School of Computer Science Engineering and Technology

Course-BTech	**Type - AI Core-1**
Course Code - CSET211	**Course Name - Statistical Machine Learning**
Year - Second	**Semester - ODD**
	Batch - CSE 3rd Semester

Lab Assignment - 6: Logistic Regression

CO- Mapping

Section	CO1	CO2	CO3
Section 1: Q1-Q6	✓		
Section 2: Q7-Q10	✓		✓

Section 1: Data Preprocessing on Dataset

1. Given a dataset ***diabetes.csv***, write a Python script to load and display the dataset.
2. Rename the columns accordingly: ***'Pregnancies'*** to ***'Pregnant'***, ***'BloodPressure'*** to ***'BP'***, ***'SkinThickness'*** to ***'Skin'*** and ***'DiabetesPedigreeFunction'*** to ***'Pedigree'*** and display again.
3. Use the describe() function to print the statistical summary of the data in the dataframe.
4. Consider the ***'Pregnant'***, ***'BP'***, ***'Insulin'***, ***'BMI'***, ***'Pedigree'*** and ***'Age'*** to be the feature columns and split the dataset into 80% train and 20% test data.
5. Create a scatterplot showing the relation between ***'BMI'*** and ***'Age'*** on the training data with ***'Outcome'*** as hue.
6. Perform Standardization using StandardScaler.

Section 2: Logistic Regression Model

7. Train a logistic regression model using an inbuilt function on train set.
8. Calculate the confusion matrix and display it using heatmap.
9. Calculate the accuracy and f1-score of the model using accuracy_score and f1_score respectively.
10. Print the classification report with the target names ***'with diabetes'*** and ***'without diabetes'***.

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

Submission Instructions:

- Submission required .ipynb file only
- Submission is through LMS only.