

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA E DE ARTES**

**RELATÓRIO TÉCNICO VI DESAFIO EM CIÊNCIA DE
DADOS: MODELAGEM DE DADOS PARA PREVENIR
ACIDENTES AÉREOS**

Marcus Vinicius Santos da Silva
Julia Souza de Mendonça
Rafaella Nascimento Agapito
Giulia Pacheco Botelho Sales
Alysson Victor Almeida Souza

GOIÂNIA
2023



1. Introdução

O mundo contemporâneo está imerso em uma abundância de dados, coletados diariamente em uma variedade de contextos. Alguns são relacionados a situações simples, como um sorveteiro que precisa prever em quais meses do ano fará mais calor e, conseqüentemente, venderá mais sorvetes. Outros dados têm um impacto mais direto na segurança e bem-estar da população, como entender os horários em que uma cidade é mais tranquila para se transitar.

Quando analisamos com maior atenção o universo da aviação, nos damos conta de que se trata de um espaço dinâmico e complexo, onde as decisões são frequentemente tomadas com base em análises cuidadosas de dados, para que a segurança se mantenha como prioridade. Nas últimas décadas, o avanço tecnológico transformou radicalmente a indústria aérea, permitindo a coleta e análise de grandes volumes de dados que antes eram inimagináveis. Essa revolução tecnológica trouxe consigo a oportunidade de compreendermos melhor os padrões de ocorrência de acidentes aéreos e, mais importante ainda, desenvolver estratégias para preveni-los.

Neste contexto, surge a necessidade iminente de explorar o potencial da modelagem de dados para prevenir acidentes aéreos. Sua análise pode revolucionar a segurança da aviação ao permitir que identifiquemos padrões de risco, desenvolvemos modelos preditivos e implementemos medidas preventivas eficazes.

Diante dessa perspectiva, nosso objetivo neste trabalho é explorar as vastas quantidades de dados fornecidos pelo Centro de Investigação e Prevenção de Acidentes Aeronáuticos do Brasil (CENIPA) e desenvolver estratégias proativas para mitigar os riscos de acidentes, buscando prever situações que colocam em risco a segurança aérea, além de orientar as empresas e contratantes e passageiros a entender o que levam os riscos de acidentes.

2. Extração e limpeza de dados

Nesta fase, concentramos nossos esforços em identificar e corrigir problemas nos dados fornecidos pela CENIPA, preparando-os para análises subsequentes. Ao longo deste processo, enfrentamos uma série de desafios, desde a identificação de linhas duplicadas até o filtramento de matrículas, para garantir que nenhuma estivesse repetida. Nesta seção, descreveremos as técnicas e abordagens utilizadas para limpar e preparar os dados para análise exploratória.

1. Após lermos os arquivos e analisarmos suas situações, nos deparamos com ocorrências sobre um mesmo caso em linhas diferentes. Agrupamos essas ocorrências, removendo as duplicadas. Além disso, eliminamos linhas que continham dados faltosos, como coordenadas nulas e linhas vazias. Dessa forma, conseguimos diminuir significativamente o número de linhas.
2. Realizamos um tratamento nas coordenadas, convertendo-as para formato numérico, para que se tornassem legíveis na criação dos mapas.
3. Excluimos colunas não significativas para o projeto.
4. Separamos acidentes, incidentes, e incidentes graves a fim de criarmos mapas de calor por cada ocorrência. Os conceitos de cada um serão explorados posteriormente neste relatório.
5. Convertemos o arquivo Excel para CSV.
6. Geramos gráficos e histogramas com os dados coletados.

A partir dessas etapas, pudemos seguir com um processo mais limpo e uniforme, e obtivemos análises mais precisas e executáveis.

3. Análise exploratória

3.1 Dos fatores por região

Ao analisarmos os mapas de calor que mostram a distribuição de acidentes aéreos por estado, é possível observar padrões interessantes e importantes para compreendermos a segurança da aviação em diferentes regiões. O objetivo principal desses mapas é identificar as áreas onde ocorrem mais frequentemente acidentes, incidentes e incidentes graves, fornecendo insights valiosos para direcionar investigações sobre as possíveis causas desses eventos.

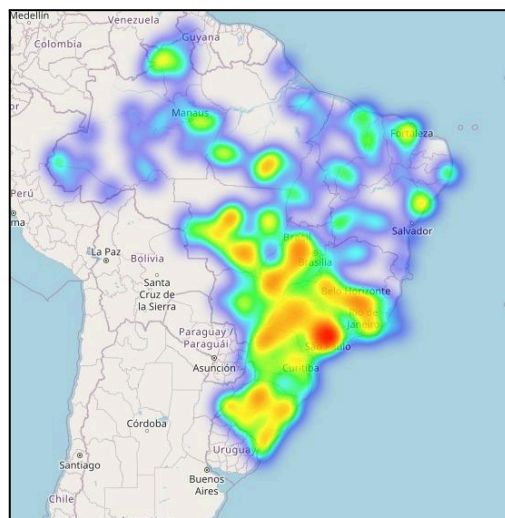
Essa análise geográfica nos permite visualizar de forma clara e intuitiva as regiões com maior concentração de ocorrências, o que pode ser fundamental para a implementação de medidas preventivas e o aprimoramento da segurança aérea.

Antes disso, é crucial entender os diferentes contextos dos acidentes, incidentes e incidentes graves na aviação:

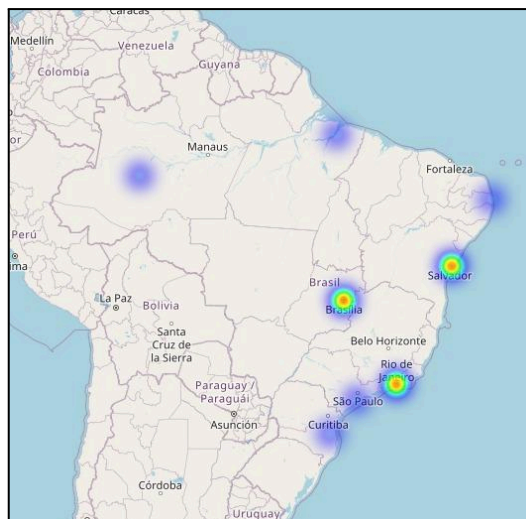
1. **Acidentes:** Ocorrem danos significativos a uma aeronave, tripulação, passageiros ou pessoas no solo, frequentemente resultando em lesões graves, fatalidades ou danos substanciais à aeronave.
2. **Incidentes:** São eventos que não resultam em danos significativos, mas ainda representam uma ameaça à segurança e operação da aeronave. Exemplos incluem falhas de motor, depressurização da cabine e colisões próximas entre aeronaves.
3. **Incidentes graves:** Possuem características que indicam um alto potencial de resultar em um acidente aéreo, mesmo que não ocorram danos significativos. São situações críticas que exigem medidas de emergência para evitar um acidente, como falhas críticas em motores ou sistemas de controle de voo, condições meteorológicas extremas ou erros de comunicação.

Com esses conceitos em mente, vamos agora explorar esses mapas de calor para entender melhor a distribuição geográfica dos eventos aeronáuticos e suas possíveis implicações.

Mapa 1: Acidentes



Mapa 2: Incidentes



Mapa 3: Incidentes graves

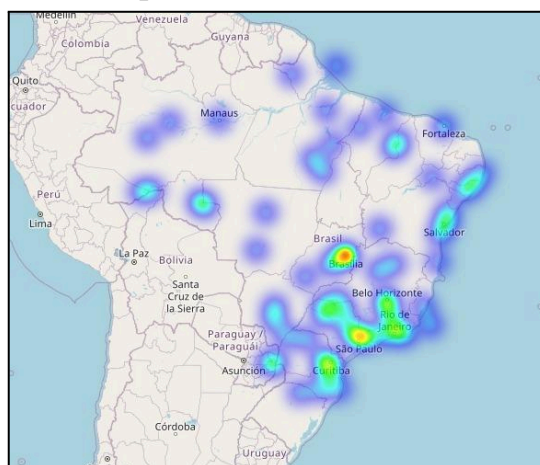
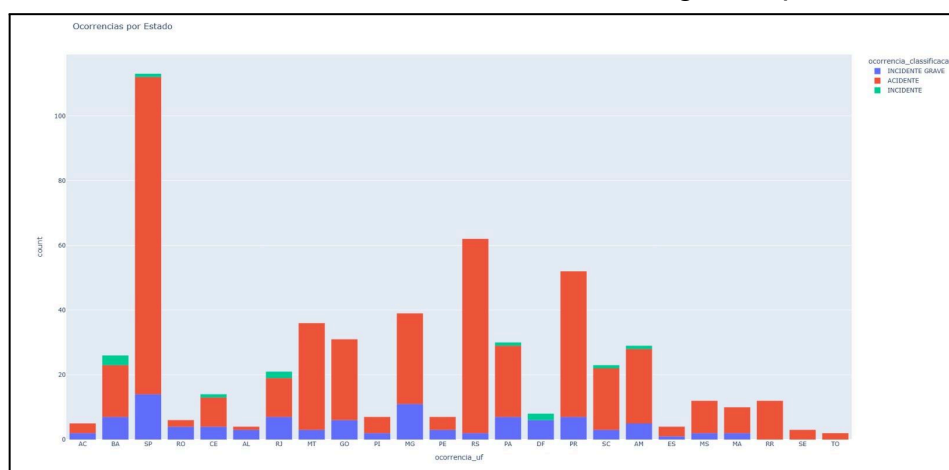
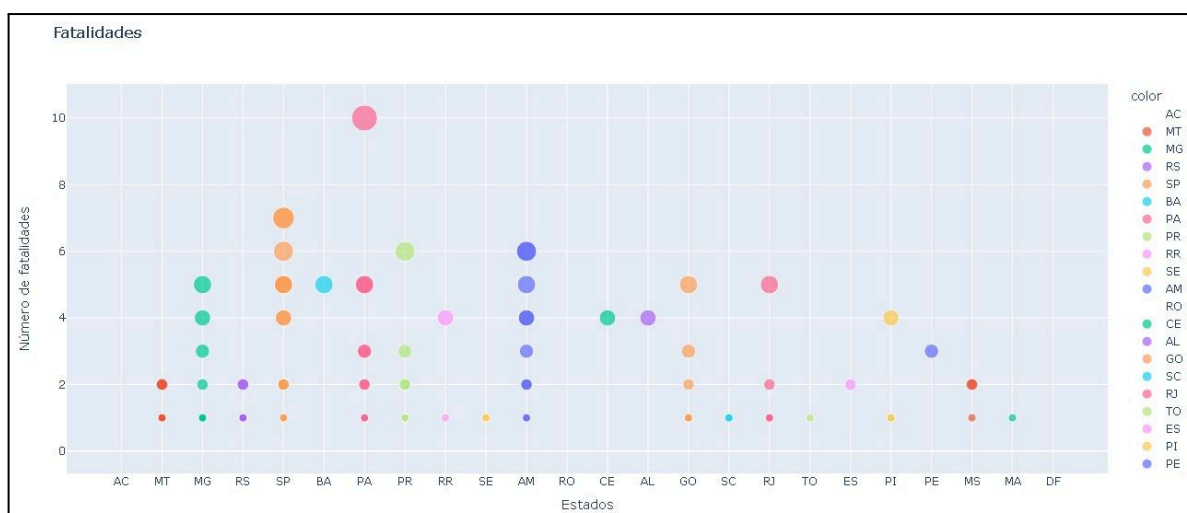


Gráfico 1: Acidentes, incidentes e incidentes graves por estado



Ao analisarmos o **mapa 1**, é evidente que as regiões sul e sudeste do país registram uma significativa concentração de ocorrências em acidentes. É ilustrado na imagem abaixo a representação dessas fatalidades decorrentes de acidentes.



Esses dados sugerem a necessidade de uma atenção especial em termos de medidas preventivas de segurança aérea. Entretanto, é importante ressaltar que tais ocorrências são, em certa medida, previsíveis, considerando se tratar da zona mais populosa do país.

Ao analisarmos o **mapa 2**, percebemos que os incidentes se tratam de casos em um número menor de ocorrências, com destaque para as capitais Brasília, Rio de Janeiro e Salvador. Essa observação ressalta a importância em se investigar as possíveis causas de seus incidentes e, ainda mais, implementar medidas corretivas para reduzir tais casos.

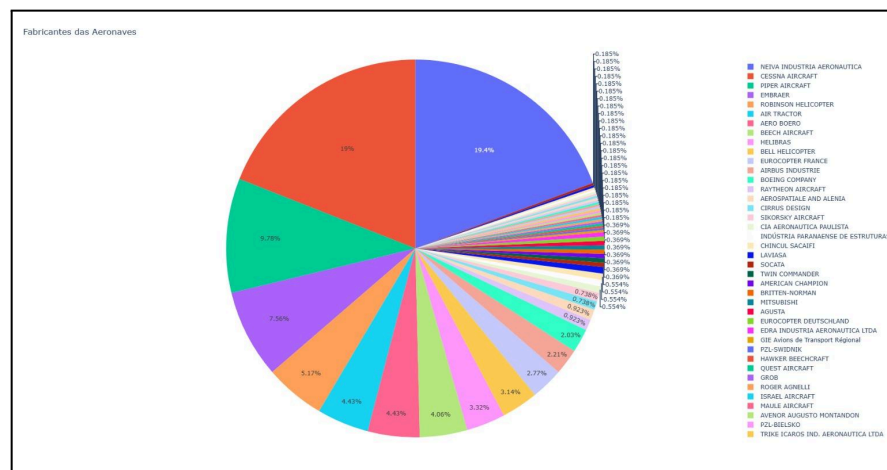
No que diz respeito ao **mapa 3**, os incidentes graves distribuem-se pelo país com alguns casos isolados que merecem um olhar mais atento. Destaca-se, novamente, a capital Brasília, mais afetada por tais incidências, juntamente de São

Por fim, ao correlacionar essas informações com as porcentagens apresentadas no **gráfico 1**, podemos direcionar melhor nossos esforços para as regiões mais afetadas, visando garantir a segurança e a integridade das operações aéreas em todo o país.

3.2 Dos fatores fabricantes

Agora, direcionaremos nosso foco para uma análise dos fatores relacionados aos fabricantes de aeronaves, buscando compreender melhor as ocorrências envolvendo diferentes motores e as fabricantes responsáveis pelas aeronaves envolvidas em acidentes, incidentes e incidentes graves. Esta etapa nos permitirá examinar os dados com o intuito de identificar padrões e tendências relevantes, fornecendo uma visão mais aprofundada sobre os fatores relacionados aos fabricantes de aeronaves. Ao analisar essas ocorrências, poderemos obter compreensões valiosas sobre a segurança e desempenho desses equipamentos.

Gráfico 2 - Fabricantes por ocorrência



[illegible]

Stacked bar chart showing the count of aircraft incidents categorized by aircraft type and engine type. The y-axis represents the count (0 to 350). The x-axis shows aircraft types: TURBOELICE, TURBOJATO, ajetone_motor_tipo, JATO, and SEM TRACAO. The legend lists 40 incident types, such as 'COM MANOBRESA / JANELA / PORTA', 'OPERAÇÃO A BAIXA ALTITUDE', 'EXCLUSÃO DE PISTA', etc.

Por outro lado, os motores turboeixo, utilizados em voos executivos e comerciais de médio e pequeno porte, apresentam uma distribuição diferente de ocorrências. Conforme evidenciado no gráfico, as principais falhas desses motores ocorrem durante o voo, em vez de durante as fases de decolagem e pouso. Isso pode ser atribuído ao fato de que as aeronaves equipadas com motores turboeixo

estão sujeitas a regulamentações mais rigorosas e são operadas em altitudes mais elevadas, onde os riscos associados às falhas de motor são diferentes.

3.3 Dos fatores técnicos e humanos

Apesar de termos adquirido um entendimento sobre a utilidade dessas aeronaves, ainda precisamos investigar outros fatores contribuintes. E quanto aos erros humanos? Agora, iremos analisar os aspectos técnicos e os fatores humanos que podem contribuir para acidentes aéreos.

Gráfico 5 - fatores técnico

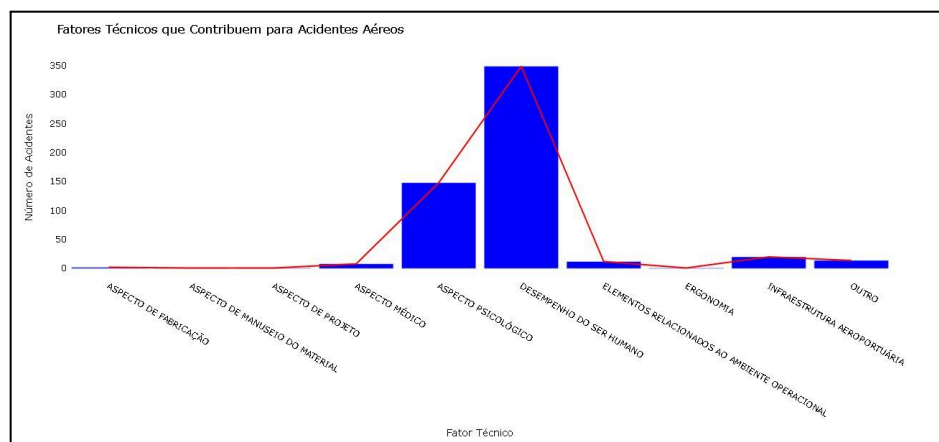
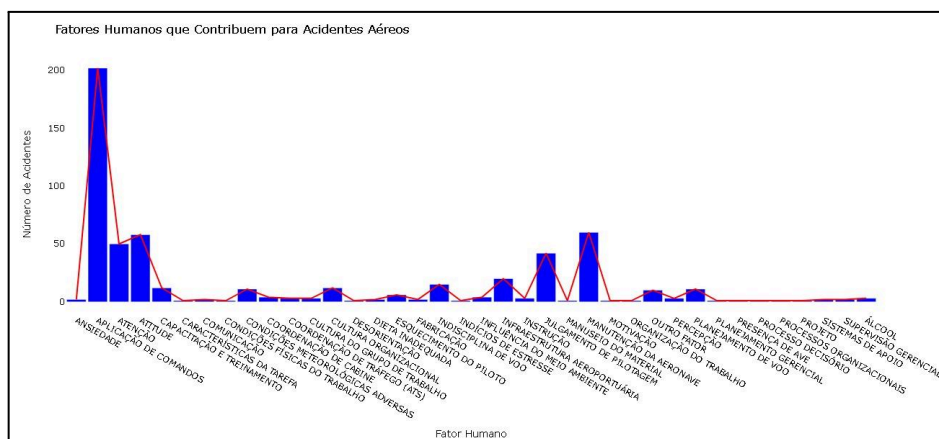


Gráfico 6 - fatores humanos



Os fatores técnicos que contribuem para os acidentes aéreos são diversos, mas é interessante notar que, no **gráfico 5**, dois deles se destacam de maneira significativa: o desempenho do ser humano e os aspectos psicológicos. Esses resultados nos sugerem que aspectos relacionados à capacidade dos pilotos podem desempenhar um papel crucial na ocorrência de acidentes. O índice elevado em

desempenho humano pode indicar que falhas na tomada de decisão, experiência de pilotagem ou treinamento inadequado estão entre alguns dos fatores que contribuem para os acidentes. Por outro lado, a ênfase nos aspectos psicológicos sugere que fatores como, por exemplo, fadiga, problemas de saúde mental ou questões pessoais dos pilotos também influenciam negativamente a segurança das operações aéreas.

No **gráfico 6**, temos a comprovação disso. O gráfico aborda diretamente os fatores humanos, com predominância no problema de aplicação de comandos. Esse achado significativo pode indicar erros na execução de procedimentos ou manipulação dos controles da aeronave. Além disso, a alta incidência de problemas relacionados ao manuseio da aeronave, atitude e atenção sugere que aspectos como coordenação motora, julgamento de situações de vigilância durante o voo são áreas críticas que merecem atenção e aprimoramento.

A correlação entre esses fatores pode ser complexa e multifacetada. Por exemplo, um piloto que enfrenta problemas de estresse ou fadiga (aspectos psicológicos) pode ser mais propenso a cometer erros na aplicação de comandos ou no manuseio da aeronave. Da mesma forma, deficiências no treinamento ou na experiência podem afetar tanto o desempenho do ser humano quanto a habilidade de aplicar corretamente os procedimentos operacionais.

Portanto, entender essas inter-relações e abordar de forma ampla os fatores técnicos e humanos é essencial para melhorar a segurança da aviação e reduzir o número de acidentes aéreos. Isso pode envolver desde a implementação de programas de treinamento mais abrangentes e eficazes até a promoção de uma cultura organizacional que valorize a saúde mental e o bem-estar dos pilotos.

4. Modelagem Preditiva

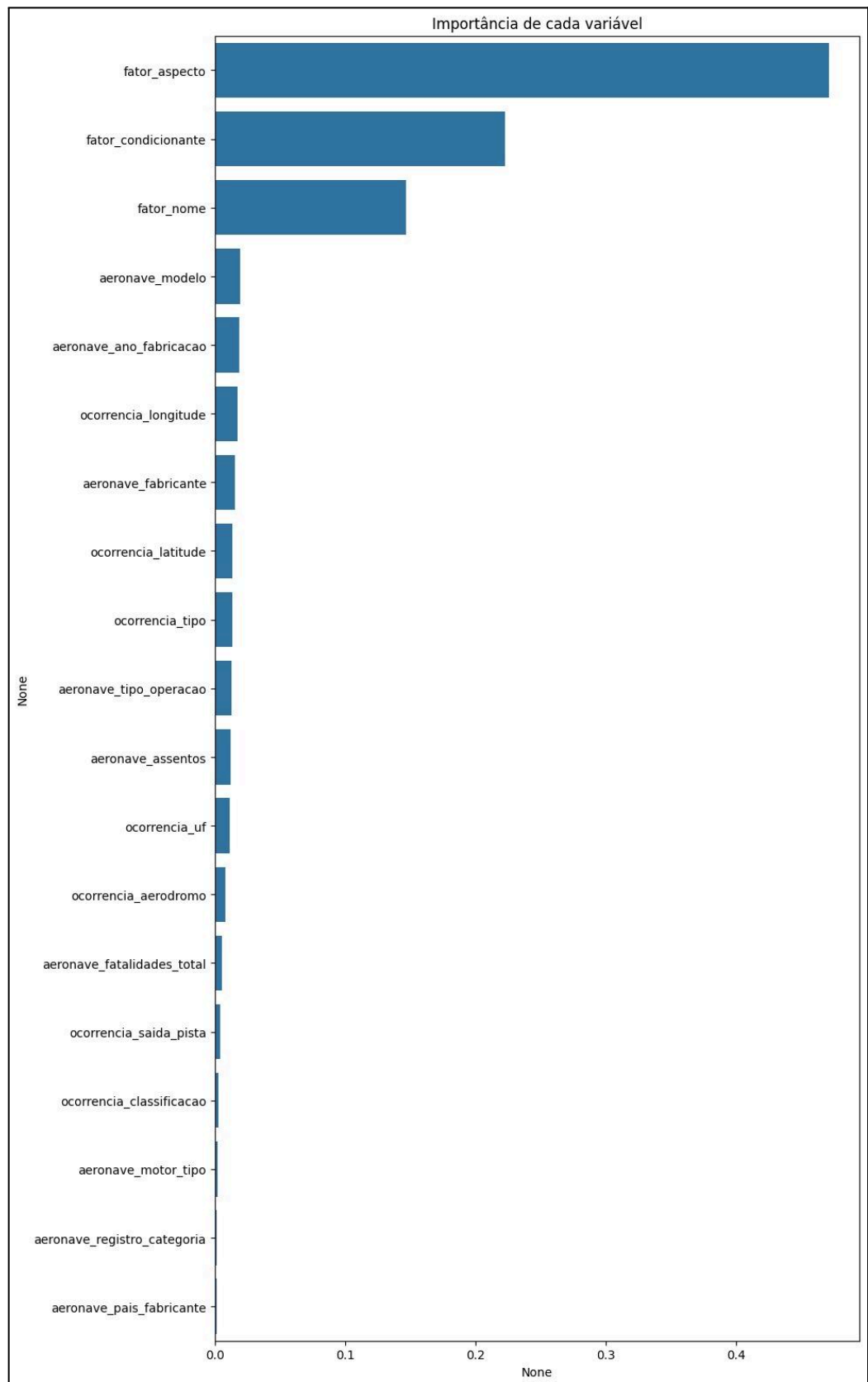
A Floresta Aleatória (Random Forest) é um algoritmo de aprendizado de máquina amplamente utilizado, desenvolvido por Leo Breiman e Adele Cutler, que combina a saída de várias árvores de decisão para produzir um resultado único. Sua popularidade e flexibilidade o tornaram uma escolha comum para lidar com problemas de classificação e regressão. A ideia da Random Forest é trabalhar com múltiplos resultados, que são gerados por diversas árvores de decisão, obtendo no final um resultado mais confiável e robusto.

A Random Forest, ao invés de utilizar o conjunto completo de dados para treinamento, gera múltiplas subamostras do conjunto original. Cada uma dessas subamostras são criadas aleatoriamente, com reposição, o que significa que alguns pontos de dados podem ser escolhidos mais de uma vez.

Para cada subamostra, é gerada uma árvore onde somente um subconjunto aleatório dessas features é selecionado. Essa aleatoriedade intencional contribui para que as árvores sejam distintas entre si.

No contexto de um modelo de aprendizado de máquina como o Random Forest para prever acidentes aéreos, a importância das características é medida com base em sua utilidade na realização de previsões precisas. Como um todo, pode-se destacar as duas variáveis que foram consideradas mais importantes pelo modelo, como é mostrado no gráfico a seguir:

Gráfico 7 - 'fator_aspecto' e 'fator_condicionante'



'Fator_aspecto': Essa característica foi considerada muito importante pelo modelo, o que significa que ela é particularmente útil para distinguir entre voos que resultam em acidentes. Isso pode indicar que o tipo de operação de voo (falha humana, falha técnica, condições climáticas, etc) está relacionado ao risco de acidente.

'Fato_condicionante': Similarmente, essa característica foi considerada importante, sugerindo que as condições que levaram ao acidente (por exemplo, a inexperiência do piloto) são relevantes para a previsão.

É crucial salientar que a importância de uma característica não implica necessariamente causalidade. Ou seja, mesmo que o *'fator_aspecto'* seja altamente relevante para o modelo, isso não significa que alterar esse fator automaticamente resultará em uma mudança na probabilidade de um acidente.

Um Score alto gerado pela random forest indica que o modelo está fazendo previsões precisas na maioria dos casos. No entanto, ela pode ser enganosa. Por exemplo, se uma classe é muito mais comum do que as outras nos dados de treinamento, o modelo pode aprender a sempre prever a classe mais comum. Isso pode resultar em um score alto, mesmo que o modelo não esteja realmente aprendendo os padrões importantes nos dados.

Conclusão

A aplicação do modelo Random Forest obteve um Score de aproximadamente 97.29%. Neste caso, o Score é normalmente calculado como a proporção de previsões corretas divididas pelo número de classes. O modelo demonstrou uma alta taxa de precisão na classificação das ocorrências. Isso nos oferece uma visão valiosa sobre os fatores que contribuem para a segurança da aviação e podem orientar futuras iniciativas de prevenção de acidentes.

Durante a correlação dos dados, exploramos diversas variáveis, desde características técnicas das aeronaves até fatores humanos e ambientais. Essa análise detalhada nos permitiu identificar padrões e tendências significativas, contribuindo para uma compreensão mais abrangente dos acidentes.

Apesar dos resultados promissores, há espaço para melhorias no modelo preditivo. Uma abordagem para aprimorar a precisão do modelo pode envolver a inclusão de dados adicionais, como dados de tempo, por exemplo. Além de testar outros modelos preditivos e um banco de dados que abranja uma maior gama de dados.

Em suma, com uma abordagem cuidadosa e contínuo refinamento, podemos continuar avançando na promoção da segurança da aviação e proteção à vida humana.

Códigos Utilizados

Os códigos utilizados neste presente relatório podem ser vistos em:
<https://github.com/Marcux777/Ci-ncia-de-Dados/tree/main/VI%20Desafio%20de%20Ciência%20de%20dados>

Fontes de Referência

<https://www.ibm.com/topics/random-forest>