Creditcard Fraud Prediction

Overview

Detecting fraud transactions is of great importance for any credit card company.Credit card fraud prediction is to predict fraudulent transactions that takes place it is very essential for the every company to overcome fraud transactions.credit card fraud detection feature related to the financial section, using user behavior and location scans to identify anomalous patterns. These patterns include user attributes such as the user spending pattern, as well as the general user's geographic location to verify their identity. If an unusual pattern is detected, you need to check the system.

In this project I have used Linear Regression tools to track the Fraud transations.

LITERATURE SURVEY

Existing problem

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

Proposed solution

In this project I have used Linear Regression tools to track the Fraud transations. We can use sigmoid function in linear regression to clasify the output that is dependent attribute which uses probability.

Hardware/Software designing

Software Requirements:

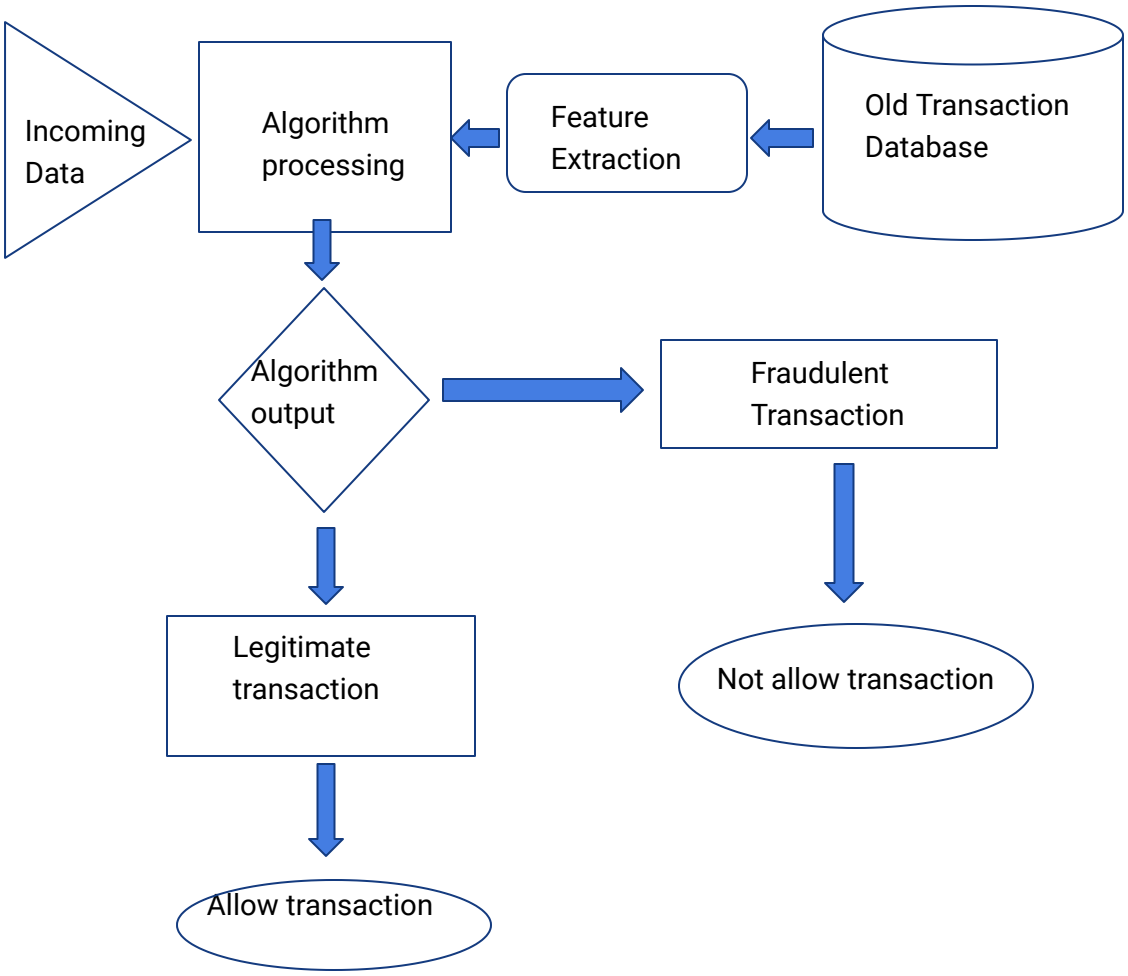
- Weka
- Eclipse IDE

Hardware Components:

laptop with basics configuration like i5,8thgen processor,4gb ram,10gb rom

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 fradulent transcations.

RESULT

Datase has: 284807 rows X 31 cols

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

Summary	Time	V1	V2	V3	V4
Count	284807	284807	284807	284807	284807
sum	27003650904	0.000000500890549703925 1	-0.000000235031160145116 05	-0.000000274480973172508 4	0.000000236852179646973 58
Mean	94813.8595750808	0.000000000001760144880358588 5	-0.000000000000824570546382875 2	-0.000000000000968940620046812 7	0.000000000000832402114333850 1
Min	0	-56.40750963	-72.71572756	-48.32558936	-5.683171198
Max	172792	2.454929991	22.05772899	9.382558433	16.87534403
Range	172792	58.862439621	94.77345655	57.708147793	22.558515228
Variance	2255124006.202241	3.8364892520273246	2.726820024675042	2.2990292407249213	2.004683821508361
Std. Dev	47488.14595456682	1.9586958038519724	1.651308579483266	1.5162550051772035	1.415868574942025

V5	V6	V7	V8	V9
284807	284807	284807	284807	284807
0.000000004534165002137436	0.000000120966317074255 87	-0.000000086871344562311 49	0.000000024757936645247725	-0.000000335989776945044 84
0.000000000000016437656061258244	0.000000000000042558831579261 77	-0.000000000000030502764857942 13	0.000000000000088067117727652 22	-0.000000000001179619208125268 5
-113.7433067	-26.16050594	-43.55724157	-73.21671846	-13.43406632
34.80166588	73.30162555	120.5894939	20.00720837	15.59499461
148.54497258	99.46213148999999	164.14673547	93.22392683	29.02906093
1.9050810467899906	1.7749462566054632	1.5304005706243582	1.4264788561215538	1.2069924674717156
1.3802467340261997	1.3322710897581855	1.2370935981664275	1.1943529026722184	1.0986320892235562

creditcard.csv

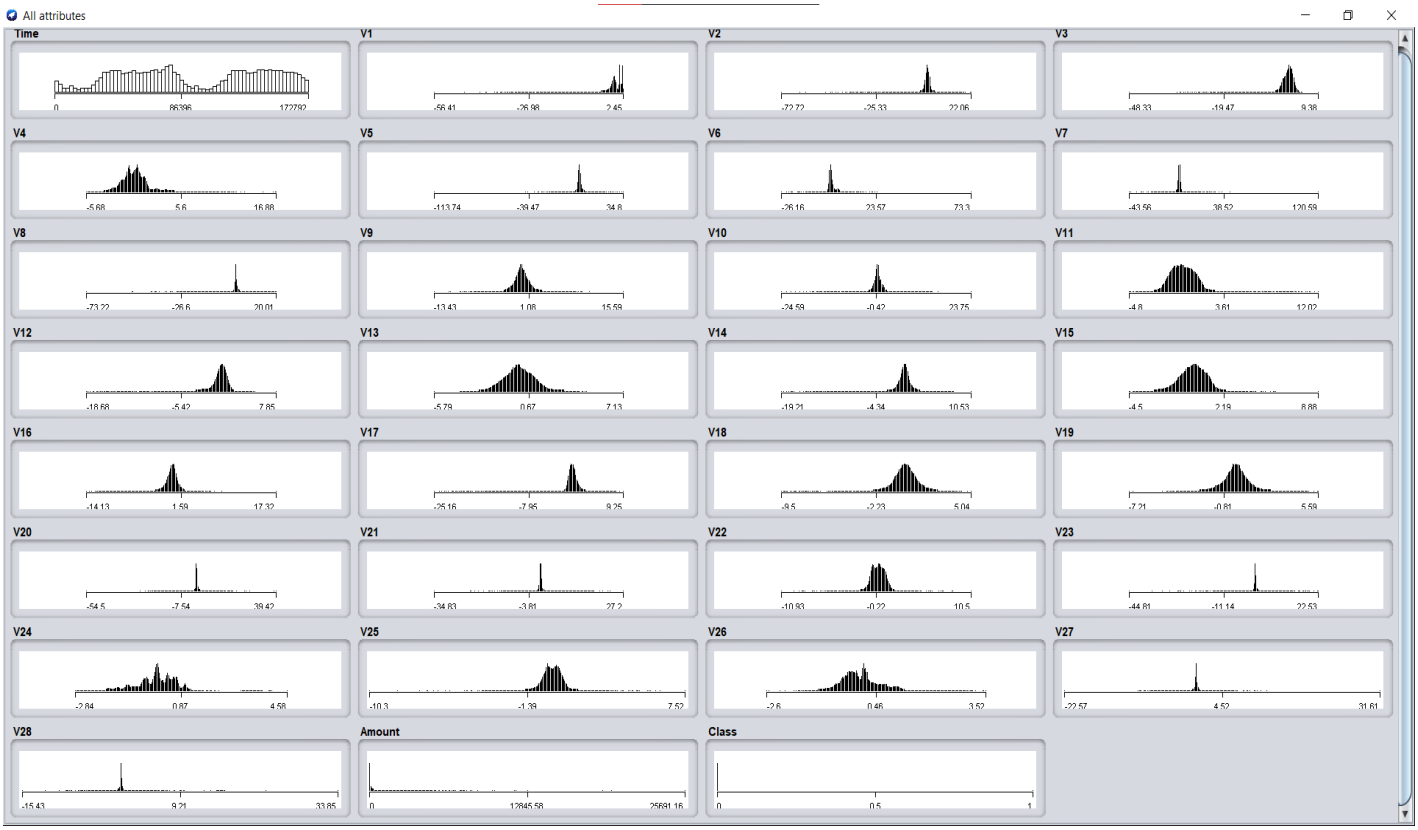
V10	V11	V12	V13	V14	V15
284807	284807	284807	284807	284807	284807
0.000000202068306309755 03	0.000000534019260811646 7	0.000000300046144860455 52	0.000000203283124639019 87	-0.000000042475238853922 9	-0.000000148838589196509 39
0.000000000007092488433817418	0.000000000001875260375941 57	0.000000000001052304796869610 7	0.000000000000713500429757 52	-0.000000000000149151984985711 38	-0.000000000000052238398172793 72
-24.58826244	-4.797473465	-18.68371463	-5.791881206	-19.21432549	-4.498944677
23.74513612	12.01891318	7.848392076	7.126882959	10.52676605	8.877741598
48.333398560000006	16.816386645	26.532106706	12.918764164999999	29.74109154	13.376686275
1.1855938116260996	1.0418550849402706	0.998403416833335	0.9905707931502421	0.918905545924414	0.837803401110015
1.088849765406642	1.0207130277116436	0.999201389527324	0.9952742301246638	0.9585956112586861	0.9153160116102061

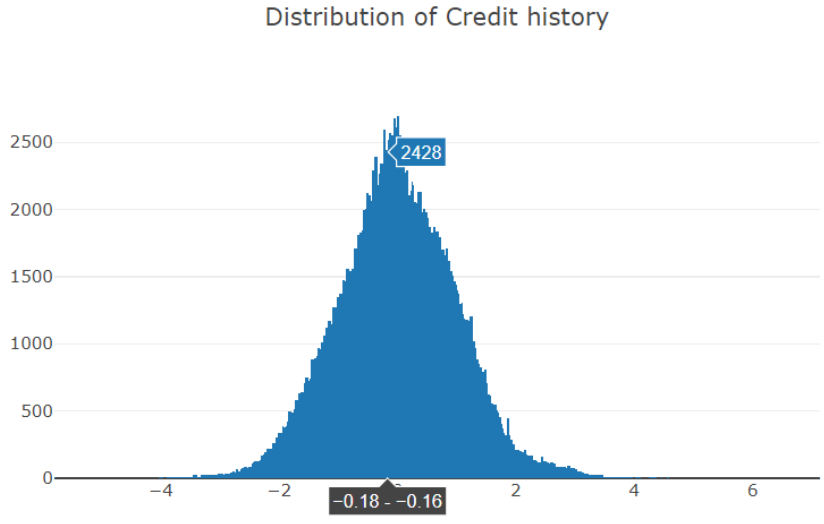
many more

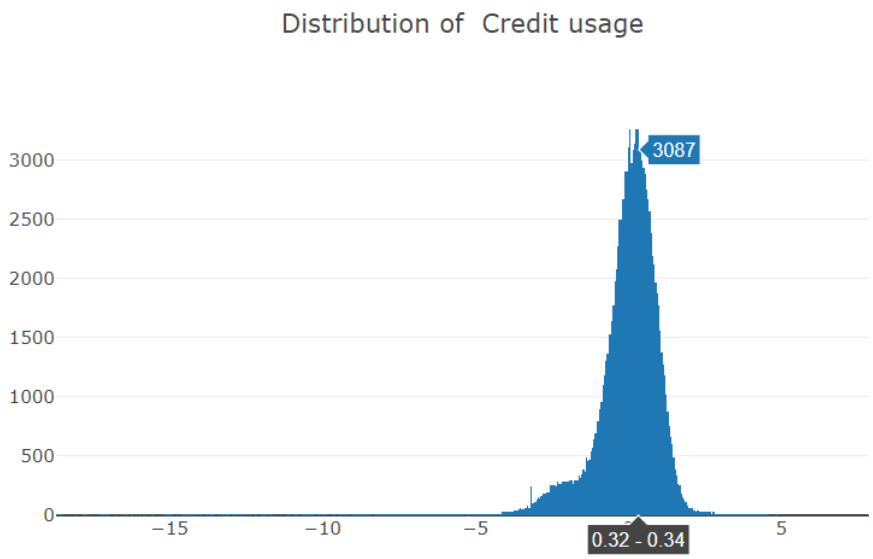
There are no missing values:

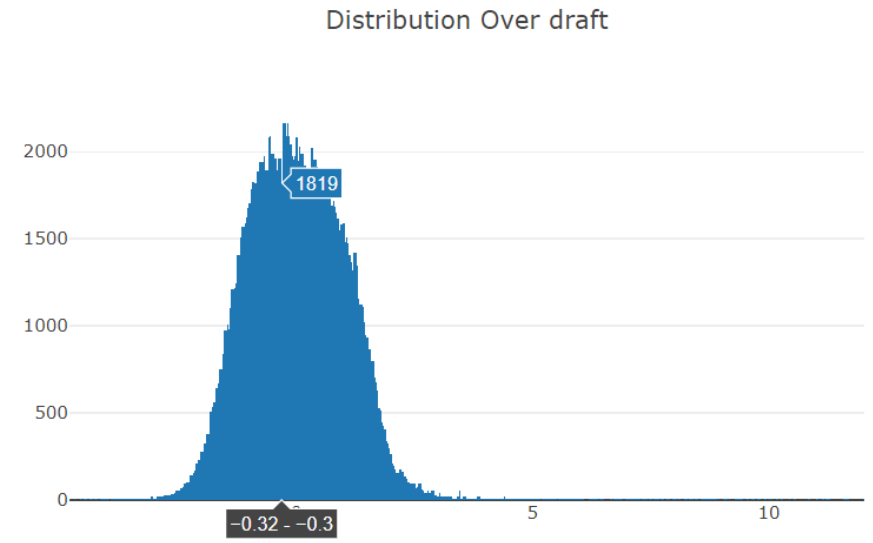
Missing Values [Time]	Missing Values [V21]	Missing Values [V20]	Missing Values [V23]	Missing Values [V22]	Missing Values [V25]	Missing Values [V24]	Missing Values [V27]
0	0	0	0	0	0	0	0

Visualization of dataset using weka:

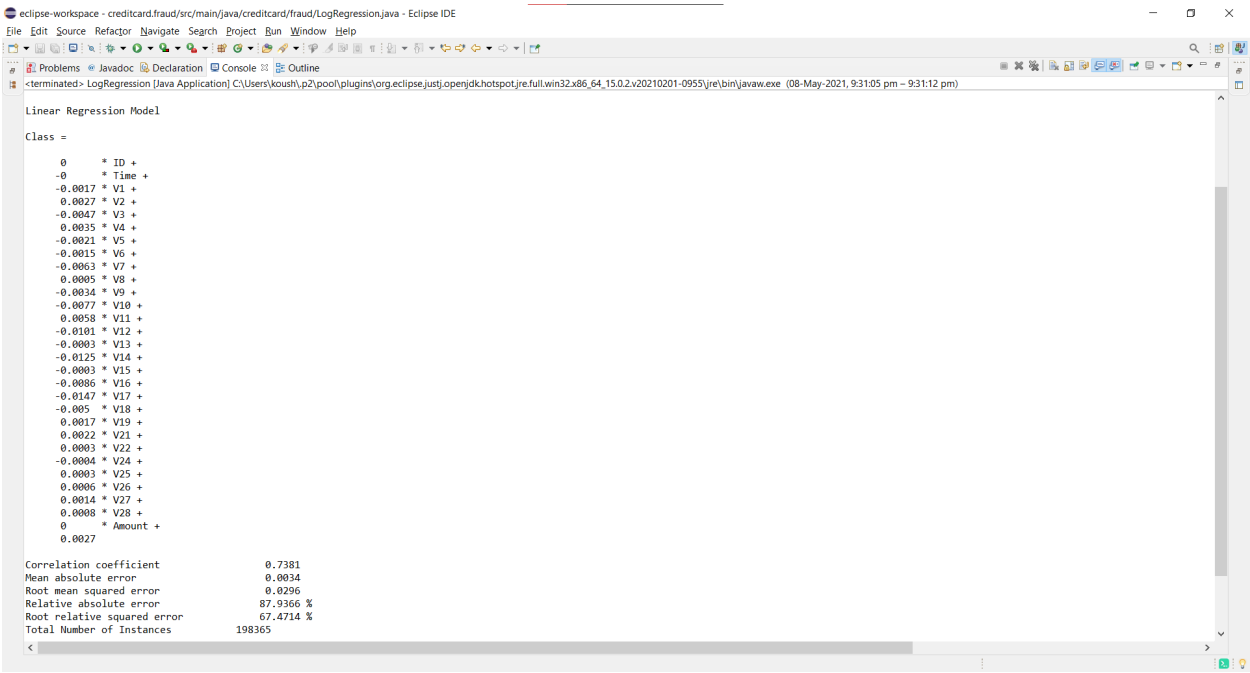





Distribution of Credit usage
3000
2500
2000
1500
1000
500
0
−15
−10
−5
5
3087
0.32 - 0.34


Distribution Over draft
2000
1500
1000
500
0
5
10
1819
−0.32 - −0.3

Linear Regression



ADVANTAGES & DISADVANTAGES

Advantages:

- Easy to interpret the data
- less complexity than other algorithms
- we can get the relationship between the independent and dependent variable to interpret the data

Disadvantages:

- Huge size of data
- imbalanced data set
- Adaptive techniqes
- Availability of data

Conclusion:

Credit card companies need to be aware of fraudulent credit card transactions, and customers may not face transaction revaluation issues. Therefore this project will help you determine if it is a transaction fraud.

References

For dataset:

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

For tablesaw methods :

https://jtablesaw.github.io/tablesaw/userguide/importing_data

For regressions:

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

https://github.com/jtablesaw/tablesaw/blob/master/docs/userguide/ml/Moneyball%20Linear%20regression.md

others:

www.javadoc.io