

## **CARACTERIZAÇÃO SEMI-AUTOMÁTICA DE HABITABILIDADE DE EXOPLANETAS VIA MÁQUINAS DE VETOR SUPORTE**

Matheus Pedro Rodrigues Ferreira (IC) e Luciano Silva (Orientador)

**Apoio: PIBIC Santander**

### **RESUMO**

O Intuito deste trabalho é pesquisar e estudar sobre a classificação de exoplanetas assim como analisar a forma mais eficiente de se executar isso.

Usando o método HITE que se utiliza da Luminosidade do Exoplaneta para extrair diversos dados para a composição de um Índice parametrizado, no qual se pode tirar probabilidade, de um planeta analisado, ser ou não habitável para vida Humana.

A tradução desse método é feita por Técnicas de Aprendizado de Máquina, na qual aumenta a eficiência da análise em relação a uma implementação ordinária. Essa implementação é feita utilizando uma técnica sofisticada de Aprendizado de Máquina chamada Redes Neurais Convolucionais (CNN) que tem desempenhos satisfatórios com grandes Volumes de Dados em multicores também apresentam ganhos com a utilização de Unidades de Processamento Neurais (NPU)

Esse Cálculo, porém, requer uma certa capacidade de máquina para ser executado assim esse trabalho também propõe uma maneira eficiente de calcular o maior número possível de planetas por tempo levando em conta o custo da máquina.

Com o decorrer do Artigo serão apresentados a teoria do por trás do método HITE, o desenvolvimento teórico do Método, a metodologia do Projeto, os Resultados obtidos, a Análise dos resultados e o as bibliografias utilizadas.

**Palavras-chave:** NPU; CNN; HITE;

## **ABSTRACT**

The Intuit of this work is to research and study the classification of exoplanets as well as to analyze the most efficient way of executing this.

Using the HITE method which uses the Luminosity of the Exoplanet to extract a range of data for the composition of a parameterized Index, in which one can take probability, of an analyzed planet, to be habitable or not for human life.

The translation of this method is done by Machine Learning Techniques, in which increases the efficiency of the analysis in relation to an ordinary implementation. This implementation is done using a sophisticated Machine Learning technique called Convolutional Neural Networks (CNN) that has satisfactory performances with large data volumes in multi-cores also present gains with the use of Neural Processing Units (NPUs)

This calculation, however, requires a certain amount of machine capacity to be executed, so this work also proposes an efficient way of calculating as many planets as possible per time considering the cost of the machine.

Throughout the article, it will be presented the theory behind the HITE method, the theoretical development of the Method, the methodology of the Project, the results obtained, the analysis of the results and the bibliographies used.

**Keywords:** NPU; CNN; HITE

## 1. INTRODUÇÃO

Exoplanetas são planetas que orbitam outras estrelas que não sejam o Sol (Perryman, 2014). A busca de exoplanetas na zona habitável é importante para o estudo do Universo como um todo, pois permite entender o processo de formação de sistemas planetários e, em particular, o próprio Sistema Solar. Exoplanetas são, geralmente, difíceis de se detectar por métodos visuais, pois a distância entre a Terra e eles é muito grande e requerem tempo de análise e comprovação de dados coletados.

Em princípio, assim como a Terra faz parte de um sistema estelar (tendo o Sol como estrela), os exoplanetas também fazem parte de seus próprios sistemas estelares. Os métodos utilizados para detectar exoplanetas variam bastante. Porém, há métodos que são mais eficazes e outros que dependem de determinadas suposições a priori sobre o sistema estrela-exoplanetas (OLIVEIRA e SARAIVA, 2014).

Os principais métodos para detecção de exoplanetas incluem astrometria, medição da velocidade radial, imageamento direto, lentes gravitacionais e trânsitos planetários. O método de trânsitos planetários, em particular, tem sido o mais bem-sucedido nos últimos anos, devido ao número de exoplanetas detectados (Perryman, 2014).

Uma vez que o exoplaneta seja detectado, parte-se para sua caracterização astrofísica e astroquímica como raio da órbita, fator de impacto, composição química, visando a sua classificação como exoplaneta dentro da zona habitável. Como a quantidade de parâmetros que definem condições de habitabilidade é grande, a classificação manual de exoplanetas dentro da zona habitável pode ser bastante custosa, o que justifica o uso de métodos de classificação como, por exemplo, o SVC disponibilizado por Máquinas de Vetor-Suporte (HAMEL, 2011).

Dentro deste contexto, o objetivo deste trabalho é aplicar técnicas de SVC com Máquinas de Vetor-Suporte para classificação semi-automática de habitabilidade para dados de exoplanetas do catálogo da missão Kepler, que monitora cerca de 150.000 estrelas na Constelação de Cisne (NASA, 2015).

As implementações das técnicas de SVC para caracterização de exoplanetas será realizada em Python, que serão disponibilizadas como um *notebook* público para uso de astrônomos e astrofísicos.

Este documento está organizado da seguinte forma:

- a Seção 1 apresenta o Referencial Teórico;
- a Seção 2 apresenta a metodologia;
- a Seção 3 apresenta os resultados e discussões;
- a Seção 4 apresenta as Considerações Finais;
- a Seção 5 as Referências usadas para o Artigo;

No final do projeto, encontram-se as referências bibliográficas iniciais para produção e realização deste projeto.

## **2. REFERENCIAL TEÓRICO**

Por definição, exoplanetas são planetas fora do Sistema Solar, que tem seu próprio sistema e esse sistema tem sua própria estrela ou estrelas. Se considerarmos que os sistemas longínquos obedecem o mesmo princípio do nosso sistema, desde sua formação ao seu funcionamento, usar o Sistema Solar como exemplo é um bom começo para estudar exoplanetas, determinar como eles compostos e o que os tornam únicos (PERRYMAN, 2014).

Para a identificação de Exoplanetas são usados 5 principais métodos, porém nesse trabalho se utilizará do Método de Trânsito Planetário que já identificou 2714 planetas.

Esse método consiste na quantidade de Luz observada na Terra emitido pelas estrelas a serem Observadas, através de monitoramento contínuo da estrela. O objetivo desse monitoramento é determinar qualquer alteração no fluxo de Luz da estrela, o que indica se um planeta a orbitando passou na frente da Terra no momento da observação.

Baseado no Método de Trânsito Planetário, Barnes, Meadows e Evans, criaram o HITE (Habitability Index for Transiting Exoplanets). Um Algoritmo que se baseia nas informações coletadas e gera um índice que Habitabilidade para os Exoplanetas, baseados em conceitos Humanos.

## **3. METODOLOGIA**

Primeiramente foi analisado o algoritmo e seus aspectos para determinar o seu desenvolvimento em código. A Figura 1 mostrar parte do cálculo do HITE.

$$p_{rocky}(R_p) = \begin{cases} 1, & R_p \leq 1.5 R_{\oplus} \\ (2.5 - R_p), & 1.5 R_{\oplus} < R_p < 2.5 R_{\oplus} \\ 0, & R_p \geq 2.5 R_{\oplus} \end{cases}$$

$$F_{max} = B\sigma \left( \frac{l}{2R \ln(P_* \sqrt{\kappa} P_0 g)} \right)^4, \quad H = \frac{\sum h_j p_j(e)}{\sum p_j(e)} p_{rocky},$$

Figura 1. Algoritmo HITE. Fonte: Barnes, Meadows, Evans - Comparative Habitability Of Transiting Exoplanets, 2015.

Em Seguida o Algoritmo foi traduzido em Python para fazer o calculo do índice por força bruta e então, analisado seu tempo de execução para um numero de planetas.

Foi analisado que o quão custoso em tempo o método era e então foi analisado com a implementação do método de Machine Learning SVM (Support Vector Machine) que usa dados já analisados e categorizados para montar um modelo. O Numero de corte definido no sistema foi de 0.7 H (HITE) sendo que igual ou acima desse valor o planeta seria classificado com Habitável e abaixo como não Habitável, logo o sistema categorizava os planetas em somente dois Tipos. Esse numero foi escolhido, pois abaixo dele os Exoplanetas necessitam adaptações para suportar vida, ou seja, ser habitável. Na Figura 2 e Figura 3, é ilustrado algumas entradas e seu impacto no calculo do índice HITE.

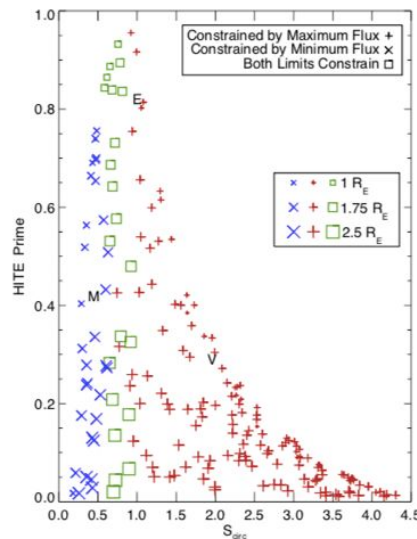


Figura 2. Esse Gráfico ilustra a relação entra a Orbits do Exoplaneta ( $S_{circ}$ ) e seu respectivo índice, onde M, E e V representam, respectivamente, Marte, Terra e Vênus. Observando o Gráfico extraímos uma relação inversamente proporcional do tamanho da Orbits e o Índice. Fonte: Barnes, Meadows, Evans - Comparative Habitability Of Transiting Exoplanets, 2015.

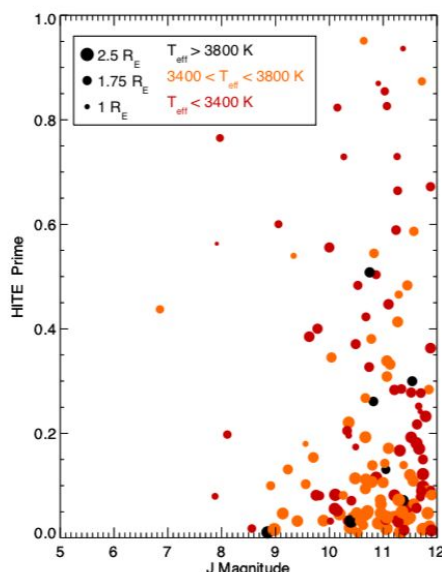


Figura 3. Esse Gráfico ilustra a relação entre a Magnitude (J) do Exoplaneta e seu respectivo índice. Observando o Gráfico extraímos que a Magnitude exerce pouca influência no cálculo do Índices. Fonte: Barnes, Meadows, Evans - Comparative Habitability Of Transiting Exoplanets, 2015.

Em seguida esse Modelo é usado para comparar inputs de novos planetas a serem classificados. Se tratando de uma comparação e não do calculo completo do índice o tempo de categorização cai e então o numero de classificação por minuto aumenta.

A base de dados gerada pelo algoritmo de força-bruta foi utilizada como a base de treinamento da SVM e depois foi utilizado novos inputs, não contidos na base de treinamento, para medir a taxa de acerto do algoritmo, comparando com os resultados obtidos através da SVM com o resultado calculado pelo algoritmo de força-bruta. O sistema foi calibrado até que chegasse a uma taxa de 95% entre a base de novos inputs e o calculo feito por script em Python.

Porém o esse método requer um investimento em maquina alto, como um Processador multicore ou uma GPU para se utilizar de paralelismo, o que torna o método custoso em quesito dinheiro. Foi então que foi comparado a utilização de um Intel core i7, GTX 1080 e a NPU Movidius, representados pelas figuras 4, 5 e 6, respectivamente.



Figura 4. Intel Core i7. Fonte: Google



Figura 5. NVidia GTX 1080. Fonte: Google



Figura 6. Intel Movidius. Fonte: Google

Para cada equipamento foi necessária uma adaptação, a GPU o desenvolvimento de Paralelismo e para NPU, a criação de CNN com Tensor Flow que a Movidius já tem suporte.

A CPU custando em torno de 750 a 800 reais, se demonstrou o hardware mais lento a ser utilizado chegando a analisar cem mil planetas em 200 segundos (3 minutos e 20 segundos), a GPU por sua vez, custando de 3000 a 3500 reais, analisou cem mil planetas em 143 segundos (2 minutos e 23 segundos) e por fim a NPU Movidius, custando entre 800 a 900 reais, analisou cem mil planetas em 185 segundos (3 minutos e 5 segundos).

Construindo uma taxa de planetas analisados por tempo a CPU desempenharia 500 planetas por segundo, a GPU 699.3 planetas por segundo e a NPU Movidius 540 planetas por segundo.

#### 4. RESULTADO E DISCUSSÃO

Os Resultados finais mostram que o Algoritmo Executado em uma GPU é o mais eficiente em quesito planetas Categorizados por segundo, seguido de NPU Movidius e CPU. A Tabela seguir traz com mais detalhes os testes feitos e os resultados obtidos.

Tempo de Execução por Hardware para Número de Planetas

Nº de Exoplanetas	CPU I7 (Octacore)	GPU GTX 1080	NPU Movidius
10	10	5	7
100	18	11	13
1000	40	29	35
10000	92	79	86
100000	200	143	185

Fonte: Própria

Com os resultados a cima fica constado nos testes executados que, para vias de efetividade, não importando o custo de hardware, a melhor escolha é o uso de GPU com paralelismo, pois é a ferramenta que categoriza mais Exoplanetas por segundo. Isso demonstra que a combinação de paralelismo e o método de SVM resulta em processo efetivo de análise em menos custo de tempo, atingindo assim o objetivo do projeto.

Em complemento ao objetivo inicial, um outro dado que se pode extrair sobre os resultados é qual dessas combinações resulta em uma taxa de categorização levando em conta o custo do hardware.

Considerando:

1. Os custos de compra dos Hardwares sendo:
  - a. CPU: R\$ 750,00.
  - b. NPU: R\$ 900,00.
  - c. GPU: R\$ 3500,00.
2. O Hardware Mínimo é a própria CPU.

Por mais R\$ 150,00 há um ganho de 40 Explanetas analisados por segundo resultando em um ganho de 8% na Taxa de Categorização. Analisando o investimento a mais por planeta, resultaria R\$ 3.75 por planeta a mais.

Por mais R\$ 2750,00 há um ganho de, aproximadamente, 200 Exoplanetas analisados por segundo resultando em um ganho de 40% na taxa de Categorização. Analisando o investimento a mais por planeta, resultaria em R\$ 13,75 por planeta a mais.

Sendo assim a NPU Movidius é o hardware custo beneficio, pois com um investimento relativamente baixo, pode-se extrair resultados interessantes e se o custo não for um limitante, a implementação em GPU é a melhor opção.

## 5. CONSIDERAÇÕES FINAIS

O objetivo inicial, descrito no projeto desse artigo, de implementação em Python foi cumprido, porém a disponibilização do código para uso publico não, contudo no processo do



desenvolvimento do trabalho foi possível fazer mais duas análises, a primeira implementação mais eficiente em quesito tempo e a segunda, qual implementação tem o melhor equilíbrio entre custo e eficiência.

## 6. REFERÊNCIAS

HAMEL, L. **Knowledge Discovery with Support Vector Machines**. New York: Wiley- Interscience, 2011.

JAIN, P. **An Introduction to Astronomy and Astrophysics**. New York: CRC Press, 2015. JOSHI, P. **Python Machine Learning Cookbook**. New York: Packt Publishing, 2016.

NASA (2015). **NASA Exoplanet Archive: A Service of NASA Exoplanet Science Institute**. Disponível em: <http://exoplanetarchive.ipac.caltech.edu/>. Acesso em 10/03/2017.

OLIVEIRA, K.; SARAIVA, M.F. **Astronomia e Astrofísica**. São Paulo: Livraria da Física, 2014.

PERRYMAN, M. **The Exoplanet Handbook**. Cambridge: Cambridge University Press, 2014. RASCHKA, S. **Python Machine Learning**. New York: Packt Publishing, 2015.

BARNES, R; MEADOWS, V. S; EVANS, N. **COMPARATIVE HABITABILITY OF TRANSITING EXOPLANETS**. Draft version October 1, 2015.

**Contatos:** matheusprf95@gmail.com e luciano.silva@mackenzie.br