Vivekanand Education Society's Institute of Technology, Chembur, Mumbai,
Department Of Artificial Intelligence and Data Science
Year:2023-24 (Even Sem)
MID TERM TEST

Class : D11AD	**Division: -**
Semester: VI	**Subject: Data Analytics & Visualization**
Date: 26th Feb 2024	**Time: 9 am to 10 am**

CO	CO1	CO2	CO5	CO6
%	**20%**	**50%**	**15%**	**15%**

Q.1)		(Attempt any five of the following)	Marks (20)	CO's Mapped
	a)	Enlist the key elements in a Chart	2M	LO5, LO6
	b)	A scientific foundation wanted to evaluate the relation between y= salary of the researcher (in thousands of dollars), x_1= number of years of experience, x_2= an index of publication quality, x_3=sex (M=1, F=0), and x_4= an index of success in obtaining grant support. A sample of 35 randomly selected researchers was used to fit the multiple regression model. Parts of the computer output appear below. Predictor Coef SE Coef T P Constant 17.846931 2.001876 8.915 0.0001 Years 1.103130 0.359573 3.068 0.0032 Papers 0.321520 0.037109 0.0002 Sex 1.593400 0.687724 2.317 0.0083 Grants 1.288941 0.298479 4.318 0.0003 s = 1.75276 R-sq = 92.3% adj R-sq = 91.4% (a) The least squares line fitted to the data is (b) How many degrees of freedom does the t* value from the previous question have?	2M	LO2
	c)	With a neat diagram explain briefly the 7 stages involved in Visualizing Data.	2M	LO5, LO6
	d)	Enlist the key roles involved in a Data Analytics Project	2M	LO1
	e)	Enlist 5 tools used for Narrative Visualization	2M	LO5, LO6
	f)	Explain the 5 elements of a Box Plot	2M	LO5, LO6

<table>
<tr><td>Q.2)</td><td>a)</td><td>Given a dataset of pass / fail in an exam of 5 students in the given table. Use Logistic Regression as a classifier to answer the following questions. Given the log(odds) = -64 + 2 * hours
<table>
<tr><th>Hours Study</th><th>Pass (1) / Fail (0)</th></tr>
<tr><td>29</td><td>0</td></tr>
<tr><td>15</td><td>0</td></tr>
<tr><td>33</td><td>1</td></tr>
<tr><td>28</td><td>1</td></tr>
<tr><td>39</td><td>1</td></tr>
</table>
(a) Calculate the probability of a pass for the student who studied for 40 hours.
(b) At least how many hours student study that helps him /her to pass the course with a probability of more than 80%</td><td>5M</td><td>LO2</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td></tr>
<tr><td></td><td>b)</td><td>A biologist assumes a linear relationship between the amount of fertilizer supplied to tomato plants and the subsequent yield of tomatoes obtained. He randomly selected 8 tomato plants of the same variety and treated them weekly with a solution in which x grams of fertilizer was dissolved in a fixed quantity of water. The yield in y kilograms of tomatoes was recorded.
<table>
<tr><td>Plant</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td>H</td></tr>
<tr><td>x</td><td>1.0</td><td>1.5</td><td>2.0</td><td>2.5</td><td>3.0</td><td>3.5</td><td>4.0</td><td>4.5</td></tr>
<tr><td>y</td><td>3.9</td><td>4.4</td><td>5.8</td><td>6.6</td><td>7.0</td><td>7.1</td><td>7.3</td><td>7.7</td></tr>
</table>
(a) Calculate the equation of the Least Squares Regression line of y on x.
(b) Estimate the yield of a plant treated, weekly with 3.2 grams of fertilizer.</td><td>5M</td><td>LO2</td></tr>
<tr><td>Q.3)</td><td>a)</td><td>Compare Linear Regression with Logistic Regression with 5 parameters</td><td>5M</td><td>LO2</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td></tr>
<tr><td></td><td>b)</td><td>With a neat diagram explain the Data Analytics Lifecycle. For each phase write down the steps involved.</td><td>5M</td><td>LO1</td></tr>
</table>