THIS PIECE OF STUDY MATERIAL HAS BEEN BROUGHT TO YOU BY

MUSTUDENTS UNITED

contributed by - Dhruv Aswani

of college - Thadomal Shahani Engineering College

(3 Hours) **Total Marks 80**

N B

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

Q1

	A	Explain in brief the taxonomy of time series forecasting	**5**
	B	Explain in brief the objectives of Data Exploration	**5**
	C	State the reasons for the outliers occurring in the dataset	**5**
	D	What is anomaly detection? Explain the process of anomaly detection	**5**
Q2	**A**	Discuss the working of the ARIMA model in detail	**10**
	B	Explain the DBSCAN algorithm to detect outliers. Give the advantages and disadvantages of the algorithm	**10**
Q3	**A**	Explain SMOTE in detail	**10**
	B	Explain the Data Science Process	**10**
Q4	**A**	Find Bowley's coefficient of skewness for the following series.	**10**

2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22

	B	State the importance of Data Visualization. State the purpose of scatter plots, quartile plots, bubble charts, density chart	**10**
Q5	**A**	Find the coefficient of skewness from the data given below	**10**

Size	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Frequency	**7**	**10**	**14**	**35**	**102**	**136**	**43**	**8**

	B	Describe how the time-series approach is useful for forecasting the demand for a product	**10**
Q6	**A**	Describe how the predictive modelling can be applied to the House price prediction recommendation	**10**
	B	Explain the significance of Volume, Dimension and Complexity to Data Science Techniques	**10**

Total Marks 80

(3 Hours)

NB

1) Question **number 1** is compulsory
2) Attempt **any three** out of the remaining **five questions**.
3) Assume suitable data if **necessary** and justify the assumptions.

Q1 Answer the following 20

a) What is the difference between data science and data analytics?

b) What are Type I and Type –II errors? Give examples.

c) Brief about SMOTE.

d) What do you mean by Time Series Decomposition?

Q2 a) Describe the terms: cross-validation, K-fold cross-validation, leave-1 out and Bootstrapping. 10

b) Explain the data science process in detail. 10

Q3 a) What are outliers? Explain different outlier detection methods. 10

b) Calculate the coefficient of correlation for the following data with Karl Pearson's method. 10

X	10	20	30	40	50	60	70	80	90	100
Y	2	4	8	5	10	15	14	20	22	50

Q4 a) Find Bowley's coefficient of skewness of the following series. 10

Size	4	4.5	5	5.5	6	6.5	7	7.5	8
F	10	18	22	25	40	15	10	8	7

b) Explain the Auto Regressive Integrated Moving Average (ARIMA) model in detail. 10

Q5 a) Brief about ANOVA and its types. How it is different from a t-test? 10

b) What is Hypothesis testing? Explain the steps involved in Hypothesis testing with an example. 10

Q6 **Write a note on any TWO :** 20

i. Data Visualization techniques
ii. Univariate Exploration and Multivariate Exploration
iii. House price Prediction or Fraud Detection

****

Time: 3hours **Marks: 80**

NB.

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

Q1 Answer the following **20**

a) Explain the significance of data science considering Volume and Dimensions of Data.

b) Explain performance evaluation with respect to Time series forecasting

c) Write a note on measure of spread

d) Write a note on Applications of Data Science

Q2 **a)** Explain Cross Validation, K-fold cross validation in detail **10**

b) Explain the Data Science Process in detail **10**

Q3 **a)** What is Data Visualization? What is Multivariate data? Explain with an example Scatter plot. **10**

b) What are Outliers and their Causes? Can Statistics be used to detect Outliers if yes, Explain. **10**

Q4 **a)** Explain SMOTE in detail **10**

b) Calculate the Bowley's coefficient of skewness from the following distribution. **10**

Income	NO. of persons
30 - 40	8
40 - 50	24
50 - 60	48
60 - 70	68
70 - 80	30
80 -90	13
90 - 100	9

Q5 **a)** Explain Smoothing methods in Time Series Forecasting **10**

b) A stenographer claims that she can type at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trails in which she demonstrates a mean of 116 words with a standard deviation of 15 words? Use 5 % level of significance, $Z_\alpha = 1.96$. **10**

Q6 **a)** What are Recommendation engines? Explain. **10**

b) Time series analysis using linear regression **10**

****

Total Marks 80

(3 Hours)

NB

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

Q1 **Attempt any four** 20

a) Explain in brief the objectives of Data Exploration

b) Explain in brief the taxonomy of time series forecasting

c) What are the outliers in the dataset? State the reasons for the outliers occurring in the dataset

d) Explain validation techniques bootstrap and cross-validation

e) State the importance of Data Visualization. State the purpose of scatter plots, quartile plots, bubble charts, density chart

Q2 a) 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 10

Company	1	2	3	4	5	6	7	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

b) Explain the data science process in detail 10

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

b) Explain SMOTE in detail 10

Q4 a) Explain the working of the Auto Regressive Integrated Moving Average Model 10

b) The data given shows salary packages (in lakhs) offered after a campus interview. Find the coefficient of skewness using Bowley's method. 10

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 decomposition technique 10

b) In certain food experiment to compare two types of baby foods A and B, the following results of the increase in weight (lbs) we observed in 8 children as follows 10

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 product. 10

b) Explain how predictive modelling can be applied to the House price prediction recommendation 10