Interstellar Explorers
CSE523 ML Project Mid 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

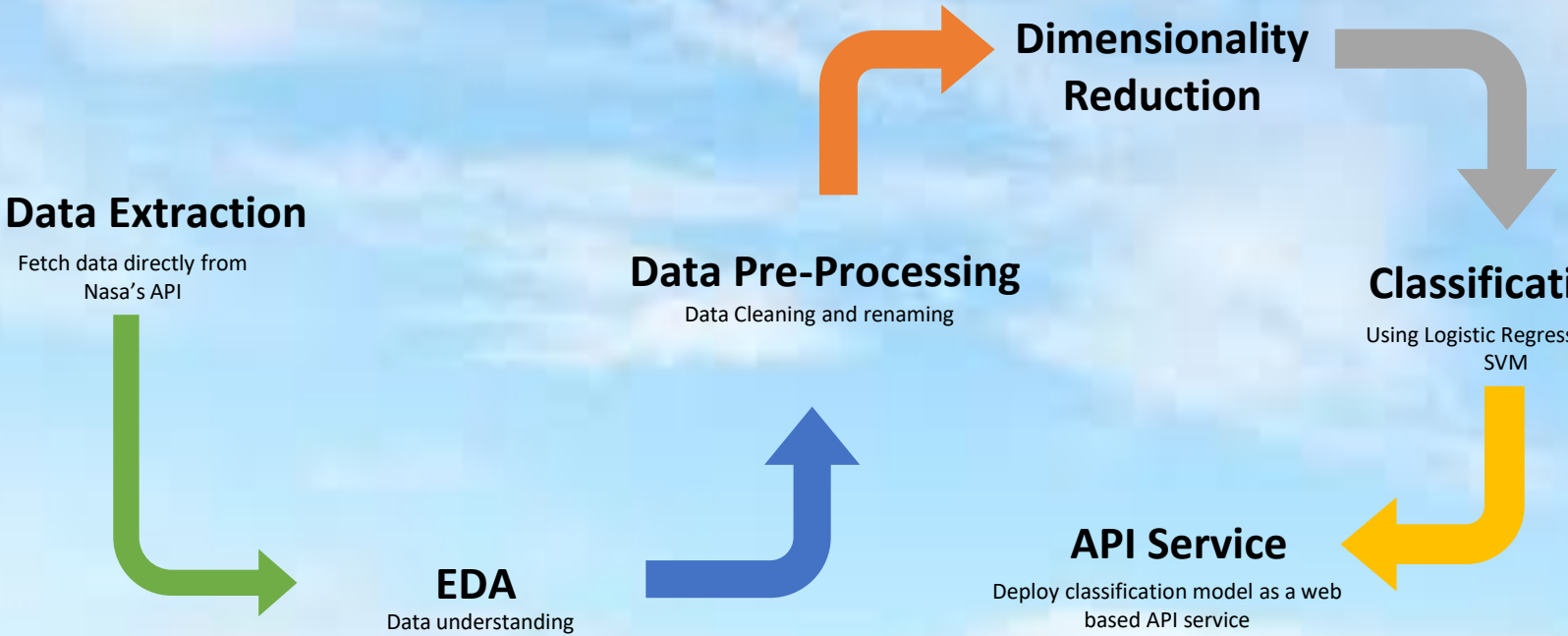
Problem Definition

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.

Background

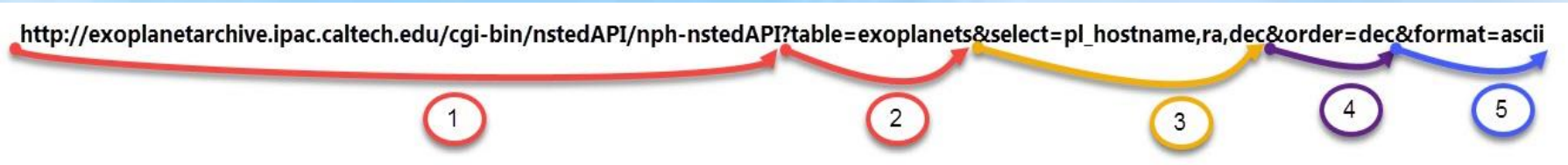
- Throughout history, humanity has shared an eternal desire to explore the unknown.
- The current search for a terrestrial, especially those in the Goldilocks (livable) zone where liquid water might exist, has been rejuvenated by technological advances in astronomy.
- As Russian space pioneer Konstantin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

Road Map


Dimensionality Reduction
Data Extraction
Fetch data directly from Nasa's API
Data Pre-Processing
Data Cleaning and renaming
Classification
Using Logistic Regression and SVM
API Service
Deploy classification model as a web based API service
EDA
Data understanding

Dataset

- 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.
- Confirmed – The planet is exoplanet
- Candidate - possibility of exoplanet
- False Positive – Not an exoplanet

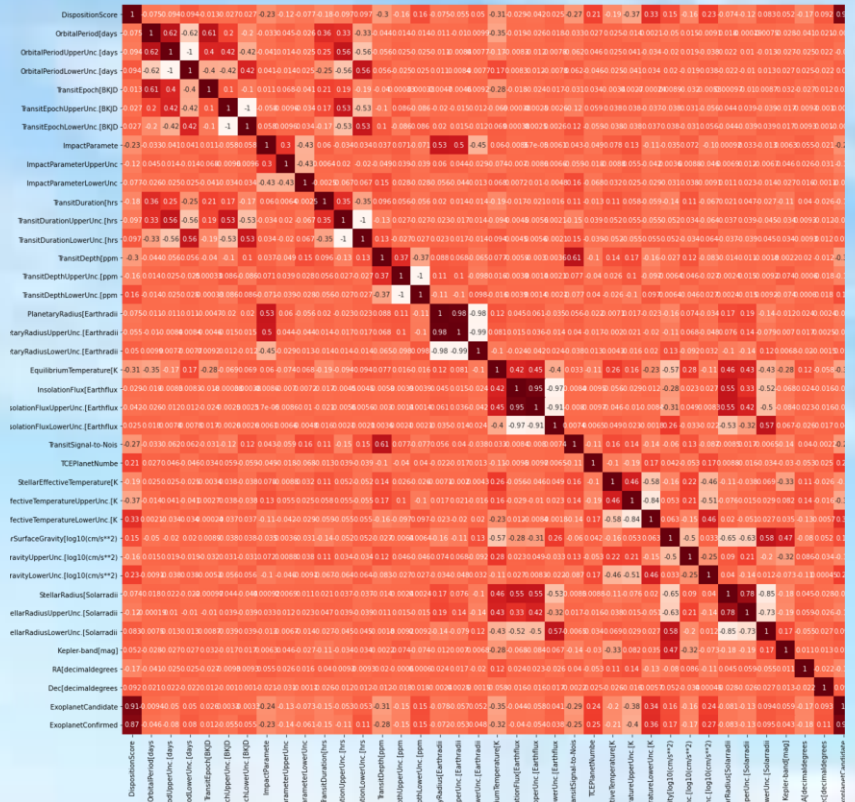
Image Source : https://exoplanetarchive.ipac.caltech.edu/docs/program_interfaces.html

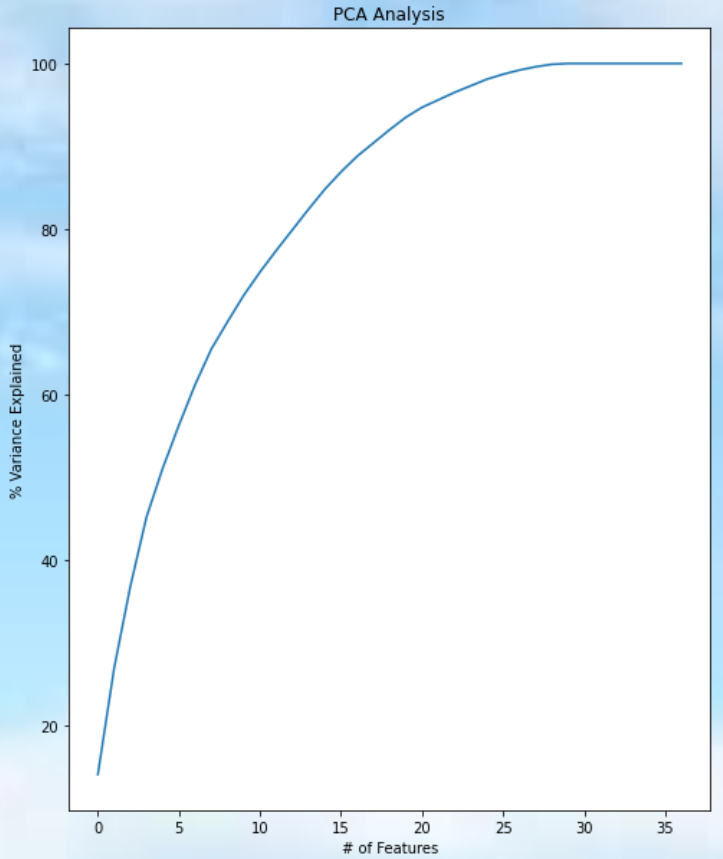
Methodology

- The project will use classical machine learning methods for the classification of planets as exoplanets or not using features extracted by the Kepler space telescope.

1. Around 8% of the data was null and hence removed.
2. Remove irrelevant columns from dataset
3. Normalize the data : Z-score normalization
4. One hot encoding of categorical data
5. Create a new column on the basis of candidate and confirmed planets which will be the target column.
6. Next, we apply dimensionality reduction techniques to reduce dimensions.

Dimensionality Reduction



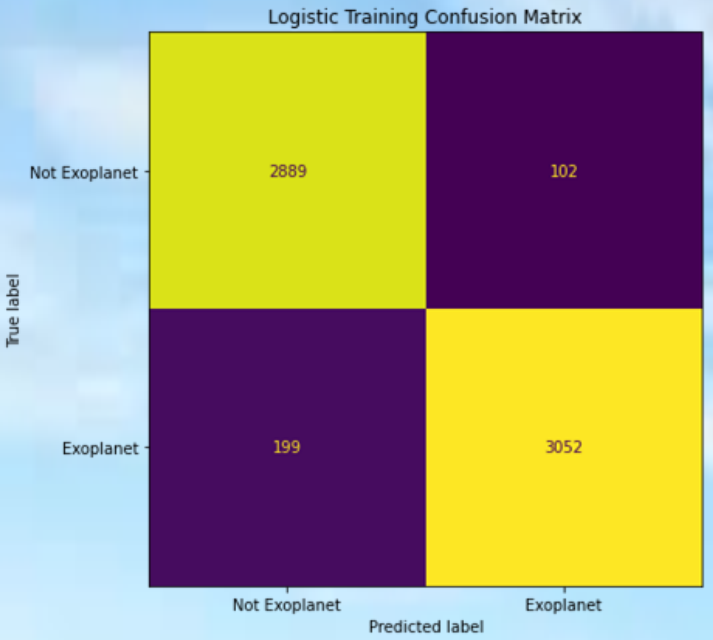


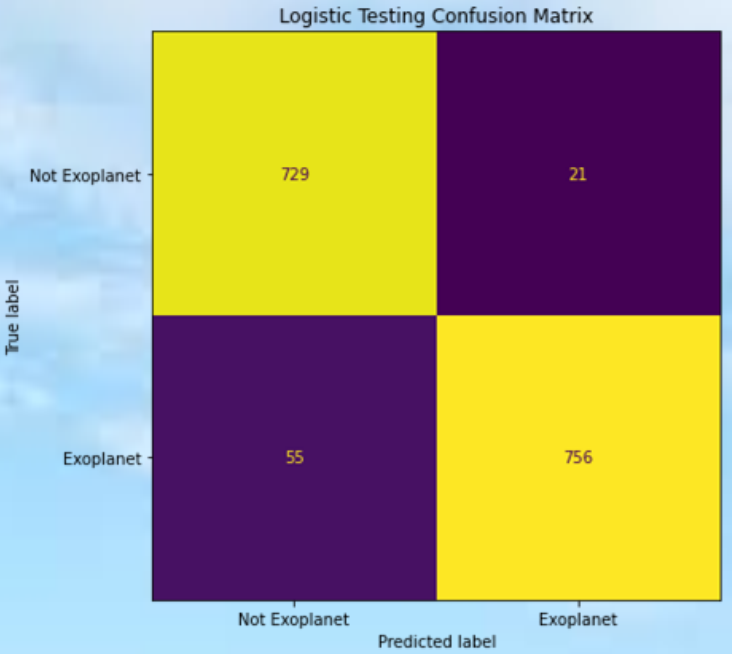
Process data for classification

- After applying dimensionality reduction, we selected five best columns
- Train-Test split : 80-20%
- Final processed data :

	DispositionScore	ImpactParameterUpper	TransitDepth[ppm]	TransitSignal-to-Noise	TCEPlanetNumber
0	-1.033552	-0.213373	3.886380	0.858828	-0.375167
1	-1.033552	-0.167855	-0.302691	-0.336844	-0.375167
2	-0.240581	-0.169386	-0.307851	-0.320646	2.710197
3	-1.033552	-0.211740	-0.309945	-0.333533	-0.375167
4	-1.033552	-0.212148	0.879915	1.733857	-0.375167

Logistic Regression Results





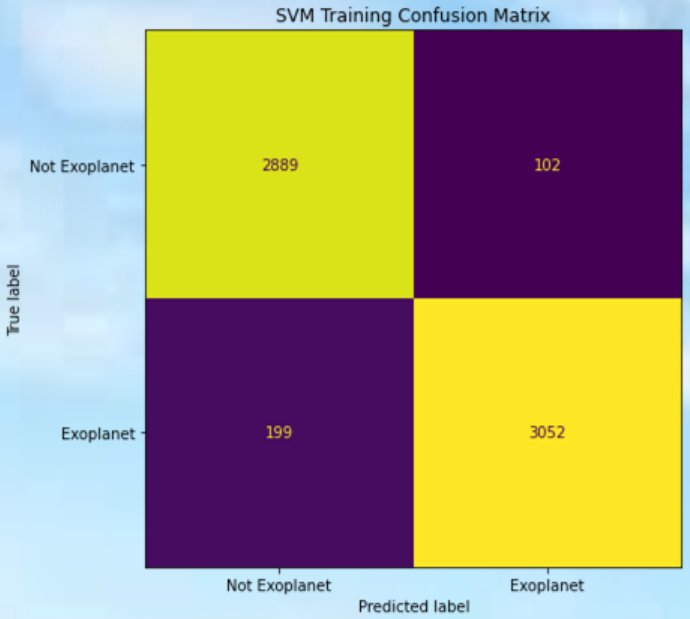
Train

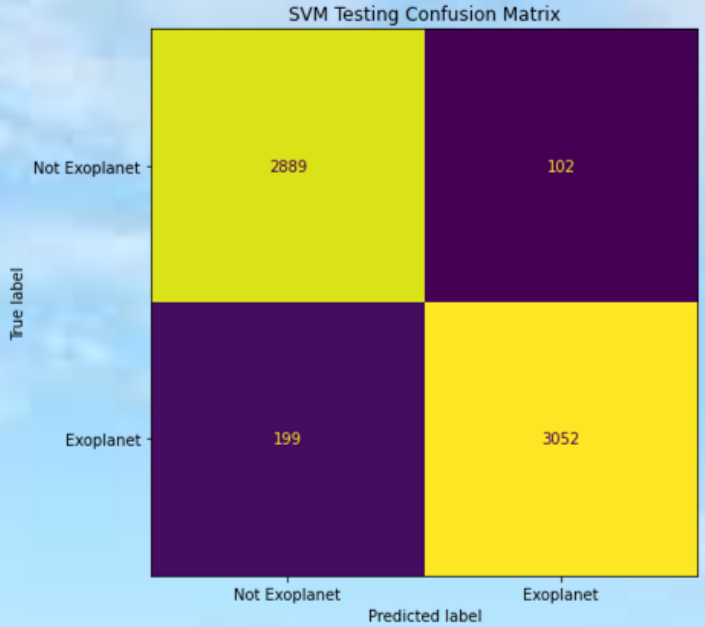
Accuracy: 95.17%
F1 Score: 95.30%

Test

Accuracy: 95.13%
F1 Score: 95.21%

Support Vector Machine Results





Train

Accuracy: 66.55%
F1 Score: 75.48%

Test

Accuracy: 66.55%
F1 Score: 74.99%

Future Work

- Create dashboard where user can key-in feature values and get instant predictions.

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

Thank you.