Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

MAIN ANSWER - BOOK

The Answer Book is bar coded. Identity of the students in any form should not be revealed in the answer book/s supplementary sheet

SUPERVISOR'S SIGNATURE WITH DATE	AF RE
04-08-25	

PROGRAMME: MECH. ENGG. (REGULAR / Re-Examination) TICK WHICHEVER IS APPLICABLE

EXAM DATE: 4/08/2025 SEMESTER: VIII

SUBJECT: BUSINESS ANALYTICS PAPER: 5

MAIN ANSWER BOOK: 1+ SUPPLEMENTARY SHEET/S 0 = 1 (TOTAL TO BE FILLED BY THE CANDIDATE)

INSTRUCTIONS TO THE CANDIDATES

1. This answer-book contains 32 pages. Check whether the relevant answer-book provided contains 32 pages and whether the pages are properly numbered.
2. Candidates should occupy their correct seats as per the seating plan displayed and write appropriate details in the space provided for the purpose.
3. Candidates will not be permitted to leave the examination hall until half an hour after the question papers are distributed.
4. Candidates, who are not in their seats by the time notified, will not as a rule, be permitted to appear for the examinations.
5. Candidates must produce their photo identity card / hall ticket provided by the college for verification to the room/ block supervisor during the examination. Candidates without the valid identity card / hall ticket will not be allowed to appear for the examination.
6. Every candidate present must sign against his /her student number on the attendance sheet provided by the room/ block supervisor.
7. Candidates should ensure that all answer-books including supplementary sheets provided to them bear the signature of the room / block supervisor and date of examination without which the answer-book will not be examined. They should tie all supplementary sheets to the main answer-book relating to the same paper and enter on the first page of the answer-book, the total number of supplementary sheets tied together.
8. Do not damage or make any stray marks on the barcodes.
9. Write on both sides of a page. Rough work, when necessary should be done only in the answer-book by specifically marking the page as Rough Work.
10. Candidates should write their answers in legible handwriting with blue / black pen only. Answers written in illegible handwriting or with pencil may not be evaluated.
11. Candidates should not write their name, roll no., student no. etc., anywhere in the answer-book and reveal their identity in any form in the answers written by them. Writing these details or putting signature is revelation of identity. Use of religious invocation or any writing that is not relevant to the answers anywhere in the answerbooks will be treated as attempts to reveal identity.
12. While underlining of answers for focusing attention is permitted. However, use of varied inks, except for illustrations and figures must be avoided. DO NOT use any symbol like encircling the question or using colour arrows for P.T.O. These will all be considered as attempts to readily identify the specific answer-books and will be treated as unfair means.
13. Candidates are forbidden to (i) bring any book, notes, scribbling paper, mobile phones or any other similar devices (ii) speak or communicate in any manner with any other candidate, while the examination is in progress (iii) take with them any answer-book, written or blank, while leaving the examination hall.
14. Exchange of writing material, stencils, mathematical instruments, question-paper, etc., is strictly prohibited. If candidates want anything, they should approach the room / block supervisor without disturbing other candidates.
15. All answer-books/supplementary sheets supplied should be returned back to the room / block Supervisor, whether written or blank. Smuggling /carrying the answer book/ supplementary sheet in/out of the examination hall will amount to adoption of unfair means.
16. No sheet should be torn from the answer-books provided nor shall additional papers be attached to them.
17. Any method to bribe the examiner/s by attaching currency notes or letters or making appeal inside the answerbook is strictly prohibited and will result in serious action being taken by the authorities.
18. Candidates should not write anything on the question-paper.
19. Candidates suspected to be guilty of any of the aforesaid acts will be allowed to write their paper only after giving an undertaking in writing that the decision of the Institution in respect of the reported act of unfair means, is binding on them.
20. The answer-books will be scrutinised before they are sent to examiners. If the college authorities are convinced that any candidate has attempted to reveal his / her identity by any means, the answer-books may NOT be sent to the examiner for evaluation and the candidate's case will be dealt with as per the provisions of University Ordinance 5050.
21. A warning bell will be given ten minutes before the close of the examination. Candidates will not be allowed to leave the examination hall during the last ten minutes. At the final bell, they must stop writing and hand over their answer-books to the room / block supervisor. They should not leave their seats until answer-books from all candidates are collected by the supervisor.
22. A candidate who disobeys any instructions issued by the Senior / Room Supervisor or who is guilty of rude or disobedient behaviour is liable for disciplinary action to be taken against him / her by the institution.
23. Candidates are requested to go through the provisions of the Maharashtra Act O. XXXI of 1982, i.e. an Act to provide for preventing malpractices at University Board and other specified examinations, published in the Maharashtra Government Gazette, dated 14th October, 1982. Every offence under this Act is a cognizable and non-bailable offence.

IT IS PRESUMED THAT CANDIDATE HAS READ ALL THE ABOVE INSTRUCTIONS.

QUESTION NOS.	MARKS AWARDED	MODERATED MARKS	MARKS AFTER REVALUATION
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
Total			
	Examiner's Signature	Moderator's Signature	Re-Valuator's Signature

P.T.O.

Space for marks	Question No.		
	Q1		
	A. i	BUSINESS INTELLIGENCE [BI]	BUSINESS ANALYTICS [BA]
		1. It helps answering What happened? and What is happening?	1. It helps answering Why did it happend? and what to do?
		2. It is a combination of discriptive analysis and real time processing	2. It is a combination of predictive analysis and diagnostic analysis.
		3. Analysis techniques include OLAP, KPI's etc.	3. Techniques include ML algorithms, optimization techniques etc.
		4. Used by executive manag-ers, production team etc for quick insights.	4. Used by data analyst & data Scientists.
		5. Tools :- Power BI, Tableau etc	5. Tools :- Python, R, Sci-kit learn etc.
		6. Examples like Average inventory per day, Customer click through rate (CTR) etc	6. Examples :- forecasting weather, inventory etc
		7. Works on real time and historical data	7. Primarily focused on historical data, learns from incoming data.

Space for marks | Question No.

Q1

B. Given :-

x	500	1000	1500	2000	1500
y	0.25	0.55	0.85	1.15	1.45

TABLE

x	y	x^2	xy	$x_i - \bar{x}$	$y_i - \bar{y}$
500	0.25	250000	125	-1000	-0.6
1000	0.55	1000000	550	-500	-0.3
1500	0.85	2250000	1275	0	0
2000	1.15	4000000	2300	500	0.3
2500	1.45	6250000	3625	1000	0.6
Σ 7500	4.25	13750000	7875		

$\bar{x} = 1500$ $\bar{y} = 0.85$

FITTING LINEAR REGRESSION

Equation $y = a + bx$

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$4.25 = 5a + 7500b$ —— equation 1

$7875 = 7500a + 13750000b$ —— equation 2

Solving equation 1 & 2

$a = -0.05$ $b = 6 \times 10^{-4}$

$\therefore y = -0.05 + 6 \times 10^{-4} x$

for amplitude at 3000 rpm

$y = -0.05 + 6 \times 10^{-4} \times 3000$

$y = 1.75$ mm.

from the observation it shows that the amplitude increases with increase in rpm.

Space for marks | Question No.

$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1000000	0.36	600
250000	0.09	150
0	0	0
250000	0.09	150
1000000	0.36	600
Σ 2500000	0.9	1500.

CORRELATION

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \, \Sigma(y_i - \bar{y})^2}}$$

$$\therefore r = \frac{1500}{\sqrt{2500000^2 \times 0.9^2}}$$

$$r = 6.67 \times 10^{-4}$$

There is a right positive correlation, indicating that the increase in rpm increase the amplitude.

Space for marks	Question No.

Q2

A

i. ~~x 23 23 25~~

x 20 21 23 23 25 26 27 28 30 32

1] Mean = $\frac{20+21+23+23+25+26+27+28+30+32}{10}$

$\mu = 25.5$ months

The average lifespan of engine belt is 25.5 months

2] Mode

Here 23 is repeated the most, i.e twice

Therefore mode = 23

3] Median

Arranging in ascending.

Median = $\frac{5^{th} \text{ term} + 6^{th} \text{ term}}{2}$

$= \frac{25+26}{2} = 25.5$ months

The median lifespan of engine belt is 25.5 months

4] Variance

$$\sigma^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$$

$$\sigma^2 = \frac{(20-25.5)^2 + (21-25.5)^2 + 2(23-25.5)^2 + (25-25.5)^2 + (26-25.5)^2 + (27-25.5)^2 + (28-25.5)^2 + (30-25.5)^2 + (32-25.5)^2}{\cancel{10}\ 9}$$

$\sigma^2 = \cancel{16.33}\ 14.94$ months^2

Space for marks	Question No.	

5] Standard deviation

$\sigma = \sqrt{\sigma^2} = \cancel{4.04}\ 3.86$ months.

The deviation from the mean is $\cancel{4.04}\ 3.86$ months

6] Coefficient of variation $= \frac{\mu}{\sigma} = \frac{25.5}{\cancel{4.04}\ 3.86} = \cancel{6.31}\ 6.60$

7] Skewness.

$Sk = \frac{3(\mu - \text{Median})}{\sigma} = \frac{3(\cancel{25.5 - 25.5})}{4.04}$

$= \frac{\mu - \text{mode}}{\sigma} = \frac{25.5 - 23}{\cancel{4.04}\ 6.60\ 3.86} = 0.6847$

The data distribution is slightly positively skewed.

8] Kurtosis

$K = \frac{\Sigma(x_i - \bar{x})^4}{n\sigma^4} =$

$= \frac{\{(-5.5)^4 + (-4.5)^4 + (-5)^4 + (-0.5)^4 + (0.5)^4 + (1.5)^4 + (2.5)^4 + (4.5)^4 + (6.5)^4}{5 \times \cancel{4.04}\ 3.86^4}$

$= 3.28$

Since Approximately 3 or greater than 3 Therefore the distribution is Mesokurtic.

Space for marks | Question No.

Q2

B.

PRIMARY DATA COLLECTION	SECONDARY DATA COLLECTION
1. Internally generated data	1. Externally generated data
2. Less Costly	2. Data involved purchase cost
3. Highly accurate type of data collection.	3. Data may not be accurate.
4. Primary data provides data of what is required	4. There mighty some unnecessary data which needs extra preprocessing
5. Not subjected to any licencing for use	5. Subjected to licencing.
6. Example - Data generated from sensors of Machine	6. Example :- Government data, WHO data etc.
7. Freshly and data collected for the first time [Ex. Interview of Sports athelet]	7. Data of COVID cases in from the year 2020 to 2024 etc.

Space for marks	Question No.	
	Q3	
	A.	
	1.2.	Given :- $X_1 = 250\,KN$ $X_2 = 1000°C$

Solution:-

Sigmoid equation : $\frac{1}{1+e^{-z}}$

Here

$$z = -5 + 0.02x_1 + 0.05x_2$$

$$\therefore P(y=1|x) = \frac{1}{1+e^{-(-5+0.02x_1+0.05x_2)}}$$

$$\therefore P(y=1) = \frac{1}{1+e^{-(-5+0.02\times250+0.05\times1000)}}$$

$$= \frac{1}{1+e^{-25}} = 1$$

Here $P(y=1) \geq 0.5$, then failed

$P(y=1) < 0.5$, then No failure.

$\therefore$ The bearing will fail under the given condition.

2. Given $\theta = 1000$ hrs, $\gamma = 0$, $\beta = 2$, $t = 100$ hrs

$R(t) = 1 - e^{-\left(\frac{t-\gamma}{\theta}\right)^{\beta}}$

for 1 parameter

$R(t) = 1 - e^{-\left(\frac{100-0}{1000}\right)^{1}} = 0.63 \sim 63\%$

$MTTF = \gamma + \theta\Gamma(1+1/\beta) = \cancel{0+1000\Gamma(1+1}$

for 2 parameter.

$R(t) = 1 - e^{-\left(\frac{100-0}{1000}\right)^{2}} = 9.95\times10^{-3}\%$

$\therefore MTTF = 0 + 1000\Gamma(1+1/2) = 1000\Gamma(1.5) = 886.23$ hrs.

Space for marks | Question No.

Q3 For 3 parameter

$R(t) = 9.95 \times 10^{-3}\%$

MTTF = 886.23 hours

B.

DECISIONS TREES	SUPPORT VECTOR MACHINE
1. It is a classification and regression type of Supervised learning algorithm.	1. It is a classification type of Supervised learning algorithm.
2. The components include Root node, leaf, decision node.	2. The components are data points plotted on the axis.
3. Uses algorithms like ID3, CART	3. Uses mathematical equations.

Space for marks | Question No.

Q4.

A

i. CREDIT CARD FRAUD DETECTION

In banking sector, unusual financial transcation needs to be report immediately to avoid financing losses. Adapting Supervised learning techniques can help analysis the incoming transaction data from the user and to flag it as spam or not scam.

following steps can be implemented to integrate supervised learning technique into the system.

Step1 :- Data Collection :- The data of the user like transaction ID, Name, location, transcation mode, Amount etc is to be collected for analysis.

Step 2 :- Data Preprocessing :- Remove duplicates [By considering various variables], remove NaN values or replace it with other values using relevant techniques, Normalizing, Standardizing and Encoding the data.

Step 3 :- Model training

The data is split into test and training data generally in a ratio of 80% training and 20% test.

The model are trained and Cross Validation is used. Algorithms like (Decision trees, SVM, Naive Bayes etc can be used].

Space for marks | Question No.

Q4.

Step 4 :- Model evaluation and Selection
The trained model can be evaluted using various evaluation metrics like accuracy, precision, recall, f1-score, AUC etc.
The best performing model is selected.

Step 5 :- Integration.
for classification type of algorithms, the output can be ~~vati~~ passed through certain rules like ? for positive - No flagging
for uncertain - Manual checking
for negative → flagging as risky.

Step 6 :- Retraining on new incoming data to improve the performance of the algorithm.

B.

FROM \ TO	ACTIVE	INACTIVE
ACTIVE	0.7	0.3
INACTIVE	0.4	0.6

Probability Matrix = $\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$

Initial probability = $\begin{bmatrix} 0.8 & 0.2 \end{bmatrix}$

• Expected proportion of active customer next months

$$= \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}' = \begin{bmatrix} 0.64 & 0.36 \end{bmatrix}$$

Active customer is 64%.

Space for marks	Question No.

- Expected proportion of active customers after two months:

$$= [0.8 \quad 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^2 = [0.592 \quad 0.408]$$

Active Customer = 59.2%.

- Expected proportion of active customer after three months

$$= [0.8 \quad 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^3 = [0.5778 \quad 0.4224]$$

Active Customer = ~~0.57~~ 57.78%.

- Long term steady state proportion of active & inactive customers

$$= [0.8 \quad 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^n = [0.5778 \quad 0.4224].$$

Space for marks	Question No.

Q5

A.

ii TEXT MINING TECHNIQUES TO EXTRACT INSIGHTS FROM A SET OF CUSTOMER REVIEWS.

STEP 1 :- Data Collections.

Collecting the data of customer reviews.
Extracting data from website using web scraping [Beautifulsoup etc] or go exporting as text data.

STEP 2 :- Data Preprocessing techniques.

Let's assume we have 2 reviews from customer.
Review 1 :- [Nice product it is easy to use]
Review 2 :- [Damaged product poor quality]

We start with tokenizing
Review 1 = ["Nice", "product", "it", "is", "easy", "to", "use"]
Same for review 2.
Next we remove stop words like it, is, to etc. Once this is done stemming / lemmitization can be applied, then, Damaged → damage. can be done.
Using BoW [Bag of Words] to count the frequency of each word.
Review 1 = [{Nice : 1}, {product : 2}, {easy : 1}]
Review 2 [{use : 1}, {poor : 1}, {damage : 1}, {quality : 1}].

STEP 3 :- Sentiment analysis.

The text data can be analysed using dictonary based algorithms like VADER etc.
Machine learning based algorithms like division trees, SVM, Random forest.

Space for marks	Question No.	

Deep learning algorithms like BERT, ROBERTa etc can be used

~~STEP~~ - STEP 4:- Classification of data
Positive, Neutral, Negative, Here Review 1 is Positive & Review 2 is Negative.
STEP 5:- Visualization & Interpretation
The positive, negative & neutral review counts can be visualized used pie charts, bar charts etc.

B. EDGE COMPUTING FUTURE POTENTIAL IN REAL-TIME ANALYTICS IN LOGISTICS OR MANUFACTURING

Edge computing can help in real-time analytics of data generated from systems In case of Manufacturing data from robots can be used to track active movement of robots, Machine data like temperature, power consumption etc can help in real-time maintenance and failure prediction. Edge computing tools like sensors, Camera, footballing tools can help reduce unpredictive movements in manufacturing.

In Case of logistics, real-time controlling of AGV's [Automated Guided Vehicles] can be easily monitored, selecting optimum route ~~for~~ based on real time data can be done. Similarly for ASRS [Automatic storage and retrival System] can be applied.

Space for marks	Question No.	

UBS uses edge computing to track customer data, vehicle route optimization.
Amazon automatic grocery store used edge computing to track customer in the store and find high density spots which can help in placement of products.

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No.	

Space for marks	Question No	

Space for marks	Question No.	

(Rough Work)

Space for marks	Question No.	