THIS PIECE OF STUDY MATERIAL HAS BEEN BROUGHT TO YOU BY

MUSTUDENTS UNITED

contributed by - Dhruv Aswani of college - Thadomal Shahani Engineering College



FOR REMOVAL OF CONTENT OR CREDITS CONTACT US AT-INSTAGRAM ID-MUSTUDENTSUNITED

OR

MAIL US AT -MUSTUDENTSUNITED@GMAIL.COM

Attem Assum The fi Expla Expla State t	on num pt any the ne suitab gures to n in brie n in brie he reaso is anoma	hree of the ef the ns for	out of the aif neodight incomments taxono objecti	he rema cessary licate for my of t	aining fi and jus ull mark ime seri	stify the	assump	tions.								
Attem Assum The fi Expla Expla State t	pt any the suitable gures to not in bries in bries the reaso	hree of the ef the ns for	out of the aif neodight incomments taxono objecti	he rema cessary licate for my of t	aining fi and jus ull mark ime seri	stify the	assump	tions.								
Assun The fi Expla Expla State t What	ne suitab gures to n in brie n in brie he reaso	le dat the ri of the of the ns for	a if neo ight inco taxono objecti	cessary licate fo my of t	and jus ull mark ime seri	stify the	assump	tions.								
Expla Expla State t	gures to n in brie n in brie he reaso	the ri	ght inc	licate for	ill mark	īs —	ge ^{ett} ge	tions.								
Expla Expla State t What	n in brie n in brie he reaso	of the of the ns for	taxono objecti	my of t	ime seri		asting									
Expla State t What	n in brie he reaso	of the	objecti	450		ies forec	asting									
Expla State t What	n in brie he reaso	of the	objecti	450		ies forec	asting									
State t What	he reaso	ns for		ves of I		Explain in brief the taxonomy of time series forecasting										
State t What	he reaso	ns for			Data Ex	ploratio	n			5						
	is anoma	ly de		tliers oc						5						
Discus			tection	? Expla	in the p	rocess o	f anoma	ly detec	ction	5						
Discus																
	s the wo	rking	of the A	ARIMA	model ir	i detail				10						
Expla	n the DI	BSCA	N algo	rithm to	o detect	outliers	. Give th	ne advai	ntages	10						
and di	sadvanta	iges o	f the al	gorithn	n											
Expla	n SMOT	ΓE in	detail							10						
Expla	n the Da	ata Sc	ience P	rocess						10						
Find F	lowlev's	coef	ficient	of skew	ness fo	r the fol	lowing s	eries.		10						
40	1500						190	40								
State	he impo							rpose o	f scatter	10						
1 TO 1	7.0	100														
Find t	ne coeffi	cient	of skev	vness fi	rom the	data giv	en belo	W		10						
Size	0	3	4	5	6	7 0	8	9	10							
Freq	uency	7	10	14	35	102	136	43	8							
				series a	ipproacl	n is use	ful for	forecas	ting the	10						
Descr	be how	the p	redictiv		elling ca	n be ap	plied to	the Hou	ise price	10						
Expla	n the sig	gnific	ance of		ne, Dim	ension a	and Con	plexity	to Data	10						
	Explain Explai	Explain SMO Explain the Da Find Bowley's State the imporplots, quartile prind the coefficient of the coeffic	Explain SMOTE in Explain the Data Sc Find Bowley's coef State the important plots, quartile plots, Find the coefficient Size 3 Frequency 7 Describe how the demand for a production recomme Explain the signific	Explain SMOTE in detail Explain the Data Science P Find Bowley's coefficient 2, 4, 6, 5 State the importance of Da plots, quartile plots, bubble Find the coefficient of skey Size 3 4 Frequency 7 10 Describe how the time-s demand for a product Describe how the predictive prediction recommendation	Explain SMOTE in detail Explain the Data Science Process Find Bowley's coefficient of skew 2, 4, 6, 8, 10, 1 State the importance of Data Visu plots, quartile plots, bubble charts, Find the coefficient of skewness find the coeff	and disadvantages of the algorithm Explain SMOTE in detail Explain the Data Science Process Find Bowley's coefficient of skewness fo 2, 4, 6, 8, 10, 12, 14, 1 State the importance of Data Visualizatio plots, quartile plots, bubble charts, density Find the coefficient of skewness from the Size 3	Explain SMOTE in detail Explain the Data Science Process Find Bowley's coefficient of skewness for the fol 2, 4, 6, 8, 10, 12, 14, 16, 18, 26 State the importance of Data Visualization. State plots, quartile plots, bubble charts, density chart Find the coefficient of skewness from the data given by the coefficient of skewness from the following by the coefficient of skewness from the coefficient of skewness from the c	Explain SMOTE in detail Explain the Data Science Process Find Bowley's coefficient of skewness for the following s 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 State the importance of Data Visualization. State the purplots, quartile plots, bubble charts, density chart Find the coefficient of skewness from the data given below Size 3	Explain SMOTE in detail Explain the Data Science Process Find Bowley's coefficient of skewness for the following series. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 State the importance of Data Visualization. State the purpose of plots, quartile plots, bubble charts, density chart Find the coefficient of skewness from the data given below Size 3 4 5 6 7 8 9 Frequency 7 10 14 35 102 136 43 Describe how the time-series approach is useful for forecast demand for a product Describe how the predictive modelling can be applied to the Houprediction recommendation Explain the significance of Volume, Dimension and Complexity	Explain SMOTE in detail Explain the Data Science Process Find Bowley's coefficient of skewness for the following series. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 State the importance of Data Visualization. State the purpose of scatter plots, quartile plots, bubble charts, density chart Find the coefficient of skewness from the data given below Size 3 4 5 6 7 8 9 10 Frequency 7 10 14 35 102 136 43 8 Describe how the time-series approach is useful for forecasting the demand for a product Describe how the predictive modelling can be applied to the House price prediction recommendation Explain the significance of Volume, Dimension and Complexity to Data						

Total Marks 80 (3 Hours) NB 1) Question **number 1** is compulsory Attempt any three out of the remaining five questions. 2) 3) Assume suitable data if necessary and justify the assumptions **Q**1 Answer the following What is the difference between data science and data analytics a) What are Type I and Type –II errors? Give examples. b) Brief about SMOTE. c) d) What do you mean by Time Series Decomposition? Q2 Describe the terms: cross-validation, K-fold cross-validation, leave-1 out and Bootstrapping. Explain the data science process in detail. b) What are outliers? Explain different outlier detection methods. Calculate the coefficient of correlation for the following data with Karl Pearson's method. 50 70 80 100 10 15 20 Find Bowley's coefficient of skewness of the following series. 10 4.5 5.5 Size 6.5 10 22 25 40 15 10 18 Explain the Auto Regressive Integrated Moving Average (ARIMA) 10 model in detail. Brief about ANOVA and its types. How it is different from a t-test? 10 What is Hypothesis testing? Explain the steps involved in Hypothesis 10 testing with an example. 20 Write a note on any TWO: Data Visualization techniques Univariate Exploration and Multivariate Exploration House price Prediction or Fraud Detection

55453

Гіте	: 3hou	rs			Marks:	80
	NB.					
	1)	Question number 1 is compu	lsorv			AF.
	2)	Attempt any three out of the		g five questions.		
	3)	Assume suitable data if nece	ssary and	justify the assump	tions.	
	4)	Figures to the right indicate	full mark	s high		
01		1 011				
Q1	-)	Answer the following		77.1	2 1 D. 9	20
	a)	Explain the significance of da Data.	ita science	considering volum	e and Dimensions	OI
	b)	Explain performance evaluation	on with res	pect to Time series	Forecasting	
	c)	Write a note on measure of spr	read			
	d)	Write a note on Applications of	of Data Sci	ence		
Q2	a)	Explain Cross Validation, K-fo	old cross v	alidation in detail		10
~ -	b)	Explain the Data Science Proc				10
	D)	Explain the Bata Science 1 for	CSS III deta			
Q3	a) b)	What is Data Visualization? V Scatter plot. What are Outliers and their Ca yes, Explain.	20		260	ple 10
Q4	a)	Explain SMOTE in detail		CO 41 C 11		10
	b)	Calculate the Bowley's coeffic	cient of ske	ewness from the foll	owing distribution.	. 10
		Inc	come	NO. of persons		
			- 40	8 0		
		40	- 50	24		
			- 60	48		
		760	- 70	68		
		\$ 70	- 800	30		
		20 80	90	13		
		90	- 100	9.5		
Q5	a)	Explain Smoothing methods in	n Time Sei	ries Forecasting		10
	b)	A stenographer claims that she			-	
		we reject her claim on the basi				
		116 words with a standard dev	nation of 1	5 words? Use 5 % 1	evel of significance	÷,
		$Z_{\alpha} = 1.96.$				
04	3	W/L D D 1 4: 8	~i~??	-1-:		4.0
Q6	a)	What are Recommendation en	gines! Exp	oiain.		10

Total Marks 80

(3 Hours)

NB

- 1) Question **number 1** is compulsory
- Attempt any three out of the remaining five questions. 2)
- Assume suitable data if **necessary** and justify the assumptions. 3)
- Figures to the **right** indicate full marks 4)

Q1 Attempt any four

10

- Explain in brief the objectives of Data Exploration a)
- Explain in brief the taxonomy of time series forecasting b)
- What are the outliers in the dataset? State the reasons for the outliers c) occurring in the dataset
- d) Explain validation techniques bootstrap and cross-validation
- State the importance of Data Visualization. State the purpose of scatter plots, e) quartile plots, bubble charts, density chart
- Q2 Given data of 10 companies. Find out the type of correlation between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method

Company	1	2	3	4	5	6	77	8	9	10
Advt expenses	11	13	14	16	16	15	15	14	13	13
Sales volume	50	50	55	60	65	65	65	60	60	50

Explain the data science process in detail

Explain the density-based outlier detection approach 10 Q3 a) b)

Explain SMOTE in detail 10

- Explain the working of the Auto Regressive Integrated Moving Average 10 Model
 - The data given shows salary packages (in lakhs) offered after a campus 10 interview. Find the coefficient of skewness using Bowley's method.

Salary	4-8	8-12	12-16	16-20	20-24
No of Candidates	4	10	15	8	3

- Q5 a) What are the attributes of time series decomposition? Explain the classical 10 decomposition technique
 - b) In certain food experiment to compare two types of baby foods A and B, 10 the following results of the increase in weight (lbs) we observed in 8 children as follows

Food A	49	53	51	52	47	50	52	53
Food B	52	55	52	53	50	54	54	53

Examine the significance of the increase in weight of children due to food B. (Given t-value at alpha=0.05 is 2.365)

Q6

- a) Explain how the time-series approach is used to forecast the demand for a 10 product.
- b) Explain how predictive modelling can be applied to the House price 10 prediction recommendation