GOA UNIVERSITY

T.Y.B.C.A. (NEP) (Semester-VI)
Subject Name: Machine Learning (Theory) (Major)

Paper Title: Machine Learning **Paper Code: CSA-306**

Duration: 2:00 Hours **Total Marks: 60**

Instructions:
i) Use of non-programmable calculators is allowed.
ii) All questions are compulsory and figures to the right indicate full marks
iii) Questions of 2 marks to be answered in not more than 40 words
iv) Questions of 3 marks to be answered in not more than 60 words
v) Questions of 4 marks to be answered in not more than 80 words
vi) Questions of 5 marks to be answered in not more than 100 words

---

Q.1. **Name the following.** **(6x2=12marks)**

(i) Name **any two** types of Machine Learning approaches.
(ii) Define 'Regression' with an example.
(iii) How is Information different from Data?
(iv) What are outliers? Give an example.
(v) Name **any two** clustering algorithms.
(vi) What is Association Rule mining? Mention the name of the algorithm that performs this task.

Q.2 A. (i) A decision tree is used to determine whether a loan application should be Approved or Rejected. The tree begins with "Credit Score". The next split is based on "Income Level". The final outcomes are "Approved" or "Rejected". Identify the decision node(s), root node, and leaf node(s). **(3 Marks)**

(ii) Differentiate between traditional programming and Machine Learning giving **two points** each. **(2 Marks)**

OR

Q.2 A. (iii) Explain Simple Linear Regression with an example of your choice. **(3 Marks)**

(iv) State your understanding as to why Random Forest algorithm works better than a single tree. **(2 Marks)**

X8Y66AD38X8Y66AD38X8Y66AD38X8Y66AD38

Q.2 B. (i) Explain Dimensionality Reduction. Compare Dimensionality Reduction with Feature Selection giving **three** points each. **(5 Marks)**

(ii) Write a short note on Naïve Bayes Classifier Algorithm. **(2 Marks)**

Q.3 A. (i) A company wants to improve the performance of a predictive model by combining several weak models that are trained one after another. **(4 Marks)**

Explain the working of Boosting technique to improve prediction accuracy in this case.

(ii) Identify the concept of statistics given the following analogy, "You taste 2 spoons of curry to decide how the entire pot tastes." **(1 Mark)**

OR

Q.3 A. (iii) A doctor wants to predict whether a patient has a disease by comparing the patient's symptoms with records of previous patients using the K-Nearest Neighbours Algorithm. **(4 Marks)**

Identify **any two** advantages and **any two** disadvantages of using the K-Nearest Neighbours Algorithm.

(iv) Compare and analyse Decision trees with Random Forest giving **one point** each. **(1 Mark)**

Q.3 B. (i) Given the dataset, A = {3, 8, 7, 15, 18, 5, 9, 21, 24} **(5 Marks)**

Let the number of clusters K = 2. Find the final clusters using K-Means clustering algorithm. Consider initial centroids c1 = 3 and c2 = 8.

(ii) You are required to solve your client's problem of Predicting loan defaults for a bank. You need to collect data for the analysis. Name any two possible sources of data for the data collection. **(2 Marks)**

Q.4 A. (i) Consider the following dataset containing two features (X, Y) for 8 data points. Suppose, K = 2 clusters, form the clusters using the K-Medoids clustering algorithm. Consider the following two points as medoids from the data. **(4 Marks)**

Medoid 1 (m1): Point 3 (4, 7)

Medoid 2 (m2): Point 6 (8, 7)

Point	X	Y
1	2	5
2	3	6
3	4	7
4	5	4
5	7	8
6	8	7
7	9	6

X8Y66AD38X8Y66AD38X8Y66AD38X8Y66AD38

(ii) Provide **any one** problem area in physical fitness domain where Data Science/Machine Learning can help with analysis. **(1 Mark)**

OR

Q.4 A. (iii) Assume you are working on predicting house prices. Explain the concept of 'Feature Selection' using the above scenario. **(4 Marks)**

(iv) Explain in one sentence why is an Artificial Neural Network considered to be inspired by the human brain. **(1 Mark)**

Q.4 B. (i) The number of customers visiting a store in 10 days were: **(5 Marks)**
15, 18, 20, 17, 19, 15, 20, 21, 20, 14.
Find:
(a) Mean
(b) Median
(c) Mode
(d) Range
(e) Variance

(ii) A rule generated from market basket analysis is: **(2 Marks)**
{Chips} → {Soft Drink}
Evaluate this rule and write what it indicates about customer buying patterns, and how could the store use this information to increase sales?

Q.5 A. (i) Compare and contrast supervised learning with unsupervised learning and elaborate on **any four** points. **(4 Marks)**

(ii) Compare Analysis vs Analytics giving any one point each. **(2 Marks)**

OR

Q.5 A. (iii) Compare and analyse the following sampling techniques: **(4 Marks)**
Stratified Sampling and Cluster Sampling with suitable examples.

(iv) Analyse and write **any two** advantages of Gradient Boosting Machines. **(2 Marks)**

Q.5 B. (i) Outline the working of Linear Discriminant Analysis algorithm with a neat and labelled diagram. **(3 Marks)**

(ii) For each of the following dependent variables, state the type of logistic regression that should be used: **(3 Marks)**
(a) "Yes", "No"
(b) "Red", "Blue", "Green", "Yellow"
(c) "Low", "Moderate", "High", "Very High"

X8Y66AD38X8Y66AD38X8Y66AD38X8Y66AD38