



UNIVERSITY EXAMINATIONS

**FOURTH YEAR EXAMINATION FOR THE AWARD OF THE DEGREE OF
BACHELOR OF SCIENCE IN APPLIED COMPUTER/COMPUTER SCIENCE**

**FIRST SEMESTER 2022/2023
[SEPTEMBER-DECEMBER, 2022]**

ACMP 496/COMP 460: MACHINE LEARNING ALGORITHMS

STREAM: Y4S1

TIME: 2 HOURS

DAY: TUESDAY, 12:00 – 2:00 PM

DATE: 06/12/2022

INSTRUCTIONS

- 1. Do not write anything on this question paper.**
- 2. Answer question ONE and any other TWO questions.**

QUESTION ONE (30 MARKS)

- How can we use your machine learning skills to generate revenue? [5marks]
- What evaluation approaches would you work to gauge the effectiveness of a machine learning model? [10marks]
- How would you handle an imbalanced dataset? [5marks]
- Write a basic Machine Learning program to check the accuracy of a model, by importing any dataset using any classifier? [10marks]

QUESTION TWO (20 MARKS)

- Explain the circumstances under which you can use Classification over Regression? [6marks]
- Explain the K Nearest Neighbor Algorithm with an aid of a diagram [6marks]
- What are the two types of problems solved by Supervised Learning? [4marks]
- What's the difference between Covariance and Correlation? [4marks]

QUESTION THREE (20 MARKS)

- What's the difference between Type I and Type II error? [5marks]
- You're asked to build a random forest model with 10000 trees. During its training, you got training error as 0.00. But, on testing the validation error was 34.23. What is going on? Haven't you trained your model perfectly? [7marks]
- Explain false negative, false positive, true negative, and true positive with a simple example. (8marks)

QUESTION FOUR (20 MARKS)

- a) You are working on a time series data set. Your manager has asked you to build a high accuracy model. You start with the decision tree algorithm since you know it works fairly well on all kinds of data. Later, you tried a time series regression model and got higher accuracy than the decision tree model. Can this happen? Why? [6marks]
- b) You are given a cancer detection data set. Let's suppose when you build a classification model you achieved an accuracy of 96%. Why shouldn't you be happy with your model performance? What can you do about it? [6marks]
- c) Write an SQL query that makes recommendations using the pages that your friends liked. Assume you have two tables: a two-column table of users and their friends, and a two-column table of users and the pages they liked. It should not recommend pages you already like. [8marks]

QUESTION FIVE (20 MARKS)

- a) Distinguish between bagging, boosting and stacking. [6marks]
- b) You are given a data set consisting of variables having more than 30% missing values? Let's say, out of 50 variables, 8 variables have missing values higher than 30%. How will you deal with them? [6marks]
- c) A jar has 1000 coins, of which 999 are fair and 1 is double headed. Pick a coin at random, and toss it 10 times. Given that you see 10 heads, what is the probability that the next toss of that coin is also a head? [8marks]