

Avaliação de Desvios de Conceitos Reais e Virtuais nos Atrasos de Voos em São Paulo nos Períodos Pré, Intra e Pós-Pandemia*

Fabiana Santos¹, Lucas Giusti¹, Diego Carvalho¹,
Eduardo Ogasawara¹, Jorge Soares¹

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

{fabiana.conceicao, lucas.giusti}@aluno.cefet-rj.br,

{d.carvalho, eogasawara}@ieee.org, jorge.soares@cefet-rj.br

Abstract. *Flight delays pose significant challenges to operational efficiency and passenger satisfaction. This study investigates the detection of real and virtual concept drifts in flight delays at São Paulo's main airport (SBSP) during the pre-pandemic, intra-pandemic, and post-pandemic periods. Using the Naive Bayes model and integrating data from the Voo Regular Ativo (ANAC) with meteorological information from NOAA, the study assesses the impact of different concept drift detection methods on predictive performance. The analysis reveals that virtual drift techniques, such as KSWIN, were more effective in stable contexts, while basic methods like Passive and Inactive excelled during the pandemic. These results underscore the importance of selecting and combining detection techniques to enhance the accuracy and adaptability of predictive models across various operational scenarios.*

Resumo. *Os atrasos nos voos representam desafios significativos para a eficiência operacional e a satisfação dos passageiros. Este estudo investiga a detecção de desvios de conceitos reais e virtuais nos atrasos de voos no principal aeroporto de São Paulo (SBSP) durante os períodos pré, intra e pós-pandemia. Utilizando o modelo Naive Bayes e integrando dados do Voo Regular Ativo (ANAC) com informações meteorológicas da NOAA, o estudo avalia o impacto de diferentes métodos de detecção de desvio de conceito no desempenho preditivo. A análise revela que técnicas de desvio virtual, como KSWIN, foram mais eficazes em contextos estáveis, enquanto métodos básicos, como Passivo e Inativo, se destacaram durante a pandemia. Esses resultados enfatizam a importância de escolher e combinar técnicas de detecção para aprimorar a precisão e a adaptabilidade dos modelos preditivos em diferentes cenários operacionais.*

1. Introdução

Os atrasos nos voos representam um desafio significativo para a indústria de aviação, impactando tanto a eficiência operacional quanto a satisfação dos passageiros. A previsão precisa desses atrasos é crucial para mitigar seus impactos adversos e otimizar a gestão aeroportuária. No entanto, o comportamento dos atrasos pode variar ao longo do tempo devido a mudanças nas condições operacionais e externas, um fenômeno conhecido como desvio de conceito (do inglês, *concept drift*). Este trabalho investiga a detecção de desvios

*Os autores agradecem à CAPES, FAPERJ e CNPq pelo financiamento parcial do trabalho

de conceitos reais e virtuais nos atrasos de voos, utilizando dados do principal aeroporto de São Paulo (SBSP) aplicando técnicas de aprendizado de máquina [Gui et al., 2020].

O principal objetivo deste estudo é explorar estratégias de detecção de desvios de conceitos e avaliar a eficácia dessas estratégias em diferentes contextos temporais: pré-pandemia (2018 e 2019), intra-pandemia (2020 e 2021) e pós-pandemia (2022 e 2023). Esta análise comparativa busca compreender como eventos globais, como a pandemia de COVID-19, influenciam os padrões de atraso de voos e como os modelos preditivos podem ser ajustados para manter sua precisão ao longo do tempo. A escolha do aeroporto de São Paulo se justifica pela sua relevância no tráfego aéreo brasileiro, sendo um dos principais concentradores de transporte aéreo no país [ANAC, 2023; Teixeira et al., 2021].

2. Fundamentos

Os atrasos nos voos representam um desafio significativo para a indústria de aviação, impactando tanto a eficiência operacional quanto a satisfação dos passageiros. A previsão de atrasos é crucial para mitigar esses impactos adversos. O uso de modelos de aprendizado de máquina para prever atrasos de voo tem se mostrado eficaz. Modelos como Naive Bayes são comumente empregados como referência devido à sua capacidade de lidar com grandes volumes de dados, identificar padrões complexos e apresentar boa interpretabilidade dos resultados [Moreira et al., 2018]. A integração de dados de voos com informações meteorológicas permite a criação de modelos mais precisos.

A aviação é uma área dinâmica, com empresas surgindo e desaparecendo. Essas, entre outras alterações, interferem na dinâmica dos voos e na perspectiva de atrasos. Nesse contexto, o desvio de conceito acaba sendo bastante comum na aviação e refere-se a percepção de uma mudança significativa na distribuição dos dados que afeta a relação entre as variáveis de entrada e saída ao longo do tempo [Giusti et al., 2022]. Este fenômeno é especialmente relevante em sistemas de previsão baseados em aprendizado de máquina, onde a eficácia do modelo pode ser comprometida pela mudança nos padrões e interrelações entre os dados [Webb et al., 2016]. Existem dois tipos principais de desvios de conceito: real e virtual [Iwashita and Papa, 2019b].

No desvio de conceito real, a relação entre a variável de entrada (X) e a variável de saída (Y) muda. Por exemplo, as pessoas podem preferir bebidas quentes no inverno, em vez de sorvetes. Isso é representado por mudanças nas probabilidades posteriores $p(Y|X)$ [Iwashita and Papa, 2019b].

Já no desvio de conceito virtual, a distribuição das variáveis de entrada muda, mas a relação com a variável de saída permanece a mesma. Por exemplo, a temperatura média aumenta gradualmente devido ao aquecimento global, mas as pessoas ainda compram sorvete apenas quando está muito quente. Isso é representado por mudanças na probabilidade condicional $p(X)$ [Iwashita and Papa, 2019a; Lu et al., 2019].

A detecção de desvio de conceito é crucial para manter a precisão dos modelos preditivos ao longo do tempo. Existem métodos baseados na distribuição de dados e na taxa de erro para identificar mudanças significativas [Lu et al., 2019]. A detecção pode ser passiva, com atualizações constantes dos modelos, ou ativa, acionando o re-treinamento apenas quando necessário [Gama et al., 2014; Iwashita and Papa, 2019b].

3. Trabalhos relacionados

Estudos como Gama et al. [2014] fornecem uma análise abrangente das metodologias de adaptação a desvio de conceito, destacando a eficácia de técnicas passivas e ativas de detecção de desvio. Lu et al. [2019] revisam os métodos de aprendizado sob desvio de conceito, enfatizando a importância de adaptar continuamente os modelos para manter a precisão em ambientes dinâmicos.

No contexto de previsão de atrasos de voos, Giusti et al. [2022] e Teixeira et al. [2021] realizaram estudos utilizando a base de dados do Voo Regular Ativo (VRA) da ANAC. Giusti et al. [2022] exploraram estratégias de detecção e tratamento de desvio de conceito virtual com Naive Bayes, enquanto Teixeira et al. [2021] mostraram que a integração de dados meteorológicos melhora a precisão dos modelos preditivos. No entanto, nenhum estudo combinou Naive Bayes com múltiplas técnicas de detecção de desvio, abordando tanto os desvios reais quanto virtuais em dados de voos e características meteorológicas. Essa lacuna justifica a realização deste estudo, que explora essas combinações no contexto operacional da aviação em São Paulo.

A previsão de atrasos de voos e a detecção de desvio de conceito têm sido amplamente investigadas na literatura recente. Rani et al. [2023] utilizaram Redes Neurais Recorrentes (RNN) para analisar dados de tráfego aéreo durante a pandemia de COVID-19, demonstrando a eficácia do modelo em se adaptar rapidamente a mudanças significativas nos padrões de dados. Este trabalho enfatiza a importância de metodologias adaptativas em cenários de crise, similar ao foco deste estudo.

Além disso, Sakthithasan and Pears [2016] abordaram a captura de conceitos recorrentes em fluxos de dados aplicando a Transformada Discreta de Fourier (DFT) a classificadores de árvores de decisão. Este método permite reutilizar classificadores previamente aprendidos sem a necessidade de reaprendizagem, melhorando a precisão durante os intervalos de recorrência de conceitos. O estudo incluiu um caso de uso com um conjunto de dados de voos, demonstrando os benefícios da aplicação da DFT em um ambiente onde os conceitos recorrentes são críticos.

4. Metodologia

Os dados utilizados neste estudo foram obtidos do Voo Regular Ativo (VRA) da Agência Nacional de Aviação Civil (ANAC), que contém informações detalhadas sobre partidas e chegadas de voos no Brasil [ANAC, 2023]. Esses dados foram enriquecidos com informações meteorológicas fornecidas pela Administração Nacional Oceânica e Atmosférica (NOAA), que registra condições climáticas em aeroportos ao redor do mundo [NOAA, 2023]. O dataset utilizado está disponível publicamente e pode ser acessado na plataforma IEEE DataPort¹.

A integração dos dados envolveu a fusão dos dados de voo do VRA com os dados meteorológicos do NOAA. Formalmente, sejam A o conjunto de dados de voos da ANAC e M o conjunto de dados meteorológicos do NOAA. O conjunto de dados integrado AM é definido como: $AM = f(A, M)$, onde f é uma função que mapeia e combina os dados de A e M com base no período e localização correspondentes.

¹<https://test.ieee-dataport.org/documents/brazilian-flights-dataset>

O conjunto de dados AM foi filtrado para incluir apenas voos partindo do aeroporto de Congonhas (SBSP), resultando no conjunto S , *i.e.*, $S = \sigma_{departure='SP'}(AM)$. Em seguida, os dados integrados foram organizados em lotes semanais para análise. Seja S o conjunto de dados integrados filtrado em São Paulo, tem-se s_i um lote de dados correspondente a uma semana i . Assim, S pode ser particionado em uma sequência de lotes semanais $\{s_1, s_2, \dots, s_n\}$, onde cada lote s_i contém os dados para a semana i .

Para a construção dos modelos preditivos, foi utilizada a técnica de aprendizado de máquina Naive Bayes devido à sua interpretabilidade e simplicidade, o que facilita a prova de conceito. Os principais atributos utilizados incluíram horário de partida, companhia aérea, temperatura, visibilidade, e outras condições meteorológicas. Esses atributos foram escolhidos devido à sua influência reconhecida nos atrasos de voos, conforme indicado em estudos anteriores. A função de treinamento $train$ aplicada a um conjunto de lotes prévios s_{i-1} de dados produz um modelo treinado m_i para uso na semana i , *i.e.*, $m_i = train(s_{i-1})$.

A distribuição de um atributo a em um lote s_i é dada por $\chi_a(s_i)$. Neste estudo, observa-se um desvio de conceito virtual quando $a \in X$ e $\chi_a(s_{i-1}) \neq \chi_a(s_i)$ com significância estatística. Para avaliar esses desvios de conceitos virtuais, foi adotado o método Kolmogorov-Smirnov Windowing (KSWIN), que detecta mudanças significativas na distribuição das variáveis de entrada [Raab et al., 2020].

Já no que tange aos desvios de conceito reais, dado um modelo m , tem-se que a distribuição dos erros das predições de m em um lote s_i é dada por $\chi_{m_i}(s_i)$. Considera-se que há um desvio de conceito real em m quando a média de erros do modelo aumenta significativamente em comparação com períodos anteriores. Para detectar esses desvios, utilizou-se o método Drift Detection Method (DDM), que monitora a média e a variância dos erros preditivos [Gama et al., 2004]. O DDM identifica desvio de conceito real ao detectar um aumento significativo nos erros, sinalizando uma mudança na distribuição dos atrasos previstos pelos modelos Naive Bayes.

A comparação dos métodos focou na eficácia de detecção e tratamento dos desvios de conceito real e virtual. O desempenho preditivo dos modelos foi avaliado utilizando métricas como F1 score, precisão e revocação. A análise detalhou o desempenho dos modelos em diferentes períodos, destacando as melhores combinações de modelos e métodos de detecção. O pacote R Heimdall² foi utilizado para monitorar e gerenciar a detecção de desvios, aplicando os métodos como KSWIN e DDM para identificar e adaptar-se a mudanças nos dados em tempo real, mantendo a eficiência preditiva dos modelos.

5. Resultados

A metodologia descrita na Seção 4 foi aplicada aos dados de voos do aeroporto de Congonhas (SBSP) durante os períodos pré-pandemia, intra-pandemia e pós-pandemia, abrangendo 132.994 registros de voos atrasados de 2018 a 2023. A Tabela 1 resume essa análise exploratória. Em 2018, 32.325 voos atrasados foram registrados, representando aproximadamente 15% do total de voos. Esses voos, operados por quatro companhias aéreas, acumularam 621.851 minutos de atraso, com uma média de 19,24 minutos por voo. Em 2019, 31.198 voos atrasados foram registrados, cerca de 14% do total, totalizando 630.814 minutos de atraso, com uma média de 20,22 minutos por voo. Em 2020, durante a pandemia de COVID-19, foram registrados 6.535 voos atrasados, aproximadamente 10% do

²<https://cran.r-project.org/web/packages/heimdall/>

total, somando 112.667 minutos de atraso e uma média de 17,24 minutos por voo. Em 2021, 8.759 voos atrasados representaram 12% do total, totalizando 160.142 minutos de atraso, com média de 18,28 minutos por voo. Em 2022, 26.478 voos atrasados (13% do total) acumularam 521.768 minutos de atraso, com uma média de 19,71 minutos por voo. Finalmente, em 2023, 34.699 voos atrasados representaram 16% do total, com 750.413 minutos de atraso, uma média de 21,63 minutos por voo. Observa-se que o aumento na média dos atrasos pode estar relacionado ao aumento do número de voos, com valores médios menores durante o período intra-pandemia.

Tabela 1. Análise dos Voos atrasados no Aeroporto de Congonhas (SBSP)

Ano	Total de Voos	Companhias Aéreas	Média de Atraso (minutos)
2018	32.325	4	19,24
2019	31.198	7	20,22
2020	6.535	6	17,24
2021	8.759	6	18,28
2022	26.478	5	19,71
2023	34.699	5	21,63

Além das variações nos números de voos e nas médias de atraso ao longo dos períodos, o impacto das condições meteorológicas também foi considerado na análise. Observou-se que, em períodos com condições meteorológicas adversas, como baixa visibilidade ou tempestades, houve um aumento significativo no número de atrasos. Este efeito foi mais pronunciado durante o período intra-pandemia, onde as condições meteorológicas adversas tiveram um impacto amplificado devido ao menor número de voos operacionais. A inclusão dos dados meteorológicos nas análises permitiu identificar correlações entre as condições climáticas e os atrasos, especialmente em cenários de desvio de conceito real.

Tabela 2. Avaliação dos Modelos e Métodos de Detecção de Desvio de Conceito

Período	Modelo	Método	F1 Score	Precisão	Revocação	% Desv. Conc.
Pré	NB	KSWIN	0,414	0,490	0,425	22,3%
	NB	Passivo	0,402	0,464	0,412	99,0%
	NB	DDM	0,384	0,389	0,432	14,6%
	NB	Inativo	0,283	0,217	0,442	0,0%
Intra	NB	KSWIN	0,239	0,252	0,197	10,0%
	NB	Passivo	0,243	0,340	0,178	99,0%
	NB	DDM	0,197	0,138	0,182	29,0%
	NB	Inativo	0,258	0,267	0,187	0,0%
Pós	NB	KSWIN	0,449	0,483	0,453	20,8%
	NB	Passivo	0,445	0,508	0,451	99,0%
	NB	DDM	0,420	0,428	0,494	13,5%
	NB	Inativo	0,391	0,343	0,500	0,0%

A Tabela 2 apresenta os resultados da avaliação dos modelos preditivos utilizando diferentes métodos de detecção de desvio de conceito, incluindo KSWIN, Passivo, DDM,

e Inativo. O método KSWIN é utilizado para detectar desvios de conceito virtual, onde a distribuição das variáveis de entrada muda, mas a relação com a variável de saída permanece a mesma. Já os métodos Passivo e DDM são empregados para detectar desvios de conceito real, onde a relação entre as variáveis de entrada e saída sofre mudanças significativas. O método Passivo assume que o desvio de conceito ocorre continuamente, enquanto o DDM é mais reativo, acionando o re-treinamento do modelo apenas quando um aumento significativo nos erros é detectado. Por outro lado, o método Inativo serve como controle, sem aplicação de detecção de desvio, permitindo comparar os resultados com e sem ajustes de conceito.

A coluna "% Desv. Conc." representa a porcentagem de instâncias em que o método detectou um desvio de conceito durante o período avaliado. Essa métrica é fundamental para entender a sensibilidade do método em identificar mudanças nos padrões dos dados. No entanto, deve-se notar que essa porcentagem não garante que todos os desvios detectados correspondam a mudanças reais no conceito subjacente; são estimativas que dependem do método aplicado. É crucial combinar essa métrica com outras, como F1 Score e Precisão, para uma avaliação abrangente da eficácia dos métodos de detecção.

Os resultados preditivos mostram que, no período pré-pandemia, a combinação de Naive Bayes com KSWIN foi mais eficaz, com um F1 score de 0,414 e precisão de 0,490, sendo particularmente adequada para detecção de desvios de conceito virtuais. Durante o período intra-pandemia, a eficácia do Naive Bayes diminuiu devido à volatilidade operacional. Técnicas básicas (Inativo e Passivo) tiveram melhor desempenho, evidenciando a dificuldade de prever atrasos em contextos incertos. No período pós-pandemia, Naive Bayes com KSWIN voltou a apresentar melhores resultados, sugerindo estabilização das condições operacionais. Métodos de detecção de desvio virtual, como KSWIN, foram superiores aos de desvio real, como DDM, em diversos cenários.

Este estudo contribui ao demonstrar o impacto da pandemia na previsão de atrasos de voos, destacando que as técnicas de desvio de conceito e retreino ativo tiveram relevância limitada durante a pandemia. Além disso, técnicas baseadas em desvio virtual mostraram-se mais eficazes que as baseadas em desvio real em vários contextos.

6. Conclusão

Este trabalho investigou a detecção de desvios de conceitos reais e virtuais nos atrasos de voos no aeroporto de São Paulo (SBSP) durante os períodos pré, intra e pós-pandemia, utilizando o modelo Naive Bayes. A combinação de Naive Bayes com o método Kolmogorov-Smirnov Windowing (KSWIN) foi eficaz na detecção de desvios de conceito virtual em períodos estáveis, enquanto técnicas passivas e inativas tiveram melhor desempenho durante a pandemia, quando as mudanças de contexto eram mais evidentes. Os valores de F1, precisão e revocação refletem a complexidade do cenário, onde variáveis externas influenciaram os atrasos, tornando a previsão mais desafiadora.

Os resultados ressaltam o impacto da pandemia na previsão de atrasos de voos, destacando a importância das técnicas de desvio virtual em cenários estáveis. A eficácia dessas abordagens em contextos de alta variabilidade sugere a necessidade de futuras investigações sobre diferentes algoritmos de aprendizado de máquina ou a inclusão de variáveis adicionais para melhorar a acurácia dos modelos em cenários semelhantes.

Referências

- ANAC (2023). The Brazilian National Civil Aviation Agency. Technical report, <http://www.anac.gov.br/>.
- Gama, J., Medas, P., Castillo, G., and Rodrigues, P. P. (2004). Learning with drift detection. In *Brazilian Symposium on Artificial Intelligence (SBIA)*, pages 286–295. Springer.
- Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4).
- Giusti, L., Carvalho, L., Gomes, A. T., Coutinho, R., Soares, J., and Ogasawara, E. (2022). Analyzing flight delay prediction under concept drift. *Evolving Systems*.
- Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., and Zhao, D. (2020). Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology*, 69(1):140 – 150.
- Iwashita, A. S. and Papa, J. P. (2019a). An Overview on Concept Drift Learning. *IEEE Access*, 7:1532 – 1547.
- Iwashita, R. and Papa, J. P. (2019b). An Overview on Concept Drift Adaptation. *Journal of Artificial Intelligence Research*.
- Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., and Zhang, G. (2019). Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12):2346 – 2363.
- Moreira, L., Dantas, C., Oliveira, L., Soares, J., and Ogasawara, E. (2018). On Evaluating Data Preprocessing Methods for Machine Learning Models for Flight Delays. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2018-July.
- NOAA (2023). Climate at a Glance Global Time Series. Technical report, <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/global/time-series>.
- Raab, C., Heusinger, M., and Schleif, F.-M. (2020). Reactive soft prototype computing for concept drift streams. *Neurocomputing*.
- Rani, S. S., Ali, A. I. A., Marie, A., El-Bannany, M., and Khedr, A. M. (2023). Air Traffic Data Analysis Using Recurrent Neural Network (RNN) Classifier During COVID-19. In *Proceedings - 17th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2023*, pages 402 – 408.
- Sakthithasan, S. and Pears, R. (2016). Capturing recurring concepts using discrete Fourier transform. *Concurrency and Computation: Practice and Experience*, 28(15):4013 – 4035.
- Teixeira, C., Giusti, L., Soares, J., dos Santos, J., Amorim, G., and Ogasawara, E. (2021). Integrated Dataset of Brazilian Flights. In *Anais do Brazilian e-Science Workshop (BreSci)*, pages 89–96. SBC.
- Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., and Petitjean, F. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery*, 30(4):964 – 994.