***Data Mining Project Proposal***

***By: Anees Azhar***

***Roll no.13***

***BS-5A***

***Teacher: Miss Uzma***

***Objective:***

Classify iris plants into three species (cluster analysis) and get a decision tree of plants.

***Dataset Link:***

Iris dataset from Kaggle

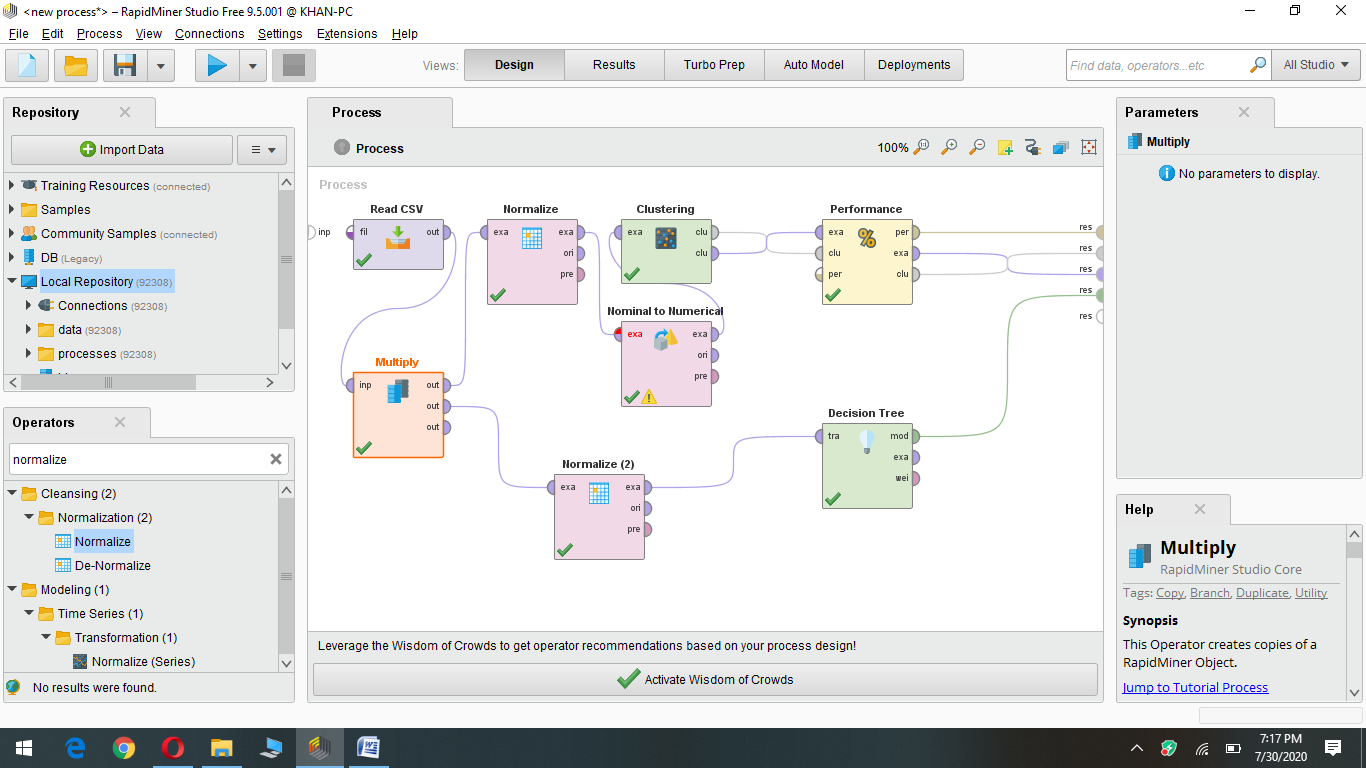
Link: https://www.kaggle.com/uciml/iris/data#

***Algorithms:***

K-Means

Decision Tree

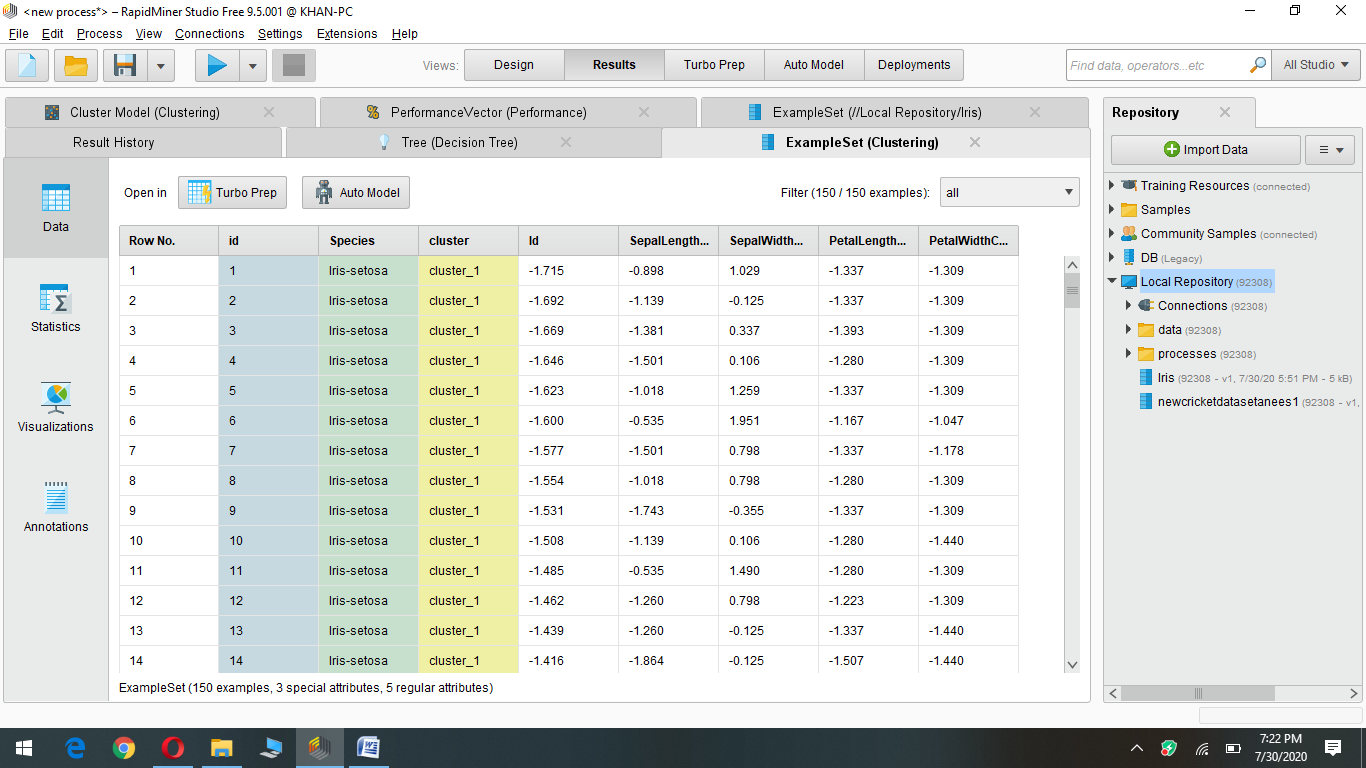
Read CSV is used to read the file containing dataset and then Multiply is used so we can use the same dataset multiple times without using Read CSV again. One output of multiply connects to Normalize so the data can be normalized and then to Nominal to Numerical conversion (it was giving error without converting for some reason saying that numerical data is required and the dataset is not numerical) and it is connected to Clustering (k=3) and then through Cluster Distance Performance and then it is connected to the res.

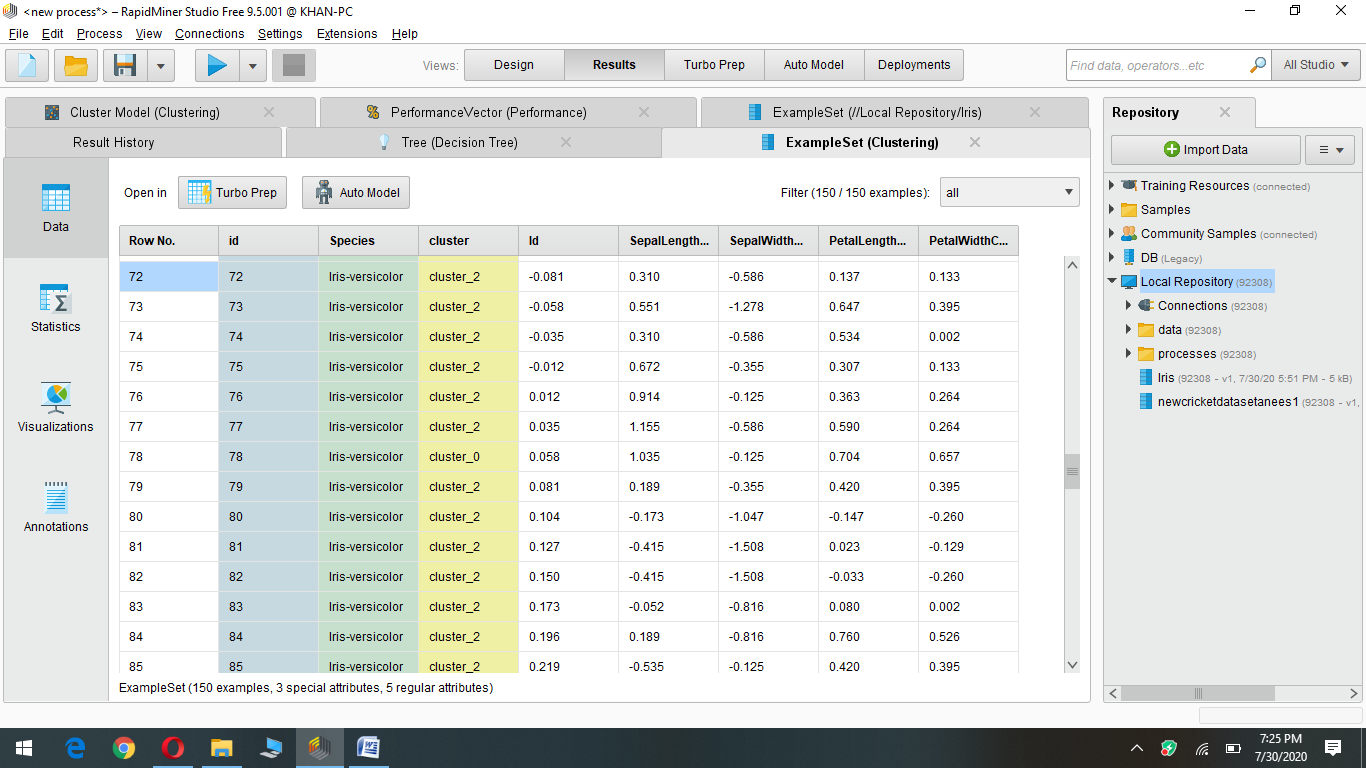
******

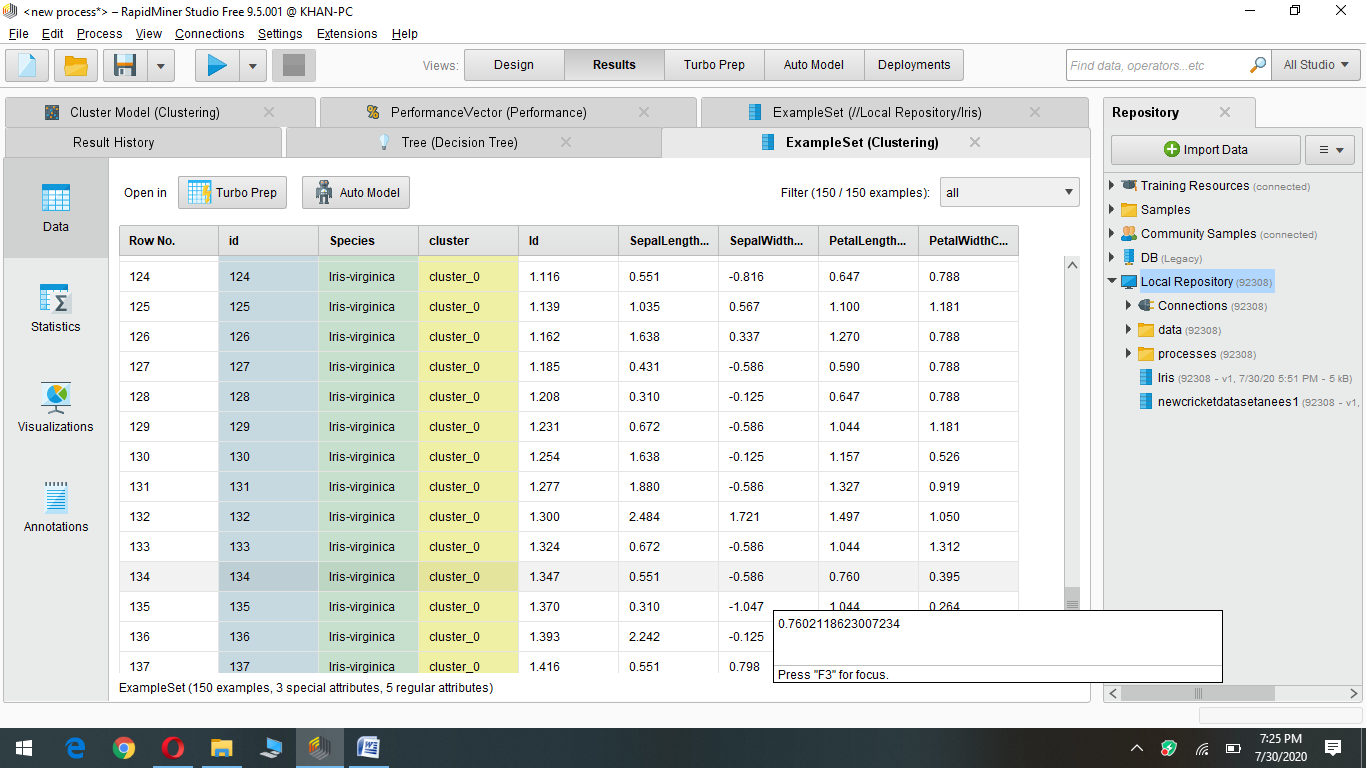
The other output of Multiply connects to Normalize and then to the Decision Tree and then to the res.

***Expected Results:***

In the data we already have the species of plants listed. But what we are trying to do is to use the data to find out the species of the plants without using the species column. We do this by cluster analysis and then we compare the clusters with the species of plants. We will find that the three clusters that we get are the three species of plants. Iris-setosa is cluster 1 and Iris-versicolor is cluster 2 and Iris-virginica is cluster 3. By dividing the data into clusters we have successfully classified the plants into the three different species.







We have also made a decision tree of the plants.

