INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND APPLICATIONS
VOLUME XVI, ISSUE VI, JUNE 22, WWW.IJCEA.COM, ISSN 2321-3469

EXOPLANETS DETECTION USING MACHINE LEARNING

Josiah Joseph, Dhiraj Thakare and Arathi Kamble

Department of Computer Engineering, New Horizon Institute of Technology and Management Thane, India

ABSTRACT:

Kepler's planetary candidate catalog, the first based catalog of all, processed similarly by Kepler's data set. This is the first automated catalog, which uses robots to test robots to accurately measure all the periodic signals received by the Q1 – Q17 Data Release 24 Kepler pipeline. Although we prioritize the same tests over the exact accuracy of each item, we find that our robotic tests are completely the same, and in most cases, the human testing procedures used by previous catalogs. This catalog is the first to use an artificial transport injection to evaluate the effectiveness of our testing procedures and to measure potential bias, which is important in accurately calculating planetary occurrence rates. In terms of the Kepler Object of Interest catalog, we select 1478 new KOIs, 402 of which are classified as PCs. Also, 237 KOIs ranked as FP points in Kepler's previous catalogs changed to PC and 118 PCs changed to FP status. This brings the total number of known KOIs to 8826 and PCs to 4696. Comparing the Q1 – Q17 DR24 KOI catalog with previous KOI catalogs, as well as the corresponding Kepler catalogs, we find a positive agreement between them. We highlight both new PCs that can be rocky and possibly in the host space, many of which revolve around the solar system. This work represents the milestone in the history of the solar system.

The transport system, one of the few methods used to detect exoplanets, detects a periodic decrease (or immersion) in starlight that reflects the movement of the planet in front of the star as seen in the observer's point of view. This approach far surpasses all other exoplanet acquisitions in terms of the number of planets discovered and, as a result, represents our best choice for discovering planetary planets with our current technological capabilities.

In addition, this method allows us to determine the composition of the planet's atmosphere by examining the spectrum of stars with a high resolution of light from a star passing through the upper atmosphere of the planet. A summary of the concept is provided along with some of the results presented for its use.

Keywords: Exoplanets, Kepler, Random Forest, Transportation, Light Therapy & Solar Flux.

[1] INTRODUCTION

Hunting for exoplanets has always been one of the main focus of space research. The need for a habitable planet grows as the Earth warms. Exoplanet is any planet outside of our solar system. Most orbits other stars, but free-floating exoplanets, called planetary orbits, orbiting the galaxy and are not connected to any other star. The Mikulski Archive for Space Telescopes (MAST) makes Kepler's pixel and flux measurements publicly available. Kepler gazed at nearly 156,000 stars in a period of 29.4 minutes as Long Cadence (LC) directed its original goal of discovering the moving planets. The mode of transport depends on taking light of the target star and comparing these variable values with other stars in the same celestial body. When a planet passes in front of a star the marked light diminishes. This fading depends on the size of the planet. This results in a decrease in the observed variable values. [1]

Several algorithms have been developed to effectively identify exoplanets. Machine Learning is one of the methods used to determine whether an object of interest is a certified exoplanet or a false positive. This study aims to understand the effectiveness of the Random Forest algorithm as a split algorithm whether a particular star can handle the exoplanet or not. This study uses Kepler data with a time series label for exoplanet hunting in a deep space. This database contains two files, exoTrain and exoTest. The training set consists of 5087 targets and 3198 columns, column 1 is a labeled vector and column 2 to 3198 values fluctuate over time. The test set contains 570 views and 3198 features. [1]

Exoplanets are planets higher than our solar system. Thousands have been discovered over the past two decades, most notably with NASA's Kepler Space Telescope. These exoplanets come in a variety of shapes and sizes. Some large planets cling to their parent stars; some with snow, some with rocks. NASA and other spacecraft are looking for a special kind of planet: the equivalent of a planet, orbiting a sun-like star in its orbit.

The habitat of the starry heavens is not too hot or too cold for liquid water on the surrounding planets. Some exoplanets may have the necessary physical characteristics (amount and quality of light from a star, temperature, atmospheric composition) in order for complex organic chemistry and possibly the development of Life, (which may be very different from Earth's Life). Nearby Exoplanets are over 4 light years. Scientists have discovered a very effective way to study these phenomena; the planets themselves do not emit light, but the stars around them do. Considering this fact, scientists at NASA developed a method they called the Transit method when using technology such as a digital camera to detect and measure small immersion in starlight as the planet cuts in front of the star. By observing the changing planets, astronomers can calculate the average planet's width at that of its star - basically the size of the planet's shadow - and at that rate, they can calculate the size of the planet. The Kepler Space Telescope's main method of searching for planets was the "Transport" method. [2]

[2] PROBLEM STATEMENT

Exoplanet is a star orbiting star outside our solar system. While orbiting the star, when Exoplanet comes between the earth and the star, the intensity of the star's light decreases as the exoplanet blocks some of its light. We need to find the presence of such exoplanet by measuring the slightest decrease in starlight. [1]

[3] OBJECTIVE

- Detection of exoplanet by measuring the slightest decrease in starlight.
- We find the database on the GitHub website.
- Cleans empty values in data.
- We should make the data normal.
- By building a model, we separate the data into a training and testing database.
- To train the model, we use algorithms – Random Forest Classifier, Fast Fourier

transform, and XG boost classifier.
- We optimize our model and calculate accuracy and loss.
- We test the model using testing dataset.
- Finally, we label the predicted databases showing the presence of exoplanets.

[4] METHODOLOGY

[4.1] FAST FOURIER TRANSFORM:

The first function used is Fourier transform on the dataset. The dataset, as we know, includes flux values. As a result, we transform the time domain values of the most highly sinusoidal waves into frequency domain values. Fourier representation has the advantage of non-linear distribution of data without requiring a kernel function. It also allows for probabilistic interpretation of the division, unlike vector support systems. It can also deal with isolated groups. Unlike Logistic regression, Fourier representation does not require an engineer feature. In general, its computational performance is also very good for large data sets and unlike other algorithms, the common problem of overloading is not detected. The ability of the algorithm to perform multiple class segments by scattered classes and more indirect class distribution is demonstrated. [2]

The transport method involves regularly measuring the brightness of the star to determine the periodic decrease in light associated with the movement of the exoplanet. The move occurs when the planet passes in front of its star. On the other hand, when the planet passes after the star, it is called an eclipse. The impact measured during transit is quite small. In a star the size of the Sun, the movement of the planet Jupiter will cause a light decrease of about 1%, while this decrease will be about 0.001% on a planet the size of Earth.

This approach makes it possible to determine the exact location of a planet and the period of its evolution. In addition, if the planet has already been discovered using the radial velocity method, then its magnitude is known and it is possible to determine the density of the planet's density.

Most of the planets discovered in the form of transit method are identified by a large field survey. The goal is to study as many stars as possible, without having to worry about the future, as there is no indication that the stars will have planets orbiting the Earth so that they can see them in orbit.

It was founded in 2009 and after completing its mission in 2018, the Kepler Space Telescope has played a key role in the search for exoplanets using the space shuttle. It is the only one that has seen 530,000 stars in the constellation Cygnus and confirmed the existence of more than 2,600 exoplanets, altering our view of the exoplanet. These and other planets have shown that there is a great diversity of planets compared to the planets in our solar system.

One of the difficulties of transit method is the geometry of the problem. Indeed, for the exoplanet to fly in our orbit around the Earth, the exoplanet orbiting plane must orbit the Earth. The probability of seeing the movement of a planet around a particular star is closely related to the radius of the star and is also related to the distance between the star and the planet. In the case of a solar star, the probability of observing a transit of a planet the size of the Earth at the distance from the Earth is 0.5%.

[4.2] RANDOM FOREST CLASSIFIER:

Several algorithms have been developed to effectively identify exoplanets. Mechanical Learning is one of the methods used to determine whether an object of interest is a verified exoplanet or a false positive [4]. This study aims to understand the effectiveness of the Random Forest algorithm as a split algorithm whether a particular star can handle the exoplanet or not. We use Precision, Recall, F1 scores, Confusion matrix, Area Under ROC Curve (AUC) and receiver feature curve (ROC) to test the functionality of the Machine Learning Model. This study uses Kepler data with a time series label for exoplanet hunting in a deep space. This database contains two files, exoTrain and exoTest. The training set consists of 5087 targets and 3198 columns, column 1 is a labeled vector and column 2 to 3198 values fluctuate over time. The test set contains 570 visuals and 3198 features. [5]

Database predictions are made using Random Forest Classifier, an Ensemble based Machine Learning algorithm that uses a full database

[4.3] XG BOOST CLASSIFIER:

The following methodology in this project is:

- Preprocessing and data merging.
- Handling missing data.
- Exploratory data analysis.
- Data preparation for machine learning models.
- Defining models.
- Evaluation and optimization.

It is not uncommon for astronomical databases to have large numbers missing, which was the case in this case. So, I start by clearing up these missing numbers.

Speed and performance: The use of advanced decision-making trees designed for speed and performance is a highly competitive machine learning.

[5] SYSTEM ARCHITECTURE

We have learned the principle of how to transit method of finding exoplanets in space. We will create scattering lines and lines to visualize fluctuations in light levels (or volatility values) of light emitted by stars. If there is an occasional immersion in light levels, then we can say that the star has at least one planet. This star is labeled 2.

If there is no clear immersion in light levels, then we can say that the star does not have a planet. In data sets, such stars are labeled 1.

ExoTrain has 5087 rows and 3198 columns. Each line means a star. The LABEL column tells the label or star rating, that is, whether it has at least one planet or at least one planet. The remaining 3197 columns, i.e., columns FLUX.1 to FLUX.3197 contain light levels per star. Light level is a floating point value that can be positive, negative or zero. In the meantime, do not worry about poor light levels.

We will create scatter plots and rows of 3 stars labeled 2 and 3 stars labeled 1. (3)

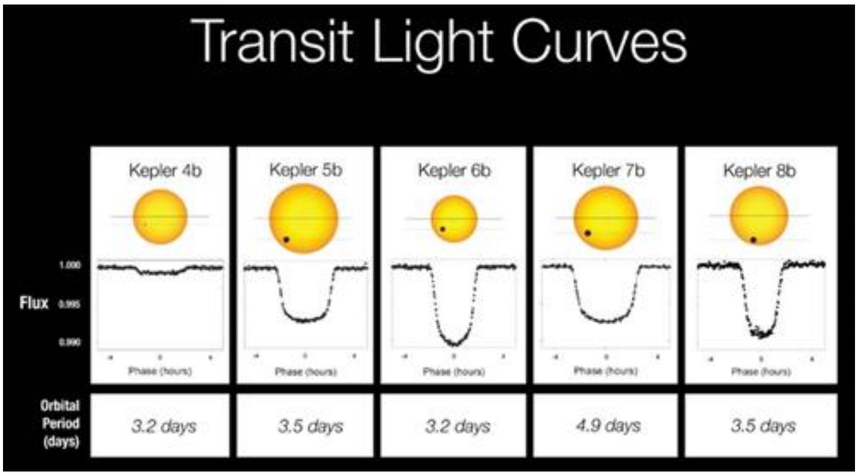

Figure 1. Change in the light intensity due to transit.

[6] IMPLEMENTATION

[6.1] DATASET

We have taken the dataset from the Kepler dataset as it is highly reliable and provides accurate measurements which is needed for accurately predicting the presence of exoplanets. The dataset already was separated into test and train dataset, for the purpose of model testing and evaluation. We have directly used the dataset as it is and didn't do any modifications on it.

[6.2] DATA PRE-PROCESSING

Data obtained from the Kepler database had already been cleared and had no null values. So more cleaning was not needed on the database. However, Fast Fourier Transform was used to convert data into the required format.

[6.3] TRAINING THE MODEL

In our project we aim to predict the presence of exoplanets. In this case we used a random forest classifier and XG boost to achieve the highest levels of accuracy in finding exoplanets. The training database was already in a separate file so we did not have to separate any database.

[6.4] RESULTS

The model ran successfully and had an accuracy of 95%.

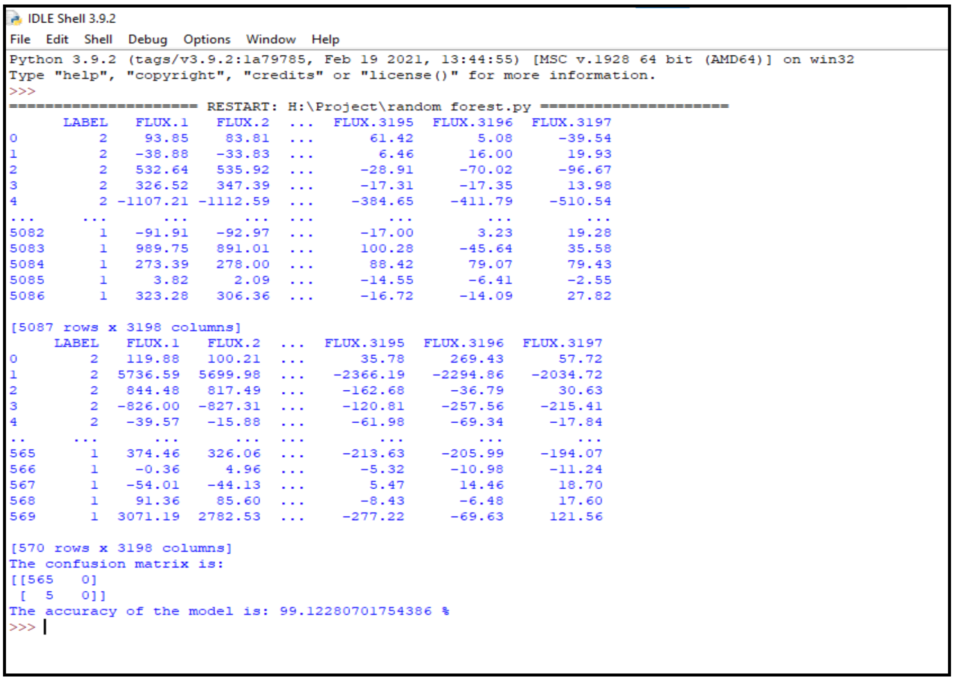

Figure 2. Output of Code

In the more distant future, huge space interferometers will make detailed maps of planets. And possibly, interstellar probes will be launched towards the nearest exoplanets to take close-up images. Engineers are already working on propulsion techniques to reach such distant targets.

[7] CONCLUSION

We can use this model to detect undiscovered exoplanets in faraway galaxies. We can use the ratio of the decrease of the intensities to find the size of the exoplanet using the diameter of the star. With the calculated diameter of the planet, we can find the distance of the exoplanet from the star and hence find whether the exoplanet lies in the goldilocks zone of the star, i.e. where the exoplanet has the possibility of sustaining life.

This project deals with the transit method of detection of exoplanets. This is an attempt to learn how we can use transit photometry to understand the characteristic properties of an exoplanet.

Being a part of this Solar System, there is a limitation to our knowledge about planetary system. Ability to observe other planetary systems can enhance our knowledge about planetary evolution. May be someday with more technological advancements, we will be able to observe the origin of life in an exoplanet or may find another intelligent species. But for all these, first step is to find an exoplanet. After that only we can solve these mysteries and who knows, maybe someday in the distant future Man will step foot on one of those exoplanets.

REFERENCES

[1] Digvijay Patil, Shraavya R. Srinivasarao, Rajat Puri, "Predicting the Presence of Exoplanets in Star-Systems using Random Forest Classifier", International Conference on Computational Intelligence and Digital Technologies (ICCIDT – 2021)

[2] Shivam Pratap Singh1, Devendra Kumar Misra, "International Journal of Research in Engineering, Science and Management". Volume-3, Issue-9, September-2020. journals.resaim.com/ijresm | ISSN (Online): 2581-5792| RESAIM Publishing

[3] Asif Amin, R. M., Talha Khan, A., Raisa, Z. T., Chisty, N., SamihaKhan, S., Khaja, M. S. and Rahman, R. M. (2018). Detection of exoplanet systems in kepler light curves using adaptive neuro-fuzzy system, 2018 International Conference on Intelligent Systems (IS), pp. 66-72.