Machine Learning for Astronomical Purposes

Final Review

17BIT003

Guided by Dr. Swati Jain

Two major objectives

Exoplanet detection

- Use Kepler’s Object of Interest cumulative Exoplanet dataset from NASA exoplanet archive.
- In addition to stellar features, time series flux values are taken STScI via lightkurve package
- Tried classifying using basic classifiers and boosting classifiers will try to use ExoNet or AstroNET for increasing efficiency.

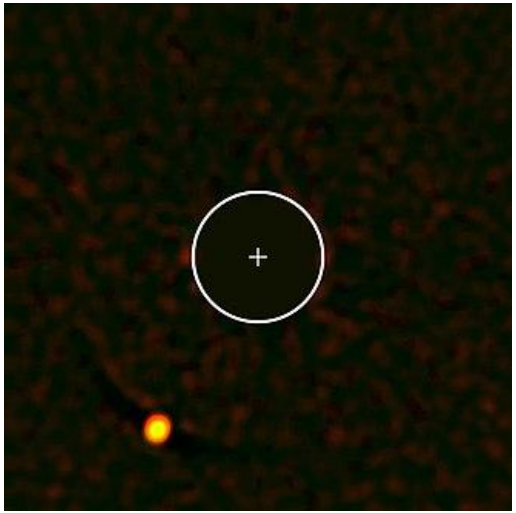
Checking the habitability

- Use PHL’s Exoplanets Catalog Data with NASA archive to predict habitable planets and handle missing data.
- Preprocess the data and oversample using SMOTE.
- Will make model to check habitability and evaluate it.

Exoplanet

What are Exoplanets?

An exoplanet or extrasolar planet is a planet outside the Solar System. The discovery of exoplanets has intensified interest in the search for extraterrestrial life. There is special interest in planets that orbit in a star's habitable zone, where it is possible for liquid water, a prerequisite for life on Earth, to exist on the surface. The study of planetary habitability also considers a wide range of other factors in determining the suitability of a planet for hosting life. Below is the image of Exoplanet HIP 65426b is the first discovered planet around star.



KOI cumulative dataset

- The data set is extracted from NASA Exoplanet archives. : https://exoplanetarchive.ipac.caltech.edu/
- In this particular model Kepler's Object of Interest Dataset is used containing stellar features of celestial bodies. : https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=koi
- Dataset description is given here : https://exoplanetarchive.ipac.caltech.edu/docs/API_kepcandidate_columns.html
- A KOI is a target identified by the Kepler Project that displays at least one transit-like sequence within Kepler time-series photometry that appears to be of astrophysical origin and initially consistent with a planetary transit hypothesis.

STScI dataset for time-series flux of exoplanets

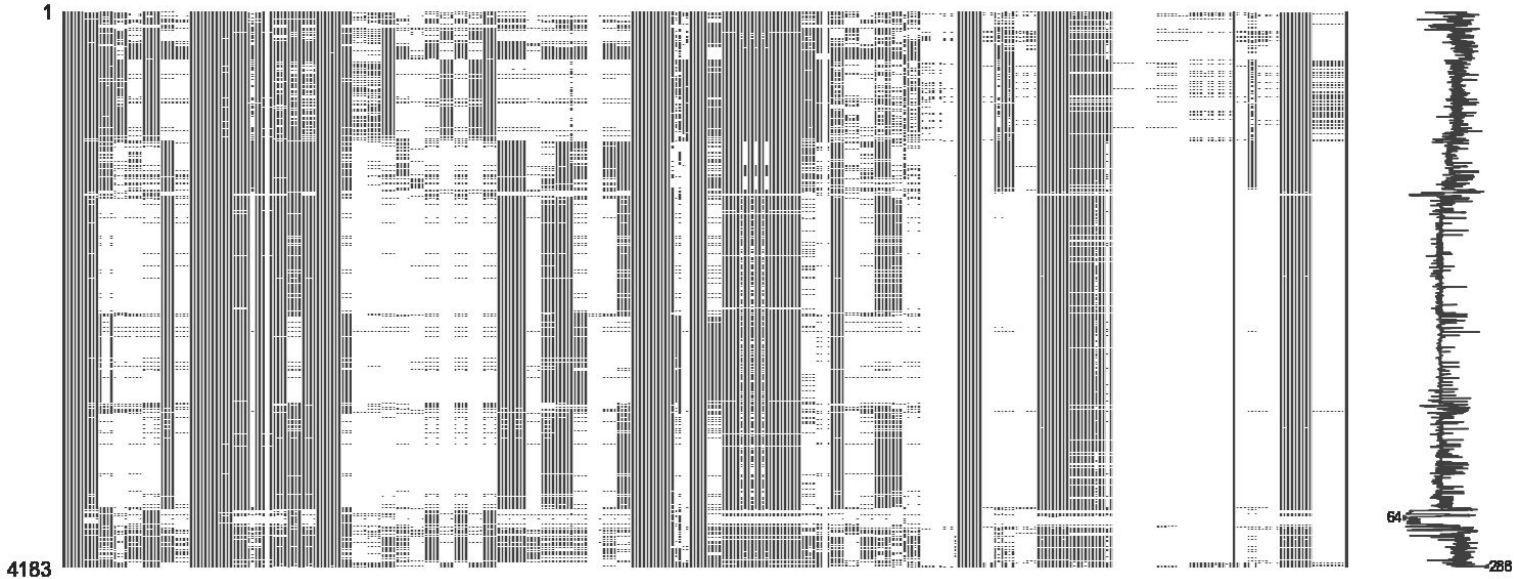
- As above dataset contained only stellar features we use api of MAST(Mikulski Archive of Space Telescope) provided by STScI (Space Telescope Science Institute) using lightkurve package. : https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html

Habitability of an exoplanet

- Around almost every “normal” star, including our sun, we can draw a band of potential habitability: the right distance and temperature for liquid water to exist. The key, of course, is a planetary surface where the water could pool. Both stars and planets come in many types and sizes, and the interplay of these factors determines the extent and influence of this “habitable zone.”
- PHL's Exoplanets Catalog was used to predict habitable planets. : http://phl.upr.edu/projects/habitable-exoplanets-catalog/data/database
- It contains observed and modeled parameters for all currently confirmed exoplanets, including those potentially habitable.

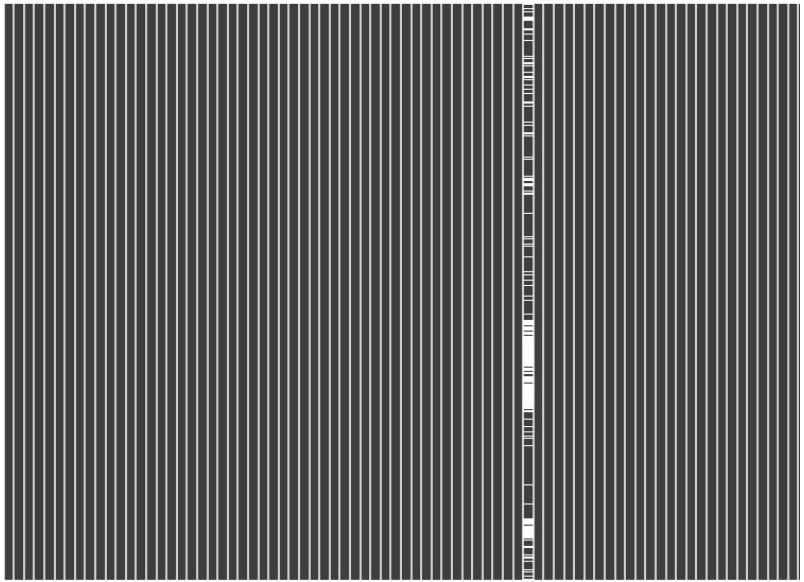
Preprocessing of data

- Now there are two datasets to handle: Nasa Exoplanet Archive and PHL data which contains habitability situations. These datasets contain the physical features of planets and stars.
- First task was to remove the missing values in the dataset.

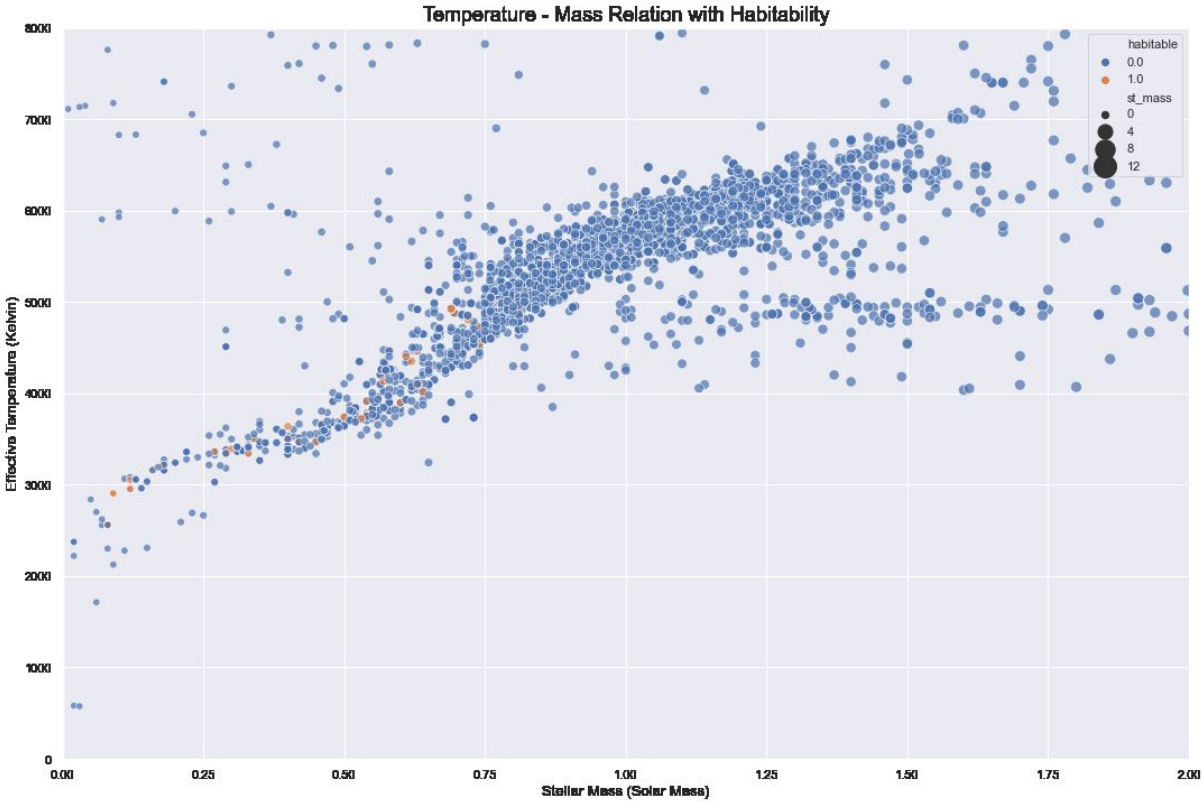


Preprocessing of data

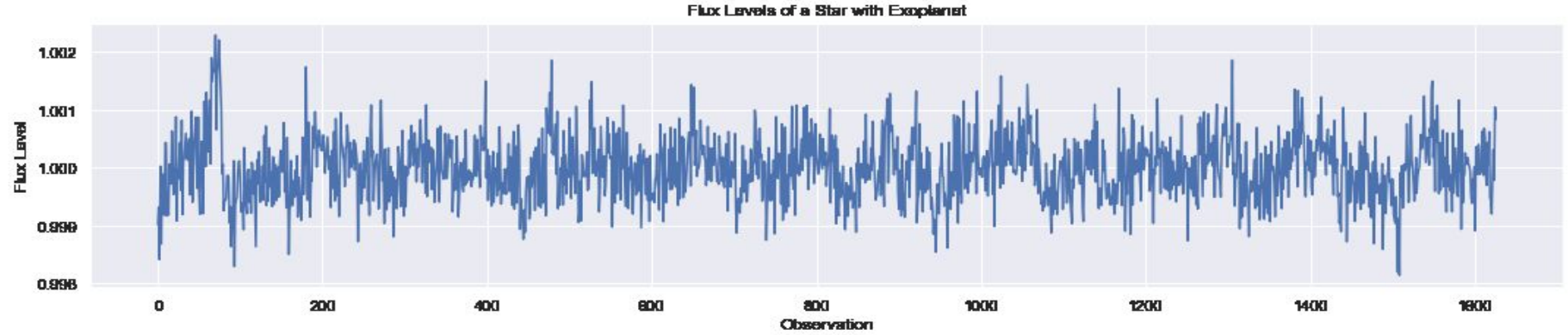
- Removing columns with more than 40% missing data.
- Removing categorical features with more than 10 unique values.
- Iterative imputing on numeric values.

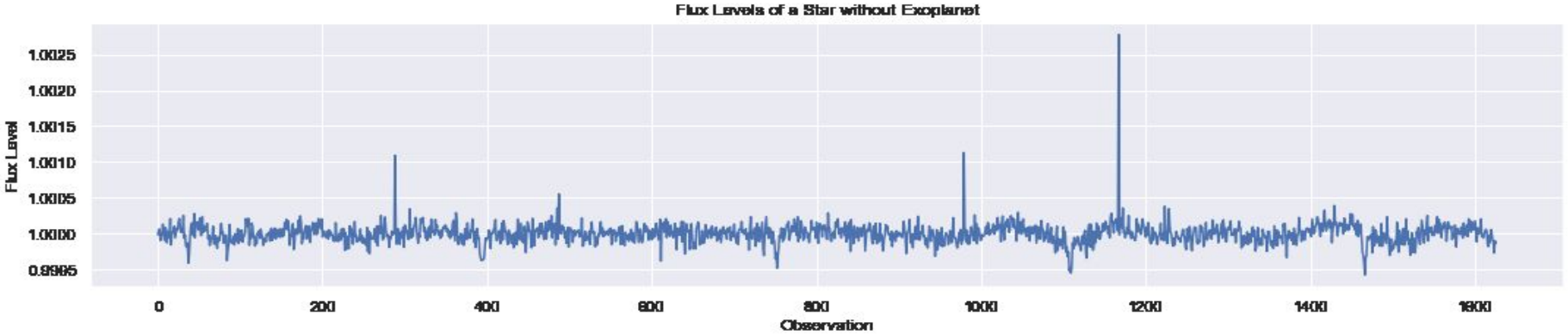


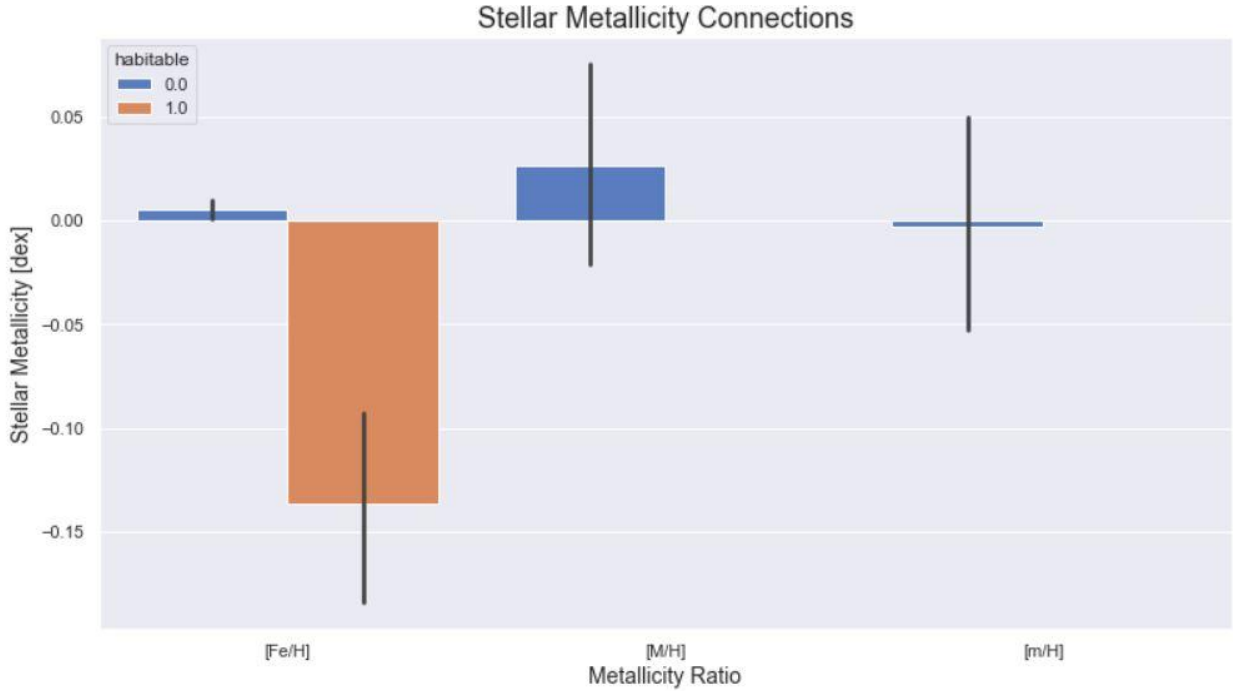
Identifying relevant features



Relation between flux and exoplanets

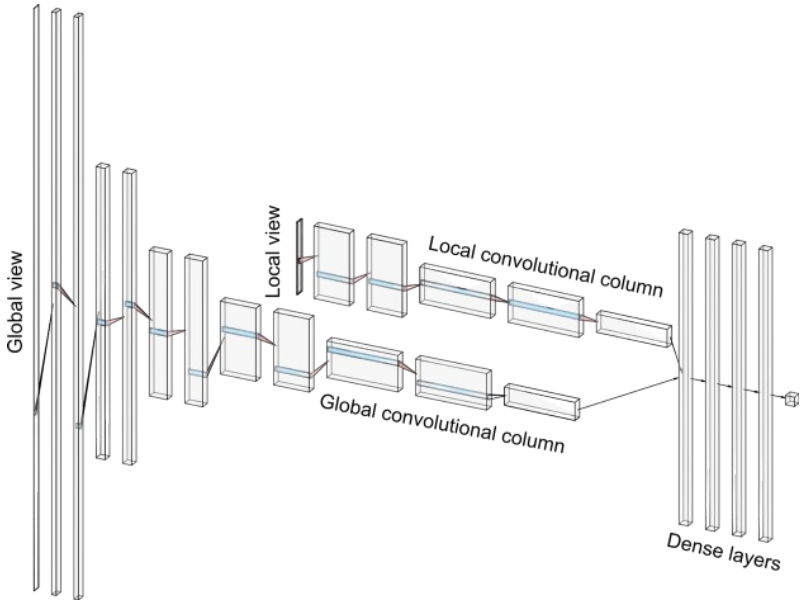





Stellar Metallicity Connections
habitable
0.0
1.0
0.05
0.00
−0.05
−0.10
−0.15
Stellar Metallicity [dex]
[Fe/H]
[M/H]
[m/H]
Metallicity Ratio

ExoNet

- The ExoNet used here is a Convolutional Neural Network(CNN) specially made for detecting exoplanets in space.
- In ExoNet two one-dimensional CNNs are incorporated, one of the views is global view and the other is local view.



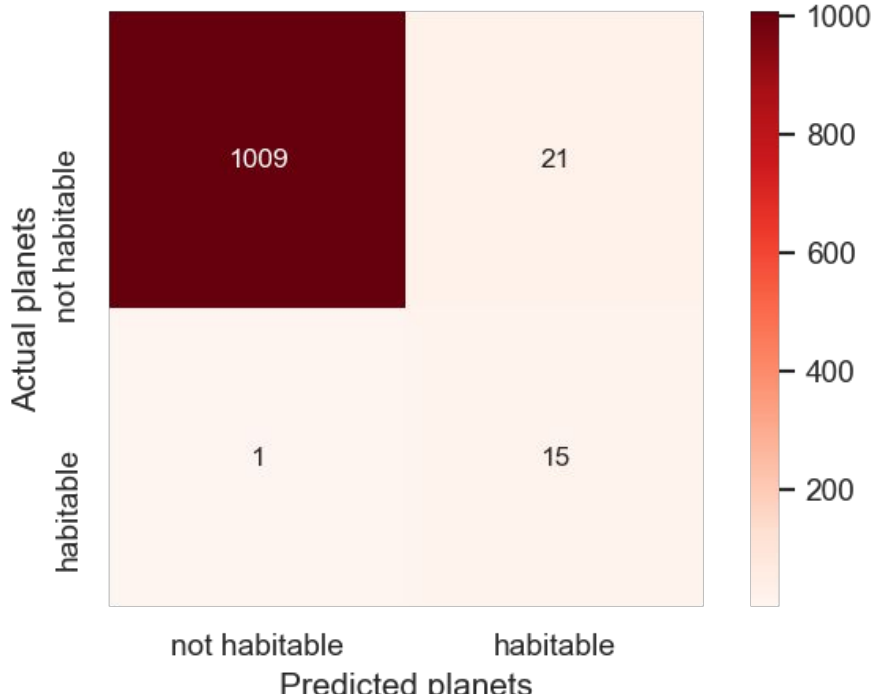
Evaluation

- Common Classifiers like SVC and KNN achieved around 60-62% accuracy on test data.
- Classifiers like Random Forest and Bagging Classifiers achieved 65-66% accuracy
- As there are lots of features training is hard, proposed to use ExoNet. Implementation of ExoNet on TESS candidates have given 94-95% accuracy.

Model	Train Accuracy	Test Accuracy	Precision	Recall	F1
Bagging Classifier	98.28	66.89	0.67	0.66	0.66
Random Forest	98.97	65.97	0.66	0.65	0.65
Gradient Boosting	64.74	64.40	0.70	0.63	0.60
Logistic Regression	61.61	62.47	0.75	0.60	0.55
Linear SVC	61.58	62.32	0.73	0.60	0.55
Decision Trees	98.97	62.02	0.62	0.62	0.62
GaussianNB	61.16	61.82	0.75	0.59	0.53
KNN	75.78	60.40	0.60	0.60	0.60

ExoNet performance

Here we can see the ExoNet performs well and shows 97.90 test accuracy, which is way better than common machine learning techniques



References

- https://www.researchgate.net/publication/344173550_Astronomical_big_data_processing_using_machine_learning_A_comprehensive_review/citations
- https://medium.com/@miguel.blanco.marcos/exonet-an-ai-saturdays-project-a1bda907bdef
- https://github.com/alpercakr/Planet-Hunting-and-Habitability-Detection-with-Machine-Learning

Thank you