21AD62-DATA SCIENCE AND APPLICATIONS
ANSWER KEY
IAT-1

USN | | | | | | | | | | |

Internal Assessment Test 1 –JUNE 2024

Sub	Data Science and Its Applications					SubCode:	21AD62	Branch:	AIML
Date	03/06/24	Duration:	90 minutes	Max Marks:	50	SEC	VI-A		OBE

	SCHEME AND SOLUTIONS	MARKS	CO	RBT
1a)	Explain the following with suitable examples i) Mode ii) Quantile iii) Standard normal distribution iv) Histograms v) Binomial Distribution Answer- Definition +Example-2M each i)Mode-It is a value with highest frequency in the data set +Example ii)Quantile-represents the value less than which a certain percentile of the data lies. +Example iii)Std normal distribution- It is of normal distribution with mean value 0 and standard deviation=1+formula for std.normal distribution iv)Histogram -A visual representation for continuous distribution +Example v)Binomial Distribution-a discrete distribution where there are 2 complementary out comes success and failure.+ Example	10	1	L2
2a)	Explain Bayes Theorem and solve the following, A diagnostic test has 99% accuracy and 60% of all people have Covid-19. If a patient tests positive, what is the probability that they have the disease? Answer- Bayes theorem 3 M It is used to determine the conditional probability of event A when event B has already occurred. -----(1) P(a|b)=(p(b|a)*p(a))/p(b)--- with each quantity explanation(2) Problem Solution- 3M P(positive|covid19) = 0.99 P(covid19) = 0.6 P(positive) = 0.6*0.99+0.4*0.01=0.598 $P(covid19\|positive) = \frac{0.99 * 0.6}{0.598} = 0.993$	6	1	L3
2b)	Describe the statement “correlation is not causation” with an example in detail. Answer- Correlation shows the relationship between the values of 2 variables. But we cannot conclude that one variable causes the change in the other. The relationship may be coincidental or a third factor may cause both the variables to change.----(3 M) Example-----(1M)	4	1	L2

3a)	Describe dispersion and variance and write the python code to compute the variance. Answer- Dispersion- it is used to measure how spread the data is +Example----(2M) Variance-Definition +Formula-----(2M) Python code:-----(2 M) def variance(x): n=len(x) d=mean(x) return sumofsquares(d)/n-1	6	1	L3
3b)	Illustrate central limit theorem with a neat diagram. Answer- Central limit theorem –2 M It states that the sample mean distribution of a random variable will assume a near normal or normal distribution if the sample size is enough though -sample size-distribution variation Diagram—2 M	4	1	L2
4a)	Explain the following with example, i) Null and Alternative Hypothesis ii) p-value iii) Type I Error. iv) Level of significance v) A/B Testing Answer- Definition +Example 2M each i) Null hypothesis -statement about a population that is believed to be true and Alternative Hypothesis -statement that is contradictory to the null hypotheses Example ii) p-value -The probability value obtained after the experiment. It is calculated from the Z value. It is compared with the level of significance to accept or reject null hypothesis-Conditions specify.+Example iii) Type I Error.-Its when your null hypothesis is true but we reject it in order to support the alternative hypothesis.+Example iv) Level of significance- parameter used in hypothesis testing to determine the threshold inorder to accept or reject the null hypothesis.+Example v) A/B Testing-Its is used to compare two products to find which is more efficient-method-example	10	2	L2
5a)	Explain the following with an example a) Named Tuple b) Data classes Answer- a) Named Tuple Definition +Example 3M b) Data classes Definition +Example 3M	6	2	L2

5b)	Write a python code for the probability density function of a Beta distribution.
	Answer-
	def B(alpha, beta):
	return math.gamma(alpha) * math.gamma(beta) / math.gamma(alpha + beta)
	
	def beta_pdf(x, alpha, beta):
	if x < 0 or x > 1:
	return 0
	return x ** (alpha - 1) * (1 - x) ** (beta - 1) / B(alpha, beta)
		4	2	L3

6a)	Explain how gradient descent is used to fit parameterized models. Answer- Use of Gradient descent-2M How it is determined-2M Linear regression-how theta are adjusted-6M	10	2	L2