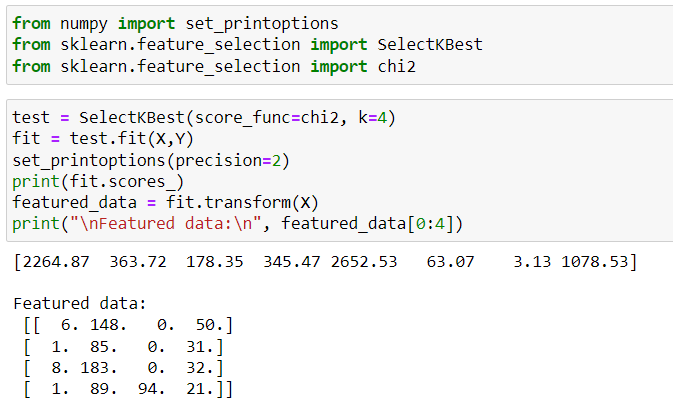
**Name:** Aravind Kumar Kaspe **Banner ID:** 001291145

**ASSIGNMENT 3**

**CSCI 5930 – Homework 3: Feature Selection**

1. **For the data of week 3 (Indian Diabetes) which features do you select based on the below methods?**
2. **Univariate Feature Selection (Chi-square for classification)**

It Selects the top k features based on chi-square statistics; we want to select the top 4 features here.

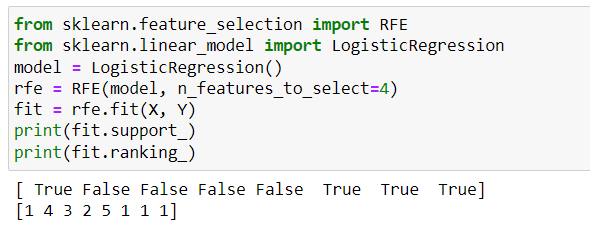
****

Based on the above data we can conclude that 'preg', 'mass', 'age’, and ‘pedi’ are the top

features.

1. **RFE**

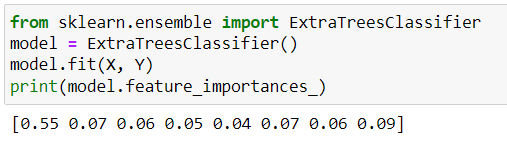
We'll use Logistic Regression as the estimator.



Based on the above data we can conclude that 'preg', 'mass', 'age’, and ‘pedi’ are the top

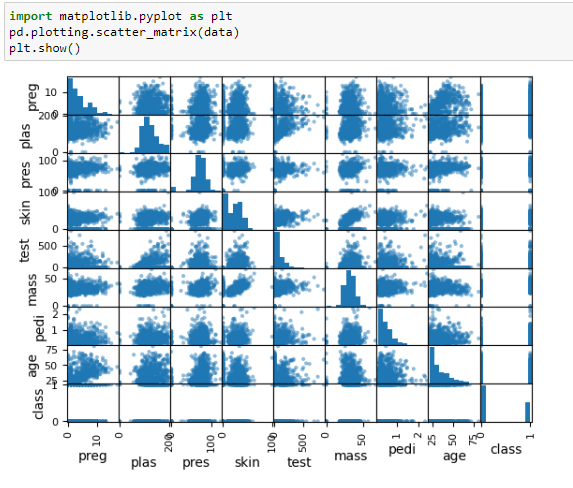
Features.

1. **Feature Importance**

****

Based on the above data we can conclude that the highest value is seen at preg followed by pedi , mass etc.

1. **Scatter Plot**

****

Based on the above data we can conclude that 'preg', 'mass', 'age’, and ‘pedi’ are the top

features.

**b) based on the common results of part a what features do you select finally?**

The most common features are ‘Preg’, ‘Pedi’, ‘Mass’ and ‘Age’.