

Informe TP Big Data

“Airline Passenger Satisfaction”

Objetivo

Este análisis se centra en identificar los factores que contribuyen a la satisfacción de los pasajeros en la industria de la aviación, permitiendo a las aerolíneas ajustar sus servicios y mejorar la experiencia de vuelo. La pregunta de investigación es: **¿Qué factores contribuyen más a la satisfacción o insatisfacción de los pasajeros, y cómo puede una aerolínea ajustar sus servicios para maximizar el nivel de satisfacción general?**

Exploración de Datos

Para realizar la Exploración de Datos, se seleccionaron variables clave relacionadas con el confort, la puntualidad y el servicio a bordo. Estas dimensiones permiten segmentar a los pasajeros y analizar cómo se distribuye la satisfacción en función de estas variables. Entre los hallazgos principales, se observa:

- **Distancia del Vuelo:** Los vuelos cortos presentan mayor insatisfacción, mientras que los vuelos largos tienden a tener una satisfacción más alta, posiblemente debido a servicios adicionales.
- **Calificaciones de Servicio a Bordo:** Las puntuaciones se concentran entre 3 y 4, indicando que la mayoría de los pasajeros consideran el servicio como “bueno” o “muy bueno”.
- **Distribución por Clase:** La clase Business representa el 48.2% de los pasajeros, seguida de la Economy con el 44.6%, y finalmente la Economy Plus con el 0.74%.

Para entrenar este modelo, ignoramos distintas columnas para agilizar la respuesta, como unnamed, id, edad, género y ubicación de la puerta de embarque.

Preprocesamiento

Para preparar los datos, se aplicaron diversas técnicas de preprocesamiento:

- **Imputación de Valores Faltantes:** Los valores faltantes se completaron utilizando el promedio.
- **Discretización de Variables:** Las variables continuas fueron discretizadas mediante intervalos iguales, simplificando el análisis y la interpretación.
- **Eliminación de Características Irrelevantes:** Se eliminaron características con una alta proporción de valores faltantes o ceros, optimizando la estructura del conjunto de datos.

Modelos de aprendizaje

Se utilizaron tres modelos de aprendizaje automático para evaluar la satisfacción del cliente:

1. **Árbol de Decisión (Tree):** Modelo intuitivo y fácil de interpretar, ideal para visualizar reglas de decisión. Este modelo obtuvo un AUC de 0.835 y una métrica F1 de 0.867. Es adecuado en contextos donde la interpretabilidad es fundamental, aunque su precisión es limitada en comparación con otros modelos.
2. **Regresión Logística:** Modelo eficiente y simple, con un AUC de 0.924, que ofrece una buena relación entre rendimiento y facilidad de interpretación. Es ideal para entender las probabilidades y patrones lineales en los datos, siendo una buena opción como baseline.
3. **Red Neuronal (Neural Network):** Este modelo es capaz de capturar patrones no lineales y complejas relaciones entre las variables. Con un AUC de 0.971 y un F1 de 0.906, la red neuronal mostró el mejor rendimiento en términos de precisión, lo que la convierte en una opción ideal para clasificaciones de alta precisión.

En este contexto, el **Recall** es la métrica más importante, ya que refleja la capacidad de identificar correctamente a los pasajeros insatisfechos. Un mayor recall permite a la aerolínea tomar acciones para mejorar sus servicios y reducir la insatisfacción.

Conclusión

La **Red Neuronal** se destaca como el modelo más efectivo para este análisis de satisfacción de pasajeros, logrando un equilibrio óptimo entre precisión y recall.