Subject Name: Machine Learning
Subject Code: 3170724

Faculties: Ms. Munira Topia, Dr. Shruti Yagnik

		QUESTIONS	
		UNIT NO- 1 :	
		Introduction to Machine Learning	
		TOPIC:1 : Human Learning and its Types, Machine Learning and its Types	
Sr. No		**SHORT QUESTIONS/MCQ**	**Marks**

1.		Identify correct groups of machine learning problems:
		   a) Given a transaction, label as fraud or regular, learn fraud detection.
		   b) Given set of movies found on a portal and group them into set of movies with same genre.
		   c) Given database of visitors, automatically discover popularity of holiday place segments and group visitors according to holiday place.
		   d) Find pattern of purchasing electronics during Diwali. **[LJIET]**
		
		1-supervised learning, 2,3,4-unsupervised
		1-classification 2-Regression, 3-4 Association rule
		1,2- supervised learning 3,4- unsupervised learning
		1,2- unsupervised learning 3,4-supervised learning
		1,3-supervised learning 2,4-unsupervised learning
		1,2,3-supervised learning 4-unsupervised learning
			01

2.		1. Training a dog with sound of a bell to pick a ball. 2. Rat learning how to carefully step on Skinner box to release food. 3. Child plays game with friends in recess after observing other adults also play games at home. a. Example 1 is type of learning under expert guidance b. Example 2 is type of learning by self c. Example 3 is type of learning under expert guidance d. Example 1,3 both are type of learning under expert guidance e. Example 2,3 both are type of learning by self f. Example 1,2 both are type of learning by self **[LJIET]** True- a,b False-c,d,e,f True- a,e False-b,c,d,f True- b,d False-a,c,e,f True- c,f False-a,b,d,e True-d False-a,b,c,e,f True-b,c False-a,d,e,f	01
3.		Identify correct groups of machine learning problems: 1. Given a geographical data, label areas that are hit by flood or earthquake. 2. Given a set of post found on a social networking site, group users into set with similar sentiments. 3. Given database of satellite images, automatically discover places segments where air pollution index is similar. Recognizing handwritten letter and converting into e-mail. **[LJIET]**	01

1,4-supervised learning, 2,3,-unsupervised 1-classification 2-Regression, 3-4 Association rule 1,2- supervised learning 3,4- unsupervised learning 1,2- unsupervised learning 3,4-supervised learning 1,3-supervised learning 2,4-unsupervised learning 1,2,3-supervised learning 4-unsupervised learning	

4.	A computer program is said to _____ from experience E with respect to some _____ of task T and performance measure P, if it's _____ at _____ as measured by performance P improves with _____. **[LJIET]** learn, class, performance, task T, experience E. learn, performance, class, task T, experience E learn, class, performance, experience E, task T performance, class, learn, task T, experience E performance, class, learn, experience E, task T performance, learn, class, experience E, task T	01
5.	The choice of model used for solving specific problem is? **[LJIET]** Human task Machine learning task Abstraction Generalization task Decision tree Mathematical equation	01

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	Explain how Human learning happens.**[LJIET]**	07
2.	Compare different types of Machine Learning. **[LJIET]**	07
3.	What are the different types of Supervised Learning approaches? Explain by giving examples. **[LJIET]**	07
4.	What are the different types of Unsupervised Learning approaches? Explain by giving examples. **[LJIET]**	07
5.	What is generalization? What role does it play in Machine Learning? **[LJIET]**	07

	TOPIC:2 :Well Posed Learning Problem, Applications of Machine Learning, Issues in Machine Learning	

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	State and describe in details any two application areas of Machine Learning.**[LJIET]**	07
2.	Which are the problems that cannot be solved using Machine Learning.**[LJIET]**	03

UNIT NO- 2 :
Preparing to Model

	TOPIC:1 : Basic Data Types, Exploring numerical data, Exploring categorical data, Exploring relationships between variables	

Sr. No	SHORT QUESTIONS/MCQ	Marks

1.	a. Divide data into two parts: training and testing b. Find potential issues in data c. Understand nature and quality of data d. Do remediation e. Train model based on training data Arrange given preprocessing activities in correct order for supervised learning. **[LJIET]** c, b, d, a, e a, c, d, b, e c, d, b, e, a c, d, b, a, e a, c, b, d, e a, b, c, d, e	01
2.	a. Discrete values can assume finite or count ably infinite number of values b. Pin code is example of Nominal attribute c. Nominal attribute may or may not have finite number of values. d. Numerical attributes can have count ably infinite values e. Nominal and Ordinal attributes are in general discrete. Which statement is/are false? **[LJIET]** only a both a,c both b,e all three c, d, e all three a, c, d none	01
3.	_____ is likely to get shifted drastically due to presence of _____. **[LJIET]** Mean, Outlier Median, Outlier Mean, Variance Mean, Median Variance, Outlier Median, Variance	01
4.	<table><tr><td>1. Mechanism for one shot view and understand nature of data</td><td>a. Whisker</td></tr><tr><td>2. Box span from first to third quartile</td><td>b. Box Plot</td></tr><tr><td>3. Can range up to 1.5 times IQR from bottom/top of box</td><td>c. Median</td></tr><tr><td>4. Denoted by line or band in box</td><td>d. Inter quartile range</td></tr></table>**[LJIET]** 1-b, 2-d, 3-a, 4-c 1-d, 2-b, 3-a, 4-c 1-b, 2-d, 3-c, 4-a 1-d, 2-b, 3-c, 4-a 1-c, 2-a, 3-d, 4-b 1-c, 2-b, 3-d, 4-a	01
5.	1. Requires at-least one independent variable 2. Requires at-least one dependent variable	01

	3. Determines future outcome of data 4. Determines behavior of data **[LJIET]** ==Supervised learning: 1,2,3 Unsupervised learnnig: 1,4== Supervised learning: 1,3 Unsupervised learnnig: 2,4 Supervised learning: 2,3 Unsupervised learnnig: 1,4 Supervised learning: 1,2 Unsupervised learnnig: 3,4 Supervised learning: 1,3,4 Unsupervised learnnig: 2 Supervised learning: 2 Unsupervised learnnig: 1,3,4	
6.	What is first and third quartile of data. 44, 35, 52, 35, 31, 37, 49, 39, 28 **[LJIET]** ==33, 46.5== 35,49 31,44 31,49 35,44 32.5,45	01
7.	To explore relationship between variables what can be used? 1. Box plot 2. Histogram 3. Scatter plot 4. Cross tab 5. PCA **[LJIET]** ==3,4== 1,2 1,2,3,4 5 1,5 1,2,4	01
8.	The height of _____ reflects total count of data elements whose value falls specifically _____. **[LJIET]** ==Bar, bin== Bin, bar Histogram, mode Mode, histogram Bar, mode Bin, mode	01
9.	1. Discrete values can assume only finite values 2. Employee ID is example of Nominal attribute 3. Nominal attribute may or may not have finite number of values. 4. Numerical attributes can have count ably infinite values 5. Nominal and Ordinal attributes are in general discrete. Which statement is/are false? **[LJIET]** ==only a== both a,c both b,e all three c, d, e all three a, c, d none	01

10	Numerical attributes having lesser possible number of values can be treated as _____.**[LJIET]** **Categorical** Nominal Ordinal Discrete Interval Ratio	01
11	1. Larger the value of variance indicates more dispersion in data. 2. Deviation between mean and median is significant high, means chances of outliers is less. 3. Larger difference between two quartile values indicates more data spread in respective quarter. Identify false statements **[LJIET]** All 1,2,3 Both 1,2 Both 2,3 Only 1 **Only 2** Only 3	01
12	What is first and third quartile of data. 44, 12, 25, 71, 27, 59, 59, 38, 66 **[LJIET]** **26,62.5** 27,59 27,44 38,59 25.5,68.5 44,59	01
13	_____ data, if have possibly only two values/ labels is called 'dichotomous'. **[LJIET]** **Nominal** Interval Ordinal Ratio Discrete Binary	01

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	What are the main activities involved when you are preparing to start with modeling in Machine Learning? **[LJIET]**	07
2.	What are the basic data types in Machine Learning? Explain by giving examples of each one of them. **[LJIET]**	07
3.	Explain in details the different components of Box Plot. State how outliers can be detected using Box Plots. **[LJIET]**	07

4.	State and explain the two ways in which we can explore the relationship between two variables (attributes). **[LJIET]**	07

TOPIC:2 : Data Issues and Remediation, Data Pre-processing

Sr. No	SHORT QUESTIONS/MCQ	Marks
1.	Issues of data quality **[LJIET]** Missing values Outliers Missing values and outliers Error in data collection Error in sample selection Error in data collection and sample selection	01
2.	Techniques for dimensionality reduction PCA stands for **[LJIET]** Principle Component Analysis Primary Coefficient Analysis Principle Coefficient Analysis Primary Component Analysis Partial Component Aggregation Partial Component Aggregation	01
3.	Dimensionality reduction helps in removing _____ and _____ features. **[LJIET]** Redundant and ir-relevant Redundant and relevant Noisy and redundant Relevant and rich Noisy and ir-relevant Rich and noisy	01
4.	1. removing columns which have too many missing values 2. removing columns which have high variance in data 3. removing columns with dissimilar data trends which of given statement(s) is/ are correct: **[LJIET]** Only 1 Only 2 Only 3 Both 1,3 Both 2,3 None of given	01

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	Explain in details the various ways to address the missing values in a dataset. **[LJIET]**	07
2.	Define Outliers. Can they be handled in a dataset? If yes, how? **[LJIET]**	05
3.	What are the different techniques for data pre-processing. Explain in brief, dimensionality reduction and feature selection.**[LJIET]**	07
	UNIT NO- 3 : **Modelling and Evaluation**	

Sr. No	SHORT QUESTIONS/MCQ	Marks
	TOPIC:1 : Selecting a model, Training a model, Model representation and interpretability	
1.	A machine learning problem that does not include target variable is called _____ and that include target variable is called _____. **[LJIET]** descriptive learning, predictive learning predictive learning, descriptive learning descriptive learning, reinforcement learning predictive learning, reinforcement learning reinforcement learning, predictive learning reinforcement learning, descriptive learning	01
2.	Summarized knowledge representation of raw data is called as? **[LJIET]** Model Abstraction Generalization Training Testing Validation	01
3.	Full Form of – SRSWS ? **[LJIET]** Simple Random Sampling With Replacement Sampling Random Simple Without Replacement Simple Random Sampling Without Replacement Simple Repeated Sampling With Replacement Simple Repeated Sampling Without Replacement Sampling Random Simple With Replacement	01
4.	1. _____ method divides data into 3partitions namely training data, validation data, test data 2. _____ is part of input data with known labels 3. _____ is used to measure performance of a model 4. _____ is used in iteration to refine a model 5. _____ is used only once on finalize models to report final performance of a model. **[LJIET]** Holdout, Test data, Validation Data, Validation Data, Test Data Holdout, Test data, Validation Data, Training Data, Test Data K-fold CV, Test data, Validation Data, Validation Data, Test Data K-fold CV, Test data, Validation Data, Training Data, Test Data 3-fold CV, Test data, Validation Data, Validation Data, Test Data 3-fold CV, Test data, Validation Data, Training Data, Test Data	01
5.	To measure the extent to which model is going wrong in estimating the relationship between Independent Variable and Dependent Variable is called as? **[LJIET]** Cost function Loss function Objective function No free lunch Target Function AUC	01

6.	_____ randomly picks data instance from input data set , with possibility of same data instance being picked up multiple times. **[LJIET]** <mark>Bootstrapping</mark> LOOCV 10-fold CV k-fold CV Holdout Stratified Random Sampling	01
7.	Less than 0.5 – Worse than random prediction 0.5-0.7 weak predictive ability 0.7-0.9 good to fair predictive ability 0.9-1.0 excellent predictive ability Given value indications are used with which measure? **[LJIET]** <mark>AUC</mark> ROC Residual F-measure Kappa-coefficient purity	01

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	Compare and contrast Descriptive and predictive Models. **[LJIET]**	05
2.	Explain in details about the methods to train a learning model. **[LJIET]**	07
3.	What is underfitting and overfitting in the context of Machine Learning Model? **[LJIET]**	07
4.	Write about the bias-variance tradeoffs in context of model fitting. **[LJIET]**	07

TOPIC:2 : Model Performance Evaluation, Performance Improvement

Sr. No	SHORT QUESTIONS/MCQ	Marks
1.	a. Follows general principle of machine learning b. Also called as non-parametric learning c. Doesn't need to refer back training data d. Decision Tree e. k-nearest neighbor **[LJIET]** <mark>Eager Learner–a,c,d Lazy Learners -b,e</mark> Eager Learner–a,d Lazy Learners –c,b,e Eager Learner–a,e Lazy Learners b,c,d Eager Learner–b,d Lazy Learners –a,c,e Eager Learner–b,e Lazy Learners -a,c,d Eager Learner–b,c,d Lazy Learners -a,e	01
2.	Model Accuracy **[LJIET]** <mark>(TP+TN)/(TP+FP+FN+TN)</mark> (TP+FN)/(TP+FP+FN+TN) (FP+TN)/(TP+FP+FN+TN) (FP+FN)/(TP+FP+FN+TN) (TP+TN)/(TP+FP-FN+TN)	01

	(TP+TN)/(TP-FP+FN-TN)	
3.	Sensitivity **[LJIET]**	01
	$$\frac{TP}{TP + FP}$$ $$\frac{TP}{TP + FN}$$ $$\frac{FP}{TP + FP}$$ $$\frac{FP}{TP + FN}$$ $$\frac{TP + TN}{TP}$$ $$\frac{TN}{TN + FP}$$	
4.	Specificity **[LJIET]**	01
	$$\frac{TP}{TP + FP}$$ $$\frac{TP}{TP + FN}$$ $$\frac{FP}{TP + FP}$$ $$\frac{FP}{TP + FN}$$ $$\frac{TP + TN}{TP}$$ $$\frac{TN}{TN + FP}$$	
5.	Precision **[LJIET]**	01
	$$\frac{TP}{TP + FP}$$ $$\frac{TP}{TP + FN}$$ $$\frac{FP}{TP + FP}$$ $$\frac{FP}{TP + FN}$$ $$\frac{TP + TN}{TP}$$ $$\frac{TN}{TN + FP}$$	
6.	_____ is a part of model preparation activity. **[LJIET]** Remediate data	01

	Cross validation	
	Confusion matrix	
	ROC curve	
	Hold-out	
	Ensemble	
7.	Identify false statement(s) 1. Mean is impacted if too many data elements are having value closer to the far end of the range. 2. Skewness/Shape of histogram depends on nature of data. 3. Height of bin keeps on decreasing as we move toward right, is called right-skew. 4. Bivariate relationship can be visualized using scatter-plot and box-plot. 5. Cross-tab allows operations like roll-up and drill-down **[LJIET]** 4 only 4,5 only None of given 1,2,3 only 1,4 only All of given	01
8.	What is not correct about Eager Learners from following 1. Follows the steps abstraction and generalization. 2. Need to refer back to training data. 3. Support Vector Machine 4. Very little time in training **[LJIET]** 2,4 only 1,2 only 3,4 only 1,3,4 only All of given None of give	01
9.	1. Arises from simplifying assumptions made by model to make target function less complex 2. Occurs from difference in training data set used to train the model 3. Due to under-fitting a model 4. Due to over-fitting a model 5. Poor performance on complex dataset Small change in training data set magnified in model **[LJIET]** Error due to bias- 1,3,5 Error due to Variance- 2,4,6 Error due to bias- 1,4,6 Error due to Variance- 2,3,5 Error due to bias- 2,4,6 Error due to Variance- 1,3,5 Error due to bias- 2,3,5 Error due to Variance- 1,4,6 Error due to bias- 1,4,5 Error due to Variance- 2,3,6 Error due to bias- 2,3,6 Error due to Variance- 1,4,5	01
10	In Kappa coefficient,p(pr) chance agreement is calculated as **[LJIET]**	01

$$\frac{TP + FP}{TP + FP + FN + TN} * \frac{TP + FN}{TP + FP + FN + TN} + \frac{TN + FP}{TP + FP + FN + TN} * \frac{TN + FN}{TP + FP + FN + TN}$$

$$\frac{TP + FP}{TP + FP + FN + TN} * \frac{TP + FN}{TP + FP + FN + TN} * \frac{TN + FP}{TP + FP + FN + TN} * \frac{TN + FN}{TP + FP + FN + TN}$$

$$\frac{TP + FP}{TP + FP + FN + TN} + \frac{TP + FN}{TP + FP + FN + TN} * \frac{TN + FP}{TP + FP + FN + TN} + \frac{TN + FN}{TP + FP + FN + TN}$$

$$\frac{TP + FP}{TP + FP + FN + TN} - \frac{TP + FN}{TP + FP + FN + TN} * \frac{TN + FP}{TP + FP + FN + TN} + \frac{TN + FN}{TP + FP + FN + TN}$$

$$\frac{TP + FP}{TP + FP + FN + TN} * \frac{TP + FN}{TP + FP + FN + TN} + \frac{TN + FP}{TP + FP + FN + TN} - \frac{TN + FN}{TP + FP + FN + TN}$$

$$\frac{TP + FP}{TP + FP + FN + TN} * \frac{TP + FN}{TP + FP + FN + TN} - \frac{TN + FP}{TP + FP + FN + TN} * \frac{TN + FN}{TP + FP + FN + TN}$$

11	a. R-squared Error b. Silhouette Width c. Pearson's Correlation Coefficient d. Cosine Similarity Value lies between 1. 0 to +1 2. -1 to +1 **[LJIET]** a,c-1 b,d-2 a-1 b,c,d-2 a-2 b,c,d-1 a,b-1 c,d-2 a,c-2 b,d-1 a,b-2 c,d-1	01

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	Can the performance of a learning model be improved? If yes, How? **[LJIET]**	07
2.	Explain how can we evaluate the performance of a Classification algorithm by measuring Specificity, precision, recall and ROC curves **[LJIET]**	07
3.	Differentiate between a) Lazy Vs Eager Learners b) Bagging Vs Boosting **[LJIET]**	05

UNIT NO- 4 :
Basics of Feature Engineering

TOPIC:1 : Feature Construction, Feature Extraction, Feature Selection

Sr. No	SHORT QUESTIONS/MCQ	Marks
1.	Meaningful attribute of data set_____. **[LJIET]** Feature Redundant feature Relevant feature Input variable Output variable Variable	01
2.	1. Feature construction	01

	2. Feature extraction 3. Feature subset selection a. Discovering missing information about relationship of features b. Creating new set of features from original features using functional mapping c. Deriving most meaningful features from given feature space. **[LJIET]** 1-a, 2-b, 3-c 1-c, 2-a, 3-b 1-a, 2-c, 3-b 1-b, 2-a, 3-c 1-c, 2-b, 3-a 1-b, 2-c, 3-a	
3.	Text data converted into numerical representation using process called as _____. In this process occurrences of a word belonging to _____ are consolidated in form of _____. **[LJIET]** Vectorization, corpus, bag-of-word Vectorization, bag-of-word, corpus Corpus, bag-of-word, vector Corpus, tokens, normalization Vectorization, token, bag-of-word Corpus, vector, token	01
4.	1. New set of features are extracted from original set of features which are similar in nature 2. n-dimensional feature space is transformed to m-dimensional space. Where n less than m. 3. New features transformed are orthogonal to each other 4. Sum of variance of new feature is equal to sum of variance of old feature In context of PCA which of following is false **[LJIET]** 1,2 1,2,4 4 All of given None of given 2,3	01
5.	If P1,P2 and P3 are three principle components What are true among following: 1. P1, P2 and P3 are parallel 2. Variance of P1 is largest 3. Variance of P3 is smallest 4. P1, P2 and P3 are orthogonal 5. Variance of P1 is smallest 6. Variance of P3 is largest **[LJIET]** 2,3,4 1,2,3 4,5,6 1,5,6 1,2 only 4,5 only	01

6.	a. PCA b. SVD c. LDA 1. Matrix factorization technique 2. Identify dimensions that are independent of each other 3. Eigen values and eigen vectors are used 4. Singular, left singular and right singular matrix is produced 5. Captures variability in data set 6. Inter-class and intra-class matrices are used **[LJIET]** a-2,3 b-1,4 c-5,6 a-1,2 b-5,6 c-3,4 a-3,4 b-5,6 c-1,2 a-4,5 b-3,6 c-1,2 a-2,4 b-1,5 c-3,6 a-1,3 b-2,5 c-4,6	01
7.	a. Average information in a variable is _____. b. Indicator of feature relevance _____. c. Features are ranked in descending order of _____. **[LJIET]** Entropy, Mutual Information, Information Gain Mutual Information, entropy, Information gain Information gain, entropy, Mutual Information Entropy, Information Gain, Mutual Information Mutual Information, Information gain, entropy Information gain, Mutual Information, entropy	01
8.	Find the value of the person's correlation coefficient from the following data: F1= (43,21,25,42,57,59) F2= (99,65,79,75,87,81) **[LJIET]** 0.5298 0.9852 0.2599 0.9985 -0.5298 -0.9852	01
9.	a. $\dfrac{n_{11}+n_{00}}{n_{01}+n_{10}+n_{11}+n_{00}}$ b. Widely used in text classification c. $\dfrac{n_{11}}{n_{01}+n_{10}+n_{11}}$ d. $\sqrt[r]{\sum_{i=1}^{n}(F1_i - F2_i)^r}$ **[LJIET]** SMC-a, Cosine-b, Jaccard-c, Minkowski-d SMC-b, Cosine-a, Jaccard-c, Minkowski-d SMC-b, Cosine-c, Jaccard-a, Minkowski-d SMC-a, Cosine-d, Jaccard-c, Minkowski-b SMC-c, Cosine-b, Jaccard-a, Minkowski-d SMC-c, Cosine-d, Jaccard-a, Minkowski-b	01
10	a. In _____ feature space is expanded. b. _____ and _____ are two variants of _____. **[LJIET]** a-feature construction b-feature construction, feature extraction, feature transformation	01

a-feature transformation b-feature construction, feature extraction, feature subset selection

a-feature transformation b-feature construction, feature extraction, feature transformation

a-feature construction b-feature selection, feature extraction, feature transformation

a-feature construction b-feature construction, feature selection, feature transformation

a-feature construction b-feature construction, feature extraction, feature selection

11	a. Text specific feature construction b. Inter-class and intra-class scatter matrix c. Eigen values and Eigen-vectors d. Singular value matrix 1- LDA 2- PCA 3- SVD 4- Vectorization **[LJIET]** ==a-4, b-1, c-2, d-3== a-4, b-2, c-3, d-1 a-4, b-2, c-1, d-3 a-4, b-1, c-3, d-2 a-4, b-3, c-2, d-1 a-4, b-3, c-1, d-2	01
12	a. $\sum_{i=1}^{c} s_i$ b. $\sum_{x \in D_i}^{n} (x - m_i)(x - m_i)^T$ c. $\sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T$ from above given equations S_W, S_i and S_B relatively are? **[LJIET]** ==a, b, c== a, c, b b, a, c b, c, a c, a, b c, b, a	01
13	Euclidean distance and Manhattan distance respectively for given data is? F1= (2,3,6,7,8,6,6,7) F2= (6,5.5,4,2.5,3,5.5,7,6) **[LJIET]** ==8.6, 20.5== 20.5, 8.6 8.8, 8.6 20.5, 8.8 7.2, 5.5 5.5, 8.6	01
14	_____ used more frequently to calculate the distance between binary vectors is Hamming distance. **[LJIET]** ==Manhattan== Minkowski	01

	Pearson	
	Euclidean	
	Jaccard	
	Cosine	
15	a. Filter approach b. Wrapper approach c. Hybrid approach d. Embedded approach 1. Induction algorithm 2. Statistical tests 3. Induction algorithm and statistical tests Simultaneous selection and classification **[LJIET]** a-2 b-1 c-3 d-4 a-1 b-2 c-3 d-4 a-1 b-2 c-4 d-3 a-2 b-1 c-4 d-3 a-1 b-3 c-2 d-4 a-2 b-3 c-1 d-4	01
16	What is true about feature selection process 1. No new feature is generated 2. Improving efficiency of a learning model 3. Better understanding of underlying model 4. Faster and cost effective model Used functional mapping **[LJIET]** 1,2,3,4 only 1,3,4,5 only 1,2,5 only All of given None of given 1,2 only	01
17	What is true if groups are well separated in LDA 1. Intra-group mean are far away from each other 2. Data points are close to intra-group mean 3. Intra-group mean from grand-mean is large. **[LJIET]** All of given 1,2 only 2,3 only 1,3 only None of given 1 only	01
18	What features are ALWAYS candidate for rejection in the process of feature subset selection a. Redundant features b. Irrelevant features c. Weakly relevant features d. Non-redundant features**[LJIET]** a,b only a,b,c only all of given	01

	c only	
	b,c only	
	b,c,d only	
19	Which is not a search strategy**[LJIET]**	01
	Sequential forward selection	
	Sequential backward selection	
	Bi-directional selection	
	Bi-directional elimination	
	Recursive elimination	
	Recursive Selection	

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	What is feature selection? Why is it needed? What are the different approaches for feature selection? **[LJIET]**	07
2.	Explain the process of feature engineering in context of a text categorization problem. **[LJIET]**	05
3.	Differentiate between SMC and Jaccard Coefficients. **[LJIET]**	05
4.	Explain in short the three methods for Feature Extraction. **[LJIET]**	07
5.	State and explain the methods to find out the similarity or redundancy aspect of the attributes in a dataset. **[LJIET]**	07
6.	What is a feature? What is feature Engineering? What are the major elements of Feature Engineering? Explain them.**[LJIET]**	05

UNIT NO- 5 :
Brief overview of Probability

TOPIC:1 : Basic concept of Probability, Random Variables, Discrete Distributions, Continuous Distribution, Central Theorem, Monte Carlo Approximations.

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	Define the Bayesian interpretation of probability. **[LJIET]**	05
2.	Define probability of a union of two events with equation. **[LJIET]**	05
3.	What is joint probability? Explain it using appropriate equations. **[LJIET]**	05
4.	What is the chain rule of probability? **[LJIET]**	05
5.	What do you mean by Conditional Probability **[LJIET]**	05
6.	Define Covariance and Correlation **[LJIET]**	05
7.	What is hypothesis? Give examples **[LJIET]**	05

UNIT NO- 6 :
Bayesian Concept Learning

TOPIC:1 : Impotence of Bayesian methods, Bayesian theorem, Bayes' theorem and concept learning, Bayesian Belief Network

Sr. No	SHORT QUESTIONS/MCQ	Marks
1.	In a bolt factory, machines A1, A2 and A3 manufactures respectively 25%, 35% and 40% of the total machines. Of these 5%, 4% and 2% are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is probability that it was manufacture by machine A2? **[LJIET]**	01

Sr. No	DESCRIPTIVE QUESTIONS	Marks

1.	Explain following with example:
	   a. Prior probability
	   b. Posterior probability
	   c. Likelihood **[LJIET]**
		07

2.	Explain Naïve Bayes classifier with example and its use in practical life. **[LJIET]**	07
3.	What are Bayesian Belief Networks? Where are they used? Can they solve all types of problem?**[LJIET]**	07
4.	How Bayes theorem support the concept learning principle? **[LJIET]**	07
5.	In an airport security checking system, the passenger are checked to find out any intruder. Let I indicate whether somebody is an intruder (I) or not (Ɨ) and A indicate alarm whether raised (A=1) or not (A̶). Alarm will be raised if an intruder is identified with probability $P(A|I)=0.98$. Alarm will be raised if non-intruder with probability $P(A|Ɨ)=0.001$, which implies error factor. The probability of intruder in population of passenger is $P(I) = 0.00001$. What is probability that alarm is raised when a person actually is an intruder? **[LJIET]**	07
6.	For preparation of the exam questions are of types A, B or C. Probability of A, B or C appearing in the exam are 30%, 20% and 50% respectively. During preparation student solved 9 of 10 problems type A, 2 of 10 problems type B, and 6 of 10 problems of type C. What is probability that the student will solve the problem will be of type A? **[LJIET]**	07
7.	A tumor can be malignant or non-malignant. Only 0.5% of population has malignant tumor. Laboratory report detects malignancy present with 98% accuracy. Also laboratory report shows malignancy not present correctly in 97% of cases. If a Lab report is showing malignancy is present, should we declare this as malignant tumor or not. **[LJIET]**	07
8.	What is optimal Bayes classifier? Explain.**[LJIET]**	07

UNIT NO- 7 :
Supervised Learning: Classification and Regression:

TOPIC:1 : Supervised Learning, Classification Model, Learning steps, Classification algorithms

Sr. No	DESCRIPTIVE QUESTIONS	Marks
1.	What is supervised learning? Why it is called so? Give three example of supervised learning.**[LJIET]**	07
2.	Explain in brief, the SVM model. Give its advantage and disadvantage.**[LJIET]**	07
3.	Write note on: Validation error in kNN, error rate in kNN choosing value of k in kNN, inductive bias in decision tree.**[LJIET]**	07
4.	Explain in brief: Decision tree algorithm.**[LJIET]**	07
5.	Explain in brief: Random forest model.**[LJIET]**	07
6.	Explain following w.r.t decision trees: node and leaf, entropy, information gain**[LJIET]**	07
7.	Discuss SVM model in detail with different scenarios. **[LJIET]**	07

8.	Discuss how to calculate distance between the test data and training data in kNN.**[LJIET]**	07
9.	Discuss OOB error and variable importance in random forest.**[LJIET]**	07
10	How to avoid over fitting decision tree? Explain in detail.**[LJIET]**	07

TOPIC:2 : Regression, Regression algorithms		
Sr. No	**DESCRIPTIVE QUESTIONS**	**Marks**
1.	What is simple linear regression? What is dependent and independent variable in linear regression?**[LJIET]**	07
2.	What are condition of a positive slope and conditions of negative slope in linear regression?**[LJIET]**	07
3.	What is multiple linear regression? Give example.**[LJIET]**	07
4.	What is multicollinearity in regression equation? Explain.**[LJIET]**	07
5.	Explain rise, run, and slope in graph. Find slope if points on line are (-3,-2) and (2,2) respectively.**[LJIET]**	07
6.	How to improve accuracy of linear regression model?**[LJIET]**	07
7.	Explain OLS algorithm with steps.**[LJIET]**	07
8.	Define sum of squares due to error in multiple linear regression.**[LJIET]**	07

UNIT NO- 8 :
Unsupervised Learning

TOPIC:1 : Supervised vs. Unsupervised Learning, Applications, Association rules		
Sr. No	**DESCRIPTIVE QUESTIONS**	**Marks**
1.	How unsupervised learning is different from supervised leaning? Explain with example.**[LJIET]**	07
2.	Mention few application areas of unsupervised learning.**[LJIET]**	07
3.	How apriori algorithm helps in reducing the calculation overhead for market basket analysis? Give example.**[LJIET]**	07
4.	What is Association rule? What are the Applications of Association rule mining? Define support and confidence in Association rule mining.**[LJIET]**	07
5.	A database has 4 transactions, shown below. <table><tr><th>TID</th><th>Date</th><th>items_bought</th></tr><tr><td>T100</td><td>10/15/04</td><td>{K, A, D, B}</td></tr><tr><td>T200</td><td>10/15/04</td><td>{D, A, C, E, B}</td></tr><tr><td>T300</td><td>10/19/04</td><td>{C, A, B, E}</td></tr><tr><td>T400</td><td>10/22/04</td><td>{B, A, D}</td></tr></table> Assuming a minimum level of support min_sup = 60% and a minimum level of confidence min_conf = 80%. a. Find all frequent item-sets using the Apriori algorithm. b. List all of the strong association rules, along with their support and confidence values for buys(item1, item2) $\Rightarrow$ buys(item3) items can be A, B etc.**[LJIET]**	07

6.	Explain how Market Basket Analysis uses the concept of association analysis.**[LJIET]**	07
7.	Explain apriori algorithm for association rule learning with an example.**[LJIET]**	07

<table>
<tr><td colspan="3" align="center">TOPIC:2 : Clustering</td></tr>
<tr><td>Sr. No</td><td align="center">SHORT QUESTIONS/MCQ</td><td>Marks</td></tr>
<tr><td>1.</td><td>In which situation k-meas clustering fails to give good results?[LJIET]</td><td>01</td></tr>
<tr><td>Sr. No</td><td align="center">DESCRIPTIVE QUESTIONS</td><td>Marks</td></tr>
<tr><td>1.</td><td>Describe how quality of clusters are measured in k-means algorithm?[LJIET]</td><td>07</td></tr>
<tr><td>2.</td><td>What are three broad categories of clustering technique? Explain characteristics of each briefly.[LJIET]</td><td>07</td></tr>
<tr><td>3.</td><td>Describe main difference in approach of k-means and k-medoids algorithm with a neat diagram.[LJIET]</td><td>07</td></tr>
<tr><td>4.</td><td>Explain k-means method with step-by-step algorithm.[LJIET]</td><td>07</td></tr>
<tr><td>5.</td><td>Given a set of one-dimensional data point: {5,10,15,20,25,30,35}. Assume that k=2 and first set of random centroid is selected as {15,32} and then it is refined with {12,30}
     a. Create two clusters with each set of centroid mentioned above following the k-means approach.
     b. Calculate SSE for each set of centroid.[LJIET]</td><td>07</td></tr>
<tr><td>6.</td><td>How to re-compute cluster centroids in k-means algorithm?[LJIET]</td><td>07</td></tr>
<tr><td>7.</td><td>Explain concept of clustering with neat diagram.[LJIET]</td><td>07</td></tr>
</table>

<table>
<tr><td colspan="3" align="center">UNIT NO- 9 :
Neural Network:</td></tr>
<tr><td colspan="3" align="center">TOPIC:1 : Introduction to neural network, Biological and Artificial Neurons, Types of Activation functions, Implementation of ANN, Architecture, Leaning process, Backpropogation, Deep Learning</td></tr>
<tr><td>Sr. No</td><td align="center">DESCRIPTIVE QUESTIONS</td><td>Marks</td></tr>
<tr><td>1.</td><td>What is function of summation junction of a neuron? What is threshold activation function?[LJIET]</td><td>07</td></tr>
<tr><td>2.</td><td>What is step function? What is the difference of step function with threshold function?[LJIET]</td><td>07</td></tr>
<tr><td>3.</td><td>Explain McCulloch-Pitts model of neuron.[LJIET]</td><td>07</td></tr>
<tr><td>4.</td><td>Explain the basic structure of multi-layer perceptron. Explain how it can solve XOR problem.[LJIET]</td><td>07</td></tr>
<tr><td>5.</td><td>What is Artificial Neural Network (ANN)? Explain some of salient highlights in different architectural option for ANN.[LJIET]</td><td>07</td></tr>
<tr><td>6.</td><td>Explain learning process of ANN. With example explain challenges in assigning synaptic weights for interconnection of neuron.[LJIET]</td><td>07</td></tr>
<tr><td>7.</td><td>Explain in details backpropagation algorithm. What are the steps? What are the limitations?[LJIET]</td><td>07</td></tr>
<tr><td>8.</td><td>Write short note on: Single-layer feed forward ANN, Recurrent network, Gradient descendent. [LJIET]</td><td>07</td></tr>
</table>

| 9. | Write Short note on: Deep learning, Multi-layer perceptron, Artificial neuron.**[LJIET]** | 07 |