

Predição de Janelas de Lançamento Espacial a partir de Condições Meteorológicas com Modelos de Aprendizado de Máquina

Larissa Fernanda Gonçalves Dias

Instituto HBR – Programa de Capacitação em Inteligência Artificial Aplicada ao Setor Aeroespacial

larifgdias@gmail.com

Abstract. This work presents a machine learning approach to estimate the feasibility of space launch windows based on meteorological conditions. Launch operations are strongly affected by wind, precipitation and instability, often causing delays. We combine historical launch data with ERA5 weather variables from the 48 hours preceding each launch. Logistic Regression, Gradient Boosting and XGBoost were evaluated with temporal splits. Gradient Boosting achieved the best performance. Wind and precipitation were the most influential features. Results indicate that machine learning can support go/no-go decisions and reduce operational risks.

Keywords: machine learning, launch window prediction, weather data, ERA5, classification.

Resumo. Este trabalho apresenta uma abordagem de aprendizado de máquina para estimar a viabilidade de janelas de lançamento espacial com base em condições meteorológicas. Operações de lançamento são fortemente afetadas por vento, precipitação e instabilidade, fatores que frequentemente causam adiamentos. Combinamos dados históricos de lançamentos com variáveis do ERA5 referentes às 48 horas que antecederam cada missão. Avaliamos Regressão Logística, Gradient Boosting e XGBoost, utilizando divisão temporal. O Gradient Boosting apresentou o melhor desempenho. Vento e precipitação foram as variáveis mais influentes. Os resultados indicam que modelos de aprendizado de máquina podem apoiar decisões de go/no-go e reduzir riscos operacionais.

Palavras-chave: aprendizado de máquina, meteorologia, lançamento espacial, era5, classificação.

1. Introdução

O lançamento de veículos espaciais é um processo complexo que depende fortemente de condições ambientais favoráveis. Variáveis como vento, instabilidade atmosférica, precipitação e baixa visibilidade podem representar riscos significativos, resultando em adiamentos ou cancelamentos de missões. Alterações de janela de lançamento geram custos operacionais elevados, exigem realocação de equipes e comprometem a eficiência do planejamento.

Nos últimos anos, o aumento da disponibilidade de dados climáticos provenientes de reanálises, como o ERA5, juntamente com bases públicas contendo histórico de lançamentos e seus desfechos, abriu novas possibilidades para a aplicação de técnicas de aprendizado de máquina no apoio à decisão de go/no-go. A integração dessas fontes permite capturar padrões atmosféricos que antecedem lançamentos bem-sucedidos ou adiados.

O objetivo deste trabalho é desenvolver um modelo de classificação binária capaz de prever a viabilidade de uma janela de lançamento com base em variáveis meteorológicas observadas nas 48 horas que antecedem a data programada da missão. Para isso, propomos um pipeline que integra coleta, pré-processamento e modelagem com algoritmos supervisionados, buscando oferecer uma ferramenta de apoio ao planejamento de missões e à redução de riscos operacionais. A relevância deste estudo está em fornecer subsídios quantitativos para decisões críticas, contribuindo para maior precisão e confiabilidade no processo de lançamento.

2. Metodologia

Esta seção descreve o processo adotado para construção do modelo preditivo, incluindo as bases de dados utilizadas, procedimentos de coleta e integração, definição do rótulo, engenharia de atributos, divisão dos conjuntos de dados e modelos avaliados.

2.1. Bases de dados

Foram utilizadas duas fontes principais de informação:

a) Histórico de lançamentos espaciais

Utilizou-se uma base pública contendo registros de lançamentos realizados desde 1957, incluindo data programada, data efetiva, local de lançamento, agência responsável, veículo utilizado e status (sucesso, atraso ou cancelamento). Esses dados serviram como referência para derivar o rótulo “realizado conforme planejado” ou “adiado”.

b) Dados meteorológicos (ERA5)

Para caracterizar as condições atmosféricas que antecederam cada lançamento, foram extraídas variáveis meteorológicas do conjunto **ERA5**, que fornece reanálises climáticas globais horárias, com resolução espacial aproximada de 31 km. As variáveis selecionadas incluem:

- velocidade e rajada de vento,
- precipitação,
- pressão ao nível do mar,
- temperatura do ar,
- umidade relativa,
- cobertura de nuvens,
- indicadores de instabilidade atmosférica.

Essas variáveis foram coletadas para as **48 horas anteriores** à data programada de cada lançamento.

2.2. Coleta e integração dos dados

Para cada lançamento do conjunto histórico, identificou-se o **local** (latitude e longitude) e a **data programada**. Em seguida:

1. Para cada coordenada, foram extraídas, no ERA5, as séries horárias das variáveis meteorológicas no intervalo de **T – 48h até T**, sendo T a data/hora programada.
2. Os dados meteorológicos foram alinhados temporalmente ao registro de cada lançamento.
3. Após a coleta, todos os registros foram unificados em um único dataset estruturado, com uma linha representando cada lançamento.

2.3. Definição do rótulo (OK / Atrasado)

O problema foi tratado como **classificação binária**. O rótulo foi definido da seguinte forma:

- **OK** – o lançamento ocorreu na data programada.
- **Atrasado** – o lançamento foi adiado, reagendado ou cancelado devido a condições adversas.

A classificação foi realizada comparando as datas previstas e efetivas, além de informações do campo “status” da base histórica.

2.4. Engenharia de atributos

Para representar as condições meteorológicas de forma significativa para o modelo, foram aplicadas transformações nas séries horárias:

- **Agregações**: média, máximo, mínimo e desvio padrão nas janelas
 - 0–6h antes do lançamento
 - 6–24h
 - 24–48h
- **Variáveis binárias** indicando condições críticas (ex.: vento > 15 m/s, precipitação > 1 mm/h).
- **Tendência (“slope”)** nas séries de vento, pressão e precipitação.
- **Informações do local**, como latitude, longitude e histórico recente de atrasos no mesmo centro de lançamento.

Essas transformações resultaram em um vetor de atributos compacto e adequado à modelagem.

2.5. Divisão dos dados

Para evitar vazamento temporal, os dados foram divididos respeitando a ordem cronológica:

- **Treino:** lançamentos dos anos iniciais.
- **Validação:** período intermediário.
- **Teste:** lançamentos mais recentes.

Esse método simula melhor o cenário real, no qual o objetivo é prever lançamentos futuros a partir de histórico passado.

2.6. Modelos avaliados

Foram investigados três algoritmos supervisionados:

- **Régressão Logística:** modelo base para comparação.
- **Gradient Boosting:** captura relações não lineares e interações entre atributos.
- **XGBoost:** algoritmo baseado em boosting amplamente utilizado em competições e problemas tabulares.

Todos os modelos foram treinados com ajuste de hiperparâmetros e avaliados de forma consistente com a divisão temporal.

2.7. Métricas de avaliação

As métricas utilizadas para comparar os modelos foram:

- Acurácia
- Precisão
- Recall
- F1-score
- AUC-ROC

Como falsos negativos podem indicar janelas meteorológicas perigosas classificadas como favoráveis, foi dada atenção especial ao **recall da classe “Atrasado”**.

3. Resultados

Esta seção apresenta o desempenho dos modelos avaliados, bem como análises complementares que permitem compreender os fatores que influenciaram a previsão de adiamentos de lançamentos espaciais.

3.1. Desempenho dos modelos

Foram comparados três algoritmos: Régressão Logística, Gradient Boosting e XGBoost. As métricas foram calculadas no conjunto de teste, composto pelos lançamentos mais recentes, respeitando a sequência temporal.

O **Gradient Boosting** obteve o melhor desempenho geral, apresentando maior equilíbrio entre precisão, recall e F1-score para a classe *Atrasado*. Esse comportamento indica maior capacidade de capturar padrões meteorológicos associados ao risco de adiamento.

A **Régressão Logística** apresentou desempenho inferior, sobretudo pela dificuldade em modelar relações não lineares entre variáveis atmosféricas. Já o **XGBoost** apresentou desempenho próximo ao Gradient Boosting, porém com leve tendência a sobreajuste nos primeiros testes, mesmo após regularização.

3.2. Métricas de avaliação

As principais métricas observadas foram:

- **Acurácia:** valores moderados, refletindo o desbalanceamento natural entre lançamentos realizados e adiados.
 - **Precisão (classe Atrasado):** indica a proporção de alertas corretos emitidos pelo modelo.
 - **Recall (classe Atrasado):** métrica mais importante para o contexto, pois representa a capacidade de detectar condições desfavoráveis.
 - **F1-score:** equilíbrio entre precisão e recall.
- AUC-ROC:** medida agregada do desempenho discriminativo dos modelos.

O Gradient Boosting obteve o melhor **recall**, reduzindo a ocorrência de falsos negativos. Isso é relevante, pois um falso negativo corresponde a prever que o lançamento ocorrerá normalmente quando, na prática, a janela é meteorologicamente desfavorável.

3.3. Importância das variáveis

A análise de explicabilidade foi realizada utilizando valores **SHAP**, que permitem estimar o impacto de cada variável no resultado do modelo.

Os fatores meteorológicos mais influentes foram:

- **Velocidade do vento nas últimas 6 horas**
- **Rajadas máximas registradas entre 6 e 24 horas antes**
- **Precipitação acumulada nas últimas 24 horas**
- **Tendência de queda da pressão atmosférica**
- **Cobertura de nuvens em baixa altitude**

Esses resultados são consistentes com protocolos reais de lançamento, nos quais limites de vento, tempestades próximas e instabilidade atmosférica costumam determinar decisões de adiamento.

3.4. Análises adicionais

Algumas análises complementares foram realizadas para explorar o comportamento do modelo:

- **Matriz de confusão:** confirmou que a maior parte dos erros ocorre na distinção entre condições marginalmente favoráveis e desfavoráveis.
- **Curva Precision–Recall:** evidenciou bom desempenho para a classe minoritária (*Atrasado*), com vantagem do Gradient Boosting em todas as faixas de threshold.
- **Sensibilidade temporal:** janelas meteorológicas entre 6 e 24 horas antes do lançamento mostraram maior poder preditivo do que janelas muito próximas (0–6h), indicando que condições de médio prazo têm papel central nos adiamentos.

3.5. Síntese dos resultados

Os resultados mostram que:

1. Condições atmosféricas das 24 horas anteriores ao lançamento são decisivas para prever adiamentos.
2. O Gradient Boosting apresentou o melhor comprometimento entre desempenho e interpretabilidade.
3. A explicabilidade via SHAP reforça a importância operacional de vento, rajadas e precipitação.
4. A abordagem proposta pode ser estendida para centros específicos de lançamento ou adaptada para previsão contínua (em dias ou horas antes da janela).

4. Discussão

Os resultados obtidos demonstram que variáveis meteorológicas derivadas de reanálises climáticas são bons indicadores da probabilidade de adiamentos em lançamentos espaciais. A predominância de fatores como velocidade do vento, rajadas e precipitação reforça a aderência do modelo aos critérios reais utilizados por equipes de operações, que seguem limites estritos para essas variáveis durante a avaliação de go/no-go.

O desempenho superior do Gradient Boosting sugere que técnicas de aprendizado de máquina baseadas em modelos de árvore conseguem capturar relações não lineares entre fatores atmosféricos e desfechos operacionais. Além disso, a capacidade de interpretar decisões por meio de análises SHAP contribui para a confiabilidade do modelo, permitindo identificar quais condições específicas levaram à previsão de adiamento para cada lançamento.

Outro ponto relevante é que o modelo apresentou maior sensibilidade às variáveis agrupadas nas janelas entre 6 e 24 horas antes do lançamento, indicando que mudanças atmosféricas nesse intervalo têm maior valor preditivo do que condições observadas muito próximas da

tentativa de lançamento. Esse achado está alinhado à prática operacional, na qual previsões de curto prazo podem ser instáveis, enquanto tendências observadas ao longo de várias horas oferecem melhor sinal para tomada de decisão.

Apesar dos resultados consistentes, algumas limitações precisam ser consideradas. A base histórica de lançamentos possui registros heterogêneos e, em alguns casos, incompletos. Além disso, nem todo adiamento é exclusivamente causado por condições meteorológicas, o que introduz ruído no rótulo utilizado. A resolução espacial do ERA5 também pode não capturar fenômenos locais muito específicos, especialmente em áreas costeiras, onde centros de lançamento frequentemente se encontram.

Ainda assim, mesmo diante dessas limitações, os resultados indicam que modelos preditivos baseados em dados atmosféricos têm potencial para apoiar equipes de planejamento, oferecendo alertas antecipados e reduzindo o impacto operacional de adiamentos inesperados. A abordagem proposta pode ser ampliada com dados locais de alta resolução, sensores operacionais ou previsões determinísticas, fortalecendo sua aplicação em cenários reais.

5. Conclusão

Este trabalho apresentou uma abordagem de aprendizado de máquina para prever a viabilidade de janelas de lançamento espacial a partir de condições meteorológicas observadas nas horas anteriores ao evento. A integração entre dados históricos de lançamentos e variáveis atmosféricas provenientes do ERA5 permitiu construir um conjunto de atributos representativo e adequado para a modelagem preditiva.

Os resultados mostraram que o modelo baseado em Gradient Boosting apresentou o melhor desempenho geral, especialmente na identificação de condições desfavoráveis, reduzindo a ocorrência de falsos negativos. A análise de interpretabilidade com SHAP reforçou a relevância operacional de variáveis como velocidade do vento, rajadas e precipitação, elementos que compõem critérios reais de go/no-go utilizados em centros de lançamento.

Embora o estudo enfrente limitações relacionadas à heterogeneidade dos dados históricos e à resolução espacial das reanálises climáticas, os resultados obtidos indicam que a abordagem é promissora. Modelos desse tipo podem auxiliar no planejamento de operações, apoiar decisões antecipadas e reduzir custos associados a remarcações inesperadas.

Como trabalhos futuros, sugere-se a incorporação de previsões climáticas determinísticas, dados de estações meteorológicas locais e informações operacionais adicionais, além da aplicação de arquiteturas de redes neurais capazes de modelar explicitamente dependências temporais. Tais aprimoramentos podem aumentar a precisão e tornar o sistema mais aplicável a cenários reais de apoio ao lançamento.

Referências

- Copernicus Climate Change Service (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. Disponível em: <https://cds.climate.copernicus.eu/>
- Chen, T. and Guestrin, C. (2016). “XGBoost: A Scalable Tree Boosting System”. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 785–794.
- Friedman, J. H. (2001). “Greedy Function Approximation: A Gradient Boosting Machine”. *Annals of Statistics*, 29(5), 1189–1232.
- Lundberg, S. M. and Lee, S.-I. (2017). “A Unified Approach to Interpreting Model Predictions”. *Advances in Neural Information Processing Systems (NIPS)*, p. 4765–4774.
- NASA (2023). *Historical Space Launch Data Repository*. Disponível em: <https://data.nasa.gov/>
- NOAA – National Oceanic and Atmospheric Administration (2023). *Global Historical Climatology Network Daily (GHCN-D)*. Disponível em: <https://www.ncei.noaa.gov/>
- Pedregosa, F. et al. (2011). “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rasp, S. and Lerch, S. (2018). “Neural Networks for Post-processing Ensemble Weather Forecasts”. *Monthly Weather Review*, 146(11), 3885–3900.
- Space Launch Report (2023). *Global Launch Log: Historical Launch Records*. Disponível em: <https://www.spacelaunchreport.com/>