

ANÁLISE DE DADOS DE VOO UTILIZANDO A TÉCNICA CLUSTERAD-FLIGHT PARA DETECÇÃO DE ANOMALIAS

Felipe Campos Vidigal

Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos - SP, 12228-900
pradelfjpb@fab.mil.br

Antonio Celio Pereira de Mesquita

Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos - SP, 12228-900
celio.mesquita@gp.ita.br

Fernando Teixeira Mendes Abrahão

Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos - SP, 12228-900
abrahao@ita.br

RESUMO

O artigo analisa os dados de voo da aeronave C-98A *Grand Caravan*, usando a técnica *ClusterAD-Flight* para identificar anomalias. Desde a incorporação dessas aeronaves pelo 7º Esquadrão de Transporte Aéreo, uma quantidade significativa de dados foi coletada, mas não foi sistematicamente aproveitada para melhorar operações ou desempenho. Este estudo visa otimizar operações e segurança mediante uma abordagem analítica, capacitando operadores para usar os dados coletados de maneira proativa. A abordagem inclui transformar dados em vetores de alta dimensão, reduzi-los e usar a análise de clusters para detectar padrões que não constituem falhas, mas são importantes para a detecção de eventuais ineficiências operacionais ou logísticas. Essa técnica tem potencial para análises que destaquem comportamentos que requeiram atenção a fim de prevenir incidentes e acidentes. A identificação desses padrões pode contribuir para operações mais seguras.

Palavras-chave: Análise de dados. Detecção de anomalias. Otimização de operações. Segurança operacional.

ABSTRACT

The article analyzes flight data from the C-98A *Grand Caravan* aircraft, using the *ClusterAD-Flight* technique to identify anomalies. Since the incorporation of these aircraft into the 7th Airlift Squadron, a significant amount of data has been collected but has not been systematically leveraged to improve operations or performance. This study aims to optimize operations

and safety through an analytical approach, enabling operators to use the data collected proactively. The approach includes transforming data into high-dimensional vectors, reducing them and using cluster analysis to detect patterns that do not constitute failures, but are important for detecting possible operational or logistical inefficiencies. This technique has the potential for analyzes that highlight behaviors that require attention to prevent incidents and accidents. Identifying these patterns can contribute to safer operations.

Keywords: Data analysis. Anomaly detection. Operational optimization. Operational safety.

1 INTRODUÇÃO

Desde a incorporação das aeronaves C-98A *Grand Caravan* ao 7º Esquadrão de Transporte Aéreo (7º ETA), dados extensivos têm sido registrados pelo sistema *Garmin 1000*, um sistema de navegação e controle de voo. Contudo, esses dados não têm sido sistematicamente analisados para melhorar operações ou desempenho. Este estudo visa preencher essa lacuna, aplicando uma técnica de análise de dados para incentivar uma melhor utilização das informações disponíveis.

1.1 Contextualização

O advento de aeronaves com mais capacidade de geração de dados de operação e manutenção pela FAB exige a adequação do treinamento dos seus operadores para atender suas demandas de maneira eficiente. A incorporação dessas novas tecnologias requer não apenas uma familiarização técnica, mas também uma abordagem analítica robusta para maximizar o aproveitamento das informações provenientes dos dados gerados. Assim, depreende-se a necessidade de desenvolvimento de procedimentos de análise dos dados de voos gerados que contribuam para o gerenciamento dos respectivos ciclos de vida (Aerospace and of Europe, 2022).

1.2 Motivação

Em virtude disso, surgiu a necessidade de criar um procedimento de análise dos dados de voos gerados por essas aeronaves. Este procedimento planeja capacitar os pilotos e equipes de manutenção a identificar padrões discrepantes e comportamentos anômalos, com o intuito de melhorar o gerenciamento da operação e da suportabilidade das aeronaves (Russell, 2003). Além disso, a análise dos dados de voo pode desempenhar um papel crucial na segurança aérea para mitigar incidentes e acidentes.

É importante ressaltar, que esses padrões eventualmente discrepantes não necessariamente constituem falhas; por isso, são difíceis de serem percebidos. Assim, com técnicas avançadas de aprendizado de máquina, supervisionadas ou não, essas anomalias podem ser identificadas

(Coelho e Silva and Murça, 2023).

1.3 Organização do Trabalho

Este trabalho foi organizado da seguinte forma: A Seção 2 expõe a fundamentação teórica, abordando os principais sistemas e programas de análise de dados de voo, a descrição detalhada da aeronave C-98A e seu sistema de aviônica, bem como os métodos e técnicas empregados para garantir a segurança e eficiência das operações aéreas.

A Seção 3 apresenta a metodologia utilizada neste estudo para identificar problemas desconhecidos a partir dos dados de voo da FAB. A abordagem proposta emprega técnicas de mineração de dados para processar as informações de maneira eficiente e interpretar os resultados e implicações operacionais. No primeiro passo, o algoritmo é utilizado para detectar voos anormais de um grande conjunto de dados do FDR do C-98A, que possuem um alto nível de similaridade devido à padronização das operações da FAB. Voos com padrões incomuns são rotulados como potencialmente interessantes para inspeção adicional, pois podem sugerir situações de risco.

No segundo passo, realiza-se uma análise detalhada dos voos detectados para constatar suas deficiências. Embora o algoritmo detecte padrões anormais de dados, a análise requer uma compreensão aprofundada de procedimentos operacionais e de mecânica para revisar os detalhes e tirar conclusões baseadas em conhecimentos operacionais e logísticos.

A Seção 4 detalha os resultados e discussões, incluindo a aplicação do algoritmo *ClusterAD-Flight* em dois conjuntos de dados do FDR do C-98A, totalizando 442 voos das aeronaves da frota do 7º Esquadrão de Transporte Aéreo. São apresentados os métodos de transformação dos dados em vetores de alta dimensão, a redução de dimensões, e a análise de *clusters*. Os resultados mostram a identificação de voos anormais, destacando os comportamentos que desviam dos padrões normais.

Finalmente, a Seção 5 apresenta as conclusões, ressaltando a importância de métodos de análise de dados para a segurança proativa dos voos da FAB.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda o principal sistema de verificação e análise de dados de voo, bem como o programa adotado na aviação moderna para garantir a segurança e a eficiência das operações aéreas. Além disso, descreve a aeronave que será objeto de estudo, incluindo seu avançado sistema. Cada seção fornece uma exploração detalhada de como esses elementos fundamentais influenciam no estudo. Visando orientar e fundamentar as discussões e conclusões, demonstrando como a integração de tecnologias avançadas e protocolos rigorosos pode significativamente melhorar a gestão da operação aérea.

2.1 *Flight Data Recorder (FDR)*

O *Flight Data Recorder (FDR)* registra parâmetros da aeronave durante um voo inteiro. Parâmetros típicos de voo incluem altitude, velocidade indicada, potência, temperatura do motor, superfícies de controle, etc. (Int, 2023). Um exemplo dos dados de potência, altitude e velocidade do *FDR* do C-98A *Grand Caravan* é mostrado na Figura 1.

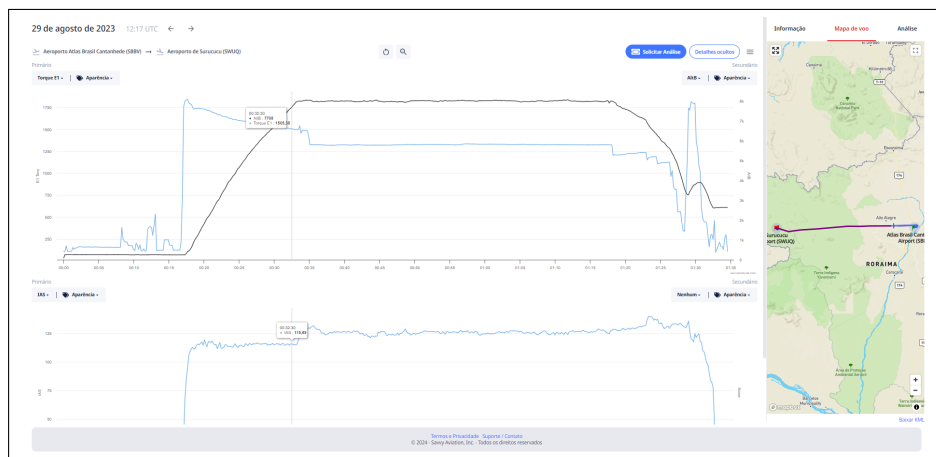


Figura 1: Dados do *FDR* do C-98A *Grand Caravan*. Fonte: SavvyAviation (2024).

Com o avanço da digitalização dos dados do *FDR*, plataformas especializadas permitem que sua análise seja realizada de maneira rápida e eficiente. Atualmente, a Força Aérea Brasileira emprega esses dados principalmente em investigações de acidentes aéreos. No entanto, existe a possibilidade de extrair e analisar esses dados de maneira periódica durante as operações cotidianas, não apenas após eventos críticos. Essa abordagem proativa possibilita a identificação e mitigação de riscos operacionais antes que resultem em incidentes, aumentando significativamente a eficiência da operação aérea.

2.2 *Flight Operations Quality Assurance (FOQA)*

O programa *FOQA* é amplamente utilizado na aviação para permitir a análises sistemáticas de dados e auxiliar as companhias aéreas a identificar desvios dos padrões operacionais, garantindo assim uma operação mais segura e eficiente. Este programa proativo envolve a coleta sistemática de dados durante os voos, utilizando o *Flight Data Recorder (FDR)* para registrar parâmetros de voo e ações do piloto (Federal Aviation Administration, 2004).

Após a coleta, os dados são analisados para identificar desvios dos procedimentos operacionais padrão e detectam eventos que podem indicar operações potencialmente inseguras ou ineficientes. Essa análise permite aos operadores reconhecerem tendências e padrões que necessitam de atenção.

Os resultados das análises são utilizados para melhorar a operação e prevenir incidentes. O *FOQA* também desempenha um papel crucial no desenvolvimento e revisão dos programas

de treinamento de pilotos, focando em áreas onde os dados indicam a necessidade de melhorias.

Em resumo, o *FOQA* é uma ferramenta vital para operadores aéreos comprometidos com práticas de gestão de segurança baseadas em evidências, proporcionando uma base sólida para a tomada de decisões informadas e a melhoria contínua das operações de voo.

2.3 Aeronave C-98A *Grand Caravan*

A aeronave C98A *Grand Caravan*, utilizada pela Força Aérea Brasileira (FAB), é uma adaptação militar do *Cessna 208 Caravan* (Company, 2015). Este modelo é equipado com um motor turboélice, conhecido por sua confiabilidade e eficiência de combustível. Além de possuir um avançado sistema de navegação e comunicação, incluindo o sistema *Garmin 1000*, que facilita a navegação e aumenta a segurança das operações.

Devido à sua robustez e habilidade para operar em pistas curtas e não pavimentadas, o C98A é amplamente utilizado em missões de transporte geral e operações em áreas remotas. Além disso, é frequentemente empregado em missões humanitárias, resgate, evacuação médica e apoio em situações de desastres naturais, aproveitando sua capacidade de acessar locais isolados.

Atualmente, na FAB, o C-98 é operado em diversas regiões do Brasil. Demonstrando assim a versatilidade e a confiabilidade dessa aeronave. O 7º ETA possui a maior frota dessas aeronaves devido às necessidades da operação na Região Amazônica.

2.4 *Garmin 1000*

O G1000 é um sistema de aviônica avançado desenvolvido pela *Garmin* (Ltd., 2015), que revolucionou as cabines de pilotagem de aeronaves de aviação geral e algumas aplicações militares. Este sistema substitui os instrumentos de voo analógicos tradicionais por uma configuração de integrada.

Desenhado para ser intuitivo, o sistema apresenta telas de alta resolução e controles acessíveis que diminuem a carga de trabalho do piloto, tornando a pilotagem mais segura e eficiente. Além disso, o *Garmin 1000* permite possui um sistema integrado de *Flight Data Recorder (FDR)*, que permite o monitoramento de diversos parâmetro do voo. Essas características fazem do *Garmin 1000* uma escolha popular, proporcionando uma plataforma confiável e segura.

2.4.1 Dados do *FDR* do *Garmin 1000*

O sistema do *Garmin 1000* registra diversos parâmetros a partir do momento que a bateria é ligada. Esses dados são essenciais para o monitoramento e análise detalhada dos voos, permitindo identificar padrões e possíveis anomalias. Os parâmetros registrados pelo *FDR* do *Garmin 1000* que foram utilizados neste estudo estão detalhados na Tabela 1, que inclui infor-

mações como data e hora do voo, altitude, velocidade, acelerações, atitudes, e diversos outros parâmetros críticos para a análise de desempenho, segurança do voo e fatores externos como: temperatura do ar e velocidade do vento.

Tabela 1: *FDR do Garmin 1000*

Parâmetro	Descrição
Lcl Date	Data do voo
Lcl Time	Horário UTC
AltB	Altitude da aeronave acima do nível do mar, medida pelo altímetro
OAT	Temperatura externa ao ar
IAS	Velocidade indicada do ar
GndSpd	Velocidade da aeronave sobre o solo
VSpd	Velocidade vertical da aeronave, indicada no velocímetro
Pitch	Atitude em relação ao eixo longitudinal
Roll	Atitude em relação ao eixo lateral
E1 OilT	Temperatura do óleo do motor
E1 OilP	Pressão do óleo do motor
E1 Torq	Torque gerado pelo motor
E1 NP	Velocidade da hélice do motor
E1 NG	Velocidade do gás no motor
E1 ITT	Temperatura total do ar de entrada no motor
WndSpd	Velocidade do vento

Fonte: Garmin (2015).

3 ABORDAGEM METODOLÓGICA

Este capítulo inicia com uma exploração detalhada da técnica de detecção de anomalias por meio da análise de *cluster*, a qual será aplicada neste estudo. Em seguida, discute-se a metodologia empregada para transformar os dados de voo em vetores de alta dimensão, facilitando a identificação de padrões normais de voo e *outliers* que sinalizam possíveis anomalias. Finalmente, será abordada a linguagem de programação utilizada neste estudo, destacando como essa ferramenta é fundamental para a análise de dados.

3.1 Detecção de Anomalias

A técnica de detecção de anomalias (Chandola et al., 2009) envolve a identificação de conjuntos de pontos de dados em que aqueles dentro do mesmo conjunto, exibem maiores semelhanças entre si do que com outros conjuntos. A identificação dessas anomalias envolve identificar objetos que não pertencem a nenhum cluster, ou estão associados a grupos muito pequenos, operando sob a suposição de que os dados padronizados estão agrupados nos grupos mais densos e de tamanho considerável.

Após a conclusão da transformação dos dados, a análise é conduzida em vetores situados em um espaço de alta dimensão. Conjuntos de vetores próximos são reconhecidos como agrupamentos, denotando padrões de dados predominantes no conjunto de dados; vetores que não

se alinham a nenhum agrupamento são classificados como discrepantes, indicando padrões de dados anormais (Chandola et al., 2009).

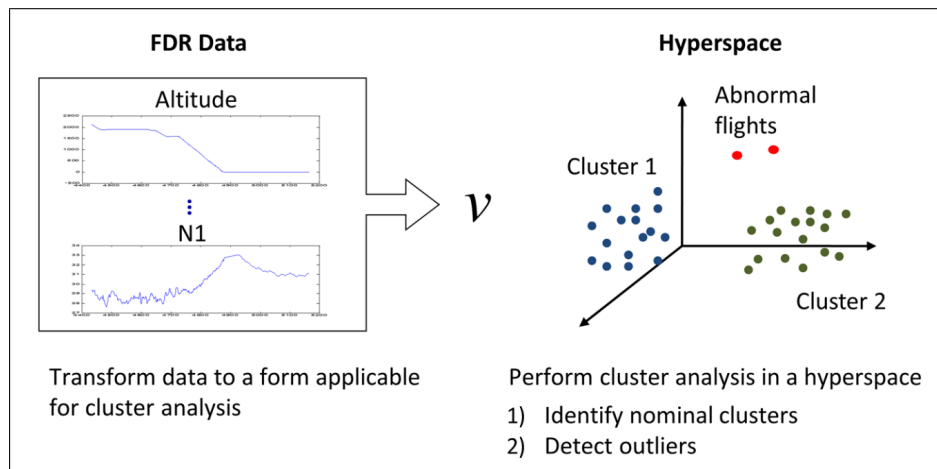


Figura 2: Detecção de Anomalias. Fonte: Li (2013).

3.2 ClusterAD-Flight

O *ClusterAD-Flight* (Li, 2013) consiste em três etapas principais:

1. Transformação dos Dados: transformando séries temporais em vetores de alta dimensão.
2. Redução de Dimensão: abordando problemas de alta dimensionalidade.
3. Análise de *Clusters*: identificando *clusters* e *outliers* no espaço de alta dimensão.

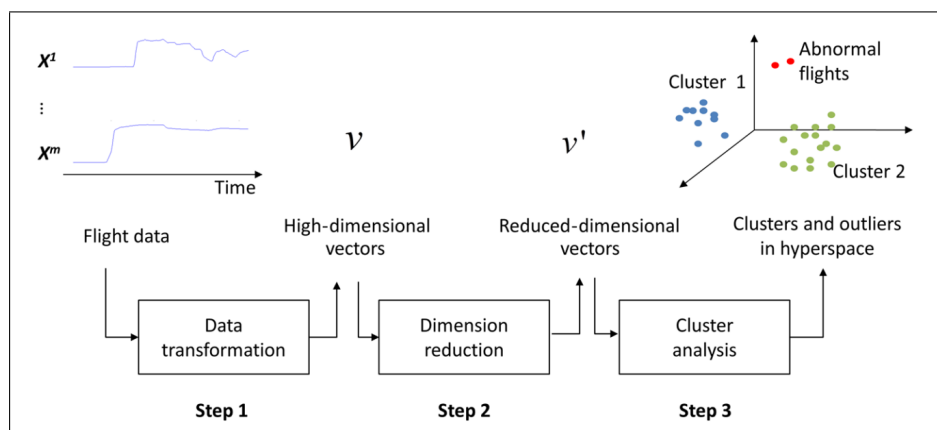


Figura 3: Etapas do *ClusterAD-Flight*. Fonte: Li (2013).

A primeira etapa transforma dados de séries temporais multivariadas em vetores de alta dimensão. A técnica de transformação ancora as séries temporais em um evento específico, como a potência de decolagem, preservando as informações temporais e torna os dados comparáveis entre diferentes voos. Então, na segunda etapa, técnicas são desenvolvidas para lidar com o problema de alta dimensionalidade. A dimensionalidade dos vetores é reduzida para viabilidade computacional, mantendo informações essenciais. Na última etapa, a análise de *clusters*

é realizada para detectar *outliers* e *clusters* de voos normais no espaço de características de dimensões reduzidas. A análise de *clusters* é uma técnica que identifica grupos de pontos de dados de forma que os pontos dentro do mesmo grupo sejam mais semelhantes entre si do que com os de outros grupos (Li, 2013).

3.2.1 Transformação dos Dados

Para alinhar dados brutos em vetores no mesmo espaço de alta dimensão, os dados provenientes de vários voos são sincronizados com base em um evento designado para garantir os mesmos padrões temporais. Posteriormente, cada parâmetro do voo é amostrado em um mesmo intervalo de tempo, distância ou outros intervalos de referência predeterminados subsequentes ao evento de referência. Os valores coletados são então estruturados para criar um vetor correspondente a cada voo individual:

$$\mathbf{v} = [x_1^1, x_2^1, \dots, x_n^1, \dots, x_j^i, \dots, x_n^m] \quad (1)$$

onde x_j^i é o valor do i -ésimo parâmetro de voo no tempo de amostra j ; m é o número de parâmetros de voo; n é o número de amostras para cada parâmetro de voo. A dimensionalidade total de cada vetor é $m * n$. Cada dimensão representa o valor de um parâmetro de voo em um determinado tempo. A similaridade entre os voos pode ser medida pela distância euclidiana entre os vetores (Li, 2013).

3.2.2 Redução de Dimensões

Devido ao aspecto temporal, os vetores formados na primeira etapa normalmente terão milhares de dimensões. É difícil identificar nuvens de dados em uma distribuição tão grande. Portanto, o *PCA* (*Principal Component Analysis*) foi usada para reduzir o número de dimensões antes de realizar a análise de *clusters*. O *PCA* é um procedimento comum para transformar dados em um sistema de coordenadas ortogonais baseado na variância dos dados. As coordenadas no novo sistema são referidas como componentes. A maior variância por qualquer projeção dos dados se encontra no primeiro componente, a segunda maior variância no segundo, e assim por diante. Como consequência, um número de últimos componentes poderia ser descartado para reduzir a dimensão do espaço hiperdimensional sem perder informações significativas. Neste algoritmo, são mantidos os primeiros componentes K que capturam 90% da variância nos dados.

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > 90\% \quad (2)$$

onde λ_i é a variância explicada pelo componente principal i . N é o número total de componentes principais, que é igual ao número original de dimensões. K é o número de componentes

principais mantidos. A magnitude da redução dimensional variará com o conjunto de dados, mas pode ser significativa. Nesta análise, as dimensões dos vetores foram reduzidas de 2917 para 9 na fase da decolagem e de 1674 para 11 na fase de pouso. A matriz de correlação é uma ferramenta essencial para identificar relações entre as variáveis do conjunto de dados. Ela mostra a força e a direção das relações lineares entre as variáveis, permitindo uma melhor compreensão das interdependências presentes. Nas Figuras 4 e 5, estão presentes as matrizes de correlação construída a partir dos dados de voo na fase de decolagem e pouso.

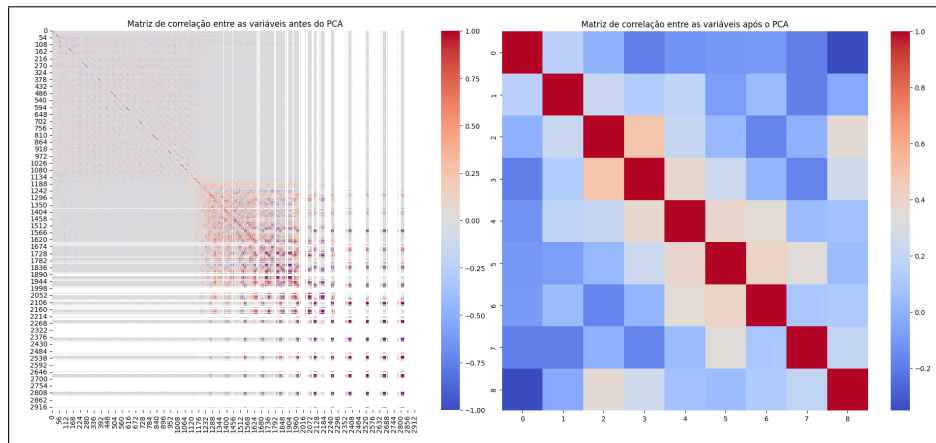


Figura 4: Matriz de Correlação - Decolagem. Fonte: O Autor.

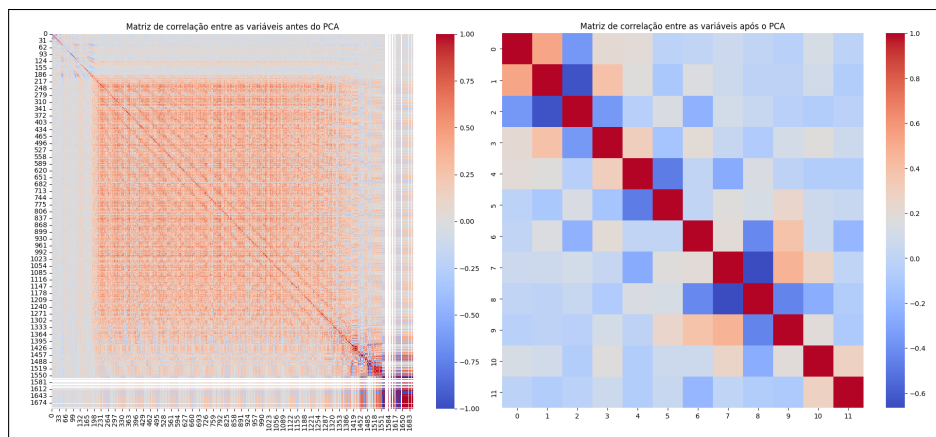


Figura 5: Matriz de Correlação - Pouso. Fonte: O Autor.

3.2.3 Análise de Clusters

A análise de agrupamento visa identificar *clusters* de dados no espaço de características. Em seguida, os *outliers* podem ser detectados com base nos *clusters* identificados. O algoritmo *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*), proposto por Ester et al. (1996), foi selecionado para agrupar os dados desta análise.

O *DBSCAN* é um algoritmo de agrupamento baseado em densidade. Ele progressivamente encontra *clusters* baseados em um critério de densidade. Um *cluster* se forma quando um número mínimo de pontos (*MinPts*) estiver em um raio ϵ de um círculo. O *cluster* cresce

encontrando os vizinhos, que também satisfazem o mesmo critério de densidade, até que nenhum outro ponto possa ser adicionado ao *cluster* existente. Neste ponto, ele começa a buscar um novo *cluster*. Os *outliers* são os pontos que não pertencem a nenhum *cluster* (Li, 2013).

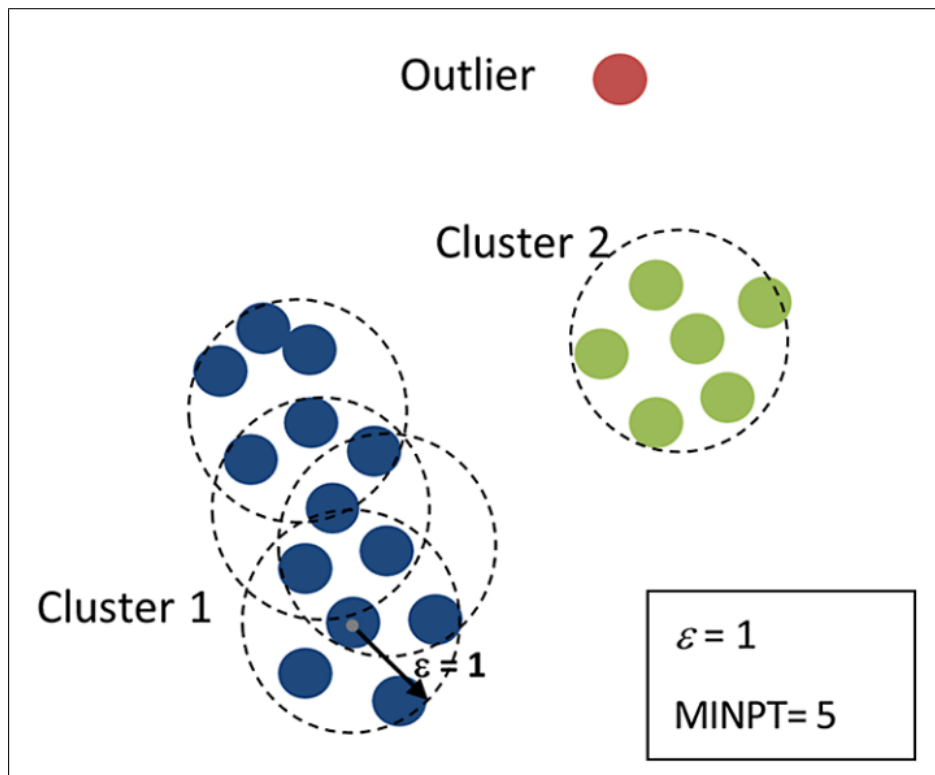


Figura 6: Algoritmo *DBSCAN*. Fonte: Li (2013).

3.3 Utilização do Python na Análise de Dados de Voo

O *Python* (Python Software Foundation, 2023) é amplamente reconhecido por sua eficiência e versatilidade no processamento de grandes conjuntos de dados, características que o tornam ideal para análise de dados na aviação. Sua sintaxe clara, com um suporte extensivo por meio de bibliotecas de ciência de dados, facilita a realização de análises complexas. Neste contexto, existe uma infinidade de ferramentas como *Pandas* (The Pandas Development Team, 2023) para manipulação e limpeza de dados, *Matplotlib* (Hunter, John D., 2023) e *Seaborn* (Waskom, Michael (2023) para visualização, e *Scikit-Learn* (Scikit-Learn Developers, 2023) para técnicas de *Machine Learning* (Mitchell, 1997), tais como redução de dimensão e clusterização. Essas ferramentas são essenciais para transformar dados brutos em formatos organizados e acessíveis, permitindo uma análise detalhada e a visualização de padrões e tendências.

A aplicação prática do *Python* na análise de dados de voo permite transformar dados brutos em informações acessíveis, melhorando significativamente a segurança e a eficiência das operações de voo. A capacidade de automação de rotinas analíticas proporcionada pelo *Python* facilita a aplicação do *ClusterAD-Flight*, permitindo uma análise detalhada e precisa das possíveis causas das anomalias com base em informações confiáveis extraídas dos dados.

4 RESULTADOS E DISCUSSÕES

4.1 Teste do *ClusterAD-Flight*

O *ClusterAD-Flight* foi testado usando dois conjuntos de dados do *FDR* do C-98A *Grand Caravan* que juntos totalizam 442 voos das 5 aeronaves da frota do 7º Esquadrão de Transporte Aéreo, são elas: FAB2740, FAB2737, FAB2736, FAB2732 e FAB2730. Para obter dados relativamente homogêneos, o conjunto foi filtrado por voos do aeroporto de Boa Vista (SBBV) para o aeródromo de Surucucu (SWUQ) e volta do aeródromo de Surucucu (SWUQ) para o aeroporto de Boa Vista (SBBV).

A detecção de *outliers* foi realizada separadamente para a fase de decolagem e para a fase de aproximação. Os parâmetros do *FDR* foram testados para identificar os principais *outliers* apontados pelo método. Todos os *outliers* identificados foram posteriormente analisados para determinar se de fato eram voos anormais e, em caso afirmativo, caracterizar os comportamentos.

“A detecção de anomalias é examinar pontos de dados específicos e detectar ocorrências raras que parecem suspeitas porque são diferentes do padrão de comportamento estabelecido. A detecção de anomalias não é nova, mas à medida que os dados aumentam, o rastreamento manual é impraticável” (Amazon Web Services - AWS).

4.2 Dados, Preparação e Resultados

Os conjuntos de dados do gravador usados para os testes do *ClusterAD-Flight* continham 57 parâmetros em cada um dos voos, incluindo parâmetros do motor, posição da aeronave, velocidades, acelerações, atitudes, velocidade do vento e temperatura do ambiente. Para facilitar a comparação entre os voos, foram selecionados apenas voos para uma mesma localidade que permitam análises mais precisas.

Para transformação de séries temporais em vetores, as observações de 1 segundo foram agrupadas em intervalos de 30 segundos, desde a marcação de Velocidade Indicada (IAS) superior a 40, que caracteriza o início do voo, até 10 minutos após para a fase de decolagem. Para a fase de aproximação, o mesmo número de observações foi agrupado até 10 minutos antes da Velocidade Indicada (IAS) ficar inferior a 40, que caracteriza o término do voo. Após realizar a *PCA*, o número de dimensões dos dados foi reduzido para 9 componentes principais na fase de decolagem e 11 na de pouso, que separadamente explicam 90% da variância total dos dados, facilitando a análise subsequente pelo *DBSCAN*.

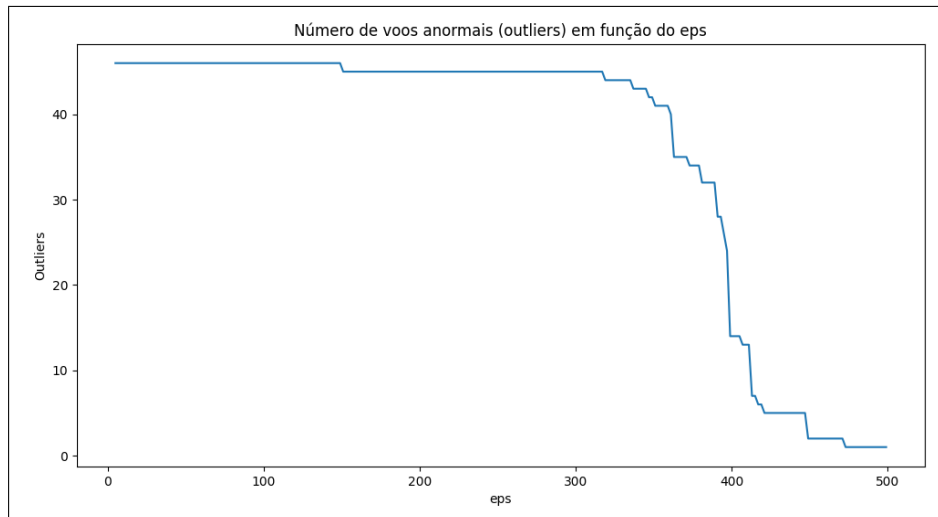


Figura 7: EPS para Decolagem. Fonte: O Autor.

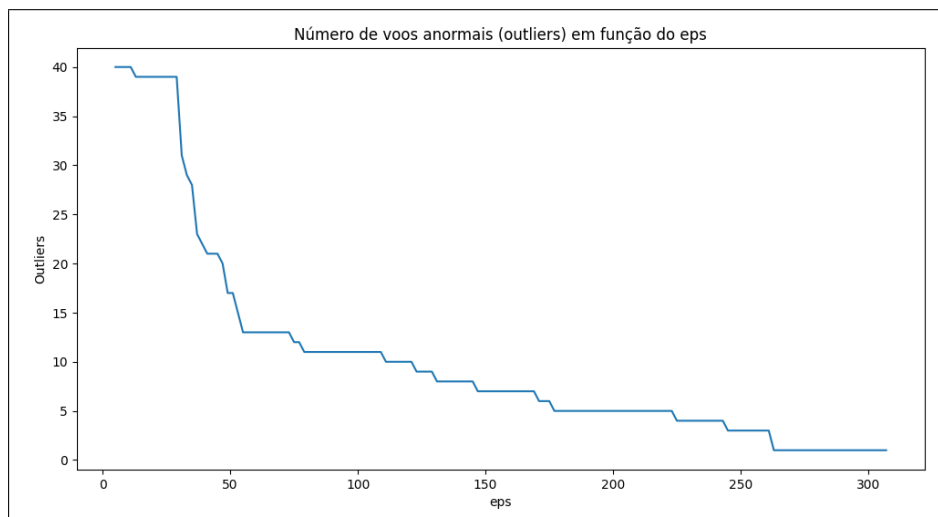


Figura 8: EPS para o Pouso. Fonte: O Autor.

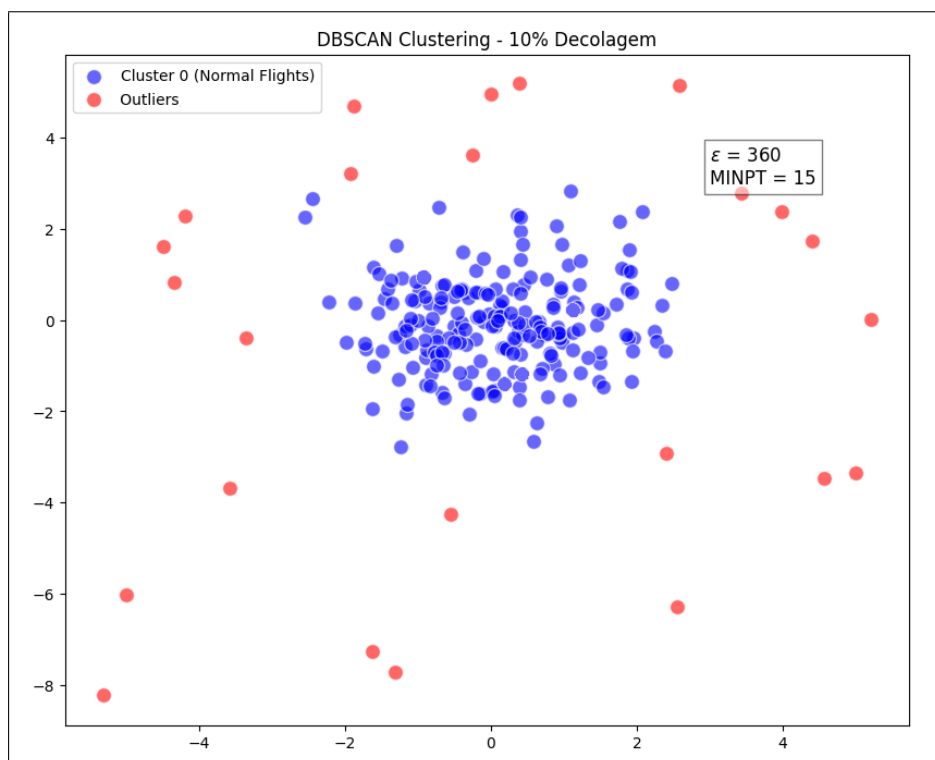
A sensibilidade aos critérios de seleção (ϵ e *MinPts*) é demonstrada para os dados de decolagem e pouso nas Figura 8 e 9, respectivamente. Foi observado que o *MinPts* mais adequado seria superior ao número de componentes principais para ambas as fases do voo, e que um aumento no ϵ resulta em menos voos identificados como *outliers*. Portanto, *MinPts* foi estabelecido em 13 para a decolagem e 15 para o pouso, enquanto o valor de ϵ foi escolhido para identificar as porcentagens de 1%, 5% e 10% de *outliers* em cada fase. Os resultados estão resumidos na Tabela 2.

Tabela 2: Outliers Detectados

% de Resultados	Configuração do DBSCAN	Outliers
1% Decolagem	$\epsilon = 450$, $MinPts = 15$	2
5% Decolagem	$\epsilon = 380$, $MinPts = 15$	12
10% Decolagem	$\epsilon = 360$, $MinPts = 15$	23
1% Pouso	$\epsilon = 300$, $MinPts = 12$	1
5% Pouso	$\epsilon = 100$, $MinPts = 12$	11
10% Pouso	$\epsilon = 40$, $MinPts = 12$	20

Fonte: O Autor.

Os resultados do algoritmo *DBSCAN* foram selecionados para identificar 10% dos *outliers* nos dados de decolagem utilizando $\epsilon = 380$ e $MinPts = 15$. Nesta configuração, os pontos azuis representam os voos classificados como normais (*Cluster 0*), enquanto os pontos vermelhos representam os *outliers*, que são voos que não pertencem a nenhum *cluster*. Esses resultados são ilustrados na Figura 9.

Figura 9: *DBSCAN Clustering* - 10% Decolagem. Fonte: O Autor.

Para os dados de pouso, o algoritmo *DBSCAN* foi aplicado com $\epsilon = 40$ e $MinPts = 12$. Na Figura 10, os pontos verdes representam os pousos classificados como normais (*Cluster 0*), enquanto os pontos amarelos indicam os *outliers*, que são pousos que não pertencem a nenhum *cluster*.

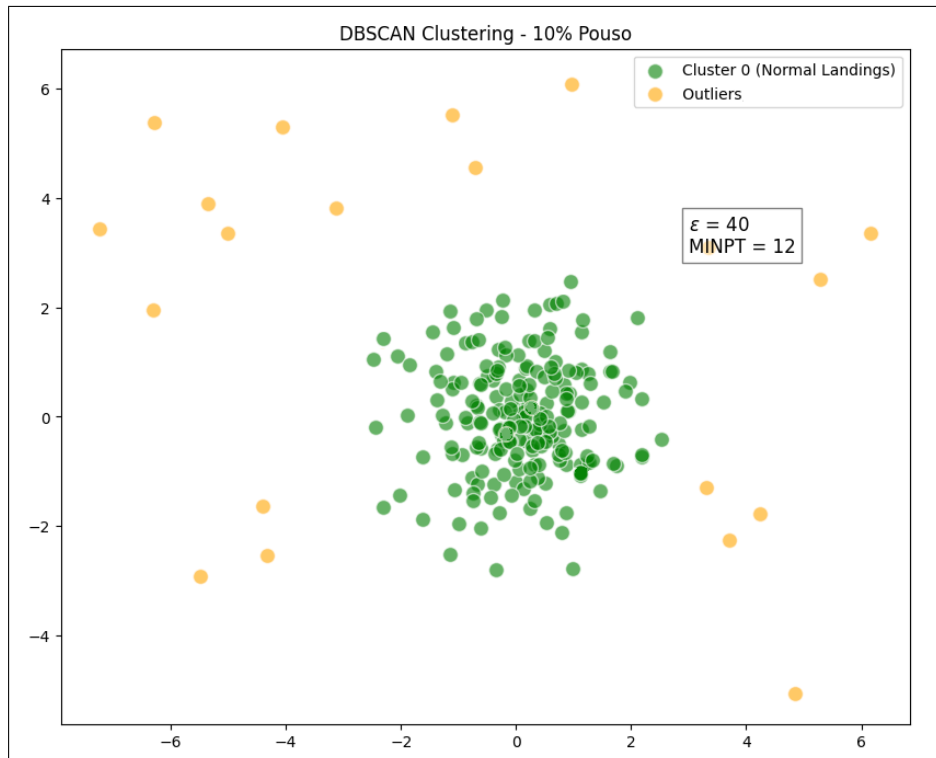


Figura 10: DBSCAN Clustering - 10% Pouso. Fonte: O Autor.

4.3 Análise dos Voos Detectados

Todos os voos classificados nos 10% de *outliers* foram analisados mais detalhadamente para determinar se exibiam comportamentos anormais. Os parâmetros de voo mais distintos para as fases de decolagem e pouso são apresentados em gráficos que utilizam o mesmo formato. Os voos anormais são representados por linhas laranjas e a média dos voos normais são representados por faixas azuis que possuem um intervalo de confiança de 90%.

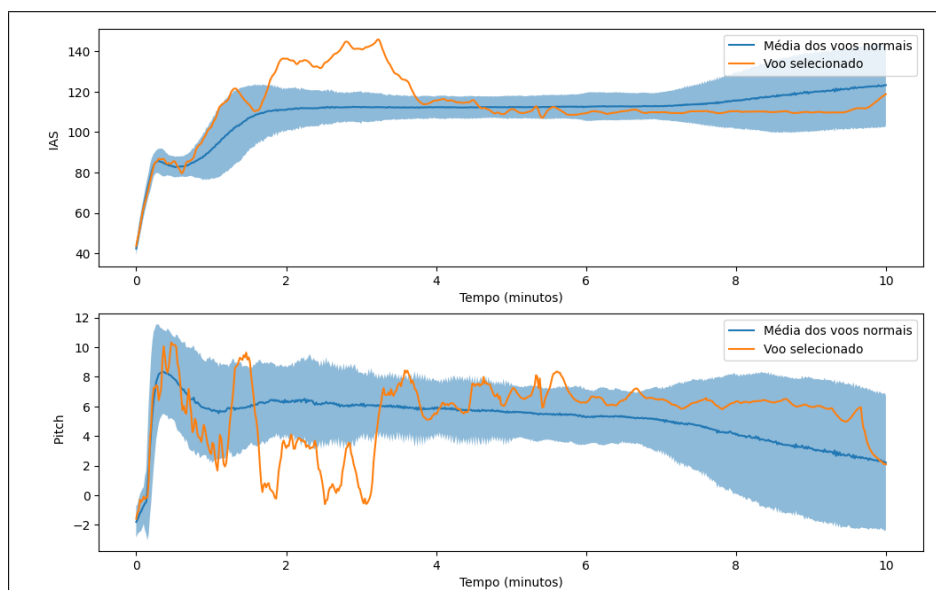


Figura 11: Voo 68 - Análise da Decolagem. Fonte: O Autor.

No voo 68, o piloto empregou Amplitudes de Arfagem (Pitch) significativamente maiores em comparação à média dos demais voos, resultando em variações atípicas dos parâmetros durante a fase de subida. As alterações notáveis na arfagem, possivelmente destinadas à correção da velocidade, estendeu-se até o final dos quatro primeiros minutos. Após a estabilização da arfagem e da velocidade, a subida começou a normalizar-se.

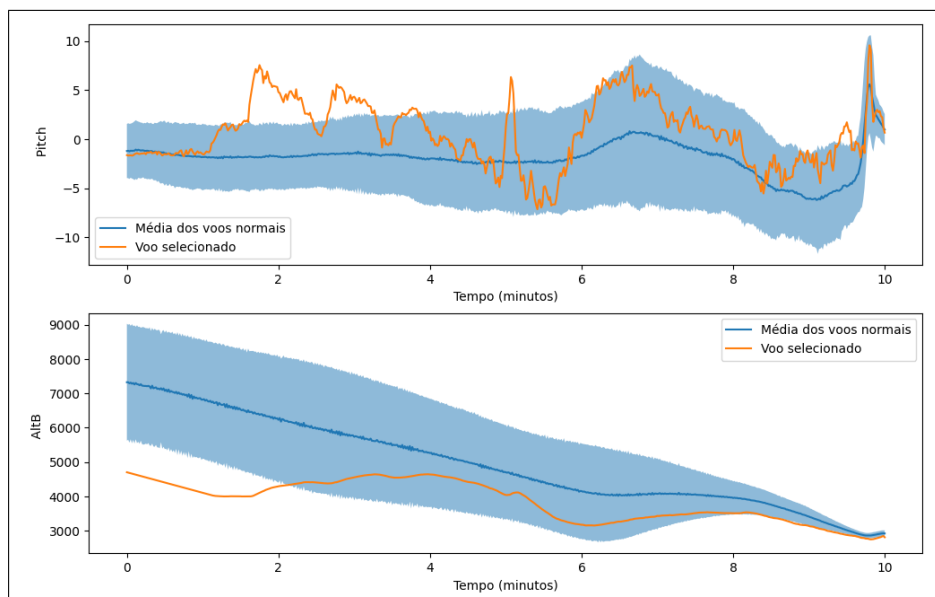


Figura 12: Voo 7 - Análise do Pouso. Fonte: O Autor.

No voo 7, o perfil da Altitude Barométrica (AltB) inicialmente se posiciona abaixo da trajetória de planeio típica de aproximação, indicando uma descida abaixo dos padrões comuns. Adicionalmente, é notável uma variação considerável na Atitude de Arfagem (Pitch) no intuito de acertar a trajetória de descida adequada.

A mesma análise foi realizada para todos os voos anormais detectados. Todos os voos identificados exibiram algum grau de anomalia. Os comportamentos anormais na decolagem são resumidos na Tabela 3. Os comportamentos mais frequentes foram decolagens com variação da potência, que frequentemente incluem outros fatores notáveis. Também foram observados: atitude de arfagem alta ou baixa durante a decolagem e padrões de temperatura do óleo do motor discrepantes na aeronave FAB2740.

Tabela 3: Decolagens Anormais Identificadas

Voo / Anormalidade
03 / Temperatura Externa Alta
07 / Ângulo de Arfagem Alto
11 / Ângulo de Arfagem Alto
12 / Ângulo de Arfagem Alto
44 / Velocidade Indicada Baixa
63 / Torque Baixo
64 / Ângulo de Arfagem Baixo
68 / Velocidade Indicada Alta
69 / Temperatura do Óleo Baixa
71 / Temperatura do Óleo Baixa
75 / Temperatura do Óleo Baixa
76 / Temperatura do Óleo Baixa
83 / Temperatura do Óleo Baixa
85 / Temperatura do Óleo Baixa
88 / Temperatura do Óleo Baixa
111 / Torque Baixo
135 / Velocidade Indicada Baixa
150 / Torque Baixo
169 / Pressão do Óleo Baixa
177 / Pressão do Óleo Baixa
182 / Pressão do Óleo Baixa
201 / Torque Baixo
216 / Ângulo de Arfagem Baixo

Fonte: O Autor.

Na fase de pouso, os comportamentos anormais foram mais diversificados (Tabela 4). Alguns voos apresentaram operações incomuns, como arfagem anormalmente alta, ângulo de rolagem defasado e variações no motor. Além disso, anomalias ambientais, como vento cruzado forte e alta temperatura atmosférica, foram encontradas em alguns dos voos anormais. Deve-se notar que nem todos os voos anormais identificados indicam preocupações com a segurança. Alguns voos foram identificados como anormais, mas eram casos benignos, como pouso em vento forte e temperatura alta.

Tabela 4: Pousos Anormais Identificados

Voo / Anormalidade
01 / Velocidade do Vento Alta
07 / Rampa Baixa
08 / Temperatura do Motor Alta
26 / Velocidade do Vento Baixa
49 / Temperatura do Motor Baixa
153 / Ângulo de Arfagem Alto
160 / Pressão do Óleo Baixa
166 / Velocidade do Vento Alta
167 / Pressão do Óleo Baixa
179 / Pressão do Óleo Baixa
194 / Temperatura do Externa Alta
195 / Temperatura do Motor Baixa
197 / Velocidade do Vento Alta
198 / Velocidade Indicada Baixa
203 / Velocidade do Vento Alta
205 / Ângulo de Rolagem Defasado
209 / Espera Prolongada
213 / Ângulo de Rolagem Defasado
214 / Temperatura do Óleo Alta
216 / Ângulo de Rolagem Defasado

Fonte: O Autor.

Em resumo, o *ClusterAD-Flight* foi aplicado a um conjunto de dados representativo do *Flight Data Recorder (FDR)* de 442 voos do C-98A. Voos anormais foram detectados, no entanto, anomalias relacionadas ao ambiente também foram encontradas, como vento cruzado forte e altas temperaturas. No voo 1, a Velocidade do Vento (WndSpd) atinge marcas superiores à normalidade na fase de pouso. Demonstrando que alguns voos foram identificados, mas eram casos relacionados a fatores externos.

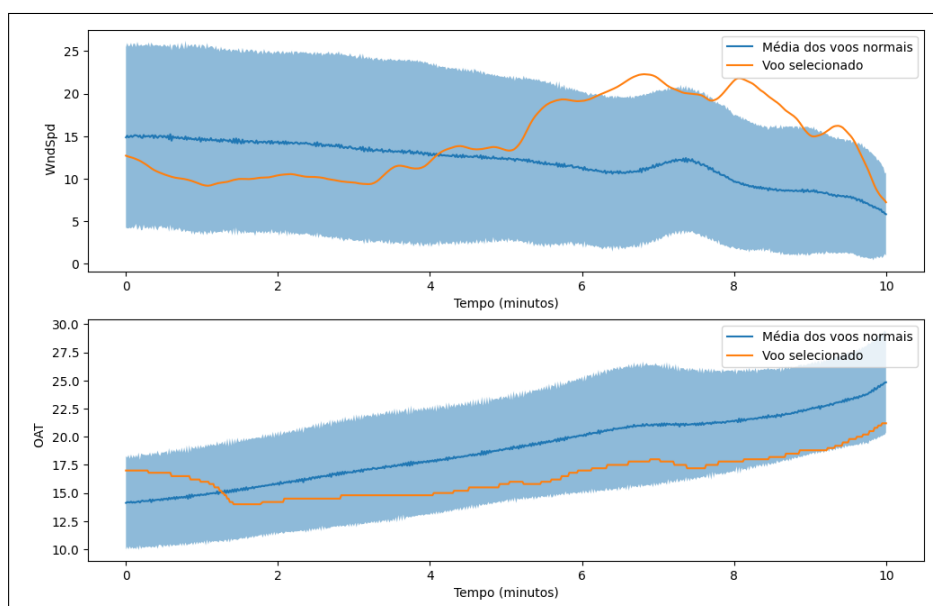


Figura 13: Voo 1 - Análise do Ambiente. Fonte: O Autor.

Além das anomalias relacionadas ao ambiente, observaram-se algumas relacionadas à mecânica do avião. Os voos do FAB 2740, por exemplo, sempre apresentavam parâmetros de Temperatura do Óleo (E1 OilT) abaixo na média demais voos normais e de Pressão do Óleo (E1 OilP) alta, mesmo que no padrão dos demais. Apontando um alerta para uma investigação minuciosa da causa dessa anomalia.

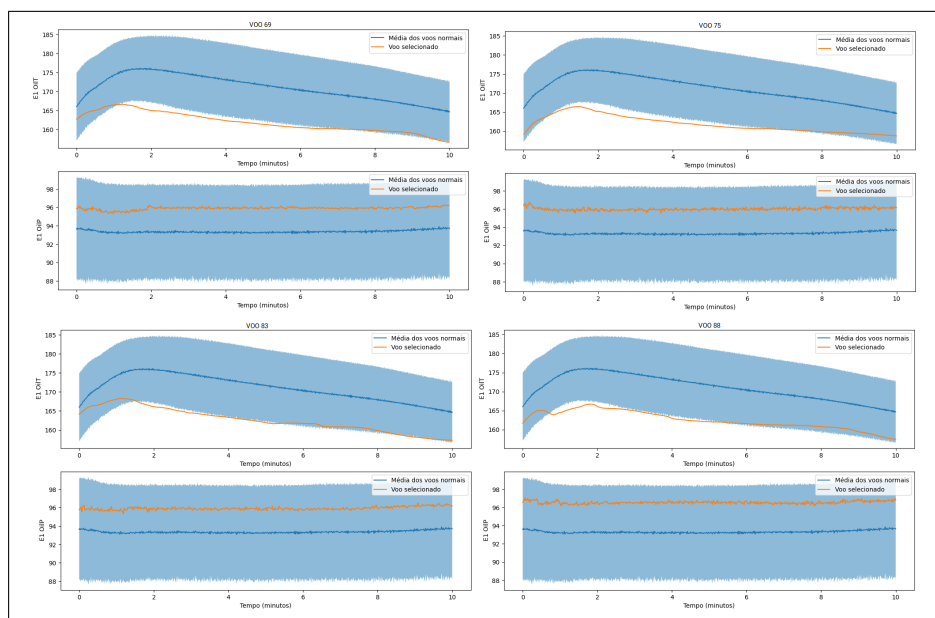


Figura 14: Voos 69/75/83/88 - Análise do Motor. Fonte: O Autor.

5 CONCLUSÃO

Atualmente, a Força Aérea Brasileira possui aeronaves com grande capacidade de armazenamento de dados de voos. No entanto, ainda não existem procedimentos bem definidos para analisar e extrair lições úteis dessas informações. Este artigo apresenta, então, um dos métodos para analisar esses dados e extrair informações para reforçar a segurança proativa dos voos da FAB.

O método de detecção de anomalias utilizado neste trabalho, o *ClusterAD-Flight*, pode detectar automaticamente voos anormais da FAB, por meio da detecção de *clusters*, agrupamentos de *outliers*. Pretende-se que esses voos sejam encaminhados para a Seção de Segurança de Voo, para a Seção de Operações e para a Seção de Manutenção dos Esquadrões para serem feitas análises dos riscos latentes. Posteriormente poderão ser estabelecidas recomendações que poderão ser transmitidas por meio de *briefings* de segurança, reuniões doutrinárias e recomendações logísticas.

O *ClusterAD-Flight* foi testado em dados de voo extraídos das aeronaves C-98A *Grand Caravan*, pertencentes ao 7º Esquadrão de Transporte Aéreo. Os testes foram realizados em um conjunto de voo estabelecidos. Os seus resultados mostraram que o algoritmo conseguiu ajudar a identificar comportamentos anormais dentre os voos, alguns dos quais podem indicar

consideráveis níveis de risco.

Um aprimoramento futuro deste trabalho é especificar as configurações do algoritmo para gerar resultados mais desejáveis na prática, considerando que o *ClusterAD-Flight* requer um limite de detecção em % de voos a ser definido com antecedência. Outra direção é estender o método proposto para outras fases do voo. Atualmente, o método desenvolvido está limitado à fase de decolagem. Dada a diversidade de padrões temporais em outras fases do voo, são necessárias novas técnicas para torná-las comparáveis.

Disponibilidade dos dados

Os dados e o código estão disponíveis em: www.aerologlab.ita.br/datafiles

Referências

- Aerospace and Defence Industries Association of Europe(2022). *SX000i - International Guide for the Integrated Logistic Support (ILS) Process*, 2022. Issue 3.1.
- Amazon Web Services - AWS. O que é detecção de anomalias? <https://aws.amazon.com/pt/what-is/anomaly-detection/>. Accessed: 2024-07-26.
- V. Chandola, A. Banerjee, and V. Kumar. Detecção de anomalias: uma pesquisa. *ACM Computing Surveys*, 41(3):58, 2009. doi: 10.1145/1541880.1541882. URL <http://doi.acm.org/10.1145/1541880.1541882>.
- Lucas Coelho e Silva and Mayara Condé Rocha Murça. A data analytics framework for anomaly detection in flight operations. *Journal of Air Transport Management*, 110:102409, 2023.
- Cessna Aircraft Company. *Cessna Model 208B G1000 Pilot's Operating Handbook (POH)*, 2015. URL <https://www.cessna.txt>. Available at Cessna Aircraft Company, Wichita, Kansas, USA. Revision 1.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231, Portland, OR, 1996. AAAI Press. URL <http://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>.
- Federal Aviation Administration. *Advisory Circular AC 120-82: Flight Operational Quality Assurance*, 2004. URL https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-82.pdf. Available at U.S. Department of Transportation, Federal Aviation Administration.

- Hunter, John D. Matplotlib: A 2d graphics environment. <https://matplotlib.org>, 2023. Accessed: data de acesso.
- Annex 6 - Operation of Aircraft, Part I - International Commercial Air Transport - Aeroplanes.* International Civil Aviation Organization, Montreal, Canada, 2023. Available from ICAO.
- Lishuai Li. *Anomaly Detection in Airline Routine Operations Using Flight Data Recorder Data.* PhD thesis, Massachusetts Institute of Technology, 2013. URL <https://dspace.mit.edu/handle/1721.1/79127>.
- Garmin Ltd. *Caravan G1000 Pilot's Guide (190-00749-06 Rev C)*, 2015. URL <https://www.garmin.com>. Available at Garmin International, Inc., Olathe, Kansas, U.S.A.
- Tom M. Mitchell. *Machine Learning.* McGraw-Hill, New York, NY, USA, 1997.
- Python Software Foundation. Python language reference, version 3.10. <https://www.python.org>, 2023. Accessed: data de acesso.
- Stephen Hays Russell. The growing world of logistics. *Air Force Journal of Logistics*, 24(4): 15–19, 2003.
- Scikit-Learn Developers. Scikit-learn: Machine learning in python. <https://scikit-learn.org>, 2023. Accessed: data de acesso.
- The Pandas Development Team. pandas: powerful python data analysis toolkit, version 1.4. <https://pandas.pydata.org>, 2023. Accessed: data de acesso.
- Waskom, Michael. seaborn: statistical data visualization. <https://seaborn.pydata.org>, 2023. Accessed: data de acesso.