

# Creditcard Fraud Prediction

## Overview

Credit Card Fraud Prediction is the process which is used to predict fraudulent transactions. As this prediction is important that is related to finance section.

The credit card fraud detection features uses user behavior and location scanning to check for unusual patterns. These patterns include user characteristics such as user spending patterns as well as usual user geographic locations to verify his identity. If any unusual pattern is detected, the system requires reverification.

Linear Regression is used in this project to know whether the transactions are fraud or not.

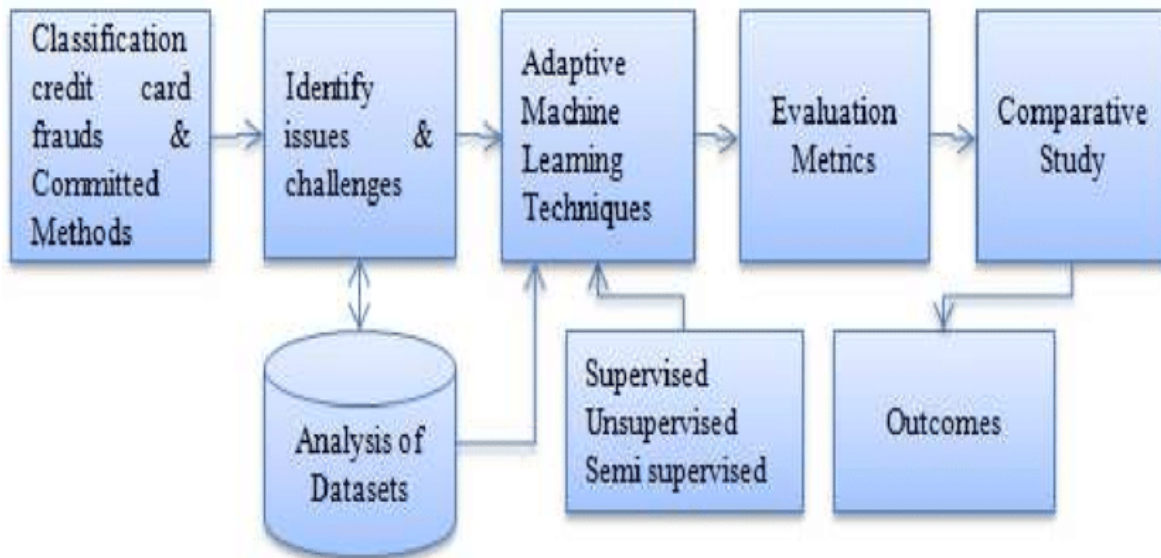
### Software Requirements:

- Windows Xp, Windows 7(ultimate, enterprise)
- Sql 2008
- Visual studio 2010

### Hardware Components:

- Processor – i3,i5
- Min Hard Disk – 4 GB
- Min Memory – 4GB RAM

## Block Diagram for the Credit Card fraud Detection



## Structure of the Credit card DataSet

OracleProject - org.ml/src/main/java/org/ml/regression.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Project Explorer

- org.ml
  - src/main/java
    - org.ml
      - ccard.java
      - regression.java
    - src/main/resources
    - src/test/java
    - src/test/resources
    - JRE System Library [J2SE-1.5]
    - Maven Dependencies
    - src
      - main
        - java
          - org
            - ml
              - ccard.java
              - regression.java
        - resources
        - test
        - target
        - testoutput
        - pom.xml

regression.java

```

31 System.out.println(train_data.size());
32 DataSource source=new DataSource("C:\\Users\\saish\\Desk
33 Instances dataset=source.getDataset();
34 dataset.setClassIndex(dataset.numAttributes()-1);
35 LinearRegression lr=new LinearRegression();
36 lr.buildClassifier(dataset);
37 System.out.println(lr);
38
39 Evaluation lreval=new Evaluation(dataset);
40 lreval.evaluateModel(lr, dataset);
41 System.out.println(lreval.toSummaryString());
42
43 Classifier classifier = new weka.classifiers.functions.S
44
45 System.out.print(" the expression for the input data as
  
```

ccard.java

```

19 import tech.tablesaw.plotly.traces.HistogramTrace;
20 import static tech.tablesaw.aggregate.AggregateFunctions.sum;
21
22 public class ccard {
23     try {
24         Table data=Table.read().csv("C:\\Users\\saish\\Desktop\\OracleProject\\ccard.csv");
25         System.out.println(data.structure());
26         System.out.println("First 6 of DataSet values : "+data.first(6));
27         System.out.println("Last 6 DataSet values : "+data.last(6));
28         System.out.println("DATASET null : "+data.dropRowsWithMissingValues() );
29         System.out.println(data.sampleN(200));//=> random 200 rows
30
31         Layout layout1 = Layout.builder().title("Distribution Over draft").build();
32
33     }
34 }
  
```

Structure of creditcard.csv

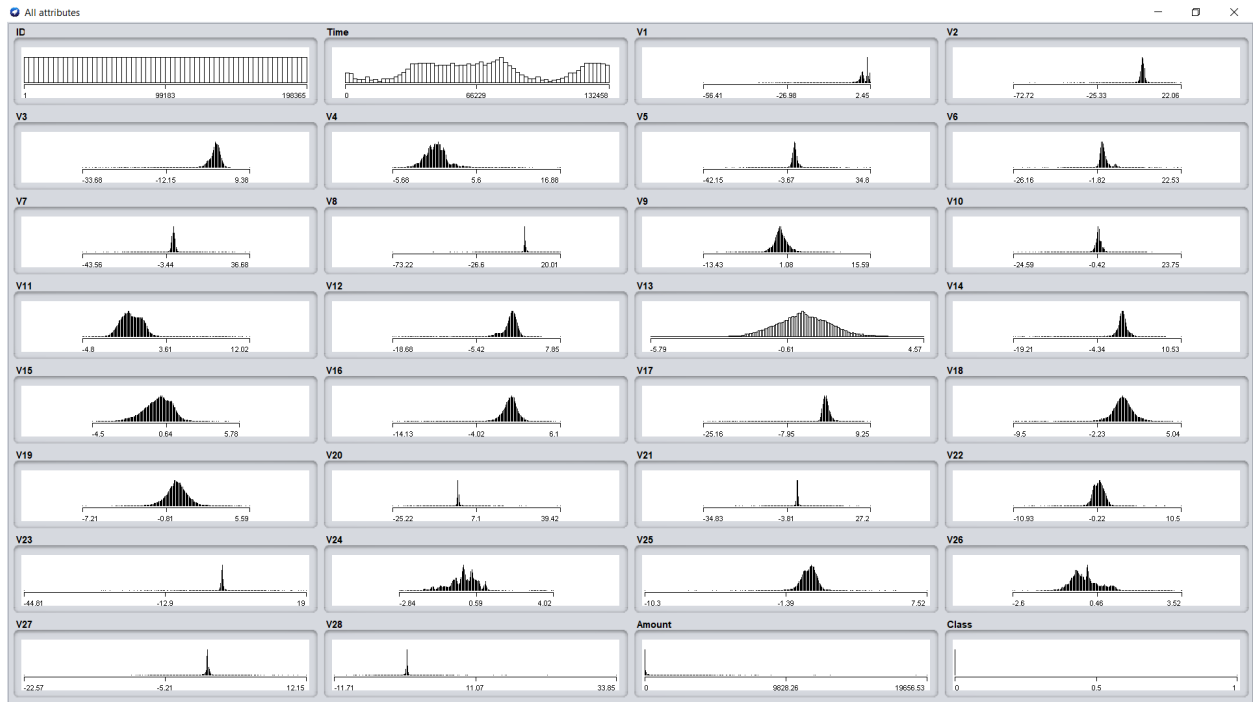
Index	Column Name	Column Type
0	Time	INTEGER
1	V1	DOUBLE
2	V2	DOUBLE
3	V3	DOUBLE
4	V4	DOUBLE
5	V5	DOUBLE
6	V6	DOUBLE
7	V7	DOUBLE
8	V8	DOUBLE
9	V9	DOUBLE
...	...	...
21	V21	DOUBLE
22	V22	DOUBLE
23	V23	DOUBLE
24	V24	DOUBLE
25	V25	DOUBLE
26	V26	DOUBLE
27	V27	DOUBLE
28	V28	DOUBLE
29	Amount	DOUBLE
30	Class	INTEGER

First 6 of DataSet values :

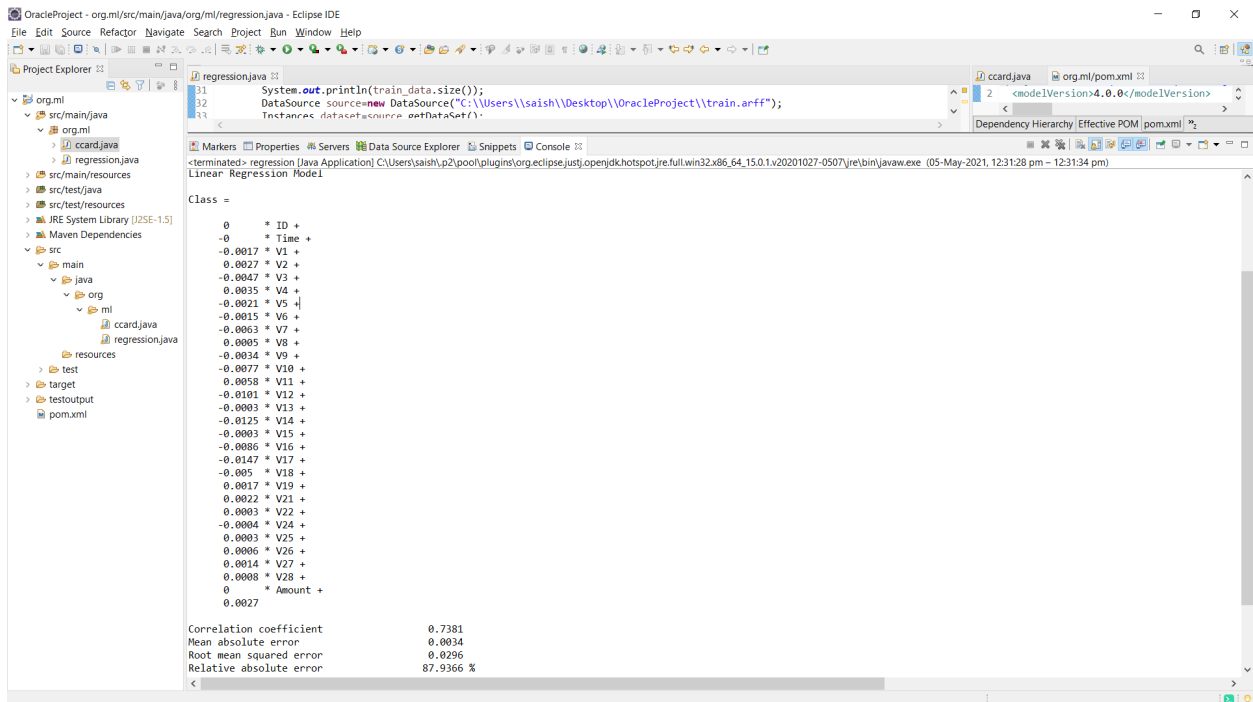
Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10

Writeable Smart Insert 47:9:1533

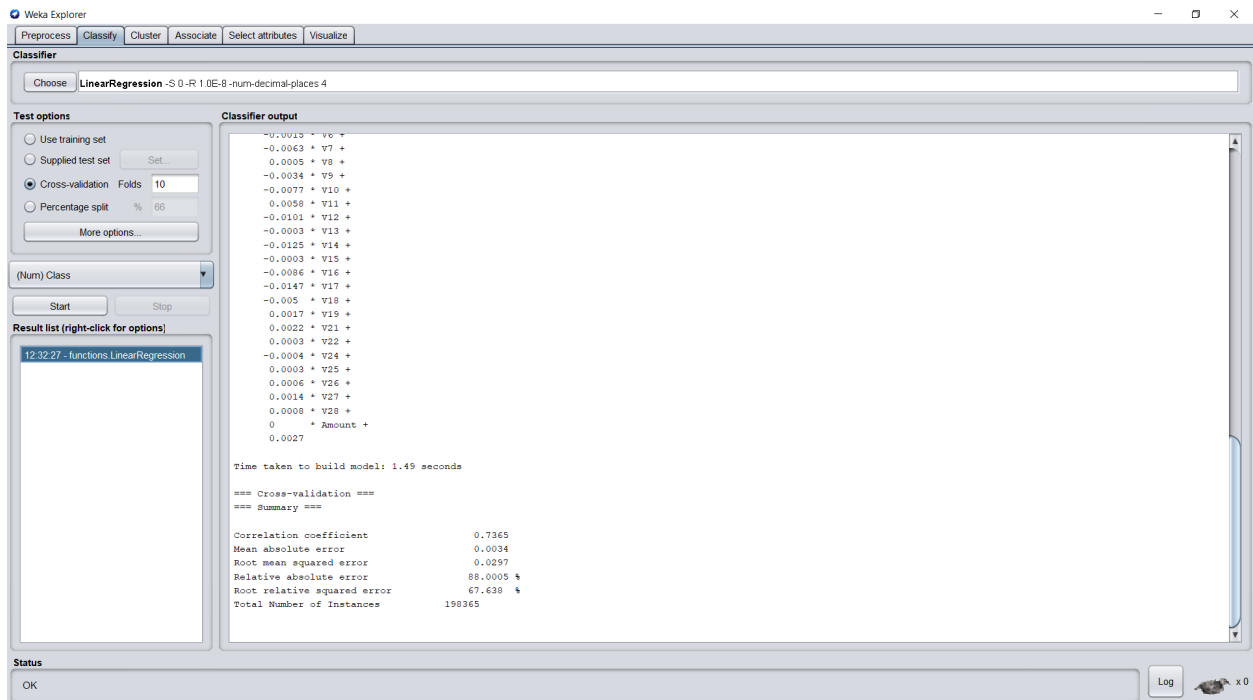
## Visualization of Dataset(All attributes)



## Linear Regression(The solution is used for this project)



## The Linear Regression solution through WEKA



## Advantages & Disadvantages:

### Advantages:

- Linear Regression is simple to implement and easier to interpret the output coefficients.
- When we know the relationship between the independent and dependent variable have a linear relationship, this 'Linear Regression' is the best to use because of its less complexity compared to the other algorithms.

### Disadvantages:

- Linear regression technique outliers can have huge effects on the regression and boundaries are linear in this technique.
- Diversely, linear regression assumes a linear relationship between dependent and independent variables. That means it assumes that there is a straight-line relationship between them. It assumes independence between attributes.

## Conclusion:

It is important for credit card companies to be able to recognize fraudulent credit card transactions so that customers may not face the issue regarding the transactions. So, this project helps to predict whether the transactions are fraud or not.

### ***References:***

Dataset for Credit card fraud prediction:

<https://www.kaggle.com/mlg-ulb/creditcardfraud>

Source code:

<https://github.com/smartinternz02/SPS-10710-Creditcard-Fraud-Prediction->