

SRM VALLIAMMAI ENGINEERING COLLEGE
(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203
DEPARTMENT OF MANAGEMENT STUDIES

QUESTION BANK

III SEMESTER

PBA402 DATA MINING FOR BUSINESS INTELLIGENCE

Regulation – 2023

Academic Year 2026 – 2027



Prepared by

Ms. G. SATHYA, AP/MBA

UNIT 1 : INTRODUCTION

Data Mining Definitions, Applications, Origins, Growth, Text Mining, Web Mining, Spatial Mining, Process Mining, Data Ware House and Data Marts.

PART- A

S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Define Data Mining.	Remembering	Level 1	CO1
2.	What is Knowledge Discovery in Databases (KDD)?	Understanding	Level 2	CO1
3.	List any four applications of Data Mining.	Remembering	Level 1	CO1
4.	Differentiate between Data Mining and Database Management.	Understanding	Level 2	CO1
5.	What are the major tasks of Data Mining?	Remembering	Level 1	CO1
6.	What is predictive data mining?	Understanding	Level 2	CO1
7.	What is descriptive data mining?	Remembering	Level 1	CO1
8.	State the origins of Data Mining.	Understanding	Level 2	CO1
9.	Mention the factors responsible for the growth of Data Mining.	Remembering	Level 1	CO1
10.	Define Text Mining.	Understanding	Level 2	CO1
11.	List any two applications of Text Mining.	Remembering	Level 1	CO1
12.	What is Web Mining?	Understanding	Level 2	CO1
13.	Name the three categories of Web Mining.	Remembering	Level 1	CO1
14.	Define Web Content Mining.	Understanding	Level 2	CO1
15.	What is Web Structure Mining?	Remembering	Level 1	CO1
16.	What is Web Usage Mining?	Understanding	Level 2	CO1
17.	Define Spatial Data Mining.	Remembering	Level 1	CO1
18.	Give two applications of Spatial Data Mining.	Understanding	Level 2	CO1
19.	What is Process Mining?	Remembering	Level 1	CO1
20.	State the importance of Process Mining in business organizations.	Understanding	Level 2	CO1
21.	Define Data Warehouse.	Remembering	Level 1	CO1

22.	What are the characteristics of a Data Warehouse?	Understanding	Level 2	CO1
23.	What is a Data Mart?	Remembering	Level 1	CO1
24.	Differentiate between Data Warehouse and Data Mart.	Understanding	Level 2	CO1

	PART- B			
S.NO	QUESTIONS	COMPETENCE	BTL	CO Level
1.	Explain the concept of Data Mining and discuss its importance in modern business environments. Illustrate with suitable applications.	Apply	Level 3	CO1
2.	Describe the Knowledge Discovery in Databases (KDD) process in detail. Explain each stage with examples.	Analyse	Level 4	CO1
3.	Discuss the origins and evolution of Data Mining. Explain the technological and business factors that contributed to its growth.	Apply	Level 3	CO1
4.	Explain various Data Mining functionalities such as classification, prediction, clustering, association analysis, and outlier detection.	Analyse	Level 4	CO1
5.	Discuss the applications of Data Mining in banking, retail, healthcare, telecommunications, and e-commerce sectors.	Apply	Level 3	CO1
6.	Define Text Mining. Explain the Text Mining process, techniques, challenges, and applications.	Analyse	Level 4	CO1
7.	Explain the architecture and working methodology of Text Mining systems with suitable examples.	Apply	Level 3	CO1
8.	What is Web Mining? Explain its categories—Web Content Mining, Web Structure Mining, and Web Usage Mining—with examples.	Analyse	Level 4	CO1
9.	Discuss the techniques, applications, advantages, and limitations of Web Mining.	Apply	Level 3	CO1
10.	Explain Spatial Data Mining in detail. Discuss spatial data types, techniques, and real-world applications.	Analyse	Level 4	CO1
11.	Describe the challenges involved in Spatial Data Mining and the methods used to address them.	Apply	Level 3	CO1

12.	Define Process Mining. Explain its types, namely Process Discovery, Conformance Checking, and Process Enhancement.	Analyse	Level 4	CO1
13.	Discuss the role of Process Mining in business process improvement with suitable case studies or examples.	Apply	Level 3	CO1
14.	Explain the concept, architecture, and characteristics of a Data Warehouse.	Analyse	Level 4	CO1
15.	Discuss the various components of a Data Warehouse and explain the ETL (Extract, Transform, Load) process in detail.	Apply	Level 3	CO1
16.	Compare Data Warehouse and Data Mart. Explain different types of Data Marts with suitable examples.	Analyse	Level 4	CO1
17.	Explain the relationship between Data Mining, Data Warehousing, Text Mining, Web Mining, Spatial Mining, and Process Mining in modern data analytics environments.	Analyse	Level 4	CO1

UNIT II : DATA MINING PROCESS

Data Mining Process – KDD, CRISP-DM, SEMMA and Domain – Specific, Classification and Prediction Performance Measures – RSME, MAD, MAP, MAPE, Confusion Matrix, Receiver Operating Characteristic Curve & AUC; Validation Techniques – Hold-Out, K-Fold Cross-Validation, LOOCV, Random Subsampling, and Bootstrapping.

PART- A				
S.NO	QUESTIONS	COMPETENCE	BTL	CO Level
1.	Define Knowledge Discovery in Databases (KDD).	Remembering	Level 1	CO2
2.	List the major steps involved in the KDD process.	Understanding	Level 2	CO2
3.	What is data preprocessing in KDD?	Remembering	Level 1	CO2
4.	Define CRISP-DM.	Understanding	Level 2	CO2
5.	Expand CRISP-DM.	Remembering	Level 1	CO2
6.	List the phases of the CRISP-DM model.	Understanding	Level 2	CO2
7.	What is SEMMA?	Remembering	Level 1	CO2
8.	Expand SEMMA.	Understanding	Level 2	CO2
9.	Name the stages of the SEMMA methodology.	Remembering	Level 1	CO2
10.	Differentiate between KDD and CRISP-DM.	Remembering	Level 1	CO2

11.	What is Domain-Specific Data Mining?	Understanding	Level 2	CO2
12.	Give two examples of Domain-Specific Data Mining applications.	Remembering	Level 1	CO2
13.	Define classification in data mining.	Understanding	Level 2	CO2
14.	What is prediction in data mining?	Remembering	Level 1	CO2
15.	Define Root Mean Squared Error.	Understanding	Level 2	CO2
16.	What is Mean Absolute Deviation (MAD)?	Remembering	Level 1	CO2
17.	Define Mean Absolute Percentage Error (MAPE).	Remembering	Level 1	CO2
18.	What is Mean Percentage Error (MPE/MAP)?	Remembering	Level 1	CO2
19.	What is a Confusion Matrix?	Remembering	Level 1	CO2
20.	What is meant by True Positive (TP) and False Positive (FP)?	Understanding	Level 2	CO2
21.	Define Receiver Operating Characteristic (ROC) Curve.	Remembering	Level 1	CO2
22.	What is Area Under the Curve (AUC)?	Understanding	Level 2	CO2
23.	What is Hold-Out Validation?	Understanding	Level 2	CO2
24.	Define K-Fold Cross-Validation.	Remembering	Level 1	CO2

PART- B				
S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Explain the KDD process in detail. Discuss each stage with suitable examples.	Analyse	Level 4	CO2
2.	Describe the CRISP-DM methodology. Explain all six phases and their significance in a data mining project.	Apply	Level 3	CO2
3.	Compare and contrast KDD and CRISP-DM frameworks. Highlight their advantages and limitations.	Analyse	Level 4	CO2
4.	Explain the SEMMA methodology in detail. Discuss each phase with a practical example.	Apply	Level 3	CO2
5.	Compare KDD, CRISP-DM, and SEMMA methodologies used in data mining projects.	Analyse	Level 4	CO2

6.	What is Domain-Specific Data Mining? Discuss its importance and applications in healthcare, finance, retail, manufacturing, and social media analytics.	Apply	Level 3	CO2
7.	Explain classification and prediction tasks in data mining. Discuss their objectives and real-world applications.	Analyse	Level 4	CO2
8.	Describe various performance measures used for prediction models. Explain RMSE, MAD, MAP, and MAPE with suitable examples.	Apply	Level 3	CO2
9.	Discuss the importance of error measures in predictive analytics. Compare RMSE, MAD, and MAPE in terms of interpretation and usage.	Analyse	Level 4	CO2
10.	Explain the concept and structure of a Confusion Matrix. Discuss accuracy, precision, recall, specificity, and F1-score derived from it.	Apply	Level 3	CO2
11.	Describe the Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC). Explain their role in classifier evaluation	Analyse	Level 4	CO2
12.	Explain how ROC curves are constructed. Discuss the interpretation of different AUC values with examples.	Apply	Level 3	CO2
13.	Discuss Hold-Out Validation as a model evaluation technique. Explain its advantages and limitations.	Analyse	Level 4	CO2
14.	Explain K-Fold Cross-Validation in detail. Compare it with Hold-Out Validation.	Apply	Level 3	CO2
15.	Describe Leave-One-Out Cross-Validation (LOOCV). Discuss its benefits and computational challenges.	Analyse	Level 4	CO2
16.	Explain Random Subsampling and Bootstrapping techniques for model validation. Compare their characteristics and applications.	Apply	Level 3	CO2
17.	Compare all major validation techniques—Hold-Out, K-Fold Cross-Validation, LOOCV, Random Subsampling, and Bootstrapping—with suitable examples and use cases.	Apply	Level 3	CO2

UNIT III : PREDICTION TECHNIQUES

Data Visualization, Time Series – ARIMA, Winter Holts. Vector Autoregressive Analysis, Multivariate Regression Analysis

	PART- A			
S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Define Data Visualization.	Remember	Level 1	CO3
2.	State the objectives of Data Visualization.	Understand	Level 2	CO3
3.	What are the characteristics of effective data visualization?	Remember	Level 1	CO3
4.	Differentiate between static and interactive visualization.	Understand	Level 2	CO3
5.	Define Time Series Data.	Remember	Level 1	CO3
6.	What are the components of a time series?	Understand	Level 2	CO3
7.	What is a trend component in a time series?	Remember	Level 1	CO3
8.	Define seasonality in time series analysis.	Understand	Level 2	CO3
9.	What is stationarity in a time series?	Remember	Level 1	CO3
10.	Expand ARIMA.	Understand	Level 2	CO3
11.	What are the three parameters of an ARIMA model?	Remember	Level 1	CO3
12.	Define the autoregressive (AR) component.	Understand	Level 2	CO3
13.	What is the moving average (MA) component?	Remember	Level 1	CO3
14.	What is differencing in ARIMA?	Understand	Level 2	CO3
15.	Define Holt-Winters Method.	Remember	Level 1	CO3
16.	What are the types of Holt-Winters models?	Understand	Level 2	CO3
17.	What is exponential smoothing?	Remember	Level 1	CO3
18.	Define Vector Autoregressive (VAR) Analysis.	Understand	Level 2	CO3
19.	What is the purpose of VAR models?	Remember	Level 1	CO3
20.	Define Multivariate Regression Analysis.	Understand	Level 2	CO3
21.	Differentiate between simple and multivariate regression.	Remember	Level 1	CO3
22.	What is a dependent variable in regression analysis?	Understand	Level 2	CO3
23.	What is a regression coefficient?	Remember	Level 1	CO3
24.	Mention any two applications of multivariate regression analysis	Understand	Level 2	CO3

	PART- B			
S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Explain the concept of Data Visualization. Discuss its principles, benefits, and applications in business intelligence and data mining.	Apply	Level 3	CO3
2.	Discuss various types of data visualization techniques such as bar charts, line charts, scatter plots, heat maps, dashboards, and geographical maps with suitable examples.	Analyse	Level 4	CO3
3.	Explain Time Series Analysis in detail. Discuss its components namely trend, seasonal, cyclical, and irregular variations.	Apply	Level 3	CO3
4.	Describe the process of time series forecasting. Explain the importance of stationarity and data preprocessing.	Analyse	Level 4	CO3
5.	Explain the ARIMA model in detail. Discuss the roles of autoregression, differencing, and moving averages in forecasting.	Apply	Level 3	CO3
6.	Describe the steps involved in building an ARIMA model. Explain model identification, parameter estimation, diagnostic checking, and forecasting.	Analyse	Level 4	CO3
7.	Discuss the advantages, limitations, and applications of ARIMA models in business and economic forecasting.	Apply	Level 3	CO3
8.	Explain the Holt-Winters forecasting method. Discuss additive and multiplicative models with suitable examples.	Analyse	Level 4	CO3
9.	Compare ARIMA and Holt-Winters models with respect to assumptions, forecasting accuracy, and applications.	Apply	Level 3	CO3
10.	Explain exponential smoothing techniques and discuss their significance in time series forecasting.	Analyse	Level 4	CO3
11.	Define Vector Autoregressive (VAR) Analysis. Explain the mathematical model and assumptions underlying VAR.	Apply	Level 3	CO3
12.	Discuss the methodology of VAR modeling. Explain lag selection, model estimation, impulse response analysis, and forecasting.	Analyse	Level 4	CO3

13.	Explain the applications of VAR models in economics, finance, marketing, and business analytics.	Apply	Level 3	CO3
14.	Define Multivariate Regression Analysis. Explain the regression model, assumptions, and parameter estimation techniques.	Analyse	Level 4	CO3
15.	Discuss the interpretation of regression coefficients, goodness-of-fit measures, and significance testing in multivariate regression.	Apply	Level 3	CO3
16.	Compare Simple Regression, Multiple Regression, and Multivariate Regression Analysis with suitable examples.	Analyse	Level 4	CO3
17.	Explain how Data Visualization, Time Series Forecasting, VAR Analysis, and Multivariate Regression can be integrated for predictive analytics and decision support systems.	Apply	Level 3	CO3

UNIT IV : CLASSIFICATION AND CLUSTERING TECHNIQUES

Classification – Decision Trees, K Nearest Neighbours, Logistic Regression, Discriminant Analysis; Clustering; Market Based Analysis.

PART- A				
S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Define Classification in Data Mining.	Understand	Level 1	CO4
2.	What is a classifier?	Understand	Level 1	CO4
3.	Define supervised learning.	Understand	Level 1	CO4
4.	What is a Decision Tree?	Understand	Level 1	CO4
5.	Name any two Decision Tree algorithms.	Understand	Level 1	CO4
6.	What is the root node in a Decision Tree?	Understand	Level 1	CO4
7.	Define entropy in Decision Tree learning.	Understand	Level 1	CO4
8.	What is information gain?	Understand	Level 1	CO4
9.	Define K-Nearest Neighbours (KNN).	Remember	Level 2	CO4
10.	What is the role of parameter K in KNN?	Remember	Level 2	CO4
11.	What is distance measure in KNN?	Remember	Level 2	CO4
12.	Name any two distance metrics used in KNN.	Understand	Level 2	CO4

13.	Define Logistic Regression.	Remember	Level 2	CO4
14.	Why is Logistic Regression used for classification?	Remember	Level 2	CO4
15.	What is the sigmoid function?	Remember	Level 2	CO4
16.	Define Discriminant Analysis.	Remember	Level 2	CO4
17.	Differentiate between classification and clustering.	Remember	Level 2	CO4
18.	Define Clustering.	Remember	Level 2	CO4
19.	What is an unsupervised learning technique?	Understand	Level 1	CO4
20.	Define centroid in clustering.	Understand	Level 1	CO4
21.	What is K-Means clustering?	Understand	Level 1	CO4
22.	Define Market Basket Analysis.	Remember	Level 2	CO4
23.	What is an association rule?	Understand	Level 1	CO4
24.	Define support and confidence.	Remember	Level 2	CO4

PART- B				
S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Explain the concept of classification in data mining. Discuss its objectives, characteristics, and applications.	Apply	Level 3	CO4
2.	Describe the structure of a Decision Tree. Explain the process of Decision Tree construction with a suitable example.	Apply	Level 3	CO4
3.	Explain entropy and information gain in Decision Trees. Illustrate how they are used for attribute selection.	Apply	Level 3	CO4
4.	Discuss the advantages, limitations, and applications of Decision Tree classifiers.	Apply	Level 3	CO4
5.	Explain the K-Nearest Neighbours (KNN) algorithm. Discuss the role of distance measures and parameter selection.	Apply	Level 3	CO4
6.	Describe the working of the KNN algorithm with a suitable example. Discuss its strengths and weaknesses.	Apply	Level 3	CO4
7.	Explain Logistic Regression in detail. Discuss the logistic	Apply	Level 3	CO4

	function, model formulation, and classification process.			
8.	Compare Linear Regression and Logistic Regression with suitable examples.	Apply	Level 3	CO4
9.	Explain the assumptions, methodology, and applications of Logistic Regression in predictive analytics.	Apply	Level 3	CO4
10.	Define Discriminant Analysis. Explain Linear Discriminant Analysis (LDA) and its role in classification.	Apply	Level 3	CO4
11.	Compare Logistic Regression and Discriminant Analysis with respect to assumptions, accuracy, and applications.	Apply	Level 3	CO4
12.	Explain the concept of Clustering. Discuss its importance and applications in data mining.	Analyze	Level 4	CO4
13.	Describe the K-Means Clustering algorithm in detail. Explain its working with suitable examples.	Analyze	Level 4	CO4
14.	Compare Classification and Clustering techniques with respect to objectives, methods, and applications.	Analyze	Level 4	CO4
15.	Explain Market Basket Analysis. Discuss its significance in retail and e-commerce industries.	Analyze	Level 4	CO4
16.	Describe association rule mining and explain the measures Support, Confidence, and Lift with examples.	Analyze	Level 4	CO4
17.	Explain how Decision Trees, KNN, Logistic Regression, Discriminant Analysis, Clustering, and Market Basket Analysis contribute to business intelligence and decision-making systems.	Analyze	Level 4	CO4

UNIT V : MACHINE LEARNING AND AI				
Genetic Algorithms, Neural Network, Fuzzy Logic, Support Vector Machine, Optimization Techniques – Ant Colony, Particle Swarm, DEA.				
	PART- A			
S.NO	QUESTIONS	COMPETENCE	T LEVEL	CO Level
1.	Define Genetic Algorithm (GA).	Remember	Level 1	CO5
2.	What are the basic components of a Genetic Algorithm?	Understand	Level 2	CO5
3.	What is a chromosome in Genetic Algorithms?	Remember	Level 1	CO5
4.	Define fitness function.	Understand	Level 2	CO5

5.	What is crossover in Genetic Algorithms?	Remember	Level 1	CO5
6.	Define mutation operation.	Understand	Level 2	CO5
7.	What is selection in Genetic Algorithms?	Remember	Level 1	CO5
8.	Define Neural Network.	Understand	Level 2	CO5
9.	What is an artificial neuron?	Remember	Level 1	CO5
10.	What are the layers of an Artificial Neural Network (ANN)?	Understand	Level 2	CO5
11.	Define activation function.	Remember	Level 1	CO5
12.	What is backpropagation?	Understand	Level 2	CO5
13.	Define Fuzzy Logic.	Remember	Level 1	CO5
14.	What is a fuzzy set?	Understand	Level 2	CO5
15.	Differentiate between crisp sets and fuzzy sets.	Remember	Level 1	CO5
16.	What is a membership function?	Understand	Level 2	CO5
17.	Define Support Vector Machine (SVM).	Remember	Level 1	CO5
18.	What is a support vector?	Understand	Level 2	CO5
19.	What is the purpose of a hyperplane in SVM?	Remember	Level 1	CO5
20.	Define Ant Colony Optimization (ACO).	Understand	Level 2	CO5
21.	What is pheromone in Ant Colony Optimization?	Remember	Level 1	CO5
22.	Define Particle Swarm Optimization (PSO).	Understand	Level 2	CO5
23.	What is Data Envelopment Analysis (DEA)?	Remember	Level 1	CO5
24.	Mention any two applications of DEA.	Understand	Level 2	CO5

PART- B				
S.NO	QUESTIONS	COMPETENCE	BT LEVEL	CO Level
1.	Explain the concept of Genetic Algorithms. Discuss the principles of evolutionary computation and their significance in optimization problems.	Apply	Level 3	CO5

2.	Describe the working of a Genetic Algorithm. Explain chromosome representation, fitness evaluation, selection, crossover, and mutation operations.	Analyse	Level 4	CO5
3.	Discuss the advantages, limitations, and applications of Genetic Algorithms in data mining and machine learning.	Apply	Level 3	CO5
4.	Explain the architecture of Artificial Neural Networks. Discuss the functions of input, hidden, and output layers.	Analyse	Level 4	CO5
5.	Describe the learning process in Neural Networks. Explain supervised learning and the backpropagation algorithm.	Apply	Level 3	CO5
6.	Discuss various activation functions used in Neural Networks and their impact on model performance.	Analyse	Level 4	CO5
7.	Explain the concept of Fuzzy Logic. Discuss fuzzy sets, membership functions, and fuzzy inference systems.	Apply	Level 3	CO5
8.	Compare classical Boolean logic and Fuzzy Logic. Explain their advantages and applications.	Analyse	Level 4	CO5
9.	Describe the structure and working of a Fuzzy Inference System with suitable examples.	Apply	Level 3	CO5
10.	Explain Support Vector Machines (SVM) in detail. Discuss hyperplanes, support vectors, and margin maximization.	Analyse	Level 4	CO5
11.	Describe linear and non-linear SVM classification. Explain the role of kernel functions in SVM.	Apply	Level 3	CO5
12.	Compare Neural Networks and Support Vector Machines with respect to methodology, accuracy, and applications.	Analyse	Level 4	CO5
13.	Explain Ant Colony Optimization (ACO). Discuss the pheromone update mechanism and solution construction process.	Apply	Level 3	CO5
14.	Describe the Particle Swarm Optimization (PSO) algorithm. Explain particle movement, velocity update, and convergence.	Analyse	Level 4	CO5
15.	Compare Genetic Algorithms, Ant Colony Optimization, and Particle Swarm Optimization techniques.	Apply	Level 3	CO5

16.	Explain Data Envelopment Analysis (DEA). Discuss its mathematical foundation, efficiency measurement process, and applications.	Analyse	Level 4	CO5
17.	Discuss the role of Genetic Algorithms, Neural Networks, Fuzzy Logic, SVM, ACO, PSO, and DEA in modern intelligent data mining and optimization systems	Apply	Level 3	CO5