

PROYECTO: Aprendizaje Supervisado, Clasificación

Planteamiento

F5 Airlines lleva un tiempo recogiendo datos relativos a la satisfacción de los clientes. Esos datos han sido utilizados en general, pero con poco éxito, para ser analizados a mano en busca de los motivos y de un plan de actuación futuro para evitar este tipo de casos. Ahora con el histórico, los han puesto en manos del departamento de datos para buscar una solución.

Vuestro jefe cree que es posible resolver este problema con inteligencia artificial. El plan tiene dos fases. La primera es entrenar un modelo que prediga de manera satisfactoria si un cliente dado estará satisfecho o no con el producto en las categorías en las que se recoge ese dato. Mediante un análisis de importancia de características del modelo, averiguar cuáles son las características que más influyen en la satisfacción del cliente y plasmar esos datos en un informe para negocio. La segunda es desarrollar una pequeña aplicación que recoja los datos de un cliente nuevo y realice una predicción sobre su grado de satisfacción.

Para validar el proyecto, será necesario también un informe del rendimiento del modelo entrenado.

Plazos

Se presupuestan dos semanas para el prototipado. El día de entrega es el jueves 12 de Septiembre.

Condiciones de entrega

Para el día de la reunión, será necesario entregar:

- Una aplicación que pida como entrada los datos de un cliente y devuelva la predicción de satisfacción de este cliente.
- El repositorio en GitHub, con el trabajo ordenado adecuadamente en ramas y mensajes de commit limpios y claros.
- Un informe de la clasificación explicado que de cuenta de la capacidad del modelo de machine learning.
- Presentación de negocio (power point, canva, prezi, etc) y presentación técnica del código.
- Overfitting menor al 5%
- Trello y herramientas organizativas usadas.

Tecnologías a usar

- Scikit-learn
- Streamlit / Dash / PyWebIO
- Pandas
- Git
- GitHub
- Docker

Niveles de Entrega

Nivel Esencial:

- ☐ Un modelo de ML funcional que prediga el grado de satisfacción de los clientes
- ☐ Un análisis exploratorio de los datos.
- ☐ Overfitting inferior al 5%
- ☐ Una solución que productivice el modelo (Una aplicación de Streamlit, GradIO, una API, un Dash o algo similar)
- ☐ Un informe del rendimiento del modelo con sus métricas y la explicación de su performance, (feature importance, curva ROC, matrices de confusión)

Nivel Medio:

- ☐ Un modelo de ML con técnicas de ensemble
- ☐ Uso de técnicas de Validación Cruzada.
- ☐ Optimización del modelo escogido con técnicas de ajuste de hiperparámetros
- ☐ Un sistema de recogida de feedback que monitorice la performance del modelo en producción.
- ☐ Un sistema de recogida de datos nuevos para futuros entrenamientos.

Nivel Avanzado:

- ☐ Una versión dockerizada del programa.
- ☐ Guardado en bases de datos de los datos recogidos por la aplicación
- ☐ Despliegue en Cloud de las soluciones aportadas.
- ☐ Incluir test unitarios.

Nivel Experto:

- ☐ Experimentos o despliegues con modelos de redes neuronales.
- ☐ Sistemas de entrenamiento y despliegue automático de nuevas versiones del modelo que se aseguren de que la nueva versión arroja mejores métricas y en función de eso sustituyen el viejo modelo por el nuevo o no. (A/B testing, Data Drifting, MLOps)

Datos

[Airlines Dataset](#)