MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

University Examinations 2020/2021

FIRST YEAR FIRST SEMESTER EXAMINATIONS FOR MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

CCD 7104: MACHINE LEARNING

DATE: JULY 2021   **TIME: 3 HOURS**

INSTRUCTIONS: Answer Question ONE and any other Two questions.

QUESTION ONE (20 MARKS)

Assume we have a set of data from patients who have visited Meru Level 5 hospital since the year 2010. A set of features (e.g., temperature, height) have been also extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or Cancer (a patient can have one or more of these diseases).

a) Supposedly we have decided to use a neural network to solve this problem and we have two choices:- either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. You are consulted to provide your feedback on this dilemma. Which method would you advise for adoption and why? (Provide a detailed comparison in addition to your stance)   (10 Marks)

b) Some patient features are expensive to collect (e.g., brain scans) whereas others are not (e.g., temperature). Therefore, we have decided to first ask our classification algorithm to predict whether a patient has a disease, and if the classifier is 80% confident that the patient has a disease, then we will do additional examinations to collect additional patient features In this case. The team you lead has the option of using either of the following classifiers:- neural networks, decision tree, or naive

Bayes. You are required to provide a report comparing the three classifiers and finally recommend one to fit the problem with justification (10 Marks)
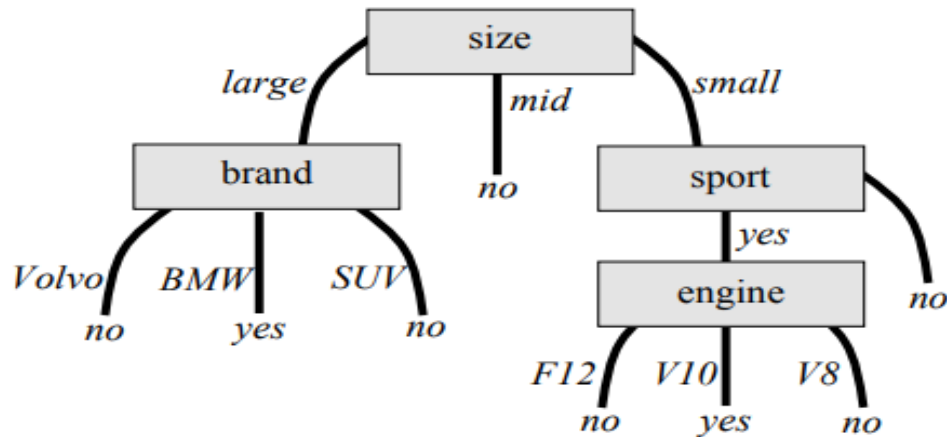
Question Two (20 Marks)

a) Compare and contrast between each of the following pairs of machine learning models;
   a. Generativevs discriminative (4 Marks)
   b. Parametric vs Non-parametric (4 Marks)

b) Classification is a predictive modeling problem that involves assigning a class label to each observation. Consequently, classification models generate a predicted class, which comes in the form of a discrete category; and thus for most practical applications, a discrete category prediction is required in order to make a decision. However, this is highly influenced by class imbalances of data
   a. Explain what you understand by the term *class imbalance* (2 Marks)
   b. Describe some consequences of using unbalanced classes in classification (4 Marks)
   c. Using appropriate examples, outline any two mechanisms of resolving the class imbalance problem (4 Marks)
   d. Cite four different application areas in which class imbalance is practical (2 Marks)

QUESTION THREE (20 MARKS)

a) Machine learning employs different strategies to enable a model to learn. Briefly explain each of the following strategies;
   a. Supervised Learning (2 Marks)
   b. Unsupervised Learning (2 Marks)
   c. Reinforcement Learning (1 Marks)

b) Machine learning is a branch of artificial intelligence that deals with the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. Explain three other domains that form part of machine learning (6 Marks)

c) The goal of machine learning to improve the performance P, on some task T over experience E. using these metrics (P, T, E). Using these metrics identify and justify any three application areas within which machine learning can be employed(9 Marks)
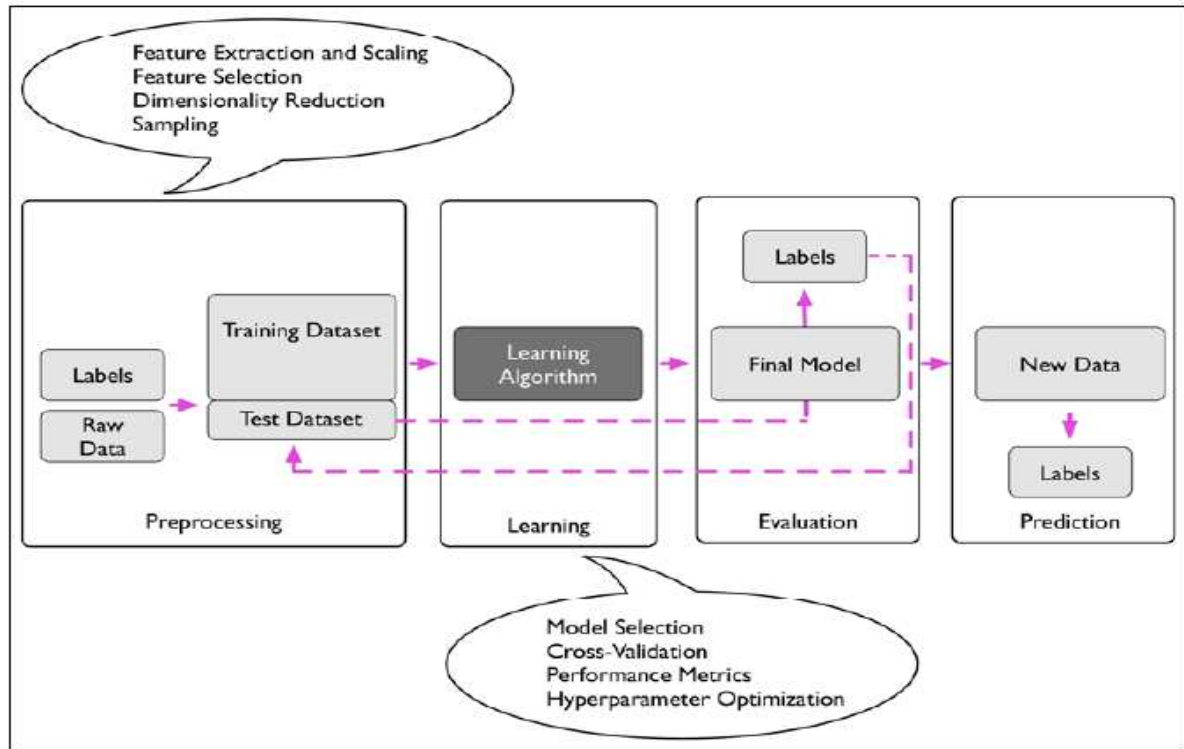
QUESTION FOUR (20 MARKS)

a) Suppose that we want to solve the problem of finding out what a good car is by using genetic algorithms. Suppose further that the solution to the problem can be represented by a decision tree as follows:

  i. Illustrate what would be the appropriate chromosome design for the given problem   (4 Marks)

  ii. Describe the Genetic Algorithm parameters that need to be defined to solve the problem   (6 Marks)

  iii. Solve the problem applying a single round of the prototypical Genetic Algorithm   (5 Marks)

  iv. Explain your answer in (iii) above by providing the pseudo code of the algorithm   (5 Marks)

QUESTION FIVE (20 MARKS)

a) The following diagram provides the necessary steps for solving a machine learning problem. Study it and answer the questions that follow;

a. Explain what you understand by the terms

    i.   Feature extraction and reduction   (3 Marks)

    ii.   Dimensionality reduction   (2 Marks)

    iii. Cross validation   (2 Marks)

    iv. Hyper-parameter Optimization   (2 marks)

b. Using an appropriate machine learning prediction problem, discuss the algorithms that you would utilize in solving the problem from the onset to prediction   (11 Marks)