**Problem 1**

Write a web scraping script which takes an input of any film actor and gives the output of filmography of that actor in descending order.

Use web scraping method

Eg :

input [Leonardo DiCaprio]

Output : Films done by Leonardo DiCaprio in descending order

Time Limit:1 Day

**Problem 2**

Given a list of planets discovered by KEPLER.

Kepler Data: <https://drive.google.com/drive/folders/1GwqC4STc_KgVPofacQUzKHBMHQsmflvY?usp=sharing>

Create an ML algorithm to classify the planets as Candidate/False positive/Confirmed etc based on the column “koi\_disposition”.

Time Limit:1 Day

**Sub-questions for Problem 2**

1. Why did you choose the particular algorithm?

The Random Forest Classifier was chosen for its ability to handle high-dimensional data, maintain good accuracy with many features, and reduce overfitting compared to a single decision tree. Its ensemble nature combines multiple trees for better performance.

1. What are the different tuning methods used for the algorithm?

Key tuning methods for the Random Forest Classifier include:

* Number of Trees (n\_estimators)
* Maximum Depth of Trees (max\_depth)
* Minimum Samples Split (min\_samples\_split)
* Minimum Samples Leaf (min\_samples\_leaf)
* Feature Selection (max\_features)

1. Did you consider any other choice of algorithm? Why or why not?

Yes, other algorithms like Support Vector Machines (SVM) or Logistic Regression could have been considered. However, The Random Forest was chosen for its reliable performance, ease of implementation, and ability to handle various data types without extensive preprocessing.

SVM: Effective for high-dimensional data, though requiring more tuning and computational resources.

Logistic Regression: Simple and interpretable, but potentially less capable of capturing complex relationships compared to ensemble methods like Random Forest.

1. What is the accuracy?

Accuracy: 0.9012023000522739

1. What are the different types of metrics that can be used to evaluate the model?

Classification Report:

precision recall f1-score support

CANDIDATE 0.83 0.77 0.80 484

CONFIRMED 0.81 0.84 0.82 490

FALSE POSITIVE 0.98 1.00 0.99 939

accuracy 0.90 1913

macro avg 0.87 0.87 0.87 1913

weighted avg 0.90 0.90 0.90 1913