Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230089

Identifying potentially habitable exoplanets: A study using the transit method and Kepler dataset

Dr Richa Gupta[1,2], Sidratul Muntaha[1,3]

[1]Department of Computer Science and Engineering, SEST, Jamia Hamdard, 110062-New Delhi, India

[2]richagupta@jamiahamdard.ac.in
[3]muntahasidratul725@gmail.com

Abstract. An exoplanet is a planet that orbits a star outside of our solar system. The study of exoplanets is an active area of research in astronomy. In this research, we aim to utilize the Kepler dataset provided by NASA EXOPLANET ACRCHIEVE to identify and classify exoplanets that could potentially support life. The Kepler dataset, which comprises of observations of over 150,000 stars, has been instrumental in the discovery of thousands of exoplanets. We will analyse the dataset using machine learning techniques to classify exoplanets as potentially habitable based on their orbital period, size, distance from their host star, and other parameters. The findings of this research will greatly enhance our understanding of the frequency of life in the universe and the use of machine learning techniques on the Kepler dataset will be an essential tool in the quest for finding potentially habitable exoplanets. Emerging Machine Learning Algorithms aid in detecting habitability of exoplanet in different stellar system. For finding an Exoplanet we have used the “transit method” which is based on the principle that when an exoplanet passes in front of its host star, it causes a temporary dip in the star's brightness. By monitoring the brightness of a star over time, scientists can detect these periodic dips and use them to infer the presence of an exoplanet. The findings of this research have the potential to significantly advance our understanding of the prevalence of life in the universe.

Keywords: exoplanets, Kepler dataset, PHL dataset, machine learning.

1. Introduction

The search for potentially habitable exoplanets has been a topic of intense interest in recent years. With the discovery of thousands of exoplanets in the last two decades, the possibility of finding a planet similar to Earth in terms of habitability has become increasingly likely. The characteristics that make a planet potentially habitable include the presence of liquid water, a stable climate, and a suitable atmosphere. In this research paper, we will explore the methods used to identify potentially habitable exoplanets and examine recent discoveries in the field. We will also discuss the potential for future discoveries and the implications of finding a truly habitable exoplanet. Based on Kepler data, scientists have estimated that there could be nearly 40 billion earth sized planet in our own Milky Way galaxy alone [1].

DOI: 10.54254/2755-2721/8/20230089

To detect or study an exoplanet, we have to work with the star. Scientist discovered these exoplanets by various method Transit, Direct Imaging, Radial Velocity, Gravitational Microlensing and Astrometry. The wobble method detects exoplanets via a Doppler shift in the star's light frequencies caused by its planets; direct imaging uses extra terrestrial telescopes to capture images from exoplanets; gravitational micro lensing analyses the distortion of the background light [2].

The transit method is one of the most successful techniques used to detect exoplanets. It is based on the principle that when an exoplanet passes in front of its host star, it causes a temporary dip in the star's brightness. By monitoring the brightness of a star over time, scientists can detect these periodic dips and use them to infer the presence of an exoplanet. One of the advantages of the transit method is that it can provide information about the exoplanet's size, orbital period, and distance from the host star. Additionally, by measuring the duration and depth of the transit, scientists can estimate the exoplanet's density, which can be used to infer its composition. The NASA's Kepler telescope has been able to detect the exoplanet by using the transit method. The first goal of the project is to find an exoplanet through Kepler Dataset and second goal is to predict whether the planet is habitable or non-habitable using PHL Data [8].

Figure 1 showing transit method looking for tiny dips in the brightness of a star when a planet orbiting it crosses in front of it, we say the planet transit the star.

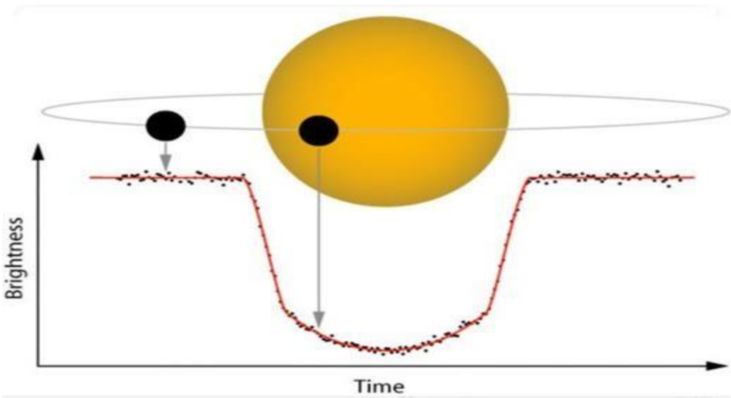

Figure 1. The Transit method, a groundbreaking astronomical technique, unveils the presence of distant exoplanets by detecting subtle changes in starlight. As an exoplanet passes in front of its parent star, it casts a tiny shadow, causing a periodic dimming of the star's brightness. This captivating image symbolizes the remarkable discoveries made possible through the transit method, as it allows scientists to uncover hidden worlds beyond our Solar System.

2. Literature survey

Table 1. This comprehensive table represents a literature survey on the remarkable discoveries of exoplanets achieved through the transit method and Kepler dataset.

Sr.No.	Paper Title and its Author	Details of Publications	Findings
1.	Identifying Exoplanets with Deep Learning By- Christopher J. Shallue and Andrew Vanderburg.	Published in The Astronomical Journal in the year 2018.	. Use of Deep Learning and Neural Networks in finding Exoplanets. . Use of CNN (Convolutional Neural Network) in Machine Learning.

DOI: 10.54254/2755-2721/8/20230089

Table 1. (continued).

Sr.No.	Paper Title and its Author	Details of Publications	Findings
2.	The NASA Exoplanet Archive: Data and Tools for Exoplanet Research By- Marcy Harbut and Stephen Kane	Publications of the Astronomical Society of the Pacific. July 2013	. How to use NASA exoplanet archive . How to extract useful data from it
3.	Exoplanet Detection Using Machine Learning By- Abhishek Malik, Ben Moster and Christian Obermeier.	Published in astrophysics paper in year 2015.	. How to use NASA's Dataset. . How to apply ML Algorithms to find Exoplanets.
4.	Machine Learning Pipeline for Exoplanet Classification By- George Clayton Sturrock, Brychan, Manry, Sohail Rafiqi.	Published in SMU Data Science Review Volume 2, Number1, Article 9 in Year 2019.	. SVM, KNN and Random Forest Classifications models were created to classify data found in the Kepler Dataset. . Classification of Exoplanets.

3. Methodology

3.1. Dataset and features

The discovery of exoplanets has been revolutionized in recent years with the use of machine learning techniques applied to data collected by NASA's Kepler telescope. In this research paper, we aim to utilize machine learning techniques to analyse the Kepler dataset in order to identify exoplanets that could potentially support life. We will explore the methods used to analyse the data, the criteria used to classify exoplanets as potentially habitable, and the application of different machine learning algorithms. The foremost data for this research was NASA EXOPLANET ACRCHIEVE. For detection of an exoplanet Kepler Dataset is used and PHL dataset for predicting whether the planet is habitable or non-habitable. Kepler dataset contains 9564 rows and 83 columns, entities like kepid, kepler_name,koi_pdispoisition and other features of these planet. For more EDA and creation models We need flux values for each star with this purpose, I use light curve library to extract time- series flux data via kepid feature [9]. We just need kepid column in this stage, for that we created a new reduced data frame.

3.2. Data pre-processing and merging

As we know dataset provided by the NASA EXOPLANET ACRCHIEVE contain lots of missing values and it is very important to clean the data by removing all the missing values from the dataset which is the first step of Data Preprocessing. Data preprocessing involves cleaning and preparing the data for analysis. This may include removing any outliers or bad data points, correcting any errors, and normalizing the data. This step is important to ensure that the data is accurate and consistent, and that any analysis performed on it will yield reliable results. After analysing the missing values, we decided to remove the column data which has value greater than 40% of missing data. Data merging involves combining multiple datasets into a single dataset. This is often necessary when using multiple methods to search for exoplanets, as each method may produce its own dataset. Once the data is preprocessed

DOI: 10.54254/2755-2721/8/20230089

and merged, it can be analysed using various algorithms and statistical methods to search for evidence of exoplanets. These may include periodograms, Bayesian analysis, machine learning algorithms, and others. The results of the analysis can then be used to confirm or disprove the presence of exoplanets, and the data can be further refined and analysed as needed.

3.3. Feature engineering

For our first part of the goal of the research, for finding an exoplanet the dataset we used uses time-series and it not good or appropriate for machine learning algorithms since it has 1626 columns, Next step, is Feature Engineering .“Feature engineering is the process of transforming raw data into features that represent the underlying problem to the predictive models resulting in improved models’ accuracy on unseen data” [4].In the case of transit method, feature engineering may involve creating a light curve of the star, which plots the star's brightness over time. This can be used to identify the characteristic dip in brightness caused by a planet passing in front of the star, and to estimate the planet's size and orbital properties. Once the features are extracted, they can be used as inputs for various algorithms and statistical methods for analysis and to identify the presence of exoplanets.

4. Discussions

Finding exoplanets is a challenging task that requires a combination of advanced technology, statistical methods, and data analysis techniques. The process of finding exoplanets can be divided into several steps, including data collection, data preprocessing, data merging, feature engineering, and data analysis. Different methods, such as radial velocity, transit, microlensing, direct imaging, and pulsar timing, can be used to search for exoplanets, and the data from these methods can be combined to increase the chances of detecting a planet. The data must be preprocessed and merged to ensure that it is accurate and consistent, and feature engineering process is used to extract useful information from the data and create new features that can be used in the analysis.

Overall, the transit method is a powerful tool for finding exoplanets, and it is expected to play a major role in future exoplanet discoveries. New space-based telescopes such as TESS and PLATO will be able to detect even more exoplanets, and the increasing sensitivity of ground-based telescopes will make it possible to study the properties of these exoplanets in greater detail. The transit method is highly sensitive to small exoplanets, even those that are smaller than Earth. This is because the transit method detects exoplanets by measuring the tiny dip in a star's brightness that occurs when a planet passes in front of the star. The transit method is less affected by noise and other sources of error than other methods, such as the radial velocity method. This is because the transit method uses the same data to confirm the existence of the exoplanet, where as radial velocity method relies on different data. The transit method can detect multiple exoplanets in a single star system. This is because each exoplanet causes a separate dip in the star's brightness, which can be identified and analysed individually.

5. Conclusion

Finding exoplanets is a complex process that involves multiple steps, including data collection, data preprocessing, data merging, and data analysis. Different methods, such as radial velocity, transit, microlensing, direct imaging, and pulsar timing, can be used to search for exoplanets, and the data from these methods can be combined to increase the chances of detecting a planet. Data preprocessing and data merging are important steps in this process as they ensure that the data is accurate and consistent, and that any analysis performed on it will yield reliable results. Once the data is preprocessed and merged, it can be analysed using various algorithms and statistical methods to search for evidence of exoplanets.

The results of the analysis can then be used to confirm or disprove the presence of exoplanets, and the data can be further refined and analysed as needed. With advancements in technology, the ability to detect exoplanets is increasing, and many new discoveries are expected in the coming years.

DOI: 10.54254/2755-2721/8/20230089

6. Future work

There are several areas of future work that can be done in the search for exoplanets, including: Developing more advanced telescopes and instruments: Advances in technology are continually improving the capabilities of telescopes, which can help to detect smaller, more distant, and more elusive exoplanets. Improving data analysis techniques: New and more sophisticated algorithms, machine learning techniques, and statistical methods will be developed to better analyse the data from telescopes and other instruments, which can increase the chances of detecting exoplanets. Searching for biomarkers: Future research will focus on identifying biomarkers that can indicate the presence of life on exoplanets. This can include looking for signs of water, oxygen, and other gases that are necessary for life as we know it. Overall, the future of exoplanet research looks promising with many new discoveries expected in the coming years as we continue to improve our ability to detect exoplanets and learn more about their properties and potential for habitability.

References

[1] Akeson, R. L., et al. "The NASA exoplanet archive: data and tools for exoplanet research."
[2] Publications of the Astronomical Society of the Pacific 125.930 (2013): 989
[3] J DeVore, John, et al. "On the detection of non-transiting exoplanets with dusty tails." Monthly Notices of the Royal Astronomical Society 461.3 (2016): 2453-2460
[4] Cooke, Benjamin F., and Don Pollacco. "specphot: a comparison of spectroscopic and photometric exoplanet follow-up methods." Monthly Notices of the Royal Astronomical Society 495.1 (2020): 734- 742.
[5] Basak, Suryoday, et al. "Habitability classification of exoplanets: a machine learning insight." The European Physical Journal Special Topics 230.10 (2021): 2221-2251.
[6] Wright, Jason, and Cullen Blake. "Atomically precise sensors could detect a planet much like our own: Earth 2.0: This artist's rendition shows how an Earth-like exoplanet might appear." IEEE Spectrum 58.3 (2021): 22-28.
[7] Breitman, Daniela (2017). “How do astronomers find exoplanets?”. EarthSky.org. Online. Retrieved from https://earthsky.org/space/how-do-astronomers-discover-exoplanets/.
[8] Santos NC. Extra-solar planets: Detection methods and results. New Astronomy Reviews. 2008 Jun 1;52(2-5):154-66.
[9] Brownlee, Jason. "Discover feature engineering, how to engineer features and how to get good at it." Machine Learning Process (2014)
[10] Mishra, Rajeev. "Predicting habitable exoplanets from NASA's Kepler mission data using Machine Learning." (2017).
[11] Moulds, V. E., et al. "Finding exoplanets orbiting young active stars–I. Technique." Monthly Notices of the Royal Astronomical Society 430.3 (2013): 1709-1721.