Index

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Experiment** | **Page** |
| 1 | Analysis of Data, know the type of data- nominal, ordinal, ratio, interval |  |
| 2 | Find the mean, median, variance and standard deviation of data |  |
| 3 | Visualize your dataset in Weka and learn about Jitter |  |
| 4 | Calculate co-relation among attributes |  |
| 5 | Calculate Information Gain of Different attributes |  |
| 6 | Perform PCA on a dataset |  |
| 7 | Train a Decision Tree |  |
| 8 | Train a Naïve Bayes Classifier |  |
| 9 | Train a Bayesian Belief Network Classifier |  |
| 10 | Train a Linear Regression Model |  |
| 11 | Train a Logistic Regression Model |  |
| 12 | Train a KNN model |  |
| 13 | Train a K Means Clustering Model |  |

EXPERIMENT-1

Analysis of Data, know the type of data- nominal, ordinal, ratio, interval

**Nominal**

A nominal scale describes a variable with categories that do not have a natural order or ranking. You can code nominal variables with numbers if you want, but the order is arbitrary and any calculations, such as computing a mean, median, or standard deviation, would be meaningless.

Examples of nominal variables include:

* genotype, blood type, zip code, gender, race, eye color, political party

**Ordinal**

An ordinal scale is one where the order matters but not the difference between values.

Examples of ordinal variables include:

* socio economic status (“low income”,”middle income”,”high income”), education level (“high school”,”BS”,”MS”,”PhD”), income level (“less than 50K”, “50K-100K”, “over 100K”), satisfaction rating (“extremely dislike”, “dislike”, “neutral”, “like”, “extremely like”).

## Interval

An interval scale is one where there is order and the difference between two values is meaningful.

Examples of interval variables include:

* temperature (Farenheit), temperature (Celcius), pH, SAT score (200-800), credit score (300-850).

## Ratio

A ratio variable, has all the properties of an interval variable, and also has a clear definition of 0.0. When the variable equals 0.0, there is none of that variable.

Examples of ratio variables include:

* enzyme activity, dose amount, reaction rate, flow rate, concentration, pulse, weight, length, temperature in Kelvin (0.0 Kelvin really does mean “no heat”), survival time.

EXPERIMENT-2

Find the mean, median, variance and standard deviation of data

#include <bits/stdc++.h>

using namespace std;

//CALCULATE MEAN

double meancal(double \*arr, int n)

{

    double sum=accumulate(arr, arr+n, 0.0);

    return sum/n;

}

//CALCULATE MEDIAN

double median(double \*arr, int n)

{

    if(n%2==1)

    return (double)(arr[n/2]);

    return (double)((arr[n/2]+arr[n/2-1])/2);

}

//CALCULATE VARIANCE

double variance(double \*arr, int n, double mean)

{

    double t=0.0;

    for(int i=0; i<n; i++)

    {

        double temp=abs(arr[i]-mean);

        t+=temp\*temp;

    }

    if(n>50)

    return t/(n-1);

    else

    return t/n;

}

int main() {

    cout<<"Enter size of data: ";

    int n;

    cin>>n;

    cout<<"\nEnter data: ";

    double arr[n];

    for(int i=0; i<n; i++)

    {

        cin>>arr[i];

    }

    sort(arr, arr+n);

    cout<<endl<<"Entered data after sorting is:\n";

    for(int i=0; i<n; i++)

    cout<<arr[i]<<" ";

    cout<<endl;

    double mean;

    //mean calculation

    mean=meancal(arr, n);

    cout<<"\nMean of dataset is: "<<mean<<endl;

    //median calculation

    cout<<"Median of dataset is: "<<median(arr, n)<<endl;

    //variance

    double var=variance(arr, n, mean);

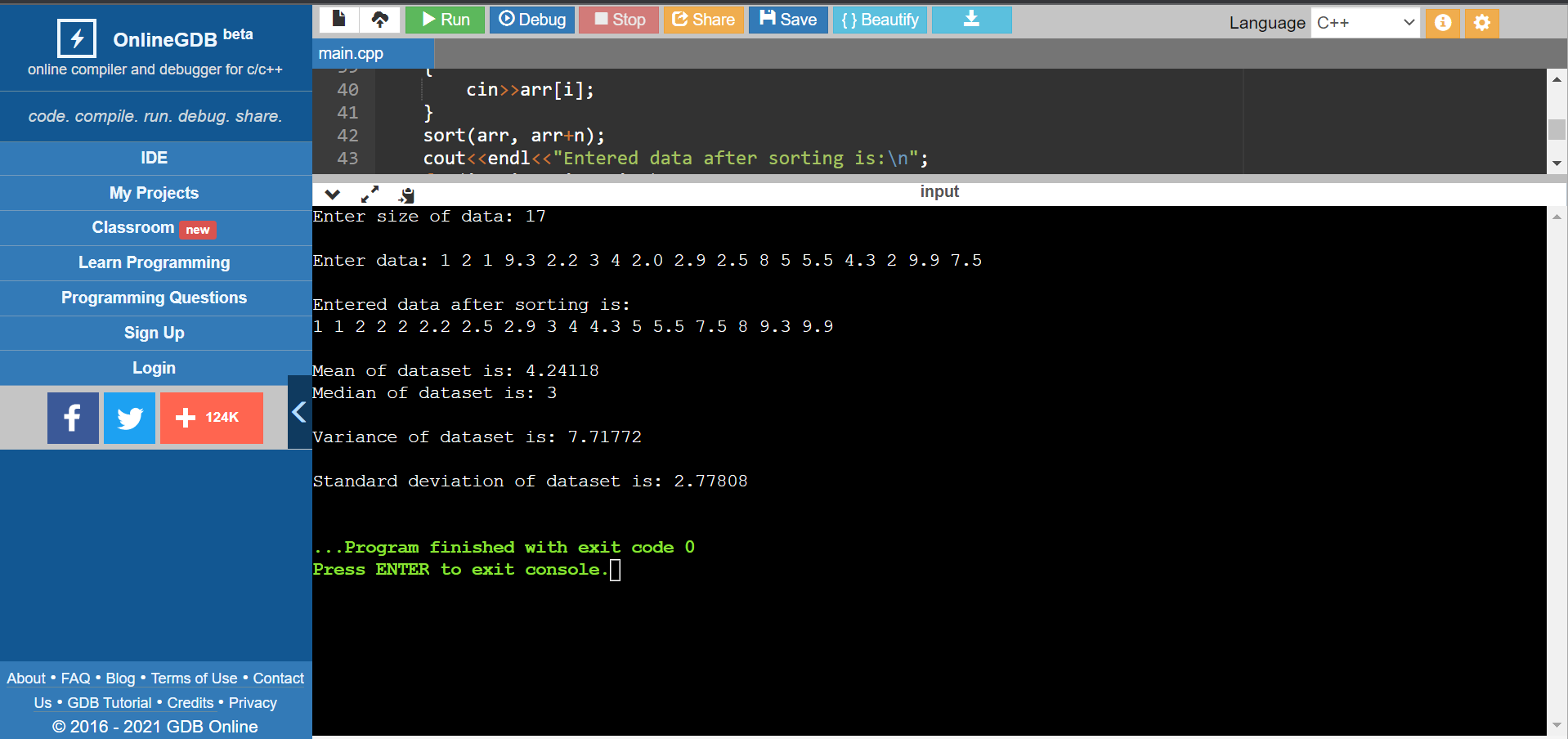
    cout<<"\nVariance of dataset is: "<<var<<endl;

    //standard deviation

    cout<<"\nStandard deviation of dataset is: "<<sqrt(var)<<endl;

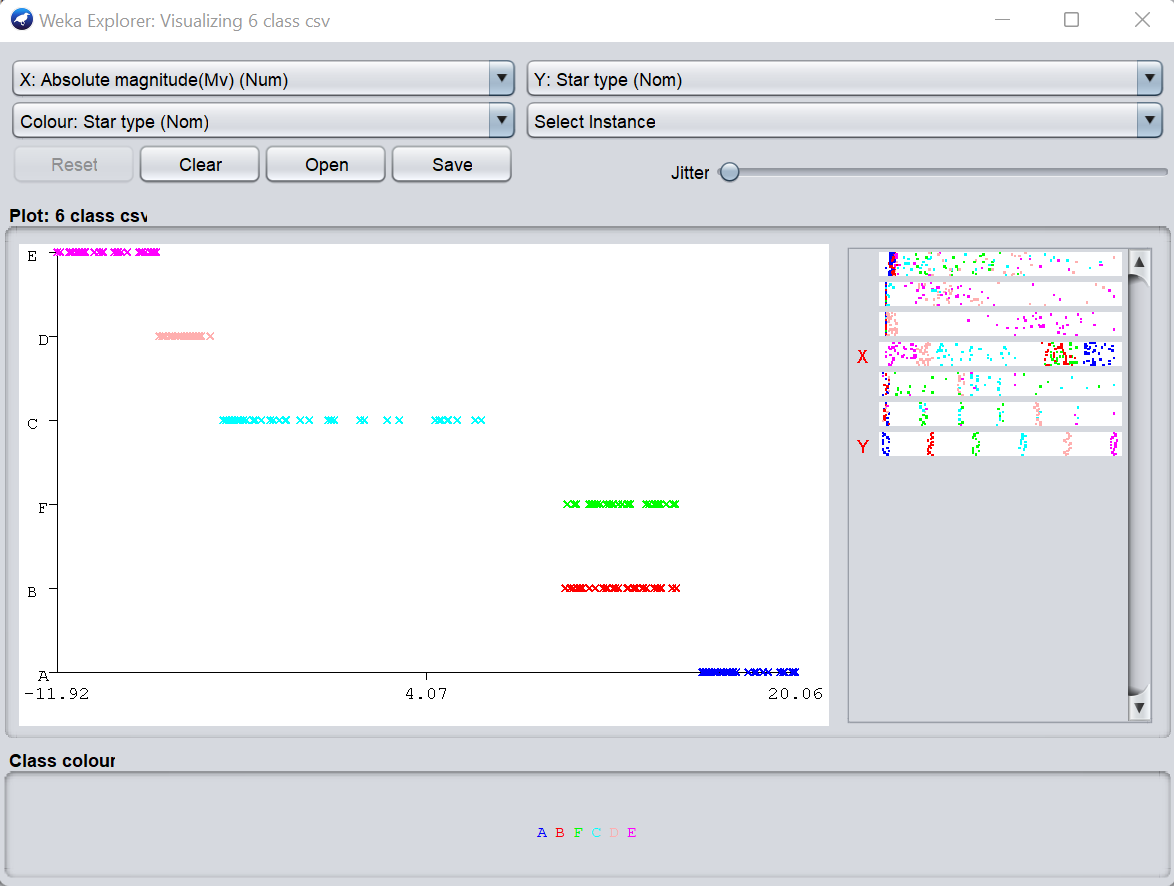
    return 0;

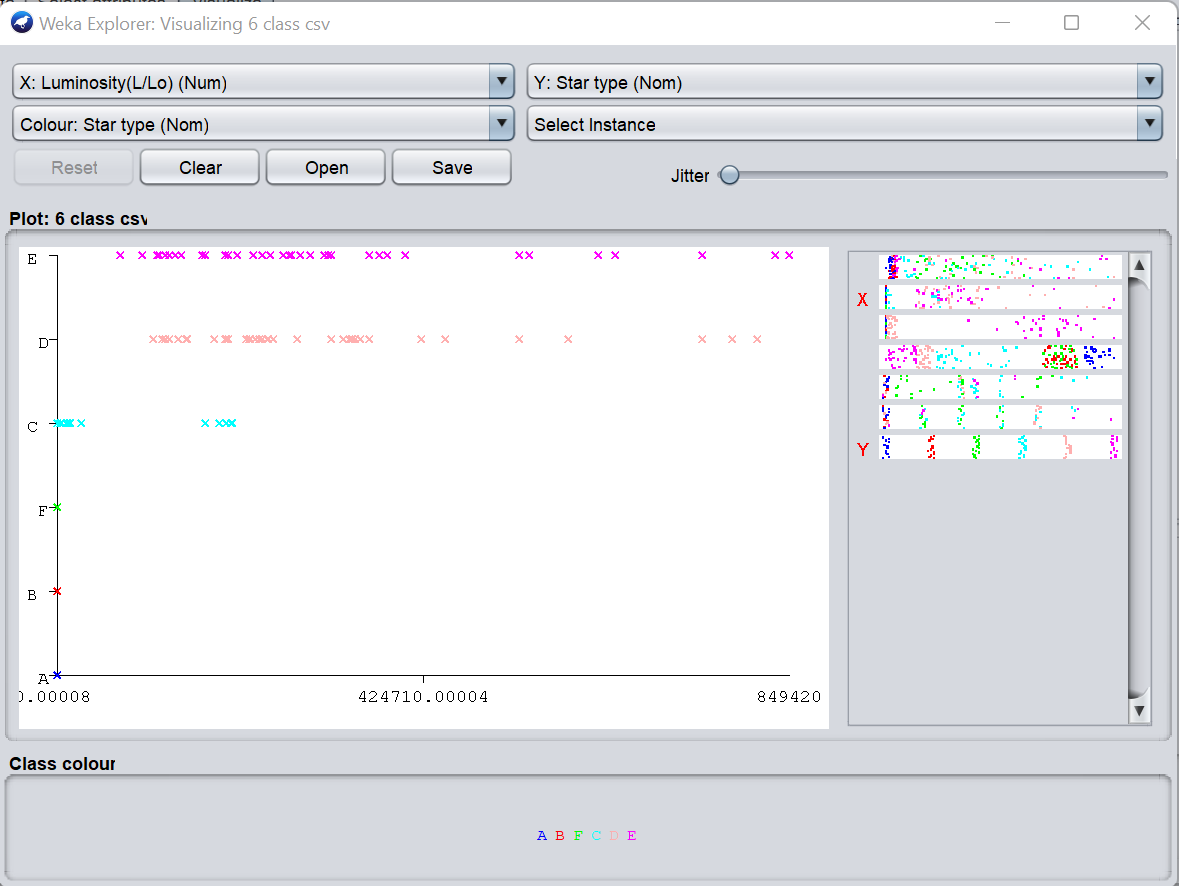
}



EXPERIMENT 3

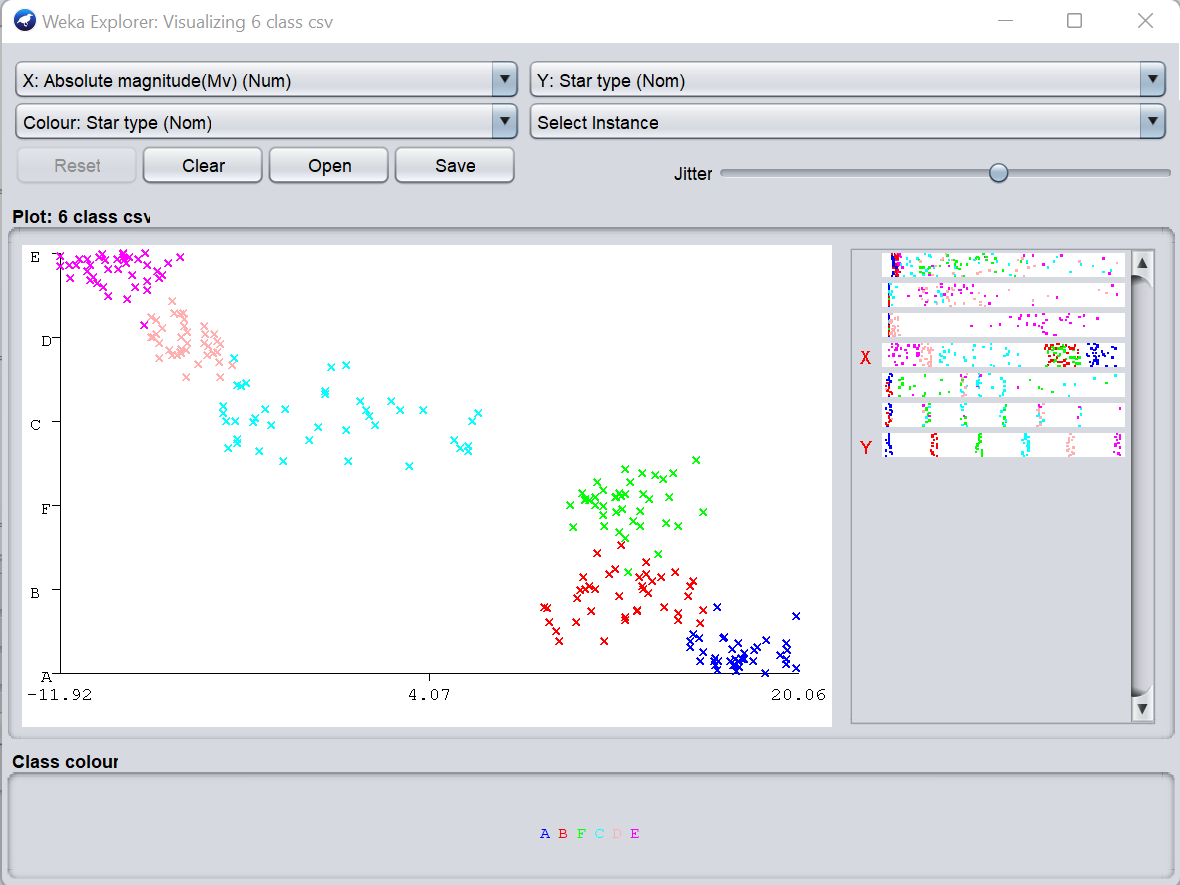
Visualize your dataset in Weka and learn about Jitter





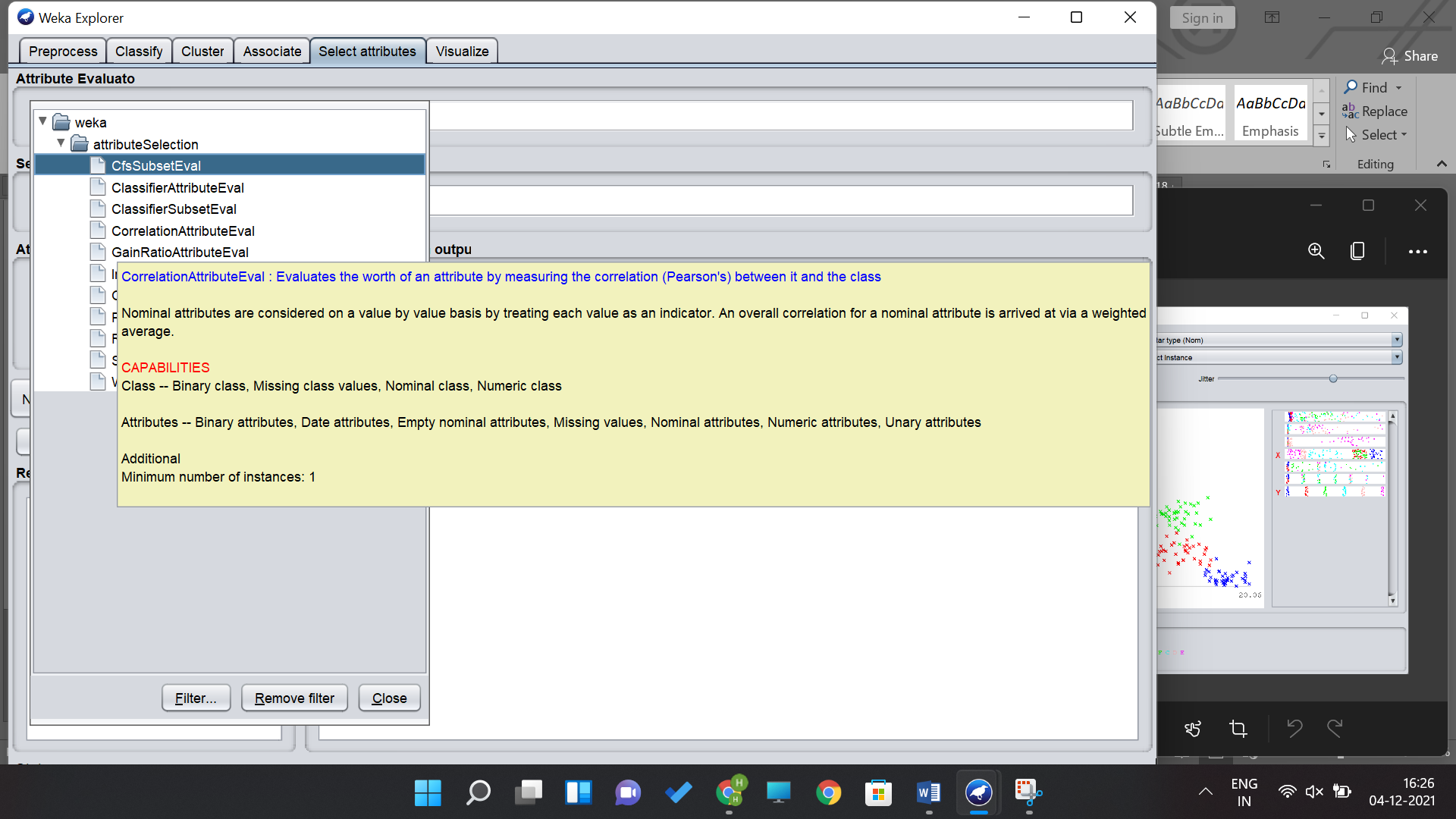
Using Jitter

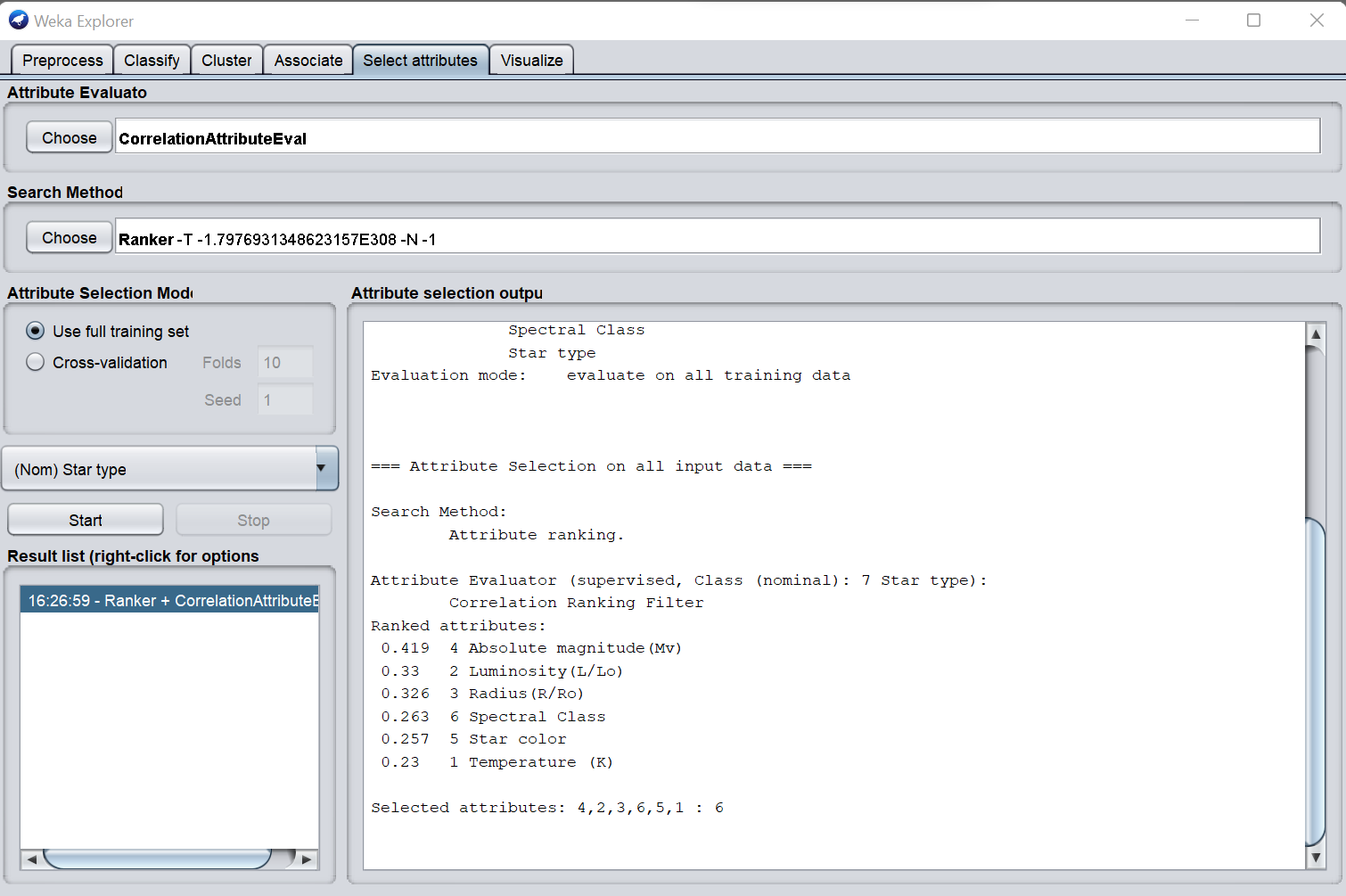
Jitter is used to introduce random noise to the dataset so as to classify datapoints (separate them)



EXPERIMENT-4

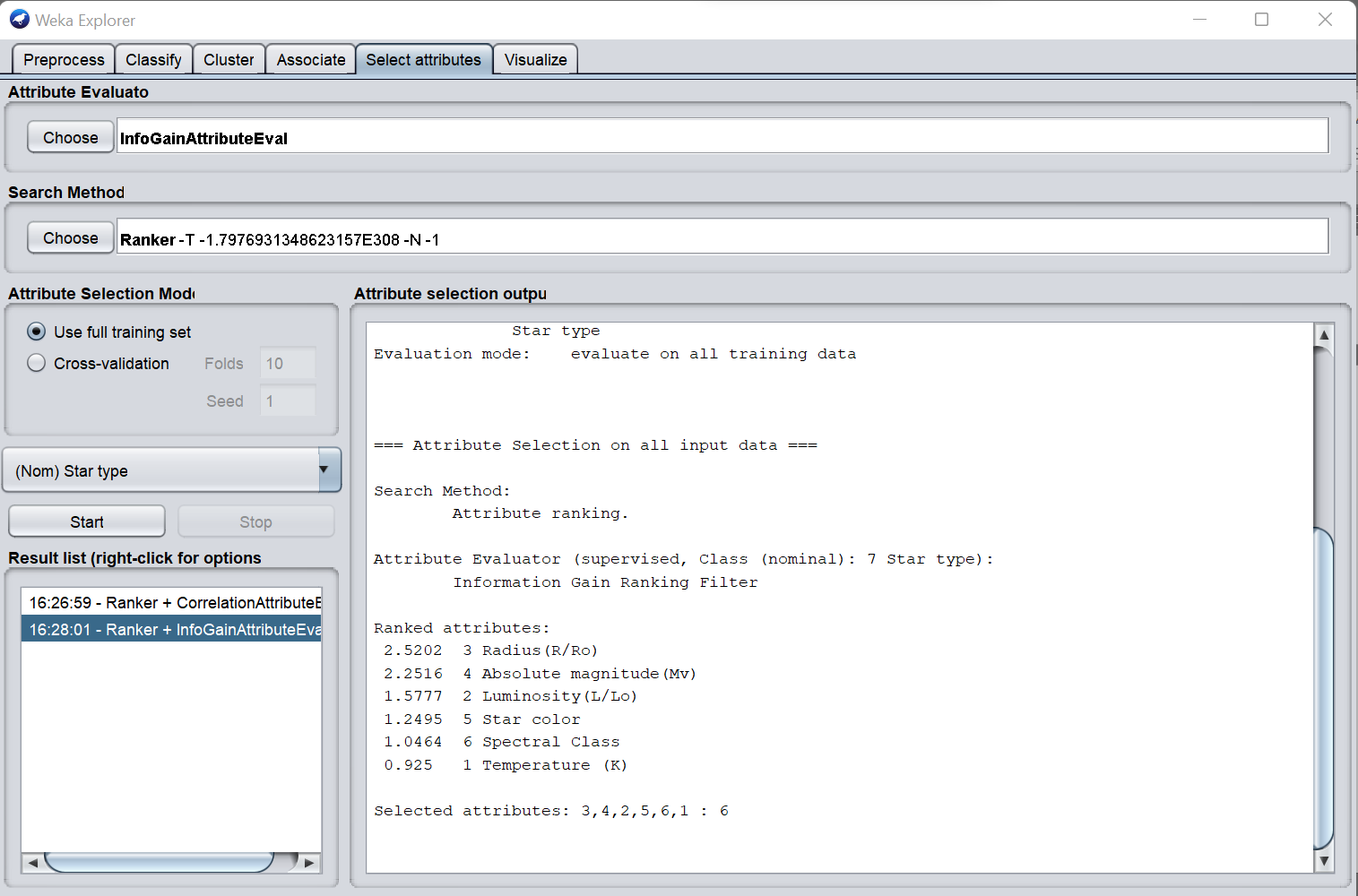
Calculate co-relation among attributes





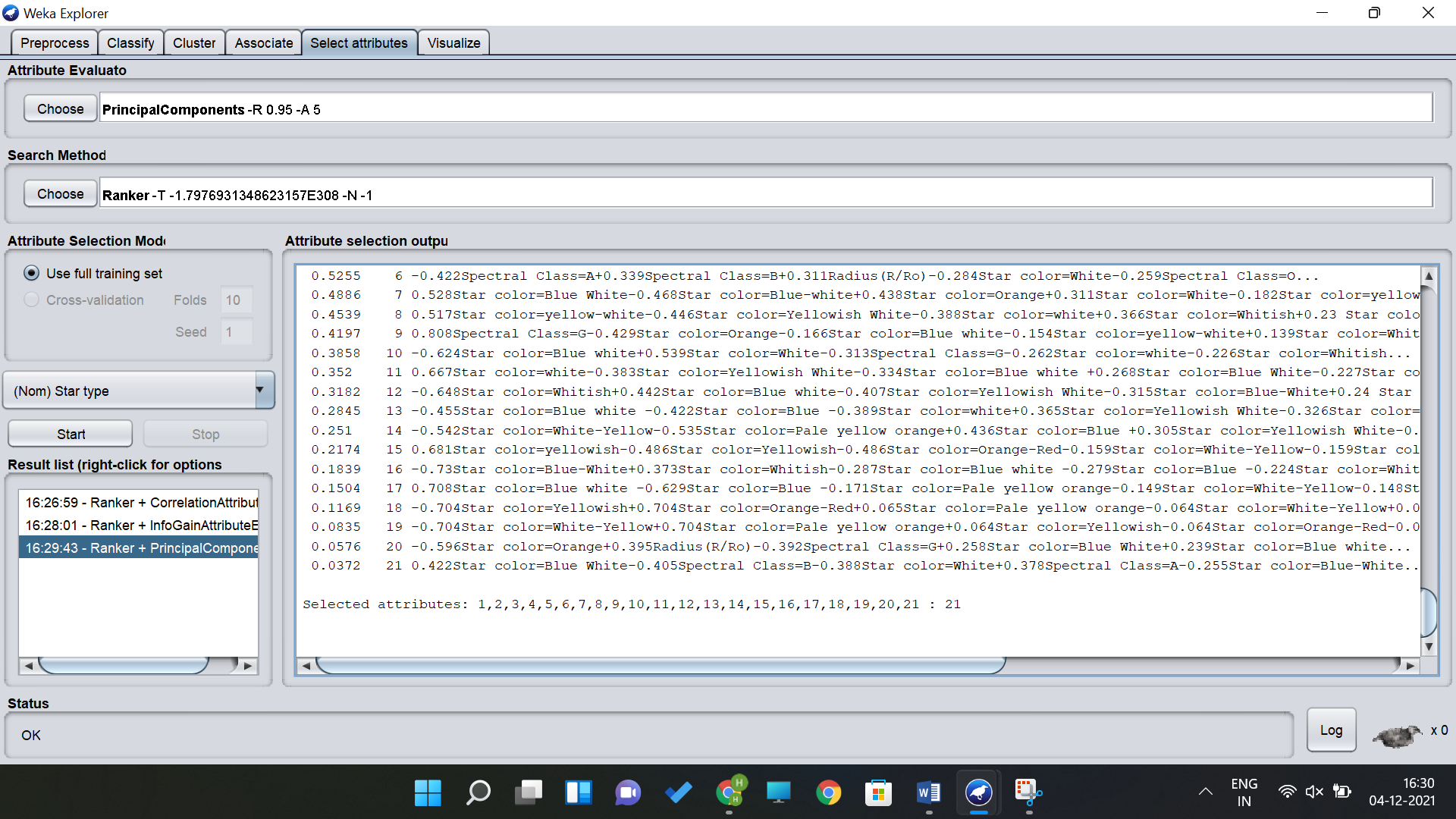
EXPERIMENT-5

Calculate Information Gain of Different attributes



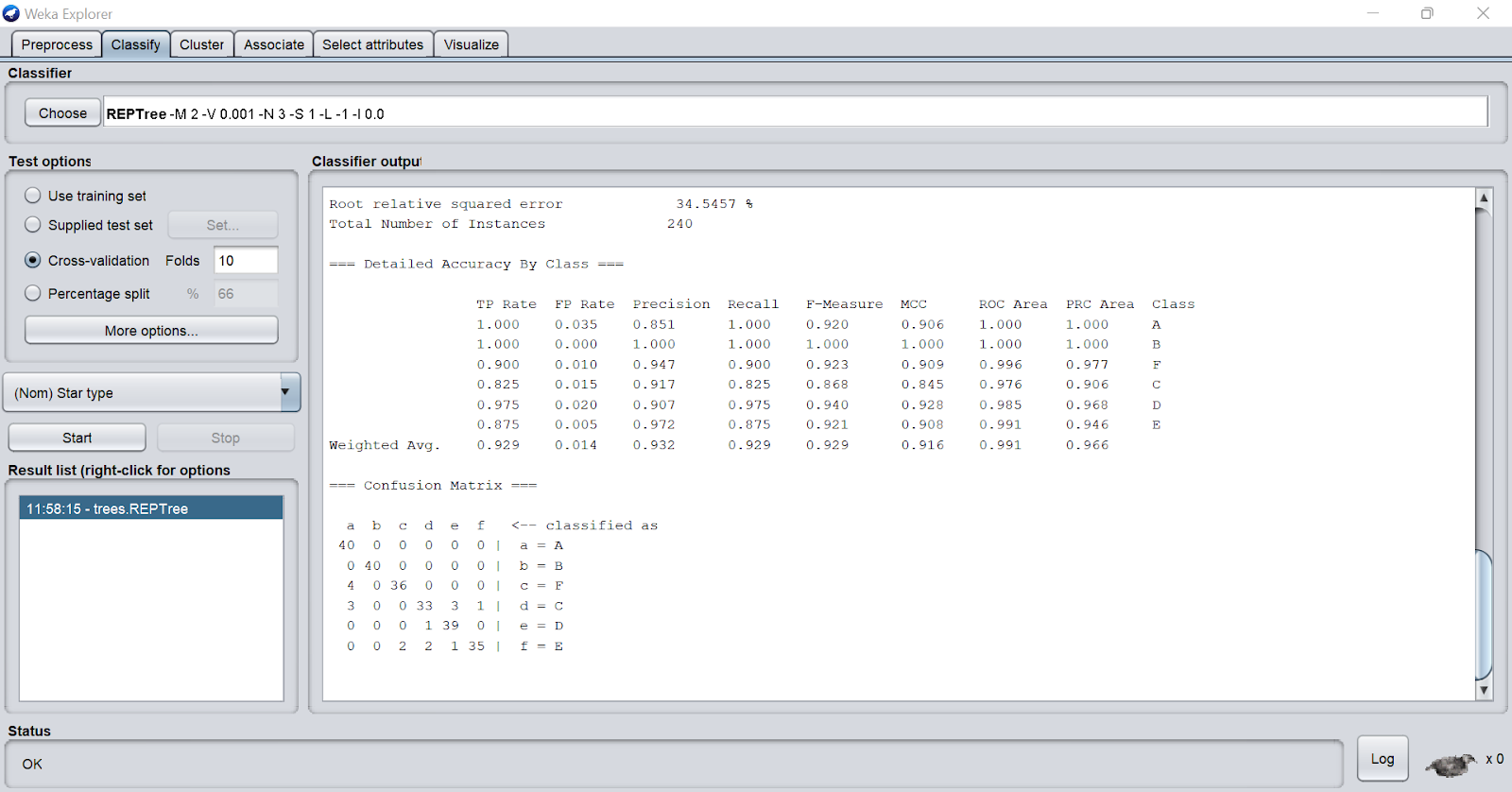
EXPERIMENT-6

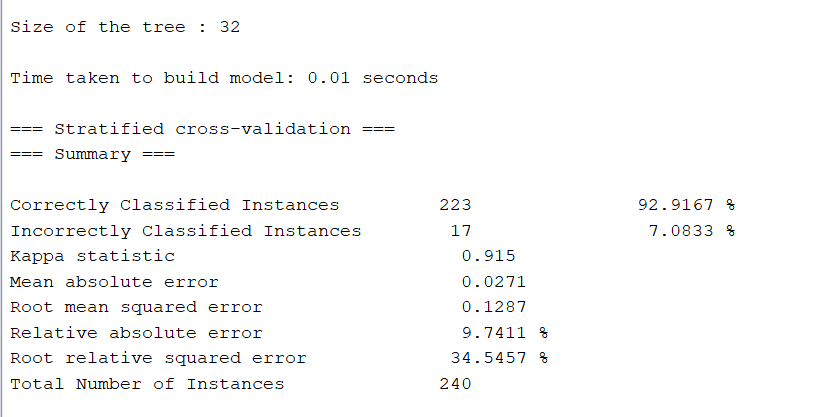
Perform PCA on a dataset

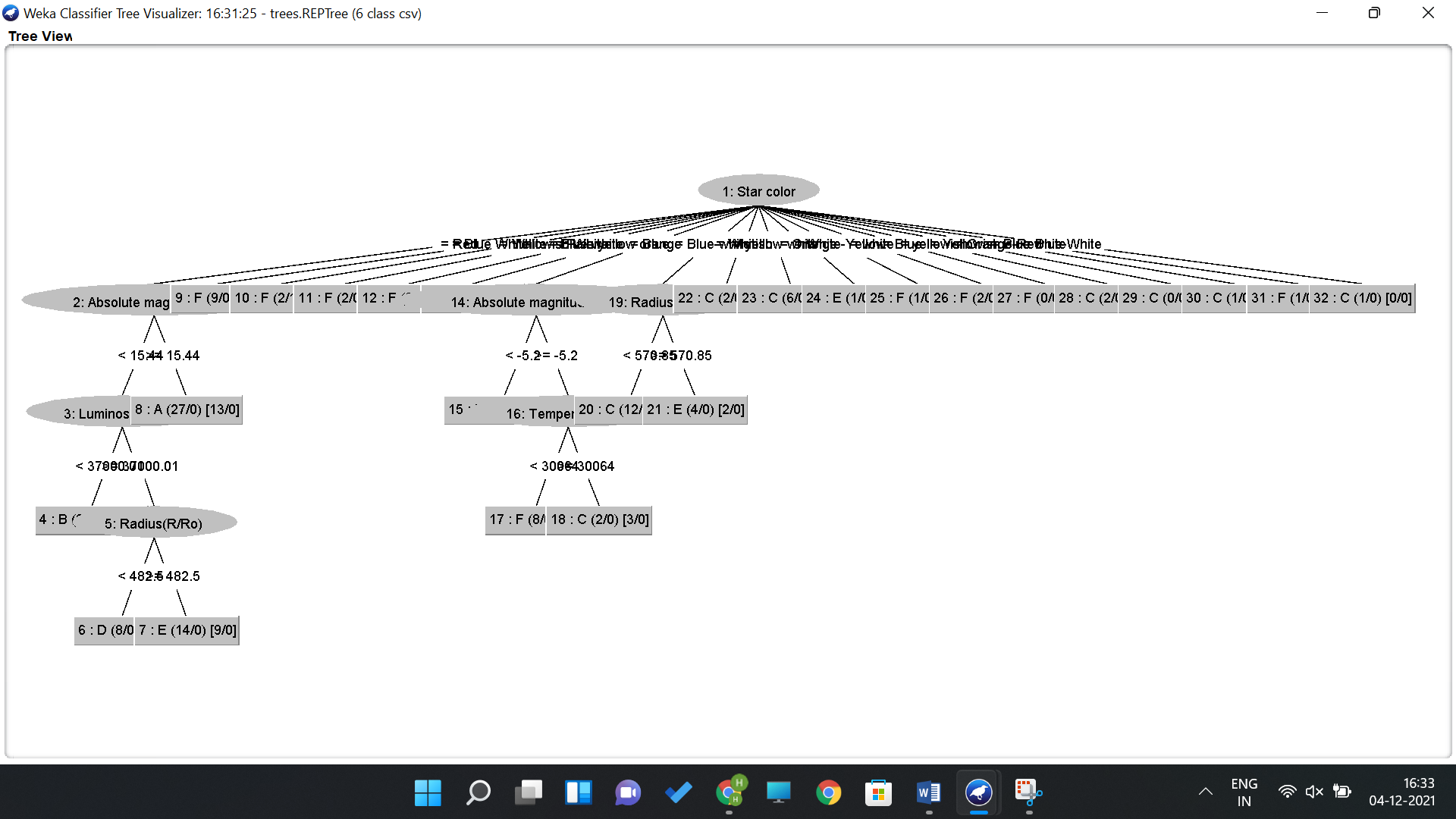


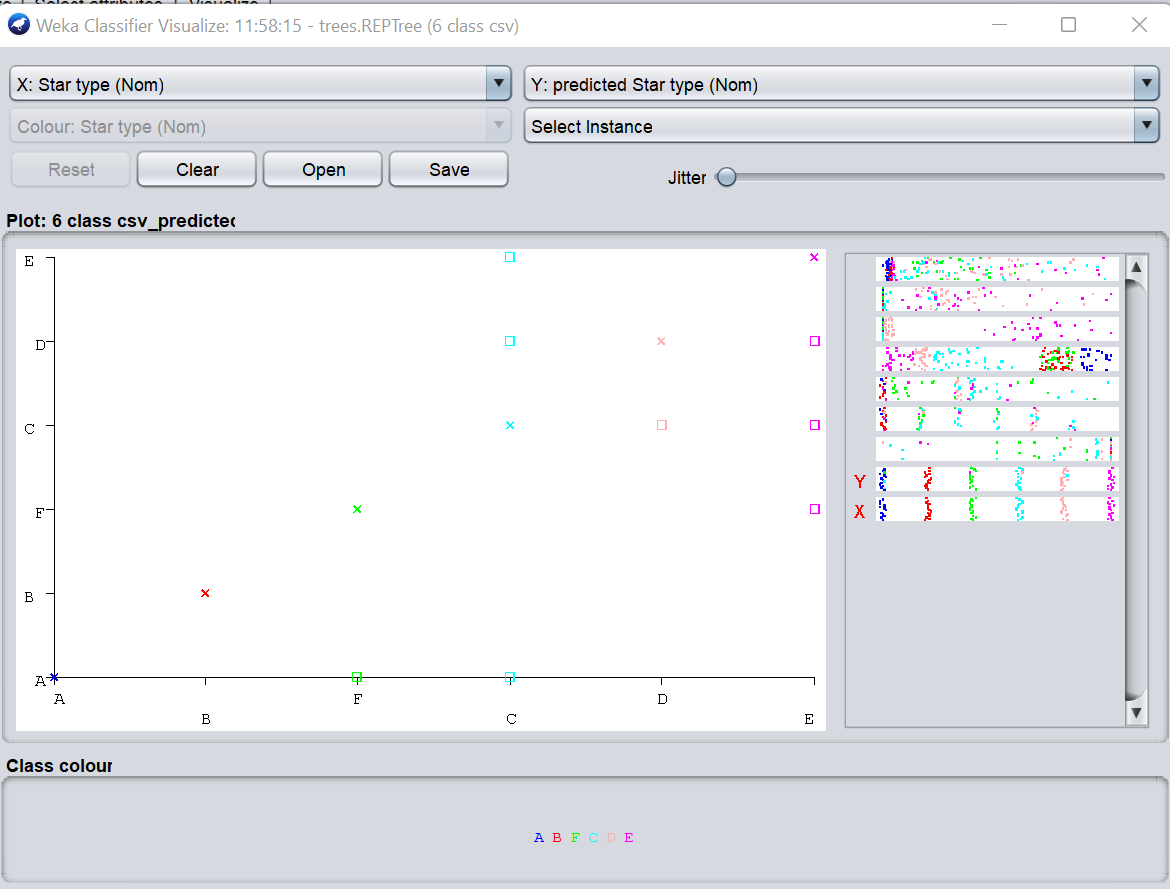
EXPERIMENT-7

Train a Decision Tree

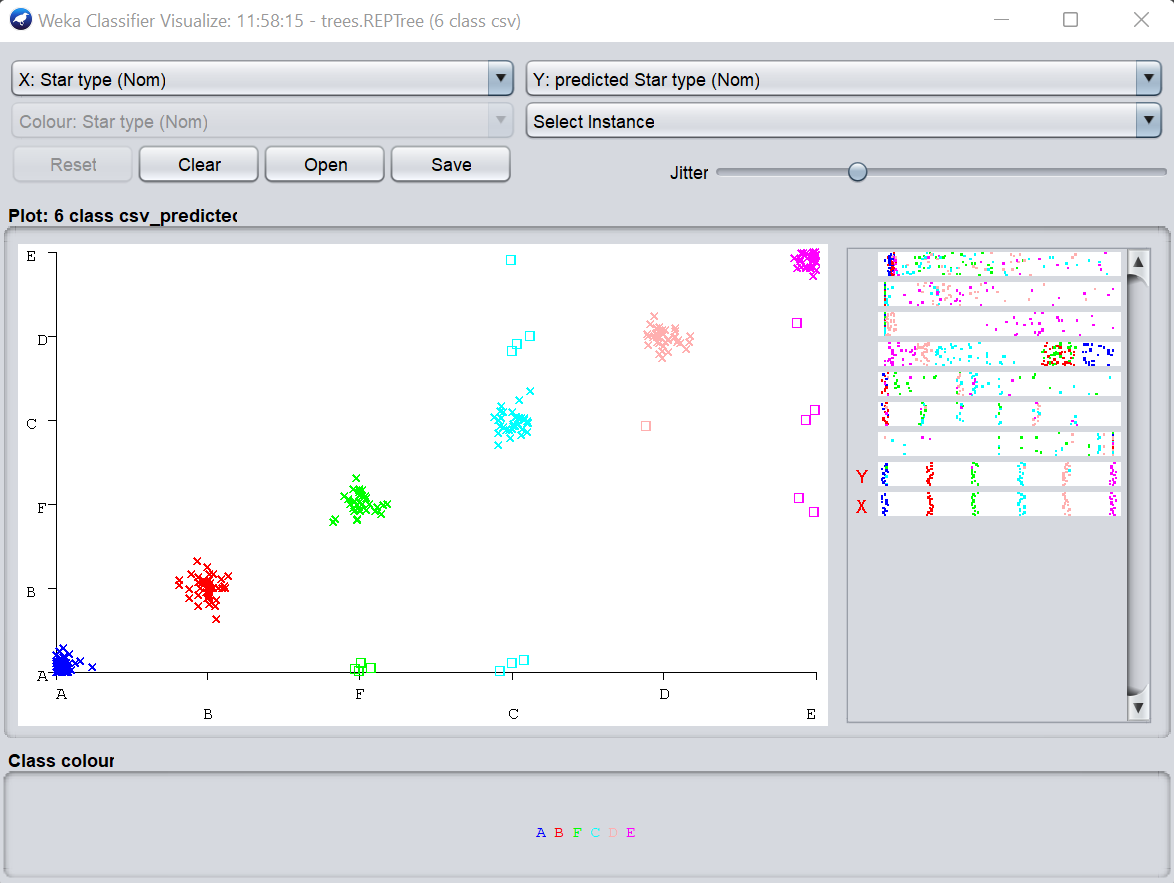






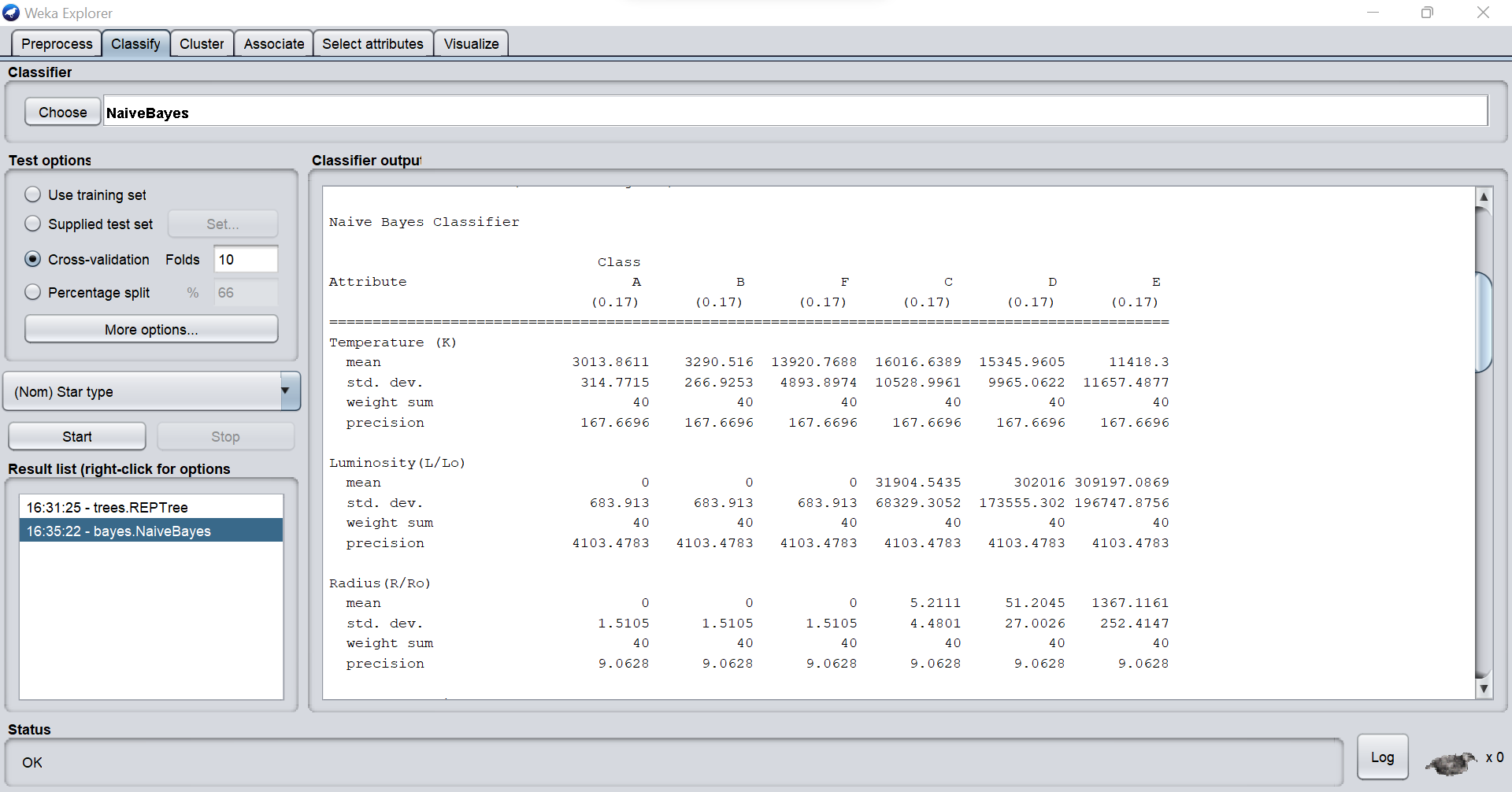


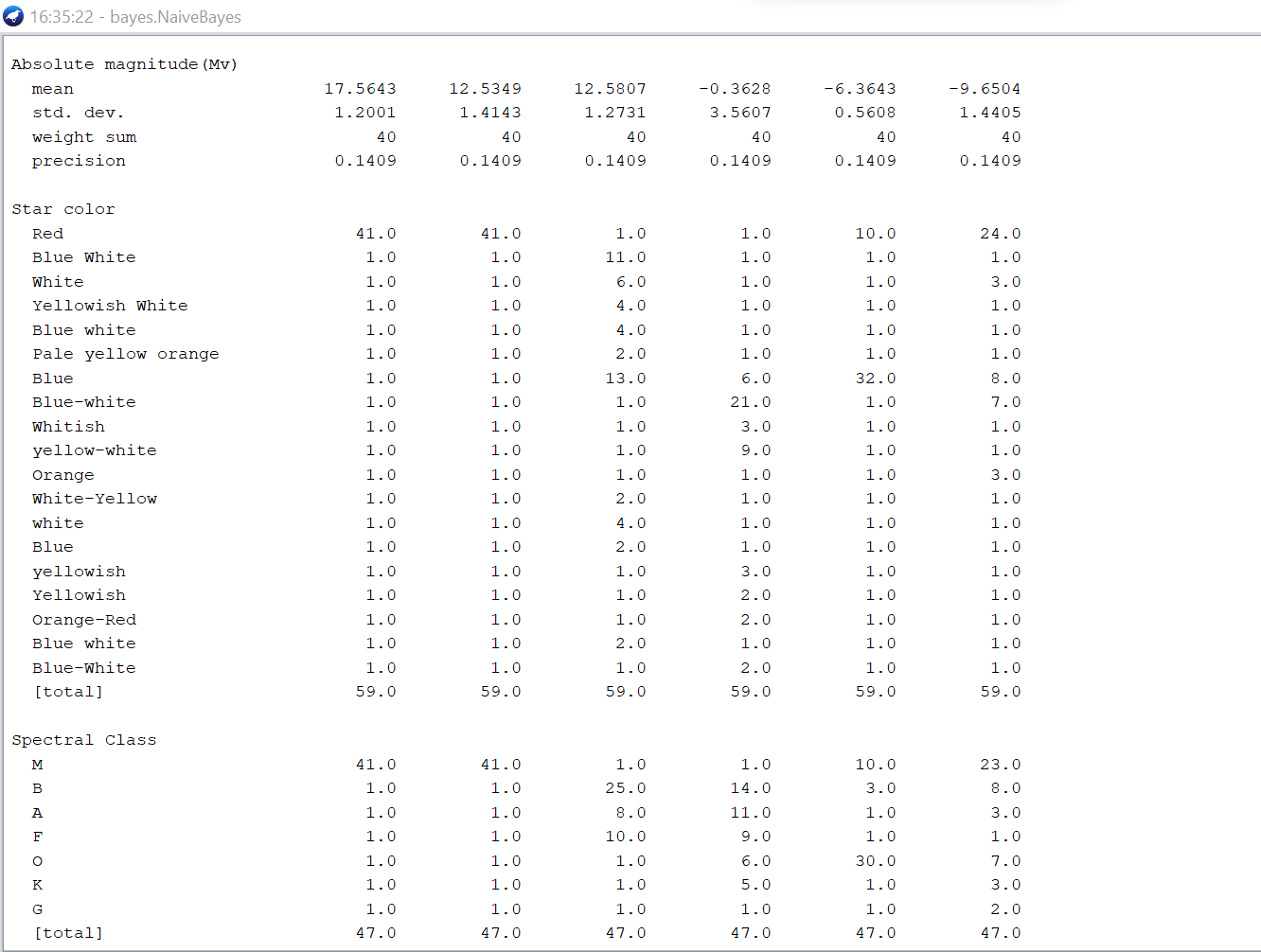
Using some Jitter

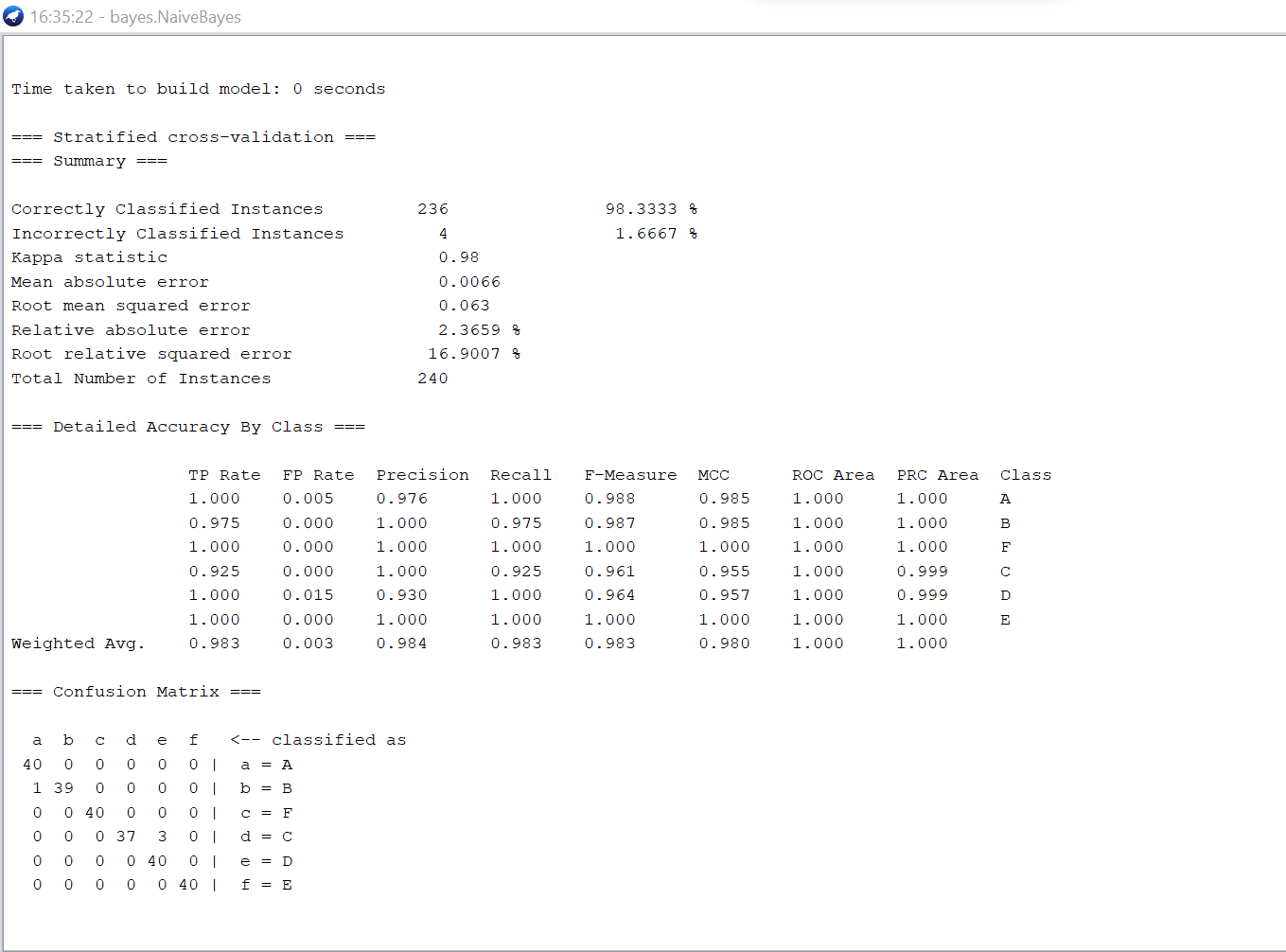


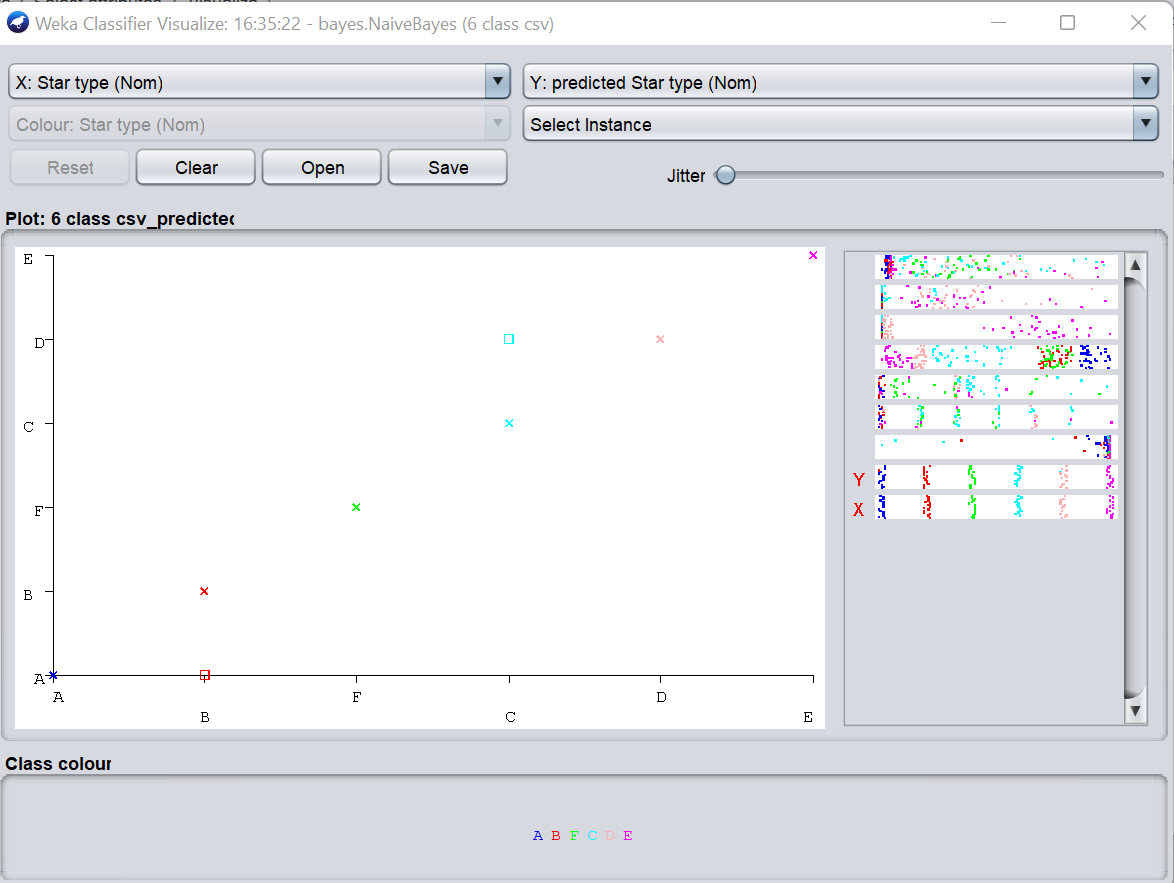
EXPERIMENT-8

Train a Naïve Bayes Classifier



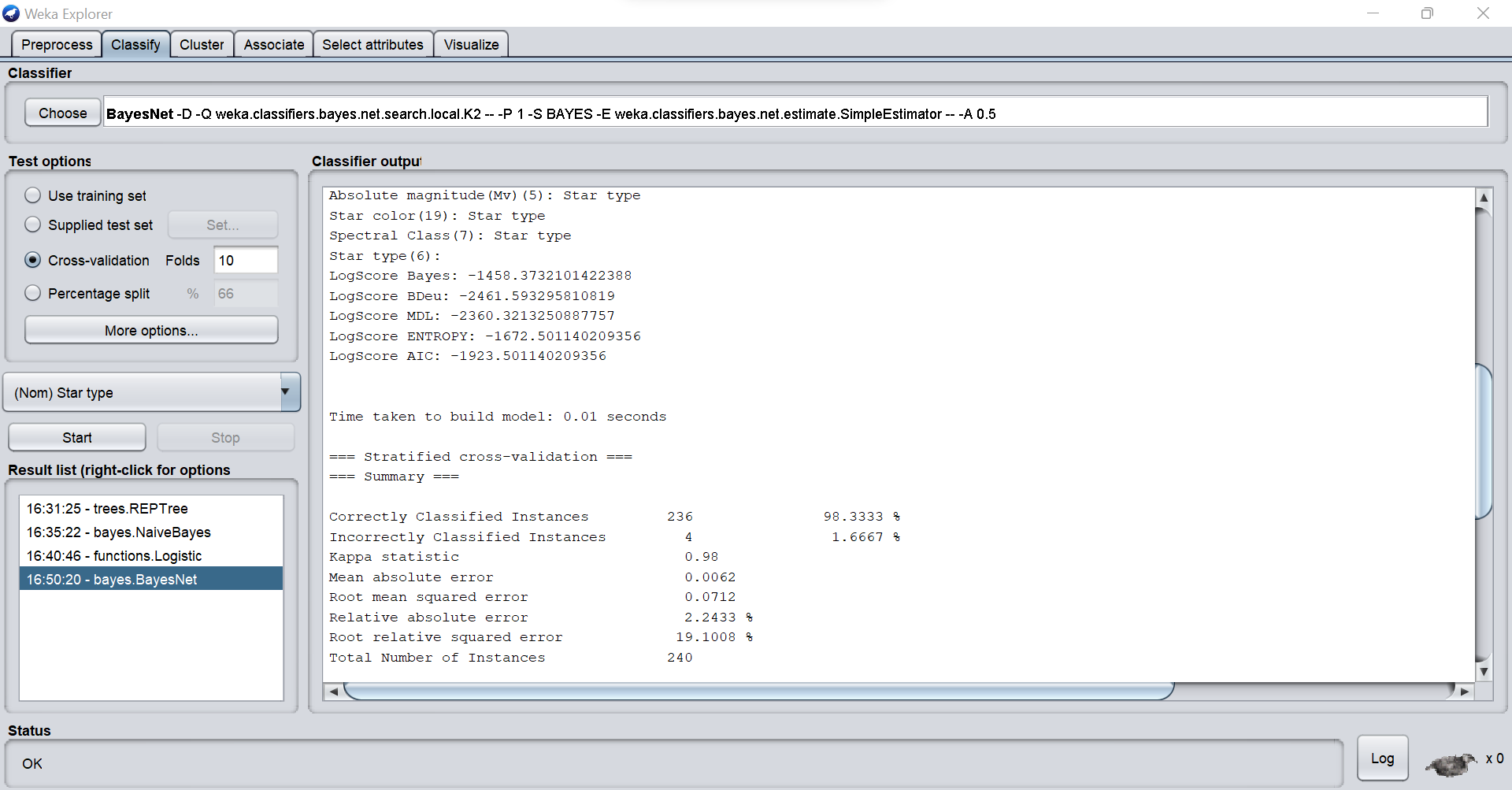


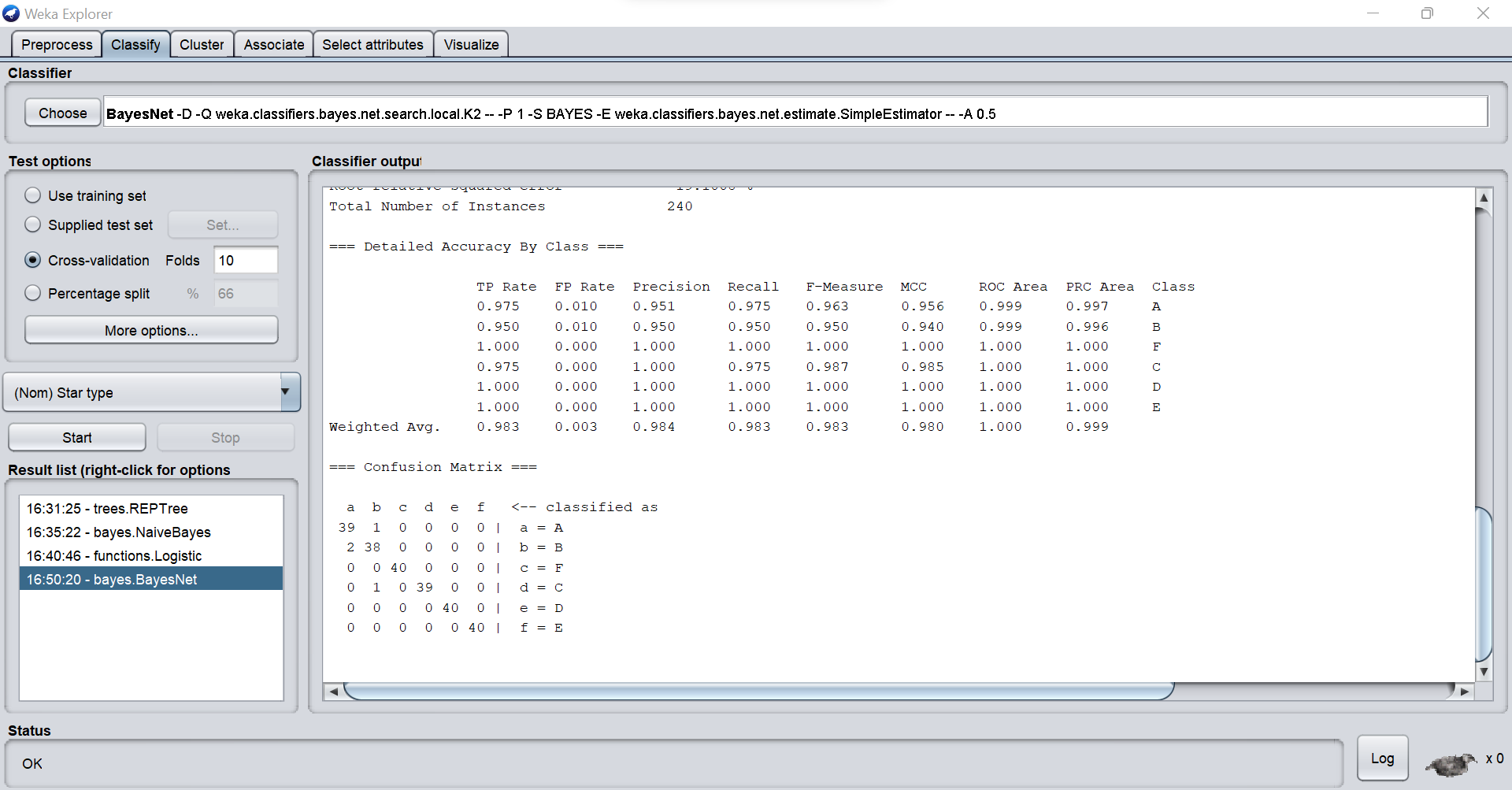


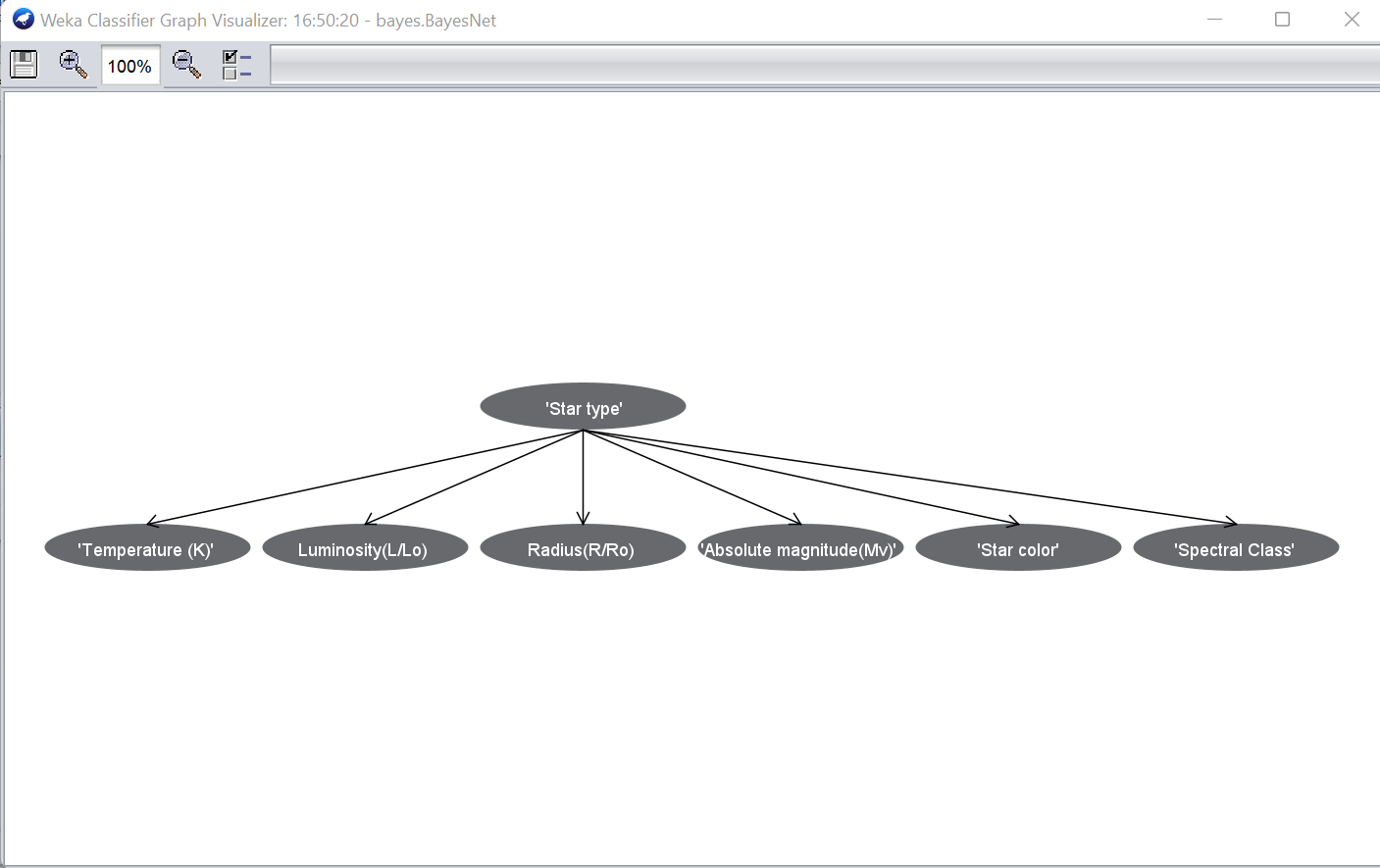


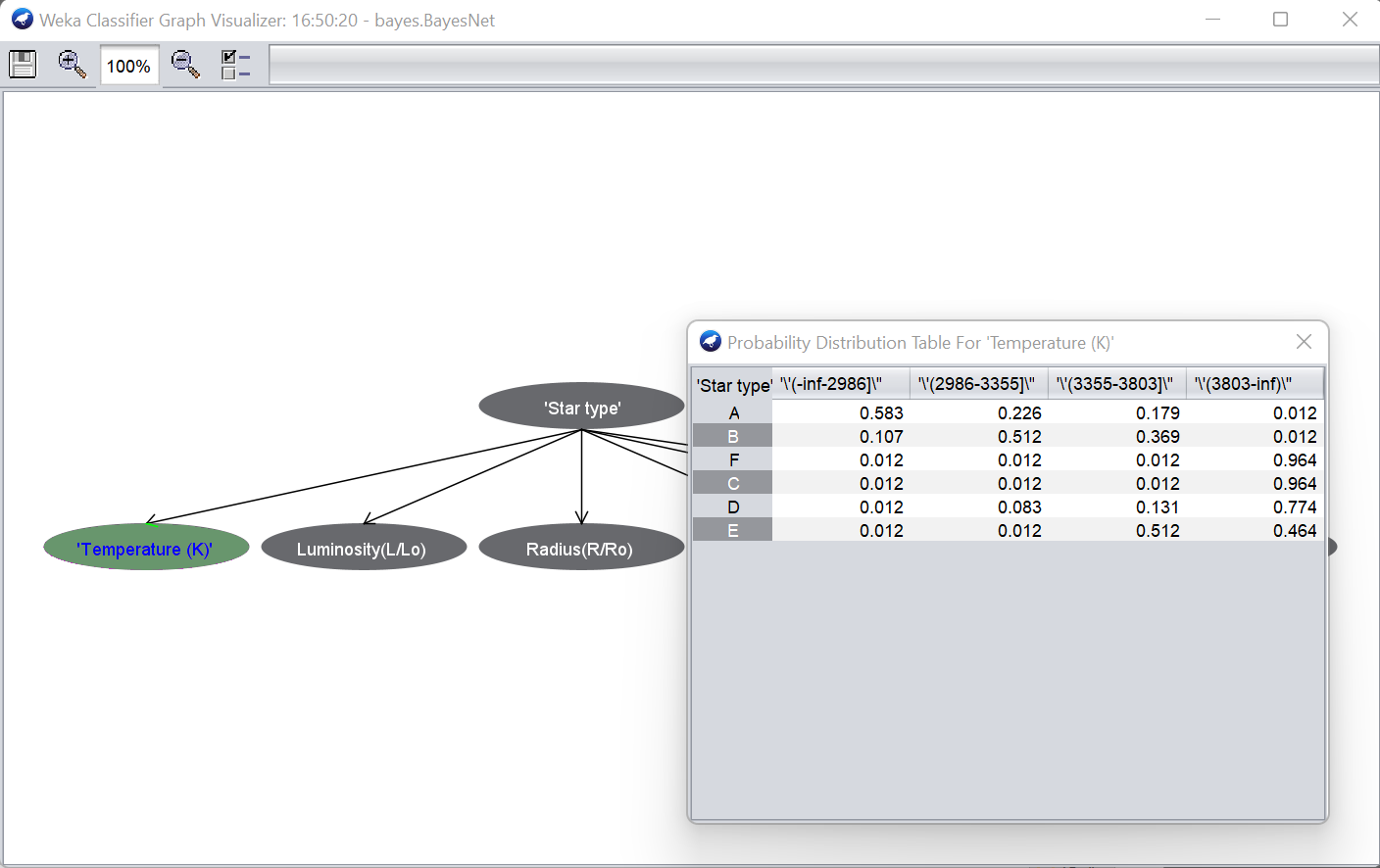
EXPERIMENT-9

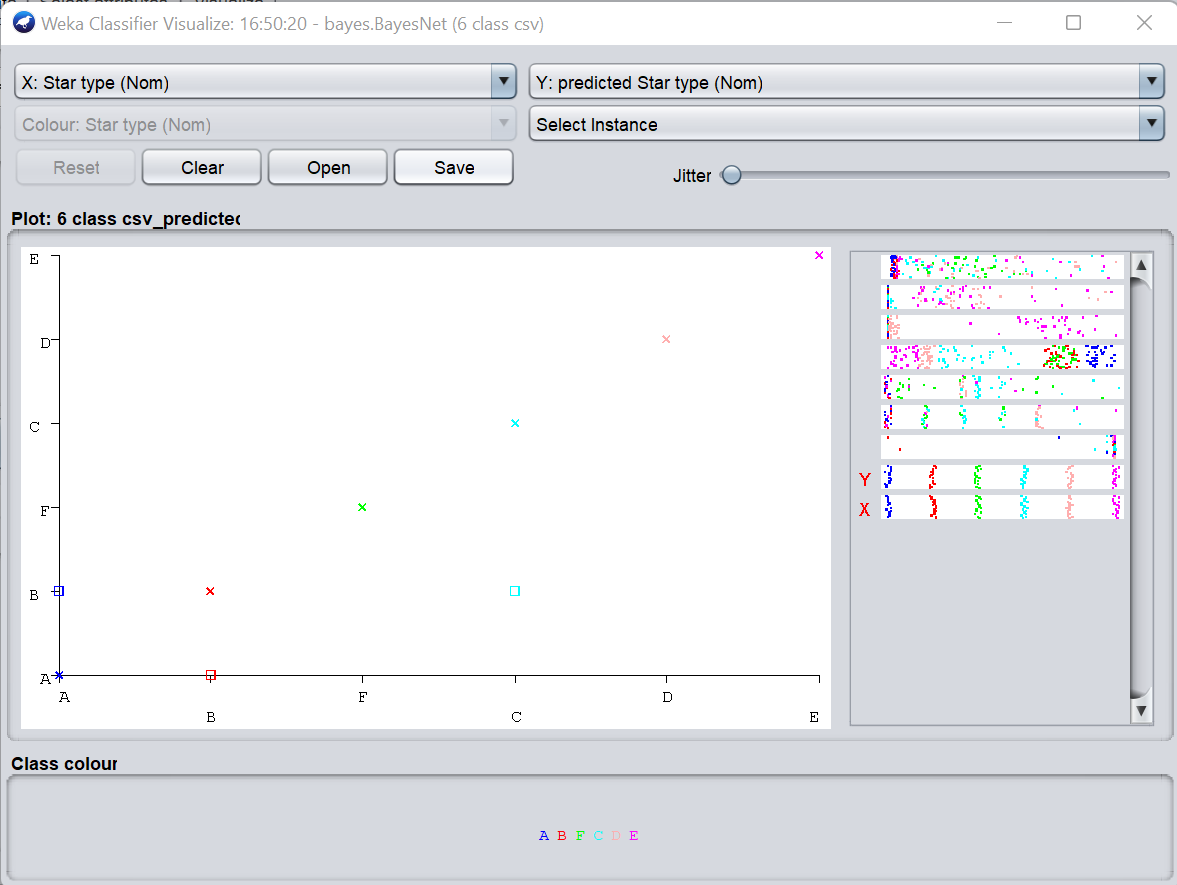
Train a Bayesian Belief Network Classifier





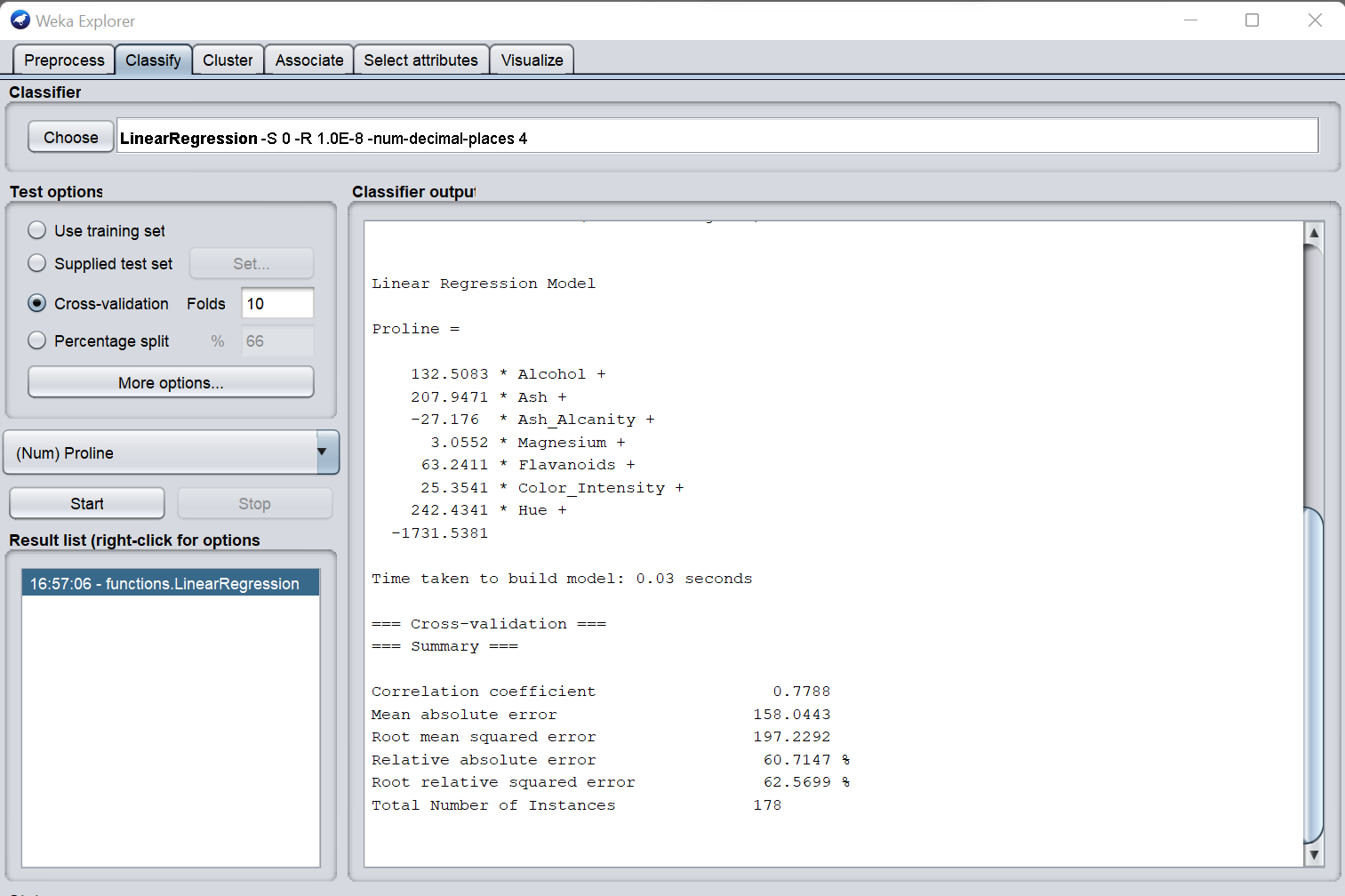


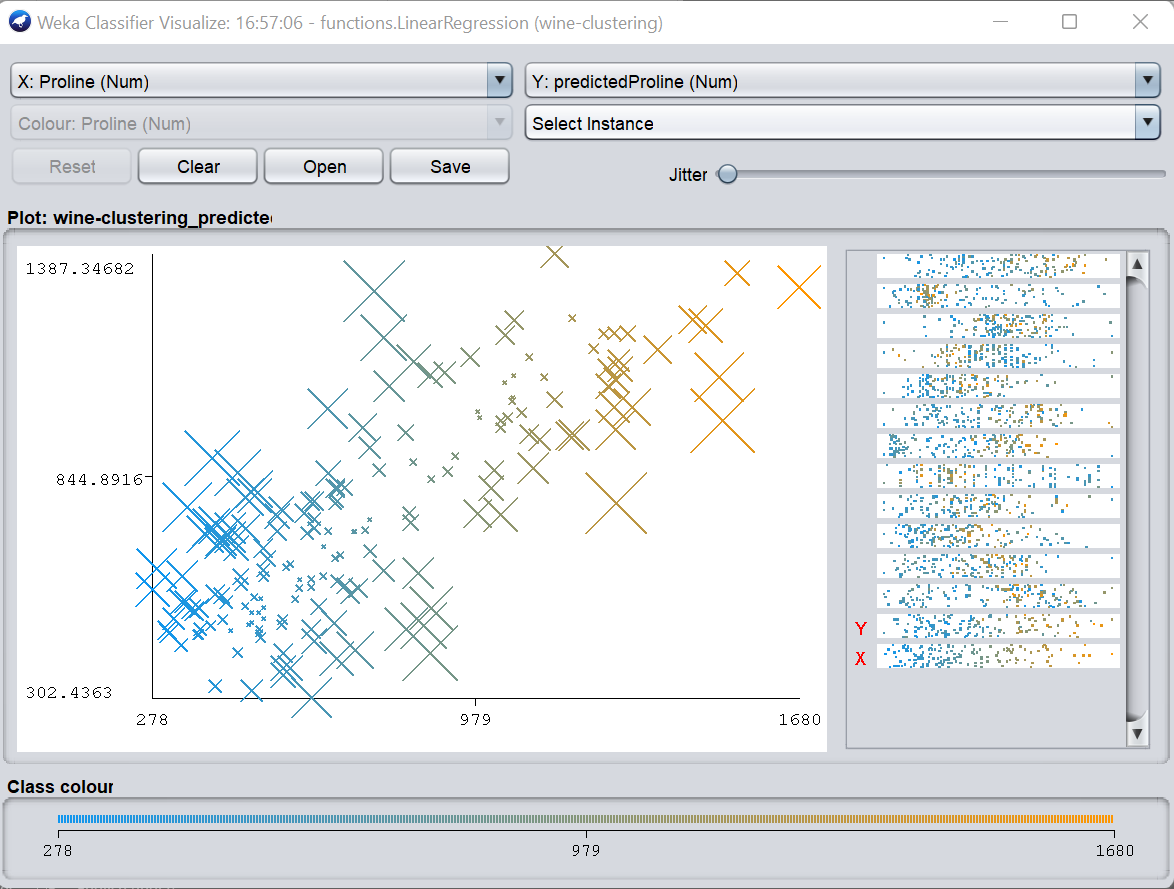




EXPERIMENT-10

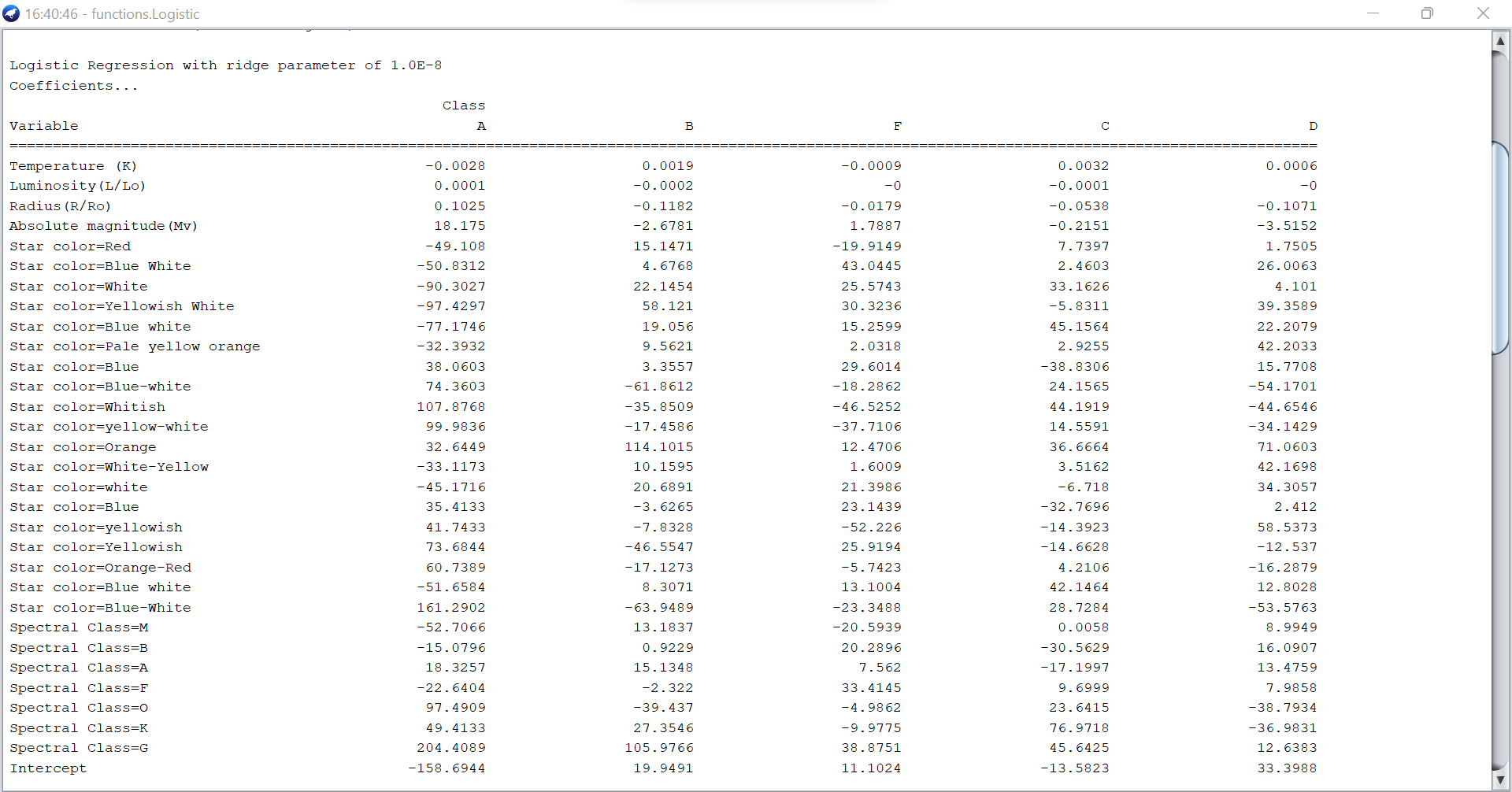
Train a Linear Regression Model

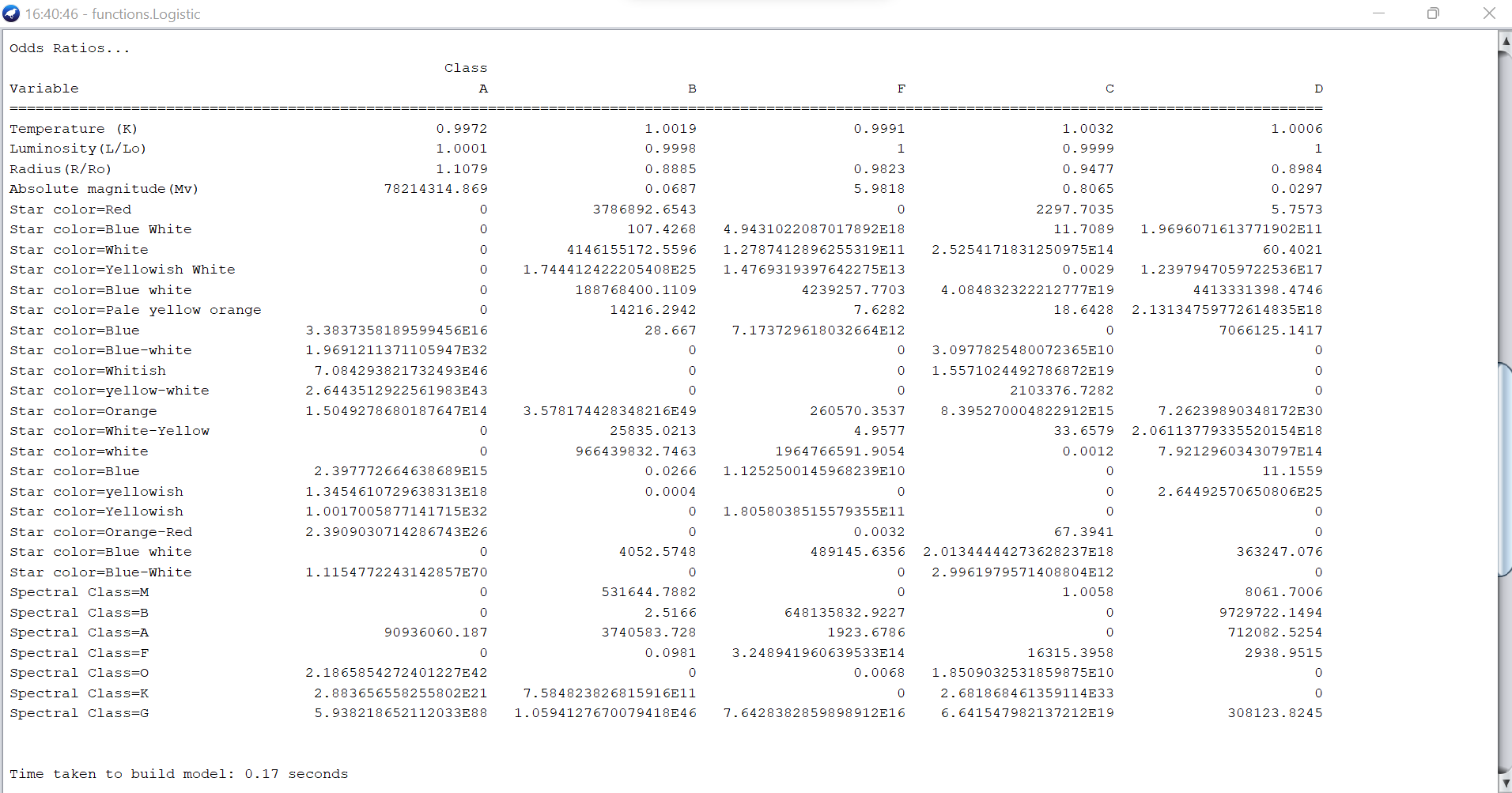


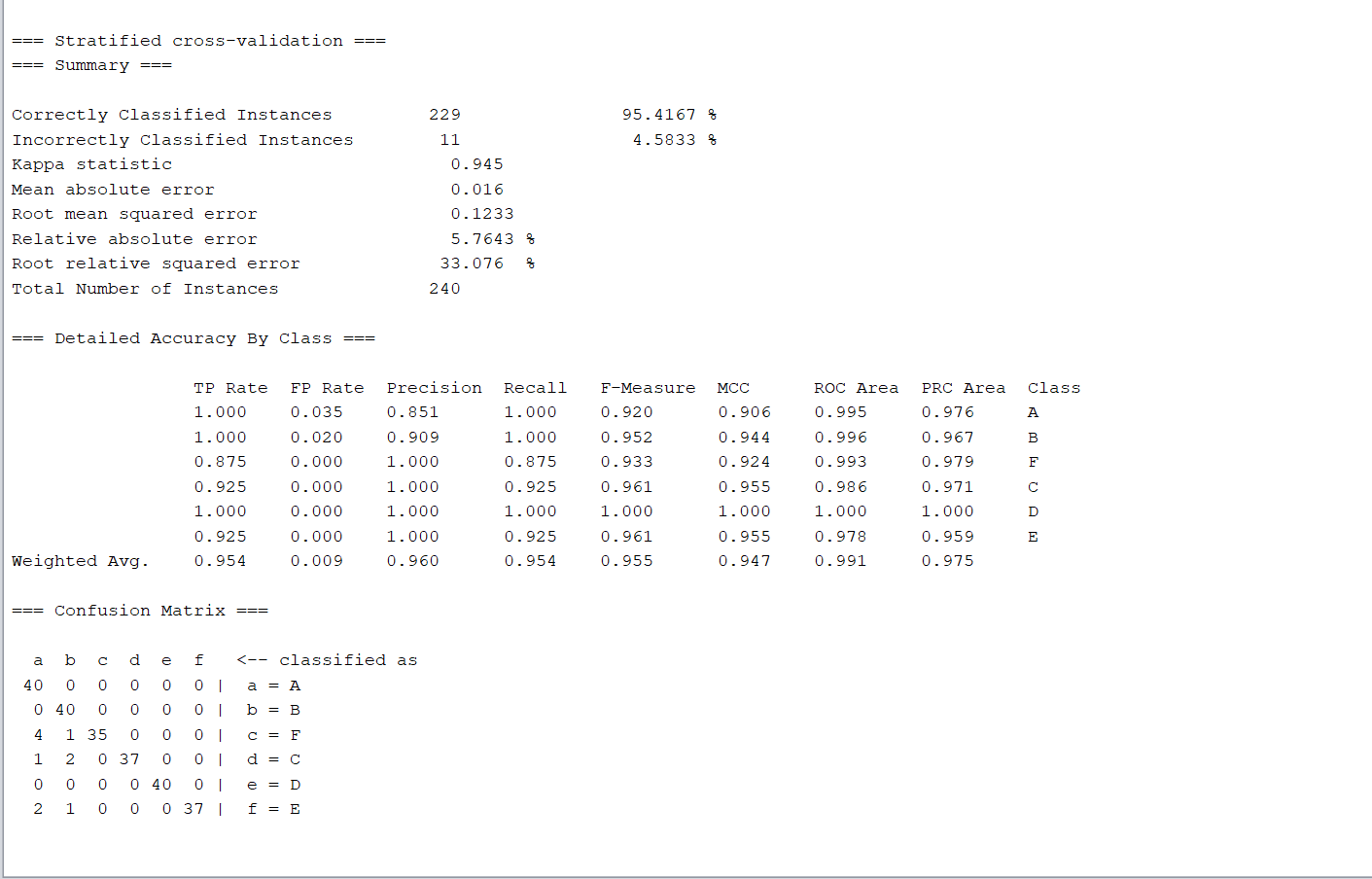


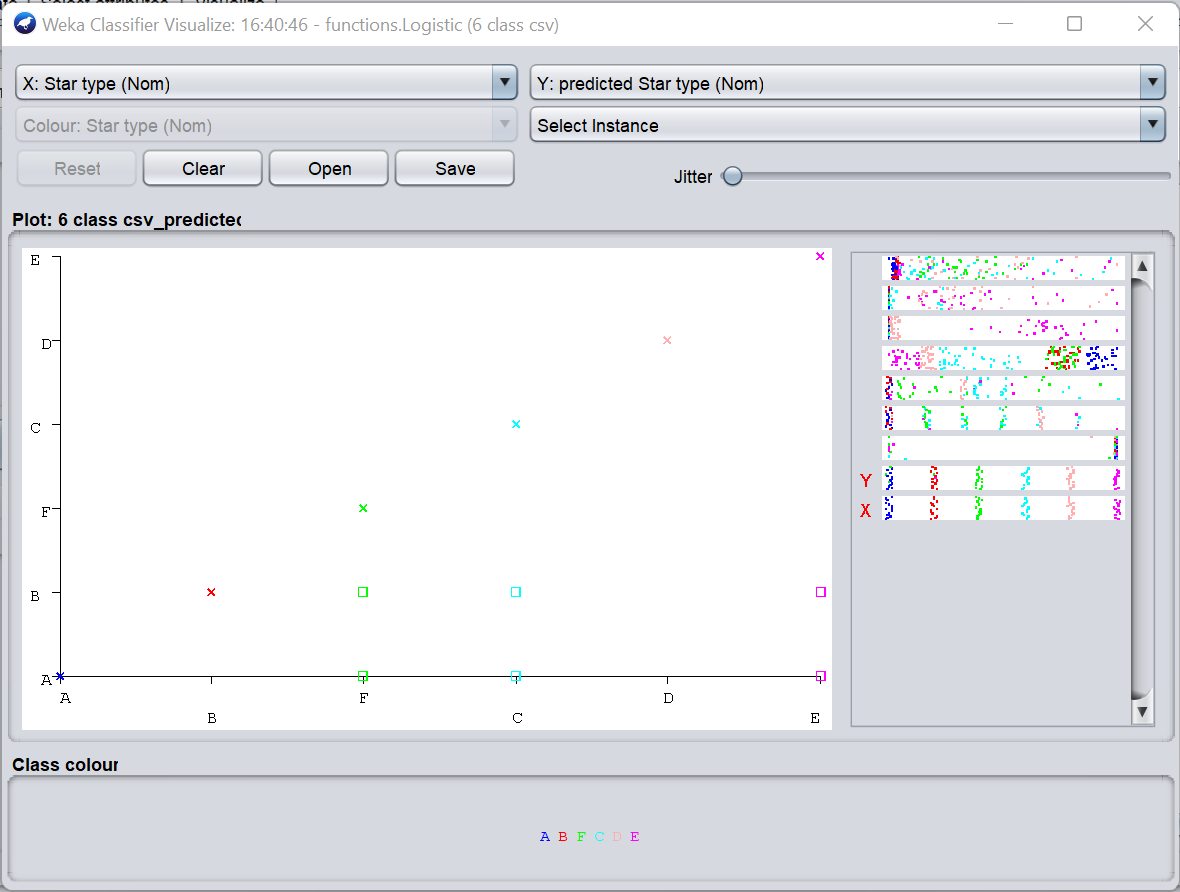
EXPERIMENT-11

Train a Logistic Regression Model



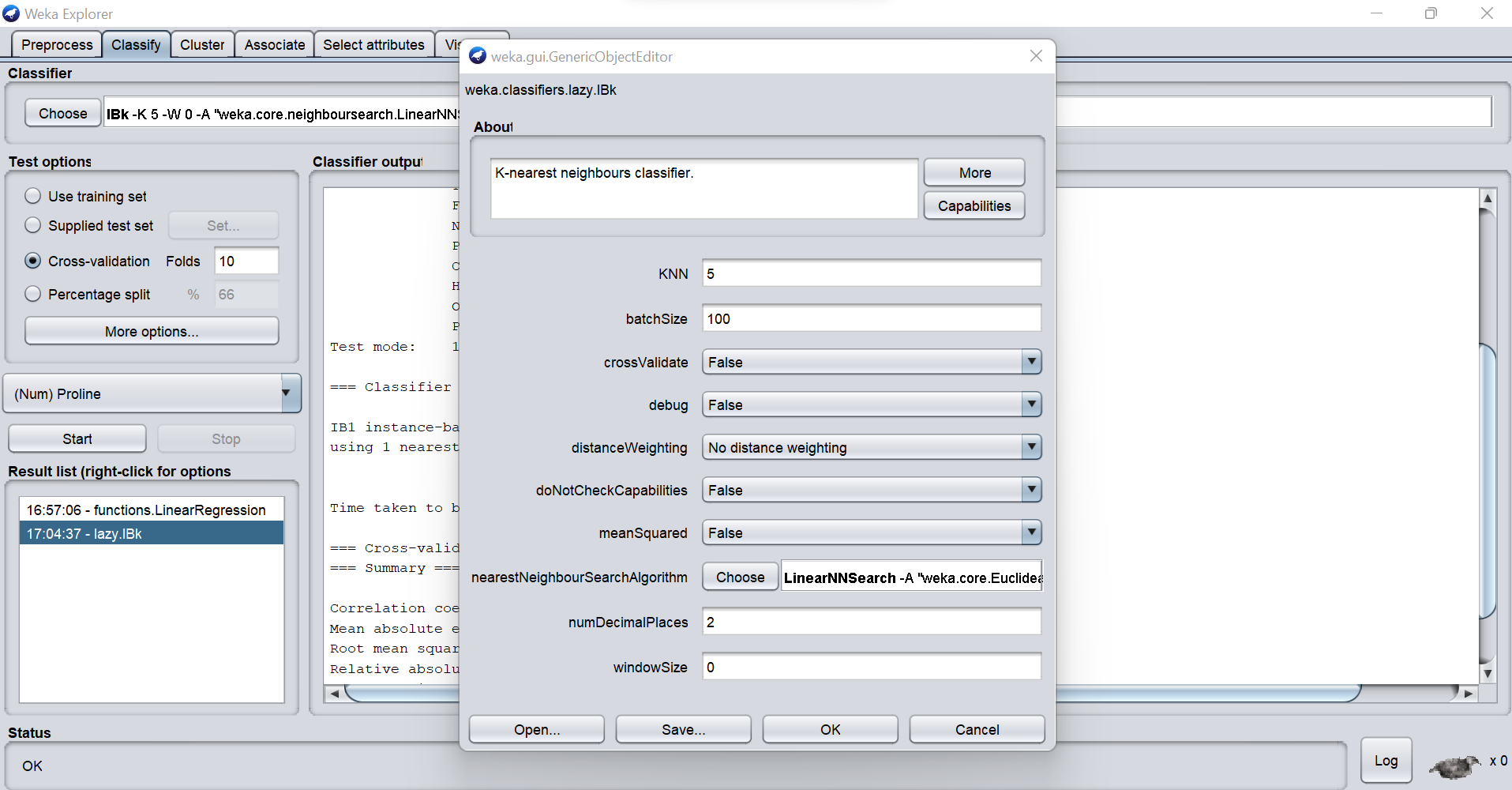


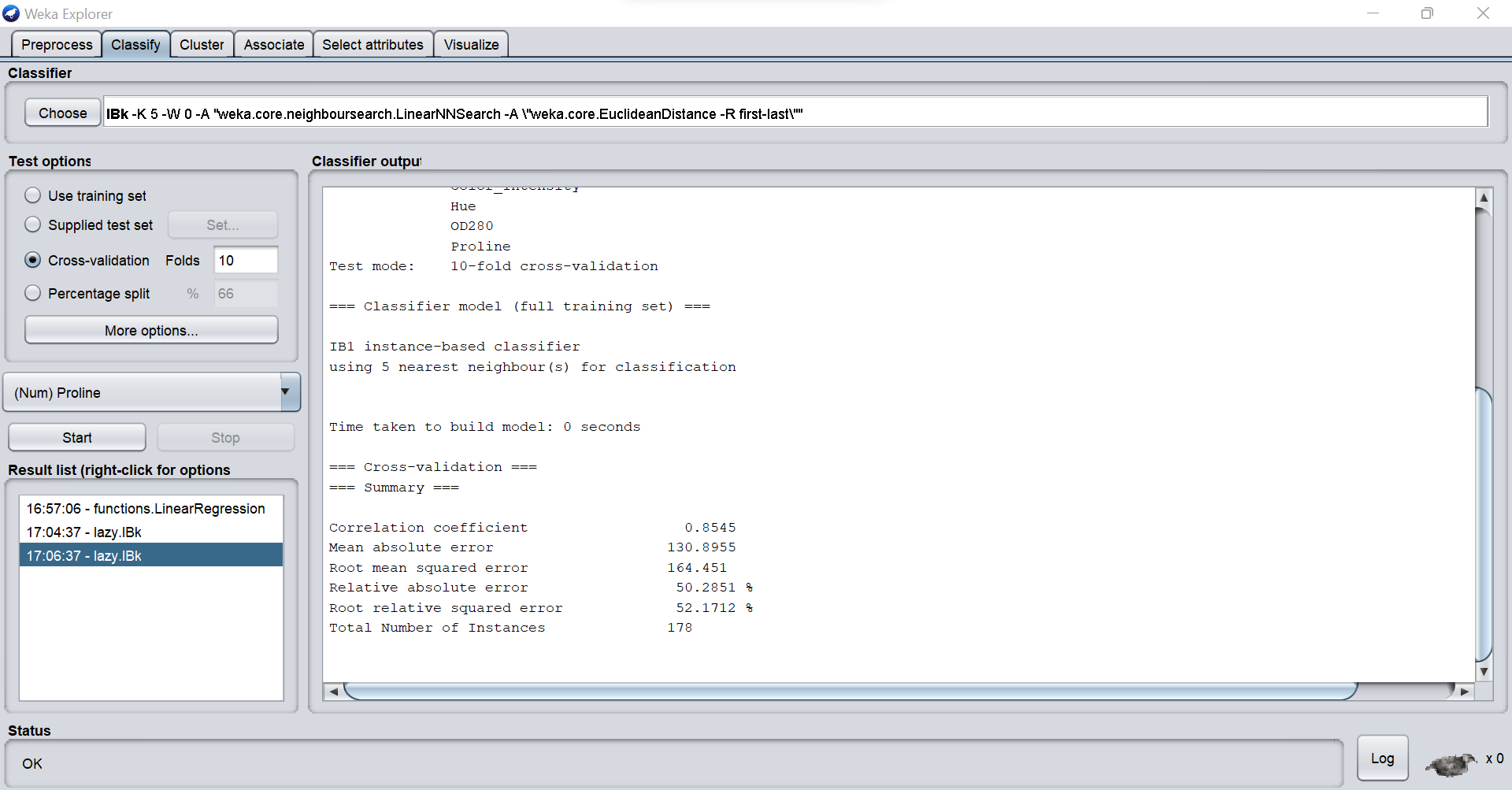


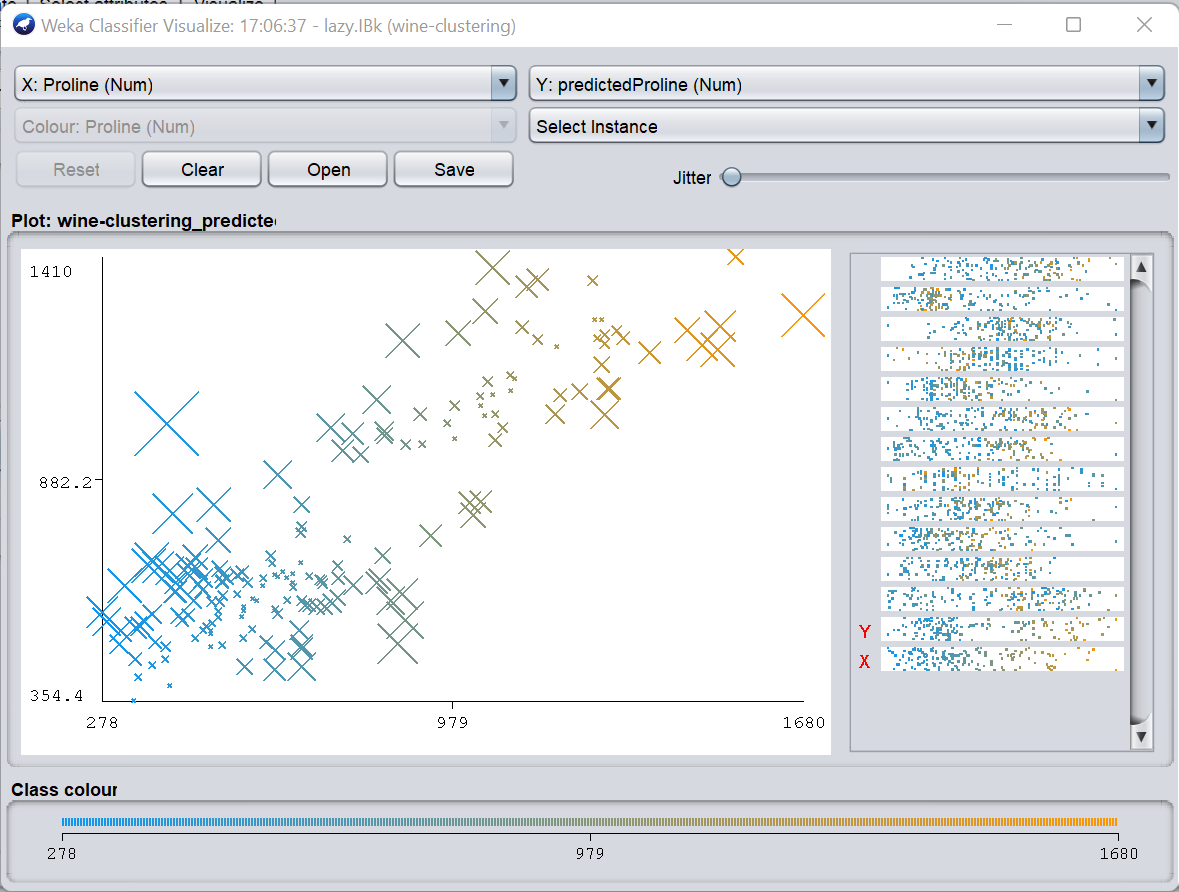


EXPERIMENT-12

Train a KNN model







EXPERIMENT-13

Train a K Means Clustering Model

