SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, RAMAPURAM

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ANSWER KEY SUBMISSION

Date of Exam & Session	**21/02/2023**	**Category of Exam**	**CLA1**
Course Name	**Applied Machine Learning**	**Course Code**	**18CSE481T**
Name of the Faculty submitting	**Dr. M. Mahasree**	**Date of submission of Answer Key**	**28/02/2023**
Department to which the faculty belongs to	**CSE**	**Total Marks**	**25**

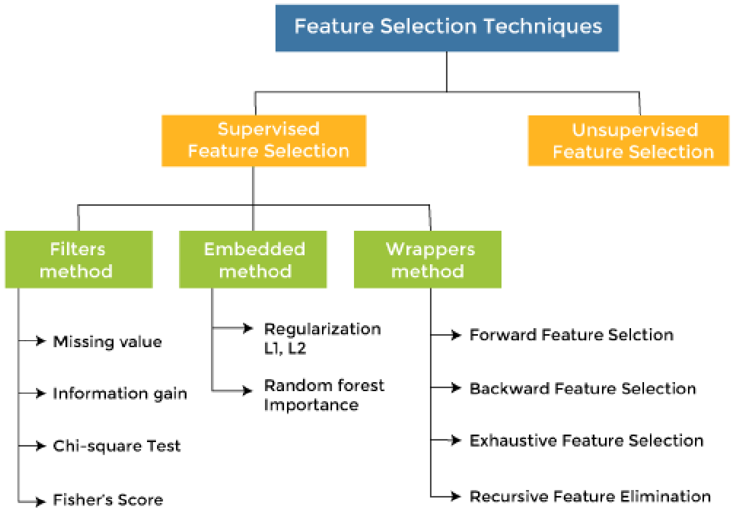
PART A (5x1= 5)
ANSWER ALL THE QUESTIONS

Q.No.	MCQ Questions	Marks	CO	BL	PI
1.	A model of language consists of the categories which does not include _____________. (A). System Unit **(B). Structural units** (C). data units (D). empirical units	1	1	1	2.5.2
2.	Different learning methods does not include? **(A). Introduction** (B). Analogy (C). Deduction (D). Memorization	1	1	2	2.5.2
3.	The model will be trained with data in one single batch is known as? (A). Batch learning (B). Offline learning **(C). Both A and B** (D). None of the above	1	1	1	1.7.1
4.	How do you handle missing or corrupted data in a dataset? (A). Drop missing rows or columns (B). Replace missing values with mean/median/mode (C). Assign a unique category to missing values **(D). All of the above**	1	1	2	1.7.1
5.	Which of the following techniques cannot be used for normalization in text mining? (A). Stemming (B). Lemmatization **(C). Stop word removal** (D). None of the above	1	1	2	1.7.1

PART B (2 X 4= 8)
ANSWER ALL THE QUESTIONS

Q. No.	Questions	Marks	CO	BL	PI
6	**Why Data Normalization is necessary for Machine Learning models?** Normalization is a data preparation technique that is frequently used in machine learning. The process of transforming the columns in a dataset to the same scale is referred to as normalization. Every dataset does not need to be normalized for machine learning. It is only required when the ranges of characteristics are different. Normalization is useful when your data has variable scales and the technique you're employing, such as k-nearest neighbors and artificial neural networks, doesn't make assumptions about the distribution of your data.	4	1	3	2.5.2
7	**What is the difference between feature extraction and feature selection? When should each one be used?** The key difference between feature selection and extraction is that feature selection keeps a subset of the original features while feature extraction algorithms transform the data onto a new feature space**.** We need to do feature selection when the dataset contains a large number of features, or when the features are highly correlated, redundant, or irrelevant.	4	1	3	2.5.2

PART C (1 X 12= 12)
ANSWER THE QUESTIONS

Q. No.	Questions	Marks	CO	BL	PI
8	**a) Summarize the best way to select features in supervised learning problems?** **Techniques (4)** **Explanation (8)**  *Forward selection* - Forward selection is an iterative process, which begins with an empty set of features. After each iteration, it keeps adding on a feature and evaluates the performance to check whether	12	1	1	1.7.1

	it is improving the performance or not. The process continues until the addition of a new variable/feature does not improve the performance of the model. *Backward elimination* - Backward elimination is also an iterative approach, but it is the opposite of forward selection. This technique begins the process by considering all the features and removes the least significant feature. This elimination process continues until removing the features does not improve the performance of the model. *Exhaustive Feature Selection*- Exhaustive feature selection is one of the best feature selection methods, which evaluates each feature set as brute-force. It means this method tries & make each possible combination of features and return the best performing feature set. *Recursive Feature Elimination*- Recursive feature elimination is a recursive greedy optimization approach, where features are selected by recursively taking a smaller and smaller subset of features. Now, an estimator is trained with each set of features, and the importance of each feature is determined using coef_attribute or through a feature_importances_attribute. *Regularization*- Regularization adds a penalty term to different parameters of the machine learning model for avoiding overfitting in the model. This penalty term is added to the coefficients; hence it shrinks some coefficients to zero. Those features with zero coefficients can be removed from the dataset. The types of regularization techniques are L1 Regularization (Lasso Regularization) or Elastic Nets (L1 and L2 regularization). *Information Gain:* Information gain determines the reduction in entropy while transforming the dataset. It can be used as a feature selection technique by calculating the information gain of each variable with respect to the target variable. *Chi-square Test:* Chi-square test is a technique to determine the relationship between the categorical variables. The chi-square value is calculated between each feature and the target variable, and the desired number of features with the best chi-square value is selected. *Fisher's Score* - Fisher's score is one of the popular supervised technique of features selection. It returns the rank of the variable on the fisher's criteria in descending order. Then we can select the variables with a large fisher's score.				
	OR				
	b) Explain in detail about what dimensionality reduction is? **Definition (2)** **Application (2)** **Advantages (2)** **Techniques (3)** **Explanation (3)**	12	1	2	5.6.2

Dimensionality reduction technique can be defined as, *"It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information."* These techniques are widely used in machine learning for obtaining a better fit predictive model while solving the classification and regression problems.

It is commonly used in the fields that deal with high-dimensional data, such as speech recognition, signal processing, bioinformatics, etc. It can also be used for data visualization, noise reduction, cluster analysis, etc.

Benefits of applying Dimensionality Reduction

- By reducing the dimensions of the features, the space required to store the dataset also gets reduced.
- Less Computation training time is required for reduced dimensions of features.
- Reduced dimensions of features of the dataset help in visualizing the data quickly.
- It removes the redundant features (if present) by taking care of multicollinearity.

Wrappers Methods

The wrapper method has the same goal as the filter method, but it takes a machine learning model for its evaluation. In this method, some features are fed to the ML model, and evaluate the performance. The performance decides whether to add those features or remove to increase the accuracy of the model. This method is more accurate than the filtering method but complex to work. Some common techniques of wrapper methods are:

Forward Selection

Backward Selection

Bi-directional Elimination

Embedded Methods:

Embedded methods check the different training iterations of the machine learning model and evaluate the importance of each feature. Some common techniques of Embedded methods are:

LASSO

Elastic Net

Ridge Regression, etc.

Common techniques of Dimensionality Reduction

Principal Component Analysis

Backward Elimination

Forward Selection

Score comparison

	Missing Value Ratio Low Variance Filter High Correlation Filter Random Forest Factor Analysis Auto-Encoder *Principal Component Analysis* is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. *The backward feature elimination* technique is mainly used while developing Linear Regression or Logistic Regression model. In this technique, by selecting the optimum performance of the model and maximum tolerable error rate, we can define the optimal number of features require for the machine learning algorithms. *Forward feature selection* follows the inverse process of the backward elimination process. It means, in this technique, we don't eliminate the feature; instead, we will find the best features that can produce the highest increase in the performance of the model.				

COURSE COORDINATOR **HOD/CSE**