

Aprendizado de Máquina para a Otimização da Obtenção de Resultados em Simulações de Defesa Aeroespacial

**Lucas Silva Lima
Rafael Hoffmann Giannico
Denys Derlian Carvalho Brito
Antonio Gustavo Silveira Dantas
João Paulo De Andrade Dantas**

*Instituto Tecnológico de Aeronáutica
ITA*

23 de julho de 2024



Roteiro

1 Introdução

- Contexto
- Objetivo

2 Metodologia

- Definição do Caso de Uso
- Coleta de Dados
- Preparação dos Dados
- Desenvolvimento dos Modelos de *ML*

3 Resultados

- Performance dos Modelos

4 Discussão

5 Considerações

- Limitações
- Sugestões

6 Contatos



Introdução



Simulação

- A simulação aeroespacial tem sido amplamente utilizada no campo da aviação e defesa há décadas.
- Desde os primórdios, a simulação desempenha um papel fundamental no treinamento de pilotos, no projeto de aeronaves e no planejamento de missões.
- *Ambiente de Simulação Aeroespacial (ASA).*

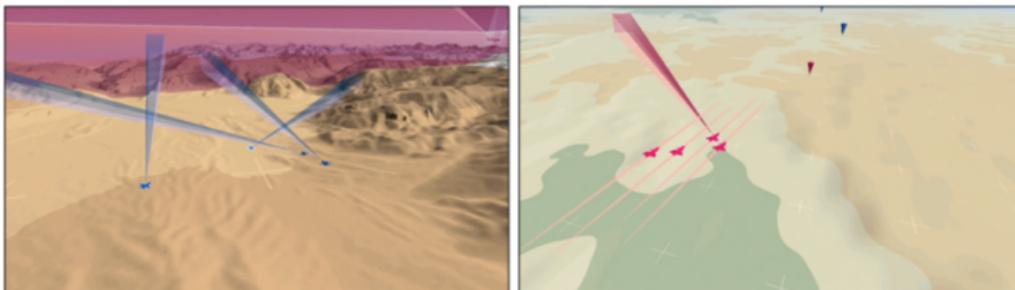


Figura: Ambiente de Simulação Aeroespacial (ASA).



Aprendizado de Máquina

- Crescimento do poder computacional.
- Aumento da disponibilidade de dados.
- Possibilidade de implementar e aprimorar de forma iterativa novos algoritmos e principalmente algoritmos propostos no passado.

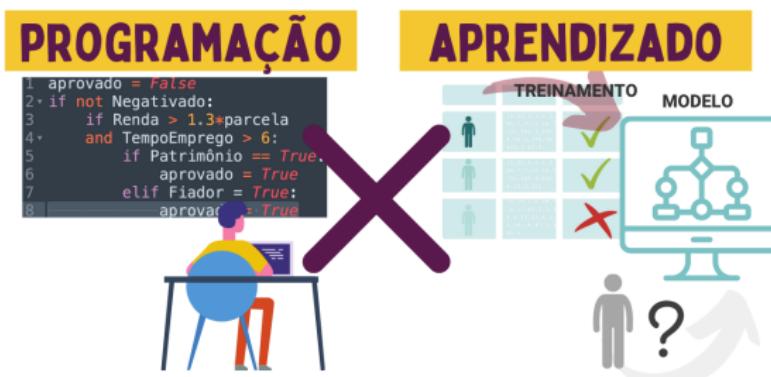


Figura: Programação x Aprendizado de Máquina.



Aprendizado de Máquina

■ Aprendizado Supervisionado.

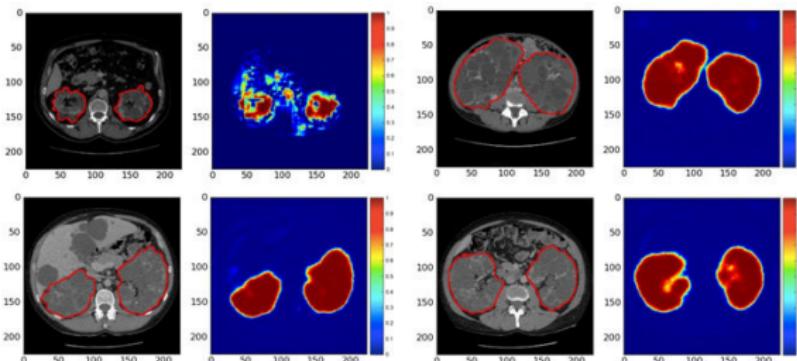


Figura: Exemplo de aprendizado supervisionado (regressão e classificação)



Aprendizado de Máquina

- Aprendizado não Supervisionado.

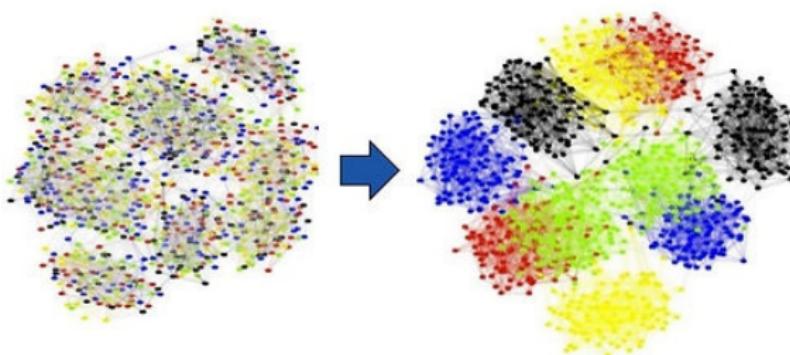


Figura: Exemplo de aprendizado não supervisionado (agrupamento)



Objetivo

- Investigar o uso de *ML* para acelerar a aquisição de novos resultados em simulações aeroespaciais previamente executadas no ASA.
- Categorias de *ML* escolhidos: supervisionado (regressão e classificação) e não supervisionado (agrupamento).



Metodologia



Caso de Uso

- Natureza reservada das informações relacionadas a modelos realísticos e cenários operacionais complexos.
- Caso de uso será restrito a um cenário simplificado.
- A análise foi estruturada para ser agnóstica ao grau de complexidade do cenário sob investigação.



Cenário: Previsão do consumo de combustível durante uma missão de navegação de uma aeronave de caça

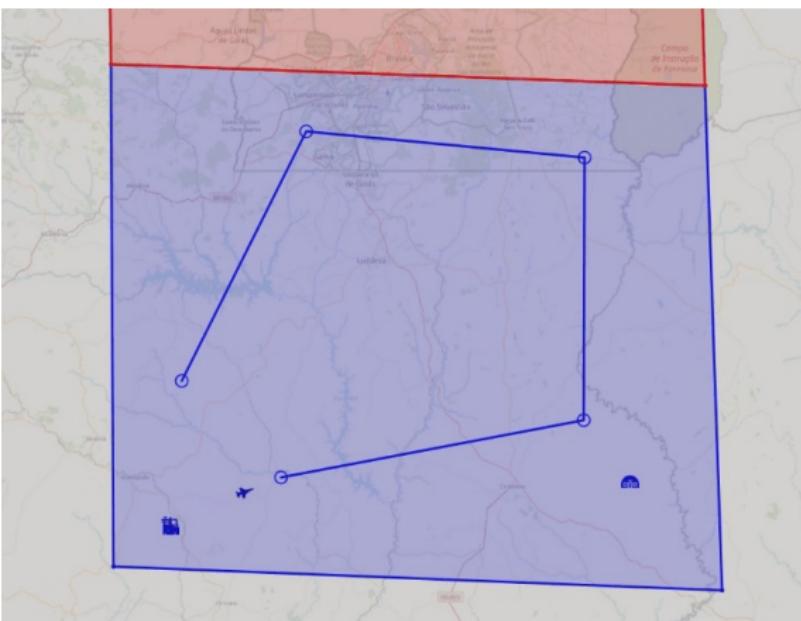


Figura: Interface gráfica do usuário para construção de cenários do ASA.



Coleta de Dados

- 4.000 episódios de simulação.
- Diversas altitudes e velocidades.
- Amostragem sob a metodologia *Latin Hypercube Sampling (LHS)*.

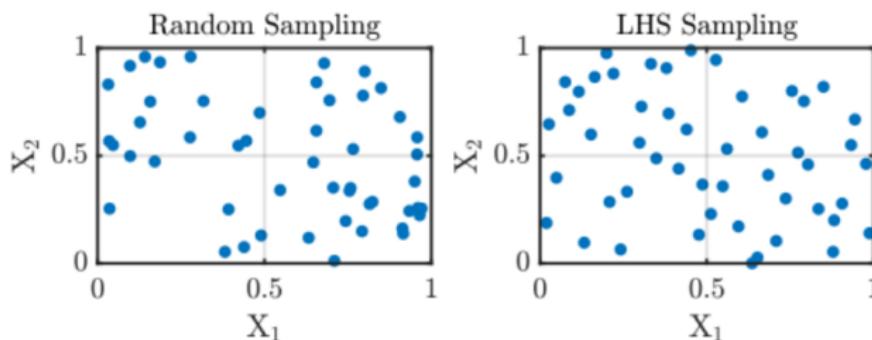


Figura: Exemplo de amostragem sem e com metodologia *LHS*.



Exemplo de Dados

Exp.	Vel. (knot)	Alt. (ft)	Comb. consumido (lb)
244	505.22	23646.51	1065.37
411	546.05	30595.64	945.54
1480	350.66	19152.46	1115.60
2015	352.43	19110.39	1103.92
3572	465.19	12089.28	1431.66

Tabela: Amostra dos Dados Obtidos da Simulação



Preparação dos Dados

- Os dados foram submetidos a uma verificação de integridade, transformação de padronização *Z-Score* e divisão em conjuntos de treinamento e validação (85%) e teste (15%).

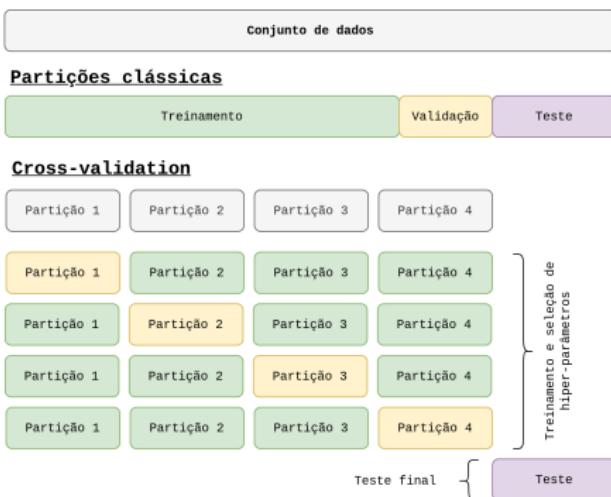


Figura: Estratégia de particionamento dos dados.



Boxplot dos Dados

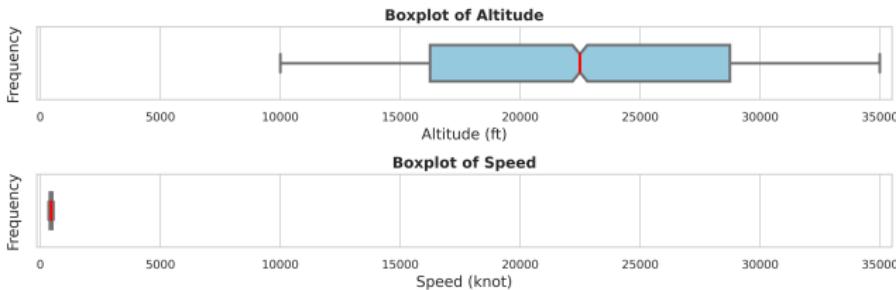


Figura: Boxplot dos dados antes de serem padronizados.



Boxplot dos Dados

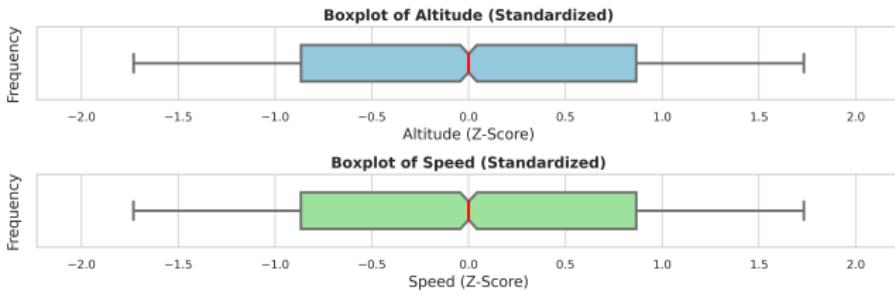


Figura: Boxplot dos dados após serem padronizados.



Modelos de Regressão

- Meta: Estimar o consumo de combustível para combinações não simuladas de altitude e velocidade.
- Previsores: 2 modelos de regressão desenvolvido com os algoritmos a) *Redes Neurais Artificiais (RNA)* e b) *Extreme Gradient Boosting (XG Boost)*.
- Métrica de avaliação: RMSE e R².



Modelo de Agrupamento

- Meta: Categorizar os registros de voo em 4 regimes distintos.
- Previsor: 1 modelo de agrupamento desenvolvido com o algoritmo a) *K-Means*.
- Métrica de avaliação: Índice de Silhueta.



Modelos de Classificação

- Meta: Dadas as categorias de regime obtidas do modelo anterior, prever a qual categoria um novo registro de voo pertence.
- Previsores: 2 modelos de classificação desenvolvidos com os algoritmos a) *RNA* e b) *XG Boost*.
- Métrica de avaliação: Entropia Cruzada Categórica.



Métricas de Avaliação

Raiz do Erro Quadrático Médio (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

Coeficiente de Determinação (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

Entropia Cruzada Categórica (CC)

$$\text{Categorical Crossentropy} = - \sum_{i=1}^M y_i \log(\hat{y}_i) \quad (3)$$



Resultados



Modelos de Regressão

- O modelo *XG Boost* apresentou *RMSE* de 0.03385 e R^2 de 0.999
- O modelo *RNA* apresentou *RMSE* de 0.23167 e R^2 de 0.947



Modelos de Regressão

■ Gráfico da evolução do treinamento (*RNA*)

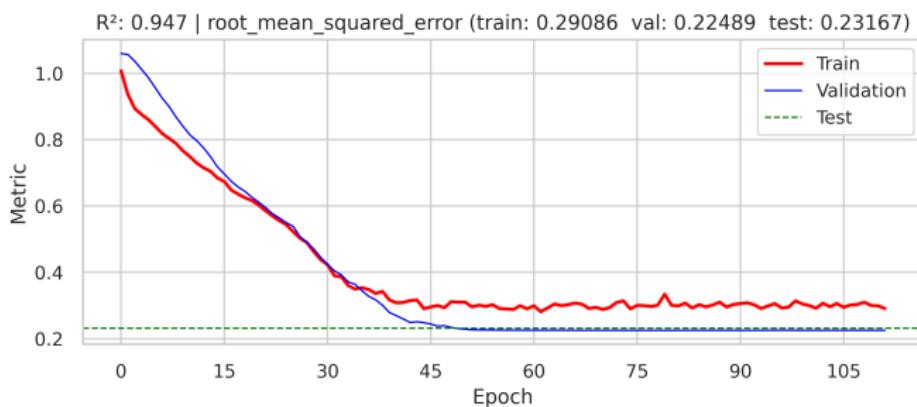


Figura: Gráfico de Bias-Variância do modelo de regressão baseado no algoritmo *RNA*.



Modelos de Regressão

- Gráfico da evolução do treinamento (*XG Boost*)

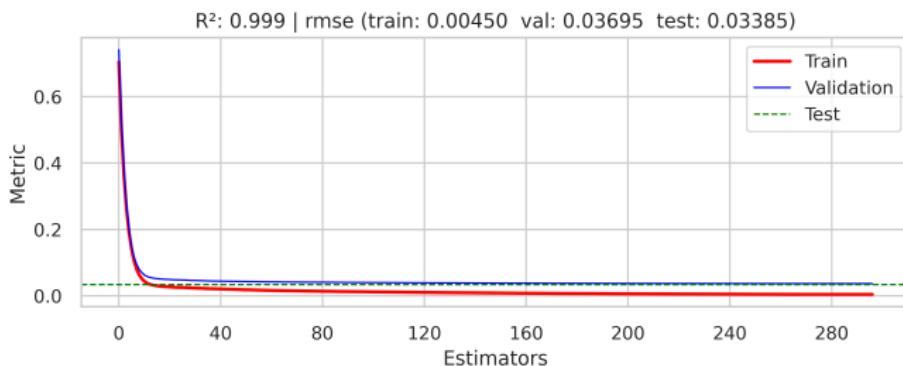


Figura: Gráfico de Bias-Variância do modelo de regressão baseado no algoritmo XG Boost.



Modelos de Regressão

■ Exemplos de Previsão

Tabela: PREVISÃO DE CONSUMO DE COMBUSTÍVEL NAS SIMULAÇÕES PREVIAMENTE AMOSTRADAS (XG Boost)

Exp.	Comb. cons. Real (lb)	Comb. cons. Previsto (lb)
244	1065.37	1075.48
411	945.54	944.78
1480	1115.60	1113.65
2015	1103.92	1095.03
3572	1431.66	1437.84



Modelo de Agrupamento

O modelo de agrupamento *K-Means* obteve um *Índice de Silhueta* de 0.39453.

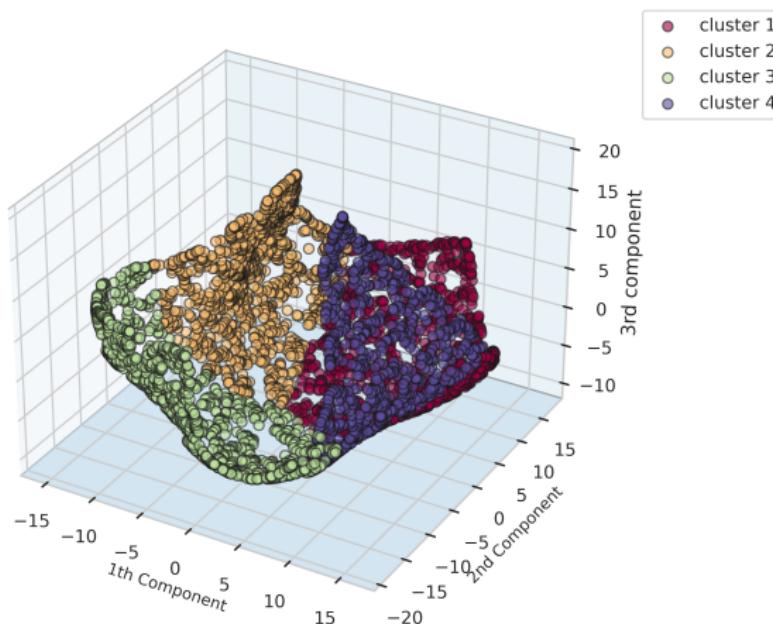


Figura: Gráfico de *Incorporação Estocástica de Vizinhos com Distribuição t*.



Modelos de Classificação

- O modelo *XG Boost* obteve *Entropia Cruzada Categórica* de 0.02985
- O modelo *RNA* obteve *Entropia Cruzada Categórica* de 0.12499



Modelos de Classificação

■ Gráfico da evolução do treinamento (*RNA*)

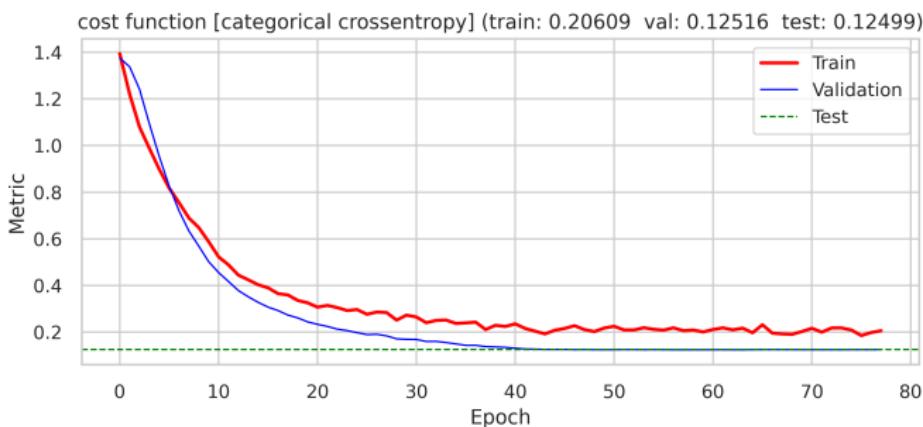


Figura: Gráfico de Bias-Variância do modelo de classificação baseado no algoritmo *RNA*.



Modelos de Classificação

■ Gráfico da evolução do treinamento (*XG Boost*)

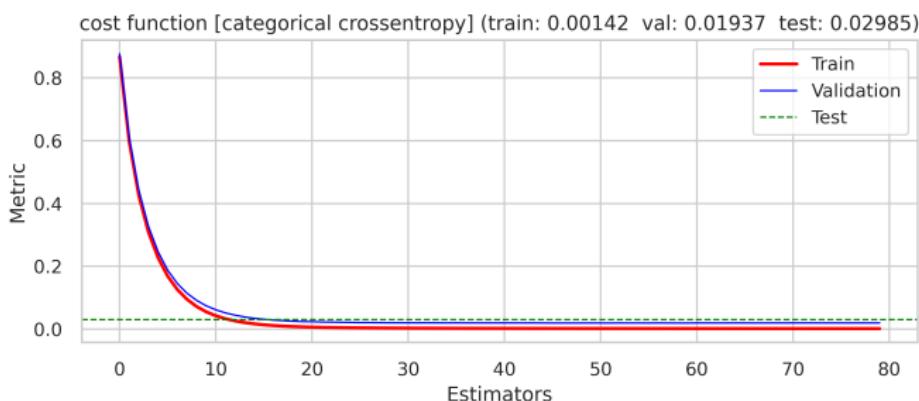


Figura: Gráfico de Bias-Variância do modelo de classificação baseado no algoritmo XG Boost.



Modelos de Classificação

■ Matriz de Confusão (*RNA*)

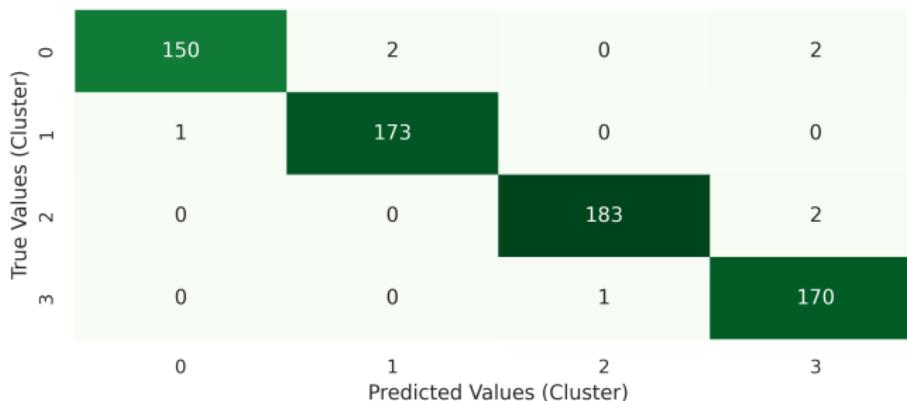


Figura: Matriz de confusão do modelo de classificação baseado no algoritmo *RNA*.



Modelos de Classificação

■ Matriz de Confusão (*XG Boost*)

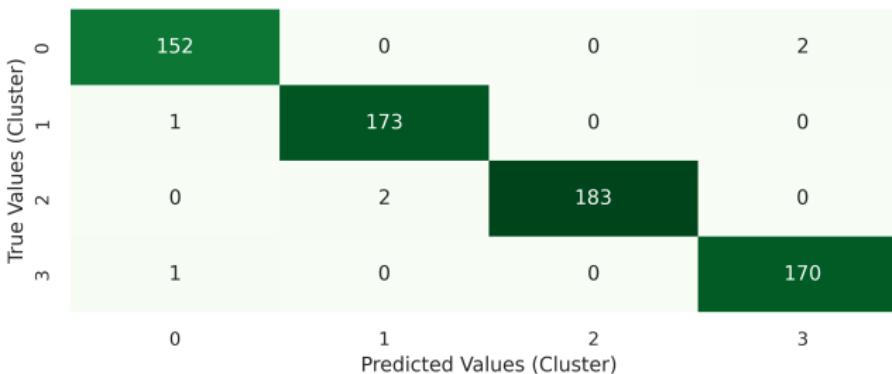


Figura: Matriz de confusão do modelo de classificação baseado no algoritmo *XG Boost*.



Tempo de Processamento

Método	T. de treino (s)	T. de Inferência (s)
Simulação (ASA)	48013.000	-
RNA (Reg.)	1189.263	0.009
XG Boost (Reg.)	225.531	0.009
K-Means (cluster.)	38.348	-
RNA (Class.)	1079.477	0.010
XG Boost (Class.)	200.894	0.004

Tabela: Tempo de Treinamento/Simulação e Inferência dos Métodos



Discussão



Discussão

- Os resultados apontam uma aplicabilidade de técnicas de *ML* para acelerar o tempo de obtenção de resultados das simulações aeroespaciais.
- A redução no tempo de simulação, sem comprometer a fidelidade dos resultados, destaca o potencial das abordagens de *ML* para tornar mais ágil e preciso o planejamento e a execução de missões aéreas.



Considerações



Limitações

- Tendência dos modelos de *ML* a se tornarem especialistas em prever cenários estritamente dentro do espectro dos dados usados durante o treinamento.
- Seleção abrangente e representativa de dados de treinamento é crucial.



Presente e Futuro

- Ampliar a aplicação dos modelos para cenários de combate com diferentes tipos de vetores e condições.
- Desenvolver e validar novas táticas operacionais utilizando modelos preditores.
- Integrar modelos baseados em *Reinforcement Learning (RL)* para decisões diante de cenários específicos.



Figura: Exemplo de cenário complexo onde é simulado um combate aéreo além do alcance visual.



Contatos



Contatos



- limalsl@fab.mil.br
- limalsl@ita.br
- www.linkedin.com/in/lucaslima25

