***Data Mining Project Proposal***

***By: Anees Azhar***

***Roll no.13***

***BS-5A***

***Teacher: Miss Uzma***

***Objective:***

Use data from an autism questionnaire to predict if the person has high chance of autism (YES) or low chance of autism (NO) by using k-means. Also use decision tree and naïve bayes to find out information about the data.

***Dataset Link:***

<https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult>

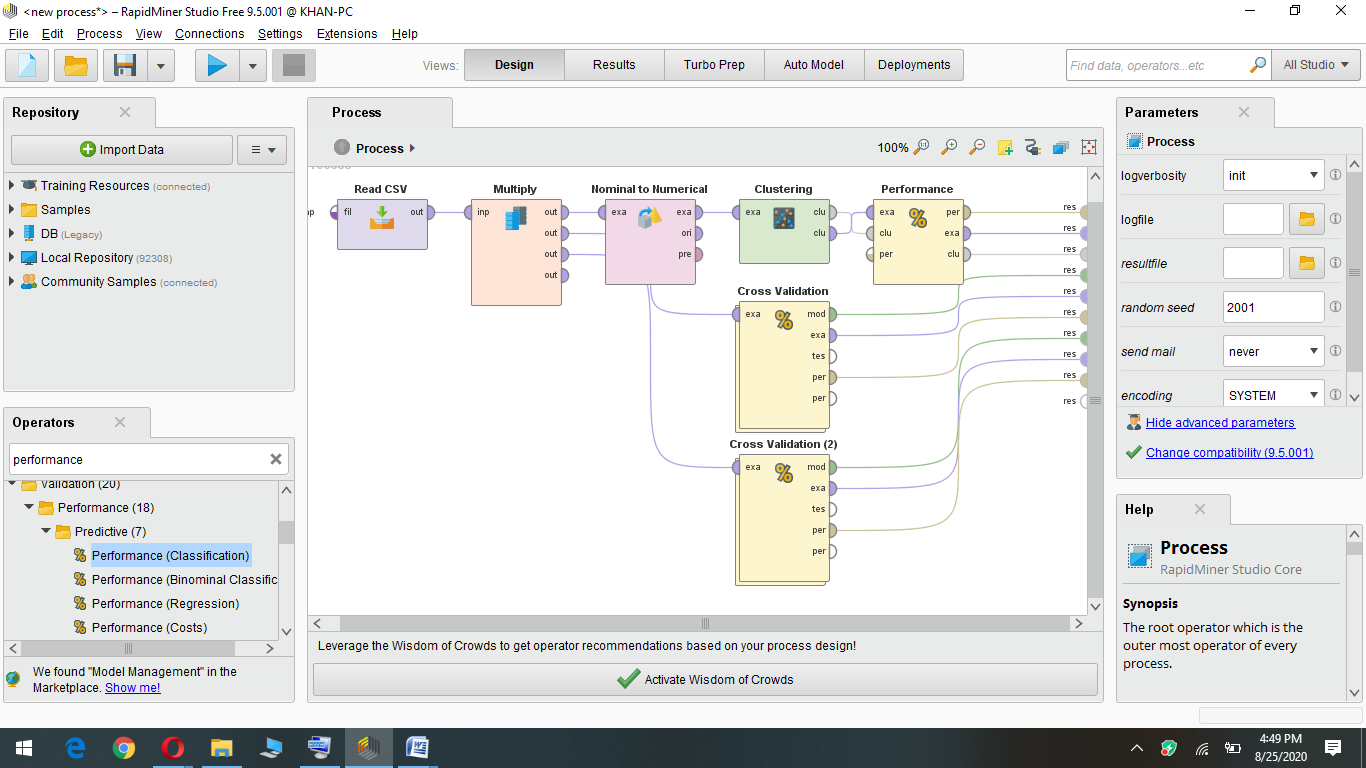
The dataset contains answers of questions of participants in a questionnaire that predicts their likelihood of having autism. The data is binary with 1 for saying yes to the question and 0 for saying no to the questions. The final column has “YES” and “NO” meaning that the participant has a high or low chance of having autism respectively. The data has been changed slightly by reducing unnecessary columns such as ethnicity and gender. The original dataset had only values and no column names, so column names have been added.

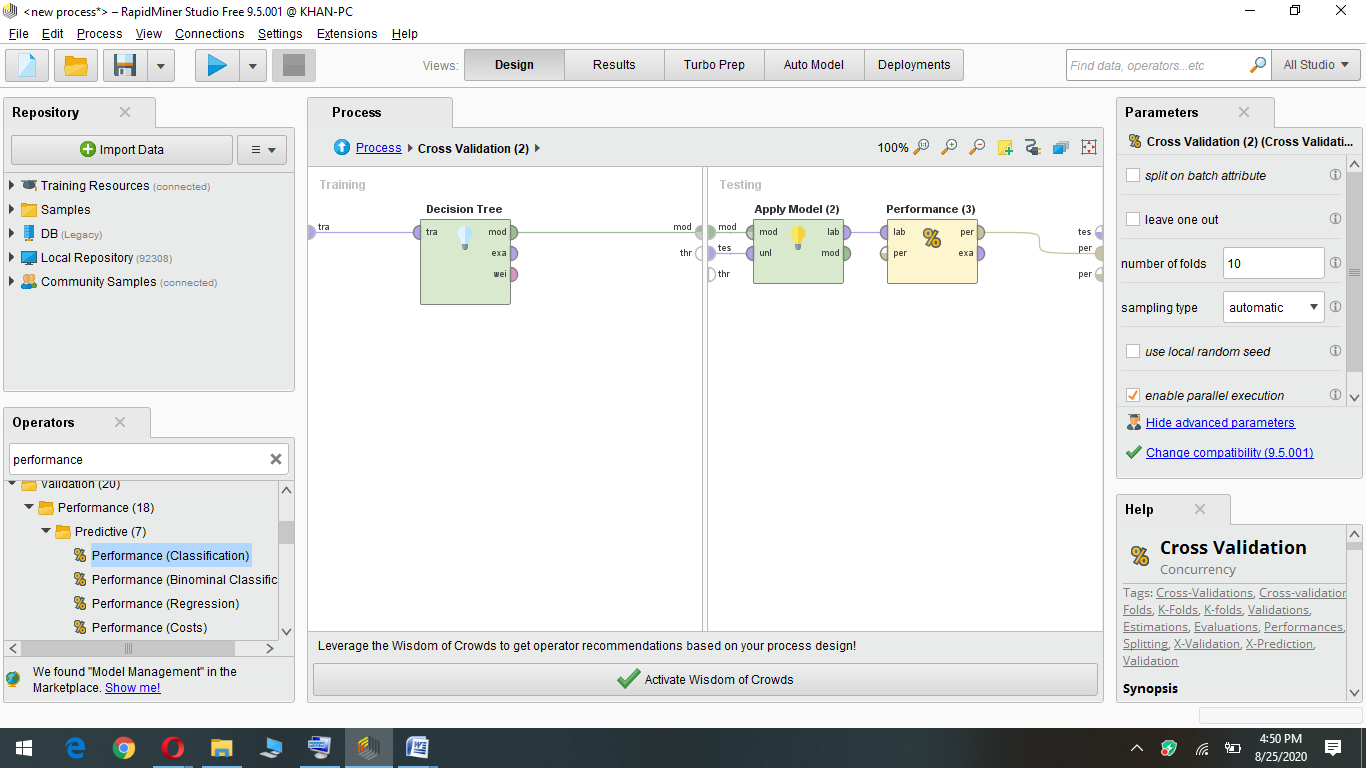
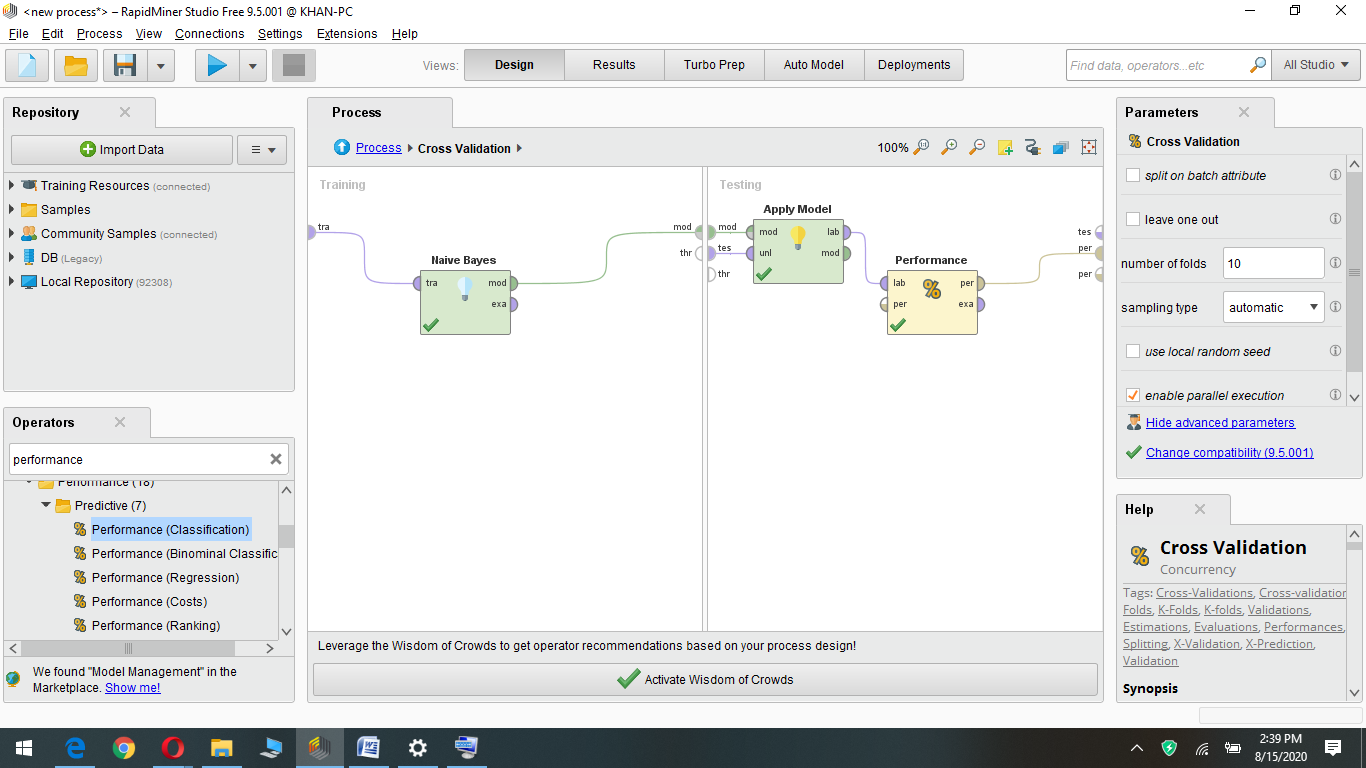
***Algorithms:***

K-Means

Decision Tree

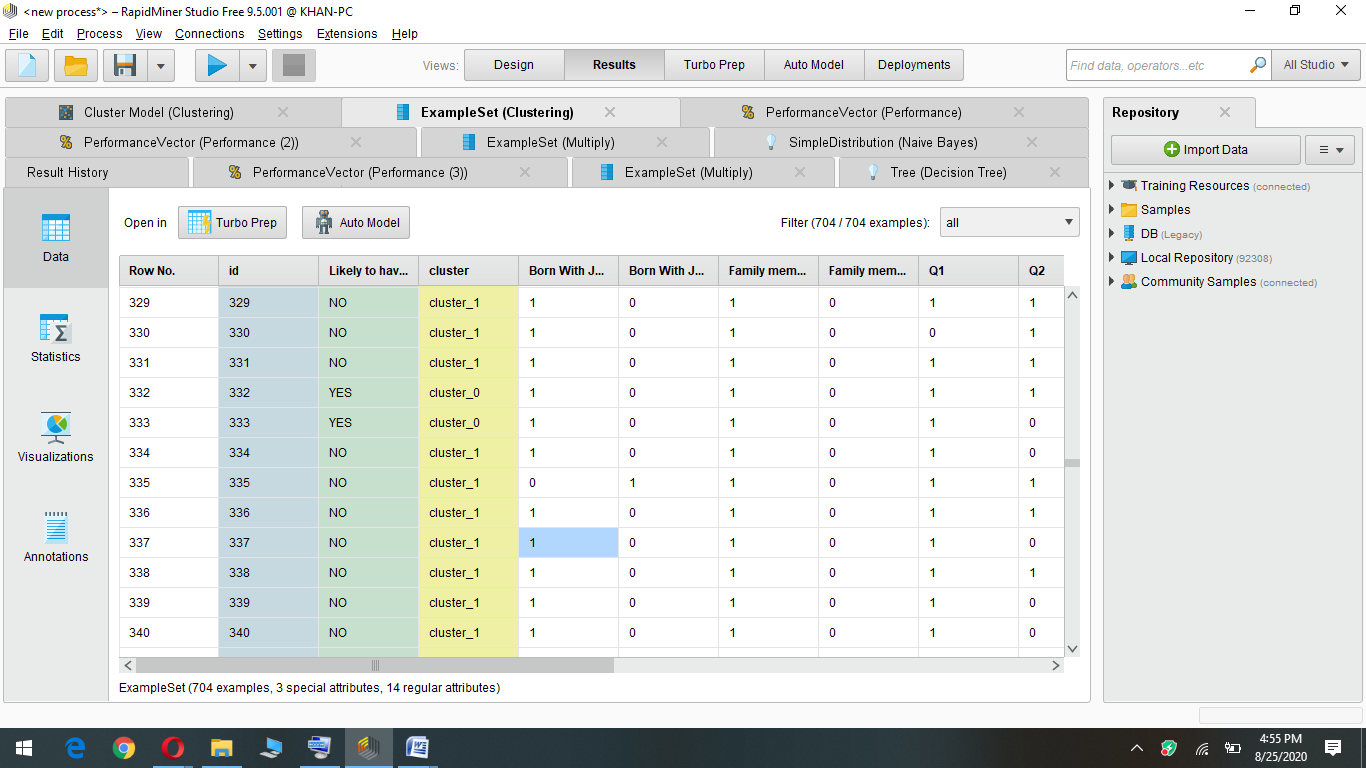
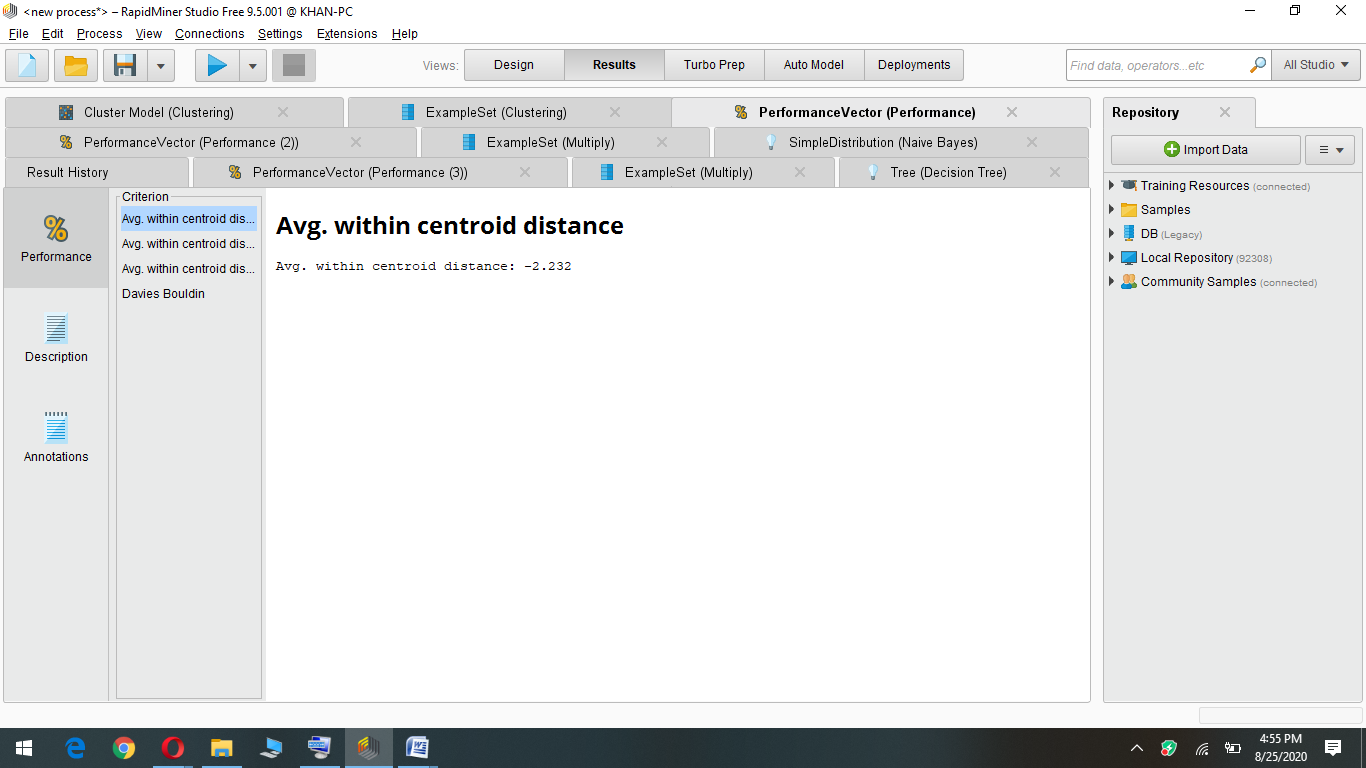
Naïve Bayes



******

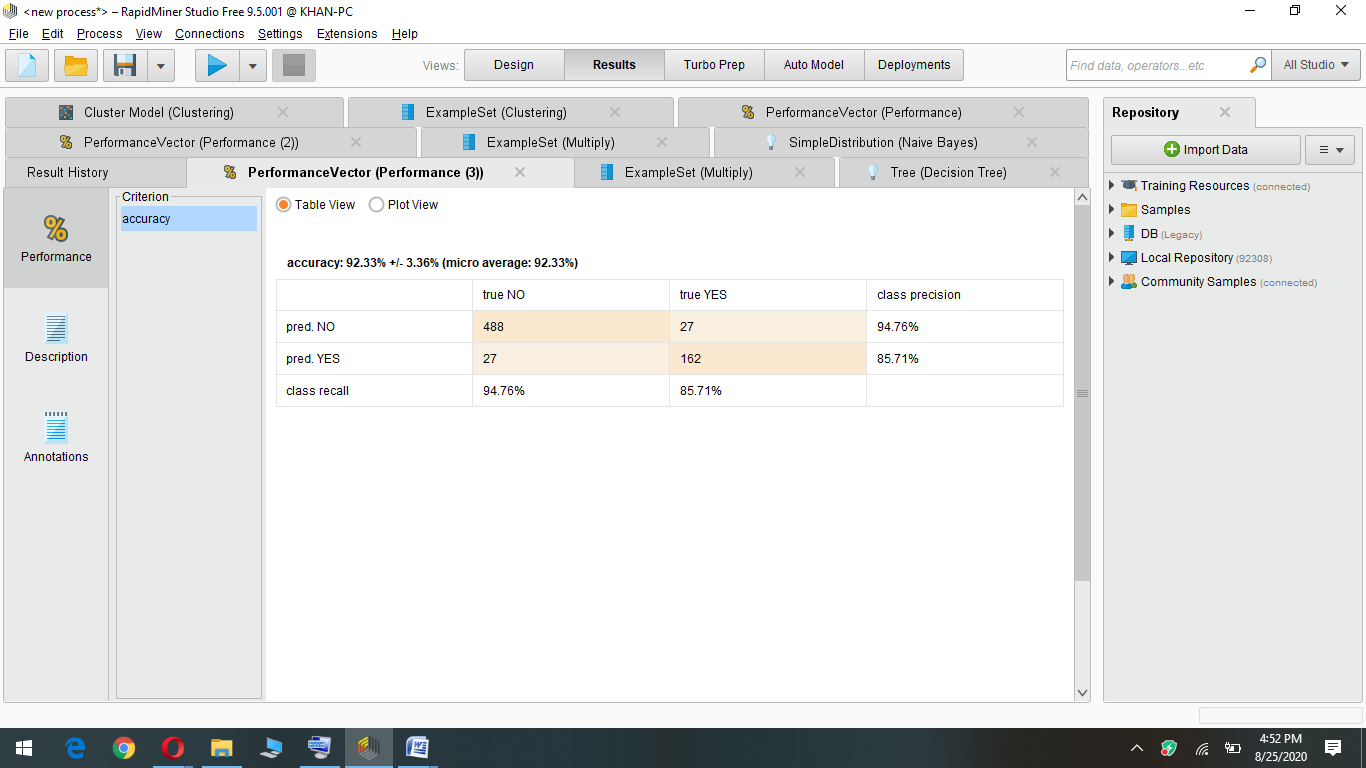
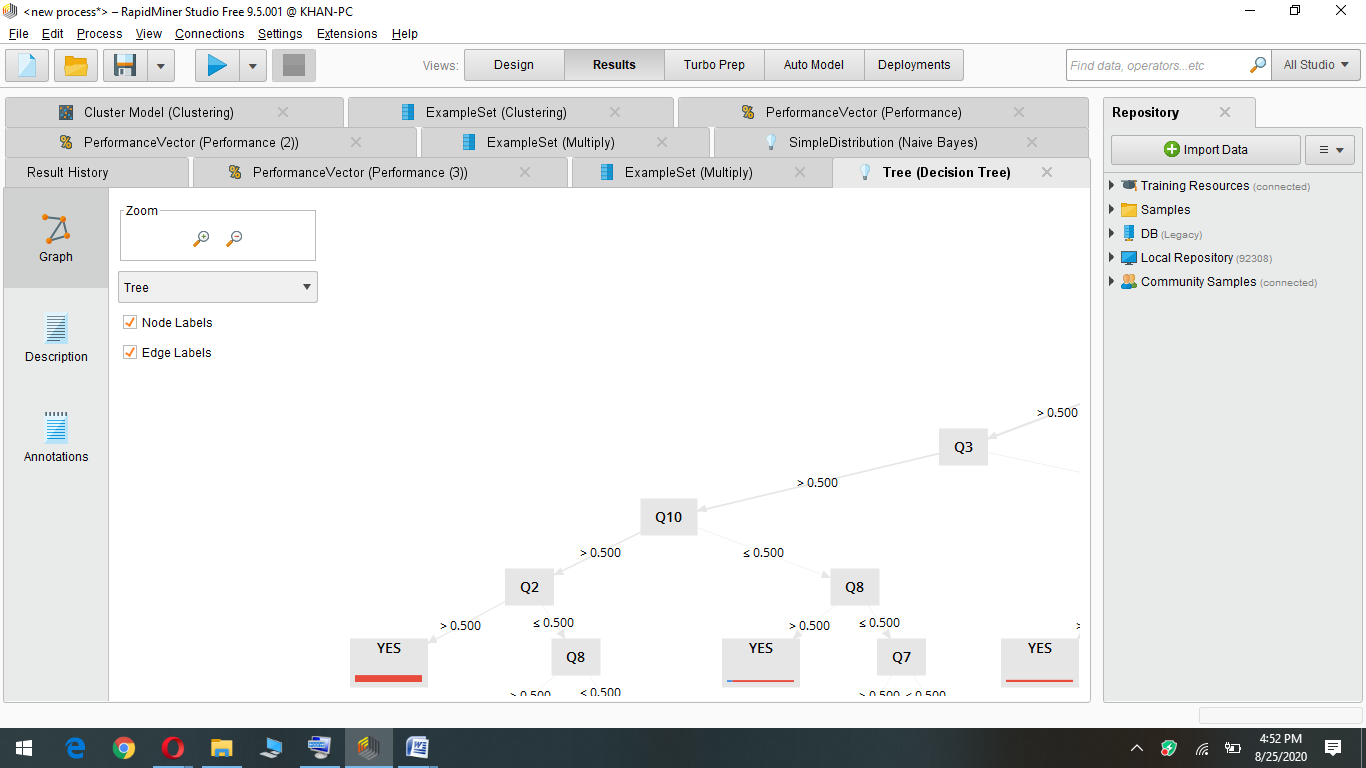
***Expected Results:***

***Output of k-means:***



In our dataset we already have the column of “Likely to have autism” with “YES” and “NO” values which is our label. Using k-means we see if we can make clusters of our data to predict the label column and predict whether the person is likely to have autism (YES) or not (NO). After using k-means with k=2, we can see that cluster 1 is “NO” and cluster 0 is “YES”. We have successfully predicted the label column with high accuracy.

***Output of decision tree:***



***Output of Naïve Bayes:***

