



UM MÉTODO BASEADO EM APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE ASTEROIDES

Tiago C. BATISTA¹; Daiane C. GARCIA²; Diego SAQUI³

RESUMO

O presente artigo apresenta um modelo obtido através da aplicação de Aprendizado de Máquina, para auxiliar na classificação de asteroides contidos em uma base de dados. Para chegar aos resultados alcançados, foram utilizados três diferentes tipos de algoritmos de Aprendizado Supervisionado, *Support Vector Machine - RBF Kernel (SVM-RBF)*, *Naive Bayes* e *K-Nearest Neighbors*. Além disso, também foram utilizadas quatro métricas de avaliação, *Accuracy*, *Precision*, *Recall* e *F1-Score*. Foram utilizados 33% dos dados para teste. Conclui-se que o modelo mais eficiente para a classificação de asteroides originou-se a partir do algoritmo *SVM-RBF* e com isso foi realizado a integração de tal modelo à uma aplicação web.

Palavras-chave:

Inteligência Artificial; Astronomia; Dados; Tecnologia; Ciência.

1. INTRODUÇÃO

Os estudos pelos corpos celestes presentes no universo vem se tornando cada vez mais presentes em contextos acadêmicos. Muito se vale pelo seu grande potencial de descoberta, já que a vasta dimensão dos Cosmos se acarreta ao impossível de se quantificar. De certo modo, o entendimento de como o universo funciona com todos seus aspectos extraordinários afeta o conhecimento sobre nós mesmos, humanos, já que vivemos em um desses corpos celestes citados anteriormente dentre os milhares existentes no universo observável: a Terra. A ciência, com todos os pontos destacados, visa a obtenção de auxílios para o conhecimento de tais corpos celestes presentes em nosso Sistema Solar e existem outros tipos dos mesmos que viajam, tangenciando o espaço em que nos encontramos.

Nos últimos anos, tem havido um crescimento significativo na lista de asteroides numerados no Sistema Solar, o que torna essencial o desenvolvimento de métodos eficientes para lidar com o grande volume de dados observacionais e aprimorar as soluções orbitais desses corpos celestes. Com o advento de novas naves espaciais e telescópios, espera-se que um enorme volume de dados adicional seja disponibilizado, exigindo um adequado processamento e análise (SMIRNOV e MARKOV, 2022). Asteroides caracterizam-se por pequenos objetos em deslocamento pelo espaço. Milhares dos mesmos entram na órbita da Terra anualmente, além também de outros corpos celestes como meteoros e cometas, porém somente uma parcela de quantidade é observável, pelo motivo de

¹ Bacharelado em Ciência da Computação, IFSULDEMINAS - *Campus* Muzambinho. E-mail: tiagod07@gmail.com

² Bacharelada em Ciência da Computação, IFSULDEMINAS - *Campus* Muzambinho. E-mail: daianemuz4@gmail.com

³ Orientador, IFSULDEMINAS - *Campus* Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br

se evaporar em pequenas partículas antes de aproximar-se da atmosfera terrestre (PRASAD *et al.*, 2020). Acontecimentos históricos inclusive, aconteceram justamente por conta de ações oriundas dos asteroides e seus comportamentos, como a extinção dos dinossauros, sendo o impacto do asteroide à Terra uma das razões para o acontecimento destacado (PRASAD *et al.*, 2020).

O Aprendizado de Máquina (AM) é uma subárea de Inteligência Artificial (IA), que de maneira simples a ser relatada, permite que computadores e/ou dispositivos “aprendam” sem a interferência direta de programações (GARCIA-CANADILLA, 2020). Os algoritmos de AM adquirem “experiência” ao melhorar seu desempenho em tarefas designadas e a “experiência” destacada, sugere algo como adaptação aos dados (BI *et al.*, 2019).

A partir de todos os fatos levantados e toda a importância dos estudos dos astros, mais especificamente se tratando de asteroide, neste trabalho é proposto um modelo de AM para classificação dos asteroides. Para isso, foram considerados três algoritmos e mais 4 métodos de avaliação, de medir sua eficácia para a escolha do melhor adequado ao cenário atual.

3. MATERIAL E MÉTODOS

Para auxílio no desenvolvimento dos algoritmos escolhidos, a linguagem de programação *Python* foi utilizada devido ao seu fácil entendimento ligada à execução. Em relação ao ambiente de execução, foi usado o *Google Collaboratory*, plataforma do *Google* para desenvolvimentos específicos com a linguagem *Python*. Os algoritmos escolhidos para o método de classificação se utilizaram da biblioteca *sklearn* e que se seguem: *Support Vector Machine - RBF Kernel* (SVM-RBF), *Naive Bayes* e *K-Nearest Neighbors* (KNN). Para a normalização dos dados, foi utilizado uma classe da *sklearn*, o *StandardScaler*. Em relação às métricas de avaliação utilizadas, incluem-se *Accuracy*, *Precision*, *Recall* e *F1-Score*.

A base de dados utilizada foi captada do site de catálogo de *datasets* diversos *Kaggle*, denominada “*Asteroid Impacts*” (JOSHI, 2021), contando com 15.634 dados de asteroides compostos por 16 colunas (ou *features*). Para o tratamento dos dados, previamente à aplicação direta dos algoritmos de AM, se dispôs nos seguintes passos:

- 1) Foram retirados dados nulos presentes na base de dados;
- 2) Foi retirada a coluna *Object Name*, visto que a mesma não possui utilidade para a classificação;
- 3) Foram convertidos dados não-numéricos em dados numéricos;
- 4) A base de dados foi separada entre dados de treino e teste, sendo 33% referente aos de teste e o restante para treino;
- 5) Os dados separados foram normalizados para uma melhor padronização nos mesmo, no pré-processamento.

No momento posterior, foram aplicados os três algoritmos em sequência em conjunto com as métricas de avaliação. A partir da execução dos algoritmos, foi possível observar o quão eficiente cada um se saiu, com base no auxílio das próprias métricas, possibilitando uma avaliação a fim de selecionar o que mais se adequou aos contextos, a partir de resultados significativos. Para fins de classificação baseado no *dataset* escolhido, um asteroide pode se encaixar em 4 tipos: *Apollo Asteroid*, *Amor Asteroid*, *Aten Asteroid* e *Apohele ou Atira Asteroid*. Segundo Mhatre e Mhatre (2020), asteroides da classe *Amor* possuem a órbita na zona entre a Terra e Marte em um espaço de tempo de milhões de anos, enquanto os classe *Apollo* e *Aten* são objetos que atravessam a Terra com um elevado fator de impacto. Já asteroides da classe *Apohele*, orbitam o Sol na órbita da Terra, passando a maior parte do seu tempo perto do próprio Sol (YE *et al*, 2020).

Contudo, o modelo escolhido foi salvo utilizando a biblioteca *pickle*, a fim de ser utilizado na construção de uma página *web*. Dessa forma, está sendo implementada uma Operação de Aprendizado de Máquina (*Machine Learning Operations* - MLOps).

4. RESULTADOS E DISCUSSÃO

A análise feita para escolha do melhor algoritmo, para a composição do modelo, com as métricas de avaliação já citadas, contou com resultados obtidos através da execução de cada algoritmo em cima da base de dados manuseada. A Tabela 1 apresenta os resultados colhidos, comparando a eficiência de cada algoritmo com cada métrica utilizada, quanto mais o valor de aproxima de 1, melhor resultado se obteve na classificação.

Tabela 1 - Resultados dos algoritmos em relação às métricas de avaliação

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>Support Vector Machines - RBF Kernel (SVM-RBF)</i>	≅ 0.96%	≅ 0.72%	≅ 0.71%	≅ 0.71%
<i>Naive Bayes</i>	≅ 0.89%	≅ 0.67%	≅ 0.81%	≅ 0.71%
<i>K-Nearest Neighbors (KNN)</i>	≅ 0.88%	≅ 0.65%	≅ 0.63%	≅ 0.64%

Fonte: Os autores (2023)

Dentre os resultados obtidos na Tabela 1, o modelo escolhido foi obtido através da aplicação do algoritmo SVM-RBF, para ser aplicado no MLOps. A Figura 1 apresenta uma prévia da aplicação web.

Classificação

Epoch (TDB): <input style="width: 90%;" type="text"/>	Orbit Axis (AU): <input style="width: 90%;" type="text"/>	Orbit Eccentricity: <input style="width: 90%;" type="text"/>	Orbit Eccentricity: <input style="width: 90%;" type="text"/>
Orbit Inclination (deg): <input style="width: 90%;" type="text"/>	Perihelion Argument (deg): <input style="width: 90%;" type="text"/>	Node Longitude (deg): <input style="width: 90%;" type="text"/>	Mean Anomaly (deg): <input style="width: 90%;" type="text"/>
Perihelion Distance (AU): <input style="width: 90%;" type="text"/>	Aphelion Distance (AU): <input style="width: 90%;" type="text"/>	Minimum Orbit Intersection Distance (AU): <input style="width: 90%;" type="text"/>	Orbital Reference: <input style="width: 90%;" type="text"/>
Asteroid Magnitude: <input style="width: 90%;" type="text"/>	Hazardous: <input style="width: 90%;" type="text"/>		

Figura 1 - Sistema web construído para classificação de asteroides

Fonte: Os autores (2023)

5. CONCLUSÃO

O atual trabalho teve como objetivo a aplicação de algoritmos de AM para classificação de asteroides. Depois de executado, foi avaliado o quão eficiente cada algoritmo se mostrou com cada uma das métricas também utilizadas. O SVM mostrou superior aos outros métodos juntamente ao valor de sua acurácia. Para futuros trabalhos com o mesmo embasamento, é proposto que se apliquem outros algoritmos de AM para classificação e avalie sua eficácia.

REFERÊNCIAS

- BI, Qifang et al. **What is machine learning? A primer for the epidemiologist**. American journal of epidemiology, v. 188, n. 12, p. 2222-2239, 2019.
- GARCIA-CANADILLA, Patricia et al. **Machine learning in fetal cardiology: what to expect**. Fetal diagnosis and therapy, v. 47, n. 5, p. 363-372, 2020.
- JOSHI, Shrushti. **Asteroid Impacts**. Kaggle, 2021. Acesso em: 01 de junho de 2023.
- MHATRE, Pradnesh; MHATRE, Pranjali. **Asteroid Mining: Future of Space Commercialization**. Population, v. 875, p. 8537, 2020.
- PRASAD, LV R. Chaitanya; REDDY, TA Shivani; KASHI, Bhuvaneshwari. **Asteroid Detection using Machine Learning Algorithm**. Communications of the Byurakan Astrophysical Observatory, v. 67, p. 329-334, 2020.
- SMIRNOV, Evgeny A.; MARKOV, Alexey B. **Identification of asteroids trapped inside three-body mean motion resonances: a machine-learning approach**. Monthly Notices of the Royal Astronomical Society, v. 469, n. 2, p. 2024-2031, 2017.
- YE, Quanzhi et al. **A Twilight Search for Atiras, Vatiras, and Co-orbital Asteroids: Preliminary Results**. The Astronomical Journal, v. 159, n. 2, p. 70, 2020.