**Challenge Intermedio: Modelado con Spark ML**

1. **Introducción:**

Este challenge lleva a los estudiantes un paso más allá en el uso de PySpark, implementando Spark ML para realizar un modelo predictivo basado en una gran base de datos.

1. **Objetivo general:**

Aplicar algoritmos de machine learning en grandes bases de datos utilizando Spark ML.

1. **Objetivo específico:**

Descargar una base de datos desde UCI Machine Learning Repository, limpiarla, y entrenar un modelo predictivo de clasificación utilizando Spark ML.

1. **Base de datos:**

Base de datos: [Wine Quality Dataset](https://archive.ics.uci.edu/ml/datasets/Wine+Quality)  
Plataforma: UCI Machine Learning Repository  
Instrucciones: Descargar y cargar los datos en Databricks para crear el modelo.

1. **API y herramientas:**

Usar Spark ML para entrenar un modelo de regresión logística que prediga la calidad del vino.

1. **Repositorio GitHub:**

Subir el código del modelo y un informe de los resultados a GitHub.

**7. Métricas de evaluación:**

* Limpieza y preparación de los datos (20%)
* Selección adecuada del modelo y justificación (20%)
* Exactitud del modelo y calidad de la evaluación (30%)
* Documentación clara y detallada en GitHub (30%)