

Análise e predição de voos no aeroporto de Cascavel-PR utilizando o Algoritmo Random Forest.

¹Darlon Vasata, ² Ricardo da Rosa, ³ João Rafael Quadros dos Santos

¹ Professor Orientador, Instituto Federal do Paraná – darlon.vasata@ifpr.edu.br

² Professor Co-orientador, Instituto Federal do Paraná – ricardo.rosa@ifpr.edu.br

³ Discente, Instituto Federal do Paraná – j.rafael@live.com

RESUMO

Dentre os desafios discutidos nos últimos anos, em que o uso de aprendizado de máquina pode auxiliar na possível solução de problemas computacionais, pode-se destacar as dificuldades encontradas por aeroportos associadas a fatores externos ligados a condições meteorológicas, seja por grandes índices pluviométricos, que podem pôr em risco tanto a decolagem, quanto pouso de aeronaves por conta da quantidade de água na pista de decolagem ou por conta das *Windshears*, contantes de vento ou cisalhamento de vento, que são bruscas mudanças de intensidade e direção do vento que pode afetar também o pouso de aeronaves. O uso de algoritmos de inteligência artificial (IA) e aprendizado de máquina são apontados como ferramentas para auxiliar na possível solução desses impasses. Utilizando o algoritmo Random Forest de classificação de dados, juntamente com bibliotecas desenvolvidas para I.A, dados disponibilizados publicamente pela Agência Nacional de Aviação Civil (ANAC) e informações de previsão do tempo, é proposto um modelo para auxílio nesse problema, que tem por objetivo prever a chance de um voo acontecer de acordo com a previsão do tempo e histórico de voos no aeroporto de Cascavel-PR.

PALAVRAS-CHAVE: Vãos. Pandas. *Machine Learning*.

ABSTRACT

Among the challenges discussed in recent years in which the use of Machine Learning can assist in the possible solution of computational problems, we can highlight the difficulties encountered by airports due to external factors related to weather conditions, either by large rainfall rates that can endanger both the takeoff and landing of aircraft because of the amount of water on the runway or because of Wind Shears or wind contours, which are sudden changes in wind intensity and direction that can also affect the landing of aircraft. The use of A.I. (artificial intelligence) algorithms and machine learning are pointed out as tools to help in the possible solution of these impasses. Using the Random Forest algorithm for data classification, which classifies the data and through a combination of prediction of multiple Decision Trees, which are separated in subdivisions, join votes from several estimators to decide the best final result, along with libraries developed for A.I., data publicly available from ANAC (National Civil Aviation Agency) and weather forecast

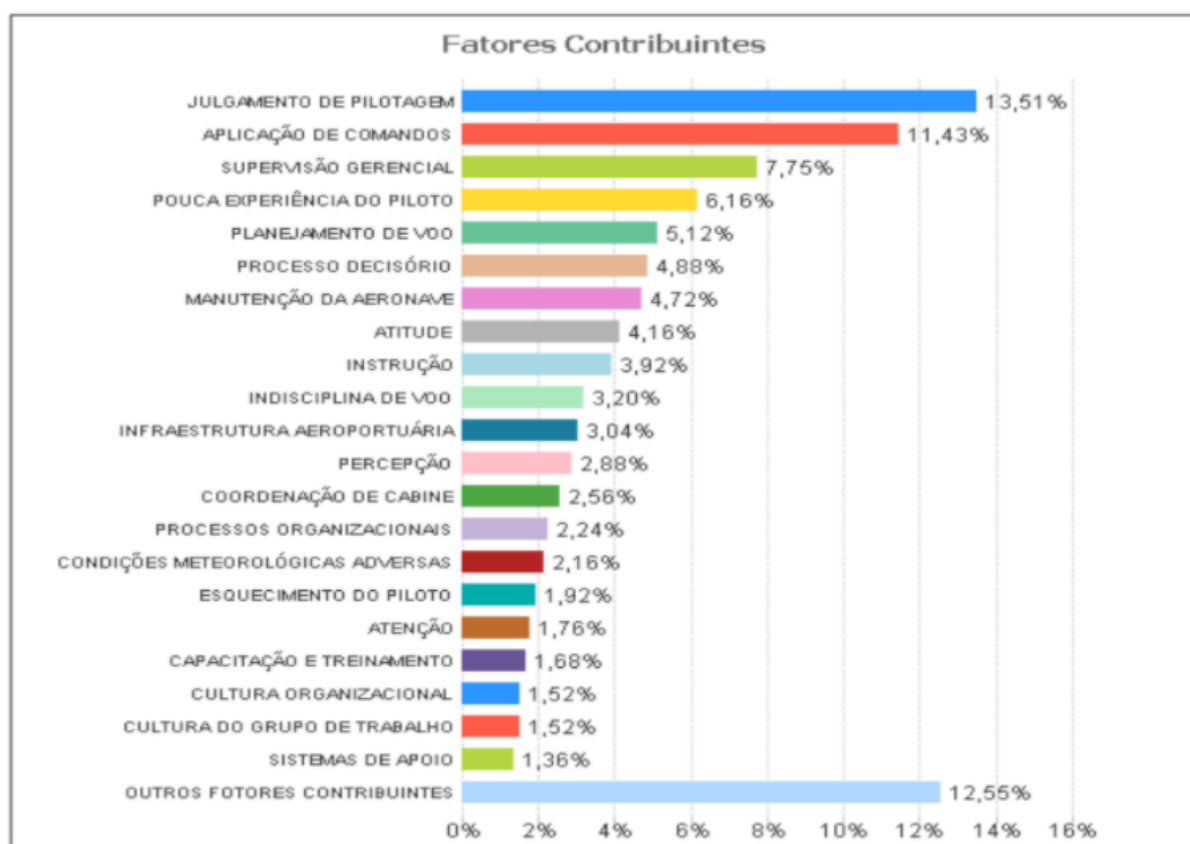
information, it is proposed a model to help in this problem, which aims to predict the chance of a flight according to the weather forecast and flight history at the Cascavel-PR airport.

KEYWORDS: *Flights. Pandas. Machine Learning.*

1. INTRODUÇÃO

É sabido que atualmente existem fatores externos ao controle humano que afetam o setor aéreo e ainda, levando em consideração que esse setor de transporte é formado por ligações entre os aeroportos, quando um desses fatores traz empecilhos para o local do embarque, chegada ou a rota que o avião utiliza, por muitas vezes esse problema não possui formas de ser solucionado de forma imediata, ocorrendo o comprometimento dessa operação e por muitas vezes o cancelamento dos voos. Pesquisas realizadas apontaram que no Brasil, 19.1% dos voos domésticos são cancelados ou sofrem atrasos de mais de 30 minutos segundo Sternberg et al. (2016, p.282), além disso, existem dados do Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA) que exemplificam as principais causas de incidentes aéreos no nosso país, dentre esses, os principais fatores externos que podem causar grande prejuízo nas operações do meio de transporte são as condições meteorológicas adversas, conforme mostrado na Figura 1:

Figura 1. Fatores contribuintes em incidentes aéreos.



Incidência dos fatores contribuintes em acidentes nos últimos 10 anos.

Fonte: Adaptada de Cenipa(2021)

Diversos estudos sobre a incidência de fenômenos meteorológicos e sua influência na aviação foram feitos, como súbitas alterações na intensidade e velocidade do vento que podem pôr em risco atividades aeronáuticas, os cisalhamentos de vento, segundo o Departamento de Controle de Tráfego Aéreo (DECEA) (2013, apud DA SILVA et al., 2021, p.8714) e (MONTEIRO et al., 2005, p.46) além de chuvas muito intensas que podem levar à suspensão dos voos por ameaçar a segurança dos pousos e decolagens na pista, como em Benedetto (2002, p. 667, apud DOS SANTOS et al., 2018, p.46). Visto isso, nota-se que as pesquisas envolvendo fatores climáticos e suas influências na rotina aérea levam em consideração pesquisa e coleta de dados de eventos que já aconteceram e não trazem análise de previsões vindouras.

Por conta disso, esse artigo tem por objetivo trazer uma possível alternativa para auxiliar na solução do problema em questão, a predição de voos saindo do aeroporto de Cascavel-PR utilizando o conceito de aprendizado de máquina com algoritmos de *Random Forest* (Floresta Aleatória) e Árvores de Decisão, sendo este um modelo de predição de dados que mapeia possíveis resultados de uma sequência de escolhas relacionadas, conforme Monard e Baranauskas (2003, p.228) “A Árvore de Decisão é um método adequado quando o objetivo da Mineração de Dados é a classificação de dados ou predição de saídas.” e *Random Forest*, que, segundo apontado por Bhatnagar (2018, apud BRITTO e PÁCIFICO, 2019, p.1086) é um conceito de combinação de predição de múltiplas Árvores de Decisão que são separadas em subdivisões retiradas dos dados treinados e juntam votos de vários estimadores para decidir o resultado final dos dados de teste.

Utilizando esse método, é possível fazer predições de voos e determinar as probabilidades desses serem cancelados de acordo com a previsão meteorológica, levando em consideração dados públicos coletados da Agência Nacional de Aviação Civil (ANAC) e dados de previsão meteorológicas, no Aeroporto de Cascavel/PR.

2. MATERIAIS E MÉTODOS

2.1. Plataforma e recursos

Tendo em vista a necessidade de uma plataforma versátil e de fácil acesso, com capacidade de oferecer ferramentas para utilização de aprendizado de máquina e ciência de dados, foi utilizado o Google Colaboratory (GOOGLE COLABORATORY, 2021), que é uma plataforma *online* que tem como base o Jupyter Notebook e oferece máquinas virtuais em ambiente Linux, sendo esta disponível gratuitamente e voltada para análises que utilizam elementos de inteligência artificial, além de permitir acesso externo e possuir suporte para execução de *scripts* em várias linguagens de programação, principalmente Python. Utilizou-se Python no trabalho, visto que essa linguagem de programação é uma linguagem de altíssimo nível orientada a objeto, de tipagem dinâmica e forte, interpretada e interativa (BORGES, 2014). Considerando essas características, esse recurso pode ser utilizado em aprendizado de máquina (CARNEIRO et al. 2018, p. 61678).

Além disso, foi utilizado a biblioteca Numpy, que possui capacidade para realizar cálculos em *arrays* multidimensionais, possuindo um conjunto de funções que auxiliam a executar esses cálculos, possibilitando desenvolver modelagens e análises mais precisas nos conjuntos de dados trabalhados (NUMPY, 2021). Também se fez necessário o uso da biblioteca Pandas(PANDAS,2021.), que destina-se a incrementar a computação científica com ferramentas que utilizam estatística e linguagem de banco de dados para implementar recursos, como dados automáticos de alinhamento e indexação hierárquica (MCKINNEY, 2011, p.1).

Com essas ferramentas é possível separar todas as informações necessárias em formas tabulares para melhor análise e manejo dos dados para utilização como dados de treinamento e os ajustando a parâmetros para utilização em modelos de aprendizado de máquina .

2.2. Dados dos Voos

Foram utilizados arquivos disponíveis publicamente pela ANAC(AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL, 2021.) contendo dados de voos entre os anos de 2000 e 2020, esses são disponibilizados em formato de arquivos csv estão disponibilizados por mês e ano. Os mesmos foram pré-processados, de maneira que fossem organizados por dia e tabulados, contendo informações sobre pousos e decolagens nos aeroportos do Brasil, bem como informações de voos realizados e cancelados e horários de saída e chegada.

A Figura 2 traz a visualização do *Data Frame* com os primeiros 5 registros da Base de dados, onde os registros das colunas trazem respectivamente as informações de - Empresa aérea, Número de voo, Código de autorização, Código de tipo de linha, Aeroporto de origem, Aeroporto de destino, Partida prevista, Partida real, Chegada prevista, Chegada real, Situação de voo e código da justificativa:

Figura 2. Exemplo da base de dados da Anac.

Número Voo	Código Autorização (DI)	Código Tipo Linha	ICAO Aeródromo Origem	ICAO Aeródromo Destino	Partida Prevista	Partida Real	Chegada Prevista	Chegada Real	Situação Voo
4660	0	R	SBCX	SBCT	01/01/2000 05:45		01/01/2000 06:37		CANCELADO
8710	0	I	MMMX	SBEG	01/01/2000 19:00		02/01/2000 01:25		REALIZADO
8250	0	N	SBBR	SBGO	01/01/2000 20:04		01/01/2000 20:30		REALIZADO
241	9	R	SBSP	SBKP		01/01/2000 22:06		01/01/2000 22:45	REALIZADO
7161	9	R	SBRJ	SBCF		01/01/2000 10:46		01/01/2000 11:35	REALIZADO

Fonte : Próprio Autor

2.4. Tratamento dos dados

Os dados coletados da base da ANAC foram pré-processados e preparados, primeiramente, com auxílio das funções da biblioteca Pandas, foram concatenados todos os arquivos referente aos meses e anos do período compreendido entre 2000 e 2020 e criando um único arquivo. Após isso , foram retiradas algumas informações não relevantes no resultado apontado pelo algoritmo *Random Forest* de maneira a manter em memória apenas os dados necessários para o processamento das tarefas. Será necessário fazer o mesmo procedimento com as informações referente a previsão do tempo para após isso, juntar toda informação coletada em uma única base de dados, que será utilizada para análise com o algoritmo de aprendizado de máquina.

Conforme a Figura 3, os dados foram tratados mantendo as colunas relevantes para a utilização no modelo de predição:

Figura 3. Exemplo da base de dados da Anac após tratamento nos dados.

ICAO	Empresa Aérea	Número Voo	ICAO Aeródromo Origem	ICAO Aeródromo Destino	Dia Partida	Mes Partida	Ano Partida	Hora Partida	Minutos	Dia Chegada	Mes Chegada	Ano Chegada	Hora Chegada	Minutos	Situação Voo
310	ITB	250	SBCA	SBCH	02	01	2000	35.0		02	01	2000		70.0	REALIZADO
1254	ITB	280	SBCA	SBGR	01	01	2000	1210.0		01	01	2000		1327.0	CANCELADO
1370	ITB	230	SBCA	SBCH	01	01	2000	680.0		01	01	2000		715.0	CANCELADO
1371	ITB	260	SBCA	SBCT	01	01	2000	945.0		01	01	2000		1000.0	CANCELADO
1402	RSL	3230	SBCA	SSTD	01	01	2000	685.0		01	01	2000		705.0	CANCELADO

Fonte : Próprio Autor

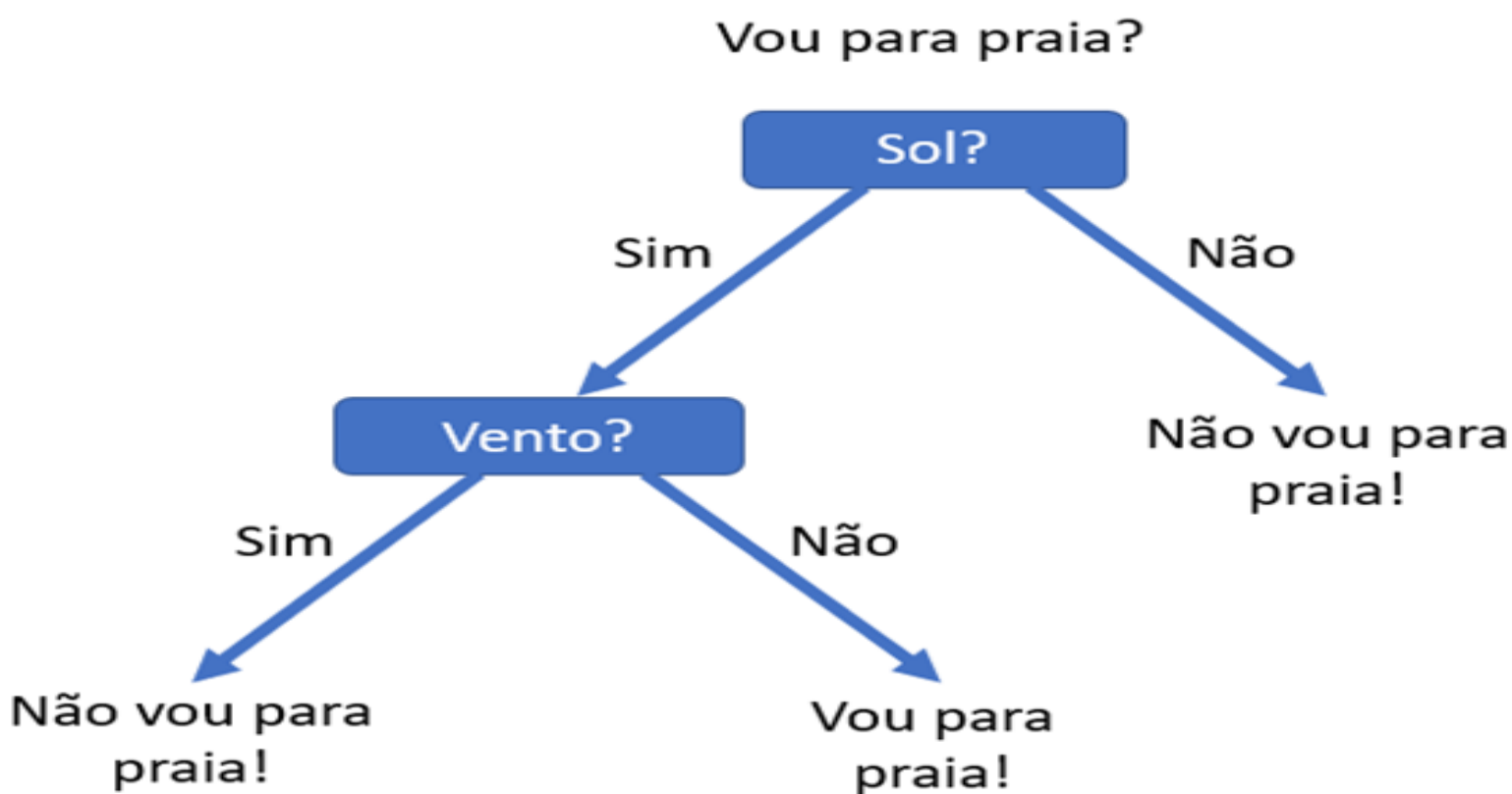
2.5 Árvores de decisão

Utilizando o conceito de dividir-para-conquistar, esse modelo tem por critério a divisão de grandes problemas em subdivisões de forma recursiva, ou seja, a

grande questão principal é dividida em menores situações, que por sua vez, sofrem o mesmo processo. De acordo com Garcia (2000), a estrutura se baseia em uma árvore, onde o problema é dividido em nós (representando os atributos) e ramos (representando os possíveis valores para esses atributos) e ainda, Conforme dito por Kassai et al. (2000, p.113), esse conceito “é utilizada para analisar o processo das decisões por meio de um diagrama, onde é possível visualizar as consequências das decisões atuais e futuras”.

Na Figura 4, está exemplificada uma estrutura de de uma árvore de decisão de classificação, que foi utilizado no desenvolvimento do modelo proposto no artigo, que mede o erro e acerto trabalhando com valores finitos e ordenados, e buscam a quantidade de classificações que diferem do valor proposto para análise. Esse exemplo mostra a tomada de decisão para ir à praia, levando em consideração as informações em cada nó, se existe sol e se existe vento, e nos ramos, os valores possíveis para esses nós. Com essas informações , é possível validar as possibilidades de ir ou não à praia e as possíveis combinações de fatores que influenciam nessa decisão.

Figura 4 - Exemplo de uma Árvore de Decisão de Classificação.



Fonte: Adaptada de Rodrigues(2021)

Para a correta parametrização dos dados e variáveis a serem transportadas para o algoritmo *Random Forest*, antes, foi necessário realizar a construção de uma árvore de decisão para avaliar as variáveis alvo (*Target* ou objetivo a ser atingido) que na abordagem proposta é a informação de cancelamento ou não de voos, variáveis qualitativas(*one hot encoding*), que são as informações das empresas aéreas, aeródromo de destino do voo, mês e dia da partida, e quantitativas, as informações de horário de partida e horário de chegada, conforme visualização da Figura 5 :

Figura 5 - Algoritmo de árvore de decisão.

```
##Implementando Arvore de Decisão !

from sklearn import tree

# Variáveis
alvo = 'Situação Voo' #variável alvo
target = baseFiltro[alvo] #Separa a variável alvo

#Variáveis Dummy
baseFiltroquali = ['ICAO Empresa Aérea', 'ICAO Aeródromo Destino', 'Mes Partida', 'Dia Partida'] #Variáveis qualitativas
dfQuali = pd.get_dummies(baseFiltro[baseFiltroquali])

baseFiltroquant = ['Hora Partida Minutos', 'Hora Chegada Minutos'] #Variáveis quantitativas
dfQuant = baseFiltro[baseFiltroquant] #Dataframe com quantitativas
dfWork = pd.concat([dfQuali, dfQuant], axis=1) #Dataframe com todas variáveis

# Árvore
arv = tree.DecisionTreeClassifier() #árvore de decisão
# arv = tree.DecisionTreeRegressor() #árvore de regressão
arv.fit(dfWork, target) #cria a árvore

# Teste

DecisionTreeClassifier()
```

Fonte: Próprio Autor

2.6. Análise dos dados com Random Forest

Para realizar o objetivo de prever as chances de voo do aeroporto de Cascavel-PR utilizando *Machine Learning*, usaremos um algoritmo de floresta aleatória, um meta estimador que ajusta vários classificadores de árvores de decisão em amostras menores do conjunto de dados e usa a média para melhorar a precisão preditiva (SCI-KIT LEARN, 2021). Já com a base de dados pronta para análise, foi utilizado 70% da base de dados para treinamento do algoritmo (27773 registros) e 30% para teste (11904 registros), e foi apresentado ao modelo variáveis e fatores para que seja possível cancelar o quanto esses influenciam no resultado esperado

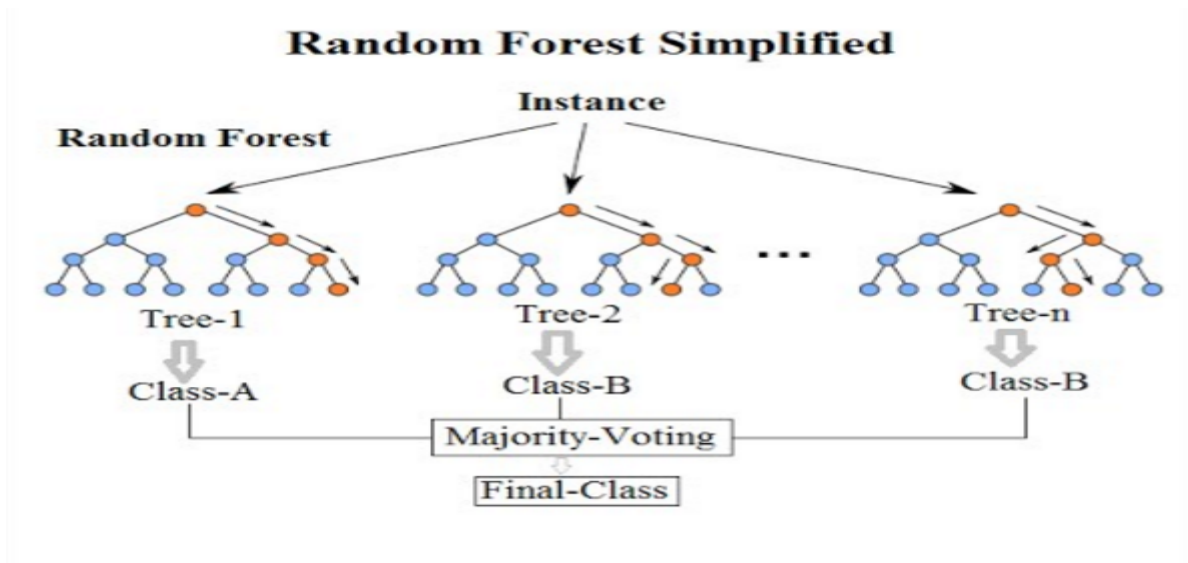
que é a chance do voo ser cancelado de acordo com a previsão meteorológica. Para visualização do desempenho do algoritmo , foi utilizada a matriz de confusão como ferramenta, que é definida como o resultado efetivo de um modelo de classificação analisando os as classificações corretas e preditivas de um conjunto de dados, segundo Monard e Baranauskas (2003, pg. 47), juntamente com um gráfico do tipo *heatmap* (Mapa de calor) utilizando a biblioteca Seaborn, uma biblioteca de estatística para visualização de dados para a linguagem de programação Python (SEABORN, 2021.).

O funcionamento dessa ferramenta de aprendizado de máquina se dá pela combinação de várias árvores de decisões, que pode ser definido conforme Dantas(2015, pg 37):

O método random forest consiste em um conjunto de árvores de decisão geradas dentro de um mesmo objeto. Cada objeto (conjunto de árvores) passa por um mecanismo de votação (*bagging*), que elege a classificação mais votada. A classificação encontra-se nos nós terminais das mesmas

Exemplificando a estrutura de uma *Random Forest* , a Figura 5 traz a união de várias árvores de decisão, que depois de terem realizado as devidas análises, definem a decisão a ser tomada de acordo com os critérios e objetivos definidos, após isso, utilizando o recurso de *bagging*, onde cada árvore define sua decisão, e essa é escolhida com uma votação majoritária e definida como o valor real de tomada de decisão do modelo.

Figura 5 - Exemplo de Random Forest.



Fonte : Adaptada de *Researchgate*(2021)

3. RESULTADOS

Ao inferir sobre o método proposto, verificou-se que após o treinamento dos dados pelo algoritmo de Random Forest, as variáveis que mais afetaram os resultados e as tomadas de decisões do algoritmo foram as quantitativas, hora de partida que obteve um peso de 18,12% nos acertos do algoritmo e hora de chegada, com um valor de 18.33%. As análises também definiram a matriz de confusão para visualização dos resultados, que, apontados pelo algoritmo, foram de 82.48% de acurácia de predição, 85.62% de precisão, que é a quantidade de acertos do algoritmo ao classificar um resultado como correto, 93.17% de *recall score*, que retorna a frequência que o classificador está acertando os valores pertencentes à classe correta (SCI-KIT LEARN, 2021), com esses dados, foi calculado o *F1 score*, que é a combinação dos estimadores de *recall* e precisão, cálculo demonstrado na Figura 6 , que foi de 89.07%.

Figura 6 - Cálculo do F1 Score.

$$F1 = \frac{2 * \text{precisão} * \text{recall}}{\text{precisão} + \text{recall}}$$

Fonte : Adaptada de Rodrigues(2021)

Com esses resultados, foi possível, à partir da visualização da matriz de confusão, valores descritos na Figura 7 :

Figura 7 - Escores e Matriz de confusão.

```

↳ A acurácia foi de 82.48%.
   A precisão foi de 85.62%.
   O Score de Recall foi de 92.81%.

   O Score de F1 foi de 89.07%.

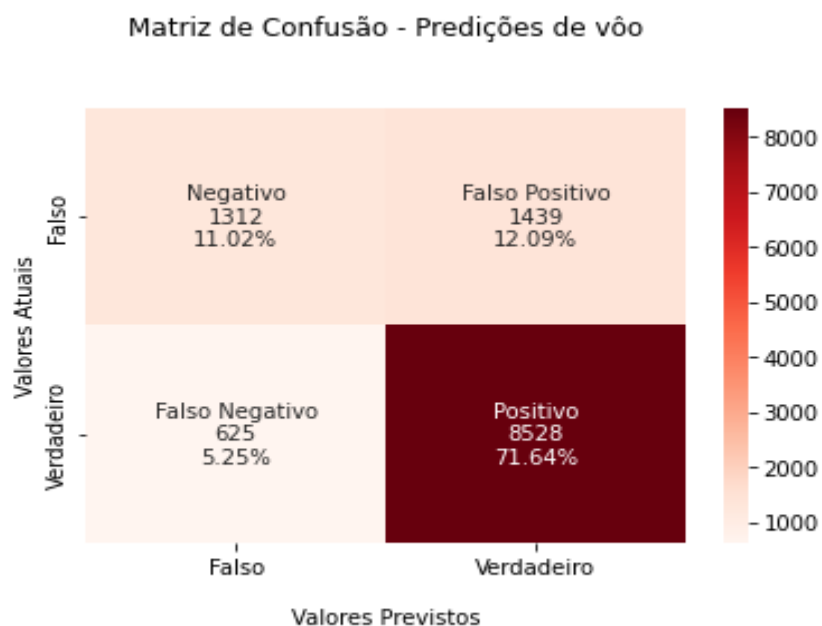
Matriz de confusão :
[[1324 1427]
 [ 658 8495]]

```

Fonte : Próprio Autor

Utilizando *heatmap* como gráfico para *dataviz* (Visualização de dados) da biblioteca Seaborn, gerando assim, a visualização contendo a relação entre os valores preditos e atuais, além da porcentagem de resultados positivos e negativos, e falsos positivos e negativos, e a sua devida quantidade, conforme Figura 8 :

Figura 8 - Visualização dos resultados com matriz de confusão.



Fonte : Próprio Autor

4. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Neste trabalho foi desenvolvido um modelo utilizando *Random Forest*, tendo em vista a sua capacidade de precisão de situações utilizando classificação de dados. A base de dados utilizada foi obtida por meio da disponibilização pública das

informações referentes a voos domésticos no Brasil. O trabalho se ateve a modelar esses dados a fim de utilizar as informações dos voos que saíram do aeroporto de Cascavel-PR no espaço de tempo compreendido entre os anos 2000 e 2020.

O objetivo principal desse modelo era prever a chance de um voo ser cancelado ou não de acordo com as informações da base da ANAC, tendo em vista os problemas causados por condições meteorológicas adversas em aeroportos, que, por conta de serem fatores externos ao controle humano, pode ser difíceis de serem solucionados e contribuem para incidentes aéreos, desde cancelamentos de voos, até atrasos de mais de 30 minutos.

Desse modo, a principal preocupação do trabalho foi no teste do algoritmo e obter seus valores preditivos utilizando a base de dados, contudo, um estudo sobre como o modelo pode se comportar ao analisar além dos dados de voos, dados sobre condições meteorológicas podem contribuir para o possível refinamento dos resultados das previsões realizadas, além de trazer novas perspectivas para os resultados obtidos, como variáveis mais influentes nos resultados.

Finalmente, as diversas vantagens na utilização do algoritmo de aprendizado de máquina pela sua capacidade de previsão mediante a classificação dos dados sugeridos trazem mais força à importância da utilização de métodos de inteligência artificial e é esperado que esse artigo possa servir de incentivo à elaboração de novos estudos. Sugere-se que, assim como os dados da ANAC são disponibilizados publicamente e de fácil acesso, o mesmo acontecesse com dados referentes a previsão do tempo e condições meteorológicas passadas, tendo em vista que, esses dados poderiam refinar a elaboração de modelos como esse e proporcionar o desenvolvimento de estudos nessa área.

5 REFERÊNCIAS

AERÓDROMOS-SUMÁRIO ESTATÍSTICO 2010-2019. Cenipa. 2021. Página Inicial. Disponível em : <https://www2.fab.mil.br/cenipa/index.php/estatisticas>. Acesso em 19 Jul.2021.

BORGES, Luiz Eduardo. Python para desenvolvedores: aborda Python 3.3. Novatec Editora, 2014.

BRITTO, Larissa; PACÍFICO, Luciano. Análise de sentimentos para revisões de aplicativos mobile em português brasileiro. In: Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional. SBC, 2019. p. 1080-1090. Disponível em: <https://sol.sbc.org.br/index.php/eniac/article/view/9359/9261> Acesso em: 04 Jul.2021.

CARNEIRO, Tiago et al. Performance analysis of google colab as a tool for accelerating deep learning applications. IEEE Access, v. 6, p. 61677-61685, 2018. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8485684> Acesso em: 15 Jul.2021.

COLABORATORY. Google Colaboratory. 2021. Página Inicial. Disponível em : <https://colab.research.google.com/> Acesso em: 15 Jul.2021.

COMO FUNCIONA O ALGORITMO DE ÁRVORE DE DECISÃO. Didatica.tech. 2021. Página Inicial. Disponível em : <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao> Acesso em: 12 Nov.2021.

DA SILVA, Letícia Pereira et al. Análise das condições atmosféricas na ocorrência de Windshear no aeroporto internacional de Belém: um estudo de caso (XX CBMET-2018). Brazilian Journal of Development, v. 7, n. 1, p. 8712-8721, 2021. Disponível em: <https://www.brazilianjournals.com/index.php/BRJD/article/view/23601> Acesso em: 02 Jul.2021.

DADOS ABERTOS. Agência Nacional de Aviação Civil. 2021. Página Inicial. Disponível em : <https://www.anac.gov.br/aceso-a-informacao/dados-abertos> Acesso em: 15 Jul.2021.

DANTAS, Guilherme Vieira. Utilização de classificador random forest na detecção de falhas em máquinas rotativas. Unpublished data, 2015. Disponível em : <http://repositorio.poli.ufrj.br/monografias/monopoli10015019.pdf> Acesso em 12 Nov.2021.

DOS SANTOS, Thiago Adriano et al. Modelo de identificação do impacto futuro de chuvas extremas nos atrasos/cancelamentos de voos. TRANSPORTES, v. 26, n. 2, p. 44-53, 2018. Disponível em: <https://www.revistatransportes.org.br/anpet/article/view/1379/693> Acesso em: 02 Jul.2021.

GARCIA, S.C. O uso de árvores de decisão na descoberta de conhecimento na área da saúde. In: SEMANA ACADÊMICA, 2000. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul, 2000. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/4703/000503532.pdf?sequence=1> Acesso em 12 Nov.2021.

LEMONS, Eliane Prezepiorski; STEINER, Maria Teresinha Arns; NIEVOLA, Julio César. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. Revista de Administração-RAUSP, v. 40, n. 3, p. 225-234, 2005. Disponível em : <https://www.redalyc.org/pdf/2234/223417392002.pdf> Acesso em 12 Nov.2021.

KASSAI, José Roberto et al. Retorno de investimento: abordagem matemática e contábil do lucro empresarial. São Paulo: Atlas, v. 2, 2000.

MACHINE LEARNING, O QUE SÃO ACCURACY, PRECISION, RECALL E F1 SCORE. vilson rodrigues.medium. 2021. Página Inicial. Disponível em : <https://vilsonrodrigues.medium.com/machine-learning-o-que-s%C3%A3o-accuracy-precision-recall-e-f1-score-f16762f165b0> Acesso em: 12 Nov.2021.

MCKINNEY, Wes et al. pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing, v. 14, n. 9, p. 1-9, 2011. Disponível em: https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf Acesso em 16 Jul.2021.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Indução de regras e árvores de decisão. Sistemas Inteligentes-Fundamentos e Aplicações, v. 1, p. 115-139, 2003. Disponível

em: <https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap5.pdf>
Acesso em: 07 Jul.2021.

MONTEIRO, Maurici A. et al. OS PERIGOS DO VENTO PARA A AVIAÇÃO. A Meteorologia e a Aeronáutica. Boletim da Sociedade Brasileira de Meteorologia, v.29, n. 3, p 45-51, 2005. Disponível em :
https://www.researchgate.net/profile/Jaci-Saraiva-2/publication/292146520_O_PROJETO_SIVAM_O_SIPAM_E_A_SUA_CONTRIBUICAO_PARA_A_SEGURANCA_DA_NAVEGACAO_AEREA_NA_AMAZONIA_LEGAL/links/56aa5e6e08aed814bde698cb/O-PROJETO-SIVAM-O-SIPAM-E-A-SUA-CONTRIBUICAO-PARA-A-SEGURANCA-DA-NAVEGACAO-AEREA-NA-AMAZONIA-LEGAL.pdf#page=47 Acesso em: 02 Jul.2021.

PANDAS. Pandas. 2021. Página Inicial. Disponível em : <https://pandas.pydata.org/> Acesso em : 15 Jul.2021.

PRINA, Bruno Zucuni; TRENTIN, Romario. GMC: Geração de Matriz de Confusão a partir de uma classificação digital de imagem do ArcGIS®. Anais do XVII Simpósio Brasileiro de Sensoriamento Remoto–SBSR, p. 137, 2015.

RANDOM FOREST CLASSIFIER. scikit-learn. 2021. Página inicial. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> Acesso em: 15 Jul.2021.

RANDOM FOREST SIMPLIFIED REPRESENTATION?. Researchgate. 2021. Página Inicial. Disponível em :
https://www.researchgate.net/figure/Random-forest-simplified-representation_fig3_333152684
Acesso em: 12 Nov.2021.

SEABORN : STATISTICAL DATA VISUALIZATION. Seaborn. 2021. Página Inicial. Disponível em: <https://seaborn.pydata.org/> Acesso em: 15 Jul.2021.

STERNBERG, Alice et al. An analysis of Brazilian flight delays based on frequent patterns. Transportation Research Part E: Logistics and Transportation Review, v. 95, p. 282-298, 2016. Disponível em:
https://www.sciencedirect.com/science/article/pii/S1366554516301740?casa_token=URgbC5dl_iAAAAA:sCGGXn7HdE75SMLtJFX9awdGO2Ygl8Ntk3wbaWuWpWBucyShm8ZGsB-hEd_ZBb-0Zdj-pHWnCQcr Acesso em: 07.Jul.2021.

WHAT IS NUMPY?. Numpy. 2021. Página Inicial. Disponível em :
<https://numpy.org/doc/stable/user/whatisnumpy.html> Acesso em: 15 Jul.2021.