



UTILIZAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA PARA A CLASSIFICAÇÃO DA SATISFAÇÃO DE USUÁRIOS NOS AEROPORTOS DO BRASIL

USE OF MACHINE LEARNING MODELS TO CLASSIFY USER SATISFACTION AT AIRPORTS IN BRAZIL

USO DE MODELOS DE APRENDIZAJE AUTOMÁTICO PARA CLASIFICAR LA SATISFACCIÓN DEL USUARIO EN AEROPUERTOS EN BRASIL

Leonardo Fernando de Oliveira ¹, Rafael Henrique Palma Lima ^{2*}, & Bruno Samways dos Santos ³

¹²³ Universidade Tecnológica Federal do Paraná, Campus Londrina

¹leo.olivaf@gmail.com ^{2*}rafaelhlma@utfpr.edu.br ³brunosantos@utfpr.edu.br

ARTIGO INFO.

Recebido: 23.04.2024

Aprovado: 15.05.2024

Disponibilizado:

PALAVRAS-CHAVE: Aeroportos; Aprendizado de Máquinas; Satisfação; Classificação.

KEYWORDS: Airports; Machine Learning; Satisfaction; Classification

PALABRAS CLAVE: Aeropuertos; Aprendizaje Automático; Satisfacción; Clasificación

*Autor Correspondente: Lima, R. H. P.

RESUMO

Este artigo descreve a aplicação de técnicas de aprendizado de máquina (AM) utilizando os dados da pesquisa de satisfação de usuários em vários aeroportos no Brasil para classificá-los de acordo com sua satisfação. Foram utilizados os métodos K-Nearest Neighbors (KNN), Naïve Bayes, Árvore de Decisões e Floresta Aleatória para classificar a satisfação dos usuários, e a regressão linear para a imputação de dados, utilizando o conjunto de dados de 2017 a 2022 como conjunto de treinamento. Os dados foram previamente processados e limpos. O conjunto de dados de 2017 a 2022 foi utilizado para treinar o modelo, enquanto o conjunto de dados mais recente de 2023 foi utilizado como conjunto para teste. Após a classificação, a técnica de hiperparâmetros foi aplicada para melhorar os resultados das métricas. Os modelos de aprendizado de máquina apresentaram resultados satisfatórios na classificação dos usuários. Além disso, a pesquisa revelou os principais fatores que afetam a satisfação dos clientes nos aeroportos, destacando o conforto acústico do aeroporto, a disponibilidade de sanitários e a quantidade e qualidade de estabelecimentos comerciais como os mais influentes.

ABSTRACT

This article describes the application of machine learning (ML) techniques using user satisfaction survey data at various airports in Brazil to classify them according to satisfaction. The K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Random Forest methods were used to classify user satisfaction, and linear regression for data imputation, using the dataset from 2017 to 2022 as a training set. The data was pre-processed and cleaned. The dataset from 2017 to 2022 was used to train the model, while the most recent dataset from 2023 was used as the test set. After classification, the hyperparameter technique was applied to improve the results of the metrics. The machine learning models showed satisfactory results in classifying users. Additionally, the research revealed the main factors that affect customer satisfaction at airports, highlighting airport acoustic comfort, restroom availability, and the quantity and quality of commercial establishments as the most influential.

RESUMEN

Este artículo describe la aplicación de técnicas de aprendizaje automático (AM) utilizando los datos de la encuesta de satisfacción de usuarios en varios aeropuertos en Brasil para clasificarlos según su satisfacción. Se utilizaron los métodos de K-Nearest Neighbors (KNN), Naïve Bayes, Árbol de Decisiones y Bosque Aleatorio para clasificar la satisfacción de los usuarios, y la regresión lineal para la imputación de datos, utilizando el conjunto de datos de 2017 a 2022 como conjunto de entrenamiento. Los datos fueron previamente procesados y limpiados. El conjunto de datos de 2017 a 2022 se utilizó para entrenar el modelo, mientras que el conjunto de datos más reciente de 2023 se utilizó como conjunto de prueba. Después de la clasificación, se aplicó la técnica de hiperparámetros para mejorar los resultados de las métricas. Los modelos de aprendizaje automático mostraron resultados satisfactorios en la clasificación de los usuarios. Además, la investigación reveló los principales factores que afectan la satisfacción del cliente en los aeropuertos, destacando el confort acústico del aeropuerto, la disponibilidad de sanitarios y la cantidad y calidad de establecimientos comerciales como los más influyentes.

1. INTRODUÇÃO

Um dos principais desafios em construir um negócio em torno dos lucros derivados da entrega de experiências aos consumidores que possam resultar em maior lealdade é o desenvolvimento de uma maneira simples de medir sua satisfação e, em seguida, vinculá-la ao desempenho organizacional (Reichheld, 2021). Essa satisfação do cliente pode ser um importante diferencial competitivo para a empresa, pois clientes mais satisfeitos podem escolher a empresa dentre os concorrentes, enquanto clientes insatisfeitos, além de acabarem optando por concorrentes, podem afetar negativamente a imagem da empresa e exigir mais recursos e atenção, o que pode aumentar os custos operacionais (Kotler, 2019).

Quando as empresas prestam um atendimento que supera as expectativas do consumidor, oferecendo produtos e/ou serviços de qualidade, os clientes não só podem atrair outros clientes, mas também podem retornar para consumir os produtos novamente (Kotler, 2019). Na área de *marketing*, os *feedbacks* dos clientes têm o importante papel de mensurar a satisfação do consumidor e com isso gerar dados que possam ser utilizados para posteriormente avaliar o impacto do *mix* de produtos e serviços de da empresa. Esses dados, após tratamento apropriado, podem ser explorados por técnicas de aprendizado de máquina (AM) para identificar padrões de comportamento e preferências dos clientes, melhorando os produtos e serviços prestados (Zaghloul et al., 2024).

Os dados se tornaram uma das principais matérias-primas de vários negócios, sendo visto como um insumo econômico vital utilizado para criar formas inovadoras de valor econômico. Com a abordagem adequada, os dados podem ser inteligentemente reutilizados para se tornarem uma fonte de inovação e novos serviços (Mayer-Schönberger & Cukier, 2013). De acordo com autores como Tene & Polonetsky (2012), vivemos na era do *big data* e os avanços em tecnologias de mineração de dados, juntamente com o aumento massivo na capacidade de computação e armazenamento de dados, expandiram o escopo das informações disponíveis e das formas de uso de tais informações por empresas e governos.

As empresas têm reconhecido a importância de criar bancos de dados com informações sobre seus clientes. A análise em tempo real do comportamento do cliente na internet para uma melhor seleção e segmentação de clientes já é uma realidade, o que permite que as empresas tomem diversas decisões, que variam desde a antecipação sobre o que os clientes provavelmente irão comprar (Gupta & Aggarwal, 2012) e otimizar a visibilidade e campanhas de divulgação de suas marcas e produtos (Ngai & Wu, 2022).

Na literatura acadêmica, diversos trabalhos têm estudado a aplicação de técnicas de AM para analisar a satisfação dos clientes. Por exemplo, o trabalho de Ruiz et al. (2024) estudou a aplicação de tais técnicas para avaliar uma base de dados de satisfação dos usuários de transporte público no Chile e prever seu grau de satisfação, revelando fatores mais relevantes no caso estudado. Em contextos como esse, a AM se destaca como uma abordagem apropriada para lidar com a complexidade e a variação dos dados, proporcionando um *framework* robusto e flexível para a pesquisa de satisfação de clientes (Wangkiat & Polprasert, 2023).

Nesse contexto, o objetivo deste trabalho foi utilizar métodos de AM para classificar a satisfação de passageiros em aeroportos brasileiros e utilizar tais modelos de classificação para identificar os fatores mais relevantes para sua satisfação.

Para atingir o objetivo proposto, métodos de imputação de dados foram empregados para preencher dados faltantes e aprimorar o desempenho dos algoritmos de classificação (Jäger *et al.*, 2021). Além disso, a avaliação da importância dos atributos disponibilizado pela técnica da Floresta Aleatória foi utilizado na análise dos fatores mais relevantes da satisfação, usando método similar ao apresentado por Zhao *et al.* (2022).

Após esta seção introdutória, o restante deste artigo está estruturado da seguinte forma. Na Seção 2, apresenta-se um breve referencial teórico sobre aprendizado de máquina e os métodos utilizados. Em seguida, a Seção 3 aborda a base de dados, descrevendo as etapas de limpeza, organização e definição da variável de saída. A Seção 4 discute os resultados obtidos e os métodos empregados. Na sequência, a Seção 5 proporciona uma visão geral sobre os resultados alcançados. Por fim, na Seção 6, sintetiza as conclusões do trabalho.

2. REFERENCIAL TEÓRICO

Esta seção apresenta o referencial teórico que embasou a realização desta pesquisa, abordando os tópicos básicos de aprendizado de máquina (AM), visão geral das técnicas de AM usadas e introdução à otimização de hiperparâmetros usando o algoritmo *Grid Search*.

2.1 Aprendizado de Máquina

O AM tem se revelado uma ferramenta poderosa para auxiliar na análise de dados e proporciona informações valiosas, revelando padrões não triviais, tendências e relacionamentos que, de outra forma, poderiam passar despercebidos (Perifanis & Kitsios, 2023). Esse aspecto também é destacado por Cranenburgh *et al.* (2022) ao afirmarem que os modelos de aprendizado de máquina possuem uma notável flexibilidade para aprender a estrutura dos dados sem serem limitados por suposições.

Existem três tipos principais de aprendizado de máquina: (i) aprendizado supervisionado, (ii) aprendizado não supervisionado e (iii) aprendizado por reforço. O aprendizado supervisionado aprende a mapear as entradas para as saídas correspondentes com base em exemplos fornecidos previamente. O objetivo é que o modelo generalize esses padrões e seja capaz de fazer previsões ou classificações precisas em dados não vistos anteriormente. Dessa forma, o algoritmo constrói um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados (Ludermir, 2021).

O aprendizado não supervisionado tem como objetivo a descoberta de padrões, estruturas ou agrupamentos nos dados de forma natural. O valor de saída ou o rótulo de cada instância não é especificado durante o aprendizado, logo o aprendizado não supervisionado auxilia na identificação de padrões não triviais na base de dados. O algoritmo agrupa os exemplos pelas similaridades dos seus atributos. O algoritmo analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters* (Witek, 2014).

Por fim, o aprendizado por reforço faz com que o modelo aprenda por meio de interações de um agente com um ambiente, a partir de recompensas e punições devido às ações do agente. O algoritmo não recebe a resposta correta, mas sim um sinal de reforço, de recompensa ou punição. O algoritmo faz uma hipótese baseada nos exemplos e determina se essa hipótese foi boa ou ruim (Ludermir, 2021)

Ainda que o AM seja uma ferramenta poderosa para a aquisição automática de conhecimento, deve ser observado que não existe um único algoritmo que apresente o melhor desempenho para todos os problemas (Monard & Baranaukas, 2003). Portanto é fundamental considerar vários critérios que podem garantir o modelo adequado para o problema em questão. Algoritmos como *K-Nearest Neighbors* (KNN), Árvore de Decisões, Floresta Aleatória e *Naive Bayes* (NB) são frequentemente utilizados e cada um possui suas próprias vantagens e limitações. Essas considerações envolvem a análise do custo computacional, a acurácia do modelo e a adequação às características do conjunto de dados. É essencial avaliar cuidadosamente esses fatores para escolher o algoritmo mais adequado e obter resultados precisos e confiáveis.

2.2 Técnicas de Aprendizado de Máquinas

O *K-Nearest Neighbors* (KNN) é um algoritmo de aprendizado de máquina não paramétrico utilizado para classificação e regressão. Ele se baseia na ideia de que objetos similares tendem a estar próximos uns dos outros no espaço de características. Em outros termos, o KNN procura os k pontos mais próximos da instância a ser classificada e usa a maioria deles para prever a classe de saída dessa instância (Alpaydin, 2010). O valor k é um parâmetro presente no algoritmo que pode ser controlado para melhorar seu desempenho. Este parâmetro deve ser atribuído como um número inteiro, o qual definirá a quantidade de vizinhos próximos a serem usados na determinação da classe da instância.

O algoritmo *Naive Bayes* (NB) baseia-se no teorema de Bayes e na suposição de independência condicional entre os recursos. De acordo com Murphy (2012), o NB é um método eficiente de aprendizado probabilístico, adequado para problemas de alta dimensionalidade. Dentro do algoritmo NB, diferentes variantes são adaptadas para diferentes distribuições de probabilidade. O método Gaussiano assume que os recursos seguem uma distribuição normal (gaussiana), enquanto o método Bernoulli é mais adequado para dados binários. As implementações computacionais do NB geralmente permitem parametrizar a suavização das estimativas de variância, tornando o modelo mais robusto em relação a dados com baixa variabilidade.

As Árvores de Decisão (do inglês, *Decision Trees* – DT) são algoritmos de aprendizado de máquina que modelam decisões e comportamentos por meio de uma estrutura em forma de árvore. Durante o aprendizado, a árvore cresce, novos ramos e folhas são adicionados com base na complexidade do problema presente nos dados (Alpaydin, 2010). Cada nó da árvore representa uma pergunta sobre os atributos dos dados, e as respostas a essas perguntas levam a ramificações adicionais na árvore. Esse processo continua até que uma decisão final seja alcançada nos nós terminais. As implementações computacionais deste algoritmo

permitem parametrizar vários aspectos como o número o número mínimo de amostras em um nó folha, o número mínimo de amostras necessário para dividir um nó e o número máximo de características a serem consideradas ao procurar a melhor divisão.

O algoritmo da Floresta Aleatória (do inglês, Random Forest – RF) constrói um conjunto de árvores de decisão, onde cada árvore fornece uma previsão com base nas características de entrada e as previsões individuais são agregadas para produzir uma previsão final. Essa agregação reduz a variância e aumenta a generalização do modelo. A floresta aleatória é conhecida por sua capacidade de lidar com problemas complexos, apresentando boa robustez ao *overfitting* e melhor desempenho em relação a uma única árvore de decisão. As implementações computacionais do RF permitem especificar a quantidade de árvores na floresta (Breiman, 2021).

A biblioteca *Scikit-learn* (Pedregosa et al., 2011) implementa os algoritmos mencionados anteriormente e permite a parametrização de diversos aspectos. A Tabela 1 lista os parâmetros associados a cada uma das técnicas de AM avaliadas neste artigo.

Tabela 1. Definição dos parâmetros utilizados nas técnicas de AM

Modelo	Parâmetros	Definição
Naive Bayes (NB)	<i>var_smoothing</i>	Quantidade de variância a cada característica
	<i>min_samples_leaf</i>	Número mínimo de amostras que um nó folha deve ter
Decision Tree (DT)	<i>min_samples_split</i>	Número mínimo de amostras necessário para que um nó interno seja dividido
	<i>max_features</i>	Número máximo de características a serem consideradas ao procurar a melhor divisão
	<i>min_samples_leaf</i>	Número mínimo de amostras que um nó folha deve ter
Random Forest (RF)	<i>min_samples_split</i>	Número mínimo de amostras necessário para que um nó interno seja dividido
	<i>max_features</i>	Número máximo de características a serem consideradas ao procurar a melhor divisão
	<i>n_estimators</i>	Número de árvores na floresta
	<i>n_neighbors</i>	Quantidade de vizinhos próximos
K-Nearest Neighbors (KNN)	<i>n_neighbors</i>	Quantidade de vizinhos próximos

2.3 Otimização dos Modelos de Aprendizado de Máquinas

Segundo Pontes Jr. & Fagundes (2023) a otimização dos hiperparâmetros é a etapa da pesquisa que visa melhorar o desempenho nas previsões dos modelos através de ajustes nas configurações padrões dos hiperparâmetros de cada modelo.

Os hiperparâmetros desempenham um papel crucial no ajuste fino e na otimização dos modelos de aprendizado de máquina. Ele permite realizar uma busca mais aprofundada para determinar quais são os melhores parâmetros para um determinado modelo (Ghate & Hemalatha, 2023). O *GridSearch*, uma técnica comumente utilizada nesse processo, permite testar sistematicamente cada parâmetro, combinando-os em cada iteração, a fim de encontrar a combinação ideal que maximize a métrica escolhida. Segundo Costa Filho *et al.* (2019) a realização da busca em grade para otimização dos parâmetros dos algoritmos de AM é importante para seleção de modelos que superaram a performance dos modelos de regressão.

O *GridSearch* avalia o modelo com diferentes valores para cada hiperparâmetro e, em seguida, combina esses valores em todas as possíveis combinações para determinar qual delas produz os melhores resultados. De acordo com Liashchynskiy & Liashchynskiy (2019) o *GridSearch* é um algoritmo de força bruta. Ele realiza uma busca completa por um determinado subconjunto do espaço de hiperparâmetros. Ao final, o *GridSearch* retorna o melhor conjunto de parâmetros, ou seja, aqueles que proporcionam a maior acurácia possível. Dessa forma, é possível tomar decisões sobre qual combinação de hiperparâmetros otimiza o modelo alcançando melhores resultados.

3. MÉTODO DE PESQUISA

Esta seção detalha os procedimentos para condução desta pesquisa. Primeiramente é descrita a base de dados analisada e, em seguida, são explicados os passos para preparação da base, análise dos resultados e otimização dos hiperparâmetros.

3.1 Descrição da base de dados

A base de dados utilizada neste trabalho foi obtida a partir do Portal Brasileiro de Dados Abertos do Governo Federal (Brasil, 2023). Essa base consiste em uma pesquisa de satisfação realizada nos 20 aeroportos mais movimentados do Brasil.

O arquivo da base de dados está no formato de planilha *OpenDocument* (.ods) e contém dados de 2013 a 2019 em um único arquivo. Porém, os dados referentes aos anos de 2020 a 2023 estão separados por trimestre. Para facilitar a manipulação dos dados, os arquivos de 2020 a 2022 foram combinados em um único arquivo, uma vez que possuíam a mesma estrutura e atributos. No entanto, os arquivos de 2023 foram combinados em um arquivo à parte, para serem usados como base de validação para o classificador desenvolvido. Dessa maneira, os valores de 2013 a 2022 foram utilizados como conjunto de treino, a qual também foi usada para imputar dados faltantes na base de 2023.

A base de dados contém vários atributos com respostas em forma de notas numéricas usando o sistema *Likert* de cinco pontos (1 = muito ruim / 2 = ruim / 3 = regular / 4 = bom / 5 = muito bom) para indicar o grau de satisfação com diversos aspectos do serviço dos aeroportos.

A base de dados no período entre 2013 e 2019 abrange um total de 483.707 instâncias, onde cada instância representa as respostas de um dos entrevistados e são armazenadas com uma única linha da base de dados. Conforme mostra a Tabela 2, o conjunto de dados é composto por 51 atributos, dos quais 38 são numéricos usando a escala *Likert*, 10 são textuais, 2 consistem em datas e 1 é representado em valor numérico sem usar a escala *Likert*.

Tabela 2. Tipos de variáveis existentes na base de dados utilizada

Tipo do Atributo	Pesquisa 2013 a 2019	Pesquisa 2020 a 2022
Númérico (1 a 5)	38	61
Texto	10	28
Data	2	2
Alfanumérico	-	2
Hora	-	2
Númérico (Maior que 5)	1	1

Fonte: Brasil (2023).

Já conjunto de dados referente ao período de 2020 a 2022 tem uma estrutura diferente, com 61 atributos numéricos usando a escala *Likert* e 28 atributos textuais, com 2 campos de data e mais 2 de hora. Ao todo, esse período conta com um total de 143.530 instâncias.

3.2 Etapas da pesquisa

Durante o processo de análise, foram realizadas etapas para preparar e tratar os dados dos períodos de 2013 a 2022 e, de forma separada, os dados de 2023.

Durante o processo de limpeza dos dados, foram removidas todas as linhas (instâncias) que continham valores nulos. Além disso, foram eliminados os valores que não eram considerados como respostas válidas, tais como "não respondeu", "não soube responder" e "não informou". A exclusão dos dados anteriores a 2017 foi realizada para garantir a relevância temporal da análise, considerando as rápidas mudanças no contexto da experiência dos clientes em aeroportos, permitindo uma investigação mais precisa e atualizada da satisfação dos passageiros.

Após a realização da limpeza, o conjunto resultante apresentou 8.734 instâncias. Foram selecionados 12 atributos para aplicação das técnicas de classificação, os quais são destacados na Tabela 3. Vale destacar que os 11 primeiros atributos são considerados como variáveis de entrada do modelo, ou seja, são os regressores do modelo. Por outro lado, o último atributo da Tabela 3 foi utilizado para derivar a variável de saída do modelo, pois indica o grau geral de satisfação do usuário com o aeroporto.

Tabela 3. Atributos selecionados para a aplicação das técnicas de classificação

Atributo	Tipo
Disponibilidade E Qualidade Das Informações Nos Painéis De Voo	Entrada
Disponibilidade De Sanitários	Entrada
Disponibilidade De Assentos Na Sala De Embarque	Entrada
Conforto Térmico Do Aeroporto	Entrada
Conforto Acústico Do Aeroporto	Entrada
Disponibilidade De Tomadas	Entrada
Limpeza Geral Do Aeroporto	Entrada
Qualidade Da Internet / Wi-Fi	Entrada
Custo-Benefício Dos Produtos De Lanchonetes/Restaurantes	Entrada
Quantidade E Qualidade De Lanchonetes/Restaurantes	Entrada
Quantidade E Qualidade De Estabelecimentos Comerciais	Entrada
Satisfação Geral Com O Aeroporto	Saída

Fonte: Autores.

Foi realizada uma classificação inicial utilizando o algoritmo *Naïve Bayes* para verificar de forma preliminar o desempenho da classificação antes da otimização dos hiperparâmetros. A Tabela 4 apresenta o desempenho obtido, com acurácia de 89%.

Tabela 4. Resultados preliminares usando a base de dados de 2017 a 2022

	Naive Bayes		
	Precisão	Recall	F1-score
Insatisfeito	0,57	0,75	0,65
Satisfeito	0,96	0,91	0,94
Acurácia	0,89		

Fonte: Autores.

As métricas apresentadas na Tabela 4 foram utilizadas para avaliar o desempenho de todos os modelos de classificação neste artigo e incluem: precisão, *recall*, *f1-score* e acurácia. Tais métricas são comumente usadas em modelos de aprendizado supervisionado e são baseadas na tabela verdade após a aplicação do classificador a todas as instâncias. Após a aplicação do classificador, as instâncias são avaliadas como Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN), sendo as quantidades de instâncias em cada categoria usadas para calcular as métricas de desempenho (Gerón, 2019). A precisão, calculada usando a Equação 1, é uma medida que quantifica a quantidade de acertos do modelo em relação ao número total de amostras avaliadas.

$$Precisão = \frac{VP}{VP + FP} \quad (1)$$

O *recall*, calculado usando a Equação 2, é uma métrica que demonstra a eficácia do modelo em classificar corretamente a classe de interesse.

$$Recall = \frac{VP}{VP + FN} \quad (2)$$

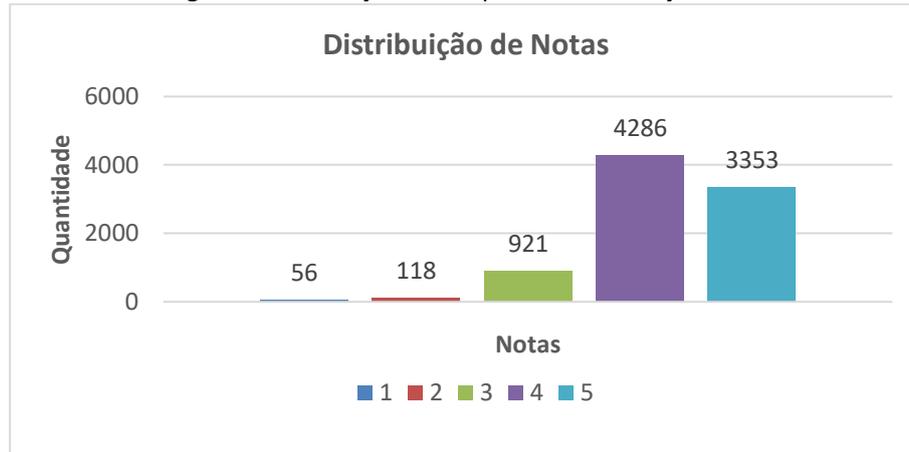
A métrica *F1-score*, calculada usando a Equação 3, combina tanto a capacidade do modelo em identificar corretamente a classe de interesse quanto a taxa de acertos em relação ao total de exemplos avaliados, fornecendo uma visão equilibrada do desempenho geral do modelo. O *f1-score* ajuda a entender quão equilibrado está o resultado, por isso é sempre importante observar quão bom é o valor.

$$F1 \text{ score} = 2 \times \frac{Precisão \times Recall}{Precisão + Recall} \quad (3)$$

A acurácia é calculada usando a Equação 4 e avalia a capacidade do modelo em classificar corretamente as instâncias em geral, representando a taxa de acertos do modelo em relação a todos os exemplos avaliados.

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (4)$$

A base de dados de 2023 foi preparada mantendo o mesmo conjunto de 12 variáveis da Tabela 3, as quais também estão presentes no banco de dados de 2017 a 2022. As instâncias com valores faltantes foram preservadas, porém instâncias com respostas como 'NS' (Não Sabe), 'NS/NR' (Não Sabe/Não Respondeu) e 'não informou', foram removidos. Essas ações resultaram em um banco de dados para treinamento com 8.734 instâncias. A Figura 1 mostra a distribuição das respostas da coluna "Satisfação Geral com o Aeroporto", utilizada como variável de saída no modelo de classificação. Pode-se notar que a base possui 87,5% de respondentes que deram nota 4 ou 5 para a satisfação geral, enquanto 10,5% deram nota 3 para a satisfação e apenas 2% responderam com notas 1 ou 2 para sua satisfação geral com o serviço do aeroporto.

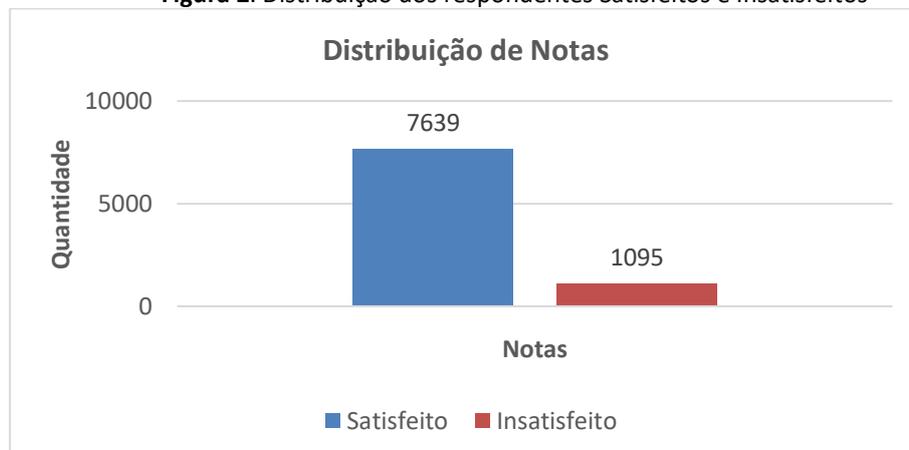
Figura 1. Distribuição das respostas de Satisfação Geral

Fonte: Autores.

Para a imputação dos dados faltantes dos atributos da base de 2023 foi utilizada a técnica de regressão linear como método de predição e a base de dados de 2017 a 2022 como referência para geração dos dados a serem imputados. A variável de saída (Satisfação Geral com o Aeroporto) foi retirada no momento da imputação, tanto no banco de dados de 2017 a 2022 quanto no banco de 2023, para evitar possíveis interferências nas análises futuras.

Após a imputação dos dados, uma nova variável foi adicionada às bases de dados para converter as notas de satisfação que originalmente usavam uma escala Likert de 1 a 5 para uma variável binária indicando se o entrevistado está satisfeito ou insatisfeito com o aeroporto. Os respondentes que atribuíram notas 1, 2 ou 3 foram considerados “Insatisfeitos” e os entrevistados que deram notas 4 ou 5 foram considerados “Satisfeitos”. A Figura 2 mostra a divisão dos respondentes após a conversão, com 87,5% dos entrevistados classificados como “Satisfeitos” e 12,5% classificados como “Insatisfeitos”.

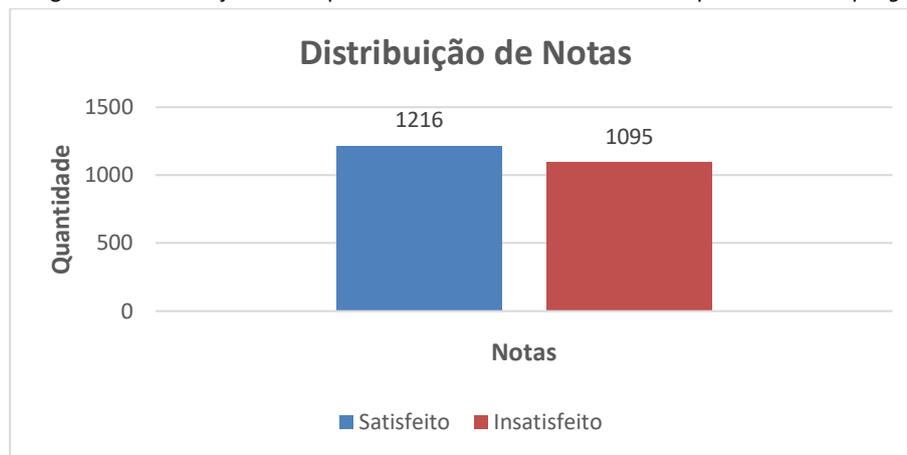
Após a imputação dos dados, foram treinados classificadores os algoritmos de aprendizado supervisionado *KNN*, *Floresta Aleatória*, *Árvore de Decisões* e *Naïve Bayes*.

Figura 2. Distribuição dos respondentes Satisfeitos e Insatisfeitos

Fonte: Autores.

Com base na Figura 2, pode-se notar que a base de dados é desbalanceada, com grande parte dos respondentes classificados como “Satisfeitos”, o que pode prejudicar a validade do classificador obtido a partir dos algoritmos de treinamento. Para superar esse problema foi empregada a técnica de *undersampling*. Essa técnica consiste em reduzir o número de exemplos da classe majoritária, que é a classe mais favorecida em termos de representação no conjunto de dados (Sun *et al.*, 2024). Nesse caso, foi aplicado um corte para que a classe “Insatisfeito” tivesse uma quantidade de instâncias igual a 90% da quantidade de instâncias da classe “Satisfeito”. A Figura 3 ilustra como ficou a distribuição das instâncias após a aplicação da técnica.

Figura 3 – Distribuição dos respondentes Satisfeitos e Insatisfeitos após o *undersampling*



Fonte: Autores.

3.3 Avaliação dos Hiperparâmetros

O algoritmo *GridSearch* foi usado para avaliar os classificadores considerando várias combinações de parâmetros. De acordo com Liashchynskiy & Liashchynskiy (2019) o *GridSearch* é um algoritmo de força bruta que testa todas as combinações de parâmetros de cada técnica e retorna o melhor classificador obtido, usando como referência o indicador de acurácia. A Tabela 5 mostra os parâmetros testados pelo *GridSearch* para cada um dos 4 algoritmos usados nesta pesquisa. Os valores dos parâmetros que resultaram no melhor desempenho estão destacados em negrito.

Tabela 5. Resultado da avaliação dos hiperparâmetros

Modelo	Parâmetros	Valores
<i>Naive Bayes</i>	<i>var_smoothing</i>	1e-1, 1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7, 1e-8
<i>Decision Tree</i>	<i>min_samples_leaf</i>	range(1, 2 , 20)
	<i>min_samples_split</i>	range(1, 38 , 150)
	<i>max_features</i>	sqrt
<i>Random Forest</i>	<i>min_samples_leaf</i>	range(1, 20)
	<i>min_samples_split</i>	range(1, 36 , 150)
	<i>max_features</i>	sqrt
	<i>n_estimators</i>	100, 110, 115, 120 , 125, 130, 135
<i>KNN</i>	<i>n_neighbors</i>	range(1, 28 , 150)

Fonte: Autores.

4. RESULTADOS

Esta seção apresenta os resultados obtidos ao aplicar as técnicas de AM conforme os passos descritos na seção anterior. Também é feita uma discussão sobre a importância dos atributos e suas implicações práticas para a gestão de aeroportos.

4.1 Aplicação das técnicas de AM

A implementação computacional das técnicas de aprendizado de máquina foi realizada usando a linguagem *Python* 3.12 no ambiente de desenvolvimento *Visual Studio Code* com a extensão *Jupyter*. Os pacotes *scikit-learn* versão 1.3.2 e *pandas* versão 2.1.4 foram utilizados para implementar os modelos. O código foi executado em um computador desktop com processador AMD *Ryzen* 5900X, 32GB de memória RAM no sistema operacional *Windows 11*. Os algoritmos *Decision Tree* (DT), *Random Forest* (RF), *K-Nearest Neighbors* (KNN) e *Naive Bayes* (NB) da biblioteca *scikit-learn* foram avaliados. Primeiramente os classificadores foram treinados usando os dados de 2017 a 2022 e o desempenho foi testado usando os dados de 2023. A Tabela 6 mostra o desempenho obtido pela técnica DT considerando a aplicação do classificador à base de 2023. A parte esquerda da tabela mostra os resultados obtidos antes da otimização dos hiperparâmetros, ou seja, usando as configurações padrão do algoritmo. A parte direita da tabela mostra os resultados obtidos após a otimização dos hiperparâmetros. O classificador otimizado obteve uma acurácia de 94%, com um índice de recall de 95% para os usuários “insatisfeitos”, o que mostra que o modelo tem boa capacidade em identificar tais usuários.

Tabela 6. Desempenho da classificação usando a técnica *Decision Tree*

	Antes da Análise dos Hiperparâmetros			Após análise dos Hiperparâmetros		
	Precisão	Recall	F1-score	Precisão	Recall	F1-score
Insatisfeito	0,38	0,48	0,43	0,56	0,95	0,70
Satisfeito	0,96	0,94	0,95	1,00	0,94	0,97
Acurácia	0,90			0,94		

Fonte: Autores.

A Tabela 7 mostra os resultados do algoritmo *Random Forest* (RF), que obteve melhor desempenho que o DT. Após a otimização dos hiperparâmetros, o modelo demonstrou acurácia de 97%, assim como valores elevados de *recall* e *F1-score*, tanto para os usuários satisfeitos e insatisfeitos.

Tabela 7. Desempenho da classificação usando a técnica *Random Forest*

	Antes da Análise dos Hiperparâmetros			Após análise dos Hiperparâmetros		
	Precisão	Recall	F1-score	Precisão	Recall	F1-score
Insatisfeito	0,64	0,98	0,77	0,70	0,99	0,82
Satisfeito	1,00	0,96	0,98	1,00	0,97	0,98
Acurácia	0,96			0,97		

Fonte: Autores.

O algoritmo KNN, cujos resultados são apresentados na Tabela 8, também obteve bons resultados, sendo comparáveis ao RF. A acurácia observada foi de 97%, com altos níveis de *recall* e *F1-score*, o que indica que esse classificador é adequado para identificar usuários satisfeitos e insatisfeitos. Pode-se notar que o KNN teve resultados de *F1-score* levemente superiores ao RF.

Tabela 8. Desempenho da classificação usando a técnica *K-Nearest Neighbors*

	Antes da Análise dos Hiperparâmetros			Após análise dos Hiperparâmetros		
	Precisão	Recall	F1-score	Precisão	Recall	F1-score
Insatisfeito	0,73	0,90	0,80	0,77	0,92	0,84
Satisfeito	0,99	0,97	0,98	0,99	0,98	0,99
Acurácia	0,97			0,97		

Fonte: Autores.

O algoritmo NB também obteve bons resultados, também estando próximo ao desempenho das técnicas RF e KNN. Foram observadas pequenas oscilações de desempenho nos indicadores de recall e F1-score com relação às demais técnicas, mas de forma geral o algoritmo NB também se demonstrou capaz de classificar corretamente usuários satisfeitos e insatisfeitos, com acurácia de 97%.

Tabela 9. Desempenho da classificação usando a técnica *Naïve Bayes*

	Antes da Análise dos Hiperparâmetros			Após análise dos Hiperparâmetros		
	Precisão	Recall	F1-score	Precisão	Recall	F1-score
Insatisfeito	0,68	0,95	0,80	0,72	0,93	0,82
Satisfeito	1,00	0,96	0,98	0,99	0,97	0,98
Acurácia	0,96			0,97		

Fonte: Autores.

Com exceção do algoritmo DT, que teve desempenho inferior, as técnicas KNN, RF e NB tiveram bom desempenho, sendo viável sua aplicação às demais etapas do estudo.

4.2 Importância dos atributos

O algoritmo *Random Forest* permite analisar o grau de importância de cada atributo na classificação das instâncias. Para cada valor é calculada a importância usando números fracionários cuja soma é igual a 1. A Tabela 10 mostra o grau de importância dos 11 atributos de entrada utilizados no estudo.

Tabela 10. Importância dos atributos obtidos a partir da Floresta Aleatória

Ordem	Atributo	Valor
1	CONFORTO ACÚSTICO DO AEROPORTO	0,193704
2	DISPONIBILIDADE DE SANITÁRIOS	0,181366
3	QUANTIDADE E QUALIDADE DE ESTABELECIMENTOS COMERCIAIS	0,099740
4	DISPONIBILIDADE E QUALIDADE DAS INFORMAÇÕES NOS PAINÉIS DE VOO	0,098731
5	QUANTIDADE E QUALIDADE DE LANCHONETES/RESTAURANTES	0,098366
6	LIMPEZA GERAL DO AEROPORTO	0,082555
7	CONFORTO TÉRMICO DO AEROPORTO	0,068635
8	DISPONIBILIDADE DE ASSENTOS NA SALA DE EMBARQUE	0,050749
9	CUSTO-BENEFÍCIO DOS PRODUTOS DE LANCHONETES/RESTAURANTES	0,050231
10	DISPONIBILIDADE DE TOMADAS	0,038661
11	QUALIDADE DA INTERNET / WI-FI	0,037261

Fonte: Autores.

4.3 Discussão

Os valores da Tabela 10 servem como referência para discutir os fatores mais relevantes para a satisfação dos usuários dos aeroportos e podem revelar pontos prioritários para a gestão dos serviços prestados.

O atributo “Conforto Acústico do Aeroporto” foi destacado com o mais relevante na análise, o que indica que os usuários se sentem desconfortáveis com o grau de ruído nos aeroportos.

É possível que a superlotação dos aeroportos de grande porte e a aglomeração dos usuários resulte em níveis maiores de ruído, e conseqüentemente, em menores índices de satisfação.

Também se destaca o atributo “Disponibilidade de Sanitários”, que foi apontado como o segundo mais importante na análise de importância dos atributos. Isso demonstra a importância de a gestão dos aeroportos focar na disponibilidade, limpeza e qualidade dos sanitários.

É importante destacar que atributos como a “Disponibilidade de Tomadas” e a “Qualidade da Internet / Wi-Fi” foram os menos relevantes na classificação da satisfação dos usuários. No caso das tomadas, apesar de sua importância para alguns usuários, é possível que apenas uma quantidade pequena de pessoas realmente se importe com este atributo, não sendo crucial para determinar a satisfação dos usuários. No caso da disponibilidade de internet, uma possível explicação pode estar relacionada ao uso de redes de internet móvel do próprio usuário, principalmente em aeroportos com boa qualidade de sinal de internet móvel 4G/5G. Esta descoberta contrasta com os resultados de Oliveira *et al.* (2023), que encontraram evidências de que avaliações favoráveis dos restaurantes do aeroporto e do serviço de Wi-Fi têm um efeito moderador relevante na relação entre atrasos e satisfação do consumidor.

5. CONCLUSÃO

Este trabalho analisou a base de dados da pesquisa de satisfação dos usuários de aeroportos brasileiros, realizada pela Secretaria de Aviação Civil utilizando técnicas de aprendizado de máquina. Foram testados 4 algoritmos de classificação com otimização de hiperparâmetros e em todos eles foram obtidos valores de acurácia superior a 90%. Dentre as técnicas utilizadas, o algoritmo *Decision Tree* obteve o pior desempenho, enquanto os algoritmos *KNN*, *Random Forest* e *Naive Bayes* obtiveram patamar semelhante de desempenho. Dessa forma, este estudo mostrou a viabilidade do estudo dessa base de dados usando técnicas de aprendizado supervisionado.

A técnica de importância dos atributos do algoritmo *Random Forest* foi utilizada para avaliar os atributos mais relevantes para classificar a satisfação dos usuários, destacando-se como mais relevantes o conforto acústico, a disponibilidade de sanitários e a quantidade e qualidade dos estabelecimentos comerciais. Destaca-se também que os atributos relacionados à disponibilidade de tomadas e qualidade da internet foram os menos relevantes dentre as variáveis selecionadas para este estudo. Este resultado pode indicar aos gestores de aeroportos as áreas prioritárias de melhoria para elevar o grau de satisfação dos usuários.

Com relação às limitações deste estudo, destaca-se os desafios no pré-processamento das bases de dados, uma vez que houve mudanças na estrutura do questionário e nem todos os atributos estavam disponíveis em ambas as versões. Além disso, houve uma quantidade de valores faltantes que precisaram ser imputados para as demais análises. Por fim, a base de dados é desbalanceada, sendo necessário o uso da técnica *undersampling*, o que acaba eliminando da base de treino milhares de instâncias para atingir o balanceamento da base. Apesar disso, os classificadores obtidos obtiveram bom desempenho, o que foi demonstrado por sua validação com os dados de 2023.

Pesquisas futuras podem focar no uso de outros algoritmos de aprendizado de máquina e aprofundar o estudo dos atributos que influenciam a satisfação dos usuários. Além disso, sugere-se que pesquisas futuras utilizem técnicas de aprendizado não supervisionado, tais como técnicas de agrupamento, para identificar novos padrões de satisfação dos usuários de aeroportos. Também se recomenda que futuramente este estudo seja replicado quando novas edições da base de dados sejam publicadas para identificar novas tendências com relação à satisfação em aeroportos.

REFERÊNCIAS

- Alpaydin, E. (2010). *Introduction to Machine Learning* (2nd ed.). Cambridge, MA: MIT Press.
- Brasil. (2023). Portal de Dados Abertos: Indicadores que avaliam a satisfação do passageiro com os processos e serviços aeroportuários a ele oferecidos, coletados nos aeroportos durante a pesquisa (20 aeroportos principais). Recuperado de <https://dados.gov.br/dados/conjuntos-dados/pesquisa-de-satisfacao-do-passageiro-em-aeroportos>. Acesso em: 20/09/2023.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Costa Filho, S. V. S., Arce, J. E., Montañó, R. N. R., & Pelissari, A. L. (2019). Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. *Ciência Florestal*, 29(4), 1501-1515. <https://doi.org/10.5902/1980509828392>
- Cranenburgh, S., Wang, S., Vij, A., Pereira, F., & Walker, J. (2022). Choice modelling in the age of machine learning - Discussion paper. *Journal of Choice Modelling*, 42. <https://doi.org/10.1016/j.jocm.2021.100340>
- Gerón, A. (2019). *Hands-on machine learning with scikit-learn, Keras, and TensorFlow*. O'Reilly.
- Ghate, V. & Hemalatha S. (2023). A comprehensive comparison of machine learning approaches with hyper-parameter tuning for smartphone sensor-based human activity recognition, *Measurement: Sensors*, Volume 30. <https://doi.org/10.1016/j.measen.2023.100925>
- Gupta, G. & Aggarwal, H. (2012). Improving customer relationship management using data mining. *International Journal of Machine Learning and Computing*, 2(6), 874-877. <https://dx.doi.org/10.7763/IJMLC.2012.V2.256>
- Jäger, S., Allhorn, A. & Bießmann, F. (2021). A benchmark for data imputation methods. *frontiers big data*, 4. <https://doi.org/10.3389/fdata.2021.693674>
- Kotler, P. (2019). *Administração de Marketing*. (15th ed). São Paulo: Prentice Hall.
- Ludermir, T. B. (2021). Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. *Inteligência Artificial: Estudos Avançados*. 35(101), 85-94. <https://doi.org/10.1590/s0103-4014.2021.35101.007>
- Liashchynskiy, P., & Liashchynskiy, P. (2019). Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. arXiv. <https://doi.org/10.48550/arXiv.1912.06059>
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- Monard, M. C. & Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes: Fundamentos e aplicações*. Editora Manole.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- Ngai, E. & Wu, Y. (2022). Machine learning in marketing: A literature review, conceptual framework, and research agenda. *Journal of Business Research*, 145. <https://doi.org/10.1016/j.jbusres.2022.02.049>
- Oliveira, A. V. M., Oliveira, B. F., & Vassallo, M. D. (2023). Airport service quality perception and flight delays: Examining the influence of psychosituational latent traits of respondents in passenger satisfaction surveys. *Research in Transportation Economics*, 102. <https://doi.org/10.1016/j.retrec.2023.101371>
- Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825-2830.
- Perifanis, N. A. & Kitsios, F. (2023). *Investigating the influence of artificial intelligence on business value in the digital era of strategy: A Literature review*. information. 14 (2). <https://doi.org/10.3390/info14020085>

Pontes Jr., A. P. & Fagundes, R. A. (2023). Aplicação de técnicas de otimização de hiperparâmetros em modelos de machine learning na tarefa de classificar bons e maus clientes. *XVI CBIC 2023*, Salvador. <https://doi.org/10.21528/CBIC2023-141>

Reichheld, F. (2011). The ultimate question 2.0: How net promoter companies thrive in a customer-driven world. *Harvard Business Review Press*.

Ruiz, E., Yushimito, W.F., Aburto, L., & de la Cruz, R. (2024). Predicting passenger satisfaction in public transportation using machine learning models. *Transportation Research Part A: Policy and Practice*, 181. <https://doi.org/10.1016/j.tra.2024.103995>

Sun, Z., Ying, W., Zhang, W., & Gong, S. (2024). Undersampling method based on minority class density for imbalanced data. *Expert Systems with Applications*, 249. <https://doi.org/10.1016/j.eswa.2024.123328>

Tene, O., Polonetsky, J. (2012). Big Data for All: Privacy and User Control in the Age of Analytics. *Journal of Technology and Intellectual Property*, 11 (5).

Wangkiat, P. & Polprasert, C. (2023). Machine learning approach to predict e-commerce customer satisfaction score, *2023 8th International Conference on Business and Industrial Research (ICBIR)*, Bangkok, Thailand.

<https://doi.org/10.1109/ICBIR57571.2023.10147542>

Witek, P. (2014). *Quantum machine learning: what quantum computing means to data mining*. Academic Press. <https://doi.org/10.1016/C2013-0-19170-2>

Zaghloul, M., Barakat, S., & Rezk, A. (2024). Predicting E-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches. *Journal of Retailing and Consumer Services*, 79. <https://doi.org/10.1016/j.jretconser.2024.103865>

Zhao, Y., Zhu, W., Wei, P., Fang, P., Zhang, X., Yan, N., Liu, W., Zhao, H., & Wu, Q. (2022). Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. *Ecological Indicators*, 135. <https://doi.org/10.1016/j.ecolind.2021.108529>