

PROYECTO ML Refuerzo: Aprendizaje Supervisado

Planteamiento del Problema

Una empresa de telecomunicaciones ha segmentado a sus clientes en cuatro grupos basados en sus patrones de uso del servicio. El objetivo es desarrollar un modelo de clasificación que, utilizando datos demográficos como región, edad, ingresos, etc; prediga el grupo al que pertenece un cliente. Este modelo permitirá personalizar las ofertas y servicios para clientes prospectivos, mejorando la efectividad de las campañas de marketing.

Impacto Esperado

El proyecto busca desarrollar un sistema que facilite la segmentación de clientes basándose en datos demográficos, permitiendo una personalización efectiva de ofertas y servicios. Esto no solo mejorará la experiencia del cliente, sino que también optimizará los recursos y estrategias de marketing de la empresa.

Objetivos del Proyecto

1. Nivel Esencial:

- Construir un modelo de clasificación multi-clase que prediga el grupo de un cliente basándose en características demográficas.
- Implementar una solución funcional que permita ingresar las características de un cliente y devuelva el grupo al que pertenece.
- Realizar un **Análisis Exploratorio de Datos (EDA)** para comprender las relaciones entre las variables y los patrones de los datos.
- Generar un informe detallado que incluya métricas de rendimiento como precisión, recall, F1-score y matriz de confusión.

2. Nivel Medio:

- Incorporar técnicas avanzadas de clasificación como **ensemble methods** (Random Forest, Gradient Boosting, etc.).
- Implementar validación cruzada para evaluar el rendimiento del modelo de manera robusta.
- Utilizar técnicas de ajuste de hiperparámetros con herramientas como **Optuna** o **GridSearchCV** para optimizar el modelo.
- Mitigar problemas de desbalanceo en las clases, si los datos lo requieren.

3. Nivel Avanzado:

- Dockerizar el programa para facilitar su despliegue en diferentes entornos.
- Desplegar el modelo en la nube y ofrecerlo como un servicio accesible mediante una API o interfaz web.
- Almacenar las predicciones en una base de datos relacional para poder medir el rendimiento del modelo en nuevos datos y poder reentrenarlo en el futuro.

- Implementar un sistema para monitorizar el rendimiento del modelo una vez en producción.
4. **Nivel Experto:**
- Desarrollar un modelo basado en redes neuronales y comparar su rendimiento con modelos tradicionales de ML.
 - Implementar un sistema de MLOps para automatizar el entrenamiento, evaluación y despliegue continuo del modelo.
 - Implementar un sistema de tracking de experimentos (parámetros, métricas, artefactos) usando herramientas como **MLFlow**.
-

Plazos

Las propuestas de solución se deberán entregar el día 6 de Diciembre (el 9 de Diciembre se presenta)

Condiciones de entrega

Para el día de la reunión, será necesario entregar:

- Presentación de 30 min (negocio + tecnica)
 - Proyecto en github
 - Artículo en Medium
-

Otros Entregables

1. **Informe de Resultados:**
 - Rendimiento del modelo, análisis de características relevantes y recomendaciones basadas en el análisis.
 2. **Control de Validación:**
 - Verificación de que las métricas de validación cruzada sean consistentes con el conjunto de test.
 3. **Herramientas Organizativas:**
 - Documentar el uso de herramientas como **Trello**, **Notion** o similares para la planificación.
-

Tecnologías Requeridas

- **Python:** Scikit-learn, Pandas, Matplotlib/Seaborn.
- **Gestión de Código:** Git, GitHub.
- **Despliegue y Prototipado:** Streamlit, Gradio, Flask o FastAPI.
- **Optimización:** Herramientas de ajuste de hiperparámetros como Optuna o GridSearchCV.
- **Docker:** Para contenedorización del proyecto.
- **MLOps (Nivel Experto):** MLFlow

Datos a Utilizar

[Telecust1000](#)

- **Variable Objetivo (custcat):**
 - 1: Basic Service.
 - 2: E-Service.
 - 3: Plus Service.
 - 4: Total Service.
-