



Name:

Enrolment No:

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, Dec 2023

Course: Predictive Modelling

Semester: V

Program: BBA ABD

Course Code: DSBA 3006

Time : 03 hrs.

Max. Marks: 100

Instructions: Attempt all sections

SECTION A
10Qx2M=20Marks

S. No.		Marks	CO
Q 1	Attempt all multiple choice questions		CO1
a.	The purpose of applying data reduction is a) to generate a larger set of variables b) to remove negative values c) to use a smaller set of variables that capture maximum information d) None of the above	2	CO1
b.	A graph that uses vertical bars to represent data is called as a) Line graph b) Bar graph c) Scatterplot d) Vertical graph	2	CO1
c.	What does a data value measure after standardization (centering and scaling)? a) accuracy b) number of standard deviations between each data point and median c) number of standard deviations between each data point and mean d) slope	2	CO1
d.	The main benefit of standardizing a dataset is a) it makes multiple variables of a dataset come to a common scale. b) eliminates negative data values c) makes data interpretation easier.	2	CO1
e.	What is an outlier? a) data point most proximal to mean b) data point that falls outside the overall pattern. c) data point above or below 3 standard deviations of the mean.	2	CO2

f.	<p>_____ are used when you want to visually examine the relationship between two quantitative variables.</p> <p>a) Bar graph b) pie graph c) line graph d) Scatterplot</p>	2	CO2
g.	<p>Financial fraud detection is an example of:</p> <p>a) Prediction problem b) Clustering problem c) Outlier detection problem d) None of these</p>	2	CO2
h.	<p>What kind of target variables are we dealing with in simple linear regression?</p> <p>a) continuous b) binary c) categorical</p>	2	CO2
i.	<p>On what stage of data exploration are the missing values handled?</p> <p>a) Data transformation b) Data reduction c) Data cleaning d) All of the above</p>	2	CO2
j.	<p>Statement 1: Data transformation works on individual variables. Statement 2: Data reduction works on a set of variables.</p> <p>a) Only statement 1 is true b) Only statement 2 is true c) Both the statements are True d) Both the statements are False</p>	2	CO1

SECTION B
4Qx5M= 20 Marks

Q2.	What do you understand by data cleaning? What is an outlier? Explain the process of outlier detection.	5	CO2
Q3.	What is dimensionality reduction? Explain the difference between feature extraction and feature selection.	5	CO1
Q4.	What is curse of dimensionality?	5	CO2
Q5.	Explain the difference between feature extraction and feature selection.	5	CO1

SECTION-C
3Qx10M=30 Marks

Q6.	Explain in detail the steps in Principle component Analysis.	10	CO2
Q7.	What do you understand by a time series? Explain in detail.	10	CO2
Q8.	What do you understand by CART and CHAID? What is the difference between the two?	10	CO2

SECTION-D
2Qx15M= 30 Marks

Q9.	<p>Considering the following confusion matrix, define and compute the following:</p> <ul style="list-style-type: none"> a) Accuracy b) Precision c) Recall d) F1 score e) Sensitivity 	15	CO3																				
<table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td colspan="2" rowspan="2" style="text-align: center;">N=165</td> <td colspan="2" style="text-align: center;">Predicted</td> <td></td> </tr> <tr> <td style="text-align: center;">No</td> <td style="text-align: center;">Yes</td> <td></td> </tr> <tr> <td rowspan="2" style="text-align: center;">Actual</td> <td style="text-align: center;">No</td> <td style="text-align: center;">50</td> <td style="text-align: center;">10</td> <td style="text-align: center;">60</td> </tr> <tr> <td style="text-align: center;">Yes</td> <td style="text-align: center;">5</td> <td style="text-align: center;">100</td> <td style="text-align: center;">105</td> </tr> <tr> <td></td> <td></td> <td style="text-align: center;">55</td> <td style="text-align: center;">110</td> <td></td> </tr> </table>				N=165		Predicted			No	Yes		Actual	No	50	10	60	Yes	5	100	105			55
N=165		Predicted																					
		No	Yes																				
Actual	No	50	10	60																			
	Yes	5	100	105																			
		55	110																				

Q10.	<p>Study the Excel regression output that follows. How many predictors are there? What is the equation of the regression model? Using the key statistics discuss the strength of the model and its predictors.</p>	15	CO3																																																				
<p>SUMMARY OUTPUT</p> <hr/> <p><i>Regression Statistics</i></p> <hr/> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">Multiple R</td> <td style="text-align: right;">0.814</td> </tr> <tr> <td>R Square</td> <td style="text-align: right;">0.663</td> </tr> <tr> <td>Adjusted R Square</td> <td style="text-align: right;">0.636</td> </tr> <tr> <td>Standard Error</td> <td style="text-align: right;">51.761</td> </tr> <tr> <td>Observations</td> <td style="text-align: right;">28</td> </tr> </table> <hr/> <p>ANOVA</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th style="text-align: center;"><i>df</i></th> <th style="text-align: center;"><i>SS</i></th> <th style="text-align: center;"><i>MS</i></th> <th style="text-align: center;"><i>F</i></th> <th style="text-align: center;"><i>Significance F</i></th> </tr> </thead> <tbody> <tr> <td>Regression</td> <td style="text-align: center;">2</td> <td style="text-align: right;">131567.02</td> <td style="text-align: right;">65783.51</td> <td style="text-align: right;">24.55</td> <td style="text-align: right;">0.0000013</td> </tr> <tr> <td>Residual</td> <td style="text-align: center;">25</td> <td style="text-align: right;">66979.65</td> <td style="text-align: right;">2679.19</td> <td></td> <td></td> </tr> <tr> <td>Total</td> <td style="text-align: center;">27</td> <td style="text-align: right;">198546.68</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <hr/> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th style="text-align: center;"><i>Coefficients</i></th> <th style="text-align: center;"><i>Standard Error</i></th> <th style="text-align: center;"><i>t Stat</i></th> <th style="text-align: center;"><i>P-value</i></th> </tr> </thead> <tbody> <tr> <td>Intercept</td> <td style="text-align: right;">203.3937</td> <td style="text-align: right;">67.518</td> <td style="text-align: right;">3.01</td> <td style="text-align: right;">0.0059</td> </tr> <tr> <td>X₁</td> <td style="text-align: right;">1.1151</td> <td style="text-align: right;">0.528</td> <td style="text-align: right;">2.11</td> <td style="text-align: right;">0.0448</td> </tr> <tr> <td>X₂</td> <td style="text-align: right;">-2.2115</td> <td style="text-align: right;">0.567</td> <td style="text-align: right;">-3.90</td> <td style="text-align: right;">0.0006</td> </tr> </tbody> </table>				Multiple R	0.814	R Square	0.663	Adjusted R Square	0.636	Standard Error	51.761	Observations	28		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	Regression	2	131567.02	65783.51	24.55	0.0000013	Residual	25	66979.65	2679.19			Total	27	198546.68					<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	Intercept	203.3937	67.518	3.01	0.0059	X ₁	1.1151	0.528	2.11	0.0448	X ₂	-2.2115	0.567
Multiple R	0.814																																																						
R Square	0.663																																																						
Adjusted R Square	0.636																																																						
Standard Error	51.761																																																						
Observations	28																																																						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>																																																		
Regression	2	131567.02	65783.51	24.55	0.0000013																																																		
Residual	25	66979.65	2679.19																																																				
Total	27	198546.68																																																					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>																																																			
Intercept	203.3937	67.518	3.01	0.0059																																																			
X ₁	1.1151	0.528	2.11	0.0448																																																			
X ₂	-2.2115	0.567	-3.90	0.0006																																																			