Predicting Habitable Exoplanets In Different Star Systems Using Machine Learning.

Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING IN

CSE (INTERNET OF THINGS) Project_Group_06

Submitted by:

Darshit Verma (17BCS4633)

Yash Patel (17BCS4635)

Aryan Raj (17BCS4640)

Lucky Verma (17BCS4645)

Rajiv Kumar (17BCS4596)

Under the Supervision of:

Mr. Krishnendu Rarhi

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB , 2021

DECLARATION

I, **'Yash Patel'**, student of **'Bachelor of Engineering in Computer Science (Internet of Things)**, **session:<u>2017-21</u>**, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled **'Predicting Habitable Exoplanets In Different Star Systems Using Machine Learning'** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Yash Patel

Candidate UID: **17BCS4635**

Date: 29 April 2021

Table of Contents

ABSTRACT

Since the beginning of cosmological expansion, the observable universe is estimated to be about the diameter of 93 billion light years. In this vast cosmic space there might be a higher probability of the existence of earth size celestial bodies or similar habitable zones. This popular interest led to embracing the vision to explore our Milky Way galaxy. In 2009, NASA launched the Kepler Space Telescope which studied a small patch of area which consists of about half-million stars. Our proposed way to detect habitability consists of statistical approach by using Machine-learning algorithms on Kepler dataset. Our prime centre is to build the increased set of parameters over which habitability of an exoplanet depends and furthermore utilizing a dataset with an increased set of the star systems which will additionally bring about expanded precision and accuracy of the prediction.

GROUP MEMBERS:

Darshit Verma (17BCS4633)

Yash Patel (17BCS4635)

Aryan Raj (17BCS4640)

Lucky Verma (17BCS4645)

1. INTRODUCTION

The recognisable universe is around 46.5 billion light a very long time across in span and contains around 200 billion to 2 trillion cosmic systems. In this inconceivability of room here's our home cosmic system called The Milky Way. Cosmologists assessed that there are about 250±150 billion stars in the Milky Way alone and stars like the Sun make up approximately 10 percent of all-stars for example around 20±15 billion stars. A rocket was propelled by NASA on March 7, 2009, which was a part of NASA's Discovery Program by a telescope named Kepler Space Telescope to find Earth-size planets circling different star system's.

In the mid-90s, the researcher's began finding planets around other habitable planets. As of late distributed paper "Planetary competitors saw by Kepler VIII" recognized a few stars, called exoplanets, utilizing Doppler spectroscopy, at times called the spiral speed technique, and ordinarily known as the wobble strategy. As of April 2016, 582 exoplanets (about 29.6% of the all-out known at that point) were found utilizing this strategy.

1.1. Software Specification

- Jupyter Notebook / PyCharm
- Python Libraries
- Kepler Dataset

2. LITERATURE SURVEY

Author	***Method***	***Parameter***	***Result***	***Future Outcome***
Rajeev Misra [2]	Support Vector Machine Model (SVM)	Planetary span, Segregation transition, Equilibrium temperature, Orbital Period, Span from the parent star, Stellar temperature, and so forth.	Accuracy=98.1% Precision=14%	Narrow down the habitable planets list more accurately.
Karan Hora [1]	Decision Tree, CART, Random Forest, Logistic Regression, Feed Forward Neural Network, Support Vector Machine, Naive Bayes.	17-Categorical feature, 51-Continuous feature.	Accuracy=99.89 % (CART)	Optimizing model and parameters that has more credibility in determining habitability.
Suryoday Basak [3]	Gaussian Naive Bayes, Linear Discriminant Analysis, SVM, Radial Basis SVM, KNN, Decision tree,	Stellar data hroiparcose database and PHL-EC database.	Accuracy=100% (GBDT)	This work seek towards achieving sustainable and automated discriminated system of exoplanets.

	Random forest, GBDT.			
Snehanshu Saha [4]	Naive Bayes, Linear Discriminant Analysis,, SVM, KNN, Decision Tree.	Parameters from PHL-EC catalogue. e.g.- Eccentricity, Mass, Geo dynamic.	Accuracy=96.46 % (Random Forest)	To discover intrinsic limitation of each learning method in different dataset paradigm.
S. Basak [5]	Neural Network based approach CESSA.	Parameters of PHL-EC catalogue.	Accuracy=97.6% (without fuzzy input) Accuracy=99.5% (with fuzzy input)	This work seek towards improving the prediction model with more dataset of exoplanets in future.
N. Schanche [6]	Random forest, classifier and Convolutional Neural Network.	Time Period of revolution, Width of transit, Number of observed transit,	Accuracy=90%	With increase in data of exoplanets in future we seek to improve accuracy of our model.

		Depth-to-width ratio, Radius of planet.		

2.1 Existing System

This section gives a brief overview of implementation of different machine learning models for predicting habitable and non-habitable exoplanets which exist in literature.

- Rajeev Misra [2] uses a support vector machine model as a method for prediction of habitable exoplanets. The parameters used as a dataset features were Planetary span, Segregation transition, Equilibrium temperature, Orbital Period, Span from the parent star, Stellar temperature, and so forth. The accuracy for the afore mentioned model is 98.1% and precision is 14%.
- Karan Hora [1] uses Decision Tree, CART, Random Forest, Logistic Regression, Feed Forward Neural Network, Support Vector Machine and Naive Bayes as a method for prediction of habitable exoplanets. The parameters used as a dataset features were 17 categorical features and 51 continuous features of habitable exoplanets catalogue containing 1952 exoplanets data[1]. The accuracy achieved using CART is 99.89%.
- Suryoday Basak [3] uses Gaussian naive bayes, LDA, SVM, radial basis SVM, KNN, decision tree random forest and GBDT as methods for predicting habitability of exoplanets. The parameters used are taken from stellar data from hiparcose database and PHL-EC dataset. The accuracy achieved is 100% using GBDT algorithm.
- Snehanshu Saha [4] uses naive bayes, LDA, SVM, KNN and Decision tree as methods for predicting habitability of exoplanets using machine learning model. The parameters used are taken from PHL-EC catalogue for e.g. eccentricity, mass, geo dynamics. The accuracy achieved in this model 96.46% using random forest algorithm.
- S. Basak [5] uses neural network based approach CESSA as the methods for determining habitability of exoplanets. The parameters are take from PHL-EC catalogue. The accuracy achieved is 97.6% without fuzzy input and 99.5% with fuzzy input.
- N. Schanche [6] uses random forest classifier and convolutional neural network as the methods for determining the habitability of exoplanets. The parameters taken into account are time period of revolution, width of transit, radius of planet using mcmc

algorithm, number of observed transit and depth to width ratio. The accuracy achieved in this model is approx. 90%.

2.1Proposed System

When identified, the planet's orbital size can be determined from the period (to what extent it takes the planet to circle once around the star) and the mass of the star utilizing Kepler's Third Law of planetary motion[2]. The size of the planet is found from the profundity of the travel (how much the brilliance of the star drops) and the size of the star[2]. From the orbital size and the temperature of the star, the planet's trademark temperature can be determined. These Planetary and Stellar boundaries are distributed utilizing Kepler's open information repository[2].

Kepler’s field of view covers 115 square degrees. During its over nine and a half years of service, Kepler observed 530,506 stars and detected close to 9,564 “CANDIDATE” exoplanets and more than 2,600 exoplanets has been “CONFIRMED” as exoplanets after scrutinising[2].

The researcher’s has examined these planets and recognized a few planets that might be habitable or reasonable forever. "Planetary Habitability Laboratory" has distributed a rundown of tenable planets [16].

This project undertaking's goal is to use the Kepler dataset as training data and testing data also. The planetary and stellar features act as attributes that help to fabricate an AI model that could foresee conceivably new habitable planets as and when increasingly "CONFIRMED" planets are distributed in Kepler's exoplanet archive[2].

The result of this paper is a computer code that fabricates an AI model that helps in foreseeing the habitability of exoplanets from Kepler exoplanets chronicle data[17].

3. PROBLEM FORMULATION

As our Kepler Mission itself collected a huge amount of exoplanet’s data and this volume of data will keeps on increasing with future exploration missions eg. James Webb Space Telescope (which is proposed to launch on October 2021) and Hubble Telescope (which is currently in orbit). Therefore, in near future there will be explosion of data collected from these space missions and hence implementing statistical analysis on such huge datasets requires complex computational manpower. So, to tackle this problem, we came up with mixed approach of astronomy and machine learning.

Thus, we implemented machine learning algorithm on the Kepler dataset to predict the habitable exoplanets with increased accuracy. Algorithms will classify the planets into 2 classes of : habitable & non-habitable. Implementation has been done using Python programming languages, using relevant packages and libraries. The dataset has been split into training and testing sets, with a ratio of 70:30.

4. METHODOLOGIES

The recognisable universe is around 46.5 billion light a very long time across in span and contains around 200 billion to 2 trillion cosmic systems. In this inconceivability of room here's our home cosmic system called The Milky Way. Cosmologists assessed that there are about 250±150 billion stars in the Milky Way alone and stars like the Sun make up approximately 10 percent of all-stars for example around 20±15 billion stars. A rocket was propelled by NASA on March 7, 2009, which was a part of NASA's Discovery Program by a telescope named Kepler Space Telescope to find Earth-size planets circling different star system's.

In the mid-90s, the researcher's began finding planets around other habitable planets. As of late distributed paper "Planetary competitors saw by Kepler VIII" recognized a few stars, called exoplanets, utilizing Doppler spectroscopy, at times called the spiral speed technique, and ordinarily known as the wobble strategy. As of April 2016, 582 exoplanets (about 29.6% of the all-out known at that point) were found utilizing this strategy.

Later on, different procedures were presented, for example,

· **Direct Imaging:** Direct imaging is an extremely troublesome and restricting strategy for finding exoplanets. Above all else, the star framework must be moderately near Earth. Next, the exoplanets in that framework must be far enough from the star with the goal that astronomers can recognize them from the star's glare.

· **Transit Method:** The travel technique depends on the perception of a star's little drop in brilliance, that happens when the circle (ran line) of one of the star's planets passes ('travels') before the star The measure of light lost- ordinarily somewhere in the range of 0.01% and 1%-relies upon the extents of the star and the planet and the duration of revolution.

· **Micro lensing:** This technique depends on the gravitational power of far off objects to twist and concentrate light originating from a star. As a planet goes before the star comparative with the onlooker (for example makes travel), the light plunges quantifiably, which would then be able to be utilized to decide the nearness of a planet.

5. REFERENCES

1. Hora, Karan. "Classifying Exoplanets as Potentially Habitable Using Machine Learning." ICT Based Innovations. Springer, Singapore, 2018. 203-212.
2. Misra, Rajeev. "Predicting habitable exoplanets from NASA's Kepler mission data using Machine Learning." (2017).
3. Basak, Suryoday, et al. "Habitability classification of exoplanets: A machine learning insight." arXiv preprint arXiv:1805.08810 (2018).
4. Saha, Snehanshu, et al. "Theoretical validation of potential habitability via analytical and boosted tree methods: An optimistic study on recently discovered exoplanets." Astronomy and computing 23 (2018): 141-150.
5. Bora, Kakoli, et al. "Cd-hpf: New habitability score via data analytic modeling." Astronomy and Computing 17 (2016): 129-143.
6. Basak, S., et al. "CEESA meets machine learning: A Constant Elasticity Earth Similarity Approach to habitability and classification of exoplanets." Astronomy and Computing 30 (2020): 100335.
7. Schanche, N., et al. "Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys." Monthly Notices of the Royal Astronomical Society 483.4 (2019): 5534-5547.
8. Agrawal, Surbhi, et al. "A Comparative Study in Classification Methods of Exoplanets: Machine Learning Exploration via Mining and Automatic Labeling of the Habitability Catalog." (2015).
9. Arbesman, Samuel, and Gregory Laughlin. "A scientometric prediction of the discovery of the first potentially habitable planet with a mass similar to earth." PloS one 5.10 (2010).
10. Beatty, Thomas G., and B. Scott Gaudi. "Predicting the yields of photometric surveys for transiting extrasolar planets." The Astrophysical Journal 686.2 (2008): 1302.
11. Barnes, Rory, et al. "Tides and the evolution of planetary habitability." Astrobiology 8.3 (2008): 557-568.
12. Barnes, Rory. "Tidal locking of habitable exoplanets." Celestial Mechanics and Dynamical Astronomy 129.4 (2017): 509-536.
13. Gaidos, E., et al. "Trumpeting M dwarfs with CONCH-SHELL: a catalogue of nearby cool host-stars for habitable exoplanets and life." Monthly Notices of the Royal Astronomical Society 443.3 (2014): 2561-2578.

14. NASA Exoplanet Archive: https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative

15. NASA Kepler Positional Probabilities: https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=koiapp&constraint=pp_host_prob_score%3E0.3

16. Planetary Habitability Laboratory "http://phl.upr.edu/projects/habitable-exoplanets-catalog/data"

17. NASA cumulative data: "https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative"