



UnB



pra fazer
a diferença

Meta A05 - Relatório Técnico de Aprendizado de Máquina

Diagnóstico sobre a Judicialização do Transporte Aéreo no Brasil: Uma aplicação de Aprendizado de Máquina.



Brasília, Novembro de 2024.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Identificação do Projeto	1
1.2	Justificativa do Projeto	1
1.3	Metas	2
1.4	Valor entregável	3
1.5	Estrutura do Relatório	3
2	APLICAÇÃO DO APRENDIZADO DE MÁQUINA NAS QUESTÕES DE JUDICIALIZAÇÃO DO TRANSPORTE AÉREO	5
3	METODOLOGIA	7
3.1	Bases de dados	7
3.2	Pré-processamento	7
3.2.1	Limpeza e organização dos dados	8
3.2.2	Preparação dos dados para os modelos	10
3.3	Algoritmos de aprendizado de máquina	12
3.3.1	<i>Naive Bayes Classifier</i>	12
3.3.2	<i>Support Vector Machine</i>	13
3.3.3	<i>Random Forest Classifier</i>	13
3.4	Avaliação das métricas de desempenho dos modelos	14
3.4.1	Importância das variáveis	15
3.5	Códigos da preparação dos dados e dos modelos	16

4	RESULTADOS	17
4.1	Resultados dos modelos para previsão do valor do dano moral .	17
4.1.1	Matrizes de confusão	18
4.1.2	Importância das variáveis	20
4.1.3	Discussão dos Resultados	25
5	CONSIDERAÇÕES FINAIS	27
	REFERÊNCIAS	28

1 INTRODUÇÃO

1.1 Identificação do Projeto

Nome do projeto: Diagnóstico sobre a Judicialização do Transporte Aéreo no Brasil: Uma aplicação de Aprendizado de Máquina.

Objetivo do projeto: Identificar e avaliar fatores influentes, motivações, causas e situações processuais das demandas judiciais contra as companhias aéreas, realizando um diagnóstico por meio de aprendizado de máquina. Como objetivos específicos, destacam-se:

- Obter dados acessíveis dos processos civis nos juizados especiais, tribunais de justiça estaduais e federais brasileiros contra companhias aéreas que operem no Brasil, nos anos de 2019 a 2022;
- Obter índices e dados internacionais acessíveis e construir uma base de dados a respeito da eficiência/qualidade dos serviços das companhias aéreas *versus* índice de judicialização;
- Realizar o pré-processamento dos dados, em relação a fatores influentes, motivações, causas e situações processuais, visando a aplicação de aprendizado de máquina;
- Aplicar aprendizado de máquina explicável para identificar e avaliar as características de interesse;
- Aplicar a inteligência artificial generativa em aprendizado ativo para avaliar qualitativamente uma pequena amostra de conveniência de processos judiciais relacionados;
- Elaborar um diagnóstico sobre a judicialização do transporte aéreo no Brasil, com base na aplicação de aprendizado de máquina.

Período de duração do projeto: Dezembro de 2023 a Dezembro de 2024

1.2 Justificativa do Projeto

A presente proposta de pesquisa objetiva contribuir para o desenvolvimento econômico e social no Brasil e para a formulação de políticas públicas adequadas a todas

os *stakeholders* envolvidos na prestação de serviços de transporte aéreo. A proposta de pesquisa objetiva identificar os fatores que influenciam o elevado índice de demandas judiciais no Brasil, levando-se em consideração o atual sistema regulatório.

O número de litígios com transportadoras no setor aéreo é bastante expressivo no Brasil. A Associação Brasileira das Empresas Aéreas - ABEAR apurou que 98,5% das ações cíveis no mundo de passageiros contra as companhias aéreas estão concentradas no Brasil. Esse percentual alarmante consta expressamente da Cartilha do Transporte Aéreo (FUX et al., 2021), publicada em 25 de maio de 2021, pelo Conselho Nacional de Justiça. Neste sentido, este projeto de pesquisa visa responder às seguintes questões:

1. Quais são os principais fatores influentes que resultam no alto índice de demandas judiciais contra as companhias aéreas no Brasil?
2. Quais as principais motivações e causas dos litígios ajuizados?
3. Quais as situações processuais resultantes da solução dos litígios ajuizados no judiciário brasileiro?
4. Como aplicar aprendizado de máquina explicável para responder as questões 1 a 3 anteriores?

Esta pesquisa contará com dados acessíveis, públicos e privados, a nível nacional e internacional, para elaborar um diagnóstico sobre a judicialização no Brasil, elucidando as aplicações das normas nos tribunais brasileiros.

1.3 Metas

Esse trabalho apresenta sete metas definidas conforme Tabela 1.

Tabela 1 – Quadro de metas

Meta 1	Relatório Estrutura Analítica do Projeto (EAP)
Meta 2	Relatório Técnico de Especificação Detalhada do Projeto
Meta 3	Minuta de artigo científico de Revisão da Literatura
Meta 4	Relatório Técnico de Bases de Dados
Meta 5	Relatório Técnico de Aprendizado de Máquina
Meta 6	Relatório Técnico de Avaliação Qualitativa por Amostragem
Meta 7	Relatório Técnico Final do Projeto, com diagnóstico da judicialização do transporte aéreo

O presente relatório refere-se a **Meta A05 - Relatório Técnico de Aprendizado de Máquina**. Esta versão visa dar um retorno às aéreas e à empresa responsável pela raspagem da jurimetria, identificando necessidades de revisão dos dados fornecidos.

1.4 Valor entregável

As principais contribuições do referido projeto podem ser entendidas como:

- Científicas: Gerar conhecimento científico aplicado ao Transporte Aéreo e à Inteligência Artificial por meio de artigos científicos, TCCs de graduação, projetos de mestrado e doutorado.
- Tecnológicas: Gerar conhecimento tecnológico, por meio da pesquisa, desenvolvimento e inovação. Haverá o desenvolvimento do ciclo completo de uma pesquisa aplicada, com a implementação de melhores práticas e tecnologias, incluindo inteligência artificial, para a análise diagnóstica da judicialização do transporte aéreo, com aplicação de aprendizado de máquina.
- Graduação e Pós-graduação da UnB: contribuir para a constituição de um *Think Tank* para formação na análise diagnóstica da Administração Pública, por meio da aplicação da inteligência artificial. Gerar oportunidade para participação e desenvolvimento de docentes e discentes (graduação e pós-graduação). O projeto de pesquisa permitirá que professores que ainda não fazem parte de programas de pós-graduação se habilitem a participar, atuando como coordenadores e/ou integrantes de projetos de pesquisa com financiamento externo.
- Social: contribuir para a boa gestão do transporte aéreo, com aumento da eficiência e da eficácia, contribuindo para a melhoria da qualidade dos serviços públicos.

1.5 Estrutura do Relatório

Este relatório está organizado da seguinte forma: a Aplicação do Aprendizado de Máquina nas Questões de Judicialização do Transporte Aéreo apresenta alguns trabalhos anteriores sobre o tema; a Metodologia apresenta informações a respeito da base de dados utilizada, sobre o pré-processamento para preparar os dados para os modelos e os

algoritmos de aprendizado de máquina utilizados; os Resultados apresenta um comparativo dos resultados obtidos por cada um dos modelos; por fim, as Considerações Finais traz uma conclusão dos resultados obtidos.

2 APLICAÇÃO DO APRENDIZADO DE MÁQUINA NAS QUESTÕES DE JUDICIALIZAÇÃO DO TRANSPORTE AÉREO

O aprendizado de máquina (*machine learning*, em inglês) é um subcampo da inteligência artificial que estuda a capacidade de resolver problemas por meio de grandes volumes de dados. Seu objetivo é treinar algoritmos para identificar regras ou parâmetros que estabeleçam uma relação entre os dados de entrada (chamados de atributos preditivos) e os dados de saída (conhecidos como atributos alvo). Isso viabiliza a execução de diversas tarefas, como classificação, previsão e agrupamento de dados [Lenz et al. 2020].

Alguns trabalhos anteriores aplicaram aprendizado de máquina em questões de judicialização e de transporte aéreo. Um problema comum é a dificuldade de agrupar os processos judiciais devido às diversas classificações de motivos e causas utilizados. Além disso, os documentos legais frequentemente estão associados a múltiplos assuntos e possuem uma semântica complexa, como constatado por De Martino, Pio e Ceci 2023. Nesse sentido, foram pesquisadas abordagens de agrupamento ou *clustering*, com o objetivo de agrupar n indivíduos em grupos homogêneos (*clusters*), a fim de possibilitar uma análise com mais clareza do conjunto de dados [Sicsú, Samartini e Barth 2023]. Raghuvver et al. 2012 utilizaram *Latent Dirichlet Allocation* (LDA) para propor um agrupamento com base em tópicos. Raghav et al. 2015 aplicaram K-means para testar três agrupamentos utilizando citações e links sobre outros julgamentos. Zhang e Zhou 2019 usaram um algoritmo de vetor de parágrafo (Doc2Vec) para atualizar os resultados do agrupamento de documentos jurídicos, comparando a similaridade do texto sem reimplementar o agrupamento.

Alguns trabalhos também estudaram o tema considerando o contexto brasileiro. Sabo et al. 2022 buscaram agrupar julgamentos judiciais do Juizado Especial Cível localizado na Universidade Federal de Santa Catarina (JEC/UFSC), nas quais consumidores reivindicavam indenizações morais e materiais às companhias aéreas por falhas no serviço. Os autores aplicaram quatro algoritmos de *clustering*: hierárquico e lingo (*clustering* suave), *K-means* e propagação de afinidade (*clustering* rígido). Como resultado, os autores identificaram que o *clustering* hierárquico demonstrou ser a abordagem mais vantajosa, porém a tarefa de fornecer os rótulos cabia exclusivamente ao especialista. Por outro lado, o algoritmo lingo demonstrou a capacidade de gerar rótulos, permitindo ao especialista apenas identificar aqueles que não deveriam ser uma variável do julgamento. De forma geral, os autores concluíram que todas as abordagens testadas permitiram identificar padrões, porém com o apoio de especialistas jurídicos, demonstrando uma limitação das

técnicas em explicar um evento jurídico.

Lima e Costa 2022 realizaram uma avaliação de seis diferentes abordagens para o agrupamento, também utilizando bases de dados de documentos jurídicos brasileiros. Eles testaram os algoritmos *K-means*, *Mini Batch K-Means* e HDBSCAN em diferentes hiperparâmetros e em conjunto ou não com o Mapa de Kohonen como técnica de pré-clustering. Além disso, o trabalho também propõe uma nova estrutura orientada para o Processamento de Linguagem Natural (PLN) para a avaliação específica de clusters textuais. Os resultados obtidos demonstraram que o *K-means* e o *Mini Batch K-Means* são as melhores escolhas, e o uso do mapa de Kohonen pode aumentar o desempenho geral do *clustering*.

Por outro lado, Torres, Guterres e Celestino 2023 compararam o desempenho de três modelos de aprendizado de máquina (*Random Forest*, *SVM* e *Naive Bayes*) e da Regressão Logística Multinomial a fim de identificar as melhores técnicas para prever os valores de indenização final pago ao consumidor. O estudo foi baseado em dados de ações judiciais ajuizadas em cidades brasileiras, no período de 2016 a 2021. Entre os modelos analisados, o *Random Forest* teve o melhor desempenho em relação aos modelos de aprendizagem de máquina, com valores semelhantes à Regressão Logística Multinomial, que se mostrou importante na classificação de conjuntos de dados categóricos. Contudo, apesar da melhor precisão, o tempo de processamento do *Random Forest* foi muito elevado, o que pode ser uma desvantagem se o poder computacional for limitado. Assim, a Regressão Logística Multinomial apresentou a melhor relação custo-benefício.

3 METODOLOGIA

Este trabalho reaplicou os modelos de aprendizado de máquina desenvolvidos por Torres, Guterres e Celestino 2023, porém com algumas diferenças. A primeira foi a utilização da técnica detalhada no relatório da **Meta A04 - Relatório Técnico de Bases de Dados** para padronização dos motivos dos processos judiciais (objeto e subobjeto principal) e das causas informadas pelas companhias aéreas. O objetivo era superar a limitação apresentada por Torres, Guterres e Celestino 2023. Segundo eles, a diferença na classificação dos motivos e das causas dos processos judiciais por cada companhia aérea pode ter influenciado no desempenho dos algoritmos, visto que eventos semelhantes podem ter sido interpretados de forma diferente na criação do banco de dados.

A segunda diferença foi a mudança da variável-alvo (*target*), isto é, aquela que se deseja prever com os modelos. No trabalho de Torres, Guterres e Celestino 2023 a variável-alvo era valor pago e nesse foi valor do dano moral, visto que o objetivo era prever os processos judiciais com altos valores de danos morais.

3.1 Bases de dados

A base de dados utilizada foi obtida após todos os tratamentos detalhados no relatório da **Meta A04 - Relatório Técnico de Bases de Dados**. A base continha as informações de 126.041 processos judiciais movidos contra as seis empresas aéreas (Azul, Copa, Gol, Latam, Tap e United), no período de 2019 a 2023, considerando apenas 16 estados brasileiros (AM, BA, CE, DF, ES, GO, MG, MS, MT, PE, PR, RJ, RO, SC, SP e TO) e que tiveram como resultado a Condenação.

3.2 Pré-processamento

Nessa etapa, foi realizado um pré-processamento para preparar os dados para a aplicação dos modelos. Primeiramente, foi feita a limpeza e a organização dos dados, que consistiu na seleção e preparação das variáveis. Em seguida, os dados foram preparados para os modelos: as variáveis categóricas foram tratadas, os dados foram divididos em dados de entrada (X) e dados de saída (y) e em dados de treino e teste.

3.2.1 Limpeza e organização dos dados

Inicialmente, as variáveis foram selecionadas, sendo elas:

- **Companhia aérea:** por fornecer entendimento sobre a frequência de reclamações que cada empresa recebe dos passageiros, o que pode influenciar a decisão final dependendo das reinvindicações;
- **Ano:** por permitir que as companhias aéreas entendam a proporção de ações judiciais ao longo do período e avaliem se o comportamento dos valores de indenização muda devido a algum fator externo, como a pandemia da COVID-19;
- **UF:** por ser uma variável essencial para compreender o comportamento das ações judiciais em todo o Brasil, onde há uma vasta extensão territorial com regiões tanto altamente desenvolvidas quanto subdesenvolvidas, assim conhecer as áreas com as maiores concentrações de reclamações auxilia as companhias aéreas a agirem de maneira eficaz nessas localidades;
- **Operação:** por indicar se o voo era nacional ou internacional;
- **Jurisdição:** por indicar a classificação do tribunal onde tramitou o processo;
- **Instância:** por indicar em qual instância o processo se encontra;
- **Situação processual e resultado da decisão (situacao_processual_setenca:** por mostrar o resultado do processo;
- **Número de autores:** por mostrar a quantidade de autores envolvidos no processo;
- **Valor do pedido inicial, valor do dano moral, e valor do dano material:** por fornecerem informações sobre o tipo de compensação que os passageiros solicitaram, seja com maior incidência de danos patrimoniais, que visam compensar perdas materiais, ou danos morais, que são mais proeminentes nas decisões judiciais;
- **Motivos e causas:** por fornecerem dados sobre os principais motivos das reclamações dos clientes e a origem dos problemas que eles mencionaram;

Com as variáveis selecionadas, algumas etapas foram realizadas para padronizar as categorias das variáveis. Em relação a UF, as informações das UFs foram reduzidas nas cinco regiões brasileiras (Centro-Oeste, Nordeste, Sul, Norte e Sudeste). Para valor do

pedido inicial, valor do dano moral e valor do dano material, os dados foram categorizados em três intervalos: Baixo, Médio e Alto. Para definir esses intervalos, foram retirados os valores acima de 1 milhão e os valores repetidos. Tendo como base os valores únicos de cada uma dessas variáveis, foram calculados os percentis 25 e 75 e definido os intervalos:

- Baixo: valores até o percentil 25;
- Médio: valores entre o percentil 25 e o 75;
- Alto: valores acima do percentil 75.

A Tabela 2 apresenta os valores dos intervalos para cada variável.

Tabela 2 – Intervalos das variáveis de indenização

Variável	Baixo	Médio	Alto
Valor do pedido inicial	$\leq 10372,59$	$>10372,59$ e $\leq 22526,68$	$>22526,68$
Valor do dano moral	$\leq 1489,05$	$>1489,05$ e $\leq 5953,00$	$>5953,00$
Valor do dano material	$\leq 447,95$	$>447,95$ e $\leq 2601,45$	$>2601,45$

Fonte: Elaborado pela autora.

A Tabela 3 apresenta um resumo final das variáveis utilizadas, suas definições e categorias.

Tabela 3 – Variáveis

Variável	Definição	Categorias
Companhia aérea	Companhias aéreas analisadas	Azul Copa Gol Latam Tap United
Ano	Ano do início do processo judicial	2019 2020 2021 2022 2023
Região	Região em que a ação foi iniciada	Centro-Oeste Nordeste Sul Norte Sudeste
Operação	Tipo de operação	Doméstico Internacional Outros
Jurisdição	Classificação do tribunal onde o processo tramitou	Juizado Especial Cível Vara Cível Outros
Instância	Instância da sentença do processo	1ª Instância 2ª Instância 3ª Instância Não informado
Situação processual e resultado da decisão	Julgamento do processo	Condenação
Número de autores	Quantidade de autores do processo	Não se aplica
Valor do pedido inicial	Valor inicial pedido pelo(s) autor(es)	Baixo Médio Alto
Valor do dano moral	Compensação por dano moral	Baixo Médio Alto
Valor do dano material	Compensação por dano material	Baixo Médio Alto
Motivo	Motivo apresentado pelo passageiro para entrar com uma ação judicial contra a companhia aérea	Problemas operacionais Bagagem Contrato Outros
Causa	Fator informado pela empresa que levou ao problema relatado pelo cliente	Aérea Cliente Terceiros Fortuitos Força maior Não informado Outros

Fonte: Elaborado pela autora.

3.2.2 Preparação dos dados para os modelos

Para preparar os dados para os modelos, primeiramente foi feita a divisão dos dados de entrada (X) e os dados de saída (y), excluindo a coluna de valor do dano moral

do conjunto de dados principal para formar o X, e mantendo-a no y. Em seguida, tendo em vista que o conjunto de dados era formado por variáveis categóricas, aplicou-se *one-hot encoding* nas variáveis do X e *label encoding* no y. O *one-hot encoding* cria uma nova coluna para cada valor único de uma coluna existente, indicando com 1 quando o valor é afirmativo e com 0 quando negativo, ele é recomendado quando se trata de variáveis nominais. Por outro lado, o *label encoding* designa um valor inteiro (0, 1, 2, ...) para cada valor único da coluna existente, sendo recomendado quando se trata de variáveis ordinais.

Por fim, a base de dados foi dividida duas partes, uma para treino e outra teste. Para isso, utilizou-se a função *train_test_split* do *scikit-learn*, com os seguintes parâmetros: `test_size=0.2`, `stratify=y`, `random_state=42`. Isso quer dizer que:

- **test_size=0.2:** 20% dos dados foram reservados para o conjunto de teste e 80% para o de treino.
- **stratify=y:** a divisão foi estratificada com base na variável y, ou seja, as proporções relativas de diferentes classes em y são mantidas nos conjuntos de treino e teste;
- **random_state=42:** garante a reprodutibilidade da divisão, ou seja, que a divisão será sempre a mesma se o código for executado várias vezes.

Ressalta-se que a base de dados utilizada é uma base desbalanceada, isto é, a quantidade das categorias baixo, médio e alto de valor de dano moral não são iguais, existindo uma menor incidência de uma classe (classe minoritária) em relação as outras (classes majoritárias). Essa distribuição desigual das classes pode levar a modelos enviesados e impactar a capacidade de generalização do modelo. Dessa forma, foi aplicada uma técnica de subamostragem (*undersampling*), que consiste em reduzir os exemplos da classe majoritária. Essa técnica foi aplicada apenas nos dados de treino do modelo, preservando a integridade dos dados de teste, a fim de possibilitar a avaliação de como o modelo se comportaria em situações reais refletidas pela base de dados desbalanceada. Ou seja, ajustar a distribuição das classes na base de teste artificialmente poderia levar a conclusões falsas e otimistas sobre o modelo.

Com base nisso, a técnica utilizada nos dados de treino foi o *RandomUnderSampler*. O parâmetro `sampling_strategy='auto'` define que a amostragem será feita de forma a reduzir as classes majoritárias até o tamanho da classe minoritária, enquanto o parâmetro `random_state=42` garante a reprodutibilidade do processo, fixando a semente do gerador de números aleatórios.

3.3 Algoritmos de aprendizado de máquina

Após a etapa de pré-processamento, os três modelos de aprendizado de máquina foram executados, utilizando os algoritmos e melhores hiperparâmetros identificados por Torres, Guterres e Celestino 2023, sendo eles:

- **Naive Bayes Classifier (NB):** algoritmo *MultinomialNB* com alpha igual a 8;
- **Support Vector Machine (SVM):** algoritmo *LinearSVC* com valor de penalidade (penalty) igual a 12, valor de perda (*loss*) igual a *hinge* e valor de regularização (C) igual a 2;
- **Random Forest Classifier (RF):** algoritmo *RandomForestClassifier* com o número de árvores (*n_estimators*) igual 1.000, medida de qualidade (*criterion*) igual a *gini*, profundidade máxima da árvore (*max_depth*) igual a 12 e número mínimo de divisões de um nó (*min_samples_split*) igual a 4.

Tendo em vista que os dados deste trabalho são multiclasse, foi necessária uma modificação interna nos algoritmos que permite dividir a classificação multiclasse em várias classificações binárias. Dessa forma, foi aplicada a abordagem *one vs one* ("ovo"), que treina um classificador binário para cada par de classes possíveis no conjunto de dados. Isso foi utilizado para obter a área sob a curva ROC de cada modelo.

3.3.1 Naive Bayes Classifier

O *Naive Bayes Classifier* (NB) é um algoritmo supervisionado fundamentado no raciocínio bayesiano. Ele é utilizado em tarefas de aprendizado em que cada instância x é descrita por um conjunto de atributos (Ex.: companhia aérea, ano, época, região, valor do dano moral e material etc.) e em que a classe alvo pode assumir qualquer valor de um conjunto V (Ex.: baixo, médio e alto). Dessa forma, o NB classifica uma instância de dados ao atribuir o valor mais provável de y para o alvo, considerando os valores de atributos [Mitchell e Mitchell 1997].

O algoritmo Naive Bayes opera em várias etapas para realizar essa classificação de instâncias. Primeiramente, calcula as probabilidades *a priori* de cada classe com base nos dados de treinamento. Em seguida, estima as probabilidades condicionais de cada valor de atributo para cada classe. Utilizando o Teorema de Bayes e assumindo independência

condicional entre os atributos, calcula a probabilidade *a posteriori* de cada classe para a instância em teste. A classe com a maior probabilidade *a posteriori* é então escolhida como a predição para a instância.

3.3.2 Support Vector Machine

A *Support Vector Machine* (SVM) é uma técnica que ajuda na seleção de um classificador adequado para um conjunto de dados específico [Faceli et al. 2021]. Assim, seu objetivo é encontrar uma fronteira que separe as possíveis classes no conjunto de dados e classificar o maior número possível de exemplos, maximizando a distância da fronteira aos pontos mais próximos desta [Géron 2019].

Além disso, como o algoritmo que apliquei foi o *LinearSVC* e ele não possui um método para calcular as probabilidades previstas de cada classe, utilizei um calibrador de modelos para classificações multiclasse chamado *CalibratedClassifierCV*. Com isso, foi possível obter o valor da área sob a curva ROC.

3.3.3 Random Forest Classifier

O *Random Forest Classifier* (RF) é um algoritmo *ensemble*, isto é, que combina o resultado de múltiplos modelos em busca de produzir um melhor modelo preditivo. Assim, ele combina várias árvores de decisão, treinadas com atributos e conjunto de dados distintos, e utiliza o método *bagging* para unir o resultado de todas as árvores geradas e identificar a classe final por meio de votação [Breiman 2001]. Ressalta-se que árvores de decisão são um tipo de estrutura de dados que possui nós internos, ramos e nós folhas. Em cada nó interno ocorre a verificação de um dos atributos. Os ramos representam os possíveis resultados da verificação. E os nós folhas representam as classes a serem atribuídas (CASTRO; FERRARI, 2016 apud ROCHA, 2020).

De forma geral, as principais etapas do algoritmo são: i) um conjunto de n amostras é criado utilizando *bootstrapping*, isto é, são selecionadas amostras aleatórias com reposição (ou seja, a mesma linha pode ser escolhida mais de uma vez); ii) uma árvore de decisão é construída por amostra, resultando na predição de uma classe; iii) as classes determinadas por cada árvore são computadas como um voto, sendo que a classe com maior quantidade de votos se torna a classe predita do modelo [Freitas 2018, Torres 2022].

3.4 Avaliação das métricas de desempenho dos modelos

Para avaliar o desempenho de cada modelo foram utilizadas as métricas avaliadas por Torres, Guterres e Celestino 2023 para fins de comparação, sendo elas: matriz de confusão, área sob a curva ROC, acurácia e tempo de processamento do modelo em minutos. Adicionalmente, também calculou-se três métricas comumente usadas para problemas de classificação multiclasse: precisão, *recall* e *F1-score* (BRADLEY; RAJENDRAN, 2021 apud RAJENDRAN; SRINIVAS; GRIMSHAW, 2021).

A matriz de confusão é uma representação tabular do desempenho de um modelo de classificação no conjunto de teste, mostrando o número de previsões corretas e incorretas para cada categoria [Ahmad et al. 2020]. A Tabela 4 apresenta um exemplo dessa matriz, em que as linhas representam as categorias reais e as colunas as previstas, resultando em quatro parâmetros:

- **verdadeiro positivo (TP):** casos em que o modelo previu corretamente a classe positiva (verdadeira);
- **falso positivo (FP):** casos em que o modelo previu incorretamente a classe positiva (falso);
- **falso negativo (FN):** casos em que o modelo previu incorretamente a classe negativa (falso);
- **verdadeiro negativo (TN):** casos em que o modelo previu corretamente a classe negativa (verdadeiro).

Tabela 4 – Matriz de Confusão

	Previsto Positivo	Previsto Negativo
Real Positivo	<i>TP</i>	<i>FN</i>
Real Negativo	<i>FP</i>	<i>TN</i>

Fonte: Elaborado pela autora.

A área sob a curva ROC (AUC-ROC) mostra se o classificador funciona melhor do que uma escolha aleatória. Para isso, as taxas de verdadeiros-positivo e falsos-positivo são relacionadas, o que permite saber se o classificador consegue diferenciar bem as

classes [Murphy 2012]. Quanto maior a AUC-ROC, melhor o modelo é em distinguir as classes.

A acurácia indica uma performance geral do modelo. Ela é definida como a taxa total de acerto do algoritmo, ou seja, calcula dentre todas as classificações, quantas ele classificou corretamente, considerando tanto as verdadeiras como as falsas [Ahmad et al. 2020]. Já a precisão mostra a proporção de verdadeiros positivos em relação a todos os classificados como positivos. Por outro lado, o *recall* mostra quantas observações positivas foram classificadas corretamente em relação ao total de observações positivas. Por fim, o *F1-score* é uma média harmônica entre a precisão e o *recall*, sendo utilizada para comparar dois modelos quando estas métricas são importantes [Sicsú, Samartini e Barth 2023]. Todas essas medidas variam de 0 a 1, sendo que quando maior o valor, melhor o desempenho de classificação. A Tabela 5 apresenta as fórmulas de cada métrica.

Tabela 5 – Fórmulas das métricas de avaliação de classificação.

Métrica	Fórmula
Acurácia	$\frac{TP+TN}{TP+TN+FP+FN}$
Precisão	$\frac{TP}{TP+FP}$
<i>Recall</i>	$\frac{TP}{TP+FN}$
<i>F1-score</i>	$2 \cdot \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}}$

Fonte: Elaborado pela autora.

3.4.1 Importância das variáveis

Para avaliar as variáveis mais importantes para cada modelo, foi aplicada o *permutation importance*. Essa técnica mede o impacto de cada variável no desempenho do modelo, através de uma alteração aleatória nos seus valores. Quando uma variável é permutada (ou seja, seus valores são embaralhados), o modelo perde parte de sua capacidade de previsão, o que reflete a importância dessa variável. Dessa forma, quanto maior a queda na performance do modelo após a permutação de uma variável, maior é a sua importância. Assim, para calcular a importância das variáveis, a função *permutation_importance* foi utilizada, repetindo o processo 10 vezes para garantir a robustez dos resultados. Em seguida, foi gerado um gráfico com as variáveis ordenadas de acordo com a sua importância, permitindo identificar as que mais contribuem para o modelo.

De forma complementar, também foi considerada a multicolinearidade e os efeitos específicos dos modelos para analisar os resultados de importância das variáveis. A multicolinearidade ocorre quando duas ou mais variáveis independentes são altamente

correlacionadas. Quando isso ocorre, as pontuações de importância das variáveis correlacionadas podem ser diluídas, já que a permutação de uma variável pode afetar a previsão do modelo de forma similar à permutação de outra variável correlacionada. Dessa forma, a correlação foi calculada por meio do coeficiente de Pearson, utilizando a função `.corr()` da biblioteca Pandas. Esse coeficiente varia entre -1 e 1, sendo que:

- 1: indica correlação positiva perfeita, ou seja, as duas variáveis aumentam ou diminuem juntas de forma linear;
- -1: indica correlação negativa perfeita, ou seja, quando uma variável aumenta, a outra diminui de forma linear;
- 0: indica que não há correlação linear;
- Valores acima de 0,8 ou abaixo de -0,8: indica uma correlação forte.

Em relação aos efeitos específicos dos modelos, diferentes tipos de modelos podem atribuir importâncias distintas às variáveis, mesmo que as variáveis sejam as mesmas. Por exemplo, no *Naive Bayes*, que assume a independência entre os recursos, a importância de permutação pode se comportar de maneira diferente se as variáveis forem fortemente correlacionadas, já que o modelo não leva em consideração essas correlações. Por outro lado, no SVM, os kernels não lineares são capazes de capturar padrões mais complexos, o que pode levar a uma classificação diferente da importância dos recursos em comparação com modelos lineares. Já em modelos como *Random Forest*, que lida bem com interações e correlações entre os recursos, a distribuição da importância pode ser distinta, refletindo a capacidade do modelo de capturar essas interações de forma mais robusta, em contraste com modelos mais simples.

3.5 Códigos da preparação dos dados e dos modelos

Para executar todas as etapas descritas foram utilizadas as bibliotecas *Pandas*, *Numpy*, *Scikit-learn*, *Imbalanced-learn*, *Matplotlib* e *Seaborn* do Python. O computador utilizado possui as seguintes características: DELL, 16 GB de RAM, processador 13th Gen Intel(R) Core(TM) i7-1360P 2.20 GHz. Os *notebooks* com os códigos em Python da preparação dos dados, assim como o de cada algoritmo de aprendizado de máquina, podem ser acessados por este link.

4 RESULTADOS

4.1 Resultados dos modelos para previsão do valor do dano moral

Aqui apresentam-se os resultados da aplicação do aprendizado de máquina para prever o valor do dano moral, nos processos onde houve condenação. A Tabela 6 apresenta os resultados das métricas globais de desempenho de cada modelo, sendo que os melhores valores estão em negrito. Ressalto que para a precisão, o *recall* e o *F1-score* foi utilizado o parâmetro *average='weighted'*, tendo em vista o problema multiclasse. Esse parâmetro calcula a média dessas métricas, ponderando cada classe pela sua distribuição no conjunto de dados. Isso significa que classes com mais amostras têm um impacto maior na métrica final do que classes com menos amostras, refletindo melhor o desempenho geral do modelo em um problema multiclasse.

Tabela 6 – Desempenho global dos modelos de aprendizado de máquina

Métricas	NB	SVM Linear	RF
Acurácia	0,6726	0,7056	0,7346
Precisão	0,6757	0,7335	0,7665
<i>Recall</i>	0,6726	0,7056	0,7346
<i>F1-score</i>	0,6732	0,7158	0,7454
AUC-ROC	0,7801	0,8302	0,8688
Tempo de processamento (min)	0,2487	0,7735	19,6942

Fonte: Elaborado pela autora.

Observa-se na Tabela 6, que o RF se destacou com os melhores resultados para todas as métricas entre os três modelos. Em relação aos tempos de processamento, o NB foi o que processou em menor tempo devido à sua simplicidade em comparação com as outras técnicas, seguido pelo SVM e o RF, o que está de acordo com o trabalho de Torres, Guterres e Celestino 2023 e outros autores como Ting et al. 2011, Tsangaratos e Ilia 2016 e Lei et al. 2017.

Analisando os resultados do RF, percebe-se que ele apresentou uma acurácia de 73,46%, o que significa que 73,46% das previsões feitas pelo modelo estão corretas em relação ao total de previsões feitas. Em outras palavras, o modelo está acertando a classificação da maioria das instâncias. Em relação a precisão dessas predições, ou seja, a proporção de predições positivas corretas em relação ao total de predições positivas, o

valor foi de 76,65%. O *recall* foi de 73,46%, o que indica que os modelos estão capturando corretamente 73,46% das instâncias positivas existentes no conjunto de dados. Isto é relevante para problemas em que é crucial identificar corretamente todas as instâncias de uma classe. E o *F1-score*, que combina precisão e *recall* em uma única métrica, também foi de 74,54%, o que sugere um bom equilíbrio entre essas duas métricas, indicando que os modelos estão acertando tanto na identificação correta das instâncias positivas quanto na minimização dos falsos positivos. Por fim, a AUC-ROC indica quão bem o modelo consegue distinguir as diferentes classes e o RF apresentou o valor de 86,88%.

4.1.1 Matrizes de confusão

Além das métricas globais apresentadas acima, também foram calculadas as matrizes de confusão de cada modelo. Por meio delas, é possível avaliar o desempenho de um classificador por meio da análise de quais classes foram classificadas corretamente. As Tabelas 7, 8 e 9 apresentam as matrizes de confusão de cada modelo.

Tabela 7 – Matriz de Confusão - Naive Bayes

	Predito: Baixo	Predito: Médio	Predito: Alto
Real: Baixo	1438	759	1418
Real: Médio	702	10721	2437
Real: Alto	1045	1891	4797

Fonte: Elaborado pela autora.

Tabela 8 – Matriz de Confusão - SVM Linear

	Predito: Baixo	Predito: Médio	Predito: Alto
Real: Baixo	2160	411	1044
Real: Médio	1133	10853	1874
Real: Alto	1870	1088	4775

Fonte: Elaborado pela autora.

Tabela 9 – Matriz de Confusão - Random Forest

	Predito: Baixo	Predito: Médio	Predito: Alto
Real: Baixo	2388	274	953
Real: Médio	951	10975	1934
Real: Alto	1851	726	5156

Fonte: Elaborado pela autora.

A partir dos valores apresentados nas Tabelas 7, 8 e 9 foi possível calcular as métricas para cada classe, conforme apresentado nas Tabelas 10, 11 e 12.

Tabela 10 – Métricas da classe Baixo

Métricas	NB	SVM Linear	RF
Precisão	45,14%	41,85%	45,94%
<i>Recall</i>	39,78%	59,70%	66,10%
<i>F1-score</i>	42,32%	49,90%	54,45%

Fonte: Elaborado pela autora.

Tabela 11 – Métricas da classe Médio

Métricas	NB	SVM Linear	RF
Precisão	80,22%	87,72%	91,61%
<i>Recall</i>	77,32%	78,20%	79,22%
<i>F1-score</i>	78,75%	82,74%	85,14%

Fonte: Elaborado pela autora.

Tabela 12 – Métricas da classe Alto

Métricas	NB	SVM Linear	RF
Precisão	73,22%	61,36%	64,81%
<i>Recall</i>	59,17%	69,12%	75,42%
<i>F1-score</i>	64,41%	64,85%	69,84%

Fonte: Elaborado pela autora.

Em relação a precisão de cada classe:

- **Baixo:** o melhor foi o RF com 45,94%, seguido pelo NB com 45,14% e o SVM Linear com 41,85%;
- **Médio:** o melhor foi o RF com 91,61%, seguido pelo SVM Linear com 87,72% e o NB com 80,22%;
- **Alto:** o melhor foi o NB com 73,22%, seguido pelo RF com 64,81% e o SVM Linear com 61,36%;

Em relação ao *recall* de cada classe:

- **Baixo:** o melhor foi o RF com 66,10%, seguido pelo SVM Linear com 59,70% e o NB com 39,78%;
- **Médio:** o melhor foi o RF com 79,22%, seguido pelo SVM Linear com 78,20% e o NB com 77,32%;

- **Alto:** o melhor foi o RF com 75,42%, seguido pelo SVM Linear com 69,12% e o NB com 59,17%;

Em relação ao *F1-score* de cada classe:

- **Baixo:** o melhor foi o RF com 54,45%, seguido pelo SVM Linear com 49,90% e o SVM Linear com 42,32%;
- **Médio:** o melhor foi o RF com 85,14%, seguido pelo SVM Linear com 82,74% e o NB com 78,75%;
- **Alto:** o melhor foi o RF com 69,84%, seguido pelo SVM Linear com 64,85% e o NB com 64,41%;

Com base nos resultados obtidos para precisão, *recall* e *F1-score*, a classe "Médio" se destacou como a melhor classificada entre as três, com o modelo RF apresentando os melhores resultados em todas as métricas para esta categoria: precisão (91,61%), *recall* (79,22%) e *F1-score* (85,14%). No entanto, a classe "Alto" teve um desempenho superior com o modelo NB, alcançando a maior precisão (73,22%), embora o RF tenha se destacado no *recall* (75,42%) e no *F1-score* (69,84%). Por fim, na classe "Baixo", o RF foi o melhor modelo, com 45,94% de precisão, 66,10% de *recall* e 54,45% de *F1-score*, destacando-se significativamente em relação aos outros modelos para essa classe. Em termos gerais, o RF foi o modelo mais consistente, alcançando bons resultados em todas as classes e métricas, mas o NB se destacou na classe "Alto" pela sua precisão.

4.1.2 Importância das variáveis

Antes de analisar os resultados da importância das variáveis, é importante observar as correlações entre elas. A Tabela 13 apresenta as correlações maiores que 0,5 e menores que -0,5 que foram identificadas.

Tabela 13 – Correlação entre Variáveis

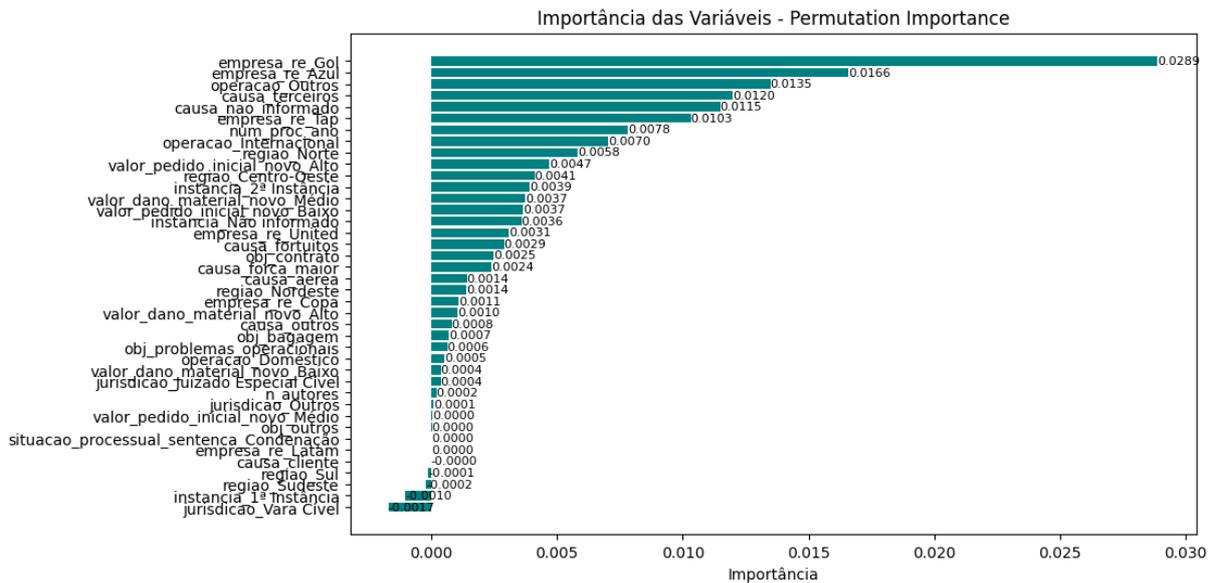
Variável 1	Variável 2	Correlação
causa_ nao_ informado	empresa_re_Azul	0,938022
jurisdicao_Vara Cível	jurisdicao_Juizado Especial Cível	0,905471
valor_dano_material_novo_Médio	valor_dano_material_novo_Baixo	0,796130
empresa_re_Gol	empresa_re_Azul	0,778577
instancia_Não informado	instancia_1ª Instância	0,740919
causa_ nao_ informado	empresa_re_Gol	0,730323
operacao_Outros	empresa_re_Gol	0,714924
operacao_Outros	operacao_Doméstico	0,680422
causa_fortuitos	empresa_re_United	0,670907
causa_ nao_ informado	instancia_Não informado	0,649546
operacao_Doméstico	empresa_re_Azul	0,636525
valor_pedido_inicial_novo_Médio	valor_pedido_inicial_novo_Baixo	0,627408
instancia_Não informado	empresa_re_Azul	0,608945
causa_ nao_ informado	operacao_Doméstico	0,592934
operacao_Internacional	empresa_re_Tap	0,589450
operacao_Outros	empresa_re_Azul	0,553199
causa_aerea	causa_ nao_ informado	0,543827
obj_bagagem	obj_problemas_operacionais	0,536848
operacao_Internacional	operacao_Doméstico	0,519364
causa_ nao_ informado	operacao_Outros	0,519250
valor_dano_material_novo_Baixo	valor_dano_material_novo_Alto	0,518663
valor_pedido_inicial_novo_Médio	valor_pedido_inicial_novo_Alto	0,501331

Fonte: Elaborado pela autora.

Na Tabela 13, percebe-se que as correlações são todas positivas, isto é, as variáveis aumentam ou diminuem juntas. Foram identificadas algumas correlações fortes, como a entre "causa_ nao_ informado" e "empresa_re_Azul" (0,938022), entre "jurisdicao_Vara Cível" e "jurisdicao_Juizado Especial Cível" (0,905471), entre "valor_dano_material_novo_Médio" e "valor_dano_material_novo_Baixo" (0,796130) e entre "empresa_re_Gol" e "empresa_re_Azul" (0,778577). Esses altos valores indicam que as variáveis estão fortemente relacionadas, o que pode sugerir que ambas influenciam o modelo de forma similar.

Com base nas correlações identificadas, as figuras 1, 2 e 3 apresentam os resultados da *permutation importance* e fornecem uma visão detalhada das variáveis que mais influenciaram a previsão do valor do dano moral em cada modelo.

Figura 1 – Importância das variáveis no Modelo NB



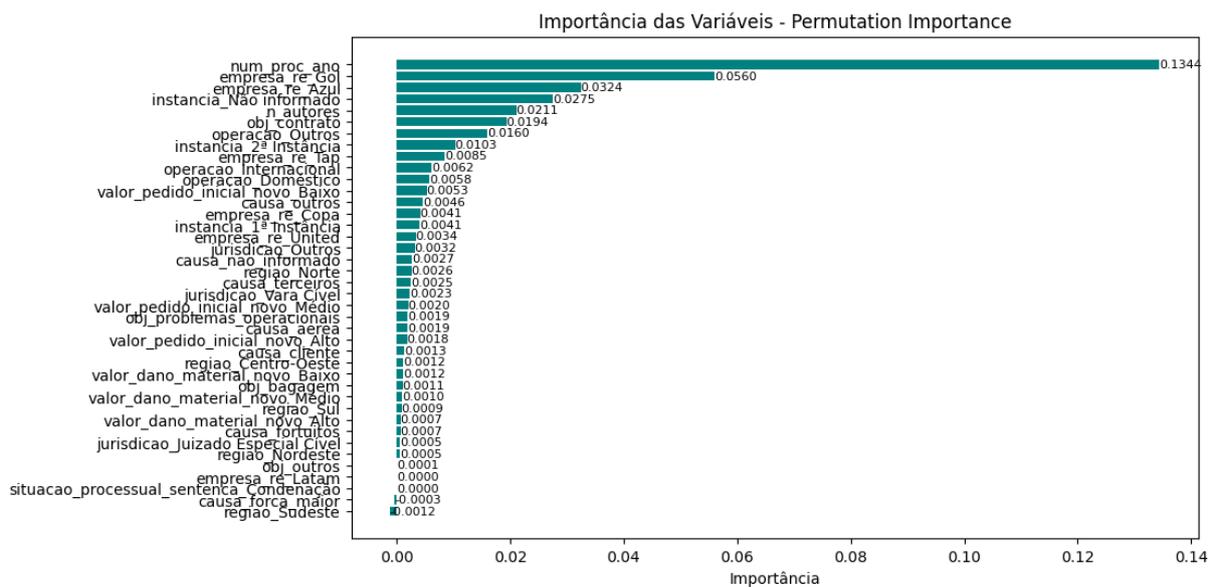
No modelo NB, as cinco variáveis com maior importância foram `empresa_re_Gol`, `empresa_re_Azul`, `operacao_Outros`, `causa_terceiros` e `causa_nao_informado`. Primeiramente, as variáveis `empresa_re_Gol` e `empresa_re_Azul` indicam que o nome da empresa aérea teve um impacto significativo na determinação do valor do dano moral neste modelo, sugerindo que características específicas dessas empresas influenciam o valor atribuído. A correlação entre essas duas variáveis também foi alta (0,778577), o que confirma a forte associação entre elas e pode indicar que as características da empresa aérea têm uma influência semelhante sobre o valor do dano moral.

A variável `operacao_Outros` também se destacou, refletindo a relevância de operações não categorizadas diretamente em nacional ou internacional. Essa variável tem uma correlação moderada com `operacao_Domestico` (0,680422), sugerindo que operações não claramente classificadas como nacionais ou internacionais podem estar, de alguma forma, relacionadas a operações domésticas, e juntas essas variáveis têm um impacto considerável no valor do dano moral.

Além disso, as variáveis relacionadas à causa do processo, como `causa_terceiros` e `causa_nao_informado`, indicam que as razões subjacentes para o processo (como ações de terceiros ou a falta de informações claras) também desempenharam um papel crucial na definição do valor do dano moral. A correlação entre `causa_nao_informado` e outras variáveis, como `causa_aerea` (0,543827), embora mais baixa, ainda sugere que a falta de

informações claras ou a causa do processo podem ser fatores inter-relacionados que afetam o valor atribuído. A correlação mais forte entre `causa_ nao_ informado` e `empresa_re_Azul` (0,938022) reforça a ideia de que, em alguns casos, a falta de informações sobre a causa pode ser associada a características específicas da empresa envolvida, influenciando diretamente a decisão sobre o valor do dano moral.

Figura 2 – Importância das variáveis no Modelo SVM Linear



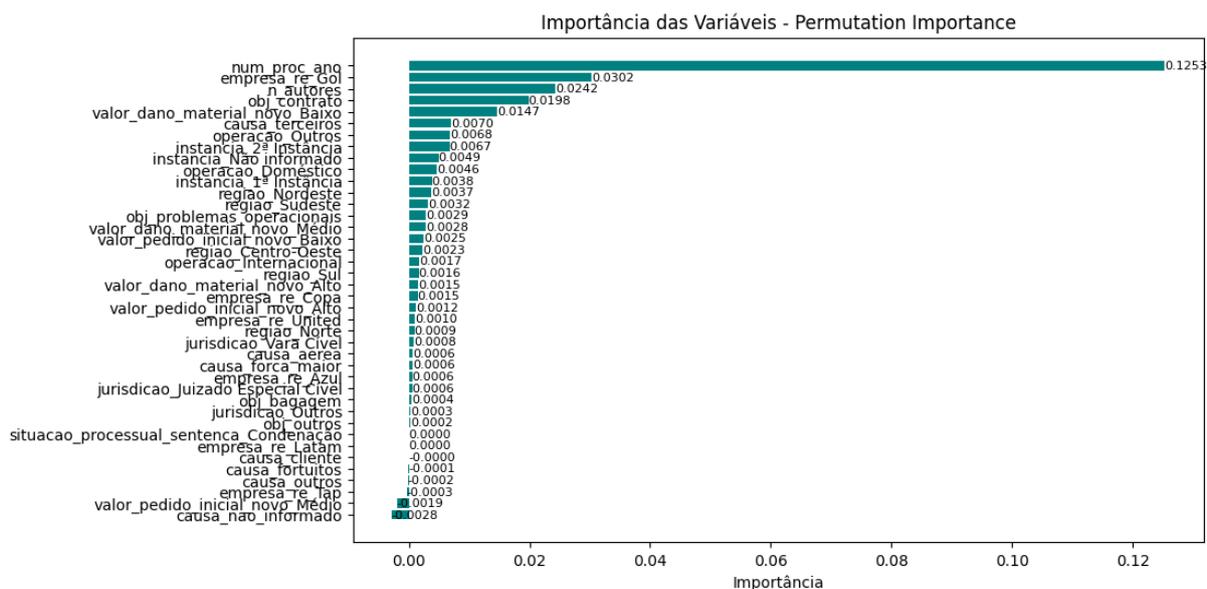
No modelo SVM Linear, as cinco variáveis mais importantes foram `num_proc_ano`, `empresa_re_Gol`, `empresa_re_Azul`, `instancia_Nao_informado` e `n_autores`. A variável `num_proc_ano` destacou-se como a mais relevante, sugerindo que o ano em que o processo ocorreu tem um papel importante na previsão do valor do dano moral. Pode-se inferir que o ano em que o processo ocorre pode estar diretamente relacionado a tendências jurídicas e sociais que impactam a avaliação do valor do dano moral, como foi o caso da Covid-19. As variáveis `empresa_re_Gol` e `empresa_re_Azul` novamente indicaram que o nome da empresa aérea influenciou significativamente o valor atribuído.

A `instancia_Nao_informado` refletiu que a falta de informações claras sobre a instância do processo também teve impacto na estimativa do dano moral. Além disso, a correlação mais alta entre `instancia_Nao_informado` e `instancia_1ª Instância` (0,740919) sugere que, quando não se sabe a instância, pode haver uma tendência de associar esse processo a uma instância inicial (1ª Instância).

Por fim, `n_autores` indicou que o número de autores envolvidos no processo também

influenciou o valor do dano moral, possivelmente sugerindo que processos com múltiplos autores tendem a resultar em valores mais elevados. Embora a correlação direta entre `n_autores` e outras variáveis não tenha sido destacada, o fato de múltiplos autores estarem envolvidos no processo pode estar relacionado a maior complexidade ou maior dano alegado, o que normalmente resulta em um valor de indenização maior.

Figura 3 – Importância das variáveis no Modelo RF



Já no modelo RF, as variáveis mais importantes foram `num_proc_ano`, `empresa_re_Gol`, `obj_contrato` e `valor_dano_material_novo_Baixo`. Assim como no modelo SVM, `num_proc_ano` e `empresa_re_Gol` se destacaram, evidenciando a relevância do ano em que o processo ocorreu e o nome da empresa na determinação do valor do dano moral. A variável `n_autores` também se manteve importante.

O `obj_contrato` indicou que os processos referentes a contrato podem influenciar a gravidade do dano moral. Além disso, a correlação entre `obj_contrato` e `obj_problemas_operacionais` (0,536848) indica que esses objetos, têm uma conexão com a natureza do processo, impactando o valor do dano moral.

A variável `valor_dano_material_novo_Baixo` mostrou-se relevante, indicando que o valor do dano material, especialmente quando baixo, pode influenciar a estimativa do dano moral. A correlação entre `valor_dano_material_novo_Baixo` e `valor_dano_material_novo_Alto` (0,518663) sugere que há uma relação entre os diferentes níveis de danos materiais e a estimativa do dano moral. Esse comportamento pode refletir

a forma como os danos materiais, especialmente os de valor mais baixo, impactam a percepção de gravidade do processo, o que influencia diretamente a avaliação do dano moral.

Em relação às variáveis mais importantes apresentadas para cada modelo, observa-se que, embora algumas variáveis se destaquem em mais de um modelo, a distribuição da importância varia conforme a arquitetura de cada um. No modelo *Naive Bayes*, as variáveis relacionadas a empresas aéreas e causas processuais têm uma relevância expressiva, mas essa atribuição não leva em conta possíveis correlações entre as variáveis, o que pode limitar a capacidade do modelo de capturar interações mais complexas. Já no SVM, a influência das variáveis, como o ano do processo e as empresas aéreas, reflete a capacidade do modelo de capturar padrões mais complexos e interações não lineares, o que pode explicar a diferença na classificação das variáveis quando comparado ao Naive Bayes. Por fim, no modelo *Random Forest*, que lida de forma mais robusta com interações e correlações entre as variáveis, a importância das variáveis, como o número de autores e o valor do dano material, é mais balanceada, refletindo melhor as relações entre as variáveis e sua interação no contexto do problema.

4.1.3 Discussão dos Resultados

Os resultados aqui apresentados confirmam os achados de Torres, Guterres e Celestino 2023 e de outros trabalhos anteriores da literatura [Ting et al. 2011, Tsangaratos e Ilia 2016, Lei et al. 2017], que consideraram a técnica de Random Forest (RF) como a mais acurada para classificar e prever se o valor na condenação seria Baixo, Médio ou Alto.

Este trabalho se diferenciou dos anteriores por aplicar o aprendizado de máquina para prever a classificação do valor do dano moral (baixo, médio ou alto) nos processos onde houve condenação. Como já mencionado anteriormente, o RF apresentou uma acurácia de 73,46%. Apesar deste valor ser inferior a outros achados anteriores, isso significa que quase três quartos das previsões do valor do dano moral feitas pelo modelo estão corretas e o modelo está acertando a classificação em 73,46% dos processos analisados.

Além da acurácia, a literatura costuma reportar outras métricas. Em relação à precisão (proporção de predições da classificação do dano moral corretas) foi de 76,65%. O *recall*, ou sensibilidade (proporção de captura de instâncias positivas das classificações de dano moral no conjunto de dados), foi de 73,46%. Finalmente, o *F1-score* (combina precisão e *recall*) foi de 74,54%, o que sugere um bom equilíbrio entre essas duas métricas,

indicando que os modelos estão acertando tanto na identificação correta das instâncias positivas quanto na minimização dos falsos positivos. Por fim, a AUC-ROC (quão bem o modelo consegue distinguir as diferentes classes) do RF foi de 86,88%.

Também como comentado anteriormente, a classe "Médio" do valor do dano moral se destacou como a melhor classificada, com as métricas do RF: precisão (91,61%), *recall* (79,22%) e *F1-score* (85,14%). Na classe "Alto", o RF teve um desempenho inferior ao modelo NB na precisão, que ficou em 64,81%, mas foi o melhor no *recall* (75,42%) e no *F1-score* (69,84%). Na classe "Baixo", o RF também foi o melhor modelo, mas com valores das métricas inferiores às outras duas categorias: 45,94% de precisão, 66,10% de *recall* e 54,45% de *F1-score*.

Na análise da importância das variáveis para prever o valor do dano moral, primeiramente, é importante ressaltar que não é possível uma análise comparativa entre modelos. Como já explicado, apesar de se observar que algumas variáveis se destacam em mais de um modelo, a importância explicativa varia com a arquitetura de cada modelo.

Entre as variáveis, merece destaque a alta correlação entre algumas variáveis explicativas incluídas como características do aprendizado de máquina: *empresa_re_Gol* e *empresa_re_Azul* (0,778577); *causa_nao_informado* e *empresa_re_Azul* (0,938022). Daí, a relevância da aplicação do RF, que lida de forma mais robusta com interações e correlações entre as variáveis, como já pontuado.

Assim, com foco no modelo *Random Forest*, destaca-se a importância das variáveis número de autores e o valor do dano material "Baixo", por serem mais balanceadas. O valor do dano material, especialmente quando baixo, pode influenciar a estimativa do dano moral. Além dessas, as variáveis mais importantes foram *num_proc_ano*, *empresa_re_Gol*, *n_autores*, *obj_contrato* e *valor_dano_material_novo_Baixo*. Isso evidencia a relevância do ano em que o processo foi distribuído e o nome da empresa (em particular, a Gol) na determinação do valor do dano moral. O *obj_contrato* indicou que os processos referentes a contrato influenciam o dano moral.

5 CONSIDERAÇÕES FINAIS

Os desafios enfrentados pelas companhias aéreas brasileiras em relação aos litígios judiciais são significativos, refletindo-se em altos custos legais e indenizações. Assim, a aplicação de técnicas de aprendizado de máquina neste contexto são úteis na previsão de valores de indenizações, porém para que os modelos tenham uma boa performance é necessário se atentar para a complexidade dos dados jurídicos. Nesse sentido, a padronização dos motivos e causas dos processos judiciais através da metodologia baseada em análise de conteúdo e revisões sistemáticas de literatura mostrou-se interessante para agrupar os temas relacionados e melhorar a precisão dos modelos.

Em relação aos modelos, o *Random Forest* se destacou como o mais eficaz neste trabalho, alcançando bons resultados em todas as classes analisadas, destacando-se especialmente na classe "Médio". No estudo anterior de Torres, Guterres e Celestino 2023, o Random Forest também alcançou os melhores resultados, o que reforça que ele pode ser útil para o planejamento estratégico das companhias aéreas diante de litígios judiciais.

De forma complementar, a análise da importância das variáveis para cada modelo ressaltou os fatores que mais influenciam a previsão do valor do dano moral. Para todos os modelos, variáveis como o nome da empresa aérea (*empresa_re_Gol* e *empresa_re_Azul*) e o ano do processo (*num_proc_ano*) mostraram ser de grande relevância. O modelo RF, em particular, destacou a importância de variáveis adicionais, como o número de autores (*n_autores*) e o valor do dano material (*valor_dano_material_novo_Baixo*), sugerindo que fatores relacionados ao contexto do processo, como o número de partes envolvidas e a natureza do dano material, podem influenciar significativamente o valor do dano moral atribuído.

Por fim, o estudo confirmou que os modelos de aprendizado de máquina podem ser altamente eficazes na previsão de valores de dano moral. O uso dessas técnicas pode representar um avanço importante na automatização e melhoria da precisão na previsão de valores de dano moral, o que não apenas contribui para a mitigação de riscos legais e financeiros enfrentados pelas companhias aéreas, mas também promove uma maior segurança jurídica no setor, potencialmente impactando positivamente a operação das empresas, a competitividade do mercado e, conseqüentemente, o bem-estar dos consumidores e *stakeholders* envolvidos.

REFERÊNCIAS

- AHMAD, I. et al. Fake news detection using machine learning ensemble methods. *Complexity*, v. 2020, n. 1, p. 8885861, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/8885861>>.
- BRADLEY, J.; RAJENDRAN, S. Increasing adoption rates at animal shelters: A two-phase approach to predict length of stay and optimal shelter allocation. *BMC Veterinary Research*, Springer, v. 17, p. 1–16, 2021. Disponível em: <<https://link.springer.com/article/10.1186/s12917-020-02728-2>>.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001. Disponível em: <<https://link.springer.com/article/10.1023/a:1010933404324>>.
- CASTRO, L. de; FERRARI, D. *Introdução à Mineração de Dados: Conceitos Básicos*. [S.l.: s.n.], 2016.
- De Martino, G.; PIO, G.; CECI, M. Multi-view overlapping clustering for the identification of the subject matter of legal judgments. *Information Sciences*, v. 638, p. 118956, 2023. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S002002552300525X>>.
- FACELI, K. et al. Inteligência artificial: uma abordagem de aprendizado de máquina. 2021.
- FREITAS, C. C. G. de. Dissertação de Mestrado em Economia, *Demanda por seguro de automóvel no Rio de Janeiro*. Rio de Janeiro, Rio de Janeiro: [s.n.], 2018. Disponível em: <<https://repositorio.fgv.br/items/ee5b7cea-d645-46b5-9b36-aebf4982d1a8>>.
- FUX, L. et al. *Cartilha do Transporte Aéreo. Fortalecendo a cidadania e reduzindo a litigiosidade*. [s.n.], 2021. Disponível em: <www.cnj.jus.br>.
- GÉRON, A. *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow*. [S.l.]: Alta Books, 2019.
- LEI, M. et al. Automatically classify chinese judgment documents utilizing machine learning algorithms. In: SPRINGER. *Database Systems for Advanced Applications: DASFAA 2017 International Workshops: BDMS, BDQM, SeCoP, and DMMOOC, Suzhou, China, March 27-30, 2017, Proceedings 22*. [S.l.], 2017. p. 3–17.
- LENZ, M. et al. *Fundamentos de aprendizagem de máquina*. [S.l.: s.n.], 2020.
- LIMA, J. P.; COSTA, J. A. Comparing clustering techniques on brazilian legal document datasets. In: SPRINGER. *International Conference on Hybrid Artificial Intelligence Systems*. [S.l.], 2022. p. 98–110.

MITCHELL, T. M.; MITCHELL, T. M. *Machine learning*. [S.l.]: McGraw-hill New York, 1997. v. 1.

MURPHY, K. P. *Machine learning: a probabilistic perspective*. [S.l.]: MIT press, 2012.

RAGHAV, K. et al. Text and citations based cluster analysis of legal judgments. In: SPRINGER. *Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3*. [S.l.], 2015. p. 449–459.

RAGHUVVEER, K. et al. Legal documents clustering using latent dirichlet allocation. *IAES Int. J. Artif. Intell*, Citeseer, v. 2, n. 1, p. 34–37, 2012.

RAJENDRAN, S.; SRINIVAS, S.; GRIMSHAW, T. Predicting demand for air taxi urban aviation services using machine learning algorithms. *Journal of Air Transport Management*, v. 92, p. 102043, 2021. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699721000260>>.

ROCHA, A. C. P. Dissertação de Mestrado Profissional em Computação Aplicada, *Mineração de textos para classificação de processos judiciais trabalhistas*. Brasília, Distrito Federal: [s.n.], 2020. Disponível em: <<http://www.realp.unb.br/jspui/handle/10482/37933>>.

SABO, I. C. et al. Clustering of brazilian legal judgments about failures in air transport service: an evaluation of different approaches. *Artificial Intelligence and Law*, Springer, v. 30, n. 1, p. 21–57, 2022. Disponível em: <<https://link.springer.com/article/10.1007/s10506-021-09287-3#citeas>>.

SICSÚ, A. L.; SAMARTINI, A.; BARTH, N. L. *Técnicas de aprendizado de máquina*. [S.l.]: Editora Blucher, 2023.

TING, S. et al. Is naive bayes a good classifier for document classification. *International Journal of Software Engineering and Its Applications*, v. 5, n. 3, p. 37–46, 2011.

TORRES, G. d. O.; GUTERRES, M. X.; CELESTINO, V. R. R. Legal actions in brazilian air transport: A machine learning and multinomial logistic regression analysis. *Frontiers in Future Transportation*, v. 4, 2023. ISSN 2673-5210. Disponível em: <<https://www.frontiersin.org/articles/10.3389/ffutr.2023.1070533>>.

TORRES, G. de O. Dissertação de Mestrado em Engenharia de Infraestrutura Aeronáutica, *Judicialização no Transporte Aéreo Brasileiro: uma Análise por meio de Aprendizado de Máquina e de Regressão Logística Multinomial*. São José dos Campos, São Paulo: [s.n.], 2022. Disponível em: <http://www.bdita.bibl.ita.br/tesesdigitais/lista_resumo.php?num_tese=78593>.

TSANGARATOS, P.; ILIA, I. Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena*, Elsevier, v. 145, p. 164–179, 2016.

ZHANG, H.; ZHOU, L. Similarity judgment of civil aviation regulations based on doc2vec deep learning algorithm. In: IEEE. *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. [S.l.], 2019. p. 1–8.