

# Detecção e Prevenção de Colisões Espaciais

**Nico Ramos**

GRR20210574

Novembro 2025



# Dataset escolhido



## Problema

- Prever o **risco final de colisão** entre um **satélite** e outro **objeto espacial** (lixo ou outro satélite).
- A base recebe uma **grande quantidade de avisos**, mas apenas uma parte **muito pequena** são de **alto risco**.



## Dados Recebidos

- Cada satélite envia periodicamente **mensagens à base** com **alertas de aproximação**, formando uma **série temporal**.
- Cada mensagem contém informações como:
  - **Data de colisão estimada**
  - **Risco estimado**
  - **Incerteza**



## Timeline de Decisão

- As **manobras de evasão** são planejadas pelo menos **2 dias antes** da data de colisão estimada.
- A **decisão final** é tomada **1 dia antes**.

# Desafio

Proposto pela **Agência Espacial Europeia (ESA)**

Treinar um **modelo** capaz de prever o **risco final** estimado pelo satélite

O **principal objetivo** é **minimizar** os **Falsos Negativos**

**Eventos de alto risco** identificados como de **baixo risco** pelo modelo

# Objetivo (inicial)



## Treinar um Modelo

Projetar e treinar um **modelo** capaz de prever o **risco estimado final**.



## Identificar Eventos

Detectar e categorizar os **eventos de alto risco**.



## Comparar Otimizadores

Avaliar o desempenho dos otimizadores **Descida de Encosta** e **Adam** em **modelos ARIMA**.

# EDA

O EDA foi o principal **desafio** enfrentado no projeto:

## Problema Complexo e Novo

A necessidade de entender simultaneamente um **problema complexo novo** e definir a melhor abordagem para o **EDA**.

## Volume de Colunas

Lidar com uma **grande quantidade de colunas**, levantando a questão: "**por onde começar?**".

# Por onde começar?

## Diminuir a Dimensionalidade

Selecionar as **colunas mais relevantes** para focar a análise inicial.

## Pesquisa e Referências

Buscar **artigos e vídeos** para entender abordagens e conceitos.

## Familiarização com o Dataset

Iniciar com a **exploração geral** para compreender a estrutura dos dados.

# Artigos de Referência

As colunas mais relevantes foram selecionadas com base nos seguintes artigos:



## **PACEcraft Collision Avoidance Challenge: Design and Results of a Machine Learning Competition**

**Ano:** 2020

**Fonte:** ArXiv

**Contribuição:** Este artigo apresenta os detalhes de um desafio de machine learning focado na prevenção de colisões de naves espaciais (PACEcraft), oferecendo insights sobre as abordagens e resultados obtidos na competição.



## **Implementation and Comparison of Data-Based Methods for Collision Avoidance in Satellite Operations**

**Ano:** 2021

**Fonte:** Journal of Spacecraft and Rockets

**Contribuição:** Explora a implementação e comparação de diferentes métodos baseados em dados para a prevenção de colisões em operações de satélites, avaliando suas eficácias e limitações.

# Divisão dos Dados

## Divisão do Conjunto

O conjunto de dados foi dividido em 80% para treino e 20% para validação.

## Proporção de Alto Risco

A mesma proporção de eventos de alto risco foi mantida em cada subconjunto .

## Super-representação

Devido à raridade dos **eventos de alto risco**, eles foram selecionados para super representar a quantidade real.

A tabela abaixo detalha a distribuição dos dados:

Conjunto	Alto Risco	Baixo/Médio Risco	Total de Eventos
Treino (Original)	2.77%	97.23%	13154
Validação	2.89%	97.11%	2630
Treino	2.75%	97.25%	10524
Teste	8.21%	91.79%	2167



# Características dos Conjuntos

Todos os eventos nos testes têm:

- A **última observação** (avaliação) a menos de um dia do **TCA** (recente)
- Todas as outras a pelo menos **2 dias** do TCA (tempo para planejar e executar a manobra)
- Os eventos no **conjunto de testes** não são **filtrados**

TCA	Última observação	Todas antes da última
Teste	< 1 dia	> 2 dias
Treino	Qualquer data	Qualquer data

# Conclusões Chave da Análise Exploratória de Dados (EDA)



## Dinâmica Temporal dos Eventos

- Eventos exibem uma **parte estável** no início.
- Forte **correlação com eventos próximos**; observações anteriores perdem importância rapidamente.
- Quando um evento muda de estado, ele tende a **permanecer estável**, sem oscilar.



## Fatores Determinantes de Risco

- Variáveis como a **incerteza da órbita** do objeto perseguidor e a **distância** são as mais correlacionadas com o risco final.
- A relevância das **incertezas aumenta** à medida que a colisão se aproxima.
- Fatores como **tamanho, arrasto, risco máximo** e **número de observações** diminuem sua importância com o tempo.

# Modelos

## Modelos Lineares para Séries Temporais

### → Arima:

- Só aceita uma variável  
(univariado)

### → Arimax:

- As séries são muito curtas e o modelo não converge
- Para prever o futuro, preciso saber o valor das outras variáveis no futuro

### → Varmax:

- Também não convergiu
- Não generalizou para outras séries

## Modelos de Séries Multivariadas e Multisséries

→ Dificuldade em unificar séries multivariadas em um modelo global de múltiplas séries

→ Os métodos encontrados atendem a apenas um dos cenários

# Objetivo (Novo)



## ~~Previsão do Valor Final~~

~~Prever o valor final do risco estimado na última observação antes da colisão~~

DESCARTADO / NÃO IMPLEMENTADO



## Classificação de Eventos

Classificar os eventos em alto/baixo nível no último evento

# Treino do Modelo

O processo de treino do modelo envolveu as seguintes etapas:

1

## 1. Pré-processamento de Séries

- **Padronização** de tamanho: Ajuste todas as séries para o tamanho da maior.
- **Padronização** de frequência entre eventos da mesma série.

2

## 2. Normalização por variância média

- Foco na forma: Ignora diferenças de amplitude e escala entre colunas.
- Comparação: Prioriza a forma dos dados.

3

## 3. Clusterização com KMeans

- Métrica **DTW**: Permite comparar séries temporais desalinhadas.
- Configuração: Máximo de 100 iterações e 3 clusters.

4

## 4. Definição das Classes de Risco

- Alto Risco: Top 5% dos riscos na última observação (acima do **threshold**).
- Baixo Risco: Todos os outros eventos classificados.
- Observação: Eventos de alto risco são intrinsecamente raros.

# Validação e Teste do Modelo de Risco

1

## Penúltima Observação

Dados utilizados para gerar as previsões e classificações de risco do modelo.

2

## Comparação

O resultado da previsão é comparado com o cenário real.

3

## Última Observação

Dados reais mais recentes do evento, usados como referência para a validação.



## Validação do Modelo

- Utilizada principalmente para ajustar o threshold de classificação, definindo o ponto de corte para identificar eventos de alto risco.
- O objetivo é otimizar a sensibilidade e especificidade do modelo, garantindo um equilíbrio entre a detecção de riscos reais e a minimização de falsos positivos.



## Teste de Desempenho

- Avaliado com o threshold de classificação previamente escolhido na fase de validação para garantir a independência.
- Tem como propósito verificar o desempenho generalizável do modelo em dados não vistos, simulando sua aplicação em um ambiente real.

# Resultados

## Análise de Risco por Cluster

Risco médio e o risco no percentil 75 para cada cluster identificado pelo modelo:

Cluster	Risco Médio	Risco 75%
0	-27.3440	-28.5429
1	-9.7131	-6.8949
2	-20.7393	-11.4110

- O **Cluster 0** foi o único classificado como **baixo risco**, apresentando os valores mais negativos de risco médio e 75%.

# Matrizes de Confusão

## Validação

Real \ Previsto	Alto Risco	Baixo/Médio Risco
Alto Risco	71	4
Baixo/Médio Risco	2117	401

## Teste

Real \ Previsto	Alto Risco	Baixo/Médio Risco
Alto Risco	172	1
Baixo/Médio Risco	1663	256



# Conclusões do Modelo de Risco



## Sobre o Modelo

O modelo prevê risco com base na **incerteza** das medidas.



## Limitações dos Métodos

Modelos de séries temporais lineares (ARIMA, etc.) **não generalizam** padrões entre séries distintas.

Modelos com múltiplos features dependem da **previsão de futuras variáveis**.

Inadequados para **séries muito curtas** ou com poucas observações.

Eficientes e confiáveis para séries com **dados suficientes**.



## Problemas Identificados

A **reamostragem** (resample) de séries temporais pode causar perda de informações.

Os **clusters** não classificaram bem os eventos, resultando em classificações extremas (todo alto ou todo baixo risco).

# Próximos Passos e Direções Futuras



## Melhorias Metodológicas

### Otimização de Clusters:

Explorar e validar a utilização de **diferentes granularidades de clusters**.



## Novas Abordagens

### Classificação & Predição Combinada:

Integrar a **classificação do risco** com a **predição do valor final** do evento.



## Refinamentos Avançados

### Predição Auxiliada por Eventos:

Refinar a **predição do valor final** utilizando informações e características de eventos classificados como similares, capitalizando sobre o aprendizado de padrões históricos.