



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)
```

	0	1	2	3	4	5	6	7
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	1	2	3	4	...
	No	No	No	No	No	...
284802	No					
284803	No					
284804	No					
284805	No					
284806	No					

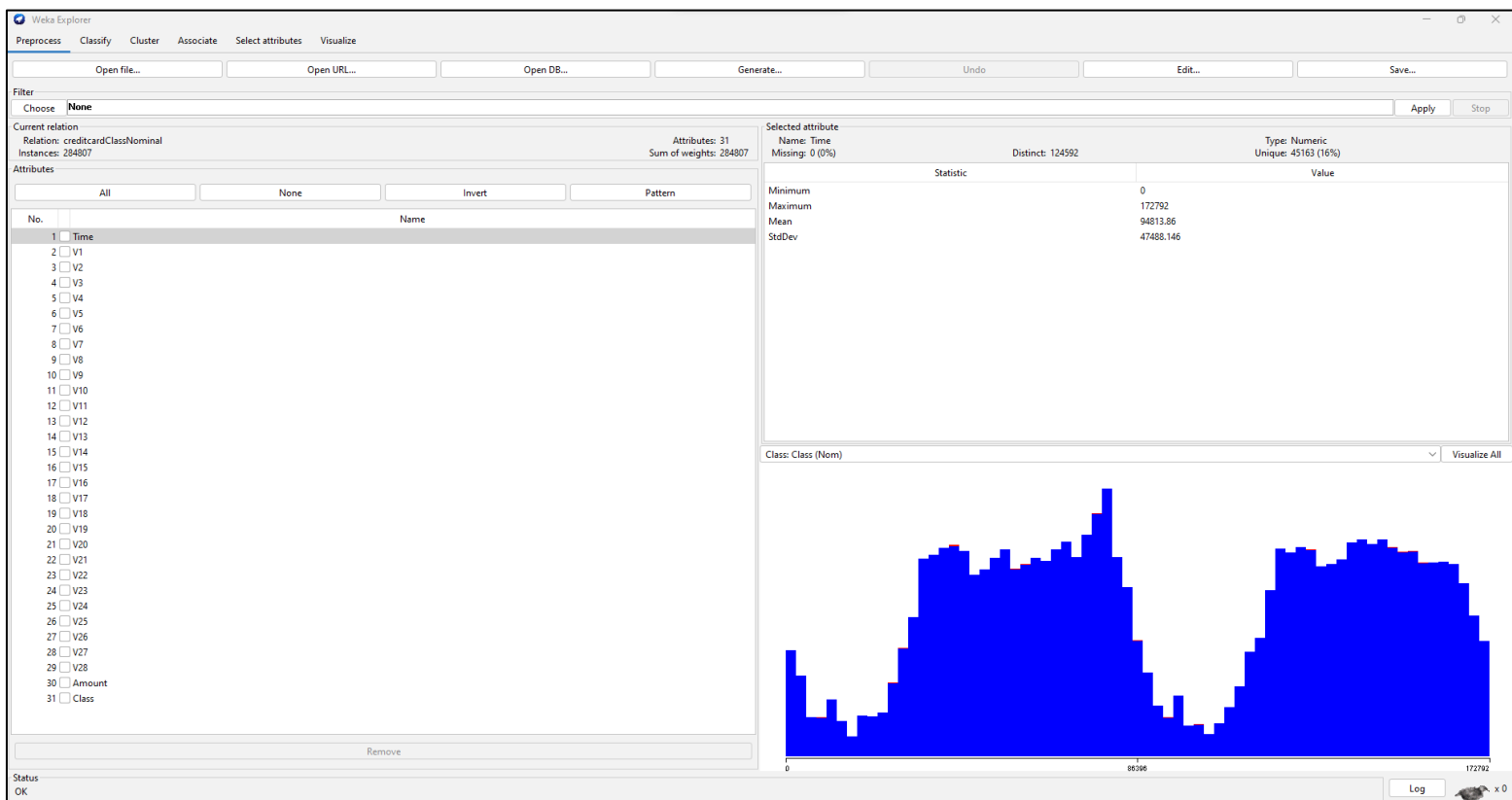
[284807 rows x 31 columns]

Data:

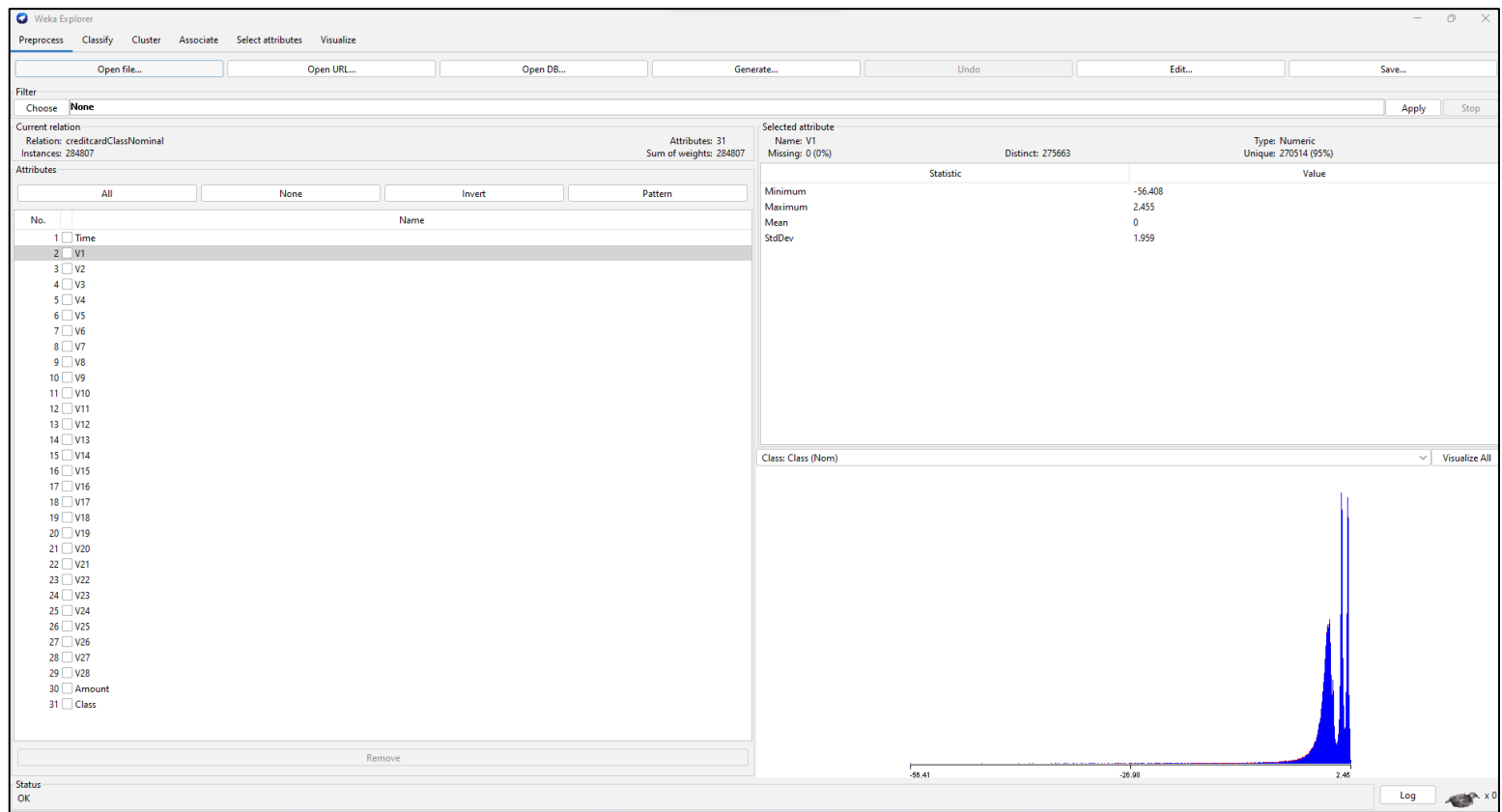
Dataset Name, Instances, Attribute, Sum of Weights

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

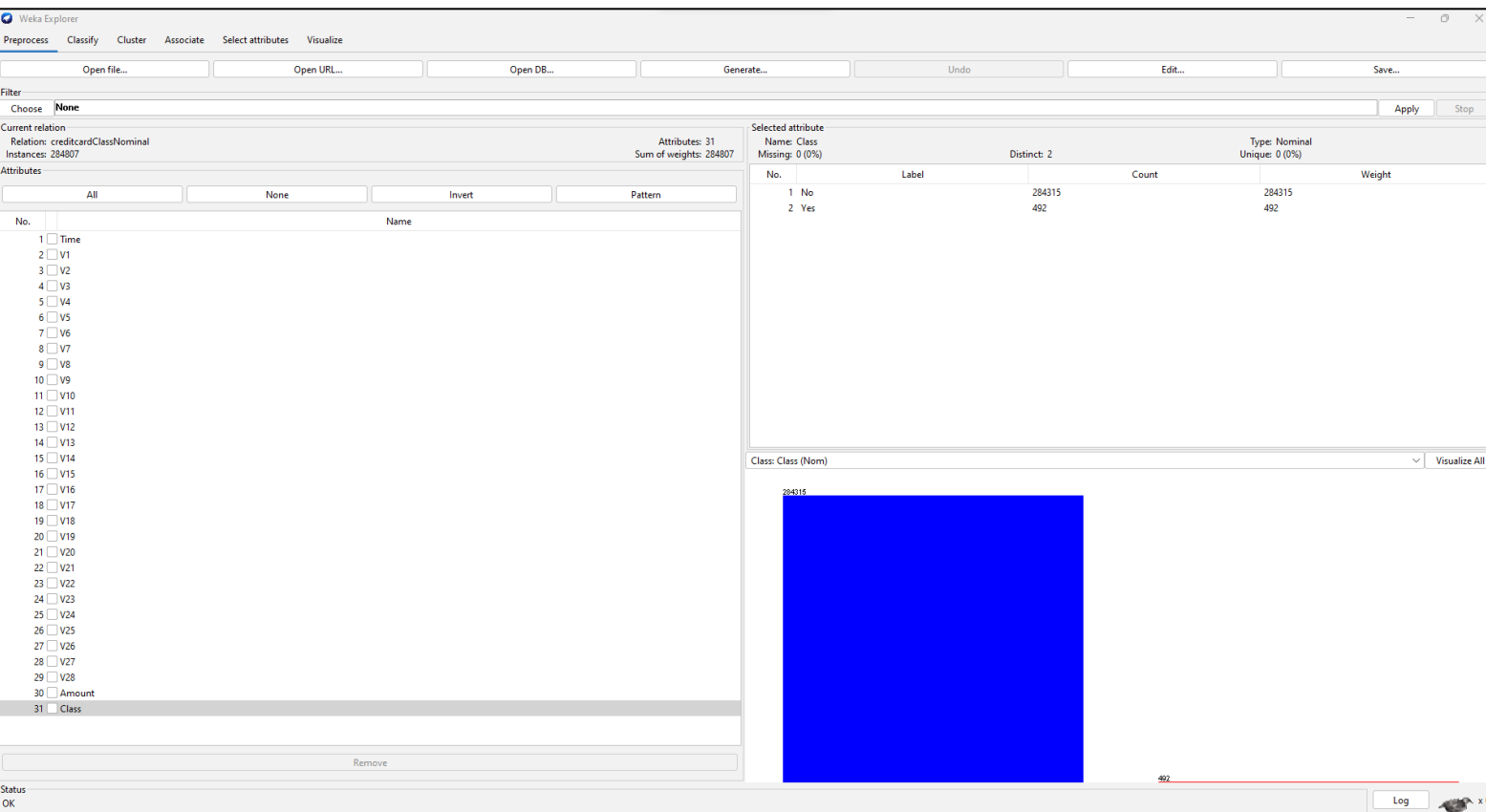
Statistics -> Minimum, Maximum, Mean & StdDev Value



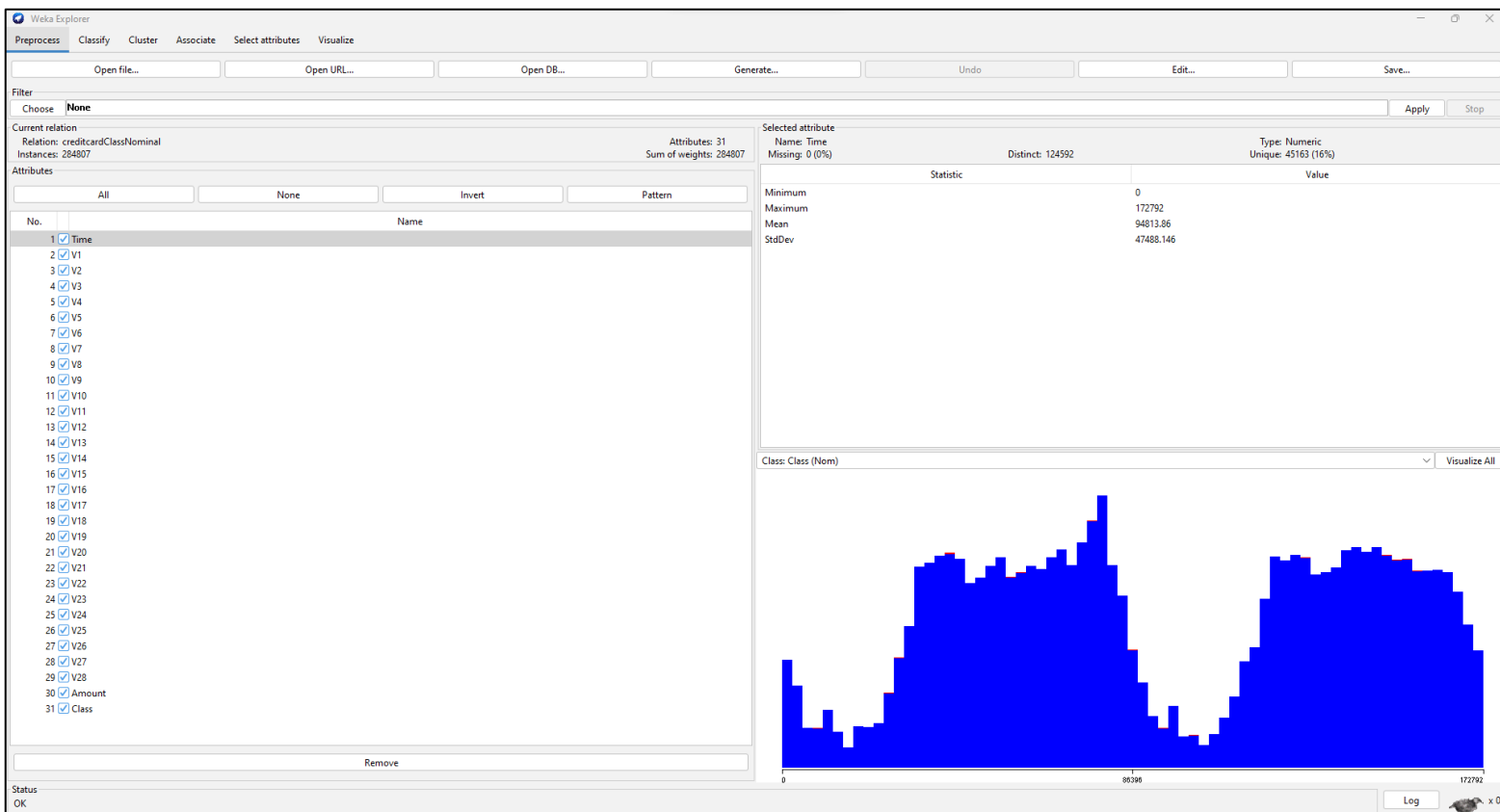
V1 Attribute:



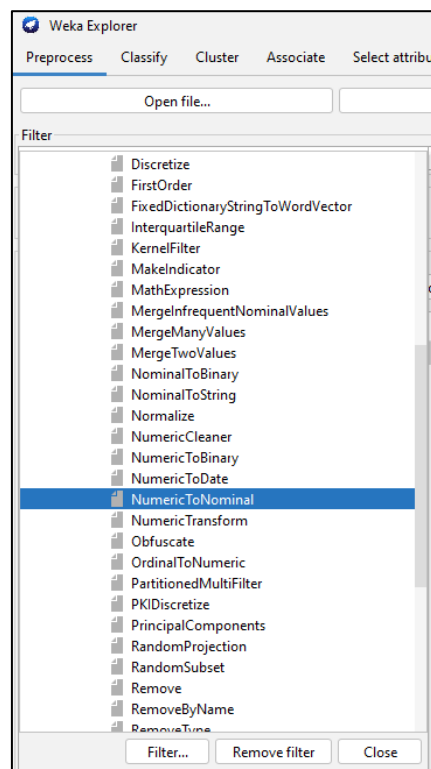
Class Attribute:



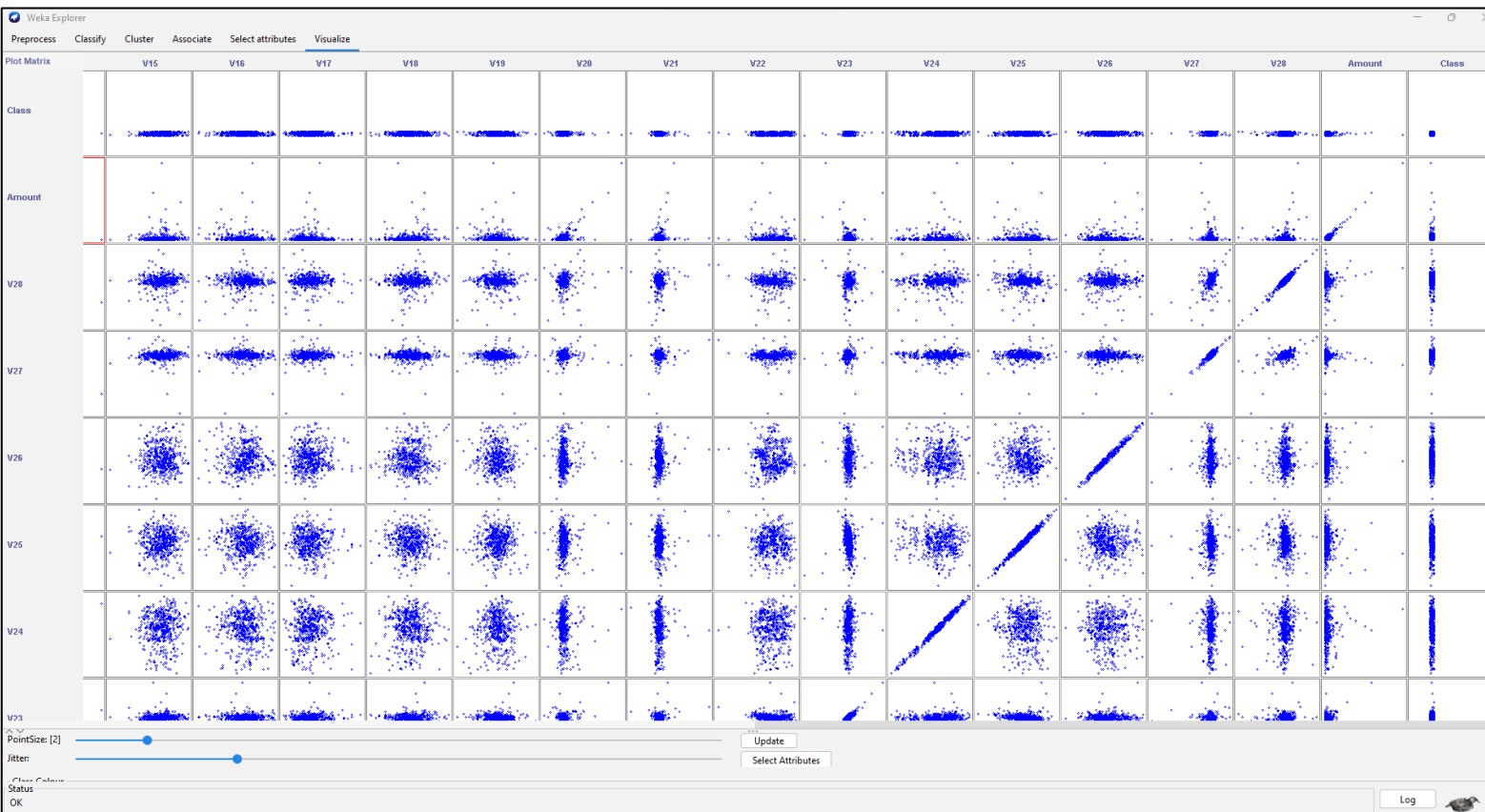
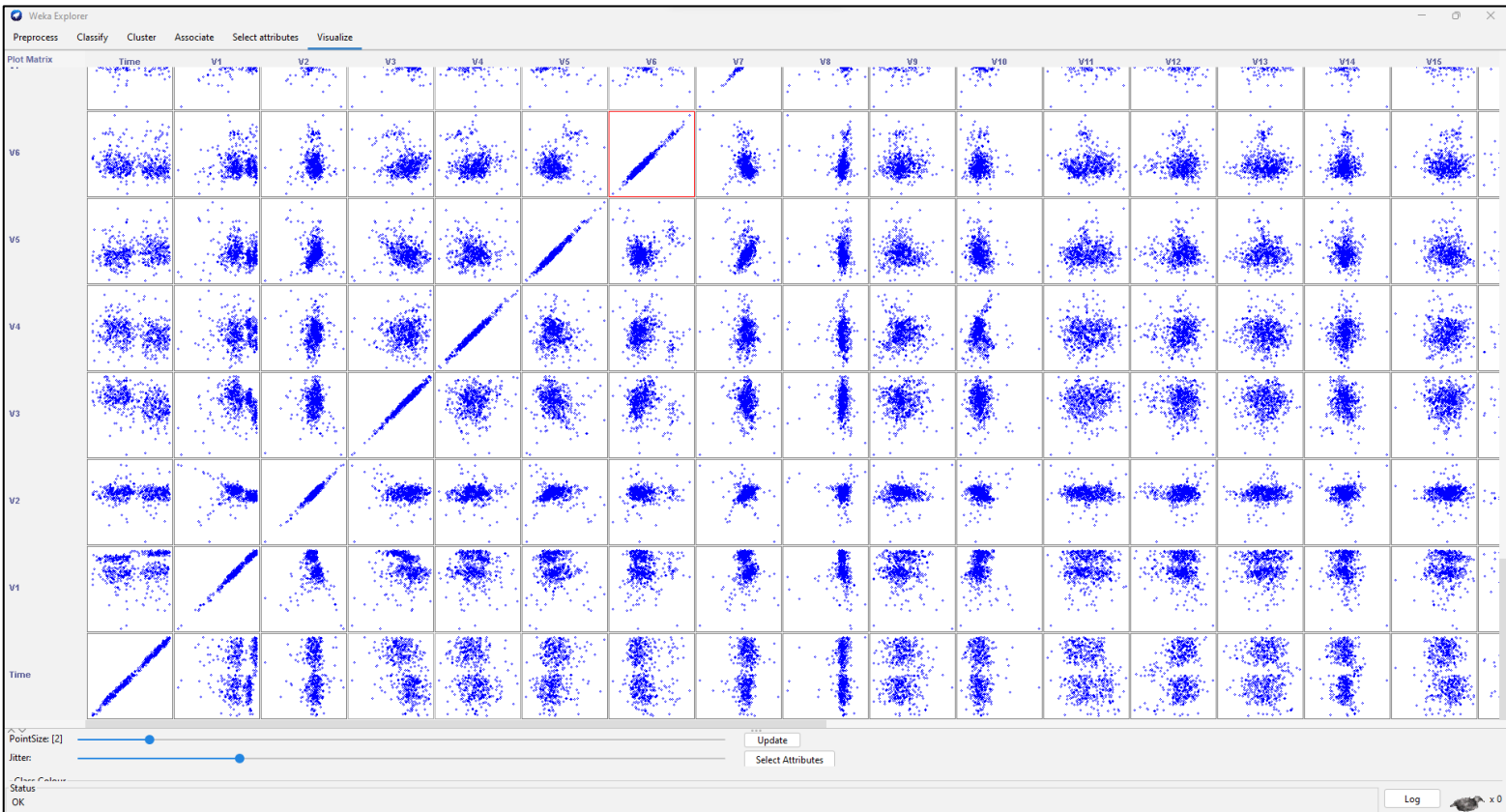
All Attribute:



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



Visualize:



Correlation

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Attribute Evaluator

Choose

CorrelationAttributeEval

Search Method

Choose

Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode

Use full training set

Cross-validation

Folds 10

Seed 1

No class

Start

Stop

Result list (right-click for options)

20:55:04 - Ranker + CorrelationAttributeEval

Attribute selection output

Class

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 31 Class):

Correlation Ranking Filter

Ranked attributes:

0.326481 18 V17

0.302544 15 V14

0.260593 13 V12

0.216883 11 V10

0.196539 17 V16

0.192961 4 V3

0.187257 8 V7

0.154876 12 V11

0.133447 5 V4

0.111485 19 V18

0.101347 2 V1

0.097733 10 V9

0.094974 6 V5

0.091289 3 V2

0.043643 7 V6

0.040413 22 V21

0.034783 20 V19

0.02009 21 V20

0.019875 9 V8

0.01758 28 V27

0.012323 1 Time

0.009536 29 V28

0.007221 25 V24

0.005632 30 Amount

0.00457 14 V13

0.004455 27 V26

0.004223 16 V15

0.003308 26 V25

0.002685 24 V23

0.000805 23 V22

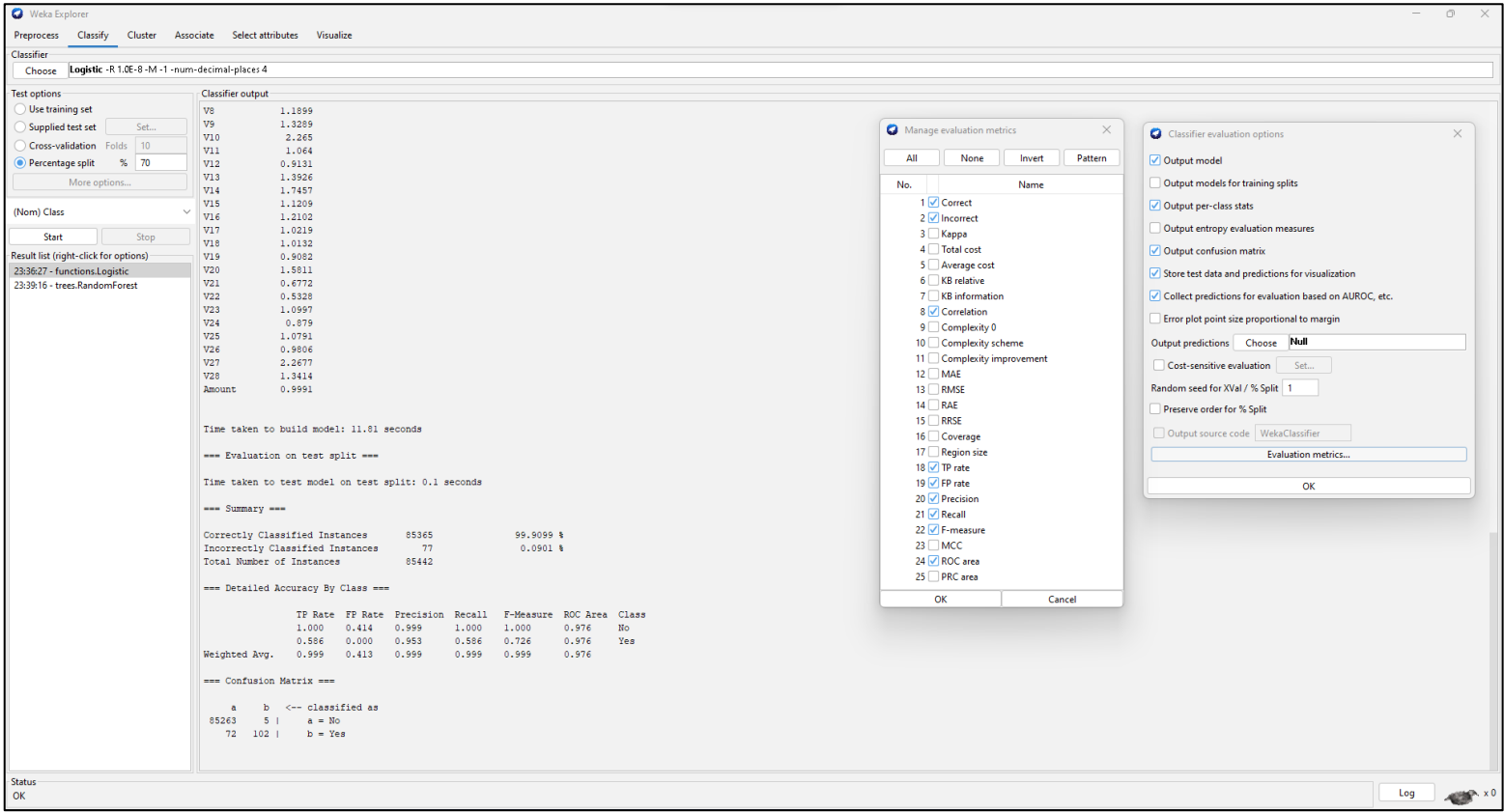
Selected attributes: 18,15,13,11,17,4,8,12,5,19,2,10,6,3,7,22,20,21,9,28,1,29,25,30,14,27,16,26,24,23 : 30

Status

OK

Classifier Results

Logistic Regression



Random Forest

