1. **Abstract**

This study looks at NASA's exoplanet exploration from 1992 to 2023 using a dataset of 5250 observations from 11 detection techniques. It investigates detection method patterns throughout discovery years, offering insights into these approaches' performance in identifying certain planet types and uncovering historical trends in exoplanet discoveries. Notably, Neptune-like exoplanets and gas giants have been identified regularly, demonstrating the sensitivity of detection technologies to their unique properties. The exploration demonstrates NASA's progress in locating planets with longer orbital periods while highlighting the difficulties in detecting those with shorter periods. Correlations between planetary properties and Chi-square statistical validation highlight strong correlations between planet kinds and detection techniques. Furthermore, machine learning models—Random Forest, KNN, and SVM—show excellent prediction skills for categorizing planet types based on particular traits, indicating interesting pathways for future research in exoplanetary research.

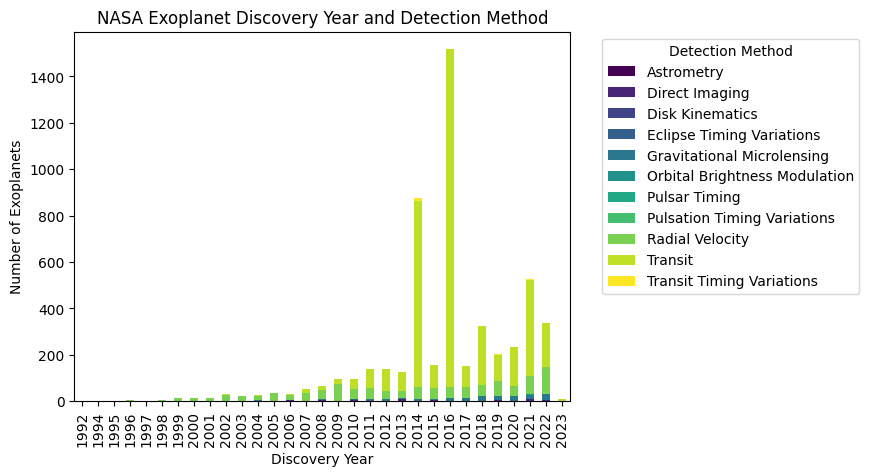
1. **Data Description**

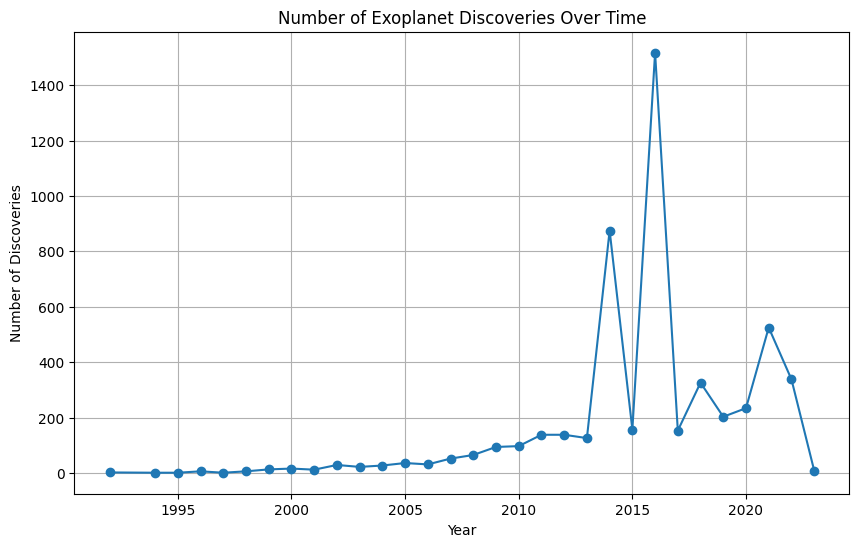
NASA's search for exoplanets includes a wide range of planetary bodies with different properties. Located at varying distances from Earth, these exoplanets display a spectrum of stellar magnitudes that enhance NASA's all-encompassing comprehension of planetary systems. The dataset sheds light on the diversity of planetary compositions by providing information about the exoplanet types, including their mass and radius multipliers with respect to Earth. NASA's ongoing endeavors are chronicled through the discovery years, which demonstrate the agency's dedication to expanding our understanding of far-off planets. Understanding the orbital dynamics of these exoplanets is vital for understanding their individual planetary systems, and information about eccentricity, orbital radius, and orbital period can be found here.The detection methods add to the technological sophistication of NASA and demonstrate the agency's ability to identify and study exoplanets using a variety of techniques, ultimately expanding our understanding of the universe.

The dataset comprises of 5250 observations and 12 features + Name column

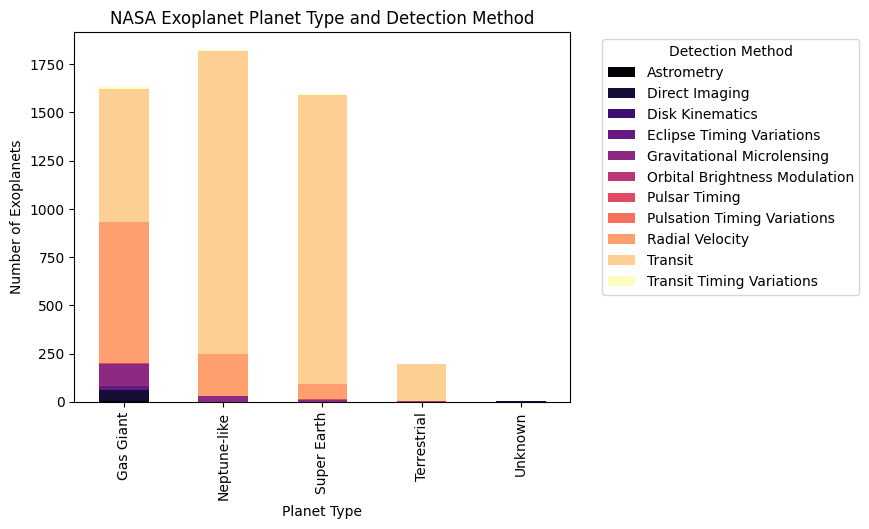
| Feature Name | Description |
| --- | --- |
| Name | Name of the planet as per given by NASA |
| Distance | Distance of the planet from earth in light years |
| Stellar Magnitude | Brightness of the planet, the brighter the planet the lower number is assigned to the planet |
| Planet Type | Type of the planet, these types are derived from our solar system planets |
| Discovery Year | Year in which planet got discovered |
| Mass Multiplier | Mass multiplier of the planet with mass\_wrt planet |
| Mass w.r.t | Mass of the planet in comparison with the mass of planets of our solar system |
| Radius Multiplier | Radius multiplier of the planet with radius\_wrt planet |
| Radius w.r.t | Radius of the planet in comparison with the radius of planets of our solar system |
| Orbital Radius | Orbital radius of planets orbiting around their sun (in AU) |
| Orbital Period | Time in years taken by those planets to complete 1 orbit of their star |
| Eccentricity | Eccentricity basically tells how circular is the orbiting path, eccentricity close to 0 means orbiting path is nearly circular |
| Detection Method | Method used by NASA to find that exoplanet |

1. **SMART Questions**
   1. Can we determine any patterns in the detection methods used for exoplanets and their respective discovery years? Are certain methods more effective at finding specific types of planets?
   2. What is the historical trend in the discovery of exoplanets based on their detection methods, and what can this tell us about the progress and limitations of our observational techniques over time?
   3. How does the discovery year of exoplanets impact the significance of their orbital periods in the context of planetary formation theories?
   4. Are there any discernible patterns or trends in the orbital characteristics of planets, such as orbital radius, orbital period, and eccentricity, and how might these relate to their host stars?
   5. What is the relationship between the distance of exoplanets from their host stars and their planetary characteristics, such as mass and radius?
   6. How do the orbital radii and orbital periods of exoplanets relate to their eccentricity, and how might this information aid in uncovering patterns in orbital dynamics within planetary systems?
2. **EDA**



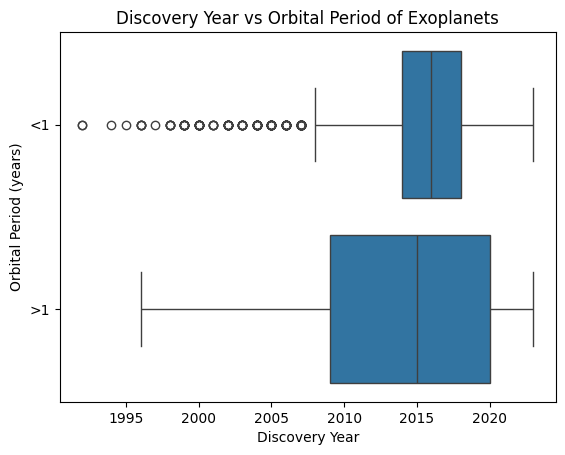


From 1992 until 2023, a total of 11 detection methods were used in the search for exoplanets. Among these methods, Transit stands out as the most productive, resulting in the identification of 1453 exoplanets in 2016 and continually contributing to important discoveries in subsequent years. Surprisingly, 2016 appears as the outstanding year for exoplanet discovery, with the largest number of fresh discoveries. Disk Kinematics, on the other hand, is mentioned as the least used detection approach, reflecting its comparatively limited application in the search for exoplanets during this time period.

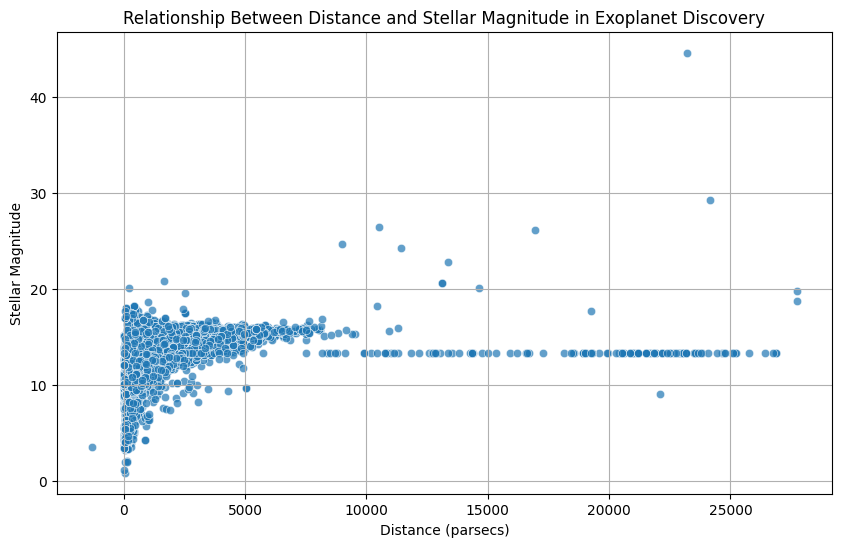


In the continuous discovery for planets outside our solar system, Neptune-like exoplanets have proven numerous. Compared to other exoplanet categories, scientists have discovered and identified a greater number of Neptune-like exoplanets. This frequency might be explained by the sensitivity of existing detection techniques to the traits and signals linked to planets similar to Neptune, which makes them easier to find in the vastness of space.

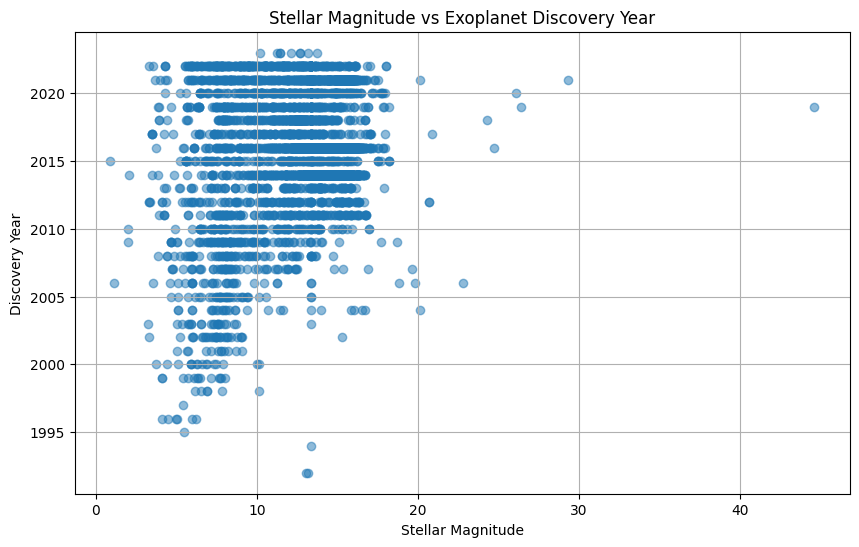
Among the major targets of discovery in the race to locate and comprehend the diversity of exoplanets, gas giants have equally been. These large planets are typically comprised mostly of hydrogen and helium, and researchers have used a range of detection techniques to locate and describe them.



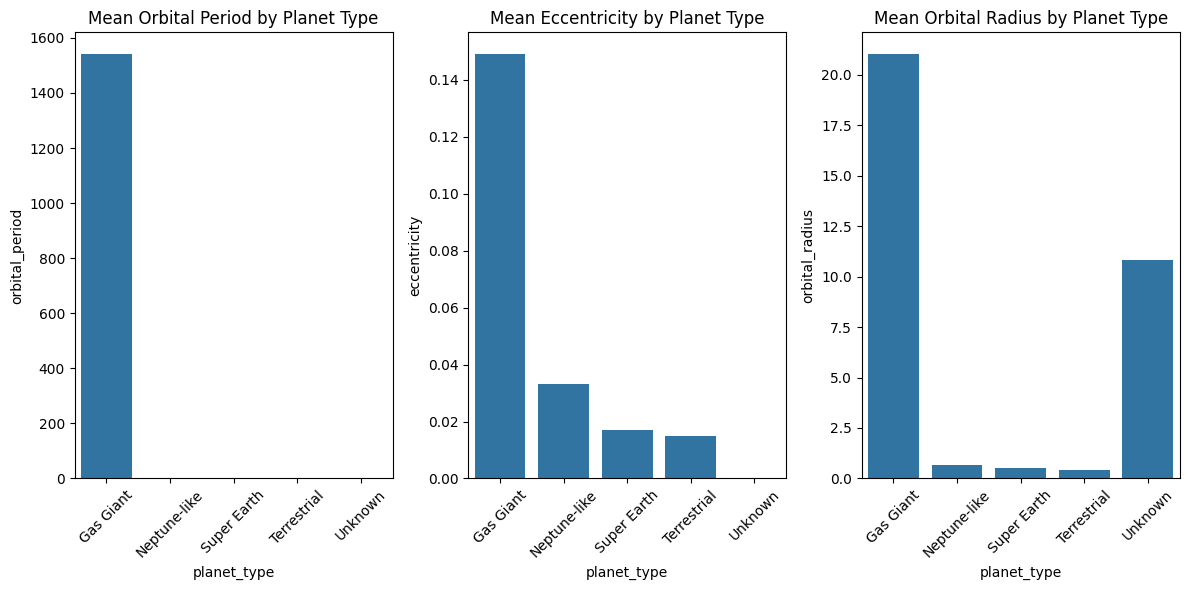
NASA has made significant progress in identifying and characterizing exoplanets that have orbital periods longer than a year, which makes them comparatively simple to detect. Numerous such exoplanets have been found as a result of their efforts, significantly improving our knowledge of distant planetary systems. There haven't been many discoveries in this category, though, because finding exoplanets with orbital periods less than a year has proven to be more difficult. Exoplanets with orbital periods longer than a year seem to be more diverse than those with shorter periods, highlighting the need for greater research to solve the mysteries surrounding both long- and short-period exoplanets in our vast cosmos.



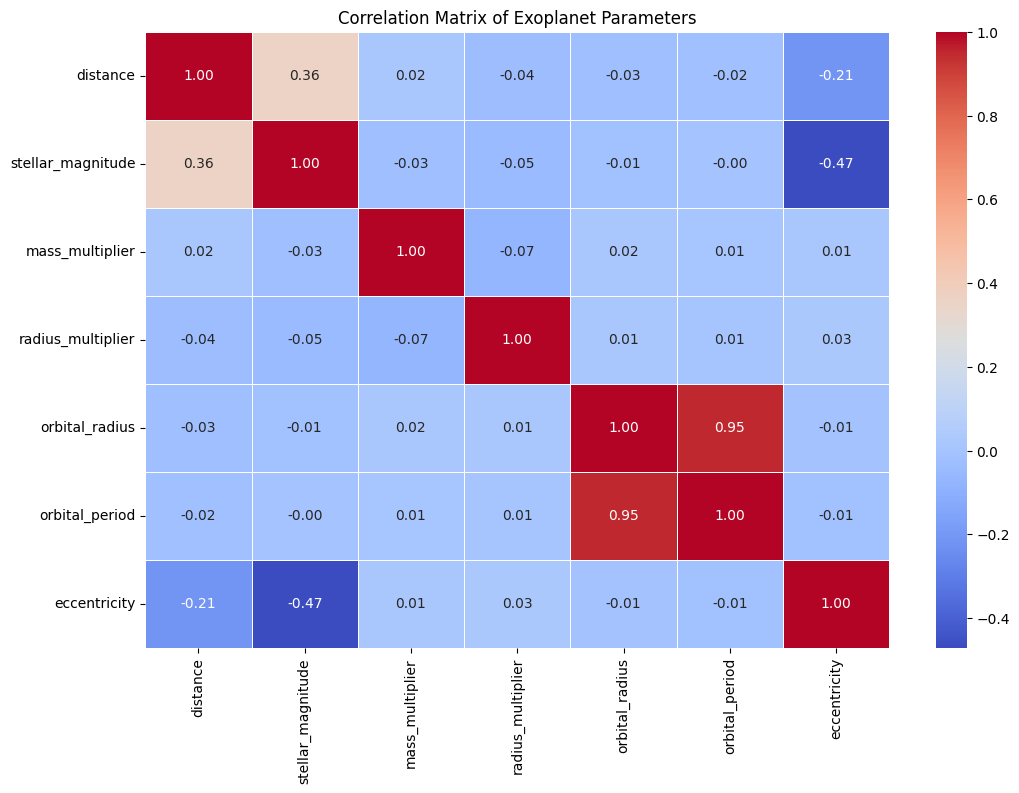
In general, exoplanets that are closer to Earth and have smaller stellar magnitude values seem brighter in the night sky. This is due to the fact that both their host stars' intrinsic brightness and their distance from Earth affect how brilliant they are. Interestingly, though, some exoplanets are able to retain their brightness in spite of their great distance from Earth. This phenomena implies that the perceived brightness of these far-off celestial bodies depends on variables other than proximity. Remarkably, a large fraction of exoplanet discoveries have been discovered nearer to Earth, and many of them have smaller stellar magnitude values, which enhances their visibility in the night sky.



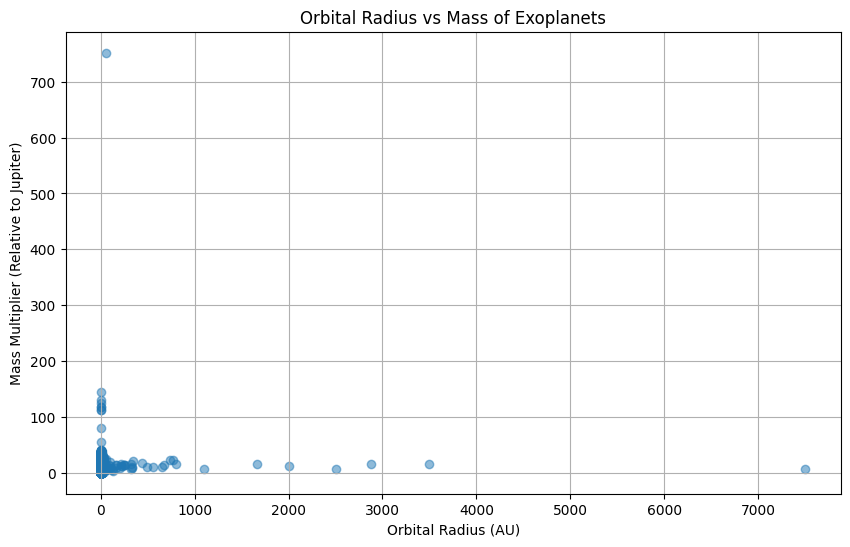
The plot makes it abundantly evident that the high number of exoplanets that NASA has found is a result of the availability of lesser stellar magnitude values, particularly those that are less than 20. This range includes the bulk of exoplanets in the dataset, demonstrating NASA's efficacy in locating celestial entities with lower stellar magnitudes. Curiously, there seems to be one notable exception, a single exoplanet with a stellar magnitude greater than 40. This remarkable finding deviates from the usual pattern and implies that NASA's exoplanet search skills can be extended to more difficult situations with greater stellar magnitudes. All things considered, the information highlights NASA's competence in exoplanet research, especially when it comes to identifying objects linked to fainter stars.



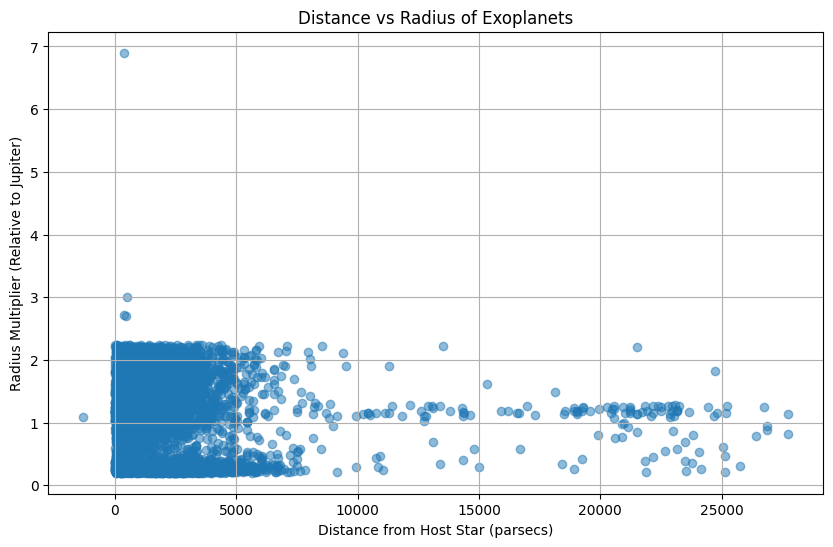
When compared to other planet types, gas giant planets have larger mean values for three major orbital features: orbital period, eccentricity, and orbital radius. This means that gas giants have longer orbital periods, more elliptical orbits (greater eccentricity), and bigger orbital radii than terrestrial planets. These distinguishing features contribute to the unique dynamics and properties of gas giants within planetary systems, distinguishing them in terms of orbital behavior and geographic distribution within a star system.



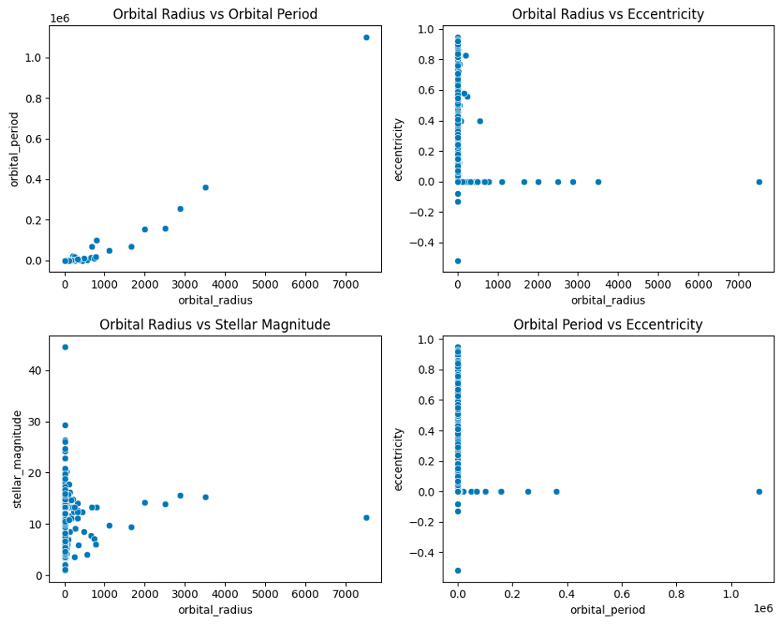
As we can see from the above plot, eccentricity and stellar magnitude have a negative correlation. When eccentricity increases, the stellar magnitude also decreases in a linear fashion and vice versa. A very strong positive correlation can be seen between the variables orbital radius and orbital period. This implies that the planets farther away from their host star take longer to complete one orbit.



The graph explores the relationship between orbital radii and mass multipliers for numerous exoplanets in compared to Jupiter. While the vast majority of exoplanets follow a predictable pattern, two outliers differ dramatically. Although one of these exoplanets is 700 times more massive than Jupiter, it has a lower orbital radius. The second outlier, on the other hand, has the longest orbital radius of the exoplanets represented yet a Mass Multiplier similar to the other planets in the dataset. These anomalies highlight the unique and fascinating character of exoplanetary systems, challenging traditional assumptions and driving deeper research into the variables controlling their orbital dynamics.



According to the findings, there is no substantial relationship between the distance of exoplanets from their host stars and their radii. However, there is a trend showing a decrease in the maximum limit of exoplanet radius as the distance from the host star grows, however there are minor outliers. This finding points to a complex link between exoplanet size and orbital distance. Notably, exoplanet sizes vary greatly between orbits, showing that a variety of factors influence their sizes. The findings highlight the complexities of the mechanisms that influence the size variation of exoplanets in different orbital configurations.



A compelling relationship emerges in celestial dynamics, revealing a positive correlation between orbital radius and orbital period. In the vast expanse of space, most celestial bodies maintain low eccentricity, showcasing a notable shift in behavior at a specific radius. As orbital radius expands, a fascinating phenomenon unfolds—the luminosity of objects intensifies, reaching a peak beyond which brightness stabilizes. This intriguing luminosity-radius relationship hints at underlying astrophysical processes governing the behavior of celestial entities. Furthermore, a distinct orbital shift is observed in the realm of exoplanets, characterized by a sharp transition from short, circular orbits to longer periods, unveiling the diverse and complex nature of planetary systems beyond our own.

1. **Hypothesis Testing/Modelling**

Chi square test to check if there a significant association between planet type and detection method

Null Hypothesis(H0) : There is a significant association between planet type and detection methods.

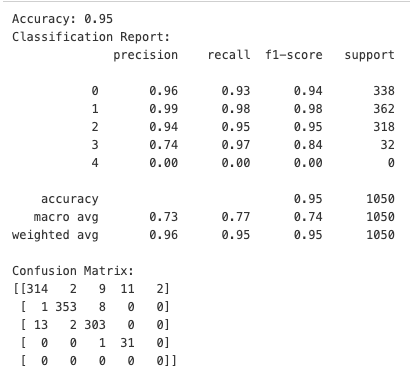
Alternative Hypothesis(Ha) : There is no significant association between planet types and detection methods

Chi-square value: 2580.385837930924

p-value: 0.0

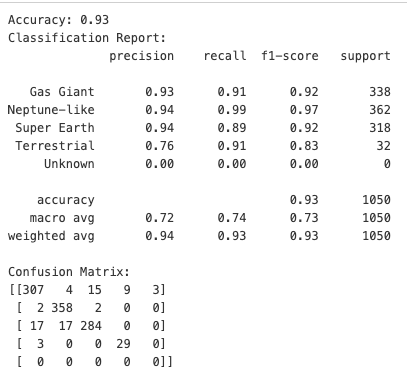
Significance test: There is a significant association between planet type and detection method.

**Random Forest**



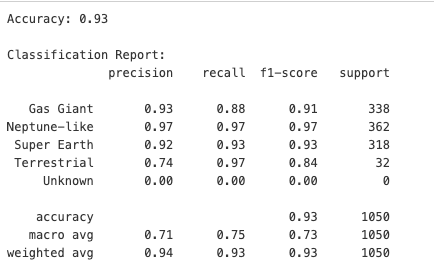
To predict the 'planet\_type,' a Random Forest classifier was trained on a dataset containing characteristics such as'mass\_multiplier,' 'radius\_multiplier,' 'eccentricity,''stellar\_magnitude,' 'orbital\_radius,' and 'orbital\_period'. The model obtained 100% training accuracy and 95% test accuracy, showing great generalization to previously unknown data. The categorization report demonstrates that each class ('Gas Giant,' 'Neptune-like,' 'Super Earth,' and 'Terrestrial') has excellent accuracy, recall, and F1-scores. The confusion matrix demonstrates the model's ability to predict accurately with few misclassifications. While the perfect training accuracy raises the possibility of overfitting, the model's strong performance on the test set suggests that the selected features provide meaningful information for predicting planet types, and the Random Forest algorithm captures the underlying patterns in the data effectively. Continued monitoring for overfitting and consideration of additional evaluation techniques, such as cross-validation, would be prudent for a comprehensive assessment of the model's robustness.

**KNN**

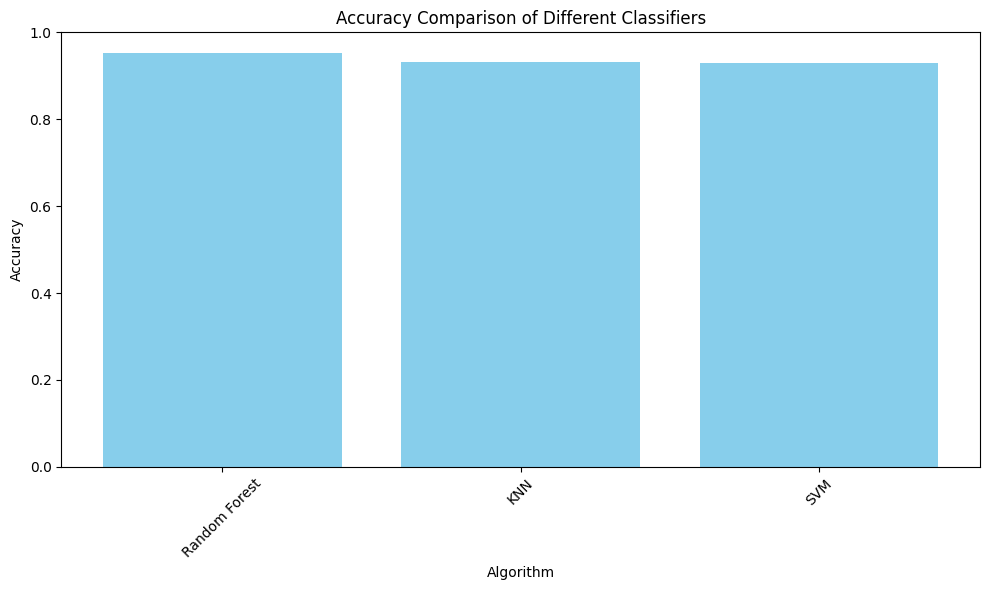


Six characteristics, including'mass\_multiplier,' 'radius\_multiplier,' 'eccentricity,''stellar\_magnitude,' 'orbital\_radius,' and 'orbital\_period,' were employed in this KNN model to predict the 'planet\_type' target variable. The dataset was divided into training and testing sets after missing value rows were removed. With k=5, the KNN model attained a high train accuracy of 0.95 and a test accuracy of 0.93. While the model did well for certain classes, such as 'Neptune-like,' it fared poorly for others, most notably 'Terrestrial' and 'Unknown.' The confusion matrix went over the model's predictions for each class in further depth, indicating areas for future improvement. To improve speed, one may experiment with different k values, think about feature scaling, and look at misclassifications to solve unique issues associated with different classes. Overall, the KNN model is promising. And refining its parameters and features could lead to improved classification results based on the dataset characteristics and application requirements.

**SVM**



The Support Vector Machine (SVM) model with an RBF kernel, regularization parameter (C) of 100, and gamma of 10 was trained on a dataset and achieved a high accuracy of 97% on the training set and 93% on the test set. The categorization report suggests that classes such as "Neptune-like," "Gas Giant," and "Super Earth" have high precision and recall, but the "Terrestrial" class has lesser precision. The "Unknown" class has accuracy, recall, and F1-score values of 0, suggesting no positive predictions. The confusion matrix gives a thorough analysis of guesses for each class. The model's slightly greater training accuracy shows the possibility of moderate overfitting, and more investigation, such as hyperparameter tweaking, might be explored for optimization. Furthermore, assessing the model's performance on various datasets and addressing potential class imbalances may contribute to a more robust assessment.



1. **Conclusion**

Conclusively, this extensive research offers significant perspectives on NASA's exoplanet investigation, exhibiting an extensive array of findings and progressions in our comprehension of far-off planetary systems. The analysis of detection methods over the discovery years has revealed the effectiveness of these techniques in identifying various planet types, with a particular emphasis on the frequent discovery of Neptune-like exoplanets and gas giants, while also highlighting the difficulties in detecting planets with shorter orbital periods. The relationships between planetary properties like orbital radius, orbital period, and eccentricity provide more insight into the intricacies of these celestial systems. Statistical validations using Chi-square tests confirm substantial relationships between planet types and detection methods, lending credibility to the findings. Furthermore, the promising performance of machine learning models—Random Forest, KNN, and SVM—in categorizing planet types using particular characteristics indicates future advances in exoplanetary research. Overall, this analysis highlights the remarkable progress made in exploring exoplanets, as well as the ongoing complexities and challenges in comprehending the diverse nature of these distant worlds, paving the way for further exploration and deeper insights into our universe's celestial marvels.

1. **Reference**
2. Ofman, L., Averbuch, A., Shliselberg, 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.
3. Jin, Y., Yang, L., & Chiang, C. E. (2022). Identifying exoplanets with machine learning methods: a preliminary study. *arXiv preprint arXiv:2204.00721*.
4. Malik, A., Moster, B. P., & Obermeier, C. (2022). Exoplanet detection using machine learning. *Monthly Notices of the Royal Astronomical Society*, *513*(4), 5505-5516.
5. 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.
6. Angerhausen, D. (2019). Big Data and Machine Learning for Exoplanets and Astrobiology: Results from NASA Frontier Development Lab. In *The Tenth Moscow Solar System Symposium* (pp. 244-245).
7. Vishwarupe, V., Bedekar, M., Pande, M., Bhatkar, V. P., Joshi, P., Zahoor, S., & Kuklani, P. (2022). Comparative Analysis of Machine Learning Algorithms for Analyzing NASA Kepler Mission Data. *Procedia Computer Science*, *204*, 945-951.
8. Sturrock, G. C., Manry, B., & Rafiqi, S. (2019). Machine learning pipeline for exoplanet classification. *SMU Data Science Review*, *2*(1), 9.
9. Nixon, M. C., & Madhusudhan, N. (2020). Assessment of supervised machine learning for atmospheric retrieval of exoplanets. *Monthly Notices of the Royal Astronomical Society*, *496*(1), 269-281.
10. 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
11. Ansdell, M., Ioannou, Y., Osborn, H. P., Sasdelli, M., Smith, J. C., Caldwell, D., ... & Angerhausen, D. (2018). Scientific domain knowledge improves exoplanet transit classification with deep learning. *The Astrophysical journal letters*, *869*(1), L7