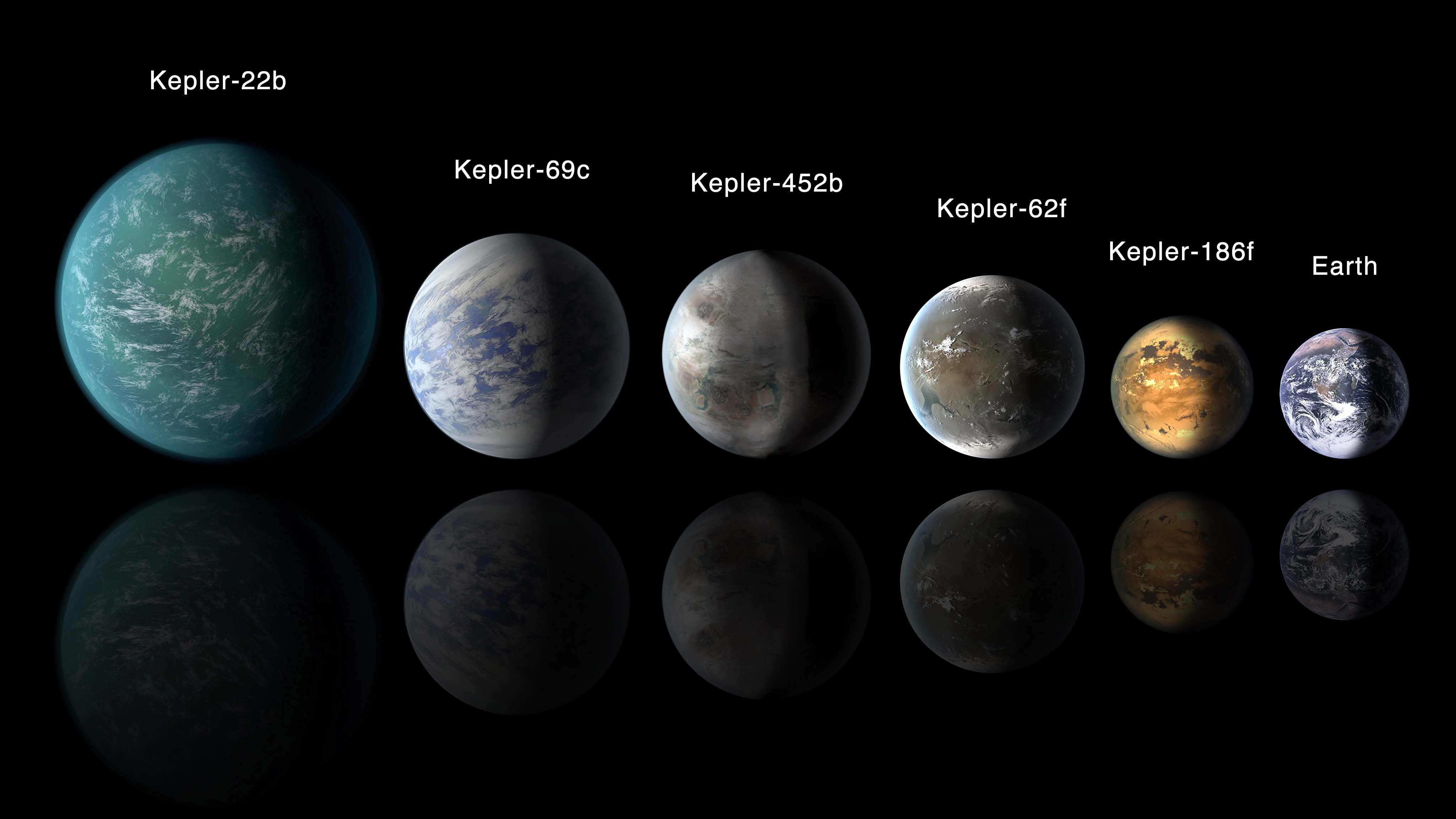
**Machine Learning Homework - Exoplanet Exploration**

[](https://github.com/Apex8/Exoplanet-machine-learning/blob/main/Images/exoplanets.jpg)

**Before You Begin**

1. Create a new repository for this project called machine-learning-challenge. **Do not add this homework to an existing repository**.
2. Clone the new repository to your computer.
3. Give each model you choose their own Jupyter notebook, **do not use more than one model per notebook.**
4. Save your best model to a file. This will be the model used to test your accuracy and used for grading.
5. Commit your Jupyter notebooks and model file and push them to GitHub.

**Note**

Keep in mind that this homework is optional! However, you will gain a much greater understanding of testing and tuning different Classification models if you do complete it.

**Background**

Over a period of nine years in deep space, the NASA Kepler space telescope has been out on a planet-hunting mission to discover hidden planets outside of our solar system.

To help process this data, you will create machine learning models capable of classifying candidate exoplanets from the raw dataset.

In this homework assignment, you will need to:

1. [Preprocess the raw data](https://github.com/Apex8/Exoplanet-machine-learning#Preprocessing)
2. [Tune the models](https://github.com/Apex8/Exoplanet-machine-learning#Tune-Model-Parameters)
3. [Compare two or more models](https://github.com/Apex8/Exoplanet-machine-learning#Evaluate-Model-Performance)

**Instructions**

**Preprocess the Data**

* Preprocess the dataset prior to fitting the model.
* Perform feature selection and remove unnecessary features.
* Use MinMaxScaler to scale the numerical data.
* Separate the data into training and testing data.

**Tune Model Parameters**

* Use GridSearch to tune model parameters.
* Tune and compare at least two different classifiers.

**Reporting**

* Create a README that reports a comparison of each model's performance as well as a summary about your findings and any assumptions you can make based on your model (is your model good enough to predict new exoplanets? Why or why not? What would make your model be better at predicting new exoplanets?).

**Resources**

* [Exoplanet Data Source](https://www.kaggle.com/nasa/kepler-exoplanet-search-results)
* [Scikit-Learn Tutorial Part 1](https://www.youtube.com/watch?v=4PXAztQtoTg)
* [Scikit-Learn Tutorial Part 2](https://www.youtube.com/watch?v=gK43gtGh49o&t=5858s)
* [Grid Search](https://scikit-learn.org/stable/modules/grid_search.html)

**Hints and Considerations**

* Start by cleaning the data, removing unnecessary columns, and scaling the data.
* Not all variables are significant be sure to remove any insignificant variables.
* Make sure your sklearn package is up to date.
* Try a simple model first, and then tune the model using GridSearch.

**Submission**

* Create a Jupyter Notebook for each model and host the notebooks on GitHub.
* Create a file for your best model and push to GitHub
* Include a README.md file that summarizes your assumptions and findings.
* Submit the link to your GitHub project to Bootcamp Spot.
* Ensure your repository has regular commits (i.e. 20+ commits) and a thorough README.md file

**© 2019 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.**