

Creditcard Fraud Prediction

Introduction

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 use 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.

LITERATURE SURVEY

Existing problem

Methods to solve this problem are:

- Logistic Regression
- Decision Tree
- Random Forest
- Naive Bayes
- ANN Model

Proposed solution

Solution that I suggest for the project is Linear Regression.

Linear Regression works with sigmoid function because the sigmoid function can be used to classify the output that is dependent feature and it uses the probability for classification of the dependent feature.

This algorithm works well with less amount of data set because of the use of sigmoid function. If the value of the sigmoid function is greater than 0.5 the output will be 1. If the output of the sigmoid function is less than 0.5 then the output is considered as 0.

Hardware/Software designing

Software Requirements:

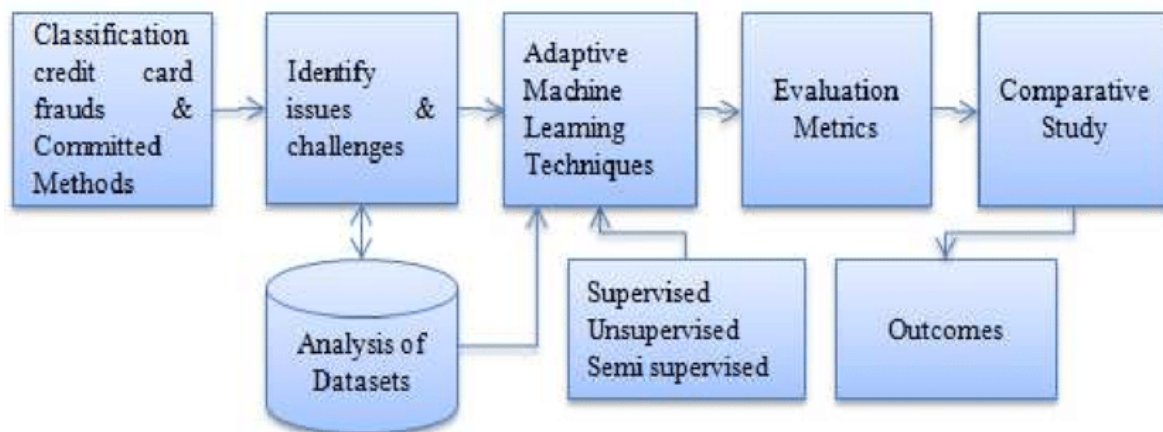
- Java JDK10
- Weka
- Eclipse IDE

Hardware Components:

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

THEORITICAL ANALYSIS

Block Diagram for the Credit Card fraud Detection



Experimental Investigations

Analysis is using in this project to make sure which method is best to know or predict the fraudulent transactions.

Result

Structure of the Credit card DataSet

The screenshot shows the Eclipse IDE with the following components:

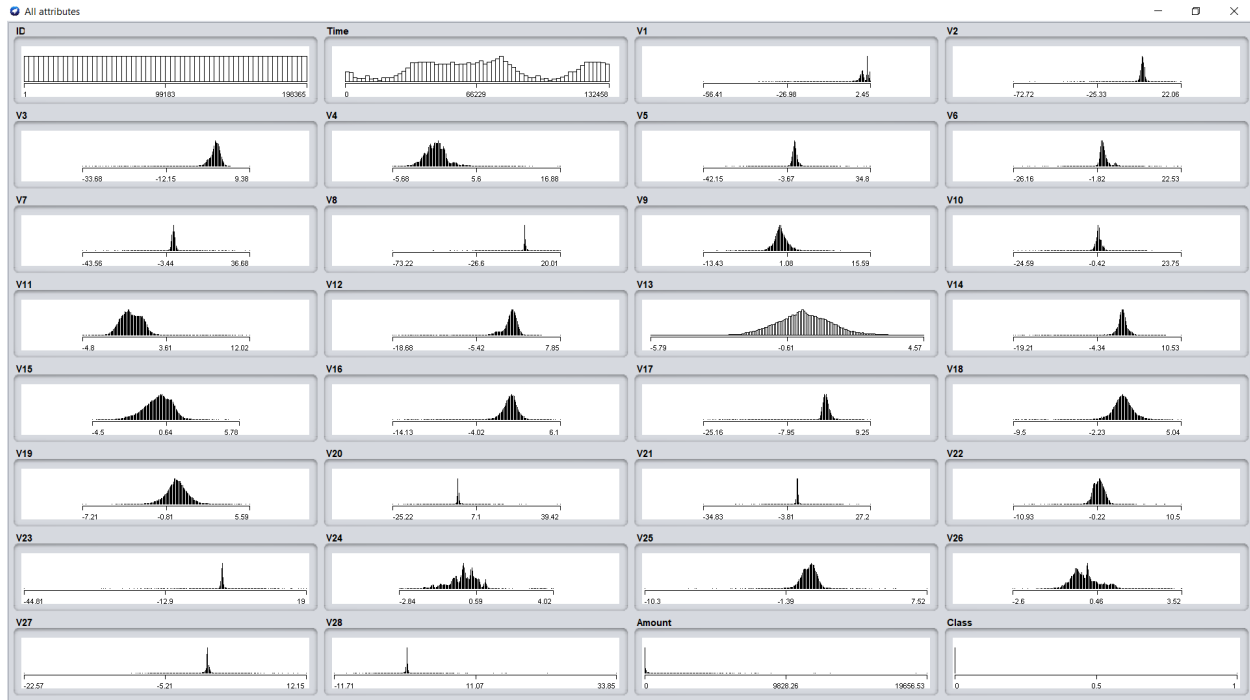
- Project Explorer:** Shows the project structure with folders like `src/main/java`, `src/main/resources`, `src/test/java`, `src/test/resources`, `JRE System Library [J2SE-1.5]`, `Maven Dependencies`, `src`, `main`, `java`, `org`, `ml`, `ccard.java`, `regression.java`, `resources`, `test`, `target`, `testoutput`, and `pom.xml`.
- regression.java:** Contains code for loading the dataset, building a linear regression model, and evaluating it.
- ccard.java:** Contains code for reading the dataset from a CSV file and printing its structure.
- Console:** Displays the output of the program, showing the first 6 values of the dataset.
- Table:** A table showing the structure of the dataset with columns for Index, Column Name, and Column Type.

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

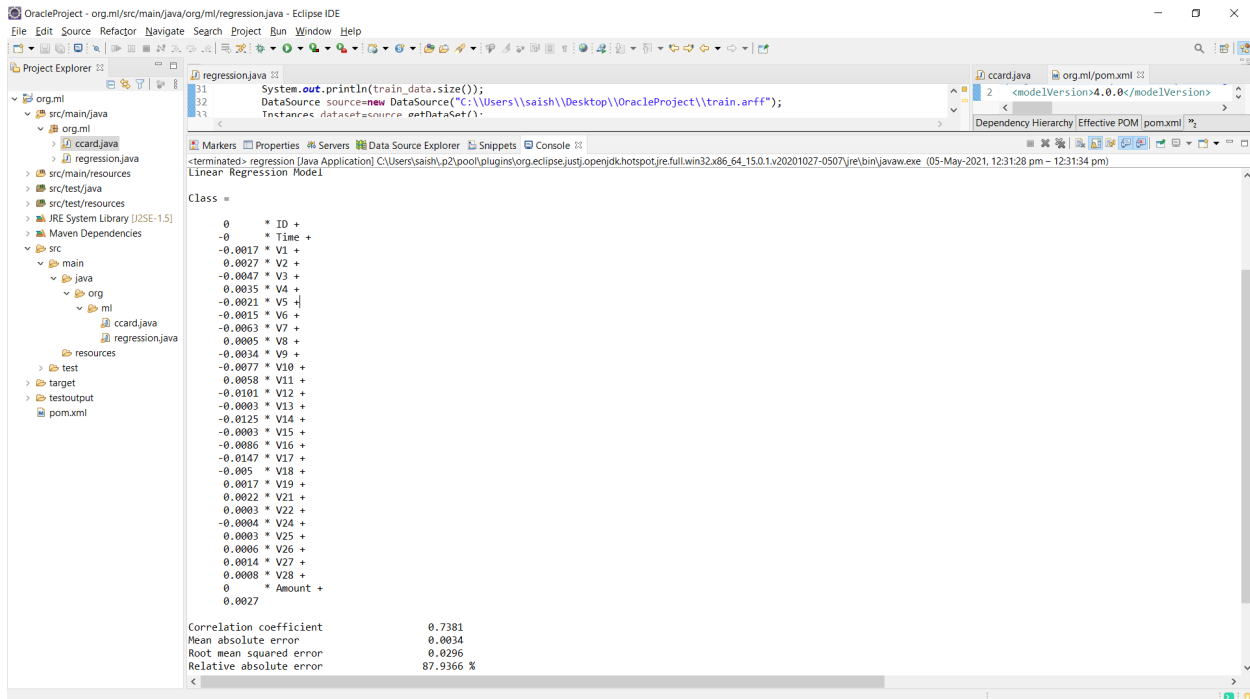
First 6 of DataSet values :

Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
...

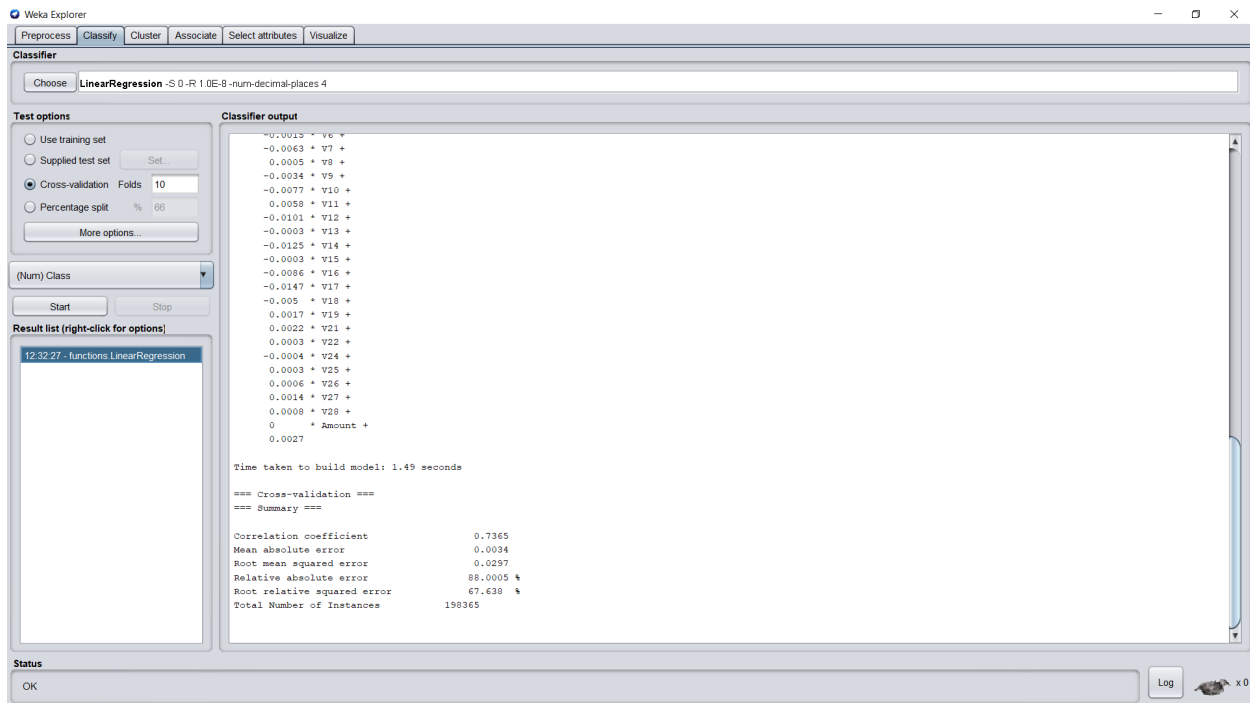
Visualization of Dataset(All attributes)



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->