BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: BTECH **SEMESTER : V**
BRANCH: AIL **SESSION : MO/2025**

SUBJECT: AI353 FEATURE ENGINEERING

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Explain the Data centric and Model centric approaches of Feature Engineering. Also explain the importance of feature engineering in terms of its advantages in machine learning domain.	[5]	1	1
Q.1(b)	Explain the role of Exploratory Data Analysis (EDA) in Feature Engineering. Which type of plots from EDA help us to observe the relationship among the features. Explain with suitable example.	[5]	1	2
Q.2(a)	The following data represents the monthly expenses (in Rs.1000) of 12 students: 13, 9, 12, 8, 42, 14, 15, 18, 15, 16, 10, 40, Using the Interquartile Range (IQR) method, identify any outliers.	[5]	2	3
Q.2(b)	The following data represents the ages of 10 persons: 18, 22, 25, 27, 35, 42, 45, 47, 52, 60 Perform equal-width binning and equal-frequency binning using 3 bins. (i) Determine the bin ranges for equal-width binning and assign each value to its bin. (ii) Divide the data into 3 equal-frequency bins (each bin with equal number of observations). (iii) Write the final binned data for both methods.	[5]	2	3
Q.3(a)	The following dataset represents the number of products sold by a shop on 8 different days: 2, 3, 4, 6, 9, 15, 25, 40 Verify the skewness of the data, whether it is positively or negatively skewed by taking suitable plot. Perform a Box-Cox Transformation on the data using $\lambda = 0.5$. And ensure whether it has reduced the skewness or not.	[5]	3	3
Q.3(b)	Explain the importance of categorical variable encoding. Explain the difference between OneHot, Dummy and Binary Encoding with suitable example.	[5]	3	2
Q.4(a)	Explain the concept of feature interaction in feature construction. How does combining two or more existing features help improve model performance? Give one example each of a linear and a non-linear feature interaction.	[5]	4	3
Q.4(b)	You are given a dataset with a categorical target variable (three classes) and one numeric feature X. The feature values for each class are shown below:	[5]	4	3

Class	Feature (X) values
C_1	2, 3, 4
C_2	5, 6, 7
C_3	8, 9, 10

Using ANOVA (Analysis of Variance), determine whether the feature X is significantly different across the classes, and hence whether it should be selected for classification. Use the Fcrit = 5.14 to compare the calculated F statistic with Fcrit.

PTO

Q.5(a)	Explain the Bag of words and Tf-IDF word encoding for document vectorization with suitable example. How the word2vec word embedding is different from these encoding techniques? Explain with suitable example.	[5]	5	2
Q.5(b)	By taking a suitable small example, describe the components a time series data decomposition. How can we extract the trend component from the time series data. Explain it in same example.	[5]	5	3

::::::21/11/2025::::::M