School of Computer Science Engineering and Technology

Course- BTech	Type- Core
Course Code- CSEL301	Course Name-AIML
Year- 2022	Semester- Odd
Date- 08-07-2022	Batch- 5th Sem

1 - Lab Assignment No. 1.2 (Part-B)

Objective: To use Pandas Python library and perform various pre-processing operations.

Download and read the Pima Indians Diabetes Dataset (https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database). Perform the following preprocessing tasks. (**60**)

a) Read the dataset

b) Convert all the column name into uppercase

c) Check the shape of the dataset

d) Check the presence of missing value

e) Handle the missing value if present.

f) Display the mean and standard deviation of "Glucose" in the following format:

Mean (std) = mean_value (Std_Vale)

g) Write a function to find the median value of "BMI" for both 0 and 1 outcome classes

h) In continuation with step g , now replace the null values present in BMI attribute by its median values.

i) Create a new feature named "Age_Category" and add it as last column based on following conditions:

If Age >= 50, Age_Category = Old
If 50> Age >21, Age_Category = Middle_Young
If Age == 21, Age_Category = Young

j) The zero (0) values presents in "Glucose", "BloodPressure", "SkinThickness" and "Insulin" attributes replace with 'NaN' and then fill all NaN entries with corresponding attribute's mean value.

Suggested Platform: Python: Azure Notebook/Google Colab Notebook, packages such as numPy and Pandas.