Course Code: 25CS5808

R25

SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY: PUTTUR
(AUTONOMOUS)
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK

Subject with Code: (25CS5808) COMPUTING FOR DATA ANALYTICS
Course & Branch: M.Tech CSE
Year &Sem: I-M.Tech & I-Sem
Regulation: R25

UNIT – I

DATA ANALYTICS LIFE CYCLE

1	**a)**	Define Big Data and explain how business analytics uses it for decision-making.	**[L1] [CO1]**	**[5M]**
	b)	Compare the four types of analytics with suitable business examples.	**[L1] [CO1]**	**[5M]**
2		Discuss how industries currently use analytics and how organizations progress to advanced analytics maturity	**[L2] [CO1]**	**[10M]**
3		Describe the role of a data scientist, outlining the key skills required and explaining how they collaborate with other team members in an analytics project.	**[L2] [CO1]**	**[10M]**
4	**a)**	List the key roles in an analytics project and briefly state what each role contributes	**[L3] [CO1]**	**[5 M]**
	b)	Analyze how collaboration among these roles ensures project success.	**[L3] [CO1]**	**[5 M]**
5		illustrate the phases of the analytics life cycle. Explain why each phase is critical for ensuring reliable	**[L4] [CO2]**	**[10 M]**
6	**a)**	What are the major challenges associated with Big Data Explain.	**[L4] [CO2]**	**[5M]**
	b)	Analyze how challenges affect data processing and decision-making.	**[L4] [CO2]**	**[5M]**
7		Discuss the importance of business understanding in the initial phase of an analytics project.	**[L1] [CO2]**	**[10M]**
8		Describe the process of data preparation and feature engineering in analytics and justify	**[L2] [CO3]**	**[10M]**
9		What is the significance of model development and evaluation in analytics projects? Describe techniques and metrics used to assess model performance.	**[L4] [CO1]**	**[10M]**
10		Develop a basic analytics project plan for a selected business issue.	**[L6] [CO4]**	**[10M]**

Course Code: 25CS5808

R25

UNIT – II
STATISTICS

1		.Explain and compare different sampling techniques used in statistical data collection.	**[L1] [CO1]**	**[10M]**
2		Discuss the methods of data classification and the role of tabulation in presenting data	**[L2] [CO2]**	**[10M]**
3		Describe how frequency distributions are constructed and represented using graphical methods.	**[L2] [CO2]**	**[10M]**
4		Explain the computation and uses of Arithmetic Mean, Geometric Mean, and Harmonic Mean	**[L3] [CO3]**	**[10M]**
5		Describe the calculation and significance of Mode and Median for grouped and ungrouped data.	**[L4] [CO3]**	**[10M]**
6		Explain Quartiles, Deciles, and Percentiles and discuss their importance in statistical analysis	**[L4] [CO3]**	**[10M]**
7		Discuss the different measures of variation, including Range, IQR, Quartile Deviation, and Mean Deviation	**[L1] [CO3]**	**[10 M]**
8		Explain Standard Deviation and Coefficient of Variation and how they help compare variability across datasets.	**[L2] [CO2]**	**[10 M]**
9		Describe the concept of skewness and explain how it is measured and interpreted using standard coefficients.	**[L5] [CO4]**	**[10M]**
10		Explain moments and discuss how kurtosis is measured and interpreted using central moments.	**[L6] [CO5]**	**[10M]**

Course Code: 25CS5808 **R25**

UNIT – III

PROBABILITY AND HYPOTHESIS TESTING

1		Define a random variable and distinguish between a discrete and a continuous random variable with examples.	**[L1] [CO1]**	**[10M]**
2		Explain the purpose of deriving a marginal density function from a joint density function and the process involved.	**[L2] [CO2]**	**[10M]**
3		State the probability mass function, mean, and variance for the Binomial and Poisson distributions	**[L2] [CO3]**	**[10M]**
4		Explain the conceptual difference between a Probability Mass Function (PMF) for discrete variables and a Probability Density Function (PDF) for continuous variables	**[L3] [CO3]**	**[10M]**
5		Explain the key characteristics of a Bernoulli trial and how they relate to the foundation of the Binomial distribution.	**[L4] [CO4]**	**[10M]**
6		The time to complete a series of tasks follows a Gamma distribution with $\alpha = 5$ and $\beta = 2$ (tasks/hour). Apply the distribution's properties to find its mean and variance.	**[L4] [CO4]**	**[10M]**
7		Explain the process and purpose of "standardizing" a normally distributed random variable.	**[L2] [CO3]**	**[10M]**
8		Elucidate the conditions under which the Poisson distribution serves as a good approximation to the Binomial distribution	**[L1] [CO4]**	**[10M]**
9		Write the probability density function and the cumulative distribution function for the Uniform distribution over an interval $[a,b]$	**[L5] [CO5]**	**[10M]**
10		Explain the relationship between the Cumulative Distribution Function (CDF) and the Probability Density Function (PDF) for a continuous random variable.	**[L6] [CO6]**	**[10M]**

Course Code: 25CS5808 **R25**

UNIT – IV

PREDICTIVE ANALYTICS

1		Differentiate between standard deviation and standard error of the mean.	**[L1] [CO1]**	**[10M]**
2		Evaluate the importance of sampling distribution in hypothesis testing.	**[L2] [CO2]**	**[10M]**
3		Differentiate between Type I and Type II errors with examples.	**[L2] [CO3]**	**[10M]**
4		Explain the properties of a good estimator: unbiasedness and efficiency	**[L3] [CO3]**	**[10M]**
5		Calculate the test statistic for a one-sample z-test.	**[L4] [CO4]**	**[10M]**
6		State the conditions for using t-distribution instead of normal distribution	**[L4] [CO4]**	**[10M]**
7	**a)**	Identify situations requiring F-distribution in statistical analysis.	**[L1] [CO3]**	**[5M]**
	b)	Evaluate the limitations of chi-square test with small expected frequencies.	**[L2] [CO3]**	**[5M]**
8	**a)**	Analyze the power of one-tailed versus two-tailed tests.	**[L1] [CO3]**	**[5M]**
	b)	Differentiate between correlation and causation.	**[L2] [CO3]**	**[5M]**
9	**a)**	Analyze the advantages of rank correlation with ordinal data.	**[L5] [CO5]**	**[5 M]**
	b)	Define Spearman's rank correlation coefficient.	**[L5] [CO5]**	**[5 M]**
10	**a)**	Analyze the relationship between R, R^2 and adjusted R^2.	**[L6] [CO6]**	**[5 M]**
	b)	Define multiple correlation coefficient in multiple regression	**[L6] [CO6]**	**[5 M]**

Course Code: 25CS5808

R25

UNIT – V

TIME SERIES FORECASTING AND DESIGN OF EXPERIMENTS

1		Explain how SES gives more weight to recent observations compared to SMA	**[L1] [CO1]**	**[10M]**
2		Compare the suitability of SMA versus SES for a trend-stationary time series	**[L2] [CO2]**	**[10M]**
3		Differentiate between experimental error and treatment effect in ANOVA.	**[L2] [CO2]**	**[10M]**
4		Define the three fundamental principles of Design of Experiments: Replication, Randomization, and Blocking, and state the purpose of each	**[L3] [CO3]**	**[10M]**
5		Explain why randomization is crucial in experimental design.	**[L4] [CO4]**	**[10M]**
6		Explain the null hypothesis tested in one-way ANOVA	**[L4] [CO4]**	**[10M]**
7		Explain the partitioning of total sum of squares in ANOVA.	**[L1] [CO4]**	**[10 M]**
8		Explain how Latin Square controls two sources of variation.	**[L2] [CO4]**	**[10M]**
9		Analyze the efficiency of factorial designs compared to one-factor-at-a-time experiments.	**[L5] [CO5]**	**[10M]**
10		Define 2^k factorial design and its components. and explain the concept of interaction in factorial experiments	**[L6] [CO6]**	**[10M]**

Prepared By:

Dr. K.ARUN KUMAR M.Tech, Ph.D.

PROFESSOR

Dept. of CSE/ SIETK.