**Diapositiva 1: Portada y Contexto**

En este mundo globalizado en el que vivimos, cada vez más gente elige como medio de transporte el avión.

Sin embargo, son habituales escuchar quejas de pasajeros de vuelo y es por eso que para facilitar a las aerolíneas la gestión de las encuestas que realizan a sus clientes he creado un modelo de machine learning capaz de predecir las satisfacción de los pasajeros de vuelo con e in e que puedan ofrecer un servicio completo y satisfactorio.

**Diapositiva 2: El Dataset - Especificaciones Técnicas**

Para este proyecto he usado un dataset de 103904 registros que contaba con 25 variables o columnas. De las cuales, me he quedado únicamente con 19 ya que las demás no eran relevantes o no aportaban ninguna información.

Tiene distintos tipos de variables:

1. **Datos Demográficos:** Edad y género.
2. **Contexto de Viaje:** Clase (Business/Eco), tipo de viaje (Personal/Negocios) y distancia del vuelo.
3. **Métricas de Servicio:** críticas sobre los servicios ofrecidos por la aerolínea. Están valoradas del 1 al 5.

Variables como, por ejemplo: Servicio wifi a bordo, Facilidad para reservar online, Comida y bebida, Limpieza, Comodidad del asiento

Como antes he dicho el objetivo el target es la satisfacción del cliente o pasajero. Y se codifica de forma binaria es decir uno. sería satisfecho y cero sería insatisfecho.

**Diapositiva 3: Limpieza o Preprocesamiento de datos**

Lo primero ha sido cargar el dataset y consultar la información de este para eso se hace un punto info y se mira que variables tiene y si hay valores nulos.

1. **Eliminar columnas irrelevantes**
2. **Valores nulos**: Como he comentado antes se han identificado los valores nulos o faltantes en columna variable 'Retraso en la llegada' y dado que representaban menos del cero, 0.3% de la muestra decidí eliminar las filas de estos registros
3. **Transformación de las variables categóricas:** object variables numéricas, usando label encoding.
4. **Traducción las columnas (***Renombrarlas***):** del inglés al castellano para realizar un análisis y poder interpretar las variables más fácilmente.
5. **Escalarizar los datos (****StandardScaler)**
   1. Establecer la y= target y la X.
   2. Dividir Train/Test
   3. Se escalan los datos con StandardScaler

para que el modelo no dé más peso a variables con rangos grandes (como la distancia del vuelo) sobre las valoraciones de servicio (1-5). Finalmente

**Diapositiva 4: Modelos de Prueba**

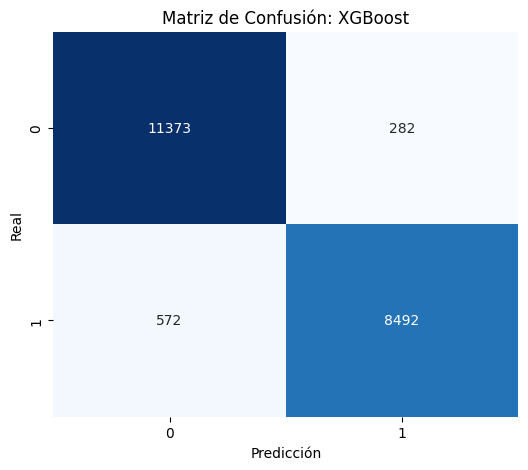
Total de 8 modelos:

*Aprendizaje Supervisado*

1. Regresión Logística (Logistic Regression)
2. Árbol de Decisión (Decision Tree)
3. Random Forest
4. Gradient Boosting
5. XGBoost (tu modelo final seleccionado)
6. KNN (K-Neighbors Classifier)
7. SVC (Support Vector Classification)

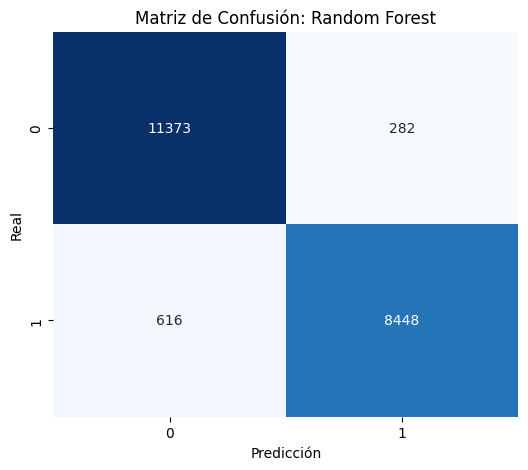
Aprendizaje No Supervisado

1. KMeans

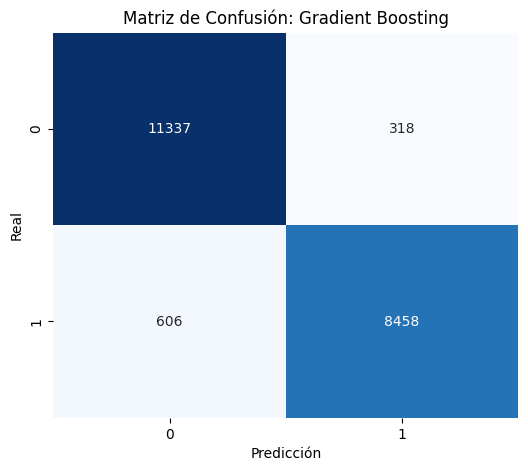
****

**Accuracy:** 0.9588

**F1\_Score:** 0.9587

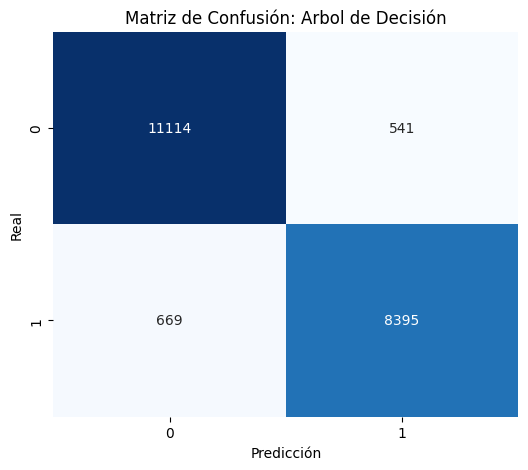
****

**Accuracy:** 0.9567

**F1\_Score:** 0.9566

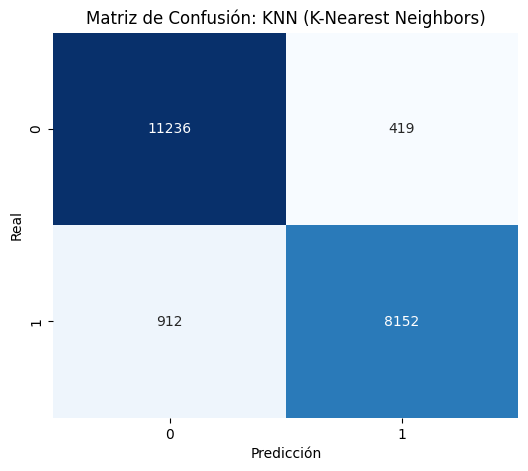
**Accuracy:** 0.9554

**F1\_Score:** 0.9553



**Accuracy:** 0.9416

**F1\_Score:** 0.9416

****

**Accuracy:** 0.9358

**F1\_Score:** 0.9355

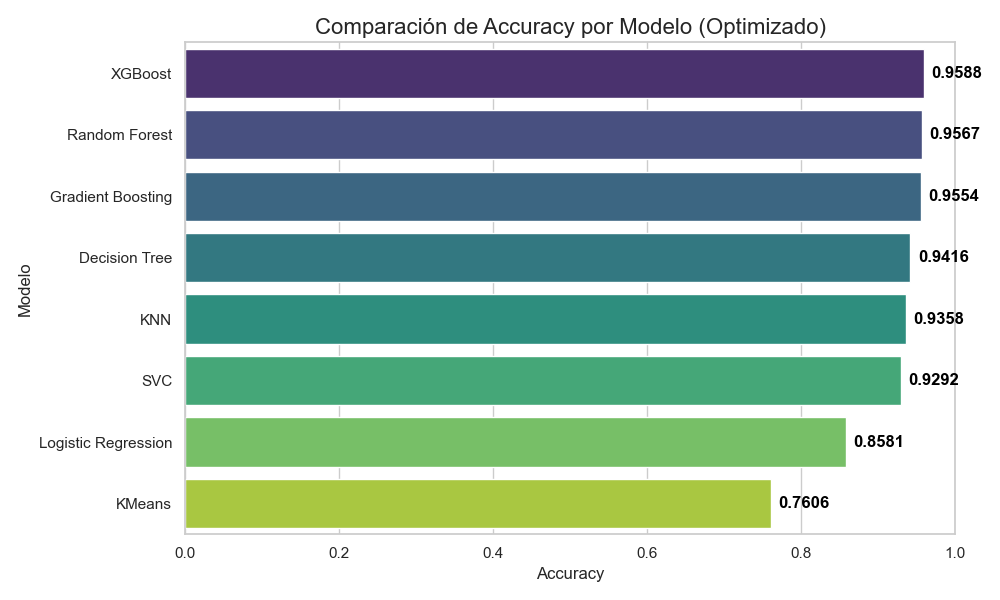
**Diapositiva 7: Evaluación y Métricas Finales**

"El modelo seleccionado ha sido **XGBoost** por su capacidad para manejar relaciones no lineales y su excelente rendimiento. Presenta un equilibrio casi perfecto entre precisión y sensibilidad (Recall), lo que minimiza el error al predecir pasajeros insatisfechos. En la matriz de confusión, se observa que el modelo identifica correctamente la gran mayoría de ambos casos, siendo una herramienta fiable para la toma de decisiones."

**Diapositiva 8: Despliegue y Conclusiones**

"Para facilitar el uso de este modelo por parte de personal no técnico, se ha desarrollado una aplicación en **Streamlit**. Esta herramienta permite cargar un listado de pasajeros y obtener instantáneamente quiénes están en riesgo de insatisfacción, permitiendo acciones comerciales proactivas. Como conclusión, el proyecto demuestra que factores digitales como el 'Embarque Online' son hoy tan determinantes para la satisfacción como la comodidad física del asiento."

Accuracy (Precision)



F1-Score ()

