International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Finding New Earths Using Machine Learning & Committee Machine

Piyush Gawade[1], Akshay Mayekar[2], Ashish Bhosale[3] , Dr. Sanjay Jadhav[4]

[1]Student, Computer Engineering, SIGCE, Navi Mumbai, Maharashtra, India
[2]Student, Computer Engineering, SIGCE, Navi Mumbai, Maharashtra, India
[3]Student, Computer Engineering, SIGCE, Navi Mumbai, Maharashtra, India
[4]Dean - Training and Placement , Professor, Computer Engineering, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India

---***---

Abstract - *Planet identification has typically been a tasked performed exclusively by teams of astronomers and astrophysicists using methods and tools accessible only to those with years of academic education and training. NASA's Exoplanet Exploration program has introduced modern satellites capable of capturing a vast array of data regarding celestial objects of interest to assist with researching these objects. The availability of satellite data has opened up the task of planet identification to individuals capable of writing and interpreting machine learning models. In this study, several classification models and datasets are utilized to assign a probability of an observation being an exoplanet. Since the start of the Wide Angle Search for Planets (WASP) program, more than 160 transiting exoplanets have been discovered in the WASP data. In the past, possible transit-like events identified by the WASP pipeline have been vetted by human inspection to eliminate false alarms and obvious false positives. The goal of the present paper is to assess the effectiveness of machine learning as a fast, automated, and reliable means of performing the same functions on ground-based wide-field transit survey data without human intervention.*

Key Words—Classification Algorithm, Exoplanets, WASP.

1. INTRODUCTION

The task of identifying planets outside of our solar system, known as exoplanets, leads to genuinely novel discoveries. Exoplanet identification has traditionally been a time-intensive task reserved for highly-trained, educated experts with access to specialized—and usually expensive—equipment. These experts relied upon their education, intelligence, diligence, and team knowledge in their painstaking search for exoplanets using images collected by terrestrial observatories and satellite-based telescopes, such as Hubble.

Machine learning techniques have been applied by citizen astronomers to classify objects of interest. One of the more notable examples of this is the work done by Shallue and Vanderberg in their 2011 study (1). Shallue and Vanderberg were two machine learning engineers at Google who trained a neural network model to scour archived data to identify planets using transit events which had gone unnoticed by other researchers (1). The "Autovetter Project" created a Navie Bayes Model to classify objects of interest based on transit data as well (1). In effect exoplanet classification has now been crowd sourced.

Test and train datasets are derived from the labeled observations in the KCOI table. KCOI data contains over eighty columns, or features, collected and preaggregated from Kepler data. This data undergoes cleansing to format the data appropriately for feature selection. Once the most prominent and influential features are identified, the support vector machine is trained, fit, and then used to assign a probability of an observation from the KCOI table being an exoplanet.

One of the satellites is new era modern planet-hunting satellites is the Kepler space telescope which was launched by NASA in 2009. To date, it has been the most successful telescope in the discovery of exoplanets [3]. As of October 2018, Kepler has identified over 9500 objects of interest; with over 2000 of these objects of interest being confirmed exoplanets7 . Kepler excels at identifying Earth-sized planets where past telescopes have only had the power to identify larger "gas giant" planets similar to Jupiter [2]. Kepler targets known stars to seek out exoplanets in that solar system's habitable zone [3]. The Kepler satellite is specifically tuned to detect star brightness [3]. A dip in a star's brightness could indicate one of its planets is passing between the star and the observing telescope. A light curve from the Kepler space telescope with a "U-shaped" dip that indicates a transiting exoplanet.

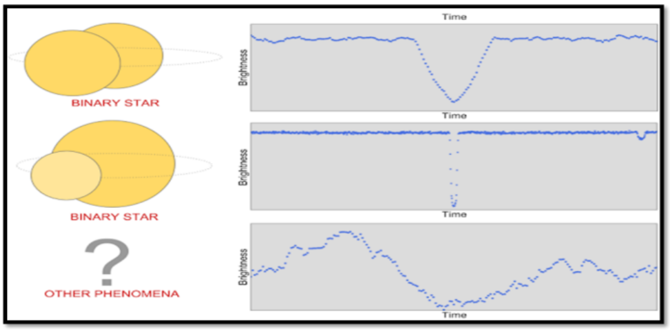


Fig -1: Transit shape

A dip in a star's brightness could indicate one of its planets is passing between the star and the observing telescope. The time it takes for the planet to pass between the start and observing telescope is the transit time and is usually measured in hours. The magnitude of the reduction in brightness and transit time can provide mathematical clues to the relative size and position of the planet relative to its star [2]. Though Kepler was technically a telescope, it is essentially a statistical mission (1). Kepler was purpose-built to collect data to support proven exoplanet identification techniques [2].

2. LITERATURE SURVEY

Exoplanet transit surveys such as the Convection Rotation and Planetary Transits (CoRoT; Auvergne et al. 2009), Hungarian-made Automated Telescope Network (HATnet; Hartman et al. 2004), HATSouth (Bakos et al. 2013), the Qatar Exoplanet Survey (QES; Alsubai et al. 2013), the Wide-angle Search for Planets (WASP; Pollacco et al. 2006), the Kilodegree Extremely Little Telescope (KELT; Pepper et al. 2007), and Kepler (Borucki et al. 2010) have been extremely prolific in detecting exoplanets, with over 2,900 confirmed transit detections as of August 9, 20181 . The majority of these surveys employ a system where catalogue-driven photometric extraction is performed on calibrated CCD images to obtain an array of light curves.

Following decorrelation of common patterns of systematic error (eg Tamuz et al. (2005)), an algorithm such as the BoxLeast Squares method (Kov´acs et al. 2002) is applied to all of the lightcurves. Objects that have signals above a certain threshold are then identified as potential planet candidates. Before a target can be flagged for follow-up observations, the phase-folded light curve is generally inspected by eye to verify that a genuine transit is present. As these surveys contain thousands of objects, the manual component quickly becomes a bottleneck that can slow down the identification of targets. Additionally, even with training it is difficult to establish consistency in the validation process across different observers. It is therefore desirable to design a system that can consistently identify large numbers of targets more quickly and accurately than the current method. Several different techniques have been used to try to automate the process of planet detection.

A common method is to apply thresholds to a variety of different data properties such as signal-to-noise ratio, stellar magnitude, number of observed transits, or measures of confidence of the signal, with items exceeding the given threshold being flagged for additional study (For WASP-specific examples, see Gaidos et al. (2014) and Christian et al. (2006)). Applying these criteria can be a fast and efficient way to find specific types of planets quickly, but they are not ideal for finding subtle signals that cover a wide range of system architectures. Machine learning has quickly been adopted as an effective and fast tool for many different learning tasks, from sound recognition to medicine (See, e.g., Lecun et al. (2015) for a review). Recently, several groups have begun to use machine learning for the task of finding patterns in astronomical data, from identifying red giant stars in asteroseismic data (Hon et al. 2017) to using photometric data to identify quasars (Carrasco et al. 2015), pulsars (Zhu et al. 2014), variable stars (Pashchenko et al. 2018; Masci et al. 2014; Naul et al. 2017; Dubath et al. 2011; Rimoldini et al. 2012), and supernovae (du Buisson et al. 2015). For exoplanet detection in particular, Navie Bayes Classifiers (McCauliff et al. 2015; Mislis et al. 2016), Artificial Neural Networks (Kipping & Lam 2017).

Convolutional Neural Networks (Shallue & Vanderburg 2018), and Self-Organizing Maps (Armstrong et al. 2017) have been used on Kepler archival data. Convolutional Neural Networks were trained on simulated Kepler data by Pearson et al. (2018). While Kepler provides an excellent data source for machine learning (regular observations, no atmospheric scatter, excellent precision, large sample size), similar techniques can also be applied to ground-based surveys, and in fact machine learning techniques have recently been incorporated by the MEarth project (Dittmann et al. 2017) and NGTS (Armstrong et al. 2018). The work of highly skilled astrophysicists or other researchers can be redirected towards more specialized exoplanet research. Light curves vary greatly, even at a large scale. Planets closer to the sun like Venus, are too hot to support life as we know it; while planets further out like Mars and beyond thought are too cold. Using a combination of transit time and other measurements collected by Kepler.

3. PROPOSED SYSTEM

3.1.Dataset:

Our main focus in this project is to analyze the features extracted from Dataset .which are required by windows loader. This contains various elements like size of code, size of data, overlay number. With help of this one can understand how a program is going to execute.

3.1.1.Training Set:

- 5087 rows or observations.
- 3198 columns or features.
- Column 1 is the label vector. Columns 2–3198 are the flux values over time.
- 37 confirmed exoplanet-stars and 5050 non-exoplanet-stars.

3.1.2.Dev Set:

- 570 rows or observations.
- 3198 columns or features.
- Column 1 is the label vector. Columns 2–3198 are the flux values over time.
- 5 confirmed exoplanet-stars and 565 non-exoplanet-stars.

3.2.Classification Algorithm:

For predicting exoplnets existence we aim at using three classifiers. By using different classification algorithm we can get different results.

3.2.1.Naive Bayes (Nb) Classifier:

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.It is mainly used in text classification that includes a high-dimensional training dataset.Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

Formula :-P(A/B)=(P(B/A)P(A))/P(B)

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

3.2.2.Support Vector Machine (Svm):

Support vector machines exist in different forms, linear and non-linear. A support vector machine is a supervised classifier. What is usual in this context, two different datasets are involved with SVM, training and a test set. In the ideal situation the classes are linearly separable. In such situation a line can be found, which splits the two classes perfectly. However not only one line splits the dataset perfectly, but a whole bunch of lines do. From these lines the best is selected as the "separating line".

3.2.2. Artificial Neural Network:

Neural Network is built by stacking together multiple neurons in layers to produce a final output. First layer is the input layer and the last is the output layer. All the layers in between is called hidden layers. Each neuron has an activation function. Some of the popular Activation functions are Sigmoid, ReLU, tanh etc. The parameters of the network are the weights and biases of each layer. The goal of the neural network is to learn the network parameters such that the predicted outcome is the same as the ground truth.

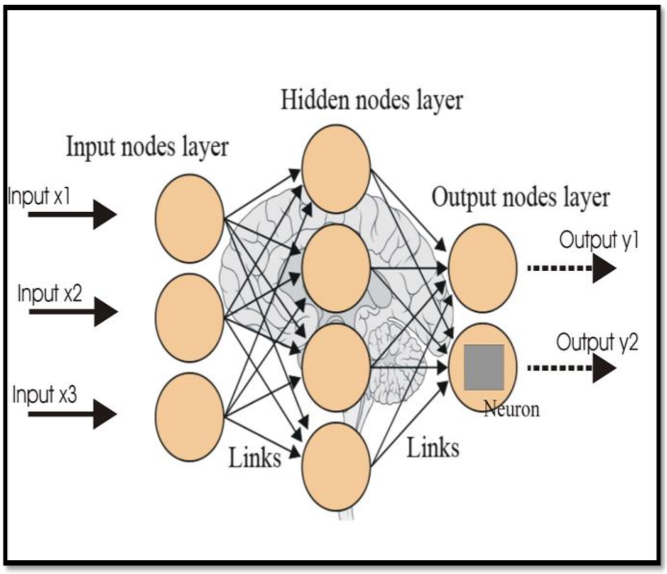
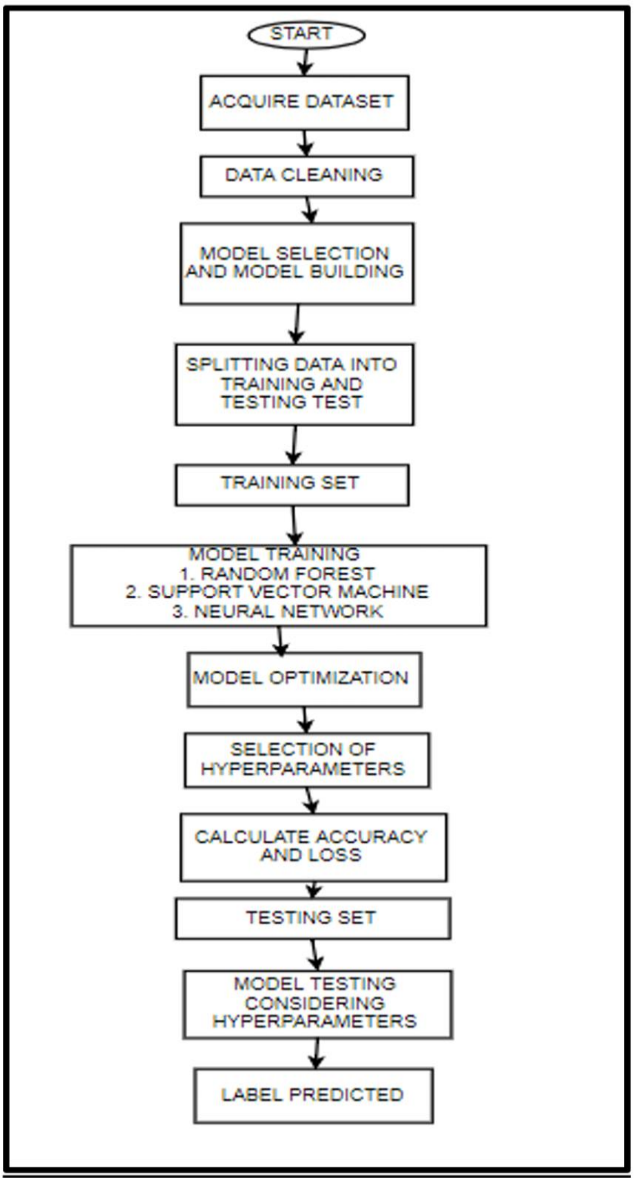


Fig -2: Multilayer Artificial Neural Network

3.3. Flowchart:



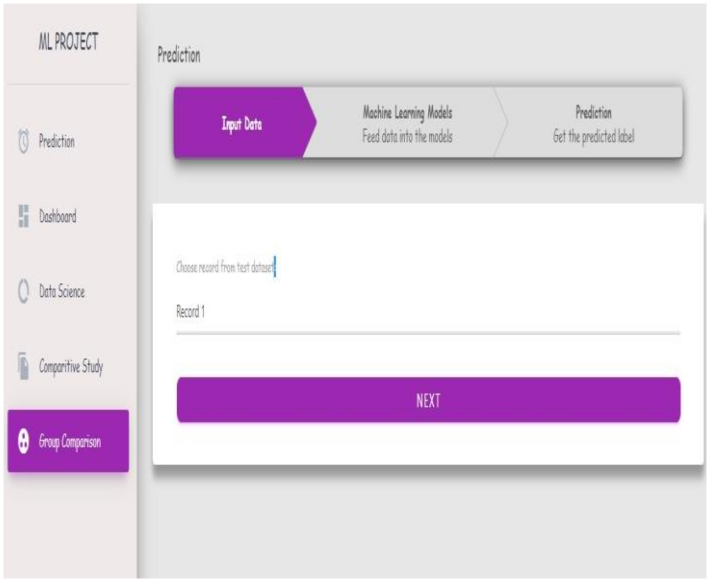
4. RESULT & CONCLUSION

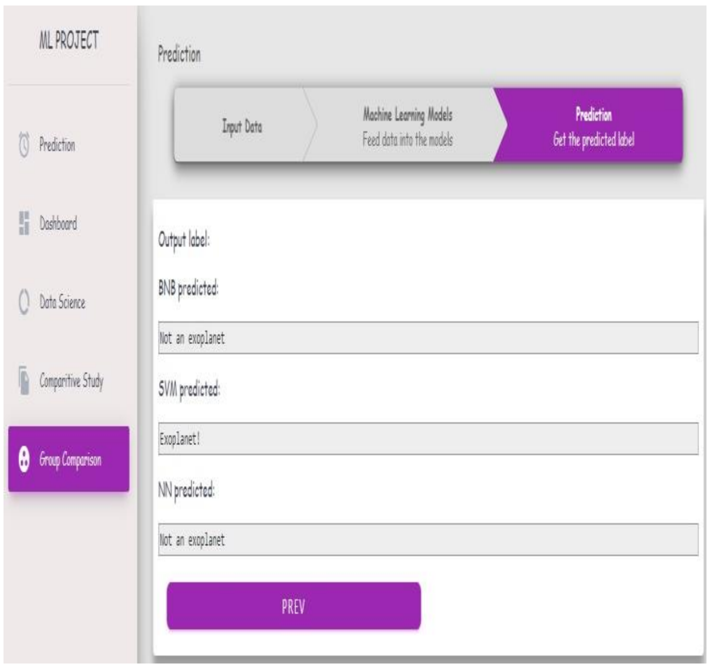
Using multiple machine learning models is an effective framework that can be modified and applied to a variety of different large-scale surveys in order to reduce the total time spent in the target identification and ranking stage of exoplanet discovery. Combining the results from additional machine learning methods could further improve the predictions. An additional advantage of this approach is that the algorithms can be quickly re-trained as new information, such as new known classifications from completed follow-up observations, become available.

It has proven to be very effective in producing new candidates for future follow-up and eventual planet status. The large size of the WASP archive makes it undesirable for human observers to manually look at each one to determine whether it is a good candidate for further study. The machine-learning framework we have created provides a tool for the observer wanting to re-examine the full set of data holdings in any WASP field, enabling fast re-classification of all targets showing transit-like behavior and identification of new targets of interest.

Website Snapshots

After entering user ID and password.





ACKNOWLEDGEMENT

The success and final outcome of this research required a lot of guidance and assistance and we are extremely privileged to have got this all. All that we have done is only due to such supervision and assistance and we would not forget to thank them.

We respect and thank DR. Sanjay Jadhav, for providing us insight and expertise that greatly assisted the research. We are extremely thankful to her for providing such a nice support and guidance.

REFERENCES

[1] http://kepler.nasa.gov/.

[2] Cassan, A., D. Kubas, J.-P. Beaulieu, M. Dominik, K. Horne, J. Greenhill, J. Wambsganss, J. Menzies, A. Williams, U. G. Jorgensen, A. Udalski, D. P. Bennett, M. D. Albrow, V. Batista, S. Brillant, J. A. R. Caldwell, A. Cole, Ch. Coutures, K. H. Cook, S. Dieters, D. Dominis Prester, J. Donatowicz, P. Fouque, K. Hill, J. Kains, S. Kane, J.- B. Marquette, R. Martin, K. R. Pollard, K. C. Sahu, C. Vinter, D. Warren, B. Watson, M. Zub, T. Sumi, M. K. Szymanki, M. Kubiak, R. Poleski, I. Soszynski, K. Ulaczyk, G. Pietrzynski, and L. Wyrzykowski. "One or More Bound Planets per Milky Way Star from Microlensing Observations." Nature 481.7380 (2012): 167-69. And https://www.space.com/19103-milky-way-100-billion-planets.html

[3] https://exoplanetarchive.ipac.caltech.edu/index.html

[4] http://archive.stsci.edu/kepler/

[5] http://fits.gsfc.nasa.gov/
and http://en.wikipedia.org/wiki/FITS.

[6] http://www.astropy.org/.

[7] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[8] http://www.numpy.org/

[9]http://archive.stsci.edu/kepler/manuals/archive_manual.pdf.

[10]http://kepler.nasa.gov/education/activities/transitTracks/

[11]http://en.wikipedia.org/wiki/Discoveries_of_exoplanets

BIOGRAPHIES

Piyush R. Gawade, Pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi College of Engineering (SIGCE), Navi Mumbai. Her current research interests include Web Designing & Machine Learning

Akshay P. Mayekar, Pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi College of Engineering (SIGCE), Navi Mumbai. Her current research interests include Image Processing & IoT.

Ashish Bhosale, Pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi College of Engineering (SIGCE), Navi Mumbai. Her current research interests include Image Processing & Web development.

Dr. Sanjay Jadhav, Dean - Training and Placement at Smt.Indira Gandhi College Of Engineering, Ghansoli, Navi Mumbai. He is also Secretary of Mahashtra TPO Association & Former Head, Training & Placement at Saraswati College and Former Assistant Professor at Bharati Vidyapeeth's College of Engineering, Navi Mumbai.