



# Prueba de Técnica de Ciencia de Datos

Readmisión en Pacientes Diabéticos  
Junio 2024

## Descripción

El siguiente dataset representa los datos de varios años de atención clínica a personas diabéticas en varios hospitales de EEUU. El dataset tiene múltiples características que representan los resultados del paciente y del hospital o centro de atención.

**El objetivo de este ejercicio:** Realizar un modelo de predicción que determine si un paciente será readmitido en el hospital en menos de 30 días.

Para ello se sugiere realizar en una primera instancia un Exploratory Data Analysis (EDA) de este dataset, para entender la información contenida y obtener insights relevantes para tareas analíticas.

Esta prueba es para determinar el nivel de las capacidades y competencias en ciencia de datos

- Descargar el dataset desde este [enlace](#)

## Tareas

Desarrollar el objetivo de la prueba en notebooks, scripts y alojarlos en un repositorio. En los notebooks se deberá cargar todas las librerías que se necesitan para explorar y procesar el dataset dado, y así realizar el modelo correspondiente para resolver el objetivo propuesto.

Se puede realizar cualquier análisis deseado, pero al final se espera encontrar realizadas las tareas del tipo "**requerido**". Además, hay algunos aspectos valorados del tipo "**deseable**" que enriquecen el trabajo, y algunos "**bonus**" extra. El desarrollo de estos últimos no es obligatorio, pero se valorará positivamente en el diagnóstico del nivel de competencias.



- **Requerido**

El código debe ser desarrollado en Python 3.x

- **Data QA:** Se debe chequear la calidad del dataset para hacer una evaluación de qué tan apropiados son los datos para tareas de Data Science. Realizar las correcciones en los datos de ser necesario.
- **Reporting:** Documente el análisis del problema, los resultados e insights obtenidos durante la exploración y describa conclusiones, así como justificando las decisiones tomadas, además soportado por gráficos / tablas / métricas.
- **Feature Engineering:** Limpieza, transformación y selección de features que podrían utilizarse para la predicción.
- **Modelo predictivo:** Realice un modelo predictivo capaz de encontrar cuando puede ocurrir una readmisión de un paciente.

- **Deseable**

- **Versionado de código con Git** (**Repositorio privado en GitHub**).
- **Feature Engineering:** Mediante librerías específicas para este fin
- **Machine Learning:** Proceso de selección entre varios modelos, hiper parametrización y por lo menos un modelo predictivo entrenado
- **Modelo:** Interpretabilidad del modelo seleccionado
- **Mostrar skills en Python:** Teniendo buenas prácticas en la estructura del código y la documentación.
- **Realizar pruebas estadísticas:** para validaciones cuando sea necesario.
- **Métricas** Definir y justificar las métricas que considere más relevantes para la problemática propuesta.

- **Bonus**

- Manejo de environment de desarrollo mediante alguna tecnología (e.g. pyenv, virtualenv, conda).
- Llevar el desarrollo del modelo de notebooks a scripts de python con buenas prácticas de desarrollo de software.
- Llevar el modelo a un demo (Ejemplo: Streamlit, [taipy](#), gradio, Dash, Flask, etc).

Una vez completado este ejercicio, por favor agregar como colaborador en el **Repositorio privado en GitHub** a [jose.zapata@mercadolibre.com.co](mailto:jose.zapata@mercadolibre.com.co) y [franciscojavier.morales@mercadolibre.com.mx](mailto:franciscojavier.morales@mercadolibre.com.mx) enviar un correo electrónico con el enlace del repositorio de GitHub

**Cómo agregar un usuario a un repo privado:** [Link de Informacion](#)

El repositorio se espera que contenga todos los recursos usados en este trabajo (e.g. Jupyter notebook, scripts, documentos, imágenes, archivos de configuración del repo, etc)