Q. Code: 376055

Reg. No. | | | | | | | | | | | | | |

B. E / B. TECH.DEGREE EXAMINATIONS, MAY 2024

Third Semester

OE18512 – MACHINE LEARNING TOOLS AND TECHNIQUES

(Computer Science and Engineering)

(Regulation 2018A)

TIME: 3 HOURS **MAX. MARKS: 100**

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	To understand the overview of different problems solved using machine learning, Learn on introduction to supervised learning techniques.	**3**
CO 2	To Learn on unsupervised learning techniques. (Including python implementation)	**4**
CO 3	To Understand various problems under NLP umbrella and learn on text mining.	**3**
CO 4	To Learn on time series analysis. (Including python implementation)	**5**
CO 5	To Learn on operationalization of machine learning models.	**3**

PART- A (10x2=20Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	What is machine learning and its classifications?	**1**	**2**
2.	Brief about Loss Function with example.	**1**	**2**
3.	Describe Principal Component Analysis (PCA).	**2**	**3**
4.	Analyze the Elbow method and its application in machine learning.	**2**	**4**
5.	What are stop words? Give Example.	**3**	**2**
6.	List any two real-life applications of Natural Language Processing.	**3**	**2**
7.	Analyze the purpose of the Train-Test split in the modeling process for time series data.	**4**	**4**
8.	Examine what plots are typically used in the exploratory data analysis (EDA) phase of time series modeling?	**4**	**4**
9.	What are some examples of commonly used cloud services?	**5**	**2**
10.	What is MLOps, and what does CI/CD stand for in the context of MLOps?	**5**	**2**

PART- B (5x 14=70Marks)

		Marks	CO	RBT LEVEL
11. (a)	Discuss in detail about multiple linear regression with example.	**(14)**	**1**	**3**
	(OR)			
(b)	Discuss in detail about Decision tree algorithm with a neat sketch and an example.	**(14)**	**1**	**3**

12. (a)	How does the K-means algorithm work, and what are its key components? Provide a detailed explanation with a practical example demonstrating its application and the process of clustering data.	**(14)**	**2**	**4**
	(OR)			
(b)	How do labels, Centroids, inertia, eigenvectors, and eigenvalues contribute to the key features of a cluster, and what is their significance in cluster analysis?	**(14)**	**2**	**4**
13. (a)	What are the fundamental techniques used in Text Mining, such as Text Cleaning, Regular Expressions, Stemming, and Lemmatization? How do these methods help in processing and analyzing textual data effectively?	**(14)**	**3**	**3**
	(OR)			
(b)	How do techniques like Bigrams & Trigrams play a role in Text Classification and Sentiment Analysis tasks? Explain their importance in understanding and categorizing text data effectively.	**(14)**	**3**	**3**
14. (a)	Analyze the significance of Exploratory Data Analysis (EDA) in the modeling process for time series data, highlighting its role in understanding data patterns and trends.	**(14)**	**4**	**4**
	(OR)			
(b)	Examine the key metrics used in model evaluation for time series forecasting, such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), discussing their interpretability and relevance in assessing model accuracy.	**(14)**	**4**	**4**
15. (a)	Discuss the relationship between cloud technology, Git, and CI/CD in MLOps, identifying how each component contributes to the overall efficiency and scalability of machine learning workflows.	**(14)**	**5**	**3**
	(OR)			
(b)	Compare and contrast the fundamentals of cloud technology with commonly used cloud services, highlighting their key features and functionalities.	**(14)**	**5**	**3**

PART- C (1x 10=10Marks)

(Q.No.16 is compulsory)

		Marks	CO	RBT LEVEL
16.	Evaluate the effectiveness of using the ARIMA model to analyze the provided time series data that illustrates the number of passengers of an airline per month from the year 1949 to 1951. Discuss the model's ability to capture the underlying patterns and fluctuations in the data, considering its predictive accuracy and any potential limitations. Additionally, assess the suitability of alternative time series forecasting methods for this specific dataset and justify your reasoning based on comparative analysis.	**(10)**	**4**	**5**

date	value	date	value	date	value
01-01-1949	112	01-01-1950	115	01-01-1951	145
01-02-1949	118	01-02-1950	126	01-02-1951	150
01-03-1949	132	01-03-1950	141	01-03-1951	178
01-04-1949	129	01-04-1950	135	01-04-1951	163
01-05-1949	121	01-05-1950	125	01-05-1951	172
01-06-1949	135	01-06-1950	149	01-06-1951	178
01-07-1949	148	01-07-1950	170	01-07-1951	199
01-08-1949	148	01-08-1950	170	01-08-1951	199
01-09-1949	136	01-09-1950	158	01-09-1951	184
01-10-1949	119	01-10-1950	133	01-10-1951	162
01-11-1949	104	01-11-1950	114	01-11-1951	146
01-12-1949	118	01-12-1950	140	01-12-1951	166

****