NORTH SOUTH UNIVERSITY

Spring 2023

CSE445: Machine Learning

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

Credit Card Fraud Detection Project

Course Initial & Section: CSE445.6
Date: 27.05.2023

Submitted to:

Professor Dr. B. M. Mainul Hossain (BMH)

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING, North South University

Submitted by:

Group No. 5

1. Mohammad Saiful Islam [ID: 1922071042]
2. Kamrul Hassan Jony [ID: 1921537042]
3. Jahangir Alam [ID: 1921863042]
4. Md Asif Hossain [ID: 1831737642]

For Credit Card Fraud Detection, we will use two popular classification algorithms of supervised machine learning techniques.

1. **Logistic Regression 2. Random Forest**

Data Pre-processing:

The class in a classification problem should be nominal, while the attributes can be nominal or numeric.

Python Code for Transform Machine Learning Data for “Class” attribute column from Numeric to Nominal and saves the dataset as 'creditcardClassNominal':

```
import pandas as pd

# Load the dataset
df = pd.read_csv('creditcard.csv')

# Print the new dataset
print(df)

# Create a new column for the nominal class
df['nominal_class'] = df['Class'].astype('str')

# Convert the values in the numeric class column to strings
df['nominal_class'] = df['nominal_class'].replace({'0': 'No', '1': 'Yes'})

# Drop the old numeric class column
df = df.drop('Class', axis=1)

# Create a new column for the nominal class
df['Class'] = df['nominal_class'].astype('str')

# Drop the old numeric class column
df = df.drop('nominal_class', axis=1)

# Print the new dataset
print(df)

# Save the DataFrame to a CSV file
df.to_csv('creditcardClassNominal.csv', index=False)
```

Output:

```
# Drop the old numeric class column
df = df.drop('Class', axis=1)

# Create a new column for the nominal class
df['Class'] = df['nominal_class'].astype('str')

# Drop the old numeric class column
df = df.drop('nominal_class', axis=1)

# Print the new dataset
print(df)

# Save the DataFrame to a CSV file
df.to_csv('creditcardClassNominal.csv', index=False)
```

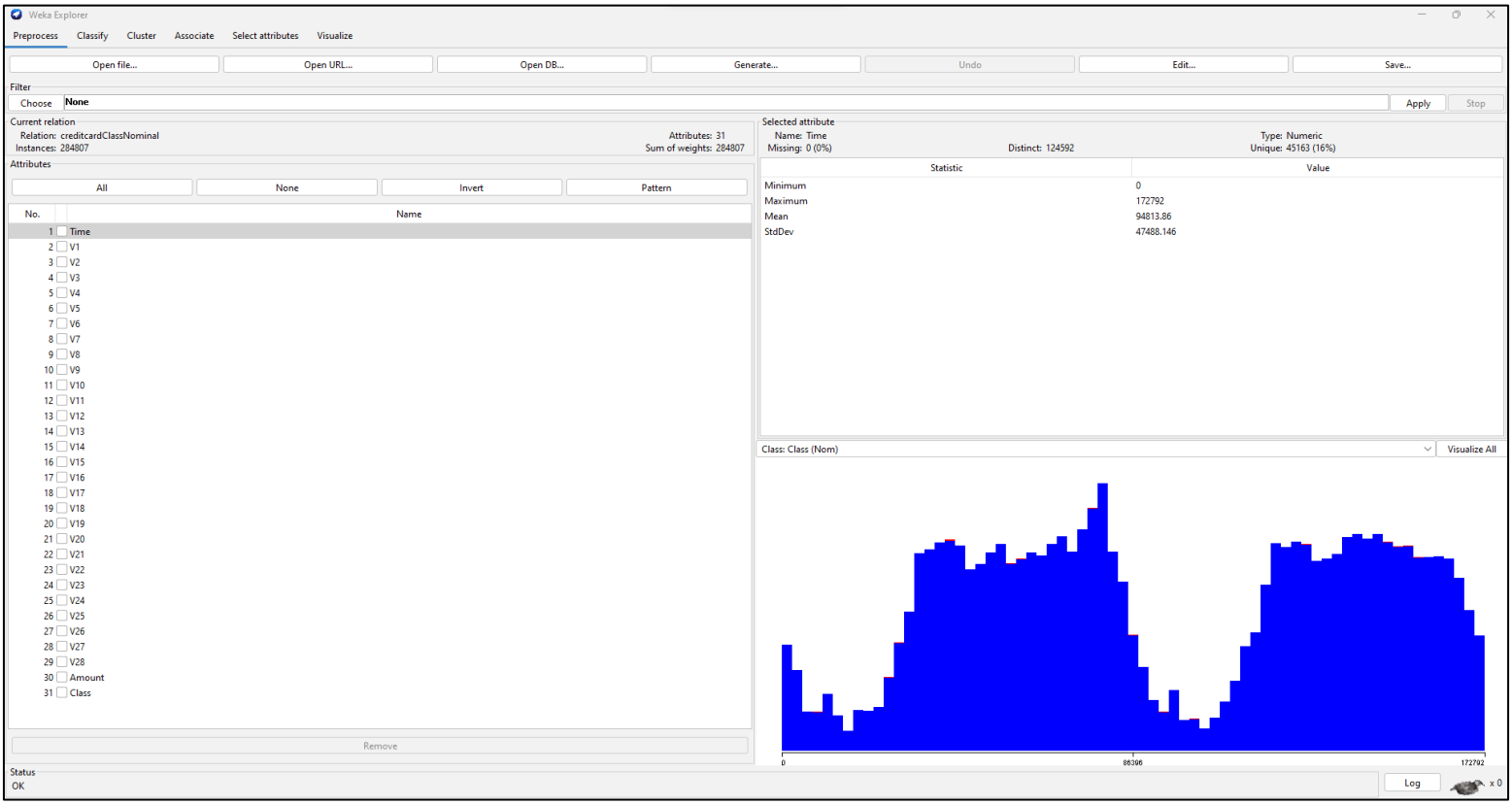
```
284803  0.012463 -1.016226 -0.606624 -0.395255  0.068472 -0.053527   24.79
284804 -0.037501  0.640134  0.265745 -0.087371  0.004455 -0.026561   67.88
284805 -0.163298  0.123205 -0.569159  0.546668  0.108821  0.104533   10.00
284806  0.376777  0.008797 -0.473649 -0.818267 -0.002415  0.013649  217.00

       Class
0         No
1         No
2         No
3         No
4         No
...      ...
284802    No
284803    No
284804    No
284805    No
284806    No

[284807 rows x 31 columns]
```

Data:

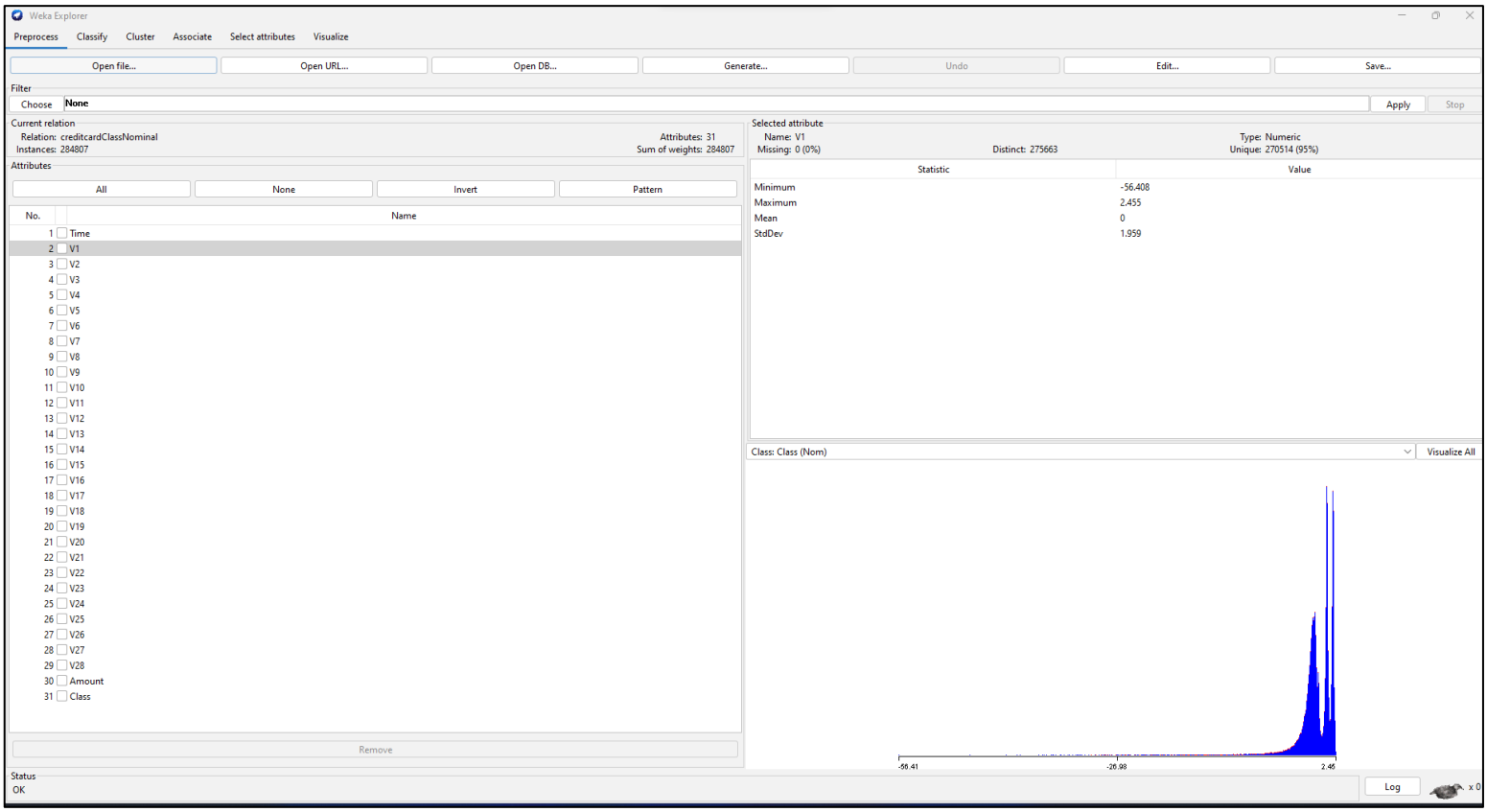
Dataset Name, Instances, Attribute, Sum of Weights

Attribute -> Name, Type, Missing Data, Distinct, Unique

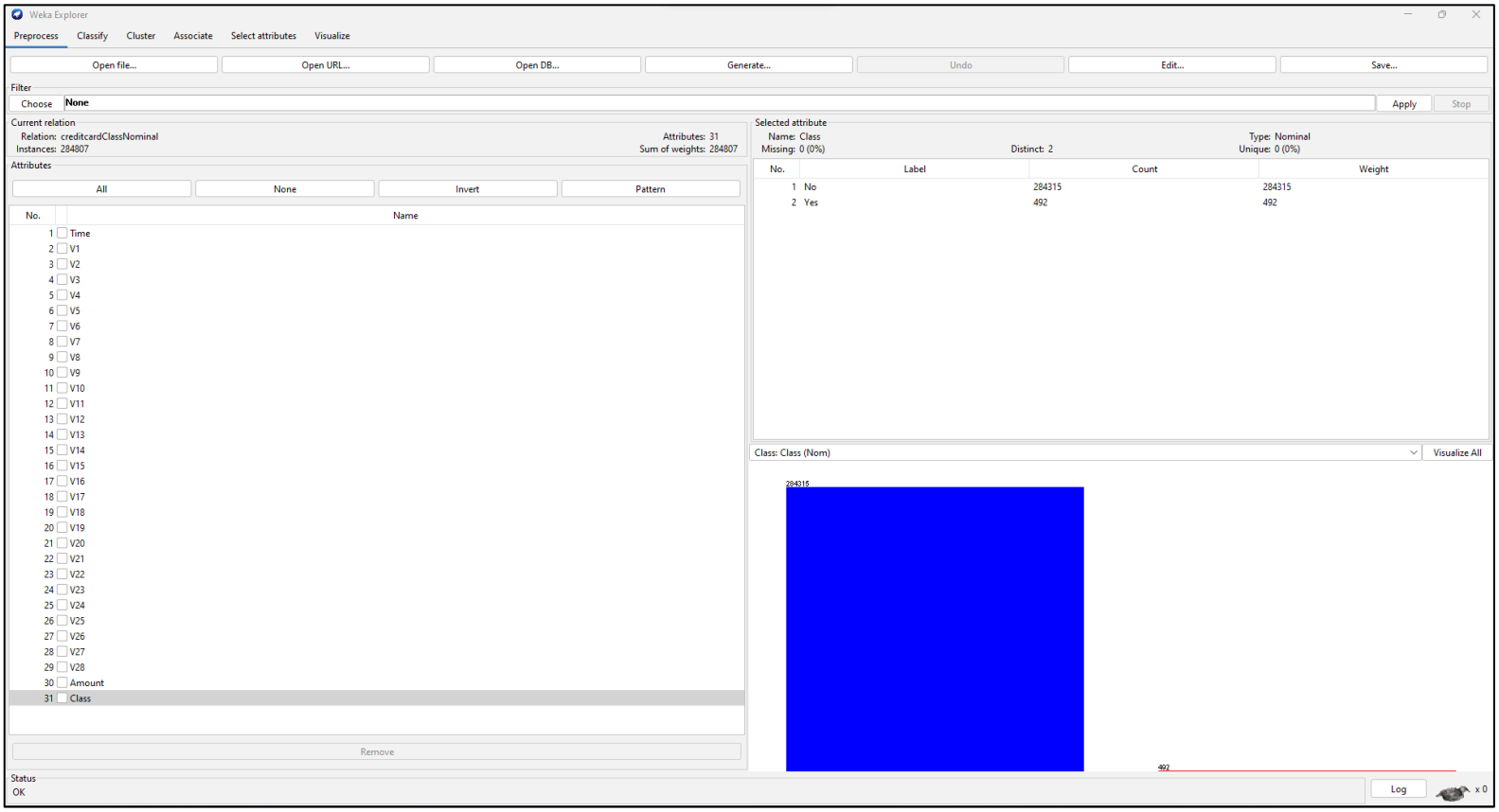
Statistics -> Minimum, Maximum, Mean & StdDev Value



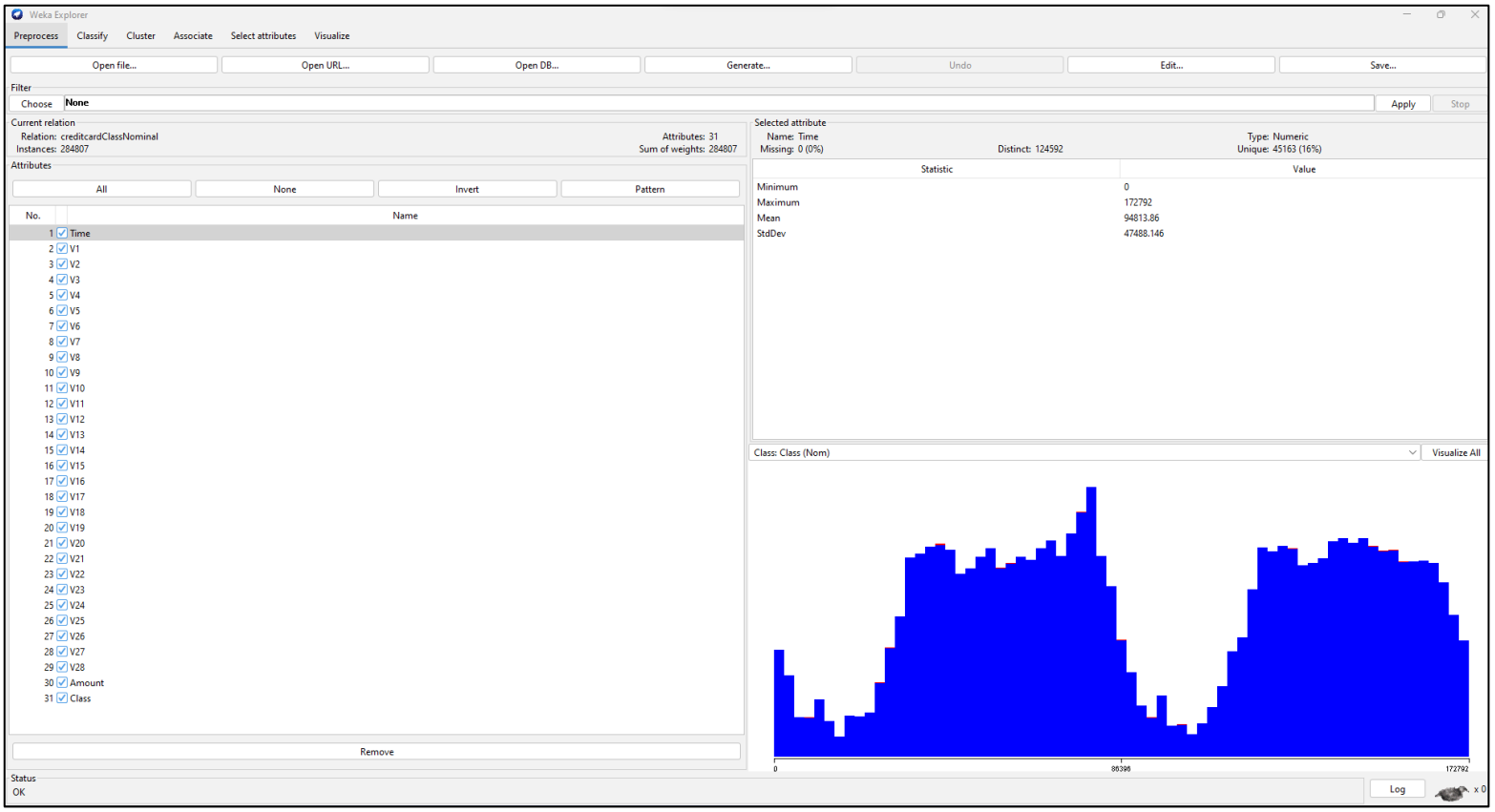
V1 Attribute:



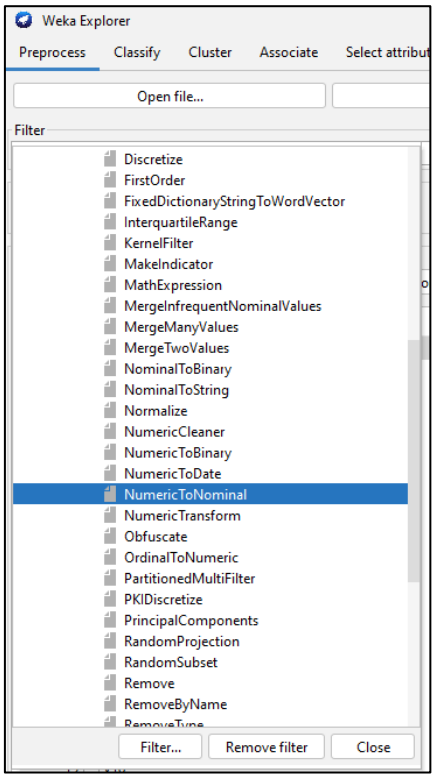
Class Attribute:



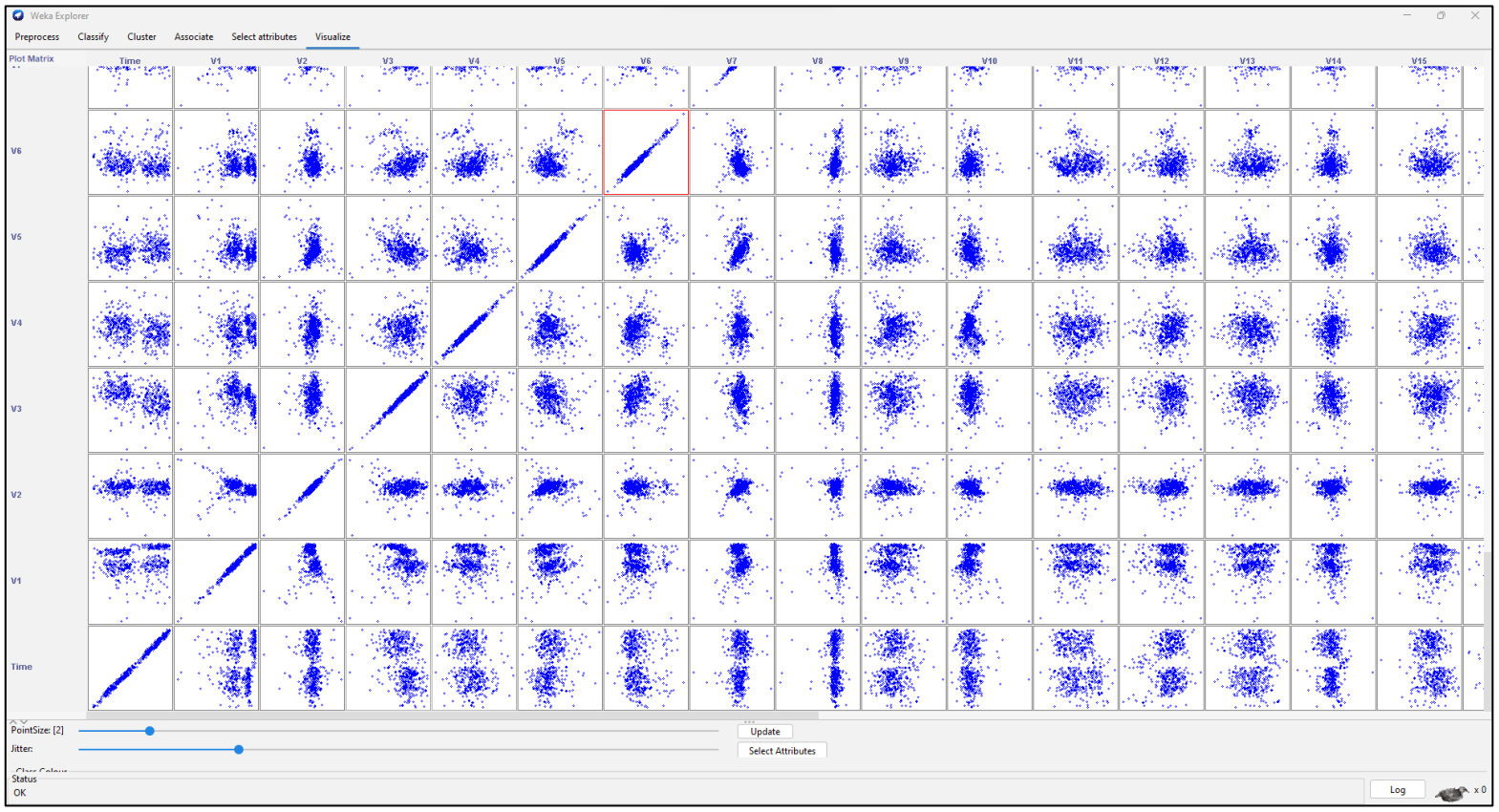
All Attribute:

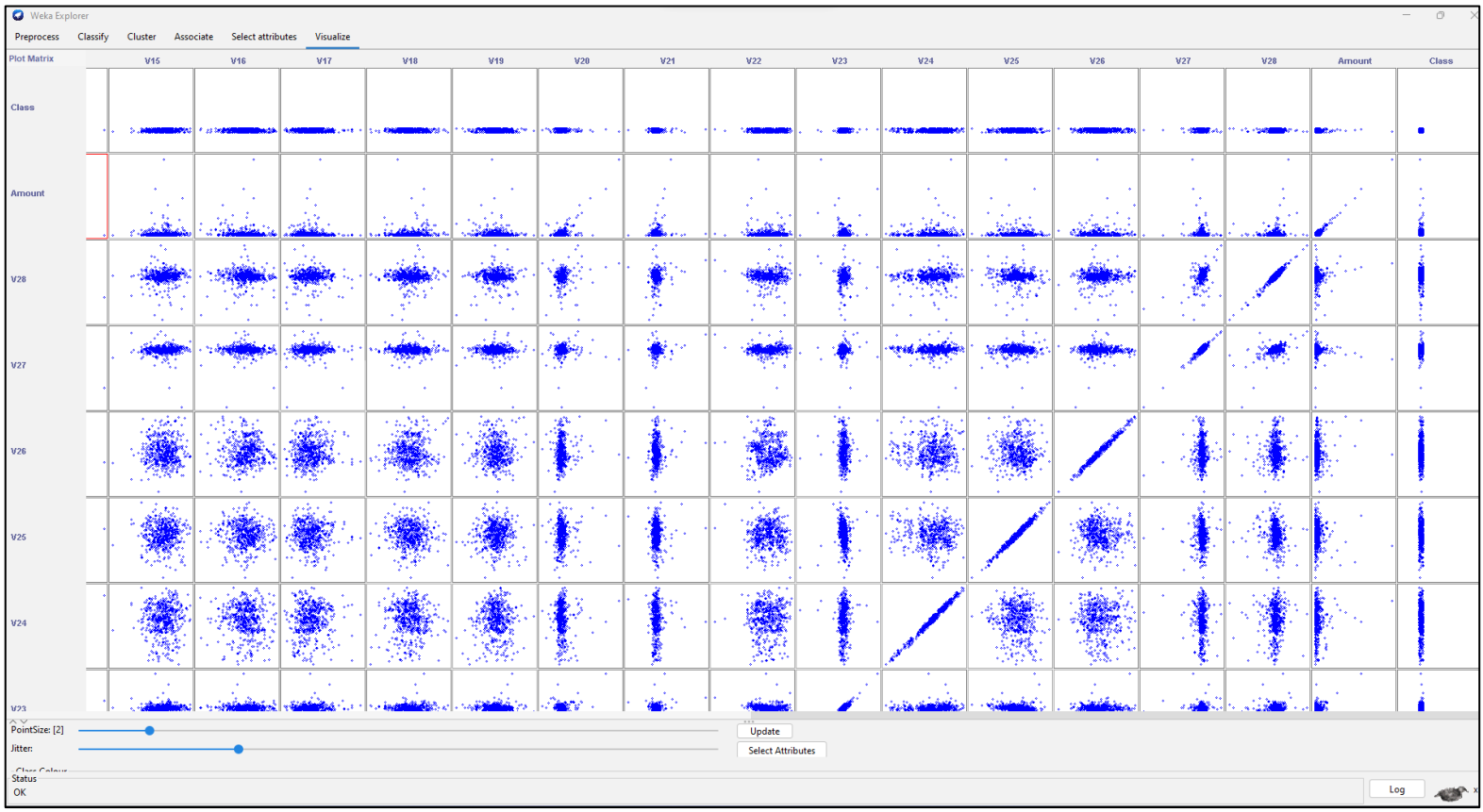


Pre-process: Class attribute data (Numeric to Nominal)

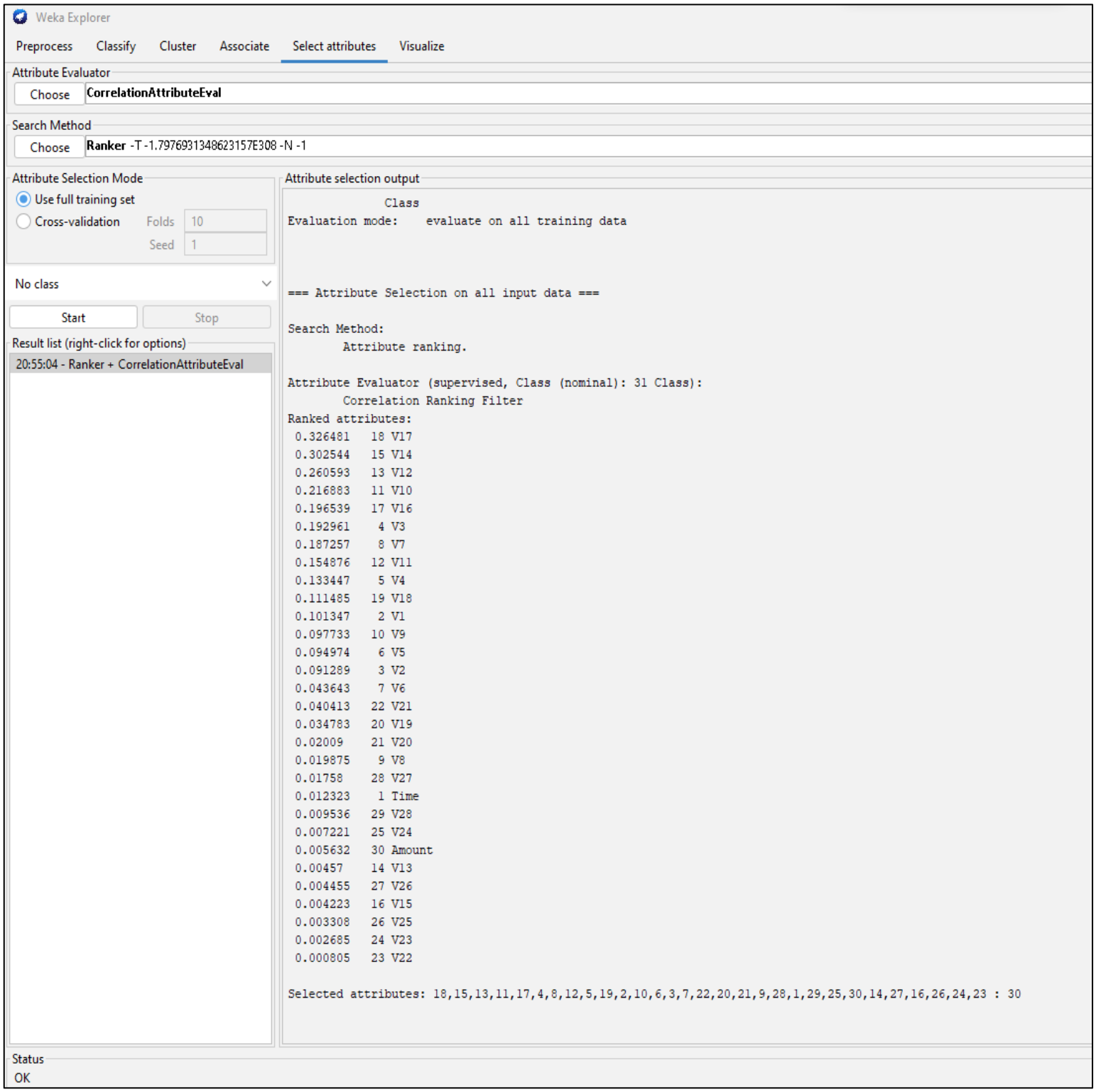


Visualize:


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Plot Matrix
Time
V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V11
V12
V13
V14
V15
PointSize: [2]
Jitter:
Update
Select Attributes
Status
OK
Log

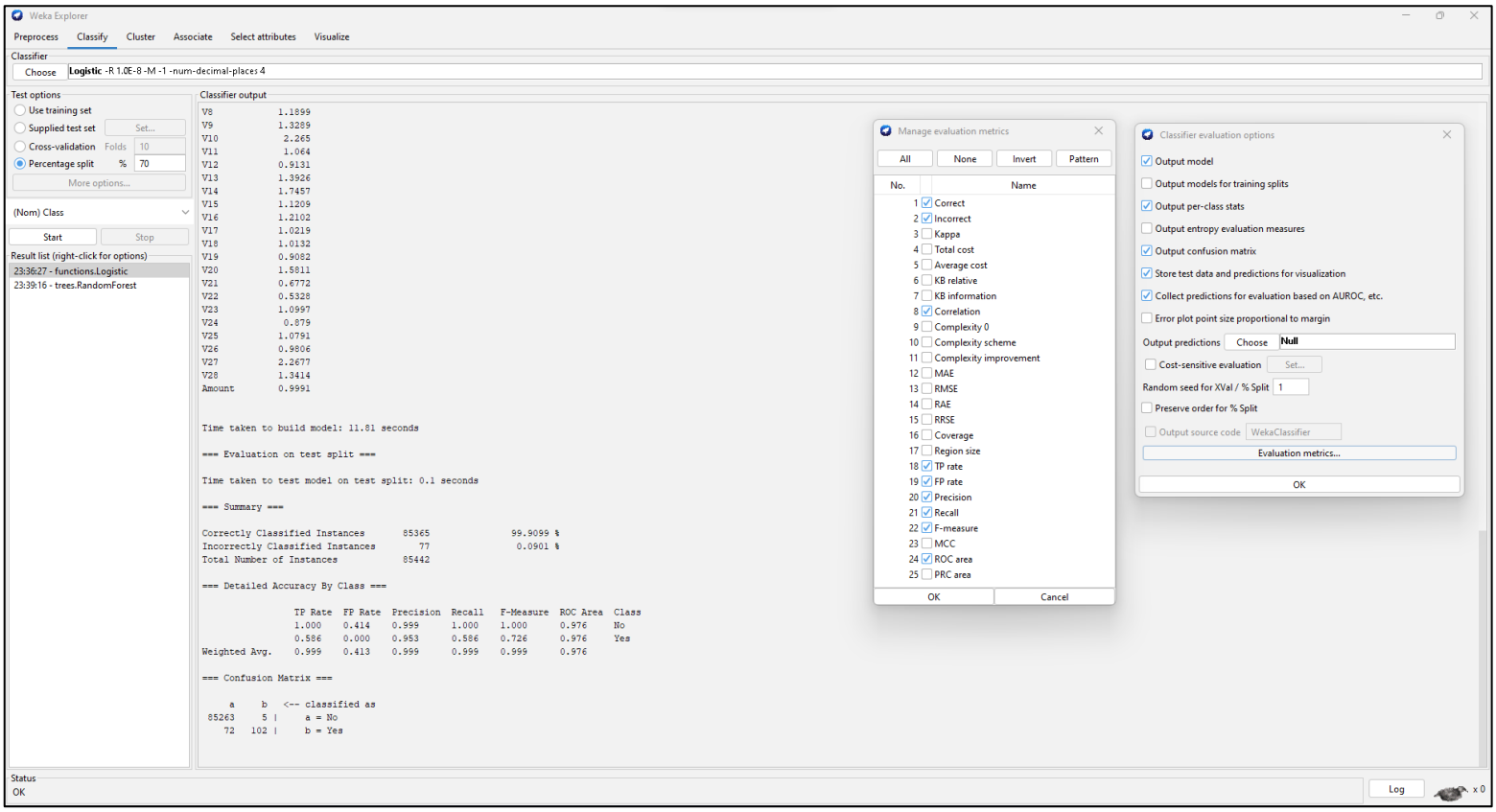

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Plot Matrix
V15
V16
V17
V18
V19
V20
V21
V22
V23
V24
V25
V26
V27
V28
Amount
Class
PointSize: [2]
Jitter:
Update
Select Attributes
Status
OK
Log

Correlation



Classifier Results

Logistic Regression



Random Forest