

Predizendo o clima espacial com aprendizado de máquina: erupções solares e tempestades geomagnéticas

Fernando de S Mello¹

¹Instituto Hardware BR

fdesmello@gmail.com

Abstract.

Resumo. o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.

1. Introdução

O clima espacial (também tempo espacial, *space weather*) refere-se às variações no ambiente espacial no Sistema Solar, mas com focal especial na interação entre o Sol e a Terra. (Denardini et al., 2016) O principal motor e agente do clima espacial é o Sol, principalmente por eventos de erupções solares, ejeções de massa coronal, vento solar, e tempestades geomagnéticas. A frequência desses eventos não é constante. O Sol tem estações de atividade seguindo um ciclo de aproximadamente 11 anos. Durante esse círculo, é possível observar um lento aumento na quantidade de manchas solares, anomalias magnéticas e na frequência de erupções solares e ejeções de massa coronal; para então decair lentamente de novo.

Regiões ativas (*active regions*) são características temporárias na atmosfera solar que apresentam um forte e complexo campo magnético. Essas regiões são frequentemente observadas como manchas escuras (manchas solares ou *sunspots*), regiões mais frias na superfície solar que aparecem ser mais escuras do que a região circundante, durando de dias a meses. Essas regiões estão frequentemente associadas com grandes e súbitas emissões de radiação eletromagnética (erupções solares ou *solar flares*), especialmente de radiação ultravioleta e X, durando da ordem de alguns minutos. Associadas às erupções solares há frequentemente a ejeção de grandes quantidades de plasma da corona solar (ejeção de massa coronal ou *coronal mass ejection – CME*). Esse plasma, quando cruza a órbita terrestre e interage com a magnetosfera de nosso planeta, provoca as auroras.

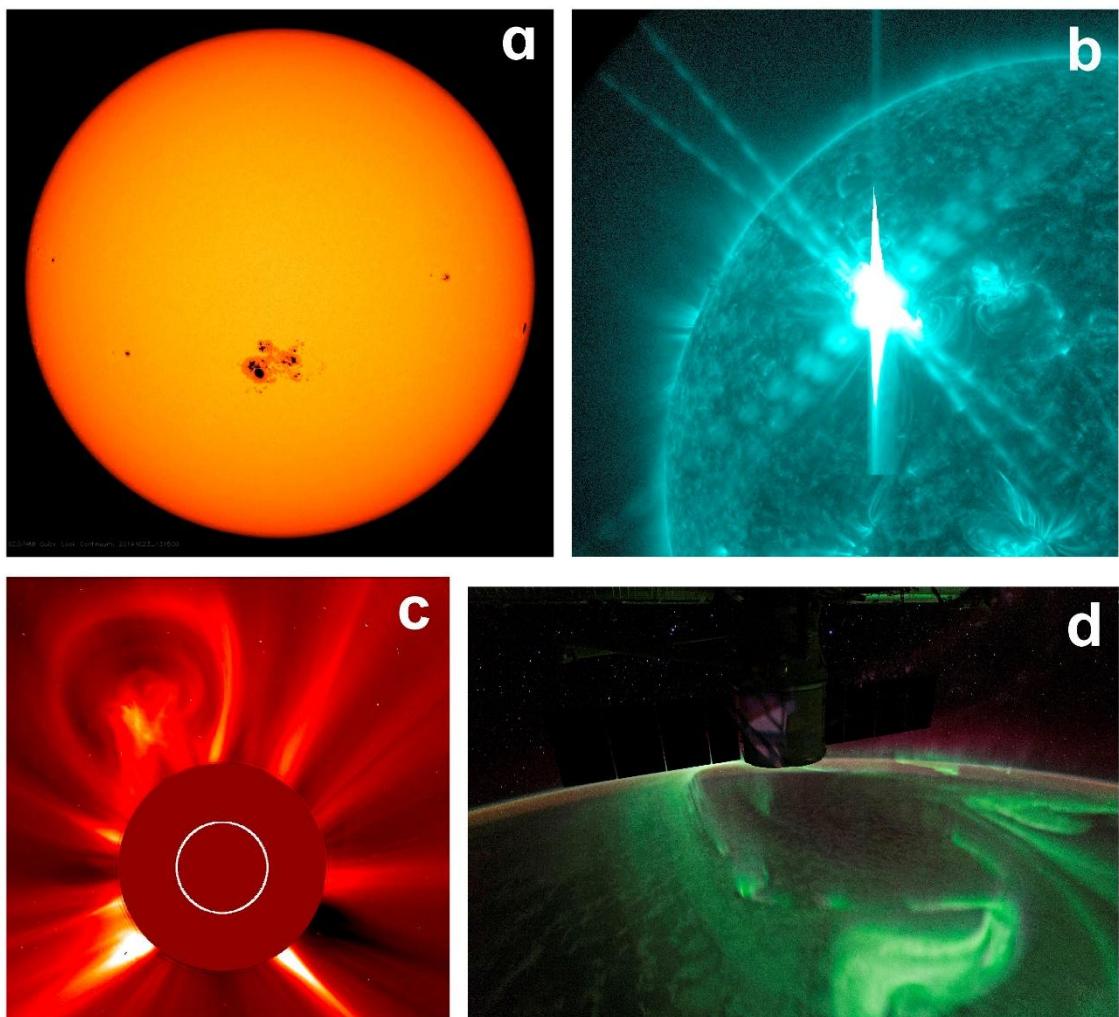


Figura 1. Eventos solares e terrestre relacionados com o tempo espacial. a) Um grande conjunto de manchas solares. Crédito: NASA/SDO/AIA. b) Uma grande erupção solar vista em raios-x. Crédito: NASA/SDO. c) Uma grande ejeção de massa coronal com coronógrafo obstruindo a vista do Sol. Crédito: NASA/GSFC/SOHO. d) Aurora austral vista da Estação Espacial Internacional. Crédito: NASA.

Mas as auroras não são os únicos efeitos do clima espacial na Terra. A grande quantidade de radiação ionizante emitida durante uma erupção solar aumenta a ionização e temperatura da ionosfera terrestre, o que pode interferir com comunicação por rádio, sistemas de localização global, e aumentar o arrasto sentido por satélites na órbita terrestre baixa, pois a atmosfera superior se expande ao ser aquecida.

Ejeções de massa coronal, por sua vez, podem perturbar o campo magnético terrestre e provocar tempestades geomagnéticas (*geomagnetic storms*). As variações magnéticas e o grande fluxo de partículas carregadas podem afetar a comunicação e funcionamento de satélites, causar picos nos sistemas de rede elétrica, podendo levar até a apagões, além de expor astronautas em órbita a maiores doses de radiação.

Um exemplo recente de um evento intenso foi a tempestade solar de 1989, que causou interferência na comunicação por rádio, apagões no Canadá e o avistamento de auroras muito mais ao Sul do que de costume. No entanto, a maior tempestade solar já registrada

foi a tempestade solar de 1859, também conhecida como Evento Carrington. Seu impacto prático se restringiu mais aos sistemas de telégrafo da época. Mas se uma tempestade semelhante acontecesse hoje, com muitos mais sistemas elétricos e eletrônicos do que no século XIX, o potencial destrutivo seria bem maior.

De maneira geral, esses eventos extremos do clima espacial podem ter consequências econômicas significativas e o potencial de afetar negativamente diversos setores da civilização humana. Daí a importância da previsão desses eventos para nos dar mais tempo de reação e mitigar seus efeitos potencialmente até catastróficos.

Ao longo dos anos, vários modelos físicos foram usados para prever os diferentes fenômenos do clima espacial. Mais recentemente, métodos de aprendizado de máquina (*machine learning*) também vêm sendo empregados com bons resultados devido ao aumento no poder de processamento de computadores e na grande quantidade de dados observacionais.

Este trabalho vem nessa esteira ao desenvolver modelos de previsão para erupções solares e tempestades geomagnéticas. Isso é feito no contexto do curso Residência em Tecnologias Aeroespaciais - Inteligência Artificial, aplicando diferentes modelos de aprendizado de máquina estudados e comparando seu desempenho com diferentes métricas. Restrições de tempo, poder de processamento e conhecimento limitaram o escopo do trabalho, mas sugestões realistas e operacionalizáveis de melhoramentos para os modelos são discutidas mais ao fim.

2. Abordagem, metodologia (software, dataset, ...)

Na literatura, há trabalhos que tentam fazer a previsão das erupções e tempestades usando desde técnicas simples de floresta aleatória (*random forest*, RF) ou vetor de suporte de máquina (*support vector machine*, SVM), até redes neurais (*neural nets*, NN) e redes neurais profundas (*deep neural networks*) e obtêm resultados satisfatórios. Por isso, eu decidi fazer um panorama e usar diferentes técnicas abordadas no curso e comparar seus resultados por diferentes métricas. Não necessariamente a técnica mais complexa e demorada traria os melhores resultados. Muito da qualidade dos resultados também residiria na qualidade dos dados e nas características (*features*) utilizadas.

Erupções solares e tempestades geomagnéticas são fenômenos infreqüentes e de curta duração. Isso cria datasets bastante desbalanceados. A natureza temporal dos dados exige cuidado especial tanto em sua ordenação quanto em evitar o vazamento de dados (*data leakage*). São fenômenos com grande potencial de danos, e os eventos mais energéticos são também os mais raros, isso torna importante a previsão do maior número possível de eventos mesmo que ao custo de alguns falsos positivos.

O desbalanceamento foi inicialmente tratado com a criação de dados sintéticos baseados na classe minoritária usando SMOTE (*Synthetic Minority Over-sampling Technique*). No entanto, devido ao maior tempo de processamento ao utilizar a técnica e o temor da possibilidade de os dados sintéticos distorcerem a série histórica, foi escolhido o método de pesos, dando maior peso aos casos de maneira inversa à sua frequência nos datasets.

Eu tratei a temporalidade não separando o dataset em treino, validação e teste de maneira aleatória, mas em fatias contínuas. E dado o desbalance nos dados e a necessidade de prever a classe minoritária, as métricas utilizadas teriam de focar em diminuir falsos negativos, nesse caso, F2, F1, Precision, Recall, AUC. Caso contrário, a previsão

constante da classe majoritário (poderiam nunca predizer evento algum) apresentaria boas métricas que derrotariam o propósito do projeto.

Como os eventos não estão distribuídos homogeneamente devido ao ciclo solar, eu tive que escolher um intervalo de datas para o dataset que contivesse

Dado o desbalance nos dados a necessidade de prever a classe minoritário, as métricas utilizadas teriam de focar em diminuir falsos negativos, nesse caso, F₂ Score (F₂), F₁ score (F₁), precisão, revocação, AUC. Caso contrário, a predição constante da classe majoritário (poderiam nunca predizer evento algum) apresentaria boas métricas que derrotariam o propósito do projeto.

2.1. Datasets

The subsection titles must be in boldface, 12pt, flush left.

2.2. Tratamento dos datasets

The subsection titles must be in boldface, 12pt, flush left.

2.3. Erupções Solares

Erupções solares são classificadas segundo o máximo de emissão de raios X no comprimento de onda de 0,1 a 0,8 nm, contendo as classes A, B, C, M e X (Messerotti et al., 2009). As de classe M e X são as mais fortes, emitindo grandes quantidades de raios X, UV e rádio.

Tabela 1. Informações do dataset para erupções solares: data_flares.csv. 33 colunas e 4480 linhas, com medidas por dia, de maio de 2010 a janeiro de 2023. Fluxos e campos quando não especificados são magnéticos e corrente é corrente elétrica.

Característica	Unidade	Descrição
DATE		Data
USFLUX	Mx	Total de fluxo magnético
MEANGAM	°	Ângulo médio de campo em relação ao radial
MEANGBT	G Mm ⁻¹	Gradiente horizontal do campo total
MEANGBZ	G Mm ⁻¹	Gradiente horizontal do campo vertical
MEANGBH	G Mm ⁻¹	Gradiente horizontal de campo horizontal
MEANJZD	mA m ⁻¹	Densidade de corrente vertical
TOTUSJZ	A	Total de corrente vertical
MEANALP	Mm ⁻¹	Parâmetro de torção característico, α
MEANJZH	G ² m ⁻¹	Helicidade da corrente (contribuição de Bz)

TOTUSJH	$\text{G}^2 \text{ m}^{-1}$	Helicidade total da corrente
ABSNJZH	$\text{G}^2 \text{ m}^{-1}$	Valor absoluto da helicidade de corrente líquida
SAVNCPP	A	Soma do módulo da corrente líquida por polaridade
MEANPOT	erg cm^{-3}	Indicador da densidade média de excesso de energia magnética na fotosfera
TOTPOT	erg cm^{-3}	Indicador da densidade total de energia livre magnética da fotosfera
MEANSHR	$^\circ$	Ângulo de cisalhamento
SHRGHT45		Área fracionária com cisalhamento $>45^\circ$
R_VALUE	Mx	Fluxo magnético próximo às linhas de inversão de polaridade
xrs_A_mean	W m^{-2}	Média de fluxo de raios-x de curto comprimento de onda
xrs_A_min	W m^{-2}	Mínimo de fluxo de raios-x de curto comprimento de onda
xrs_A_max	W m^{-2}	Máximo de fluxo de raios-x de curto comprimento de onda
xrs_B_mean	W m^{-2}	Média de fluxo de raios-x de longo comprimento de onda
xrs_B_min	W m^{-2}	Mínimo de fluxo de raios-x de longo comprimento de onda
xrs_B_max	W m^{-2}	Máximo de fluxo de raios-x de longo comprimento de onda
Radio Flux 10.7cm	$\text{s.f.u. (10}^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}\text{)}$	Fluxo de rádio em comprimento de onda de 10,7 cm
Sunspot Number		Número de manchas solares
Sunspot Area	MSH (milionésimo de hemisfério solar)	Área somada das manchas solares
New Regions		Número de novas manchas solares
Flares: C		Número de erupções classe C
Flares: M		Número de erupções classe M
Flares: X		Número de erupções classe X
flare_today		Flag para erupções
flare_missing		Flag para algum dado faltando na fonte

2.3. Tempestades Geomagnéticas

Por sua vez, as tempestades geomagnéticas são classificadas segundo o índice de distúrbio-tempo de tempestade (disturbance storm-time index, Dst), um índice de atividade magnética derivado de uma rede de observatórios geomagnéticos na Terra. Quanto maior a diferença entre o observado e o valor típico, maior a interferência magnética pela tempestade geomagnética. Valores de Dst inferiores a -50 nT indicam

uma tempestade geomagnética moderada, quanto valores menores do que -250 nT indicam uma supertempestade.

Tabela 2. Informações do dataset para tempestades geomagnéticas: data_storms.csv. 19 colunas e 244535 linhas, com medidas por hora, de janeiro de 1995 a dezembro de 2024. Componentes e campos são magnéticos. Quando não especificado, outras grandezas são referentes ao plasma.

Característica	Unidade	Descrição
datetime		Data
ABS_B	nT	Magnitude do vetor de campo médio, $ $
F	nT	Magnitude média do campo, $<F>$
BX_GSE	nT	Componente X do campo magnético interplanetário usando o sistema de coordenadas GSE
BY_GSE	nT	Componente Y do campo magnético interplanetário usando o sistema de coordenadas GSE
BZ_GSE	nT	Componente Z do campo magnético interplanetário usando o sistema de coordenadas GSE
SIGMA-ABS_B	nT	Desvio Padrão RMS na magnitude média
SIGMA-B	nT	Desvio Padrão RMS no vetor de campo
SIGMA-Bx	nT	Desvio Padrão RMS da componente X média no GSE
SIGMA-By	nT	Desvio Padrão RMS da componente Y média no GSE
SIGMA-Bz	nT	Desvio Padrão RMS da componente Z média no GSE
T	K	Temperatura
N	cm^{-3}	Densidade
V	km s^{-1}	Velocidade
Ratio		Razão entre partículas alfa e prótons
Pressure	nPa	Pressão do fluxo
R		Número de novas manchas solares por dia
DST	nT	Índice Dst
storm_now		Flag de hora com tempestade

3. Resultados

Ghkghkghkhkjjkjk

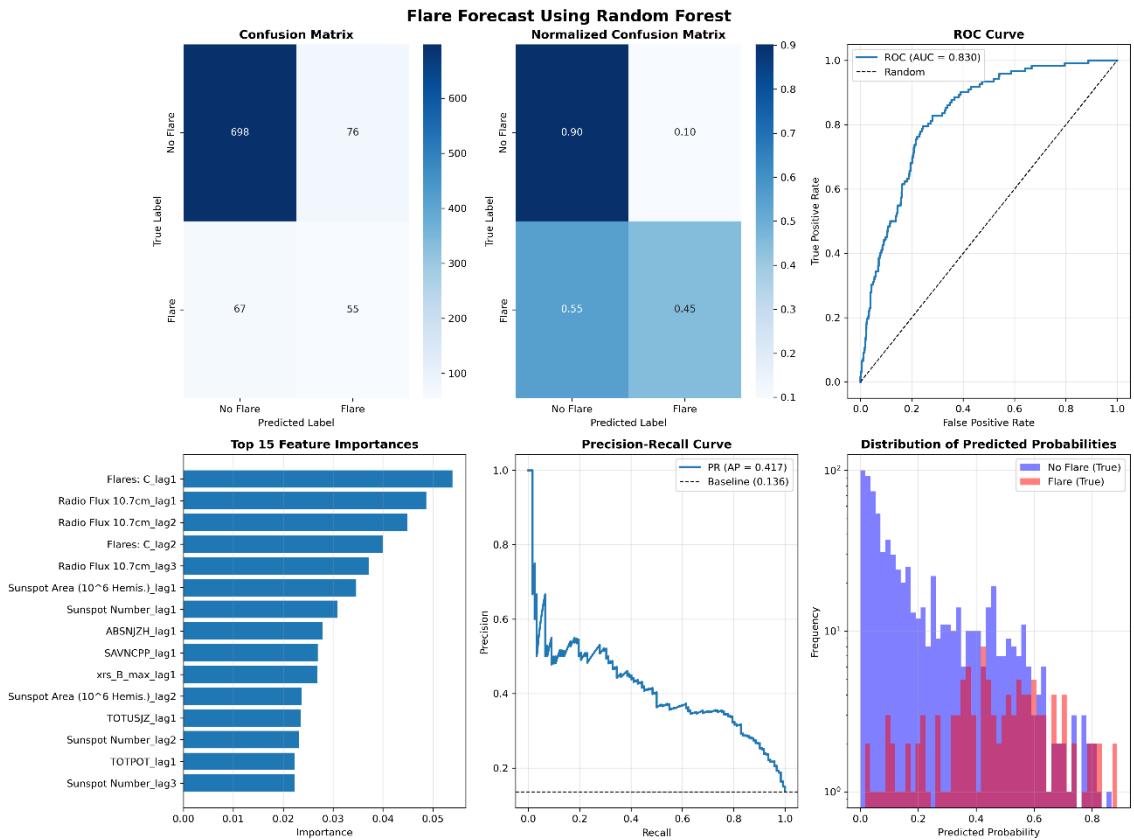


Figure 2. A typical figure

4. Análise e discussão dos resultados

Section titles must be in boldface, 13pt, flush left. There should be an extra 12 pt of space before each title. Section numbering is optional. The first paragraph of each section should not be indented, while the first lines of subsequent paragraphs should be indented by 1.27 cm.

4.1. Subsections

The subsection titles must be in boldface, 12pt, flush left.

5. Conclusão

Figure and table captions should be centered if less than one line (Figure 1), otherwise justified and indented by 0.8cm on both margins, as shown in Figure 2. The caption font must be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.



*"No, you weren't downloaded.
You were born."*

Figure 3. A typical figure

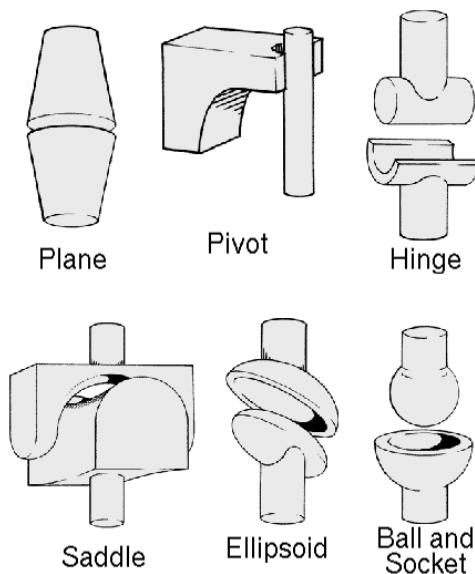


Figure 2. This figure is an example of a figure caption taking more than one line and justified considering margins mentioned in Section 5.

In tables, try to avoid the use of colored or shaded backgrounds, and avoid thick, doubled, or unnecessary framing lines. When reporting empirical data, do not use more decimal digits than warranted by their precision and reproducibility. Table caption must be placed before the table (see Table 1) and the font used must also be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.

6. Images

All images and illustrations should be in black-and-white, or gray tones, excepting for the papers that will be electronically available (on CD-ROMs, internet, etc.). The image resolution on paper should be about 600 dpi for black-and-white images, and 150-300 dpi for grayscale images. Do not include images with excessive resolution, as they may take hours to print, without any visible difference in the result.

7. References

Bibliographic references must be unambiguous and uniform. We recommend giving the author names references in brackets, e.g. [Knuth 1984], [Boulic and Renault 1991]; or dates in parentheses, e.g. Knuth (1984), Smith and Jones (1999).

The references must be listed using 12 point font size, with 6 points of space before each reference. The first line of each reference should not be indented, while the subsequent should be indented by 0.5 cm.

Referências

Boulic, R. and Renault, O. (1991) “3D Hierarchies for Animation”, In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons Ltd., England.

Dyer, S., Martin, J. and Zulauf, J. (1995) “Motion Capture White Paper”, http://reality.sgi.com/employees/jam_sb/mocap/MoCapWP_v2.0.html, December.

Holton, M. and Alexander, S. (1995) “Soft Cellular Modeling: A Technique for the Simulation of Non-rigid Materials”, Computer Graphics: Developments in Virtual Environments, R. A. Earnshaw and J. A. Vince, England, Academic Press Ltd., p. 449-460.

Knuth, D. E. (1984), The TeXbook, Addison Wesley, 15th edition.

Smith, A. and Jones, B. (1999). On the complexity of computing. In *Advances in Computer Science*, pages 555–566. Publishing Press.