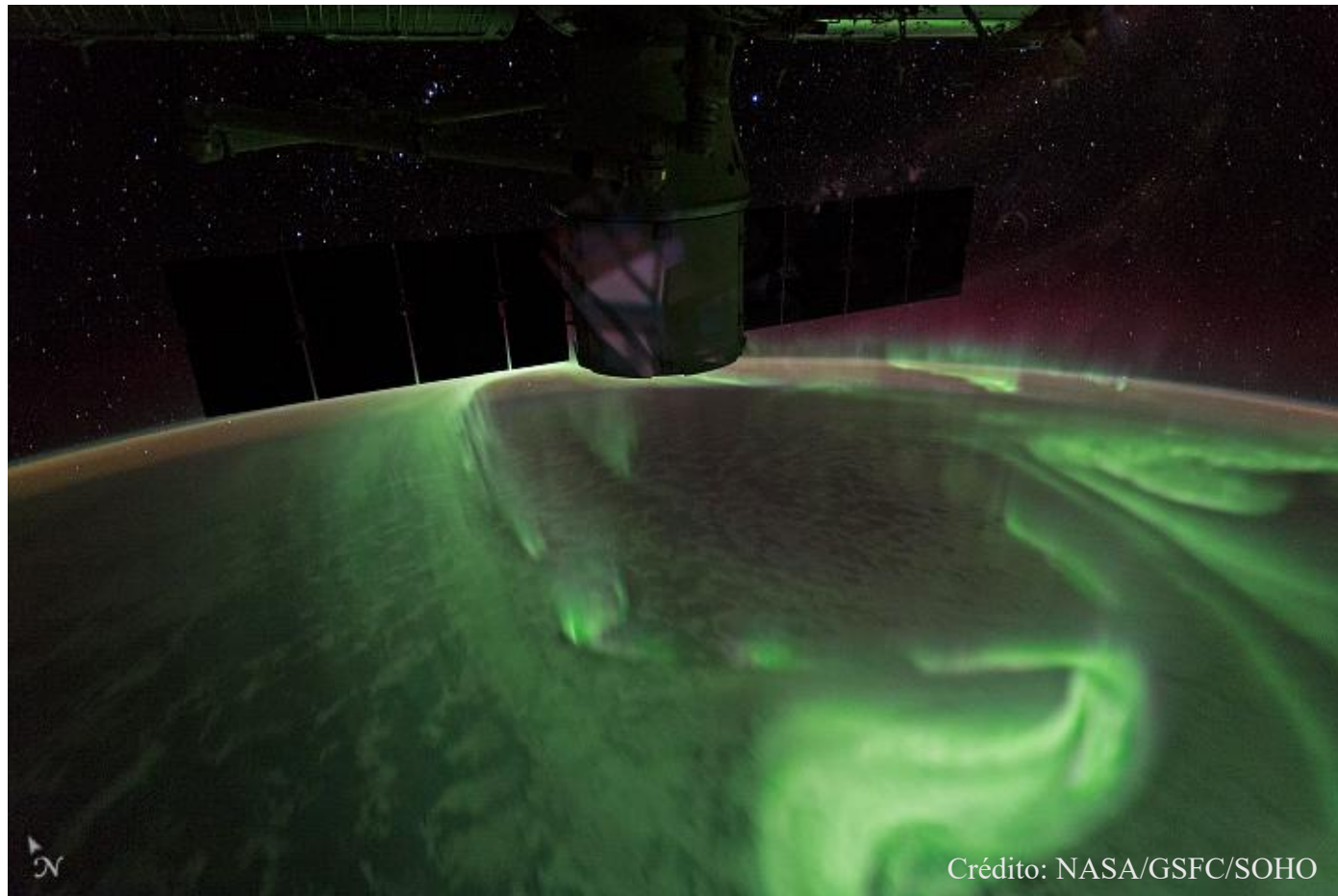


Comparação de algoritmos de aprendizado de máquina para previsão de clima espacial: erupções solares e tempestades geomagnéticas



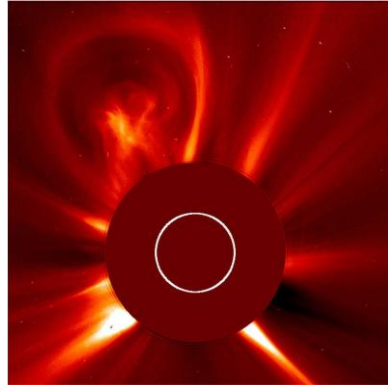
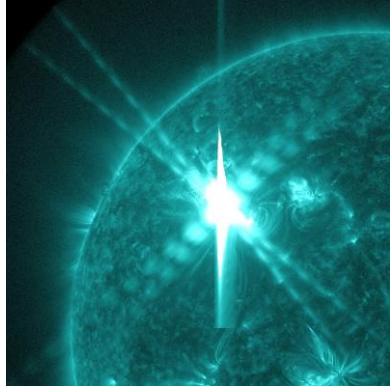
Fernando de S. Mello¹

¹Instituto Hardware BR

fdesmello@gmail.com

Crédito: NASA/GSFC/SOHO

O clima espacial e seus efeitos aeroespaciais

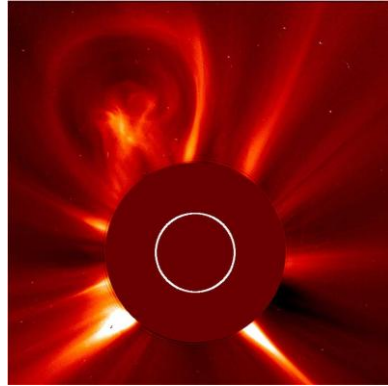
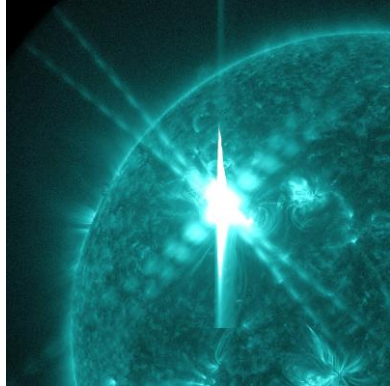


O Sol é o culpado

- Erupções solares
- Ejeções de massa coronal
- Tempestades geomagnéticas

- Disrupções de comunicações por rádio;
- Alta dose de radiação para astronautas;
- Rede elétrica sobrecarregada;
- Sistemas de localização prejudicados;
- Satélites fora de curso;

O objetivo



O Sol é o culpado

- Erupções solares
- Ejeções de massa coronal
- Tempestades geomagnéticas

Predizer:

- Erupções solares;
 - Tempestades geomagnéticas;
- ...para ter tempo para se preparar.

Os dados

Datasets

Erupções solares

(33 colunas, 4480 linhas)

- SHARP
- GOES
- Kaggle

Tempestades geomagnéticas

(19 colunas 244.535 linhas)

- OMNI

Dados

- Campo magnético do vento solar;
- Campo magnético das regiões complexas;
- Fluxo de raios X;
- Parâmetros do plasma;
- ...

Problemas... ou desafios

- Eventos incomuns;
- Datasets desbalanceados;
- Estrutura temporal dos eventos;
- Risco de vazamento de dados;

Métodos



Alvos:

- Classe de erupção
Flare: C/M/X
- Índice Dst

Características atrasadas
(RF, CVM, NN)

X

Sequência
(RNN-GRU)

Soluções para datasets desbalanceados:

- Pesar características;
- Respeitar temporalidade;
- Usar métricas adequadas;

Quatro algoritmos:

- Floresta aleatória (RF)
- Máquina de vetor de suporte (SVM)
- Rede neural (NN)
- Rede neural recorrente GRU (RNN-GRU)

Por quê?

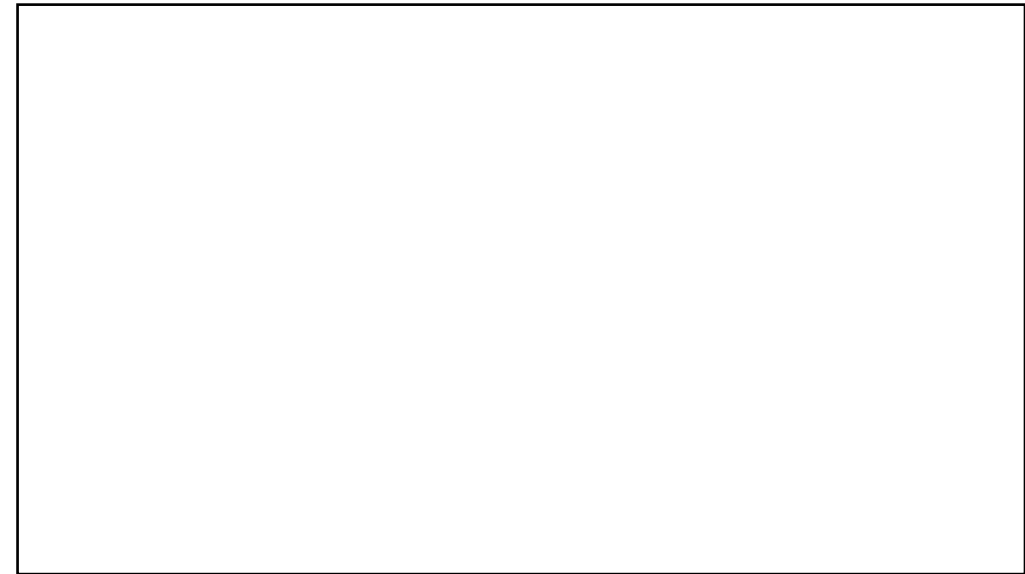
- Testar diferentes complexidades, arquiteturas e habilidades;
- Comparar algoritmos com mesmo objetivo e dados;
- Não necessariamente o mais complexo seria o melhor;
- “Assim no projeto como no curso.”

Avaliação

Métricas

- **Revocação**
(recall)
- **F₂ score**
- **F₁ score**
- **Precisão média**
(*average precision*)
- Área sob a curva característica de operação do receptor
(*area under the receiver operating characteristic curve*,
ROC-AUC)
- Avaliar matrizes de confusão

Predizer positivos verdadeiros é mais importante do que evitar falsos positivos.
...até certo ponto.



Resultados

Erupções solares

	F ₂	F ₁	Revocação	Precisão média	ROC-AUC
RF	0,518	0,452	0,574	0,448	0,825
SVM	0,614	0,476	0,762	0,434	0,824
NN	0,555	0,435	0,680	0,422	0,820
RNN-GRU	0,636	0,472	0,828	0,537	0,854

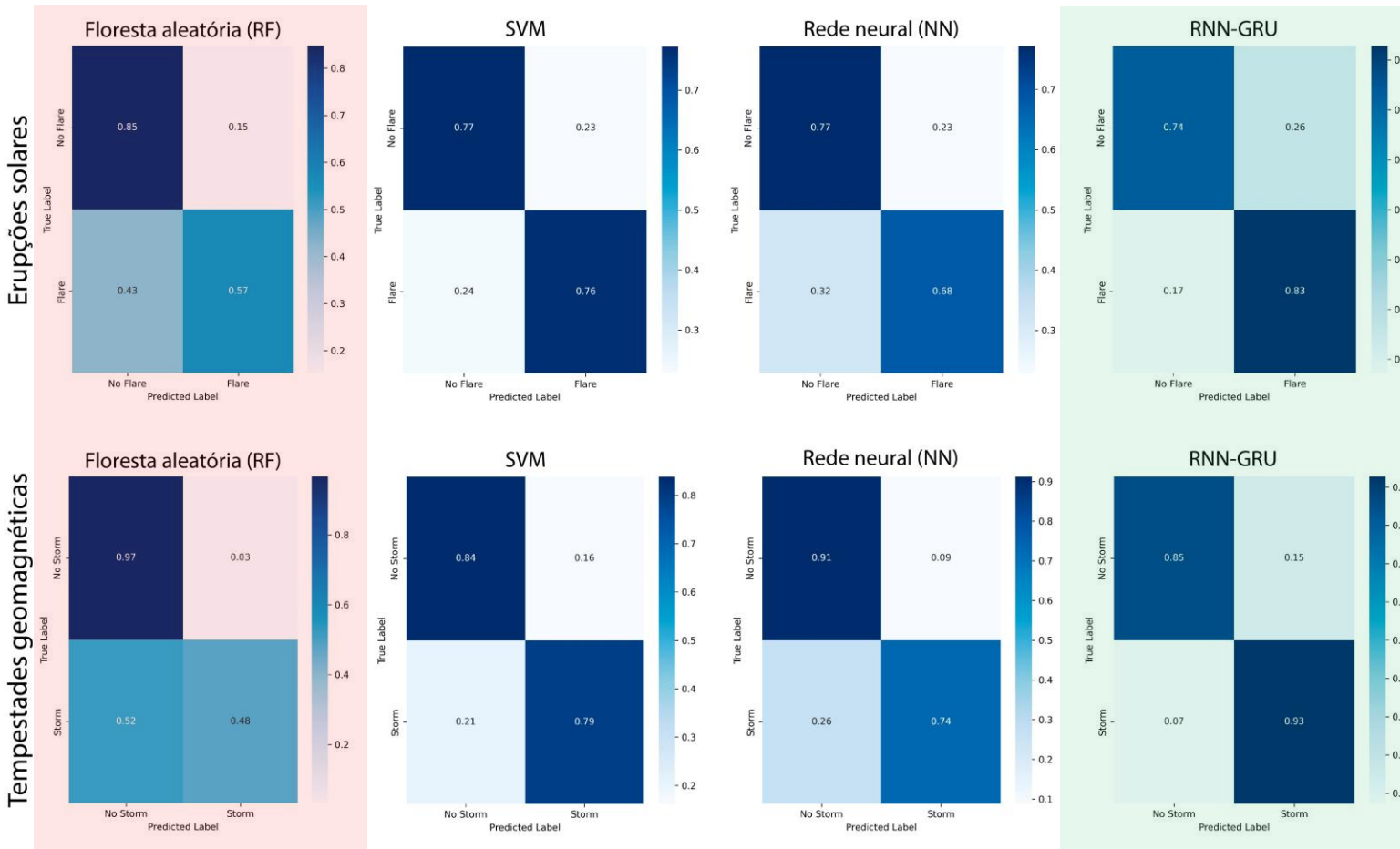
- **Floresta aleatória:**
A pior, mas a mais rápida.
(minutos)
- **RNN-GRU:**
A melhor, mas a mais lenta.
(horas)

Tempestades geomagnéticas

	F ₂	F ₁	Revocação	Precisão média	ROC-AUC
RF	0,449	0,411	0,478	0,410	0,917
SVM	0,429	0,254	0,792	0,355	0,895
NN	0,450	0,284	0,737	0,477	0,930
RNN-GRU	0,436	0,243	0,928	0,511	0,959

Resultados

Matrizes de confusão



- **Floresta aleatória:**
A pior, mas a mais rápida.
(minutos)
- **RNN-GRU:**
A melhor, mas a mais lenta.
(horas)

Limitações e perspectivas

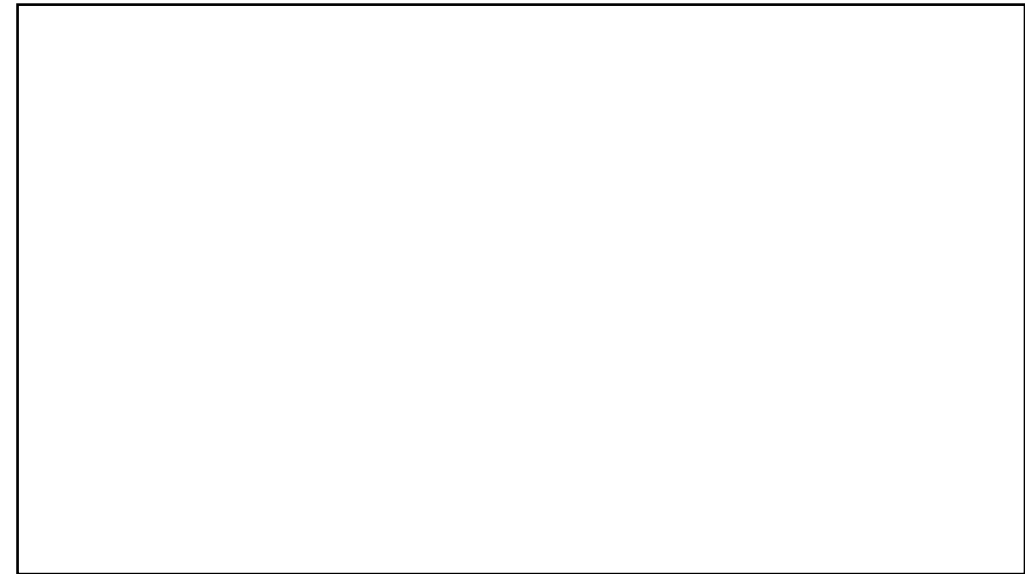
- **Característica**

- Mais características;
- Melhores características;

- **Alvos**

- Classificação multi-classe de erupções solares;
- Regressão linear em Dst para previsão de tempestades geomagnéticas;

- **Previsão em tempo real**



Fim

Obrigado!

github.com/ferdesmello/hbr-solar-forecast-project