

Inteligencia Artificial Aplicada a la Ingeniería Eléctrica

Indicaciones generales para el proyecto 1 del curso

2024

Prof. Marvin Coto Jiménez

Contenidos

1	Agrupamiento	5
1.1	Contexto	5
1.2	Descripción	7
1.3	Objetivos	7
1.4	Referencias de código sugeridas	7
1.5	Elementos de análisis y discusión	8
1.6	Etapas y entregables	9
1.7	Uso de LLM	10
1.8	Rúbrica de evaluación	11

Presentación

Este documento presente el plantamiento de un proyecto por desarrollar en el curso curso IE-0435 “Inteligencia Artificial Aplicada a la Ingeniería Eléctrica” de la Universidad de Costa Rica. Se muestran los lineamientos generales de la propuesta, la cual podrá ser adaptada de acuerdo con las posibilidades de experimentación, así como de los intereses del estudiantado participante.

Es importante destacar que los aspectos propios de la organización (establecimiento de los grupos, cronograma, fechas y modalidades de entrega) serán establecidos durante las lecciones del curso. Se pretende con este proyecto brindar un acercamiento a la solución de problemas reales relacionados con la Inteligencia Artificial y la Ingeniería Eléctrica.

Planteamiento de proyecto 1

Agrupamiento de circuitos de distribución eléctrica

1.1 Contexto

El suministro de electricidad en Costa Rica se lleva a cabo principalmente a partir de los lugares fuentes de generación, y se transporta mediante las redes de transmisión y distribución, con el fin de llevarla a los diferentes sectores que requieren su consumo (cargas).

Todos los puntos de generación conforman un sistema de generación de energía eléctrica del país, en el cual no solo participan empresas públicas, sino también municipales y cooperativas de electrificación rural. La generación privada también es permitida con fuentes renovables, siempre y cuando sea a pequeña escala y exclusivamente para vender al Instituto Costarricense de Electricidad (ICE), según lo establecido en la Ley 7200.

Por otro lado, es importante destacar que el proceso de transmisión de electricidad en Costa Rica es llevado a cabo exclusivamente por el ICE. La tarea de distribución de energía es desempeñada por un total de 8 empresas, que son: ICE, Compañía Nacional de Fuerza y Luz, S.A (CNFL), Empresa de Servicios Públicos de Heredia, S.A. (ESPH), Junta Administrativa del Servicio Eléctrico de Cartago, S.A. (JASEC), Cooperativa de Electrificación de Rural de Guanacaste COOPEGUANACASTE), Cooperativa de Electrificación Rural de Alfaro Ruiz (COOPEALFARORUIZ), Cooperativa de Electrificación Rural de Los Santos (COOPESANTOS) y Cooperativa de Electrificación Rural de San Carlos (COOPELESCA).

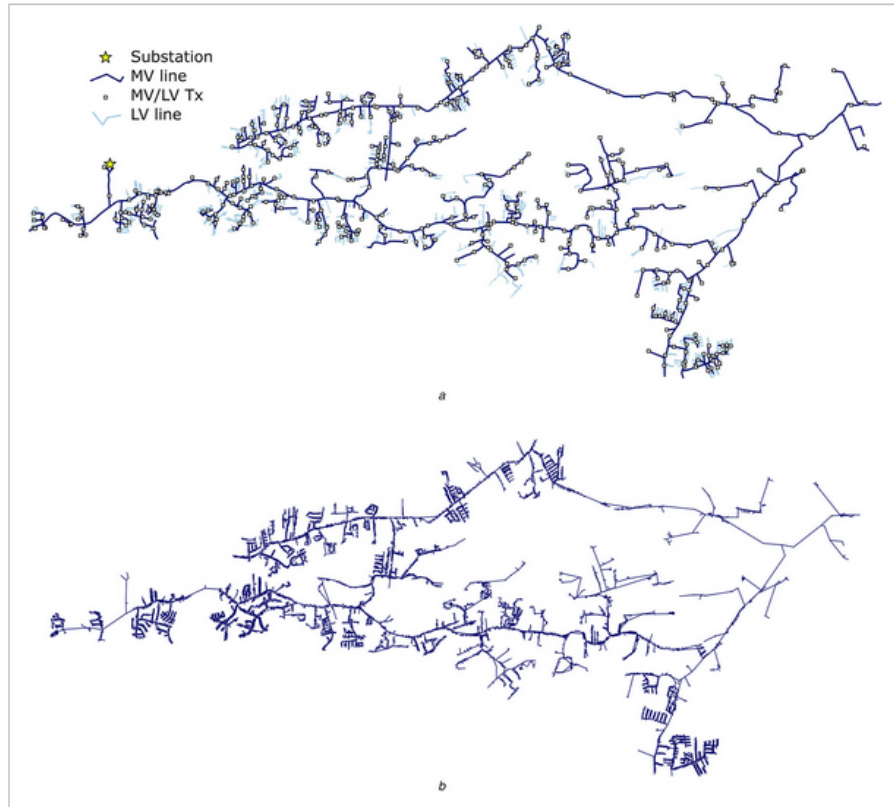


Figura 1.1: Modelo de circuito de distribución

En la Figura 1.1 se muestra un ejemplo de circuito de distribución.

En la etapa de distribución a usuarios y de comercialización, la calidad se evalúa en la Autoridad Reguladora de Servicios Públicos (ARESEP), para velar que las empresas eléctricas cumplan criterios de calidad, continuidad y confiabilidad en el suministro de energía.

Anualmente, ARESEP publica un Informe de Calidad del Suministro de Electricidad, en el cual utiliza métricas comunes de calidad como DPI y FPI para determinar aquellos sistemas que incumplen con los estándares. Específicamente la información se registra por cada circuito de las empresas distribuidoras, y está contenido en el “Informe de calidad eléctrica del servicio de distribución.”

En el presente proyecto se espera que los circuitos contenidos en el informe se puedan agrupar, para tratar de descubrir patrones y semejanzas entre ellos.

1.2 Descripción

En el presente proyecto se espera complementar la información contenida en los informes y datos abiertos disponibles en la página web de ARESEP: https://aresep.go.cr/datos-abiertos/#info_datos_estadisticaelectricidad.

Cada estudiante debería seleccionar uno de los grupos de indicadores: Indicadores de continuidad por empresa, Indicadores comerciales y de quejas del sistema de distribución, Indicadores de generación eléctrica, Indicadores de transmisión eléctrica, Infraestructura del Sistema Eléctrico Nacional, Fiscalización calidad de la tensión eléctrica. Posteriormente, deberá analizar la información suministrada de datos, seleccionar las variables que puedan ser de interés para analizar en términos de agrupamiento, y proponer un agrupamiento de datos que brinde información complementaria para el aprovechamiento de los datos publicados.

1.3 Objetivos

El objetivo principal de este proyecto es aplicar y evaluar un sistema de agrupamiento para el aprovechamiento de la información abierta del sistema eléctrico nacional de Costa Rica.

Al finalizarlo, la persona estudiante estará en la capacidad de:

1. Explicar la aplicabilidad de los sistemas de agrupamiento para aplicaciones en sistemas de ingeniería eléctrica.
2. Procesar un conjunto de datos apto para la aplicación de algoritmos de agrupamiento.
3. Implementar un sistema de agrupamiento sobre un conjunto de datos procesado.
4. Interpretar y evaluar los resultados en el contexto del problema de generación, transmisión, distribución o calidad de energía.

1.4 Referencias de código sugeridas

Para desarrollar el proyecto, se sugiere desarrollar un cuaderno de Google Colab a partir de la adaptación de código ya generado para sistemas de agrupamiento. Se cuenta, por ejemplo con las implementaciones de:

- K-means: <https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/05.11-K-Means.ipynb>
- Agrupamiento jerárquico: https://github.com/sandipanpaul21/Clustering-in-Python/blob/master/09_Hierarchical_Clustering.ipynb
- Coeficiente silueta: https://github.com/sandipanpaul21/Clustering-in-Python/blob/master/06_K_Means_Clustering.ipynb
- Con visualización del coeficiente silueta: <https://www.scikit-yb.org/en/latest/api/cluster/silhouette.html> o bien: https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html
- Uso de DBSCAN: [https://github.com/Gopizone/DataScience/blob/main/DBSCAN\(Wholesale_customers_data\).ipynb](https://github.com/Gopizone/DataScience/blob/main/DBSCAN(Wholesale_customers_data).ipynb)
- Uso de mini-batch KNN: <https://www.geeksforgeeks.org/ml-mini-batch-k-means-clustering-algorithm/>

Nota: Se deben seleccionar y aplicar al menos dos algoritmos.

1.5 Elementos de análisis y discusión

Para complementar el análisis sobre los resultados del proyecto, se sugieren abordar las siguientes preguntas, las cuales deben responderse en el informe escrito y durante la presentación a realizar:

1. ¿Por qué se requiere una autoridad como ARESEP para evaluar el sistema de generación, transmisión, distribución y la calidad de la energía eléctrica nacional?
2. ¿Cuál es el propósito de utilizar varios métodos de agrupamiento para este problema?
3. Mencione y describa un caso de métodos de agrupamiento que se encuentre en la literatura científica reciente relacionada con el sector eléctrico¹.

¹Se recomienda realizar una búsqueda en un sistema como scholar.google.com

4. ¿Existe alguna anomalía en estos conjuntos de datos? ¿Cómo pueden detectarse anomalías con otros métodos además de agrupamiento?
5. ¿Qué papel juega la normalización de los datos en este problema? ¿Existen otras formas de realizar normalización de datos? ¿Cambian los resultados si no se aplica la normalización?

1.6 Etapas y entregables

Para el cumplimiento de los objetivos del proyecto, se requiere que cada grupo de trabajo asuma una de las posibilidades listadas en la Sección 1.4. Cada grupo se organiza para distribuir y realizar cada una de las actividades requeridas de forma equitativa. Cada semana se entregará una bitácora de trabajo individual con el proyecto, y la distribución que se realice de las actividades, las cuales inicialmente son:

1. Buscar y descargar el informe del año 2021 (o el más reciente) de ARESEP en Internet.
2. Extraer el Anexo A y convertirlo en una tabla en formato texto.
3. Extraer el Anexo B y convertirlo en una tabla en formato texto.
4. Extraer la información sobre las variables DPI, FPI y abonados, e investigar en la literatura su significado exacto y forma de calcularse.
5. Preprocesar los datos (separadores) extraídos de las tablas, para darles formato de tabla csv con el cual se pueda proceder a realizar un agrupamiento con la terna (abonados, dpi, fpi).
6. Probar la implementación del código sugerido y asignado al grupo, con el ejemplo dada en la referencia.
7. Comentar en español el código en Python dado en la referencia.
8. Adaptar el código de Python para el conjunto de datos.
9. Utilizar Python para explorar los datos, y brindar estadísticas sobre los mismos.

10. Aplicar algoritmo de agrupamiento para la Tabla A:
 - (a) Decidir el número de grupos.
 - (b) Analizar los grupos, para determinar qué información puede aportar cada uno.
11. Escribir un reporte que incluye una comparación de la información extraída, incluyendo las estadísticas principales de los datos, con la información contenido en el reporte de ARESEP.
12. Elementos del informe:
 - Portada
 - Resumen
 - Introducción: Presentación del problema de clustering y de calidad eléctrica de servicio de distribución, así como la descripción precisa de las variables involucradas.
 - Procedimiento: Descripción de los pasos seguidos.
 - Resultados, incluyendo estadísticas principales de los datos, gráficas obtenidas, agrupamiento, análisis de la información generada en el agrupamiento y de las preguntas propuestas.
 - Conclusiones.
 - **Entregable 1:** Informe escrito, que incorpore el código y los resultados, adecuadamente documentos. Como todo informe debe tener al menos portada, resumen, introducción, presentación del problema, resultados, conclusiones.
 - **Entregable 2:** Presentación en clase con los resultados obtenidos.

1.7 Uso de LLM

Como complemento a las tareas desarrolladas, se debe analizar la factibilidad de sistemas LLM (ChatGPT, Gemini, etc.) para pre-procesar los datos, visualizarlos y en general como una herramienta de desarrollo del proyecto. **Importante:** En ningún momento debe usarse como herramienta de redacción.

1.8 Rúbrica de evaluación

Rúbrica de evaluación del Entregable 1

Rubro	2 puntos	1 punto	0 puntos
El documento contiene portada y todos los elementos de un reporte			
Contempla un resumen completo y adecuado			
Presenta a tiempo el reporte			
Claridad, orden y redacción adecuados en el documento entregado			
Descripción detallada de los procedimientos utilizados			
Incluye el código utilizado			
Muestra manejo adecuado de los datos			
Aplica adecuadamente el algoritmo correspondiente			
Reporta completa y adecuadamente los resultados del agrupamiento			
Presenta conclusiones adecuadas, acorde con el cumplimiento de los objetivos			

Importante: La rúbrica se utiliza para una entrega completa, con calidad identificable de un reporte profesional en todos sus elementos. Si la entrega no cumple con estos requisitos, podrá ser devuelta para su mejora antes de la evaluación.