Project Title : Creditcard Fraud Prediction

Submitted By:

ANJUM FIRDOS (CVR COLLEGE OF ENGINEERING)

1.INTRODUCTION

1.1 Overview

Detecting fraud transactions is of great importance for any credit card company. We are tasked by a well-known company to detect potential frauds so that customers are not charged for items that they did not purchase.

1.2 Purpose

The key objective of any **credit card fraud detection** system is to identify suspicious events and report them to an analyst while letting normal transactions be automatically processed. For years, financial institutions have been entrusting this task to rule-based systems that employ rule sets written by experts

2.LITERATURE SURVEY

2.1 Existing problem

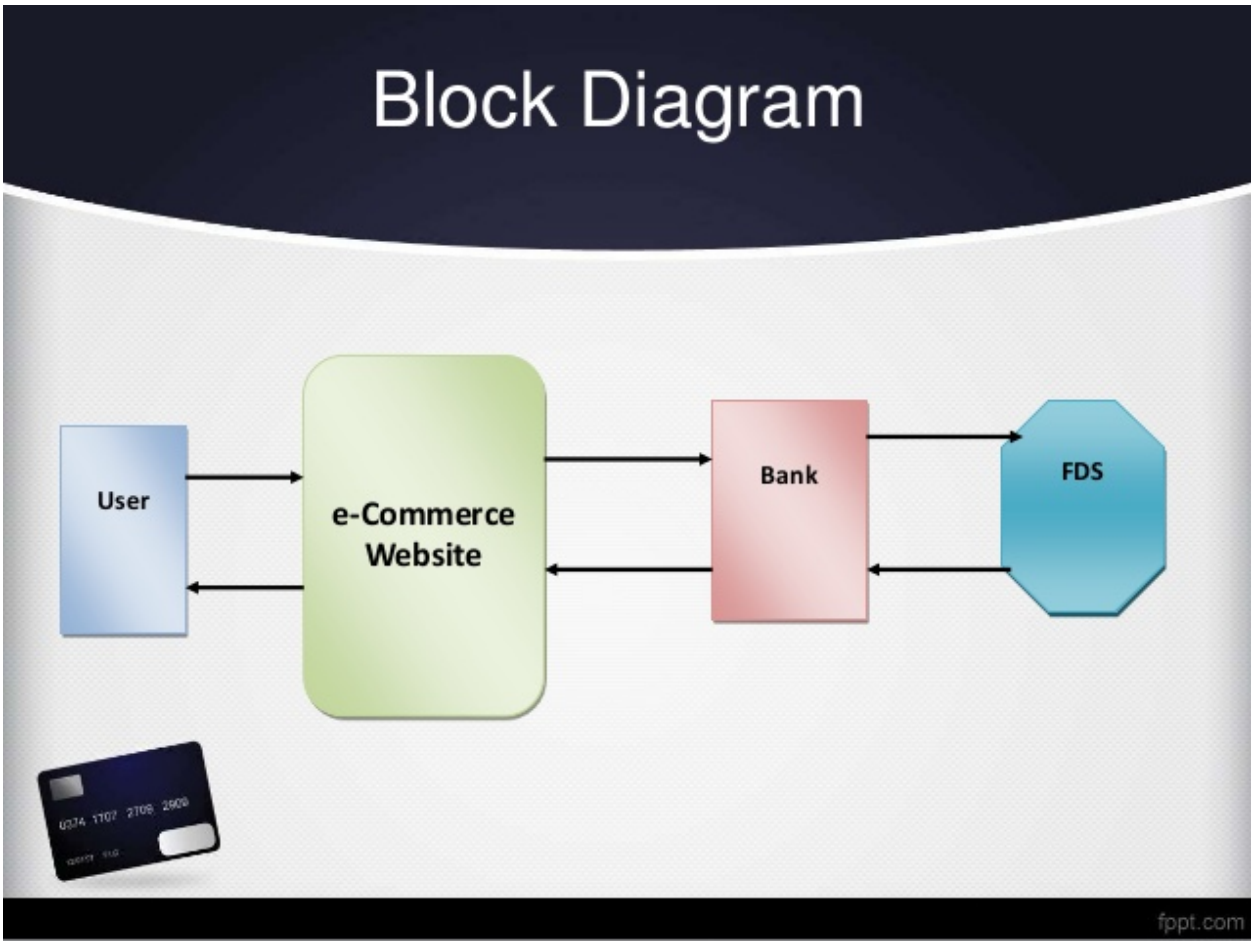
fraud detection solutions use two ML **techniques** — supervised or unsupervised learning. Supervised learning means that **a** model learns from previous examples and is trained on labeled data.

2.2 Proposed Solution

I suggest you to use Weka tool for resample,removeDuplicate for data preprocessing and data visualization and apply Linear Regression Algorithm.

3.THEORITICAL ANALYSIS

3.1 Block Daigram

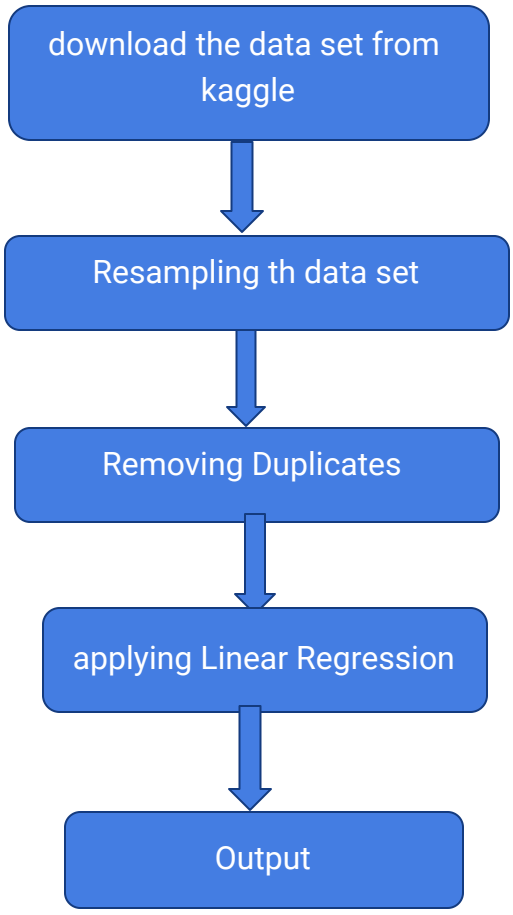


3.2 Hardware/ software designing

1.Eclipse

2.WEKA tool

4.FLOW CHART


download the data set from kaggle
Resampling th data set
Removing Duplicates
applying Linear Regression
Output

5.Result:

output in weka

=== Run information ===

Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation: creditcardfinalmod
Instances: 179503
Attributes: 31

Time
V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V11
V12
V13
V14
V15
V16
V17
V18
V19
V20
V21
V22
V23
V24
V25
V26

V27
V28
Amount
Class
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Linear Regression Model

Class =

-0 * Time +
-0.002 * V1 +
0.0029 * V2 +
-0.0052 * V3 +
0.0038 * V4 +
-0.0024 * V5 +
-0.0017 * V6 +
-0.007 * V7 +
0.001 * V8 +
-0.0036 * V9 +
-0.0081 * V10 +
0.0063 * V11 +
-0.0109 * V12 +
-0.0001 * V13 +
-0.0131 * V14 +
-0.0002 * V15 +
-0.0095 * V16 +
-0.0163 * V17 +
-0.0057 * V18 +
0.002 * V19 +
0.0004 * V20 +
0.0018 * V21 +
0.0003 * V22 +
-0.0004 * V24 +

0.0002 * V25 +
0.0003 * V26 +
0.0016 * V27 +
0.0011 * V28 +
0 * Amount +
0.0015

Time taken to build model: 1.4 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.7216
Mean absolute error	0.0031
Root mean squared error	0.0276
Relative absolute error	97.2163 %
Root relative squared error	69.2268 %
Total Number of Instances	179503

6.ADVANTAGES & DISADVANTAGES

Adv:

>**Linear regression** performs exceptionally well for linearly separable data

>Easier to implement, interpret and efficient to train

>**Weka** has a lot of **machine learning** algorithms. This is **great**, it is one of the large benefits of using **Weka** as a platform for **machine learning**. A down side is that it can be a little overwhelming to know which algorithms to use, and when

dis adv:

>**Linear Regression** Only Looks at the Mean of the Dependent Variable.

>**Linear regression** looks at a relationship between the mean of the dependent variable and the independent variables.

>it can only handle small datasets. Whenever a set is bigger than a few megabytes an OutOfMemory error occurs. The object of this thesis is to alter **Weka** in such a way that it can handle ”all” datasets, up until a few gigabytes.

7. APPLICATIONS

-->Applications for Marchants and Retailer

-->Applications for issuing Banks

8.CONCLUSION

It has detected those transaction as fraud where user belongs to low category and high category is made or vice versa.This mechanism require atleast 10 transactions to determine accurately the transaction is fraud or not

9.FUTURE SCOPE

The most popular area of current **fraud detection** research has been in **credit card**, but we see online bots and Ad click **fraud** as growing concerns for the **future**. With rapid reduction in the cost of computing power, publishers can exploit vulnerabilities by creating bots to click on Ads to generate more revenue.

10.BIBILOGRAPHY

**https://lahirumadushankablog.wordpress.com/2018/02/11/data-preparation-in-weka/#:~:text=%20Data%20Preparation%20in%20Weka%20%201%20Data,unsupervised%20-%3E%20attribute%20-%3E%20PrincipleComponents%0AOriginal%20iris...%20More%20

**https://machinelearningmastery.com/use-regression-machine-learning-algorithms-weka

**https://www.oreilly.com/library/view/data-science-with/9781491934104/ch04.html

Source code:

```
package fp.ml;


import java.io.IOException;

//import javax.sql.DataSource;
import weka.classifiers.Evaluation;
import weka.classifiers.functions.LinearRegression;
import weka.core.Instances;
import weka.core.converters.ConverterUtils.DataSource;
```

```
public class Linear {
		public static void main(String[] args) throws Exception {

			DataSource source =new
DataSource("C:\\Users\\anjumfirdous\\eclipse-workspace\\fp.ml\\src\\main\\java\\fp\\ml\\c
reditcardfinalmod.csv");
			Instances dataset=source.getDataSet();
			dataset.setClassIndex(dataset.numAttributes()-1);
			//linear Regression
			LinearRegression lr=new LinearRegression();
			lr.buildClassifier(dataset);

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


		}

	}
```