Birla Institute of Technology and Science, Pilani, Hyderabad Campus

MACHINE LEARNING FOR ELECTRONICS ENGINEERS LAB

Experiment 3

Understanding Data Loading and Processing

Q1. For the data given below,

Student_Name	Roll_No	Digital_elec	Sig_Sys	ML	Dept
Amit	101	85	90	78	CSE
Priya	102	92	88	85	EEE
Rahul	103	78	85	80	EEE
Sneha	104	88	92	88	EEE
Vikram	105	95	89	82	CSE

Write a Python program to generate the above dataframe, and perform the following using the pandas library

a) After generating the above dataframe, convert the dataframe into a pandas dataframe

b) Print the shape of the dataframe

c) Print the column names (feature names) of the dataframe

d) Print the first two rows of the dataframe

e) Print the last two rows of the dataframe

f) Print the descriptive statistics of the dataframe

Q2. Download the .csv file (filename: "heart_failure_clinical_records_dataset.csv") from the LMS and perform the following

a) Mount the drive and upload the .csv file into the Google Colab directory.

b) Write a Python code to read the csv file using the pandas library

c) After reading the file, print the first two rows of the dataset.

d) Display the descriptive statistics of all the features of the dataset.

e) Import the matplotlib library and plot a bar chart (use plt.bar command). Consider the target variable, death event = 1 (dead), from the dataset. Plot a bar graph showing the number of males and females for this given event.

Note: In the data file, the Gender feature column uses '1' for male and '0' for female. A Death event of 1 indicates death occurred, and 0 indicates alive.

Q3. Download the image file (name: "Lenna.png") from the LMS for the given problem:

a) Mount and upload the image file into the Google Colab directory

b) Import the OpenCV library to read the image file

c) Convert the input image to RGB (by default, OpenCV loads images in BGR)

d) Convert the obtained RGB image to a grayscale image

e) Convert the above grayscale image to a binary image by considering a threshold value of 127.

f) Import the matplotlib library and use the subplot command to plot the converted RGB image, grayscale, and binary image in a single figure.

Q4. Some built-in(example) datasets are available in scikit learn library.

a) From this scikit learn library, import the "California Housing dataset".

Hint: You can use the following to import the data:

"from sklearn.datasets import fetch_california_dataset".

After importing, load the data.

b) Using the pandas library, read the features (predictors) and target variables (response) from the loaded data. Store the features in the variable X and the target in the variable Y.

c) Print the dimensions of the feature variable in the form m × n, where m is the number of samples and n is the number of features in the dataset.

d) Determine whether the house pricing estimation is a regression or classification task.

Q5. Learn to access, download, and load data from the data explorer of Google Colab, and perform the following

a) Use the data explorer of Google Colab and download the "Credit Card Fraud Detection" dataset.

b) Import the pandas library to read the features and target variables from the dataset. Save the features/predictor variables in the X variable and the target/response variable in the variable Y.

c) Identify whether this dataset is for a regression or classification task.

Hint: Learn the context of the data from the link provided in the data explorer while accessing this dataset.