School of Computer Science and Engineering
Fall Semester-2024-25

Course Code : CBS3007
Course: Data Mining and Analytics

Class Number: VL2024250103849 / 50

Slot: L1+L2 / L49+L50

Experiment-3 [CO5 & BL4]

Instructions:

- *Aim, packages used, Sample input, Output and results must be recorded.*
- *Coding and Output screenshot must be included.*
- *Prepare a document as instructed and upload (Final Version) in VTOP student's log in on-time. (Dataset link must be provided in the document)*

1. Collect the student attendance and performance dataset of your classroom to identify students who are likely to drop out or fail early. Implement the KNN to classify the above cases and display the list of students and their classes as per classifications.

2. Linear regression of 2 variables is to use one variable to forecast another variable value. Collect the DEMAT account counts of Indians for the past 60 months. Implement the Linear regression Technique to predict what will be count in JAN2025 in future. Collect the real time sample data from news sources to perform the algorithm.

3. Implement the Random Forest Supervised Machine Learning Algorithm that is used widely in multi-Classifications in Fruits dataset. (Assume own dataset at least 50 entries). [Lab purpose question]