Nitanshu Shahane Div-A Roll no.51

Mini Project DMW Report

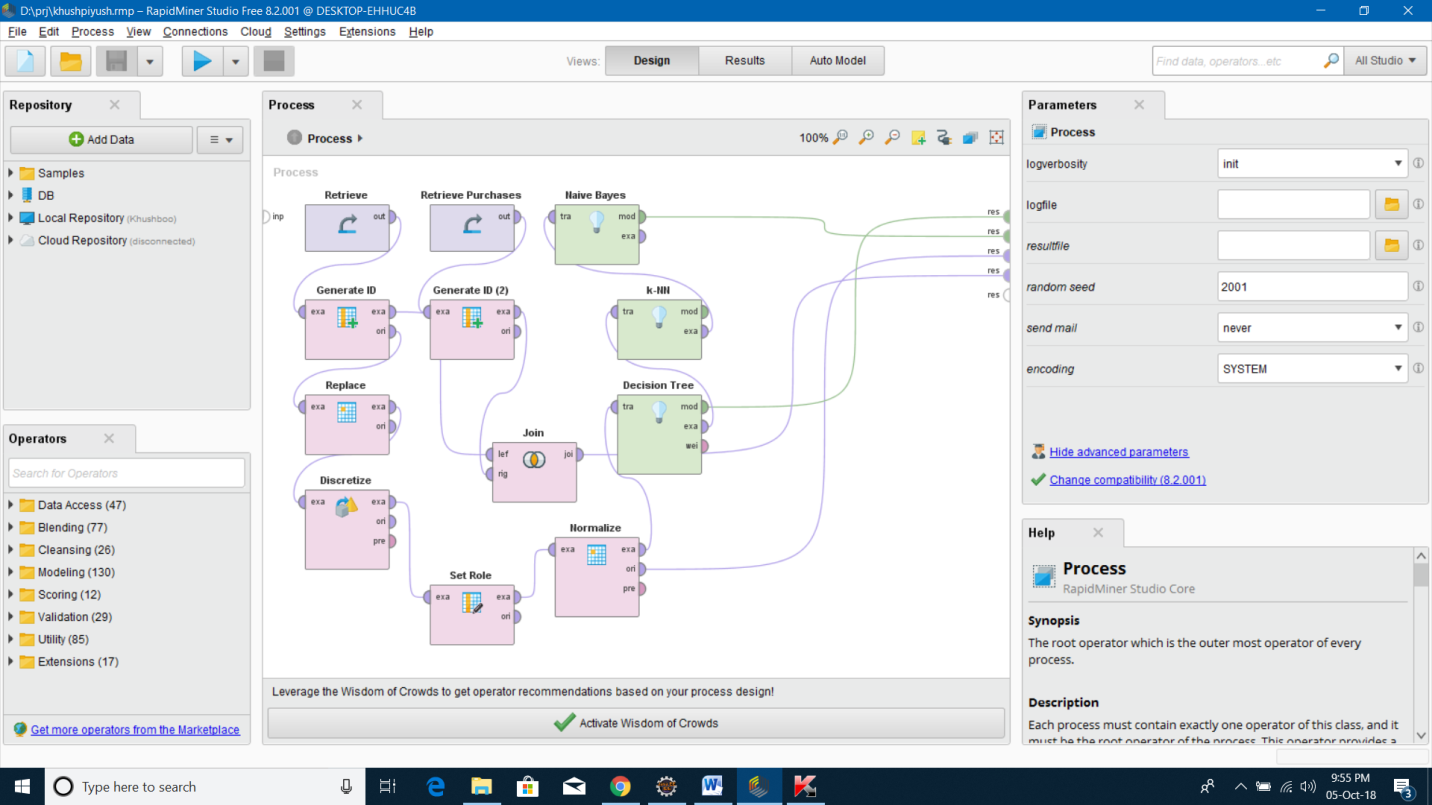
**Title: -** Mini project on classification: Consider a labeled dataset belonging to an application domain. Apply suitable data preprocessing steps such as handling of null values, data reduction, and discretization. For prediction of class labels of given data instances, build classifier models using different techniques (minimum 3), analyze the confusion matrix and compare these models. Also apply cross validation while preparing the training and testing datasets.

**Software Requirements:** Rapid Miner

**Hardware Requirement:** PIV, 2GB RAM, 500 GB HDD, Lenovo A13-4089Model.

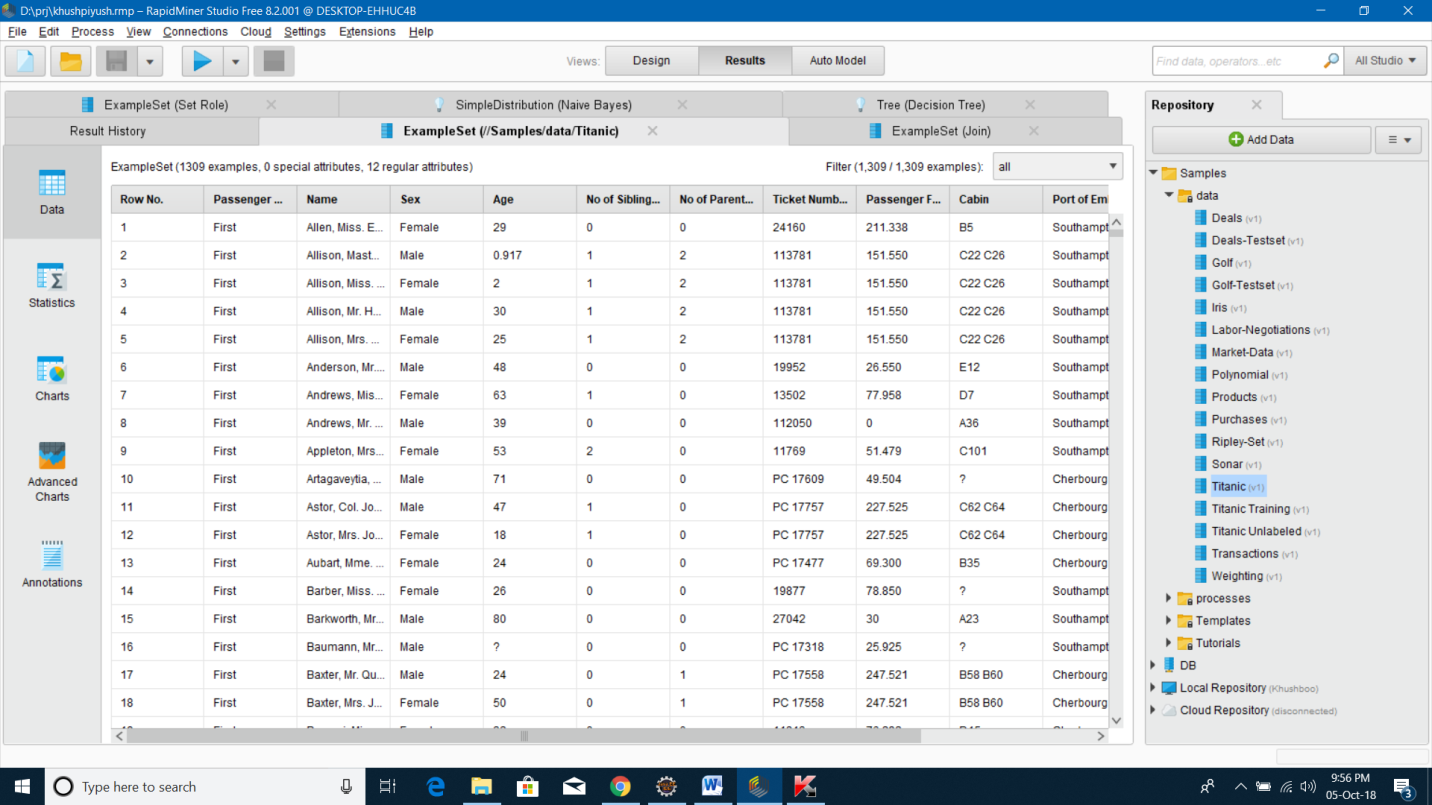
Theory:-

Project view

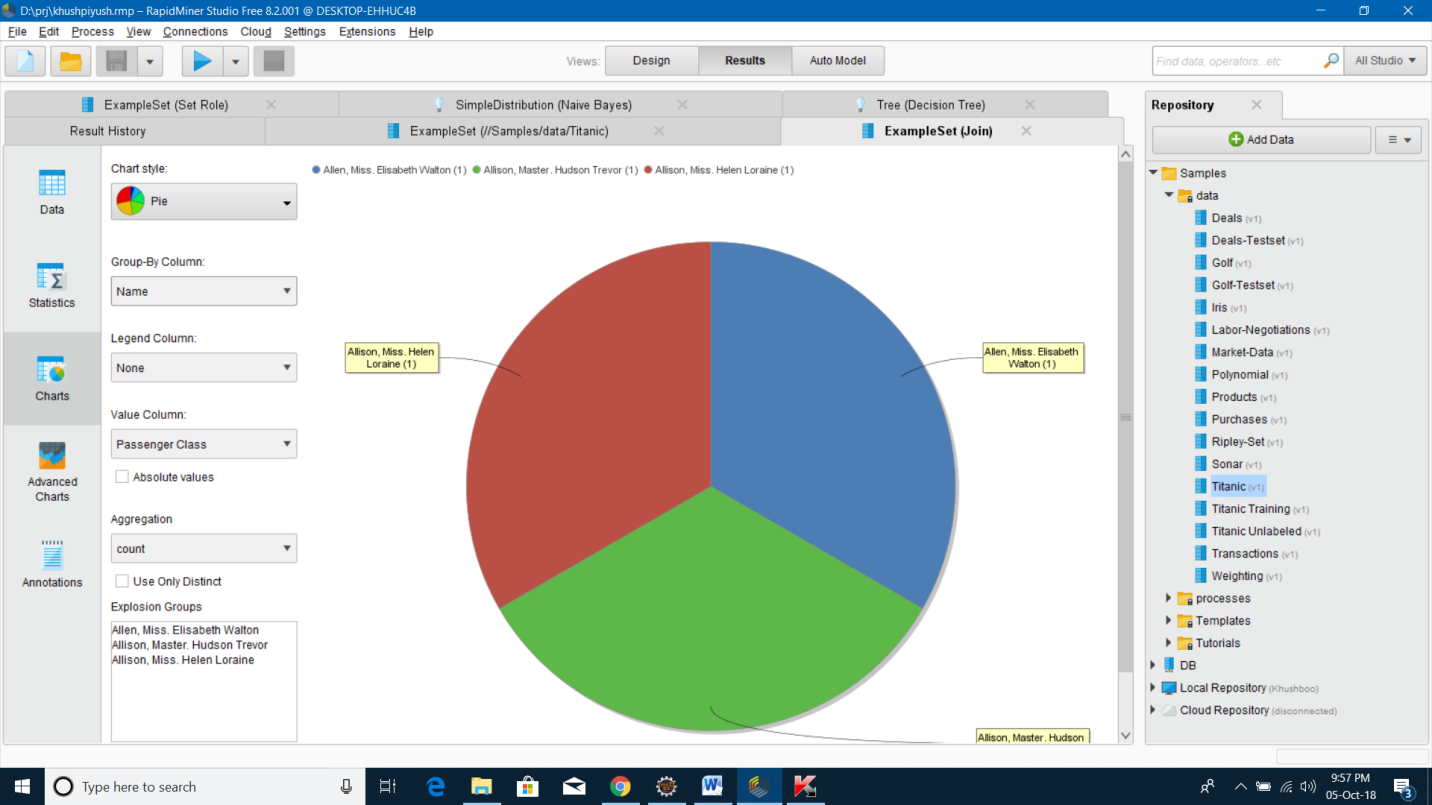
****

**Project Output:-**

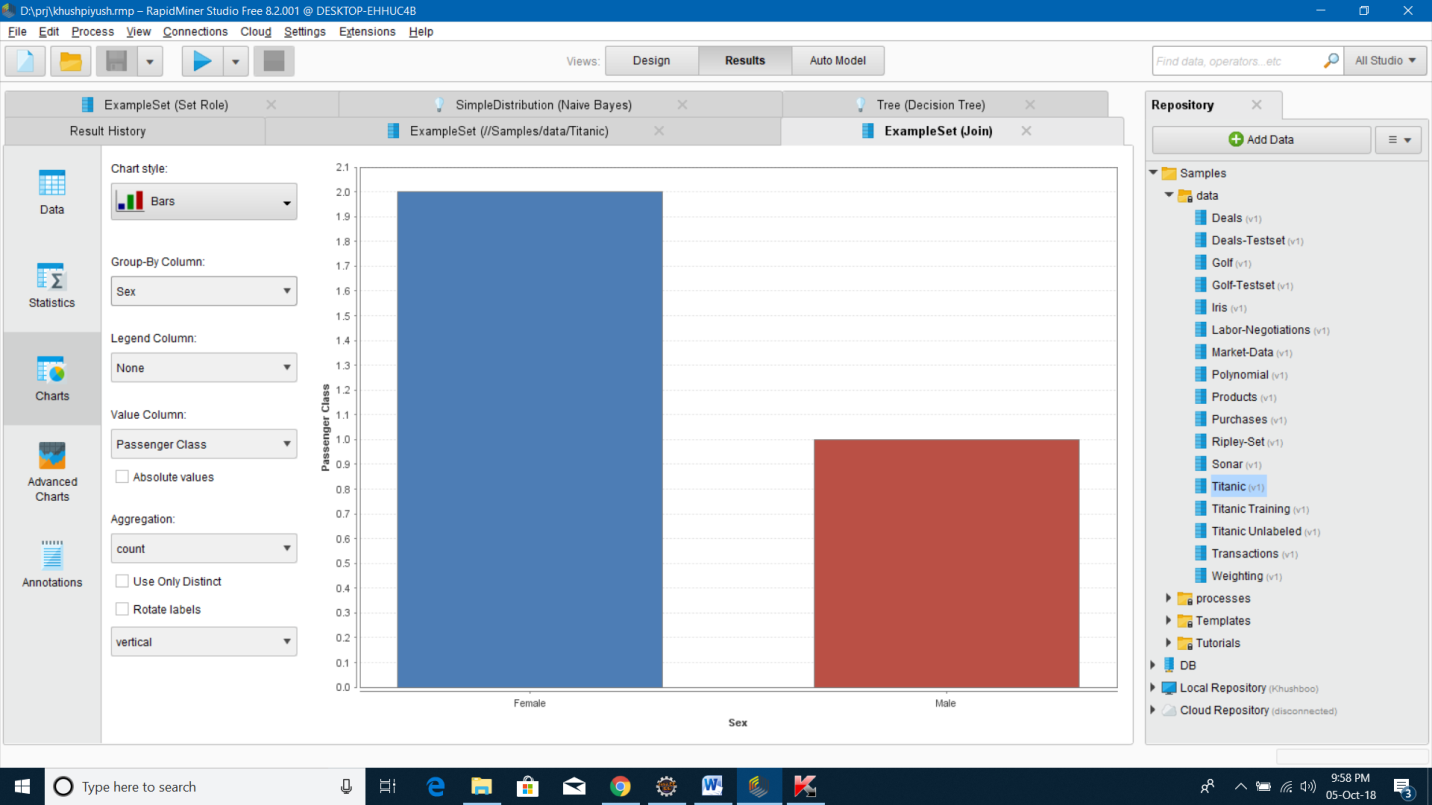
**1)Database output**

****

2) Pie Diagram: On Name



3) Bars : On sex

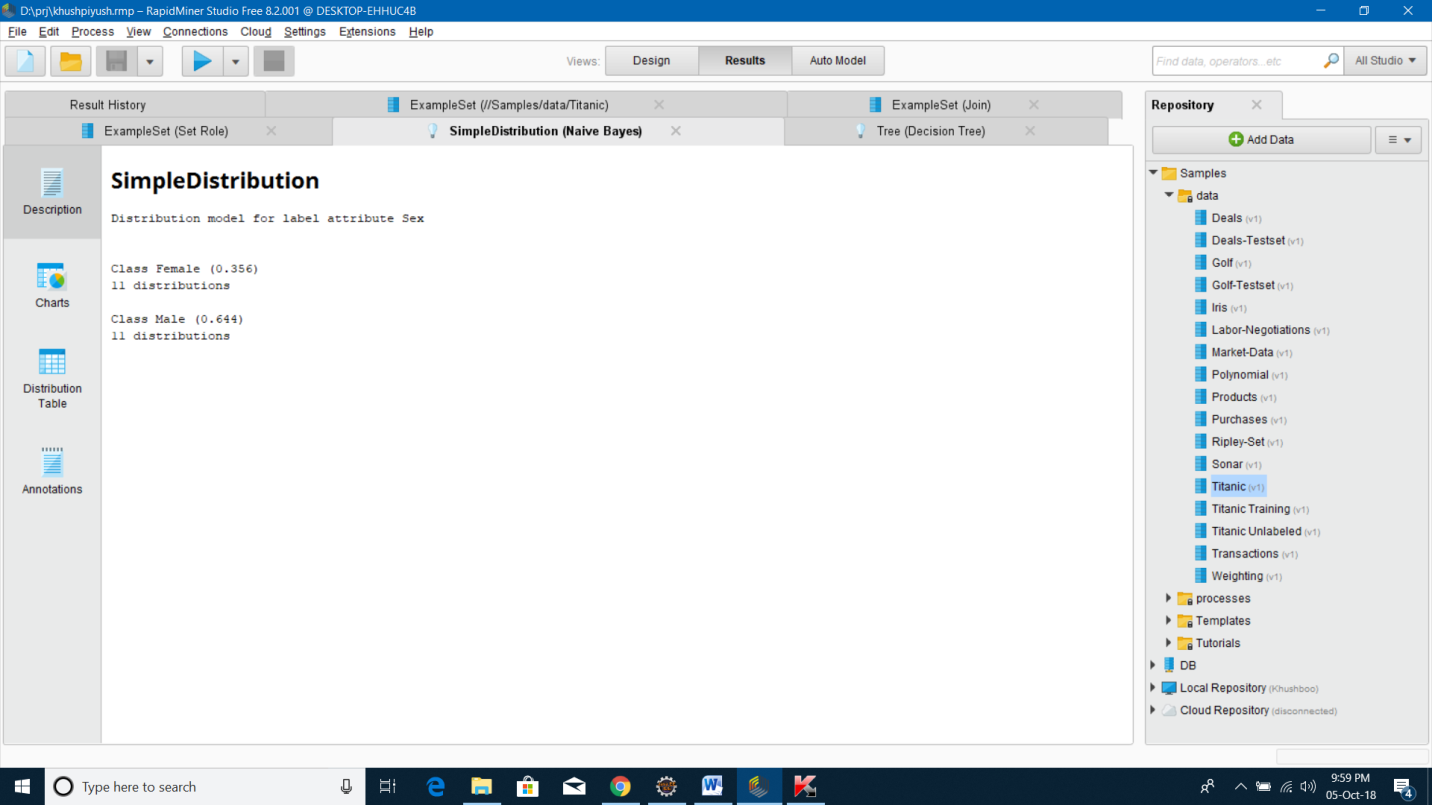


Operations perform on Project

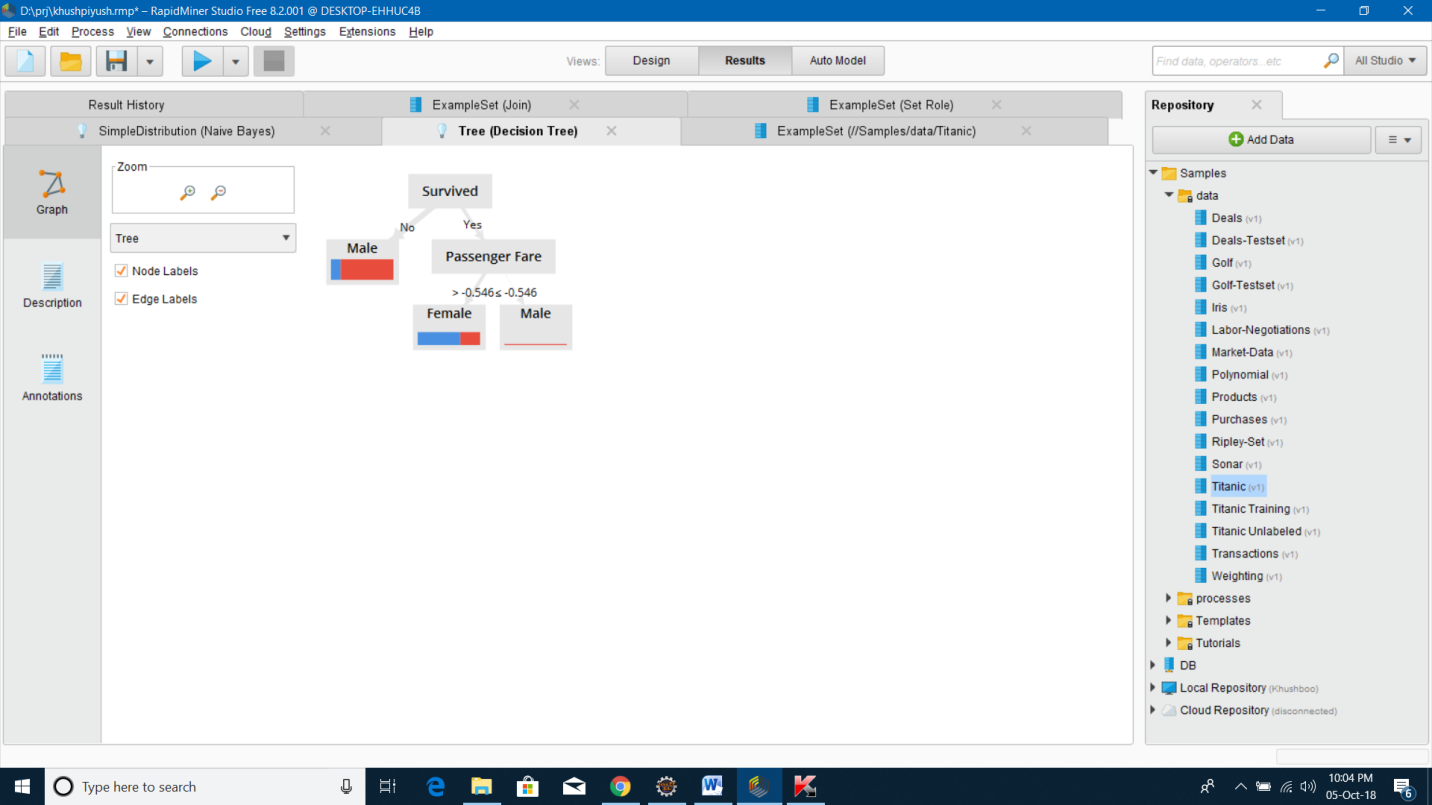
1. Simple Distribution (Naive Bays):-

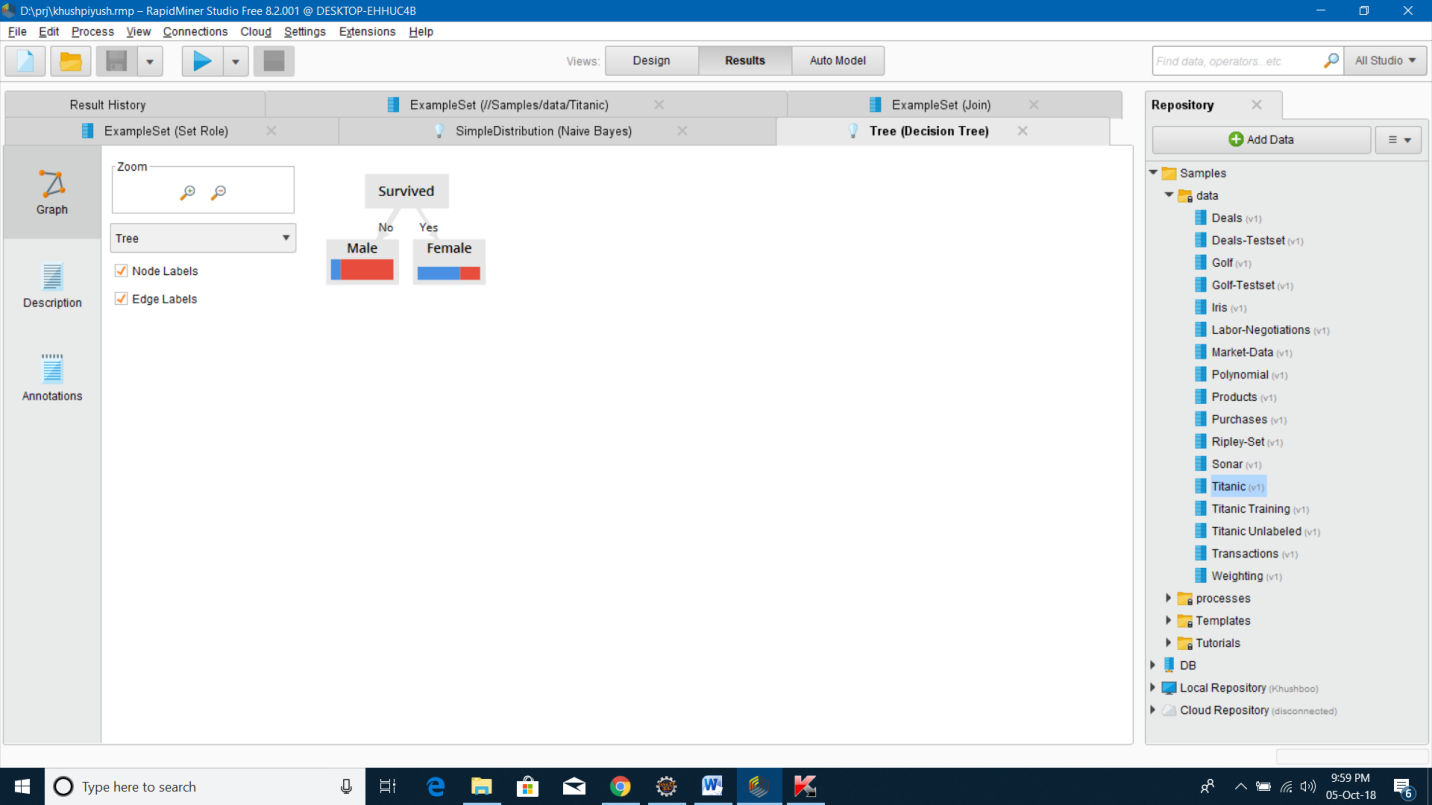
It shows label distribution.

It represents the density of probability of the TF-IDF of the words

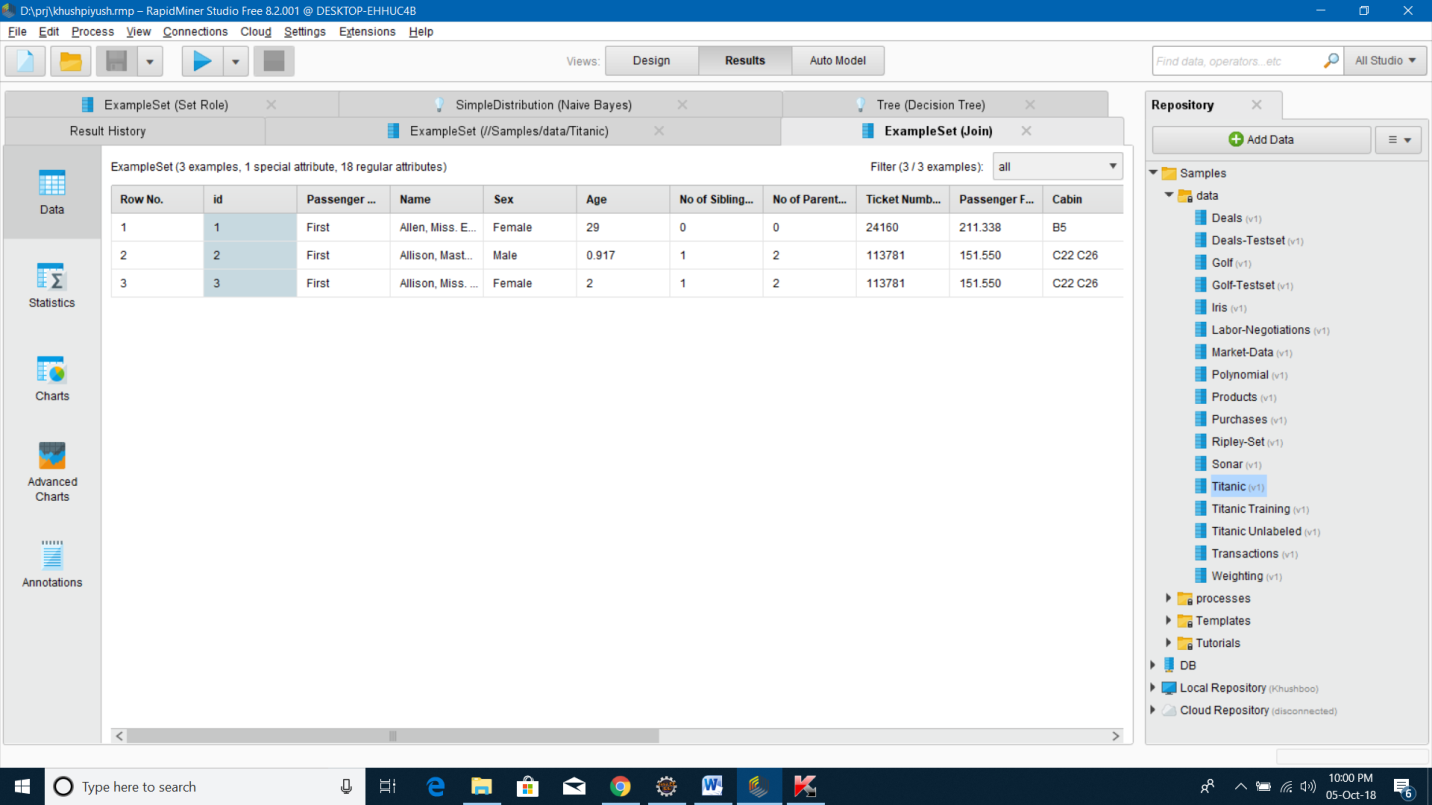


1. Decision Tree :- This Operator generates a decision tree model, which can be used for classification and regression.

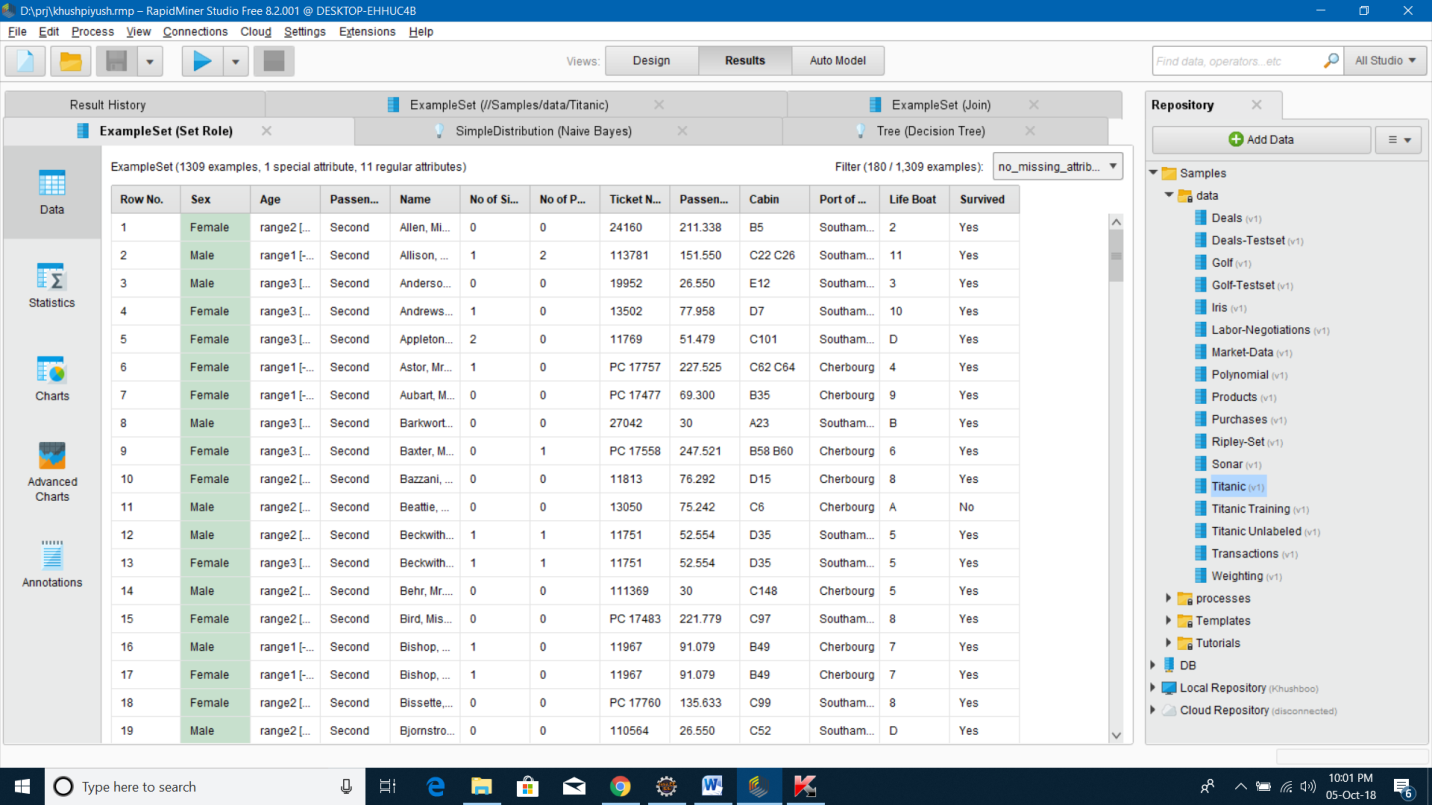




3) Join:- This Operator joins two Example Sets using one or more Attributes of the input Example Sets as key attributes.



4) Set Role :- This Operator is used to change the role of one or more Attributes.



5) Discretizes:-This operator discretizes the selected numerical attributes into user-specified number of bins. Bins of equal range are automatically generated; the number of the values in different bins may vary.

6) K-NN: - This Operator generates a k-Nearest Neighbor model, which is used for classification or regression.

7) Replace: - This operator replaces parts of the values of selected nominal attributes matching a specified regular expression by a specified replacement.