

variable indicating whether the customer churned (1 for "Churn" and 0 for "No Churn").

Build a suitable regression model to determine whether customers would churn or not. Write down the steps with python/R codes for building the model using the above dataset. (15)

6. What is a confusion matrix and what is its purpose in classification problem? Consider the following confusion matrix :

Prediction	No	50	5
	Yes	10	100
		No	Yes
	Actual		

Regarding the above table, calculate the following :

- Precision
 - Recall
 - F score
 - Accuracy
 - Specificity (15)
7. Write short notes on any **three** :
- Social media analytics
 - Multicollinearity
 - ROC and Cut-off probability
 - Decision Tree and Random Forest (15)

(500)

(M.) 22/12/2023
[This question paper contains 4 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 504

G

Unique Paper Code : 2922062301

Name of the Paper : Introduction to Business Analytics

Name of the Course : **Bachelor of Management Studies (BMS)**

Semester : III

Duration : 3 Hours

Maximum Marks : 90

Instructions for Candidates

- Write your Roll No. on the top immediately on receipt of this question paper.
- Attempt any **six** questions.
- All** questions carry equal marks.
- Show your working clearly in your answer sheet.

- 'Data analytics is essential for businesses to make informed decisions'. In the light of the current statement explain the role of data analytics and its various types. (15)

- Differentiate between Supervised and Unsupervised machine learning with the help of suitable examples. (7)

P.T.O.

(M)

(b) How does descriptive analytics differ from other types of analytics. Consider there is a dataset of 'salary' with one of the variables as 'years_education'. Write suitable Python/R code to determine range, standard deviation, and variance of 'years_education'. (8)

3. (a) What is regression and its types? State the assumptions in a linear regression model. (7)

(b) How do you know that linear regression is suitable for any given data? What are the different types of metrics used to determine the accuracy of a regression model? (8)

4. Consider the output summary of a regression model that aims to predict 'Employee Performance Rating'. The model uses 'Years of Experience', 'Education Level', and 'Job Satisfaction' as predictor variables.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.5	5.681	5.752	0.0004
Years of Experience	0.4	0.014	3.397	0.001
Education Level	0.2	0.925	2.993	0.003
Job Satisfaction	0.3	0.066	0.461	0.645

Degree of freedom (Residuals): 390

R-squared: 0.767, Adjusted R-squared: 0.759

F-statistic: 96.39 on 3 and 391 DF, p-value: < 1.2e-11

(a) Formulate the regression equation, specifying the predictors, the target variable, and the size of the dataset.

(b) Explain the meaning of the intercept in the context of this employee performance model and determine its statistical significance.

(c) Interpret the coefficients for 'Years of Experience,' 'Education Level,' and 'Job Satisfaction.' Assess their statistical significance and provide insights into the relative importance of these predictors in explaining employee performance.

(d) Evaluate the quality of the model's fit using R-squared and the F-statistic. Offer interpretations on the model's overall effectiveness in predicting employee performance ratings. (15)

5. What is a classification problem? In a dataset 'ChumData', the aim is to predict customer chum based on the provided customer related variables.

CustomerID: A unique identifier for each customer (categorical variable).

ContractDuration: The duration of the customer's subscription contract in months (continuous variable).

TotalCharges: The total charges incurred by the customer (continuous variable).

ContractType: The type of contract the customer has (categorical variable with options like "Month-to-Month," "One Year," "Two Year").

MonthlyUsage: The average monthly data/usage by the customer (continuous variable). Chum: The target