INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING(AI & ML)

QUESTION BANK

Department	**COMPUTER SCIENCE AND ENGINEERING(AI & ML)**				
Course Title	**STATISTICAL FOUNDATIONS OF DATA SCIENCE**				
Course Code	ACAC07				
Program	B.Tech				
Semester	V	CSE(AI & ML)			
Course Type	Elective				
Regulation	IARE - UG20				
Course Structure	Theory			Practical	
	Lecture	Tutorials	Credits	Laboratory	Credits
	3	-	3	-	-
Course Coordinator	Ms. K. Anjali, Assistant Professor				

COURSE OBJECTIVES:

The students will try to learn:

I	The fundamental knowledge on basics of data science.
II	The basic principles of data acquisition, exploring and modeling data efficiently.
III	The foundations of probability and statistics for data science.
IV	The current scope, potential applications of data science.

COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	**Recall** the categories and levels of data using steps involved in data science.	Remember
CO 2	**Demonstrate** the data pre-processing terms for improving the quality of dataset using processes such as feature generation and feature selection	Understand
CO 3	**Solve** mathematical problems using various arithmetic and more challenging forms of math.	Apply
CO 4	**Apply** probability theorems and approaches for calculating the number of outcomes of the events.	Apply
CO 5	**Illustrate** the obtaining and sampling data in statistics to quantify and visualize our data.	Understand
CO 6	**Summarize** the concepts of communication by using the visualization and presenting strategies.	Understand

QUESTION BANK:

Q.No	QUESTION	Taxonomy	How does this subsume the level	CO's
MODULE I				
FLAVORS OF DATA				
PART A-PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				
1	What does AI have to do with unstructured data?	Understand	The learner can **Remember** the data management and data quality	CO 1
2	Why is Unstructured Text Data Important in Decision Making?	Understand	The learner will get to lnow about the unstructured data in decisiion making	CO 1
3	How can we change our processes and technology to eliminate or cut across silos of information?	Understand	The learner will get to know how to eliminate silos in information	CO 1
4	There are some steps for qualitative data analysis. Write those steps in detail.	Understand	The learner can understand the proces of quantitative data analysis.	CO 1
5	How Can we Transform Unstructured Data Into Structured Data?	Understand	The learner get to know the process of transformation form unstructured to structured data.	CO 1
6	How do you approach solving any data analytics based project?	Understand	The learner can get the idea that how to tackle the issues at the time of handllinf the data science based project	CO 1
7	Are the necessary assumptions and conditions for your chosen statistical method likely to be met by your data, given your data collection plan (e.g., quantitative versus qualitative data)?	Understand	The learner can understand the statistical methods	CO 1
8	How will you determine whether any data points should be excluded from your analysis?	Understand	The learner get to know about analyzing of data	CO 1

9	You are given a data set consisting of variables with more than 30 percent missing values. How will you deal with them?	Understand	learner will get to know how to deal with the variables	CO 1
10	People who bought this also bought…' recommendations seen on Amazon are a result of which algorithm?	Understand	The learner can understand the filtering method	CO 1
11	What are some common Machine Learning problems that Unsupervised Learning can help with?	Understand	The learner will get the knowledge of machine learning with unsupervised learning	CO 1
PART-B LONG ANSWER QUESTIONS				
1	Discuss the four levels of data. Give examples of each and explain in detail.	Understand	The learner can gain a knowledge about the levels of data.	CO 1
2	What is the difference between data analytics and data science?	Understand	The learner can predict the difference of these terms.	CO 1
3	Give explaination about data science process with data science life cycle.	Understand	The learner can get the detail knowledge about the data science process	CO 1
4	Explain about data exploration? Why is data exploration important?How data exploration works?	Understand	The learner will learn about exploring the data	CO 1
5	How Would You Approach a Dataset That's Missing More Than 30 Percent of Its Values?	Understand	The learner get the process of handling missing values	CO 1
6	Desognate the qualitative data? Present with importance of qualitative data and methods of qualitative data collection.	Understand	The learner can know the importance of qualitative data and concepts of qualitative data collection methods.	CO 1
7	Give the examples of oevrfitting and underfitting with defining it. Explain the methods to avoid overfirttign and underfitting in machine learning.	Understand	The learner will get to know about the concepts of overfittig and underfitting	CO 1

8	Construe a term quantitative data. Confer the analysis methods of quantitative data.	Understand	The learner will get the information about analysisi methods of quantitative data.	CO 1
9	Explain nominal level in detail with its measures of center Annotate characteristics of it.	Understand	The learner can understand the nominal level of data in detail	CO 1
10	Why are levels of measurement important? Categorize different levels of data in detail with its examples.	Understand	The learner can analyze the data of level of meeasurement.	CO 1
11	How Can we Transform Unstructured Data Into Structured Data? What is the need for Unstructured Data to Structured Data Conversion?	Understand	The learner will get the idea how to handle the structured and unstructured data	CO 1
12	How do you analyze unstructured data? Differentiate between quantitative and qualitative data in detail with example?	Understand	The learner will get to know the data concepts	CO 1
13	Present the ordinal data with its definition, examples. Explain how to analyze ordinal data.	Understand	The learner can analyze the ordinal data with its analysis.	CO 1
14	Explain data science with its venn diagram. While working on a data set, how can you select important variables? Explain.	Understand	The learner can clear the data science concepts using venn diagram	CO 1
15	Characterize the term qualitative data collection. Present it with qualitative data collection methods.	Understand	The learner get the idea of handling the qualitative data.	CO 1
16	What is data science life cycle? Explore different stages of it in detail.	Understand	The learner will get to know the stages of data science life cycle	CO 1
17	Explore interval data with its characteristics. Give the interval data collection method.	Understand	The learner will get the basic information about interval data	CO 1

18	Explain five P's of data science. Differentiate between data science and data analytics.	Understand	The learner will get the knowledge of different component of data science	CO 1
19	Signify the term ratio level data with its characteristics and explain the different ways to calculate ratio data.	Understand	The learner can understand the concept of ratio level data	CO 1
20	How to Analyse & Interpret Interval Data? Give the uses of interval data.	Understand	The learner can get to know about the proces of analyzing and interpreting the interval data	CO1
PART-C SHORT ANSWER QUESTIONS				
1	What is meant by data science and enlist steps of data science.	Remember	Learner can get the basic knowledge of data science	CO 1
2	Explain in brief Data Model?	Remember	Learner can get the basic knowledge of data model	CO 1
3	Enlist the types of data model. Explain each.	Remember	The learner to know about the concepts of data model	CO 1
4	Distinguish between structured and unstructured data.	Remember	The learner can understand the difference between these terms	CO 1
5	Give an atleast three examples of quantitative and qualitative data?	Remember	the learner can understand the concepts of quantitative and qualitative data	CO 1
6	Discuss term explore data.	Remember	the learner can able to know the concepts of exploring the data	CO 1
7	Annotate the term in short interval level.	Remember	The learner will able to understand the interval level of data	CO 1
8	Demonstrate the applications of data scienece.	Remember	Learner will get the knowldge of data science	CO 1
9	Define nominal level and ratio level.	Remember	-	CO 1
10	Discuss an interesting question step of data science in detail.	Remember	The learner can understand the data science in detail	CO 1
11	Write short notes on data exploration in Analytics?	Understand	The learner will llearn that how to explore the data.	CO 1

12	Give an example of ordinal level of measurement?	Remember	The learner can understand the ordinal level of measurement	CO 1
13	Explain the latest trends in Data Science?	Remember	The learner will get to know about the trends in data science	CO 1
14	Give the 3 main concepts of Data Science?	Remember	The learner can able yo understand the different algorithms with various techniques	CO 1
15	Describe three types of structured data?	Remember	The learner first to know the concept of structured data	CO 1
16	What statistics is used in data science?	Remember	The learner will able to understand the relation if statistics and data science	CO 1
17	Give the 4 major components of data science?Explain each.	Remember	Learner can able to understand the different strategies, model and analysis techniques.	CO 1
18	Write short notes on mode, median and mean.	Remember	The learner get to know about the mode, median and mean	CO 1
19	List out the disadvantages of Ordinal Data	Remember	The learner first to know the disadvantages of ordinal data	CO 1
20	Explain the steps which involved in data science process?	Remember	the learner get to know the strps of data science process	CO 1

MODULE II				
DATA PRE-PROCESSING AND FEATURE SELECTION				
PART-A PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				
1	Should your Test Data be Cleaned the same way that the Training Data is?	Understand	The learner to know about the data cleaning process	CO 2
2	When you sample, what potential Sampling Biases could you be inflicting?	Understand	The learner able to understand the sampling concepts	CO 2
3	Can data cleaning worsen the results of statistical analysis	Understand	The learner can understand the statistical analysis	CO 2
4	You have to train a 10GB dataset using a neural network and Support Vector Machines with a machine that has only 4GB RAM. How would you go about it?	Understand	-	CO 3
5	If it takes one hour to train a Decision Tree on a training set containing 1 million instances, roughly how much time will it take to train another Decision Tree on a training set containing 10 million instances.	Understand	The learner must know the concepts of requirements review	CO 2
6	Does the strategy of the organisation dictate a data transformation? If so, what needs to change and why?	Understand	The learner must know the concepts of data transformation strategies	CO 3
7	How can one handle suspicious or missing data in a dataset while performing the analysis?	Understand	The learner must know the concepts of data transformation	CO 2
8	If we have a date column in our dataset, then how will you perform Feature Engineering?	Understand	The learner will get to know the feature engineering	CO 2
9	Explain the general principle of an ensemble method and what is bagging and boosting in the ensemble method?	Understand	The learner must know the concepts of bagging and boosting	CO 2

10	How can less Training Data give Higher Accuracy?	Understand	The learner will get the information about the data for higher accuracy	CO 2
PART-B LONG ANSWER QUESTIONS				
1	What is decision tree analysis used for?How to create a decision tree?	Understand	The learner can analyze the decision tree in proper way	CO 2
2	Discuss Data Preprocessing & What Are The Steps Involved?	Understand	The learner will learn the data preprocessing steps	CO 2
3	Explore the term data transformation with its steps.	Understand	The learner can understand the knowledge of data transformation	CO 2
4	Explain data preprocessing with its types in detail.?	Understand	The learner to know about the types of data preprocessing	CO 2
5	Discuss feature selection algorithms in detail with feature selection models.	Understand	The learner will get to know about the algorithms related to feature selection	CO 2
6	Give detail information about Data Discretization? What are techniques of Data Discretization?.	Understand	The learner will get to know the data discretization techniques	CO 2
7	Explain each step of data preprocessing in detail.	Understand	The learner can clear the basic concepts of data preprocessing	CO 1
8	Give comparison between data cleaning and data transformation? How to clean your data (step-by-step) explain in detail?	Understand	The learner will get to know the process of data cleaning	CO 2
9	Describe the Random Forest. How does Random Forest algorithm work? Give the applications of random forest.	Understand	The learner can understand the random forest algorithm	CO 2
10	Give the importance of Random Forest. List down the advantages and disadvantages of the Random Forest Algorithm.	Understand	The learner will get know the importance of random forest	CO 2

11	How are Decision Trees used in Classification?	Understand	The learner will get to know that how to use decision trees for the classification	CO 2
12	Give the benefits of using feature selection. How do you perform feature selection with Categorical data?	Understand	The learner get to know the steps of performing feature selection	CO 2
13	Give the advantages, limitaions and applications of decision tree.	Understand	The learner can clear the concepts of decision tree	CO 2
14	What are the steps involved in Data Transformation Process? Why Need to Transform Data?	Understand	The learner get to define the functions of data transformation	CO 2
15	What are the techniques used in data reduction? Explain.	Understand	The learner to know the concept of data reduction techniques in data preprocessing	CO 2
16	What is data reduction and why is it important?	Understand	The learner will get to know the concept of data reduction	CO 2
17	What do you understand about Information Gain? Also, explain the mathematical formulation associated with it.	Understand	The learner will able to understand the concept of information gain	CO 2
18	Explain the techniques used for data integration in detail.	Understand	The learner get to know about the data integration	CO 2
19	Explain data cleaning techniques in detail and give the benefits of data cleaning.	Understand	The learner will get the detail information about the data cleaning	CO 2
20	What are challenges of data preprocessing?	Understand	The learner will get to know that which problems are facing by users ata the time of data preprocessing	CO 2
PART-C SHORT ANSWER QUESTIONS				
1	Explain the operations which can be used for data pre processing?	Understand	The learner to know about the operations of data preprocessing	CO 2
2	Annotate the 5 major steps of data preprocessing?	Understand	The learner get to know that how to process the data	CO 2

3	Explain the feature selection methods used to select the right variables?	Understand	The learner to know about the feature selection methods	CO 2
4	Describe the preprocessing techniques?	Understand	The learner will get the information about the preprocessing technique	CO 2
5	Explain the problems in data cleaning?	Understand	The learner will get the information about problems in data cleaning	CO 2
6	Can Random Forest Algorithm be used both for Continuous and Categorical Target Variables?	Understand	The learner will get the knowledge of random forest	CO 2
7	Name some benefits of Feature Selection?	Understand	The learner will get to know the names of fdature selection	CO 2
8	Can random forest use logistic regression?	Undesrtand	The learner can get to know about the logistic regression	CO 2
9	Give the purpose of data filtering?	Understand	The learner can get to know about the data filtering	CO 2
10	Give the difference between Data Processing and Data Mining	Understand	The learner can get to know the difference between these terms	CO 2
11	What is random forest used for?	Understand	The learner will get the information abot the random forest	CO 2
12	Where are filters used?	Understand	The learner to know the concept of system requirements	CO 4
13	Describe feature extraction in machine learning?	Understnad	The learner can get to know the feature extraction in machine learning	CO 2
14	Explain the Importance of Data Reduction?	Understand	The learner will get to know about an importance of data reduction	CO 2
15	List the different techniques used in data transformation.	Understand	The learner can understand the techniques of data transformation	CO 2
16	Define feature selection.	Understand	the learner can get to know the feature selection	CO 2
17	Annonate the feature generation and why feature generation used in data science?	Undesrtand	The learner will able to understand feature generation	CO 2

18	Give the applications of filters.	Understand	The learner get to know the uses of filters	CO 2
19	Why is random forest better than linear regression?	Understand	The learner can compare between random forest and linear regression	CO 2
20	Why is it necessary to preprocess data?	Understand	the learner will get the knowledge of preprocessing data	CO 2

MODULE III				
BASIC MATHEMATICS AND PROBABILITY FOR DATA SCIENCE				
PART A-PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				
1	Let A and B be events on the same sample space, with P (A) = 0.6 and P (B) = 0.7. Can these two events be disjoint?	Apply	The learner to know the basic concepts of probability	CO 3
2	Alice has 2 kids and one of them is a girl. What is the probability that the other child is also a girl? You can assume that there are an equal number of males and females in the world.	Apply	The learner can understand the concepts related to probability	CO 3
3	A fair six-sided die is rolled 6 times. What is the probability of getting all outcomes as unique?	Apply	the leraner will get the result of probability	CO 3
4	Find the vector joining the points P(2, 3, 0) and Q(– 1, – 2, – 4) directed from P to Q.	Apply	The learner get to know the concepts of vector	CO 4
5	How would you explain a confidence interval to a non-technical audience?	Understand	The learner will get to know the concepts of statistics	CO 3
6	A certain couple tells you that they have two children, at least one of which is a girl. What is the probability that they have two girls?	Understand	The learner to know about the concepts of probabiliity	CO 3

7	One hundred people line up to board an airplane. Each has a boarding pass with assigned seat. However, the first person to board has lost his boarding pass and takes a random seat. After that, each person takes the assigned seat if it is unoccupied, and one of unoccupied seats at random otherwise. What is the probability that the last person to board gets to sit in his assigned seat?	Understand	The learner to know the probability concepts	CO 3
8	A fair six-sided die is rolled twice. What is the probability of getting 2 on the first roll and not getting 4 on the second roll?	Apply	The learner to know the concepts of probability using die	CO 4
9	In any 15-minute interval, there is a 20% probability that you will see at least one shooting star. What is the probability that you see at least one shooting star in the period of an hour?	Apply	The learner to know the concepts of probability	CO 3
10	Find the probability of drawing two cards that belong to different colors or different shapes (suits) from a shuffled deck of 52 cards?	Apply	The learner to know how to find the probabbility using cards	CO 4
PART-B LONG ANSWER QUESTIONS				
1	How can you calculate accuracy using a confusion matrix?	Apply	Ther can able to understand the confusion matrix	CO 3
2	A. What will be the probability of getting odd numbers if a dice is thrown? B. From a pack of 52 cards, two cards are drawn together at random. What is the probability of both the cards being kings?	Apply	The learner get to know abot the probability	CO 3

3	Explain the applications of probability in detail. Two coins are tossed 500 times, and we get: Two heads: 105 times One head: 275 times No head: 120 times Find the probability of each event to occur.	Apply	the learner will get to know about the applications of probability	CO 4
4	The entries in a 2 x 2 matrix are integers that are independently chosen for each entry. The probability that the entry is odd is p. If the probability that the value of the determinant is even is 0.5, find p.	Apply	The learner get to know about the probability	CO 2
5	Three persons A, B and C have applied for a job in a private company. The chance of their selections is in the ratio 1 : 2 : 4. The probabilities that A, B and C can introduce changes to improve the profits of the company are 0.8, 0.5 and 0.3, respectively. If the change does not take place, find the probability that it is due to the appointment of C.	Apply	The learner get help understand the concept of bayes theorem	CO 3
6	Prove the bayes theorem. Give the difference between Bayesian versus Frequentist	Apply	The learner get the knowledge of Bayesian and Frequentist approach	CO 4
7	A.What are different types of vectors in R? Explain each. B. Find the vector and its magnitude which joins the point A(4, 5, 6) to point B(10, 11, 12).	Apply	The learner will get the guidance on vector	CO 3

8	A. Let X be a random variable with probability distribution function f (x)=0.2 for |x|<1 = 0.1 for $1 < \|x\| < 4 = 0$ otherwise The probability P ($0.5 < x < 5$) is B. Consider a dice with the property that that probability of a face with n dots showing up is proportional to n. The probability of face showing 4 dots is?	Apply	The learner get to know about the concepts of probability with random variable	CO 4
9	A.Explain in detail Independent and Dependent Events of probability. B.Solve the following. A coin is tossed three times, consider the following events. P: 'No head appears', Q: 'Exactly one head appears' and R: 'At Least two heads appear'. Check whether they form a set of mutually exclusive and exhaustive events.	Apply	The learner get to know about different events of probability	CO 3
10	In a neighbourhood, 90% children were falling sick due flu and 10% due to measles and no other disease. The probability of observing rashes for measles is 0.95 and for flu is 0.08. If a child develops rashes, find the child's probability of having flu.	Apply	the learner will get to know about the concepts of bayes theorem	CO 3
11	A.Describe Probability Distributions with its types? B. Find the binomial distribution of getting a six in three tosses of an unbiased dice.	Apply	The learner can able to define the probability distributions	CO 3

12	List out a few examples of Discrete Probability Distribution? Explain any four.	Understand	The learner will get to know the discrete probability distribution	CO 3
13	A.Write short notes on functions of random variables. B. Let X be a discrete random variable with Px(k)=1/5 for k=−1,0,1,2,3. Let Y=2|X|. Find the range and PMF of Y	Apply	The learner get to know the information about the random variable	CO 4
14	Give the difference between Frequentist approach and Bayesian approach?	Understand	the learner can understand the difference betweeen these terms	CO 3
15	Explain the rules of probability in detail.	Undesrtand	The learner to define the rules of probability	CO 3
16	A jar has 1000 coins, of which 999 are fair and 1 is double headed. Pick a coin at random, and toss it 10 times. Given that you see 10 heads, what is the probability that the next toss of that coin is also a head?	Understand	The learner to know the probability concepts	CO 3
17	Explain data transformation functions which serves within the data analytics stack.	Understand	The learner get to define the functions of data transformation	CO 3
18	A. Explain the term collectively exhaustive event. B. In an experiment, three coins are tossed at a time, consider the following events. A: 'No tail appears' B: 'Exactly one tail appears' C: 'At least two tails appear' Do they form a set of exhaustive events?	Apply	The learner can able to know the exhaustive events	CO 4

19	A.Discuss set theory with its operations and properties. B. In an experiment, three coins are tossed at a time, consider the following events. A: 'No tail appears' B: 'Exactly one tail appears' C: 'At least two tails appear' Do they form a set of exhaustive events? C. If A = 3, 5, 7, 9, 11, B = 7, 9, 11, 13, C = 11, 13, 15. Find A (B C).	Apply	The learner will get to know different set theory operations with properties	CO 2
20	A.Explain basic operations of matrix. B. Write R program for matrix addition, matrix subtraction and matrix multiplication.	Understand	The learner get to know various opeartions which are performed on matrix	CO 3
PART-C SHORT ANSWER QUESTIONS				
1	Explain the relationship between matrices and vectors?	Understand	The learner can compare the matrices and vectors	CO 3
2	Which are the arithmetic symbols in data science?	Understand	The learner get to know the arithmetic symbols	CO 3
3	How are graphs used in data science?	Understand	The learner get to know about the concepts of graphs related to data science	CO 4
4	What is logarithm in data science?	Understand	The learner can understand the basic concepts of logarithm	CO 3
5	Give the application of set theory?	Understand	The learner will get to know that uses of set thoery	CO 4
6	Why linear algebra is used in machine learning?	Understand	The learner will get to know the concepts of linear algebra using machine learning	CO 3
7	Describe the probability in statistics?	Remember	The learner can understand the concepts of probability in statistics	CO 4

8	Give the difference between frequentist and Bayesian approaches?	Remember	The lewraner get to know the compraison between them	CO 3
9	Explain compound event in probability?	Remember	The learner get to know the compound event in probability	CO 3
10	Explain conditional probability with real life examples?	Undedrstand	The learner get to know about the conditional probability	CO 4
11	Annonate the term collectively exhaustive?	Understand	The learner to explain the collectively exhaustive	CO 3
12	How do you know if events are collectively exhaustive?	Understand	The learner will able to learn the concepts of collectively exhaustive	CO 3
13	What is Bayes Theorem explain with example?	Remember	The learner can able to learn the Bayes theorem	CO 3
14	Give the difference between probability and conditional probability?	Understand	The learner will able to understand the difference of above terms	CO 3
15	Why is Bayes Theorem important in data science?	Understand	The learner to know the import5ance of bayes theorem in data science	CO 4
16	How do you identify a random variable?	Remember	The learner get to know about the random variables	CO 3
17	Are matrices set of vectors?	Understand	The learner get to know the relation of matrix and vectors	CO 3
18	Give the purpose of Bayesian analysis in decision making?	Understand	The learner to know the purpose of Bayesian analysis	CO 3
19	Why is Bayesian statistics better than frequentist?	Understand	The learner get to know the comparison of Baysian statistics and frequentist	CO 4
20	Elaborate the term random variable with example?	Understand	The learner get to define the random variable	CO 3

MODULE IV				
STATISTICS FOR DATA SCIENCE				
PART A- PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				
1	How important is the standard deviation to understanding relationships between data points? Why do you think students find it so hard to understand this concept? What could students do to better understand the standard deviation?	Understand	The learner get to know about standard deviation	CO 5
2	If a distribution is skewed to the right and has a median of 20, will the mean be greater than or less than 20?	Understand	Learner to know the concepts ofsoftware architectures	CO 5
3	When creating a statistical model, how do we detect overfitting?	Understand	The learner get to know about statistical model	CO 5
4	To identify confidence interval about mean when population variance is known, which distribution is suitable?	Understand	the learner will get to know about the confidence interval	CO 5
5	What is meant by mean imputation for missing data? Why is it bad?	Understand	The learner will get to know the mean imputation	CO 5
6	State the case where the median is a better measure when compared to the mean.	Understand	The learner get to know the importance of median	CO 5
7	The standard normal curve has a total area to be under one, and it is symmetric around zero. True or False?	Understand	The learner get to know the condition of given statement	CO 5
8	Does a symmetric distribution need to be unimodal?	Understand	The learner will try to recall symmetric distribution	CO 5
9	Can You Avoid Overfitting Your Model? If Yes, Then How?	Understand	The learner will try to gain the knowledge of overfitting	CO 5

10	How to find the mean length of all fishes in the sea?	Understand	The learner will try to fing out the mean length using confidence level concepts	CO 5
PART-B LONG ANSWER QUESTIONS				
1	Explain confidence intervals in detail and how do you calculate it?	Understand	The Learner get the detailing of confidence intervals	CO 5
2	Describe a few methods or techniques used in statistics for analyzing the data.	Understand	The Learner get to know about the techniques used in statistics	CO 5
3	Explain sampling data with its types in deatil.	Understand	The learner will get to know the basic concepts of sampling data	CO 5
4	A. Explore empirical rule in detail. B. A sulfide of iron was formed by combining 1.926g of sulfur(S) with 2.233g of iron (Fe). What is the compound's empirical formula?	Apply	The learner can get knowledge of empirical rule	CO 5
5	What are the different types of Correlation?Draw suitable disagram each.	Understand	The Learner to know about correlation	CO 5
6	What are the Properties of Point Estimators? What are the Formulae that Can be Used to Measure Point Estimators? What are the Values Needed to Calculate Point Estimators?	Understand	The Learner get to know the properties, formulae of point estimators	CO 5
7	In a tree, there are hundreds of apples. You are randomly choosing 46 apples with a mean of 86 and a standard deviation of 6.2. Determine that the apples are big enough.	Apply	The learner will able to understand the conceepts of confidence interval	CO 5
8	Four friends were comparing their scores on a recent essay. Calculate the standard deviation of their scores: 6, 2, 3, 1	Apply	The Learner will get to know that hoew to calculate the standard deviation	CO 5

9	How to use empirical rule? Explain with example.	Understand	The learner get to know about the empirical rule with example	CO 5
10	What are the 7 steps in hypothesis testing? Explain each.	Understand	The learner get to know steps of hypothesis testing	CO 5
11	What are some of the techniques to reduce underfitting and overfitting during model training?	Understand	The learner het to know the techniques for reducing underfitting and overfitting	CO 5
12	A Telecom service provider claims that individual customers pay on an average 400 rs. per month with standard deviation of 25 rs. A random sample of 50 customers bills during a given month is taken with a mean of 250 and standard deviation of 15. What to say with respect to the claim made by the service provider?	Apply	The learner get to know the purpose of hypothesis testing	CO 5
13	The following data represents shear strength (X) of the test spot weld 392, 376, 401, 367, 389, 362, 409, 415, 358, 375. (a) Assuming that X is normally distributed, estimate the true average shear strength and standard deviation of shear strength using the method of maximum likelihood. (b) Obtain the MLE of P(X $\leq$400).	Apply	The learner get to know that how to find out MLE in point estimation	CO 5
14	Give the details on statistical analysis of data and explain Properties and scales of measurement	Understand	The learner will get to know the analysis of data and some scales of measurement	CO 5
15	What is Statistics for Data Analytics? Explain its types.	Understand	The learner will get to know the data analytics	CO 5

16	Explain Benefits of Statistics for Data Analytics. Give difference between obtain data and collect data.	Understand	The learner get to know the benefits of statistics	CO 5
17	A group of 10 foot surgery patients had a mean weight of 240 pounds. The sample standard deviation was 25 pounds. Find a confidence interval for a sample for the true mean weight of all foot surgery patients. Find a 95% CI.	Apply	The learner will get to know about the confidencce interval	CO 5
18	What is the use of a Confidence Interval in data science? How to calculate and interpret confidence interval?	Understand	The learner to know the concepts of confidence interval in detail	CO 5
19	Describe properties of Point Estimate.	Understand	The learner get to know the concepts pf point estimates	CO 5
20	What are the effects of the width of confidence interval?	Understand	The learner get to know the effects of confidence interval	CO 5
PART C - SHORT ANSWER QUESTIONS				
1	Give the importance of empirical rule.	Understand	The Learner to know the importance of empirical rule	CO 5
2	Annonate the term obtaining data? Explain two ways of collecting data.	Remember	The learner can learn that how to analyze the data	CO 5
3	Write short notes on point estimates.	Remember	the learner will get the information about the point estimates	CO 5
4	Describe the sampling in statistics? How many sampling methods are there?	Remember	The learner get to know about the concepts of sampling in statistics	CO 5
5	Annonate an alternative hypothesis?	Understand	The learner gte to define the alternative hypothesis	CO 5
6	Explain the properties of a normal distribution?	Understand	The learner get to know the properties of normal distribution	CO 5
7	How is missing data handled in statistics?	Understand	The learner get to know about how to handle data in statistics	CO 5

8	Explain the term P-value? And what is the use of it in machine learning?	Understand	The learner get to know abouut the P-value	CO 5
9	Give the the difference between normal distribution, standard normal distribution and uniform distribution?	Understand	The learner can get the overview of the types of distributions	CO 5
10	Where you have used Hypothesis Testing in your machine learning solution?	Understand	The learner can get to know about the hypothesis testing	CO 5
11	Define confidence interval.	Understand	The learner can get to know about the confidence interval	CO 5
12	Write short nptes on null hypothesis	Understand	The learner can get to know about the null hypothesis	CO 5
13	Give the difference between Covariance and Correlation?	Understand	The learner can verify the difference between the terms	CO 5
14	Explain the 6 points of estimation?	Remember	the learner will learn about point estimation	CO 5
15	Describe the term Collaborative filtering?	Understand	The learner can get to know about the collaborative filtering	CO 5
16	Give the difference between one tail and two tail hypothesis testing?	Understand	The learner get to know the basics of hyppothesis testing	CO5
17	Explain the formula of Logistic Regression?	Understand	The learner get to know about the logistic regression	CO 5
18	Why do we use sampling distributions?	Understand	The learner will get the information about the sampling distributions	CO 5
19	Can the empirical rule be negative?	Understand	The learner can get to know about the empoirical rule	CO 5
20	Why we need sampling distribution in big data analytics?	Understand	The learner can understand the need of sampling distribution	CO 5

MODULE V				
COMMUNICATING DATA				
PART A-PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)				
1	Is there a way we can measure or assess effective communication?	Understand	The learner get to know the concepts of effective communication	CO 6
2	What are some of the problems that may arise when trying to visualize data using charts and graphs?	Understand	The leaner get to know the problem related to visualization	CO 6
3	Has the development of the internet and social media caused a change in the way we communicate (quality, quantity, style etc.)?	Understand	The learner can develop their communicating skills	CO 6
4	What are the critical elements of visualization that we can use to illustrate data stories	Understand	The learner get to knpw the elements of visualization	CO 6
5	Do you think we should always visualize our data before performing analysis on it? If yes, then why? If not, then why not?	Understand	The learner get to know concepts of data visualization	CO 6
6	How does correlation different than causation?	Understand	The learner can get to know the difference of correlation and causation	CO 6
7	Why is Simpson's paradox a problem?	Understand	The learner to know the concepts Simpson's paradox	CO 6
8	Jenny has been requested her manager to make an oral presentation on the proposed good governance in business performance during the company's annual general meeting. Summarize the factors that Jenny should consider when planning for each of the following stages of oral presentation.	Understand	The learner will get the idea about the facors of presenetation	CO 6

9	What is the role of brain in interpreting data visuals?	Understand	The learner to know the role of brain at the time of data visuals	CO 6
10	Identify six factor that an entrepreneur should consider to make audio visual communication effective.	Understand	The learner get to kknow the some effective communixation factors	CO 6
PART-B LONG ANSWER QUESTIONS				
1	What makes an effective graph data visualization?	Understand	The learner can get the information about the effective graph data visualization	CO 6
2	Describe the 7 C's of communication.	Understand	The learner will get to know about importance of communication	CO 6
3	Why are line graphs used for data?	Understand	The learner to know about graphs	CO 6
4	Explain in detail advantages and disadvantages of verbal communication.	Understand	The learner will get to know about verbal communnication in deatil.	CO 6
5	Explain the presenting strategy with 'Wh'?	Understand	The learner can learn the presenting skills	CO 6
6	Give the importance of verbal communication?	Understand	The learner can get to know importance of verbal communication	CO 6
7	Define communiation. Explain its types. Describe the importance of communication.	Understand	The learner get the knowledge of verbal communication with its types	CO 6
8	Explain the components of communication process.	Understand	The learner will able to understand the communication process	CO 6
9	Explain three ways to improve your data science communication skills.	Understand	The learner get to know that how to improve your communication skills	CO 6
10	Explain different graphs and charts used in data science	Understand	The learner will get basic idea about graphs and charts	CO 6
11	Write down the six steps you have to follow at the time of data presenting?	Remember	The learner get to know the steps steps for presenting	CO 6

<table>
<tr><td>12</td><td>Give the detailing on verbal communication. What is the most crucial factor in verbal communication?</td><td>Understand</td><td>The learner will get to know the detail of verbal communication</td><td>CO 6</td></tr>
<tr><td>13</td><td>Find the value of the correlation coefficient from the data given in the following table which mentioned subject, age and glucose level:<table><tr><td>SUB</td><td>AGE(x)</td><td>GL(y)</td></tr><tr><td>1</td><td>43</td><td>49</td></tr><tr><td>2</td><td>21</td><td>65</td></tr><tr><td>3</td><td>25</td><td>79</td></tr><tr><td>4</td><td>42</td><td>75</td></tr><tr><td>5</td><td>57</td><td>87</td></tr><tr><td>6</td><td>59</td><td>81</td></tr></table></td><td>Apply</td><td>The learner get to know about correlation coefficient</td><td>CO 6</td></tr>
<tr><td>14</td><td>Give the importance of presentatin. What are the 5 P's of presentation?</td><td>Understand</td><td>The learner will get to know the 5 P's of presentation</td><td>CO 6</td></tr>
<tr><td>15</td><td>Explain the characteristics of an effective speech.</td><td>Understand</td><td>The learner can get to know about he effective speech</td><td>CO 6</td></tr>
<tr><td>16</td><td>Illustrate the term data visualization. Write steps for creating effective data visualization.</td><td>Understand</td><td>The learner can get to know the steps for creating the effective data visualization</td><td>CO 6</td></tr>
<tr><td>17</td><td>Differentiate between correlation and causation? How can we determine if variables are correlated?</td><td>Understand</td><td>The learner get to know the keywords related to correlation and causation</td><td>CO 6</td></tr>
<tr><td>18</td><td>What is correlation? Explain it's types in detail.</td><td>Understand</td><td>The learner can able to understand the correlation with its types</td><td>CO 6</td></tr>
<tr><td>19</td><td>Describe the effective data presentation techniques</td><td>Understand</td><td>The learner will get to know about the presenting skills for data presentation</td><td>CO 6</td></tr>
<tr><td>20</td><td>Explain different parts of Line Graph</td><td>Understand</td><td>The learner get to know the different parts of line graph</td><td>CO 6</td></tr>
<tr><td>21</td><td>Calculate the correlation coefficient of the given data.<table><tr><td>x</td><td>12</td><td>15</td><td>18</td><td>21</td><td>27</td></tr><tr><td>y</td><td>2</td><td>4</td><td>6</td><td>8</td><td>12</td></tr></table></td><td>Apply</td><td>The learner can able to find the vale for correlation coefficient</td><td>CO 6</td></tr>
</table>

21	Calculate the correlation coefficient for the following data. X = 4, 8 ,12, 16 and Y = 5, 10, 15, 20.	Understand	The learner will get to know about the correlation coefficient	CO 6
22	Explore how to present data. Give details.	Understand	The learner can get to know how to present data with different skills	CO 6
PART-C SHORT ANSWER QUESTIONS				
1	What makes a visualization ineffective?	Understand	The Learner can get the knowledge about the ineffective visualization	CO 6
2	Annonate the scatter plot with example?	Remember	The learner will learn the scatter plots in data visualization	CO 6
3	Why are line graphs used for data?	Understand	The learner to know about graphs	CO 6
4	Explain the bar charts in data science?	Apply	The learners can clear the concepts of bar charts	CO 6
5	Describe Simpson's Paradox in short?	Understand	The learner will get the basic information about the Simpson's Paradox	CO 6
6	Give the importance of graphs in data science?	Understand	the learner will get the information about the importance of graphs in data science	CO 6
7	What makes an effective Visualisation?	Understand	The learner will able to identify the effective visualization	CO 6
8	Explain three basic visualization considerations?	Understand	The learner can get the information about the basic visualization considerations	CO 6
9	Why is Simpson a paradox?	Understand	The Learner get to know the idea about Simpson's paradox	CO 6
10	Explain the purpose of a box plot?	Understand	The learner can get to know about the purpose box plot	CO 6
11	Why does communication matter in statistical foundations of data science?	Understand	The learner can get to know about the communication matter	CO 6
12	Why are visuals more effective than words?	Understand	The Learner can compare data using words and visuals	CO 6

13	How do you evaluate visualization?	Understand	The learner get to know the evaluation of visuslizatiom	CO 6
14	Describe Simpson's paradox and how does it pertain to confounding?	Understand	The learner can get to know about the Simpson's paradox	CO 6
15	How are graphs and statistics used in data analysis?	Understand	The learner get to know the knowledge of the graphs with statistics	CO 6
16	Write short notes on histograms	Understand	The learner can learn the basicsn of histograms in statistics	CO 6
17	Why is it called the Simpson's paradox?	Remember	The learner can get to know about the Simpson's paradox	CO 6
18	Give the examples of correlation and causation?	Understand	The learner can clear the concepts related toi correlation and causation	CO 6
19	Give the uses of verbal communication	Understand	The leaner will get the information about ht everbal communication	CO 6
20	Annonate the elements of effective data visualization?	Understand	the learner will get to know about the effective data visualization	CO 6

Course Coordinator: **HOD,CSE(AI&ML)**
Ms. K. Anjali, Assistant Professor