COMP 428 MACHINE LEARNING CAT ONE

a) Suppose you are given the following data set in CSV format

Index	Country	Age	Salary	Purchased
0	India	38	68000	No
1	France	43	45000	Yes
2	Germany	30	54000	No
3	France	48	65000	No
4	Germany	40	nan	Yes
5	India	35	58000	Yes
6	Germany	nan	53000	No
7	France	49	79000	Yes
8	India	50	88000	No
9	France	37	77000	Yes

i) Write a python program to load the data set make assumption where necessary (3**Marks**)

ii) Write a program to extract dependent and independent variable (2**Marks**)

iii) Write a program to handle missing data (2**Marks**)

b) One of the application areas of naïve bayes algorithm is sentiment analysis, using an example explain the meaning of the term sentimental analysis. (4**Marks**)

c) Discuss the types of the machine learning techniques (6**Marks**)

a) You want to train a neural network to drive a car. Your training data consists of grayscale 64 ×64 pixel images. The training labels include the human driver's steering wheel angle in degrees and the human driver's speed in miles per hour. Your neural network consists of an input layer with 64 × 64 = 4,096 units, a hidden layer with 2,048 units, and an output layer with 2 units (one for steering angle, one for speed). You use the ReLU activation function for the hidden units and no activation function for the outputs (or inputs).

Calculate the number of parameters (weights) in this network. (6**Marks**)

d) A 4 - input neuron has weight 1 2 3 4 , The transfer function is linear with a constant of proportionality of 2 , the inputs are , 4, 10 , 5 and 20, Calculate the output. (3**Marks**)

e) Some patient features are expensive to collect (e.g., brain scans, heart etc) whereas others are not (e.g., temperature, age ,Bp) and want our classification algorithm to predict whether a patient has a given disease based on the less complicated features, if the classifier is 80% confident then the patient has that disease, afterwards we can do additional examinations to collect additional

patient features to improve the accuracy. In this case, which classification methods do you recommend for such a task , neural networks, decision tree, or naive Bayes? Justify your choice. (4**Marks**)

f) A recruitment agency faces a challenge of determining the salary of an employee after several years of experience, write a program in python that will help the agency achieve the following tasks based on a data set

i) Import the necessary libraries (2**Marks**)

ii) Load data training set (2**Marks**)

iii) Find out how many rows and columns are in the dataset (2**Marks**)

iv) Find out the statistical summary (2**Marks**)

v) Visualize the data set (2**Marks**)