**Predictive Analytics for YouTube Trending Videos**

**Marks 40**

This assessment will enable you achieve essential experience of machine learning using sklearn to program advanced predictive models. **(Deep Learning not allowed)**

The focus of this assessment is to explore what makes videos popular on various platforms, e.g., Netflix and YouTube. The assessment will focus on YouTube 's data.

PLEASE NOTE There can be multiple solutions possible for this assessment. various students will end up implementing different solutions.

**Section 1. Machine Learning with Sklearn**

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression, and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Now we will train some machine learning models using sklearn to predict views, rather than predicting views directly we will predict views\_log to avoid numerical instability issues

**Task 1: Project-II (Sklearn )** (10 marks)

1. Implement two "Ensemble Methods", not already implemented in this notebook or in your technical project). The performance of these models should be compared on single plot in Step 2.

Refer to the following link: https://scikit-learn.org/stable/supervised\_learning.html

2. you must tune the models and show the accuracy of both models using high-quality.

3. Record a presentation to explain the implementation of the models in the notebook, not longer than 10 minutes.

**Task 2: In-class Presentation and Viva (30 marks)**

The presentation should include:

* Title
* Introduction of problem/domain/Significance
* Dataset Description
* Questions
* Selection of Models/Motivation to Model Section
* Architecture of First Model (Sketch/draw, no copy from sources allowed)
* Architecture of Second Model (Sketch/draw, no copy from sources allowed)
* Comparison of Results of Model 1 and Model 2
* Discussions
* Conclusions

**Submission:**

* Source code with highlighted modifications added under appropriate section in the same notebook, it must be highlighted/sections with clear Label Task-1.
* Recorded presentation of the notebook
* Presentation as PPT