CreditCard Fraud Prediction

Overview:

Building a machine learning model to predict the credit card fraud data.
Here we are provinding a dataset having the previous creditcard fraud data.using that data we are training a machine by writing the code for machine in java programing.Data is processed to the model using Linear Regression, Logistic Regression , Clustering.... After Training the model we are going to test the model using related data.

purpose:

The credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase

Literature Survey:

here is the some of the data from the dataset regarding fraud prediction

Gender,Married,Dependents,Education,Self_Employed,ApplicantIncome,CoapplicantIncome,LoanAmount,Loan_Term,Credit_History_Available,Housing,Locality,Fraud_Risk
1,0,0,1,0,5849,0,146,360,1,1,1,0
1,1,1,1,1,4583,1508,128,360,1,1,3,1
1,1,0,1,1,3000,0,66,360,1,1,1,1
1,1,0,0,1,2583,2358,120,360,1,1,1,1
1,0,0,1,0,6000,0,141,360,1,1,1,0
1,1,2,1,1,5417,4196,267,360,1,0,1,1
1,1,0,0,1,2333,1516,95,360,1,1,1,1
1,1,3,1,1,3036,2504,158,360,0,1,2,1
1,1,2,1,1,4006,1526,168,360,1,1,1,1
1,1,1,1,1,12841,10968,349,360,1,0,2,1
1,1,2,1,1,3200,700,70,360,1,0,1,1
1,1,2,1,1,2500,1840,109,360,1,0,1,1
1,1,2,1,1,3073,8106,200,360,1,0,1,1
1,0,0,1,0,1853,2840,114,360,1,1,3,1
1,1,2,1,1,1299,1086,17,120,1,1,1,1
1,0,0,1,0,4950,0,125,360,1,1,1,0
1,0,1,0,0,3596,0,100,240,1,1,1,0

0,0,0,1,0,3510,0,76,360,0,1,1,1
1,1,0,0,1,4887,0,133,360,1,1,3,1
1,1,0,1,1,2600,3500,115,12,1,1,1,1
1,1,0,0,1,7660,0,104,360,0,1,1,1
1,1,1,1,1,5955,5625,315,360,1,0,1,1
1,1,0,0,1,2600,1911,116,360,0,1,2,1
0,1,2,0,1,3365,1917,112,360,0,0,3,1
1,1,1,1,1,3717,2925,151,360,1,0,2,1
1,1,0,1,1,9560,0,191,360,1,1,2,1
1,1,0,1,1,2799,2253,122,360,1,0,2,1
1,1,2,0,1,4226,1040,110,360,1,0,1,1
1,0,0,0,0,1442,0,35,360,1,1,1,1
0,0,2,1,1,3750,2083,120,360,1,1,2,0
1,1,1,1,1,4166,3369,201,360,1,1,1,1
1,0,0,1,0,3167,0,74,360,1,1,1,1
1,0,1,1,1,4692,0,106,360,1,0,3,1
1,1,0,1,1,3500,1667,114,360,1,1,2,1
1,0,3,1,0,12500,3000,320,360,1,1,3,1
1,1,0,1,1,2275,2067,146,360,1,0,1,1
1,1,0,1,1,1828,1330,100,12,0,1,1,1
0,1,0,1,1,3667,1459,144,360,1,1,2,1
1,0,0,1,0,4166,7210,184,360,1,1,1,0
1,0,0,0,0,3748,1668,110,360,1,0,2,0
1,0,0,1,0,3600,0,80,360,1,1,1,1
1,0,0,1,0,1800,1213,47,360,1,1,1,0
1,1,0,1,1,2400,0,75,360,1,0,1,1
1,1,0,1,1,3941,2336,134,360,1,0,2,1
1,1,0,0,1,4695,0,96,12,1,1,1,1
0,0,0,1,0,3410,0,88,12,1,0,1,0
1,1,1,1,1,5649,0,44,360,1,0,1,1
1,1,0,1,1,5821,0,144,360,1,1,1,1
0,1,0,1,1,2645,3440,120,360,0,1,1,1
0,0,0,1,0,4000,2275,144,360,1,1,2,0
0,1,0,0,1,1928,1644,100,360,1,0,2,1
0,0,0,1,0,3086,0,120,360,1,0,2,0
0,0,0,1,0,4230,0,112,360,1,1,2,1
1,1,2,1,1,4616,0,134,360,1,0,1,1
0,1,1,1,1,11500,0,286,360,0,0,1,1
1,1,2,1,1,2708,1167,97,360,1,1,2,1
1,1,0,1,1,2132,1591,96,360,1,1,2,1
1,1,0,1,1,3366,2200,135,360,1,1,3,1

1,1,1,1,1,8080,2250,180,360,1,1,1,1
1,1,2,0,1,3357,2859,144,360,1,1,1,1
1,1,0,1,1,2500,3796,120,360,1,0,1,1
1,1,3,1,1,3029,0,99,360,1,1,1,1
1,1,0,0,1,2609,3449,165,180,0,1,3,1
1,1,1,1,1,4945,0,146,360,0,0,3,1
0,0,0,1,0,4166,0,116,360,0,1,2,1
1,1,0,1,1,5726,4595,258,360,1,1,2,1

Existing Problem:

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

proposed solution:

Using Artificial Intelligence , Training a machine with previous data regarding fraud data there might be a very useful chances to companies to recognize fradulent credit transactions.

Theoretical Analysis:

I have used Eclipse IDE and Weka software to train the machine and vizualize the data.

Experimental Analysis:

```
package org1.ml;

        import java.io.IOException;

       import tech.tablesaw.api.Table;
        import tech.tablesaw.plotly.Plot;
         import tech.tablesaw.plotly.components.Figure;
        import tech.tablesaw.plotly.components.Layout;
        import tech.tablesaw.plotly.traces.BoxTrace;
        import tech.tablesaw.plotly.traces.HistogramTrace;

        public class Analysis {
IN:       public static void main(String args[]) {
```

```
        System.out.println("Creditcard data Analysis");
    }   }
```

OUT: Creditcard data Analysis

IN:
```
Table Creditcard_data =
Table.read().csv("C:\\Users\\Saketh\\eclipse-workspace\\org1.ml\\src\\main\\java\\org1\\ml\
\fraud_dataset.csv");
                System.out.println(Creditcard_data.structure());

                System.out.println(Creditcard_data.summary());
```

OUT:

Index	Column Name	Column Type
0	Gender	INTEGER
1	Married	INTEGER
2	Dependents	INTEGER
3	Education	INTEGER
4	Self_Employed	INTEGER
5	ApplicantIncome	INTEGER
6	CoapplicantIncome	INTEGER
7	LoanAmount	INTEGER
8	Loan_Term	INTEGER
9	Credit_History_Available	INTEGER
10	Housing	INTEGER
11	Locality	INTEGER
12	Fraud_Risk	INTEGER

fraud_dataset.csv

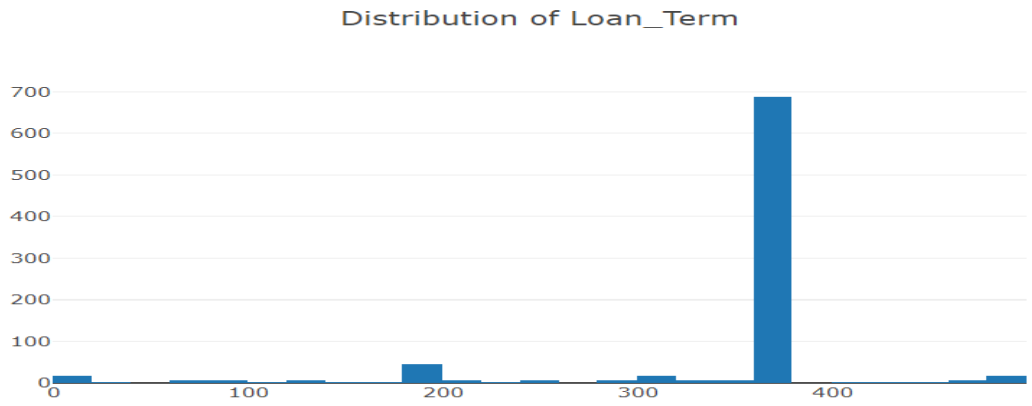
Summary	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_T
Count	827	827	827	827	827	827	827	827	
sum	607	398	540	654	475	4311127	1228964	116518	
Mean	0.73397823458283	0.48125755743651766	0.6529625151148738	0.7908101571946797	0.5743651753325266	5212.970979443772	1486.0507859733966	140.89238210399026	338.12817
Min	0	0	0	0	0	150	0	9	
Max	1	1	3	1	1	81000	41667	700	
Range	1	1	3	1	1	80850	41667	691	
Variance	0.1954905709542645	0.24995095900758577	0.8757872177215122	0.16562973025990263	0.24476578900369195	31289628.52700187	7855956.815819599	6371.3043849966675	5678.097
Std. Dev	0.44214315662946146	0.4999509566023309	0.9358350376650322	0.4069763264121178	0.49473810142710045	5593.713303969186	2802.8479830022175	79.82045091952736	75.35156

IN: //// Histogram

```
                Layout layout1=Layout.builder().title("Distribution of
```

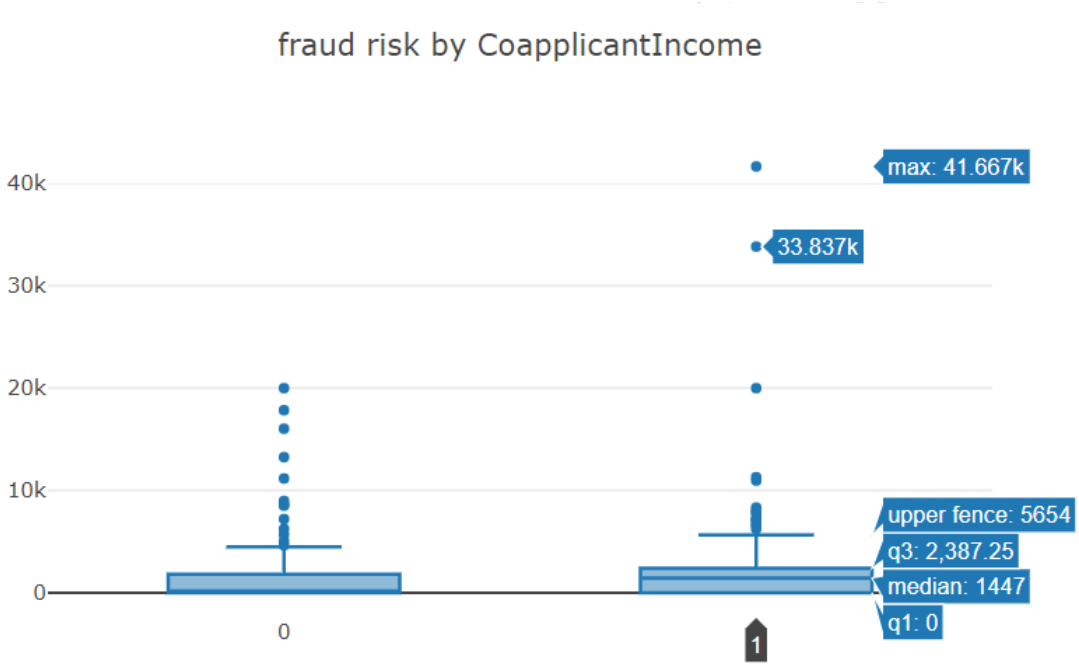
Loan_Term").build();
HistogramTrace trace1= HistogramTrace.builder(Creditcard_data.nCol("Loan_Term")).build();
Plot.show(new Figure(layout1, trace1));

OUT:



IN: Layout layout3 = Layout.builder().title("fraud risk by ApplicantIncome").build();
BoxTrace trace3 =BoxTrace.builder(Creditcard_data.categoricalColumn("Fraud_Risk"), Creditcard_data.nCol("ApplicantIncome")).build();
Plot.show(new Figure(layout3, trace3));

OUT:



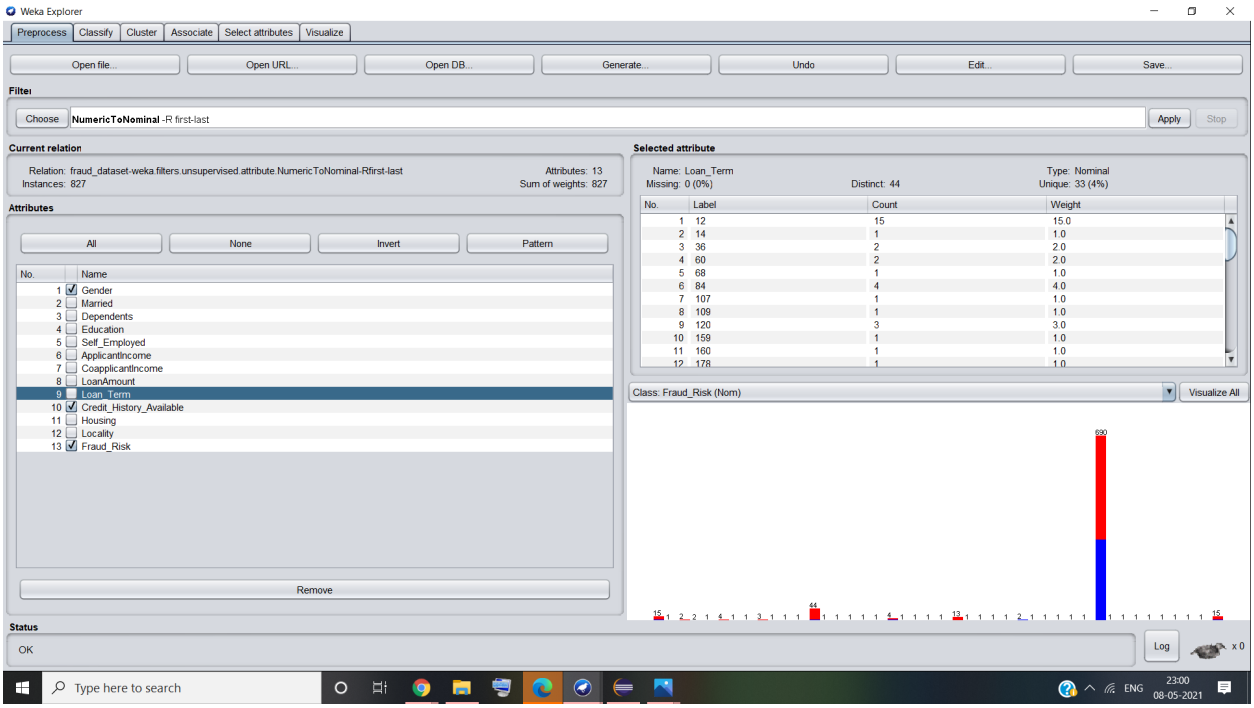
IN: //linear Regression

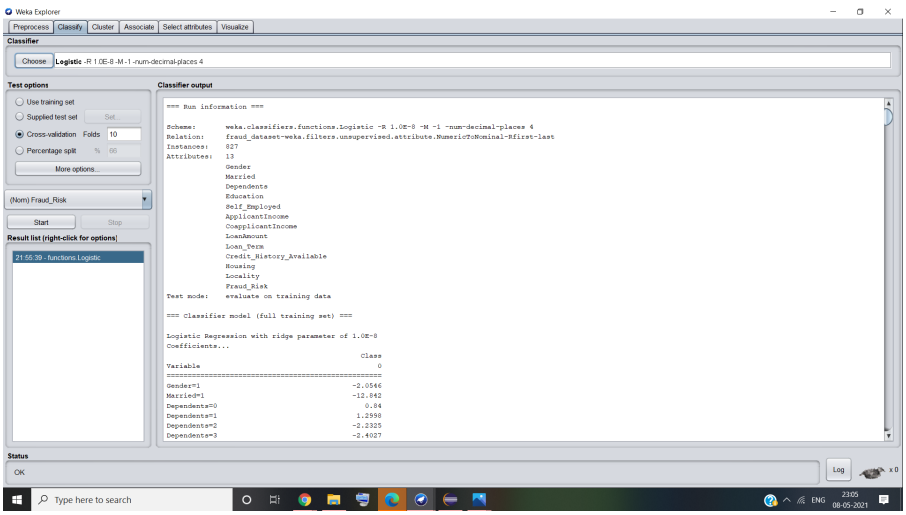
```
LinearRegression lr=new LinearRegression();
lr.buildClassifier(dataset);

Evaluation lreval =new Evaluation(dataset);
lreval.evaluateModel(lr,dataset);
System.out.println(lreval.toSummaryString());
```

OUT:

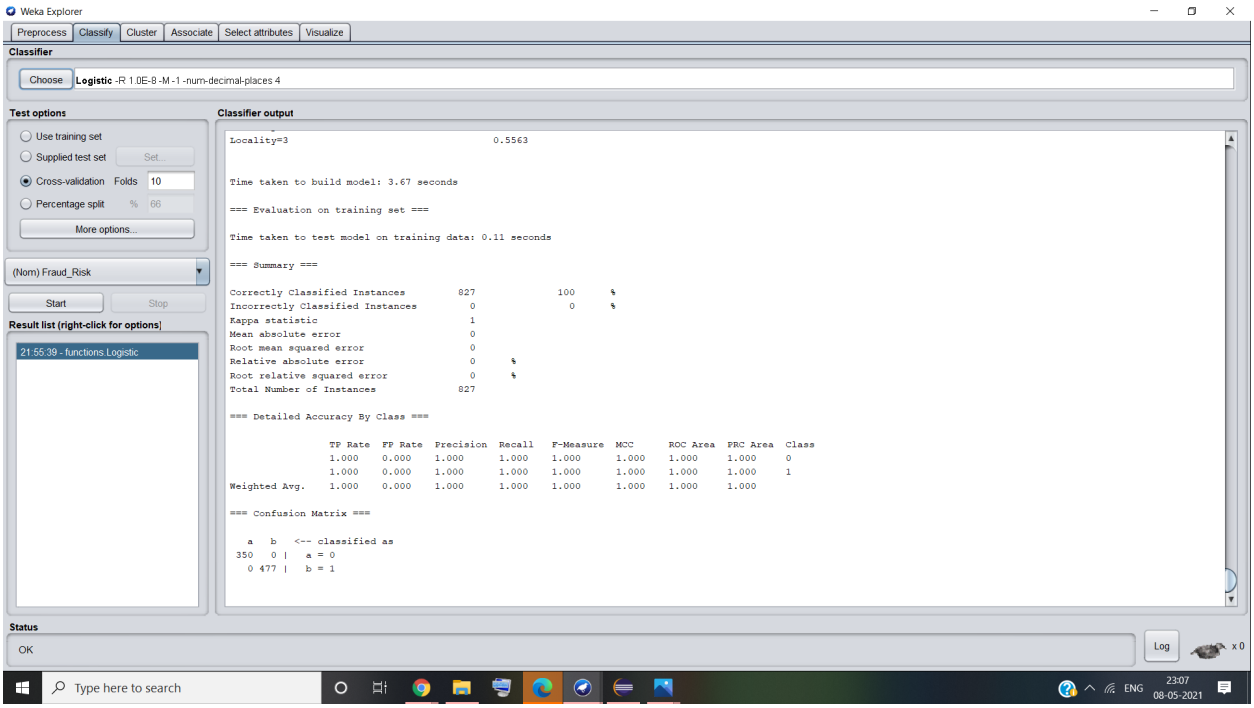
```
Correlation coefficient                  0.8428
Mean absolute error                      0.1727
Root mean squared error                  0.2659
Relative absolute error                 35.3768 %
Root relative squared error             53.827  %
Total Number of Instances              827
```

// **Log Regression**

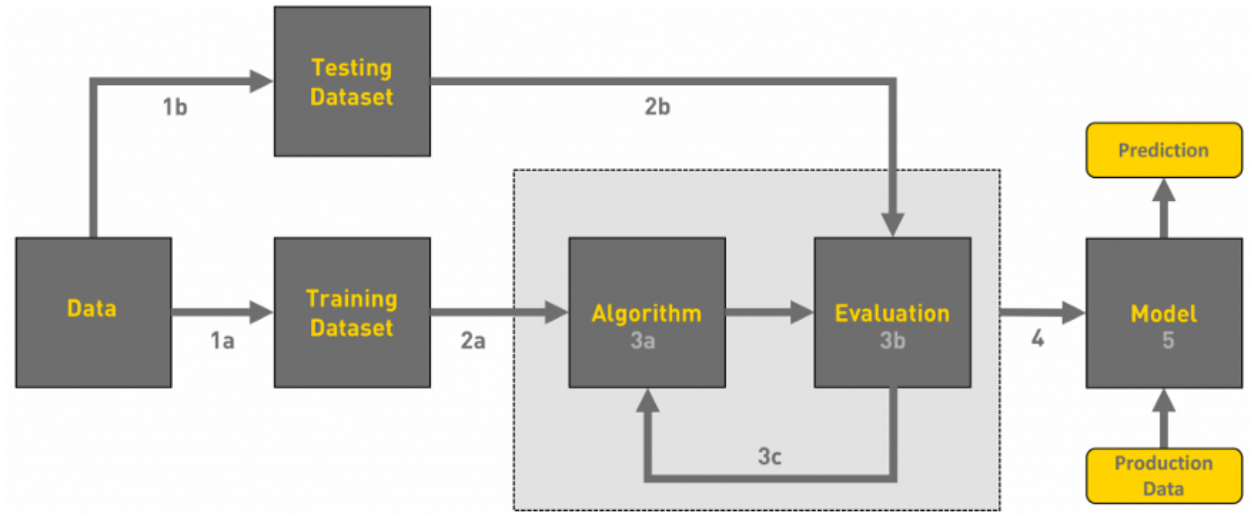




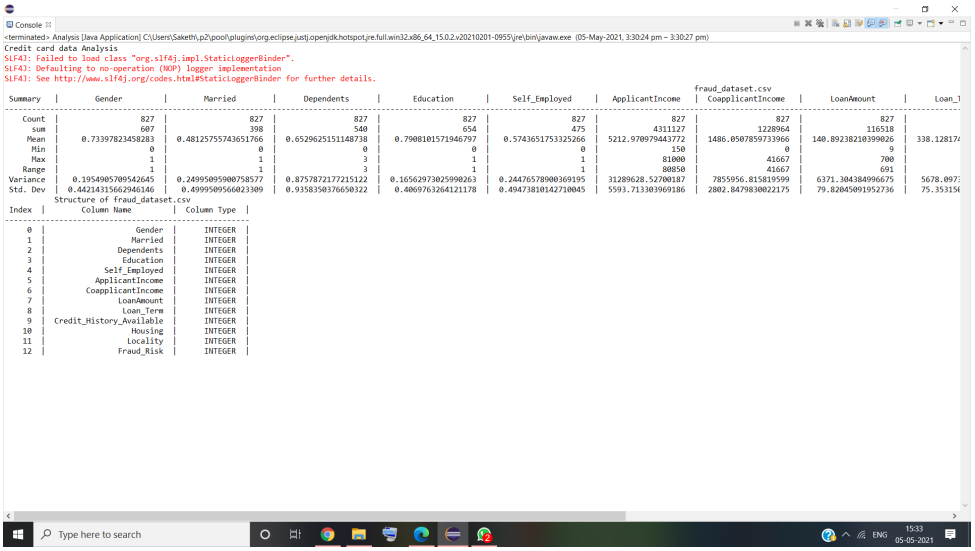
//Logistic using weka explorer

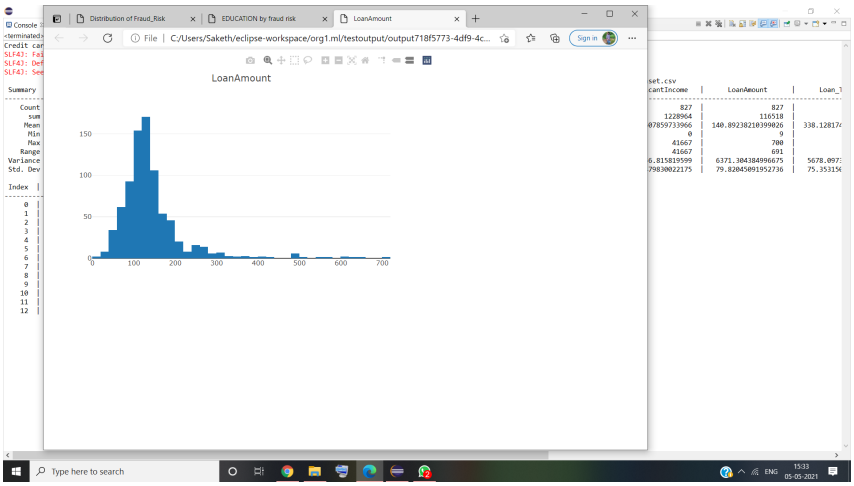


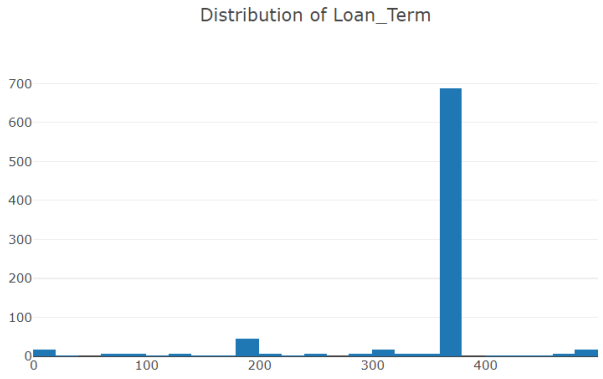
FLOW CHART:

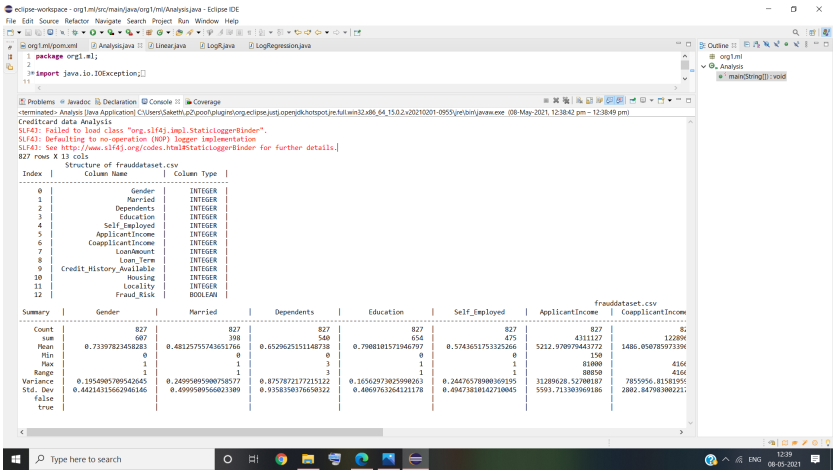


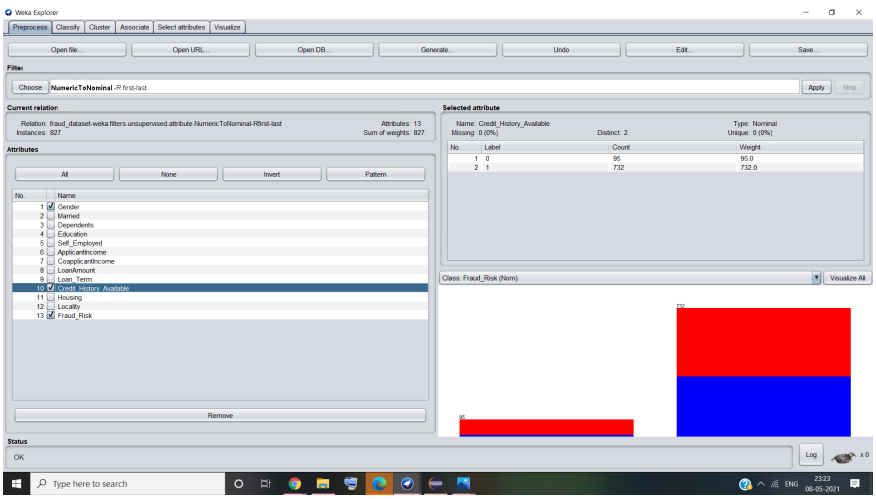
Result:

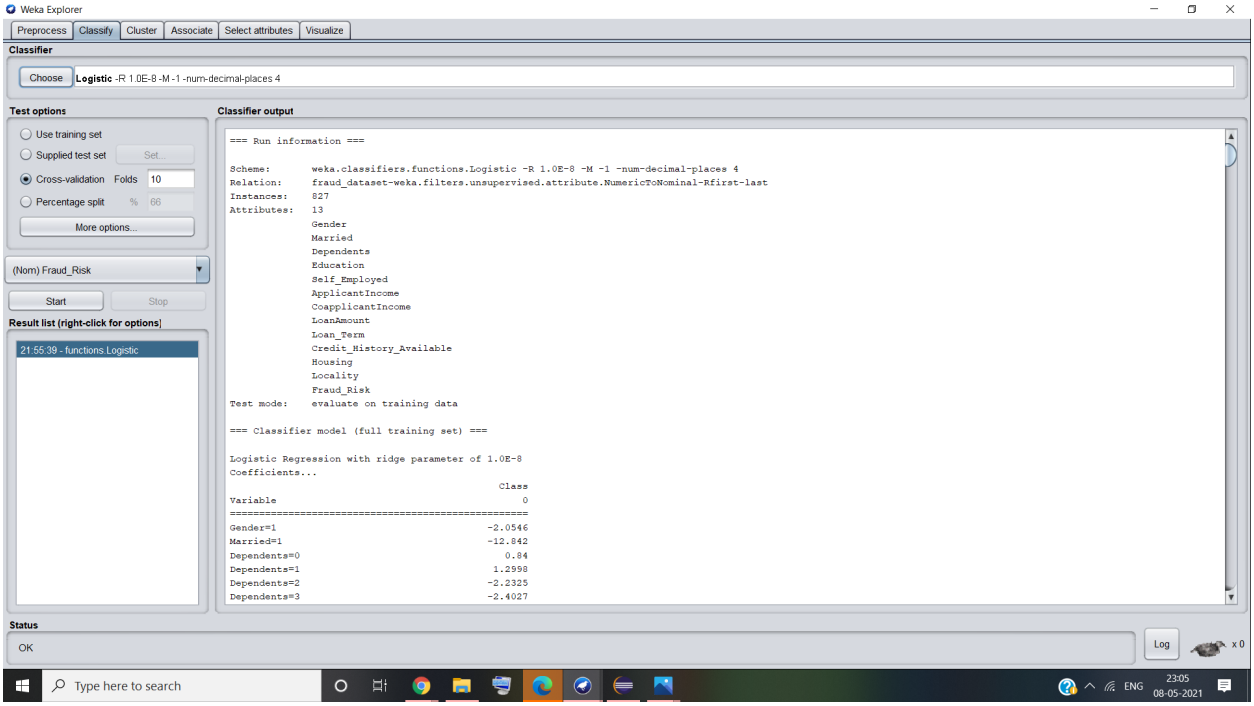

Console
<terminated> Analysis [Java Application] C:\Users\Saketh\.p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full.win32.x86_64_15.0.2.v20210201-0955\jre\bin\javaw.exe (05-May-2021, 3:30:24 pm – 3:30:27 pm)
Credit card data Analysis
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
fraud_dataset.csv
Summary | Gender | Married | Dependents | Education | Self_Employed | ApplicantIncome | CoapplicantIncome | LoanAmount | Loan_T
Count | 827 | 827 | 827 | 827 | 827 | 827 | 827 | 827 |
sum | 607 | 398 | 540 | 654 | 475 | 4311127 | 1228964 | 116518 |
Mean | 0.73397823458283 | 0.48125755743651766 | 0.6529625151148738 | 0.7908101571946797 | 0.5743651753325266 | 5212.970979443772 | 1486.0507859733966 | 140.89238210399026 | 338.12817
Min | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 9 |
Max | 1 | 1 | 3 | 1 | 1 | 81000 | 41667 | 700 |
Range | 1 | 1 | 3 | 1 | 1 | 80850 | 41667 | 691 |
Variance | 0.1954905709542645 | 0.2499509590758577 | 0.8757872177215122 | 0.16562973025990263 | 0.24476578900369195 | 31289628.52700187 | 7855956.815819599 | 6371.304384996675 | 5678.097
Std. Dev | 0.44214315662946146 | 0.4999509566023309 | 0.9358350376650322 | 0.4069763264121178 | 0.49473810142710045 | 5593.713303969186 | 2802.8479830022175 | 79.82045091952736 | 75.35315
Structure of fraud_dataset.csv
Index | Column Name | Column Type
0 | Gender | INTEGER
1 | Married | INTEGER
2 | Dependents | INTEGER
3 | Education | INTEGER
4 | Self_Employed | INTEGER
5 | ApplicantIncome | INTEGER
6 | CoapplicantIncome | INTEGER
7 | LoanAmount | INTEGER
8 | Loan_Term | INTEGER
9 | Credit_History_Available | INTEGER
10 | Housing | INTEGER
11 | Locality | INTEGER
12 | Fraud_Risk | INTEGER
Type here to search
ENG
15:33
05-05-2021

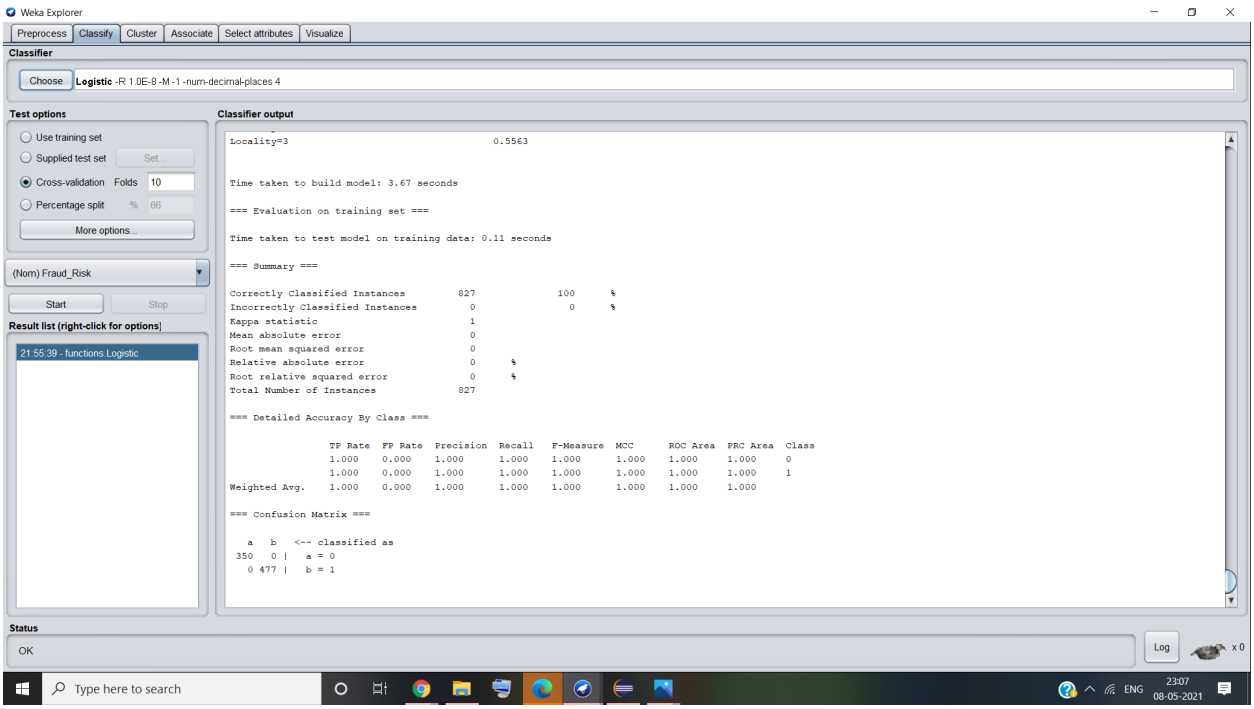

Distribution of Fraud_Risk
EDUCATION by fraud risk
LoanAmount
File | C:/Users/Saketh/eclipse-workspace/org1.ml/testoutput/output718f5773-4df9-4c...
Sign in
LoanAmount
150
100
50
0
0
100
200
300
400
500
600
700
Type here to search
ENG
15:33
05-05-2021


Distribution of Loan_Term
700
600
500
400
300
200
100
0
0
100
200
300
400

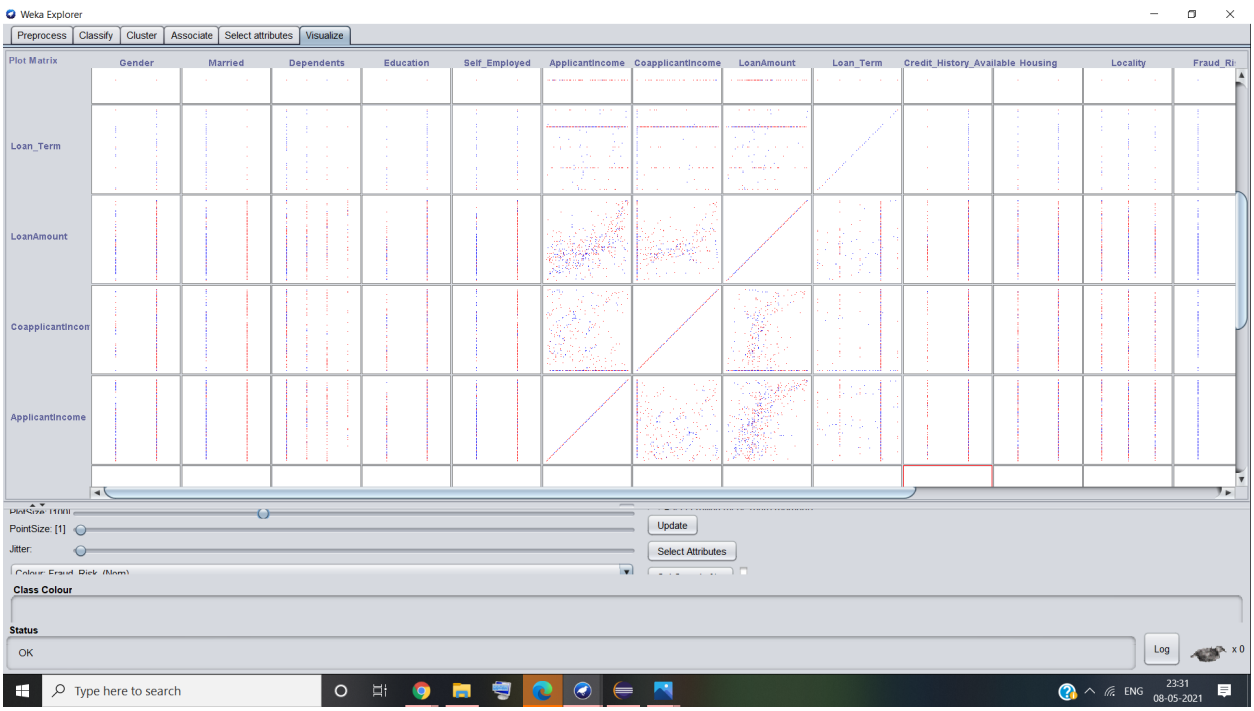

eclipse-workspace - org1.ml/src/main/java/org1/ml/Analysis.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Creditcard data Analysis
827 rows X 13 cols
Structure of frauddataset.csv
Index | Column Name | Column Type
0 | Gender | INTEGER
1 | Married | INTEGER
2 | Dependents | INTEGER
3 | Education | INTEGER
4 | Self_Employed | INTEGER
5 | ApplicantIncome | INTEGER
6 | CoapplicantIncome | INTEGER
7 | LoanAmount | INTEGER
8 | Loan_Term | INTEGER
9 | Credit_History_Available | INTEGER
10 | Housing | INTEGER
11 | Locality | INTEGER
12 | Fraud_Risk | BOOLEAN
frauddataset.csv
Summary | Gender | Married | Dependents | Education | Self_Employed | ApplicantIncome | CoapplicantIncome
Count | 827 | 827 | 827 | 827 | 827 | 827 |
sum | 607 | 398 | 540 | 654 | 475 | 4311127 |
Mean | 0.7339782459283 | 0.48125755743651766 | 0.6529625151148738 | 0.7908101571946797 | 0.5743651753325266 | 5212.970970443772 | 1486.0507859733
Min | 0 | 0 | 0 | 0 | 0 | 150 |
Max | 1 | 1 | 1 | 1 | 1 | 81000 |
Range | 1 | 1 | 1 | 1 | 1 | 80850 |
Variance | 0.1954905709542645 | 0.2499509590758577 | 0.875787217215122 | 0.16562973025990263 | 0.24476578900369195 | 31289628.527001187 | 7855956.81501595
Std. Dev | 0.44214315662946146 | 0.4999509566023309 | 0.9358350376650322 | 0.4069763264121178 | 0.4947381014271004 | 5593.713303969186 | 2802.847983002217
false | | | | | | |
true | | | | | | |
Type here to search
ENG 12:39 08-05-2021


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Open file... Open URL... Open DB... Generate... Undo Edit... Save...
Filter
Choose NumericToNominal -R first-last
Apply Stop
Current relation
Relation: fraud_dataset-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 827
Attributes: 13
Sum of weights: 827
Selected attribute
Name: Credit_History_Available
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No. Label Count Weight
1 0 95 95.0
2 1 732 732.0
Attributes
All None Invert Pattern
No. Name
1 Gender
2 Married
3 Dependents
4 Education
5 Self_Employed
6 ApplicantIncome
7 CoapplicantIncome
8 LoanAmount
9 Loan_Term
10 Credit_History_Available
11 Housing
12 Locality
13 Fraud_Risk
Remove
Class: Fraud_Risk (Nom)
Visualize All
Status
OK
Log
x 0
Type here to search
ENG 23:23 08-05-2021





Visualization using weka..



Advantages and Disadvantages:

Using this model not always give the accurate data. There is no thinking power for a machine. If the trained data is different from present appplicable data there would be lot of issues in dealing with the Amount.

But by using this models large data can be calculated in a short period of time.

Applications:

This data is applicable for banking sectors in the case of predicting fraud transactions

Conclusion:

here we have chosen a dataset and trained the machine with the data. by the the model can understand what is the fraud transaction and which is approved transaction.
from the data provided I concluded That data is not linearly regressed.

Future scope:

In future machines are going to lead the world.Man power is going to replace with machine intelligence....