|  |  |
| --- | --- |
|  |  |
| Project 2 |  |
|  |  |
|  | Data Mining |
|  | Ahmed Salah Hammad120201578 |

**Title of Dataset:**

-Airline Passenger Satisfaction

**Description of Dataset:**

-This dataset contains an airline passenger satisfaction survey. What factors are highly correlated to a satisfied (or dissatisfied) passenger.

**Objective of Data Mining project:**

- Predict new passengers’ satisfaction to keep them as Loyal Customer’s

- Predict the passengers who are likely to be disloyal Customer to make them Loyal Customers.

- Improve services that can be increase satisfaction for passengers.

**Name of Attributes:**

**-** **Customer Type:** The customer type (Loyal customer, disloyal customer).

**-Age:** The actual age of the passengers

**-Type of Travel:**Purpose of the flight of the passengers (Personal Travel, Business Travel).

**-Class:** Travel class in the plane of the passengers (Business, Eco, Eco Plus)

**-Flight distance:** The flight distance of this journey

**-Inflight Wi-Fi service**: Satisfaction level of the inflight Wi-Fi service (0:Not Applicable;1-5)

**-Ease of Online booking:** Satisfaction level of online booking

**-Food and drink:** Satisfaction level of Food and drink

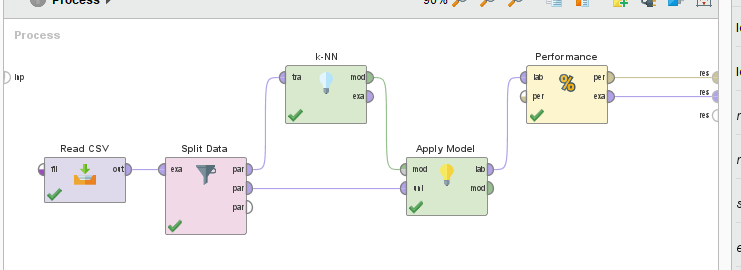
**-Satisfaction:** Airline satisfaction level (Satisfaction, neutral or dissatisfaction), **and its class (Label)**

**Note:** the dataset has been cleaned by the owner, and I’ve deleted some rows because the large number in original dataset.

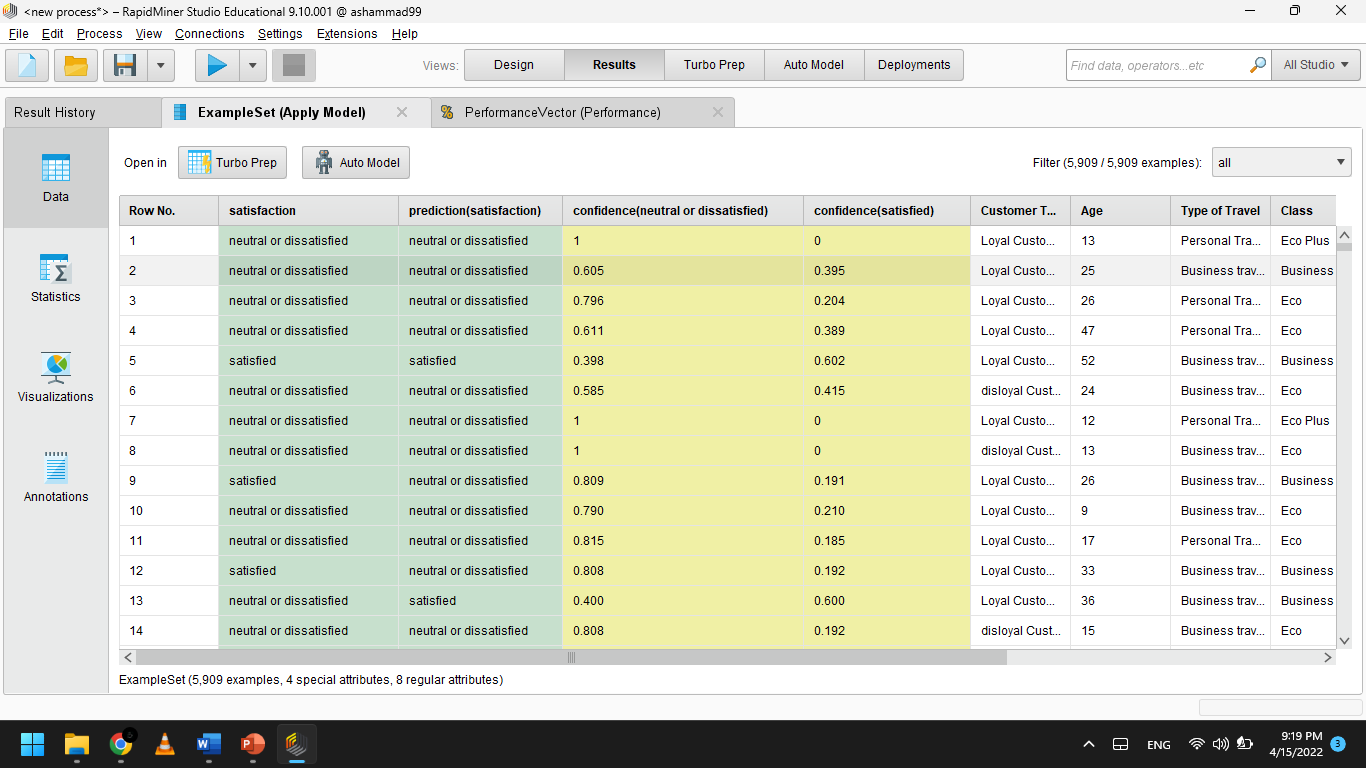
1. **KNN method:**

-The data was split to 0.7 for training data and 0.3 for testing data.

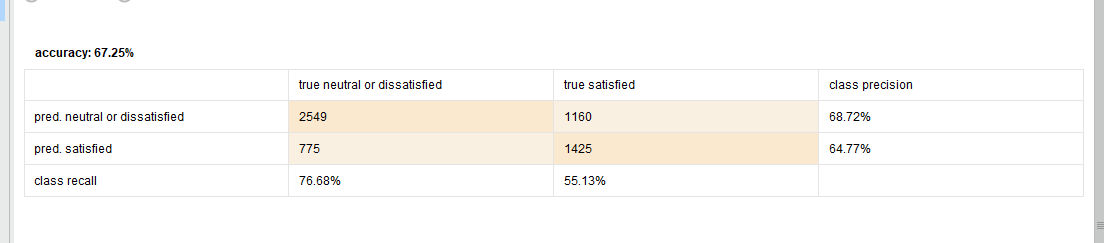
- k (user-defined) =5, that’s mean calculate the distance for the 5 nearest neighbors to predicate class of new instance.

Screenshots of process:

**Result of KNN:**



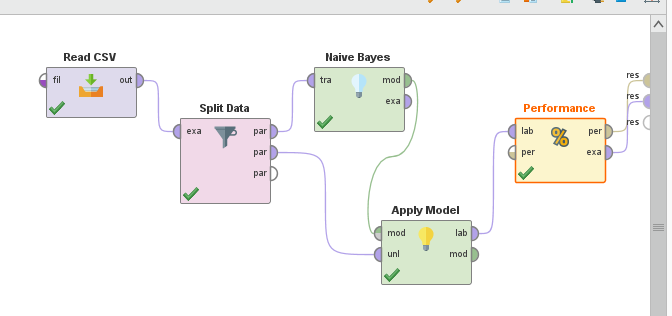
Accuracy with confusion matrix:

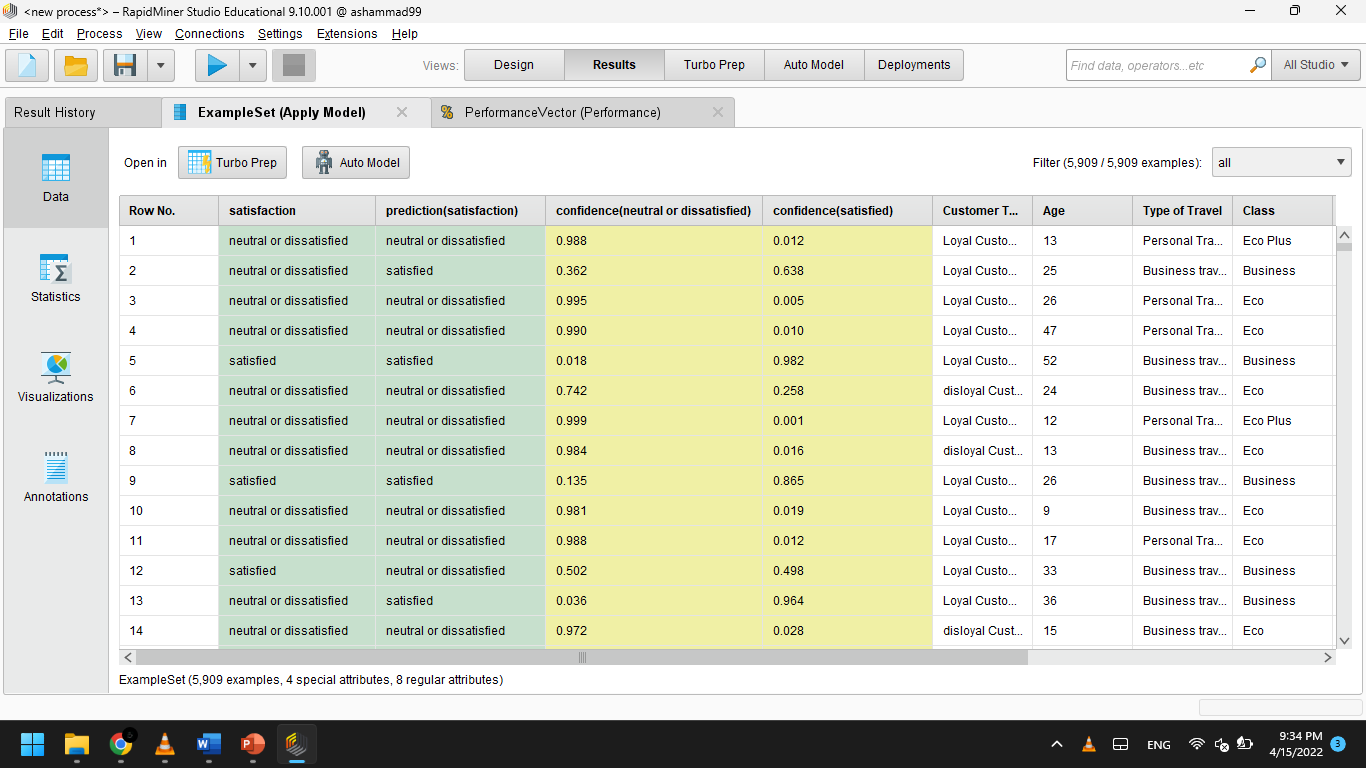


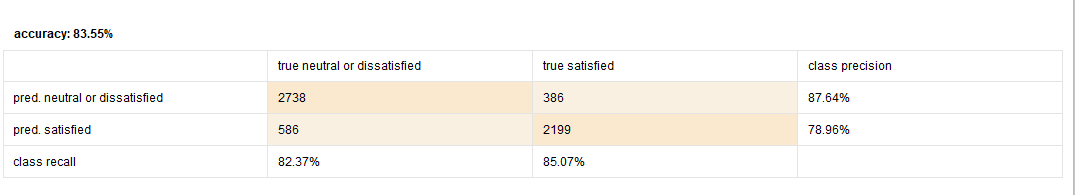
1. **Naïve Bias method:**

-The data was split to 0.7 for training data and 0.3 for testing data.

Screenshots of process:

****

**Result of Naïve Bias:**

Accuracy with confusion matrix:****

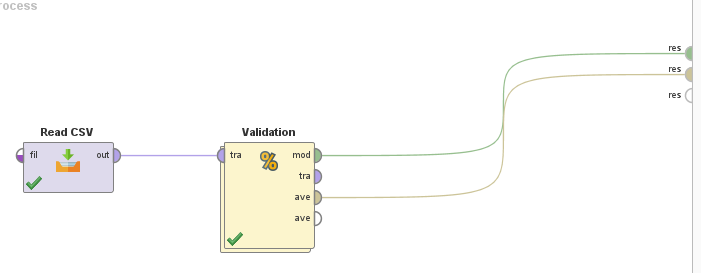
1. **Decision Tree method:**

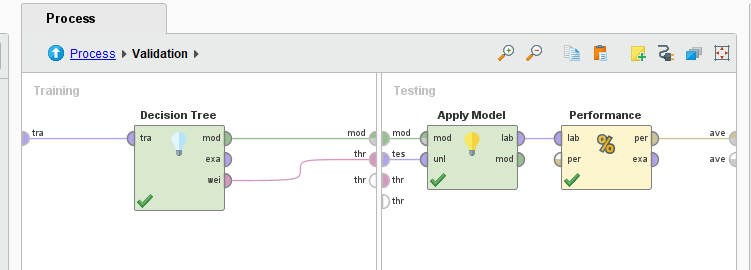
-Maximal depth: 10 that’s mean the max number of edges from far node to the tree's root node.

-apply pruning: true, that’s mean remove the weak branches that cause poor accuracy

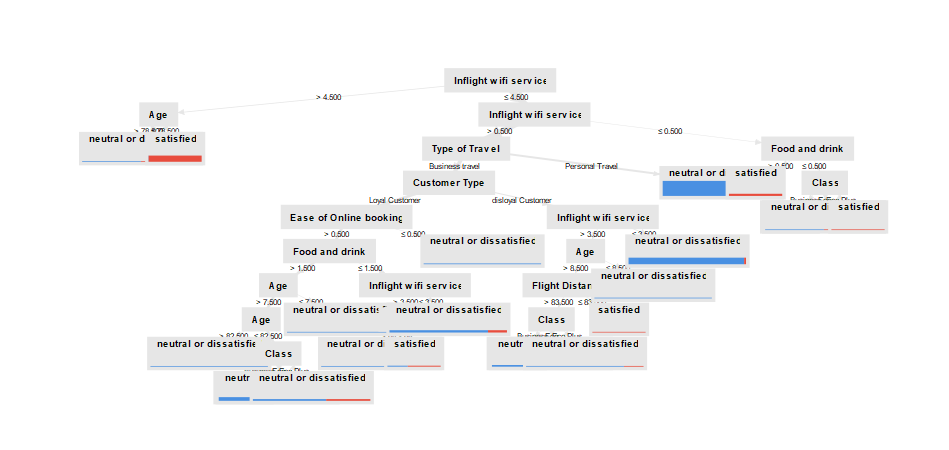
-The data was split to 0.7 for training data and 0.3 for testing data.

Screenshots of process:

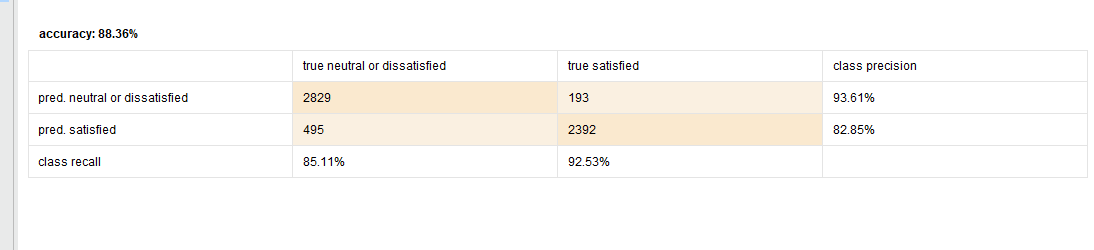




**Result of Decision Tree:**

****

Accuracy with confusion matrix:

****

1. **Random Forest method:**

-Number of trees: 50, that’s mean number of models (decision tress) that created from that same dataset, then select tree that has the more accuracy.

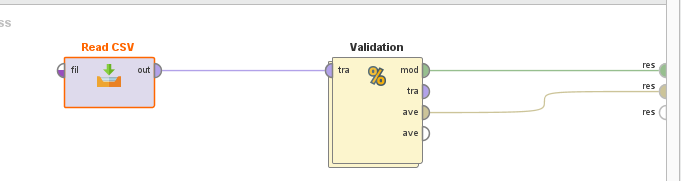
-Maximal depth: 10 that’s mean the max number of edges from far node to the tree's root node.

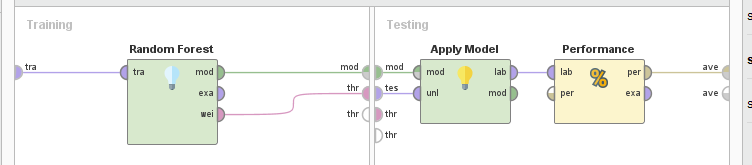
apply pruning: true , that’s mean remove the weak branches that cause poor accuracy.

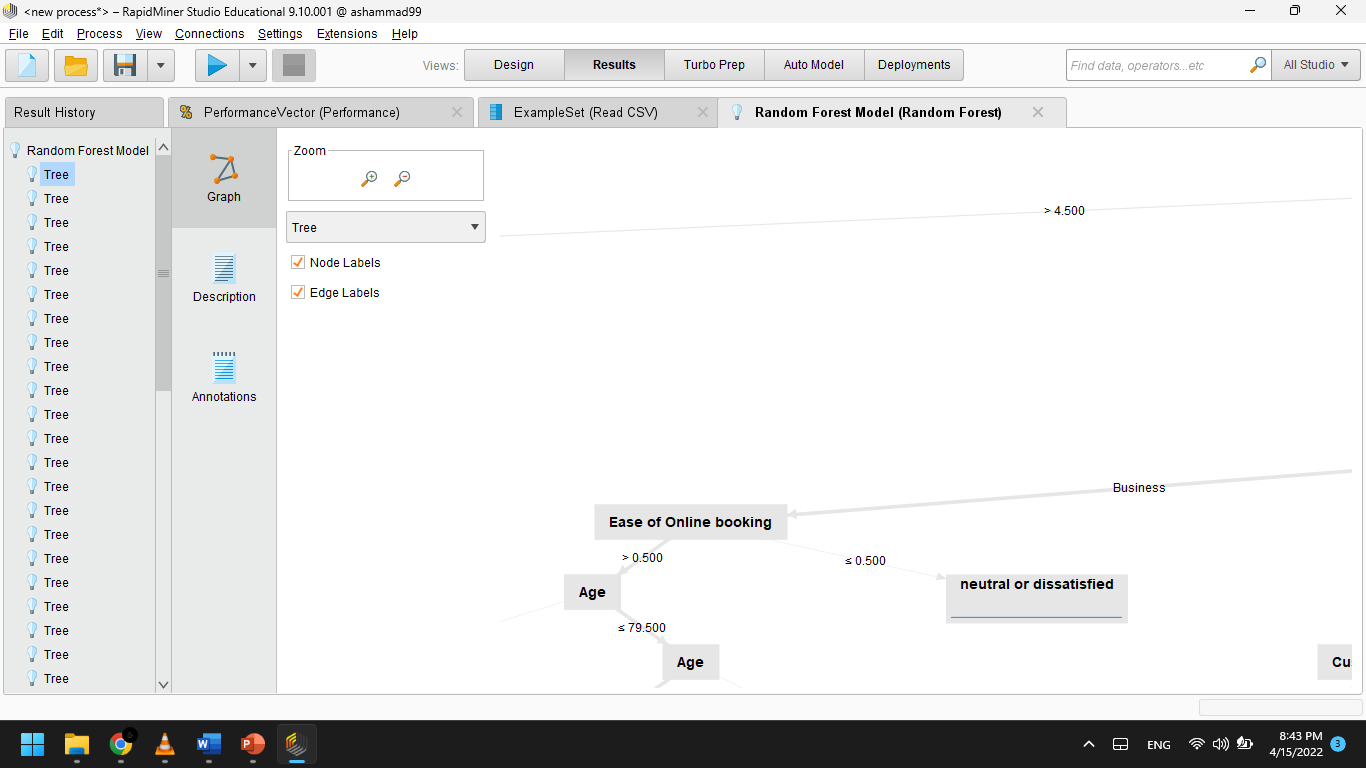
Apply pre-pruning: do not split a node if this would result in the goodness measure falling below a threshold

-The data was split to 0.7 for training data and 0.3 for testing data.

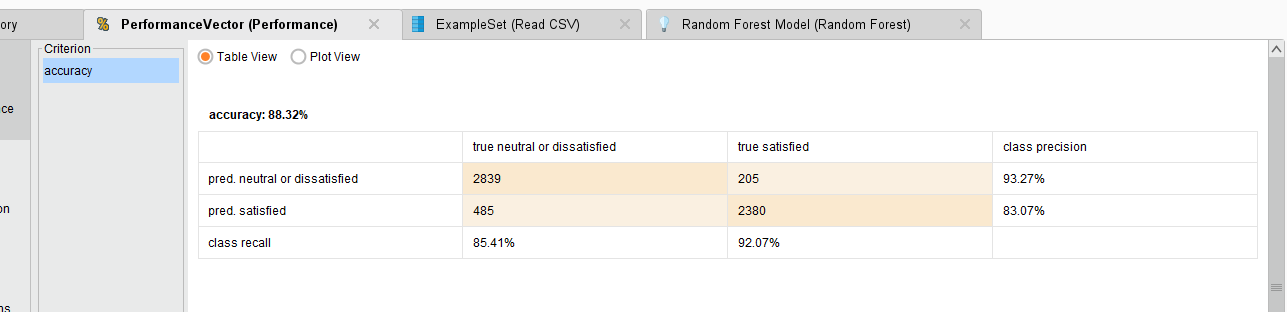
Screenshots of process:





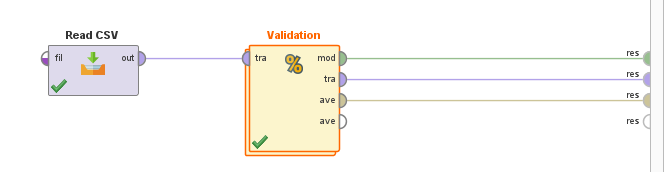
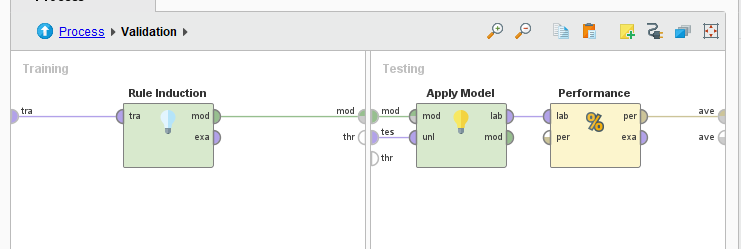


Accuracy with confusion matrix:

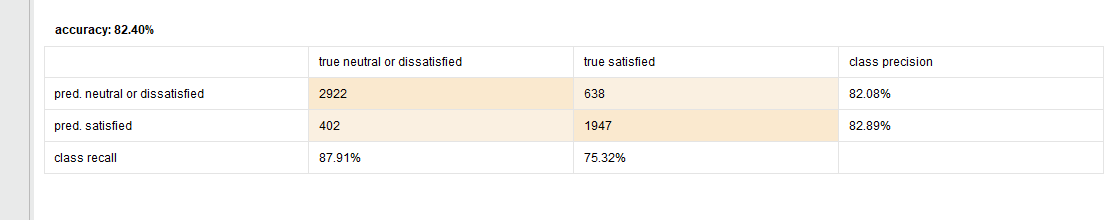
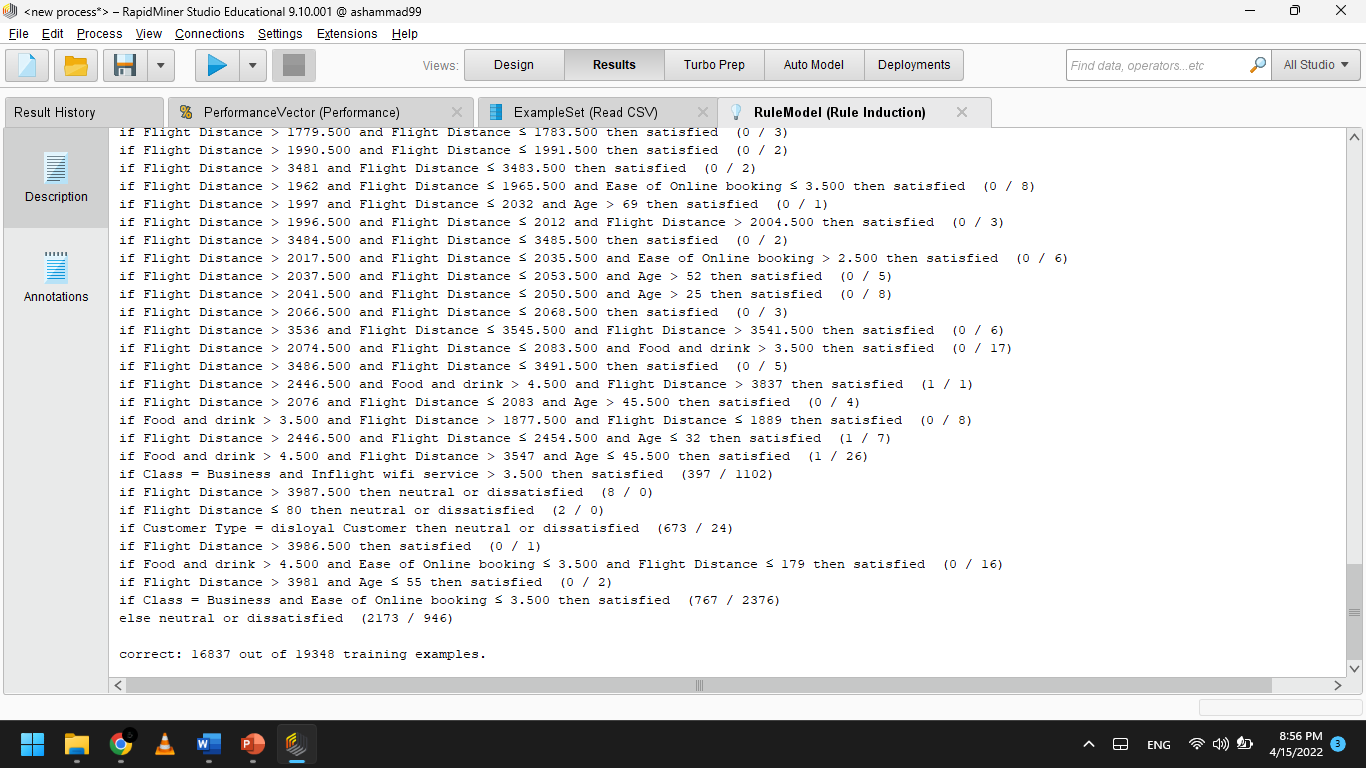
****

1. **Rule Based classifier:**

-The data was split to 0.7 for training data and 0.3 for testing data.

****Screenshots of process:

**Result of Rule Based Classifier**

****

Accuracy with confusion matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification method** | **KNN** | **Naïve Bias** | **Decision Tree** | **Random Forest** | **Rule Based classifier** |
| **Accuracy** | **67.25%** | **83.55%** | **88.36%** | **88.32%** | **82.40%** |