

Task-1

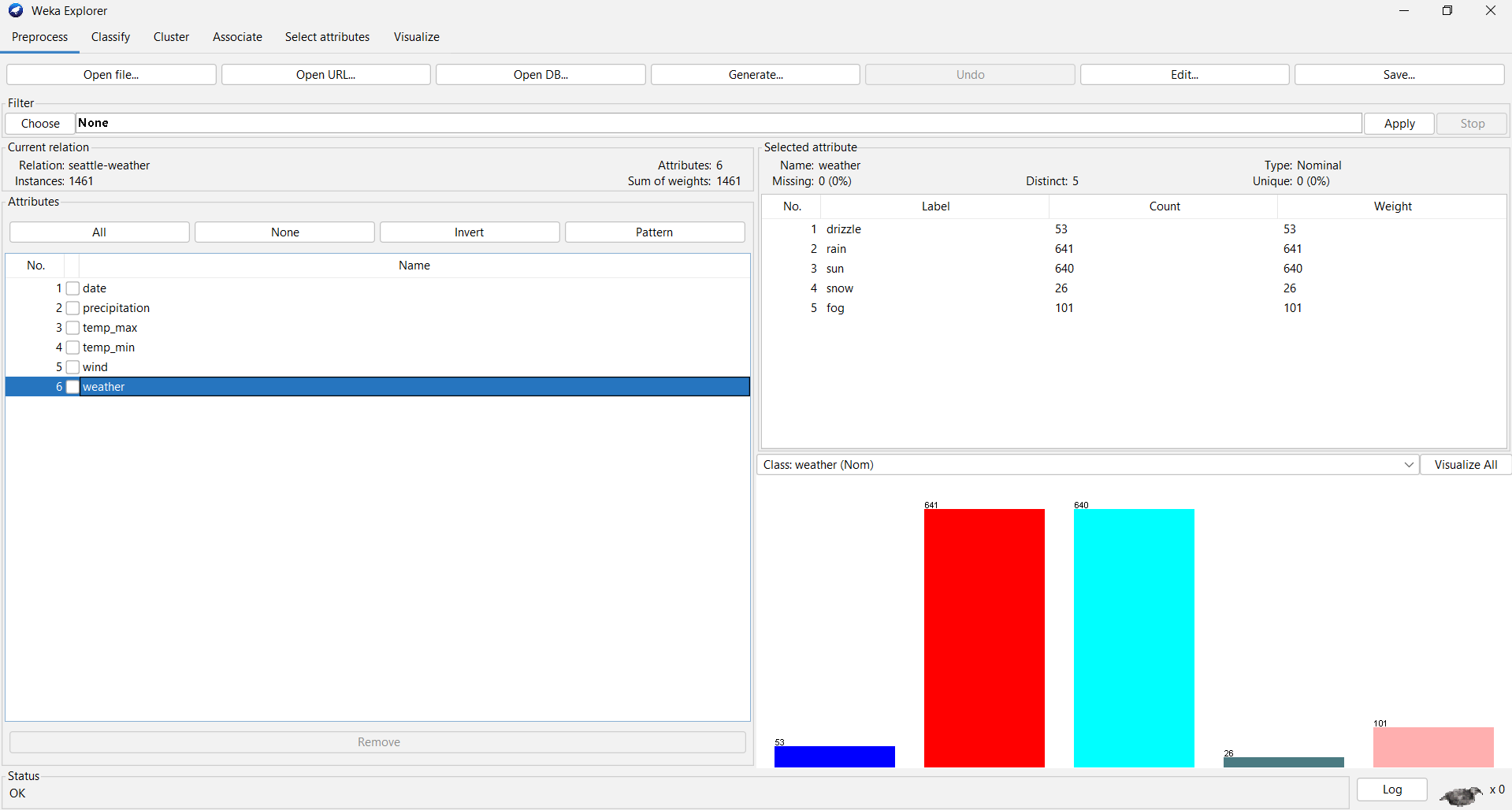
**I. Introduction**: The goal of the task-1 is to predict weather condition of a year based on date to measure precipitation, temperature minimum, temperature maximum and wind. Two different classification models were used to predict the outcome and comparison between the models were done.

Classification in machine learning and statistics is a supervised learning approach in which the computer program learns from the data given to it and makes new observations or classifications. Here, 2 classification models were used - Naïve Bayes and Decision Tree(J48).

The dataset is collected from Kaggle. Here is the link of the dataset: [WEATHER PREDICTION | Kaggle](https://www.kaggle.com/datasets/ananthr1/weather-prediction).

This dataset represents the prediction weather of 1th January 2012 to 31th December 2015.

There is a total of 1461 instances and 6 attributes.

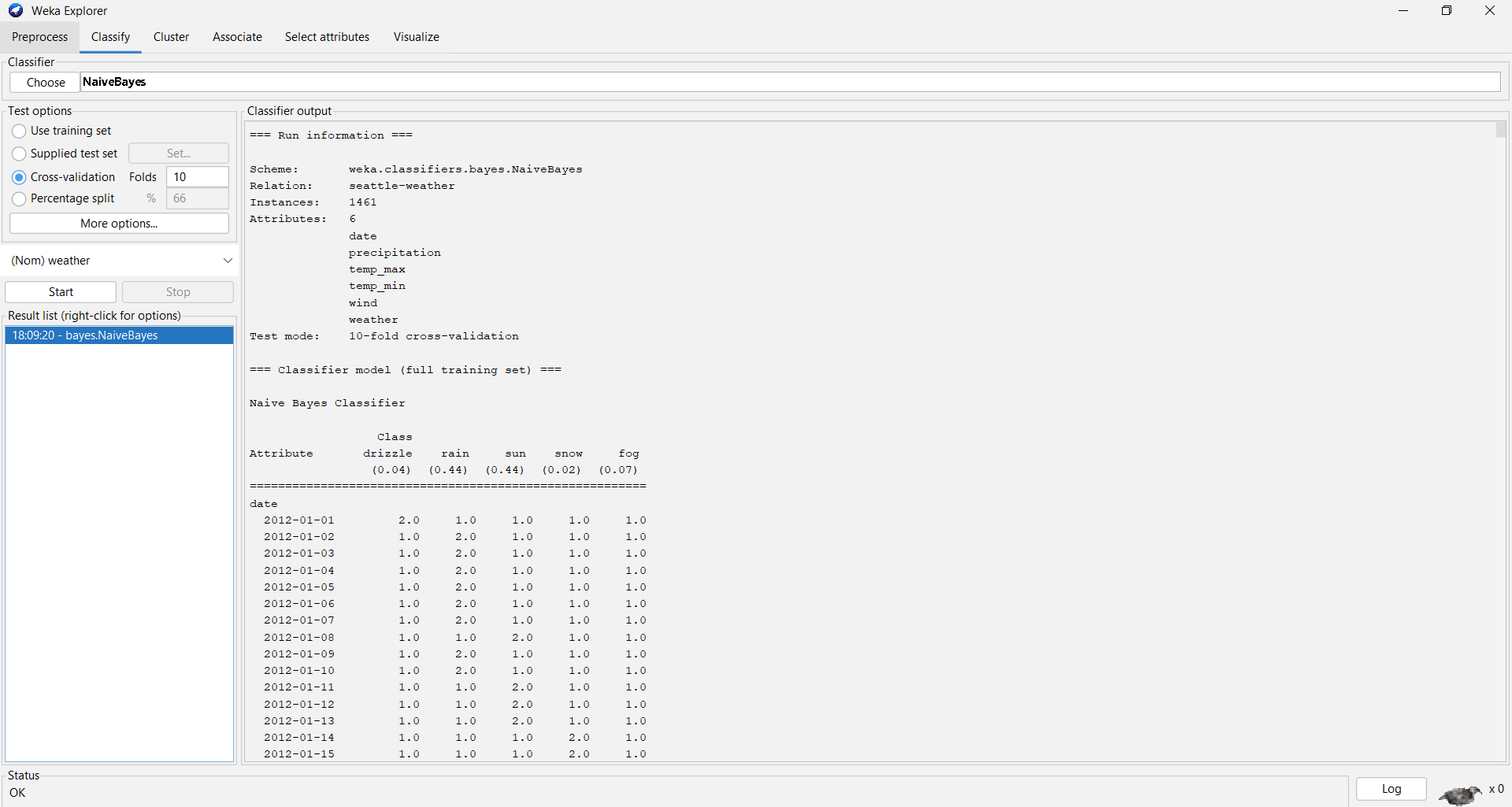


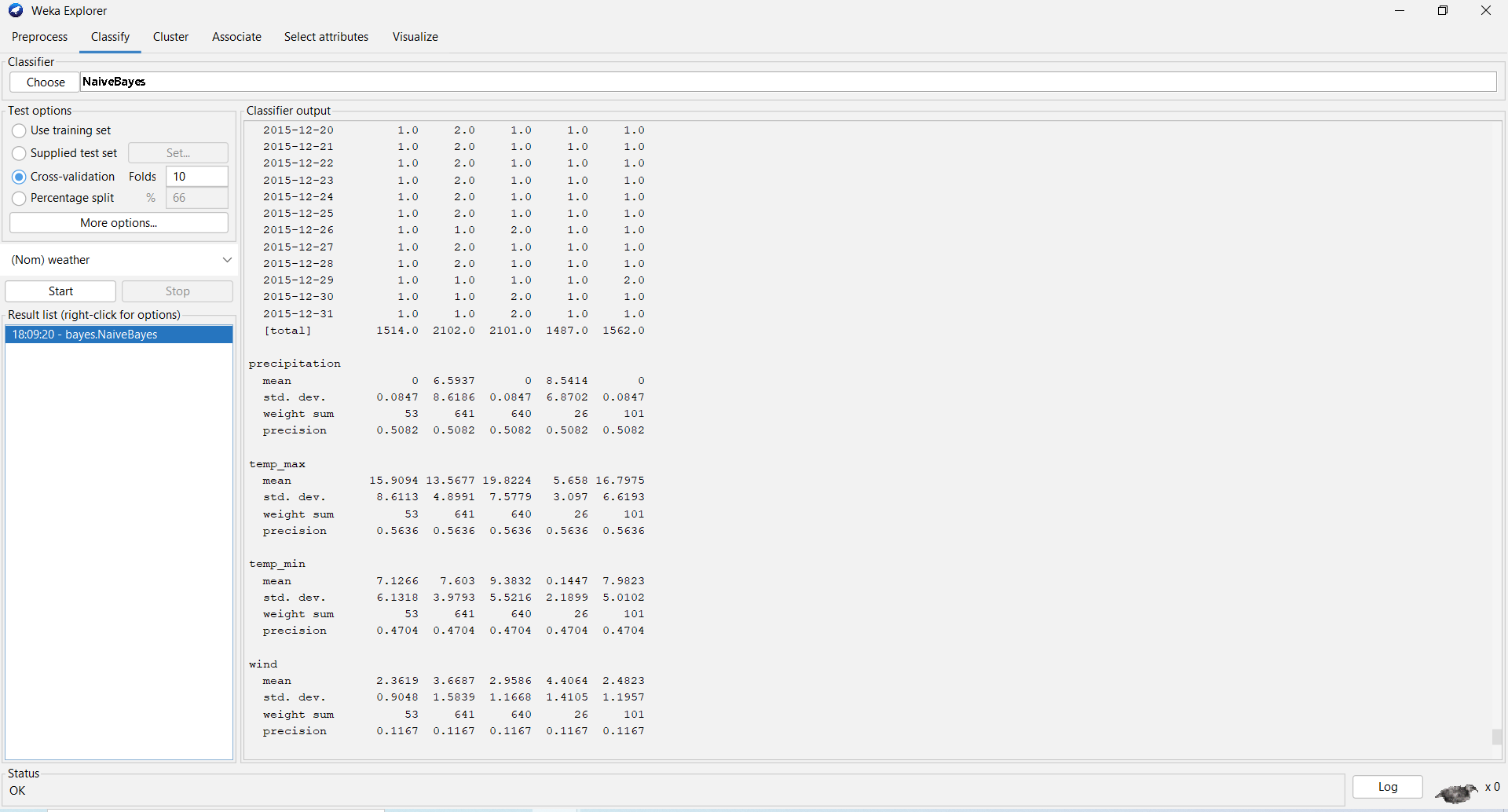
**II. Result:**

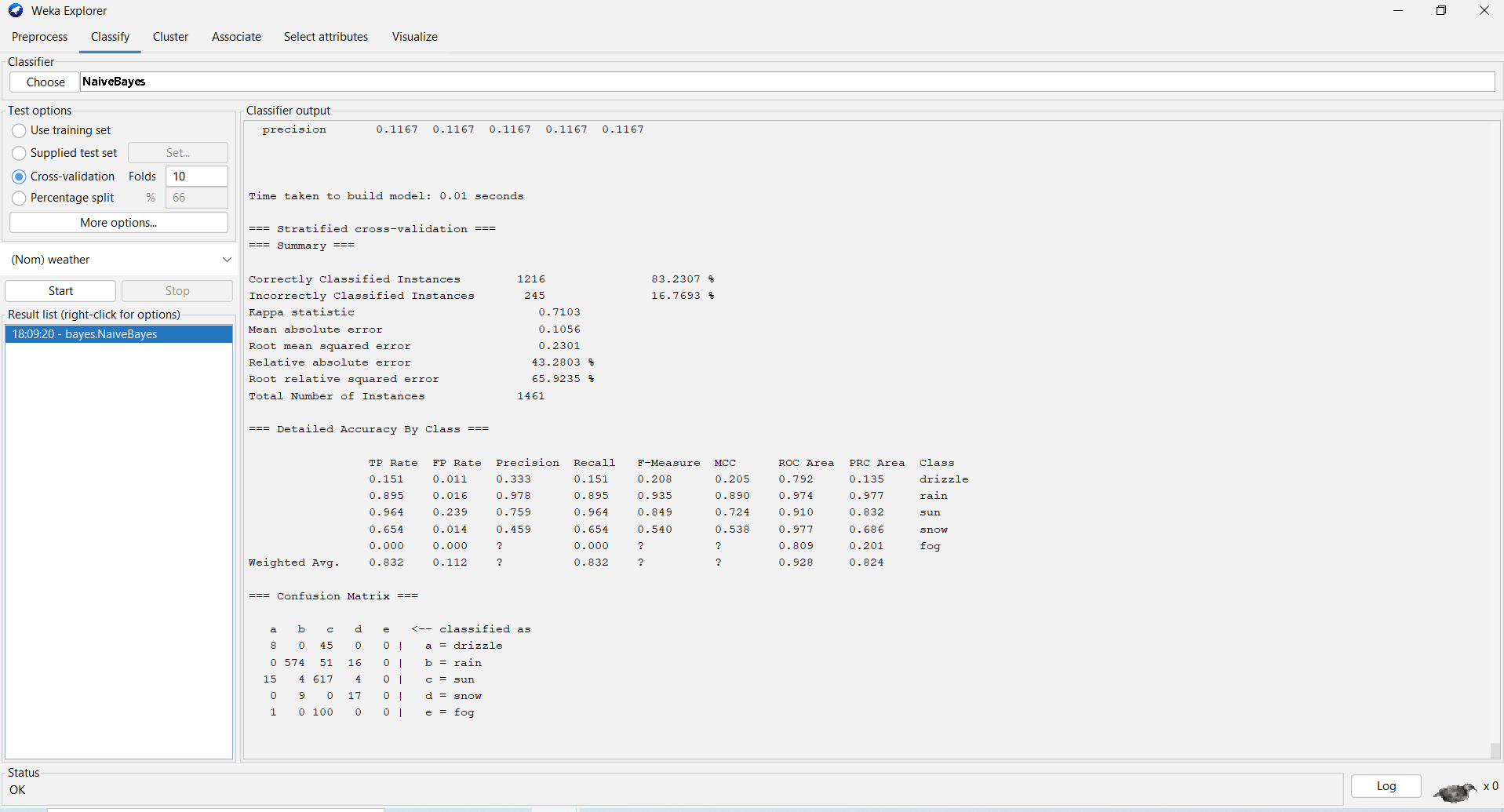
Model Development:

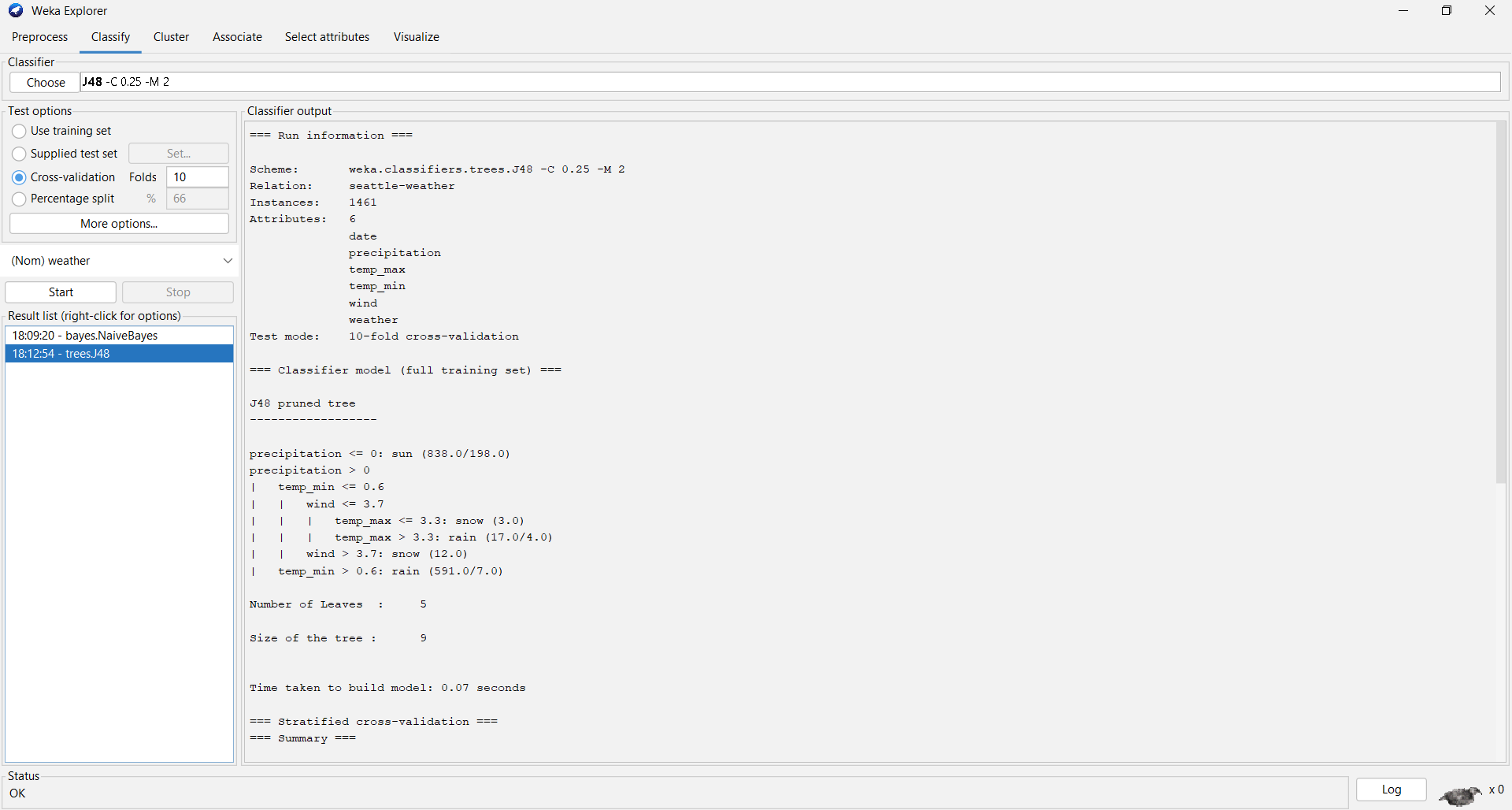
Model used in this project are Naïve bayes and Decision Tree. One type test option was applied on the dataset that is Cross-validation.

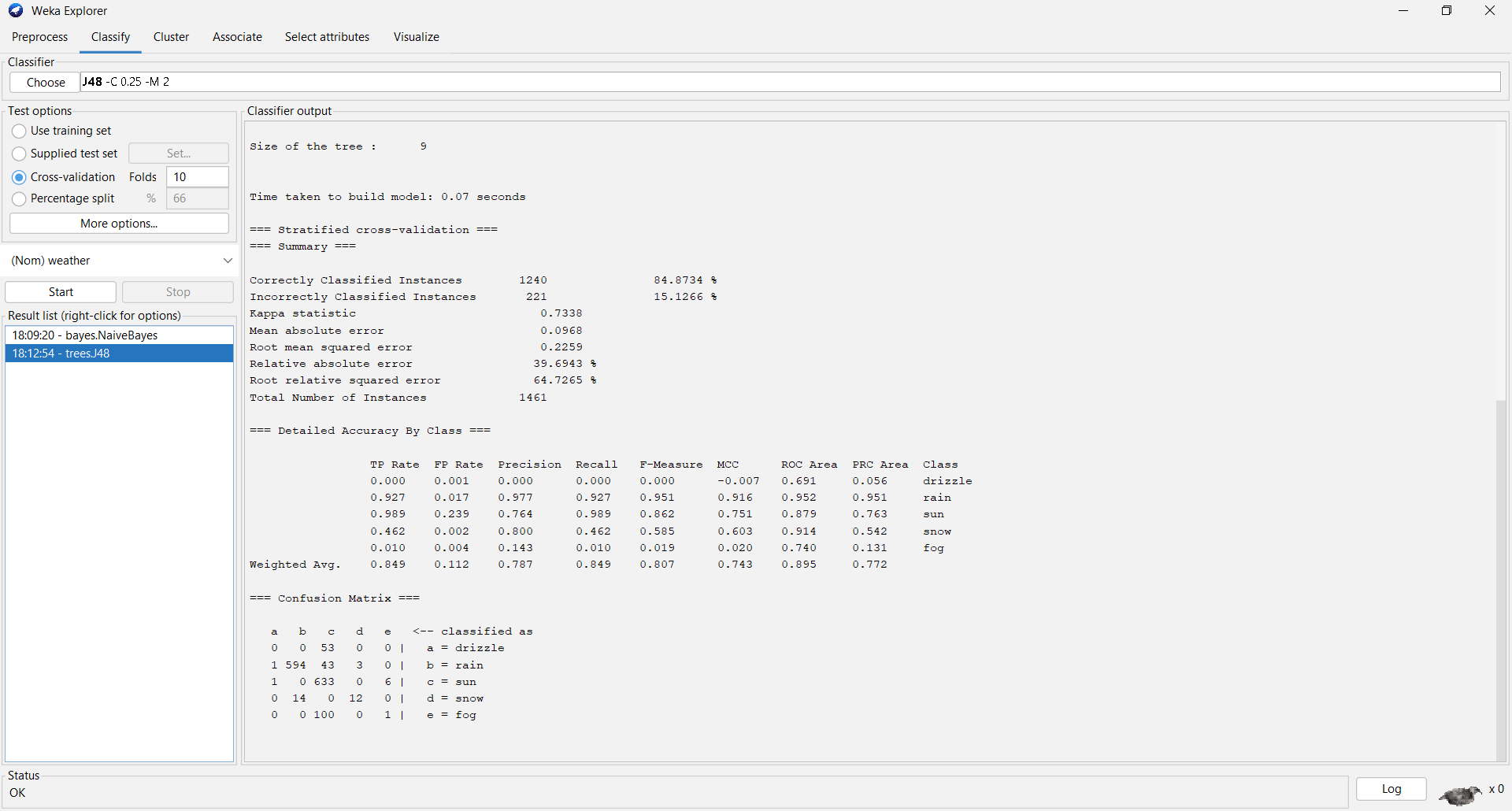
To create these three models, a data analysis program called "WEKA" is used. To begin, the dataset is uploaded and reviewed to ensure that all instances are valid. We chose the classify option since this task is classification-based. “<=50” attribute is selected as the class attribute and for test options two methods were used for all the selected models. One is called "Percentage Split (Train 66%, Test 34%)," while the other is called cross validation (10 folds). After training and testing the dataset, we obtained a statistical report on the model's prediction accuracy.











**III. Discussion:**

This paper presents a study of weather prediction dataset using a total of two classification models (Naïve Bayes and Decision Tree) and comparison was done to find out which model is more efficient. These models are based on classification supervised learning approaches.

For Naïve Bayes correctly classified instances are 1216 and accuracy is 83.2307% and incorrectly classified instances are 245 and accuracy is 16.7693%.

For Decision Tree correctly classified instances are 1240 and accuracy is 84.8734% and incorrectly classified instances are 221 and accuracy is 15.1266%.

So, we can see that Decision Tree is the best.

Task-2

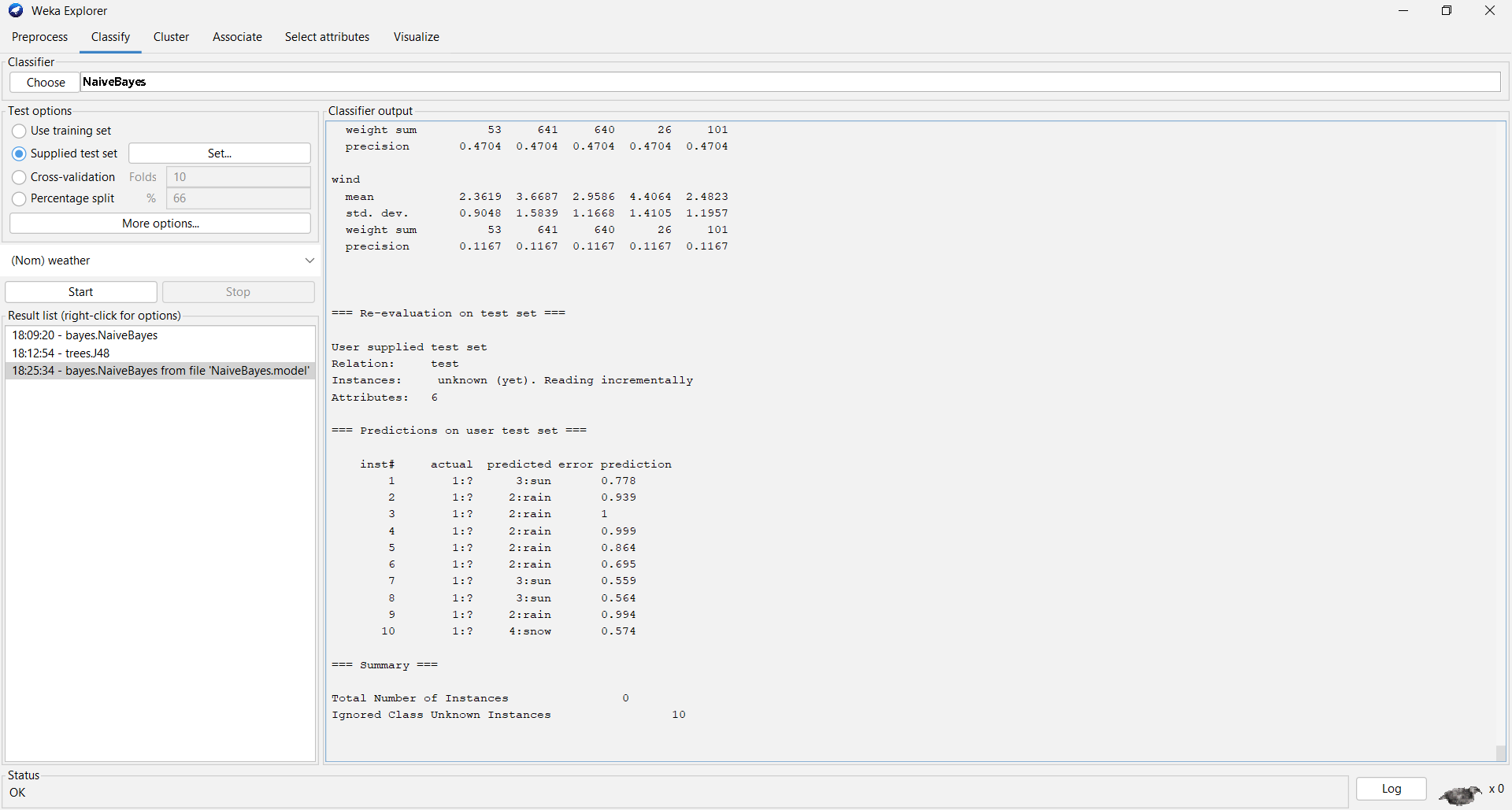
**I: Introduction**: We need to test a test dataset from the Naïve Bayes classification model which we are selected for testing. We selected first 10 instances from the weather dataset for creating the test dataset. The actual test data set is given below:

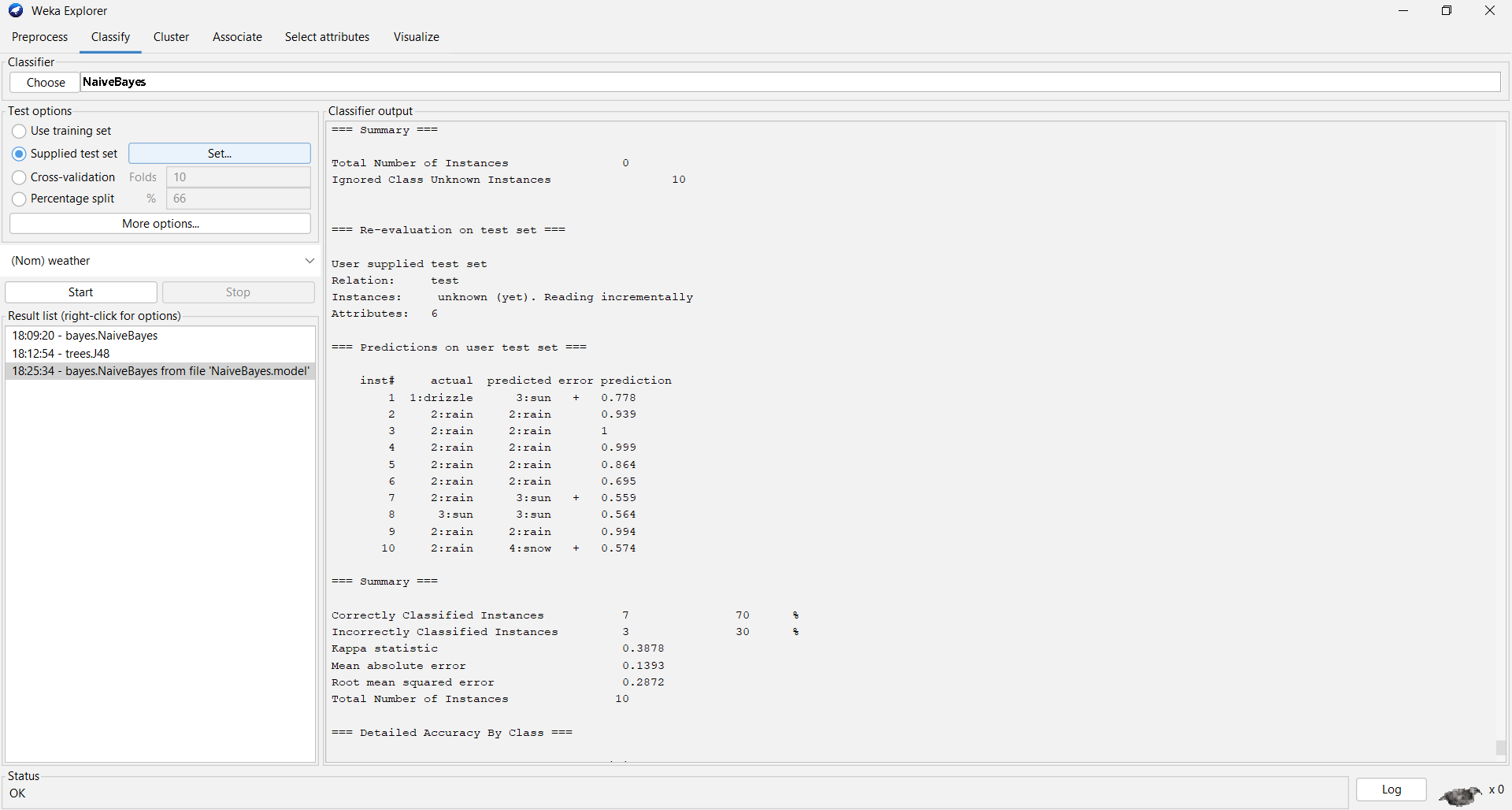
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| date | precipitation | temp\_max | temp\_min | wind | weather |
| 1/1/2012 | 0 | 12.8 | 5 | 4.7 | drizzle |
| 1/2/2012 | 10.9 | 10.6 | 2.8 | 4.5 | rain |
| 1/3/2012 | 0.8 | 11.7 | 7.2 | 2.3 | rain |
| 1/4/2012 | 20.3 | 12.2 | 5.6 | 4.7 | rain |
| 1/5/2012 | 1.3 | 8.9 | 2.8 | 6.1 | rain |
| 1/6/2012 | 2.5 | 4.4 | 2.2 | 2.2 | rain |
| 1/7/2012 | 0 | 7.2 | 2.8 | 2.3 | rain |
| 1/8/2012 | 0 | 10 | 2.8 | 2 | sun |
| 1/9/2012 | 4.3 | 9.4 | 5 | 3.4 | rain |
| 1/10/2012 | 1 | 6.1 | 0.6 | 3.4 | rain |

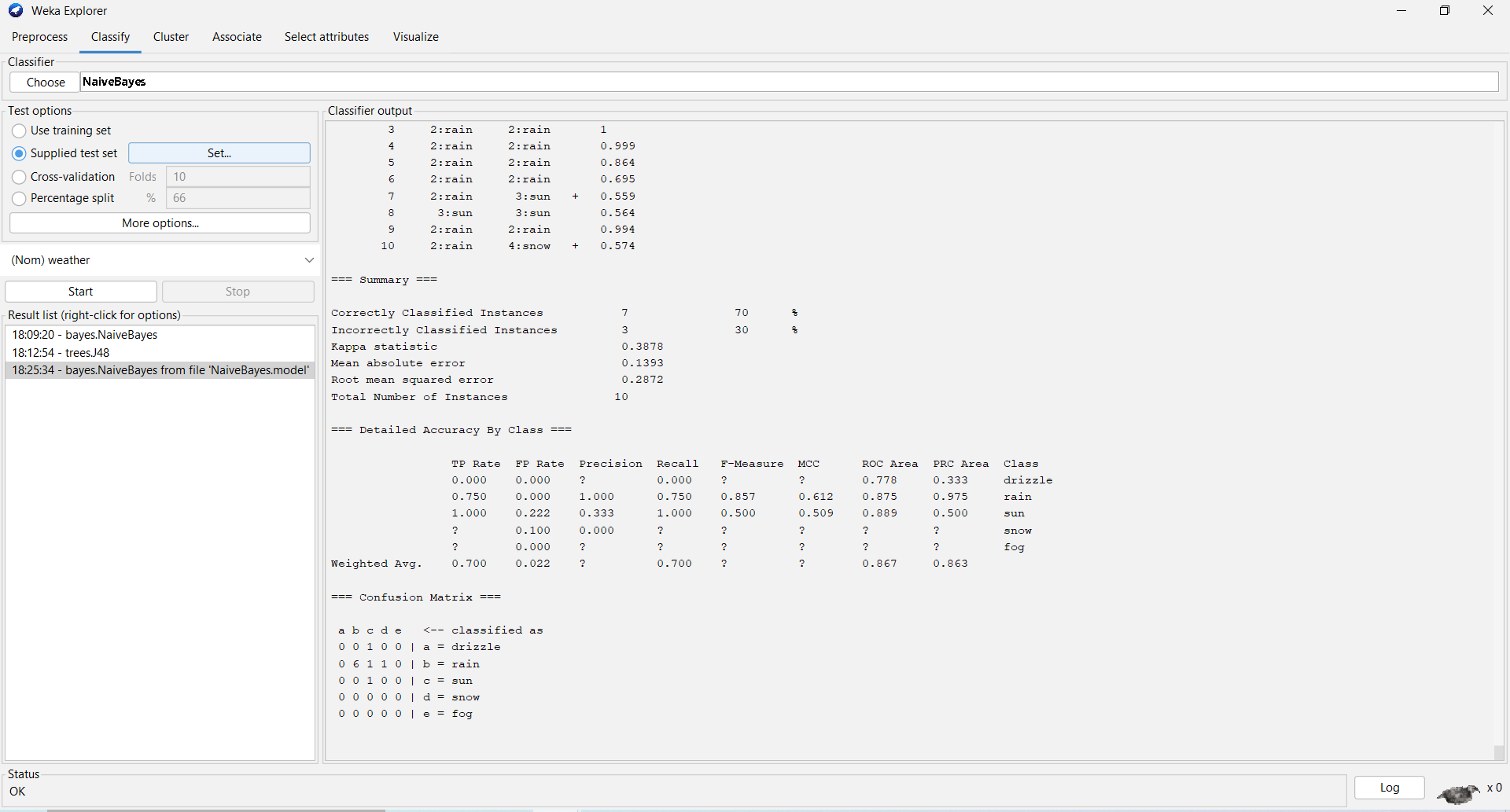
Now we need to test this data set using Naïve Bayes classification model.

**II. Result:**

Model Development:







**III. Discussion:**

We can see that correctly classified instances are 7 and accuracy is 70% and incorrectly classified instances are 3 and accuracy is 30%. Instances number 1, 7 and 10 are incorrectly classified.

Task-3

**I. Introduction:** We made a model for the price of stock market with the available independent variables. It will be used by the investors to understand prediction how the stock prices vary with the independent variables. Many investors are interested in knowing the stock market's future scenario in order to make a smart and successful investment. By giving helpful information like the stock market's future direction, good and successful stock market prediction systems assist traders, investors, and analysts. This model will be a good way for traders, investors, and analysts to understand the pricing of NETFLIX stock market.

Number of Instances: 1009

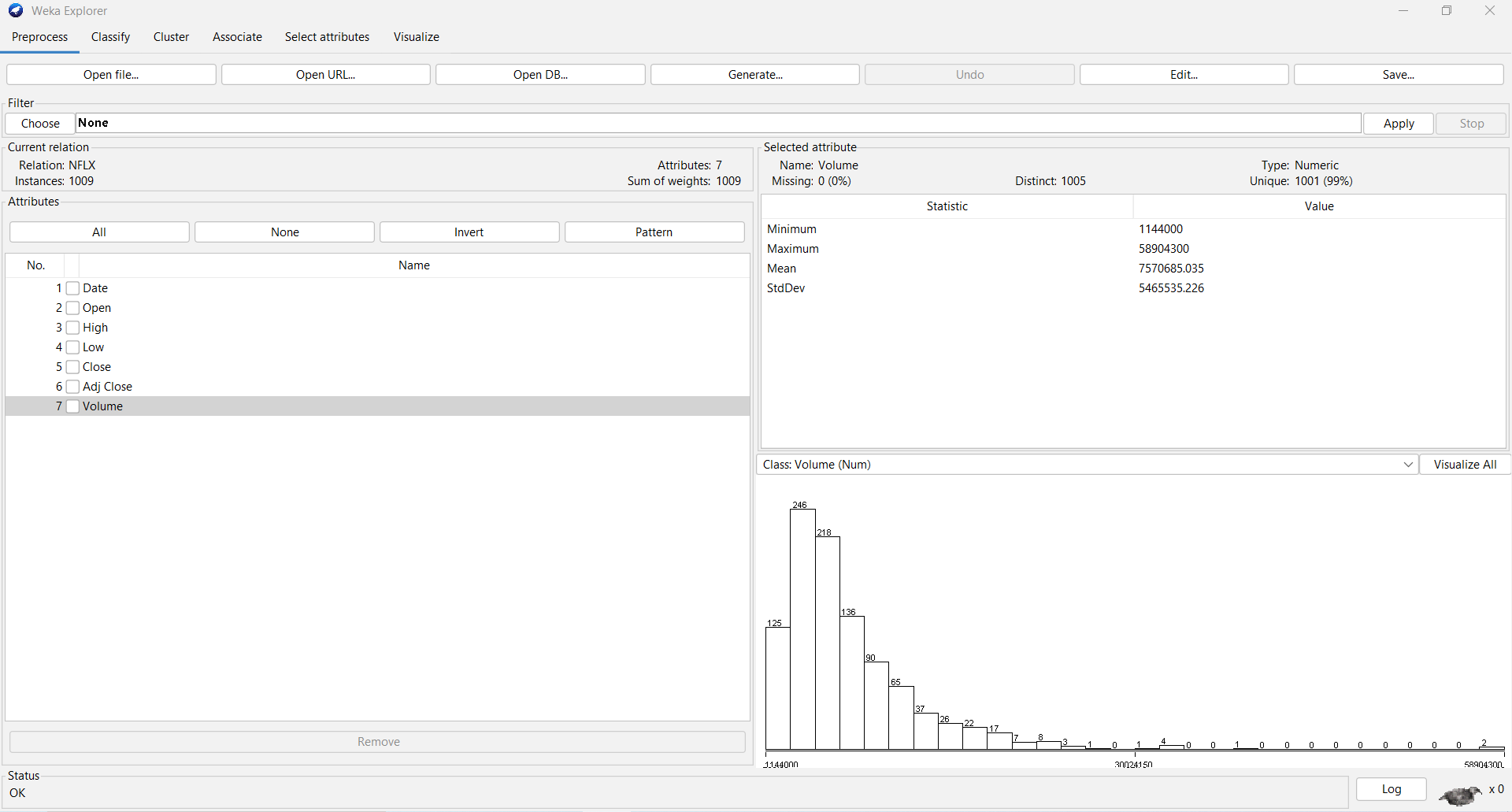
Number of Attributes: 7 numeric predictive.

Attribute Information (in order):

* Date: Everyday price.
* Open: Price at which stock open.
* High: Today’s high.
* Low: Today’s low.
* Close: Close price adjusted for splits.
* Adj close: Adjusted close price adjusted for splits and dividend and/or capital gain distributions.
* Volume: Volume of stocks.

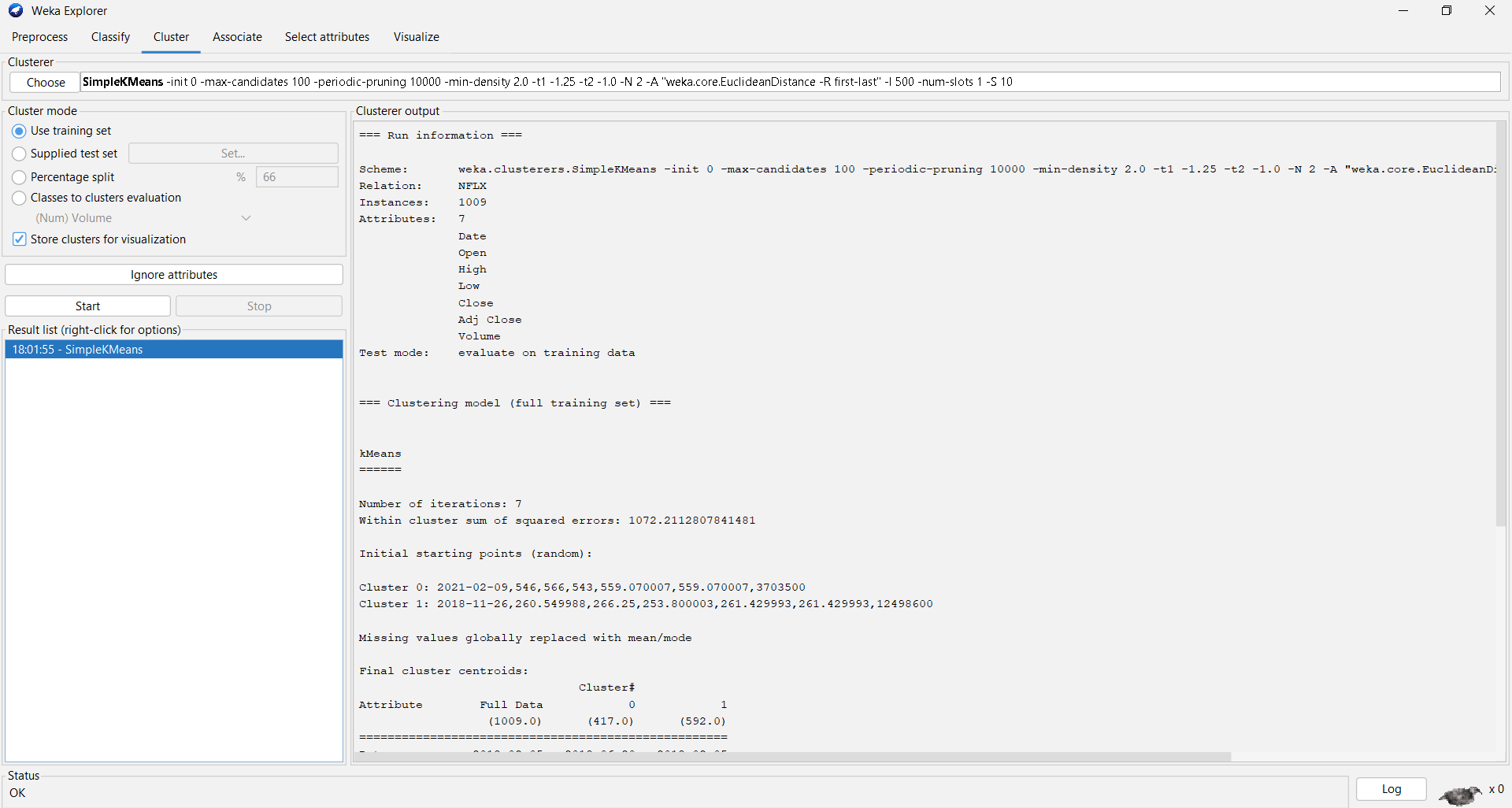
This dataset is taken from Kaggle.com. The name of the dataset is Netflix stock price prediction. The Dataset contains data for 5 years (from 5th Feb 2018 to 4th Feb 2022)

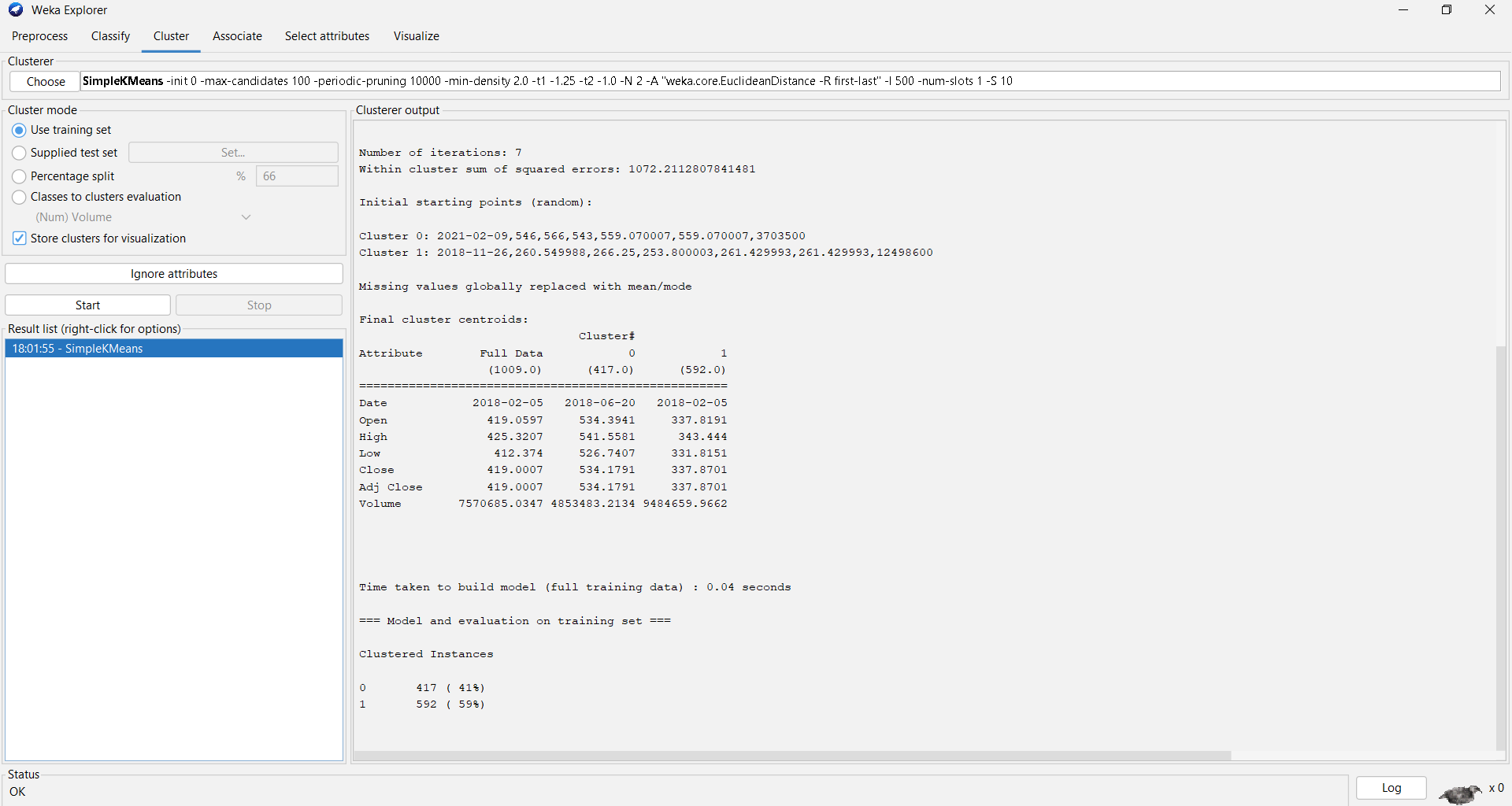
This is a copy of Netflix stock price prediction dataset:  [Netflix Stock Price Prediction | Kaggle](https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction)



II. Result:

Model Development:





**III. Discussion:** K means clustering is a simple cluster analysis method. The number of clusters can be set using the setting tab. We set numClusters 2. The centroid of each cluster is calculated as the mean of all points within the clusters. After applying Simple KMeans clustering We can see that using number of clusters two we get two centroids. Number of iterations are seven. Number of squared errors are 1072. Cluster Instances are 41% & 59%. With the increase in the number of clusters, the sum of square errors is reduced. The objects within the cluster exhibit similar characteristics and properties. The clusters represent the class labels. We visualized the clustering.