International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 3, May-June 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Credit Card Fraud Detection Using XG Boost Algorithms

B.Birunda*, P.Tamilselvi**

*(Asst Professor CSE, Sembodai Rukmani Varatharajan Engineering College, and Sembodai
Email: birunda1212@gmail.com)

** (PG Scholar CSE, Sembodai Rukmani Varatharajan Engineering College, and Sembodai
Email: tamilselvi88.ps@gmail.com)

----------**********----------

Abstract:

Banking industry has the main activity of lending money to those who are in require ofmoney. In order to payback the principle borrowed from the depositor bank collects theinterest made by the principle borrowers. Credit risk analysis is becoming an important fieldinfinancialriskmanagement.Manycreditriskanalysisproceduresareusedfortheevaluation of credit risk of the customer dataset. The evaluation of the credit risk datasetsleads to the decision to issue the loan of the customer or reject the application of the customeris the hard task which involves the deep analysis of the customer credit dataset or the dataprovided by the customer. In this paper we are surveying different procedures for the creditrisk analysis which are used for the evaluation for the credit risk datasets. Credit card fraud isa serious problem in financial services. Machine learning algorithm based fraud detectionscheme is implemented for detect the fraud card. Hybrid methods which use AdaBoost andmajority voting methods are useful. To estimate the model efficiency, a publicly availablecredit card data set is used. Then, a real-world credit card data set from a financial institutionisevaluated.

Keywords **—Machine learning,AdaBoost,Credit risk analysis,XGBoost (Extreme Gradient Boosting).**

----------**********----------

I. INTRODUCTION

Creditcardfraud isabroadrangingwordfor theft and fraud committedusingorinvolving a payment cards, such as a credit card or debit card, as a fake source of funds in atransaction. The purpose may be to attain goods without paying, or to attain unofficial fundsfrom an account.Creditcard fraudisalsoasupplementto identity theft.Accordingtothe United States Federal Trade Commission, While the speed of identity theft had been holdingstable during the middle-2000s, it increased by 21 percent in 2008. However, credit card fraud,thatcrimewhichthe majoritypeoplejoinwithIDrobberydecreasedasape rcentageofevery oneIDtheftcomplaintsforthesixthyearinarow.

Invention of creditcards hasmadeonlinetransactions seamless,easier,comfortableand suitable. However, it has as well provided original fraud opportunities for criminals, and in turn,increased fraud rate. The global impact of credit card fraud is alarming, millions of US dollarscontainbeenlostbya lot ofcompaniesandindividuals.Furthermore,cybercri minalsareinnovating sophisticated techniques on a regularbasis,hence, there isan vitaltask todevelop improved and dynamic techniques capable of adapting to rapidly evolving fraudulentpatterns. Achieving this task is very challenging, mainly due to the dynamic nature of fraudand also due to lack of dataset for

researchers. This paper offerings a review of better-qualitycreditcardfrauddetectiontechniques.Exactly,thispaperfocusedonrecentMachineLearning based and Nature Inspired based credit card fraud detection techniques proposed inliterature. This paper offers a picture of recent trend in credit card fraud finding.Moreover,this review outlines some limitations and contributions of existing credit card fraud detectiontechniques,italsoprovidesnecessarybackgroundinformationforresearchers inthisdomain.

Machine learning(ML) is the technical study of algorithms and statisticalmodels that computer systems use to effectively perform a specific task without using explicitinstructions, relying on models and interpretation instead. It is seen as a subclass of artificialintelligence. Machine learning algorithms build a calculated model of sample data, known astraining data, in order to make predictions or decisions without being obviously programmedtoperform the task. Machinelearning algorithmsare used in the applicationslike emailfiltering, detection of network invaders, and computer vision, where it is infeasible to developan algorithm of specific instructions for execution the task. Machine learning is thoroughlyrelatedtocomputationalstatistics,which emphasesonmakingpredictionsusingcomputers.Machine learning tasks are classified into more than a little broad categories. In supervised learning, thealgorithm builds a mathematical model of a locate of data that contains both the inputs and thedesired outputs. For example, if the task were determining whether an picture contained adefinite object, the training data for a supervisedlearning algorithm wouldcontainimageswith and without that object (the input), and each image would have a label (the output)designating whether it contained the object. In special cases, the input may be only partiallyavailable,orrestrictedtospecialfeedback,Semi-supervisedlearning algorithmsexpandmathematical models from unfinished training data, where a piece of the example inputs aremissingthedesiredoutput.

Data mining, the abstraction of hidden predictive information from large databases, is ainfluentialnewtechnologywithenormouspotential tohelpcompaniesfocal pointonthemostimportant information in their data warehouses. Data mining tools predict future trends andbehaviors,allowingbusinessestomakepositive, knowledge-drivendecisions.Theautomated, potential analyses offered by data miningmove beyond the analyses of pastevents provided by reconsidering tools typical of decision support systems. Data mining toolscan answer business questions that conservatively were too time overriding to determination.They scour databases for hidden patterns, finding predictive information that specialists maymissbecauseitliesoutsidetheiropportunities.

ArecentMETAGroupsurveyofdatawarehouseprojectsfoundthat19%ofdefendants are beyond the 50 gigabyte level, while 59% expect to be there by second quarterof1996.1.Insomeindustries,suchasretail,thesenumberscanbemuchlarger.Thesupplementaryneedforimproved

Computationalenginescannowbemetin acost-effectivemannerwith parallelmultiprocessor computer technology. Data mining algorithms represent techniques that haveoccurred for atleast10 years,buthave only recently been implemented asestablished,reliable,understandabletoolsthatconsistentlyoutclassolderstatisticalmethods.

2.LITRATURE REVIEW

2.1 ReviewOfCredit CardFraudDetection Techniques

Nilson discussing about Credit card fraud is one of the most important threats that affectpeople as well as companies across the world, particularly with thegrowingvolume offinancialtransactionsusingcreditcardseveryday.

Thisputsthesecurityoffinancialtransactions at serious risk and calls for a essential solution. From this paper, we discussvarious methods of credit card fraud detection systems that provide improved protection forcredit card systems against a variety of scams. We also compare these techniques in terms ofaccuracy,time,andcost,and outlinedprobablestrengthsandweaknessestoprovid eaguideline tochoosetherighttechnique.

Today, the credit card system is widely used to settle payments in modern economies tofacilitate business transactions around the world. Given the popularity of the credit cardsystem, it became a target for cyber-attacks and fraud worldwide. This calls for better securitytodealwithpotentialbreachesandillegaluse rs.Inparticular, themostrecognizedcreditcardthreats come from database breaches and identity theft issues. The credit card system looksvulnerable to various risks, hence the pressing need for a more secure financial transactionworldwide.Inthispaper, differentcreditcard frauddetectionmethodsarecensoriouslydiscussed, addressingtheirpotentialstrengths andweaknesses.

Credit card fraud [1] occurs when someone uses a credit card of someone else illegally orsteals information from someone's credit card to make illegal purchases or steals money fromsomeone's bank account. Fraudsters or thieves, who usually find illegal ways to breach creditcardsystems,oftenmakeunauthorizedorillicit transactions.

In thispaper we discusses some statistics of credit card breaches that happened in the lastdecade.

According to"TheNilson Report" [2] that was published on NOV 18, 2019, financial lossesdue to fraud worldwide were $27.85 billion in 2018 and are expected to reach $35.67billionin five years and $40.63 billion in 10 years' time. The losses bigger by more than 15% from2017to2018.TheUnitedStatesscoredthehighe stglobalpercentageasitreached38.6%ofall reported credit card fraud losses in 2018.According to thesamereport,UnitedStateslossesreached$9.47bi llionin2018associatedto$8.98billionin2017.

Credit carddetectiontechniques

Themaintechniquesthatareusedforcreditcardf rauddetectionaresixtechniquesandareidentifiedbelo w:

Neuralnetworks

Neural networks can be defined as a set of interconnected nodes designed to represent theoperative of the human brain [1]. Each node has a weighted connection to several otherlinked nodes in adjacent layers. Sushmi to ghosh and Douglas L. Reilly introduced anapproach that explains most credit card fraud types via neural networks [11]. They used alarge sample of labeled credit card account transactions from a credit card issuer company.Theytestedonaholdoutdataset thatrepresentedallaccountactivityoverasucceedin gtwo-monthperiod.Thenetworkdetectedmorefraudaccoun tswithsignificantlyscarcerfalse-positive transactions than other credit card fraud techniques that can be terminated based onfrauddetectionprocedures.

Theneuralnetworkfrauddetectiontechnique isbasedon thehumanbrainworkingprinciple.Itlearnsfromthe previousactivitiesortransactionsmadebythegenui necardholderto determinewhether theupcomingtransactionislegitimateornot.Neural networksdevelopapattern of uses for each credit card account to help the card issuer decide to stop anysuspiciousorunauthorizedtransaction.

Hiddenmarkovmodel(hmm)

HMMhasmanyadvantages;HMMhasanext remedecreaseinthe False Positivetransactionsrecognized asafraudulent transactionbytheHMMfraud

detectionsystem[13].Theoutcomeis generatedaccordingtoan associatedprobability distribution.

Geneticalgorithms

The genetic algorithm has been introduced by John Holland [18]. Genetic algorithms havebeen inspired by natural evolution. Genetic algorithms search for an ideal solution with apopulation of possible solutions that are always represented as a binary string calledchromosomes. This algorithm is used for predictive purposes, so it is considered one of thesolutions to detect any possible credit card fraud. Genetic algorithms classify credit cardtransactions into two categories: suspicious transactions and non suspicious transactions toensurehighsecurityforbothconsumerandcardissuercompanies.

Fuzzylogicbasedsystem

Fuzzylogic-basedsysteminspiredbyL.AZadeh[21]whothefirstintroducedFuzzylogicin1988. Fuzzy logic systems address the doubt of the input and output variables by sets andnumbersdefined byfuzzylogic to express valuesin an reasonable form. For example, thesystem gives a particular operation as a low, sensible risk, high risk. As we discuss credit cardfraud, fuzzy logic can ensure a highly secured credit card system because it can categorizetransactionsintoalow,sensible,orhightransaction.Hence, afuzzylogic-basedsystemcanstopany potentially fraudulent transactions. The fuzzy logic-based system has two main types,fuzzyneuralnetwork,fuzzyDarwiniansystem .Thecommoncriterionusedinalmostallthestate- of-the-artapproachesoffrauddetection is substantially based on the comparison between the set of previous legitimatetransactions of a user and the new transactions under evaluation. This is a rather trivialcriterionthatinmanycases,duetothehighheterogeneityofdata,leadstowardmisclassifications. In order to overcome this problem, a fraud detection approach should beable to use as much as possible in- formation about the transactions during the evaluationprocess, but this is not always possible due to the inability of some approaches to managesome information (e.g., *Random Forests*, one of the most performing approaches, is notable tomanage typesofdata thatinvolvea largenumberofcategories).

2.2.A Survey Of Machine-Learning AndNature-InspiredBasedCreditCardFraudDetectionTechniques

A. O. Adewumi and A. A. Akinyelu discussed the Credit card fraud can be defined as illegaluseofcreditcard informationforonlinepurchase.Creditcardtransactionsaredonephysically or virtually (Zareapoor et al. 2012). Physical transactions refers to transactionsinvolving physical interaction with seller. Users are required topre- senta physical cardatthepointofpurchase (Zareapoor et al. 2012). Virtual transactions denote to transactionsperformed over the internet or telephone . It require users to provide certain card information(such as CVV number, password, security question, etc.) for online purchase. The discoveryofcreditcardshasnotonlymadeonlinetransactionsunified,easier,comfortableandappropriate, it has also provided new fraud opportunities for criminals, and increased the rateof fraud (Maes et al.2002) (Pun 2011).Every credit card user view the risk of falling victimtocardnotpresentfraud andmerchantsallowsthecostofirregulartransactions (FFA2015).

Credit card fraud detection is a categorization problem (Wong et al. 2012). Credit cardnumbers are generated via Luhnalgorithm.The algorithm does notcategorically protect users from online fraud, it essentially helps in authenticating data inputfrom users (Wong et al. 2012). Somegauge companies use manual authentication methods,including:validationofphonenumbers,physicaladdress,secretquestionandanswer.However, thismethodsmay not be feasible for large scale companies, they are expensiveandinefficient(Wongetal.2012).

Additionally,most onlinemerchants now use Card VerificationValue (CVV2) as anadditional security measure for approval of card-not-present transactions (Wong et al. 2012).Although,thisadditionalsecuritymeasureha sreducedcard-not-presentfraudtoareasonable minimum, it does not prevent fraud that occurs due to lost or stolen card (Wong etal.2012).Address VerificationServicecanbeusedtocombatcard-not-presentfraud.Itis an electronic service that verifies transactions by using shipping address details of card owners(Wong et al. 2012). This method reduces fraud, however, it lead to loss in sales, because, notall customers are willing to ship purchased items to their billing address (Wong et al. 2012).Furthermore, MasterCard and VISA card haspresenteda 3-D secured protocol for onlinebanking, they include MasterCard Secure CodeandVerified byVISA(Wonget al. 2012).These protocols use a digital certificate to authenticate online merchants and password toauthenticatecustomers(Wongetal.2012).

Fraud detection isa datamining problem with an aim of segregating transactionsintotwo classes – legitimate and fraudulent (Duman and Ozcelik 2011). Current fraud detectionsystems used by merchants and banks are planned to confirm transactions by checkingspending patterns and behavior of customers (Quah and Sriganesh 2008). To achieve this,fraud detection systems use predictionalgorithmstoclassify pattern observations(Maeset al. 2002). A transaction will be labeled fraudulent if the system observes a deviation inthenormalspendingpattern ofauser.

Two major methods used to handle fraud include: fraud prevention and fraud detectionFraud prevention aims to stop fraudulent activity from taking place. Fraudrecognition toidentify fake transactions, and in turn prevent authorization of the transaction (Sahin andDuman2011).Frauddetectionoriginatesafter asystemfailstostophoaxerfrominitiatingatransaction,thatis,after the hoaxer has started the transaction. Most of theexisting fraud detection mechanism, such as Chip and PINhas failed to handle fraudmeritoriously.

3. SYSTEM ANALYSIS

The obtained model can be useful in anti-fraud monitoring systems, or a alike modeldevelopment procedure can be performed in connected business areas to sense fraud and reducethe rateofsuchbehaviors.

3.1 EXISTING SYSTEM

Our existing has made a detail study on fraud detection using themethod of naturalobservation of the events happened from the customer side. The existing has worked on themethods of collecting the data from the social media and framing them in terms of big datamodelsandworkingonthechallengesexistedthe field.

Implementedasystemwhichsupports in the detection of the scams or frauds in the field of the business by recording thetransactions and there by building a model using data mining models.In existing systemAdaBoost algorithm has been implemented. The Random Forest (RF)algorithm has beenimplemented. Loss from credit card fraud affects the merchants, where they bear all costs,includingcardissuerfees,charges,andadministrative charges.

3.1.1.DRAWBACKS

- Difficulttoidentify
- Insecure

3.2 PROPOSED SYSTEM

The proposed system first step is data collection in this step data collection fromKaggle. After collecting data to processing using machine learning algorithm. The processingdataisconvertedintoprinciplecomponentanalysis.Inthisstepusing10differentcomponent.

Data splitting process to split a data from principle component analysis and k-told cross validation used. Model development process using machine learning algorithm totrain a data to

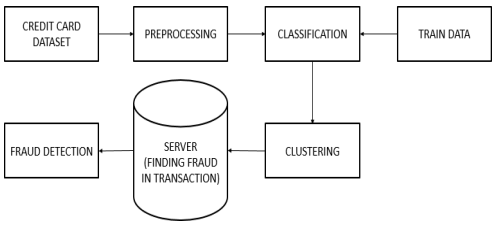
send next step of the process. Performance evaluation implemented this projectusingformulas.Finallywedeployedtheproje ctusingmodelselection.Majorityvotingis frequently used in data classification, which involves a combined model with at least twoalgorithms.Machinelearningbasedalgorithm isimplementedforfrauddetection.Eachalgorithm makes its own prediction for every testsample. Itwill be extended to onlinelearning models.

In addition, other online learning models will be investigated. The use ofonline learning will enable rapid detection of fraud cases, potentially in real-time. In theproposed system the KNN algorithm is used. The KNN algorithm is a non-parametric methodused for classification and regression. By using K Nearest Neighbor the processing time iscompactandalsoprocessesthelargerdatasets.

3.2.1 ADVANTAGES

- Stablesystem
- More accuracy
- Lesstimetopredictthefraud

4. ARCHITECTURE DIAGRAM



.

Fig 4.1 Architecture diagram

The System architecture is the hypothetical model that defines the organization, performance, andextra visions of a system. A system architecture can occupy of system components and thesub-systemsdeveloped,thatwillemploymenttogethertoi mplement thecompletesystem**.**

5.SYSTEM IMPLEMENTATION

Implementation is the phase in the project wherever the theoretical design is turned into a working system. The implementation phase constructs, installs and operates the new system. The most crucial phase in achieving a new successful system is that it will work efficiently and effectively.

5.1 MODULES

1. AdminModule
2. BankAdmin
3. UserModule
4. ProductSearchModule
5. Paymentmodule

5.2 Module Description

5.1.1 AdminModule

- Inthismoduleadminverifyandauthorizethe bothregisteredbankadminandregistereduse raccount.
- AdminaddandviewProducts.Userpurchase dhistoryalsoadminviewinthismodule.
- TheRandomForest(RF)createsanensemble ofrandomtreesitwillclassifythedata'sintoC VV fraudandExpiryDate problems.
- DependsuponRFalgorithmAdminviewthe userfraudulentactivitiesgraphicallyinthism odule.

5.2.2 BankAdmin

- Inthismodulebankadminmanagetheuserrequ irementsandfraudulentactivities.
- Acceptttheusercreditcardrequestandshowthe usagedetails also.
- SendthecardCVVandexpirydatefornewlya pprovedcreditcarduserinthemodule.

5.2.3 UserModule

- Inthis moduleuserneedsregistrationandlogin.
- Newlyregisteredusersendcreditcardrequestf

orbankadmin.

- Afterapprovalofyourcardyouwillusethatcardforpurchasingandtransactions.
- Usercheckhiscardusageforhisaccount.

5.2.4 Product search module

- Theycan findproductseasilyword search.
- Afterviewingthedescriptionsselectingtheir requiredbrandsandonconfirmationtheycan boughtproducts.
- Theusercaneasilygothroughthesitebyjusth avingtheminimumknowledgeofcomputeris sufficienttousethissite.

5.2.5 Payment module

- The module takes care of the all the secured payments that should happen for thepurchases that happens online, so to implement a security algorithm is one of themajorconceptsofthepayment.
- VerifytheCVV detailsaswellasbalanceofyourcard also.
- IfuserenterwrongCVVorlowbalanceaccou ntmeansitwillterminatethetransactionproc ess.
- Herewearecalculatetheexpirydateofourcredi t cardalso.
- If fraud user try to used expired card means it will analyze and terminate the paymentprocess.

6. CONCLUSION

It is worth keep in mind that objective of the paper is to surveying on the differentclassifierwhichareusedinthecreditriskevalu ation.Inthispaperdifferenttypesofclassifiers are discussed and also different types of ensemble classifiers are briefed. Thedataset which are used in the classifier is discussed in the paper. We have analysed andcompare their accuracies using different types classifiers andfrom comparison table wefound that the ELM classifier gives better accuracies compare to other classifiers that is ELMgives96.33(%)inGermandatasetand96.32(%)in Australiandataset.

7. FUTURE ENHANCEMENT

The methods studied in this paper will be extended to online learning models.Inaddition, other online learning models will be investigated. Theuseofonline learningwillenablerapiddetectionoffraudcases, potentiallyinreal-time.Thisinturnwillhelpdetectand prevent fraudulent transactions before they take place, which willreducethenumberoflossesincurredeverydayinth efinancialsector.

References

[1] Y. Sahin, S. Bulkan, and E. Duman, “A cost-sensitive decision tree approach for frauddetection,”ExpertSystemswith Applications,vol.40,no.15,pp.5916–5923,2013.

[2] A. O. Adewumi and A. A. Akinyelu, “A survey of machine-learning andnature-inspiredbasedcreditcardfrauddetectiontechniques, ”InternationalJournalof SystemAssuranceEngineeringandManagement,vo l.8,pp.937–953,2017.

[3] TheNilsonReport(October2016)[Online]. Available:https://www.nilsonreport.com/upload/c ontent_promo/The_Nilson_Report_10-17-2016.pdf

[4] N. S. Halvaiee and M. K. Akbari, “A novel model for credit card fraud detectionusingArtificialImmuneSystems,”Applie dSoftComputing,vol.24,pp.40–49,2014.

[5] N.MahmoudiandE.Duman,“Detectingcredit cardfraudbymodifiedFisherdiscriminant analysis,” Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516,2015.

[6] R.SaiaandS.Carta,“EvaluatingCreditCardTra nsactionsinthe FrequencyDomainforaProactiveFraudDetectionA pproach,” In Proceedingsofthe14thInternationalJointConferenc

eone- Business and Telecommunications, vol. 4, pp. 335–342,2017.

[7] E. Duman, A. Buyukkaya, and I. Elikucuk, "A novel and successful credit card frauddetection system Implemented in a Turkish Bank," In IEEE 13th International Conference onData MiningWorkshops(ICDMW),pp.162–171,2013.

[8] M. Seera, C. P. Lim, K. S. Tan, and W. S. Liew, "Classification of transcranial Dopplersignals using individual and ensemble recurrent neural networks," Neurocomputing, vol. 249,pp.337-344,2017.

[9] Y. Li, C. Yan, W. Liu, and M. Li, "A principle component analysis- based random forestwith the potential nearest neighbor method for automobile insurance fraud identification,"Applied SoftComputing,tobe published.DOI: 10.1016/j.asoc.2017.07.027.

[10] F. H. Chen, D. J. Chi, and J. Y. Zhu, "Application of Random Forest, RoughSetTheory,DecisionTreeandNeuralNetwor ktoDetect FinancialStatementFraud– TakingCorporateGovernanceinto Consideration,"In InternationalConferenceonIntelligentComputing, pp.221–234,Springer,2014.

[11] C. F. Tsai, "Combining cluster analysis with classifier ensembles to predictfinancialdistress" InformationFusion,vol.16,pp.46–58,2014.

[12] Chan, Philip K., et al. "Distributed data mining in credit card fraud detection." IEEEIntelligentSystems andTheirApplications 14.6(1999):67-74

[13] Gaikwad, Jyoti R., et al. "Credit Card Fraud Detection using Decision Tree InductionAlgorithm."InternationalJournal of InnovativeTechnology andExploringEngineering(IJITEE) 4.6(2014).

[14] I.T.Christou,M.Bakopoulos,T.Dimitriou,E.Amolochitis,S.Tsekeridou,andC.Dimitridis,"Detectingfraudinonlinegamesofchanceandlotteries,"ExpertSystemswith Applications,vol.38,no.10,pp.13158–13169,2011.

[15] D. Olszewski, "Fraud detection using self-organizing map visualizing the userprofiles,"Knowledge-BasedSystems,vol.70,pp.324–334,2014.

