School of Computer Science Engineering and Technology

Course- B.TECH	Type- AI Core-1
Course Code- **CSET211**	Course Name- **Statistical Machine Learning**
Year- Second	Semester- ODD
Date- 14/11/2022	Batch- CSE 3rd Semester

Lab Assignment (14th Nov – 18th Nov 2022)

Lab 11– Principal Component Analysis

Objective: Student will be able to learn how to implement decision tree classifier.

Exp. No.	Name	CO1	CO2	CO3
11	PCA			✔

Question -1: Marks: 2, Time: 45 min

Consider the dataset breast_cancer.csv. Dimensions actual data is (569, 30) because dataset has 569 data items with 30 input attributes. If the number of dimensions needs to reduce to 2 from 30 by applying PCA, we need to choose n_components=2, so that the dimensions would be reduced to 2.

1. Load the breast cancer dataset from sklearn
2. Do all the necessary pre-processing on the data.
3. Apply PCA with required number of principal component.
4. Print the different component values.
5. Calculate the variance ratio.

Practice Question: Marks: 2, Time: 45 min

Consider the dataset Wine.csv apply all the required pre-processing steps on the dataset.

1. Reduce the dimension of the dataset using Principal component analysis.
2. Visualise the different principal components.
3. Calculate the explained variance ratio.
4. Apply linear or logistic regression to predict the quality of wine along with the accuracy rate.