|  |  |
| --- | --- |
| **TY BTech** | **Project Based Learning 4 (ML)** |
| **Experiment No: 1** | Feature Selection |

|  |  |
| --- | --- |
| **Title** | Select any data set with high dimensions (such as Bostan dataset, breast cancer dataset) from any repository of data such as SK-Learn, UCI library, Kaggle dataset library etc. Write a program to perform the following operations on the selected dataset and display the result.   1. Reduce dimensions using SelectKBest method 2. Reduce dimensions using SelectPercentile method 3. Reduce dimensions using PCA techniques |
| **Outcomes** | * Use different feature selection techniques to prepare training and testing data sets. |
| **Theory** | Explanation about Feature selection techniques.  What is dimensionality reduction?  Advantages and disadvantages of dimensionality reduction.  Different techniques of dimensionality reduction.  Explanation about 1. SelectKBest, 2. SelectPercentile 3. PCA. |
| **Algorithm/Program:** | Syntax and uses of different functions used in the program.  Algorithm of your program. |
| **Conclusion** |  |
| **Output:** Program print with output**.**  **Additional:** Apply other preprocessing techniques such as   1. Handling missing values 2. Normalize and data scaling 3. Handling categorical value 4. Partition data into train and test split   **Suggested dataset:**  Download and used the dataset from   1. UCI repository 2. Kaggle | |
| **References** | <https://scikit-learn.org/stable/modules/preprocessing.html>  https://scikit-learn.org/stable/modules/feature\_selection.html  <https://scikit-learn.org/stable/modules/unsupervised_reduction.html>  https://scikit-learn.org/stable/modules/generated/sklearn.feature\_selection.SelectKBest.html |