo. Finalmente, se evalúan los resultados de cada algoritmo utilizando diversas métricas, como Silhouette Score, SSE, Davies-Bouldin Index y Calinski-Harabasz Index, para seleccionar el algoritmo que ofrezca las mejores agrupaciones. Los resultados obtenidos serán presentados tanto de manera visual como cuantitativa, permitiendo una interpretación clara y precisa de las agrupaciones logradas.

3.1. Objetivo general

Implementar un modelo de aprendizaje no supervisado a partir de un proyecto de minería de datos, para identificar patrones de consumo energético en los clientes no regulados a partir de sus curvas características anuales.

3.2. Objetivos específicos

- **Objetivo específico 1**

  Implementar un proceso ETL en conjunto con el marco de trabajo Apache Airflow para la extracción, transformación y carga de los datos de la curva característica representativa anual de cada cliente.

- **Objetivo específico 2**

  Determinar el número óptimo de agrupaciones mediante la consulta con partes interesadas y la aplicación de métodos de validación para el número óptimo de agrupaciones.

- **Objetivo específico 3**

  Seleccionar diferentes algoritmos de aprendizaje no supervisado y optimizar sus

hiperparámetros para segmentar a los clientes en base a la similitud en sus curvas características representativas de consumo.

- **Objetivo específico 4**

  Aplicar técnicas de reducción de ajuste y reducción de dimensionalidad en los datos de las curvas características anuales para garantizar su comparabilidad.

- **Objetivo específico 5**

  Presentar resultados visuales de las agrupaciones obtenidas por cada tipo de algoritmo mediante gráficos representativos que permitan visualizar tanto las agrupaciones como la tendencia de consumo energético de cada agrupación.

- **Objetivo específico 6**

  Evaluar y comparar el desempeño de los algoritmos de clustering utilizando métricas que validan la calidad de las agrupaciones, con el fin de identificar y elegir la agrupación que contenga la mayor calidad y homogeneidad.

3.3. Alcance

El componente se desarrolla siguiendo la metodología CRISP-DM, con una modificación en su fase final:

1. **Comprensión del negocio**

   El desarrollo del proyecto empieza con entender el objetivo de la segmentación de los clientes. Se identifican los principales desafíos y expectativas de las partes interesadas.

2. **Comprensión de los datos**

   Se realiza un análisis exploratorio de los datos que se tienen inicialmente. Evaluando necesidades como la normalización y escalado de los datos de consumo.

3. **Preparación de los datos**

   En esta parte se llevará a cabo el proceso ETL bajo el marco de trabajo de Apache Airflow, este proceso nos permitirá estructurar y preparar los datos de consumo de los clientes en curvas anuales características para su posterior análisis de agrupación.

4. **Modelado**

   En este apartado se va a elegir y evaluar el número de agrupaciones que se desean obtener. Se seleccionarán diferentes modelos de aprendizaje no supervisado a utilizar con el fin de segmentar a los clientes según la similitud de sus curvas características de consumo anual. Para cada uno de los algoritmos escogidos, se realizará una hiperparametrización, con el fin de escoger aquellos parámetros que ofrezcan los resultados óptimos dado nuestro conjunto de datos.

5. **Evaluación** Durante esta fase, se llevará a cabo la evaluación de cada una de las agrupaciones generadas por cada algoritmo escogido, mediante la construcción de una tabla comparativa, utilizando métricas que cuantifican la calidad de las agrupaciones.

6. **Validación y selección de resultados** A diferencia de la fase original de la metodología CRISP-DM, que se enfoca en el despliegue del modelo, en el presente componente, la fase final se centra en analizar y seleccionar aquella agrupación que contenga la mejor calidad y homogeneidad. Los resultados de agrupaciones obtenidas son presentados de manera gráfica, lo que permitirá observar la cohesión y separabilidad dentro de cada grupo, facilitando el análisis.

3.4. Marco Teórico

Para comprender este trabajo y su contexto, es de gran importancia tener bases sólidas sobre los principios subyacentes que sustentan el análisis y agrupación de los clientes en función de su curva de carga. Los apartados siguientes explicarán conceptos claves dentro del desarrollo del presente componente.

3.4.1. Sobre el sector eléctrico

3.4.1.1. Clientes no regulados

Los clientes no regulados en el sector eléctrico son aquellos cuya facturación por el suministro de energía se rige estrictamente por un contrato a término, el cual es realizado entre la empresa que suministra la energía y la empresa que recibe dicha energía.

Los contratos mencionados anteriormente son bilaterales[1].

Debido a la naturaleza de los contratos que se suscriben con este tipo de clientes, los patrones de consumo de energía que poseen son bastantes variados respecto a los clientes regulados [1].

3.4.1.2. Curvas típicas (curva de carga)

Una curva de carga o también llamada curva típica es un registro gráfico que indica la demanda eléctrica que ha tenido un cliente en cada instante durante un intervalo de tiempo determinado[2].

Estas curvas de carga reflejan el patrón de consumo cotidiano que poseen los clientes, dicho patrón está directamente relacionado con las máquinas o aparatos que utilizan, así como la energía que consumen durante sus actividades[3].

3.4.1.3. Importancia de segmentar a los clientes

Para comprender la importancia de segmentar a los clientes no regulados en grupos homogéneos donde la forma de sus curvas de carga características sea lo más parecida posible con respecto a las demás, primero hay que tener muy clara la razón por la cual las curvas de carga son tan importantes.

La importancia de las curvas de carga en el sector eléctrico radica en la información que estas proporcionan, la cual ayuda a los planificadores en la toma de decisiones respecto al tamaño de la capacidad instalada de la central eléctrica. Respecto a lo económico, permiten realizar una estimación del coste que tendrá la generación y de esta manera facilitar la toma de decisiones sobre el funcionamiento de la central eléctrica respecto al número de unidades que deben funcionar y durante cuánto tiempo [2].

Debido a la naturaleza de los clientes no regulados y, agregando el hecho de que

en su mayoría son grandes clientes, segmentarlos en grupos homogéneos permite optimizar la gestión de la demanda y mejorar la planificación del suministro eléctrico. Al agrupar clientes con patrones de consumo similares, es posible diseñar estrategias más eficientes para la contratación de energía, desarrollar y optimizar modelos tarifarios y, mejorar la predicción de la demanda a futuro. Además, esta segmentación ayuda a evitar el sobredimensionamiento o subdimensionamiento de la capacidad de generación y distribución, garantizando un uso más eficiente de los recursos y optimizando los costos operativos.

3.4.2. Metodología CRISP-DM

CRISP-DM, cuyas siglas corresponden a Cross-Industry Standard Process for Data Mining, es un método probado utilizado para orientar proyectos de minería de datos. Ofrece una serie de fases que resúmen el ciclo vital de minería de datos, a la vez que incluye descripciones y tareas necesarias en cada fase, ayudando a estructurar un flujo de trabajo ordenado cuya secuencia no es estricta, donde se puede avanzar y retroceder entre fases de ser necesario [4].

El modelo CRISP-DM es sumamente flexible, y sus fases pueden ser personalizadas en función de los objetivos del proyecto, pudiendo crear un modelo de minería de datos que se adapte a necesidades concretas[4].

3.4.2.1. Fases de CRISP-DM

CRISP-DM contiene un total de seis fases descritas a continuación[5]:

1. **Comprensión del negocio**

   Esta fase inicial se enfoca en analizar y comprender tanto los objetivos como los requerimientos del proyecto desde la perspectiva del negocio. Posteriormente todo este conocimiento es plasmado en un proyecto de minería de datos enfocado en alcanzar los objetivos.

2. **Comprensión de los datos**

   La fase de comprensión de datos tiene como principal objetivo la 'familiarización'

con los datos. Para lograr esto se realiza una recolección inicial de los datos y se procede a realizar un pequeño análisis exploratorio de los datos con el fin de comprender los datos que se tienen e identificar problemas con la calidad de los mismos.

3. **Preparación de los datos**

   Esta fase es crucial en CRISP-DM, debido a que abarca todas las actividades requeridas hasta la construcción final del conjunto de datos, los cuales servirán posteriormente para la fase de modelado. Esta fase incluye tareas como la limpieza, transformación y normalización de los datos, con el fin de asegurar la calidad de estos.

4. **Modelado**

   Varias herramientas de modelamiento son seleccionadas con el fin de ser aplicadas sobre nuestro conjunto de datos preparados. Los parámetros de dichas herramientas deben ser calibrados hasta obtener los valores óptimos que ofrezcan los mejores resultados.

5. **Evaluación**

   En esta penúltima fase del proyecto, ya se tiene construido uno o varios modelos que aparentemente ofrecen resultados de calidad. Antes de proceder a la fase del despliegue, se realiza una evaluación del modelo, revisando cada paso ejecutado hasta la construcción final del mismo con el fin de determinar si existe algún objetivo que no haya sido abordado lo suficiente.

6. **Despliegue**

   La construcción del modelo no es el final del proyecto. En función de los requerimientos, la fase de despliegue puede ser tan simple como la generación de un reporte o tan complejo como su respectiva implementación en otros proyectos de minería de datos.

3.4.3. Minería de datos

Según [6], la minería de datos es el proceso de obtener información relevante dentro de grandes repositorios de datos. Este proceso es empleado para explorar grandes bases

de datos con el fin de encontrar patrones interesantes que sean útiles, los cuales de otro modo habrían pasado desapercibidos.

Adicionalmente, la minería de datos es considerada una tecnología que combina métodos tradicionales de análisis de datos con algoritmos cuyo fin es procesar grandes volúmenes de datos.

3.4.3.1. Importancia de la minería de datos

La importancia de la minería de datos radica en las funcionalidades que posee [7]:

1. **Caracterización/Discriminación**

   Los datos de entrada pueden ser asociados con clases o conceptos, los cuales son útiles para describir clases individuales y conceptos de una manera resumida, concisa y precisa.

2. **Patrones frecuentes, asociaciones y correlaciones**

   Pueden existir patrones que ocurren con mucha frecuencia dentro de los datos, generando estructuras frecuentes dentro de los mismos. En consecuencia, dentro de los datos se pueden establecer reglas que definan como ciertos artículos o clases se relacionan entre sí.

3. **Análisis de predicción: clasificación y regresión**

   Con los datos que se tienen, es posible construir un modelo derivado de un subconjunto de datos que se usa como entrenamiento, con la finalidad de que dicho modelo sea usado para predecir etiquetas de clase, o valores numéricos según los datos de entrada que reciba.

   Mientras que los modelos de clasificación se enfocan en predecir etiquetas categóricas o clases, los modelos de regresión son utilizados para la predicción de valores numéricos.

4. **Análisis de agrupación**

   A diferencia de los modelos de clasificación y regresión antes mencionados, los

cuales analizan datos con etiquetas o clases con el fin de realizar una predicción, el análisis de agrupación analiza todo nuestro conjunto de datos (sin tomar en cuenta etiquetas ni clases) con el objetivo de generar etiquetas para nuestros datos.

El proceso de asignar etiquetas a datos que no las tienen es llamado 'agrupación', y mediante este proceso los objetos serán agrupados bajo el principio de 'maximizar la similitud intraclase y minimizar la similitud interclase'. Dicho de una forma sencilla, los objetos asignados a una misma agrupación probablemente comparten alguna similitud entre sí, y los objetos asignados a diferentes agrupaciones probablemente no comparten similitud alguna.

5. **Análisis de valores atípicos**

   Existe la posibilidad que en nuestro conjunto de datos existan objetos que no se ajusten a ningún comportamiento general o modelo, estos son los denominados datos atípicos, los cuales en la mayoría de los proyectos de minería de datos son descartados debido a que pueden conducir a conclusiones erróneas y modelos sesgados.

3.4.3.2. Proceso ETL

3.4.4. Aprendizaje no supervisado

3.4.4.1. Clustering

3.4.4.2. Número de agrupaciones

3.4.4.3. Algoritmos no supervisados

3.4.4.4. Hiperparametrización de algoritmos

3.4.5. Métricas de evaluación de agrupaciones

3.4.5.1. Suma de errores al cuadrado (SSE)

3.4.5.2. Puntaje de silueta (Silhouette Score)

3.4.5.3. Índice de Davies-Bouldin (DBI)

3.4.5.4. Índice de Calinski-Harabasz (CHI)

3.4.6. Herramientas utilizadas

3.4.6.1. Apache Airflow

3.4.6.2. Docker

3.4.6.3. Visual Studio Code

3.4.6.4. Python

3.4.6.5. MongoDB

$$\mathrm{f}_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

4. Metodología

Describir el diseño o el planteamiento utilizado...

4.0.1. Flujo de trabajo propuesto

- **Extracción, Transformación y Carga de los Datos (ETL):** La primera fase consiste en la extracción, transformación y carga (ETL) de los datos de consumo energético. Los datos iniciales provienen de archivos de consumo mensual por cliente. En este paso, se construye un archivo anual para cada cliente, en el cual se agregan y consolidan los datos correspondientes a cada año. Además, los datos son escalados y normalizados para garantizar su consistencia y comparabilidad. Este proceso se lleva a cabo con la ayuda de **Apache Airflow**, el cual permite automatizar el flujo de trabajo y garantizar su ejecución eficiente. Finalmente, los datos transformados son cargados en **MongoDB**, asegurando su disponibilidad para las fases siguientes del análisis.

- **Segmentación de Clientes:** En esta fase, se procede a definir el número óptimo de grupos de clientes a través de un análisis conversacional con las partes interesadas y la aplicación de métodos como el **método del codo**. Una vez definido el número de grupos, se seleccionan y optimizan los algoritmos de agrupación más adecuados para el análisis, tales como **KMeans**, **GaussianMixture**, **Birch** y **Spectral Clustering**. Estos algoritmos se utilizan para agrupar a los clientes según la similitud de sus curvas de consumo energético anual. Los resultados obtenidos se presentan visualmente, permitiendo observar las agrupaciones y patrones emergentes en el consumo de energía de los clientes.

- **Evaluación Comparativa de los Algoritmos:** Finalmente, se realiza una evaluación comparativa de los algoritmos de agrupación aplicados, utilizando diversas métricas para medir la calidad de las agrupaciones. Entre las métricas utilizadas se encuentran el **Silhouette Score**, **SSE (Suma de Errores al Cuadrado)**, **DBI (Índice de la Diferencia de Davies-Bouldin)** y **CHI (Índice**

de Calinski-Harabasz). Estas métricas permiten analizar el rendimiento de los algoritmos y seleccionar el que mejor se adapte a los datos de consumo energético de los clientes.

1. **Proceso ETL**

   El primer proceso consiste en el desarrollo de un flujo ETL bajo el marco de trabajo de Apache Airflow, este proceso nos permitirá estructurar y preparar los datos de consumo de los clientes en curvas anuales características para su posterior análisis de agrupación. La Figura 1 ilustra de manera detallada todas las etapas que abarca este proceso.

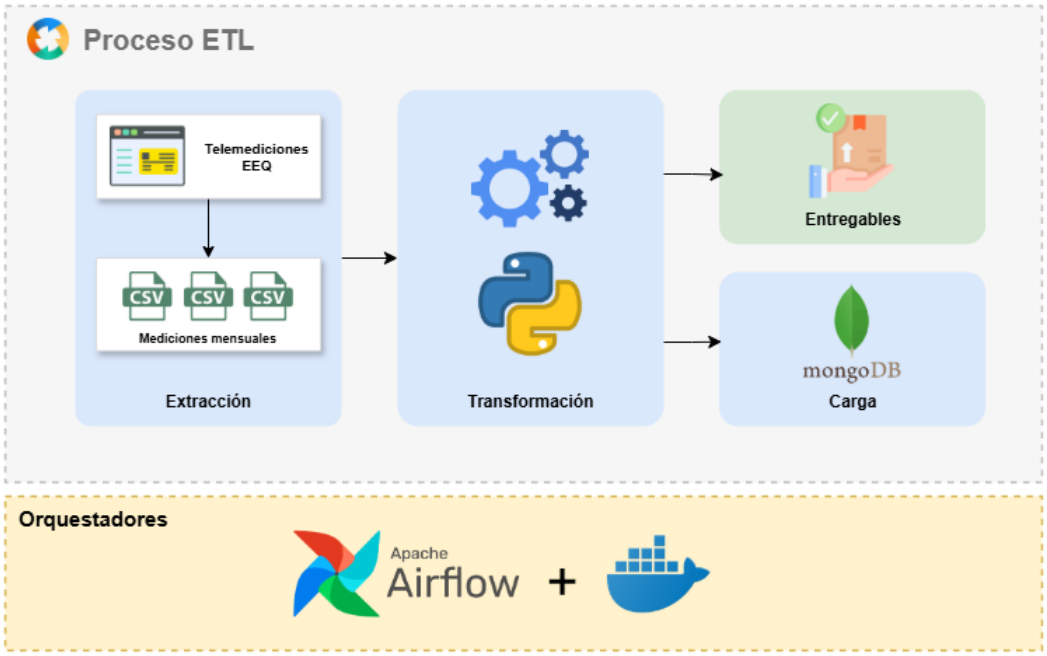

Figura 1: Proceso ETL con sus etapas

2. **Proceso de agrupación**

   El segundo proceso comprende todo el proceso de agrupamiento, en este apartado se elegirá el número de agrupaciones deseadas, se seleccionarán y aplicarán diversos algoritmos de agrupamiento para finalmente evaluar la calidad de las agrupaciones obtenidas por cada algoritmo. La Figura 2 ilustra de manera detallada todas las etapas que abarca este proceso.

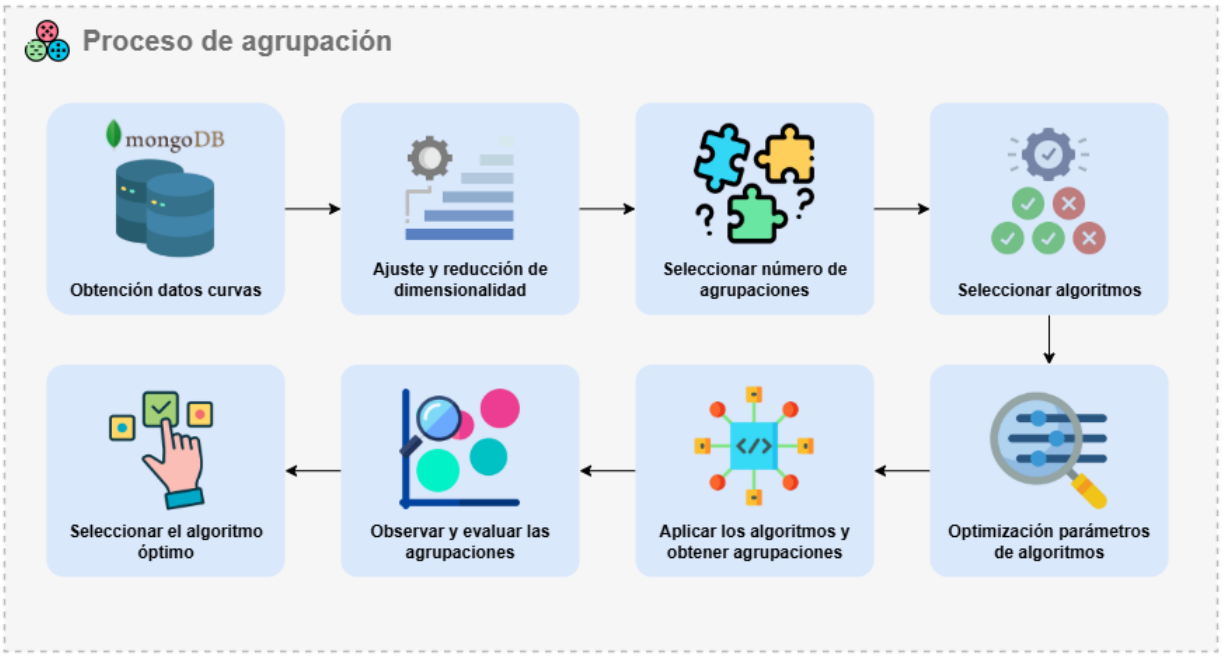

Figura 2: Proceso de agrupación con sus etapas

Los datos utilizados para el presente componente comprenden todas las mediciones mensuales del año 2024 por cada cliente, las cuales han sido obtenidos de la página de telemediciones de la Empresa Eléctrica de Quito.

Por otro lado, la segmentación de clientes se realiza exclusivamente en función de la forma de su curva característica anual, obtenida al final del proceso ETL descrito en la Figura 1. No se consideran otros factores, como las tarifas o la geolocalización, ya que el objetivo de la parte interesada es agrupar a los clientes estrictamente según el patrón de consumo de energía reflejado en su curva característica anual.

5. Resultados, Conclusiones y Recomendaciones

5.1. Resultados

Ejemplo de tabla:

No. Prueba	Resultado	Tiempo [s]
1	10	0.9
2	5	0.5

Cuadro 5.1: Resultados de las pruebas realizadas

5.2. Conclusiones

5.3. Recomendaciones

6. Referencias Bibliográficas

[1] CONELEC, *Estadística del sector eléctrico Ecuatoriano*, 2012. Obtenido de: https://www.controlrecursosyenergia.gob.ec/wp-content/uploads/downloads/2021/03/Folleto-Resumen-Estad%C3%ADsticas-2011.pdf.

[2] B. Moses y O. Akanni, "The Load Curve and Load Duration Curves in Generation Planning," *Proceedings of the Second Australian International Conference on Industrial Engineering and Operations Management, Melbourne, Australia*, 2023. Obtenido de: https://ieomsociety.org/proceedings/2023australia/245.pdf.

[3] T. Teeraratkul, D. O'Neill y S. Lall, "Shape-Based Approach to Household Load Curve Clustering and Prediction," *Stanford University*, 2017. Obtenido de: https://arxiv.org/pdf/1702.01414.

[4] IBM, "Guía de CRISP-DM de IBM SPSS Modeler," *International Business Machines Corporation*, 2018. Obtenido de: https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf.

[5] P. Chapman, R. Kerber, J. Clinton, T. Khabaza, T. Reinartz y R. Wirth, "The CRISP-DM Process Model," *CRISP-DM consortium*, 1999. Obtenido de: https://mineracaodedados.wordpress.com/wp-content/uploads/2012/12/crisp-dm-no-brand.pdf.

[6] P.-N. Tan, M. Steinbach y V. Kumar, *Introduction to Data Mining*. Pearson Education Limited, 2014. Obtenido de: https://www.ceom.ou.edu/media/docs/upload/Pang-Ning_Tan_Michael_Steinbach_Vipin_Kumar_-_Introduction_to_Data_Mining-Pe_NRDK4fi.pdf.

[7] J. Han, M. Kamber y J. Pei, *DATA MINING Concepts and Techniques*. Morgan Kaufmann, 2012. Obtenido de: https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf.

Ejemplo IEEE:

- [**1**] L. Carvajal, *Metodología de la Investigación Científica*. Santiago de Cali: U.S.C., 2006.

https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/ccs2.12080

https://hpi.de/fileadmin/user$_upload/fachgebiete/naumann/publications/2008/ETL_Manageme$

https://www.researchgate.net/profile/Sameer$_Shukla3/publication/369899578_Developing_Pragma$
$Pragmatic-Data-Pipelines-using-Apache-Airflow-on-Google-Cloud-Platform.pdf?origin=journalDetail_tp=eyJwYWdlIjoiam91cm5hbERldGFpbCJ9$

https://www.controlrecursosyenergia.gob.ec/wp-content/uploads/downloads/2021/03/Folleto-Resumen-Estad

7. Anexos

Anexo I. Conjunto de Datos Extensos

Anexo II. Formato de Entrevista

Anexo III. Enlaces