SRM Institute of Science and Technology

School of Computing

DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS

SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamilnadu

Academic Year: 2024-25 (ODD)

Test: CLAT-2 [Unit 3 and 4] **Date: Oct 9, 2024**

Course Code & Title: 21CSE426T & Financial Machine Learning **Duration:** 55 Minutes

Year & Sem: III Year & VI Sem **Max. Marks:** 30

Answer all / 15 * 2 = 30 Marks]

Q No	Question	BL	CO	PO	Marks
1.	(a) In a financial context, you're tasked with building a model to predict loan default probabilities. Outline the steps you would take to clean the dataset, identify the most important features, and build the classification model. Justify your choice of evaluation metrics and provide the Python code for implementing the model. **OR** (b) Fraud detection models often face the issue of imbalanced datasets. Explain how you would handle this imbalance while building a classification-based model for fraud detection. Write a Python code snippet to demonstrate model tuning and improving recall for fraudulent transactions using Logistic Regression.	3	3	2	15
2.	**(a)** A trading strategy requires clustering a set of financial stocks into homogeneous groups based on historical price and volatility. Explain how you would use the k-means clustering algorithm for this purpose and demonstrate how you would determine the optimal number of clusters. Include relevant Python code snippets. **OR** **(b)** You're given a dataset of 500 investors with various risk profiles and financial backgrounds. Discuss how you would use hierarchical clustering to segment these investors into distinct groups. Illustrate the process using a dendrogram and explain how you would evaluate the effectiveness of the clusters.	3	3	2	15