# GURU NANAK INSTITUTE OF TECHNOLOGY
## An Autonomous Institute under MAKAUT
## 2022
## MACHINE LEARNING
## IT802I

**TIME ALLOTTED: 3 HOURS** **FULL MARKS: 70**

*The figures in the margin indicate full marks.*
*Candidates are required to give their answers in their own words as far as practicable*

### GROUP – A
### (Multiple Choice Type Questions)

Answer any ***ten*** from the following, choosing the correct alternative of each question: **10×1=10**

| | | | Marks | CO No |
|---|---|---|---|---|
| 1. | (i) | Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?<br>a) AUC-ROC<br>b) Confusion Matrix<br>c) Accuracy<br>d) Mean-Squared-Error | 1 | CO1 |
| | (ii) | If the CT scan of a person shows that his tumour is benign, but it is actually malignatic. What type of error is this?<br>a) Type 1 Error<br>b) Type 2 Error<br>c) Type 3 Error<br>d) Type 4 Error | 1 | CO1 |
| | (iii) | Which of the following algorithms is most sensitive to outliers?<br>a) K-means clustering algorithm<br>b) K-medians clustering algorithm<br>c) K-modes clustering algorithm<br>d) K-medoids clustering algorithm | 1 | CO2 |
| | (iv) | Which of the following statements is TRUE?<br>a) Outliers should be identified and removed from a dataset<br>b) Outliers can never be present in the testing dataset<br>c) Outliers is a data point that is significantly close to other data points<br>d) The nature of our business problem determines how outliers are used. | 1 | CO2 |
| | (v) | Some telecommunication company wants to segment their customers into distinct groups in order to send appropriate subscription offers, this is an example of -<br>a) Supervised learning<br>b) Reinforcement learning<br>c) Unsupervised learning<br>d) Semi Supervised learning | 1 | CO1 |

| | | | | |
|---|---|---|---|---|
| (vi) | If we increase the k value in k-nearest neighbor, the model will _____ the bias and _____ the variance<br>a) Decrease, Decrease<br>b) Decrease, Increase<br>c) Increase, Decrease<br>d) Increase, Increase | | 1 | CO3 |
| (vii) | How do you choose the right node while constructing a decision tree?<br>a) An attribute having highest gini index<br>b) An attribute having the highest information gain<br>c) An attribute having the high entropy<br>d) None of this | | 1 | CO3 |
| (viii) | Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?<br>a) Decision Tree<br>b) Regression<br>c) Classification<br>d) Random Forest | | 1 | CO2 |
| (ix) | Which of the following algorithms do we use for Variable Selection?<br>a) Lasso<br>b) Ridge<br>c) Elastic Net<br>d) None of this | | 1 | CO4 |
| (x) | The most widely used metrics and tools to assess a classification model are<br>a) Confusion Matrix<br>b) Area under the ROC curve<br>c) Cost-sensitive accuracy<br>d) All of the above | | 1 | CO1 |
| (xi) | Which of the following model models include a backwards elimination feature election routine?<br>a) MCV<br>b) MCRS<br>c) MARS<br>d) All of the mentioned | | 1 | CO4 |
| (xii) | Which among the below methods encompass the benefits of both the wrapper and filter methods -<br>a) hybrid<br>b) Embedded<br>c) Wrapper<br>d) Recursive Method | | 1 | CO4 |

## GROUP – B
## (Short Answer Type Questions)
## (Answer any three of the following) 3 x 5 = 15

| | | | Marks | CO No |
|---|---|---|---|---|
| 2. | (a) | What is the difference between supervised learning and unsupervised learning? | 2 | CO1 |
| | (b) | What are the different assumptions made for logistic regression? | 2 | CO1 |

| | | | | |
|---|---|---|---|---|
| | (c) | What is bias- variance tradeoff? | 1 | CO1 |
| 3. | (a) | Why is Random Forest considered better than decision tree? | 2 | CO2 |
| | (b) | What is OLS regression? | 2 | CO2 |
| | (c) | What is the difference between R-square and Adjusted R-square value? | 1 | CO2 |
| 4. | (a) | When should you use classification over regression? | 2 | CO3 |
| | (b) | What is the need of regularization in the machine learning world? | 3 | CO3 |
| 5. | (a) | What is the difference between correlation and covariance? | 3 | CO4 |
| | (b) | What is the difference between grid search and randomized search? | 1 | CO4 |
| | (c) | What is regression? | 1 | CO4 |
| 6. | (a) | What is feature Selection? | 3 | CO4 |
| | (b) | What are classifications? | 2 | CO4 |

## GROUP – C
## (Long Answer Type Questions)
## (Answer any three of the following) 3 x 15 = 45

| | | | Marks | CO No |
|---|---|---|---|---|
| 7. | (a) | What is Bayes' Theorem? How is it useful in a machine learning context? | 2 | CO1 |
| | (b) | Why is Naive Bayes, "Naive" in nature? | 6 | CO1 |
| | (c) | What is the need of probability distributions? What are the different probability distributions present? | 7 | CO1 |
| 8. | (a) | Explain one probability distribution in detail, with a diagram. | 5 | CO2 |
| | (b) | What is bagging? Give some examples of bagging. | 5 | CO2 |
| | (c) | Explain one bagging algorithm with a proper diagram. | 5 | CO2 |
| 9. | (a) | What is boosting? Give some examples of Boosting. | 6 | CO3 |
| | (b) | Explain one boosting algorithm in detail with a proper diagram. | 4 | CO3 |
| | (c) | What is feature engineering? Why do we need feature engineering in machine learning? | 5 | CO3 |
| 10. | (a) | What are the different feature engineering techniques available? Explain any 5 of them with a proper example. | 4 | CO4 |
| | (b) | What is regression? explain with a real-world example. What are the different types of regression techniques available? | 4 | CO4 |
| | (c) | Explain one regression algorithm with details. | 7 | CO4 |
| 11. | | What is classification? explain with a real-world example. What are the different types of classification techniques available? | 15 | CO3 |