

## Prueba Técnica - Especialista en Machine Learning

### Contexto

Las redes de telecomunicaciones móviles están compuestas por diversas estaciones base que permiten la comunicación inalámbrica entre dispositivos móviles, como los smartphones, y la infraestructura central de la red que provee servicios como acceso a internet y llamadas. Cada estación base cuenta con varias celdas que operan en diferentes bandas de frecuencia y gestionan el tráfico de datos.



Las **bandas de frecuencia** en redes 4G LTE se refieren a los rangos de espectro radioeléctrico asignados para la transmisión de datos y voz. Cada banda tiene características específicas que influyen en la capacidad y cobertura de la señal. En esta prueba, se trabajará con las bandas **AWS (1700/2100 MHz)** y **1900 MHz**, que se clasifican de la siguiente manera:

- **Bajas frecuencias (<1 GHz):** Mayor cobertura, mejor penetración en interiores, pero menor capacidad de transmisión de datos.
- **Medias frecuencias (1-3 GHz):** Equilibrio entre cobertura y capacidad. La banda **1900 MHz** se encuentra en este rango y es utilizada ampliamente para servicios 4G debido a su balance entre alcance y rendimiento. La banda **AWS (1700/2100 MHz)** proporciona un buen rendimiento en áreas con alta demanda de tráfico, aunque con menor cobertura que las frecuencias más bajas.
- **Altas frecuencias (>3 GHz):** Mayor capacidad y velocidades más altas, pero con menor cobertura y mayor sensibilidad a obstáculos.

En una configuración básica, una estación base tiene 3 sectores, y en cada sector una celda por cada banda de frecuencia. Aunque no todas las estaciones base cuentan con todas las

bandas de frecuencia. A modo de ejemplo, se podría tener una estación base con 2 frecuencias, lo que daría una configuración de 3 sectores por 2 celdas, es decir, 6 celdas en total. O si se tiene solo una banda de frecuencia se tendrían 3 sectores por 1 banda, es decir, 3 celdas en total.

Entre los indicadores clave de rendimiento (KPIs) de estas redes, encontramos:

- **Throughput Downlink (THP\_DL):** Velocidad efectiva percibida por los usuarios en la descarga de datos.
- **PRB Utilization Downlink (PRB\_UTILIZACION\_DL):** Nivel de uso de los recursos de radio disponibles en el enlace descendente.

Estos indicadores se miden para cada celda.

Una red eficiente debe equilibrar la disponibilidad de recursos con la demanda de tráfico, minimizando congestiones y garantizando una buena experiencia de usuario (buena velocidad)

Si usamos una autopista como símil con las redes móviles; los carriles de esta podrían ser análogos a las celdas, la cantidad de vehículos a la carga (PRB), y la velocidad de estos al Throughput percibido



---

## Prueba

Se te proporciona un conjunto de datos con los siguientes campos:

DATE_SK	CELL_SK	BANDA	THP_DL	PRB_UTILIZACION_DL
YYYY-MM-DD	Identificador de la celda	Banda de frecuencia	Throughput Downlink	Uso de recursos PRB en DL

Con base en estos datos, se espera que desarrolles un análisis exploratorio y apliques técnicas de Machine Learning para abordar los siguientes retos:

## Ejercicios

### 1. Análisis Exploratorio de Datos (EDA):

- Identifica valores atípicos, datos faltantes y patrones generales en la distribución de las variables. Presta atención a posibles degradaciones de servicio
- Analiza si existen diferencias en los patrones de comportamiento de **THP\_DL** y **PRB\_UTILIZACION\_DL** en función de la **banda** de frecuencia.
- Genera visualizaciones clave que ayuden a comprender la relación entre THP\_DL y PRB\_UTILIZACION\_DL según la banda.

### 2. Detección de Anomalías:

- Construye un modelo de detección de anomalías para identificar eventos que puedan representar **riesgos de reducción de velocidad o altos crecimientos en la carga**.
- Explica la elección del enfoque

### 3. Automatización de Alertas:

- Diseña una estrategia de **automatización** para la generación de alertas a los equipos de operación cuando se detecte una anomalía que comprometa la calidad del servicio.
  - Explica el enfoque técnico que utilizarías para la integración del modelo en un sistema de monitoreo.
  - Detalla las herramientas y tecnologías que emplearías
  - **Nota: en este punto se espera el diseño y conceptualización, pero no el desarrollo del código**
- 

## Notas:

- Para efectos prácticos, se puede asumir que cada celda es independiente. Es decir, no es necesario inferir relaciones entre celdas ubicadas en la misma estación base.
- Las celdas pertenecen a una misma zona geográfica

### **Entregable**

- Un informe detallado en formato notebook de Jupyter con el código, gráficos y explicaciones del EDA y del modelo de detección de anomalías
- Una página o diapositiva resumen que presente las conclusiones más relevantes del análisis exploratorio, el enfoque y resultados del Modelo de detección de anomalías, y la propuesta de alertas automáticas.

Se evaluará la claridad del análisis, la justificación de las elecciones técnicas, la calidad del código y las conclusiones obtenidas.