

SRM VALLIAMMAI ENGINEERING COLLEGE

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

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

(COMMON TO DEPARTMENT OF INFORMATION TECHNOLOGY)



VI SEMESTER

1904607- DATA SCIENCE

Regulation – 2019

Academic Year 2024 – 2025

Prepared by

Ms.U.Chindiyababy AP(OG)
Ms.M.Kanmani AP(OG)

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur-603203

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

SUBJECT : 1904607- DATA SCIENCE

SEM/YEAR : VI / III

UNIT - I: INTRODUCTION TO DATA SCIENCE			
Introduction to Data Science-Concept of Data Science-Traits of Big data-Web Scraping- Analysis vs Reporting			
PART – A			
Q. No	Question	BT Level	Competence
1	What is the primary goal of Data Science?	BTL-1	Remember
2	Differentiate Business Intelligence (BI) and Data Science.	BTL-2	Understand
3	Compare Data Science and Statistics.	BTL-2	Understand
4	Define Data Science.	BTL-1	Remember
5	List out the areas in which Data Science can be applied.	BTL-1	Remember
6	What are the 4 V's of Big Data?	BTL-1	Remember
7	Compare Big Data with Data Science.	BTL-2	Understand
8	State the purpose or reporting and analysis.	BTL-1	Remember
9	List out the advantages of web scraping.	BTL-1	Remember
10	Can Data Science Predict the Stock Market? Examine.	BTL-2	Understand
11	Discuss about analysis and reporting.	BTL-2	Understand
12	Give Drew Conway's Venn diagram of Data Science.	BTL-2	Understand
13	Specify the life cycle of Data Science.	BTL-2	Understand
14	Illustrate the use of Data Science with an example.	BTL-2	Understand
15	Show the ways in which decision making and predictions are made in Data Science.	BTL-2	Understand
16	Differentiate Data Mining and Data Science.	BTL-2	Understand
17	Analyze Data Science ethics.	BTL-2	Understand
18	Analyze the roles of Data Science.	BTL-2	Understand

19	Bootstrap is more thorough in terms of the magnitude of replication. Justify.	BTL-2	Understand
20	Develop a general algorithm for Data Science process.	BTL-2	Understand
21	Define Velocity in the context of Big Data.	BTL-1	Remember
22	Name two tools commonly used for Web Scraping.	BTL-2	Understand
23	List two industries where Big Data is widely used.	BTL-2	Understand
24	Define the Data Science lifecycle.	BTL-1	Remember
PART – B			
1	i. What is Bigdata? (3) ii. Describe the main features of a big data in detail.(10)	BTL-3	Analyze
2	Describe life cycle of Data Science with neat diagram. (13)	BTL-3	Analyze
3	List the main characteristics of Big Data.	BTL-3	Analyze
4	i. Discuss nature of data.(7) ii. Give detail description of applications of data. (6)	BTL-2	Understand
5	i. Give the Difference between Traditional Business Intelligence (BI) versus Big Data.(7) ii. Give the various drawbacks of using Traditional system approach. (6)	BTL-2	Understand
6	i. Demonstrate the ETL (Extract, Transform and Load) system? (7) ii. Explain Big Data Technology Landscape. (6)	BTL-3	Analyze
7	Analyze and write short notes on the following. i. Hadoop Distributed File System (HDFS). (6) ii. YARN.(7)	BTL-3	Analyze
8	Explain the following in detail. i. Map Reduce. (7) ii. YARN.(6)	BTL-4	Apply
9	i. Assess the difference between analysis and analytics. (6) ii. Discuss the importance of big data analytics? (7)	BTL-3	Analyze
10	Extrapolate big data analytics and Develop a summary of various applications in the real world scenario. (13)	BTL-4	Apply
11	Describe the roles and stages in data science project. (13)	BTL-3	Analyze
12	i. Illustrate the importance of big data. (6) ii. List out the various challenges faced in big data in detail. (7)	BTL-3	Analyze
13	Explain storage consideration in Big Data. (13)	BTL-4	Apply
14	Discuss Data Cleaning and Sampling. (13)	BTL-4	Apply
15	Discuss the relationship between the traits of Big Data and the challenges of managing it. Explain how the 4 V's impact data storage, processing, and analysis.	BTL-4	Apply

16	Analyze the importance of Big Data in decision-making. Discuss how industries like healthcare, retail, and transportation utilize Big Data for predictive analytics.	BTL-3	Analyze
17	What are the ethical considerations in Web Scraping? Discuss the implications of privacy violations and how to adhere to ethical guidelines while collecting data.	BTL-3	Analyze
PART – C			
1	Create a brief summary about the challenges faced in processing big data now a day. (15)	BTL-3	Analyze
2	Evaluate in detail about the case study of big data solutions. (15)	BTL-4	Apply
3	Explain Traditional Vs Big data business approach with its drawbacks. (15)	BTL-3	Analyze
4	Evaluate the various formats of data and illustrate with a real time examples. (15)	BTL-3	Analyze
5	Illustrate the process of Web Scraping with a case study or example. Include details about the tools, techniques, and data usage.	BTL-3	Analyze
UNIT II MATHEMATICAL FOUNDATIONS			
Linear Algebra: Vectors, Matrices- Statistics: Describing a Single Set of Data, Correlation, Simpson's Paradox-Correlation and Causation- Probability: Dependence and Independence, Conditional Probability, Bayes's-Theorem, Random Variables-Continuous Distributions- The Normal Distribution-The Central Limit Theorem.			
PART – A			
Q. No	Questions	BT Level	Competence
1	What is a vector in linear algebra?	BTL-1	Remember
2	Point out the rules for dot product of two vectors.	BTL-1	Remember
3	Compare variance and covariance.	BTL-2	Understand
4	Develop a matrix to demonstrate binary relationship.	BTL-2	Understand
5	What is statistics? What are the ways to describe single set of data?	BTL-1	Remember
6	List applications of matrices.	BTL-1	Remember
7	Given single set of data, explain central tendencies of the data.	BTL-2	Understand
8	Describe dispersion in single set of data.	BTL-2	Understand
9	Give example of a continuous distribution.	BTL-2	Understand
10	Define Bayes's Theorem.	BTL-1	Remember
11	List some applications of conditional probability.	BTL-1	Remember
12	What way we can think of probability with respect to Data Science?	BTL-1	Remember

13	What is correlation?	BTL-1	Remember
14	What is Simpson's Paradox?	BTL-2	Understand
15	Classify the different distribution of values of random variables.	BTL-1	Remember
16	Illustrate normal distribution with diagram.	BTL-1	Remember
17	Complete a routine to display a histogram for sample number people and respective number of friends for them.	BTL-1	Remember
18	Analyze and write the importance of matrices in representing data sets.	BTL-2	Understand
19	Reason for importance of normal distribution is central limit theorem – Justify.	BTL-1	Remember
20	Develop a routine to plot Probability Density Function.	BTL-2	Understand
21	Differentiate between correlation and causation.	BTL-2	Understand
22	Why is correlation not causation?	BTL-1	Remember
23	Give the formula for conditional probability.	BTL-1	Remember
24	State the properties of the normal distribution.	BTL-1	Remember
PART – B			
1	Describe vectors and various operations on vectors with routines, example and diagram. (13)	BTL-3	Apply
2	Explain matrices with respect to Data Science. (6) Explain statistics and single set of Data. (7)	BTL-3	Apply
3	i. Describe about correlation in detail.(7) ii. Explain any one application of correlation.(6)	BTL-3	Apply
4	Explain normal distribution with an example. (13)	BTL-3	Apply
5	i. Explain conditional probability.(8) ii. Justify the need for normal distribution. (5)	BTL-4	Analyze
6	i. Give routine to display a histogram. (7) ii. Discuss about Dependence and Independence. (6)	BTL-4	Analyze
7	i. Describe application of matrices to represent binary relationship an example. (7) ii. Describe Bayes's Theorem. (6)	BTL-4	Analyze
8	i. Write a routine to plot Probability Density Function and illustrate with an example. (7) ii. Write a routine to plot a Histogram that compares Binomial Distribution and Normal Distribution. (6)	BTL-4	Analyze
9	i. Describe Normal Distribution in detail. (7) ii. Explain any one application of Bayes's theorem. (6)	BTL-3	Apply
10	Briefly describe the use of statistics in Data Science. (13)	BTL-3	Apply
11	Analyze and write a routine to implement various Probability Functions with example. (13)	BTL-4	Analyze
12	Develop a data set and demonstrate correlation. (13)	BTL-3	Apply

13	Discuss in detail about the variance, covariance, and correlation. (13)	BTL-3	Apply
14	Illustrate various distributions of values of random variables. (13)	BTL-4	Analyze
15	Describe the properties of the Normal Distribution and its applications in statistics.(13)	BTL-3	Apply
16	Discuss the concept of Conditional Probability, its formula, and applications with examples.(13)	BTL-3	Apply
17	Explain the relationship between dependence, independence, and conditional probability with suitable examples.(13)	BTL-4	Analyze
PART – C			
1	Develop a routine to demonstrate Binomial Distribution and Normal Distribution. (15)	BTL-4	Analyze
2	Assess the routines to implement various random variable distribution functions. (15)	BTL-5	Evaluate
3	Assess the difference between variance and covariance. Show a data set of values and demonstrate its correlation. (15)	BTL-5	Evaluate
4	Develop your own scenarios to demonstrate use of Vectors and Matrices in Data Science. (15)	BTL-6	Create
5	Discuss the use of the Normal Distribution in hypothesis testing and its role in inferential statistics.	BTL-5	Evaluate

UNIT - III: MACHINE LEARNING

Overview of Machine learning concepts –Types of Machine learning - Linear Regression- model assumptions- Classification and Regression algorithms- Naïve Bayes, K-Nearest Neighbors, logistic regression- support vector machines (SVM), decision trees, and random forest.

Q. No	Questions	BT Level	Competence
PART – A			
1	A common danger in machine learning is overfitting Justify	BTL 5	Evaluate
2	Usually the choice of a model involves a trade-off between precision and recall. Justify .	BTL5	Evaluate
3	What is Machine Learning?	BTL 1	Remember
4	Create a chart that demonstrates overfitting.	BTL 6	Create
5	How supervised models differ from unsupervised models?	BTL 4	Analyze
6	What is the reason for the word “Naïve” in Naïve Bayes classification?	BTL 1	Understand
7	List the major categories of Machine Learning.	BTL 1	Remember
8	What is a model with respect to Machine Learning? Give example.	BTL 2	Understand
9	Define simple linear Regression.	BTL 1	Remember
10	How to find the hyper plane dimension given the dimension of data in Support Vector Machine classification?	BTL 2	Understand

11	Simulate the idea behind nearest neighbor's classification.	BTL 6	Create
12	Discuss about random forests.	BTL 2	Understand
13	Give the formula for Conditional probability.	BTL 2	Understand
14	Explain Bayes's theorem.	BTL 4	Analyze
15	How we get random trees in Random Forest classification?	BTL 3	Apply
16	List major categories of supervised learning.	BTL 1	Remember
17	List out various regression models under supervised learning.	BTL 1	Remember
18	<p>Illustrate all possible decisions that can be made by the following decision tree.</p> <pre> graph TD A[Is a Person Physically Fit?] --> B[Age < 30?] B -- Yes --> C[Eat's a lot of Pizzas?] B -- No --> D[Exercises in the morning?] C -- Yes --> E[Unfit!] C -- No --> F[fit!] D -- Yes --> G[fit!] D -- No --> H[Unfit!] </pre>	BTL 3	Apply
19	Differentiate regression and Classification.	BTL 4	Analyze
20	Show formula for maximum likelihood estimation given a sample data v_1, \dots, v_n that comes from a distribution that depends on some unknown parameter Θ .	BTL 3	Apply
21	What is overfitting in machine learning, and how can it be prevented?	BTL 1	Remember
22	Define the role of entropy and information gain in decision trees.	BTL 2	Understand
23	What is the difference between a classification problem and a regression problem in machine learning?	BTL 1	Remember
24	What is a hyperplane in the context of Support Vector Machines?	BTL 1	Remember
PART – B			
1	Write routine for logistic regression and explain with necessary data and charts. (13)	BTL 1	Remember
2	Explain the following with suitable example. i. Simple Linear Regression. (6) ii. Multiple Regression. (7)	BTL 4	Analyze
3	Describe K-Nearest predictive model with suitable routine and example. (13)	BTL 1	Remember
4	Write the formula for Bayes Theorem and explain Naïve Bayes classifier with necessary routine.(13)	BTL 1	Remember
5	Discuss in detail the various Supervised Machine Learning techniques. (13)	BTL 2	Understand
6	Construct a decision tree for the following data: Explain various path in the tree that leads to various decisions. (13)	BTL 5	Evaluate

	Outlook	Temp	Humidity	Windy	Play Golf		
	Rainy	Hot	High	False	No		
	Rainy	Hot	High	True	No		
	Over roast	Hot	High	False	Yes		
	Sunny	Mild	High	False	Yes		
	Sunny	Cool	Normal	False	Yes		
	Sunny	Cool	Normal	True	No		
	Over roast	Cool	Normal	True	Yes		
	Rainy	Mild	High	False	No		
	Rainy	Cool	Normal	False	Yes		
	Sunny	Mild	Normal	False	Yes		
	Rainy	Mild	Normal	True	Yes		
	Overroast	Mild	High	True	Yes		
	Overroast	Hot	Normal	False	Yes		
	Sunny	Mild	High	True	No		
7	Discuss random forest with suitable algorithms and examples? (13)					BTL 2	Understand
8	Develop a routine for Support Vector Machine for a two dimension data. Validate the algorithm with suitable example. (13)					BTL 6	Create
9	Describe in detail about the following. i. Support Vector Machine. (7) ii. Hyper Plane. (6)					BTL 1	Remember
10	Differentiate classification model and regression model of machine learning with suitable examples. (13)					BTL 4	Analyze
11	i. Write short notes Random Trees. (6) ii. Explain random forest with example. (7)					BTL 2	Remember
12	Explain the Support Vector Machine classification for three dimensional data with necessary routine.(13)					BTL 4	Analyze
13	Illustrate decision trees with suitable examples. (13)					BTL 3	Apply
14	Explain the Different Types of Machine Learning with Suitable Examples.					BTL 3	Apply
15	Explain the Working of K-Nearest Neighbors (KNN) Algorithm and Discuss its Advantages and Disadvantages.					BTL 4	Analyze
16	Discuss the Assumptions and Mathematical Representation of Linear Regression.					BTL 3	Apply
17	i. Show a formula for maximum likelihood estimation (6) ii. Prove the working of Naïve Bayes classifier with necessary routine. (7)					BTL 3	Apply
PART – C							
1	Construct a decision tree for sample data of your own and evaluate various decision that can be arrived based on the decision tree. (15)					BTL 5	Evaluate
2	Create your own three-dimensional data and classify them using Support Vector Machine. (15)					BTL 6	Create
3	Evaluate random trees and explain random forest. (15)					BTL 5	Evaluate

4	Evaluate the K-Nearest Neighbors (KNN) algorithm in detail. Discuss how the choice of 'k' affects model performance, and explain its advantages and limitations in classification tasks.	BTL 5	Evaluate
5	Describe the role of the root node in a decision tree, and how does it impact the tree's structure and decision-making process?	BTL 5	Evaluate

UNIT - IV: PROGRAMMING TOOLS FOR DATA SCIENCE

Introduction to Programming Tools for Data Science-Toolkits using Python: Matplotlib, NumPy, Scikit-learn, NLTK-Visualizing Data: Bar Charts, Line Charts and Scatterplots-Working with data: Reading Files, Scraping the Web, Using APIs (Example: Using the Twitter APIs).

PART – A

Q.No	Questions	BT Level	Competence
1	What is SAS?	BTL1	Remember
2	Define data visualization in machine learning.	BTL1	Remember
3	Give the features of Numpy.	BTL2	Understand
4	What is meant Matplotlib? Give features of Matplotlib.	BTL1	Remember
5	Give the expansion for NLTK in machine learning and explain.	BTL2	Understand
6	List any four data science tools.	BTL1	Remember
7	Describe about Apache Spark.	BTL1	Remember
8	Predict the features of Scikit.	BTL2	Understand
9	Compare R and Python.	BTL4	Analyze
10	Distinguish Statistics and Data Science.	BTL2	Understand
11	Classify the different visualization tools.	BTL3	Apply
12	Develop line chart for the following data. years = [1950, 1960, 1970, 1980, 1990, 2000, 2010] gdp = [300.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 14958.3].	BTL6	Create
13	Which language is best for learning data science? Illustrate why?	BTL3	Apply
14	Summarize the MATLAB.	BTL5	Evaluate
15	Point out the components of Data Science.	BTL4	Analyze
16	Compare various data science languages.	BTL4	Analyze
17	Select the best tool or language for data science and give justification.	BTL5	Evaluate

18	Illustrate line charts with an example.	BTL3	Apply
19	Identify the tools for Data Science.	BTL1	Remember
20	Develop a bar chart for the following data. movies = ["Annie Hall", "Ben-Hur", "Casablanca", "Gandhi", "West Side Story"] num_oscars = [5, 11, 3, 8, 10].	BTL6	Create
21	What is NumPy, and why is it used in data science?	BTL-2	Understand
22	What's the difference between a Python list and a NumPy array?	BTL-1	Remember
23	What is the expansion for NLTK in machine learning and explain?	BTL-1	Remember
24	What is the spaCy library in Python?	BTL-1	Remember
PART – B			
1	i. Describe Numpy in detail. (6) ii. Write a python program that uses numpy and explain it. (7)	BTL1	Remember
2	Describe the following. i. Numpy. (7) ii. Scikit. (6)	BTL1	Remember
3	i. List the different types of charts? (7) ii. Explain any one chart in detail with an Example.(6)	BTL1	Remember
4	Discuss various Toolkits in Python in detail.(13)	BTL2	Understand
5	Describe various web scraping methods in detail.(13)	BTL2	Understand
6	Illustrate Matplotlib with an example. (13)	BTL3	Apply
7	Explain different visualization tools in detail with an example. (13)	BTL4	Analyze
8	Point out various features of Toolkits that can be used with Python. (13)	BTL4	Analyze
9	Write about estimators and Explain how it can be fitted to some data using its fit method. (13)	BTL5	Evaluate
10	Write a program by loading the Iris dataset, split it into train and test sets, and compute the accuracy score of a pipeline on the test data. (13)	BTL6	Create
11	i. Write a python program to read a file. (7) ii. Illustrate the flow of the program.(6)	BTL3	Apply
12	Describe the following. i. MaTLAB. (7) ii. Python. (6)	BTL1	Remember
13	Explain in detail about the following. i. Line chart. (6) ii. Bar chart .(7)	BTL4	Analyze

14	Describe NLTK. Explain the steps to use it in Python. (13)	BTL2	Understand
15	Explain in detail about Reading file with example. (13)	BTL 4	Analyze
16	How can graphical data be projected using matplotlib? Explain with examples. (13)	BTL 3	Apply
17	Write python program to plot line chart by assuming your own data and explain various attribute of line chart. (13)	BTL 3	Apply
PART – C			
1	Develop a line chart to visualize a data set of your choice and give the detailed explanation of observations from chart. (15)	BTL6	Create
2	Analyze how to construct a bar chart for a data set and explain it in detail.(15)	BTL5	Evaluate
3	Explain a various methods of Scraping the web in detail. (15)	BTL5	Evaluate
4	Prepare a program to read a file and discuss its working.(15)	BTL6	Create
5	Consider that, an E-Commerce organization like Amazon. have different regions sales as NorthSales, SouthSalea, WestSales. EastSales.csv files. They want to combine North and West region sales and South and East sales to find the aggregate sales of these collaborating regions. Help them to do so using Python code. (15)	BTL5	Evaluate

UNIT - V: CASE STUDIES OF DATA SCIENCE APPLICATION

Weather forecasting-Stock market prediction-Object recognition- Real Time Sentiment Analysis.

PART – A

Q. No	Questions	BT Level	Competence
1	What is weather forecasting?	BTL1	Remember
2	Define precipitation.	BTL1	Remember
3	Give the advantages of weather forecasting.	BTL2	Understand
4	What is Object Recognition?	BTL1	Remember
5	Give need for opinion mining.	BTL2	Understand
6	Name the applications of Sentiment Analysis.	BTL1	Remember
7	Name the applications of Object Detection.	BTL1	Remember
8	Predict the importance of opinions.	BTL2	Understand
9	Point out the role of the web in Sentiment Analysis.	BTL4	Analyze
10	Distinguish between computer vision tasks: Image Classification and Object Localization.	BTL2	Understand

11	Classify the different computer vision tasks.	BTL3	Apply
12	Develop sample input and output for Object Detection.	BTL6	Create
13	Which is said to be primary source of atmospheric science?	BTL3	Apply
14	Summarize the Role of Modeling to Predict Stock Prices.	BTL5	Evaluate
15	Point out the importance of Stock Market.	BTL4	Analyze
16	Compare different computer vision tasks.	BTL4	Analyze
17	Can Data Science be used in Stock Market Analysis? Justify .	BTL5	Evaluate
18	How weather forecasts are made?	BTL3	Apply
19	List three modules of R-CNN.	BTL1	Remember
20	Develop sample input and output for Object Localization.	BTL6	Create
21	Identify the key factors influencing accurate weather forecasting and briefly explain how advancements in technology have improved prediction accuracy in recent years.	BTL 3	Apply
22	Examine the role of historical stock market data in stock market prediction models and provide a brief justification for its significance in developing reliable forecasts.	BTL-4	Analyze
23	Explain data presentation using statistics, discuss histograms and frequency polygons, and outline a deep learning approach for object recognition.	BTL 6	Create
24	Compare the effectiveness of machine learning and traditional methods in object recognition, highlighting the strengths and limitations of each approach in real-world scenarios.	BTL5	Evaluate
PART – B			
1	i. Describe data is a crucial part of Weather Predictions. (6) ii. How weather Data is an aid for many Events. (7)	BTL1	Remember
2	Describe the following i. Image Classification. (6) ii. Object Localization. (7)	BTL1	Remember
3	Describe the following. i, A Twitter NLP chain,. (5) ii, NL processor and Ad-hoc NL processor. (8)	BTL1	Remember
4	Discuss various sub processes involved in the complete process of data science for weather prediction. (13)	BTL2	Understand
5	Describe YOLO Model Family. (13)	BTL2	Understand
6	Write in detail about R-CNN Model Family. (13)	BTL3	Apply
7	Explain schema, which shows the process of the water cycle and precipitation occurrence. (13)	BTL4	Analyze
8	Compare the following computer vision tasks and discuss about each task in a very detailed Manner. i, Object Localization. (6) ii. Object Detection. (7)	BTL4	Analyze
9	Summarize of Predictions made by YOLO Model. (13)	BTL5	Evaluate
10	Develop a code to Prepare the Input for the LSTM Model. (13)	BTL6	Create
11	i. Write short notes on R-CNN.(7)	BTL4	Analyze

	ii. Illustrate Satellite Imagery and Sensor Data. (6)		
12	Describe the following i. Image Classification. (6) ii. Object Localization. (7)	BTL1	Remember
13	i. Discuss in detail about Satellite Imagery and Sensor Data in weather forecasting. (7) i. Explain the Stock Market with suitable example. (6)	BTL3	Apply
14	Describe various computer vision tasks in object recognition. (13)	BTL2	Understand
15	Define sentiment analysis and explain with example. (13)	BTL4	Analyze
16	What is object recognition and explain in detail with their example. (13)	BTL4	Analyze
17	What is weather forecasting and explain in detail with their example. (13)	BTL5	Evaluate
PART – C			
1	Develop a case study of Sentiment Analysis in Twitter. (15)	BTL6	Create
2	Explain Condensation and coalescence are important parts of the water cycle and how data collected from it. (15)	BTL5	Evaluate
3	Explain Fast R-CNN. (15)	BTL5	Evaluate
4	Develop a case study on Google Stock Price Prediction Using LSTM. (15)	BTL6	Create
5	How natural language processing applied in machine learning? Why we used in machine learning. (15)	BTL6	Create