BOWEN UNIVERSITY IWO, OSUN STATE

COLLEGE OF COMPUTING AND COMMUNICATION STUDIES
COMPUTER SCIENCE PROGRAMME
B. SC. DEGREE SECOND SEMESTER EXAMINATION   SESSION: 2022/2023
COURSE CODE: CIT 404   COURSE TITLE: MACHINE LEARNING WITH PYTHON
COURSE CREDITS: 3 DATE: FRIDAY 19th JUNE 2023   TIME ALLOWED: 2½ HOURS
INSTRUCTION: Answer question one(compulsory) and three other questions

1. a. i. Define machine learning.   **2marks**
      ii. Consider this list
          item = [43,23,1,67,54,2,34,56,23,65,9,87]
          Sort the list item Perform the following operation on list item
          item[:], item[0:4], item[6:10], item[2:7:3]   **10marks**
   b. Study the table below showing students usage of the social media platform

S/N	Fullname	Email	Gender	Age	Whatsapp	Facebook	Twitter	Ticktok	Instagram	Remark
1	Daniella Joe	danjoe@gmail.com	F	18	1	0	1	1	1	Over used
2	Gabriel H	gabh@gmail.com	M	19	1	1	0	1	0	Averagely Used
3	Esther O	esthero@gmail.com	F	17	1	0	0	0	0	Under used
4	Israel K	israelk@gmail.com	M	12	1	1	1	1	0	Over used
5	Ojo h	ojoh@gmail.com	F	40	1	1	0	0	0	Under used
6	Kemi O	kemo@gmail.com	F	14	1	1	1	1	1	Extremely used
7	David G	davidd@gmail.com	M	30	1	1	0	0	1	Averagely used

The remark column in the table is determined based on age of the students and the number of social media platform the students visited everyday.

   i.   Use the data in the table above to create a dummy dataset (ddf) for each of the
         column   **10marks**
   ii.   Load the ddf into dataframe   **8marks**
   iii. Export the dataframe data into csv format name as studsoc.csv   **3marks**
   iv.   Read the saved dataset into dataframe   **3marks**

2. a. i. Define Normalization   **2marks**
      ii. Discuss three normalization techniques   **6marks**
   b. i. List three different methods used for data collection   **6marks**
      ii. Itemize four data preparation process and discuss any two   **8marks**

3. a. i. Define Feature Selection, Feature Extraction   **4marks**
      ii. State three important of Feature Selection   **6marks**
   b. i. Mention three techniques use for feature selection and discuss any two
                                                                          6marks
      ii. Itemize three ways to handle missing data   **6marks**

4. i. Define DataFrame   **2marks**
    ii. Itemize with code example five different way to read dataset with pandas
                                                                                        10marks
    b. i. Define Lambda function   **2marks**
        ii. List three types of Lambda function and discuss any two of them with code
            example   **8marks**

5. a. i. Define categorical features   **2marks**
        ii. Discuss two methods used in handling categorical features   **4marks**
    b. i. Define data visualization and state two uses of data visualization   **6marks**
        ii. List and discuss three data visualization techniques and state their important
            where necessary   **10marks**

6. a. i. With the aid of diagram, define Regression   **5marks**
        ii. List three model assessment techniques and discuss any two of them
                                                                                        7marks
    b   i. With the aid of diagram, discuss Confusion Matrix   **6marks**
        ii. Explain two performance measures for evaluating model   **4marks**