Modelos Supervisionados para Previsão de Atrasos em Voos: Impacto da SMOTE no Aprendizado

Lucas Almeida Soares da Silva[1], Ramiro de Paiva Fagundes[1], Bruno Santos Luiz[1], Camile da Costa Ramos[1], Daniel Alberto Pamplona[1] e Mateus Habermann[1]

[1]Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos/SP - Brasil

***Resumo*— Atrasos em voos domésticos no Brasil representam um problema crítico, com severos impactos operacionais e financeiros. Embora o *Machine Learning* ofereça ferramentas para prever esses eventos, o desbalanceamento natural dos dados, onde voos pontuais são a maioria, cria modelos com alto viés, incapazes de identificar as classes minoritárias de atraso. Este trabalho aborda essa lacuna ao analisar o impacto da técnica de sobreamostragem SMOTE em três algoritmos: Regressão Logística Ordinal, *Decision Tree* e *Random Forest*. Os modelos foram treinados para classificar atrasos em categorias distintas. Os resultados demonstram que, sem o balanceamento, os algoritmos são ineficazes. Após a aplicação da SMOTE, a capacidade de prever atrasos aumenta drasticamente, com o *Random Forest* mostrando o melhor desempenho. O estudo conclui que a SMOTE é um pré-processamento essencial para a utilidade prática de modelos preditivos no setor aéreo.**

***Palavras-Chave*— *Machine Learning*, Balanceamento de Classes, Transporte Aéreo.**

I. Introdução

O transporte aéreo de passageiros e cargas no Brasil apresenta um crescimento consistente, impulsionado pela demanda por mobilidade e logística eficiente. Em 2024, essa tendência foi confirmada pelos dados da Agência Nacional de Aviação Civil (ANAC), que registraram um movimento de 93,4 milhões de passageiros no mercado doméstico e 488 mil toneladas de carga, representando aumentos de 2,1% e 9,7% em relação ao ano anterior, respectivamente [1]. O otimismo do setor é reforçado por projeções que indicam a continuidade dessa expansão, como a estimativa de aumento de receita no segmento de cargas [2].

Esse ritmo acelerado de crescimento, contudo, contrasta com as limitações estruturais do sistema aeroviário e aeroportuário nacional. Como consequência, um dos principais desafios operacionais do setor são os atrasos nas operações, que comprometem não apenas a eficiência logística das companhias aéreas, mas também a experiência e a satisfação dos passageiros.

A predição de atrasos em voos é um tema consolidado na literatura, impulsionado pelo seu significativo impacto financeiro e operacional. Diversos autores aplicam técnicas de aprendizado de máquina (*Machine Learning*) para analisar os fatores que influenciam a pontualidade. Modelos baseados em Árvore de Decisão (*Decision Tree*) e *Random Forest*, por exemplo, são frequentemente empregados para identificar a importância de variáveis como rotas e volume de tráfego aéreo [3], [4]. Outras abordagens exploram o uso de Redes Neurais para capturar as relações não-lineares complexas entre múltiplos fatores, incluindo condições meteorológicas adversas [5]. Modelos de Regressão Logística também são comumente utilizados para estimar a probabilidade de ocorrência de um atraso com base em características operacionais do voo [6].

Apesar dos avanços, uma parcela significativa desses estudos aborda o problema por meio de uma classificação binária — isto é, prevendo apenas se um voo irá atrasar ou não, geralmente em relação a um limiar fixo como 15 minutos [3]. Essa simplificação, embora útil para certas aplicações, trata um atraso de 16 minutos da mesma forma que um de duas horas, o que limita o potencial de análise e a utilidade prática dos modelos para um gerenciamento operacional mais refinado. Ao não diferenciar a severidade do evento, perde-se a capacidade de priorizar ações e alocar recursos de forma mais eficaz. Essa limitação evidencia uma oportunidade para pesquisas que explorem a natureza multi-nível dos atrasos.

Além disso, segundo Chawla et al. (2002) [7] conjuntos de dados do mundo real são predominantemente compostos por exemplos ”normais”, com apenas uma pequena porcentagem de exemplos ”anormais”ou ”interessantes”, sendo o custo de classificar incorretamente um exemplo anormal (interessante) como um exemplo normal muito maior do que o custo do erro inverso. Nesse cenário, mesmo algoritmos mais robustos, como *Random Forest*, necessitam do apoio de técnicas de balanceamento de classes durante seu processo de treinamento.

Dessa forma, o presente estudo propõe uma análise preditiva de atrasos em voos utilizando Regressão Logística Ordinal, *Decision Tree* e *Random Forest*. O trabalho não se limita a comparar o desempenho final desses algoritmos, mas foca, principalmente, em evidenciar o impacto da *Synthetic Minority Oversampling Technique* (SMOTE) na capacidade de aprendizado dos modelos. Com isso, busca-se reforçar a importância do tratamento de classes desbalanceadas para o aprimoramento operacional das companhias aéreas.

II. Referencial Teórico

Segundo Ditzel (2022) [8], *Machine Learning* representa um modelo que trabalha com a estimativa de determinada tarefa por meio de indicadores. Essa definição é consideravelmente ampla e pode ser aplicada em diversas áreas, desde que se considere certo grau de imprecisão. O processo de construção de modelos preditivos utilizando esse tipo de técnica não apresenta alto grau de complexidade; porém, deve considerar os seguintes elementos: os dados precisam

L.A.S. Silva, soareslass@fab.mil.br; R.P. Fagundes, ramirorpf@fab.mil.br; B.S. Luiz, santosbsl@fab.mil.br; C.C. Ramos, camileramosccr@fab.mil.br; D.A. Pamplona, pamplonadefesa@hotmail.com; M. Habermann, habermann@ita.br. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) Código de Financiamento 001.

ser inicialmente processados para evitar ruídos; as entradas pré-processadas também devem ser utilizadas no treinamento dos modelos; e o algoritmo deve ser testado com uma parte do banco de dados, com o intuito de verificar a eficácia do modelo.

A dificuldade de treinar algoritmos com dados desbalanceados é um desafio recorrente. Uma solução proeminente é a técnica SMOTE, introduzida por Chawla et al. (2002) [7]. A SMOTE é uma abordagem de pré-processamento que equilibra os dados, gerando amostras sintéticas da classe minoritária a partir da interpolação linear entre uma instância e seus vizinhos mais próximos, criando um conjunto de dados mais robusto para o treinamento.

Conforme descrito por Ge, Yue e Chen (2017) [9], a SMOTE opera identificando os vizinhos mais próximos para cada instância da classe minoritária e, em seguida, seleciona aleatoriamente amostras dessas amostras vizinhas mais próximas para realizar a interpolação linear. Assim, para cada uma das instâncias minoritárias originais, novas amostras sintéticas são produzidas até que haja o equilíbrio entre todas as classes. A aplicação dessa técnica durante o pré-processamento dos dados pretende viabilizar a melhoria na capacidade de classificação dos modelos de *Machine Learning*.

Nesse sentido, a literatura sobre análise de dados aéreos comumente utiliza conceitos relacionados a (*Machine Learning*), sendo comum encontrar modelos que exploram os métodos de *Decision Tree* e *Random Forest*[10].

Dessa forma, Choi et al. (2016) [3] utilizaram algoritmos como *Decision Tree*, *Random Forest*, *AdaBoost* e *KNN* para prever dados de 45 aeroportos dos Estados Unidos, entre os anos de 2005 e 2015. Já Sternberg et al. (2016) [4] empregaram técnicas de indexação e associação para identificar padrões nos dados de voos e de condições meteorológicas dos 17 maiores aeroportos brasileiros, no ano de 2014. O estudo revelou que há períodos do ano em que o aumento da demanda acarreta uma propagação de atrasos em cadeia. Além disso, observou-se que fatores como a região geográfica do aeroporto, o nível de demanda e as condições meteorológicas possuem relação direta com a ocorrência de atrasos.

Complementando essas abordagens, o trabalho de Barros et al. (2010) [6] utilizou a Análise Envoltória de Dados (DEA) para mensurar a pontualidade de voos, considerando as variáveis de atraso na chegada e na partida, tempo de voo e distância entre origem e destino, com base na proporção do tempo de atraso em relação ao tempo total de voo.

Embora existam vários estudos que se propõem a analisar o problema dos atrasos em voo por meio de *Machine Learning*, a maioria não aborda o uso de técnicas de pré-processamento dos dados para a melhor eficiência de seus algoritmos, o que pode tender a encontrar resultados tendenciosos para o conjunto de dados majoritários. Diante disso, observa-se a oportunidade de pesquisa ao combinar técnicas de balanceamento de classes com modelos de classificação supervisionada, como Regressão Logística, *Decision Tree* e *Random Forest*.

III. MÉTODO

O método implementado neste estudo utilizou como base três etapas: 1) recebimento da base de dados; 2) processamento e engenharia de atributos (*feature engineering*); e 3) modelagem e treinamento. Cada etapa será descrita a seguir.

1) Base de Dados: Este estudo utilizou como fonte de dados o banco de informações disponibilizado pela ANAC, referente aos voos domésticos realizados no Brasil nos anos de 2023 e 2024. A base inicial, após a retirada dos valores ausentes (NA), apresentou 183.883 registros com variáveis de Tempo de voo, Dia da semana, Rota, Tipo de aeronave, dentre outras.

2) Processamento e Engenharia de Atributos: Após a coleta e limpeza inicial, foi necessário transformar os dados brutos em variáveis informativas adequadas para o uso nos modelos preditivos. Com base em todas as variáveis disponíveis, iniciou-se uma análise exploratória dos dados, com o objetivo de analisar quais variáveis seriam relevantes e com maior correlação aos atrasos de voo. Optou-se por analisar aquelas que representam características operacionais ou do planejamento do voo, ou seja, as variáveis : Tempo_Voo_Prev_min, Dia_Semana, Rota, Periodo_Dia.

Posteriormente, foram categorizados os atrasos em três faixas: ”Pontual- zero a quinze minutos, ”Atraso Leve- dezesseis a trinta e cinco minutos; e ”Atraso Moderado- trinta e seis a sessenta minutos. Também foram desconsiderados os outliers dos tempos de atrasos superiores a sessenta minutos e os voos adiantados (atraso negativo). Essa abordagem foi adotada ao analisar que atrasos acima de uma hora compreendem apenas 67 ocorrências, ou seja, 0,03644% do banco de dados , representando apenas situações excepcionais, fora do padrão operacional analisado.

Por fim, foi criada a categorização dos dados brutos dos períodos do dia em madrugada, manhã, tarde e noite, e ajustados os dias da semana e as variáveis categóricas convertidas em fatores e *dummies* possibilitando a aplicação dessas informações nos algoritmos de aprendizagem de máquina.

3) Modelagem e Treinamento: Para a modelagem dos algoritmos, dividiu-se o banco de dados em dois subconjuntos utilizando 70% das informações para treinamento e 30% para teste, garantindo que, posteriormente, fosse possível avaliar o desempenho final do teste com 30% dos dados nunca usados. Essa estratégia visa obter uma estimativa não enviesada do desempenho do modelo em dados inéditos, além de permitir a comparação entre diferentes métodos.

Ao analisar integralmente os atrasos categorizados, notou-se um considerável desbalanceamento entre as categorias classificadas, conforme a tabela abaixo:

TABELA I
DISTRIBUIÇÃO DAS CLASSES DE ATRASO

Categoria	Contagem absoluta	Proporção (%)
Pontual	62.006	89%
Atraso Leve	6.922	9,8%
Atraso Moderado	683	1,2%

Dessa forma, observou-se a possibilidade de aplicação de técnicas para o balanceamento das classes antes do uso dos algoritmos, uma vez que, caso os dados sejam utilizados sem o balanceamento, os modelos podem tender a priorizar a classe majoritária dos dados, apresentando um desempenho insatisfatório na identificação dos atrasos leves e moderados. Segundo Gameng, Gerardo e Medina (2019) [11], o balanceamento de classes pode ser realizado por meio de técnicas de sobreamostragem e subamostragem aleatória. No entanto, a subamostragem apresenta como desvantagem o risco de

descartar informações potencialmente relevantes, que podem ser cruciais para o processo de aprendizagem.

Nesse sentido, optou-se pelo uso da sobreamostragem proposta por Chawla (2002) [7] com o uso das amostras da classe minoritária e a introdução de exemplos sintéticos ao longo dos segmentos de reta que unem os cinco vizinhos mais próximos da classe. Dependendo da quantidade de sobreamostragem necessária, os cinco vizinhos mais próximos são escolhidos aleatoriamente. Os exemplos sintéticos fazem com que o classificador crie regiões de decisão maiores e menos específicas, em vez de regiões menores e mais específicas.

Dessa forma, foi possível ampliar o número total de exemplos das classes de voos com atrasos leves e moderados a partir da interpolação de pontos de dados vizinhos, replicando as amostras selecionadas das classes minoritárias [5]. Importante ressaltar que, nessa etapa, embora tenha ocorrido um aumento na representatividade das classes no conjunto de treinamento (após a aplicação da SMOTE), foram mantidas as proporções originais no conjunto de teste. Essa abordagem assegura uma avaliação imparcial e rigorosa do desempenho do modelo preditivo.

4) Modelos Utilizados: A Regressão Logística Ordinal foi aplicada para explicar respostas ordenadas dentro da região do escopo do modelo, de maneira interpolada, evitando a generalização do modelo para fora da região investigada. Enquanto a *Decision Tree* foi aplicada pela sua capacidade de capturar padrões não lineares e interações entre as variáveis, que podem ser categóricas e numéricas, criando ramificações, com nós de decisão, ramos e folhas até obter um critério de parada em que cada folha representa uma previsão final. O modelo de aprendizado de *Random Forest* pertence à classe de métodos ensemble, combinando inúmeras árvores de decisão independentes com valores de hiperparâmetros (*tuning*) comumente usados na literatura, visando equilibrar o desempenho e a eficiência computacional, considerando o tempo de processamento e os recursos disponíveis. Como vantagem, ao combinar diversas árvores, o modelo consegue analisar a importância de cada variável para a previsão, além de possibilitar a redução do *overfitting*, gerando previsões que tendem a ser mais estáveis, maximizando o desempenho preditivo e reduzindo a variância, evitando o custo computacional elevado associado à validação cruzada em grades extensas de parâmetros.

Para possibilitar a comparação entre os modelos, foi adotada a estratégia de validação cruzada *10-fold* dentro do conjunto de dados de treinamento, conforme práticas recomendadas por James et al. (2014) [12]. Esse método consiste em dividir os dados em dez partes iguais chamadas de *folds*, a cada rodada nove *folds* são utilizados para o treino do modelo e um *fold* é utilizado para a validação interna do modelo treinado. Ao final da realização de dez rodadas, é possível calcular a média das métricas de desempenho e obter uma avaliação justa de cada algoritmo.

Após a etapa de treinamento e da validação interna, são utilizados os 30% dos dados do conjunto teste, que foram preservados durante todo o processo, para garantir uma nova estimativa imparcial e realista do desempenho de cada método.

IV. Resultados

1) Modelos Avaliados: Neste trabalho, foram utilizados os algoritmos de Regressão Logística Ordinal, *(Decision Tree)* e *Random Forest*. No uso de cada modelo, foi utilizada uma versão sem SMOTE (V1) e outra com SMOTE (V2). Tal alternância teve como objetivo prever a categoria de atraso em voo e, posteriormente, comparar os resultados dos métodos mencionados com e sem o balanceamento via SMOTE.

2) Métricas Globais dos Modelos: Após realizar os aprendizados de máquina, obtiveram-se as métricas *Accuracy*, Kappa e F1 Macro de cada modelo. A Tabela 2 representa um comparativo desses valores.

TABELA II

Métricas de desempenho dos modelos

Modelo	**SMOTE**	***Accuracy***	**Kappa**	**F1 (Macro)**
Regressão Logística Ordinal V1	Não	0,888	0	0,941
Decision Tree V1	Não	0,888	0	1,000
Random Forest V1	Não	0,888	0	0,941
Regressão Logística Ordinal V2	Sim	0,554	0,027	0,290
Decision Tree V2	Sim	0,564	0,027	0,387
Random Forest V2	Sim	0,592	0,052	0,322

Para facilitar a visualização e o contraste entre os modelos, criou-se o gráfico comparativo de métricas por modelo, Fig. 1.

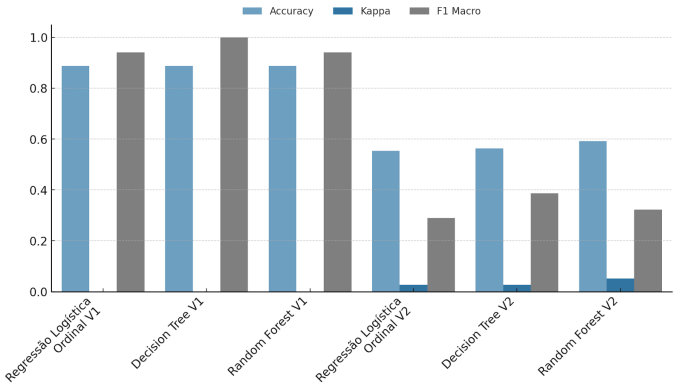

Fig. 1. Gráfico comparativo de métricas por modelo (com e sem SMOTE).

As medidas de acurácia, Kappa e F1 Macro são comumente usadas em avaliações de modelos de *Machine Learning*, representando o desempenho do modelo como um todo.

Accuracy é a proporção de previsões corretas (tanto positivas quanto negativas) em relação ao total de previsões feitas. É muito útil em previsões com dados balanceados, mas pode ser enganosa caso contrário.

O índice Kappa mede o quanto o modelo está melhor do que o acaso, considerando a chance de acertos aleatórios. E F1 Macro é o desempenho médio entre as classes, independentemente da frequência com que ocorrem.

O valor de F1 Macro para *Decision Tree* V2 é aproximado porque o modelo não previu nenhuma ocorrência da classe "Atraso Leve", ou seja, essa medida foi calculada em função apenas das classes que o modelo realmente acertou.

3) Desempenho por Classe: Para explorar mais a fundo a importância de cada classe no *Machine Learning*, estruturou-se a Tabela 3 com as medidas de Precisão, *Recall* e F1 em cada modelo por classe.

TABELA III

DESEMPENHO DOS MODELOS POR CLASSE

Modelo	Métrica	Pontual	Atraso Leve	Atraso Moderado
Regressão Logística Ordinal V1	Precisão	1	0	0
	Recall	1	0	0
	F1	1	–	–
Decision Tree V1	Precisão	1	0	0
	Recall	1	0	0
	F1	1	–	–
Random Forest V1	Precisão	1	0	0
	Recall	1	0	0
	F1	1	–	–
Regressão Logística Ordinal V2	Precisão	0,333	0,333	0,333
	Recall	0,604	0,130	0,417
	F1	0,753	0,230	0,588
Decision Tree V2	Precisão	0,905	–	0,016
	Recall	0,628	0	0,592
	F1	∼0,742	–	∼0,032
Random Forest V2	Precisão	0,333	0,333	0,333
	Recall	0,929	0,573	0,408
	F1	0,963	0,728	0,580

Da análise da Tabela 3, obteve-se o gráfico da Fig. 2, para analisar o F1 por classe e modelo.

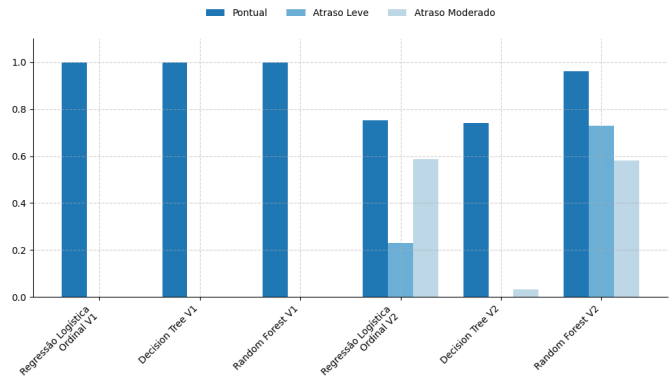

Fig. 2. Gráfico comparativo de F1 por classe e modelo.

A precisão por classes mede o quão confiável é o modelo quando ele prevê determinada classe. Já o *Recall* por classe mede quanto daquela classe foi identificado corretamente. E F1 *Score* por classe balanceia a precisão e o recall de uma classe.

Os valores NaN da Tabela 3 indicam que a métrica não pôde ser calculada porque o modelo não fez nenhuma previsão naquela classe. Enquanto que os valores 0 mostram que houve previsão, mas o modelo errou todas. Nos casos em que uma das classes teve resultado NaN, o F1 Macro foi calculado como uma média entre as demais categorias, desconsiderando o valor NaN, por isso a indicação de aproximação na Tabela.

4) Matriz de Confusão: Outra ferramenta utilizada para o exame dos modelos foram as matrizes de confusão. Elas mostram como o modelo classificou corretamente ou incorretamente cada classe, comparando as previsões do modelo com os valores reais. Nesta seção são apresentadas em formato de *Heatmap* nas Figuras 3, 4 e 5.

Em uma matriz de confusão, as linhas representam as previsões do modelo, enquanto as colunas são os valores reais.

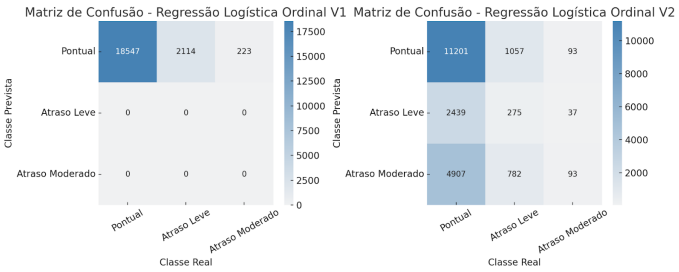

Fig. 3. Matriz de confusão de Regressão Logística Ordinal (V1 e V2).

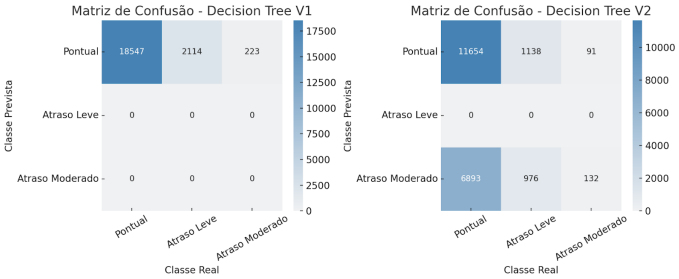

Fig. 4. Matriz de confusão de *Decision Tree* (V1 e V2).

O ideal é ver números altos na diagonal principal (acertos) e baixos fora dela (erros).

A comparação dos gráficos mostra que os modelos sem SMOTE preveem apenas a classe “Pontual”, gerando matrizes de confusão completamente vazias nas linhas “Atraso Leve” e “Atraso Moderado”. Reafirmando que os valores elevados de *Accuracy* desses modelos têm utilidade prática nula.

Além disso, os modelos com SMOTE são os únicos que conseguem prever atrasos em todas as classes. Mesmo com alguns erros, estes algoritmos apresentam presença significativa de acertos em ”Atraso Leve”e ”Atraso Moderado”. Por mostrarem capacidade preditiva nas classes minoritárias, são os únicos que podem ser considerados úteis para a tomada de decisão em atrasos.

Analisando os valores das matrizes de confusão em percentis, temos o gráfico da Fig. 6.

A inspeção da Fig. 6, demonstra que o *Random Forest* V2 se qualificou como o modelo mais eficaz, com acertos de aproximadamente 58% de ”Pontual”, 26% de ”Atraso Leve”e 2% de ”Atraso Moderado”. Em seguida, a Regressão Logística e *Decision Tree*, ambas em suas versões balanceadas, apresentaram desempenho moderado na classe ”Pontual”e mais fraco nas demais categorias.

5) Importância das Variáveis (Random Forest): Outra informação importante que se pode extrair especificamente

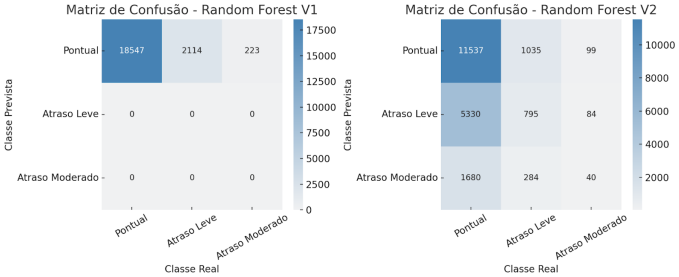

Fig. 5. Matriz de confusão de *Random Forest* (V1 e V2).

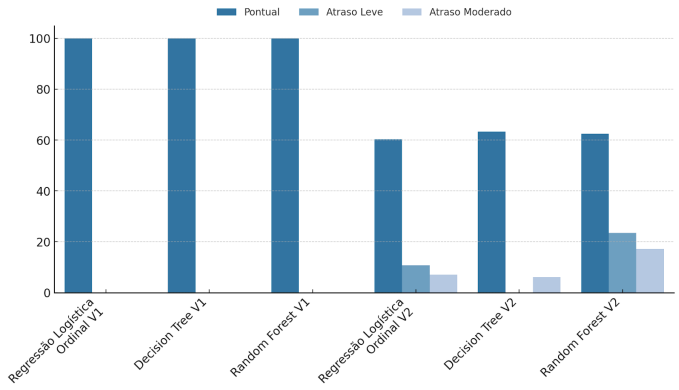

Fig. 6. Gráfico de percentual de acertos por classe e modelo.

dos modelos de *Random Forest*, é o gráfico de importância das variáveis, Fig. 7, que representa quais características (*features*) do conjunto de dados foram mais relevantes para o modelo tomar decisões.

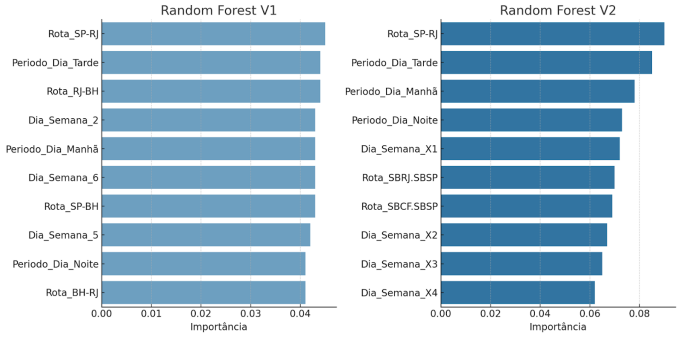

Fig. 7. Importância das variáveis no modelo de *Random Forest* (V1 e V2).

A comparação entre os modelos *Random Forest* com e sem SMOTE evidencia de forma clara os impactos do balanceamento de classes não apenas na performance, mas também na interpretabilidade do modelo. Enquanto o modelo balanceado apresentou uma hierarquia bem definida de variáveis mais relevantes, o modelo sem balanceamento das classes demonstrou uma distribuição achatada e pouco informativa na importância das variáveis.

A distribuição quase uniforme do modelo sem SMOTE indica que o modelo só aprendeu a prever a classe ”Pontual”, ignorando as demais categorias. Como consequência, o modelo não precisa de informação contextual — qualquer variável serve, já que todas as instâncias são da mesma classe.

Já no modelo V2, a hierarquia clara de variáveis relevantes indica que o modelo conseguiu aprender padrões específicos associados a diferentes tipos de atraso. Essa diversidade de variáveis importantes mostra que o modelo analisou rotas, horários e dias da semana para diferenciar as classes.

V. Análise dos Resultados

Neste trabalho, utilizamos os métodos de Regressão Logística Ordinal, *Decision Tree* e *Random Forest*. Como visto no tópico anterior sobre os resultados, em todos eles tivemos que balancear os dados para o devido treinamento do algoritmo no intuito de diminuir o erro de viés do mesmo. Abaixo analisaremos os impactos da técnica nestes três algoritmos em perspectivas diferentes.

1) Impacto da SMOTE na Performance Global: Pela Tabela 2 e Fig. 1, podemos observar que os modelos sem SMOTE têm *Accuracy* e F1 Macro elevados, mas Kappa = 0; enquanto os modelos que sofreram balanceamento no treinamento têm *Accuracy* menor, mas Kappa >0 e F1 Macro mais realista, refletindo um desempenho mais balanceado entre as classes.

Os valores elevados de F1 Macro nos modelos sem SMOTE geram interpretações dúbias à primeira vista, uma vez que representam o *score* perfeito na classe majoritária e ausência total de acertos nas demais, como será discutido mais adiante.

2) Impacto da Smote nas Classes Minoritárias: Pelo estudo da Tabela 3 e Fig. 2, vemos que todos os modelos sem SMOTE têm precisão e *recall* igual a 1, indicando que o modelo só acerta quando a classe é majoritária (”Pontual”), mas nas demais classes *Recall* = 0 e F1 = NaN, indicando que essas classes nunca foram corretamente previstas.

Observando os modelos com SMOTE, vemos que todas as classes são reconhecidas em algum grau. Nesses casos, a métrica *Recall* nas classes minoritárias deixa de ser 0. E o modelo *Random Forest* V2 se destaca com F1 razoável nas três classes e melhor equilíbrio geral entre precisão e recall.

Por isso constatamos que os modelos sem SMOTE só têm F1 alto na classe Pontual. Enquanto que os modelos com SMOTE apresentam desempenho mais balanceado, com F1 considerável também nas classes “Atraso Leve” e “Atraso Moderado”. Isso indica que as versões sem balanceamento das classes sofreram overfitting no padrão da majoritária, com total negligência às minoritárias.

3) Aprendizado das Regras de Decisão: Nesse tópico estuda-se a forma como os modelos aprenderam os padrões úteis dos dados. No modelo de Regressão Logística Ordinal sem o balanceamento das classes houve aprendizado apenas da classe majoritária e as probabilidades acumuladas entre categorias ficaram concentradas, ignorando o ordenamento das classes. Demonstrando alto enviesamento. Já o balanceamento das classes permitiu gerar probabilidades distintas entre todas as classes, demonstrando um aprendizado mais equilibrado entre as classes.

Com base nos resultados da aplicação da *Decision Tree* com o uso da técnica SMOTE, a primeira impressão foi que houve uma queda geral da acurácia. Porém, o modelo com SMOTE foi capaz de diferenciar alguns atrasos moderados do restante das classes, mas ainda não conseguiu diferenciar a classe ”Atraso Leve”das demais classes. Grande parte disso se deve à própria similaridade da classe “Pontual” e “Atraso Moderado” e também às limitações do modelo de *Decision Tree*.

Por fim, percebe-se que no modelo de *Random Forest* sem SMOTE o modelo inteiro converge para a classe majoritária e, pela análise da Fig. 7, a importância das variáveis ficou achatada, sem destaque para nenhuma *feature*. Em contrapartida, modelo com SMOTE houve um melhor aprendizado, pois conseguiu separar as classes com base em múltiplos fatores, como visível na hierarquia clara entre as variáveis relevantes.

4) Comparação entre Algoritmos: Dos 3 algoritmos testados, o *Random Forest* demonstrou um desempenho geral melhor após o uso da SMOTE, porque apresentou maior quantidade de acertos totais e por classe, maior F1 Macro entre os modelos com SMOTE e capacidade de aprendizado profundo.

A *Decision Tree* mostrou comportamento instável mesmo após o balanceamento das classes, pois mesmo com a aplicação da SMOTE ainda não conseguiu prever a classe "Atraso Leve". Além disso, demonstrou ser mais propensa a overfitting de padrões artificiais da SMOTE.

Já o Modelo de Regressão Logística apresentou um maior equilíbrio entre sensibilidade e especificidade, devidamente observável no índice de *Recall* e nas matrizes de Confusão. A sua versão sem SMOTE falhou da mesma forma que os demais, mas o balanceamento melhorou significativamente a distribuição de acertos entre as classes.

Conclui-se que o uso da SMOTE como técnica de balanceamento foi crucial para permitir que os algoritmos aprendessem a distinguir entre todas as categorias.

VI. Conclusões

O objetivo principal deste trabalho foi analisar o impacto da técnica de balanceamento SMOTE na eficácia de modelos de *machine learning* para a previsão de atrasos em voos domésticos. O estudo, conduzido com variáveis operacionais como rota e período do dia, demonstrou que a principal contribuição não reside na performance isolada dos algoritmos, mas na comprovação de que o pré-processamento com SMOTE é um passo crucial. A técnica foi de fundamental importância para viabilizar o aprendizado das classes minoritárias, como os atrasos leves e moderados, permitindo o reconhecimento de padrões relevantes.

Os resultados obtidos demonstram que, sem a aplicação da SMOTE, os três modelos testados tiveram dificuldade em prever, de fato, as ocorrências de atrasos, focando as previsões nas classes majoritárias. Com o balanceamento de classes, foi possível observar uma melhora significativa dos resultados, principalmente nas métricas de F1 para as classes minoritárias, especialmente no modelo de *Random Forest*, que teve o desempenho mais robusto e consistente entre os algoritmos testados. Tal fato reforça a importância das estratégias envolvendo o pré-processamento dos dados, num contexto de análise de classes desbalanceadas, sendo possível a ampliação da capacidade e da utilidade prática dos modelos.

Com base nas contribuições alcançadas, o presente estudo abre espaço para trabalhos futuros que possam explorar outras variáveis, como condições climáticas e a intensidade de tráfego nas aerovias , como forma de ampliar a fonte de dados. Uma opção sugerida seria a utilização de modelos mais avançados, que sejam capazes de identificar padrões, mesmo que sutis, nas observações, objetivando uma aplicação ampla e dinâmica no contexto prático.

Referências

[1] Agência Nacional de Aviação Civil (ANAC), "Com 118 milhões de passageiros transportados em 2024, setor aéreo tem segundo melhor desempenho da história," Agência Nacional de Aviação Civil, 1 2025, disponível em: https://www.gov.br/anac/pt-br/noticias/2025/com-118-milhoes-de-passageiros-transportados-em-2024-setor-aereo-tem-segundo-melhor-desempenhos-da-historia. Acesso em: 29/04/2025.

[2] G. Benevides, "Azul Cargo projeta alta de até 30% em 2025 com novos cargueiros," AERO Magazine, 4 2025, disponível em: https://aeromagazine.uol.com.br/artigo/azul-cargo-projeta-alta-de-ate-30-em-2025-com-novos-cargueiros.html. Acesso em: 29/04/2025.

[3] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *Proceedings of the IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. Sacramento, CA, USA: IEEE, sep 2016, pp. 1–6.

[4] A. Sternberg *et al.*, "An analysis of Brazilian flight delays based on frequent patterns," *Transportation Research Part E: Logistics and Transportation Review*, vol. 95, pp. 282–298, 2016.

[5] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[6] T. D. Barros, T. G. Ramos, J. C. C. B. S. Mello, and L. A. Meza, "Avaliação dos atrasos em transporte aéreo com um modelo DEA," *Produção*, vol. 20, pp. 601–611, 2010.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[8] G. G. Ditzel, "Técnicas de mineração de dados aplicadas à predição de atraso em voos comerciais," Trabalho de Conclusão de Curso, Universidade Federal de Santa Catarina, Florianópolis, 2022.

[9] Y. Ge, D. Yue, and L. Chen, "Prediction of wind turbine blades icing based on MBK-SMOTE and Random Forest in imbalanced data set," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. Beijing, China: IEEE, 2017, pp. 1–6.

[10] M. Grellert, "Machine learning mode decision for complexity reduction and scaling in video applications," Ph.D. dissertation, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

[11] H. A. Gameng, B. B. Gerardo, and R. P. Medina, "Modified Adaptive Synthetic SMOTE to Improve Classification Performance in Imbalanced Datasets," in *Proceedings of the 6th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)*. Kuala Lumpur: IEEE, 2019.

[12] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. New York: Springer, 2014.