School of Computer Science Engineering and Technology

Course- BCA   Type- Elective
Course Code- BCA355L   Course Name - Data Mining and Predictive Analysis
Year-   2022   Semester- Even
Date- 16-02-2022   Batch- NA

Lab Assignment 3

Objective: Working with missing values and Data Normalization techniques.

Part1: Perform the operation on user-created data (55)

1. Use the following data to create the Data Frame and store it in a variable named "Data" (5)

X	Y
103	18976
207	45789
315	NaN
400	78964
NaN	45683
605	NaN
355	78965
677	NaN
197	68546
869	20015

2. Check the presence of missing values (5)
3. Use the imputation approach to handle missing values (10)
4. Perform Normalization (30)
    a) Using min-max normalization with or without an inbuilt library
    b) Using Z score standardization with or without an inbuilt library
5. Stored the pre-processed data in the external file "preprocessed123.csv" (5)

Part2: Perform the following operation on Social Network Advertisement dataset (60)

1. Download the dataset from the following link (5)
    https://drive.google.com/drive/folders/1ScmcOb1UAR5LYUAZkUEAOeepKCOjDUto?usp=sharing
2. Check the presence of missing values (5)
3. Use the imputation approach to handle missing values (15)
4. Perform Normalization (30)
    a) Using min-max normalization with or without an inbuilt library
    b) Using Z score standardization with or without an inbuilt library
5. Stored the pre-processed data in the external file "preprocessed123.csv" (5)

Platform Required: Anaconda, Editor: Jupyter/Spyder/Pycharm/Google Colab

Submission Instructions:

*   Submission required 3 things 1) Python file (roll_no.ipynb/.py) 2) Dataset 3) PDF of .ipynb file. All these files are in a single zip folder (Use the naming convention: RollNo_LabNo.docx (Example: 123_Lab1))
*   Submission is through LMS only.