

# DESENVOLVIMENTO DE UM PIPELINE PARA A ANÁLISE DE CURVAS DE LUZ COM TÉCNICAS DE MACHINE LEARNING

Pedro Wilian Palumbo Bevilacqua<sup>1</sup>; Roberto Bertoldo Menezes<sup>2</sup>

<sup>1</sup> Aluno de Iniciação Científica do Instituto Mauá de Tecnologia (IMT);

<sup>2</sup> Professor do Instituto Mauá de Tecnologia (IMT).

**Resumo.** *Técnicas de machine learning têm sido cada vez mais utilizadas na astronomia, por conta do volume e complexidade dos dados observacionais obtidos com diferentes instrumentos. Uma possível aplicação de tais técnicas envolve análises das chamadas curvas de luz, que são conjuntos de dados utilizados para monitorar a variabilidade do brilho das estrelas. A análise das curvas de luz permite identificar a presença de possíveis eclipses, que ocorrem quando um objeto em órbita da estrela observada passa entre a estrela e o observador. Tal corpo orbitante pode ser um exoplaneta ou uma outra estrela, que é chamada de binária eclipsante. Nesse trabalho, avaliamos, utilizando dados obtidos com o telescópio espacial Kepler, quais as técnicas de machine learning mais eficazes para a classificação de curvas de luz em: curvas contendo eclipses causados por exoplanetas, curvas com eclipses causados por binárias eclipsantes e curvas sem eclipses. A partir disso, elaboramos um pipeline para a análise de curvas de luz obtidas com diferentes instrumentos. Tal pipeline inclui a aplicação de técnicas de filtragem, a fim de remover ruídos de alta frequência, e a classificação das curvas de luz, usando técnicas de machine learning diversas. Ao final, constatou-se que a filtragem espacial de Butterworth foi a mais adequada para ser utilizada no pipeline e a técnica de machine learning mais eficaz foi o XGBoost. Vale ressaltar que o pipeline aqui desenvolvido será disponibilizado publicamente, para ser utilizado pela comunidade astronômica nas análises de futuros dados a serem obtidos com diferentes instrumentos.*

## Introdução

Curvas de luz são dados astronômicos frequentemente utilizados no estudo da variabilidade do brilho das estrelas e consistem em gráficos do brilho estelar em função do tempo. Um exame detalhado desses dados permite detectar, entre outros fenômenos, os eclipses, que ocorrem quando um corpo, em órbita ao redor da estrela observada, passa entre a estrela e o observador, resultando numa diminuição do brilho estelar, que pode ser detectada na curva de luz. Esses eclipses podem ser causados por exoplanetas ou mesmo por outras estrelas orbitando a estrela observada. Nesse último caso, têm-se as chamadas binárias eclipsantes. Observações feitas com telescópios espaciais forneceram curvas de luz para um número consideravelmente elevado de estrelas, o que resultou na identificação de milhares de exoplanetas e de binárias eclipsantes. O telescópio espacial CoRoT (*Convection, Rotation and planetary Transits*) (Auvergne et al., 2009), lançado em 27 de dezembro de 2006, por exemplo, que já não se encontra mais ativo, produziu curvas de luz para mais de 160000 estrelas. Ao final, foram detectados 34 exoplanetas e 2269 binárias eclipsantes (Deleuil et al., 2018). Outro telescópio espacial relevante para esse tipo de pesquisa foi o Kepler (Koch et al., 2010), lançado em 7 de março de 2009, que também já não está mais ativo, mas que produziu curvas de luz para mais de 200000 estrelas. Essa grande quantidade de dados resultou na detecção de 4034 candidatos a exoplanetas (Thompson et al., 2018), dos quais 2335 foram confirmados como exoplanetas, e de mais de 2400 binárias eclipsantes (Slawson et al., 2011; LaCourse et al., 2015).

Os mais recentes instrumentos astronômicos têm fornecido dados observacionais, inclusive curvas de luz, com volumes e complexidades cada vez maiores. Por conta disso, o uso de técnicas de *machine learning* para a análise de dados astronômicos tem crescido significativamente nos últimos anos. Os algoritmos de *machine learning* costumam ser divididos em supervisionados e não supervisionados. Os algoritmos supervisionados (que são os mais utilizados na astronomia) são usados para aprender relações entre conjuntos de variáveis, a partir de exemplos fornecidos. Uma vez que esse aprendizado esteja concluído, os algoritmos podem ser utilizados para prever o

comportamento de dados ainda não vistos. Os algoritmos de *machine learning* não supervisionados, por outro lado, não se baseiam em exemplos fornecidos para aprender relações entre conjuntos de variáveis. Na verdade, tais algoritmos consistem em um conjunto grande de ferramentas estatísticas utilizadas para diferentes tipos de análises de dados, como *clustering* (na qual os dados são divididos em grupos diferentes com base em suas propriedades), redução de dimensionalidade, detecção de *outliers*, entre outras. Diferentes técnicas de *machine learning* já foram utilizadas para a análise de curvas de luz (e.g. Thompson et al., 2015; McCauliff et al., 2015; Armstrong et al., 2016; Ansdell et al., 2018; Armstrong et al., 2018; Shallue & Vanderburg, 2018; Zucker & Giryes, 2018; Dattilo et al., 2019; Schanche et al., 2019; Yu et al., 2019; Osborn et al., 2020; Rao et al., 2021).

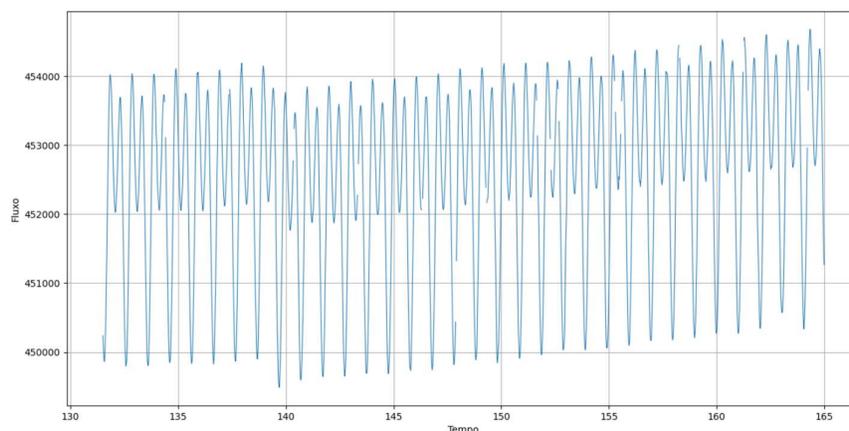
Este estudo apresenta o desenvolvimento de um *pipeline* para a análise de curvas de luz obtidas por diferentes instrumentos astronômicos, com o objetivo de realizar o pré-processamento e a classificação automática dessas curvas em três categorias: curvas com eclipses causados por exoplanetas, curvas com eclipses resultantes de binárias eclipsantes e curvas sem evidências de eclipses. Inicialmente, os dados passam por um tratamento prévio que inclui a remoção de ruído de alta frequência utilizando filtragem de Butterworth.

A análise concentra-se especialmente em dados do satélite Kepler (Kepler.,2024; KeplerSearch), buscando identificar as técnicas de *machine learning* mais eficazes para a classificação dessas curvas e também os parâmetros das curvas de luz mais relevantes para a classificação. Este *pipeline* representa uma ferramenta essencial para a análise de dados astronômicos, facilitando o tratamento e a interpretação dos dados de curvas de luz e promovendo avanços na classificação automática desses dados astronômicos.

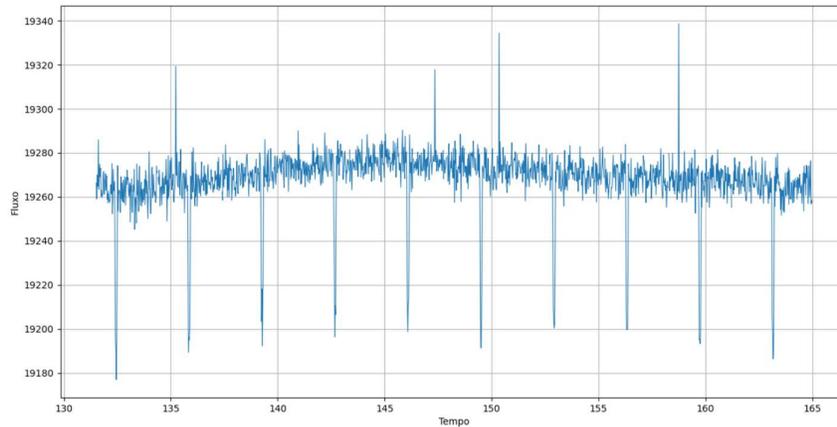
## Material e Métodos

Para iniciar este estudo a primeira etapa foi obter a amostra de curvas de luz observadas pelo satélite Kepler, categorizadas em três grupos principais: exoplanetas confirmados, binárias eclipsantes confirmadas e curvas sem evidências de eclipses. Para identificar as curvas com exoplanetas, utilizamos o catálogo de arquivos da NASA, que disponibiliza as curvas de luz associadas a esses objetos. Já para as binárias eclipsantes, utilizamos o *website* VizieR (VizieR,2024). As curvas sem eclipses foram selecionadas com base na ausência de identificação de fenômenos de eclipse. Inicialmente, foram selecionadas 100 curvas de luz para cada categoria, totalizando 300 curvas. Todas as curvas foram obtidas a partir do *Mikulski Archive for Space Telescopes* (MAST;KeplerSearch), um banco de dados público que armazena as observações do Kepler e de outros telescópios. Para garantir a consistência, selecionamos curvas de luz de longa cadência pertencentes ao primeiro quarto de observação, evitando a repetição de curvas em diferentes períodos de observação. As figuras 1,2 e 3 mostram exemplos de curvas de luz de cada um dos três grupos:

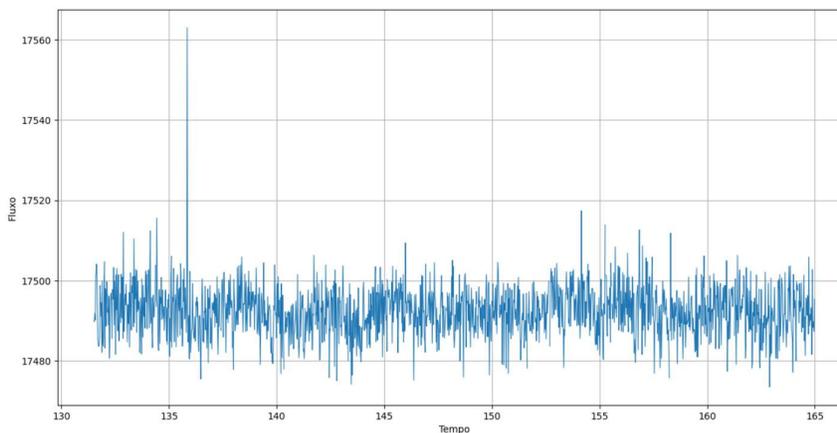
**Figura 1** – Exemplo de curva de luz contendo eclipses gerados por uma binária eclipsante



**Figura 2** – Exemplo de curva de luz contendo eclipses gerados por um exoplaneta



**Figura 3** – Exemplo de curva de luz que não contém eclipses



O pré-processamento dos dados foi a etapa subsequente ao *download* das amostras. O primeiro passo consistiu na conversão dos arquivos do formato fits (arquivo.fits), comumente utilizado em aplicações astronômicas, para o formato texto (arquivo.txt). Essa conversão permitiu a criação de uma base de dados estruturada, contendo colunas de tempo e fluxo das estrelas observadas, facilitando a manipulação dos dados nas etapas posteriores. Durante essa fase, foram identificados valores nulos ou vazios em algumas linhas, os quais poderiam impactar as análises nos processos de reamostragem e na geração de periodogramas. Para resolver esse problema, foram desenvolvidos e aplicados dois métodos (com a intenção de se avaliar qual deles teria melhor eficácia). O primeiro método consistiu na remoção de todas as linhas que continham valores nulos, eliminando assim qualquer dado incompleto. O segundo método aplicou uma interpolação linear nos valores ausentes, utilizando dados adjacentes para preencher as lacunas. Essa abordagem permitiu preservar a sequência temporal e reduzir a perda de dados, garantindo que as curvas de luz mantivessem sua integridade para as análises subsequentes.

Em seguida, foi realizada a reamostragem das curvas de luz para garantir que o intervalo de tempo entre os pontos consecutivos fosse uniforme e a cobertura temporal fosse igual em todas as curvas. Essa padronização foi fundamental, pois diferenças nos intervalos temporais poderiam impactar os resultados dos modelos de *machine learning*. Após a reamostragem, as curvas foram transformadas em periodogramas, visto que a análise em frequência pode auxiliar os modelos de *machine learning* a identificar padrões periódicos, tendências e anomalias, características essenciais para a classificação das curvas de luz. Um periodograma é uma representação gráfica ou uma técnica analítica que exhibe a distribuição da potência de um sinal em função da frequência. Em termos

matemáticos, o periodograma estima a densidade espectral de potência (DSP) de um sinal, que quantifica quanta "energia" está presente em cada frequência. A ideia principal é decompor um sinal no domínio do tempo em componentes no domínio da frequência usando a transformada de Fourier.

Cada etapa do processo gerou um arquivo de saída individual para cada curva de luz, permitindo uma análise detalhada dos resultados intermediários e otimizando o fluxo de trabalho. Essa abordagem garantiu maior organização e facilitou o rastreamento das transformações realizadas nos dados ao longo do *pipeline*. Todas as etapas foram implementadas utilizando a linguagem Python, com suporte de diversas bibliotecas especializadas. A biblioteca *astropy.io* foi utilizada para leitura e manipulação de arquivos no formato FITS; *numpy* e *pandas* foram empregadas no tratamento e análise de dados estruturados, incluindo métodos de filtragem dos valores ausentes; *matplotlib.pyplot* auxiliou na visualização gráfica dos dados; *scipy* foi aplicada para a reamostragem; *os* foi utilizada para manipulação de arquivos e diretórios, permitindo a automação de tarefas no ambiente computacional.

Na etapa final, antes da aplicação dos modelos de *machine learning*, foi realizada a normalização dos dados, consolidando os periodogramas em uma única base de dados contendo as 300 curvas iniciais, onde cada linha representa uma curva de luz e as colunas correspondem aos valores do módulo da transformada de Fourier ao quadrado obtido para cada frequência, constituindo a matriz de entrada  $X$ . A partir dessa base, foi criado um *array* com 300 linhas, com os valores 0, 1 e 2, para classificar as curvas de luz: 0 para "Nulo," 1 para "Exoplaneta," e 2 para "Eclipse Binário". Esse *array* representa a matriz de saída  $Y$ . Para garantir a reprodutibilidade, foi definida uma semente aleatória com valor fixo de 7, partindo o conjunto de dados em 75% para treino e 25% para teste, sendo usada principalmente a biblioteca *sklearn*.

Para a primeira aplicação de *machine learning*, optamos pelo modelo *XGBoost*, devido à sua robustez e eficiência. Nos testes iniciais, obtivemos uma acurácia média de 60% ao utilizar o método de interpolação para lidar com os valores nulos, em comparação com uma acurácia de 55% ao remover os valores ausentes, indicando que a interpolação era a melhor abordagem. No entanto, a acurácia de 60% sugeriu a necessidade de refinamento na amostra de dados. Após uma análise detalhada das curvas de luz, observamos que o modelo frequentemente confundia as curvas de exoplanetas com as curvas sem eclipses. Como solução, plotamos todas as 300 curvas e efetuamos um refinamento manual, descartando aquelas com evidência dúbia de eclipses ou com padrões anômalos.

Esse refinamento resultou na exclusão de várias curvas, reduzindo o tamanho da amostra. Para compensar essa perda, retomamos o processo de captação de dados, desta vez verificando previamente a qualidade das curvas antes de baixá-las. Como resultado, obtivemos um total de 330 curvas, distribuídas igualmente entre as três categorias (110 curvas por categoria). Após realizar o procedimento com o novo banco de dados, obtivemos uma melhora significativa na acurácia do modelo de *XGBoost*. Para otimizar o desempenho, utilizamos a biblioteca *Optuna*, que permite a realização de uma busca eficiente por hiperparâmetros, e um processo de repetição de testes com diferentes combinações de parâmetros, identificando tendências e áreas onde o modelo apresentou melhor desempenho. Finalmente, a combinação das estratégias de busca e ajuste permitiu alcançar picos de acurácia, resultando na melhor configuração possível para o modelo.

Em seguida, foi aplicado e testado um método de filtragem de ruídos nas curvas de luz, inserido logo após a reamostragem dos dados e antes da transformação em periodogramas, com o objetivo de verificar se a filtragem contribuiria para uma melhora na acurácia final do modelo. Utilizou-se a filtragem de Butterworth, que requer a definição de uma frequência de corte. Para identificar as melhores frequências de corte, o *pipeline* foi executado repetidamente para frequências de corte de 0 a 2 Ny (Ny representa a frequência de Nyquist)(Gonzalez & Woods, 2002), seguido pela aplicação do modelo *XGBoost* para analisar a eficiência das frequências de corte em cada caso. Observou-se que as frequências de corte que promoveram uma ligeira melhora na acurácia final dos modelos situam-se entre 0,6 e 1,0 Ny; para este estudo, foi selecionada a frequência de corte de 0,65 Ny.

Após a aplicação das técnicas de filtragem, foram testados diferentes modelos além do *XGBoost*, incluindo *LightGBM* e *Random Forest*, aos quais foram aplicadas buscas de hiperparâmetros para otimização. Em seguida, foi realizado o último passo, que consistiu na aplicação de modelos de *machine learning* não supervisionados, ou modelos de clusterização. Nesse caso, em vez de uma entrada e saída predefinidas, utilizou-se apenas a entrada, e o modelo foi responsável por determinar a divisão dos dados nas três categorias desejadas.

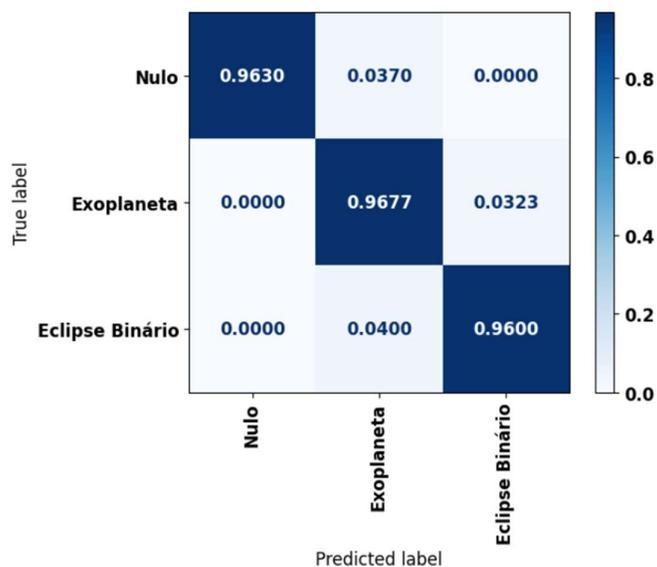
Foram aplicados os modelos *Gaussian Mixture Model* e *K-means*, cujos hiperparâmetros e configurações ideais foram determinados para obter o melhor desempenho, resultando em um *rand-score* que ofereceu uma primeira avaliação do comportamento dos modelos. Com o objetivo de obter uma métrica de acurácia em porcentagem, foi desenvolvido um método para mapear artificialmente as classes dos *clusters* com as classes reais definidas na etapa de normalização. Esse mapeamento foi realizado por meio do Algoritmo Húngaro (Algoritmo Húngaro., 2024) usando a biblioteca *scipy.optimize*, que mapeia as classes de modo a maximizar a acurácia, permitindo, assim, uma medida quantitativa dos resultados para análise. Com esses resultados, tornou-se possível comparar o desempenho entre os modelos supervisionados e não supervisionados, avaliando qual abordagem apresenta maior acurácia.

## Resultados e Discussão

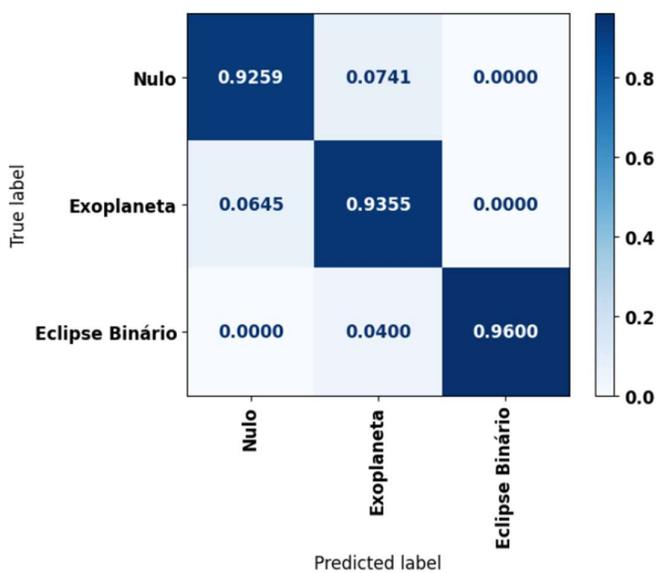
Inicialmente, verificou-se que, para este caso de estudo, substituir os valores nulos por interpolação mostrou-se mais eficaz do que simplesmente removê-los, pois essa abordagem preserva o tamanho das curvas, possibilitando uma reamostragem padronizada e minimizando a redução do comprimento da curva. Além disso, a análise revelou que a aplicação do método de filtragem de Butterworth foi eficaz para melhorar a acurácia total do modelo. No entanto, ressalta-se a importância de definir a frequência de corte com cautela, para evitar a remoção das próprias assinaturas de eclipses que se deseja identificar. A acurácia do modelo *XGBoost*, sem filtragem de ruído, foi de 93%, enquanto com a filtragem de Butterworth aumentou para 96,4%. Adicionalmente, após testes, verificou-se que, para o conjunto de dados restrito a 330 curvas de luz, a divisão de 75/25 entre treino e teste foi a mais adequada para este banco de dados.

Com a aplicação do *pipeline* proposto, foram obtidos resultados excepcionais para todos os modelos supervisionados após a otimização dos hiperparâmetros. No modelo *XGBoost*, a acurácia atingiu 96,4%, com os hiperparâmetros  $n\_estimators = 48$  e  $max\_depth = 1$ ; para o modelo *LightGBM*, a acurácia foi de 93,98%, com  $n\_estimators = 25$  e  $max\_depth = 1$ ; para o modelo *RandomForest*, a acurácia foi de 92,77%, com  $n\_estimators = 12$ . As Figuras 4, 5 e 6 a seguir mostram as matrizes de confusão dos modelos supervisionados testados.

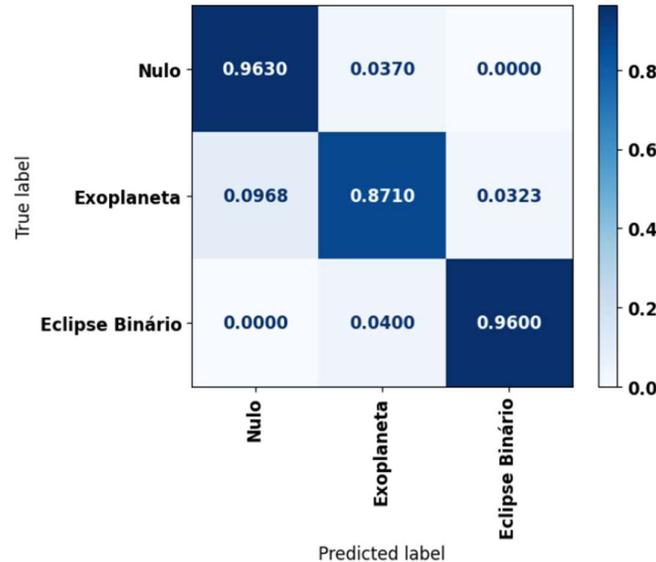
**Figura 4** – Matriz de confusão de teste do *XGBoost*



**Figura 5** – Matriz de confusão de teste do *LightGBM*



**Figura 6** – Matriz de confusão de teste do *RandomForest*



Os resultados obtidos para todos os modelos foram positivos, com destaque para o modelo *XGBoost*, que apresentou a maior acurácia entre as alternativas testadas. Assim, o *XGBoost* mostrou-se o modelo mais adequado para este estudo de curvas de luz, revelando-se ligeiramente mais eficiente do que os demais.

Além disso, foram testados modelos não supervisionados, que operam sem dados de saída predefinidos. Utilizando um algoritmo desenvolvido para mapear artificialmente as classes de saída, obtivemos uma acurácia de 57,83% com o *Gaussian Mixture Model*, utilizando o hiperparâmetro, *covariance\_type = diag*, e 62,65% com o modelo *K-means*, com o hiperparâmetro *n\_init = 10*. Esses resultados indicam que, embora os modelos não supervisionados apresentem um desempenho razoável, considerando a ausência de dados de saída, ainda são significativamente inferiores aos modelos supervisionados neste caso de estudo. Portanto, a aplicação de modelos não supervisionados para a análise de curvas de luz não é recomendada para o modelo final, sendo mais adequada para testes preliminares, com o objetivo de definir as classes e verificar a capacidade dos *clusters* em identificar padrões de forma eficiente.

## Conclusões

Este estudo apresentou o desenvolvimento e a implementação de um *pipeline* para análise e classificação automática de curvas de luz, com o objetivo de identificar assinaturas de eclipses causados por exoplanetas, binárias eclipsantes e curvas sem eclipses. Utilizando uma abordagem que combinou a filtragem de Butterworth para remoção de ruídos e modelos de *machine learning* supervisionados, obteve-se uma acurácia elevada, destacando o modelo *XGBoost* com uma taxa de acerto de 96,4%. Os resultados indicaram que a filtragem de ruídos pode melhorar a precisão, mas exige cautela para que a frequência de corte não interfira nas assinaturas de eclipses, evitando a perda de informações relevantes.

Adicionalmente, foi avaliada a eficácia de modelos não supervisionados, como *Gaussian Mixture Model* e *K-means*, para a análise exploratória das curvas de luz. Embora esses modelos tenham mostrado um desempenho razoável na ausência de dados de saída, com acurácia de até 62,65% no *K-means*, os resultados foram inferiores aos dos modelos supervisionados. Essa análise indica que os modelos supervisionados são mais adequados para a tarefa de classificação neste estudo, enquanto os não supervisionados podem ser úteis em etapas iniciais de exploração de dados.

Assim, conclui-se que o *pipeline* proposto, envolvendo o uso de técnicas de *machine learning* supervisionado, oferece uma abordagem robusta e eficiente para a classificação de curvas de luz,

permitindo a identificação de fenômenos astrofísicos com elevada acurácia. Este trabalho fornece uma base sólida para estudos futuros que pretendam aprimorar a classificação de curvas de luz e explorar outras fontes de dados, contribuindo para o avanço na pesquisa de exoplanetas e binárias eclipsantes. Além disso, o pipeline desenvolvido será disponibilizado publicamente em uma plataforma como o *GitHub*, incentivando a reprodutibilidade dos resultados e facilitando sua aplicação em novos contextos e investigações científicas.

## Referências Bibliográficas

Algoritmo Húngaro. Disponível em: <https://elemarjr.com/clube-de-estudos/artigos/o-algoritmo-hungaro-e-a-solucao-de-problemas-de-atribuicao/>. Acesso em: 03 dez. 2024.

ANSDELL, M.; IOANNOU, Y.; OSBORN, H. P.; SASDELLI, M.; 2018 NASA Frontier Development Lab Exoplanet Team et al. Exoplanet detection with machine learning: exploring the Kepler data set. *The Astrophysical Journal*, v. 869, n. L7, 2018.

ARMSTRONG, D. J.; GÜNTHER, M. N.; McCORMAC, J.; SMITH, A. M. S.; BAYLISS, D. et al. New planets from the Next Generation Transit Survey. *Monthly Notices of the Royal Astronomical Society*, v. 478, n. 4225, 2018.

ARMSTRONG, D. J.; KIRK, J.; LAM, K. W. F.; McCORMAC, J.; OSBORN, H. P. et al. Transit timing variations in Kepler planets: a uniform search. *Monthly Notices of the Royal Astronomical Society*, v. 456, n. 2260, 2016.

AUVERGNE, M.; BODIN, P.; BOISNARD, L.; BUEY, J.-T.; CHAINTREUIL, S. et al. The CoRoT satellite in flight: description and performance. *Astronomy & Astrophysics*, v. 506, p. 411, 2009.

DATTILO, A.; VANDERBURG, A.; SHALLUE, C. J.; MAYO, A. W.; BERLIND, P. et al. Identifying exoplanets with machine learning. *The Astronomical Journal*, v. 157, n. 169, 2019.

DELEUIL, M.; AIGRAIN, S.; MOUTOU, C.; CABRERA, J.; BOUCHY, F. et al. The CoRoT space mission. *Astronomy & Astrophysics*, v. 619, n. 97, 2018.

GONZALEZ, R. C.; WOODS, R. E. Digital image processing. 2. ed. Upper Saddle River: Prentice Hall, 2002.

KEPLER Search. Mikulski Archive for Space Telescopes. Disponível em: [https://archive.stsci.edu/kepler/data\\_search/search.php](https://archive.stsci.edu/kepler/data_search/search.php). Acesso em: 21 nov. 2024.

Kepler Objects of Interest. Disponível em: [https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=q1\\_q17\\_dr25\\_koi](https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=q1_q17_dr25_koi). Acesso em: 21 nov. 2024.

KIRK, B. et al. Kepler Eclipsing Binary Catalog. *The Astronomical Journal*, v. 151, n. 68, 2016.

KOCH, D. G.; BORUCKI, W. J.; BASRI, G.; BATALHA, N. M.; BROWN, T. M. et al. Overview of the Kepler Mission. *The Astrophysical Journal*, v. 713, n. L79, 2010.

LACOURSE, D. M.; JEK, K. J.; JACOBS, T. L.; WINARSKI, T.; BOYAJIAN, T. S. et al. Planetary candidates from the Kepler mission. *Monthly Notices of the Royal Astronomical Society*, v. 452, p. 3561, 2015.

McCAULIFF, S. D.; JENKINS, J. M.; CATANZARITE, J.; BURKE, C. J.; COUGHLIN, J. L. et al. Automatic vetting of Kepler planetary candidates. *The Astrophysical Journal*, v. 806, n. 6, 2015.

OSBORN, H. P.; ANSDELL, M.; IOANNOU, Y.; SASDELLI, M.; ANGERHAUSEN, D. et al. Exoplanet detection in CoRoT light curves with machine learning. *Astronomy & Astrophysics*, v. 633, n. 53, 2020.

RAO, S.; MAHABAL, A.; RAO, N.; RAGHAVENDRA, C. Machine learning applications in exoplanet discovery. *Monthly Notices of the Royal Astronomical Society*, v. 502, n. 2845, 2021.

SCHANCHE, N.; COLLIER CAMERON, A.; HÉBRARD, G.; NIELSEN, L.; TRIAUD, A. H. M. J. et al. The NGTS survey: new planetary candidates. *Monthly Notices of the Royal Astronomical Society*, v. 483, n. 5534, 2019.

SHALLUE, C. J.; VANDERBURG, A. Identifying exoplanets with deep learning. *The Astronomical Journal*, v. 155, n. 94, 2018.

SLAWSON, R. W.; PRSA, A.; WELSH, W. F.; OROSZ, J. A.; RUCKER, M. et al. Kepler eclipsing binary catalog. *The Astronomical Journal*, v. 142, n. 160, 2011.

THOMPSON, S. E.; COUGHLIN, J. L.; HOFFMAN, K.; MULLALLY, F.; CHRISTIANSEN, J. L. et al. The Kepler data release 25: planetary candidates. *The Astrophysical Journal Supplement Series*, v. 235, n. 38, 2018.

THOMPSON, S. E.; MULLALLY, F.; COUGHLIN, J.; CHRISTIANSEN, J. L.; HENZE, C. E. et al. Kepler's extended mission discoveries. *The Astrophysical Journal*, v. 812, n. 46, 2015.

VizieR. Disponível em: [https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=J/AJ/151/68/catalog&-out.max=50&-out.form=HTML%20Table&-out.add=\\_r&-out.add=\\_RAJ](https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=J/AJ/151/68/catalog&-out.max=50&-out.form=HTML%20Table&-out.add=_r&-out.add=_RAJ). Acesso em: 21 nov. 2024.

YU, L.; VANDERBURG, A.; HUANG, C.; SHALLUE, C. J.; CROSSFIELD, I. J. M. et al. Machine learning detection of exoplanets in TESS data. *The Astronomical Journal*, v. 158, n. 25, 2019.

ZUCKER, S.; GIRYES, R. Machine learning techniques in astrophysics. *The Astronomical Journal*, v. 155, n. 147, 2018.