



UNIVERSIDADE FEDERAL DO PERNAMBUCO
CENTRO DE TECNOLOGIAS E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

HÉLIO DA SILVA QUEIROZ JÚNIOR

**PREVISÃO DA OCORRÊNCIA DE ATRASOS EM DECOLAGENS
DO AEROPORTO DE GUARULHOS**

Recife

2022

HÉLIO DA SILVA QUEIROZ JÚNIOR

**PREVISÃO DA OCORRÊNCIA DE ATRASOS EM DECOLAGENS
DO AEROPORTO DE GUARULHOS**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Civil da Universidade Federal de Pernambuco – UFPE, como requisito parcial para obtenção do título de mestre em Engenharia Civil.

Área de concentração: Transportes.

Orientadora: Prof^ª. Dr^ª. Viviane Adriano Falcão

Recife

2022

Q3p

Queiroz Júnior, Hélio da Silva.

Previsão da ocorrência de atrasos em decolagens do Aeroporto de Guarulhos / Hélio da Silva Queiroz Júnior. 2022.

94 f: figs., tabs.

Orientadora: Profa. Dra. Viviane Adriano Falcão.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG.
Programa de Pós-Graduação em Engenharia Civil, Recife, 2022.

Inclui referências.

1. Engenharia civil. 2. Atraso de voo. 3. Previsão. 4. Metanálise. 5. Machine learning. 6. RNA. I. Falcão, Viviane Adriano (Orientadora). II. Título.

UFPE

624 CDD (22. ed.)

BCTG / 2023 - 13

HÉLIO DA SILVA QUEIROZ JÚNIOR

**PREVISÃO DA OCORRÊNCIA DE ATRASOS EM DECOLAGENS
DO AEROPORTO DE GUARULHOS**

Dissertação em Engenharia Civil da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito para obtenção do título de Mestre em Engenharia Civil, Área de Concentração Transporte e Gestão das Infraestruturas Urbanas.

Aprovada em 15/12/2022

Orientadora: Prof.^a Dr.^a Viviane Adriano Falcão.

BANCA EXAMINADORA

participação por videoconferência
Prof. Dr. Enilson Medeiros dos Santos (examinador interno)
Universidade Federal do Rio Grande do Norte

participação por videoconferência
Prof. Dr. Evandro José da Silva (examinador externo)
Instituto Tecnológico de Aeronáutica

participação por videoconferência
Prof. Dr. Francisco Gildemir Ferreira da Silva (examinador externo)
Universidade Federal do Ceará

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

Um problema comum em aeroportos de todo o mundo são os atrasos em voos comerciais. A crescente demanda por transporte aéreo torna esses atrasos cada vez mais recorrentes, somando custos e exigindo ajustes constantes na gestão de voos. Definir o método mais eficaz para prever a ocorrência desses atrasos é um tema recorrente em pesquisas de operação de tráfego aéreo. Nesses estudos, a especificidade da área analisada (seja uma companhia aérea, um aeroporto ou toda a operação de um país), a complexidade da saída buscada (previsão por regressão ou classificação) ou o tamanho do banco de dados utilizado exigem métodos de análise mais robustos, sendo o aprendizado de máquina uma alternativa comum ao uso de métodos estatísticos clássicos. Dentre os modelos de previsão por aprendizado de máquina, as Redes Neurais Artificiais se destacam pela capacidade modular do método, permitindo uma adequação a finalidade aplicada. Entretanto, a divergência entre os cenários estudados atribui diferentes verificações de precisão entre as respostas obtidas nos estudos realizados. Assim, este estudo tem como objetivo definir o cenário adequado de previsão e classificação de atrasos (área de análise, as variáveis de influência nos atrasos, uma faixa de acurácia aceitável e o modelo de previsão mais eficaz) em partidas regulares de um aeroporto. Além de propor um sistema preditivo em classes, capaz de ser implementado a realidade de operação do Aeroporto Internacional de São Paulo – SBGR e da companhia aérea nacional Gol Linhas Aéreas - GLO. Portanto, métodos de revisão bibliométrica e meta-análise foram utilizados para definir uma referência de base para avaliar a acurácia dos métodos. Em seguida, os estudos pesquisados foram classificados através da Análise Envoltória de Dados por metafronteira, definindo os cenários ideais para uma melhor resposta de previsão. As análises indicam que a Rede Neural do tipo *MultiLayer Perceptron* (MLP), têm melhor eficácia nas respostas preditivas para análises em rotas ou companhias aéreas, independentemente do motivo do atraso. Assim, foi aplicado um estudo de caso comparativo para o cenário de previsão definido, sendo verificada a capacidade preditiva do modelo para o Aeroporto de Guarulhos, SBGR, e uma companhia aérea nacional operante no mesmo. Os resultados indicam que, a previsão de atrasos em decolagens através de um sistema classificatório do tempo de atraso se mostrou o método com maior acurácia, independente do motivo de ocorrência. Logo, constatou-se que, a especificidade dos cenários e os motivos dos atrasos, dada uma área de análise, possuem maior relação com os resultados de acurácia de uma previsão de atraso de voo do que a massa de dados obtidos.

Palavras-chave: atraso de voo; previsão; metanálise; *machine learning*; RNA.

ABSTRACT

A widespread problem at airports around the world is delays on commercial flights. The growing demand for air transport makes these delays increasingly recurrent, adding costs and requiring constant adjustments in flight management. Defining the most effective method for predicting the occurrence of these delays is a recurring theme in air traffic operations searches. In these studies, the specificity of the analyzed area (be it an airline, an airport, or the entire operation of a country), the complexity of the output sought (regression or classification prediction) or the size of the database used require more robust analysis methods, and machine learning is a common alternative to the use of classical statistical methods. Among the models of prediction by machine learning, the Artificial Neural Networks stand out for the modular capacity of the method, allowing an adaptation to the applied purpose. However, the divergence between the scenarios studied attributes different precision checks between the responses obtained in the studies performed. Thus, this study aims to define the appropriate scenario of forecasting and classification of delays (area of analysis, parameters of influence on delays, an acceptable accuracy range and the most effective forecasting model) in regular departures from an airport. In addition to proposing a predictive system in classes, capable of implementing the operation reality of the São Paulo International Airport - SBGR and the national airline Gol Linhas Aéreas - GLO. Therefore, methods of bibliometric review and meta-analysis were used to define a baseline reference to assess the accuracy of the methods. Then, the studies were classified through metafrontier Data Envelopment Analysis, defining the ideal scenarios for a better prediction response. The analyses indicate that the *Perceptron MultiLayer* Neural Network (MLP) has better efficacy in predictive responses for analysis on routes or airlines, regardless of the reason for the delay. Thus, a comparative case study was applied to the defined forecast scenario, verifying the predictive capacity of the model for Guarulhos Airport, SBGR, and a national airline operating in it. The results indicate that the prediction of delays in takeoffs through a system classificatory of the delay time proved to be the most accuracy dutily method, regardless of the reason for occurrence. Therefore, it was found that the specificity of the scenarios and the reasons for the delays, given an area of analise, are more related to the results of accuracy of a flight delay forecast than the mass of data obtained.

Keywords: flight delay; forecast; meta-analysis; *machine learning*; RNA.

LISTA DE FIGURAS

Figura 1: Fronteiras Técnicas dos Clusters K e a Metafronteira Mm'.	28
Figura 2 - Fronteiras Técnicas por K Clusters e a Metafronteira do Tipo Côncava Formada.	29
Figura 3 - Um Modelo Geral do Processo de Aprendizagem.	31
Figura 4 - Ilustração de uma Rede Neural Artificial <i>Perceptron</i> .	34
Figura 5 - Fases <i>forward</i> e <i>backward</i> do Algoritmo <i>Backpropagation</i> .	36
Figura 6 - Pousos e decolagens em GRU (2017-2021).	38
Figura 7 - Quantidade de Assentos Registrados em GRU (2017-2021).	39
Figura 8 - Relação entre os IPCAs no Brasil e em São Paulo com o número de voos em GRU em 2019, a) ano anterior aos efeitos da pandemia por SARS-COVID-19 iniciarem no país, em 2020, b) durante o início e pico de casos da pandemia, e em 2021, c) início da recuperação do modo aéreo pela diminuição no número de casos.	40
Figura 9 - Ocorrências de atrasos em SBGR para 2019. A) Voos atrasados totais (Domésticos e Internacionais), B) Voos Domésticos atrasados, C) Voos Internacionais atrasados.	42
Figura 10 - Distribuição dos atrasos ao longo do dia de maior pico em SBGR para 2019	44
Figura 11 - Distribuição percentual dos voos por Companhia Aérea em SBGR para 2019. a) Em relação aos cancelamentos, b) aos atrasos superiores à 30 minutos e c) aos atrasos superiores à 60 minutos	46
Figura 12 - Etapas de seleção dos artigos analisados	51
Figura 13 - Organização das lógicas de previsão aplicadas no estudo	57
Figura 14 - Lógica de união dos dados aplicada aos dados para construção do banco de informações para a pesquisa	60
Figura 15 - Dispersão dos dados levantados: análise espacial por classe IATA/ANAC dos atrasos	61
Figura 16 - Modelo de arquitetura aplicadas as RNAs.	62
Figura 17 - Resumo da metodologia aplicada na segunda parte da pesquisa	63
Figura 18 - Nuvem de palavras com as palavras-chave de artigos preliminares selecionados	65
Figura 19 - Cenário A - cenário de regressões por predição classificatória binária (0 = voo on-time, 1 = voo atrasado).	67
Figura 20 - Cenário B - cenário de regressão dos atrasos por classes definidas.	68
Figura 21 - Gráfico de túnel, método de Freeman-Turkey.	69
Figura 22 - Proporções de acurácia e tamanho da amostra tendo como subgrupos os países estudados pelos artigos do cenário C	70

Figura 23 - Proporções de acurácia e tamanho da amostra tendo como subgrupos os países estudados pelos artigos do cenário D.....	71
Figura 24 - Causas dos atrasos da GLO em SBGR, ano 2019.....	78
Figura 25 - Relação entre o tempo de atraso registrado e as decolagens, a), e pousos simultâneos registrados no intervalo de hora.....	80
Figura 26 - Classificação dos atrasos em SBGR por grupo IATA/ANAC.....	80
Figura 27 - Classificação dos atrasos em SBGR por grupo IATA/ANAC, exceto atrasos iguais ou inferiores a 15 minutos.....	81
Figura 28 - Balanceamento das instâncias para a segunda lógica de previsão.....	82
Figura 29 - Proporção de distribuição dos dados, após o pré-processamento.....	83
Figura 30 - Balanceamento das instâncias para a primeira lógica de previsão.....	84

LISTA DE TABELAS

Tabela 1 - Códigos a serem usados nas mensagens de movimento e desvio de aeronaves.	16
Tabela 2 - Estudos de Previsão de Atrasos em Aeroportos.	19
Tabela 3 - Grupos de Variáveis utilizadas pelos Estudos Analisados e os seus respectivos Parâmetros em Comum.	20
Tabela 4 - Atrasos semanais registrados em SBGR para janeiro de 2019.	43
Tabela 5 - Atrasos diários registrados em SBGR para a 4ª semana de janeiro de 2019.	43
Tabela 6 - Companhias Aéreas em atividade em SBGR em 2019.	44
Tabela 7 - Grupos de palavras-chave aplicadas na pesquisa nas bases de pesquisas.	49
Tabela 8 - Clusters de efeito considerados para a metafronteira.	53
Tabela 9 - Cenários de metafronteira aplicados.	54
Tabela 10 - Grupos de atrasos por intervalos, uso da segunda lógica de previsão.	56
Tabela 11 - Conjunto de variáveis aplicadas no estudo de previsão.	58
Tabela 12 - Arquiteturas aplicadas no treino das RNAs.	61
Tabela 13 - Estatísticas principais dos estudos levantados para o Grupo 02. Análise por classe de cluster.	67
Tabela 14 - Cenário C - acurácia dos modelos de previsão por classificação binária para caracterizar os estudos.	70
Tabela 15 - Cenário D - análise de subgrupo com a área de estudo.	70
Tabela 16 - Classe de cluster a qual cada estudo analisado pertence. Artigos dos grupos C e D.	72
Tabela 17 - Resultados dos cenários de DEA Metafronteira. Para cada cenário são apresentadas as eficiências por subgrupo $D_k(x,y)$, a eficiência para o modelo DEA total $D(x,y)$ e a razão de meta-eficiência $MRT_k(x,y)$	74
Tabela 18 - Ranqueamento dos principais estudos que atendem as condições de desempenho demandadas.	76
Tabela 19 - Ocorrência de atrasos de voos por companhia aérea em SBGR, ano 2019.	77
Tabela 20 - Estatísticas descritivas das variáveis explicativas selecionadas.	79
Tabela 21 - Divisão de instâncias para treino e teste.	84
Tabela 22 - Resumo dos resultados obtidos pelas RNAs testadas, para performance do treino e acurácia do teste.	86

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	Objetivos.....	14
<i>1.1.1</i>	<i>Objetivos Específicos.....</i>	<i>14</i>
2	PRINCIPAIS CAUSAS DE ATRASOS DE VOOS.....	15
2.1	Previsibilidade e informações dos voos: dados conhecidos previamente	17
2.2	Variáveis de influência nos atrasos	18
3	CENÁRIOS APLICADOS PARA A PREVISÃO DE ATRASOS.....	23
3.1	Meta-Análises	23
<i>3.1.1</i>	<i>Meta-Análises aplicadas a estudos de transportes.....</i>	<i>25</i>
3.2	Análise Envoltória de Dados – DEA	26
4	MÉTODOS DE APRENDIZADO DE MÁQUINA – MACHINE LEARNING	31
4.1	Redes Neurais Artificiais - RNA.....	33
<i>4.1.1</i>	<i>RNAs aplicadas em estudos de previsão de atrasos de voos.....</i>	<i>37</i>
5	ESTUDO DE CASO: AEROPORTO INTERNACIONAL DE SÃO PAULO	38
5.1	Análise de atrasos em SBGR	41
6	METODOLOGIA.....	48
6.1	Primeira parte: Definição do cenário ideal de previsão.....	48
<i>6.1.1</i>	<i>1ª Etapa: Revisão Sistemática</i>	<i>49</i>
<i>6.1.2</i>	<i>2ª Etapa: Meta-Análise.....</i>	<i>50</i>
<i>6.1.3</i>	<i>3ª Etapa: Análise Envoltória de Dados - DEA.....</i>	<i>52</i>
6.2	Segunda parte: Aplicação do cenário de previsão definido em um modelo de previsão em classes	54
<i>6.2.1</i>	<i>Etapa preliminar 0: Definição das áreas de análise.....</i>	<i>55</i>
<i>6.2.2</i>	<i>1ª Etapa: Definição da lógica de previsão</i>	<i>55</i>
<i>6.2.3</i>	<i>2ª Etapa: Seleção das variáveis para o modelo</i>	<i>57</i>
<i>6.2.4</i>	<i>3ª Etapa: Formação do banco de dados.....</i>	<i>59</i>
<i>6.2.5</i>	<i>4ª Etapa: Pré-processamento dos dados</i>	<i>60</i>
<i>6.2.6</i>	<i>5ª Etapa: Treino e definição das arquiteturas.....</i>	<i>61</i>
<i>6.2.7</i>	<i>6ª Etapa: Teste da saída e análise dos resultados.....</i>	<i>62</i>

7	RESULTADOS E DISCUSSÕES	64
7.1	Primeira parte: Definição do cenário de previsão ideal	64
7.1.1	<i>Revisão Sistemática de Literatura</i>	64
7.1.2	<i>Acurácia padrão dos métodos de machine learning</i>	66
7.1.3	<i>Fatores de influência a acurácia</i>	69
7.1.4	<i>Seleção do método mais eficaz por cenário</i>	71
7.2	Segunda parte: Aplicação do cenário de previsão definido em um modelo de previsão em classes	77
7.2.1	<i>Etapa preliminar 0 e 1ª Etapa: Definição das áreas de análise e definição da lógica de previsão</i>	77
7.2.2	<i>2ª e 3ª Etapas: Seleção das variáveis para o modelo e Formação do banco de dados</i>	78
7.2.3	<i>3ª e 4ª Etapas: Formação do banco de dados e Pré-processamento dos Dados</i> 81	
7.2.4	<i>5ª e 6ª Etapas: Treino e definição das arquiteturas e Teste da saída e análise dos resultados</i>	84
8	CONCLUSÕES	89
	REFERÊNCIAS	91

1 INTRODUÇÃO

Uma problemática comum nos aeroportos em todo o mundo são os atrasos de voos comerciais (WANG & WANG, 2019). A crescente demanda pelo modal aéreo faz com que estes atrasos (*delays*) sejam cada vez mais recorrentes, ocasionando custos adicionais e exigindo ajustes constantes no gerenciamento dos voos (LIU, YIN & HANSEN, 2019).

Na Europa, em média 24,2% dos voos atrasam, segundo relatório anual da Organização Europeia para a Segurança da Navegação Aérea para o ano de 2019 (EUROCONTROL, 2019). Nos Estados Unidos, o custo estimado para as companhias aéreas americanas por minuto de atraso foi de US\$ 68,48, em 2017 (A4A, 2017).

Para o Brasil, segundo informações disponibilizadas pela Agência Nacional de Aviação Civil – ANAC, para o ano de 2019, aproximadamente metade dos voos brasileiros atrasaram (ANAC, 2019).

A ANAC disponibiliza ainda dados sobre voos e operações aéreas no país, dentre eles os percentuais de atrasos e de cancelamentos de voos consolidados por par de aeroportos de origem e de destino. Assim é possível observar que o Brasil concentra uma média anual de atrasos de 6,97% superiores a trinta minutos e 2,49% superiores a uma hora para um conjunto de aproximadamente cinquenta e quatro etapas de voos previstas, correspondendo a em torno de 10 a cada 100 voos com atrasos sendo maiores a meia hora (ANAC, 2021). Estima-se para as companhias uma perda de aproximadamente R\$10.670 em cada voo com mais de trinta minutos de atraso, com base nos valores americanos à uma cotação de R\$5,21 para o dólar.

Para o aeroporto nacional com a maior movimentação anual de passageiros e carga, o Aeroporto Internacional de Guarulhos, SGBR, uma média mensal de atrasos entre pousos e decolagens realizados atinge o valor de 14% (ANAC, 2022).

A aquisição de dados facilitada pela disponibilidade de novas e acessíveis tecnologias favorece a obtenção de informações úteis à aplicação em estudos de aviação em diversas áreas críticas, como em casos de análise de big data, gerenciamento de rede de transporte aéreo ou em previsão dos possíveis atrasos de voos utilizando para isso os métodos de aprendizado de máquina- *machine learning* (CHU *et al.*, 2020).

Logo, devido ao impacto gerado em toda logística operacional do tráfego aéreo, além dos custos atrelados às ocorrências, a previsão de atrasos permite, por exemplo, para uma companhia aérea planejar algumas outras alternativas, com outra rota possível, índice de custo mais alto etc., (KHAN *et al.*, 2021).

Assim, a definição de metodologias que consigam prever a ocorrência destes atrasos em voos comerciais, utilizando bancos de dados reais disponibilizados pelas companhias aéreas ou pelas agências nacionais de aviação, são objetos de pesquisa recorrentes. Contudo, a complexidade dos dados do setor aéreo, aliada à diversidade dos modelos disponíveis para realizar as previsões formam a principal barreira na definição do cenário com maior acurácia nas respostas obtidas.

Com isso, algumas barreiras são encontradas quando se busca estabelecer padrões que sirvam como base a aplicabilidade destes métodos de previsão diretamente no mercado, seja por uma companhia aérea ou ao analisar as operações de um aeroporto completo.

A área de análise, as variáveis de influência nos atrasos, uma faixa de acurácia aceitável ou o modelo de previsão mais eficaz, compõem os principais pontos a serem determinados ao se estabelecer um cenário adequado ao estudo de previsão.

Dentre os métodos utilizados atualmente, o aprendizado de máquina (*machine learning*) serve como uma ferramenta de análise comum, seja para verificar a influência de fatores meteorológicos em atrasos nos aeroportos chineses, CHEN *et al.* (2021), ou até prevendo a trajetória terrestre de uma aeronave em aeroportos americanos (SCHULTZ & REITMANN, 2019).

Diante disto, este trabalho foi desenvolvido em duas partes principais, na primeira parte foram analisados os distintos modelos de metodologias e métodos de previsão de atrasos com a finalidade de definir o cenário padrão de previsibilidade necessária para um modelo ser julgado razoável ou bom para tomada de decisão.

Para isso, uma proposta metodológica foi desenvolvida, na qual foi realizada uma revisão bibliométrica para compreender o comportamento das pesquisas sobre o tema, a fim de que se levantasse as principais técnicas aplicadas. Em seguida foi aplicada uma meta-análise para definir um valor padrão de acurácia a ser seguido pelo mercado na previsão de atrasos. Por fim foram definidas as variáveis de influência e métodos de *machine learning* mais eficientes para a realização de previsões de atrasos de voos, através do ranqueamento dos trabalhos por uma Análise Envoltória de Dados - DEA.

A segunda parte da pesquisa realizou um estudo de caso para estimar os possíveis atrasos em decolagens regulares no SBGR. As respostas para o maior aeroporto nacional em movimentações anuais de passageiros, carga e aeronaves caracterizam a eficácia do cenário obtido para o modelo operacional brasileiro. Para tanto, foi aplicado o cenário ideal de previsões definidos na primeira etapa, com o intuito de obter a maior acurácia possível e que melhor representasse a situação real do aeroporto. Portanto, além de

apontar os fatores que mais influenciaram as suas ocorrências, permitindo um ajuste do controle de tráfego aéreo, reduzindo os custos e a insatisfação gerada aos usuários do modal, a metodologia de definição de cenários pode ser testada.

1.1 Objetivos

O objetivo deste trabalho é definir o cenário adequado de previsão e classificação de atrasos (área de análise, as variáveis de influência nos atrasos, uma faixa de acurácia aceitável e o modelo de previsão mais eficaz) em partidas regulares de um aeroporto. Além de propor um sistema preditivo em classes, capaz de ser implementado a realidade de operação do Aeroporto Internacional de São Paulo – SBGR e da companhia aérea nacional Gol Linhas Aéreas - GLO.

1.1.1 Objetivos Específicos

Dentre os objetivos específicos da pesquisa, se destacam:

- Compreender a tendência de métodos de previsão aplicados para atrasos de voos;
- Identificar a relação entre a acurácia da análise preditiva com o cenário de estudo selecionado;
- Verificar os principais fatores de influência dos atrasos numa análise em grupos;
- Estimar a capacidade preditiva com base em informações antecipadas ao horário estimado de cada voo.

2 PRINCIPAIS CAUSAS DE ATRASOS DE VOOS

Para caracterizar a causa dos atrasos em um aeroporto é preciso, primeiro, entender a situação atual do mercado aéreo, quanto à capacidade diante da movimentação existente, o crescimento esperado de demanda e as possíveis consequências disto, como o aumento da frequência de atrasos dos voos.

Além disso, informações específicas à funcionalidade de cada aeroporto devem ser incorporadas, assim como os outros fatores-chave do sistema de aviação, como atrasos nas filas e atrasos em rota no modelo, uma vez que esses dados estejam disponíveis, como ressaltam Chen *et al.* (2021).

As operações terrestres de um aeroporto são também evidenciadas por Schultz & Reitmann (2019) quanto o atraso, uma vez que as incertezas na fase aerotransportada têm um impacto significativamente menor na pontualidade do voo do que os desvios nas operações terrestres da própria aeronave.

Por outro lado, apesar de serem problemas críticos para as companhias aéreas, o planejamento das operações dos voos atribui condições favoráveis a atrasos a depender do tipo de voo ofertado, uma vez que, em comparação com os voos domésticos, os voos internacionais envolvem vários requisitos adicionais e podem enfrentar problemas, como operações terrestres complexas e operações de rota (KHAN *et al.*, 2021).

Além dos fatores de funcionalidade e operações aéreas, condições climáticas e de sazonalidade podem interferir nos agendamentos previstos de partidas nos aeroportos.

Chen *et al.* (2021) definem que devem ser consideradas não apenas as condições climáticas atuais, mas também as condições climáticas anteriores (nas últimas duas horas), para os aeroportos chineses. Feriados e períodos de pico de demanda também devem ser considerados, devido à variação observada na movimentação de aeronaves e passageiros em relação aos dias comuns (CHEN *et al.*, 2021).

O Aeroporto Internacional de Pequim – ZZBA, na China, possui um levantamento em tempo real de dados e realizava estudos de atraso dos voos aplicando algumas variáveis macro, sendo: clima, efeitos sazonais, propagação de atrasos e controle do tráfego aéreo, como apresentaram Yu *et al.* (2019).

A diversidade de informações conhecidas necessárias, apontadas pelos estudos de previsão apresentados, é fruto dos inúmeros motivos classificáveis para ocorrência de atraso em um voo já agendado.

Por conta disto, os principais motivos de atraso de aeronaves foram enumerados em oito categorias principais pela Associação Internacional de Transportes Aéreos – IATA, com o intuito de permitir a comunicação do atraso quanto ao motivo ou outras da aeronave e outras mensagens (IATA, 2016).

Para o Brasil, a ANAC utiliza a mesma classificação dos atrasos conforme descrita pela IATA, em códigos para cada justificativa levantada após a ocorrência do atraso, sendo então distribuídas entre motivos devido a passageiros e bagagens, equipamentos técnicos e aeronáuticos ou até condições climáticas, Tabela 1.

Tabela 1 - códigos a serem usados nas mensagens de movimento e desvio de aeronaves.

Grupo IATA	Justificativas	Códigos IATA/ANAC	Fonte
1. Passageiro e Bagagem (Carga e Correio)	Facilidades do aeroporto - restrições de apoio	AF	IATA AHM-730
	Migração/alfândega/saúde	AG	IATA AHM-730
	Segurança/pax/carga/alarme	AS	IATA AHM-730
	Liberação serv. Tráfego aéreo/antecipação	AT	IATA AHM-730
2. Manuseio de aeronaves e rampas	Abastecimento/destaqueio	GF	IATA AHM-730
3. Equipamento Técnico e Aeronáutico	Troca de aeronave	TC	IATA AHM-730
	Defeitos da aeronave	TD	IATA AHM-730
4. Danos à Aeronave ou Falha no Equipamento Automatizado	Avaria durante operações em voo	DF	IATA AHM-730
	Avaria durante operações em solo	DG	IATA AHM-730
5. Operações de Voo e Tripulação	Falha equipo automotivo e de atendimento de pax	MA	AIC1504 A
	Autorizado	AO	IATA AHM-730
6. Clima	Plano de voo - aprovação	FP	IATA AHM-730
	Alternativa abaixo dos limites	WA	AIC1504 A
	Degelo e remoção de neve e/ou lama em aeronave	WI	IATA AHM-730
	Aeroporto origem abaixo dos limites	WO	IATA AHM-730
	Atraso devido retorno - condições meteorológicas	WR	IATA AHM-730
	Remoção gelo/água/lama/areia-em aeroporto	WS	IATA AHM-730
	Atraso aeroporto de alternativa - condições meteorológicas	AM	IATA AHM-730
Aeroporto destino abaixo dos limites	WT	IATA AHM-730	
7. Restrição do Fluxo de Tráfego Aéreo e Autoridades Governamentais	Aeroporto de origem interdito	AI	AIC1504 A
	Aeroporto de destino interdito	AJ	AIC1504 A
	Aeroporto com restrições operacionais	AR	AIC1504 A
8. Reacionário e Diversos	Atrasos não específicos, outros	MX	IATA AHM-730
	Conexão de aeronave	RA	IATA AHM-730
	Conexão aeronave/volta - voo de ida não penalizado aeroporto interdito	RI	AIC1504 A
	Conexão aeronave/volta - voo de ida não penalizados condições meteorológicas	RM	AIC1504 A

Fonte: Adaptado de IATA (2016), ANAC (2000).

Em decorrência disso, o efeito em cadeia de atrasos, ou efeito cascata, pode ser resultado verificado com antecedência, tomando como base ocorrências anteriores. Contudo, no momento de definição de um modelo preditivo, é evidente a necessidade de um sistema que preveja corretamente os atrasos, não apenas indicando sua ocorrência ou possuindo ambiguidades ao classificá-los em categorias (KHAN *et al.*, 2021).

2.1 Previsibilidade e informações dos voos: dados conhecidos previamente

De acordo com Khan *et al.* (2021), antes de cada voo é elaborado um plano detalhando vários parâmetros operacionais necessários para o bom funcionamento da aeronave. Para as companhias aéreas, o plano de voo é elaborado quatro horas antes de cada voo internacional.

Estas informações permitem ao controlador acompanhar as etapas de voos previstas, medindo qualquer variação no tempo estimado em vários departamentos funcionais distintos: como condições climáticas, fluxo de tráfego aéreo, desempenho da aeronave, além de outros fatores necessários a definir a trajetória de voo ideal e garantir uma operação de voo segura.

Somado a complexidade das etapas a serem realizadas, a demanda aérea comum dos aeroportos analisados é um fator de impacto considerável a manutenção da pontualidade dos voos.

Dentre os problemas encontrados ao se realizar uma análise para prever atrasos, Yu *et al.* (2019) caracterizaram como um, dentre os principais, a enorme quantidade de informações levantadas pelos planos de voo, sejam em níveis macro ou micro, que podem influenciar negativamente na capacidade do sistema em prever os atrasos ou ainda gerar colinearidade entre fatores redundantes. Esta incapacidade de generalização das respostas acaba levando a um segundo problema enfrentando, o de que uma grande massa de dados apresenta muitas informações redundantes.

Para caracterizar os efeitos dos atrasos num aeroporto é preciso, primeiro, entender a situação atual da área de análise (seja aeroporto, linha aérea, rota, etc.) quanto a capacidade existente e o crescimento esperado de demanda. Atribuir os fatores-chave causadores do aumento na frequência de atrasos dos voos, além dos já citados sobre fluxo e controle aéreo, tem-se os mais específicos e pontuais, como os atrasos nas filas e atrasos em rota no modelo, feriados e períodos de pico de demanda também são considerados, apesar da variação em relação aos dias comuns, uma vez que esses dados estejam disponíveis CHEN *et al.* (2021).

Diante disto, ao buscar uma integração dos modelos de previsão de atrasos no mercado aéreo, Khan *et al.* (2021) ainda apresentam a necessidade de se desenvolver um sistema que preveja corretamente os atrasos, não apenas indicando sua ocorrência ou possuindo ambiguidades ao classificá-los em categorias. Este pensamento é concordante com o que apresenta Truong (2021), ao ressaltar ainda a importância de se analisar a relação entre as variáveis selecionadas como entrada para os sistemas de previsão de atrasos com sua ocorrência, de modo que, grande parte dos estudos atuais tendem ou buscar uma previsão através de inúmeros métodos de *machine learning* e a combinação destes, ou a correlação entre as variáveis de entrada e o atraso se dá de forma teórica apenas, não se aprofundando aos dados ou na possibilidade de uma aplicação real dos modelos.

Assim, a complexidade na definição dos modelos de previsão atrasos de voos engloba então para sua definição a seleção das variáveis de influência a serem aplicadas, o conhecimento prévio dos dados necessários, a escolha do modelo de previsão a ser aplicado além de garantir uma correlação entre as causas e efeitos com uma margem de acurácia significativa.

2.2 Variáveis de influência nos atrasos

De modo a identificar quais os cenários comumente aplicados em previsões, alguns estudos foram levantados, Tabela 2, permitindo a comparação entre os métodos, sua acurácia e quais as variáveis aplicadas.

Tabela 2 - Estudos de Previsão de Atrasos em Aeroportos.

Nº	Autores	Métodos empregados	R ^{2*}	Grupo de Variáveis
1	Wei e Vazea (2018)	Simulação.	0,8805	A, B, I, L, C, G, O.
2	Lambelho et al. (2020)	<i>LightGBM</i> , rede neural do tipo multilayer perceptron (<i>MLP</i>) e florestas aleatórias (<i>RF</i>).	0,794 0,772 0,771	A, B, E, F, H, K, O.
3	Chen e Wang (2019)	Regressão usando dados de séries temporais.	0,718	A, B, N, O.
4	Chen, Whang e Zhou (2021)	Regressão linear, árvore de decisão, método ensemble e <i>support vector machine</i> (<i>SVM</i>).	0,69 0,65 0,74 0,71	A, B, F, G, J, N, O.
5	Liu et al. (2019 ¹)	<i>Support Vector Machine</i> (<i>SVM</i>), Regressão Logística (<i>LR</i>) e <i>Random Forest</i> (<i>RF</i>)	0,7841 0,8356 0,8732	A, B, F, N, O.
6	Liu, Yin e Hansen (2019 ²)	Modelo de regressão hedônica log-linear.	0,716	A, B, F, G, H, L, O.
7	Schultz e Reitmann (2019)	Modelo Long Short-Term Memory (<i>LSTM</i> , Hochreiter e Schmidhuber (1997)) subtipo de redes neurais recorrentes.	0,75	A, B, K, M, O.
8	Khan et al. (2021)	<i>SL-BPNN</i> (<i>BPNN</i> é um tipo de rede neural feedforward (<i>FNN</i>) de única camada – <i>single layer</i>), <i>DL-BPNN</i> (camadas intermediárias densas – <i>deep leayers</i>), <i>9SVM</i> (<i>Support vector machine</i>), <i>hyp-free CPCLS</i> (Rede Neural de Mínimos Quadrados de Componente Principal em Cascata), Ensembles (Aprendizagem em conjunto), <i>RF</i> (Random Forest), <i>GBDT</i> (Árvore de decisão de aumento de gradiente) e <i>XGBoost</i> (Aumento de gradiente extremo).	0,01 0,16 0,16 0,19 0,15 0,16 0,16 0,17	A, F, G, H, K, M, N, O.
9	Wei e Vazea (2018)	Estatística Qui-quadrado e Kolmogorov-Smirnov.	0,8773 0,8805	A, B, E, G, M, O.
10	Yua et al. (2019)	<i>DBN-SVR</i> (<i>Deep belief network</i> (<i>DBN</i>)), <i>Support Vector Regression</i> (<i>SVR</i>), <i>k-nearest neighbors</i> (<i>k-NN</i>), <i>Support vector machine</i> (<i>SVM</i>) e regressão linear (<i>LR</i>).	0,93 0,87 0,87 0,82	A, B, E, F, G, I, K, L, M, O.
11	Oliveira et al. (2021)	Modelo de regressão logit.	0,4714	A, B, N, O.
12	Chen e Lin (2021)	Regressões de dados em painel.	0,1954	A, B, G, H, I, O.
13	Arora e Mathur (2020)	Regressão logística multinomial.	0,3	A, B, E, G, O.
14	Scarpel e Pelicioni (2018)	Árvores de classificação e regressão (<i>CART</i>), regressão linear múltipla e modelo de regressão de série temporal.	0,665 0,271 0,753	A, B, E, F, G, O.
15	Truong (2021)	Rede supervisionada do tipo <i>Bayesian Network augmented Naïve Bayes classifier</i> (<i>BNAN</i>).	0,9706	A, B, C, D, I, J, K, N, O.

*Acurácia do sistema desenvolvido, sua capacidade de previsão.

As categorias de motivos de atraso estabelecidas pela IATA/ANAC conseguem indicar a considerável quantidade de razões distintas para que o planejamento de um voo saia do previsto. Essa diversidade de motivos prováveis de atrasos acaba demandando uma quantidade considerável de dados explicativos, ou variáveis descritivas. A TABELA 3 identifica quais as principais variáveis aplicadas nos estudos levantados, dividindo-as em grupos quanto à similaridade, TABELA 2.

Tabela 3 - Grupos de Variáveis utilizadas pelos Estudos Analisados e os seus respectivos Parâmetros em Comum.

Grupo de Variáveis	Justificativa	Exemplos de Parâmetros
A (Data)	Referente às informações de data do voo e da movimentação do aeroporto, como a previsão dos planos de voos e a sua real data de ocorrência. Permite acompanhar a sazonalidade e os efeitos dos picos semanais no possível atraso do voo.	Ano, mês e dia data de partida/chegada prevista e realizada para os voos, voo em período de sazonalidade na região, etc.
B (Hora)	A fim de analisar a ocorrência de atrasos notificáveis, as horas programadas e a variação real ocorrida são necessários.	Hora programada e realizada para cada partida/chegada no aeroporto.
C (Embarque)	Específico a cada portão, o embarque sofre com efeitos de passageiros, carga, tripulação e quaisquer influências da estrutura do aeroporto.	Tempo estimado e real para embarque em cada portão, tempo médio de inspeção de segurança, quantidade de áreas de inspeção por portão de embarque/desembarque, etc.
D (Desembarque)	Assim como o embarque, o tempo de desembarque atua ao perpetuar os efeitos de atrasos em todos os setores do aeroporto.	Tempo estimado e real para desembarque em cada portão, tipo de embarque realizado no portão para a companhia (<i>finger</i> , remoto, etc.), número da fase de voo a ser realizada, etc.
E (Empresa aérea)	As informações da empresa aérea responsável, assim como a logística permitida e acessível pelo pessoal desta garantem atribuir os pesos de responsabilidade sobre cada etapa realizada durante a programação e execução de um voo.	A responsável pelo voo, taxa de atraso da companhia para a rota, hubs da empresa no aeroporto, tempo de espera programado quando a tripulação troca de aeronave, tempo de descanso da tripulação entre conexões, etc.
F (Programação dos voos)	Cada decolagem parte de um conjunto de fatores previstos e programados que garantem a funcionalidade de aeroporto com sua capacidade máxima atingida. Os fatores considerados nesses planos e as configurações adotadas definem o sucesso da execução on-time da decolagem ou seu atraso.	Número total de voos programados para decolagem e partida para os próximos 15 minutos (ou a cada 15 minutos), 30 minutos, uma hora, entre pares de aeroportos, número de rotas programadas em uma hora, número de voos noturnos para o dia, etc.
G (Tempo de voo e distância)	Informações aéreas do voo permitem acompanhar o desenvolvimento do plano a partir do tráfego existente. A distância e a velocidade executada acabam permitindo a ocorrência de atrasos.	Tempo de voo estimado e realizado, Tempo médio diário entre dois movimentos programados consecutivos, Tempo médio entre dois movimentos consecutivos programados (AM/PM), etc.

H (Local de destino)	As características do aeroporto de destino, assim como a região atendida por este, permitem a formação de atrasos.	Destino do voo (país, estado, cidade), densidade populacional da área estatística metropolitana associada a cada aeroporto de destino, PIB do destino, etc.
I (Capacidade do aeroporto)	Informações específicas da estrutura fornecida pelo aeroporto de origem no desenvolvimento das decolagens, desde a infraestrutura utilizada ao pessoal existente. As normas de funcionalidade locais e suas variações ao longo do dia vão garantir visibilidade sobre possíveis causas de atrasos, permitindo sua mitigação.	Capacidade média diária de pousos e decolagens, capacidade horária máximo de decolagens (ou por outra unidade de tempo), capacidade horária máxima de pousos (ou por outra unidade de tempo), capacidade de aeronaves em pista (taxiamento), capacidade máxima de aeronaves em calço (block on time), etc.
J (Horário de pico)	Períodos específicos conhecidos e aguardados pela gestão do aeroporto, assim como as decisões e mudanças no funcionamento quando esta estiver em ocorrência.	Número médio de voos na hora-pico, volume de tráfego intenso no aeroporto em unidades de aeronave/minuto.
K (Aeronave)	Dados da frota em operação, assim como limitações do próprio aeroporto sobre o funcionamento destas. O tempo de cada etapa acompanhado pelo controle de tráfego e as decisões de prioridade em cada etapa.	Tipo de aeronave, número de assentos da aeronave atribuídos a um voo, carga da aeronave (TOW, LW, MLW, MTOW, MTW), etc.
L (Financeiro e taxas)	Fatores econômicos garantem a prioridade do fluxo de carga, de modo que o embarque de alguns bens específicos possam exigir maior dedicação.	Taxa de serviço de decolagem, taxa de serviço de pouso, concentração de mercado (lojas existentes no embarque nacional, internacional e no lado terra do aeroporto), Custo de envio de carga, presença de carga médica, tempo de embarque de carga..
M (Estrutura do aeroporto)	Informações da infraestrutura, dados comuns ao funcionamento do aeroporto a partir da infraestrutura existente.	Configuração da pista Runway Direction, comprimento da fila de passageiros na check-in, comprimento da fila de embarque de passageiros, acesso ao terminal de voo (distância do terminal pelo acesso principal e do mais distante), Tempo decorrido em groundbound, Altitude das pistas.
N (Meteorologia)	Dados meteorológicos previstos e os reais, assim como as restrições e procedimentos adotados em situações severas (tráfego, pistas, embarque de carga, embarque de passageiros, groundbound, abastecimento, etc.).	Configuração da pista Runway Direction, comprimento da fila de passageiros na check-in, comprimento da fila de embarque de passageiros, acesso ao terminal de voo (distância do terminal pelo acesso principal e do mais distante), etc.
O (Atrasos levantados)	Informações de atraso já conhecidas, esperadas pelo aeroporto e pelas companhias e as suas decisões diante destas. Quais as regras de dissipação de atrasos ou diminuição de tempos em etapas controláveis.	Voos em conexão com atrasos já existentes, atrasos por problemas na pista, atrasos por tráfego aéreo intenso, atrasos por falha de sistema, atraso de embarque, atraso de desembarque, atraso devido a carga, atraso devido abastecimento de aeronave, outros atrasos.

Os grupos de variáveis utilizadas nos estudos apresentados na Tabela 2 buscaram identificar e descrever as razões de atrasos, mas para isso conjuntos distintos de modelos de previsão foram aplicados.

Cada cenário, conjunto de variáveis e modelos selecionados, buscavam identificar e descrever motivos distintos de atrasos, seja devido a tripulação e embarque da cabine (WEI & VAZEA (2018), SCHULTZ & REITMANN (2019)); seja devido a impactos climáticos severos ou inesperados (CHEN & WANG (2019), LIU *et al.* (2019)^a, OLIVEIRA *et al.* (2021)); relacionados a entregas e embarque de carga (LIU *et al.* (2019)^b); por falhas de conduta das companhias aéreas (ARORA & MATHUR (2020)); devido a problemas de capacidade em dias congestionados no aeroporto (SCARPEL & PELICIONI (2018)), ou até modelos que englobassem a universalidade dos motivos (LAMBELHO *et al.* (2020), KHAN *et al.* (2021), YU *et al.* (2019), CHEN & LIN (2021), TRUONG (2021)).

Em destaque, nos trabalhos analisados, os modelos definidos como aprendizado de máquina, *machine learning*, são maiorias em relação aos métodos clássicos de previsão, como as regressões *logit*, em painel e logística multinomial utilizadas por Oliveira *et al.* (2021), Chen & Lin (2021) e Arora & Mathur (2020), respectivamente.

Esta situação vai diretamente a favor do que apresentaram Karlaftis & Vlahogianni (2011) ao levantarem uma comparação entre o trabalho estatístico dos dados e o uso de inteligência artificial, especificamente as redes neurais, para pesquisas em transportes ressaltando que, apesar das similaridades, quando a ênfase do tratamento das informações for obter previsões corretas, é sugerido como o método mais adequado o uso de *machine learning*.

Este princípio é concordante com o que apresentaram Chen *et al.* (2021), ao definirem o uso da regressão linear como método de previsão eficaz e aplicável devido a menor dimensão do banco de dados utilizado.

3 CENÁRIOS APLICADOS PARA A PREVISÃO DE ATRASOS

Entretanto, ao realizar um estudo mais próximo da realidade de operação de um aeroporto ou linha aérea, a precisão da resposta obtida pelo modelo de previsão está diretamente ligada ao cenário de análise aplicado.

Truong (2021) ressalta a importância de se analisar a relação entre as variáveis selecionadas como entrada para os sistemas de previsão de atrasos, com sua ocorrência, de modo que, grande parte dos estudos atuais tendem ou buscar uma previsão através de inúmeros métodos de *machine learning* e a combinação destes, ou a correlação entre as variáveis de entrada e o atraso se dá de forma teórica apenas, não se aprofundando aos dados.

Portanto, é necessário definir o cenário mais adequado a ser utilizado no estudo de previsão, de modo que, a correlação entre as variáveis de entrada e o método aplicado permitam a obtenção de uma resposta o mais acurada possível.

Contudo, primeiramente é necessário identificar as tendências em uso de técnicas de previsão de atrasos de voos na literatura, além de estipular um intervalo de confiança aceitável para as acurácias observadas por estes métodos.

3.1 Meta-Análises

Um método de análise sistemática capaz de desenvolver faixas parâmetro ao comparar estudos de caso, é a meta-análise. As análises sistemáticas por meta-análise foram inicialmente desenvolvidas com o objetivo de definir padrões proporcionais a serem seguidos por tratamentos biomedicinas. Para Oliveira Filho (2015), a apresentação comum da meta-análise é no formato de *forest plot*. Neste tipo de apresentação, a meta-análise indica, para cada estudo listado, os efeitos correspondentes a cada análise realizada.

Stivali (2022) indica que a formulação mais simples da meta-análise assume que a quantidade de interesse, I_j , é homogênea para a população, isto é, entre os estudos analisados, divergindo apenas por erro de medida e erro amostral. Essa formulação é chamada de modelo de efeitos fixos e assume que o $\beta_1 = \beta_2 = \dots = \beta$, (I).

$$I_j = \beta + e_j \text{ com } j = 1, \dots, m; \text{ e com } e_j \sim N(0, \sigma^2) \quad (\text{I})$$

Assim, o autor presume que todos os estudos considerados foram realizados de forma semelhante e que cada um produz apenas uma estimativa não viesada, sendo

conhecidas as estimativas da variância de cada estudo (s_j^2). Assim, a estimativa do interesse, ou da acurácia de previsão de um estudo, pode ser obtida por uma média ponderada, considerando o inverso das variâncias como peso para ser calculada (II).

$$\bar{\beta} = (\sum w_j I_j) / (\sum w_j) \quad (\text{II})$$

$$w_j = 1/(s_j^2) \quad (\text{III})$$

$$\sigma^2 = 1/(\sum w_j) \quad (\text{IV})$$

Contudo, o modelo de meta-análise pode ser estimado com efeitos fixos ou aleatórios, como defende YEN & LI (2022). Stivali (2022) esclarece que, para o modelo de efeitos aleatórios, considera-se que a quantidade de interesse β (o I_j “real” no presente caso) é ele próprio uma variável aleatória, de forma que cada estudo específico identifica uma realização dessa variável aleatória. Essa formulação pode ser descrita considerando que a realização β_j corresponde à média entre os estudos (β_0) e um termo de erro (u_j) com média zero e variância (σ_u^2), tal que:

$$I_j = \beta_j + e_j \quad (\text{V})$$

$$I_j = \beta_0 + u_j + e_j \quad (\text{VI})$$

$$j = 1, \dots, m; u_j \sim N(0, \sigma_u^2); e_j \sim N(0, \sigma_e^2) \quad (\text{VII})$$

Logo, como aponta Nelson (2015), o valor real ($\beta_0 + u_j$) varia de estudo para estudo, com o termo de erro u_j captando a heterogeneidade não mensurada entre os estudos e o termo de erro e_j captando erros de medida e erros amostrais, como no modelo de efeitos fixos. A variância σ_u^2 é chamada de variância entre estudos ou de variância dos efeitos aleatórios e a variância σ_e^2 é chamada de variância interna ao estudo, e ambas são refletidas.

$$Var(I_j) = \sigma_u^2 + \sigma_e^2 \quad (\text{VIII})$$

Pode-se obter uma estimativa da variância entre estudos (σ_u^2) pela estimativa amostral (sem ponderação) considerando a amostra de estudos (IX). E com essa estimativa é possível obter a média ponderada considerando o modelo de efeitos aleatórios (X).

$$s_u^2 = \sum_{j=1}^m (I_j - \bar{I})^2 / (M - 1) \quad (\text{IX})$$

$$\bar{\beta} = \sum w_j I_j / \sum w_j \quad (\text{X})$$

$$w_j = 1/(s_j^2 + s_u^2) \quad (\text{XI})$$

$$\sigma^2 = 1/(\sum w_j) \quad (\text{XII})$$

Apesar de inicialmente ser desenvolvido como método de análise biomédico, as análises sistemáticas por meta-análise têm apresentado uma tendência de uso em estudos de transportes. Além disso, mesmo sendo distintos os modelos de análise utilizados nestes estudos, definir o padrão a ser seguido em análises deste tipo é um objetivo comum observado entre as pesquisas.

3.1.1 Meta-Análises aplicadas a estudos de transportes

Dentre as aplicações de meta-análises em transportes, Garrow et al. (2021) empregaram meta-análises em um conjunto de estudos publicados, destacando a necessidade de refinar os modelos de demanda, a fim de incorporar a mudança de modal adotada por alguns usuários de transportes sob demanda diante a implantação de infraestruturas para acomodar a mobilidade aérea urbana – UAM, aguardadas para os principais centros urbanos internacionais.

Ainda sobre transportes urbanos, Kaye et al. (2021) revisaram sistematicamente estudos sobre a aceitação dos usuários de veículos automatizados condicionais a completos, sobre quando esses veículos devem ser usados de forma privada. As correlações meta-analíticas revelaram que o controle comportamental percebido, os benefícios/utilidade percebidos, a facilidade de uso percebida e as normas subjetivas/sociais tiveram resultados positivos significativos. Assim, os resultados da revisão sistemática e meta-análise fornecem suporte para modelos psicossociais que auxiliam na compreensão da aceitação dos usuários de veículos automáticos individuais.

Enquanto Rennert (2022) empregou meta-análises, introduzindo o primeiro exame de como os elementos do serviço de trânsito – por exemplo, frequência, tarifa, confiabilidade – impacto na elevação do valor da propriedade residencial induzida por trilhos. Os resultados revelam que fatores de geografia, tipo de dados de habitação, raça e etnia, políticas de controle de aluguel, tipo de trilho, custo de trânsito e extensão da rede de trânsito afetam significativamente a magnitude do aumento do acesso ferroviário, variando de efeitos de depreciação de 7,4 pontos percentuais a efeitos de apreciação de 9,6 por cento pontos.

Na economia dos transportes, Wardman (2022) apresentou uma meta-análise de elasticidade-preço da demanda de viagens de todos os modos de transportes, com base nas elasticidades de 2023, extraídas de 204 estudos britânicos publicados entre 1968 e

2020. O principal desafio que a pesquisa buscou alcançar é o de servir como um complemento importante para os padrões de elasticidade de preços a serem recomendados de forma oficial.

Uma das principais respostas obtidas por Odeck & Bråthen (2012), ao utilizarem a meta-análise para verificar as variações nas pontuações de Eficiência Técnica Média (MTE) dos portos marítimos, foi de que o modelo de meta-análise de efeitos aleatórios supera o modelo de efeitos fixos na explicação das variações nos MTEs, visto que, o uso de modelos de efeitos aleatórios em estudos de meta-análise leva em conta os efeitos específicos de cada estudo individualmente.

A questão da sustentabilidade foi abordada por Singh et al. (2020), ao investigarem empiricamente os fatores que influenciam a intenção do consumidor de adotar uns veículos elétricos (VEs). Através de uma simples meta-análise, verificou-se uma tendência de estudos sobre os fatores que influenciam a adoção de VEs aumentando significativamente na última década, os quais podem ajudar os governos e a indústria automobilística a aumentar a participação no uso de VEs.

Assim, o uso da meta-análise em estudos de transportes, com áreas específicas desde validação das infraestruturas urbanas, questões econômicas e de sustentabilidade, conseguem identificar tendências de uso, além de caracterizar os cenários observados para cada situação analisada.

Entretanto, o emprego da análise sistemática por meta-análise não permite uma comparação entre modelos distintos aplicados em estudos distintos, mas que possuem um mesmo objetivo. Desta forma, faz-se necessário a utilização de um método de hierarquização e equivalência para estipular a escolha de um método de previsão, dentre os verificados nas pesquisas analisadas.

3.2 Análise Envoltória de Dados – DEA

A técnica denominada Análise Envoltória de Dados (*Data Envelopment Analysis – DEA*) foi proposta inicialmente por Charnes, Cooper e Rhodes (1978) e por Banker, Charnes e Cooper (1984). Trata-se de um modelo de programação linear, não paramétrico, capaz de medir a eficiência relativa de cada Unidade Tomadora de Decisão (DMU) de um conjunto analisado, por meio da construção de uma fronteira de eficiência. Dessa forma, a DMU é a unidade pertencente ao conjunto no qual será avaliada a

eficiência produtiva. No caso do presente estudo, cada sistema de transportes selecionado para determinação da eficiência irá compor uma DMU.

A vantagem desta técnica é a consideração, simultânea, de entradas e saídas múltiplas, que caracterizam determinado processo produtivo de uma DMU (CAMIOTO, 2013; MARIANO, 2008). Além disso, essa técnica possibilita a análise da eficiência dos sistemas de transportes, por meio da construção de uma fronteira de eficiência.

A eficiência pode ser definida como a razão entre os produtos (*outputs*), e os máximos produtos que poderiam ser gerados, dados os insumos (*inputs*) considerados (AZAMBUJA, 2002). Dessa forma, tem-se que a eficiência produtiva está relacionada à capacidade que um sistema produtivo possui de gerar tantos produtos quanto for possível, utilizando os recursos disponíveis.

A metodologia utilizada pela DEA mede a eficiência de uma unidade por vez e compara o desempenho relativo das demais DMUs com aquelas localizadas sobre a fronteira, estabelecendo assim, as que possuem um desempenho eficiente e as que precisam modificar seus *inputs* e/ou *outputs* para atingirem a fronteira. Sendo assim, a técnica não fornece uma eficiência absoluta, apenas relativa às unidades presentes no conjunto de análise (FALCÃO & CORREIA, 2012).

A técnica DEA pode apresentar diferentes modelos para avaliação da eficiência. È necessário escolher aquele que representa melhor a eficiência do sistema produtivo analisado (CASTRO *et al.*, 2016).

Entretanto, assim como apresentado por Arjomandi *et al.* (2018), o uso do DEA em unidades tomadoras de decisão (DMUs) heterogêneas, os resultados podem se tornar imprecisos e viesados. Assim, os autores incidem ao uso do conceito de meta-fronteira desenvolvido por O'Donnell *et al.* (2008).

Segundo o estudo de O'Donnell *et al.* (2008), a meta-fronteira pode ser definida como o limite de um conjunto de tecnologia irrestrito. Nela, grupos de DMUs operam em ambientes distintos ou com poucos recursos e restritos, sendo os limites desses conjuntos de DMUs as fronteiras de cada grupo. Desta forma, considera-se a existência de sub-tecnologias que representam as possibilidades de produção para cada DMU.

A análise em sub-grupo, ou clusters, do DEA, proposto pela meta-fronteira, considera uma divisão das DMUs em $K (>1)$ grupos, nos quais as restrições devido a distinção entre os grupos podem impedir que as DMUs escolham, a partir de toda a gama de combinações possíveis de insumo-produto, o conjunto referente ao modelo buscado de *input-output*, T (O'DONNELL *et al.*, 2008).

Assim, o autor define que as combinações de inputs-outputs (x,y) disponíveis para as DMUs do k -ésimo grupo estão contidas no conjunto de tecnologia específico do grupo, em que x pode ser usado por DMUs em um grupo k para produzir y , (XIII).

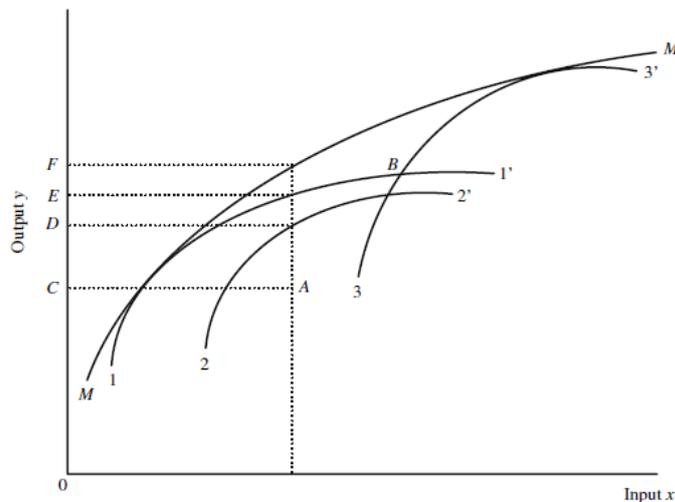
$$T^k = \{(x, y): x \geq 0; y \geq 0\} \quad (\text{XIII})$$

Desta forma, as DMUs heterogêneas (estudos levantados) apresentam como resposta fronteiras específicas para cada cluster definido. O conjunto de fronteiras provenientes de cada *output* gerado P , (XIV), têm suas funções de distância D , (XV), ajustadas a uma fronteira universal entre os grupos, denominada meta-fronteira MM' , Figura 1.

$$P^k(x) = \{y: (x, y) \in T^k\}, \quad k = 1, 2, \dots, K \quad (\text{XIV})$$

$$D^k = \inf_{\theta} \left\{ \theta > 0 : \left(\frac{y}{\theta} \right) \in P^k(x) \right\}, \quad k = 1, 2, \dots, K \quad (\text{XV})$$

Figura 1: Fronteiras Técnicas dos Clusters K e a Metafronteira Mm' .

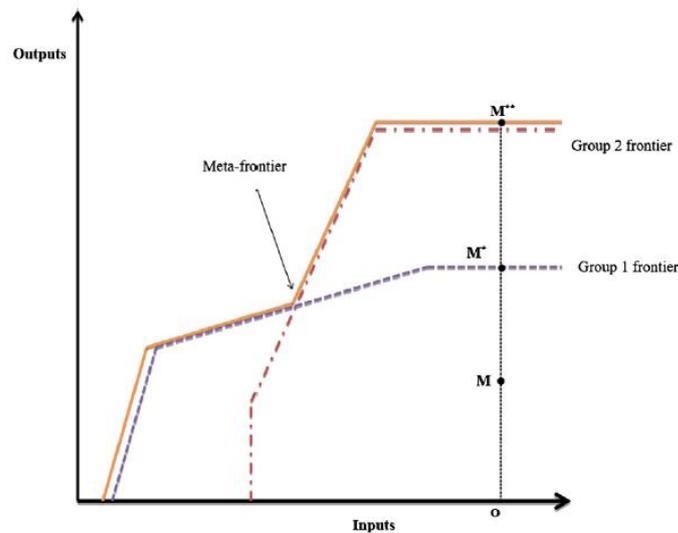


Fonte: O'Donnell *et al.* (2008).

Vale salientar que, assim como os modelos clássicos do DEA, quando a função distância $D(x,y) = 1$ assumimos que há uma relação eficiente entre os inputs-outputs gerados pelo grupo k .

No entanto, o modelo de metafronteira proposta por O'Donnell *et al.* (2008) usa como base a distribuição convexa da curva de eficiência universal, Figura 1. Assim, Arjomandi *et al.* (2018) descrevem que, para análises mais precisas de eficiência em análises de subgrupos, o modelo de curva de fronteira côncava, Figura 2, acaba apresentando resultados de eficiência mais próximos da realidade de cada cluster k , evitando assim a formação de vieses positivos ou negativos para cada DMU.

Figura 2 - Fronteiras Técnicas por K Clusters e a Metafronteira do Tipo Côncava Formada



Fonte: Arjomandi *et al.* (2008).

Outra medida que pode ser obtida pela análise de metafronteira é a eficiência técnica em relação à fronteira de um cluster k , TE^k . Então, a eficiência técnica obtida é uma medida de eficiência técnica orientada para *output* em relação à tecnologia do grupo k , (XVI).

$$TE^k(x, y) = D^k(x, y) \quad (\text{XVI})$$

Contudo, O'Donnell *et al.* (2008) ainda esclarece que, devido a função distância ou eficiência de um cluster k , $D^k(x, y)$, não poder apresentar um valor menor do que o observado de eficiência descrita pela metafronteira, ou meta-eficiência $D(x, y)$; assume-se que a eficiência metatecnológica para todo o modelo $TE(x, y)$, é dada pelo produto, (XVII), da eficiência tecnológica de um subgrupo apenas, TE^k , com o coeficiente de razão metatecnológico $MTR^k(x, y)$. Essa é outra maneira de dizer que a metafronteira envolve a fronteira de um cluster k .

$$TE(x, y) = TE^k * MTR^k(x, y) \quad (\text{XVII})$$

$$MTR^k(x, y) = \frac{D(x, y)}{D^k(x, y)} = TE(x, y) / TE^k(x, y) \quad (\text{XVIII})$$

Assim, os modelos DEA de metafronteira conseguem classificar e hierarquizar padrões distintos de DMUs. Essa funcionalidade comparativa entre um grupo selecionado de unidades faz com que aplicações da Análise Envoltória de Dados seja cada vez mais comum em estudos de transportes.

A definição de metodologia que utiliza o DEA meta-fronteira e as análises sistemáticas por meta-análises, em estudos de transporte aéreo são confirmadas por Yen & Li (2022). Os autores realizaram uma avaliação de desempenho baseada em rotas pode refletir melhor a variedade de características operacionais, considerando as diferentes características das rotas. Para tanto, foram aplicadas as variáveis-chave obtidas pela meta-análise para medir o desempenho baseado em rotas com modelagem de análise de envolvimento de dados (DEA) por meta-fronteira.

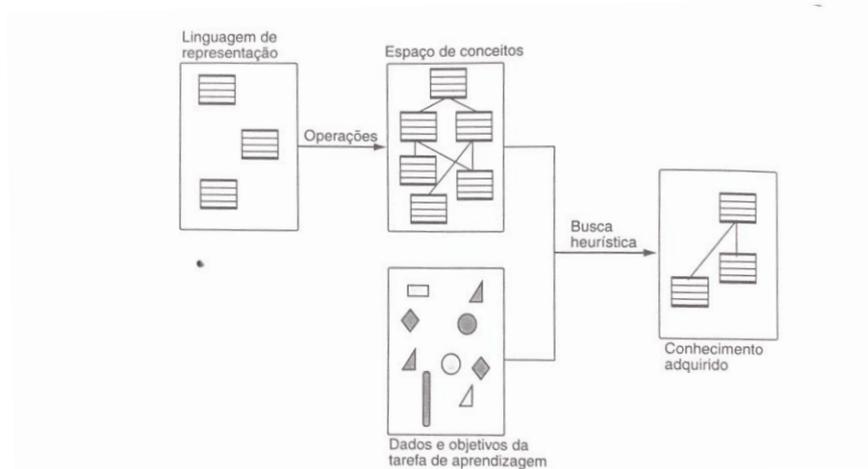
Os resultados destes estudos visam auxiliar os operadores e formuladores de políticas a melhorar o desempenho operacional e identificar os principais fatores para o desenvolvimento e planejamento do transporte aéreo.

4 MÉTODOS DE APRENDIZADO DE MÁQUINA – MACHINE LEARNING

Simon (1983) definiu aprendizado de máquina como “qualquer mudança num sistema que melhore o seu desempenho na segunda vez que ele repetir a mesma tarefa, ou uma outra tarefa da mesma população”.

De uma maneira generalista, os algoritmos que utilizam o aprendizado de máquina seguem um princípio, o objetivo de inferir uma definição geral que permitirá que o algoritmo reconheça ocorrências futuras de uma classe, como descreve Luge (2004), Figura 1.

Figura 3 - Um Modelo Geral do Processo de Aprendizagem.



Fonte: LUGE (2004).

Luge (2004) indica então que, a partir da construção de conceitos e da análise representativas de um fato a ser previsto, aliado aos dados e objetivos a serem retratados pelo modelo, o sistema dito aprendizado de máquina reconhecerá padrões de comportamento, chamado de potencial de generalização, além de ser capaz de atribuir o conhecimento adquirido a casos similares.

O autor indica que, a capacidade de adquirir um conhecimento por parte de um sistema de aprendizado de máquina decorre de duas formas principais. A primeira, chamada de aprendizado supervisionado, funciona através de alguma medida de adequação ou paralelo a outro método externo de classificação de exemplos. A outra forma, não utiliza uma referência externa ou conhecida de resultados para adquirir o conhecimento de padrão entre os grupos previstos, requerendo do próprio algoritmo de aprendizado que avalie os conceitos.

Devido à complexidade em analisar as respostas do aprendizado não supervisionado, costuma-se optar pelo uso do modelo supervisionado de generalização dos dados em estudos que preveem atrasos em voos. Esta situação é debatida por Khan *et al.* (2021), ressaltando ainda que, no aprendizado supervisionado, são dois os tipos principais de mecanismos de estimativa comumente aplicados, a regressão e a classificação. Na regressão, a saída de atraso de voo é prevista de forma contínua, como o tempo de atraso aguardado ou grupos de atraso. Enquanto, na classificação, a saída de atraso de voo é classificada de forma binária (1 = voo atrasado, 0 = voo “*on-time*”).

Truong (2021) define que uma estrutura de rede causal supervisionada é necessária para fins de previsão. Além disso, quando a análise de previsão de risco for definida como a variável de destino, a estrutura da rede pode ser usada para prever o risco (probabilidade de um incidente) com base em outras variáveis (preditores) na rede. Se o risco for uma variável discreta, o modelo de previsão será um modelo de classificação e a variável alvo também é chamada de classificador.

Contudo, um ponto considerável na análise de respostas de métodos de previsão por *machine learning* é a real dimensão da acurácia obtida. As saídas possíveis de respostas podem se enquadrar em positivos reais que são positivos corretamente previstos são chamados de positivos verdadeiros (TP), ou em positivos reais que são negativos preditos erroneamente são chamados de falsos negativos (FN), os negativos reais que são negativos previstos corretamente são chamados de negativos verdadeiros (TN) e os negativos reais que são positivos preditos erroneamente são chamados de falsos positivos (FP), como apresentaram Chicco & Jurman (2020).

Portanto, uma maneira de definir a acurácia é através da relação entre as respostas possíveis, sendo uma das mais empregadas a medida de precisão de teste *F1-Score*, ou *F-score* ou *F-measure* (XIX).

$$F1 - Score (Acurácia) = \frac{TP+TN}{TP+TN+FP+FN} \quad (XIX)$$

Além da definição da acurácia dos modelos preditivos por *machine learning*, algumas outras considerações devem ser tomadas quando aplicados trabalhos com alta densidade de dados. Em aplicações reais, os dados coletados são altamente ruidosos, desequilibrados, dispersos e distorcidos, o que pode afetar muito a capacidade de previsão dos algoritmos de aprendizado de máquina. Os dados altamente distorcidos e dispersos podem tornar difícil para os regressores preverem o atraso do voo, enquanto os dados

altamente desequilibrados e ruidosos podem dificultar a classificação dos atrasos dos voos pelos classificadores.

Diante disto, Khan *et al.* (2021) estabelecem que a natureza da incerteza nos dados de atraso de voo pode evitar a suposição de normalidade na regressão e equilíbrio de classe na classificação. As técnicas de pré-processamento e transformação para melhorar a distribuição de dados altamente distorcida e dispersa para o mecanismo de regressão não têm uma vantagem em melhorar o desempenho dos regressores. No entanto, a aplicação de várias técnicas de amostragem melhora o desempenho dos classificadores, equilibrando as classes e tornando os limites de decisão mais suaves.

Portanto, o uso de mecanismos de ajuste da amostra de dados deve ser implementado, ao serem realizadas análises com categorias diversas. Dentre os tipos de técnicas de amostragem para desequilíbrio de classe e sobreposição de limites de decisão, Khan *et al.* (2021) conseguem indicar como o mais eficaz para estudos de previsão de atraso de voos o uso da sobre-amostragem da classe minoritária pela Técnica de Minoridade Sintética sobre Amostragem (SMOTE) e, em seguida, remove a classe majoritária barulhenta e as classes majoritária e minoritária limítrofe por um método de sub-amostragem e de limpeza de dados Tomek para tornar o limite de decisão mais suave.

Assim, a estrutura do método mais eficaz de previsão de atrasos em voos por *machine learning* deve ser capaz de atingir os melhores valores de acurácia definidos pelo F1-Score e que melhor se comportem quando exposto as técnicas de amostragem para grandes massas de dados. A capacidade do método então permitirá a indicação de uso em situações reais de controle e manutenção do tráfego aéreo de um aeroporto ou linha aérea.

4.1 Redes Neurais Artificiais - RNA

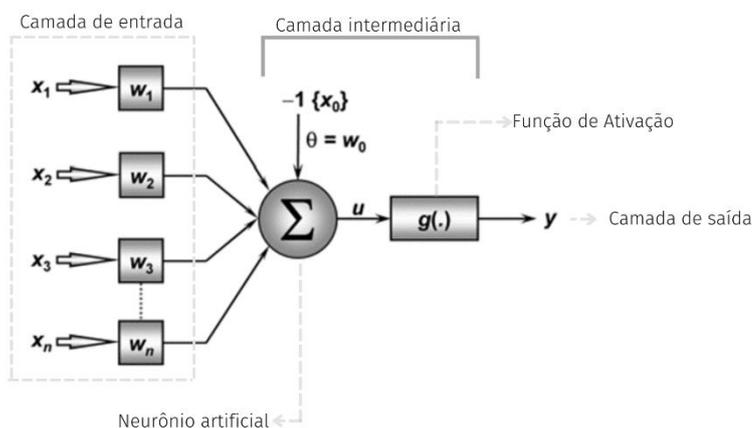
Dentre o conjunto de métodos de inteligência artificial, classificadas como aprendizado de máquina, uma que vem apresentando destaque em análises mais complexas de dados são as Redes Neurais Artificiais (RNAs) em suas diversificações.

As RNAs são definidas como sistemas paralelos e distribuídos, compostos por unidades de processamento simples interligadas, chamadas neurônios artificiais, capazes de armazenar informações, reconhecer e classificar padrões e fazer previsões (MARCY, 2017).

Um dos modelos mais simples de RNAs, o primeiro a ser formulado, foi o modelo *perceptron*, cujo propósito, como destacam Silva et al. (2010), focava em implementar um modelo computacional inspirado na retina, objetivando-se então um elemento de percepção eletrônica de sinais.

Em questão de estrutura, as RNAs são compostas por camadas, Figura 4, divididas em: a camada responsável pela entrada das variáveis explicativas (inputs) selecionadas para o modelo (camada de entrada), a camada contendo as unidades de computação e tratamento dos dados, chamados de neurônios, a qual é atribuída um modelo de função para o tratamento dos dados, a função de ativação, (camada intermediária ou escondida), além da saída da rede, com as respostas a serem observadas pelo modelo (camada de saída).

Figura 4 - Ilustração de uma Rede Neural Artificial *Perceptron*.



Fonte: Adaptado de Silva *et al.* (2010).

O conjunto de neurônios presentes na camada intermediária, o número de camadas escondidas, assim como a escolha da função de ativação e das variáveis do modelo, recebe o nome de arquitetura da rede.

O funcionamento das redes é descrito por Haykin (2001) com semelhança ao cérebro humano, dando nome ao método. De forma que, as sinapses de comunicação provenientes de um estímulo nervoso, são equivalentes as entradas x_n como estímulos sendo atribuídos pesos sinápticos w_n , processados e conectados em cada neurônio Σ , que resultará em: ou uma conexão com uma rede de neurônios seguintes ou a realização da ação demandada, y .

O autor ainda ressalta que, diferente do cérebro humano, as redes artificiais podem encontrar respostas num intervalo positivo ou negativo. Além disso, as RNAs são

compostas por um somador de sinais de entrada, ponderados pelas sinapses realizados nas camadas de neurônios, tendo uma função restritiva, função de ativação ($g(\cdot)$), que limita o intervalo permissível de amplitude do sinal de saída a um valor finito real.

De maneira geral, a RNA pode ser matematicamente descrita, como definiu Silva *et al.* (2010), por XX, numa função de limitação linear e uma saída possível de classificação em 1 (verdadeiro) ou -1 (falso).

$$y = \begin{cases} 1, & \text{se } \sum w_i * x_i - \theta \geq 0 \leftrightarrow w_1 * x_1 + w_2 * x_2 - \theta \geq 0 \\ -1, & \text{se } \sum w_i * x_i - \theta < 0 \leftrightarrow w_1 * x_1 + w_2 * x_2 - \theta < 0 \end{cases} \quad (\text{XX})$$

No modelo de Silva *et al.* (2010), caso o somatório das entradas x_n , com os pesos atribuídos em cada sinapse w_n , forem maiores ou iguais ao valor limiar θ , assume-se como positiva a resposta da rede. Caso esta diferença seja negativa, a resposta é falsa.

O limiar (θ), seria então o fator de decisão ao modelo, uma vez que as respostas obtidas pelas interações realizadas nas camadas de neurônios, devem ser comparadas a este fator conhecido. Desta maneira, o limiar corresponde ao valor real previamente atribuído a rede, como uma resposta já conhecida e esperada para um conjunto de variáveis x_n .

Sendo características de modelos de previsão supervisionados, a capacidade de generalização de uma rede neural provém da realização de um treino. No treino, valores conhecidos de entrada (*inputs*) e as respectivas saídas esperadas para essas entradas (*outputs*), são previamente inseridos na rede, como forma de calibrar a rede e o conjunto de pesos atribuídos a cada interação, permitindo que a resposta obtida seja a mais próxima do valor inicialmente conhecido θ .

Considera-se então, generalista, a rede neural capaz de, diante de um conjunto de *inputs* novos, desconhecidos, atribuir um *output* condizente com as respostas observadas no treino. Esta segunda etapa do processamento da RNA se chama teste da rede.

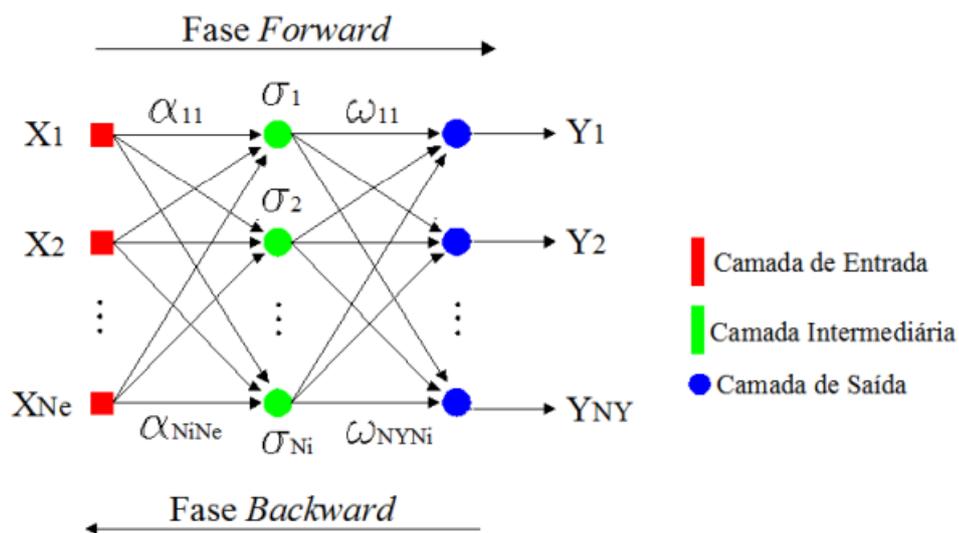
Com um conjunto de dados inicial, pesquisas que utilizam RNA tendem a dividir o banco de dados entre estas etapas da rede, permitindo que, para o teste, o maior número de situações possíveis seja verificado e aprendido pela rede, enquanto, no teste, a capacidade de predição da rede seja verificada em todas as possibilidades de ocorrências

Mesmo não havendo uma definição na literatura quanto a distribuição percentual dos dados para treino e teste, uma distribuição entre 80% para treino e 20 % para teste, como fez Chen *et al.* (2021), ou 60% para treino e 40% para teste, assim como Yu *et al.* (2019), são observadas para previsões de atrasos em voos.

Entretanto, a busca por um conjunto ideal de pesos, w_n , capazes de identificar uma RNA como generalista, possui processos distintos. Um destes processos é denominado algoritmo de retropropagação, *backpropagation*, que segundo Haykin (2001) o algoritmo define o comportamento da rede quanto o seu aprendizado, sendo o objetivo do *backpropagation* a redução do erro entre o limiar e o valor previsto em duas etapas, ainda durante o treino.

A primeira etapa do algoritmo de retropropagação segue o andamento padrão de aprendizado da rede, Figura 5, a fase *forward* então é a responsável pela aplicação de pesos as variáveis de entrada, o processamento nas camadas intermediárias com o intuito de atingir a resposta conhecida, o limiar. Enquanto, a fase seguinte, *backward*, busca verificar se, ao seguir o caminho contrário ao realizado normalmente, os pesos atribuídos para as variáveis buscando atingir o valor conhecido de limiar, seriam os mesmos caso o limiar fosse a entrada para encontrar os *inputs* aplicados.

Figura 5 - Fases *forward* e *backward* do Algoritmo *Backpropagation*



Fonte: MARCY (2017).

Desta forma, as RNAs, com uma configuração *perceptron*, com um número múltiplo de camadas, *multilayer*, além de um algoritmo de aprendizado mais robusto, o *backpropagation*, pode ser considerada uma técnica potencial a ser aplicada em previsões de atrasos em voos.

4.1.1 RNAs aplicadas em estudos de previsão de atrasos de voos

Alguns estudos implementam as RNAs com a configuração *multilayer perceptron* do tipo *backpropagation*, como Khan *et al.* (2021) previram atrasos e duração de partidas de voos, sendo capazes de alcançar uma precisão média de previsão de 65,5%, 61,5% e 59% para classificar o status de atraso e prever a duração do atraso em limites de 60 min e 30 min, respectivamente.

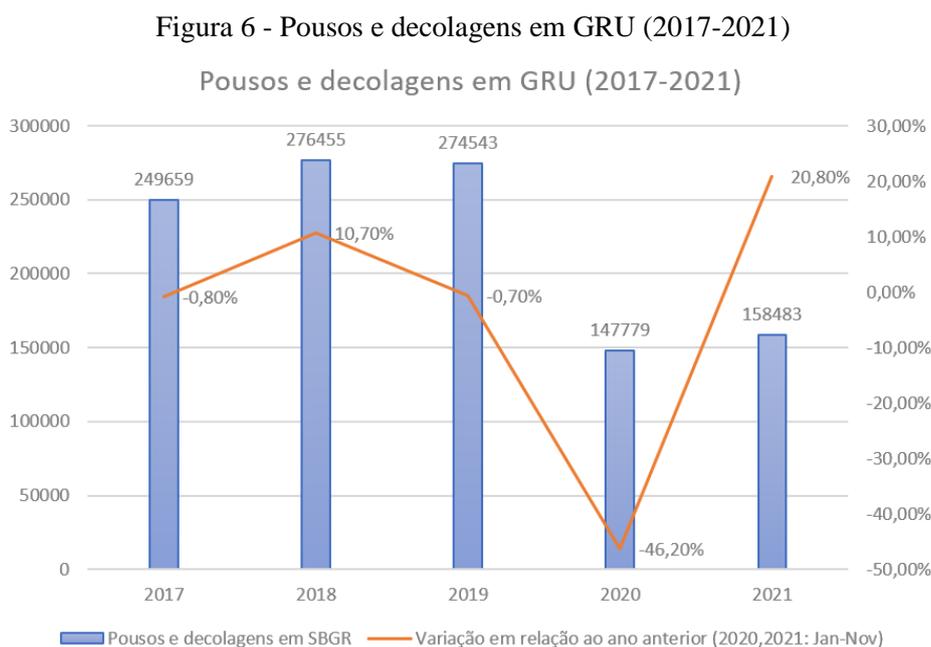
O sistema proposto por Yu *et al.* (2019), utilizando redes neurais do tipo *deep learning*, ou aprendizado robusto devido à alta densidade de dados, consegue ser altamente capaz de lidar com os desafios de grandes conjuntos de dados e pode capturar os principais fatores que influenciam os atrasos. Sendo apontados novos fatores influentes, além dos comumente empregados como *delay* de voos anteriores, número de passageiros por voo, controle do tráfego aéreo, etc., como a situação da rota aérea e o grau de lotação dos aeroportos (com base no número de voos e fluxos de passageiros).

Outro estudo que utilizou as RNAs foi o de Truong (2021), o qual propõe o emprego da metodologia de rede neural *USELEI*, correspondente em tradução livre as etapas de: compreensão (U), amostragem (S), exploração (E), aprendizagem (L), avaliação (E) e inferência (I), para prever o risco de casos de *delay* e não-*delay* em voos do espaço aéreo americano (*NAS – National Airspace System*), obtendo resultados de valores preditivos positivo e negativo de 91,56% e 95,45%, respectivamente.

Portanto, o emprego de RNAs para prever a ocorrência de atrasos é uma realidade em diversos cenários. A massa considerável de dados e de motivos possíveis para a classificação de um atraso acabam exigindo modelos com maior capacidade computacional de dados e generalização das respostas.

5 ESTUDO DE CASO: AEROPORTO INTERNACIONAL DE SÃO PAULO

O aeroporto internacional de São Paulo, o Guarulhos, SBGR, atende a maior metrópole do Brasil e também concentra a maior movimentação aérea da América Latina, com mais de 270 mil pousos e decolagens contabilizados em 2019, o equivalente a 15,75% da movimentação aérea no país para o ano, chegando a 17,03% em 2020, segundo dados da ANAC (2022), Figura 6.

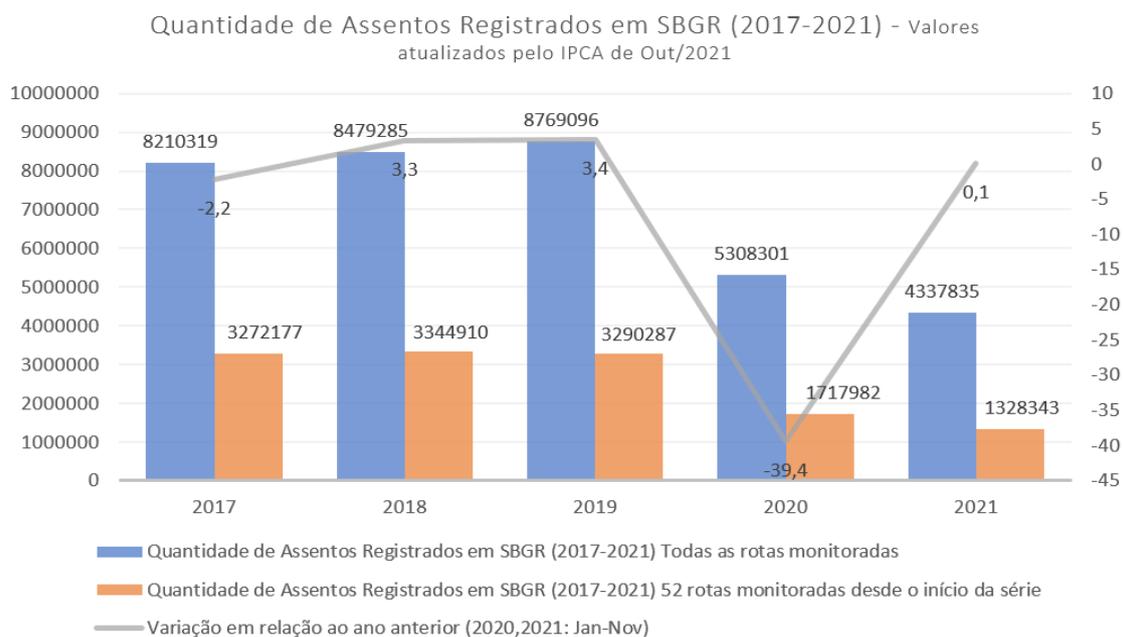


Fonte: Adaptado de ANAC (2022).

Apesar dos efeitos decorrentes da pandemia do SARS-COVID-19, em 2020 SBGR ainda se manteve com mais de 140 mil voos realizados, representando ainda uma queda de 46,20% em relação ao ano anterior.

O transporte aéreo comercial em SBGR pode ser definido majoritariamente como transporte de passageiros, visto que o número de assentos registrados para os períodos de 2017 a 2021, Figura 2, segue o mesmo comportamento verificado para toda a movimentação registrada para o aeroporto no mesmo ano, Figura 7.

Figura 7 - Quantidade de Assentos Registrados em GRU (2017-2021)

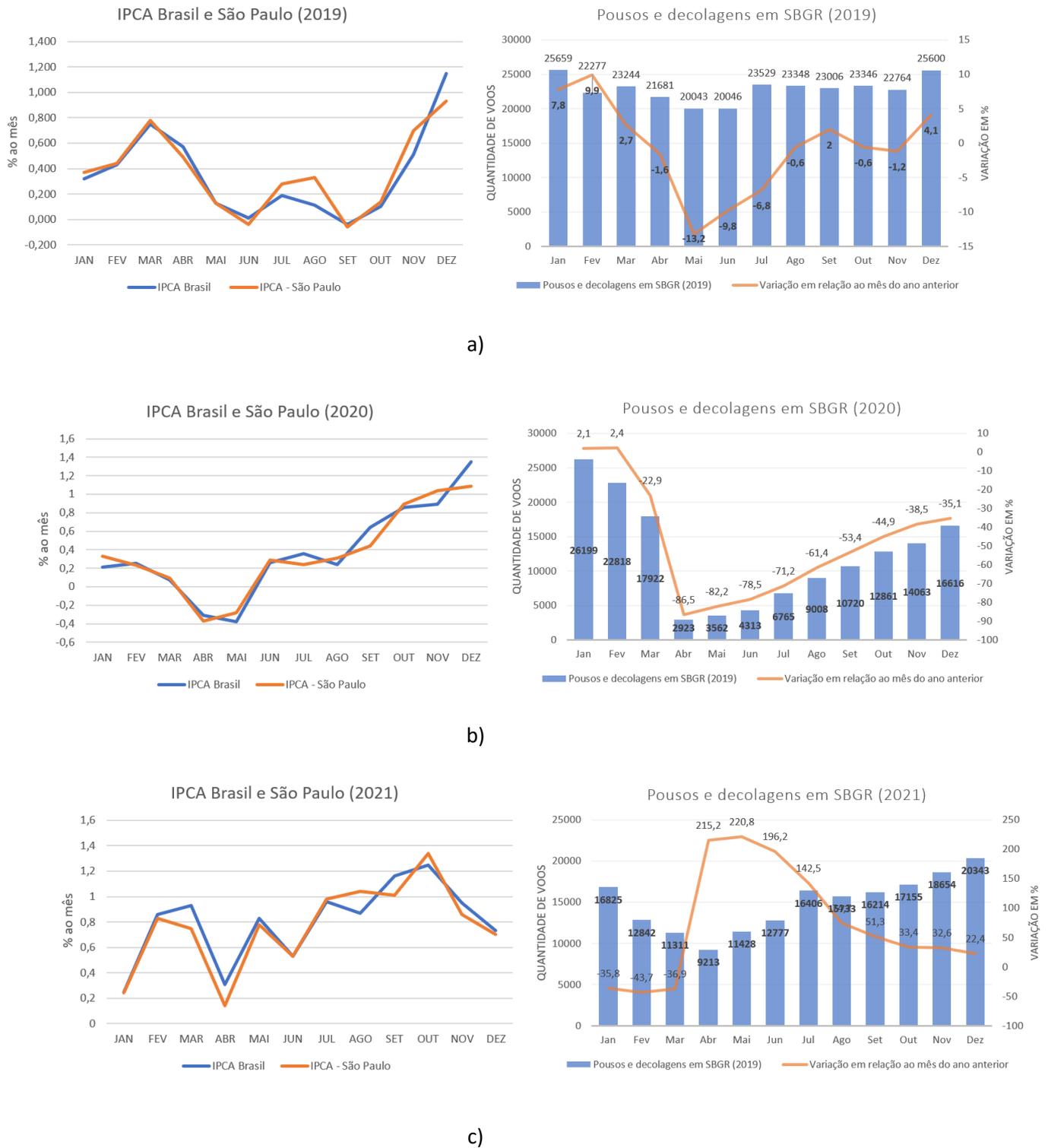


Fonte: Adaptado de ANAC (2022).

O fator econômico faz então parte do conjunto de interferências a movimentação anual em SBGR. O poder de compra dos consumidores é medido pelo Índice Nacional de Preços ao Consumidor Amplo, o IPCA do IBGE, o qual determina o possível crescimento ou decréscimo ocorrido ao longo do ano, sendo um dos fatores que caracterizam a sazonalidade do modo de transporte aéreo na região.

Quando comparadas, as séries históricas dos IPCA para o Brasil e para a região metropolitana de São Paulo-SP, com as ocorrências de pousos e decolagens ao longo dos meses, apresentam um comportamento similar independente do período verificado, sendo que o período anterior a pandemia, Figura 8 a), durante o início da pandemia e crescimento dos casos no país, b), e após a diminuição dos casos e com uma maioria efetiva vacinada, c).

Figura 8 - Relação entre os IPCAs no Brasil e em São Paulo com o número de voos em GRU em 2019, a) ano anterior aos efeitos da pandemia por SARS-COVID-19 iniciarem no país, em 2020, b) durante o início e pico de casos da pandemia, e em 2021, c) início da recuperação do modo aéreo pela diminuição no número de casos.



A movimentação mensal de voos no aeroporto em um ano anterior aos efeitos da pandemia, Figura 8 a), permitem verificar a questão da sazonalidade na demanda aérea de

SBGR. Os meses de dezembro, janeiro e fevereiro são os de maior movimentação no aeroporto, em concordância com as festividades de final de ano e férias escolares, cabendo ao verão o pico anual de voos.

A presença de sazonalidade nos voos de um aeroporto, indica que há uma maior possibilidade de generalização do comportamento observado para cada situação, seja esta benéfica ou não ao fluxo aéreo neste, como ressalta Truong (2021). Em outras palavras, análises preditivas de situações adversas podem ser aplicadas visando a prevenção de possíveis ocorrências de atrasos, por exemplo.

Logo, dada a importância econômica de SBGR para a região metropolitana de São Paulo - SP, além da dimensão considerável das movimentações aéreas anuais registradas; o Aeroporto Internacional de São Paulo, o Guarulhos, se apresenta como uma opção plausível na aplicação de um cenário de previsão da ocorrência de atrasos por categorias.

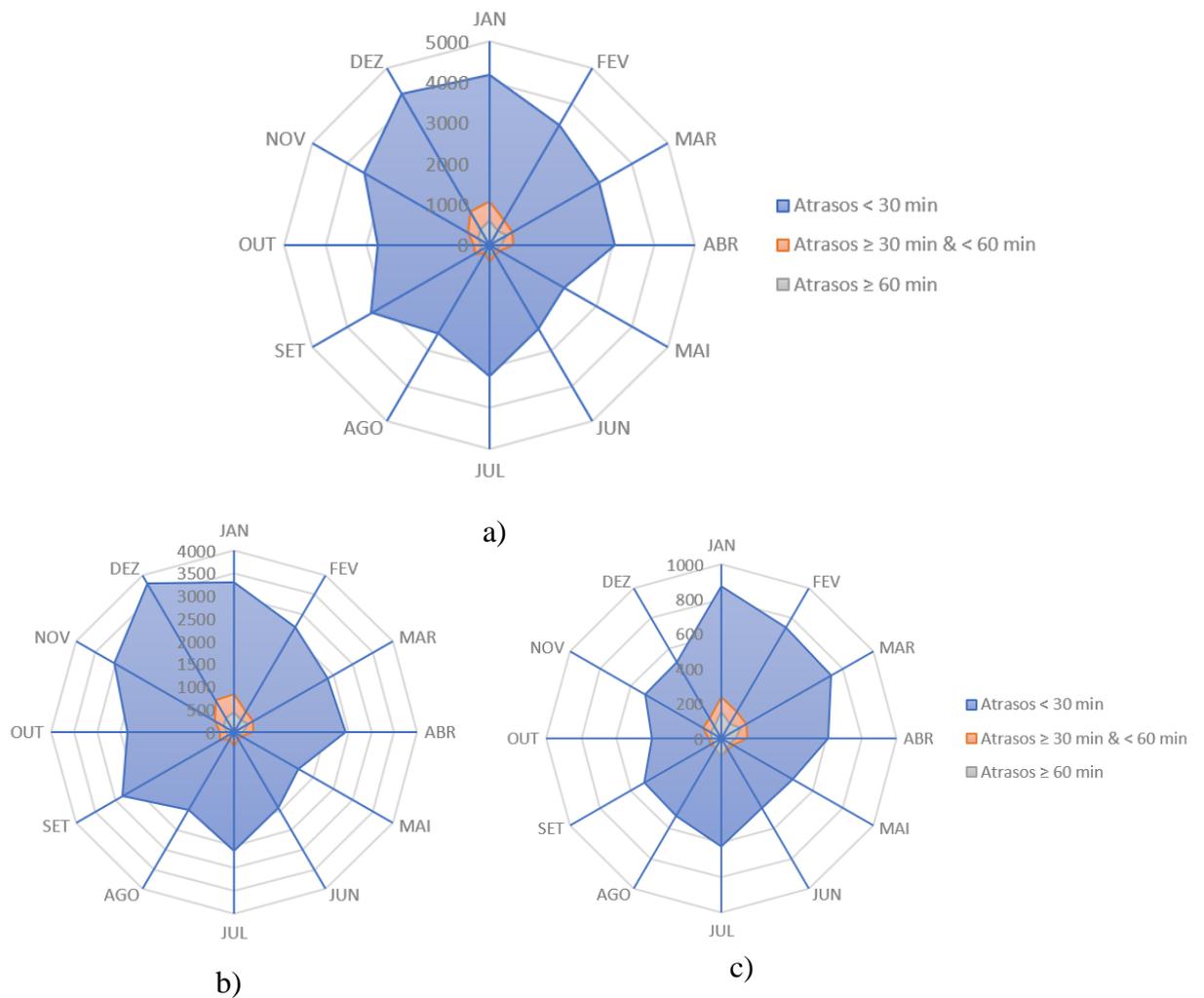
5.1 Análise de atrasos em SBGR

Para a ANAC, um voo pode ser considerado atrasado quando sua partida real for quinze minutos depois do previsto no plano de voo. Para um aeroporto, a condição de atrasos inferiores a quinze minutos pode acarretar um impacto nas atividades menor, economicamente, do que para uma companhia aérea, por exemplo.

No aeroporto SBGR, assim como para todo o Brasil, o ano de 2019 registrou a maior movimentação de aeronaves e passageiros da história da aviação nacional. Conseqüentemente, o número de atrasos decorrentes dessa grande movimentação também foi maior.

Em comparação ao número de pousos e decolagens, Figura 8, a sazonalidade observada para os atrasos se dá de forma similar, de modo que os meses picos de aeronaves registradas também são os meses com o maior número de atrasos, Figura 9 a). Ao agrupar cada atraso em razão do tempo total de ocorrência do atraso, apesar de predominantes os atrasos menores à 30 minutos em ambas as situações, para os voos domésticos, Figura 9 b), voos com atrasos entre 30 e 60 minutos são menos recorrentes do que voos superiores a uma hora, caso contrário na situação de voos internacionais, Figura 9 c).

Figura 9 - Ocorrências de atrasos em SBGR para 2019. A) Voos atrasados totais (Domésticos e Internacionais), B) Voos Domésticos atrasados, C) Voos Internacionais atrasados.



De uma forma mais específica, janeiro foi o mês com a maior movimentação registrada e conseqüentemente o maior número de atrasos verificados em SBGR para 2019. Em números, eram mais de 1200 voos atrasando por semana em média, Tabela 4.

Tabela 4 - Atrasos semanais registrados em SBGR para janeiro de 2019.

Semana-pico de atrasos em janeiro - 2019					
SEMANA	Todos os atrasos	Atrasos < 30 min	Atrasos ≥ 30 min & < 60 min	Atrasos ≥ 60 min	
1 ^a	30/12--05/01	976	724	187	65
2 ^a	06/01--12/01	1404	1014	256	134
3 ^a	13/01--19/01	1159	903	191	65
4^a	20/01--26/01	1348	809	278	261
5 ^a	20/01--02/02	1319	994	219	106

Fonte: Adaptado de ANAC (2022).

Dentre as semanas verificadas, a correspondente aos dias 06 a 12 de janeiro de 2019 foi a com o maior número total de atrasos registrados no aeroporto naquele ano. Entretanto, na quarta semana neste mesmo mês, em SBGR foram registrados os maiores números de atrasos superiores a 30 minutos em todo o ano.

A consequência disto foi uma sexta-feira registrando um total de 88 atrasos superiores a uma hora, Tabela 5. Em comparação com a distribuição sazonal mensal dos atrasos em SBGR para 2019, Figura 9 c), pode-se deduzir que o acréscimo nessa categoria de atraso para este mês se deu pela ocorrência de atrasos em voos internacionais.

Tabela 5 - Atrasos diários registrados em SBGR para a 4^a semana de janeiro de 2019.

Dia-pico de atrasos em janeiro - 2019						
	DIA	Todos os atrasos	Atrasos < 30 min	Atrasos ≥ 30 min & < 60 min	Atrasos ≥ 60 min	
	Domingo	20/01/2019	132	108	12	12
	Segunda-feira	21/01/2019	167	131	27	9
	Terça-feira	22/01/2019	164	124	26	14
	Quarta-feira	23/01/2019	200	118	47	35
	Quinta-feira	24/01/2019	224	117	54	53
	Sexta-feira	25/01/2019	236	104	44	88
	Sábado	26/01/2019	225	107	68	50

Fonte: Adaptado de ANAC (2022).

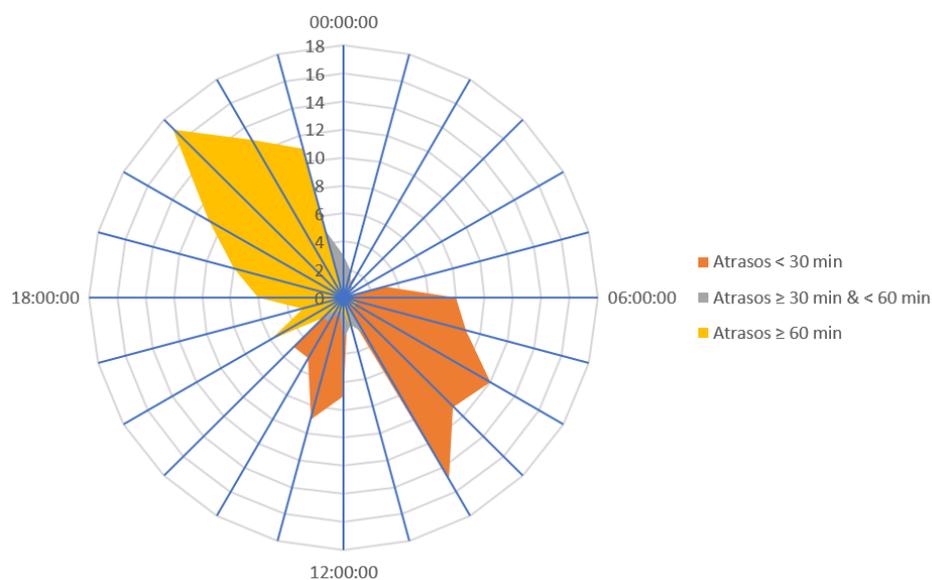
A consequência disto foi uma sexta-feira registrando um total de 88 atrasos superiores a uma hora, Tabela 5. Em comparação com a distribuição sazonal mensal dos atrasos em SBGR para 2019, Figura 9 c), pode-se deduzir que o acréscimo nessa categoria de atraso para este mês se deu pela ocorrência de atrasos em voos internacionais.

Numa distribuição horária ao longo do dia pico, Figura 10, o agrupamento realizado de acordo com o tempo verificado de atraso indicou que os voos concentrados no período noturno foram os que tiveram um maior número de atrasos graves, ou seja, atrasos com

mais de uma hora em relação ao previsto inicialmente. Enquanto atrasos menores se concentram no horário da manhã.

De uma maneira geral, a descontinuidade dos atrasos registrados em SGBR ao longo de um ano pôde ser averiguada, porém, a relação entre as ocorrências destes atrasos com a movimentação de aeronaves verificada para o mesmo período se demonstra positiva. A sazonalidade então atribui um caráter variável aos dados de atrasos em SBGR, ao mesmo tempo em que a capacidade do aeroporto, considerando pousos e decolagens realizados como uma variável representativa de capacidade, influencia diretamente na dimensão destas ocorrências.

Figura 10 - Distribuição dos atrasos ao longo do dia de maior pico em SBGR para 2019



Contudo, existem fatores externos que podem atuar nos números observados para o aeroporto em relação aos atrasos. Ainda em 2019, segundo a ANAC (2022), existiam mais de vinte operadoras aéreas atuantes em SBGR,

. Cada empresa realizou durante o ano voos de caráter doméstico, internacional, regular, não-regular, de passageiros, de carga e outras inúmeras operações possíveis.

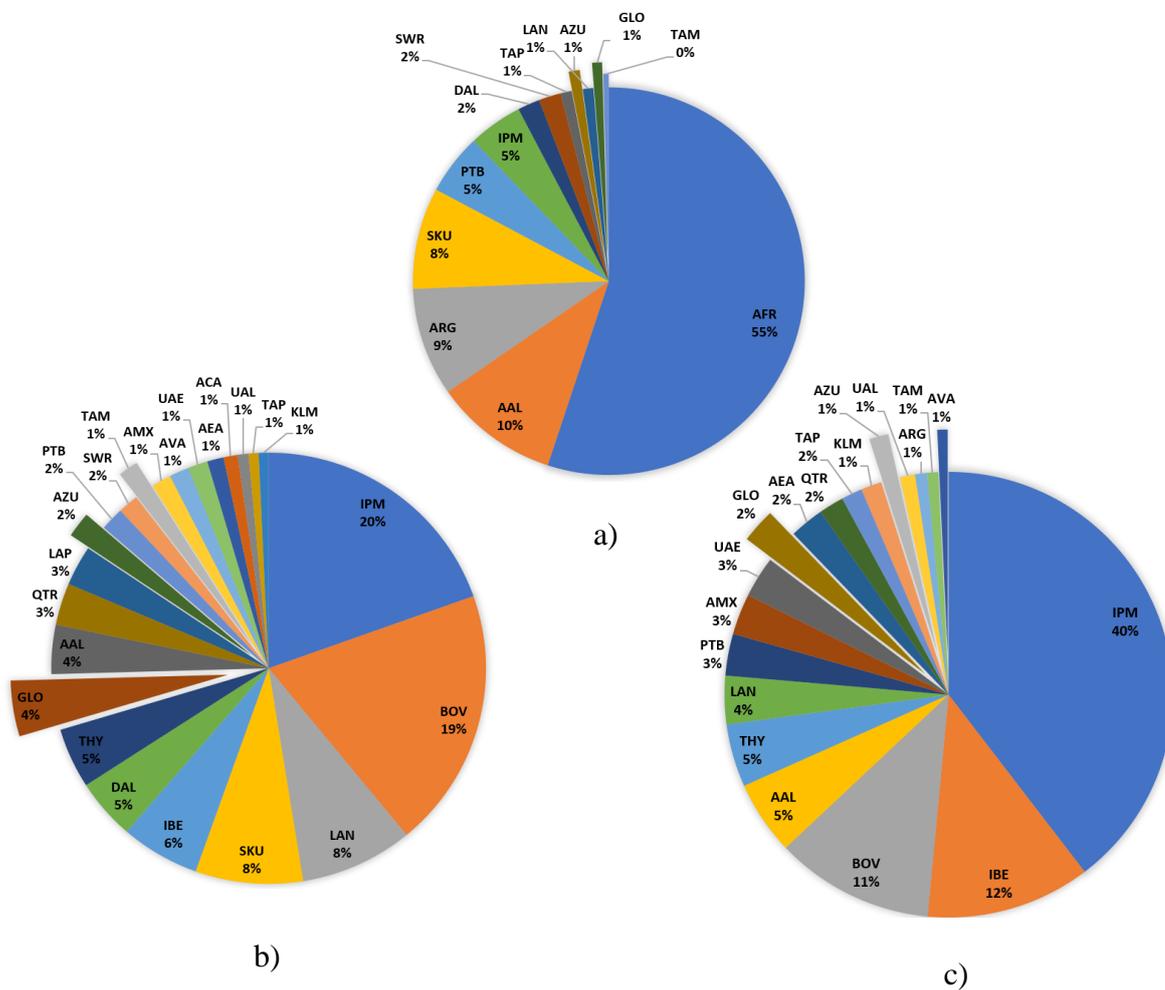
Tabela 6 - Companhias Aéreas em atividade em SBGR em 2019.

Companhias aéreas em atividade em SBGR - 2019		Companhias aéreas em atividade em SBGR - 2019	
SIGLA	Companhia Aérea	SIGLA	Companhia Aérea
	AAL American Airlines, Inc.		IPM Itapemirim Transportes Aereos Ltda
	ACA Air Canada		KLM Klm Cia. Real Holandesa De Aviação
	AEA Air Europa Lineas Aereas Sociedade Anonima		LAN Latam Airlines Group
	AFR Air France		LAP Transporte Aéreos Del Mercosur S.A. (Tam Mercosur)
	AMX Aerovã As De México S.A De C.V		LPE Lan Peru S.A.
	ARG Aerolineas Argentinas S/A		PTB Passaredo Transportes Aéreos S.A.
	AVA Aerovias Del Continente Americano S.A. Avianca		QTR Qatar Airways Group
	AZU Azul Linhas Aéreas Brasileiras S/A		SKU Sky Airlines S.A.
	BOV Boa Boliviana De Aviación		SWR Swiss International Air Lines Ltd.
	CMP Compañia Panameña De Aviacion S.A. (Copa Airlines)		TAM Tam Linhas Aéreas S.A.
	DAL Delta Air Lines Inc.		TAP Tap
	DLH Deutsche Lufthansa A.G.		THY Turkish Airlines Inc
	ETH Ethiopian Airlines Group		UAE Emirates
	GLO Gol Linhas Aéreas S.A. (Ex		UAL United Airlines, Inc
	IBE Ibéria Lineas Aereas De España Sociedad Anonima Operadora		

Fonte: Adaptado de ANAC (2022).

Cada companhia apresentou uma resposta quanto a pontualidade dos serviços prestados em SBGR em 2019. Os dados da ANAC (2022) indicam que, tanto a taxa de cancelamento de voos, Figura L a), assim como os atrasos superiores a 30 minutos verificados, Figura L b), e os superiores a 60 minutos, Figura L c), as companhias internacionais foram as principais responsáveis pelos maiores impactos devido a pontualidade e ocorrência dos voos registrados no ano.

Figura 11 - Distribuição percentual dos voos por Companhia Aérea em SBGR para 2019. a) Em relação aos cancelamentos, b) aos atrasos superiores à 30 minutos e c) aos atrasos superiores à 60 minutos



Fonte: Adaptado de ANAC (2022).

Portanto, apesar da necessidade em especificar as operações de um aeroporto numa análise de previsão de atrasos, a categoria do voo, a sazonalidade ou a própria logística de cada companhia aérea são fatores que podem interferir no resultado obtido. Assim, SBGR pode ser considerado uma área de análise ideal a ser utilizada para validar previsão de atrasos por categorização, visto que, a abrangência devido ao elevado número de movimentações diárias e a diversidade de operações dão ao aeroporto a robustez necessária para teste. Contudo, a análise isolada de uma companhia aérea atuante em SBGR pode indicar não só a influência dessa nos números de atraso contabilizados para o aeroporto, quanto permite um modelo específico de causalidade, ao considerar as distinções entre os modelos de operação e logística utilizados por cada companhia.

6 METODOLOGIA

Com a finalidade de atender os objetivos definidos para a pesquisa, a metodologia foi dividida em duas partes principais. Cada parte descrita neste capítulo apresenta os métodos e considerações aplicadas para a obtenção dos resultados esperados a validação do modelo.

6.1 Primeira parte: Definição do cenário ideal de previsão

Foi considerado como cenário de previsão o conjunto de informações utilizadas quanto a à área de análise, as variáveis de influência nos atrasos, uma faixa de acurácia aceitável e o modelo de previsão mais eficaz a serem utilizados na previsão dos atrasos. A fim de definir o cenário com a configuração mais adequada a realização de previsões de atrasos em voos, a presente pesquisa foi dividida em quatro etapas.

A primeira etapa é a de revisão sistemática de literatura, a fim de que os principais trabalhos publicados a nível nacional e internacional pudessem ser coletados e avaliados. Para a realização desta etapa, foram consultados um dos principais periódicos (Scopus) sendo selecionados os artigos publicados em revistas com alto impacto e recentes.

A etapa seguinte é a de definição da acurácia universal dos métodos de *machine learning* utilizados nos estudos coletados. Nesta etapa foram utilizados os métodos de análise sistêmica de proporção e meta-análise por meta-regressão, a fim de que fossem apontados os modelos com maior acurácia nas previsões de atraso por classificação (definição binária de resposta, em que “0” significa voo sem atraso e “1” voo atrasado) e por regressão (na qual o atraso é classificado de acordo com o tempo esperado de duração deste).

A divisão em cenários permite que a revisão sistemática por proporção e a meta-regressão associem o número de ocorrências de voos em um aeroporto com a proporção positiva de acertos do método de previsão aplicado, indicar a eficiência universal dos modelos de *machine learning* para previsão da ocorrência.

Baseado nas definições de Borenstein *et al.* (2010), essas meta-análises, consideradas “clássicas”, utilizam as proporções (não) transformadas e os erros padrão correspondentes no método de variância inversa, tendo o *logit* definindo o método de argumento "inverso".

As etapas três e quatro são responsáveis por levantar os fatores de influência na acurácia obtida em cada estudo, além de selecionar o método mais eficaz para os cenários definidos, respectivamente.

A complexidade dos dados necessários, as áreas a serem abordadas, seja aeroporto ou linha aérea numa escala universal ou com enfoque climático, foram avaliadas nestas etapas, sendo empregada a Análise Envoltória de Dados, DEA, a fim de caracterizar a eficiência de cada estudo levantado.

6.1.1 1ª Etapa: Revisão Sistemática

A primeira etapa consiste no levantamento e análise dos principais estudos de previsão de atrasos em voos. Com a finalidade de acompanhar as tendências de variáveis de influência aplicadas, as áreas limitantes usadas no estudo, a dimensão dos dados e assim definir os métodos mais utilizados.

Para isso, foram empregados três grupos de palavras-chave de pesquisa, Tabela 7, na base de publicações Scopus, permitindo um levantamento bibliométrico das previsões de atrasos no transporte aéreo.

Tabela 7 - Grupos de palavras-chave aplicadas na pesquisa nas bases de pesquisas.

<i>Grupo</i>	<i>Palavras-chave e strings</i>
01	"Delay" AND "Air transportation" AND "prediction" AND ("Departure" OR "Arrival") OR "Airplane" OR "Airport"
02	"Delay" AND ("Air Transportation" OR "Airport" OR "Airline" OR " Airplane") AND ("Prediction" OR "Prevision") AND "Machine Learning"
03	"Delay" AND ("Air Transportation" OR "Airport" OR "Airline" OR " Airplane") AND ("Forecast" OR "Predict" OR "Estimate") AND "Machine Learning"

O primeiro grupo, de maior dimensão de resultados não especifica o emprego das técnicas de previsão, permitindo verificar a tendência de uso destas técnicas e quais as mais recorrentes ao longo dos anos. Os outros grupos se baseiam na identificação de tendências de uso das técnicas de *machine learning* na previsão de atrasos.

Todas as pesquisas foram coletadas, sendo limitadas as pesquisas oriundas de revistas do tipo engenharia e com ano de publicação num intervalo de cinco anteriores ao atual.

A partir dos resultados obtidos, foi realizada uma análise bibliométrica dos artigos levantados por grupo de palavras-chave. A análise do primeiro grupo, tendo como parâmetro a recorrência simultânea de no mínimo cinco pesquisas, buscou identificar, numa visualização sobreposta ao longo dos anos, o uso de dos métodos predominantes de previsão.

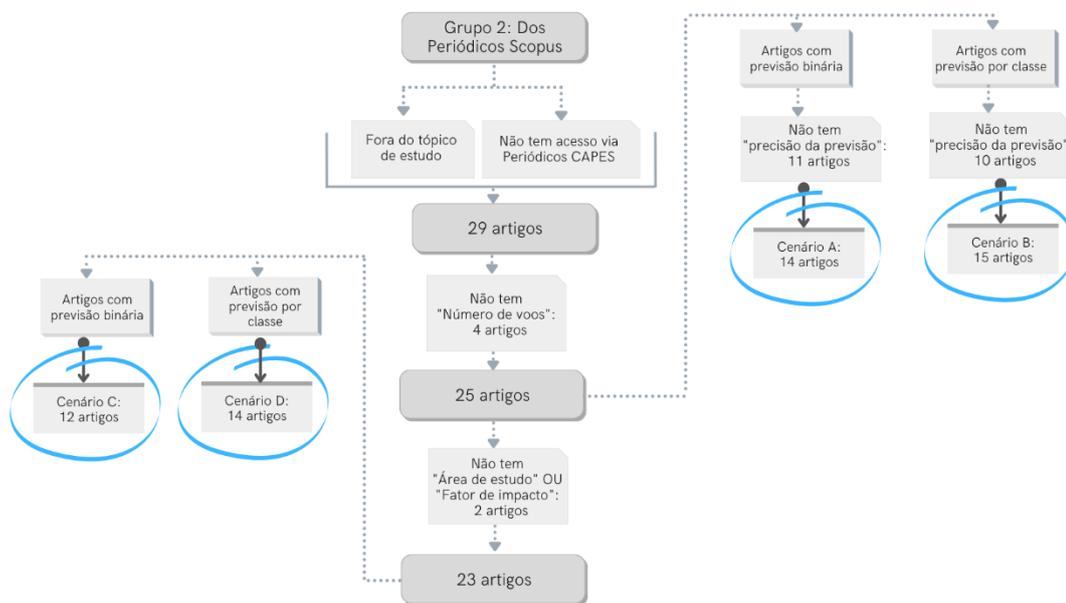
Com a determinação de uma tendência ne uso das técnicas de *machine learning* cada vez mais presente nos estudos de previsão, as análises para os grupos 02 e 03 de palavras-chave buscaram caracterizar quais as linhas de análise mais comuns que utilizem estas técnicas. Sendo selecionados apenas as publicações em revistas, foram separados 35 artigos para o grupo 2. Desta maneira foi possível agrupar os autores co-citados com maior representação nas análises de previsão, mais de 10 citações recorrentes, utilizando o software de análise bibliométrica *VOSViewer*.

6.1.2 2ª Etapa: Meta-Análise

A segunda etapa compreende a definição de um valor de acurácia padrão que deve ser seguido em análises de previsão de atrasos de voos. Para estimar esta eficiência padrão entre os métodos aplicados nos estudos levantados na primeira etapa, sendo estes predominantemente técnicas de *machine learning*, foi aplicada uma meta-análise.

Para tanto, os estudos levantados foram divididos em quatro cenários (A, B, C e D) de acordo com o tipo de previsão aplicada, podendo ser esta do tipo classificação ou do tipo regressão, e a existência de dados sobre o número de voos utilizados as especificidades da área analisada e o impacto do estudo, Figura 12.

Figura 12 - Etapas de seleção dos artigos analisados



Desta forma, os dois primeiros cenários utilizam análises de previsão do tipo classificação e regressão, na qual o cenário A estão contidos os estudos que aplicaram regressões por predição classificatória binária (0 = voo *on-time*, 1 = voo atrasado), contabilizando 14 artigos. Enquanto para o cenário B, as previsões analisadas foram do tipo regressão, ou seja, o modelo previa e atribuía os atrasos classes definidas pelos autores, de acordo com o tempo de atraso, contendo 15 artigos analisados.

Contudo, para a aplicação de uma meta-análise para medir a proporção de acertabilidade do método de previsão aplicado em cada estudo, foi estimado o número de instância previstas corretamente, seguindo os padrões de erro tolerável utilizados por cada autor, a qual sendo definida pela razão entre acurácia do método de maior eficácia do estudo com a quantidade de voos levantados para os bancos de dados (XXI).

$$N^{\circ}_{previsões\ corretas} = Acurácia\ do\ modelo\ aplicado * N^{\circ}_{voos\ aplicados\ no\ estudo}$$

(XXI)

Assim, os cenários A e B buscaram definir, através da construção de uma meta-análise por proporção clássica, quais os parâmetros base de acurácia esperada pelo uso de métodos de *machine learning* na previsão de atrasos em voos.

Para isso, a meta-análise por proporção aplicada se baseou no tipo *random effects*, visto que foi esperada heterogeneidade entre os métodos de previsão aplicado nos estudos. Foram consideradas como instâncias de entrada, assumindo uma distribuição aleatória

para a meta-análise, o número de voos considerados na previsão e o número de previsões corretas realizadas, obtidos em XXI. Desta maneira o prisma obtido pela meta-análise dos cenários A e B descrevem a faixa padrão a ser assumida em previsões por classificação e regressão em classes dos atrasos.

Contudo, a influência da dimensão da área estudada, seja um aeroporto, linha aérea ou rota específica, não são especificadas no modelo de meta-análise por proporção clássica. Desta forma, através de uma meta-análise de proporção por subgrupos, os efeitos externos na acurácia observada pelas previsões foram verificados, além do peso da confiança que os periódicos de origem dos artigos atribuem aos estudos, medido através do fator de impacto.

Logo, a meta-análise de proporção por subgrupos indicou os ambientes mais adequados a atingir melhores resultados de acurácia dos modelos preditivos sendo, então, definidos cenários de artigos em que a área de análise (aeroporto ou linha aérea) e de estudo (região analisada, neste caso o país) fossem conhecidos, cenários C e D, Figura 12. O cenário C utiliza a acurácia dos modelos de previsão por classificação binária para caracterizar os estudos, assim como o cenário A. Enquanto, o cenário D apresenta artigos de previsão de atrasos por regressão, como o cenário B.

6.1.3 3ª Etapa: Análise Envoltória de Dados - DEA

Após especificar o parâmetro base de acurácia para cada área de análise estabelecida, a terceira etapa buscou definir o método e as variáveis que auxiliem na obtenção de maiores acurácias em ambos os tipos de previsão dos atrasos, por regressão ou classificação.

A Análise Envoltória de Dados (*Data Envelopment Analysis* – DEA, em inglês) foi adotada visando, principalmente, estabelecer o ranqueamento dos trabalhos pertencentes aos grupos apresentados na segunda etapa. Os fatores considerados no ranqueamento foram então o grau de acurácia dos estudos, através do número total de voos considerados na pesquisa e a quantidade de previsões corretas; além dos fatores de validação e impacto dos artigos, através do número de citações observadas para cada estudo, o ano de publicação e o fator de impacto de cada periódico.

Entretanto, assim como apresentado por Arjomandi *et al.* (2018), o uso do DEA em unidades tomadoras de decisão (DMUs) heterogêneas, os resultados podem se tornar imprecisos e viesados. Assim, os autores incidem ao uso da metafronteira desenvolvida por O'Donnell *et al.* (2008).

Para o caso da análise dos métodos de previsão dos voos, a distinção entre cada método aplicado nos estudos, as áreas analisadas e o tipo de previsão utilizados concordam para a heterogeneidade observada, sendo então estabelecidos grupos K_n de metatecnologias que afetam as relações descritas de inputs-outputs, Tabela 8.

Tabela 8 - Clusters de efeito considerados para a metafronteira

<i>Cluster</i>	<i>Descrição</i>	<i>Alternativas</i>	<i>Valor inteiro considerado por alternativa</i>
K_1	Área analisada	Aeroporto ou Linha Aérea	1 ou 2
K_2	Tipo de atrasos considerados	Atrasos de todos os tipos (Universal) ou atrasos devido a efeitos climáticos	1 ou 2
K_3	Tipo de método de previsão aplicado	Previsão do tipo Classificação ou Regressão	1 ou 2
K_4	Método de previsão utilizado	Árvore de decisão, <i>Deep belief network</i> (DBN), <i>Ensemble learning</i> , Floresta randômica, <i>Gradient Boosting Trees</i> (GBT), SVM - <i>Support vector machine</i> , Memória de curto prazo longa (LSTM), Rede Neural de Convolutacional Gráfica, Rede Neural <i>MultiLayer Perceptron</i> (MLP), Redes Neurais Recorrentes (RNN) ou Simulação.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ou 12

Desta forma, foram desenvolvidos modelos de ranqueamento por DEA metafronteira para todos os artigos levantados pelos modelos C e D, pois estes apresentam o maior número de informações conhecidas dos estudos.

As variáveis selecionadas para o DEA metafronteira definiram a acurácia obtida em cada estudo de acordo com a influência de cada cluster de efeitos verificado na Tabela 8. Logo, foram gerados cenários de análise de eficiência por *cluster*, Tabela 9, permitindo acompanhar os efeitos de fator considerado nos artigos para a acurácia observada.

Tabela 9 - Cenários de metafronteira aplicados

<i>Variável aplicada</i>	<i>Cenários de Eficiência</i>			
	<i>Ef-K₁</i>	<i>Ef-K₂</i>	<i>Ef-K₃</i>	<i>Ef-K₄</i>
Cluster	<i>K₁</i>	<i>K₂</i>	<i>K₃</i>	<i>K₄</i>
Autores	DMU	DMU	DMU	DMU
Citações da pesquisa	Output	Output	Output	Output
Fator de Impacto da revista	Output	Output	Output	Output
Ano de publicação	Output	Output	Output	Output
Tamanho da amostra	Input	Input	Input	Input
Número de unidades previstas corretamente	Output	Output	Output	Output

A entrada dos cenários desenvolvidos, o *input*, caracteriza cada técnica de *machine learning*, pelo tamanho da amostra utilizada. O número de voos previstos corretamente pelo sistema define as DMU's pelas proporções analisadas anteriormente nas meta-análises desenvolvidas. Enquanto ano da publicação e o fator de impacto definem o peso da originalidade e fator inédito da pesquisa.

6.2 Segunda parte: Aplicação do cenário de previsão definido em um modelo de previsão em classes

Após a definição do cenário ideal de previsão de atrasos, buscou-se realizar um estudo de caso para comparar a eficiência do cenário proposto diante da operação de um aeroporto brasileiro e de uma companhia aérea. Esta comparação permitirá indicar a capacidade preditiva do cenário encontrado para uma aplicação o mais próximo da prática de análise e controle de tráfego.

Sendo assim, conforme indicado pela primeira parte da metodologia, as acurácias do modelo buscadas são provenientes do método de *machine learning* denominado Redes Neurais Artificiais, RNAs, sendo verificadas as respostas obtidas para as duas áreas de análise especificadas (previsão universal dos atrasos para um aeroporto e para uma companhia aérea). De forma que, em ambas as situações, a saída da RNA foi a categorização da ocorrência de um possível atraso quanto o tipo de atraso ou o intervalo de tempo de atraso aguardado para o voo.

Desta forma, assim como na fase anterior, a segunda parte da pesquisa foi dividida em etapas, de acordo com os procedimentos necessários a atingir as respostas de saída definidas.

6.2.1 *Etapa preliminar 0: Definição das áreas de análise*

As dimensões das áreas analisadas no caso de estudo indicam a capacidade preditiva do modelo aplicado ao aeroporto e a companhia aérea, distinguindo o grau de competência da lógica operacional de voo aplicada por cada entidade na ocorrência dos atrasos.

Apesar das movimentações e da dimensão apresentada para o Aeroporto de Guarulhos-SP, Capítulo 5, a justificativa apresentada em aplicar as respostas de definição de cenários em um caso brasileiro, fazem com que sejam verificadas as companhias nacionais de maior atuação em SBGR, além do grau de atraso anual apresentado.

Dentre as companhias atuantes em SBGR no ano de 2019,

, três companhias nacionais são listadas com as maiores movimentações em número de passageiros e aeronaves: Latam (TAM), Gol (GLO) e Azul (AZU).

Assim, foram realizadas análises de ocorrência dos atrasos para as três companhias aéreas nacionais atuantes em SBGR (TAM, GLO e AZU), sendo selecionada a empresa com a maior ocorrência de atrasos superiores a 15 minutos. Os dados coletados são de decolagens regulares em SBGR da aviação doméstica e internacional, independente se foram voos de transporte de passageiros e/ou carga no ano de 2019.

6.2.2 *1ª Etapa: Definição da lógica de previsão*

A partir do método de aprendizado de máquina selecionado, as Redes Neurais Artificiais *multilayer perceptron* com algoritmo *backpropagation*, primeiramente, é necessário definir a lógica de previsão a ser adotada.

Com o *output* esperado pela rede sendo a classificação dos possíveis atrasos de voos em SBGR, o tipo de classes que serão atribuídas aos atrasos serão duas. A primeira lógica empregada classificará os atrasos pelos grupos IATA/ANAC de causas possíveis para a ocorrência destes atrasos.

Desta forma, a saída da RNA prevista para a primeira lógica corresponderá a um número inteiro correspondente aos números atribuídos a cada grupo apresentado na Tabela 1. Contudo, a fim de abranger os casos de não atraso, ou seja, quando os voos partem no horário previsto anteriormente para sua ocorrência, houve a necessidade de inclusão de um décimo grupo, com número correspondente de saída 10, referente aos voos não atrasados.

Porém, uma análise do tipo regressão por grupo de atraso, apesar de aplicada a uma situação de previsões para um aeroporto e para uma companhia aérea apenas, a maior abrangência dos casos pode resultar numa razão mais complexa de definição do atraso, devido a amplitude das causas.

Portanto, uma segunda lógica de previsão, com uma dimensão única de classificação sendo adotada (neste caso o tempo), foi necessária. Assim, a resposta aguardada pela rede nesta segunda lógica foi o intervalo de tempo no qual o atraso se enquadra.

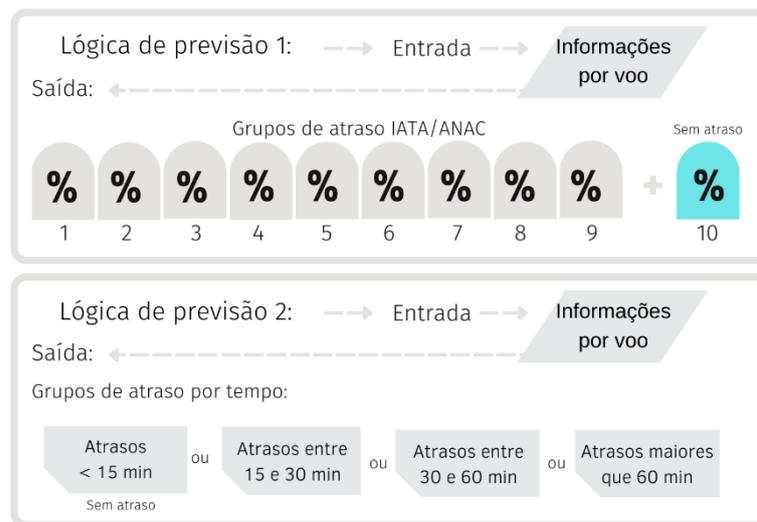
A classificação em intervalo de tempo é empregada por Khan *et al.* (2021), como parte da denominada “análise hierárquica” proposta em seu estudo. Nesta classificação, os autores particionaram o banco de dados para que a previsão do voo ocorra de acordo com o universo que mais favorece seu acerto. Neste caso, o sistema preditivo agrupa os atrasos de acordo com o intervalo de ocorrência, como descrito na Tabela 10.

Tabela 10 - Grupos de atrasos por intervalos, uso da segunda lógica de previsão.

<i>Intervalo de Atraso</i>	<i>Nomenclatura</i>	<i>Descrição</i>
1	Atrasos < 15 min	Atrasos menores do que 15 minutos do previsto. Nesta categoria são incluídos os casos de não atraso também.
2	Atrasos entre 15 e 30 min	Atrasos maiores ou iguais a 15 minutos e com menos do que 30 minutos de ocorrência.
3	Atrasos entre 30 e 60 min	Atrasos maiores ou iguais a 30 minutos e com menos do que 60 minutos de ocorrência.
4	Atrasos \geq 60	Atrasos superiores a uma hora do previsto para a realização.

Assim, os modelos lógicos de previsão podem ser resumidos de acordo com a Figura 13.

Figura 13 - Organização das lógicas de previsão aplicadas no estudo



6.2.3 2ª Etapa: Seleção das variáveis para o modelo

A primeira problemática encontrada ao definir as variáveis explicativas em um modelo de previsão, é a existência de dados compatíveis e que sejam disponibilizados com antecedência a realização da decolagem. De certa forma, utiliza-se valores previstos de operação, meteorologia, carga e passageiros esperados para o voo.

O uso de dados já previstos acaba atribuindo mais incertezas ao modelo. Entretanto, sejam provenientes do operador de tráfego, da linha aérea ou do próprio aeroporto, as informações que alimentarão o sistema devem ser capazes de aproximar as respostas previstas a realidade diária da operação, além de garantir eficácia ao modelo proposto.

Desta forma, foram levantadas as variáveis mais recorrentes em estudos de previsão de atraso, Tabela 3, de forma que os dados utilizados foram limitados as informações disponibilizadas abertamente pelos órgãos de gerência e controle do transporte aéreo brasileiro, como ANAC.

Assim, com base nos grupos de variáveis de influência já utilizados em estudos de previsão de atrasos em voos comerciais, foram selecionadas as variáveis explicativas ao *output* definido para a rede.

Desta forma, foram coletados dados dos relatórios, estudos e informações sobre as condições de mercado aéreo brasileiro disponibilizados pela ANAC. A série histórica dos dados estatísticos do transporte aéreo do Brasil, com elevado grau de detalhamento permitiu a captação das informações por voo realizado no aeroporto de análise.

Para complementar o grupo de dados necessários à categorização dos atrasos por condições climáticas locais, os informes meteorológicos regulares (*METEorological Aerodrome Report - METAR*) para SBGR foram coletados do banco de informações meteorológicas aéreas, o *Iowa Environmental Mesonet*, da Universidade Estadual de Iowa nos Estados Unidos.

Logo, o conjunto de variáveis disponibilizadas e calculadas que irão compor o estudo é apresentado na Tabela 11, junto com a descrição e o tipo de cada variável aplicada.

Tabela 11 - Conjunto de variáveis aplicadas no estudo de previsão.

<i>Grupo</i>	<i>Autor</i>	<i>Variável</i>	<i>Descrição da variável</i>	<i>Unidade</i>	<i>Fonte</i>
Plano de voo	Lambelho <i>et al.</i> (2018); Schultz & Reitman (2019); Khan <i>et al.</i> (2021); Yua <i>et al.</i> (2019); Arora & Mathur (2020); Truong (2021)	Mês_part_pr ev	Mês de partida previsto	-	Microdados ANAC
		Dia_part_pre v	Dia de partida previsto	-	
		Hora_part_pr ev	Hora de partida prevista	-	
		Min_part_pr ev	Minutos previstos para a partida	-	
		Mês_cheg_p rev	Mês de chegada previsto	-	
		Dia_cheg_pr ev	Dia de chegada previsto	-	
		Hora_cheg_p rev	Hora de chegada prevista	-	
		Min_cheg_pr ev	Minutos previstos para a chegada	-	
		Dia_semana	Dia da semana em que ocorreu o voo	-	
		Dist_voo	Distância a ser voada	Km	
		Operadora	Empresa aérea responsável pelo voo	-	
		Ass_ofertados	Número de assentos ofertados para o voo	-	
		Tráfego Aéreo	Yua <i>et al.</i> (2019); Chen <i>et al.</i> (2021); Liu <i>et al.</i> (2019); Chen & Lin (2021); Truong (2021)	Simul_dec	
Simul_pous	Pousos simultâneos previstos para ocorrer no intervalo inteiro de hora no qual o voo está previsto para decolagem			-	

Operações de voo	Chen <i>et al.</i> (2021); Liu <i>et al.</i> (2019); Yua <i>et al.</i> (2019); Truong (2021)	Payload	Payload do modelo de aeronave a ser utilizada no voo	Kg	
		Peso_max_p rev	Peso máximo de decolagem previsto	Kg	
		Veloc_media	Velocidade média de operação do voo	Km/s	
		Horas_voadas	Horas previstas de voo	-	
Meteorológicas	Chen & Wang (2019); Liu <i>et al.</i> (2019); Oliveira <i>et al.</i> (2021)	Temp_orv	Temperatura do ar no horário do voo, medida a 2 m do solo (°F)	Fahrenheit	
		Temp_ap	Resfriamento do Vento ou Índice de Calor, em Fahrenheit	Fahrenheit	<i>Iowa State University:</i>
		Umid_rel	Umidade relativa do ar no momento do voo	%	<i>Iowa Environmental Mesonet - Iowa State University</i>
		Dir_vento	Direção do vento em graus em relação ao norte verdadeiro	Graus	
		Veloc_vento	Velocidade do vento	Nós	
		Atm	Pressão atmosférica	Polegadas	
	Visib	Visibilidade no momento do voo	Milhas		

6.2.4 3ª Etapa: Formação do banco de dados

Após a coleta de dados referente ao conjunto de variáveis selecionadas para o modelo de previsão, Tabela 11, foi necessário verificar as ocorrências de atrasos ao longo do ano de 2019.

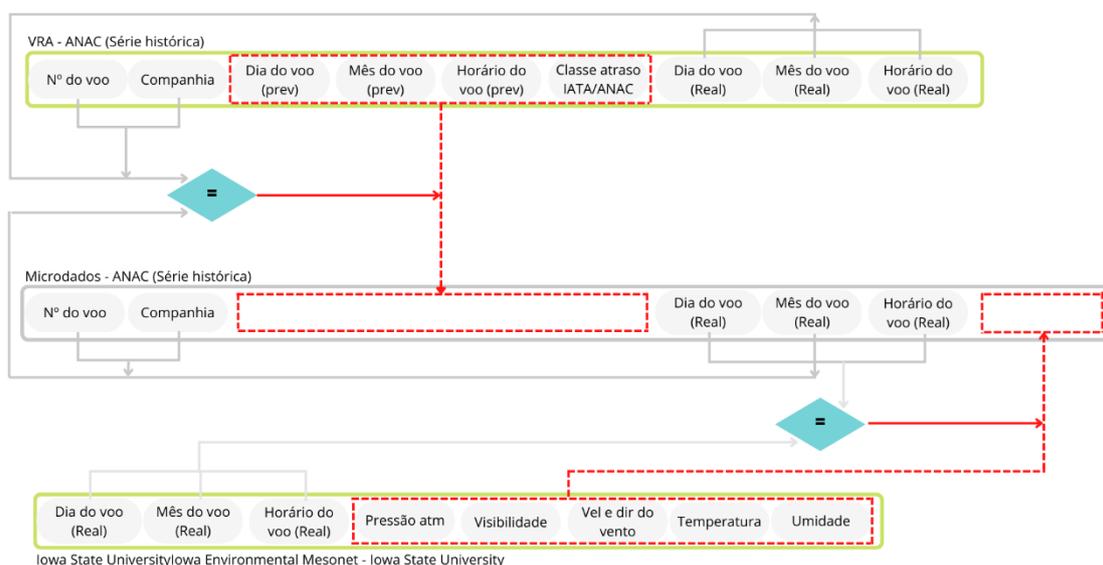
Para isso, os dados foram tratados mês a mês, buscando definir uma relação entre as ocorrências verificadas de atraso, a classe IATA/ANAC do atraso, além do intervalo de tempo observado para o atraso numa perspectiva de SBGR e da companhia aérea selecionada, a GLO.

Porém, para o cálculo do tempo de atraso, a ANAC disponibiliza obtidos através do histórico de voos regulares ativos – VRA, o tempo previsto para a realização das decolagens em SBGR e a classe de atraso IATA/ANAC. Enquanto, a série histórica dos voos disponibilizadas pela companhia possuem os horários de realização concreta de cada partida.

Assim como acontece para os dados meteorológicos, os quais precisam de uma conexão com os dados da série histórica de voos através da data e hora de realização dos voos.

Logo, a construção do banco de dados decorreu da união das informações contidas nos três conjuntos de dados originais levantados, de forma que a similaridade entre elas foi aplicada como o ponto chave para a união dos dados. A metodologia de união dos dados aplicada é apresentada na Figura 14.

Figura 14 - Lógica de união dos dados aplicada aos dados para construção do banco de informações para a pesquisa



6.2.5 4ª Etapa: Pré-processamento dos Dados

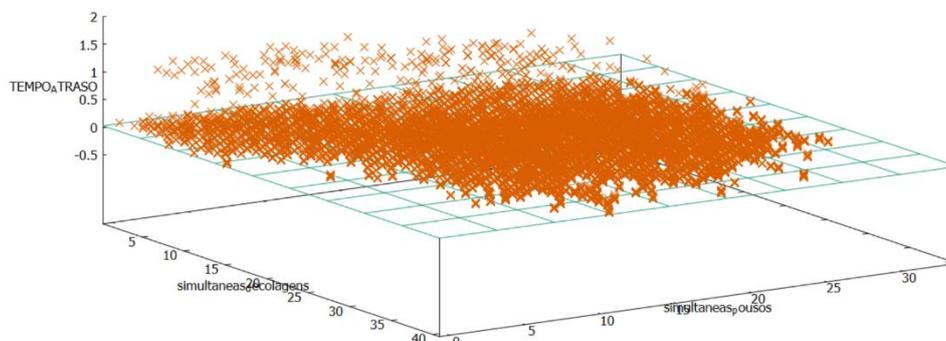
Nesta etapa, o banco de dados formado na etapa anterior é pré-processado, com a retirada de valores nulos (N/A), *outliers* identificados nas variáveis e distribuídos num formato matricial, garantindo uma maior eficácia na computação dos dados.

Os *outliers* definidos seguiram as regras definidas por Yu et al. (2019), nas quais as entradas de dados com valores extremos de atrasos para o 1% superior e inferior são consideradas *outliers* e, portanto, eliminadas do conjunto de dados. Os 98% restantes das instâncias são usados diretamente para análise.

Como exigência do método selecionado, as RNAs com treinamento supervisionado, os dados devem ser distribuídos entre treino e teste. Para isso, foi replicada a proporção utilizada por Chen *et al.* (2021), na qual 80% dos dados são destinados para a fase de treino da rede e 20% para teste.

Entretanto, a elevada dispersão dos dados, visto a dimensão do banco de dados formado, Figura 15, evidenciaram a necessidade de uma redistribuição das classes para a primeira e segunda lógica de previsão.

Figura 15 - Dispersão dos dados levantados: análise espacial por classe IATA/ANAC dos atrasos



Então, foi aplicado o método de balanceamento SMOTE-TOMEK no banco de dados formado, tanto para a distribuição nas classes de atraso IATA/ANAC, balanceamento da primeira lógica, quanto para as classes de intervalo de atraso, balanceamento da segunda lógica.

6.2.6 5ª Etapa: Treino e definição das arquiteturas

Após o pré-processamento do banco de dados formado, a etapa de treinamento da rede é responsável pela generalização dos resultados conhecidos de modo que os fatores de causa aproximem as saídas das respostas aguardadas.

Para isso, as arquiteturas aplicadas as RNAs devem ser capaz de exprimir a maior acurácia na previsão, uma vez que o conjunto: número de camadas escondidas, número de neurônios por camada, funções de ativação aplicadas aliadas ao algoritmo de programação devem buscar replicar, matematicamente, o comportamento dos dados.

Assim, para a pesquisa, foram definidas arquiteturas de rede com base nas áreas de análise selecionadas,

Tabela 12. Os treinos então foram realizados para cada arquitetura, assumindo um total de 500 épocas máximas em cada treino.

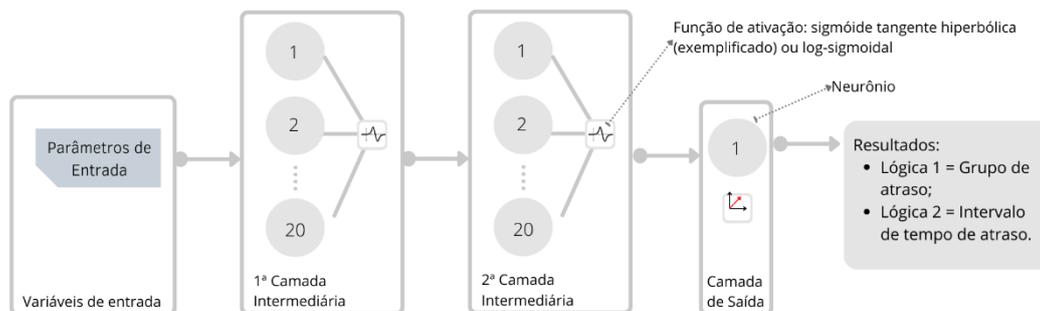
Tabela 12 - Arquiteturas aplicadas no treino das RNAs.

ARÉA	ARQUITETURA DA REDE			
	Nome	Camadas	Neurônios por camada	Função de Ativação
A (Aeroporto - SBGR)	RNA - A1	3	20	Sigmóide tangente hiperbólica

A (Aeroporto - SBGR)	RNA - A2	3	20	Log-sigmóide Log-sigmóide
B (Companhia - GLO)	RNA - B1	3	20	Sigmóide tangente hiperbólica
B (Companhia - GLO)	RNA - B2	3	20	Log-sigmóide

Desta forma, a configuração adotada pelas arquiteturas de rede definidas corresponde ao apresentado na Figura 16.

Figura 16 - Modelo de arquitetura aplicadas as RNAs.



6.2.7 6ª Etapa: Teste da saída e análise dos resultados

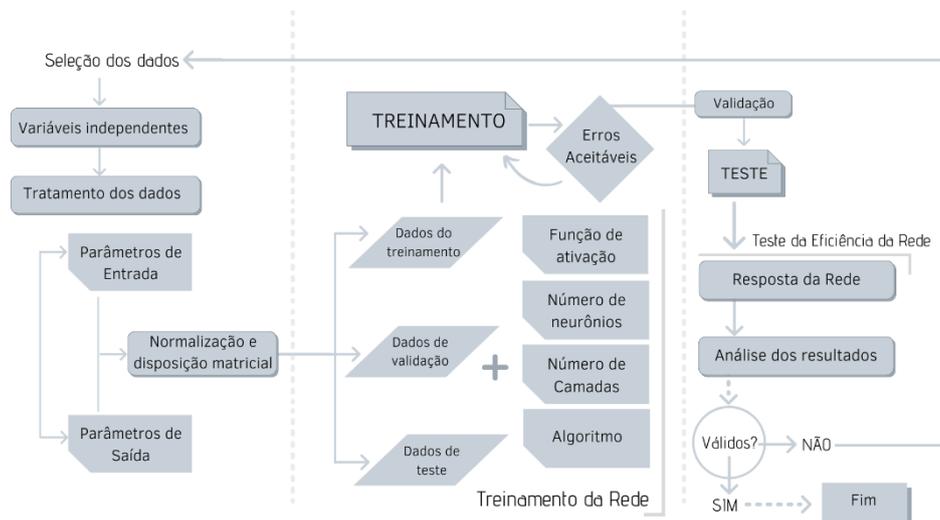
Após a realização do treino da rede, a etapa final da modelagem por rede neural, o teste do sistema com dados desconhecidos aos aplicados no treinamento, é responsável pela avaliação da capacidade de previsão da RNA criada a partir das arquiteturas construídas.

Os resultados de acurácia da rede são medidos pelo F1-Score, de forma que, para a primeira lógica as previsões corretas dos grupos de atraso (grupos 1 à 9) são consideradas os positivos verdadeiros (TP), enquanto que as previsões corretas para o grupo 10 são as falsas verdadeiros (TN).

Na segunda lógica, considerou-se que, todos os atrasos inferiores a 15 minutos são classificados como não-atrasos, devido à baixa dimensão do atraso e pelas considerações da própria ANAC em classificar o atraso de um voo apenas em casos superiores a este tempo.

Assim, um resumo da metodologia para a construção das Redes Neurais Artificiais (RNAs), proposta nesta segunda parte, é apresentado na Figura 17.

Figura 17 - Resumo da metodologia aplicada na segunda parte da pesquisa



As RNAs foram modelagens na linguagem de programação *Python*, com auxílio do ambiente de desenvolvimento de algoritmos *PyCharm*. As linguagens utilizadas foram *Numpy*, para o tratamento numérico dos dados e o *Tensorflow* para a construção das arquiteturas e testes das redes.

7 RESULTADOS E DISCUSSÕES

A partir da metodologia definida, as duas partes foram aplicadas, de modo que os resultados são apresentados conforme as etapas previstas para cada parte.

Para a primeira parte, buscou-se definir o cenário ideal de previsão para atrasos em decolagens, de forma que, os métodos de ranqueamento e validação selecionados foram aplicados em um conjunto de estudos levantados sobre o assunto.

Na segunda parte, o cenário definido (intervalo aceitável de acurácia, área de estudo, tipos de atrasos e modelo de previsão aplicado) foi testado em duas situações comparativas. Para verificar a capacidade de generalização dos dados, além da possibilidade de aplicação do método no mercado, o estudo de caso se baseou em um aeroporto modelo, SBGR, em comparação com as respostas do estudo quando aplicado para uma companhia aérea, GLO.

Assim, os resultados obtidos são apresentados conforme estabelecido para cada etapa, além das discussões levantadas sobre estes.

7.1 Primeira parte: Definição do cenário de previsão ideal

7.1.1 *Revisão Sistemática de Literatura*

A partir do levantamento realizado na base Scopus, utilizando para isso os grupos de palavras-chave e *strings* definidos na Tabela 7, os artigos coletados no Grupo 01 foram distribuídos em nuvem de palavras, Figura 18. Desta forma, foi possível identificar os conjuntos de palavras-chave que obtiveram o maior número de ocorrências nos estudos de previsão de atrasos em voos, além dos métodos utilizados com maior frequência.

O segundo grupo, composto por Kim, Choi, Brinceno e Mavris (2016) definem estudos técnicos de aplicação de métodos de aprendizado para bancos de dados densos, as chamadas *deep learning*, no tráfego aéreo americano. Entre os métodos aplicados, são destaques do grupo as redes neurais recorrentes (*Recurrent Neural Networks* - RNN, em inglês) e o K-ésimo vizinho mais próximo (*k-nearest neighbors algorithm*, em inglês).

Por fim, o grupo de maior impacto de co-ocorrências, com 39 citações, e maior densidade de correlações (medidas por conexões), 982 conexões existentes, é o grupo liderado por Hansen. O terceiro grupo tem enfoque em análises de eficiência de rota (Liu, Hansen, Ball, & Lovell, 2021), na sazonalidade e nas condições climáticas (Gurkaran & Hansen, 2013).

Desta forma, caracterizar a eficiência média dos modelos de *machine learning* em realizar previsões de atrasos dos voos, além de definir uma situação adequada de dados e zonas analisadas, permitirão servir como base a estudos futuros e a aproximar as respostas obtidas a situação real de um aeródromo ou linha aérea.

7.1.2 *Acurácia padrão dos métodos de machine learning*

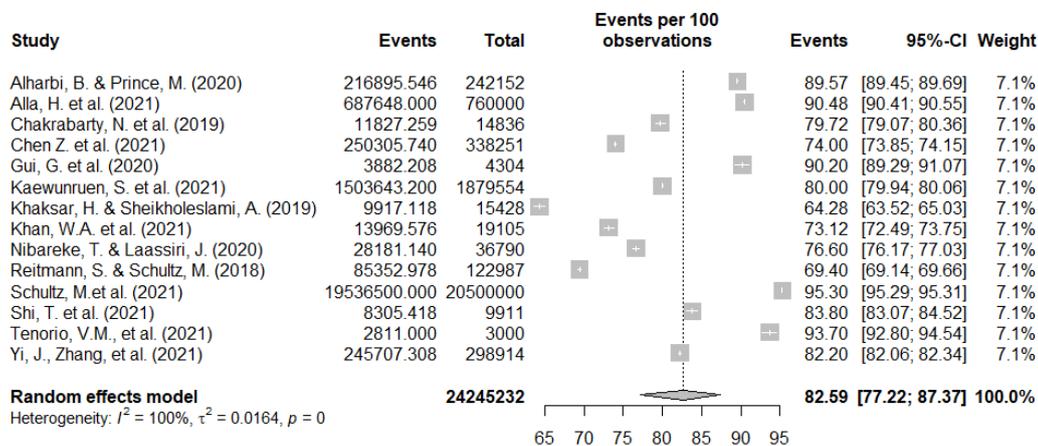
A partir da distribuição dos estudos levantados para o grupo 02, a Tabela 13 demonstra a distribuição estatística para cada cluster estipulado.

Tabela 13 - Estatísticas principais dos estudos levantados para o Grupo 02. Análise por classe de cluster

Cluster	Classe do cluster	Número de autores	Citações (média)	Fator de impacto (médio)	Média dos anos	Total de voos aplicados (Média)	Previsões corretas (média)
K_1	1	16	5,25	6,12	2020	4222741,167	3972479,696
	2	7	17,57	2,04	2020	149947	140028
K_2	1	18	9,33	4,39	2020	2062129	1947442
	2	5	7,8	5,26	2020	4570544	4277154
K_3	1	13	11,92	4,31	2020	1932193	1805966
	2	10	5,2	5,15	2020	4076601	3854052
K_4	1	3	9	5,71	2020	94073	73573
	2	1	2	4,69	2019	517279	481069
	3	2	8,5	4,67	2021	338251	250306
	4	3	5	5,15	2020	945742	756806
	5	4	4,5	4,26	2020	12279	9961
	6	1	8	1,69	2020	33548	26822
	7	2	52	3,59	2019	63646	44618
	8	1	2	10,15	2021	27080431	25680373
	9	3	2,33	2,34	2021	220063	188662
	10	2	3	6,325	2021	10630000	10131530
	11	1	1	1	4,117	2018	2000

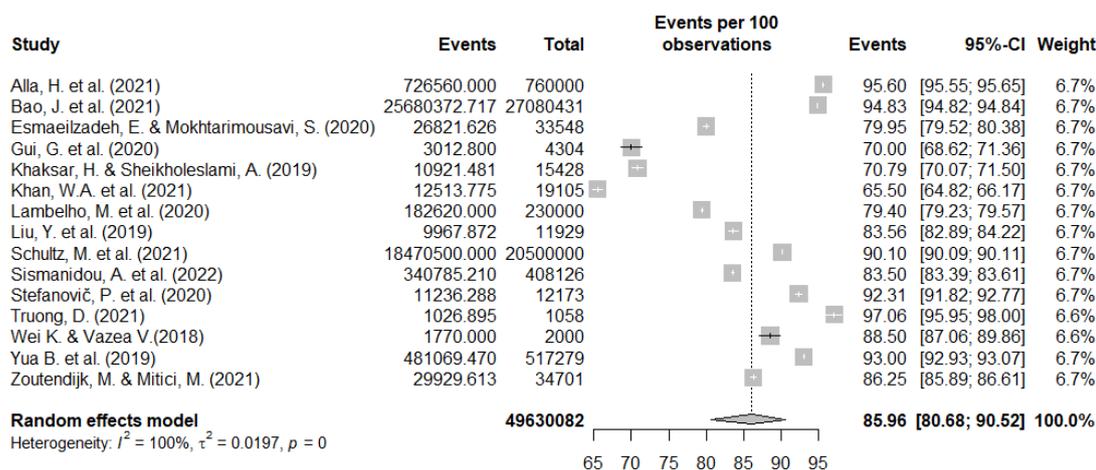
Logo, o resultado do modelo revisão sistêmica randômica para o cenário A, Figura 19, indicam uma dispersão entre os resultados de previsão, devido a distinção entre os métodos aplicados em cada estudo. Além disso, define-se como resultado de acurácia padrão aceitável para os métodos de *machine learning* um valor de 82,59%, com variações entre 77,22% e 87,37 %.

Figura 19 - Cenário A - cenário de regressões por predição classificatória binária (0 = voo on-time, 1 = voo atrasado).



O resultado da meta-análise por proporção para o cenário B, Figura 20, indica uma melhora na acurácia dos métodos de previsão do tipo regressão aplicados nos estudos, em comparação com os métodos de classificação binária do cenário A. O valor de acurácia padrão aceitável para este tipo de modelo de previsão é de 85,96% de precisão, com uma faixa de variação de 4,56% e 5,28% para mais e para menos, respectivamente.

Figura 20 - Cenário B - cenário de regressão dos atrasos por classes definidas



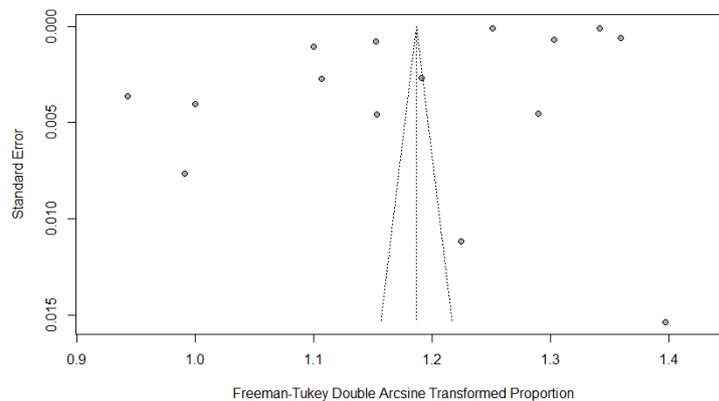
Contudo, como foi observado em ambos os cenários, a alta heterogeneidade entre os estudos é considerável, sendo então necessária a realização da distribuição dos artigos pela assimetria de gráfico de funil, com base no método de regressão linear.

Para isso foi aplicado um teste estatístico baseado no método FPV de *Macaskill et al.* (2001), o qual utiliza uma regressão linear ponderada do efeito do tratamento no tamanho total da amostra com pesos recíprocos à variância da probabilidade média do evento.

A alta similaridade entre os artigos levantados para os dois cenários fazem com que os resultados sejam iguais as duas análises. Desta forma, com uma distribuição t modular de 1,05 e p-valor de 0,3123, a estatística define o descarte da hipótese nula de existência de vieses de publicação entre as pesquisas analisadas.

Contudo, ao se verificar a distribuição do erro padrão pela prevalência normalizada pelo método da proporção transformada em arco duplo de Freeman-Turkey no gráfico de túnel, Figura 21, é possível tomar a conclusão contrária do teste estatístico, já que, apenas um trabalho se enquadra dentro do intervalo de confiança de 95% representado pelo funil.

Figura 21 - Gráfico de túnel, método de Freeman-Turkey.



Portanto, pode se concluir que, apesar da apresentação dispersa dos estudos no gráfico de funil, as análises estatísticas revelam que não há uma tendência viesada dos estudos, sendo este um comportamento característico de estudos de prevalência, os quais não obedecem a uma distribuição normal, podendo ser adotadas as acurácias universais obtidas.

A transferibilidade, ou a capacidade de generalização das pesquisas utilizadas em uma meta-análise para o universo de análise proposto deve ser considerada, visto que, em um grupo disperso e heterogêneo como o apresentado devem, segundo Drummond et al. (2009), se equiparar em relação aos requisitos do tomador de decisão e a disponibilidade de dados para a execução.

Desta forma, ao ser proposta uma análise que servirá como base inicial aos estudos de previsão, além da baixa disponibilidade de dados para o objeto estudado, vale assumir, junto ao teste de Freeman-Turkey, que a generalização é tolerável na aplicação da meta-análise para o presente objetivo.

7.1.3 Fatores de influência a acurácia

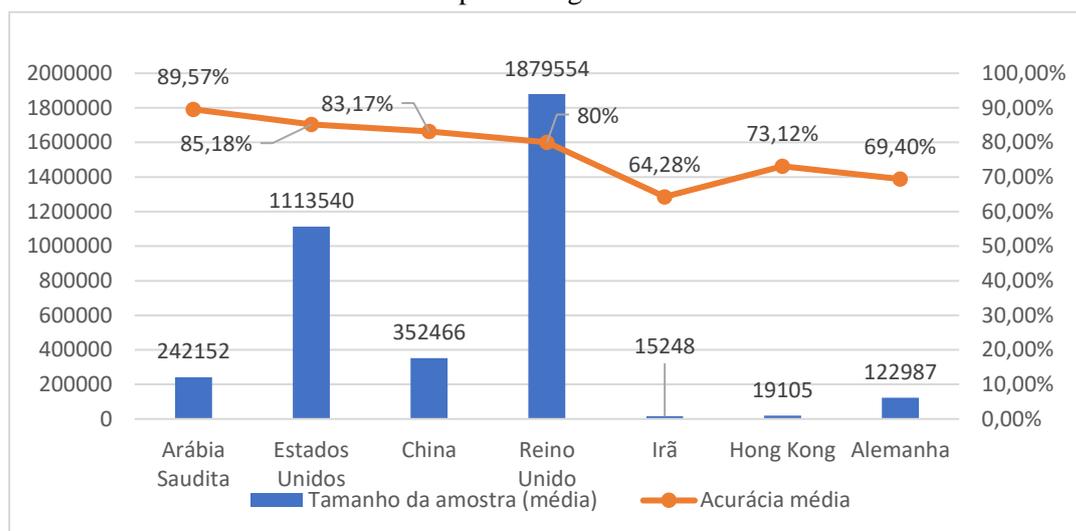
Para as análises do cenário C, os efeitos das áreas utilizadas nos estudos levantados, Tabela 14, demonstram que áreas mais amplas, como prever atrasos para todo um aeroporto, possuem em média menor capacidade assertiva.

Tabela 14 - Cenário C - acurácia dos modelos de previsão por classificação binária para caracterizar os estudos.

Área de análise	Nº de artigos	Acurácia média	Intervalo de variação positiva	Intervalo de variação negativa
Linhas Aéreas	7	85,47%	6,12%	7,44%
Aeroportos	7	79,50%	6,77%	7,72%

Da mesma forma, ao serem verificadas as proporções tendo como subgrupos os países estudados pelos artigos do cenário C, Figura 22. Os resultados apresentam indicativos de que, as análises mais específicas, como a de Alharbi & Prince (2020) que utilizaram dados de companhias aéreas específicas da Arábia Saudita e uma maior quantidade de variáveis conhecidas, obtiveram uma maior eficácia na previsão dos atrasos.

Figura 22 - Proporções de acurácia e tamanho da amostra tendo como subgrupos os países estudados pelos artigos do cenário C



Os resultados das meta-análises de proporção por subgrupos com a área de análise e estudo para o cenário D,

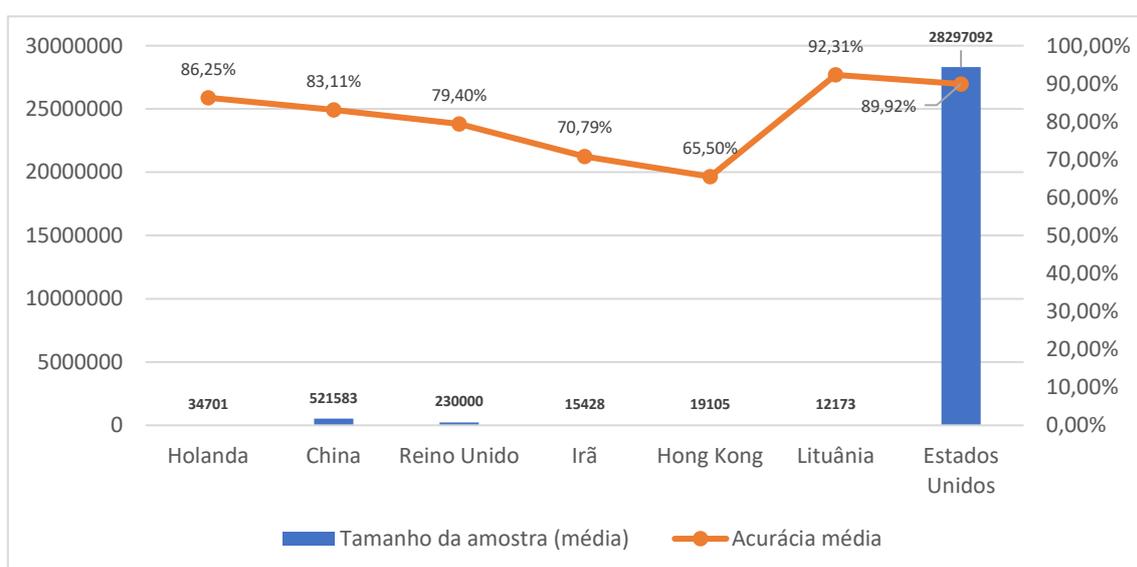
Tabela 15, indicam que os estudos limitados a linhas aéreas apresentaram uma maior margem de erro para a previsão em classes de atrasos (previsão do tipo regressão). Por outro lado, as análises de aeroportos obtiveram uma maior acurácia média e um erro variante tolerável. Uma hipótese para este comportamento é a de que quanto maior a especificidade das informações do voo para a obtenção de uma previsão em classes de atrasos, melhor a capacidade de generalização da resposta pelo modelo.

Tabela 15 - Cenário D - análise de subgrupo com a área de estudo.

Área de análise	Nº de artigos	Acurácia média	Intervalo de variação positiva	Intervalo de variação negativa
Linhas Aéreas	3	80,77%	14,64%	21,47%
Aeroportos	12	85,96%	4,32%	5,05%

Assim como observado para o cenário C, Figura 23, para a análise de subgrupo com a área de estudo, o cenário D apresenta um comportamento no qual as especificidades do país analisado são mais importantes do que a quantidade de dados.

Figura 23 - Proporções de acurácia e tamanho da amostra tendo como subgrupos os países estudados pelos artigos do cenário D



Um indicativo observado em ambas as situações é o de que em países com movimentação aérea inferior a americana, como o caso da Holanda e da Arábia Saudita, o nível de generalização das respostas dos modelos de previsão aplicados é mais eficaz em relação ao tamanho amostral analisado.

Logo, caracterizar o cenário quanto à área estudada está mais relacionada a disponibilização e capacidade dos dados obtidos em descrever o fenômeno, no caso o atraso em um voo específico, do que ao tamanho da amostra ou da região de investigação.

7.1.4 Seleção do método mais eficaz por cenário

Para verificar a eficácia de cada método de previsão aplicados nos estudos levantados nos grupos C e D, primeiramente foi levantado a qual classe dos clusters listados na Tabela 8 cada artigo pertence, Tabela 16.

Tabela 16: Classe de cluster a qual cada estudo analisado pertence. Artigos dos grupos C e D.

<i>Autores</i>	K_1	K_2	K_3	K_4
Alharbi & Prince (2020)	2	1	1	10
Alla et al. (2021)	2	1	2	10
Bao et al. (2021)	1	1	2	9
Chakrabarty et al. (2019)	2	1	1	5
Chen et al. (2021).	1	2	1	3
Esmailzadeh & Mokhtarimousavi (2020).	1	1	2	6
Gui et al. (2020)	2	1	1	7
Kaewunruen et al. (2021)	1	2	1	4
Khaksar & Sheikholeslami (2019)	2	1	1	1
Khan et al. (2021)	1	1	1	5
Lambelho et al. (2020)	1	1	2	1
Liu et al. (2019)	1	2	2	4
Nibareke & Laassiri (2020).	1	1	1	1
Reitmann & Schultz (2018)	1	2	1	7
Schultz et al. (2021)	1	2	1	11
Shi et al. (2021)	2	1	1	10
Sismanidou et al. (2022)	1	1	2	10
Stefanovič et al. (2020)	1	1	2	5
Tenorio et al. (2021)	2	1	1	5
Wei & Vazea (2018)	1	1	2	12
Yi et al. (2021)	1	1	1	3
Yua et al. (2019)	1	1	2	2
Zoutendijk & Mitici (2021)	1	1	2	4

A partir do conjunto de estudos levantados foram desenvolvidos os cenários de DEA metafronteira apresentados na Tabela 9. Os resultados obtidos pelos cenários de DEA metafronteira, Tabela 17, apresentam pouca variação nos resultados nas eficiências por subgrupo, $D_k(x,y)$, e nas eficiências para o modelo DEA total, $D(x,y)$.

Entretanto, ao serem consideradas as razões de meta-eficiência para cada cluster definido, $MRT_k(x,y)$, assume-se, através da relação estabelecida na (XVIII), que a eficiência técnica proveniente de cada grupo $TE_k(x,y)$ seja igual ou inferior à meta-eficiência técnica $TE(x,y)$.

Assim, considerando que a eficiência técnica estará diretamente relacionada a eficiência do método de produção, ou neste caso, o método de previsão aplicado no estudo; os valores de $MRT_k(x,y)$ são utilizados como parâmetro de análise para definir a acurácia adequada em cada cenário.

Desta forma, ao verificar o ranqueamento dos valores de $MRT_k(x,y)$ para o cluster da área de análise K_1 , os estudos realizados na previsão de atrasos em aeroportos inteiros são majoritariamente os menos eficientes tecnicamente. Em contrapartida, estudos que utilizaram análises em linhas aéreas ou rotas específicas, compreenderam mais de 60% dos estudos eficientes.

Esta analogia entre dimensão do cenário de análise e resultados obtidos coincide com o comportamento observado nas meta-análises de proporção por subgrupos aplicados no capítulo 7.1.3.

Para o segundo cluster, K_2 , o tipo de atrasos considerados na previsão demonstra uma descontinuidade do que foi observado para o K_1 . A razão técnica para motivos diversos ou universais de atrasos apresentaram respostas mais satisfatórias de previsão acurada, visto que, apesar de mais específicos, os atrasos devido a condições climáticas podem sofrer com sazonalidade ou variações consideráveis.

O método de previsão aplicado, ou o tipo de previsão, cluster K_3 , caracterizam a classificação binária dos atrasos como menos eficaz, metodologicamente, em comparação a regressão em classes de atrasos.

Nibareke & Laassiri (2020).	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
Reitmann & Schultz (2018)	0,998528	0,998636	0,999892	0,998395	0,999168	0,999227	0,998516	0,998516	1,000000	1,000000	1,000000	1,000000
Schultz et al. (2021)	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
Shi et al. (2021)	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
Sismanidou et al. (2022)	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
Stefanovič et al. (2020)	0,999810	1,000000	0,999810	0,999750	0,999750	1,000000	0,999757	1,000000	0,999757	1,000000	1,000000	1,000000
Tenorio et al. (2021)	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
Wei & Vazea (2018)	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	0,999874	1,000000	0,999874
Yi et al. (2021)	0,999653	0,999653	1,000000	0,999640	0,999640	1,000000	0,999657	1,000000	0,999657	0,999641	1,000000	0,999641
Yua et al. (2019)	0,998922	0,998922	1,000000	0,998890	0,998890	1,000000	0,999010	0,999192	0,999819	0,999578	1,000000	0,999578
Zoutendijk & Mitici (2021)	1,000000	1,000000	1,000000	0,999976	0,999976	1,000000	1,000000	1,000000	1,000000	0,999010	1,000000	0,999010

Por fim, ao serem verificados os métodos de *machine learning* aplicados pelos estudos para prever os atrasos, de maneira geral, o $MRT_k(x,y)$ para as *Gradient Boosting Trees* (GBT), Ensemble learning, Árvore de decisão e Memória de curto prazo longa (LSTM), a acurácia máxima obtida é tecnicamente inferior aos outros métodos descritos.

Logo, para ranquear os métodos mais eficientes para cada cenário estabelecido, Tabela 9, as considerações de eficiência técnica de cada cluster foi somada. Assim, os artigos com o maior número de condições de desempenho atendidas foram listados e ranqueados, Tabela 18.

Tabela 18 - Ranqueamento dos principais estudos que atendem as condições de desempenho demandadas

DMUs	Método aplicado	Condições atendidas			
		$K_1 = 2$ (Linha aérea)	$K_2 = 1$ (Atrasos universais)	$K_3 = 2$ (Regressão em classes)	Acurácia padrão $\geq 80,68\%$
Alla <i>et al.</i> (2021)	Rede Neural <i>MultiLayer Perceptron</i> (MLP)	X	X	X	X
Sismanidou <i>et al.</i> (2022)	Rede Neural <i>MultiLayer Perceptron</i> (MLP)		X	X	X
Bao <i>et al.</i> (2021)	Rede Neural de Convolutacional Gráfica		X	X	X
Yua <i>et al.</i> (2019)	Deep belief network (DBN)		X	X	X
Stefanovič <i>et al.</i> (2020)	Gradient Boosting Trees (GBT)		X	X	X
Wei & Vazea (2018)	Simulação		X	X	X
Zoutendijk & Mitici (2021)	Floresta Randômica		X	X	X
Liu <i>et al.</i> (2019)	Floresta Randômica			X	X
Schultz <i>et al.</i> (2021)	Redes Neurais Recorrentes (RNN)				X

Portanto, de acordo com critérios definidos pelas meta-eficiências técnicas dos trabalhos analisados, somados aos padrões base de acurácia e a dimensão restrita da área de análise, os métodos de previsão por Rede Neural *MultiLayer Perceptron* (MLP) podem ser indicados como os mais favoráveis quanto a acurácia do modelo.

7.2 Segunda parte: Aplicação do cenário de previsão definido em um modelo de previsão em classes

O cenário ideal de previsão definido na primeira parte foi testado em duas áreas de análise, aeroporto e companhia aérea. A distinção de dimensão entre as áreas analisadas permite verificar a eficácia do cenário proposto em aplicações próximas da realidade do mercado aéreo brasileiro.

Para tanto, os resultados obtidos para as etapas metodológicas previstas para a segunda parte são apresentados a seguir.

7.2.1 Etapa preliminar 0 e 1ª Etapa: Definição das áreas de análise e definição da lógica de previsão

Na etapa preliminar, além do aeroporto de Guarulhos, buscou-se definir uma dentre as companhias aéreas brasileiras atuantes no aeroporto para a aplicação do cenário de análise.

Segundos dados disponibilizados pela ANAC (2022), para o ano com o maior registro de movimentações de passageiros e aeronaves em SBGR, 2019, as três companhias principais companhias, Azul, Latam e Gol, totalizaram mais de 35.000 voos atrasados. Destes, 50,22% eram TAM, 37,73% GLO e 12,05% AZU.

Contudo, apesar de ser a operadora com a maior porcentagem de atrasos em 2019, a TAM registra um total de voos com atrasos superiores a 15 minutos menor do que a GLO, Tabela 19.

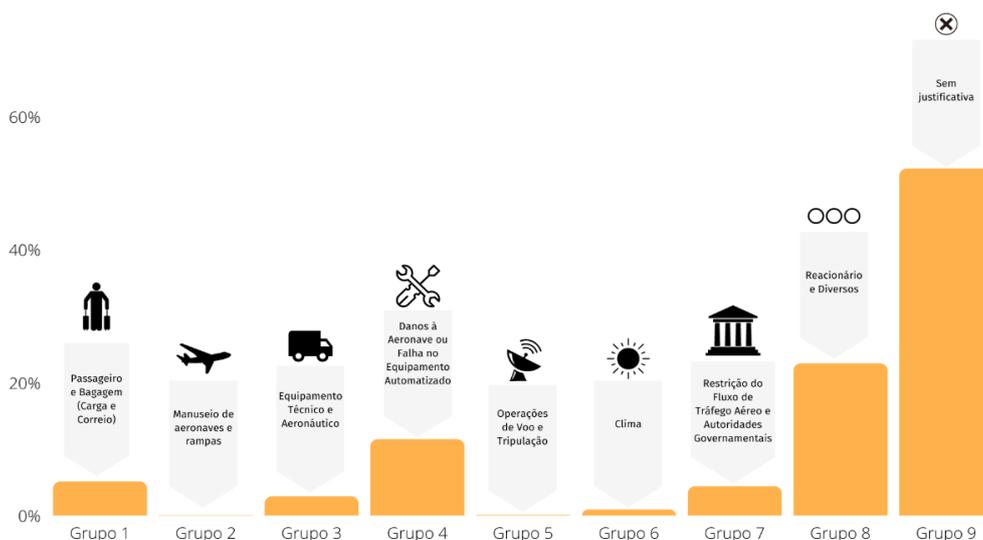
Tabela 19 - Ocorrência de atrasos de voos por companhia aérea em SBGR, ano 2019.

<i>Código da companhia (ANAC)</i>	<i>Sigla ICAO</i>	<i>On-time</i>	<i>Atrasos < 15 min</i>	<i>Atrasos entre 15 e 30 min</i>	<i>Atrasos entre 30 e 60 min</i>	<i>Atrasos > 60</i>
1000909	AZU	64,07%	17,04%	9,17%	6,62%	3,09%
1001093	TAM	59,11%	28,72%	6,18%	3,60%	2,39%
1000854	GLO	47,49%	30,02%	10,57%	7,64%	4,29%

Fonte: Adaptado de ANAC (2022).

Na classificação dos atrasos observados em 2019 para a GLO em SBGR, os grupos da IATA/ANAC predominantes os referentes a problemas com a logística e a frota da companhia, Figura 24. Os casos referentes a reacionários e diversos, grupo 8, especificam situações com motivos não esclarecidos para o atraso, mas também problemas de conexão e gestão do voo, enquanto o grupo 4 se refere a danos da aeronave em solo ou em voo que acabam gerando a ocorrência.

Figura 24 - Causas dos atrasos da GLO em SBGR, ano 2019.



Logo, além do SBGR, a dimensão e as consequências provenientes dos atrasos fazem com que para as análises de companhia aérea, a GLO seja a selecionada para estudo.

7.2.2 2ª e 3ª Etapas: Seleção das variáveis para o modelo e Formação do banco de dados

Após definidas as áreas e lógicas de análise para o modelo de previsão, as variáveis descritivas listadas na Tabela 11 foram coletadas. Com o banco de dados formados, a distribuição estatística das informações de decolagens em SBGR, Tabela 20, demonstra que as movimentações simultâneas no momento das decolagens podem indicar um problema de capacidade como causador de atrasos.

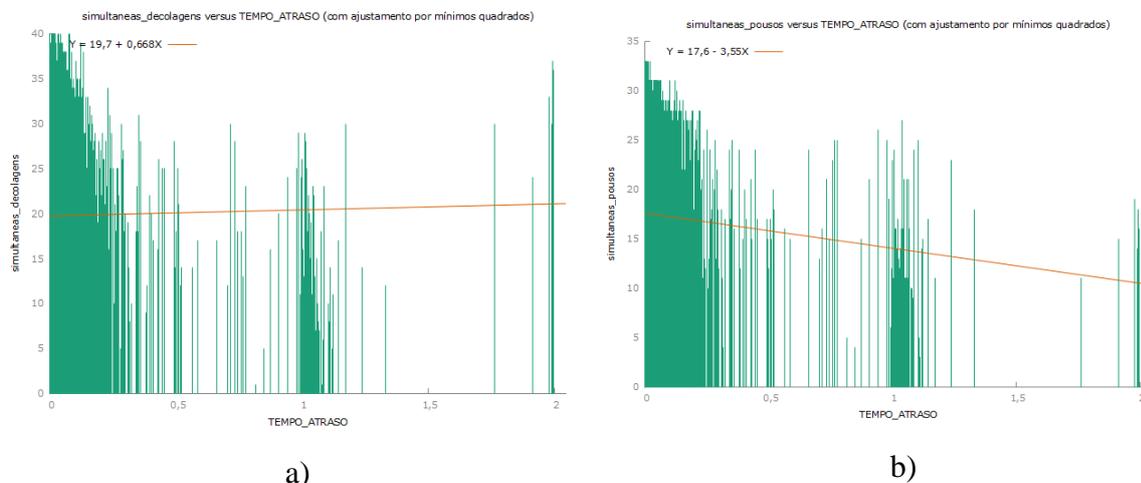
Tabela 20 - Estatísticas descritivas das variáveis explicativas selecionadas.

<i>Variável</i>	<i>Mediana</i>		<i>Desv. Padrão</i>		<i>5% dos dados</i>		<i>95% dos dados</i>		<i>C. V.</i>	
	<i>SBGR</i>	<i>GLO</i>	<i>SBGR</i>	<i>GLO</i>	<i>SBGR</i>	<i>GLO</i>	<i>SBGR</i>	<i>GLO</i>	<i>SBGR</i>	<i>GLO</i>
Mês_part_prev	6	6	3,69	3,77	1	1	12	12	0,56	0,59
Dia_part_prev	15	15	8,68	8,71	2	2	29	29	0,56	0,56
Hora_part_prev	14	14	6,12	5,49	5	7	23	22	0,44	0,38
Min_part_prev	28	31	17,50	17,71	0	0	55	55	0,63	0,58
Mês_cheg_prev	0	6	3,65	3,77	1	1	12	12	0,56	0,59
Dia_cheg_prev	15	15	8,69	8,71	2	2	29	29	0,56	0,56
Hora_cheg_prev	6	12	6,60	6,88	1	0	22	22	0,54	0,54
Min_cheg_prev	12	29	17,32	17,35	0	2	55	56	0,64	0,61
Dia_semana	4	4	1,95	1,97	1	1	7	7	0,49	0,49
Dist_voo	1920	1271	2132,8	808,98	343	343	7738	2474	1,11	0,64
Ass_ofertados	187	174	54,15	19,80	118	138	300	186	0,29	0,11
Simul_dec	20	21	6,28	5,99	10	11	30	31	0,32	0,28
Simul_pous	18	18	5,51	5,29	9	10	26	26	0,31	0,30
Payload	22920	18944	11490	1601,8	12124	15947	52309	19800	0,50	0,08
Peso_max_prev	14971	12708	7630,9	2578,6	7647	8187,6	29280	16371	0,51	0,20
Veloc_media	48234	49044	24857	19943	481	587,92	79166	71787	0,52	0,41
Horas_voadas	2,90	2,15	2,48	1,011	0,93	0,93	9,80	3,742	0,86	0,47
Temp_orv	71	71,15	8,93	8,84	57,20	57,20	87,80	87,80	0,13	0,12
Temp_ap	64	64,19	5,45	5,34	53,60	53,60	71,60	71,60	0,08	0,08
Umid_rel	81,5	81,43	18,21	18,29	46,14	46,14	100	100	0,22	0,22
Dir_vento	132,60	135,42	98,16	98,90	0	0	320	320	0,74	0,73
Veloc_vento	4,90	4,98	2,99	2,97	0	0	10	10	0,61	0,59
Atm	30,07	30,06	0,11	0,11	29,91	29,91	30,27	30,27	0,004	0,004
Visib	5,65	5,58	010	1,28	2,80	2,49	6,21	6,21	0,21	0,23

A relação entre capacidade e os atrasos nas decolagens em Guarulhos pode ser verificado quando correlacionados os tempos de atrasos registrados no banco de dados e o número de voos simultâneos esperados para a hora de ocorrência do voo.

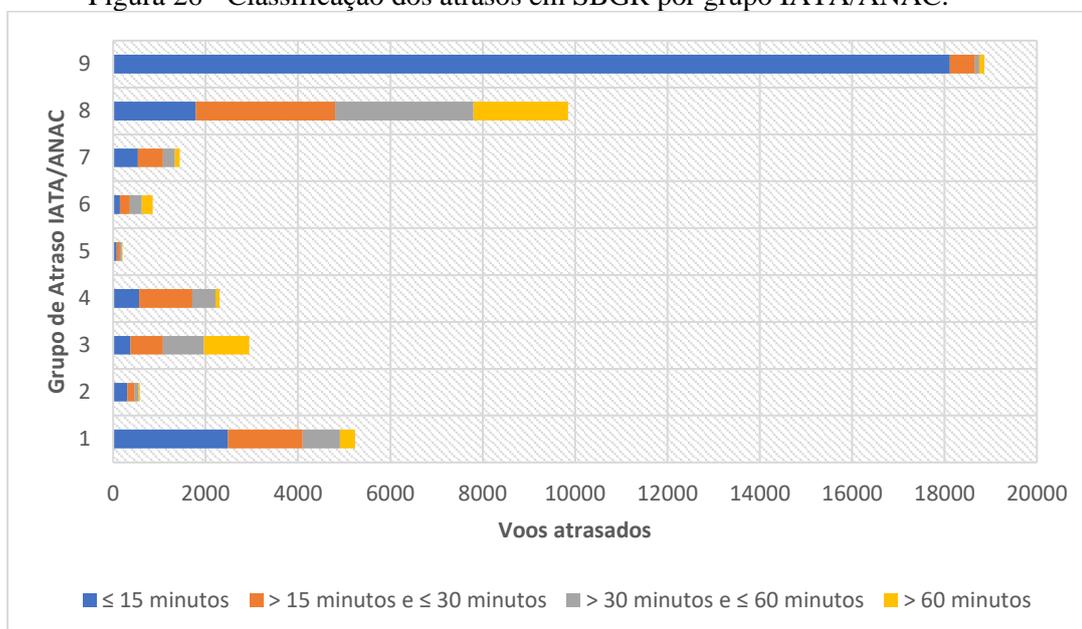
Desta forma, devido o maior tempo necessário para a operação, quanto maior o número de decolagens simultâneas, Figura 25 a), a tendência esperada é o crescimento do tempo de atraso verificado. O contrário ocorre quando os pousos são comparados, Figura 25 b), onde a tendência é decrescente.

Figura 25 - Relação entre o tempo de atraso registrado e as decolagens, a), e pousos simultâneos registrados no intervalo de hora.



A partir da distribuição em grupos de atrasos, Figura 26, mesmo com a predominância das ocorrências sem justificativas, grupo 9, as causas descritas por problemas com conexão, passageiros e bagagens e problemas mecânicos com aeronaves mostram que, a hipótese de falhas na capacidade do aeroporto para a movimentação de aeronaves possa não ser ter significância.

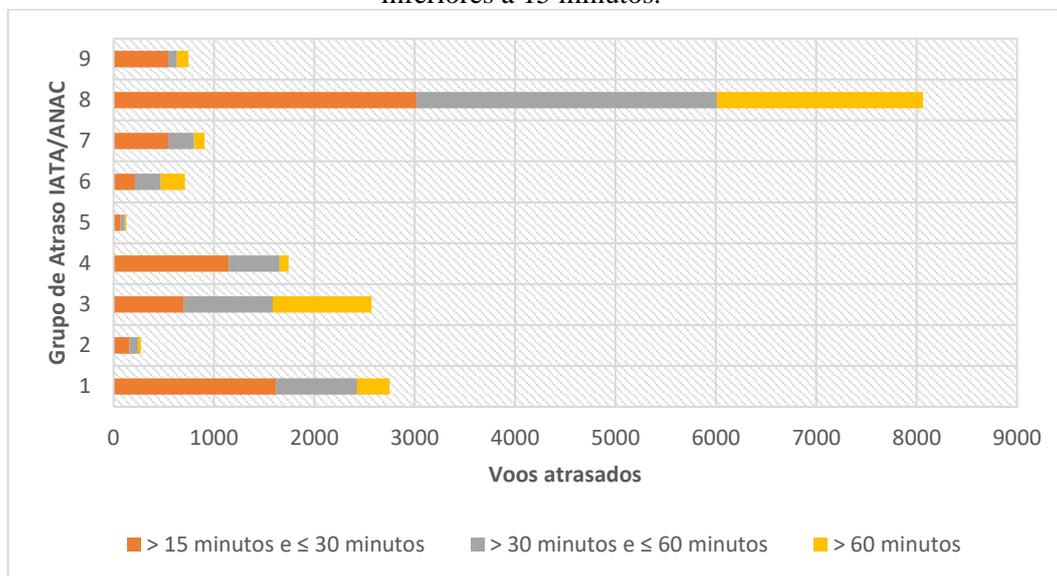
Figura 26 - Classificação dos atrasos em SBGR por grupo IATA/ANAC.



Ao não serem considerados os atrasos de menor dimensão, inferiores a 15 minutos, Figura 27, a mesma distribuição ocorre, porém, evidenciando as principais causas de atraso no aeroporto sendo por motivos diversos ou falhas de conexão, grupo 8, devido ao embarque e

despacho de passageiros e bagagens, grupo 1, e devido a problemas com equipamentos técnicos e aeronáuticos, grupo 3.

Figura 27 - Classificação dos atrasos em SBGR por grupo IATA/ANAC, exceto atrasos iguais ou inferiores a 15 minutos.



Outro ponto apresentado na Tabela 20 são as estatísticas dos dados coletados para a companhia aérea selecionada para análise, a GLO. Em comparação ao comportamento das variáveis em todo o aeroporto, para a GLO as distâncias atendidas demonstram uma predominância em voos domésticos.

Em consequência de operações domésticas predominantes, o uso de aeronaves com menor capacidade de assentos é verificada nas estatísticas dos dados. Entretanto, a frota se apresenta como um dos grupos de maior ocorrência nos atrasos registrados para a GLO em SBGR, Figura 24.

Contudo, o conjunto de dados meteorológicos possui valores similares em ambas as áreas de análise, SBGR e GLO. Em paralelo, o grupo 6, referente aos atrasos devido a problemas climáticos em SBGR no momento da decolagem, possui uma pequena margem de ocorrência em comparação aos demais grupos. Desta maneira, a logística adotada pelo aeroporto e pela companhia possuem maior impacto nos resultados observados.

7.2.3 3ª e 4ª Etapas: Formação do banco de dados e Pré-processamento dos Dados

O banco de dados formado pelo conjunto de variáveis descritivas selecionadas, Tabela 11, junto aos *outputs* definidos para as lógicas de previsão em classes construídas, Figura 13, possuiu um total de 189.547 voos inicialmente.

Através do processo de pré-processamento pela eliminação de *outliers*, conforme descreve Yu et al. (2019), o número de instâncias no banco de dados final foi de 94.381 decolagens.

Os dados após a eliminação dos *outliers* manteve a relação de distribuição entre classes observada para SBGR, Figura 26. Desta forma, uma problemática apresentada por Khan *et al.* (2021) é evidenciada, para estudos de previsão em classe a capacidade de generalização das respostas pode ser influenciada pela distribuição balanceada das saídas possíveis.

A distribuição do banco então é então caracterizada, Figura 29, de modo que, é perceptível a distribuição com predominância de atrasos de menor dimensão, menores ou iguais a 15 minutos. Para tanto, o uso do método de balanceamento misto SMOTE-TOMEK indicado como o mais adequado em estudos de previsão de atrasos de voos por Khan *et al.* (2021) foi aplicado.

Figura 28 - Balanceamento das instâncias para a segunda lógica de previsão.

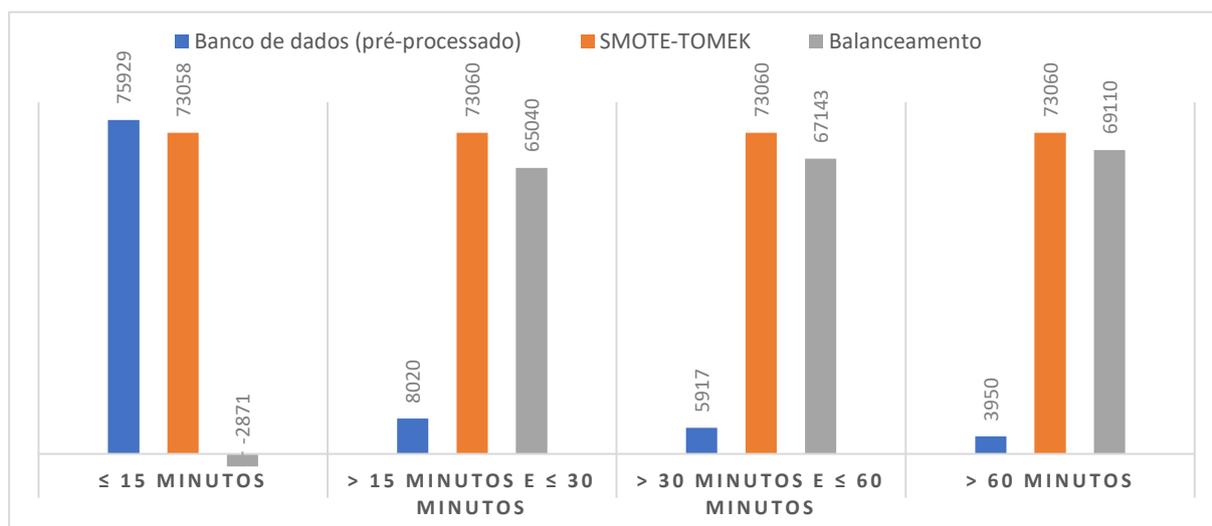
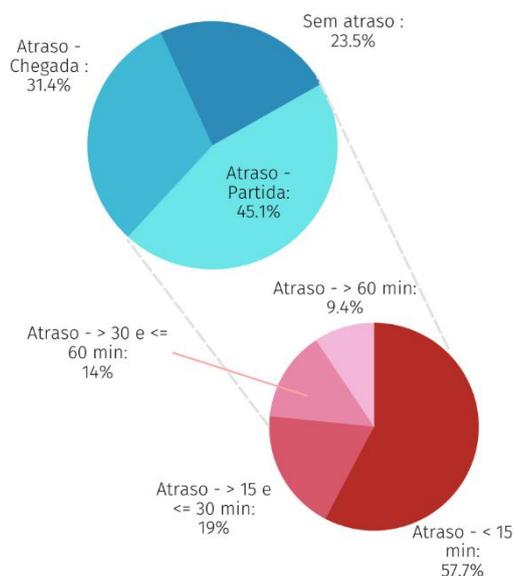


Figura 29 - Proporção de distribuição dos dados, após o pré-processamento.



Com isso, para a segunda lógica de programação, Figura 29, o resultado do balanceamento atribuiu uma quantidade igual de instância para cada classe, de modo que a predominante (≤ 15 minutos) sofreu uma redução no número de decolagens enquanto as demais receberam valores virtuais para compensar o resultado.

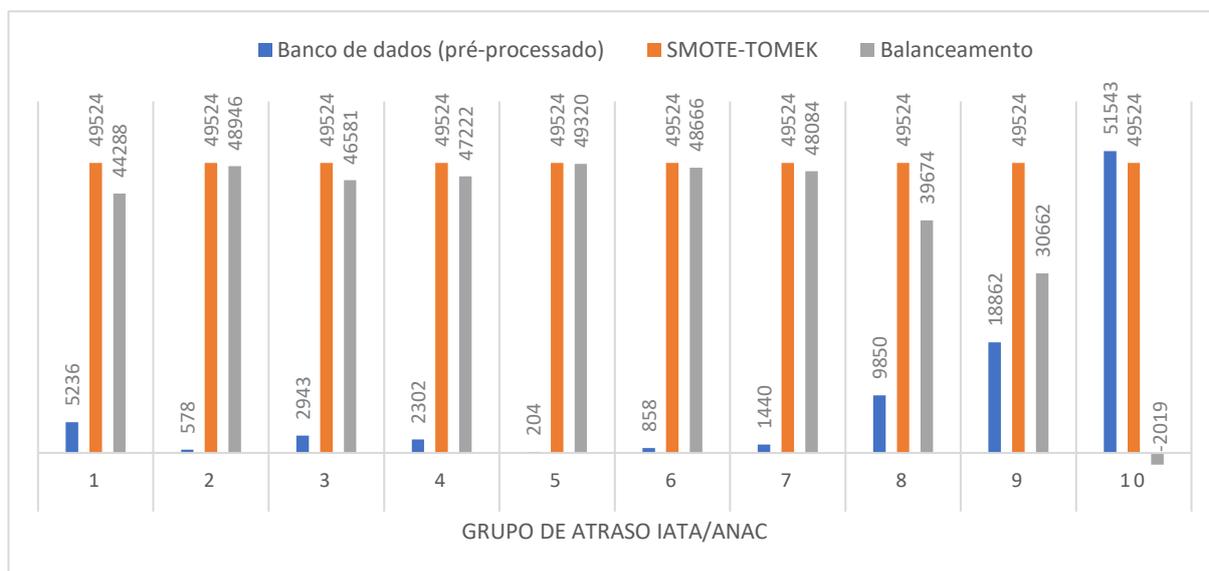
Para a lógica de previsão mais complexa, as classes dos atrasos conforme os grupos IATA/ANAC exigem um balanceamento, pois, apesar de menor o número de ocorrências de atrasos em grupos específicos, o intuito da análise em um aeroporto é de ser capaz de diferenciar o atraso entre as causas possíveis.

Portanto, o resultado para este grupo apresenta balanceamentos consideráveis em números de instâncias virtuais geradas para as classes menos recorrentes, Figura G. O grupo das decolagens *on-time* (sem atraso), grupo 10, é o único com redução no número de instâncias. A escolha em definir igualmente a classificação das decolagens não atrasadas com os grupos de atraso, pode tanto permitir com que a rede atribua corretamente a causa numa análise por voo, mas pode gerar viés falso positivo para os atrasos, visto que a proporção comum entre atraso e voo é positiva para voos *on-time*.

Assim, os bancos de dados balanceados para cada lógica foram selecionados para a aplicação nos modelos de previsão. Como requisito do método de *machine learning* apontado na primeira parte do estudo como o mais adequado para um cenário de previsão com maior

acurácia, as Redes Neurais Artificiais, os conjuntos de dados foram divididos entre treino e teste com as proporções definidas no subcapítulo 6.2.5.

Figura 30 - Balanceamento das instâncias para a primeira lógica de previsão.



Logo, o conjunto de instâncias adotadas para treino e teste em ambas as áreas de análise são apresentadas na Tabela 21.

Tabela 21 - Divisão de instâncias para treino e teste.

Área de Análise	Lógica de previsão	Instâncias usadas	
		Treino (80%)	Teste (20%)
A (Aeroporto – SBGR)	1ª (Grupos IATA/ANAC)	396192	99048
	2ª (Intervalo de atraso)	233790	58448
B (Companhia – GLO)	1ª (Grupos IATA/ANAC)	63981	15995
	2ª (Intervalo de atraso)	63919	15980

7.2.4 5ª e 6ª Etapas: Treino e definição das arquiteturas e Teste da saída e análise dos resultados

As redes foram modeladas em *Python*, utilizando o ambiente de desenvolvimento *PyCharm*. O conjunto de dados divididos em teste e treino foram aplicados as arquiteturas definidas,

Tabela 12, de modo que as respostas do treino puderam ser observadas. Em seguida, os resultados de teste foram divididos de acordo com as considerações de acerto e erro para a

matriz de correlação ser formada. Desta forma, o resumo dos resultados para treino e teste, de acordo com as áreas e lógicas analisadas é apresentado na Tabela 22.

Entre as arquiteturas testadas, Tabela 12, a função de ativação que melhor descreve as relações entre as variáveis selecionadas e a ocorrência de atrasos é a do tipo sigmoide tangente hiperbólica (TANSIG). A variação numérica proporcionada pela função sigmoide permite ao sistema atribuir mais classes respostas à rede, devido à amplitude positiva e negativa da função.

Como o modelo proposto atribui de quatro a 10 classes distintas, quanto maior o intervalo numérico permitido pela amplitude da onda gerada, maior a possibilidade de distinção entre classes.

Essa hipótese pode ser acompanhada pela comparação com as respostas do modelo log-sigmoide, com um intervalo de amplitude apenas negativo. Os testes além de apresentarem uma menor acurácia, segundo o F1-Score calculado, em análise específica dos resultados, também indicaram vieses para os grupos de atrasos em ambas as lógicas.

Outro ponto a ser considerado é a divisão das lógicas de previsão. Como indicado pela primeira parte, grupos mais específicos de análise conseguem atribuir maiores acurácias na resposta do sistema.

Ao serem analisadas ambas as áreas, aeroporto e companhia aérea, os resultados indicam duas situações opostas para cada lógica aplicada. A primeira lógica se baseia na distribuição dos voos em grupos de atrasos, essa divisão, além de possuir o maior número de grupos, agrega variáveis e condições de ocorrência específicas a cada situação. Em suma, pode-se concluir que a primeira lógica de previsão possui uma variedade de unidades consideradas para cada possibilidade de atraso.

Tabela 22 - Resumo dos resultados obtidos pelas RNAs testadas, para performance do treino e acurácia do teste.

<i>LÓGICA DE PREVISÃO</i>	<i>ÁREA</i>	<i>ARQUITETURA DA REDE</i>				<i>TREINO (80%)</i>			<i>TESTE (20%)</i>		<i>ACURÁCIA</i>	
		<i>Nome</i>	<i>Camadas</i>	<i>Neurônios por camada</i>	<i>Função de Ativação</i>	<i>Dados</i>	<i>Épocas</i>	<i>Performance</i>	<i>Dados</i>	<i>Matriz de Confusão</i>	<i>F1-Score</i>	
1ª: Grupo de Atraso IATA/ANAC	A (Aeroporto - SBGR)	RNA - A1	3	20	TANSIG	396192	785	54%	99048	12918 4820	80897 413	0,134591
	A (Aeroporto - SBGR)	RNA - A2	3	20	LOSIG	396192	500	54%	99048	11733 4594	82082 639	0,124909
	B (Companhia - GLO)	RNA - B1	3	20	TANSIG	63981	554	54%	15995	2227 1299	12155 314	0,158862
	B (Companhia - GLO)	RNA - B2	3	20	LOSIG	63981	500	54%	15995	1864 1182	12518 431	0,143482
2ª: Intervalo de atraso	A (Aeroporto - SBGR)	RNA - A1	3	20	TANSIG	233790	500	69%	58448	11905 18011	14453 14079	0,444566
	A (Aeroporto - GRU)	RNA - A2	3	20	LOSIG	233790	500	56%	58448	11716 32090	14642 0	0,200452
	B (Companhia - GLO)	RNA - B1	3	20	TANSIG	63919	500	64%	15980	4037 5119	4190 2634	0,417459
	B (Companhia - GLO)	RNA - B2	3	20	LOSIG	63919	500	58%	15980	4606 7753	3621 0	0,288235

Por outro lado, a segunda lógica de previsão utiliza apenas uma unidade para classificar as entradas do sistema, o tempo de atraso. Somado ao menor número de grupos de resposta possíveis, quatro ao invés de dez, os resultados da Tabela 22 demonstram a maior capacidade preditiva do modelo segundo essa lógica.

Contudo, mesmo a segunda lógica de previsão se apresentando como a mais adequada, devido a acurácia observada entre os testes, a área de análise com maior acurácia é a análise do aeroporto e não da companhia aérea, como preconizou os resultados da primeira parte.

Esse resultado corrobora com o proposto pela primeira parte ao definir o cenário ideal de previsão, uma vez que caracterizar o cenário quanto à área estudada está mais relacionada a disponibilização e capacidade dos dados obtidos em descrever o fenômeno, no caso o atraso em um voo específico, do que ao tamanho da amostra ou da região de investigação.

Entretanto, o conjunto de variáveis selecionadas atribuem mais informações ao aeroporto do que a companhia especificamente. A lacuna quanto a especificidade das variáveis é resultado da falta de disponibilidade das informações sobre a operação realizada em SBGR ou pela GLO. O uso das informações provenientes da ANAC ou da Universidade do Iowa demonstram uma visão macro de cada voo. Além disso, com um resultado de apenas 4% de diferença para a acurácia, pode-se dizer que os modelos praticamente empataram estatisticamente.

Para a acurácia especificamente, ambas as lógicas de previsão em classes não atingem o intervalo aceitável de acurácia definidos pela meta-análise. Com a máxima de 44,46% de acertos pelo F1-Score, a análise do tipo regressão ainda está abaixo dos 77,22% mínimos definidos na primeira parte.

Ao buscar encontrar um sistema preditivo aplicável à realidade de operação vivenciada em um aeroporto e em uma companhia aérea, as variáveis que fossem conhecidas com uma antecedência e que ainda fossem disponibilizadas para estudo, foram os principais limitantes observados.

Em comparação a especificidade das variáveis adotadas, Wei & Vazea (2018), por exemplo, utilizaram os tempos verificados por etapa para o embarque da tripulação de uma companhia aérea, enquanto Scarpel & Pelicioni (2018) utilizaram os tempos estimados para cada operação do voo e a capacidade do aeroporto no momento da decolagem.

Assim, uma hipótese ainda se baseia nas definições realizadas na primeira parte, de que a especificidade dos dados possui uma maior influência na resposta do que a dimensão da amostra.

A distribuição dos resultados nas matrizes de confusão demonstra ainda que, para a primeira lógica de previsão a predominância de falsos negativos (voos atrasados sendo classificados como não-atrasados) ocorrem devido um problema de balanceamento entre os grupos de classificação adotados. De maneira geral, a distribuição em classes acaba priorizando a ocorrência dos atrasos quando são definidos nove grupos de atraso enquanto há apenas um grupo para os casos de não-atraso.

Apesar de ainda significativa para a segunda lógica de previsão, a redução no número de falsos negativos levantados representa um maior equilíbrio das classes adotadas na previsão. Contudo, mesmo com apenas uma dimensão aplicada nas classes adotadas (o tempo de atraso), os grupos destinados a situação de “atraso” para o voo são maiores do que o para casos de não atraso.

Desta forma, mesmo sendo as variáveis de influência selecionadas uma possível causa para os modelos não atingirem as faixas aceitáveis de acurácia obtidas na primeira parte do estudo, a forma de distribuição das classes adotadas em ambas as lógicas pode também ser apontada como uma consequência as respostas observadas, uma vez que foram priorizando os atrasos diante dos casos de não-atrasos.

8 CONCLUSÕES

A partir das discussões apresentadas, é clara a tendência em aplicar métodos mais robustos de análise e regressão de dados para a previsão de atrasos na aviação mundial.

Diante da infinidade de métodos classificados como *machine learning*, alguns se destacam frente à complexidade e disponibilidade das informações necessárias a realização destas previsões. Entretanto, o estudo consegue hierarquizar e definir um modelo de avaliação entre as técnicas empregadas atualmente na literatura.

Como proposta a classificação dos cenários de previsão, o emprego conjunto de meta-análise com o DEA metafronteira permitiu a análise de estudos heterogêneos, desde que a finalidade ou objetivo geral da pesquisa tenha variáveis em comum. Mesmo sendo uma métrica comumente aplicada em uma área distinta, para as análises de engenharia o modelo da meta-análise e metafronteira conseguiram exprimir as respostas esperadas.

Vale salientar que, apesar de indicar o conjunto ideal de área de análise, tipos de atrasos analisados, métodos de previsão e metodologia aplicada, para caracterizar um atraso, a dimensão buscada dos dados deve ser mais próxima as especificidades das etapas a serem realizadas pelos voos do que da quantidade de dados.

Assim, o conjunto de métodos atribuídos à definição de um cenário de previsão indicaram que a área de análise, a especificidade dos dados obtidos e a capacidade de generalização do método empregado corroboram a resposta de acurácia observada.

No estudo de caso realizado, as análises para SBGR se mostraram mais eficientes do que para a companhia aérea selecionada, GLO. Contudo, ao contrário do especificado pelo cenário proposto, a acurácia obtida está fora dos limites aceitáveis. Assume-se que a antecedência exigida para os dados, além da disponibilização de dados menos específicos acabaram auxiliando a esse resultado.

Esta situação evidencia a necessidade de uma análise específica as variáveis, visto que, ao enumerar as causas, especificar as repostas por grupo de atraso pode ser uma alternativa a primeira lógica de previsão utilizada.

Apesar da massa de dados considerável utilizada, ao definir modelos isolados de previsão por tipo de atraso, além de serem utilizadas variáveis específicas para cada saída esperada, a lógica de classificação por grupo e causa de atraso possa obter maiores acurácias tanto no emprego em aeroportos como em companhias aéreas.

Além disso, outro fator limitante a ser considerado no estudo é o emprego de apenas um método de previsão. Mesmo com a capacidade adaptativa das arquiteturas das RNAs, outros

modelos definidos pela primeira parte podem ser testados quanto a faixa de acurácia para o conjunto de variáveis aplicadas.

Portanto, o estudo consegue comprovar que a acurácia dos métodos está atrelada a definição específica da área a ser analisada e dos principais fatores que geram essas ocorrências.

Diante disso, a partir da metodologia aplicada, algumas indicações de trabalhos podem ser enumeradas a fim de ampliar a eficácia da solução, como a obtenção de informações técnicas específicas ao aeroporto analisado, seja com o conhecimento das etapas anteriores e posteriores a cada voo, ou a indicação da instrumentação utilizada para a decolagem ou a pista e cabeceira utilizada.

Além disso, a especificação do uso da meta-análise e do DEA como ponto de partida para os estudos de previsão por aprendizado de máquina pode ser explorado, de forma que se utilize um parâmetro mais específico a cada estudo na definição dos cenários de previsão.

O uso de variáveis de redes para melhorar a estimativa dos modelos, sendo definidas as interações entre estas e o fator de influência. Como também o uso de séries temporais em dados com alta dimensão e frequência buscando a especificidade a cada voo.

Por fim, para a situação da atribuição de classes de previsão, uma alternativa a ser testada é a comparação em série, em que cada grupo de atraso é testado individualmente com o grupo de não atraso para que o o problema da dimensão aplicada seja minimizado.

REFERÊNCIAS

- AENA. (2016). *AENA BRASIL*. Fonte: <https://www.aenabrasil.com.br/pt/aeroportos/aeroporto-internacional-do-recife-guararapes-gilberto-freyre/Caracteristicas.html>.
- ANAC. (04 de 02 de 2021). *Metadados do conjunto de dados: Percentuais de atrasos e cancelamentos*. Fonte: Agência Nacional de Aviação Civil: <https://www.anac.gov.br/acesso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/percentuais-de-atrasos-e-cancelamentos>
- Arjomandi, A., Dakpo, K., & Seufert, J. (2018). Have Asian airlines caught up with European Airlines? A by-production efficiency analysis. *Transportation Research Part A*, pp. 389-403.
- Arora, S. D., & Mathur, S. (2020). Effect of airline choice and temporality on flight delays. *Journal of Air Transportation Management*.
- Button, K. (2019). Applied economics and understanding trends in air transportation policy. *Transport Policy*, 80, 78-85.
- Button, K. (2019). Applied economics and understanding trends in air transportation policy. *Transport Policy*.
- Chen, Y.-H., & Lin, J.-J. (2021). Determinants of flight delays at East Asian airports from an airport, route and network perspective . *Journal of Air Transport Management*.
- Chen, Z., & Wang, Y. (2019). Impacts of severe weather events on high-speed rail and aviation delays. *Transportation Research Part D*, 69, 168-183.
- Chen, Z., Wang, Y., & Zhou, L. (2021). Predicting weather-induced delays of high-speed rail and aviation in China. *Transport Policy*, 101, 1-13.
- Chen, Z., Wang, Y., & Zhou, L. (2021). Predicting weather-induced delays of high-speed rail and aviation in China. *Transport Policy*, 101, 1-13.
- Chicco, D., & Jurman, G. (21 de Junho de 2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*.
- Christopher J. , O., D. S. Prasada , R., & George E., B. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*, pp. 231-255.
- Chung, H. M., Kwon, O. K., Han, O. S., & Kim, H.-J. (2020). Evolving network characteristics if the asian international aviation market: A weighted network approach. *Transport Policy*.
- Chung, S.-H., Ma, H.-L., Hansen, M., & Choi, T.-M. (2020). Data science and analytics in aviation. *Transportation Research Part E*, 134.

- Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value Health*. 2009;12(4):409–18.
- Garrow, L., German, B., & Leonard, C. (2021). Urban air mobility: A comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research. *Transportation Research Part A*.
- IBGE, I. (2020). *Censo 2010: Índice de desenvolvimento humano municipal*.
- Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 387-399.
- Kaye, S.-A., Somoray, K., Rodwell, D., & Lewis, I. (2021). Users' acceptance of private automated vehicles: A systematic review and meta-analysis. *Journal of Safety Research*, pp. 352-367.
- Khan, W. A., Ma, H.-L., Chung, S.-H., & Wen, X. (2021). Hierarchical integrated machine learning model for predicting flight departure delays and duration in series. *Transportation Research Part C*, 129.
- Lambelho, M., Mitici, M., Pickup, S., & Marsden, A. (2020). Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. *Journal of Air Transport Management*.
- Liu, Y., Liu, Y., Hansen, M., Pozdnukhov, A., & Zhang, D. (2019). Using machine learning to analyze air traffic management actions: Ground delay program case study. *Transportation Research Part E*, 131, 80-95.
- Liu, Y., Yin, M., & Hansen, M. (2019). Economic costs of air cargo flight delays related to late package deliveries. *Transportation Research Part E*, 125, 388-401.
- Luger, G. F. (2004). *Inteligência Artificial: Estruturas e Estratégias para a resolução de problemas complexos*. Porto Alegre: Bookmann.
- Marcy, M. C. (2017). *METODOLOGIA BASEADA EM REDES NEURAIS ARTIFICIAIS PARA A DETECÇÃO DE DANOS ESTRUTURAIIS*. Brasília: Tese de Doutorado apresentada a Universidade de Brasília - UnB.
- Odeck, J., & Bråthen, S. (2012). A meta-analysis of DEA and SFA studies of the technical efficiency of seaports: A comparison of fixed and random-effects regression models. *Transportation Research Part A*, pp. 1574-1585.
- Oliveira Filho, P. (2015). *Epidemiologia e bioestatística: Fundamentos para a leitura crítica*. Rio de Janeiro: Editora Rubio.
- Oliveira, M., Eufrásio, A. R., Guterres, M. X., Murça, M. C., & Gomes, R. (2021). Analysis of airport weather impact on on-time performance of arrival flights for the Brazilian domestic air transportation system. *Journal of Air Transport Management*.

Rennert, L. (2022). meta-analysis of the impact of rail stations on property values: Applying a transit planning lens. *Transportation Research Part A*, pp. 165-180.

Scarpel, R., & Pelicioni, L. (2018). A data analytics approach for anticipating congested days at the São Paulo International Airport. *Journal of Air Transport Management*.

Schultz, M., & Reitmann, S. (2019). Machine learning approach to predict aircraft boarding. *Transportation Research Part C*, 98, 391–408.

Simon, H. (1983). Why should machines learn? Em R. S. Michalski, J. G. Carbonell, & T. M. Mitchell, *Machine Learning: An Artificial Intelligence Approach* (pp. 25-37). Berlin : Springer-Verlag.

Singh, V., Singh, V., & Vaibhav, S. (2020). A review and simple meta-analysis of factors influencing adoption of electric vehicles. *Transportation Research Part D*.

Stivali, M. (2022). *Valor de uma vida estatística: uma revisão da literatura empírica para o Brasil*. Brasília - DF. Brasil.: IPEA – Instituto de Pesquisa Econômica Aplicada. Programa de Pesquisa para o Desenvolvimento Nacional - PNPD.

Truong, D. (2021). Using causal machine learning for predicting the risk of flight delays in air transportation. *Journal of Air Transport Management*, 91.

Wang, A., Xu, J., Tu, R., Saleh, M., & Hatzopoulou, M. (2020). Potential of machine learning for prediction of traffic related air pollution. *Transportation Research Part D*.

Wang, C., & Wang, X. (2019). Airport congestion delays and airline networks. *Transportation Research Part E*, 122, 328–349.

Wardman, M. (2022). Meta-analysis of price elasticities of travel demand in great britain: Update and extension. *Transportation Research Part A*, pp. 1-18.

Wei, K., & Vazea, V. (2018). Modeling Crew Itineraries and Delays in the National Air Transportation System. *Transportation Science*.

Yen, B., & Li, J.-S. (2022). Route-based performance evaluation for airlines – A metafrontier data envelopment analysis approach. *Transportation Research Part E*.

Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E*, 125, 203-221.