MACHINE LEARNING

WORKSHOP

DAY 03

TOPIC NAME: EXOPLANET HUNTING USING ML

INSTRUCTORS NAME: PRAGYA AND SAHAS

ROADMAP OF WHAT WE WILL LEARN TODAY:

What is Machine Learning

How to hunt for exoplanets

How to process data

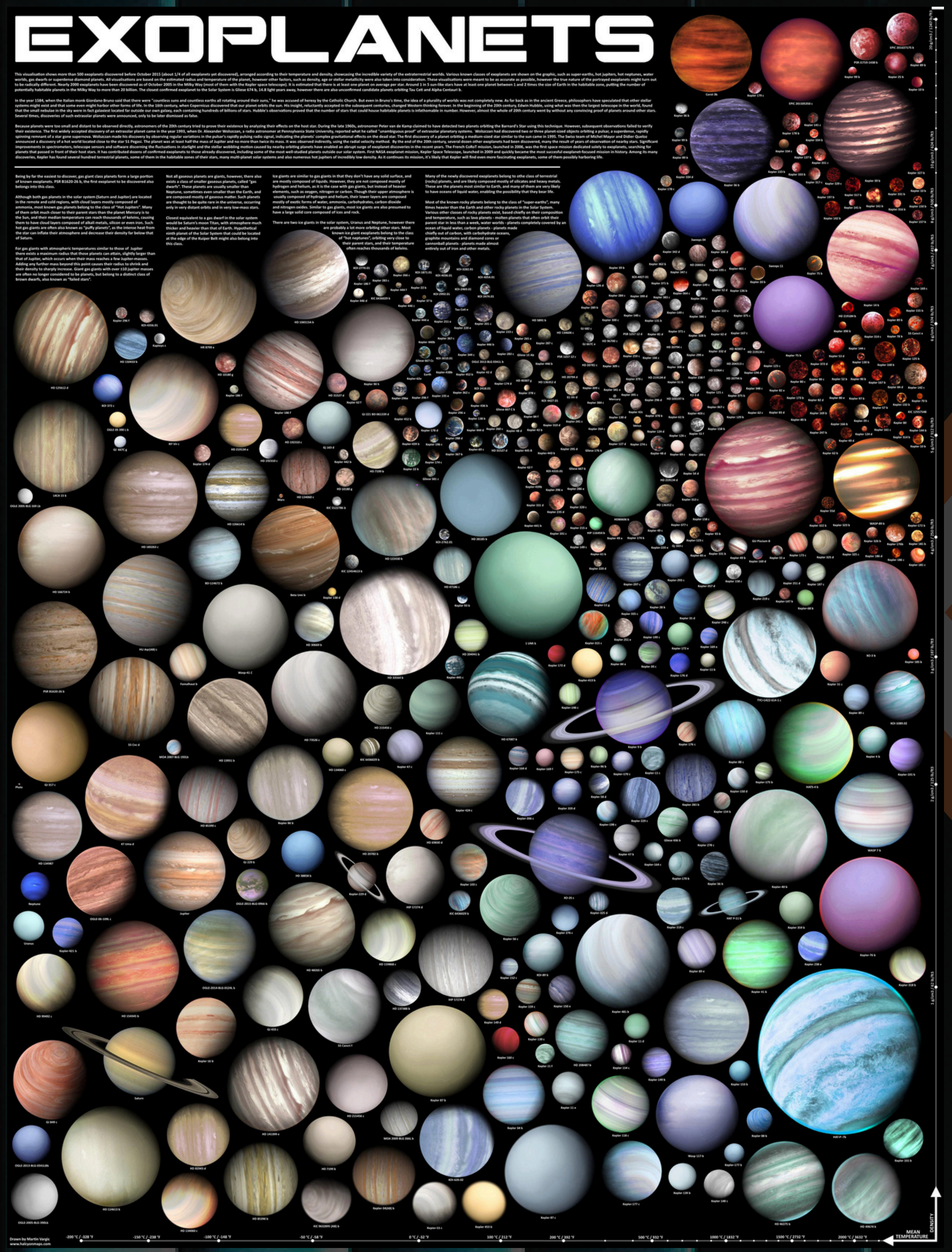
How to visualize the data

Train and make your predictions !!

LIBRARIES WE NEED,

- Numpy
- Pandas
- Matplotlib
- Scikit-learn
- Lightkurve
- Scipy
- lmlearn
- XGBoost

WHY WE USE ML IN EXOPLANET DETECTION:

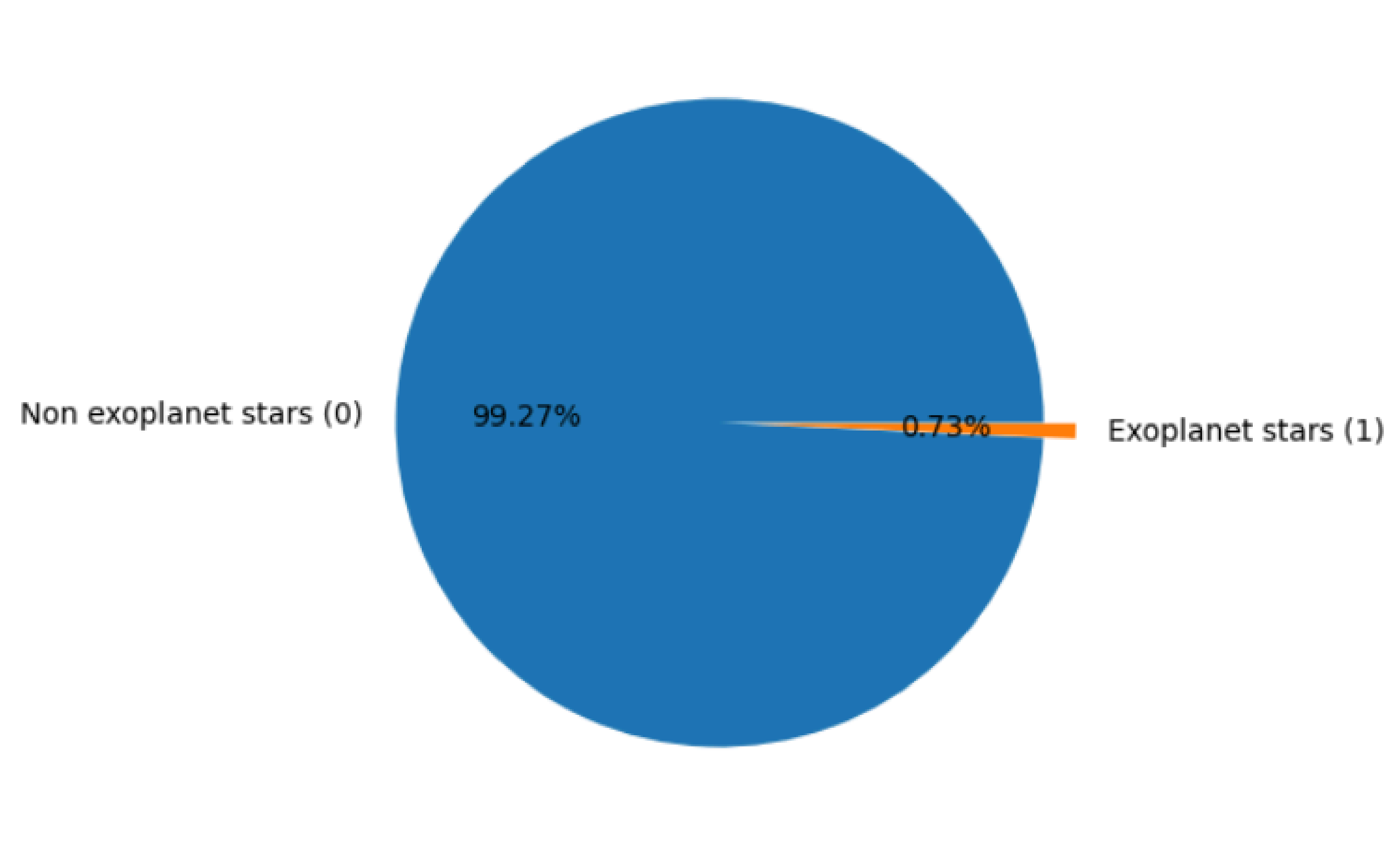


Machine learning helps us find new planets outside our solar system by quickly and accurately analyzing huge amounts of space data that would take humans a very long time to go through

CODE SNIPPETS

EXPLORATORY DATA ANALYSIS

```
[ ] label_counts = train_data['LABEL'].value_counts()
    labels = ['Non exoplanet stars (0)','Exoplanet stars (1)']
    explode = [0,0.1]
    plt.figure(figsize=(5,5))
    plt.pie(label_counts,labels=labels,autopct='%1.2f%%',explode=explode)
    plt.show()
```



class imbalance!
solve using SMOTE

MODEL CODE

Model Training - Using XGBoost Classifier

```
[ ] from xgboost import XGBClassifier
    from sklearn.metrics import classification_report
    from imblearn.over_sampling import SMOTE
    from sklearn.preprocessing import StandardScaler
    from sklearn.decomposition import PCA
```

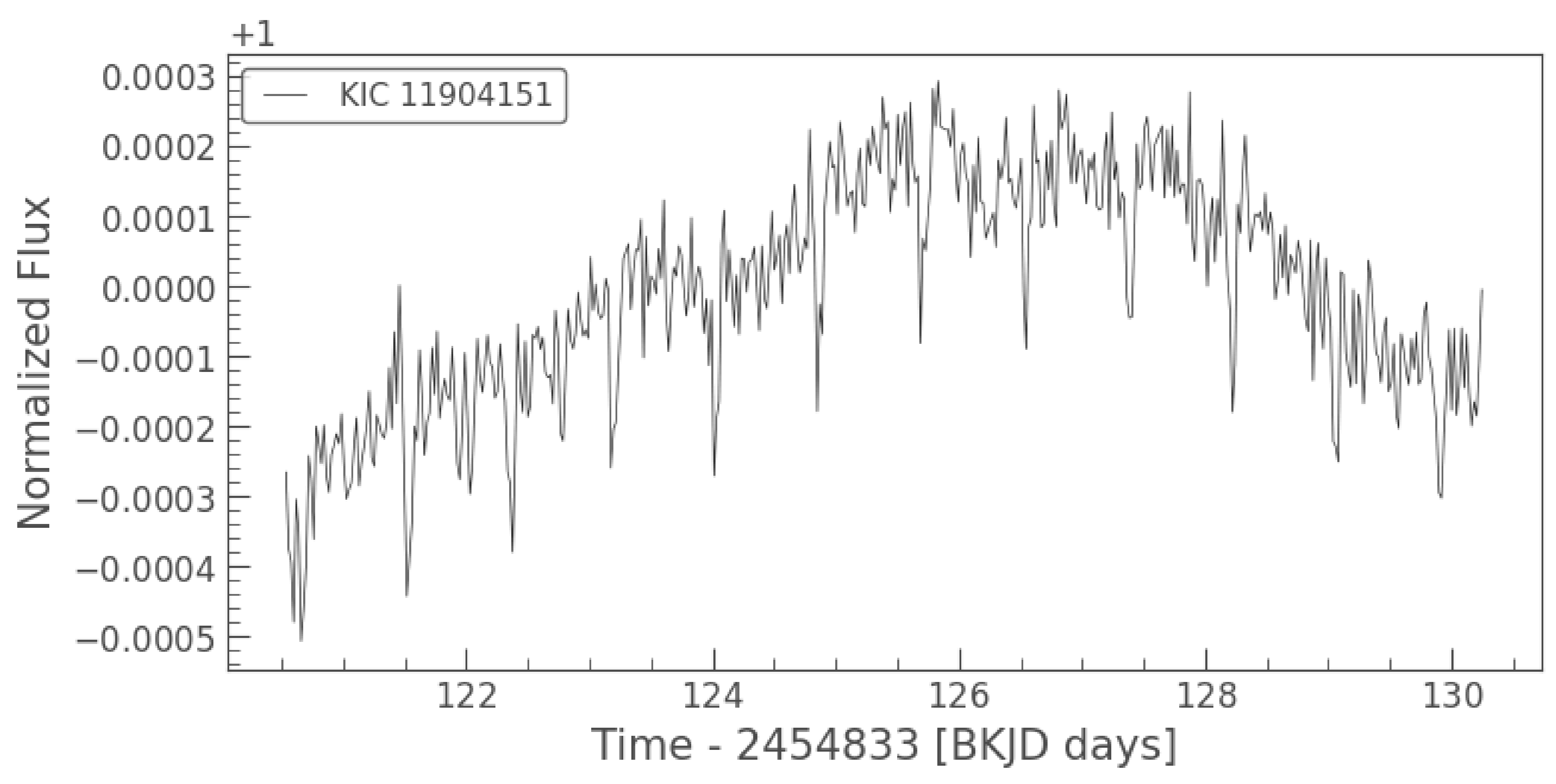
```
[ ] # train
    xgb = XGBClassifier(use_label_encoder=False, eval_metric='logloss')
    xgb.fit(X_train_feat, y_train)

    # eval
    y_pred = xgb.predict(X_test_feat)
    print(classification_report(y_test, y_pred))
```

IMPORTING LIGHTKURVE

```
!pip install lightkurve
from lightkurve import search_lightcurve
import matplotlib.pyplot as plt
```

```
search_result = search_lightcurve("Kepler-10", author="Kepler", cadence="long")
lc = search_result.download()
lc = lc.normalize().remove_nans()
lc.plot()
```



PROBLEM STATEMENT 1:

Stellar Object Classification using Sloan Digital Sky Survey DR17

Dataset: https://www.kaggle.com/datasets/fedesoriano/stellar-classification-dataset-sdss17

Description: Develop a ML model to classify celestial objects into stars, galaxies, and quasars based on their spectral characteristics. Try to cover EDA, feature extraction, Model training, classification reports.

Bonus:

For objects classified as galaxies, develop a secondary model to distinguish between elliptical, spiral, and irregular galaxy types using photometric and structural parameters.

PROBLEM STATEMENT 2:

CLASSIFICATION OF TRANSIENT AND NON TRANSIENT OBJECTS BASED ON LIGHT CURVES

Dataset: https://github.com/MachineLearningUniandes/MANTRA

Description: Develop a ML model to classify transient object and non transient object light curves with any suitable algorithm. Try to cover EDA, feature extraction, Model training, classification reports.

Bonus:

The dataset can be used for both binary classification (above) and multi-class classification (8 classes - Some of the classes included in the dataset are: supernovae, cataclysmic variables, active galactic nuclei, high proper motion stars, blazars and flares.). Try developing a multi-class classification ML model that can classify transient astronomical events based on the light curves.

SUBMISSION GUIDELINES

- The codes must shared as a google colab link.
- The notebook should be named according to the format - ‘yourRegNo_yourName’ Eg. 24BEC1234_JohnDoe
- Make sure you enable public access while sharing the notebook.
- Practices like adding comments or markdowns for better understanding is highly encouraged.