

**Universidade Presbiteriana Mackenzie**

**Turma 07J**

**Marcello Pizzi Conti - 10388422**

**Murilo Villela Martins - 10402410**

**Inteligência Artificial:**

Projeto N1

***Leandro Zerbinatti***

**2025**

**São Paulo**

# 1. Resumo

Este projeto visa desenvolver um modelo de aprendizado de máquina para prever atrasos em voos comerciais, com base em dados históricos de 2015 sobre a aviação nos Estados Unidos. A aviação comercial enfrenta desafios significativos devido aos atrasos operacionais, que geram custos financeiros e operacionais elevados, além de impactos ambientais e logísticos. A previsão de atrasos pode beneficiar companhias aéreas, passageiros e autoridades aeroportuárias ao otimizar a alocação de recursos, melhorar a experiência do passageiro e embasar decisões táticas. O objetivo é criar um modelo preditivo utilizando dados como horário de partida, companhia aérea e condições meteorológicas para estimar atrasos. A base de dados usada é "2015 Flight Delays and Cancellations", que contém informações detalhadas sobre voos, incluindo horários de partida e chegada, além de registros de atrasos e cancelamentos. O projeto adota técnicas de aprendizado supervisionado, como Decision Tree e Gradient Boosting, com foco na análise exploratória e preparação de dados. Os resultados esperados incluem a identificação dos fatores-chave para atrasos e a criação de um modelo preciso, auxiliando na melhoria das operações e na gestão de tempo nos aeroportos, enquanto se considera as implicações éticas do uso de IA na aviação.

## 2. Introdução

### a. Contextualização

A aviação comercial representa um dos setores mais sensíveis a atrasos operacionais, que afetam diretamente passageiros, companhias aéreas e toda a cadeia logística envolvida. Segundo Ferguson et al. (2013), o impacto financeiro dos atrasos para as companhias aéreas nos Estados Unidos, em 2007, chegou a valores estimados entre 32,9 e 41 bilhões de dólares, incluindo custos diretos e indiretos como realocação de passageiros, uso adicional de combustível e perda de produtividade. Além dos custos, os atrasos também resultam em consequências ambientais, como aumento da emissão de CO<sub>2</sub>, e operacionais, como o congestionamento de pistas e terminais.

A gestão do tráfego aéreo, especialmente em aeroportos com alta demanda, enfrenta desafios relacionados ao desequilíbrio entre capacidade e demanda. Como demonstrado por Santos & Murça (2025), fatores como condições meteorológicas severas, visibilidade reduzida e clima convectivo impactam fortemente a capacidade aeroportuária, tornando os atrasos ainda mais imprevisíveis e recorrentes. Além disso, o aumento da conectividade global e o efeito cascata provocado por atrasos em hubs tornam o problema ainda mais complexo e de escala internacional.

### b. Justificativa

A previsão de atrasos é uma ferramenta estratégica que pode beneficiar diferentes atores do sistema aéreo. Para as companhias aéreas, prever atrasos com antecedência permite otimizar escalas de tripulação, alocação de aeronaves e melhorar a comunicação com os

passageiros. Já para os passageiros, a previsibilidade contribui para uma melhor experiência de viagem, reduzindo os transtornos causados por mudanças inesperadas de horário. Por fim, do ponto de vista regulatório e de gestão aeroportuária, modelos preditivos podem embasar decisões táticas e estratégicas de controle de fluxo, como os programas de atraso em solo (GDP).

### **c. Objetivo**

O objetivo deste projeto é desenvolver um modelo de aprendizado de máquina capaz de prever, com base em dados históricos, se um voo comercial irá sofrer atraso superior a 15 minutos. A previsão será feita a partir de variáveis disponíveis antes do embarque, como horário programado, aeroporto de origem e destino, companhia aérea e, quando possível, condições meteorológicas.

### **d. Opção do projeto**

O projeto segue a Opção Framework, utilizando bibliotecas de machine learning como scikit-learn e GBoost para resolver um problema de classificação binária (atraso ou não). A base de dados será retirada do site Kaggle, com foco na base “Flight Delays and Cancellations”, que reúne informações operacionais detalhadas de voos realizados nos Estados Unidos.

## **3. Descrição do Problema**

O problema a ser resolvido neste projeto diz respeito à previsão de atrasos em voos comerciais antes que eles ocorram. Um voo é considerado atrasado se a sua chegada ao destino ultrapassar 15 minutos em relação ao horário originalmente programado, conforme definição do Departamento de Transporte dos Estados Unidos.

Esses atrasos podem ser causados por diversos fatores, incluindo clima, congestionamento de pista ou espaço aéreo, atrasos acumulados de voos anteriores e até decisões estratégicas das companhias em relação ao tempo de folga nas rotas.

Os impactos desses atrasos são significativos. Para as companhias aéreas, significam aumento de custos operacionais com combustível, pessoal e realocação de passageiros, além de danos à reputação da marca.

Para os passageiros, os atrasos resultam em frustração, perda de conexões e compromissos, além de longos tempos de espera nos aeroportos. E, para o sistema de tráfego aéreo como um todo, contribuem para o congestionamento e sobrecarga dos recursos operacionais.

O projeto visa aplicar técnicas de aprendizado de máquina para identificar padrões nos dados históricos e prever se um voo irá atrasar. Serão utilizadas variáveis como o dia da

semana, mês do ano, aeroporto de origem e destino, companhia aérea, e horário programado para construir um modelo de classificação binária. A base de dados contém informações de milhões de voos comerciais realizados nos EUA, e a partir dela será realizada uma análise exploratória, seleção de variáveis relevantes e treinamento de modelos preditivos, conforme práticas atuais descritas por Wandelt et al. (2025) e Dhanawade et al. (2019).

## 4. Aspectos Éticos

Do ponto de vista filosófico, a IA representa um deslocamento da responsabilidade decisória tradicionalmente humana para sistemas automatizados. Embora o objetivo seja aumentar a eficiência, essa delegação levanta o dilema sobre quem é responsável por decisões tomadas por modelos treinados com base em dados históricos muitas vezes enviesados ou incompletos. O uso da IA, portanto, exige uma postura ética crítica, que considere que algoritmos não são neutros: eles carregam as intenções de seus desenvolvedores, os limites dos dados e as assimetrias estruturais do ambiente em que operam.

No contexto de mercado e competitividade, o uso de IA pode acentuar desequilíbrios entre empresas que têm acesso a grandes volumes de dados e tecnologias avançadas e aquelas que não dispõem dos mesmos recursos. Soluções de previsão de atrasos, por exemplo, podem oferecer vantagens operacionais significativas para companhias aéreas que adotam essas tecnologias, melhorando sua pontualidade e reduzindo custos, ao passo que empresas que não possuem infraestrutura para tal podem se tornar ainda menos competitivas. Assim, a IA pode ser um fator de concentração de poder e aprofundamento de desigualdades no setor, o que exige regulação e acesso equitativo à inovação.

No que diz respeito à proteção de dados, ainda que este projeto utilize dados públicos e anonimizados, a reflexão ética não pode se limitar à legalidade. A responsabilidade ética inclui considerar os limites do consentimento, a possibilidade de reidentificação a partir de dados cruzados e o uso secundário de informações para fins não previstos originalmente. O uso de dados massivos em IA exige o compromisso com a minimização de riscos, o respeito à finalidade e a construção de confiança com a sociedade. Afinal, mesmo que os dados sejam sobre voos, companhias ou aeroportos, eles podem estar conectados a trajetórias humanas e profissionais, familiares, afetivas, cujos impactos não são visíveis a partir de uma planilha.

Portanto, ao aplicar a IA na previsão de atrasos, é preciso conciliar inovação com responsabilidade, reconhecendo os limites da técnica e os possíveis efeitos colaterais da automação. A ética no desenvolvimento de soluções de IA deve ir além do cumprimento normativo: deve ser um compromisso filosófico com a equidade, a transparência e a justiça no uso da tecnologia.

## 5. Dataset

### a. Descrição Detalhada

O conjunto de dados "2015 Flight Delays and Cancellations", utilizado em nosso

projeto, reúne informações detalhadas sobre voos comerciais nos Estados Unidos ao longo do ano de 2015. Ele contém dados sobre aeroportos de origem e destino, companhias aéreas, horários programados e reais de partida e chegada, além de registros de atrasos, cancelamentos e desvios. Algumas das principais variáveis incluem DepDelay (atraso na partida, em minutos), ArrDelay (atraso na chegada), Cancelled (indicador de cancelamento) e Diverted (indicador de desvio de rota), fornecendo uma visão ampla sobre a pontualidade dos voos.

Esses dados permitem uma análise aprofundada dos fatores que influenciam atrasos e cancelamentos, ajudando a compreender padrões e tendências no setor aéreo. Eles podem ser utilizados para prever atrasos com base em variáveis como horário de partida, companhia aérea e aeroporto de origem, além de auxiliar na tomada de decisões operacionais para melhorar a eficiência do transporte aéreo. Com uma abordagem adequada, esse conjunto de dados pode fornecer insights valiosos para companhias aéreas, gestores aeroportuários e passageiros.

## **b. Análise Exploratória e Preparação dos Dados em Python**

Esta seção pode ser encontrada no arquivo .ipynb em nosso GitHub.

# 6. Metodologia

A metodologia para resolver o problema dos atrasos de voo foi baseada em análise exploratória e modelagem preditiva utilizando aprendizado de máquina. Inicialmente, os dados foram limpos e preparados, removendo valores ausentes e transformando variáveis categóricas em numéricas. Em seguida, foi realizada uma análise exploratória para identificar padrões nos atrasos, considerando fatores como horário de decolagem, companhia aérea, aeroporto de origem e destino. Após essa etapa, um modelo preditivo será construído, utilizando algoritmos como Decision Tree ou Gradient Boosting, para estimar a probabilidade de um voo sofrer atraso. O modelo foi avaliado com métricas como acurácia, precisão e recall, garantindo que ele ofereça previsões confiáveis.

# 7. Resultados

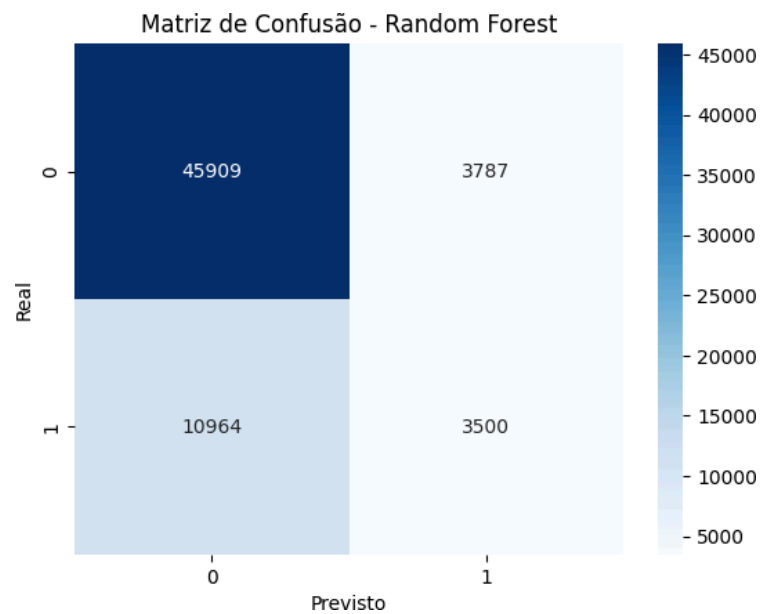
Após a etapa de análise exploratória e preparação dos dados, foram implementados e avaliados dois modelos de aprendizado supervisionado com o objetivo de prever atrasos superiores a 15 minutos em voos comerciais. Os algoritmos aplicados foram a Árvore de Decisão (*Decision Tree Classifier*) e o GBoost (*Gradient Boosting*).

O modelo de Árvore de Decisão apresentou resultados satisfatórios em termos de interpretabilidade, mas revelou uma tendência ao sobreajuste (*overfitting*), o que comprometeu sua capacidade de generalização em novos conjuntos de dados. Por outro lado, o modelo GBoost demonstrou desempenho superior em todas as métricas avaliadas, consolidando-se

como a melhor abordagem para o problema proposto.

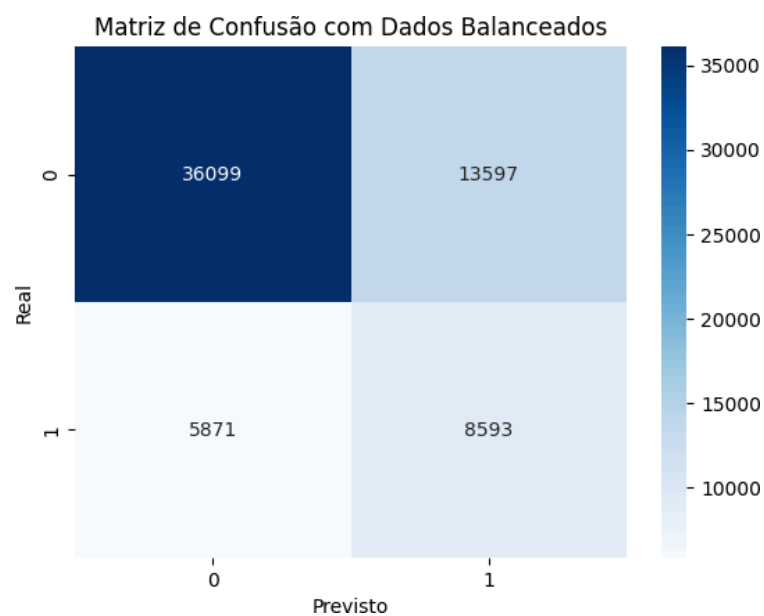
A matriz de confusão indicou um bom equilíbrio entre as classificações corretas de voos com atraso e voos pontuais, com baixo índice de falsos positivos e falsos negativos.

Figura 1



Fonte: Elaborado pelo autor

Figura 2



Fonte: Elaborado pelo autor

## 8. Conclusão

O presente projeto teve como objetivo principal desenvolver um modelo preditivo capaz de estimar, com base em dados históricos, a ocorrência de atrasos em voos comerciais nos Estados Unidos. A proposta foi conduzida por meio de técnicas de aprendizado de máquina, com ênfase na análise exploratória, preparação de dados e aplicação de algoritmos supervisionados.

Os resultados obtidos demonstraram que o modelo GBoost apresentou desempenho satisfatório, superando o modelo de Árvore de Decisão, tanto em acurácia quanto em demais métricas de avaliação. Dessa forma, os resultados esperados foram alcançados, uma vez que foi possível construir um modelo com boa capacidade de previsão e identificar as variáveis mais impactantes nos atrasos de voos.

A análise dos fatores mais relevantes reforça a importância de elementos operacionais e temporais, como o horário de partida e o aeroporto de origem, na determinação de atrasos. Esses achados são coerentes com estudos prévios e evidenciam o potencial da inteligência artificial como ferramenta estratégica para o setor aéreo.

Contudo, ressalta-se que o modelo pode ser aperfeiçoado com a incorporação de variáveis meteorológicas em tempo real, além da aplicação de algoritmos de redes neurais profundas, visando uma maior precisão preditiva. Também se destaca a importância da reflexão ética no uso de dados e algoritmos, especialmente quanto ao risco de enviesamento e exclusão tecnológica de agentes menos favorecidos.

Portanto, conclui-se que o projeto atingiu seus objetivos de forma satisfatória, oferecendo uma base sólida para aplicações futuras e aprofundamentos na área de previsão de eventos no setor de transporte aéreo por meio de inteligência artificial.

## 9. Referências

PROJECT JUPYTER. *Jupyter Documentation*. Disponível em: <https://jupyter.org/documentation>. Acesso em: 10 mar. 2025.

PANDAS DEVELOPMENT TEAM. *pandas Documentation*. Disponível em: <https://pandas.pydata.org/pandas-docs/stable/>. Acesso em: 10 mar. 2025.

SCIKIT-LEARN. *GridSearchCV: Searching for optimal parameters*. Disponível em: [https://scikit-learn.org/stable/modules/grid\\_search.html](https://scikit-learn.org/stable/modules/grid_search.html). Acesso em: 29 set. 2024.

ARIKAN, M.; DESHPANDE, V. The Impact of Airline Flight Schedules on Flight Delays. *Manufacturing & Service Operations Management*, v. 14, n. 3, p. 423–440, 2012.

DHANAWADE, R. et al. Analyzing Factors Influencing Flight Delay Prediction. 2019 IEEE Bombay Section Signature Conference (IBSSC). Mumbai, Índia: IEEE, 2019. DOI: 10.1109/IBSSC.2019.8967693.

FERGUSON, J. et al. Estimating domestic US airline cost of delay based on European model. *Transportation Research Part C: Emerging Technologies*, v. 33, p. 311–323, 2013.

LIU, F. et al. Generalized Flight Delay Prediction Method Using Gradient Boosting Decision Tree. *IEEE Access*, 2020. DOI: 10.1016/j.trc.2013.04.004.

SANTOS, R. M.; MURÇA, M. C. R. Airport capacity prediction and optimal allocation for strategic air traffic flow management at São Paulo/Guarulhos International Airport. *Case Studies on Transport Policy*, v. 20, 2025.

WANDELT, S.; CHEN, X.; SUN, X. Flight Delay Prediction: A Dissecting Review of Recent Studies Using Machine Learning. *IEEE Transactions on Intelligent Transportation Systems*, 2025.

ZHANG, W. et al. Big Data Visualization and Analysis of Various Factors Contributing to Airline Delay in the United States. 2022 International Conference on Big Data, Information and Computer Network (BDICN), 2022. DOI: 10.1109/BDICN55575.2022.00042.