Interstellar Explorers
CSE523 ML Project End Sem Presentation

Title: Classification of Exoplanets

Team members:

Rudri Jani - AU2044002

Dhaval Deshkar - AU2044003

Riya Shah - AU2044004

Supan Shah -AU2044011

Guided by: Dr. Mehul Raval

TA: Jay Patel, Arpit Patel

Ahmedabad University

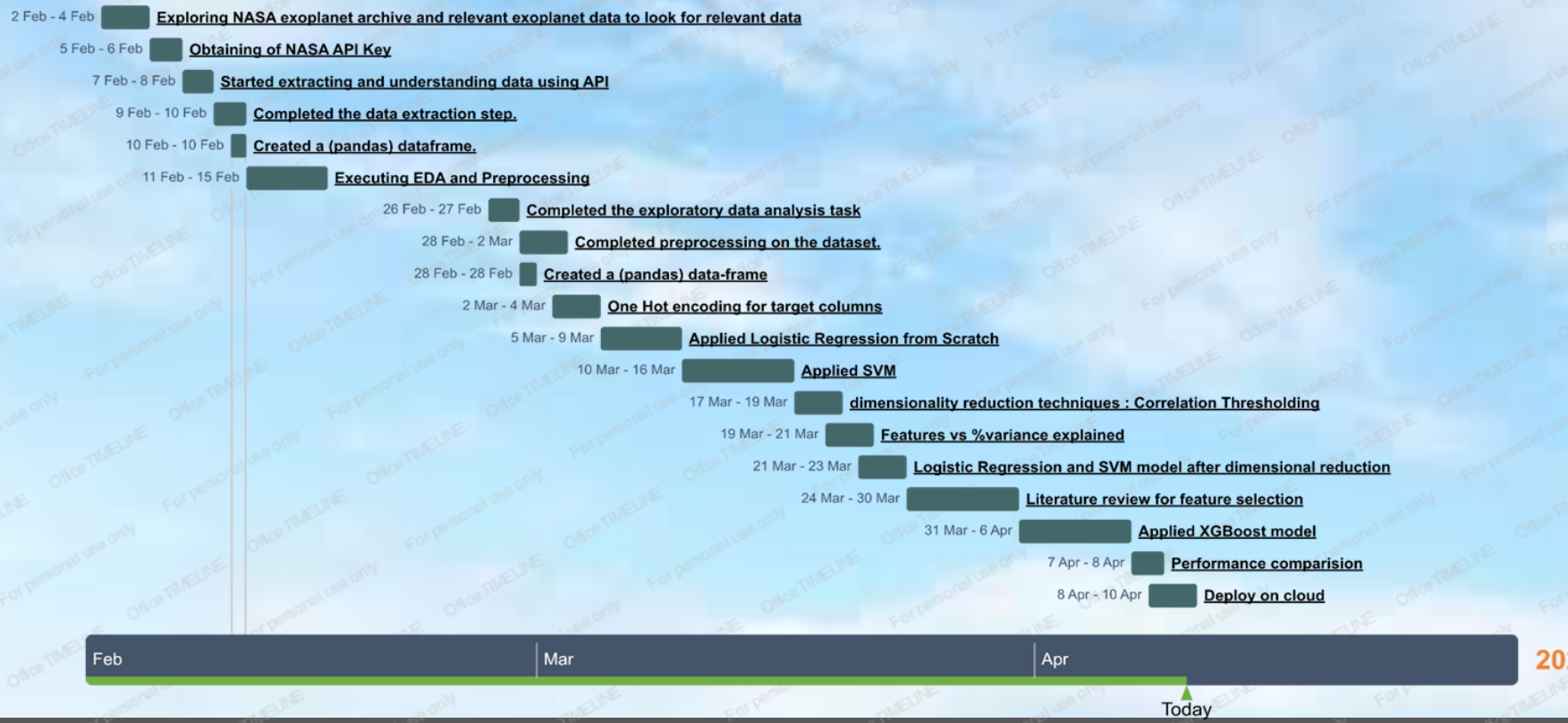
Introduction

- Throughout history, humanity has shared an eternal desire to explore the unknown.
- The current search for a terrestrial, especially those in the Goldilocks (livable) zone where liquid water might exist, has been rejuvenated by technological advances in astronomy.
- As Russian space pioneer Konstantin Tsiolkovsky said, "The Earth is the cradle of humanity, but one cannot live in a cradle forever."

Problem Definition

This project will aim to classify possible exoplanets from the data retrieved from NASA's Kepler mission that was aimed at exploring the structure and diversity of planetary systems.

Gantt Chart


2 Feb - 4 Feb
Exploring NASA exoplanet archive and relevant exoplanet data to look for relevant data
5 Feb - 6 Feb
Obtaining of NASA API Key
7 Feb - 8 Feb
Started extracting and understanding data using API
9 Feb - 10 Feb
Completed the data extraction step.
10 Feb - 10 Feb
Created a (pandas) dataframe.
11 Feb - 15 Feb
Executing EDA and Preprocessing
26 Feb - 27 Feb
Completed the exploratory data analysis task
28 Feb - 2 Mar
Completed preprocessing on the dataset.
28 Feb - 28 Feb
Created a (pandas) data-frame
2 Mar - 4 Mar
One Hot encoding for target columns
5 Mar - 9 Mar
Applied Logistic Regression from Scratch
10 Mar - 16 Mar
Applied SVM
17 Mar - 19 Mar
dimensionality reduction techniques : Correlation Thresholding
19 Mar - 21 Mar
Features vs %variance explained
21 Mar - 23 Mar
Logistic Regression and SVM model after dimensional reduction
24 Mar - 30 Mar
Literature review for feature selection
31 Mar - 6 Apr
Applied XGBoost model
7 Apr - 8 Apr
Performance comparision
8 Apr - 10 Apr
Deploy on cloud
Feb
Mar
Apr
2021
Today

Existing Body of Work

A new era has unfolded in the chase for exoplanets in any case.The Modern generation of the new satellites, such as a Kepler have been launched in recent times with the objective of in partially automating logical perceptions and data generation with exoplanet identification proof.

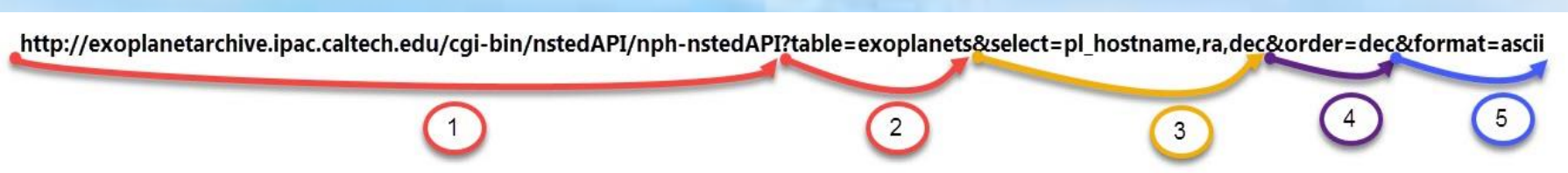
Engineering of these satellites are not for only takes the picture but also do the process of those images using higher astronomical techniques to create such a huge collection of data with proper variety of features for identification of exoplanets.

The authors Sturrock & George Clayton describes in their paper titled“""Machine Learning Pipeline for Exoplanet Classification” The simulation of model deployment is apex of data pipeline, which prepares, train and test for machine learning model.

The final result of machine learning modeling is introduced and summarised to show the classification of the exoplanets dimensionally over the different models constructed in the study.

Approach

- The dataset is taken from NASA's exoplanet archives using its live API.

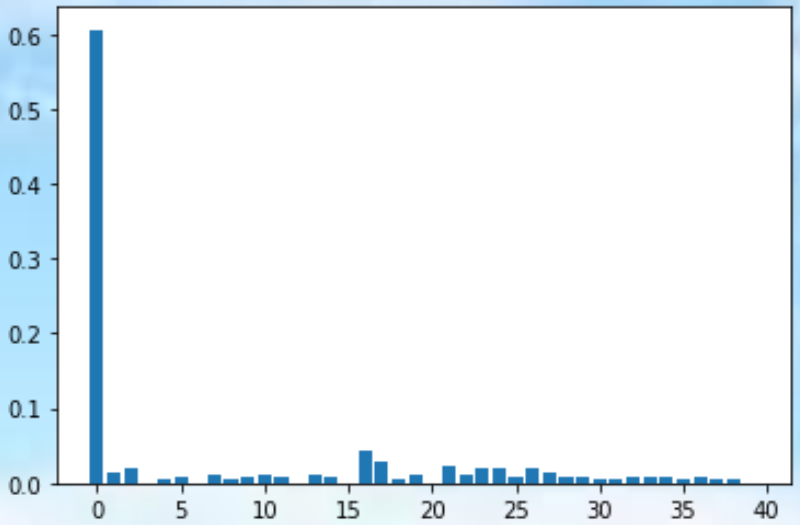


- The data identifies the three categories of exoplanets like CONFIRMED, CANDIDATE and FALSE POSITIVES. The goal of the project is to identify this using classification algorithms.
- After extracting the data, we did data preprocessing tasks and created a new column for target. We used dimensionality reduction and got 5 columns as features and applied Logistic Regression and SVM on it.

Image Source : https://exoplanetarchive.ipac.caltech.edu/docs/program_interfaces.html

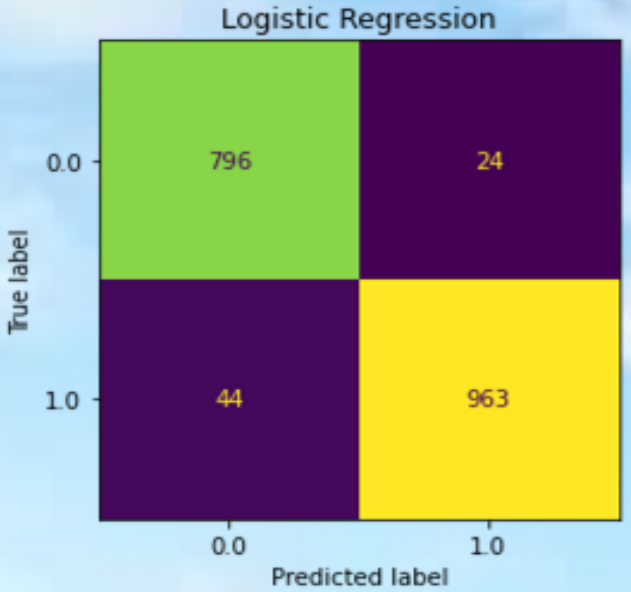
XGBoost

- XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework.
- When it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best
- <u>Selected Features</u>
  disposition score,
  planetary radius,
  TEC planet number,
  Stellar Temperature,
  Transit Signal-to-noise ratio.



source: Google

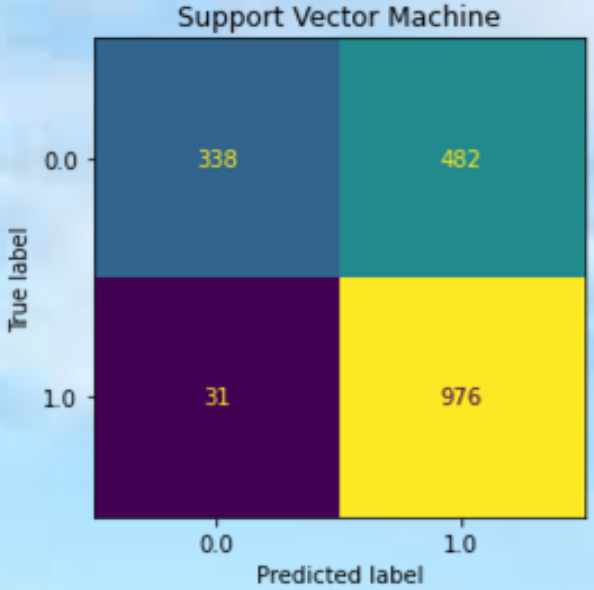
Confusion Matrix After Feature Selection Using XGBoost



Test

Accuracy: 96.28%

F1 Score: 96.59%

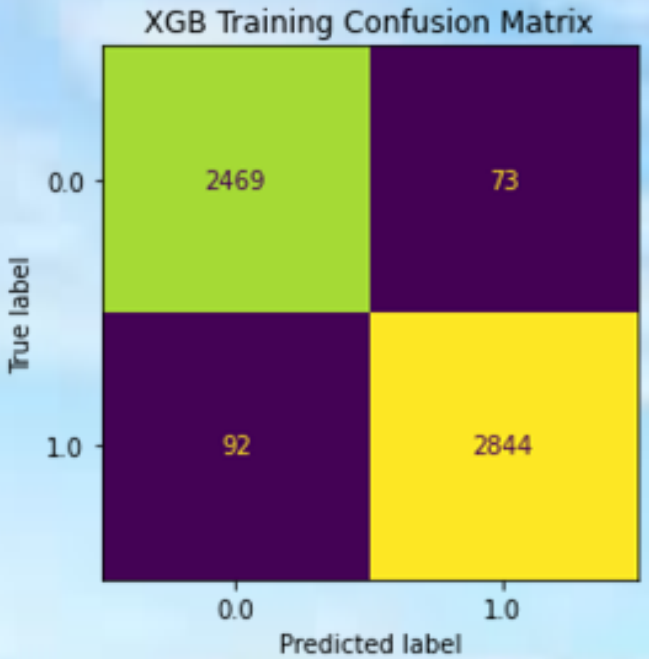


Test

Accuracy: 71.92%

F1 Score: 79.18%

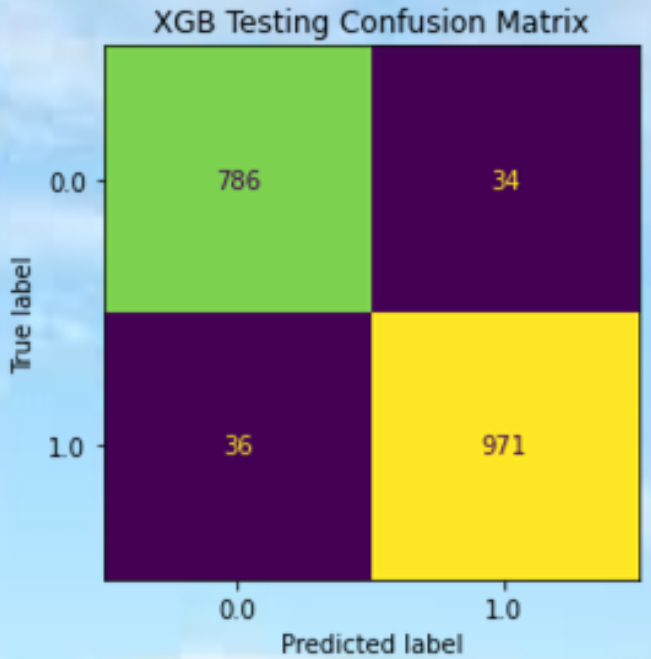
XGBoost Results



Train

Accuracy: 96.98%

F1 Score: 97.18%

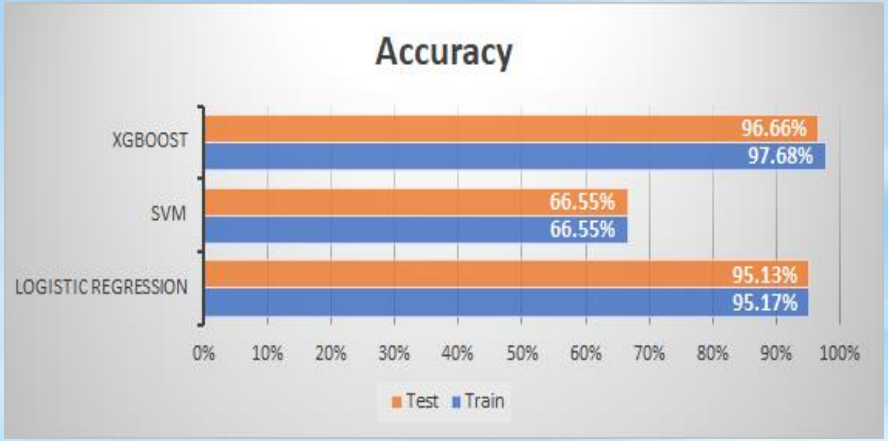


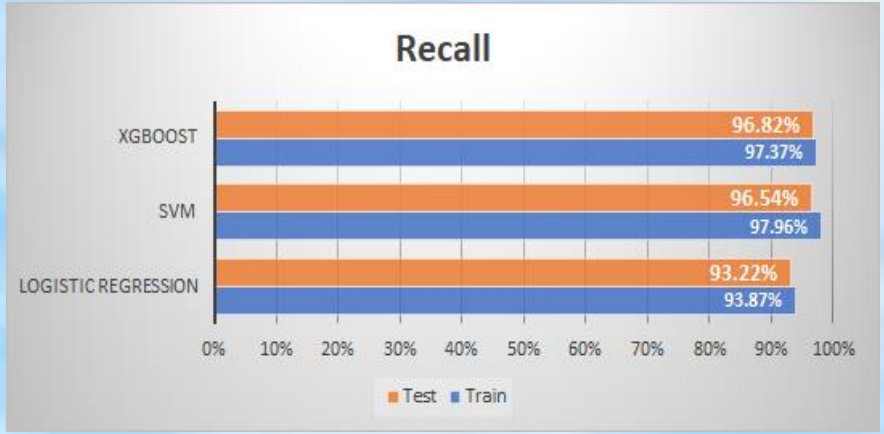
Test

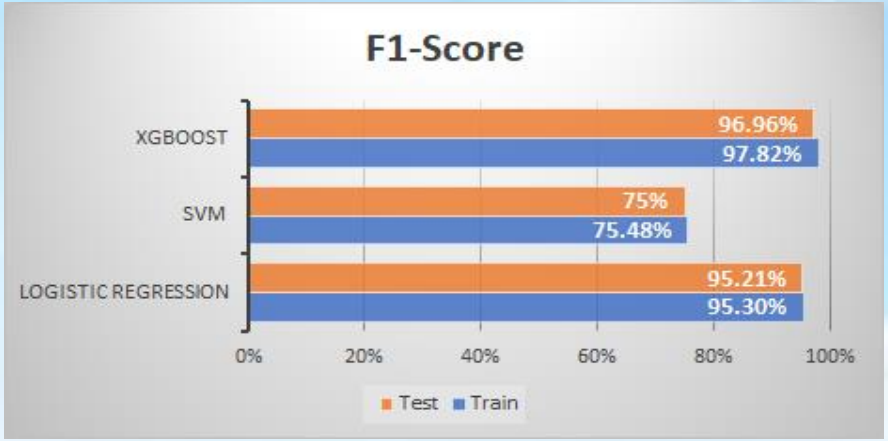
Accuracy: 96.16%

F1 Score: 96.52%

Performance Comparison



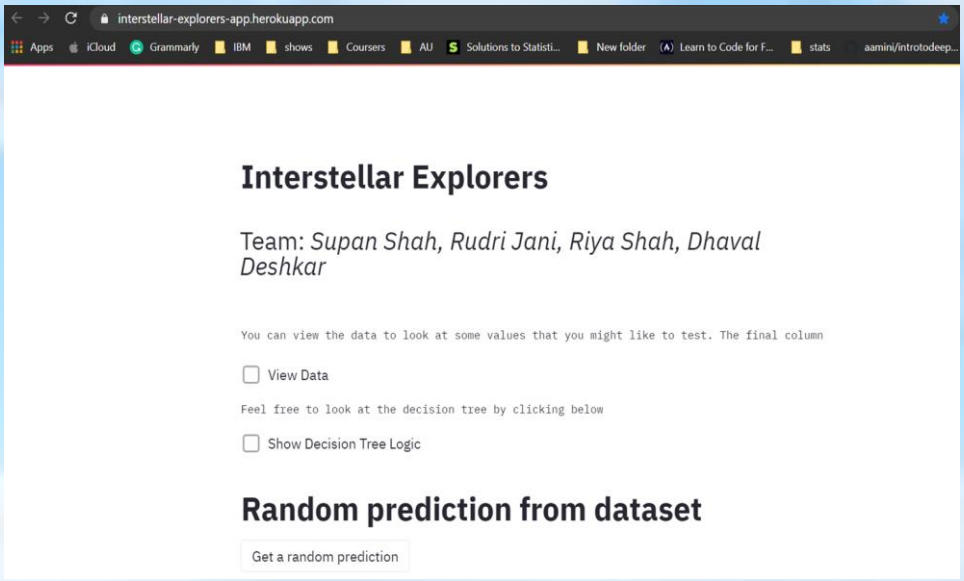




Final Result

- After deploying the trained model on the IBM cloud, it returns the probability of any planet being an exoplanet and we constructed the final web based service.
- https://interstellar-explorers-app.herokuapp.com/
- https://youtu.be/mEqn2KPAOjg

```
get_random_predictions(10)

Model Prediction:
There is 98.7 % chance of given planet being an exoplanet.
Reality:
The planet is an exoplanet.


 ---------------------------------

Model Prediction:
There is 4.3 % chance of given planet being an exoplanet.
Reality:
The planet is NOT an exoplanet.


 ---------------------------------

Model Prediction:
There is 99.4 % chance of given planet being an exoplanet.
Reality:
The planet is an exoplanet.


 ---------------------------------
```



Conclusion

- In this project we used different models to get better accuracy and f1-score so we can achieve better results for classifying exoplanets.

- To conclude, XGBoost model performs way better on unseen data as it gives 96.66% test accuracy and 96.96% test f1-score.

- As final part, we deployed the model on IBM cloud. We generated deployed model through API inference and we constructed the final web based service.

Role of each member

Name	Task performed
Rudri Jani	EDA, Preprocessing, Support vector Machine, Literature Review, feature selection using xgboost, SVM with xgboost feature selection
Dhaval Deshkar	Dimensionality Reduction, Logistic Regression, Literature Review, Logistic Regression with Xgboost features, Decision Tree
Riya Shah	Exploratory Data Analysis , Preprocessing, Support vector Machine, Literature Review, XGboost modelling, Performance Comparison
Supan Shah	Data extraction, Dimensionality Reduction, Logistic Regression, Literature Review,Deployment on IBM cloud,herokuapp

References

- API:

  https://exoplanetarchive.ipac.caltech.edu/docs/program_interfaces.html

- Dataset:

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

- Documentation :

  https://exoplanetarchive.ipac.caltech.edu/docs/API_kepcandidate_columns.html

- Literature Review :

  G. C. Sturrock, B. Manry, and S. Rafiqi, “Machine learning pipeline for exoplanet classification,” SMU Data Science Review, vol. 2, no.1,p. 9,2019

Thank you.