

TERCERA EVALUACIÓN DEL CURSO

PROYECTO INTEGRADO DE MACHINE LEARNING PARA MARKETING BANCARIO

Universidad del Desarrollo – Facultad de Ingeniería

1. CONTEXTO DEL PROYECTO

El Banco de Portugal ejecuta campañas de marketing para promover depósitos a plazo. Contactar a toda la base sin segmentación genera ineficiencias. Se solicita un análisis de machine learning que permita identificar segmentos de clientes y predecir la probabilidad de contratación.

2. OBJETIVOS DEL PROYECTO

- Analizar el dataset Bank Marketing (UCI).
- Construir un pipeline completo: EDA, preprocesamiento, modelamiento y validación.
- Aplicar métodos no supervisados (PCA, K-Means o similares).
- Entrenar y optimizar SVM, Random Forest, GBM y XGBoost mediante GridSearchCV.
- Comparar modelos con métricas profesionales (AUC, F1).
- Entregar recomendaciones ejecutivas basadas en evidencia.

3. TAREAS A REALIZAR

- Justificar el problema y el dataset.
- EDA completa.
- Preprocesar datos (encoding, escalamiento, división train/val/test).
- Aplicar al menos un método no supervisado.
- Construir baseline (Regresión Logística).
- Entrenar modelos avanzados (SVM, RF, GBM, XGBoost) con GridSearchCV.
- Comparar modelos y elegir uno.
- Presentar una conclusión ejecutiva.

4. RÚBRICA DE EVALUACIÓN (100 puntos totales)

- Planteamiento del problema y contexto (10 pts)
 - Claridad y relevancia del problema.
 - Justificación del dataset.
- EDA y análisis descriptivo (15 pts)
 - Profundidad y calidad del análisis.
 - Uso correcto de gráficos y hallazgos.
- Preprocesamiento (10 pts)
 - Encoding, escalamiento y división correctos.
 - Justificación técnica adecuada.
- Modelos no supervisados (15 pts)
 - Implementación de PCA/K-Means/etc.
 - Interpretación útil para marketing.
- Modelo baseline (5 pts)
 - Implementación correcta e interpretación.
- SVM optimizado con GridSearchCV (10 pts)

- Kernels, ajuste, análisis crítico.
- Random Forest con GridSearchCV (10 pts)
 - Ajuste adecuado, importancia de variables.
- Gradient Boosting con GridSearchCV (10 pts)
 - Parámetros bien ajustados y explicados.
- XGBoost con GridSearchCV (10 pts)
 - Ajuste, interpretación, tiempos de entrenamiento.
- Comparación final de modelos (10 pts)
 - Uso de AUC, F1, matriz de confusión.
 - Justificación del modelo ganador.
- Conclusión ejecutiva (5 pts)
 - Claridad, coherencia, utilidad para gerencia.

Total: 100 puntos.