SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

SYLLABUS

UNIT –1: OVERVIEW OF PREDICTIVE ANALYTICS

What Is Analytics? What Is Predictive Analytics? Business Intelligence Predictive Analytics vs. Business Intelligence, Predictive Analytics vs. Statistics, Predictive Analytics vs. Data Mining, Who Uses Predictive Analytics? , Challenges in Using Predictive Analytics, What Educational Background Is Needed to Become a Predictive Modeler? Setting Up the Problem: Predictive Analytics Processing Steps: CRISP-DM, Business Understanding, Defining Data for Predictive Modelling, Defining the Target Variable, Defining Measures of Success for Predictive Models, Doing Predictive Modelling Out of Order, Case studyRecovering Lapsed Donors, Fraud Detection

UNIT-II

Data Understanding: What the Data Looks Like, Single Variable Summaries, Data Visualization in One Dimension, Histograms, Multiple Variable Summaries, Data Visualization, Two or Higher Dimensions, The Value of Statistical Significance, Pulling It All Together into a Data Audit. Data Preparation: Variable Cleaning, Feature Creation.

UNIT-III

Item sets and Association Rules: Terminology, Parameter Settings, How the Data Is Organized, Measures of Interesting Rules, Deploying Association Rules, Problems with Association Rules, Building Classification Rules from Association Rules. Descriptive Modeling: Data Preparation Issues with Descriptive Modelling, Principal Component Analysis, Clustering Algorithms. Interpreting Descriptive Models: Standard Cluster Model Interpretation.

UNIT-IV

Predictive Modelling: Decision Trees, Logistic Regression, Neural Networks, K-Nearest Neighbour, Naïve Bayes, Regression Models, Linear Regression, Other Regression Algorithms. Assessing Predictive Models: Batch Approach to Model Assessment, Assessing Regression Models.

UNIT-V

Model Ensembles: Motivation for Ensembles, Bagging, Boosting, Improvements to Bagging and Boosting, Model Ensembles and Occam‘s Razor, Interpreting Model Ensembles. Text Mining: Motivation for Text Mining, A Predictive Modelling Approach to Text Mining, structured vs. Unstructured Data, Why Text Mining Is Hard, Data Preparation Steps, Text Mining Features, Modelling with Text Mining Features, Regular Expressions. Model Deployment: General Deployment Considerations. Case Studies: Survey Analysis Case Study, Help Desk Case Study.

Prepared by XXXXXX

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

Max Marks: 10

S.No.	CO	Questions	BT
Unit I: (OVERVIEW OF PREDICTIVE ANALYTICS)			
1	1	**a.** What is the main difference between Predictive Analytics and Business Intelligence? **b.** Give two real-life examples where Predictive Analytics is better than BI.	L4
2	1	**a.** Define Predictive Analytics and explain how it is different from traditional Statistics. **b.** How does Predictive Analytics differ from Data Mining? Provide one real-world example for each difference.	L3
3	1	**a.** Who uses Predictive Analytics? Discuss its applications in at least three different industries with one specific example for each. **b.** List and briefly explain any four major challenges commonly faced in Predictive Analytics projects.	L4
4	1	**a.** What educational background and key skills are typically required to become a successful Predictive Modeler? Mention relevant degrees, programming tools, and soft skills. **b.** What are the measures of success for predictive models?	L3
5	1	a. Explain the six phases of CRISP-DM in sequence, and briefly describe the key activities in the Business Understanding phase? b. Why is CRISP- DM considered an iterative process?	L5
6	1	What is the importance of clearly defining the data requirements during the setup of a Predictive Analytics project?	L4
7	1	Name and briefly describe two common success measures for classification models and two for regression models.	L3
8	1	Give one example of how poor data selection can impact model performance.	L5
10	1	Describe how Predictive Analytics is used in the case study of recovering lapsed donors in a nonprofit organization.	L3
11	1	How is Predictive Analytics applied for fraud detection in banks?	L3

Prepared by XXXXXX

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

S.No.	CO	Questions	BT
		Unit II:	
1	2	**a.** List four single-variable summary statistics. **b.** When do we prefer median over mean?	L4
2	2	**a.** Explain how to construct a histogram for a continuous variable, including bin selection, and discuss its role in understanding data distribution. **b.** Create a one-dimensional visualization (e.g., box plot) for the same variable and interpret what it reveals about the data's distribution that summaries alone might miss.	L3
3	2	**a.** For two variables compute and interpret a multiple variable summary like correlation analysis and contingency tables. Explain using example. **b.** Design a two-dimensional visualization (e.g., scatter plot) for exploring the relationship between two variables.	L4
4	2	Identify common issues in variable cleaning and describe methods to address them.	L3
5	2	What is meant by feature creation? How many ways we can create new features and explain how they enhance data understanding or preparation for analysis.	L5
6	2	Describe the process of conducting a full data audit, starting from initial data inspection through summaries and one/two-dimensional visualizations, to identifying issues requiring variable cleaning.	L4
7	2	Explain common variable cleaning .Then describe feature creation techniques and explain how they enhance model accuracy and interpretability. Use simple examples to illustrate.	L5
8	2	**Dataset:** **22, 25, 25, 28, 30, 32, 35, 38, 40, 42, 45, 48, 50, 55, 60, 65, 70, 75, 78, 80** a) Compute and list the following single-variable summary statistics for this dataset: a) Mean **b)**Median c) Mode d)Standard deviation b) Construct a histogram for this dataset. Consider 5 bins.	L4

Prepared by XXXXXX

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

9	**2**	Consider the following small sample dataset for 10 students showing Hours Studied per week and their final Exam Score. a) Construct a scatter plot. b) Describe the pattern: Does it look linear?	**L4**

Student	Hours Studied	Exam Score
1	4.4	63
2	9.6	100
3	7.6	82
4	6.4	75
5	2.4	56
6	2.4	39
7	1.5	35
8	8.8	88
9	6.4	70
10	7.4	87

S.No.	CO	Questions	BT
		Unit III:	
1	**3**	**a.** Define support, confidence, and lift for association rules. **b.** Explain how lift helps measure interesting rules.	**L4**
2	**3**	a. Name the main differences between standard predictive data format and transactional format for association rules. b. Why do we need to convert data to transactional format?	**L3**
3	**3**	a. List four common problems with association rules. b. .Describe how to build classification rules from association rules.	**L4**
4	**3**	a. Explain the basic steps of Principal Component Analysis (PCA). b. How does PCA help in data interpretation?	**L3**
5	**3**	a. Describe the K-Means clustering algorithm in simple steps. b. How do we choose the number of clusters (K)?	**L5**

Prepared by XXXXXX

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

<table>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>6</td><td>3</td><td>a.What is a cluster prototype?

b. How do we use cluster prototypes to interpret results?</td><td>L4</td></tr>
<tr><td>7</td><td>3</td><td>a.What does "deploying association rules" mean in predictive analytics?

b. Give two real-world business examples of deploying association rules.</td><td>L3</td></tr>
<tr><td>8</td><td>3</td><td>Consider the following transactional dataset. Define the following association rule mining terms clearly and briefly:
• Item
• Itemset
• Support
• Association rule
• Support
• Confidence
• Lift

<table><tr><th>T.ID</th><th>Items Purchased</th></tr><tr><td>T1</td><td>Bread, Milk, Butter</td></tr><tr><td>T2</td><td>Bread, Milk, Jam</td></tr><tr><td>T3</td><td>Milk, Butter, Curd</td></tr><tr><td>T4</td><td>Bread, Butter, Jam, Juice</td></tr><tr><td>T5</td><td>Milk, Curd, Juice</td></tr></table></td><td>L5</td></tr>
<tr><td>9</td><td>3</td><td>Consider the following transactional dataset.

Solve the problem using apriori algorithm.

<table><tr><th>T. ID</th><th>Items Purchased</th></tr><tr><td>T1</td><td>Bread, Milk, Butter</td></tr><tr><td>T2</td><td>Bread, Milk, Jam</td></tr><tr><td>T3</td><td>Milk, Butter, Curd</td></tr><tr><td>T4</td><td>Bread, Butter, Jam, Juice</td></tr><tr><td>T5</td><td>Milk, Curd, Juice</td></tr></table></td><td>L4</td></tr>
</table>

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

10	**3**	<table><tr><td>x1</td><td>2.5</td><td>0.5</td><td>2.2</td><td>1.9</td><td>3.1</td><td>2.3</td><td>2.0</td><td>1.0</td><td>1.5</td><td>1.1</td></tr><tr><td>x2</td><td>2.4</td><td>0.7</td><td>2.9</td><td>2.2</td><td>3.0</td><td>2.7</td><td>1.6</td><td>1.1</td><td>1.6</td><td>0.9</td></tr></table> Consider the above dataset and generate the PCA components?	**L3**
11	**3**	Apply K(=3)-Means algorithm over the data (2,10), (2,5), (8,4), (5,8), (7,5), (6,4),(1,2),(4,9) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.	**L3**

S.No.	**CO**	**Questions**	**BT**
		Unit IV: (Unit Name)	
1	**4**	a. Explain the working principle of a decision tree in predictive modeling**.** b. Explain entropy, Gain ratio, Gini index with example dataset.	**L4**
2	**4**	**a.** Describe the complete process of building a logistic regression model for binary classification. b.How to interpret logistic regression models?	**L3**
3	**4**	**a.** Discuss the architecture and training process of a basic feed-forward neural network for classification. Explain back propagation and the role of activation functions, and mention two common challenges (overfitting, vanishing gradient) with solutions.	**L4**
4	**4**	**a.** Explain how the K-Nearest Neighbor (KNN) algorithm makes predictions. Describe the effect of choosing different values of K, the importance of distance metrics, and why feature scaling is critical before applying KNN.	**L3**
5	**4**	**a.** Describe the working of Naïve Bayes classifier assuming Gaussian distribution with an example. **b**. state bayes theorem**.**	**L5**
6	**4**	Compare linear regression and logistic regression in terms of their objective, output type, loss function, and suitable problem types. Provide one real-world example for each.	**L4**
7	**4**	Discuss three common types of non-linear regression algorithms (e.g., polynomial regression, ridge regression, decision tree-based regression).	**L3**

Prepared by XXXXXX

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

8	**4**	Describe the batch approach to model assessment in predictive analytics	**L5**
9	**4**	Explain how to assess regression models using the following metrics: R^2, Adjusted R^2, RMSE, MAE, and MAPE	**L4**
10	**4**	Compare classification model assessment metrics: accuracy, precision, recall, F1-score, and AUC-ROC.	**L3**

S.No.	CO	Questions	BT
Unit V: (Unit Name)			
1	**5**	**a.** **b.**	**L4**
2	**5**		**L3**
3	**5**		**L4**
4	**5**		**L3**
5	**5**		**L5**
6	**5**		**L4**
7	**5**		**L3**
8	**5**		**L5**

Prepared by XXXXXX

9	**5**		**L4**
10	**5**		**L3**
11	**5**		**L3**

Note: L1-Remembering, L2-Understanding, L3-Applying, L4-Analyzing, L5-Evaluating, and L6-Creating

Instruction to Faculty Members:

The Six Levels of Bloom's Taxonomy:

1. **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from long-term memory (e.g., list, define, name, locate).
2. **Understanding:** Constructing meaning, explaining ideas, or concepts (e.g., summarize, interpret, classify, compare).
3. **Applying:** Using information in new situations or implementing procedures to solve problems (e.g., solve, use, demonstrate, implement).
4. **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure (e.g., contrast, categorize, distinguish, diagram).
5. **Evaluating:** Making judgments based on criteria and standards through checking and critiquing (e.g., judge, critique, justify, defend, argue).
6. **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure (e.g., design, construct, develop, formulate).

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester **Regulation: R23**

Subject and Code: PREDICTIVE ANALYTICS &23 CSD 361T

Prepared by XXXXXX