APRENDIZADO DE MÁQUINA NÃO SUPERVISIONADO APLICADO À IDENTIFICAÇÃO DE SETORES DE ENTRADA EM ÁREA TERMINAL

Jean P. O. Lima*, Pollyanne E. Silva, Alef G. Quintanilha and Andre L. E. Melo
Airspace Control Institute - ICEA, Research Division

*** Corresponding author e-mail adress: jeanjpol@decea.mil.br**

PAPER ID: SIT2258

RESUMO

Aprimorar a eficiência do espaço aéreo é uma prioridade contínua na gestão do tráfego aéreo. Um indicador fundamental de desempenho nesse contexto é o tempo adicional gasto nas áreas terminais. Ao monitorarmos esse indicador de forma adequada, podemos tomar decisões que trarão benefícios tanto econômicos quanto ambientais. Um requisito para calcular esse indicador é identificar os setores de entrada na área terminal. Com base no cenário apresentado, o objetivo do estudo foi implementar um modelo para identificação automática de setores de entrada nas regiões a 40 NM (Milhas Náuticas) e 100 NM de distância de cada um dos quatro aeroporto mais movimentados do Brasil. Para isso, foram testados os métodos de aprendizado não supervisionado K-Means e DBSCAN, para a tarefa de clusterizar movimentos e, portanto, identificar regiões com maior densidade de entrada de voos em área terminal. A avaliação foi conduzida de forma qualitativa, levando em consideração a densidade de voos nas áreas terminais em comparação com os procedimentos de aproximação disponíveis nas cartas aeronáuticas dos aeroportos. Além disso, uma análise de ruído foi realizada, indicando a presença de ruídos entre 0,06% e 6,59% em cada setor, considerados como *outliers* ou movimentos atípicos. Esses resultados apontaram que os *clusters* foram bem identificados, mantendo um conjunto significativo de dados. As principais vantagens apresentadas neste artigo foram: a utilização de um modelo para identificação automática dos setores de entrada em área terminal, que, com um único conjunto de parâmetros, demonstrou robustez nos experimentos para os diferentes aeroportos, indicando a escalabilidade computacional da solução proposta; e a possibilidade de calcular o indicador de tempo adicional em área terminal de forma mais assertiva.

Palavras-Chave: Tráfego Aéreo, Setores Entrada, Área Terminal, DBSCAN, K-Means.

1. INTRODUÇÃO

O crescimento do transporte aéreo tem acarretado sucessivos problemas de congestionamento e pontualidade, especialmente nos terminais de aeroportos movimentados. Isso tem ocasionado um aumento significativo nos custos operacionais das companhias aéreas como citado por Zhao et al. (2008). Logo, os estudos Xia et al. (2014) e Murça et al. (2018) revelam que o sistema de gerenciamento de tráfego aéreo (ATM) tem buscado desenvolver ferramentas e metodologias para monitorar e aprimorar a operação da aviação comercial, com o objetivo de atender às metas cada vez mais exigentes em termos de eficiência, segurança e impacto ambiental, além de lidar com o contínuo aumento do tráfego aéreo.

Nesse contexto, é fundamental realizar uma adequada mensuração e avaliação do desempenho a fim de identificar áreas passíveis de otimização, implementar medidas corretivas e embasar as tomadas de decisão. Diante disso, o Manual do Comando da Aeronáutica 100-22 (MCA 100-22) DECEA (2020) estabelece metodologias para medir o desempenho do sistema aéreo brasileiro em diversos aspectos, utilizando os Indicadores-Chave de Desempenho (KPIs).

Os padrões espaço-temporais em conjuntos de dados de trajetórias têm despertado cada vez mais interesse na literatura. Sabhnani et al. (2010) e Murça et al. (2018) citam que os estudos com técnicas de agrupamento têm sido empregados para identificar padrões de trajetórias em voo. Da mesma forma, a identificação de setores de entrada em área terminal pode resultar em melhorias no ATM, desempenhando um papel crucial na redução do consumo de combustível na aviação, como citado pelo Painel Intergovernamental sobre Mudanças Climáticas (2022), visto que essas melhorias podem levar a uma diminuição de 6% a 12% nas emissões CO_2 na aviação. Além disso, estudos de agrupamentos de setores podem aprimorar a mensuração do KPI-08, que refere-se à quantificação do tempo adicional nas áreas terminais.

No Brasil, o Centro de Gerenciamento da Navegação Aérea (CGNA), para o cálculo do KPI-08, adota a prática de identificação de setores de área terminal realizando uma divisão do espaço aéreo em setores fixos de 60° em torno dos aeródromos. Por outro lado, na Europa, especialistas da Eurocontrol realizam periodicamente uma setorização manual, conforme citado no relatório do Desempenho Operacional do ANS, disponibilizado por Eurocontrol & DECEA (2022). A Figura 1 mostra o histograma dos movimentos em função do ângulo em relação ao aeroporto e os setores identificados segundo a metodologia do CGNA (Figura 1.a) e Eurocontrol (Figura 1.b).

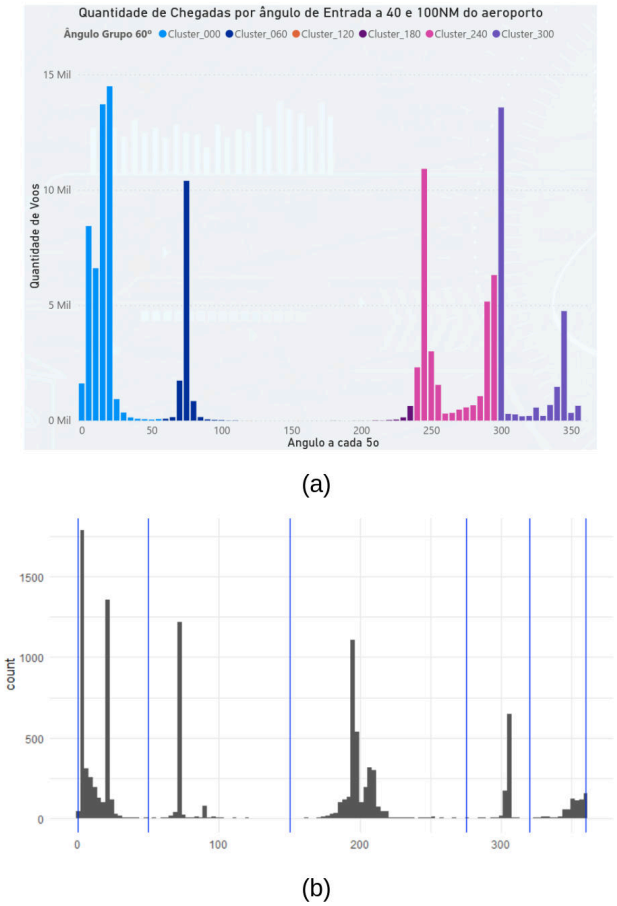


Figure 1 Identificação de setores de entrada em área terminal em Guarulhos (SBGR) de acordo com as metodologias do (a) CGNA e (b) Eurocontrol.

Fonte: Adaptado de Eurocontrol & DECEA (2022) e CGNA (2023).

A metodologia do CGNA, por realizar cortes fixos em ângulos de 60°, é facilmente replicável a demais aeroportos, entretanto possui uma precisão baixa quando comparado ao método da Eurocontrol. Na parte direita do histograma da Figura 1.a, por exemplo, podemos observar 3 picos de frequência que podem ser caracterizados como setores, entretanto apenas dois são identificados pela metodologia. A metodologia da Eurocontrol, entretanto, é mais precisa quanto a

identificação dos setores, porém, pelo fato de precisar de uma análise humana para cada aeroporto periodicamente, a aplicabilidade dessa solução torna-se difícil.

Estudos de agrupamentos possibilitam analisar a eficiência do tráfego aéreo ao combinar movimentos de acordo com a posição, velocidade e/ou proa. É importante mencionar que a identificação setores pode envolver alguns conceitos, como posição, velocidade e rota, a depender da aplicação, conforme citado por Guo et al. (2006). Diante disso, o presente estudo tem como objetivo propor um modelo para identificar automaticamente os setores de entrada em área terminal, considerando o raio de 40 NM e 100 NM em relação aos principais aeroportos do Brasil utilizando técnicas de aprendizado de máquina não supervisionado.

2. MATERIAIS E MÉTODOS

Nesta seção serão apresentadas informações referentes aos Dados Experimentais; Proposição de Modelos; e Métodos de Avaliação.

2.1. Dados Experimentais

Para o escopo deste trabalho, foram selecionados os Aeroportos de Guarulhos (SBGR), Congonhas (SBSP), Campinas (SBKP) e Brasília (SBBR). A seleção desses aeroportos ocorreu por registrarem os 4 maiores movimentos de pousos e decolagens no país em 2022 de acordo com o Anuário Estatístico, disponibilizado por CGNA (2023). O período de análise corresponde ao mês de agosto de 2022, por ser um período considerado pós-pandêmico na aviação, possuindo uma movimentação típica.

Os dados utilizados no estudo foram da Base de Dados ASMA (*Arrival Sequencing and Metering Area*), calculada a partir de dados de vigilância e mantida pelo Instituto de Controle do Espaço Aéreo (ICEA), composta pelas coordenadas geográficas de cada voo comercial apenas no instante da entrada em Áreas Terminais. Ressalta-se que, para este estudo, as áreas terminais foram simplificadas para círculos de 40 NM (C40) e 100 NM (C100) com centro em cada aeroporto, como sugerido na MCA 100-22 DECEA (2020).

O único pré-processamento adicional aplicado aos dados foi de padronização de escala baseado na média igual a 0 e desvios padrão igual 1 (Morettin & Bussab (2017)). A equação da padronização pode ser definida da seguinte forma:

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

em que Z_i é o valor padronizado; x_i é cada valor observado; μ é a média; e σ o desvio padrão. A utilização da padronização ocorreu devido a dispersão dos dados, além de possibilitar construir bons modelos de representação para *clusters* de maneira invariante a localização, isto é, permitir a reutilização de parâmetros para vários aeroportos.

2.2. Proposição de Modelos

A base de dados foi utilizada para a condução de um aprendizado de máquina não supervisionado para a tarefa de clusterização. Assim, para cada aeroporto, foi treinado um modelo para cada círculo (C40 e C100), utilizando latitude e longitude dos voos, a fim de agrupá-los e identificar os setores de entrada nas áreas terminais. Foram selecionadas duas das principais técnicas de aprendizado não supervisionado: *K-Means*, Johnson & Wichern (2002); e DBSCAN, Ester et al. (1996), descritas a seguir.

O *K-Means* e DBSCAN (*Density Based Spatial Clustering of Applications with Noise*) são dois algoritmos de aprendizado de máquina não supervisionado utilizados para tarefas de clusterização. O *K-Means* é um algoritmo de agrupamento baseado em centroides ou partição. Este algoritmo divide todos pontos do espaço amostral em K grupos de similaridade. A similaridade pode ser medida por uma métrica de distância Johnson & Wichern (2002). O DBSCAN, por outro lado, é um algoritmo baseado em densidade. A principal característica do algoritmo é que a vizinhança ponto a ponto em um *cluster* deve estar contida em um raio (ε) e cada *cluster* deve ter um número mínimo de pontos (*MinPts*). Além disso, o algoritmo indica ser extremamente eficiente na detecção de *outliers* e no

tratamento de ruídos Ester et al. (1996). O DBSCAN tem sido a principal escolha de algoritmo para tarefa de clusterização de voos quanto a dados de trajetória Murça et al. (2018).

Para o *K-Means*, o parâmetro K foi definido a partir do Método do Cotovelo, apresentado em Joshi & Nalwade (2013), em que o algoritmo foi executado repetidas vezes com K variando entre 2 e 10. O K que obteve o equilíbrio entre maior homogeneidade dentro do *cluster* e a maior diferença entre *clusters*, foi assumido como K ótimo.

Para o DBSCAN, os parâmetros definidos foram a distância máxima entre vizinhos de um *cluster* (ε) e o número mínimo de pontos (*MinPts*), que um conjunto precisa ter para ser considerado um *cluster*. Para estabelecer o ε foi utilizado a função *kNNdistplot* que exibe um gráfico de distância e para *MinPts* foi determinado o valor de 5% do total amostra.

2.3. Métodos de Avaliação

A avaliação dos resultados foi feita analiticamente, observando os *clusters* formados sobre C100 e C40 e comparando-os com os procedimentos de aproximação de acordo com as cartas aeronáuticas de cada aeroporto. Foram observados também os histogramas de voos por ângulo de entrada nas áreas terminais para cada aeroporto e comparado com os *clusters* formado a partir das coordenadas geográficas. Essa é uma importante visualização a fim de comparar com as metodologias atualmente utilizadas no CGNA e Eurocontrol (Figura 1). Além disso, foi feita uma análise de proporção de ruídos sobre os resultados do algoritmo DBSCAN.

3. RESULTADOS E DISCUSSÕES

Nesta seção serão apresentados os resultados da clusterização dos setores de entrada em C40 e C100 para os aeroportos analisados utilizando os algoritmos K-Means e DBSCAN. As subseções estão organizadas da seguinte forma: Resultado da Clusterização, que apresenta os *clusters* identificados por ambos os métodos; Validação de Clusters, que mostra a comparação entre os *clusters* identificados com os procedimentos de aproximação de cada aeroporto; Análise de Ruído, que mostra a proporção de movimentos considerados ruídos em cada *cluster* apenas para o DBSCAN, visto que o método *K-Means* não possibilita esse tipo de análise; e Comparação com Trabalhos Relacionados.

3.1. Resultado da Clusterização

Os resultados para o *K-means* constam na Figura 2. Os histogramas exibem a distribuição de movimentos por ângulo em relação a cada aeroporto. Está assinalado também, por cores, os *clusters* resultantes do *K-Means*. O método *K-Means* clusterizou os movimentos dos aeroportos em até 6 *clusters*, seguindo o método do cotovelo, conforme a Tabela 1, que mostra a quantidade de *clusters* resultantes dos métodos K-Means e DBSCAN. O aeroporto SBKP foi o que apresentou o menor número de *clusters*, sendo $k = 3$ para C40.

Table 1 Número de *clusters* resultantes dos métodos *K-Means* e DBSCAN.

Aero.	K-Means		DBSCAN	
	Clusters (C40)	Clusters (C100)	Clusters (C40)	Clusters (C100)
SBGR	4	4	3	4
SBKP	3	5	4	4
SBBR	4	4	4	4
SBSP	6	4	3	4

A Figura 2 permite, ainda, identificar que alguns setores não foram alocados devidamente. Em C40 para SBKP, por exemplo, há a altas frequências em dois ângulos distintos (próximos a 45° e 120°), que sugere a existência de dois setores, entretanto o método identificou apenas um, bem como em C100 de SBSP, próximos a 320° e 355°. Além disso, percebe-se que o método é sensível a ruídos, por não conseguir identificá-los, como acontece em C40 de SBSP, no *cluster* assinalado pela cor vermelha, onde não há uma densidade alta para caracterizar um setor. Similarmente, próximo a 320° em C40 de SBSP, há apenas uma frequência que justifica a identificação de um setor, no entanto 3 setores foram identificados.

O número de *clusters* identificados pelo DBSCAN foi de 3 para C40 em SBGR e SBSP. Para os demais cenários, foram identificados 4 *clusters*. A Figura 3 apresenta uma distribuição

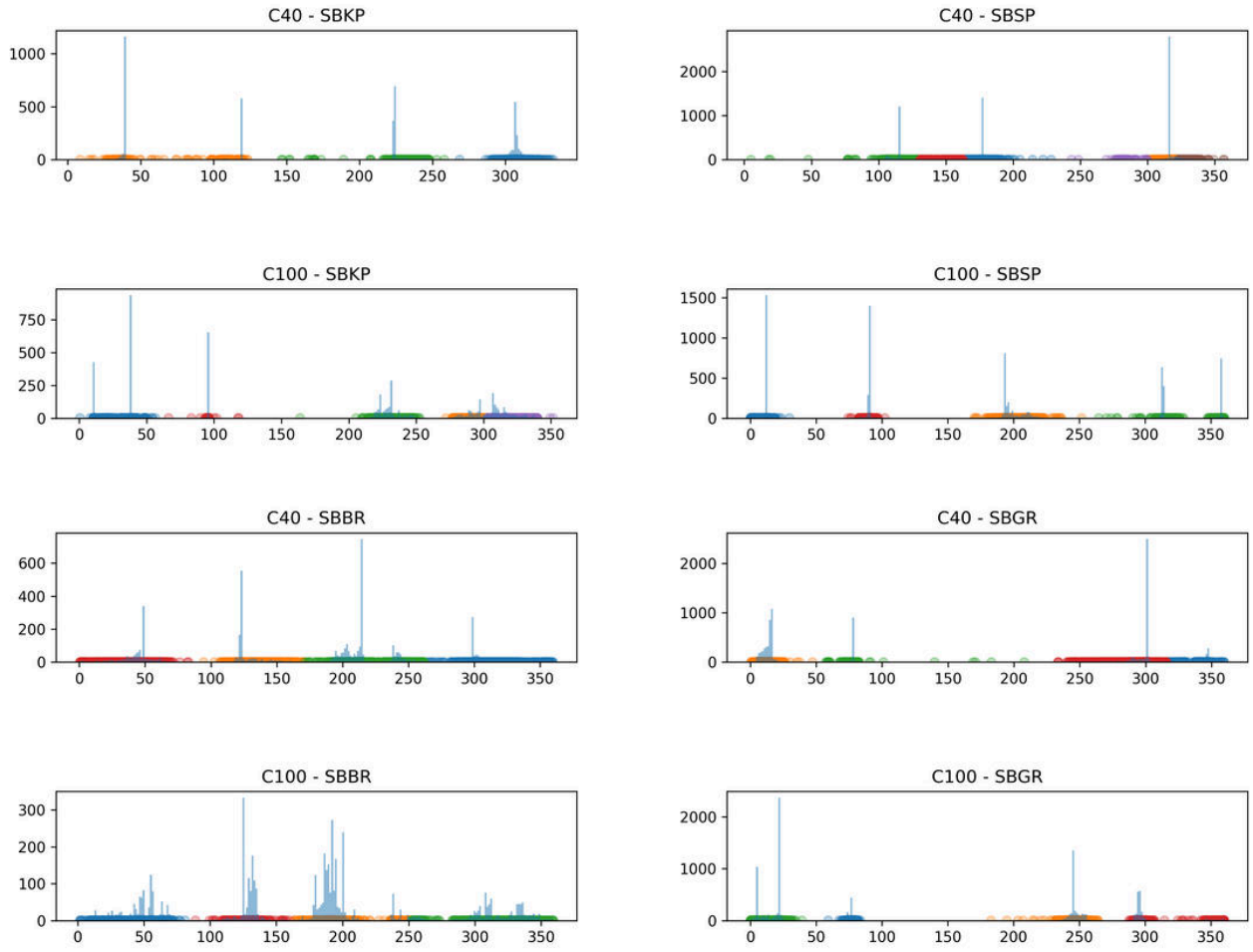


Figure 2 Distribuição de movimentos por ângulo em relação a cada aeroporto. As cores na parte inferior de cada histograma correspondem aos *clusters* identificados para os movimentos a partir do *K-Means*.

para a quantidade de *clusters*, identificados por cores na parte inferior dos histogramas, além de identificar os ruídos presentes na amostra, com a cor cinza. O DBSCAN mostrou-se eficaz na identificação dos setores, por considerar a densidade de movimentos para clusterização. Desta forma, os setores identificados são referentes aos ângulos com maior densidade de movimentos, enquanto movimentos em regiões com baixa densidade foram considerados ruídos.

Avaliando os resultados, foi possível observar que o método *K-Means* exibiu a maior quantidade de *clusters*, entretanto o método não mostrou-se eficaz, por não conseguir identificar movimentos com características atípicas, prejudicando assim a explicação dos *clusters*, o que levou a uma análise não precisa e não coerente com o esperado. O DBSCAN, por outro lado, mostrou-se preciso e explicável na identificação dos *clusters* e dos movimentos atípicos, considerados como ruídos. Isso ocorreu devido o método considerar não só a posição dos movimentos para a clusterização, mas também a densidade por setor.

3.2. Validação de *Clusters*

Para validar os resultados do DBSCAN, que obteve melhor desempenho na identificação dos setores, realizou-se uma análise georreferenciada. Essa análise permite comparar os setores identificados com os procedimentos de aproximação dos aeroportos, como mostra a Figura 4.

Na figura, as linhas azuis representam os procedimentos de aproximação disponíveis nas cartas dos aeroportos, enquanto as regiões coloridas simbolizam os movimentos clusterizados em C100 e C40, em que cada cor ilustra o setor a que cada voo foi clusterizado. Os procedimentos ilustram o conjunto de pontos de re-

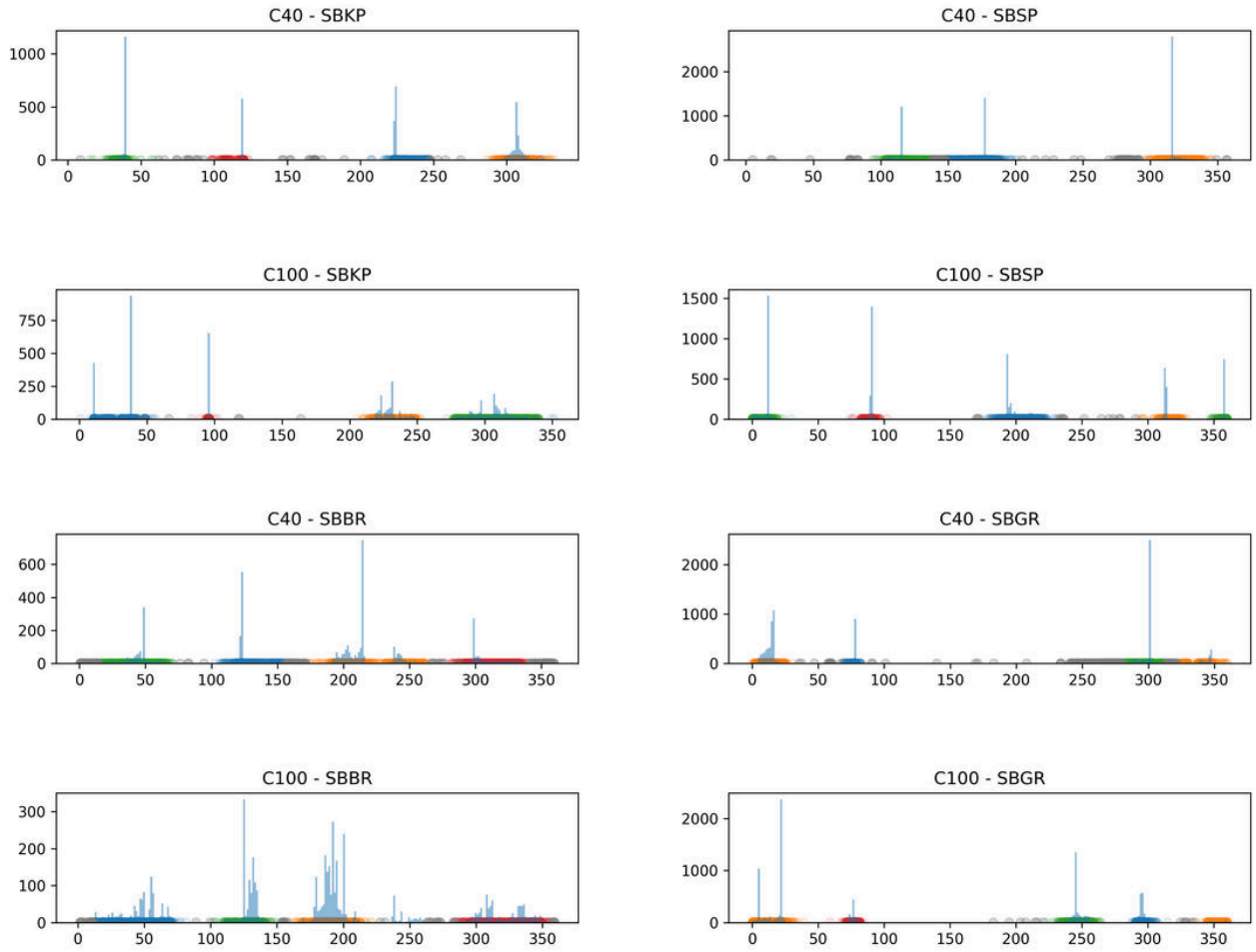


Figure 3 Distribuição de movimentos por ângulo em relação a cada aeroporto. As cores na parte inferior de cada histograma correspondem aos *clusters* identificados para os movimentos a partir do DBSCAN.

ferência para uma aeronave realizar uma manobra de aproximação ao aeroporto. Embora não seja obrigatório cada aeronave seguir sempre todos esses pontos de referência, esses procedimentos são importantes para validar empiricamente os *clusters* formados. Foi observado também o alinhamento entre os procedimentos e os setores identificados, o que valida o desempenho do método DBSCAN.

No geral, a identificação dos setores possibilitam análises mais assertivas em área terminal, o que pode contribuir para melhorias na capacidade do espaço aéreo; redução no atraso dos voos; e na estimação de indicadores de performance, como o KPI-08, mais confiáveis. Além disso, o método utilizado para identificação de setores de entrada em área terminal demonstrou-se computacionalmente escalável, utilizando os mesmos parâmetros para todos aeroportos, o que possibilita a aplicação da solução em outros cenários sem a necessidade de reajuste de parâmetros.

3.3. Análise de Ruído

O DBSCAN se mostrou robusto em relação aos ruídos, o que permitiu uma clusterização mais precisa dos setores. Ao aplicar o DBSCAN, os setores foram agrupados com base na densidade de voos em cada região. Isso significa que regiões com maior densidade de voos foram agrupadas em *clusters*, enquanto regiões com menor densidade de voos foram consideradas ruídos ou movimentos atípicos.

A Tabela 2 mostra as quantidades de movimentos que cruzaram C100 e C40 e as respectivas proporções de ruído, identificado pelo DBSCAN, para cada um dos aeroportos estudados. Foram identificados de 0,06% a 6,59% de ruídos nos dados de C100 e C40. Embora a proporção de ruídos não seja impactante para o estudo, é sufici-

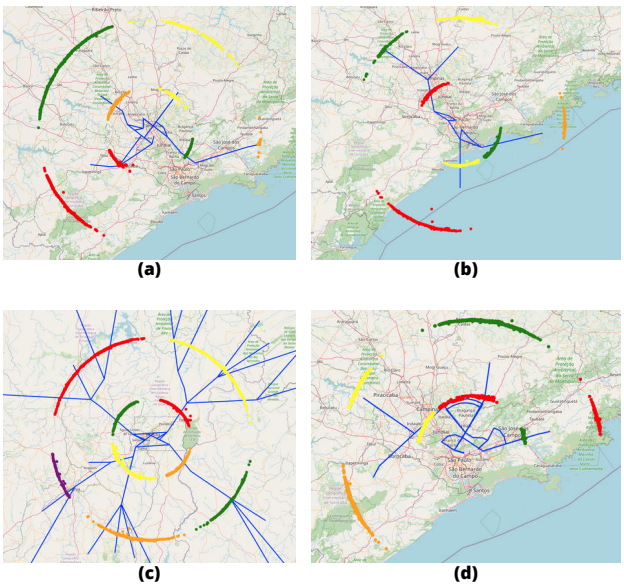


Figure 4 Movimentos clusterizados em C100 e C40 em (a) SBKP; (b) SBSP; (c) SBBR; e (d) SBGR. As linhas em azul indicam os procedimentos de pouso de acordo com as cartas dos aeródromos.

ente para prejudicar o desempenho do K-Means. O reconhecimento do ruído pelo DBSCAN faz melhorar a identificação dos *clusters*, preservando ainda um conjunto de dados significativo.

A capacidade do DBSCAN de separar ruídos da análise é uma vantagem expressiva em comparação ao K-Means. O K-Means torna-se menos eficaz por não considerar os movimentos atípicos em grupos distintos, levando a resultados não adequados. Portanto, o uso do DBSCAN na clusterização dos setores de entrada em C40 e C100 mostrou-se acurado e eficaz.

3.4. Comparação com trabalhos relacionados

Comparando o desempenho do método DBSCAN com os métodos aplicados pelo CGNA e Eurocontrol, podemos concluir que o método proposto foi capaz de identificar os setores com a precisão esperada, além de encontrar achados similares com aos resultados da Eurocontrol. Entretanto, o método proposto dispensa a utilização de um indivíduo para a identificação de setores, ao contrário da Eurocontrol, agregando a eficiência na execução do método e permitindo que ele seja replicado e escalado para outros cenários.

A metodologia implantada pelo CGNA, por sua vez, não se mostrou ajustável na identificação de setores, pois realiza a divisão por ângulo fixo de 60°, não considerando variáveis como densidade de movimentos e posicionamento das aeronaves para a identificação dos setores.

Um outro detalhe importante de ressaltar é que o modelo proposto considera, por voo, um único ponto em área terminal, pertencente a C100 e C40, para realizar a clusterização. Isso implica em a eficiência do algoritmo ser proporcional ao número de movimentos. O mesmo não ocorre em trabalhos de clusterização de fluxo, por exemplo, que levam em consideração as trajetórias completas do voo, como em Murça et al. (2018). Nesse caso, a complexidade computacional aumenta consideravelmente, uma vez que a solução passa a ser proporcional ao número de pontos das trajetórias dos voos analisados, sendo esses dados provenientes de bases de vigilância, que têm taxa de atualização na ordem de segundos.

4. CONSIDERAÇÕES FINAIS

O presente estudo realizou a aplicação de algoritmos de aprendizado de máquina não supervisionado, pelos métodos *K-Means* e DBSCAN, para a tarefa de clusterização de setores de entrada em área terminal nos aeroportos SBBR, SBKP, SBSP e SBGR. O DBSCAN obteve melhores resultados em comparação ao *K-Means*, devido o método considerar a densidade dos movimentos para a identificação dos setores, além de identificar os movimentos atípicos, considerando-os como ruído, o que não ocorre com o *K-Means*.

O DBSCAN também possibilitou a utilização dos mesmos parâmetros independente do raio do círculo (C40 e C100) e do aeroporto, tornando-se eficiente do ponto de vista da implantação da solução. Além disso, foi feita uma análise de ruído na amostra, que identificou uma proporção média de 4,13% e 1,71% para C40 e C100, respectivamente. O que não tem impacto para o estudo, porém influencia negativamente o desempenho do *K-Means*.

O método DBSCAN pode auxiliar na identificação de padrões e características comuns entre os setores de entrada, permitindo uma melhor compreensão da distribuição de fluxo de aeronaves nos aeroportos; otimização dos processos

Table 2 Proporção de Ruídos nas amostras de movimentos.

Aeroporto	Movimentos C40	Ruídos - C40	Movimentos C100	Ruídos - C100
SBGR	7142	166 (2,32%)	7901	7 (0,09%)
SBKP	4618	11 (0,24%)	4922	3 (0,06%)
SBBR	4080	161 (3,99%)	4066	268 (6,59%)
SBSP	5806	56 (0,96%)	6198	5 (0,08%)

de planejamento e gerenciamento; e na estimação do KPI-08 de maneira mais confiável e assertiva.

Os achados mostraram-se superiores às soluções utilizadas pelo CGNA e Eurocontrol. Isso se deu pelo fato da metodologia proposta realizar a identificação dos setores automaticamente, sem a necessidade de intervenção humana, ao contrário da Eurocontrol, e levando em consideração a densidade de movimentos, ao contrário do CGNA.

Para trabalhos futuros, sugere-se utilização e comparação com outros métodos de aprendizado não supervisionado; testes com outros aeródromos, a fim de validar o método em outros cenários; e validar estatisticamente a explicabilidade dos setores encontrados. Sugere-se também comparar o desempenho da clusterização apresentada neste trabalho, que utiliza apenas um ponto da trajetória por voo, com trabalhos baseado na trajetória completa, para verificar se há uma diferença significativa no resultado da clusterização, ainda que a performance computacional deste trabalho mantenha-se superior. Por fim, é importante ressaltar que a identificação dos setores de entrada na área terminal pode ser uma tarefa complexa e exigir a colaboração de diferentes partes interessadas, incluindo serviços de navegação aérea e controladores de tráfego aéreo.

Referências

CGNA (2023). Anuário estatístico de tráfego aéreo 2022, Departamento de Controle do Espaço Aéreo.

DECEA (2020). Metodologia de indicadores atm do sisceab, Manual do Comando da Aeronáutica - MCA 100-22.

Ester, M., Kriegel, H., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Vol. 96 pp. 226—231.

Eurocontrol & DECEA (2022). Brazil/europe comparison of operational ans performance, Eurocontrol e Departamento de Controle do Espaço Aéreo.

Guo, J., Qin, Z., He, S. & Huang, H. (2006). Simulation of space group formation for maneuvering target, Hsi-An Chiao Tung Ta Hsueh/Journal of Xi'an Jiaotong University 40, 1378–1382.

Johnson, R. & Wichern, D. (2002). Applied Multivariate Statistical Analysis, Prentice Hall,.

Joshi, K. D. & Nalwade, P. S. (2013). Modified k-means for better initial cluster centres, International Journal of Computer Science and Mobile Computing. 2-7, 219–223.

Morettin, P. A. & Bussab, W. O. (2017). Estatística básica, Saraiva Educação SA,.

Murça, M. C. R., Hansman, R. J., Li, L. & Ren, P. (2018). Flight trajectory data analytics for characterization of air traffic flows: A comparative analysis of terminal area operations between new york, hong kong and sao paulo, Transportation Research Part C: Emerging Technologies 97, 324–347.

Sabhnani, G., Yousefi, A. & Mitchell, J. S. (2010). Flow conforming operational airspace sector design Anais do 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, p. 9377.

Xia, Q., Li, X., Song, Y. & Zhang, B. (2014). Aircraft grouping based on improved divisive hierarchical clustering algorithm, Journal of Air Transport Management 40, 157–162.

Zhao, T., Man, Z., Wan, Z. & Bi, G. (2008). A cgs-msm parallel genetic algorithm based on