SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF COMPUTER APPLICATIONS

QUESTION BANK

III SEMESTER M.C.A.

PMC402 - PREDICTIVE MODELLING

Regulation – 2024

Academic Year 2025 – 2026 (Odd Semester)

Prepared by

Mr. M.ASAN NAINAR (AP Sel.G/M.C.A.)

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)
SRM Nagar, Kattankulathur-603203
DEPARTMENT OF COMPUTER APPLICATIONS

QUESTION BANK

SUBJECT :PMC402 - PREDICTIVE MODELLING
SEM/YEAR: III/II

UNIT - I: INTRODUCTION TO PREDICTIVE MODELING
Core ideas in data mining - Supervised and unsupervised learning - Classification vs. Prediction - Steps in data mining- SEMMA Approach - Sampling -Pre-processing – Data cleaning - Data Partitioning - Building a model - Statistical models - Statistical models for predictive analytics.

UNIT - I: PART – A

Q.No	Question	BT Level	Competence	Course Outcome
1	Define Data Mining.	BTL-1	Remember	CO1
2	What is meant by "data transformation" in data mining?	BTL-1	Remember	CO1
3	State the role of validation in supervised learning.	BTL-1	Remember	CO1
4	What is mean by Sampling?	BTL-1	Remember	CO1
5	Define a decision tree in the context of classification.	BTL-1	Remember	CO1
6	Infer the term “clustering” in unsupervised learning.	BTL-1	Remember	CO1
7	Identify the term "support vector machine" stands for in machine learning.	BTL-1	Remember	CO1
8	List the main advantage of using neural networks for classification.	BTL-1	Remember	CO1
9	Name the any two techniques of supervised learning.	BTL-1	Remember	CO1
10	Show the three types of unsupervised learning tasks.	BTL-1	Remember	CO1
11	Name two common applications of data mining.	BTL-1	Remember	CO1
12	Give one reason for using stratified sampling.	BTL-1	Remember	CO1
13	What does "overfitting" mean in the context of model building?	BTL-1	Remember	CO1
14	Expand the term SEMMA.	BTL-2	Understand	CO1
15	Write the tasks performed in Pre-processing.	BTL-2	Understand	CO1
16	Mention the role of Data cleaning.	BTL-2	Understand	CO1
17	How classifications vary from Prediction?	BTL-2	Understand	CO1
18	Differentiate between Supervised and unsupervised learning.	BTL-2	Understand	CO1

19	Specify the different ways of partitioning the Data.	BTL-2	Understand	CO1
20	Distinguish between classification and regression models.	BTL-2	Understand	CO1
21	Specify the differences between training and testing datasets.	BTL-2	Understand	CO1
22	How does k-means cluster work in unsupervised learning?	BTL-2	Understand	CO1
23	Write the purpose of feature selection in data mining.	BTL-2	Understand	CO1
24	Why do we plot a confusion matrix?	BTL-2	Understand	CO1
25	Why is it important to handle outliers?	BTL-2	Understand	CO1

UNIT - I: PART – B

Q.No	Question	Mark	BT Level	Competence	Course Outcome
1	Illustrate the types of Sampling techniques with an example.	16	BTL-3	Apply	CO1
2	Discuss the data splitting methods with neat sketch.	16	BTL-3	Apply	CO1
3	(i) Describe the process of "Data Partitioning" in model building. (ii) Discuss common partitioning ratios and the importance of using a separate test set for unbiased evaluation.	8 8	BTL-3	Apply	CO1
4	(i) Discuss various types of Statistical Models commonly employed in data mining. (ii) Explain the underlying principles, assumptions and typical applications of at least three statistical models.	8 8	BTL-3	Apply	CO1
5	Explain the concept of overfitting and underfitting in the context of model building.	16	BTL-3	Apply	CO1
6	Illustrate the challenges in handling imbalanced datasets in classification problems.	16	BTL-3	Apply	CO1
7	Explain the data transformation techniques of Standardization and Normalization.	16	BTL-3	Apply	CO1
8	Explain the role of ensemble methods like Random Forest and Boosting in improving the accuracy of predictive models.	16	BTL-3	Apply	CO1
9	Explain the steps in data mining process with neat diagram.	16	BTL-4	Analyze	CO1
10	Explain data cleaning tasks in detail and provide examples of each step.	16	BTL-4	Analyze	CO1
11	Analyzes the overfitting and underfitting causes, consequences and various techniques to prevent these issues.	16	BTL-4	Analyze	CO1
12	Describe the Cross-Industry Standard Process for Data Mining methodology and compare it with the SEMMA approach. Highlight their similarities and differences.	16	BTL-4	Analyze	CO1
13	(i) Explain in detail the various types of classification algorithms. (ii) Compare decision trees, k-nearest neighbors and support vector machines.	8 8	BTL-4	Apply	CO1

14	Summarizes the working principles of supervised and unsupervised learning.	16	BTL-5	Evaluate	CO1
15	Elaborate on the necessity and various techniques of "Data Pre-processing." Provide examples of issues addressed by each technique.	16	BTL-5	Evaluate	CO1
16	Compare Classification vs. Prediction with suitable example.	16	BTL-6	Create	CO1
17	Describe the process to build the statistical models for predictive analytics.	16	BTL-6	Create	CO1
18	Formulate with neat sketch the various phases involved in SEMMA Approach.	16	BTL-5	Evaluate	CO1

UNIT II PREDICTIVE MODELING BASICS

Data splitting – Balancing- Over fitting –Oversampling –Multiple Regression Artificial neural networks (MLP) - Variable importance- Profit/loss/prior probabilities – Model specification - Model selection - Multivariate Analysis.

UNIT II PART – A

Q.No	Questions	BT Level	Competence	Course Outcome
1	Define data splitting in machine learning.	BTL-1	Remember	CO2
2	Mention the purpose of balancing a dataset.	BTL-1	Remember	CO2
3	Define overfitting in the context of model training.	BTL-1	Remember	CO2
4	Specify the two techniques to prevent overfitting.	BTL-1	Remember	CO2
5	State the two advantages of using a multilayer perceptron.	BTL-1	Remember	CO2
6	What is multivariate analysis?	BTL-1	Remember	CO2
7	List the activation functions are commonly used in MLP.	BTL-1	Remember	CO2
8	What is backpropagation in neural networks?	BTL-1	Remember	CO2
9	Specify the meaning of R-squared in multiple regression.	BTL-1	Remember	CO2
10	State any two types of multivariate analysis techniques.	BTL-1	Remember	CO2
11	Name one advantage and one disadvantage of the SMOTE over simple random oversampling.	BTL-1	Remember	CO2
12	Write the basic equation for a multiple linear regression model with 'n' independent variables.	BTL-1	Remember	CO2
13	Name two criteria used for model selection.	BTL-1	Remember	CO2
14	What is oversampling and when is it used?	BTL-2	Understand	CO2
15	How is profit/loss matrix used in model evaluation?	BTL-2	Understand	CO2
16	Differentiate between overfitting and underfitting.	BTL-2	Understand	CO2
17	Distinguish between oversampling and undersampling.	BTL-2	Understand	CO2
18	Write the purpose of the test set.	BTL-2	Understand	CO2

19	How dependent variable vary from independent variables in regression?	BTL-2	Understand	CO2
20	Why is balancing important in classification problems?	BTL-2	Understand	CO2
21	What is the key difference between univariate and multivariate analysis?	BTL-2	Understand	CO2
22	How can you use a validation curve to detect overfitting?	BTL-2	Understand	CO2
23	Differentiate between AIC and BIC.	BTL-2	Understand	CO2
24	How does MLP handle non-linear relationships?	BTL-2	Understand	CO2
25	What is model fitting and model selection?	BTL-2	Understand	CO2

UNIT II PART – B					
Q.No	**Question**	**Mark**	**BT Level**	**Competence**	**Course Outcome**
1	Explain the process of data splitting and its significance in model development.	16	BTL-3	Apply	CO2
2	Describe multiple regression analysis with assumptions, formula, and a practical example.	16	BTL-3	Apply	CO2
3	Describe the architecture and working of a Multilayer Perceptron neural network.	16	BTL-3	Apply	CO2
4	(i) Describe the architecture of a Multi-Layer Perceptron. (ii) Explain the roles of input, hidden, and output layers.	8 8	BTL-3	Apply	CO2
5	Discuss the importance of profit/loss/prior probabilities in decision-making under uncertainty.	16	BTL-3	Apply	CO2
6	Describe regularization techniques L1 and L2 and their role in reducing overfitting.	16	BTL-3	Apply	CO2
7	Explain multivariate analysis techniques MANOVA in detail.	16	BTL-3	Apply	CO2
8	In many business applications, the cost of misclassification is not equal for all classes. Explain how to incorporate profit and loss considerations into the evaluation of a classification model.	16	BTL-3	Apply	CO2
9	(i) Discuss various techniques of balancing datasets. (ii) Compare oversampling and undersampling methods.	8 8	BTL-4	Analyze	CO2
10	Explain the concept of variable importance. How is it calculated in different models?	16	BTL-4	Analyze	CO2
11	Compare and contrast Artificial Neural Networks MLPs with traditional statistical models like logistic regression and linear regression.	16	BTL-4	Analyze	CO2
12	(i) Compare Principal Component Analysis and Factor Analysis. (ii) Explain the step-by-step process of performing PCA on a dataset.	6 10	BTL-4	Analyze	CO2

13	Explain how categorical variables can be incorporated into a multiple regression model.	16	BTL-4	Analyze	CO2
14	Evaluate various model selection techniques AIC, BIC and Cross-validation.	16	BTL-5	Evaluate	CO2
15	(i) Explain with the help of a diagram showing training and validation error. (ii) Elaborate on at least two distinct regularization techniques.	8 8	BTL-5	Evaluate	CO2
16	Describe the process of building a multiple linear regression model.	16	BTL-6	Create	CO2
17	Describe the process of building and validating a machine learning model.	16	BTL-6	Create	CO2
18	Discuss the applications of multivariate analysis in real-world problems.	16	BTL-6	Create	CO2

UNIT – III PREDICTIVE MODELS

Association Rules-Clustering Models –Decision Trees- Ruleset Models- KNearest Neighbors – Naive Bayes - Neural Network Model – Regression Models – Regression Trees – Classification & Regression Trees (CART) – Logistic Regression – Multiple Linear Regression Scorecards – Support Vector Machines – Time Series Models - Comparison between models - Lift chart Assessment of a single model.

Q.No	Questions	BT Level	Competence	Course Outcome
	UNIT III PART – A			
1	Mention the primary goal of Association Rule mining.	BTL-1	Remember	CO3
2	Define Support and Confidence in the context of Association Rules.	BTL-1	Remember	CO3
3	What is a Ruleset Model?	BTL-1	Remember	CO3
4	What is the 'curse of dimensionality' in the context of KNN?	BTL-1	Remember	CO3
5	State the primary purpose of a Regression Model.	BTL-1	Remember	CO3
6	List the three main components of a Time Series data.	BTL-1	Remember	CO3
7	State the key difference between Logistic Regression and Linear Regression.	BTL-1	Remember	CO3
8	What is the use of a validation set in model building?	BTL-1	Remember	CO3
9	Differentiate between partitioning and hierarchical clustering.	BTL-2	Understand	CO3
10	How does a Decision Tree handle a categorical predictor?	BTL-2	Understand	CO3
11	What is the core assumption of the Naive Bayes classifier?	BTL-1	Remember	CO3
12	What is the role of the 'k' in K-Means clustering?	BTL-1	Remember	CO3
13	Identify the 'k' represents in K-Means clustering?	BTL-1	Remember	CO3
14	Write the role of activation function in a Neural Network.	BTL-2	Understand	CO3
15	Distinguish between a neuron and a layer in a Neural Network.	BTL-2	Understand	CO3
16	How does a Classification and Regression Tree work?	BTL-2	Understand	CO3
17	What does the logistic function represent in Logistic Regression?	BTL-2	Understand	CO3

18	Specify the purpose of a scorecard in a business context.	BTL-2	Understand	CO3
19	How do you assess a single model's performance for classification?	BTL-2	Understand	CO3
20	What is a Confusion Matrix?	BTL-1	Remember	CO3
21	What is an ensemble model?	BTL-1	Remember	CO3
22	Write the purpose of Gini impurity used for in a CART.	BTL-2	Understand	CO3
23	Distinguish between supervised and unsupervised learning.	BTL-2	Understand	CO3
24	Differentiate between a hard margin and a soft margin SVM.	BTL-2	Understand	CO3
25	Infer the term 'Laplace smoothing' in Naive Bayes.	BTL-1	Remember	CO3

UNIT III PART – B					
Q.No	**Question**	**Mark**	**BT Level**	**Competence**	**Course Outcome**
1	Illustrate the architecture of a multi-layer perceptron with neat diagram.	16	BTL-3	Apply	CO3
2	Illustrate with an example how a regression tree makes a prediction.	16	BTL-3	Apply	CO3
3	(i) Discuss the Classification and Regression Trees algorithm in detail. (ii) Write briefly about the key features and advantages of the CART algorithm.	8 8	BTL-3	Apply	CO3
4	Explain the Autoregressive Integrated Moving Average modeling technique and its components.	16	BTL-3	Apply	CO3
5	Explain the concepts of bias and variance in the context of predictive modeling. Discuss the bias-variance tradeoff and its importance in model selection.	16	BTL-3	Apply	CO3
6	(i) Explain the Apriori algorithm for Association Rule mining in detail with a suitable example. (i) Discuss its advantages and disadvantages.	8 8	BTL-4	Analyze	CO3
7	Compare and contrast Decision Trees and Ruleset Models. What are the relative strengths and weaknesses of each approach in terms of interpretability, handling of different data types and computational complexity?	16	BTL-4	Analyze	CO3
8	(i) Describe the K-Means clustering algorithm. Explain the steps involved. (ii) Discuss the challenges associated with it and determining the optimal number of clusters.	8 8	BTL-5	Evaluate	CO3
9	Explain the complete process of building a Decision Tree model.	16	BTL-6	Create	CO3
10	Discuss the challenges and techniques for handling imbalanced datasets in classification problems.	16	BTL-3	Apply	CO3
11	Explain the theory behind Support Vector Machines. Provide a graphical illustration.	16	BTL-3	Apply	CO3

12	Describe the key components of a time series. Provide the different time series models.	16	BTL-3	Apply	CO3
13	Compare and contrast any four classification models.	16	BTL-4	Analyze	CO3
14	Describe the steps you would take to develop and implement a credit scorecard to predict the probability of loan default using a predictive modeling technique. Explain how the final scorecard would be used by a loan officer.	16	BTL-4	Analyze	CO3
15	Explain the purpose and construction of a lift chart. How do these charts help in assessing the performance of a predictive model? Provide a hypothetical example with a lift chart and interpret its meaning.	16	BTL-4	Apply	CO3
16	Formulate the K-Nearest Neighbors (KNN) algorithm in detail. Provide a scenario where KNN would be a suitable choice and one where it would not.	16	BTL-5	Evaluate	CO3
17	(i) Derive the Bayes' theorem and explain the 'naive' assumption of conditional independence. (ii) Discuss how to handle numerical features and the zero-frequency problem.	10 6	BTL-6	Create	CO3
18	(i) Derive the logistic function and explain how it is used to model the probability of a binary outcome. (ii) Discuss the interpretation of the coefficients in a logistic regression model.	9 7	BTL-6	Create	CO3

UNIT – IV PREDICTIVE MODELING MARKUP LANGUAGE				
Introduction to PMML – PMML Converter - PMML Structure – Data Manipulation in PMML – PMML Modeling Techniques - Multiple Model Support – Model Verification.				
Q.No	**Questions**	**BT Level**	**Competence**	**Course Outcome**
UNIT IVPART – A				
1	What does PMML stand for?	BTL-1	Remember	CO4
2	Name the XML-based standard that PMML is built upon.	BTL-1	Remember	CO4
3	What is a PMML converter?	BTL-1	Remember	CO4
4	List the two main components of a PMML file's header.	BTL-1	Remember	CO4
5	Define the role of the DataField element in the Data Dictionary.	BTL-1	Remember	CO4
6	State the role of TransformationDictionary in PMML.	BTL-1	Remember	CO4
7	Name two common data pre-processing operations that can be defined in PMML.	BTL-1	Remember	CO4
8	Mention the function of the MiningSchema within a PMML model.	BTL-1	Remember	CO4
9	List three modeling techniques supported by PMML.	BTL-1	Remember	CO4
10	Specify the MiningModel element used for in PMML.	BTL-1	Remember	CO4
11	Name one benefit of using PMML for model deployment.	BTL-1	Remember	CO4

12	What is the use of Extension element in PMML?	BTL-1	Remember	CO4
13	What information is stored in the Data Dictionary of a PMML file?	BTL-1	Remember	CO4
14	Write the primary purpose of PMML.	BTL-2	Understand	CO4
15	Differentiate between an active and a predicted field in the MiningSchema.	BTL-2	Understand	CO4
16	How does PMML support model ensembles?	BTL-2	Understand	CO4
17	Can PMML handle text data? If so, which element is used?	BTL-2	Understand	CO4
18	Distinguish between a DerivedField and a DataField.	BTL-2	Understand	CO4
19	Write a PMML Snippet for conditional transformation.	BTL-2	Understand	CO4
20	How are missing values handled in PMML?	BTL-2	Understand	CO4
21	Write a PMML Snippet for Concatenate Strings.	BTL-2	Understand	CO4
22	Draw a simple PMML Structure.	BTL-2	Understand	CO4
23	How does PMML ensure that a model is interpreted consistently across different platforms?	BTL-2	Understand	CO4
24	Can a single PMML file contain multiple models? Justify the answer.	BTL-2	Understand	CO4
25	How does a PMML consumer use the ModelVerification data?	BTL-2	Understand	CO4

UNIT IV PART – B					
Q.No	**Question**	**Mark**	**BT Level**	**Competence**	**Course Outcome**
1	Illustrate the overall structure of a PMML file.	16	BTL-3	Apply	CO4
2	Discuss the different strategies for imputing missing values and how they are represented in the PMML structure.	16	BTL-3	Apply	CO4
3	Discuss the different techniques for handle categorical variables and processing text data and how these are represented in PMML.	16	BTL-3	Apply	CO4
4	Discuss the role of PMML in enabling interoperability between different data mining tools and platforms.	16	BTL-3	Apply	CO4
5	Explain the key components of a typical PMML document with an example.	16	BTL-3	Apply	CO4
6	(i) Explain the working of a PMML converter. (ii) Briefly write about the challenges in converting complex models?	10 6	BTL-3	Apply	CO4
7	Discuss the role of following libraries in PMML conversion. (i) Nyoka (ii) Jpmml (iii) sklearn2pmml	 5 5 6	BTL-3	Apply	CO4
8	(i) Discuss about the purpose of the <ModelVerification> element in PMML. (ii) Describe how it can be used to store information about the model's performance and validation.	6 10	BTL-3	Apply	CO4

9	Describe the various data transformation techniques that can be expressed in PMML. Provide examples of how to perform normalization, discretization and value mapping using PMML elements.	16	BTL-4	Analyze	CO4
10	Compare between MiningField and DataField elements in PMML. Why are both required?	16	BTL-4	Analyze	CO4
11	Analyze the compatibility and performance issues encountered when converting models to PMML.	16	BTL-4	Analyze	CO4
12	Examine the steps involved in exporting a decision tree model to PMML using Python.	16	BTL-4	Analyze	CO4
13	Compare and contrast PMML with other model deployment formats ONNX and PFA.	16	BTL-4	Apply	CO4
14	Explain the representation in PMML for any two of modeling techniques.	16	BTL-5	Evaluate	CO4
15	Formulate the different data types and operational types supported by PMML and provide examples of how they are defined.	16	BTL-5	Evaluate	CO4
16	Describe the process of converting a trained model into a PMML file. Provide a code example.	16	BTL-6	Create	CO4
17	Explain how conditional logic if-then-else is implemented using the Apply element in PMML.	16	BTL-6	Create	CO4
18	Describe how multiple models can be combined using techniques and provide a use case for this approach.	16	BTL-4	Analyze	CO4

UNIT –V TECHNOLOGIES AND CASE STUDIES				
Weka – Rapid Miner – IBM SPSS Statistics- IBM SPSS Modeler – SAS Enterprise Miner – Apache Mahout – R Programming Language.-Real time case study with modeling and analysis.				
Q.No	**Questions**	**BT Level**	**Competence**	**Course Outcome**
UNIT VPART – A				
1	What is the primary use of Weka in data mining?	BTL-1	Remember	CO5
2	Define data preprocessing in RapidMiner.	BTL-1	Remember	CO5
3	Mention two types of models that can be created using IBM SPSS Modeler.	BTL-1	Remember	CO5
4	State the use of 'Filter' in Weka.	BTL-1	Remember	CO5
5	What is a Decision Tree in SAS Enterprise Miner?	BTL-1	Remember	CO5
6	List any two statistical tests available in IBM SPSS Statistics.	BTL-1	Remember	CO5
7	Specify the use of the "rpart" package in R.	BTL-1	Remember	CO5
8	Name two machine learning algorithms supported by Apache Mahout.	BTL-1	Remember	CO5
9	What is CRISP-DM in SPSS Modeler?	BTL-1	Remember	CO5
10	What kind of data format does Weka accept?	BTL-1	Remember	CO5

11	Mention any two visualization techniques in R.	BTL-1	Remember	CO5
12	State one advantage and one limitation of Apache Mahout.	BTL-1	Remember	CO5
13	Mention one use of the "caret" package in R.	BTL-1	Remember	CO5
14	Differentiate between supervised and unsupervised learning in RapidMiner.	BTL-2	Understand	CO5
15	How can clustering be implemented in Weka?	BTL-2	Understand	CO5
16	Write any two data transformation operators in RapidMiner.	BTL-2	Understand	CO5
17	Distinguish between linear and logistic regression in IBM SPSS Statistics.	BTL-2	Understand	CO5
18	How does SAS Enterprise Miner handle missing values?	BTL-2	Understand	CO5
19	Write the two types of regression models available in SPSS Statistics.	BTL-2	Understand	CO5
20	How does RapidMiner's visual workflow paradigm aid in the data mining process?	BTL-2	Understand	CO5
21	How does RapidMiner support the deployment of machine learning models?	BTL-2	Understand	CO5
22	How does the "drag-and-drop" functionality of RapidMiner?	BTL-2	Understand	CO5
23	Identify one statistical analysis technique that IBM SPSS Statistics is particularly well-suited for.	BTL-2	Understand	CO5
24	Write the primary difference between IBM SPSS Statistics and IBM SPSS Modeler.	BTL-2	Understand	CO5
25	Give an example of a descriptive statistic that can be calculated using IBM SPSS Statistics.	BTL-2	Understand	CO5

UNIT V PART – B					
Q.No	**Question**	**Mark**	**BT Level**	**Competence**	**Course Outcome**
1	Explain how IBM SPSS Modeler supports the CRISP-DM methodology. Include a sample use case from the finance domain.	16	BTL-3	Apply	CO5
2	Discuss the role of SAS Enterprise Miner in predictive analytics. Include a healthcare industry case study.	16	BTL-3	Apply	CO5
3	Explain collaborative filtering and content-based filtering in Apache Mahout. Provide code snippets or pseudo-code.	16	BTL-3	Apply	CO5
4	(i) Describe the architecture of Apache Mahout and its integration with Hadoop. (ii) Include a real-time recommendation system example.	10 6	BTL-3	Apply	CO5
5	Illustrate time series forecasting using IBM SPSS Statistics. Use a stock price dataset.	16	BTL-3	Apply	CO5
6	Discuss the process of data cleansing and feature selection in SAS Enterprise Miner using an insurance claims dataset.	16	BTL-3	Apply	CO5
7	Explain the differences in workflow-based modeling between RapidMiner and SPSS Modeler.	16	BTL-3	Apply	CO5
8	Discuss the significance of ensemble models. Build them using Weka and R.	16	BTL-3	Apply	CO5

9	Analyze a customer churn dataset using Weka. Detail all the steps from data import to model evaluation.	16	BTL-4	Analyze	CO5
10	Compare and contrast the data mining capabilities of Weka and RapidMiner with appropriate examples.	16	BTL-4	Analyze	CO5
11	Perform sentiment analysis on Twitter data using R. Discuss data preprocessing, modeling, and results interpretation.	16	BTL-4	Analyze	CO5
12	(i) Compare classification algorithms across Weka, R, and SPSS. (ii) Include performance metrics and visualizations.	10 6	BTL-4	Analyze	CO5
13	Analyze customer segmentation using K-means clustering in IBM SPSS Modeler. Include visualizations and actionable insights.	16	BTL-4	Analyze	CO5
14	Use RapidMiner to detect fraud in banking transactions. Explain each phase from ETL to model validation.	16	BTL-5	Evaluate	CO5
15	Evaluate different clustering techniques in Weka and IBM SPSS Modeler using the same dataset.	16	BTL-5	Evaluate	CO5
16	Create a marketing campaign response model using SAS Enterprise Miner. Explain variable selection, model building, and deployment.	16	BTL-6	Create	CO5
17	(i) Design and implement a credit scoring model using IBM SPSS Statistics. (ii) Describe the types of regression used and their suitability.	10 6	BTL-6	Create	CO5
18	Describe the integration of R with Weka for advanced modeling techniques.	16	BTL-5	Evaluate	CO5