

# Contenido

|  |    |
|--|----|
| Problema por resolver.....   | 3  |
| Datos Disponibles.....   | 4  |
| Enfoque de Solución.....   | 4  |
| Desafíos Técnicos .....  | 5  |
| Impacto y Aplicación del Modelo .....  | 5  |
| Mapa de calor (correlación entre variables WOE) .....  | 6  |
| Justificación ¿Por qué usar un modelo supervisado para resolver esa problemática? .....              | 7  |
| Resumen EDA.....   | 9  |
| Diccionario de variables .....   | 9  |
| Histograma de frecuencias para variables continuas:.....   | 10 |
| Métricas de los modelos .....  | 13 |
| Criterios de selección del mejor modelo.....   | 13 |
| Modelo LDA.....  | 13 |
| Neural Networks .....  | 14 |
| Decision Trees .....   | 14 |
| XGBoost.....   | 15 |
| Hiperparámetros implementados para el Modelo XGBoost (con RandomizedSearch): .....                   | 15 |
| Interpretación Detallada de los hiperparámetros del modelo XGBoost .....                             | 16 |
| Interpretación General.....  | 17 |
| Interpretación de la Curva ROC.....  | 17 |
| Matriz de confusión de las predicciones en conjunto “validate” (de octubre a diciembre de 2019)..... | 19 |
| Reporte de estabilidad del modelo .....  | 21 |
| Interpretación de la Gráfica de PSI.....   | 21 |
| Interpretación de los valores de PSI: .....  | 22 |
| Poder predictivo.....  | 23 |
| Caso de uso del modelo (enfoque al usuario final) .....  | 28 |
| Descripción General del Caso de Uso.....   | 28 |
| Implementación del Modelo (Para Usuarios Finales).....   | 28 |
| Punto de corte.....  | 32 |
| Umbral óptimo (F-1 score) .....  | 32 |
| Matriz de confusión (sin threshold): .....   | 33 |
| Matriz de Confusión aplicada con el umbral de 0.2: .....   | 34 |
| Interpretabilidad/SHAP VALUES .....  | 36 |
| ¿Cómo interpretar los SHAP values en el gráfico? .....   | 38 |

# Modelo Predictivo de Retrasos en Vuelos Internacionales: Análisis y Pronóstico de Puntualidad Aérea

## Contexto

El objetivo principal es desarrollar un modelo de aprendizaje automático capaz de predecir si un vuelo tendrá un retraso, proporcionando información valiosa para las aerolíneas, aeropuertos y pasajeros. Este modelo ayudaría a mejorar la toma de decisiones operativas, la asignación de recursos, y la experiencia del usuario al anticipar posibles retrasos en vuelos comerciales. La predicción de retrasos es un desafío que implica comprender múltiples factores y su influencia en el comportamiento de un vuelo.

## Problema por resolver

El retraso en vuelos es un problema frecuente en la industria de la aviación, con múltiples causas que pueden impactar directamente la eficiencia de los vuelos y la satisfacción de los pasajeros. Los retrasos pueden ocurrir por varios factores, como:

1. **Condiciones meteorológicas:** El clima adverso es una de las causas principales de retraso, ya que puede afectar la seguridad del vuelo y la visibilidad. Esto incluye tormentas, nevadas y fuertes vientos, que afectan tanto el despegue como el aterrizaje.
2. **Operaciones aeroportuarias:** Factores como el congestionamiento en pistas y puertas de embarque, tiempos de carga de equipaje y el tráfico aéreo en general pueden retrasar los vuelos.
3. **Factores operacionales de las aerolíneas:** Cambios en la tripulación, asignación de aviones, mantenimiento inesperado, y logística interna afectan los tiempos de salida y llegada de los vuelos.
4. **Factores de la demanda:** Los días y horarios de alta demanda, como los viernes por la tarde y las temporadas festivas, tienden a tener más probabilidades de retrasos.

El objetivo de esta tarea es predecir estos retrasos con precisión utilizando un modelo de aprendizaje automático que pueda aprender patrones en los datos históricos de vuelos.

## Datos Disponibles

Para abordar este problema, se trabajó con un conjunto de datos que contiene varias variables importantes. Entre las variables las más relevantes están:

- **Día de la semana y Mes:** Información temporal que permite captar patrones estacionales y semanales, como el hecho de que algunos días pueden tener una mayor incidencia de retrasos debido a la demanda.
- **Número de asistentes de vuelo por pasajero (FLT\_ATTENDANTS\_PER\_PASS) y Promedio mensual de pasajeros por aerolínea (AVG\_MONTHLY\_PASS\_AIRLINE):** Estas variables ofrecen una perspectiva sobre el nivel de servicio y densidad del vuelo, lo cual puede influir en los tiempos de carga y descarga de pasajeros.
- **Segmentación de vuelos (SEGMENT\_NUMBER):** Número de segmentos en el vuelo, lo que ayuda a determinar vuelos directos y de conexión, siendo estos últimos más propensos a sufrir retrasos.
- **Variables continuas y discretas:** Se manejaron diferentes tipos de variables que requieren preprocesamiento adecuado, como discretización y escalado, para ser utilizadas eficazmente en modelos de predicción.

## Enfoque de Solución

Para predecir los retrasos, se realizaron una serie de procesos para poder generar un modelo confiable:

1. **Preprocesamiento de Datos:** Esto incluye el tratamiento de valores perdidos, la eliminación de variables con poca varianza (variables unarias) y la transformación de variables relevantes en su versión WOE (Weight of Evidence) para mejorar la capacidad predictiva del modelo.
2. **Análisis de Importancia de Variables:** Se implementó el Information Value (IV) para seleccionar las variables más relevantes y descartar aquellas que tienen baja capacidad predictiva o son redundantes. Esto asegura que el modelo sea más eficiente y tenga menos ruido.

3. **Modelado y Selección de Hiperparámetros:** Se probaron varios modelos, como DecisionTreeClassifier, XGBClassifier, y MLPClassifier con una búsqueda de hiperparámetros, como Grid Search y Randomized Search, para mejorar la precisión. Estos modelos son adecuados para capturar las relaciones no lineales y las interacciones entre variables.
4. **Evaluación y Métricas:** La métrica de evaluación seleccionada es ROC\_AUC, que es apropiada para problemas de clasificación en los que el balance entre clases es una preocupación, como suele ser el caso con los vuelos retrasados.

## Desafíos Técnicos

La construcción de este modelo presenta varios desafíos técnicos:

- **Volumen de datos y procesamiento:** Con un conjunto de datos grande, existen limitaciones de memoria y procesamiento que pueden hacer que el modelo sea lento o incluso imposible de entrenar. Se implementaron optimizaciones como el ajuste de los parámetros de las búsquedas aleatorias y el uso de RandomizedSearch/GridSearch con métricas seleccionadas para reducir el tiempo de ajuste.
- **Transformaciones de variables categóricas:** Al transformar las variables categóricas en WOE, es importante preservar el sentido y evitar errores de transformación que podrían llevar a una pérdida de información o sesgo en los datos.
- **Multicolinealidad:** Muchas variables pueden estar correlacionadas entre sí (por ejemplo, variables de tiempo o relacionadas con el número de pasajeros), lo que puede degradar el rendimiento del modelo si no se manejan adecuadamente.
- **Sesgos en los datos de entrenamiento:** Es posible que los datos históricos no representen todos los escenarios futuros, lo que podría limitar la generalización del modelo.

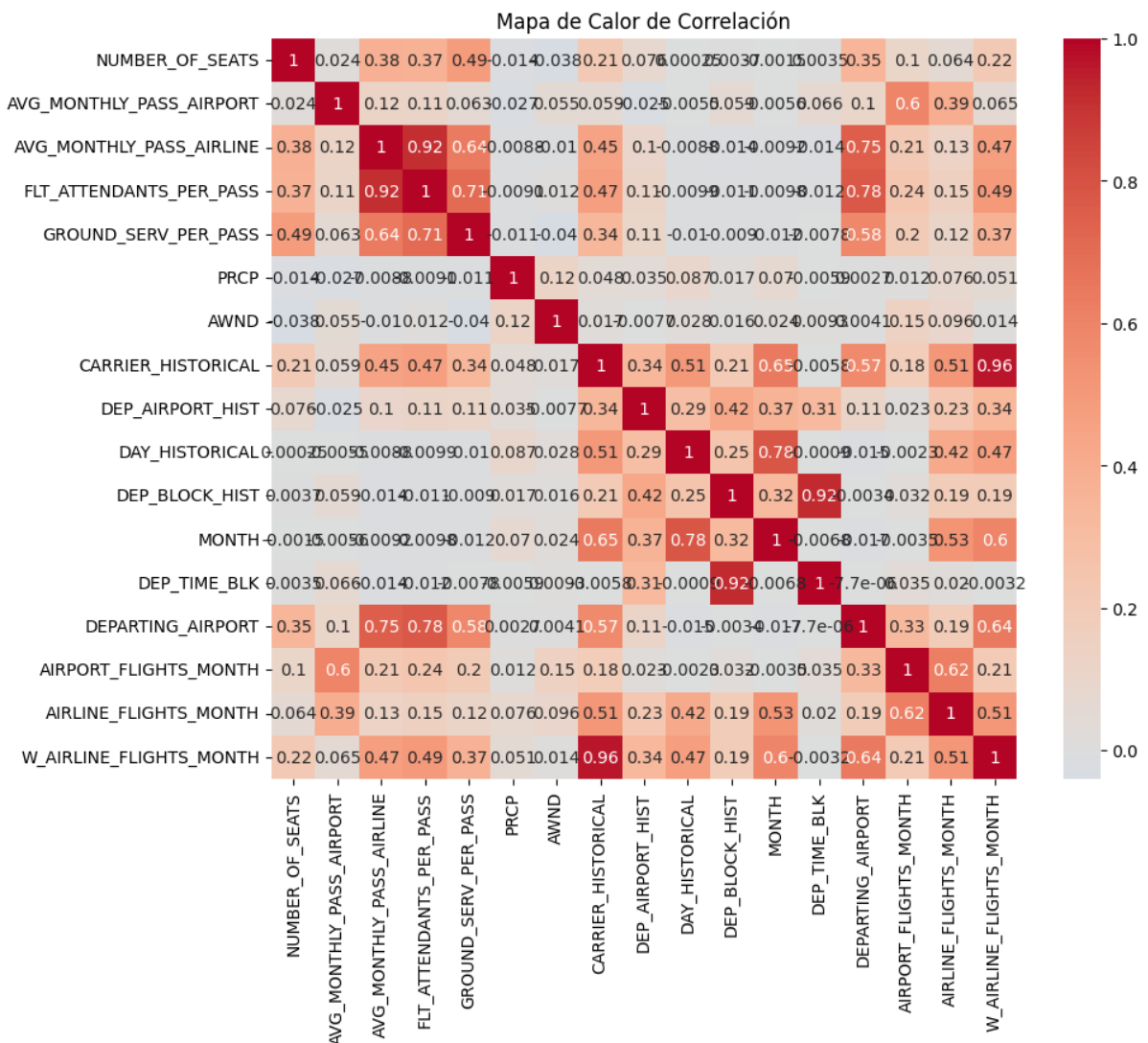
## Impacto y Aplicación del Modelo

El modelo predice retrasos en vuelos, proporcionando información valiosa en los siguientes aspectos:

1. **Mejora de la eficiencia operativa:** Al prever retrasos, las aerolíneas pueden realizar ajustes operativos, como reprogramación de vuelos y asignación de personal, para mitigar el impacto.

2. **Satisfacción del cliente:** Los pasajeros pueden recibir información anticipada sobre posibles retrasos, lo que les permite hacer arreglos alternativos.
3. **Reducción de costos:** Las aerolíneas y aeropuertos pueden reducir costos asociados con la reprogramación y reembolsos debido a retrasos no previstos.
4. **Gestión de recursos aeroportuarios:** Los aeropuertos pueden anticiparse a periodos de alta carga y administrar eficientemente los recursos como las puertas de embarque y el espacio de la pista.

## Mapa de calor (correlación entre variables WOE)



# Justificación ¿Por qué usar un modelo supervisado para resolver esa problemática?

Para abordar el problema de la predicción de retrasos en vuelos, el uso de un modelo supervisado es altamente adecuado y justificado debido a las características inherentes de la problemática y los datos disponibles. A continuación, se exponen las razones clave:

## **1. Disponibilidad de Datos Etiquetados**

En la problemática de los retrasos de vuelos, los datos históricos incluyen registros de vuelos con múltiples variables, como fecha y hora de salida, número de pasajeros, condiciones meteorológicas, y en particular, una etiqueta de si el vuelo se retrasó o no (o por cuánto tiempo se retrasó). Dado que tenemos un conjunto de datos etiquetado con un atributo objetivo claro (por ejemplo, el tiempo de retraso o una etiqueta binaria de retraso sí/no), es posible aplicar técnicas de aprendizaje supervisado. Esto facilita la construcción de un modelo que aprenda a relacionar los datos de entrada con el resultado observado en el pasado.

## **2. Patrones Complejos y No Lineales**

Los retrasos de vuelos suelen estar influenciados por múltiples factores, algunos de los cuales pueden interactuar de manera compleja y no lineal. Los modelos supervisados avanzados, como los árboles de decisión, redes neuronales o XGBoost, son capaces de capturar estas relaciones complejas entre las características y el resultado objetivo. La capacidad de los modelos supervisados de aprender patrones complejos en los datos permite que el modelo tenga un mejor desempeño al predecir futuros retrasos en base a los datos históricos.

## **3. Precisión y Evaluación de Desempeño**

Los modelos supervisados ofrecen métricas claras para evaluar su precisión y eficacia en la predicción de la variable objetivo. En este caso, métricas como el AUC-ROC, la precisión, el recall, o el F1-score permiten evaluar y afinar el modelo, brindando la capacidad de interpretar el desempeño del modelo en un contexto real de predicción de retrasos. Estas métricas son esenciales en un sistema en el que una predicción incorrecta podría afectar la logística de aerolíneas y la experiencia de los pasajeros.

#### **4. Capacidad de Generalización a Nuevos Datos**

Al entrenar un modelo supervisado con datos históricos de vuelos, el modelo aprende a reconocer patrones y características comunes en los datos que han llevado a retrasos en el pasado. Una vez entrenado adecuadamente, el modelo supervisado puede generalizar y aplicar estos conocimientos a nuevos datos de vuelos futuros, permitiendo predecir retrasos con un nivel de precisión confiable. Esto es particularmente útil en el sector de la aviación, donde los patrones históricos a menudo son buenos predictores de eventos futuros.

#### **5. Aplicabilidad en Toma de Decisiones Operativas**

La capacidad de predecir retrasos en vuelos de manera precisa permite a las aerolíneas y aeropuertos implementar estrategias para mitigar los impactos de estos retrasos. Un modelo supervisado que brinde predicciones precisas puede respaldar la toma de decisiones en tiempo real, tales como ajustes en los horarios de despegue y aterrizaje, reorganización de recursos, o comunicación proactiva con los pasajeros sobre posibles retrasos.

#### **6. Optimización del Proceso de Predicción**

Un modelo supervisado permite implementar un proceso de predicción de forma automática, facilitando un sistema de predicciones rápidas y precisas en tiempo real. Esto resulta en un flujo de trabajo más eficiente en comparación con los métodos no supervisados o manuales, permitiendo una escalabilidad mucho mayor y una integración en sistemas de apoyo de decisiones en la aviación.

En conclusión, el uso de un modelo supervisado es esencial en esta problemática debido a la disponibilidad de datos etiquetados, la necesidad de capturar patrones complejos, y la importancia de realizar predicciones precisas y escalables. Esta elección facilita la construcción de un sistema predictivo robusto que puede integrarse en operaciones de vuelo, ofreciendo un valor considerable para la planificación y optimización de recursos en la aviación.

# Resumen EDA

## Diccionario de variables

### - *Variables Continuas:*

- LATITUDE: Latitud del aeropuerto de salida.
- LONGITUDE: Longitud del aeropuerto de salida.
- PRCP: Precipitación en el aeropuerto de salida.
- SNOW: Nivel de nieve en el aeropuerto de salida.
- SNWD: Profundidad de la nieve en el aeropuerto.
- TMAX: Temperatura máxima en el aeropuerto de salida.
- AWND: Velocidad del viento promedio en el aeropuerto de salida.
- AVG\_MONTHLY\_PASS\_AIRPORT: Promedio mensual de pasajeros en el aeropuerto de salida.
- AVG\_MONTHLY\_PASS\_AIRLINE: Promedio mensual de pasajeros de la aerolínea.
- FLT\_ATTENDANTS\_PER\_PASS: Promedio de asistentes de vuelo por pasajero.
- GROUND\_SERV\_PER\_PASS: Número promedio de empleados de servicios de tierra por pasajero.
- PLANE\_AGE: Edad del avión en años.
- AIRPORT\_FLIGHTS\_MONTH: Número de vuelos desde el aeropuerto de salida en el mes.
- AIRLINE\_FLIGHTS\_MONTH: Número de vuelos de la aerolínea en el mes.
- AIRLINE\_AIRPORT\_FLIGHTS\_MONTH: Número de vuelos de la aerolínea desde ese aeropuerto en el mes.

### - *Variables Discretas:*

- MONTH: Mes del año (1 a 12).
- DAY\_OF\_WEEK: Día de la semana (1 a 7).
- DEP\_TIME\_BLK: Bloque de tiempo de la salida programada.
- DISTANCE\_GROUP: Grupo de distancia del vuelo.
- SEGMENT\_NUMBER: Número de segmentos en el vuelo.
- CONCURRENT\_FLIGHTS: Número de vuelos concurrentes desde el aeropuerto en ese momento.
- NUMBER\_OF\_SEATS: Número de asientos en el vuelo.
- CARRIER\_NAME: Nombre de la aerolínea.
- DEPARTING\_AIRPORT: Código del aeropuerto de salida



- PREVIOUS\_AIRPORT: Aeropuerto de procedencia en caso de ser un vuelo en conexión.
- CARRIER\_HISTORICAL: Promedio histórico de retrasos de la aerolínea.
- DEP\_AIRPORT\_HIST: Promedio histórico de retrasos en el aeropuerto de salida.
- DAY\_HISTORICAL: Promedio histórico de retrasos para ese día de la semana.
- DEP\_BLOCK\_HIST: Promedio histórico de retrasos para ese bloque de tiempo de salida.

## Histograma de frecuencias para variables continuas:



## 1. Capacidad y Operaciones de las Aerolíneas

- **Número de Asientos (NUMBER\_OF\_SEATS):** Observamos que hay una alta concentración de vuelos con 150 asientos, lo cual puede indicar que este tamaño de avión es el más común en la flota. Esto sugiere una estandarización que podría ayudar a las aerolíneas en términos de optimización de costos operativos y de mantenimiento, pero también plantea la pregunta de si se están maximizando los ingresos por vuelo.
- **Concurrencia de Vuelos (CONCURRENT\_FLIGHTS):** El número de vuelos concurrentes muestra una alta densidad en los valores más bajos, lo que sugiere que hay momentos de baja demanda o capacidad en ciertos aeropuertos o rutas. Esto podría indicar oportunidades para ajustar la programación y maximizar la eficiencia.

## 2. Utilización de Recursos y Productividad

- **Promedio de Pasajeros Mensuales (AVG\_MONTHLY\_PASS\_AIRLINE y AVG\_MONTHLY\_PASS\_AIRPORT):** La variabilidad en los pasajeros mensuales sugiere que las aerolíneas y aeropuertos tienen una fluctuación importante en la demanda. Identificar las causas detrás de estas fluctuaciones (temporada, días específicos de la semana, etc.) puede ayudar en la planificación de recursos, contratación de personal y mantenimiento.
- **Asistentes y Personal de Tierra (FLT\_ATTENDANTS\_PER\_PASS y GROUND\_SERV\_PER\_PASS):** Los histogramas sugieren una distribución altamente dispersa, lo que podría indicar diferencias en la productividad del personal entre diferentes rutas o aerolíneas. Ajustar la asignación de tripulación según las necesidades podría reducir costos sin afectar la calidad del servicio.

## 3. Factores de Mantenimiento y Durabilidad

- **Edad de los Aviones (PLANE\_AGE):** Hay una dispersión en la edad de las aeronaves, con muchos aviones de hasta 30 años en servicio. Este dato es crucial, ya que los aviones más antiguos pueden tener mayores probabilidades de presentar problemas técnicos, generando más retrasos. Las aerolíneas podrían considerar renovar o actualizar su flota para reducir el riesgo de demoras por mantenimiento.

#### 4. Condiciones Meteorológicas

- **Temperatura Máxima (TMAX), Precipitación (PRCP), Viento (AWND), y Nieve (SNOW):** Las condiciones meteorológicas muestran patrones significativos, especialmente en la temperatura y el viento. Las aerolíneas y aeropuertos pueden beneficiarse de sistemas predictivos que anticipen retrasos causados por el clima y permitan ajustar operaciones. Por ejemplo, tener mayor flexibilidad en la programación durante días con alta probabilidad de condiciones adversas puede mejorar la eficiencia operativa.

#### 5. Patrones Históricos de Desempeño

- **Retraso Histórico por Aerolínea (CARRIER\_HISTORICAL) y Aeropuerto (DEP\_AIRPORT\_HIST):** Los datos históricos de retrasos por aerolínea y aeropuerto reflejan patrones claros que podrían ser útiles para implementar estrategias de mejora específicas. Si ciertas aerolíneas o aeropuertos tienen constantemente mayores retrasos, puede ser indicativo de problemas estructurales que necesitan abordarse.
- **Día y Bloque de Retraso Histórico (DAY\_HISTORICAL y DEP\_BLOCK\_HIST):** La variabilidad en los retrasos históricos según el día y el bloque de salida es clave. Las aerolíneas y aeropuertos pueden emplear estos datos para ajustar horarios, reducir cuellos de botella y mejorar la experiencia del cliente en momentos pico.

#### 6. Oportunidades de Mejora en la Planificación

- **Saturación en el Uso de Recursos (Porcentaje de Concurrencia y Personal por Pasajero):** Los datos sugieren variabilidad en la utilización de los recursos de personal de vuelo y tierra, lo que podría indicar áreas donde se podrían implementar políticas de optimización de personal. Esto ayudaría a mejorar la relación costo-beneficio y reducir los retrasos provocados por falta de recursos en momentos de alta demanda.

---

### Conclusión

La exploración de estos datos proporciona insights valiosos que pueden ayudar a las aerolíneas y aeropuertos a optimizar sus operaciones, anticipar y mitigar retrasos y, en última instancia, mejorar la experiencia del pasajero. Estos insights sugieren

oportunidades para la planificación estratégica, la optimización de la asignación de recursos y la implementación de sistemas predictivos en el manejo de vuelos y operaciones.

## Métricas de los modelos

| Modelo             | LDA    | Neural Networks | XGBoost | Decision trees |
|--------------------|--------|-----------------|---------|----------------|
| ROC_score en train | 0.6821 | 0.6910          | 0.6961  | 0.6700         |
| ROC score en test  | 0.6797 | 0.6882          | 0.6958  | 0.6711         |
| Accuracy en train  | 0.8180 | 0.8188          | 0.8209  | 0.8212         |
| Accuracy en test   | 0.8181 | 0.8190          | 0.8206  | 0.8165         |

Los resultados de los modelos entrenados muestran los puntajes tanto de ROC como de accuracy en los conjuntos de entrenamiento y prueba. A continuación, se detalla el desempeño de cada modelo, analizando las métricas en ambos conjuntos y las razones para seleccionar XGBoost como el modelo óptimo.

### Criterios de selección del mejor modelo

#### Modelo LDA

- **ROC score en train:** 0.6821
- **ROC score en test:** 0.6797
- **Accuracy en train:** 0.8180
- **Accuracy en test:** 0.8181

El modelo de **Análisis Discriminante Lineal (LDA)** presenta el puntaje más bajo en el ROC en ambos conjuntos de datos. El ROC es una métrica que evalúa la capacidad del modelo para clasificar correctamente en una escala de probabilidad, siendo sensible a la calidad del modelo en tareas de clasificación binaria. Los puntajes de LDA indican que el modelo no se desempeña tan bien en la distinción entre clases como los otros modelos. En términos de **accuracy**, los resultados son ligeramente inferiores a los otros modelos, tanto en el conjunto de entrenamiento como en el de prueba.

## Neural Networks

- **ROC score en train:** 0.6910
- **ROC score en test:** 0.6882
- **Accuracy en train:** 0.8188
- **Accuracy en test:** 0.8190

El modelo de **Redes Neuronales** muestra una mejora en el ROC en comparación con LDA, con un puntaje de aproximadamente 0.6910 en el entrenamiento y 0.6882 en el conjunto de prueba. Esto indica una mejor capacidad de clasificación comparado con LDA, pero aún no alcanza los niveles de XGBoost. En cuanto a **accuracy**, los resultados en el conjunto de entrenamiento (0.8188) y en el conjunto de prueba (0.8190) son ligeramente superiores a los de LDA, pero aún por debajo de XGBoost.

## Decision Trees

- **ROC score en train:** 0.6700
- **ROC score en test:** 0.6711
- **Accuracy en train:** 0.8212
- **Accuracy en test:** 0.8165

El modelo de **Árboles de Decisión** muestra un puntaje ROC bajo, siendo uno de los más bajos de los cuatro modelos, con 0.6700 en el conjunto de entrenamiento y 0.6711 en el conjunto de prueba. Esto sugiere que el modelo es menos efectivo en la clasificación de probabilidades entre las clases. A pesar de tener un **accuracy** alto en el conjunto de entrenamiento (0.8212), el accuracy disminuye en el conjunto de prueba a 0.8165, lo cual indica un posible sobreajuste.

## XGBoost

- **ROC score en train:** 0.6961
- **ROC score en test:** 0.6958
- **Accuracy en train:** 0.8209
- **Accuracy en test:** 0.8206

Finalmente, el modelo de **XGBoost** tiene el mejor desempeño en ambas métricas. Su **ROC score** es el más alto, con 0.6961 en el conjunto de entrenamiento y 0.6958 en el de prueba. Esto sugiere que XGBoost es más efectivo en la clasificación de probabilidades y en la diferenciación entre clases en comparación con los otros modelos. En términos de **accuracy**, XGBoost también alcanza los valores más altos de 0.8209 en entrenamiento y 0.8206 en prueba, manteniendo una consistencia en ambos conjuntos y mostrando una buena generalización.

## Conclusión

Dado el desempeño superior de XGBoost en ambas métricas, tanto en entrenamiento como en prueba, este modelo se considera el más adecuado para el problema de predicción de retrasos en vuelos. Su alto **ROC score** sugiere que el modelo clasifica de manera efectiva y su **accuracy** consistente en ambos conjuntos indica una buena capacidad de generalización sin caer en el sobreajuste. Por estas razones, XGBoost es el modelo seleccionado, ya que ofrece la mejor combinación de precisión y discriminación en el contexto del problema actual.

## Hiperparámetros implementados para el Modelo XGBoost (con RandomizedSearch):

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
               colsample_bylevel=None, colsample_bynode=None,
               colsample_bytrees=0.8, device=None, early_stopping_rounds=None,
               enable_categorical=False, eval_metric=None, feature_types=None,
               gamma=None, grow_policy=None, importance_type=None,
               interaction_constraints=None, learning_rate=0.1, max_bin=None,
               max_cat_threshold=None, max_cat_to_onehot=None,
               max_delta_step=None, max_depth=5, max_leaves=None,
               min_child_weight=2, missing=nan, monotone_constraints=None,
               multi_strategy=None, n_estimators=190, n_jobs=None,
               num_parallel_tree=None, random_state=None, ...)
```

## Interpretación Detallada de los hiperparámetros del modelo XGBoost

### Resumen de los Hiperparámetros Clave

Podemos resumir los principales hiperparámetros y sus implicaciones de la siguiente manera:

- **colsample\_bytree=0.8:** Esto significa que, en cada árbol de decisión, solo el 80% de las características (columnas) serán consideradas de forma aleatoria. Esta técnica, conocida como *submuestreo de características*, ayuda a prevenir el sobreajuste (overfitting) y mejora la capacidad de generalización del modelo.
- **max\_depth=5:** Este parámetro limita la profundidad máxima de cada árbol de decisión a 5 niveles. Árboles más profundos pueden capturar patrones más complejos, pero también son más propensos a sobreajustarse. Un valor de 5 indica un equilibrio razonable entre complejidad y generalización.
- **learning\_rate=0.1:** La tasa de aprendizaje controla el tamaño de los pasos que el modelo da en cada iteración de entrenamiento. Un valor más bajo (como 0.1) indica pasos más pequeños y un entrenamiento más conservador, lo que puede ayudar a prevenir la divergencia y mejorar la precisión.
- **n\_estimators=190:** Este parámetro especifica el número total de árboles de decisión en el modelo. Un número mayor de árboles generalmente mejora el rendimiento, pero también aumenta el tiempo de entrenamiento. Un valor de 190 sugiere un modelo de tamaño moderado.
- **min\_child\_weight=2:** Este parámetro establece un umbral mínimo para la suma de los pesos de las instancias en un nodo hoja. Un valor mayor significa que el modelo requerirá más datos en un nodo hoja antes de realizar una división. Esto ayuda a prevenir la creación de nodos demasiado específicos y reduce el riesgo de sobreajuste.

### Otros Hiperparámetros Importantes

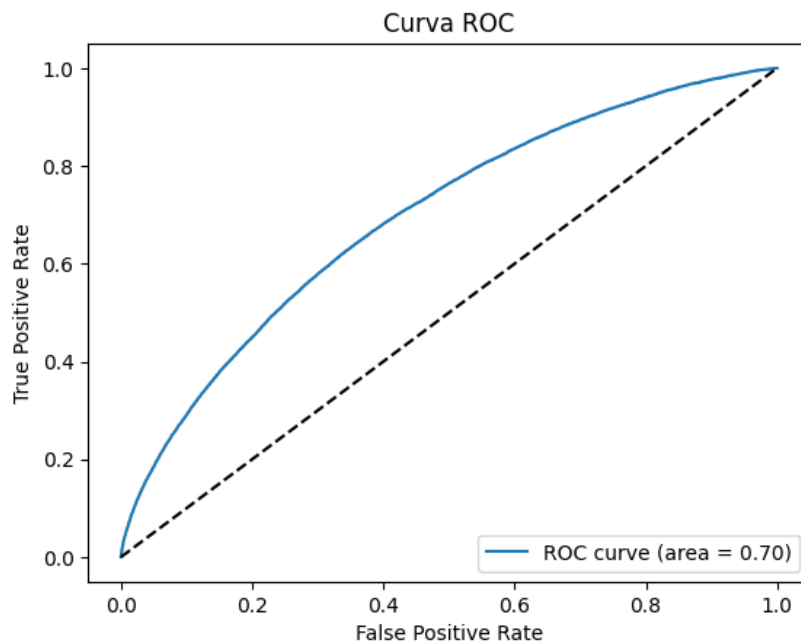
- **gamma:** Controla la regularización del modelo, penalizando modelos complejos y ayudando a prevenir el sobreajuste.
- **subsample:** Similar a colsample\_bytree, pero se aplica al submuestreo de las instancias (filas) en lugar de las características.

- **reg\_alpha y reg\_lambda:** Son parámetros de regularización L1 y L2, respectivamente, que también ayudan a prevenir el sobreajuste.

## Interpretación General

La configuración del modelo XGBoost busca un equilibrio entre complejidad y generalización. Al utilizar técnicas como el submuestreo de características e instancias, limitar la profundidad de los árboles y aplicar regularización, el modelo se vuelve más robusto y menos propenso a sobreajustarse a los datos de entrenamiento.

### *ROC CURVE del modelo XGBoost (hiperparametrizado)*



## Interpretación de la Curva ROC

### ¿Qué es una Curva ROC?

La Curva ROC (Receiver Operating Characteristic) es una herramienta gráfica que se utiliza para evaluar el rendimiento de un modelo de clasificación binaria, como el que



estás utilizando para predecir retrasos de vuelos. En el eje x se representa la tasa de falsos positivos (FPR) y en el eje y la tasa de verdaderos positivos (TPR).

- **Tasa de Verdaderos Positivos (TPR):** Proporción de casos positivos que el modelo clasifica correctamente como positivos.
- **Tasa de Falsos Positivos (FPR):** Proporción de casos negativos que el modelo clasifica erróneamente como positivos.

### Interpretación de la Curva ROC (XGBoost)

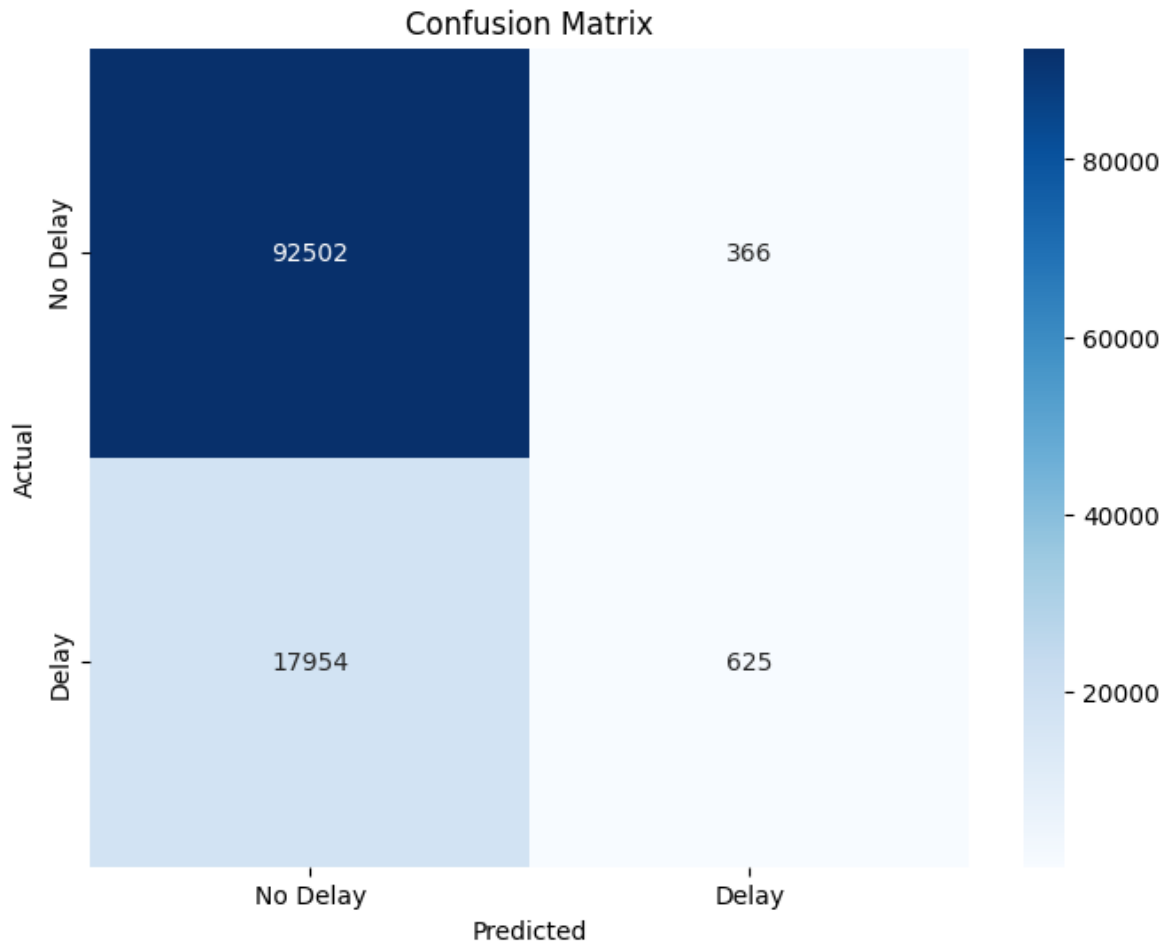
La curva ROC en el gráfico muestra un rendimiento moderado del modelo XGBoost. Aquí hay algunos puntos clave a destacar:

- **Área bajo la curva (AUC):** El valor de AUC es de 0.70. Este valor indica la probabilidad de que el modelo clasifique correctamente un ejemplo positivo elegido al azar más alto que un ejemplo negativo elegido al azar. Un valor de AUC de 0.5 indica un modelo aleatorio, mientras que un valor de 1 indica un modelo perfecto. En este caso, un AUC de 0.70 sugiere que el modelo tiene una capacidad razonable para distinguir entre vuelos retrasados y no retrasados, pero hay margen de mejora.
- **Forma de la curva:** La curva se encuentra por encima de la línea diagonal, lo que indica que el modelo está superando a un clasificador aleatorio. Sin embargo, la curva no está muy cerca de la esquina superior izquierda, lo que sugiere que hay algunos casos en los que el modelo tiene dificultades para distinguir entre las dos clases.

### ¿Qué significa esto para el modelo?

- **Buen desempeño general:** El modelo tiene una capacidad razonable para predecir los retrasos en vuelos internacionales.

## Matriz de confusión de las predicciones en conjunto “validate” (de octubre a diciembre de 2019)



### ¿Qué es una Matriz de Confusión?

Una matriz de confusión es una herramienta visual que nos permite evaluar el desempeño de un modelo de clasificación. En este caso, el modelo está tratando de predecir si un vuelo se retrasará o no.

### Interpretación de la Matriz Presentada

#### Elementos de la Matriz:

- **Verdaderos Negativos (VN):** 92502. Estos son los casos en los que el modelo predijo correctamente que el vuelo no se retrasaría (el vuelo realmente no se retrasó y el modelo también lo predijo así).

- **Falsos Positivos (FP):** 366. Estos son los casos en los que el modelo predijo incorrectamente que el vuelo se retrasaría (el vuelo realmente no se retrasó pero el modelo dijo que sí).
- **Falsos Negativos (FN):** 17954. Estos son los casos en los que el modelo predijo incorrectamente que el vuelo no se retrasaría (el vuelo realmente se retrasó pero el modelo dijo que no).
- **Verdaderos Positivos (VP):** 625. Estos son los casos en los que el métricas de evaluación del modelo predijo correctamente que el vuelo se retrasaría (el vuelo realmente se retrasó y el modelo también lo predijo así).

#### **Interpretación General:**

- **Alto porcentaje de aciertos en la predicción de vuelos que no se retrasan:** El número de Verdaderos Negativos es muy alto, lo que indica que el modelo es bastante bueno en identificar vuelos que no se retrasarán.
- **Desempeño menos preciso en la predicción de retrasos:** El número de Falsos Negativos es considerablemente mayor que el de Verdaderos Positivos. Esto sugiere que el modelo tiene dificultades para predecir correctamente los vuelos que se retrasarán.

#### **Implicaciones:**

- **Costos de Falsos Positivos y Negativos:**
  - **Falsos Positivos:** Si se alertan muchos vuelos como retrasados cuando en realidad no lo están, podría generar innecesaria preocupación en los pasajeros y aerolíneas.
  - **Falsos Negativos:** Si no se detectan muchos vuelos que realmente se retrasan, los pasajeros podrían llegar al aeropuerto sin suficiente antelación, lo que podría generar inconvenientes.
- **Mejora del Modelo:**
  - **Enfocarse en los Falsos Negativos:** Se podrían agregar o ajustar características en el modelo para mejorar la capacidad de predecir correctamente los vuelos que se retrasarán.
  - **Balancear las clases:** Si hay un desequilibrio significativo entre la clase de vuelos que se retrasan y la que no se retrasa, se podrían utilizar técnicas de balanceo de clases para mejorar el rendimiento del modelo.
- **Aplicaciones Prácticas:**
  - **Alertas tempranas:** El modelo podría utilizarse para enviar alertas a los pasajeros con suficiente antelación en caso de un posible retraso.

- **Planificación de operaciones:** Las aerolíneas podrían utilizar estas predicciones para optimizar la asignación de recursos y mejorar la eficiencia operativa.

## Reporte de estabilidad del modelo

### Interpretación de la Gráfica de PSI

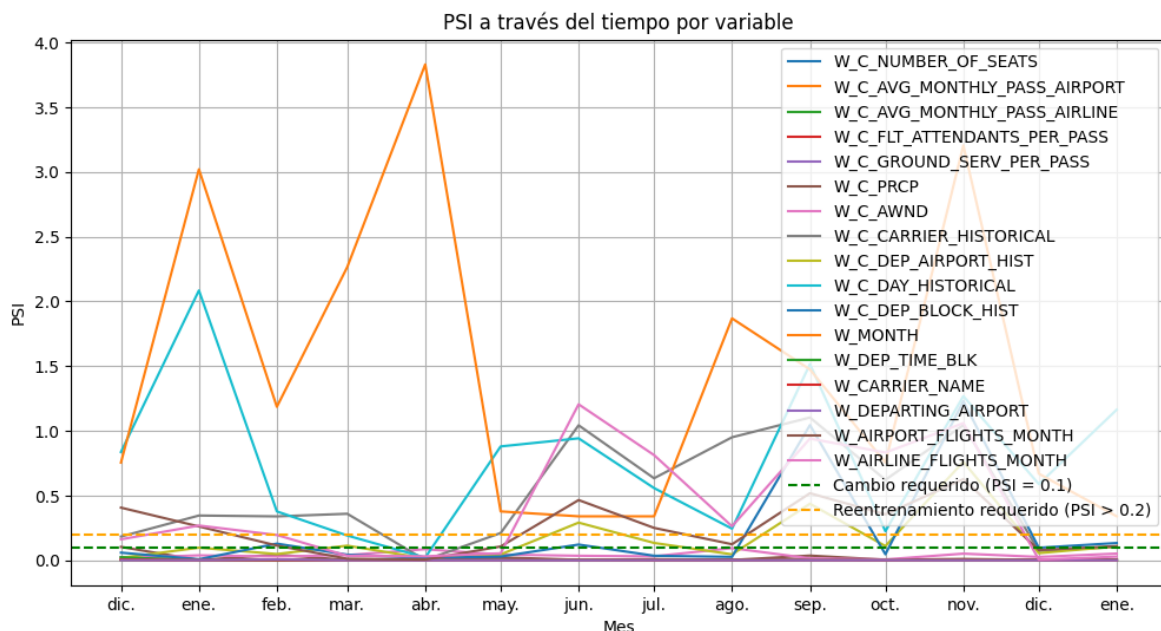
#### ¿Qué es el PSI?

El Índice de Estabilidad de la Población (PSI, por sus siglas en inglés) es una métrica utilizada en modelado estadístico para medir el cambio en la distribución de una variable entre dos poblaciones o en el tiempo. Un PSI alto indica un cambio significativo en la distribución, lo que podría sugerir que el modelo utilizado para predecir o clasificar basado en esa variable ya no es tan preciso como antes.

#### Interpretación de la Gráfica:

En la siguiente gráfica, se muestra la evolución del PSI para diferentes variables a lo largo del tiempo (meses). Cada línea representa una variable y su valor de PSI en un momento dado. Las líneas horizontales representan los umbrales de PSI:

- **PSI = 0.1:** Se considera un cambio moderado en la distribución.
- **PSI > 0.2:** Se considera un cambio significativo en la distribución, lo que podría requerir un reentrenamiento del modelo.



## Interpretación de los valores de PSI:

1. **Valores de PSI bajos ( $< 0.1$ ):** Estos indican que no hay cambios significativos en la distribución de la variable entre el conjunto de referencia y el conjunto de validación. La estabilidad de estas variables sugiere que el entorno de los datos no ha cambiado drásticamente y que estas características siguen comportándose de manera similar entre periodos.

Ejemplos en el conjunto:

- W\_C\_NUMBER\_OF\_SEATS en todos los meses (máximo PSI de 0.000527 en noviembre).
- W\_C\_AVG\_MONTHLY\_PASS\_AIRPORT y W\_C\_AVG\_MONTHLY\_PASS\_AIRLINE también tienen valores bajos de PSI en la mayoría de los meses.
- W\_C\_FLT\_ATTENDANTS\_PER\_PASS y W\_C\_GROUND\_SERV\_PER\_PASS tienen valores de PSI bajos (0.000964 y 0.000519 en noviembre, respectivamente).

Esto sugiere que estas variables tienen distribuciones bastante estables, lo cual es deseable, ya que implica que el modelo no necesitaría ajustarse para cambios en estas características.

2. **Valores de PSI moderados (entre 0.1 y 0.25):** Estos valores sugieren algunos cambios en la distribución de la variable, lo cual puede ser relevante si estos cambios persisten o aumentan. Un PSI en este rango no necesariamente implica un riesgo inmediato para el rendimiento del modelo, pero es una señal de que podría ser necesario un monitoreo más cercano.

Ejemplos en el conjunto:

- W\_C\_PRCP tiene un valor de PSI de 0.010 en noviembre, lo cual está en un rango moderado.
- W\_AIRLINE\_FLIGHTS\_MONTH presenta un valor en octubre de 0.83, lo que podría ser un indicador de cambio, aunque no crítico.

Estos cambios podrían reflejar factores estacionales o variaciones naturales en estos meses específicos que afectan la distribución de la variable.

3. **Valores de PSI altos ( $> 0.25$ ):** Un PSI alto indica cambios sustanciales en la distribución de la variable, lo cual puede impactar negativamente el

rendimiento del modelo si este cambio persiste. Es importante analizar por qué estos cambios ocurren y considerar ajustar el modelo si las variables con PSI alto son significativas para las predicciones.

Ejemplos en el conjunto:

- W\_C\_CARRIER\_HISTORICAL y W\_C\_DEP\_AIRPORT\_HIST tienen PSI elevados en noviembre, de hasta 1.05 y 0.75, respectivamente, indicando cambios sustanciales.
- W\_MONTH presenta un PSI de 3.2 en noviembre y valores altos en otros meses, sugiriendo una inestabilidad significativa en la distribución de esta variable.
- W\_DEP\_TIME\_BLK y W\_CARRIER\_NAME muestran valores bajos, pero W\_AIRLINE\_FLIGHTS\_MONTH en diciembre muestra un cambio que podría afectar la variable en el futuro.

## Poder predictivo

Mes: 2019-08-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-08-01: 18.77919335750151  
Predictive power of NUMBER\_OF\_SEATS in 2019-08-01: 86.56823360498353  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-08-01: 59.95386587216575  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-08-01: 38.70560668772252  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-08-01: 0.4532912625368021  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-08-01: 197.2655044086626  
Predictive power of PLANE\_AGE in 2019-08-01: 0.7697637067856874  
Predictive power of PRCP in 2019-08-01: 228.10890452847005  
Predictive power of TMAX in 2019-08-01: 390.4743242820126  
Predictive power of AWND in 2019-08-01: 114.9490412252305  
Predictive power of CARRIER\_HISTORICAL in 2019-08-01: 1271.6190431768473  
Predictive power of DEP\_AIRPORT\_HIST in 2019-08-01: 2.7791719568181383  
Predictive power of DAY\_HISTORICAL in 2019-08-01: 1237.3701424692385  
Predictive power of DEP\_BLOCK\_HIST in 2019-08-01: 137.27432187193168  
Predictive power of SNOW in 2019-08-01: 219.7204729966807  
Predictive power of SNWD in 2019-08-01: 370.9789862167699

Mes: 2019-01-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-01-01: 0.9264730731190918  
Predictive power of NUMBER\_OF\_SEATS in 2019-01-01: 42.05828749950962  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-01-01: 7.045318198978205  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-01-01: 4.477460719568802  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-01-01: 8.757424396252638  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-01-01: 102.35619837814045  
Predictive power of PLANE\_AGE in 2019-01-01: 159.46842905782512  
Predictive power of PRCP in 2019-01-01: 169.34127676553518  
Predictive power of TMAX in 2019-01-01: 217.8382987584677  
Predictive power of AWND in 2019-01-01: 66.1579026511354  
Predictive power of CARRIER\_HISTORICAL in 2019-01-01: 470.72051569080173  
Predictive power of DEP\_AIRPORT\_HIST in 2019-01-01: 16.92618420471988  
Predictive power of DAY\_HISTORICAL in 2019-01-01: 470.87430980778936  
Predictive power of DEP\_BLOCK\_HIST in 2019-01-01: 40.660151975629525  
Predictive power of SNOW in 2019-01-01: 117.7113572662299  
Predictive power of SNWD in 2019-01-01: 284.7720229325725

Mes: 2019-03-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-03-01: 70.34622301455664  
Predictive power of NUMBER\_OF\_SEATS in 2019-03-01: 84.61177812976081  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-03-01: 122.70078986044707

Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-03-01: 144.84100187493172  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-03-01: 52.814013019162566  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-03-01: 128.92022162531214  
Predictive power of PLANE\_AGE in 2019-03-01: 32.824972901760376  
Predictive power of PRCP in 2019-03-01: 188.48365989362085  
Predictive power of TMAX in 2019-03-01: 199.44192880909748  
Predictive power of AWND in 2019-03-01: 83.14383111001524  
Predictive power of CARRIER\_HISTORICAL in 2019-03-01: 612.9543266541874  
Predictive power of DEP\_AIRPORT\_HIST in 2019-03-01: 1.3923267224334392  
Predictive power of DAY\_HISTORICAL in 2019-03-01: 615.6095757682295  
Predictive power of DEP\_BLOCK\_HIST in 2019-03-01: 190.67257719732  
Predictive power of SNOW in 2019-03-01: 130.0285513726043  
Predictive power of SNWD in 2019-03-01: 261.21769483322527

Mes: 2019-11-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-11-01: 53.37141701868994  
Predictive power of NUMBER\_OF\_SEATS in 2019-11-01: 77.21819837089284  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-11-01: 123.1963320479059  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-11-01: 76.09100923108164  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-11-01: 90.37443259543004  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-11-01: 66.90191349092768  
Predictive power of PLANE\_AGE in 2019-11-01: 52.172903620792304  
Predictive power of PRCP in 2019-11-01: 206.50102883744114  
Predictive power of TMAX in 2019-11-01: 114.69025898124607  
Predictive power of AWND in 2019-11-01: 31.499436365964904  
Predictive power of CARRIER\_HISTORICAL in 2019-11-01: 402.17815972556747  
Predictive power of DEP\_AIRPORT\_HIST in 2019-11-01: 8.42765836417359  
Predictive power of DAY\_HISTORICAL in 2019-11-01: 401.70773105985904  
Predictive power of DEP\_BLOCK\_HIST in 2019-11-01: 167.9109254468215  
Predictive power of SNOW in 2019-11-01: 130.53899085115947  
Predictive power of SNWD in 2019-11-01: 254.9677551307356

Mes: 2019-06-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-06-01: 20.272926611202497  
Predictive power of NUMBER\_OF\_SEATS in 2019-06-01: 147.64644314874045  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-06-01: 86.87461607211043  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-06-01: 113.81824274443939  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-06-01: 64.28260940419422  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-06-01: 192.03203844841386  
Predictive power of PLANE\_AGE in 2019-06-01: 10.47426787634595  
Predictive power of PRCP in 2019-06-01: 197.9112786125799  
Predictive power of TMAX in 2019-06-01: 334.5028206082396  
Predictive power of AWND in 2019-06-01: 45.13588602934761  
Predictive power of CARRIER\_HISTORICAL in 2019-06-01: 1389.123475885755  
Predictive power of DEP\_AIRPORT\_HIST in 2019-06-01: 10.10562445003983  
Predictive power of DAY\_HISTORICAL in 2019-06-01: 1386.1529506546838  
Predictive power of DEP\_BLOCK\_HIST in 2019-06-01: 162.5615663688828  
Predictive power of SNOW in 2019-06-01: 327.6878048201906  
Predictive power of SNWD in 2019-06-01: 454.31430633507694

Mes: 2019-02-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-02-01: 29.55713563797683  
Predictive power of NUMBER\_OF\_SEATS in 2019-02-01: 47.26205956480494  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-02-01: 73.06879147414796  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-02-01: 44.1268035465009  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-02-01: 58.60788642788632  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-02-01: 45.57393663639355  
Predictive power of PLANE\_AGE in 2019-02-01: 32.16557463185144  
Predictive power of PRCP in 2019-02-01: 153.71599851865983  
Predictive power of TMAX in 2019-02-01: 240.32938735956446  
Predictive power of AWND in 2019-02-01: 116.96416340983775  
Predictive power of CARRIER\_HISTORICAL in 2019-02-01: 710.8548708759234  
Predictive power of DEP\_AIRPORT\_HIST in 2019-02-01: 28.690342148160966  
Predictive power of DAY\_HISTORICAL in 2019-02-01: 680.3975560941319  
Predictive power of DEP\_BLOCK\_HIST in 2019-02-01: 78.31064289375679  
Predictive power of SNOW in 2019-02-01: 99.6358511195312  
Predictive power of SNWD in 2019-02-01: 272.37201815141816

Mes: 2019-10-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-10-01: 28.40862329791662  
Predictive power of NUMBER\_OF\_SEATS in 2019-10-01: 44.838379602083506  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-10-01: 74.55535838509905  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-10-01: 57.0218441338413  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-10-01: 39.71345716253884  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-10-01: 242.8948056032709  
Predictive power of PLANE\_AGE in 2019-10-01: 145.96833681602834  
Predictive power of PRCP in 2019-10-01: 129.7274120681207  
Predictive power of TMAX in 2019-10-01: 277.4218473551717  
Predictive power of AWND in 2019-10-01: 138.11711626128442  
Predictive power of CARRIER\_HISTORICAL in 2019-10-01: 551.8615161753049  
Predictive power of DEP\_AIRPORT\_HIST in 2019-10-01: 3.5160361208323367  
Predictive power of DAY\_HISTORICAL in 2019-10-01: 542.262894008024  
Predictive power of DEP\_BLOCK\_HIST in 2019-10-01: 105.30838156211783  
Predictive power of SNOW in 2019-10-01: 129.25512014212094  
Predictive power of SNWD in 2019-10-01: 262.0335728284411

Mes: 2019-07-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-07-01: 39.550740336858084  
Predictive power of NUMBER\_OF\_SEATS in 2019-07-01: 73.68534289291824  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-07-01: 40.91169739642237  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-07-01: 43.62782183704946  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-07-01: 8.010892990909808  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-07-01: 291.51718981556286  
Predictive power of PLANE\_AGE in 2019-07-01: 1.6962287277029482  
Predictive power of PRCP in 2019-07-01: 169.14088999749112  
Predictive power of TMAX in 2019-07-01: 387.17354789595504  
Predictive power of AWND in 2019-07-01: 13.895790947986745  
Predictive power of CARRIER\_HISTORICAL in 2019-07-01: 1593.712455577321  
Predictive power of DEP\_AIRPORT\_HIST in 2019-07-01: 3.5202791120286197  
Predictive power of DAY\_HISTORICAL in 2019-07-01: 1510.025330759972  
Predictive power of DEP\_BLOCK\_HIST in 2019-07-01: 123.59272571247828  
Predictive power of SNOW in 2019-07-01: 210.54084709872498  
Predictive power of SNWD in 2019-07-01: 391.5692025488498

Mes: 2019-12-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-12-01: 113.09065682325854  
Predictive power of NUMBER\_OF\_SEATS in 2019-12-01: 86.68602190587623  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-12-01: 190.7743845631784  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-12-01: 161.19114208274812  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-12-01: 146.98630718786856  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-12-01: 183.66791583088715  
Predictive power of PLANE\_AGE in 2019-12-01: 34.028444957726784  
Predictive power of PRCP in 2019-12-01: 407.4996743647628  
Predictive power of TMAX in 2019-12-01: 370.04935530990525  
Predictive power of AWND in 2019-12-01: 150.92670133242248  
Predictive power of CARRIER\_HISTORICAL in 2019-12-01: 749.7875715283137  
Predictive power of DEP\_AIRPORT\_HIST in 2019-12-01: 66.82699547510761  
Predictive power of DAY\_HISTORICAL in 2019-12-01: 683.1723283906465  
Predictive power of DEP\_BLOCK\_HIST in 2019-12-01: 222.31161718544124  
Predictive power of SNOW in 2019-12-01: 183.3978834997246  
Predictive power of SNWD in 2019-12-01: 448.2693874206898

Mes: 2019-05-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-05-01: 105.92084632818934  
Predictive power of NUMBER\_OF\_SEATS in 2019-05-01: 153.99345521678438  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-05-01: 246.71256876963548  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-05-01: 236.93469269396016  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-05-01: 172.47220072182756  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-05-01: 258.92081950710246  
Predictive power of PLANE\_AGE in 2019-05-01: 44.68670866964524  
Predictive power of PRCP in 2019-05-01: 415.0793281351098  
Predictive power of TMAX in 2019-05-01: 440.87744407933656  
Predictive power of AWND in 2019-05-01: 112.45319205342741  
Predictive power of CARRIER\_HISTORICAL in 2019-05-01: 1121.9162608728766  
Predictive power of DEP\_AIRPORT\_HIST in 2019-05-01: 11.760950543902572  
Predictive power of DAY\_HISTORICAL in 2019-05-01: 1105.5737982088208  
Predictive power of DEP\_BLOCK\_HIST in 2019-05-01: 337.6335306961129  
Predictive power of SNOW in 2019-05-01: 389.1726123771992



Predictive power of SNWD in 2019-05-01: 505.324755963218

Mes: 2019-09-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-09-01: 2.9694455138774254  
Predictive power of NUMBER\_OF\_SEATS in 2019-09-01: 80.33381759146985  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-09-01: 42.047624887161874  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-09-01: 22.087510042357458  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-09-01: 0.5579088331589095  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-09-01: 167.04742706338632  
Predictive power of PLANE\_AGE in 2019-09-01: 43.971704708163635  
Predictive power of PRCP in 2019-09-01: 215.18691933141346  
Predictive power of TMAX in 2019-09-01: 155.75255895903123  
Predictive power of AWND in 2019-09-01: 130.5207904928851  
Predictive power of CARRIER\_HISTORICAL in 2019-09-01: 580.1055844213366  
Predictive power of DEP\_AIRPORT\_HIST in 2019-09-01: 47.99756772520315  
Predictive power of DAY\_HISTORICAL in 2019-09-01: 530.0316105838708  
Predictive power of DEP\_BLOCK\_HIST in 2019-09-01: 80.10174304972357  
Predictive power of SNOW in 2019-09-01: 125.73040227248619  
Predictive power of SNWD in 2019-09-01: 307.4649634422305

Mes: 2019-04-01

Predictive power of CONCURRENT\_FLIGHTS in 2019-04-01: 80.13639167543084  
Predictive power of NUMBER\_OF\_SEATS in 2019-04-01: 72.34694612138178  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2019-04-01: 140.56794159033788  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2019-04-01: 128.63299879193312  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2019-04-01: 68.3275709986355  
Predictive power of GROUND\_SERV\_PER\_PASS in 2019-04-01: 155.99795065749404  
Predictive power of PLANE\_AGE in 2019-04-01: 113.80949870610051  
Predictive power of PRCP in 2019-04-01: 236.29690894459932  
Predictive power of TMAX in 2019-04-01: 264.0263289300184  
Predictive power of AWND in 2019-04-01: 293.4621340033725  
Predictive power of CARRIER\_HISTORICAL in 2019-04-01: 648.9543842032774  
Predictive power of DEP\_AIRPORT\_HIST in 2019-04-01: 14.80539746687969  
Predictive power of DAY\_HISTORICAL in 2019-04-01: 651.5223232702161  
Predictive power of DEP\_BLOCK\_HIST in 2019-04-01: 207.9385802237747  
Predictive power of SNOW in 2019-04-01: 219.60725934412628  
Predictive power of SNWD in 2019-04-01: 301.12405376332276

Mes: 2020-01-01

Predictive power of CONCURRENT\_FLIGHTS in 2020-01-01: 2.666996689893284  
Predictive power of NUMBER\_OF\_SEATS in 2020-01-01: 8.282565216826747  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2020-01-01: 7.091472967475954  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2020-01-01: 3.9406671999220695  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2020-01-01: 2.697016047607645  
Predictive power of GROUND\_SERV\_PER\_PASS in 2020-01-01: 21.957619761146614  
Predictive power of PLANE\_AGE in 2020-01-01: 0.1040427105673951  
Predictive power of PRCP in 2020-01-01: 12.500779537068835  
Predictive power of TMAX in 2020-01-01: 32.17900489926162  
Predictive power of AWND in 2020-01-01: 3.405959034368875  
Predictive power of CARRIER\_HISTORICAL in 2020-01-01: 56.400940414979225  
Predictive power of DEP\_AIRPORT\_HIST in 2020-01-01: 4.142857142857142  
Predictive power of DAY\_HISTORICAL in 2020-01-01: 56.9881095831545  
Predictive power of DEP\_BLOCK\_HIST in 2020-01-01: 11.539954909359729  
Predictive power of SNOW in 2020-01-01: 7.497363137972131  
Predictive power of SNWD in 2020-01-01: 25.42126116622356

Mes: 2018-12-01

Predictive power of CONCURRENT\_FLIGHTS in 2018-12-01: 4.337003211488326  
Predictive power of NUMBER\_OF\_SEATS in 2018-12-01: 1.1916657058438507  
Predictive power of AVG\_MONTHLY\_PASS\_AIRPORT in 2018-12-01: 0.0799124592077258  
Predictive power of AVG\_MONTHLY\_PASS\_AIRLINE in 2018-12-01: 0.05629850959019169  
Predictive power of FLT\_ATTENDANTS\_PER\_PASS in 2018-12-01: 0.1416341490100836  
Predictive power of GROUND\_SERV\_PER\_PASS in 2018-12-01: 0.3479191541002094  
Predictive power of PLANE\_AGE in 2018-12-01: 3.8979761538166064  
Predictive power of PRCP in 2018-12-01: 0.31807223684213176  
Predictive power of TMAX in 2018-12-01: 0.4246112863744765  
Predictive power of AWND in 2018-12-01: -23.0  
Predictive power of CARRIER\_HISTORICAL in 2018-12-01: 0.1580723821277754  
Predictive power of DEP\_AIRPORT\_HIST in 2018-12-01: 92.0

Predictive power of DAY\_HISTORICAL in 2018-12-01: 0.20785918321785787  
Predictive power of DEP\_BLOCK\_HIST in 2018-12-01: 0.0531106294931258  
Predictive power of SNOW in 2018-12-01: 3.6081376335400823  
Predictive power of SNWD in 2018-12-01: 2.4448382154567545

## Principales observaciones:

1. **Variables históricas (como CARRIER\_HISTORICAL, DAY\_HISTORICAL, y DEP\_BLOCK\_HIST):** Estas variables presentan consistentemente altos valores de poder predictivo en casi todos los meses. Esto sugiere que los datos históricos de operaciones de vuelos y bloques de horarios son predictivos para las variables objetivo en los modelos, independientemente del mes.
2. **Condiciones meteorológicas (PRCP, TMAX, SNOW, SNWD):** Las variables relacionadas con la precipitación, la temperatura máxima y la nieve también muestran un poder predictivo elevado, especialmente en los meses de invierno (enero y febrero), lo cual es consistente con el impacto esperado de condiciones meteorológicas severas en la operación de vuelos.
3. **Capacidad y servicio (NUMBER\_OF\_SEATS, GROUND\_SERV\_PER\_PASS, FLT\_ATTENDANTS\_PER\_PASS):** La capacidad del avión y los servicios proporcionados a los pasajeros parecen tener poder predictivo moderado, especialmente en meses de alta demanda como agosto y noviembre. Esto podría indicar que la demanda de asientos y la calidad del servicio influye en la predicción durante temporadas de mayor tráfico aéreo.
4. **Años de la aeronave (PLANE\_AGE):** Esta variable tiene un bajo poder predictivo en la mayoría de los meses, con algunos picos en meses específicos como enero. Esto podría sugerir que la antigüedad de la aeronave tiene un impacto menor en las predicciones, posiblemente porque los modelos tienden a ser más predictivos cuando se enfocan en operaciones y datos históricos en lugar de las características físicas de los aviones.
5. **Fluctuaciones en las variables meteorológicas:** Los valores de poder predictivo de PRCP, TMAX, SNOW, y SNWD varían mucho entre meses, lo que sugiere que el impacto de las condiciones meteorológicas depende de la estación y, por tanto, del clima típico de cada mes.

6. **Promedio mensual de pasajeros (AVG\_MONTHLY\_PASS\_AIRPORT y AVG\_MONTHLY\_PASS\_AIRLINE):** Estas variables muestran un poder predictivo relativamente alto en meses de mayor demanda (como marzo y junio), lo que puede indicar que el flujo de pasajeros en ciertos meses afecta la predictibilidad de los retrasos o problemas operativos.

#### **Resumen de los meses:**

- **Enero y febrero:** Las variables históricas y meteorológicas tienen un papel importante, probablemente debido al clima invernal y al comportamiento histórico de vuelos en condiciones difíciles.
- **Marzo y junio:** Meses de alta demanda en los que el número de pasajeros y la cantidad de asientos disponibles juegan un rol predictivo significativo.
- **Agosto y noviembre:** Se observa un aumento en el poder predictivo de variables como el CONCURRENT\_FLIGHTS y GROUND\_SERV\_PER\_PASS, lo cual podría reflejar los retos operativos en temporadas de alta actividad.

## **Caso de uso del modelo (enfoque al usuario final)**

### **Descripción General del Caso de Uso**

Este modelo de predicción de retrasos en vuelos está diseñado para apoyar la planificación y toma de decisiones de las aerolíneas, aeropuertos, y otros actores de la industria. El modelo predice si un vuelo tendrá un retraso mayor a 15 minutos (DEP\_DEL15) basándose en múltiples características del vuelo, tales como el número de asientos, el horario del vuelo, la aerolínea, y datos históricos de rendimiento. El objetivo es permitir a los usuarios finales tomar medidas proactivas para reducir el impacto de los retrasos en la operación y mejorar la satisfacción de los pasajeros.

### **Implementación del Modelo (Para Usuarios Finales)**

#### **1. Entrada de Datos (Input Data)**

- Los datos de entrada consisten en una serie de características relacionadas con el vuelo (ej., horario, ruta, número de asientos) y condiciones actuales (ej., clima, tráfico en el aeropuerto).

- Estos datos se recopilan a partir de sistemas de información de vuelos (Flight Information Systems), fuentes de datos meteorológicos, y sistemas de tráfico aeroportuario.
- La mayoría de los datos se obtienen automáticamente, pero también existe la opción de ingreso manual de información adicional si fuera necesario.

## 2. Predicción del Modelo

- Una vez que se recopilan los datos, estos se envían al modelo XGBoost, el cual genera una predicción binaria (0 o 1) sobre si el vuelo tendrá un retraso mayor a 15 minutos.
- El modelo utiliza una arquitectura de gradiente de árboles (XGBoost), optimizada para precisión y eficiencia, y se ha entrenado y validado previamente para asegurar la máxima precisión en este contexto.

## 3. Visualización de Resultados

- Los resultados de las predicciones se visualizan en una interfaz de usuario intuitiva, idealmente integrada en el sistema de gestión de vuelos (Flight Management System).
- Los vuelos pronosticados con un retraso se resaltan en la pantalla de operaciones, permitiendo a los operadores enfocarse rápidamente en estos casos.
- Se proporciona una “matriz de confusión” y otras métricas clave para el equipo de operaciones para evaluar el desempeño general del modelo en tiempo real.

## 4. Acciones Basadas en la Predicción

- Si el modelo predice un retraso:
  - **Reasignación de Recursos:** Los operadores pueden reasignar puertas, personal de atención, y otros recursos clave para minimizar el impacto del retraso.
  - **Notificaciones a Pasajeros:** A través de la aplicación de la aerolínea o mensajes SMS, los pasajeros pueden ser notificados del posible retraso, ayudando a reducir el estrés y mejorar la experiencia del cliente.

- **Decisiones Operativas:** Los gerentes de operaciones pueden decidir priorizar ciertos vuelos o redirigir aviones y personal para reducir la acumulación de retrasos en vuelos posteriores.
- Si el modelo no predice un retraso:
  - El vuelo procede sin modificaciones, aunque el sistema sigue monitoreando las condiciones en tiempo real.

## 5. Monitoreo Continuo y Retraining del Modelo

- Se lleva un seguimiento de las predicciones y su precisión para asegurarse de que el modelo mantiene su efectividad en condiciones cambiantes.
- Cada cierto tiempo, el modelo se reentrena utilizando datos más recientes para reflejar patrones nuevos o condiciones cambiantes en los retrasos de vuelo.
- Además, el sistema tiene una funcionalidad de retroalimentación, permitiendo a los usuarios marcar casos en los que el modelo falló, para que los ingenieros de datos puedan investigar y ajustar el modelo si es necesario.

## Interfaz de Usuario Final

La interfaz de usuario final debería incluir los siguientes componentes:

- **Dashboard Principal:** Un tablero donde se ve el estado en tiempo real de todos los vuelos monitoreados, con alertas visuales para vuelos con alto riesgo de retraso.
- **Detalles del Vuelo:** Al hacer clic en un vuelo específico, los operadores pueden ver una lista detallada de las características que influyeron en la predicción del retraso, permitiendo entender mejor los factores detrás de cada predicción.
- **Historial y Reportes:** Un módulo que muestra el historial de precisión del modelo y reportes sobre el desempeño de las predicciones, incluyendo las métricas clave como precisión, sensibilidad y especificidad.

## Consideraciones para el Usuario Final

1. **Facilidad de Uso:** La interfaz debe ser simple e intuitiva, evitando sobrecargar al usuario con información técnica.
2. **Interpretabilidad:** Para fomentar la confianza en el modelo, sería útil implementar métodos de interpretabilidad como SHAP (Shapley Additive Explanations) para mostrar a los usuarios cómo cada característica contribuye a la predicción.
3. **Escalabilidad:** Dado el volumen de vuelos que puede manejar una aerolínea o aeropuerto, el sistema debe estar preparado para procesar una gran cantidad de predicciones en tiempo real.
4. **Tiempo de Respuesta:** El modelo y la infraestructura deben estar optimizados para entregar predicciones de manera rápida (menos de un segundo por vuelo), ya que las decisiones operativas a menudo son sensibles al tiempo.

### Beneficios Esperados

Implementar este modelo predictivo de retrasos permitirá:

- **Reducción de Costos:** Minimizar los costos asociados con los retrasos, como el combustible adicional para mantener a los pasajeros a bordo y los costos de reorganización de tripulación.
- **Mejor Experiencia para el Pasajero:** Al notificar a los pasajeros con anticipación, se reduce la incertidumbre y mejora la percepción del servicio.
- **Optimización Operacional:** Permite una asignación de recursos más eficiente y ayuda a evitar la acumulación de retrasos en vuelos posteriores.

### Conclusión

El modelo de predicción de retrasos basado en XGBoost es una herramienta poderosa para optimizar las operaciones de vuelo en tiempo real. Con la implementación adecuada, proporciona información valiosa para tomar decisiones informadas y mejorar la eficiencia operativa. Además, su flexibilidad para el retraining y su integración en sistemas de información de vuelos aseguran que se mantenga actualizado frente a cambios en los patrones de retraso.

## Punto de corte

```
import numpy as np
from sklearn.metrics import precision_recall_curve, roc_auc_score, f1_score

# Generar las probabilidades predichas por el modelo
probabilidades = search1.predict_proba(validate[ls_woe])[:, 1]

# Definir el punto de corte óptimo con base en F1-Score, Precision y Recall
precision, recall, thresholds = precision_recall_curve(validate['DEP_DEL15'], probabilidades)
f1_scores = 2 * (precision * recall) / (precision + recall)

# Encuentra el threshold óptimo
optimal_idx = np.argmax(f1_scores)
optimal_threshold = thresholds[optimal_idx]

print(f'Umbral óptimo basado en F1-Score: {optimal_threshold}')

# Aplicar el threshold para generar predicciones binarizadas
predicciones = (probabilidades >= optimal_threshold).astype(int)

# Generar una matriz de confusión para evaluar
from sklearn.metrics import confusion_matrix

# Calcular la matriz de confusión
cm = confusion_matrix(y_true, y_pred)
print("Matriz de Confusión:\n", confusion_matrix(y_true, y_pred))
# Visualizar la matriz de confusión
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", xticklabels=["No Delay", "Delay"], yticklabels=["No Delay", "Delay"])
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix")
plt.show()
```

El código realiza una evaluación del modelo de predicción de retrasos en vuelos utilizando métricas como **precision**, **recall** y **F1-Score**. Primero, genera las probabilidades de predicción para cada vuelo y luego calcula una **curva de precisión-recall** para diferentes umbrales de decisión. A partir de esta curva, se selecciona el umbral óptimo que maximiza el **F1-Score**, equilibrando así precisión y recall.

Con el umbral óptimo, las probabilidades predichas se convierten en predicciones binarias (retraso o no retraso). Luego, se calcula la **matriz de confusión**, que muestra la cantidad de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Finalmente, se visualiza esta matriz en un gráfico de calor, lo que facilita la interpretación de cómo el modelo clasifica correctamente o incorrectamente los vuelos según el retraso.

## Umbral óptimo (F-1 score)

El **umbral óptimo basado en F1-Score** que se calculó fue *0.2006*. Esto significa que para considerar una predicción como positiva (es decir, que el vuelo se retrase, DEP\_DEL15 = 1), el modelo debe predecir una probabilidad mayor o igual a 0.2006. Este umbral es el que maximiza el **F1-Score**, que es una métrica que balancea la precisión y

el recall, y es útil especialmente cuando tienes un desbalance en las clases (por ejemplo, cuando los retrasos en los vuelos son menos frecuentes).

### Matriz de confusión (sin threshold):

[[92502 366]

[17954 625]]

Esta matriz tiene dos filas y dos columnas que corresponden a las predicciones y las etiquetas reales. Se interpreta de la siguiente manera:

- **Verdaderos Negativos (TN) = 92502:** El modelo predijo que no habría retraso y efectivamente no hubo retraso.
- **Falsos Positivos (FP) = 366:** El modelo predijo que habría retraso, pero no hubo retraso real.
- **Falsos Negativos (FN) = 17954:** El modelo predijo que no habría retraso, pero sí ocurrió un retraso real.
- **Verdaderos Positivos (TP) = 625:** El modelo predijo que habría retraso y efectivamente hubo retraso.

### Interpretación en el contexto del modelo:

1. **Precisión** (Precision): Es la proporción de predicciones positivas correctas (retraso) entre todas las predicciones positivas del modelo (retraso y no retraso predicho).

$$\text{Precisión} = \frac{TP}{TP + FP} = \frac{625}{625 + 366} \approx 0.63$$

2. **Recall** (Sensibilidad): Es la proporción de casos positivos reales (retrasos) que fueron correctamente identificados por el modelo.

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{625}{625 + 17954} \approx 0.034$$

Esto indica que el modelo detecta solo una pequeña fracción de los retrasos reales.



3. **F1-Score:** Es la media armónica entre precisión y recall, siendo más relevante cuando se busca un balance entre ambas métricas.

$$F1 = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}} \approx 0.065$$

#### Puntos clave para la aplicación final:

- **F1-Score bajo:** Esto puede indicar que el modelo es muy conservador en predecir los retrasos y, por lo tanto, tiene un bajo **recall**. En términos de negocio, esto significa que el modelo está perdiendo muchos retrasos reales, lo cual puede ser problemático si se necesita detectar la mayor cantidad posible de vuelos retrasados.

En resumen, el modelo está funcionando con una **alta precisión**, pero **bajo recall**, lo que significa que, aunque cuando predice un retraso es correcto en la mayoría de los casos, no está detectando muchos de los casos de retrasos reales.

#### Matriz de Confusión aplicada con el umbral de 0.2:

[[70830 22038]

[ 9190 9389]]

#### Cálculos:

1. **Precisión** (Precision) es la proporción de predicciones positivas correctas entre todas las predicciones positivas realizadas:

$$\text{Precisión} = \frac{TP}{TP + FP} = \frac{9,389}{9,389 + 22,038} = \frac{9,389}{31,427} \approx 0.298$$

Esto significa que aproximadamente el **29.8%** de los vuelos predichos como retrasados efectivamente tuvieron un retraso.

2. **Recall** (también conocido como Sensibilidad o Tasa de Verdaderos Positivos) es la proporción de los casos positivos reales que fueron correctamente identificados por el modelo:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{9,389}{9,389 + 9,190} = \frac{9,389}{18,579} \approx 0.505$$

Esto indica que el modelo identificó correctamente el **50.5%** de los vuelos que realmente tuvieron retraso.

3. **F1-Score** es la media armónica entre precisión y recall, y proporciona un balance entre ambos. Se calcula como:

$$\text{F1-Score} = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}} = 2 \times \frac{0.298 \times 0.505}{0.298 + 0.505} \approx 0.373$$

El **F1-Score** es **0.373**, lo que sugiere que el modelo tiene un balance entre precisión y recall, aunque sigue siendo bajo debido a la mayor cantidad de **falsos positivos**.

#### Comparación con la matriz anterior (umbral 0.5):

La matriz de confusión anterior era:

[[92502 366]  
[17954 625]]

Y los cálculos previos para **precisión**, **recall** y **F1-score** fueron:

1. **Precisión: 0.630**
2. **Recall: 0.034**
3. **F1-Score: 0.065**

#### Análisis Comparativo:

- **Precisión** ha disminuido considerablemente (de **0.630** a **0.298**). Esto significa que el modelo ahora clasifica muchos más vuelos como "Delay" cuando no lo son, generando más **falsos positivos**. Esta es una **pérdida de precisión**, lo cual es un **punto de mejora**.
- **Recall** ha aumentado significativamente (de **0.033** a **0.505**). Esto implica que el modelo ha mejorado mucho en detectar los vuelos retrasados, pero esto también genera un mayor número de **falsos positivos**, lo que es una **compensación** para evitar que se pierdan vuelos con retraso.

- **F1-Score** ha subido (de **0.062** a **0.373**), lo que refleja un balance mejorado entre precisión y recall, aunque aún está lejos de un valor óptimo. Este aumento es positivo, ya que muestra que el modelo está teniendo un mejor rendimiento global, pero el F1-Score aún no es alto.

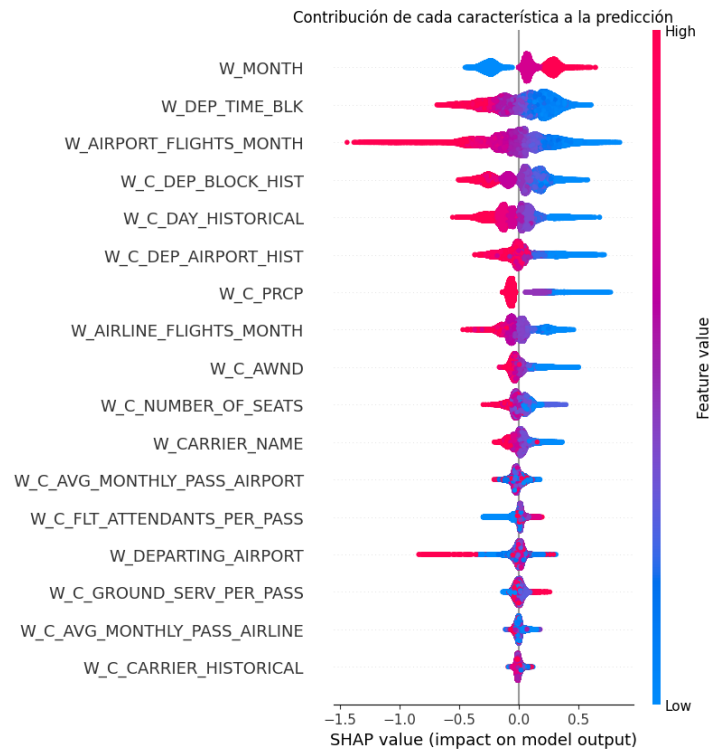
### Conclusión:

1. **Mejoras:** El ajuste del umbral a **0.2** mejora significativamente el **recall**, permitiendo que el modelo identifique muchos más vuelos con retraso, lo cual es valioso si el objetivo es detectar casi todos los vuelos retrasados, aunque se asuma el costo de más falsas alarmas.
2. **Áreas de mejora:** La **precisión** se ve afectada negativamente con el umbral bajo, lo que lleva a un aumento de **falsos positivos**. Si los falsos positivos (predicciones incorrectas de "Delay") tienen un alto costo en términos operativos o logísticos, se puede considerar ajustar el umbral hacia un valor más alto para reducir esta tasa, aunque esto podría reducir el recall.
3. **Recomendación:** Sería útil probar con un **umbral intermedio** entre 0.2 y 0.5 para encontrar un mejor balance entre precisión, recall y F1-score, dependiendo de los objetivos del modelo (más énfasis en detectar retrasos o evitar falsas alarmas)

## Interpretabilidad/SHAP VALUES

Los **SHAP values** (Shapley Additive exPlanations) son una técnica de interpretación de modelos que se utiliza para desglosar y comprender cómo cada característica de un modelo influye en la predicción de una instancia específica. Inspirados en la teoría de juegos, los valores SHAP asignan una "contribución justa" a cada característica en el resultado de la predicción, basándose en el concepto de valores de Shapley, que buscan distribuir el impacto de cada factor de manera equitativa entre todos los jugadores (en este caso, las características).

En el contexto de modelos de aprendizaje automático, los SHAP values proporcionan una explicación clara y cuantificable de por qué un modelo predice un resultado particular, lo que permite a los analistas y expertos en datos entender la relación entre las variables de entrada y las decisiones del modelo. Estos valores son especialmente útiles para modelos complejos, como los árboles de decisión o las redes neuronales, donde la interpretabilidad suele ser limitada. Al visualizar los SHAP values, se puede identificar qué características son las más influyentes, cómo impactan en las predicciones y de qué manera contribuyen al comportamiento del modelo en general.



## ¿Qué muestra el gráfico?

### 1. Eje X (Valores SHAP):

- El eje X muestra los valores SHAP, que representan el impacto de cada característica sobre la predicción del modelo. Un valor SHAP positivo significa que la característica está empujando la predicción hacia una clase positiva (por ejemplo, 1 si DEP\_DEL15 es binaria y refleja si hubo un retraso o no), y un valor negativo indica que la característica está empujando la predicción hacia la clase negativa (por ejemplo, 0, si el vuelo no tuvo retraso).
- Los valores SHAP reflejan el cambio en la probabilidad de la predicción del modelo, en comparación con la predicción base (cuando todas las características son 0 o el valor promedio).

### 2. Eje Y (Características):

- En el eje Y, las características del modelo están listadas, generalmente ordenadas por su importancia. Las características en la parte superior son las que tienen un mayor impacto en las predicciones del modelo.

- Las características que tienen un gran rango de valores SHAP son generalmente las que tienen una mayor influencia en la predicción.

### 3. Distribución de los puntos:

- Cada punto en el gráfico corresponde a un **registro individual** en el conjunto de datos. Estos puntos se muestran en un gráfico de dispersión a lo largo del eje X (valores SHAP).
- Los puntos pueden estar dispersos de izquierda a derecha, indicando la amplitud de la contribución de cada característica. Si una característica tiene un rango amplio de valores SHAP, es más probable que tenga un mayor impacto en la predicción.

### 4. Colores:

- Generalmente indican la **distribución de los valores de esa característica**. Por ejemplo, los puntos de color azul y rojo, podrían indicar que la variable tiene una **distribución de valores bajos (azul) y altos (rojo)**. Esto puede ayudarte a entender cómo las contribuciones cambian según el valor de la característica.

## ¿Cómo interpretar los SHAP values en el gráfico?

### 1. Dirección de los valores SHAP:

- Si una característica tiene un valor SHAP mayor que 0 (a la derecha), eso significa que la característica está **empujando la predicción hacia una clase positiva** (en tu caso, podría estar relacionado con la probabilidad de que un vuelo se retrase).
- Si el valor SHAP es menor que 0 (a la izquierda), esa característica está empujando la predicción hacia una clase negativa.

### 2. Tamaño de la dispersión:

- Las características que tienen una **dispersión amplia** en el eje X tienen un **mayor rango de influencia**. Esto puede significar que esta característica tiene un efecto significativo sobre las predicciones en diferentes niveles.

- Si los puntos están más concentrados en un área estrecha, eso indica que la característica tiene un impacto más **constante o pequeño**.

### 3. Colores:

- Los puntos se distribuyen de manera diferente en colores (por ejemplo, de azul a rojo), eso indica que el impacto de la característica cambia dependiendo de sus valores. Si hay una tendencia clara de color con los valores SHAP, eso muestra cómo el impacto de esa característica varía con su valor

### Variables con dispersión hacia la derecha (impacto positivo en la predicción de retrasos):

Las variables que tienen una dispersión **hacia la derecha** (valores SHAP positivos) son aquellas que, cuando tienen valores más altos, tienden a aumentar la probabilidad de que un vuelo se retrase (es decir,  $DEP\_DEL15 = 1$ ).

#### 1. 'W\_C\_DAY\_HISTORICAL':

- **Interpretación:** Esta variable probablemente está relacionada con los **retrasos históricos del día**. Si esta variable tiene un valor alto (indicando que el vuelo en ese día tiene más retrasos históricos), su valor SHAP positivo indica que **más retrasos en el pasado en el mismo día aumentan la probabilidad de un retraso en el futuro**. Es una característica que probablemente indica un patrón de retrasos recurrentes en ese día de la semana.

#### 2. 'W\_C\_PRCP' (Precipitación):

3. **Interpretación:** El valor SHAP positivo aquí sugiere que una **mayor cantidad de precipitación (lluvia, tormentas)** está asociada con un aumento en la probabilidad de que el vuelo se retrase. Esto tiene sentido, ya que las condiciones meteorológicas adversas pueden afectar la puntualidad de los vuelos.

#### 4. 'W\_C\_DEP\_AIRPORT\_HIST':

- **Interpretación:** Relacionado con los **retrasos históricos en el aeropuerto de salida**. Si este valor es alto, probablemente indica que el aeropuerto tiene un historial de retrasos. Esto está relacionado con el **efecto acumulativo de retrasos previos**, lo que sugiere que un aeropuerto con retrasos previos aumenta la probabilidad de que se repitan esos retrasos.

#### 5. 'W\_C\_AWMD' (Posiblemente relacionado con el viento o la fuerza del viento):

- **Interpretación:** Un valor SHAP positivo aquí sugiere que **condiciones meteorológicas extremas, como vientos fuertes**, podrían aumentar la probabilidad de que los vuelos se retrasen, ya que el mal tiempo en el aire o en la pista afecta las operaciones de los vuelos.

#### 6. 'W\_CARRIER\_NAME':

- **Interpretación:** Si el valor SHAP es positivo, eso indica que la **aerolínea** asociada con el vuelo tiene una alta probabilidad de causar retrasos. Algunas aerolíneas pueden tener un historial de retrasos más frecuente debido a sus rutas o flotas.

### **Variables con dispersión hacia la izquierda (impacto negativo en la predicción de retrasos):**

Las variables que tienen una dispersión **hacia la izquierda** (valores SHAP negativos) son aquellas que, cuando tienen valores más altos, tienden a **reducir la probabilidad de que el vuelo se retrase** (es decir, empujan la predicción hacia la clase negativa: DEP\_DEL15 = 0).

#### 1. 'W\_DEP\_TIME\_BLK':

- **Interpretación:** Esta variable está relacionada con la **franja horaria de salida** del vuelo. Si el valor SHAP es negativo, eso indica que los vuelos que salen en ciertas franjas horarias (por ejemplo, en la mañana o en horas menos concurridas) tienen menos probabilidad de retrasarse. Esto es consistente con la idea de que los vuelos en horarios menos congestionados son más puntuales.

#### 2. 'W\_AIRPORT\_FLIGHTS\_MONTH':

- **Interpretación:** Esta variable está relacionada con la cantidad de **vuelos en un mes específico**. Si los valores SHAP para esta variable son negativos, esto sugiere que **más vuelos en un mes** (probablemente debido a una mayor cantidad de vuelos operados en ese mes) podrían estar asociados con **menos retrasos**. Esto podría ser un efecto de mayor experiencia o eficiencia en la gestión de vuelos en los meses con más actividad.

#### 3. 'W\_DEPARTING\_AIRPORT':

- **Interpretación:** Si esta variable tiene un valor SHAP negativo, eso sugiere que algunos **aeropuertos** específicos están asociados con menos retrasos en los vuelos. Es posible que algunos aeropuertos tengan mejor infraestructura, menor congestión o procesos operativos más eficientes, lo que reduce la probabilidad de que los vuelos sufran retrasos.

### **Resumen de la interpretación:**

- **Variables que aumentan la probabilidad de retrasos (hacia la derecha):**
  - Variables como **precipitación** (W\_C\_PRCP), **retrasos históricos** en el día o aeropuerto (W\_C\_DAY\_HISTORICAL, W\_C\_DEP\_AIRPORT\_HIST), y condiciones meteorológicas adversas (como el viento W\_C\_AWMD) están asociadas con un mayor riesgo de retraso.
  - **Aerolíneas** (W\_CARRIER\_NAME) con un historial de retrasos también parecen tener un mayor impacto en la probabilidad de que los vuelos se retrasen.
- **Variables que reducen la probabilidad de retrasos (hacia la izquierda):**
  - **Horarios de salida** (W\_DEP\_TIME\_BLK) en horas menos congestionadas están asociadas con menos retrasos.
  - **Mayor cantidad de vuelos** en un mes (W\_AIRPORT\_FLIGHTS\_MONTH) y la **infraestructura del aeropuerto** (W\_DEPARTING\_AIRPORT) podrían ser factores que contribuyen a una mejor puntualidad, reduciendo la probabilidad de retrasos.

Estos patrones proporcionan una visión útil para entender qué factores están influenciando más las predicciones de retrasos en los vuelos. Las características que empujan las predicciones hacia la derecha son factores de riesgo, mientras que las que las empujan hacia la izquierda son factores que ayudan a prevenir los retrasos.



