

Relatório Modelo preditivo de interceptação de defesa aérea

Autor: Enzo Zanatta

Matrícula: 186665

Url github do projeto: <https://github.com/Enzolinn/projeto-final-data-science-2025-01>

1. Resumo do Projeto

O projeto teve como objetivos principais a criação de um processo de análise da capacidade e eficácia de uma estrutura de defesa aérea, podendo ser de um país ou região de menor escala. Para isso utilizamos o contexto da guerra russo-ucraniana. Os passos para se chegar a esse processo de análise começaram com a aplicação de um modelo matemático probabilístico para a análise da eficácia da defesa aérea ucraniana. Também o desenvolvimento de um modelo para a previsão de interceptação de um ataque iminente. E por último uma predição e análise dos valores necessários de pontos importantes para atingir uma defesa mais bem sucedida. Um ponto importante a ser destacado é que por mais que são usados dados de ataques reais, as especificações das defesas aéreas ucranianas como local de posicionamento, estoque de munição e prioridade de alvos são dados confidenciais que não puderam ser acessados portanto foi utilizado generalizações para tentar chegar em valores coerentes.

2. Dataset Utilizado

- <https://www.kaggle.com/datasets/piterfm/massive-missile-attacks-on-ukraine>
- **Logs de Ataques:** O arquivo `missile_attacks_daily.csv`, uma compilação de dados públicos (OSINT) sobre ataques aéreos russos.
- **Especificações de Armas:** O arquivo `missiles_and_uav.csv` e um inventário visual da defesa aérea ucraniana, detalhando as características técnicas das armas e sistemas envolvidos.

2.1. Variáveis e Transformações

Pré-processamento dos dados:

- **Limpeza de Dados:** Conversão de colunas para tipos de dados corretos (numérico, datetime) e tratamento de valores ausentes.
- **Engenharia de Features:**
 - **P(track) Ponderado:** A Probabilidade de Rastreamento ($P(\text{track})$) foi estimada em **0.90**, calculada como uma média ponderada baseada no inventário de sistemas de defesa, atribuindo maior peso aos sistemas da era soviética, mais numerosos.
 - **Padronização Geográfica:** A coluna de alvos (`target`) foi limpa e padronizada. Nomes de locais com variações (ex: "kyivska oblast", "kyiv region") foram

consolidados em um único padrão (ex: "kyiv") usando um dicionário de mapeamento, o que foi crucial para a precisão da análise geográfica e do modelo de ML.

- **Criação da Variável Alvo:** A variável `interception_rate` foi calculada (`destroyed / launched`) para servir como o alvo de previsão do nosso modelo de Machine Learning.

3. Modelos

Dois modelos foram centrais para este projeto: um modelo probabilístico do paper "A Simple Model for Calculating Ballistic Missile Defense Effectiveness" de Dean A. Wilkening, que forneceu o framework matemático. E foi desenvolvido um modelo preditivo.

3.1. Modelo Probabilístico

Utilizamos a fórmula de Wilkening de forma inversa para diagnosticar a eficácia atual dos sistemas de defesa.

- **Fórmula Base:** $K_w = P(\text{track}) \times (1 - (1 - \text{SSPK})^n)$
- **Aplicação:** Ao calcular a eficácia total histórica (K_w) a partir dos dados e definir as premissas de $P(\text{track})$ e n (número de disparos), nós isolamos e calculamos o **SSPK Implícito** (Single-Shot Kill Probability) para cada categoria de arma.

3.2. Modelo Preditivo

- **Algoritmo:** `GradientBoostingRegressor` do Scikit-learn, por ele ser bom para a captura de relações complexas e interações entre variáveis, além de ser um modelo de *ensemble*, então ele não cria apenas uma árvore de decisão, mas centenas delas em sequência, onde cada nova árvore tenta corrigir os erros da anterior. Esse processo de "aprendizado gradual" resulta em um modelo final extremamente preciso e robusto. Ele também é muito flexível com relações não-lineares.
- **Objetivo:** Prever a `interception_rate` de um ataque.
- **Features Finais:** O modelo foi treinado com as seguintes variáveis:
 - `launched` (tamanho do ataque)
 - `year`, `month` (fatores temporais)
 - `model`, `weapon_category` (características da arma)
 - `target_standardized` (local do alvo)

4. Resultados e Conclusões

4.1. Análise de Eficácia (SSPK)

- **Hierarquia de Ameaças:** A análise quantificou a dificuldade de interceptar cada

ameaça, resultando em um SSPK implícito de **~9%** para mísseis balísticos, **~46%** para mísseis de cruzeiro e **~64%** para drones UAV.

- **Análise Geográfica:** Foi comprovado que a eficácia da defesa não é uniforme, com regiões como Kyiv apresentando taxas de interceptação significativamente mais altas do que outras, especialmente aquelas mais próximas da fronteira.

4.2. Análise de Cenários

- **Limites da Defesa:** As simulações para atingir um objetivo de defesa de 90% contra mísseis balísticos mostraram que a baixa eficácia atual (SSPK de 9%) exigiria um número impraticável de interceptadores. Para ter sucesso com uma doutrina de 2 disparos por alvo, o SSPK precisaria saltar de 9% para mais de 56%.
- **Insight Estratégico:** A análise demonstrou que o fator limitante de uma defesa quase perfeita é a capacidade de rastreamento ($P(\text{track})$). Nenhum aumento na quantidade ou qualidade dos interceptadores pode compensar a falha em rastrear um alvo de forma confiável.

4.3. Conclusão Final

Este projeto demonstrou a possibilidade de se construir um modelo preditivo com impactos reais no cenário de defesa nacional.