**Avance propuesta proyecto final (semana 2)**

**Grupo 4**

Iván Camilo Barriga Gómez Código: 202121106

Andrés Felipe Arteta Isaacs. Código: 201124652

David Santiago Muñoz Carrillo. Código: 202121443

Edison Arcángel Giraldo Martínez. Código: 202124624

1. **Proponer al menos tres preguntas/problemas que estarían interesados en responder.**

**1.1. Monitoreo de Red de telefonía celular**

La red de telefonía móvil celular se encuentra en constante cambio principalmente por los trabajos de expansión de infraestructura, optimización de la red y cambios en la demanda de servicios de voz y datos, por lo que realizar labores de monitoreo y control son fundamentales para garantizar el correcto funcionamiento y aseguramiento de la calidad del servicio por parte del operador.

Propuesta 1: Identificación de anomalías, outliers o celdas que pudieran tener errores de configuración.

Propuesta 2: Descubrimiento o segmentación de grupos de celdas basado en indicadores de calidad del servicio como la velocidad de navegación en datos.

* **Motivación**: La red LTE tiene más de 70mil celdas para dar cobertura a nivel nacional y por lo tanto se requieren herramientas que puedan procesar y generar información valiosa para los ingenieros de radiofrecuencia, que les permita trabajar de manera más eficiente.
* **Relación con aprendizaje No supervisado**: Detección de anomalías, Clustering.
* **Potencial cliente interesado**: Sector Privado, empresa de telecomunicaciones.

**1.2. Clasificación de Municipios o Empresas con información pública**

Propuesta 1: El DNP produce un indicador de desempeño fiscal para los más de 1.100 municipios del país basado en promedios de algunos indicadores financieros y de gestión. Con base en este indicador, clasifica a los municipios en categorías (Sostenible, solvente, vulnerable, riesgo o deterioro). Podríamos, con estos datos, validar si las categorizaciones son apropiadas de acuerdo con un análisis de clústeres y formular recomendaciones para generar nuevas y/o mejores categorías, así como diferentes ponderaciones en los pesos de los indicadores basados en un análisis de componentes principales.

Propuesta 2: El DANE tiene encuestas a empresas de los sectores industrial y servicios (la EAM - Encuesta Anual Manufacturera y la EAS – Encuesta Anual de Servicios). Como en el punto anterior, podríamos generar categorías para las empresas según la información disponible.

* **Motivación**: Relacionar información pública disponible para encontrar (o mejorar) categorías por fuera de las comúnmente usadas.
* **Relación con aprendizaje No supervisado**: Análisis de clústeres.
* **Potencial cliente interesado**: Sector Público. Como una herramienta de clasificación y priorización de medidas a implementar según las características de los municipios.

**1.3. Topic Modelling**

Muchas veces contamos con documentos o bases de datos de texto que quisiéramos clasificar o agrupar para tener una mejor comprensión de los mismos. El modelado de temas o “topic modelling”, es una herramienta de análisis textual que nos permite responder preguntas como qué tipo de palabras aparecen más a menudo en una base de textos, o si es posible agrupar nuestros datos a partir de palabras clave o relaciones entre las palabras con el fin de descubrir los temas más relevantes de un documento. De esta forma, podríamos, por ejemplo, teniendo en cuenta el script de una llamada de un cliente a una empresa de servicios de telecomunicaciones, agrupar los clientes por temas de llamadas como preguntas generales, quejas, reclamos u otro tipo de solicitud. Por otro lado, sería posible también descubrir temas de una base de datos de comentarios, reseñas o tweets para descubrir sobre qué hablan u opinan los clientes acerca del servicio de una compañía.

* **Motivación**: Descubrir y agrupar los temas en una colección de documentos (bases de datos de texto).
* **Relación con aprendizaje No supervisado**: Este tipo de algoritmos están relacionados con el aprendizaje no supervisado porque no contamos con los datos de las etiquetas, es decir, el tema al que se refiere un texto o parrafo y necesitamos encontrarlas a partir del documento.
* **Potencial cliente interesado**: Sector Privado, empresas de telecomunicaciones, bancos, áreas de servicio al cliente.

1. **Examinar la factibilidad de conseguir datos para responder cada una de las preguntas:**

**1.1.** Base de datos privada en donde se puede extraer la información de la red movil a través de consultas SQL. Repositorio privado con información preestablecida. Estas dos fuentes de información no tienen barreras, sin embargo, al ser información privada debe ser utilizada únicamente para el ejercicio académico y sin compartir información sensible.

**1.2.** En las plataformas del DNP y el DANE se encuentran los microdatos anonimizados y son información pública, por lo que no existen barreras para obtenerlos.

**1.3.** Sería posible encontrar datasets que contengan texto sobre diferentes temas como noticias, opiniones, comentarios, entre otros.

1. **Definir el rol que cada uno de los miembros del grupo ocupará en el desarrollo del proyecto.**

* Exploración de datos, preprocesamiento: Andrés Arteta y Edison Giraldo
* Conceptualización del modelo a implementar: Todos
* Modelación e implementación en Python: Iván Barriga y Santiago Muñoz
* Análisis de resultados: Todos

Referencias

[Colombia, Julio 2022, Experiencia de la Red Móvil Report | Opensignal](https://www.opensignal.com/es/reports/2022/07/colombia/mobile-network-experience)

[¿Qué es LTE? Diferencias entre LTE y 4G // SoftwareLab](https://softwarelab.org/es/lte-4g/#:~:text=En%20concreto%2C%20el%20LTE%20permite,y%20cuatro%20en%20el%20equipo.)

[What is a Radio Access Network (RAN)? (techtarget.com)](https://www.techtarget.com/searchnetworking/definition/radio-access-network-RAN)

<https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Territorial/Guia-Corta-Metodlogia-Medicion-Nuevo-IDF.pdf>

<https://terridata.dnp.gov.co/>

<https://www.dane.gov.co/index.php/estadisticas-por-tema/servicios/encuesta-anual-de-servicios-eas>

<https://www.dane.gov.co/index.php/estadisticas-por-tema/industria/encuesta-anual-manufacturera-enam>

<https://es.statista.com/estadisticas/712971/publicacion-de-opiniones-en-internet-tras-las-vacaciones-en-espana/>

<https://www.kaggle.com/datasets/jcatumba/tweetsfromelespectador>

<https://www.kaggle.com/datasets/mmaguero/la-mision-hotel-boutiqueasuncionpyges>