International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 7 ǁ 2021 ǁ PP. 44-49

An Approach to CreditCardFraudDetection

DivyaKV[1],SwastiChoudhary[2],ThakurKiranSingh[2],NarendraKumarReddy[2],VishalSBalan[2]

[1]*SeniorAssistantProfessor,Department of InformationScienceand Engineering,New HorizonCollegeofEngineering,Bengaluru-560103,Karnataka,India.*

[2]*Students,DepartmentofInformationScienceandEngineering,NewHorizonCollegeofEngineering,Bengaluru-560103, Karnataka,India.*

ABSTRACT
Due to the rapid advancement in the electronic commerce technology in today's world the use of credit cards hasgradually increased. Since credit card is the most favored mode of payment, the number of fraud cases which areassociated with it is also rising every day. Fraud detection means that identifying the fraud as quickly as possibleonce it has been performed. Fraud detection methods are continually developed to defend the criminals in adaptingto their strategies used. The transaction made are classified as normal, abnormal or suspicious depending on theinitial belief. Once a transaction is found to be suspicious, belief is further strengthened or weakened according tothe similarity with fraudulent or genuine transaction history using various algorithms. It is necessary that the creditcard companies are able to identify fraudulent credit card transactions so that the customers are not charged for theitems that they didn't purchase. Such problems can be tackled with various technologies of Machine Learning anditsrelevantareas.
The Credit Fraud Detection project plans to demonstrate the modelling of a dataset making use of machine learningwith Credit Card Fraud Detection. The Credit Card Fraud Detection Problem consists of modelling past credit cardtransactions with the data of the ones that turned out to be fraud among that. This model is then used so as torecognize whether a new transaction is fraudulent or not in all. The objective here is to detect 100% ofthefraudulenttransactionsinthedatasetwhileminimizing theincorrectfraud classifications.CreditCardFraudDetection is a classic sample of classification. In this process, it is focused on analyzing and pre processing data setsaswellasthedeploymentofmultipleanomalydetectionalgorithmssuchasRandomForestClassifier,NaiveBayes,etc.onthe givenCreditCard Transactiondata.
Keywords-*MachineLearning,DataMining,CreditCardFraud,NormalandFraudTransaction.*

Date of Submission: 06-07-2021 Date of acceptance: 20-07-2021

I. INTRODUCTION

The popularity of on-line shopping has grown day by day in today's world. According to a 2005 ACNielsen survey,1 in 10 of the world's population buys online. Nowadays, credit card is the most popular method of payment. As thenumber of credit card users increases worldwide, identity theft is on the rise and fraud is on the rise. Credit card-based purchases can be divided into two types: 1) physical card purchases and 2) visual card purchases. At thepurchaseofaphysicalcard,thecardholderpersonallypresentsthecardtomakeapayment.Whilebuyingaphysical card, the attacker needs to steal the credit card and build a signature to buy it. For visual card purchases,only card details are required such as card number, expiration date, secure code, etc. Such purchases are usuallymade online or over the phone. To commit fraud in these types of purchases, a person simply needs to know thedetails of his or her card. The online shopping mode is made mostly by credit card. Credit card fraud has beenincreasing day by day. The amount of financial losses due to credit card fraud increases as the use of credit cardsbecomes more common. Security means using your credit card safely and avoiding fraudulent appearances. Thepurpose of security is to prevent the use of fraudulent credit cards. In cases of fraud there are issues such as lostcards,lostcards,stolencards,applicationfraud,fraudulentmail,postalfraudandunpaidfraud(NRI).Toreducethis risk, credit card security is required. 'Fraud' in credit card transactions is the unauthorized and unwanted use ofan account by a person other than the owner of that account. Necessary preventive measures can be taken to preventthisabuseandcanbestudiedinsuchfraudulentbehaviourtoreduceandpreventsimilarincidentsinthefuture. Someoneelse'screditcardforpersonalreasonswhiletheownerandtheissuingauthoritiesdonotknowthatthecard isinuse.

II. LITERATURESURVEY

1. ASurveyonDifferentDataMining& MachineLearningMethodsforCreditCardFraudDetection

In the above paper, different methods of fraud detection in credit cards were investigated. Firstly, significance of thesubject was stated and existing deficiencies in traditional systems were mentioned. Fake transactions have alsovarying degrees of risks and ways should be found for finding transactions with highest risk in quicker and moreaccurate manner. For identification of these transactions, common data mining methods alone do not suffice.Innovative algorithms should be used for the effective results. Due to rapid growth in cashless transaction, thechancesofnumberoffraudulenttransactionscanalsoincreasing.AFraudtransactioncanbeidentifiedbyanalyzing various behaviors of credit card customers from the previous transaction history datasets. If any deviationis noticed in spending behavior from available patterns, it is possibly of fraudulent transaction. Data mining andmachinelearningtechniquesare widelyused increditcard frauddetection.

2. CreditCardFraudDetectionUsingRandomForestAlgorithm

In this paper it mainly focuses on credit card fraud detection in real world. Here the credit card fraud detection isbased on fraudulent transactions. Generally credit card fraud activities can happen in both online and offline. But intoday's world online fraud transaction activities are increasing day by day. So in order to find the online fraudtransactionsvariousmethodshave beenused.

Inthepapermentioned,theauthorsmadeuseofRandomForestAlgorithmforfindingthefraudulenttransactions andtheaccuracyofthosetransactions.Thisalgorithmisbasedonsupervisedlearningalgorithmwhereitusesdecisiontree sforclassificationofthedataset.Afterclassificationofdatasetaconfusionmatrixwasobtained.Theperformance of Random Forest Algorithm is evaluated based on the confusion matrix. The results obtained fromprocessingthedatasetgivesaccuracyofabout90%.Thus,usingthisRandomforestalgorithmanddecisiontreesalgo rithmtheyhaveextractedtheaccuratepercentageofdetectionoffraudfromthegivendatasetbystudyingitsbehavior.Aco nfusionmatrixisbasicallyasummaryofpredictionresultsoratablewhichisusedtodescribetheperformanceoftheclassifi eronasetoftestdatawheretruevaluesareknown.Itprovidesvisualizationofanalgorithm'sperformanceandallowseasyi dentificationofclasses.Thus,resultinginthecomputingofmostperformancemeasuresbygivinginsightsnotonlytheerro rsbeingmadebytheclassificationmodelbutalsotellsthetype of errorsbeingmade.

3. CreditCardFraudDetectionUsingPredictiveModelling

In this paper author proposed that fraud detection is a critical problem affecting large financial companies that haveincreased due to the growth in credit card transactions. This paper presents detection of frauds in credit cardtransactions, using data mining techniques of Predictive modeling, logistic Regression, and Decision Tree. The datasetishighly unbalanced, thepositive class(frauds)Accountfor 0.172%of alltransactions.

Decision trees are used to choose between several courses of action. It provides effective structure to investigate thepossible outcomes. Decision trees use tree structure to build classification or regression model. A decision tree is aflowchart like tree structure, where non leaf node denotes a test on attribute. In theresults, the decision tree willhave a decision node and leaf nodes. Predictive modeling is used to analyze the data and predict the outcome.Predictive modeling used to predict the unknown event which may occur in the future. In this process, we to create,test and validate the model. There are different methods in predictive modeling. They are learning, artificialintelligenceandstatistics.Oncecreatedamodel,itcanusemanytimes,todeterminetheprobabilityofoutcomes. So predict model is reusable. Historical data is used to train an algorithm. The predictive modeling process is aniterativeprocessandoften involvestraining themodel,using multiple modelsonthesamedataset.

4. CreditCardFraudDetectionandPreventionusingMachineLearning

Inthissurvey,areviewofacontextualinvestigationincludingtheidentificationofCreditCardmisrepresentatio nwhereinformation standardizationisapplied priortocluster analysisandwithresultsobtainedfrom theuseof Cluster Analysis and Artificial Neural Network on the discovery of extortion has indicated that neuronal dataSources may be limited by bundling properties. What's more, encouraging outcomes can be gotten by utilizingstandardized information and information ought to be MLP prepared. This examination depended on solo learning.Noteworthiness of this paper was discovering an estimate and reducing the measure of costs. The result was 23%andthe calculationtheyfoundwasthe minimumchanceofBayes.Inthissystem,acollectivereplacementcomparison measure is proposed that represents profits and losses due to fraud detection. Using the existing costmeasure, a costsensitive method that dependson the Bayesminimum riskisused.

5. CreditCardFraudDetectionusingMachineLearningAlgorithms

In the above mentioned survey, the authors focused on - Card transactions are always unfamiliar when compared toprevious transactions madethecustomer. This unfamiliarityis avery difficult probleminreal-worldwhen arecalled concept drift problems. Concept drift can be said as a variable which changes over time and in unforeseenways.Thesevariablescauseahighimbalanceindata.Themainaimoftheresearchwastoovercometheproble mof Concept drift to implement on real-world scenario. They aimed to overcome three main challenges with cardfrauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabelled samples, and to increasetheabilitytoprocessa large numberoftransactions.

III. SYSTEMDESIGN

System design is the process of the defining thearchitecture, components, modules, interfaces, and data for asystem to satisfy specified requirements. Systems design could be seen as the application of systems theory toproductdevelopment.

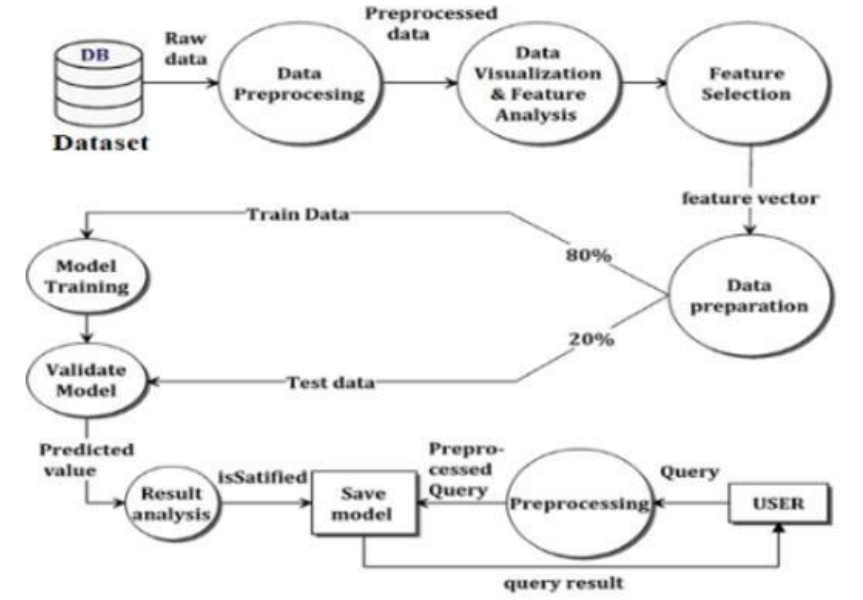

Figure1.System Architecture

The architecture shows the flow of the process of the system. It consists of various levels; from the dataset beingprovidedwithitspre-processing,extraction,trainingandtestingmodeltotheevaluationofresultsofthetransaction. It also showsthe percentage of accuracy of the training model by the specific algorithm used in theimplementation.

IV. IMPLEMENTATION

In computer science, an implementation is a realization of a technical specification or algorithm as a program,softwarecomponent,orothercomputersystemthroughcomputerprogramminganddeployment.Manyimple mentations may exist for a given specification. In this system, python is used to build the functionality of thesystem and graphical user interface (GUI). It makes use of different libraries of Python, like tkinter, numpy, panda,matplotlibandseaborntobuildthe system.

ThedifferentModulesoftheCredit CardFraudDetectionProject:

Data Collection - Data is the most important part when we work on prediction systems. It plays a specific role inwhole project i.e., the system depends on that data. So, collection of data and then select is the first and the criticalstep which should be performed properly. The data which is used in the process is a set of transactions collectedfrom the credit card transactions records. This step is basically concerned with selecting the subset of all theavailabledata that isplayingitsrole indetection.

Data Preprocessing – The next step after the selected data is organize it by the process of by formatting, cleaningand finally,samplingfromit.The datapre-processingstepsincludes:

- Formatting: The data have been selected may not be in a format which is suitable to work with further.Like, the data may be in a relational database and it should be in a flat file or it may be in a proprietary fileformatbutit shouldbe inarelational database or anytextfile.
- Cleaning: Cleaning data means here that it is the removal or fixing of missing data. There may be the datainstances which are incomplete and do not carry the data that is required or needed to address the givenproblem.These instancesmayneedtobe removedfromthedataset.
- Sampling: The final step in data pre processing is Sampling which deals there may be far more selecteddata available than required to be used. The more the data, it can result in much longer running times foralgorithmsused and larger computationalaswellasmemory requirementswithinthesystem.
- Feature Extraction and Algorithm – Nowadays, it is becoming quite common to work with datasets ofhundreds or even thousands of features. If the number of features becomes similar or even bigger than thenumber of observations stored in a dataset, then this can further lead to a machine learning model sufferingfromoverfitting.So,toavoidsuchascenario,itisnecessarytoapplyeitherregularizationordimensionality reductiontechniqueswhichisfeatureextraction.

Data Visualisation - Data visualization is the step wherein all the data will be transformed into some form of plotsand analyzed further from that instead of any such tables. As a human being, we are more likely to take a lot ofinformation from diagrammatic representation thanits substitutes. If we want to convert the data from aboringtable into an interesting pictorial form like a scatter plotting, then it can be done by making use of very goodpackageswhich areavailablefrom availablepopular programming languageswhich areused commonly.

Feature Extraction and Algorithm – Nowadays, it is becoming quite common to work with datasets of hundredsor even thousands of features. If the number of features becomes similar or even bigger than the number ofobservations stored in a dataset, then this can further lead to a machine learning model suffering from overfitting.So, to avoid such a scenario, it is necessary to apply either regularization or dimensionality reduction techniqueswhich is feature extraction. In machine learning, the dimensionality of the dataset is same as thenumber ofvariableswhichareusedtorepresentit.Featureextractiontechniquescanalsoleadtoothertypesofadvantagessuchas accuracy improvements, overfitting riskreduction, speed upintraining, improved data visualization,increase in efficiency of the model. Feature Extraction is used to reduce the number of features in a dataset bycreating the new features from the existing ones. It discards the original features. The algorithm used are RandomForest, Decision Tree and Naïve Bayes. These algorithm are stable and works well. So, they contribute in reducingtheoverall performanceof thesystem.

V. RESULTS

Out[3]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V2
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.06692
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.33984
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.68928
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.17557
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.14126
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656	-5.364473	-2.606837	-4.918215	7.305334	1.914428	...	0.213454	0.111864	1.014480	-0.50934
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589	0.868229	1.058415	0.024330	0.294869	0.584800	...	0.214205	0.924384	0.012463	-1.01622
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828	2.630515	3.031260	-0.296827	0.708417	0.432454	...	0.232045	0.578229	-0.037501	0.64013
284805	172788.0	-0.240440	0.530483	0.702510	0.689799	-0.377961	0.623708	-0.686180	0.679145	0.392087	...	0.265245	0.800049	-0.163298	0.12320
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271	-0.012546	-0.649617	1.577006	-0.414650	0.486180	...	0.261057	0.643078	0.376777	0.00879

284807 rows × 31 columns

Figure1.Displaysthedatasetstoredinthesystem

Figure2.HeatmapofCorrelation

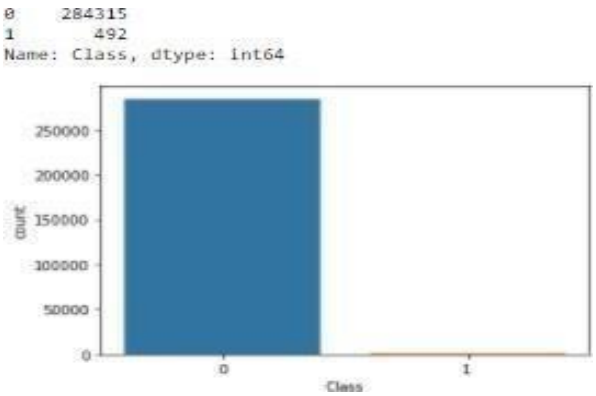

Figure3.Displaysthenumberoffraudandnormaltransactiondata

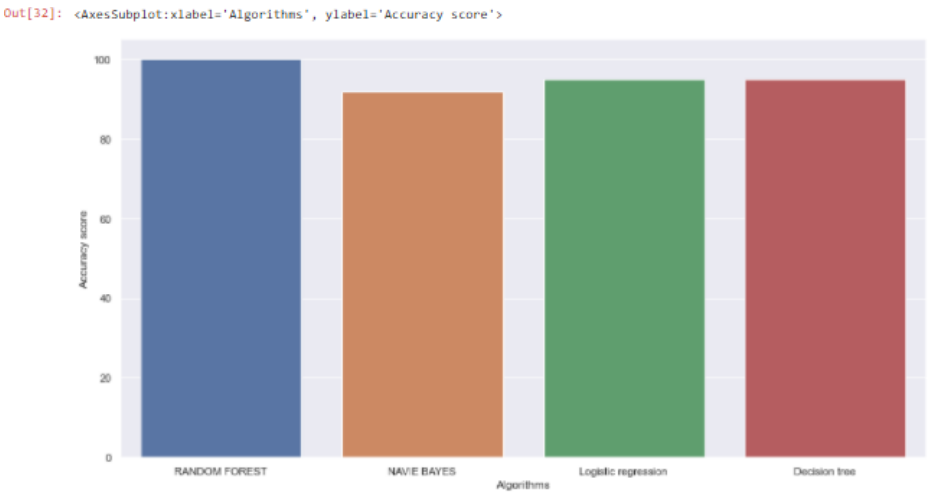

Figure4.Accuracyscoreobtainedbydifferentalgorithms

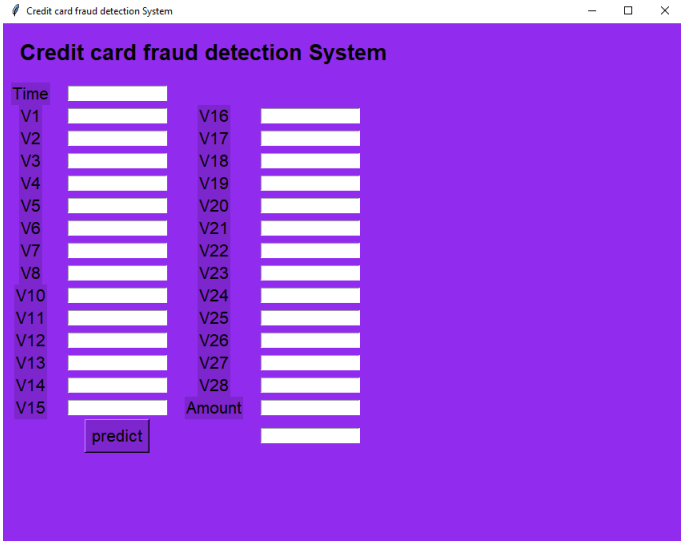

Figure5. GUIofCreditCardFraudDetection

VI. CONCLUSION

Credit Card Fraud Detection is undoubtedly an act of criminal dishonesty. This system has listed out the mostcommon methods of fraud along with their detection methods; also reviewed recent findings in this field.

Thissystemhasalsoexplainedindetailthathowmachinelearningtechniquescanbeappliedtogetbetterresultsinfraud detection consisting of algorithm, code, explanation its implementation and theresults. When the entiredataset was fed into the algorithm then the precision rose to 33%. This high percentage is as per the huge imbalancebetween the number of provided valid and genuine transactions. Since it is based on machine learning algorithms sotheprogramwillonly increaseitsefficiency over time whenmoredata isputintoit.

REFERENCES

[1]. A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection -By VipulPatil, Dr. Umesh Kumar Lilhore Published by-International Journal of Scientific Research in Computer Science,Engineering and InformationTechnology2018IJSRCSEIT|Volume3|Issue5|ISSN:2456-3307.

[2]. Credit Card Fraud Detection Using Random Forest Algorithm– By M. Suresh Kumar, V. Soundarya, S.Kavitha,E.S.Keerthika,E.AswiniPublishedby-20193rdInternationalConferenceonComputingandCommunicationsTechnologies(ICCCT),05September 2019.

[3]. CreditCardFraudDetectionUsingPredictiveModelling–ByVarrePerantalu,BhargavKiranPublishedby-February2017|IJIRT |Volume 3Issue 9|ISSN: 2349-6002.

[4]. Credit Card Fraud Detection and Prevention using Machine Learning – By S. Abinayaa, H. Sangeetha, R. A.Karthikeyan,K.SaranSriram,D.PiyushPublishedBy-InternationalJournalofEngineeringandAdvancedTechnology (IJEAT) ISSN: 2249–8958, Volume-9Issue-4, April, 2020.

[5]. Credit Card Fraud Detection using Machine Learning Algorithm By- Vaishnavi Nath Dornadula, Geetha SPublished By–InternalConference On RecentTrendsIn Advanced Computing2019,ICRTAC2019.

[6]. Credit Card Fraud Detection using Machine Learning and Data Science By- S P Maniraj ,Aditya Saini, SwarnaDeep Sarkar Published By- International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181Vol.8Issue 09, September-2019.

[7]. Performance Evaluation of Class Balancing Techniques for Credit Card Fraud Detection By- Dilip SinghSisodia, Nerella Keerthana , Shivangi Bhandari Published by- IEEE International Conference on Power, Control,Signals&InstrumentationEngineering(ICPCSI-2017).

[8]. A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective By- SamanehSorournejad ,Zahra Zojaji,RezaEbrahimiAtani,Amir Hassan Monadjemi,PublishedBy- IEEE2016.

[9]. ASurveyOfSignature-BasedMethodsForFinancialFraudDetectionBy-E.MichaelandS.Pedro,PublishedBy- Computerandsecurity,vol. vol28, no. 6, pp.381–394.

[10]. Random Forest For Credit Card Fraud Detection By= S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N.Surname, Published By-IEEE15thInternationalConferenceon Networking, Sensing andControl(ICNSC),2018.

[11]. Credit Card Fraud Detection Using Machine Learning Algorithim By- Varun Kumar K S , Vijaya Kumar V ,Vijayshankar A , Pratibha K Published By- IJERT IJERTV9IS070649 Volume 09, Issue 07 (July 2020)05-08-2021ISSN2278-0181.

[12]. Machine Learning For Credit Card Fraud Detection System By- Lakshmi S V S S1 ,Selvani Deepthi Kavila2Published By-International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 24(2018)pp. 16819-16824.

[13]. Detecting and Preventing Fraud with Data Analytics By- B. Adrian Pulished By- Procedia Economics andFinance, vol.32,no.15,pp.1827–1836, 2015.

[14]. Credit Card Fraud Detection Using Random Forest By- Devi Meenakshi , Janani, Gayathri., Mrs. Indira. NPublished By-International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume:06 Issue: 03Mar 2019p-ISSN: 2395-0072.

[15]. Fraud Detection in Credit Cards using Logistic Regression By- Hala Z Alenzi1 , Nojood O Aljehane PublishedBy- (IJACSA) International Journalof AdvancedComputerScience andApplications,Vol. 11, No.12,2020.