

Enhancement of Stellar Enlightenment on Star Categorization Using Radial Velocity

Ch. Sandeep¹, Martha Sheshikala², Chennuri Srijanya³, Mummaneni Charishma⁴

^{1,2,3,4}School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India.

Abstract—This research focuses on creating and assessing a machine learning model that hope to improve our knowledge of the universe by connecting the detection of planetary systems with the categorization of celestial bodies through the combination of these efforts. A more comprehensive approach to exploring the vastness of space is made possible by the synergy between spectral analysis and flux monitoring, which throws light on the nature of distant objects as well as the possibility of alien planetary systems.

INTRODUCTION

Exoplanet exploration is a global phenomenon that is discussed and studied at various academic levels and in a variety of scholarly domains. Exoplanets have emerged as the poster child for multidisciplinary collaborative work in the obvious field of science, which includes astronomy, astrobiology, biology, astrophysics, geology, and planetary science, to mention a few prominent participants. However, other academic disciplines like philosophy and theology are also involved. Exoplanet research is popular and widely regarded because of our long-standing curiosity about the night sky. The age-old human query, "Are we alone?" probably originated when we first raised our heads to the night sky and began to wonder.

The transit technique has proven successful in the majority of exoplanet discoveries to date. Exoplanets are located by taking exact photometric measurements from Earth or space, then waiting and observing to spot the slight, distinctive dip in the light curve that occurs when an exoplanet crosses its parent star. The planet's semi-major axis, orbital period, and size can all be estimated from transit observations. As a result of the planet's area relative to its host star, the planet's size is directly determined by the depth of the transit. If the host star is a part of a bound or line-of-sight close stellar pair, this straightforward plan may become complicated, such as its radius and surroundings, are essential to comprehending and describing any exoplanets that are found. An ongoing and crucial aspect of exoplanet research is hence the collection of exact star characteristics through ground-based follow-up observations.

Numerous isolated transit occurrences, probably resulting from massive, far-off exoplanets, have been identified by space missions like Kepler. Single-transit exoplanets suggest planets with longer orbital periods in the outer system, which may be connected to smaller inner exoplanets that have been found before in the same system. Even while a single transit cannot reveal every intricate element of the orbit, the statistical makeup of a group of far-off, massive planets might be a helpful indicator for comprehending worlds outside of our solar system.

Customary techniques for locating exoplanets are Another useful technique for finding exoplanets is the discovery of their radial velocity (RV). Although currently restricted to planets with masses greater than Earth, this method offers data that would not be accessible otherwise. Doppler observations can yield the mass (or mass limit) of the planet as well as the characteristics of its orbit. The upcoming generation of huge ground-based telescopes and new instrumentation will enable RV to discover more exoplanets and significantly reduce the mass limit that is currently heading toward Earth-mass objects.

Through the classification resultsthe original approach for categorizing stars' low efficiency and accuracy can be improved by this research, which can also teach how various training models perform on various objects, which can be used to select the best classifier model for an ensemble learning algorithm. It is known that the binary percentage of exoplanet host stars is approximately 46%. This results in binary star systems where planets are found orbiting one of the stars, sometimes leaving uncertainty about which star the planet is orbiting. We speculate that the presence of a distant star companion in certain hot Jupiter systems could play a crucial role in the equation. These are but a handful of the many fascinating topics regarding the architectures of planetary systems that remain unanswered. Research programs concentrating on exoplanet creation, dynamics, and evolution are taking on the challenge of providing answers to these problems, as theoretical and observational study moves from simple detection to detailed characterization.

I. RELATED WORK

Numerous strategies have been proposed and published in various journals. Every study we mentioned in this section takes a different angle on the problems. Below, a variety of existing styles are discussed.

[1]The Kepler Space Telescope observed a specific region of the Milky Way to discover Earth-sized exoplanets in or near the habitable zone of their parent stars. It has identified thousands of exoplanet candidates. One of Kepler's primary goals was to identify Earth-sized exoplanets located within or near the habitable zone of their parent stars. The habitable zone is the region around a star where conditions might be suitable for liquid water to exist on the surface of a planet, making it potentially conducive to life as we know it.

[2]Several projects focus on studying the atmospheres of exoplanets to determine their composition, temperature, and potential habitability. Techniques such as transmission spectroscopy, emission spectroscopy, and phase curve analysis are employed to characterize exoplanet atmospheres. The Hubble Space Telescope and upcoming missions like the James Webb Space Telescope (JWST) are crucial for these studies.

[3]Exoplanet Characterization: Efforts are underway to characterize the physical properties of exoplanets, including their masses, radii, densities, and orbital dynamics. This information helps scientists understand the formation and evolution of planetary systems. The Kepler Follow-up Program and ground-based observatories play important roles in these efforts.

[4]Researchers investigate the potential habitability of exoplanets by studying factors such as surface conditions, atmospheric composition, and the presence of liquid water. The field of astrobiology explores the possibility of life beyond Earth and considers the implications of exoplanet discoveries for the search for extraterrestrial life.

[5]The training and evaluation sets were chosen from the Q1-Q17 DR24 Autovetter Planet Candidate Catalog. They compared models based on linear logistic regression and fully connected neural networks and achieved 95% recall, 90% accuracy, and 96% precision. Automated Triage and Vetting of TESS Candidates” by Liang Yu et al [14] to Identify exoplanets with Deep Learning technique. The deep neural network was developed and tested using actual TESS data, and it has an average precision and accuracy of 97.0% and 97.4% for identifying transit-like signals from instrumental noise in triage mode.

[6] There have been previous attempts to automatically classify Super WASP variable stars (SVS) from the raw data using machine literacy styles and artificial neural networks, frequently known as neural networks (NN). These NNs only incompletely succeeded, rightly detecting 75 of the light angles.

[7]Stellar Activity and its Implications for Exoplanet Habitability” by Guinan et al. (2019): Investigates how stellar activity, including flares and stellar winds, can affect the habitability of exoplanets by influencing their atmospheres and surface condition data. Delicacy, resistance to outliers, and a comparatively high position of forbearance to features with minimum or spare class information. Using cross-validation analyses, we achieved bracket edge of 80.7, 82.7, and 84.5%.

[8]Characterization of Stellar and Exoplanetary Environments” by Sing et al. (2016): Reviews the techniques and challenges associated with characterizing stellar atmospheres and their impact on exoplanetary observations, including transmission spectroscopy.

[9] "Implications of Stellar Binarity on Exoplanet Formation Mechanisms" by Moe & Kratter (2018): Investigates how the presence of binary companions influences the formation mechanisms and architectures of exoplanetary systems, considering dynamical interactions and migration processes.

[10] "The TESS" is examined in this essay. The information from the VSX (The International Variable Star Index) can be updated and verified using data from short exposure (single target, 2 min exposure) and high delicacy observations made by a spacecraft. variable star, which traditionally provided periodic observations. For this

investigation, the TESS-recorded variable information for each star in sector 21 was obtained. periodogram system, which was used for data processing.

[11] The goals of this study included creating a predictive model to categorize stars, planets, and quasars as well as evaluating several eye-catching categorization schemes to determine which was the most fashionable. The stylish scheme had a delicacy score of 0.8664, an across-validation delicacy score of 0.87, and weighted average perfection, recall, and f-1 scores of 0.87.

II. METHODOLOGY

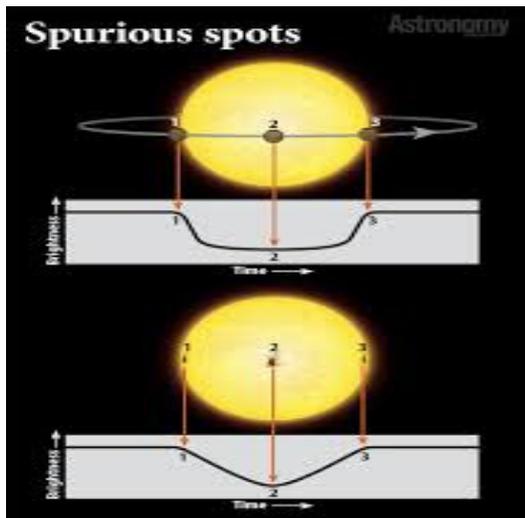
In our model, we focused on the division of the stars.

The data which is taken from the SDSS (Sky Digital Solar Survey) as our primary source. To enhance the dataset's quality and diversity, we employed various preprocessing and augmentation techniques. Initially, we collected the data from online sources. On the light curves of stars, our technique applies feature extraction, and wavelet modification, to get fully accurate results. By cross-matching observations from several space-based observatories with databases of known binary and exoplanet stars, we have also assembled a dataset of binary and exoplanet stars for training and validation.

The data describe the change in flux (light intensity) of several thousand stars. Each star has a binary label of 2 or 1. 2 indicates that the star is confirmed to have at least one exoplanet in orbit; some observations are multi-planet systems.

As you can imagine, planets themselves do not emit light, but the stars that they orbit do. If said star is watched over several months or years, there may be a regular 'dimming' of the flux (the light intensity). This is evidence that there may be an orbiting body around the star; such a star could be considered to be a 'candidate' system. Further study of our candidate system, for example by a satellite that captures light at a different wavelength, could solidify the belief that the candidate can in fact be 'confirmed'.

In the above diagram, a star is orbited by a blue planet. At $t = 1$, the starlight intensity drops because it is partially obscured by the planet, given our position. The starlight rises back to its original value at $t = 2$. The graph in each box shows the measured flux (light intensity) at each time interval.



Light Curve

Fig. 1.

The light curve created by an exoplanet passing in front of its star (top) is different from the one caused by a starspot. These subtle inequalities tell astronomers which is which.

Even though every item on your list produces starlight drops as it passes through its stellar host, an exoplanet's signal is frequently very defined.

As they revolve in and out of view, starspots can take up the same amount of space as an exoplanet, or they can be larger or smaller. They differ from a planet transit, though, in a few ways.

Although spots lack round forms and sharp edges, they do arise with every star revolution and have periods that can resemble planets. However, as the star spins, spots move in both latitude and longitude; as a result, after a few rotations, their period changes slightly. Additionally, stars with spots typically exhibit bigger amplitude changes in their light. Stars with spots also tend to have larger amplitude variations in their light, which are usually planet transit depth and lead to a sine-like light curve.

Because of their darkness as well as the host star's limb darkening, which occurs when the star's center is brighter than its periphery, exoplanet transits are typically U-shaped. Additionally, they often maintain the same period (exoplanets in extremely elliptical orbits being the exception). Given that we know the radius of the star, we may calculate the size of the planet based on the depth of its transit.

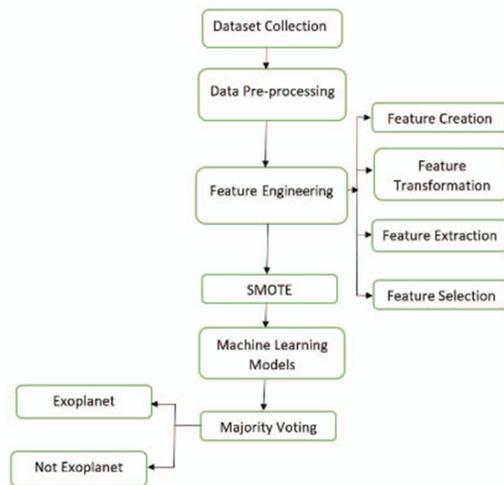


Fig.2. Architecture of Proposed Model

A. DataSet

The flux (light intensity) variation of many thousand stars is described by the data. The binary designation for each star is either 1 or 2. 2 showed that at least one exoplanet is confirmed to be in orbit around the star; however, some sightings are multi-planetsystems. As you might expect, planets circle stars that emit light; planets themselves do not. Observations of the star over a period of months or years may reveal a consistent 'dimming' of the flux or intensity of light. This suggests that the star could have an orbiting body; such a star could be referred to as a "candidate" system. Additional research on our potential system, such as using a satellite that records light at adifferent wavelength, could solidify the belief that the candidate can be 'confirmed'.

A binary label is given to each candidate system: 1 if it is still a candidate pending more confirmation and 2 if at least one exoplanet is proven to be in orbit. Comprehensive observations have occasionally revealed complete multi-planet systems circling a single star in addition to individual planets.

Further research is required to confirm the existence of exoplanets and improve our comprehension of these far-off worlds. Subsequent research employing cutting-edge technology, like as space-based observatories that can record light at various wavelengths, can offer vital proof to validate the existence of exoplanets in candidate systems. Astronomers continue to find new exoplanets and broaden our understanding of the universe beyond our cosmic neighborhood by merging data from several sources and using advanced analytical tools.

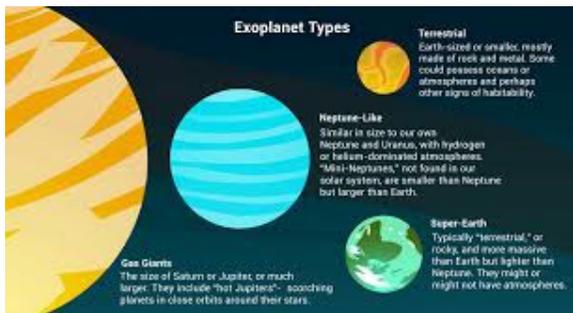


Fig.3. Exoplanet types show the exoplanet to the solar planet

III. IMPLEMENTATION

The data above was collected from each survey finding the new data points that are existing, and which are above to diminish the exoplanet or non-exoplanet.

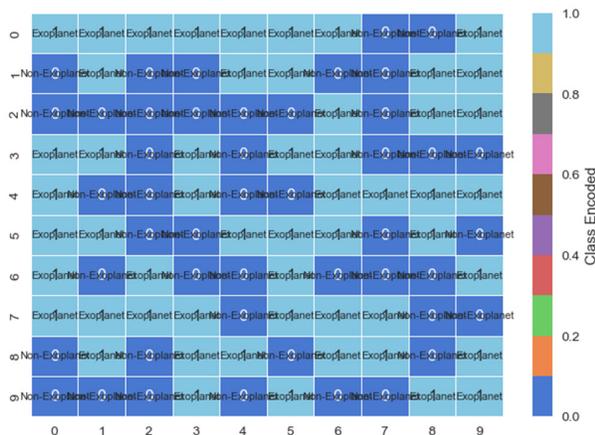


Fig.4. Visual image of all features of the data

The above image i.e., (Fig.4.) is the representation of the attribute class which consists of target variables "Exoplanet", "Non-Exo Planet"

The celestial sphere is a fictitious sphere with a vast radius that revolves around the Earth. All celestial objects can be thought of as being projected onto the celestial sphere's inner surface, which can be centered on Earth or the observer.

All the maps are different from each other all the values are taken randomly from the dataset. Each class has a legitimate colormap assigned to it, and photos from that class are displayed using this colormap. Grid Layout: A 2x5 grid is used to organize the photographs. A mapping between classes and colormaps is defined in the code. Each class has a unique colormap assigned to it, such as LABEL 1 with "Viridis," LABEL 2 with "plasma," Image Display: A grid of images is created by the code. There are ten randomly selected photographs in this sample. Each image's colormap is determined by its related class, and the photos are shown in a 2x5 grid.

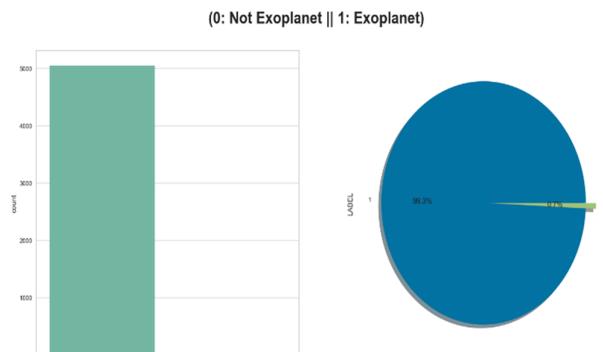


Fig.5. Distribution of Target Variable of each Label division features of data.

creates a count plot (bar plot) of the 'LABEL' column in the DataFrame df. It counts the occurrences of each unique value ('0' for Not Exoplanet and '1' for Exoplanet) and displays them as bars. The palette parameter sets the color palette for the bars. ax=ax[0] specifies that this plot should be placed in the first subplot.

Here 99.3% belongs to division 1 which is exoplanet and likewise, the remaining 0.7% belongs to division 2 which is non-Exo Planet.

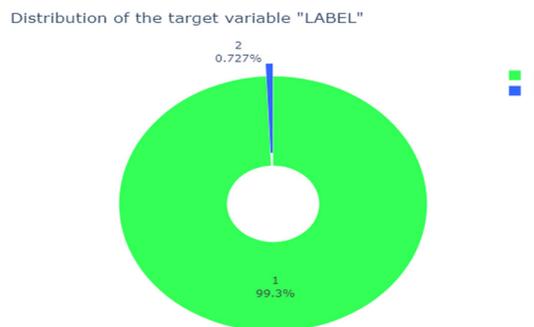


Fig.6. Pie chart

The above Fig.6 pie chart shows the distribution of the target variable "LABEL" which has Exo planet and non-exoplanet

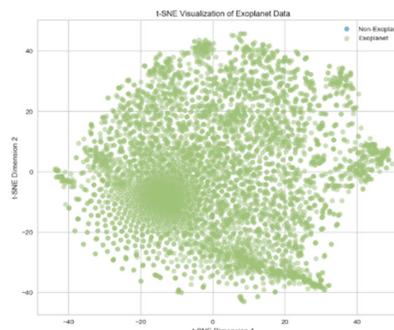


Fig.7. TSNE t-SNE (t-distributed Stochastic Neighbor Embedding) chart

The above plot shows that (Fig.7.) In a t-SNE chart, each data point represents an original high-dimensional data point, and the distance between points in the t-SNE space reflects their similarity in the original high-dimensional space. Points that are similar in the original space are often clustered together in the t-SNE plot. chart is a visualization technique used for dimensionality reduction. It helps in visualizing high-dimensional data in lower-dimensional space, typically 2D or 3D, while preserving the local structure of the data points as much as possible. The above figure shows the exoplanet visualization in 2d format.

IV. FEATURE EXTRACTION:

As the model is based on the Multiclass classification data suitable features should be extracted to train the model. In this project, we used advanced visualization techniques to extract and understand the data based on the appropriate features.

Decision Boundary Visualization: Decision boundary visualization is a technique used to visualize how a machine learning model separates different classes or categories in a dataset. The decision boundary represents the dividing line or surface that separates the points belonging to different classes in the feature space.

For binary classification problems, the decision boundary is typically a line in 2D space or a plane in 3D space. In higher-dimensional spaces, the decision boundary can be a hyperplane. The goal of the machine learning model is to learn the decision boundary that best separates the classes based on the training data.

Visualizing the decision boundary can provide insights into how the model makes predictions and how well it generalizes to unseen data. It helps in understanding the model's behavior and performance, as well as identifying regions of uncertainty.

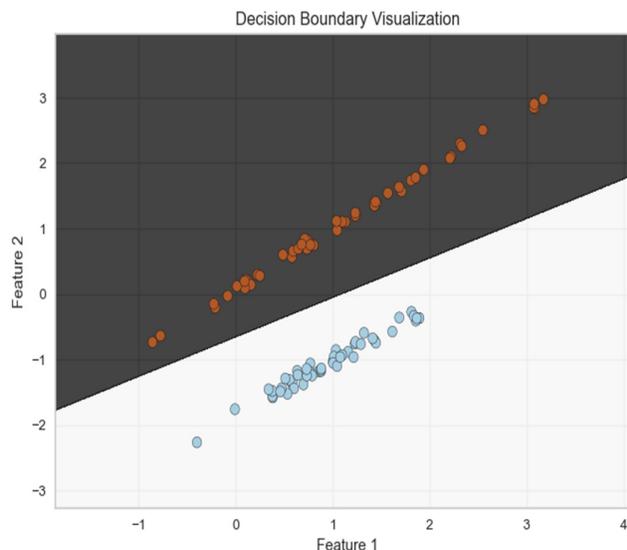


Fig.8. DB Visualization of target variable features are taken with the target variable and fixed the range of random here feature1 is non-exoplanet and feature 2 is exoplanet on the Decision Boundary visualization

To visualize the decision boundary, the code defines a meshgrid covering the range of feature values. It predicts the class for each point in the meshgrid using the trained classifier. Finally, it plots the decision boundary and the data points on a contour plot.

As a 2-D figure Using three or more numerical factors displayed A radar chart, often called a spider chart or a kite chart, is a graphical depiction of multivariate data on axes that begin from the same point. Radar charts are incredibly useful when you want to visualize and compare multiple variables across numerous categories or groupings.

The data is huge consisting of 3198 columns WC cannot use all the data that are not useable and can be removed from feature extraction techniques.

Boxplot of Outliers:

A box plot, also known as a box-and-whisker plot, is a graphical representation of the distribution of a dataset. It displays several important statistical properties, including the median, quartiles, and potential outliers.

Box plots are useful for visually summarizing the distribution of a dataset and identifying potential outliers. They provide insights into the central tendency, spread, and skewness of the data. Additionally, they are particularly effective for comparing the distributions of multiple groups or categories within a dataset.

Whiskers: The lines extending from the box, called whiskers, show the range of the data outside the interquartile range. By default, whiskers extend to the minimum and maximum values within a certain range from the quartiles.

Outliers: Individual data points that fall beyond the whiskers are considered outliers and are plotted as individual points. They can indicate unusual or extreme values in the dataset.

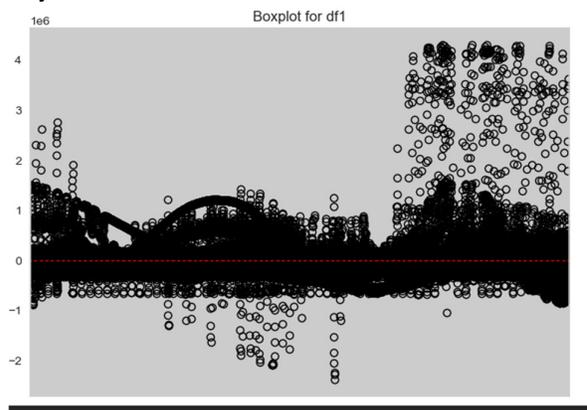


Fig.9 Boxplot of all the data points

There are many outliers in the exoplanet dataset where there are 3198 columns and 50000 rows and there are need to be removed to get the best accuracy for the model to evaluate.

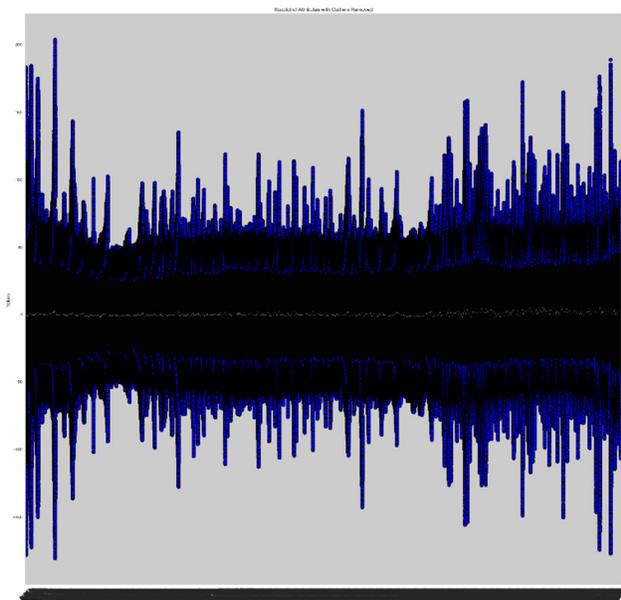


Fig.10 Box-and-Whisker plot

The explained defines a function that removes outliers that take a DataFrame as input and removes outliers from it using the interquartile range (IQR) method. Here's a summary of the steps:

Calculate Quartiles and IQR: The function calculates the first quartile (Q1), third quartile (Q3), and the interquartile range (IQR) for each column in the data frame.

Define Lower and Upper Bounds: Based on the IQR, lower and upper bounds.

Overall, this code aims to visually summarize the distribution of attributes in the data frame after removing outliers, making it easier to identify central tendencies and variability in the data.

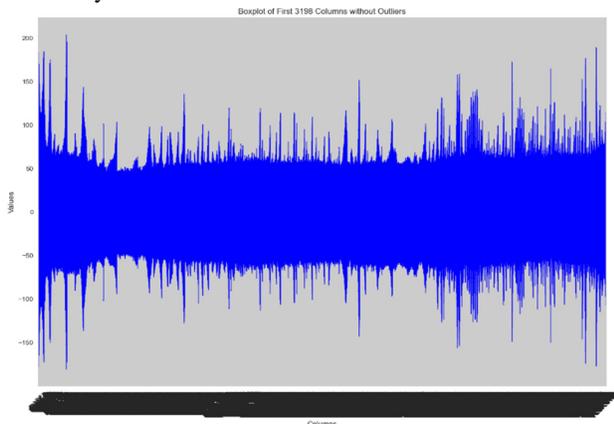


Fig.11 Box- and- Whisker plot without outliers

Distributions Of Flux:

At the outset, the code sets the size of the plot to ensure optimal visualization. A title is added to the plot, indicating its purpose of displaying the distribution of flux values in the dataset. The x-axis is labeled as "Flux values," denoting the values being measured, while the y-axis is labeled as "Flux intensity," representing the intensity of the flux values.

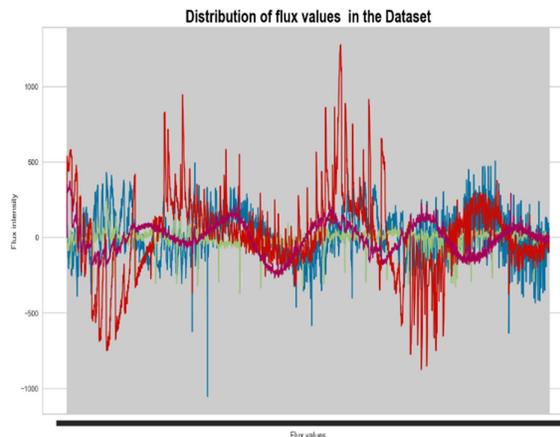


Fig.12.Distribution of Flux values of the dataset

The title of the plot is set as 'Distribution of flux values in the Dataset', with a font size of 20, bold weight, and black color. The x-axis label is 'Flux values', and the y-axis label is 'Flux intensity', providing context to the plot.

Plotting Lines: The code plots the flux values for each row in the DataFrame. It uses `.iloc[]` to access rows by index. The lines are plotted using `plt.plot()`, where each line represents the flux values of a specific row in the DataFrame.

Visualization: By plotting each row separately, the plot provides a visual representation of how flux values vary across different observations or samples in the dataset. The lines show the intensity of flux values over some measurement or observation range.

Display: Finally, the plot is displayed using `plt.show()`, allowing users to visualize the distribution of flux values effectively.

Overall, this plot helps in understanding the overall trend and variability of flux values across different observations in the dataset.

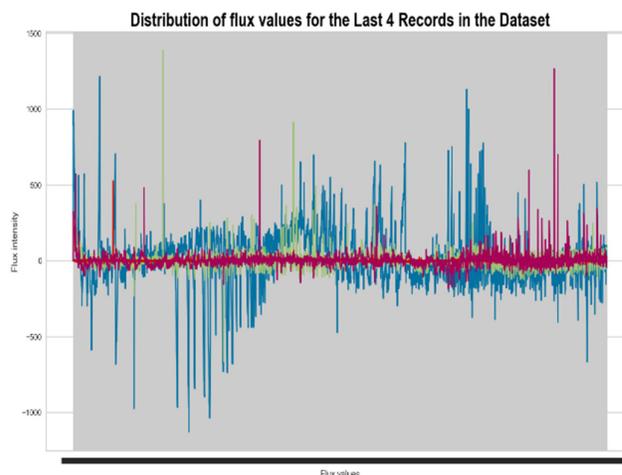


Fig.13Distribution of 4 records in datasets

The main functionality of the code involves plotting the flux values for each of the last four records in the dataset using a line plot. This is achieved by indexing the DataFrame to

select the rows corresponding to these records and plotting their flux values as separate lines on the plot.

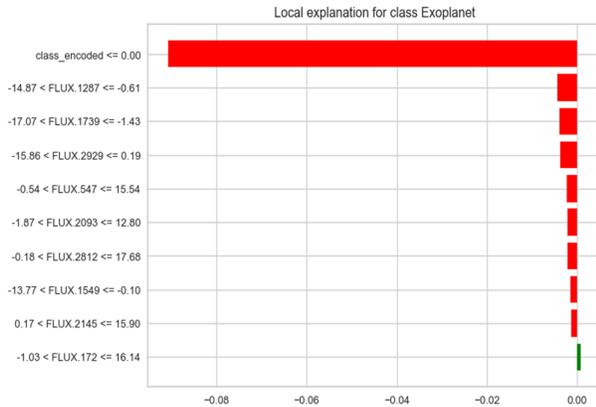


Fig.14 LimeTabularExplainer of the Data Points

demonstrates how to use LIME to provide interpretable explanations for individual predictions made by a machine learning model, helping to understand the model's behavior and predictions on specific instances.

IME generates an explanation for the selected instance using the explaininstance method. This process involves perturbing the instance's features and observing the corresponding changes in the model's predictions to approximate local behavior around the instance.

The explanation provides insights into which features contributed the most to the model's prediction for the chosen instance and how they influenced the prediction.

Transit Photometry Simulation

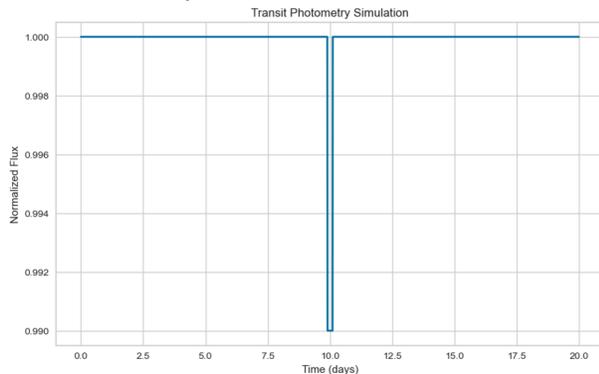


Fig.15TPS of all Data Values

The flux of the star is initially set to 1 (normalized flux) for all time points. During the transit event, the flux decreases due to the planet blocking a portion of the star's light.

The start and end times of the transit event are calculated based on the transit_duration and mid_transit_time.

Within the transit duration, the flux is reduced based on the ratio of the planet's area to the star's area.

Plotting:

The simulated flux data (flux) is plotted against time using Matplotlib.

The transit event is depicted as a dip in the flux curve, representing the decrease in brightness during the planet's passage in front of the star.

The plot is labeled with appropriate axes labels and a title, providing context for the simulation.

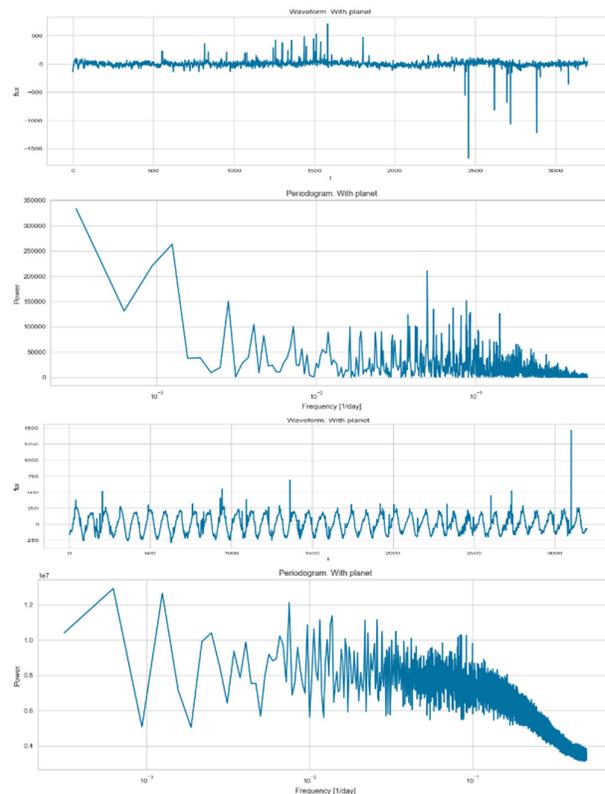


Fig.16 Periodogram of the flux values

A periodogram is a tool commonly used in signal processing and time series analysis to identify periodic components or frequencies within a signal. In the context of machine learning (ML), a periodogram can be applied to analyze periodic patterns in time series data, which can then be used for various ML tasks such as classification, regression, or anomaly detection.

The periodogram is calculated using a mathematical technique called the Fourier transform or its variants, such as the Lomb-Scargle periodogram for unevenly sampled data.

The periodogram represents the power spectral density of the signal, indicating the strength of periodic components at different frequencies.

Higher peaks in the periodogram indicate dominant frequencies or periodic patterns in the data.

Additional features derived from the periodogram, such as the amplitude or phase of dominant frequencies, can also be extracted and incorporated into the ML model.

Max power at frequency: 0.0016

Corresponding period 625.0 days

Max power to median power ratio 146.74

Max power at frequency: 0.0006

Corresponding period 1666.7 days

Max power to median power ratio 2.43

V. MODELIMPLEMENTATION:

a) Support Vector Classification (SVC)

Support Vector Classification (SVC) is a supervised learning algorithm used for classification tasks in machine learning. It belongs to the family of Support Vector Machines (SVM), which is widely used for both classification and regression tasks. Here's an overview of SVC:

Initializes a Support Vector Classifier (SVC) with a linear kernel using SVC(kernel='linear').

Fits the SVC model to the scaled training data (train_scaled) and corresponding target variable (trainy['LABEL']) using svcclassifier.fit(). Uses the trained SVC model to predict the labels for the scaled testing data (test_scaled) using svcclassifier.predict() and assigns the predictions to y_pred.

Evaluation:

Computes and prints the confusion matrix and classification report using confusion_matrix() and classification_report() from sklearn.metrics.

```
[[5050  0][ 31  6]]
Precision recall f1-score support 1
0.99 1.00 1.00 5
050
2 1.00 0.16 0.28
37
accuracy 0.99 5087
macro avg 1.00 0.58 0.64
5087
weighted avg 0.99 0.99
0.99 5087
```

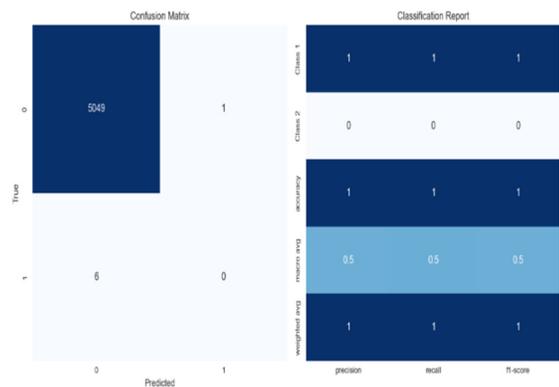


Fig.17. Confusion matrix and f1 score of data points

Here (Fig.17)The classification report is calculated using the classification_report() function from sklearn.metrics. The results are then visualized using another heatmap generated with Seaborn. Both visualizations offer insights into the model's accuracy, precision, recall, and other performance metrics for each class in the classification problem. They help in evaluating the model's strengths and weaknesses and identifying any potential issues, such as class imbalance or misclassifications.

b) Multi-Layer Perceptron Classifier:

MLPClassifier stands for Multi-Layer Perceptron Classifier. It is a type of artificial neural network model that belongs to the class of feedforward neural networks. Here's a brief explanation:

Multi-Layer Perceptron (MLP): MLP is a type of artificial neural network that consists of multiple layers of nodes, including an input layer, one or more hidden layers, and an output layer. Each node in a layer is connected to every node in the subsequent layer. MLPs are often used for supervised learning tasks, including classification and regression.

Classifier: In the context of machine learning, a classifier is an algorithm that takes input data and assigns it to one of several predefined classes or categories based on features extracted from the data

Parameters: The MLPClassifier allows you to specify various parameters, such as the number of hidden layers, the number of nodes in each hidden layer, the activation function used in each layer, the optimizer for training, and regularization techniques like L2 regularization.

Training: To train an MLPClassifier, you provide it with a labeled dataset consisting of input features and corresponding class labels. The model learns to map input features to class labels by adjusting its internal parameters during training.

Confusion Matrix:

```
[[1009  1]
 [  8  0]]
```

Classification Report:precision recall f1-score support

```
1 0.99 1.00 1.00 1010
2 0.00 0.00 0.00 8
accuracy 0.99 1018
macro avg 0.50 0.50 0.50
1018
weighted avg 0.98 0.99 0.99
1018
```

```
from sklearn.metrics import accuracy_score
```

```
# Calculate accuracy
accuracy = accuracy_score(V_y, predictions)
# Print accuracy
print("Accuracy:", accuracy)
Accuracy: 0.9911591355599214
```

VI. RESULTANALYSIS

The Receiver Operating Characteristic (ROC) curve is a graphical representation of the performance of a classification model across different thresholds. It illustrates the trade-off between the true positive rate (TPR), also known as sensitivity, and the false positive rate (FPR), which is equal to 1 minus specificity.

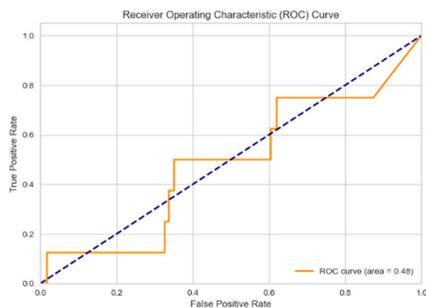


Fig.18 Receiver Operating Characteristics (ROC) of SVM of TPR and FPR.

A curve that hugs the top-left corner of the plot indicates a model with high sensitivity and low FPR, implying excellent performance.

A curve that follows the diagonal (from bottom-left to top-right) suggests a random classifier with no predictive power.

A curve that falls below the diagonal indicates a model that performs worse than random guessing.

VII. CONCLUSION

Machine learning methods have seen very active development in the last decade and now they are an essential part of our way of working. In fact, for everyone who interacts with a computer or smartphone, it is highly likely to interact with some sort of machine learning program. They are also widely used in sciences for several use cases such as detecting diseases, creating new samples without long simulations, and finding string models in a string theory landscape etc. In astronomy, we have seen several applications such as galaxy zoo, predicting galaxy morphology, identifying gravitational waves, and gravitational lensing with new and advanced telescopes. Data in astronomy are growing at a fast pace. Conventional methods that involve human judgements are not efficient and prone to variability depending on the investigating expert. For example, the commonly used method for exoplanet detection, *bls*, produces a large number of false positives in case of noisy data, which has to be reviewed manually.

SVC, and MLP may be able to improve the accuracy of individual identification tasks. Overall, the study's findings show that using multiple machine learning models and majority voting can help with individual recognition tasks by increasing accuracy levels. The present study employed various techniques, including exploratory data analysis, feature engineering, SMOTE, and ensemble learning, to create a highly accurate machine learning model. The study emphasized the significance of preparing the data and using appropriate methods to handle imbalances in the data. The ensemble learning technique was found to be effective in enhancing the model's performance. In the future, researchers could explore alternative feature engineering and ensemble learning techniques to further improve model accuracy. Moreover, research efforts could concentrate on developing new strategies to address imbalanced data, which could lead to even more robust model

REFERENCES

[1] Bugueno, M., Mena, F., & Araya, M. (2018). Refining exoplanet detection using supervised learning and feature

engineering. In 2018 XLIV Latin American Computer Conference (CLEI) (pp. 278-287). IEEE.

[2] Jara-Maldonado, M., Alarcon-Aquino, V., Rosas-Romero, R., Starostenko, O., & Ramirez-Cortes, J. M. (2020). Transiting exoplanet discovery using machine learning techniques: a survey. *Earth Science Informatics*, 13, 573-600.

[3] Schanche, N., Cameron, A. C., Hébrard, G., Nielsen, L., Triaud, A. H. M. J., Almenara, J. M., ... & Wheatley, P. J. (2019). 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), 5534-5547.

[4] Armstrong, D. J., Gamper, J., & Damoulas, T. (2021). Exoplanet validation with machine learning: 50 new validated Kepler planets. *Monthly Notices of the Royal Astronomical Society*, 504(4), 5327-5344.

[5] Priyadarshini, I., & Puri, V. (2021). A convolutional neural network (CNN) based ensemble model for exoplanet detection. *Earth Science Informatics*, 14, 735-747.

[6] Jin, Y., Yang, L., & Chiang, C. E. (2022). Identifying exoplanets with machine learning methods: a preliminary study. *arXiv preprint arXiv:2204.00721*.

[7] Dong, G., & Liu, H. (Eds.). (2018). *Feature engineering for machine learning and data analytics*. CRC Press.

[8] Matchev, K. T., Matcheva, K., & Roman, A. (2022). Unsupervised Machine Learning for Exploratory Data Analysis of Exoplanet Transmission Spectra. *The Planetary Science Journal*, 3(9), 205. SDSS sources without spectra. *Astron. Astrophys.* 2020, 639, A84.

[9] Ofman, L., Averbuch, A., Shlisselberg, A., Benaun, I., Segev, D., & Rissman, A. (2022). Automated identification of transiting exoplanet candidates in NASA Transiting Exoplanets Survey Satellite (TESS) data with machine learning methods. *New Astronomy*, 91, 101693.

[10] Aydoğan, Koray. "Exoplanet Detection by Machine Learning with Data Augmentation." *arXiv preprint arXiv:2211.15577* (2022). 121. Authorized licensed use limited to: VIT-Amaravathi campus. Downloaded on October 05, 2023 at 10:05:35 UTC from IEEE Xplore. Restrictions apply.

[11] Malik, Abhishek, Benjamin P. Moster, and Christian Obermeier. "Exoplanet detection using machine learning." *Monthly Notices of the Royal Astronomical Society* 513.4 (2022): pp- 5505-5516.

[12] Sturrock, G.C., Manry, B. and Rafiqi, S., 2019. Machine learning pipeline for exoplanet classification. *SMU Data Science Review*, 2(1), p.9.

[13] Shallue, C.J. and Vanderburg, A., 2018. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and a eighth planet around kepler-90. *The Astronomical Journal*, 155(2), p.94.

[14] Yu, L., Vanderburg, A., Huang, C., Shallue, C.J., Crossfield, I.J., Gaudi, B.S., Daylan, T., Dattilo, A., Armstrong, D.J., Ricker, G.R. and Vanderspek, R.K., 2019. Identifying exoplanets with deep learning. III. Automated triage and vetting of TESS candidates. *The Astronomical Journal*, 158(1), p.25.

[15] Chintarunguangchai, P. and Jiang, G., 2019. Detecting exoplanet transits through machine learning techniques with convolutional neural networks. *Publications of the Astronomical Society of the Pacific*, 131(1000), p.064502