ELSEVIER Scopus doi

Journal of Vibration Engineering

ISSN:1004-4523

Registered

- SCOPUS
- GOOGLE SCHOLAR
- DIGITAL OBJECT IDENTIFIER (DOI)
- IMPACT FACTOR 6.1

Our Website
www.jove.science

CreditCardFraudDetectionUsingDataMining

PranjalSingh[1],LokendraSinghUmrao[2]

[1]ComputerScience&Engineering,I.E.T,Dr.RammanoharLohiaAvadhUniversity,Ayodhya,UttarPradesh,India.
[2]AssistantProfessor,Dept.of[1] ComputerScience&Engineering,I.E.T,Dr.RMLAvadhUniversity,Ayodhya,UttarPradesh,India

Abstract— **Now days credit card fraud is a seriousproblemin financial services. Billions of dollars are lostdue to credit card fraud every year. There is a lack ofresearch studies on analyzing real-world credit card dataowingtoconfidentialityissues.Inthispaper,machinelearning algorithms are used to detect credit card fraud.Standardmodelsarefirstused.Then,hybridmethodswhichuseAdaBoostandmajorityvotingmethodsareapplied.Toevaluatethemodelefficacy,apubliclyavailable credit card data set is used. Then, a real-worldcredit card data set from a financial institution is analyzed.In addition, noise is added to the data samples to furtherassess the robustness of the algorithms. The experimentalresults positively indicate that the majority voting methodachieves good accuracy rates in detecting fraud cases incredit cards.**

Keywords:creditcard,AdaBoost,detectingfraud,accuracyrates,robustness.

1. INTRODUCTION

As per Global Payments Report 2015, Mastercard is the mostnoteworthy utilized installment technique around the world in2014contrastedwithdifferentstrategies,forexample,e-wallet and Bank Transfer [1]. The tremendous value-basedadministrations are frequently looked at by digital crooks todirect false exercises utilizing the Mastercard administrations.Visa extortion is characterized as the unapproved use of card,surprisingexchangeconduct,orexchangeson anidlecard[2].By and large, there are three classifications of charge cardextortionspecifically,ordinarycheats(forexampletaken,phony and fake), online cheats (for example bogus/counterfeitdealer destinations), and shipper related fakes (for exampleshipper arrangement and triangulation) [3]. In the past two orthreetheyears,Mastercardbreakshavebeenmovingalarmingly. As per Nilson Report, the worldwide Mastercardextortion misfortunes came to $16.31 billion out of 2014 andit is assessed that it will surpass $35 billion of every 2020 [4].Inthismanner,itisimportanttofosterVisaextortionidentificationstrategiesasthecountermeasuretobattlecriminaloperations.Byandlarge,Mastercardextortionidentification has been known as the method involved withdistinguishing whether exchanges are certifiable or fake. Astheinformation miningand AImethods are immeasurablyusedtocounterdigitallawbreakercases,researchersfrequentlyembracedthosewaystodealwithstudyand distinguishchargecardextortionexercises.Informationminingis knownasthemostcommonwayofacquiringfascinating,noveland cannyexamplesaswellasfindingreasonable,illustrativeandprescient modelsfromhugesize ofinformation assortments [5], [6]. The capacity of informationmining methods to remove productive data from enormoussize of information utilizing factual and numerical strategieswouldhelpMastercardextortionrecognitioninlightofseparatingtheattributesofnormalanddubiousVisaexchanges. While information mining zeroed in on findingsignificantknowledge,AIisestablishedinlearningtheknowledge and fostering its own model with the end goal ofgrouping, bunching or so on. The utilization of AI proceduresspreadsgenerallyallthroughPCsciencesspaces,forexample,spam sifting, web looking, promotion position, recommenderframeworks,creditscoring,drugplan,misrepresentationlocation,stockexchanging,andnumerousdifferentapplications. AI classifiers work by building a model frommodelinformationsourcesandutilizingthattosettleonforecastsorchoices,asopposedtoadheringtorigorouslystaticprogramguidelines.TherearevariouskindsofAIapproachesaccessiblewiththeexpectationstotackleheterogeneousissues.Becauseoftheideaof this reviewwhichwascenteredaroundorder,theconversationthatfollowsdependsonthissubject.AIorderalludestothemethod involved with figuring out how to appoint cases topredefinedclasses.Officially,thereareafewsortsoflearning,for example,directed,semi-managed,solo,support,transduction and figuring out how to learn [7]. As the interestof this review was to lead managed based AI grouping, theconversations about the other strategies are disposed of fromadditionalelaboration.Inmostgroupingreview,supervisedbasedlearningisinclinedtowardmorethandifferentstrategiesbecauseofthecapacitytocontroltheclasses of the occurrences with the mediations of human. Inregulated learning, the classes of the occurrences would bemarked preceding taking care of into classifiers. Then, at thatpoint,byutilizingspecificassessmentmeasurements,theexhibitionsoftheclassifierscouldbeestimated.

In this paper section I contains the introduction, section IIcontains the literature review details, section III contains thedetails about methodologies, section IV shows architecturedetails,VdescribetheresultandsectionVIIprovideconclusionofthispaper.

2. LITERATUREREVIEW

A formative information mining and AI are famous strategiesto study and battle the charge card misrepresentation cases.There is countless examinations that took advantage of thestrength of information mining and AI to forestall the chargecardfakeexercises.InlightofSelf-OrganizingMapandNeuralNetwork,theinvestigationof[8]gotReceiverOperatingCurve(ROC)morethan95.00%ofmisrepresentationcaseswithoutphonyproblemsrate.TheHidden Markov Model (HMM) additionally has been appliedin charge card misrepresentation recognition with low level ofdeceptionrates[9].Nonetheless,changecycleofvariousstates and ascertaining the likelihood in HMM are exorbitantand escalated. Besides, instead of utilizing single classifiers, aportion of the Mastercard extortion identification concentrateson utilized metalearning students in view of directed learning.Stolfoetal.exploredMastercardmisrepresentationidentification framework utilizing four sorts of calculations tobe specific Iterative Dichotomiser 3 (ID3), Classification andRegression Tree (CART), Ripper and Bayes as base studentsand tried with heterogeneous information circulations [10]. Inviewofhalf/halfdisseminationofoccurrences(misrepresentationandnon-extortion),theinvestigationdiscoveredthatmetalearninginvolvingBayesasabasestudentgotahighergenuinepositiveratecontrasted withothermetastudents.Inany case,despitethefact thatthedissemination of half/half yields great outcomes, it doesn'treflectcertifiableconditionswhereveritableMastercardexchangesareveryhigherthannon-authenticexchanges.Scientists have likewise tried different sorts of meta learningclassifiers, for example, Adaboost, Logitboost, Bagging andDaggingandyieldedintriguingresults[11].

Through our writing review, Bayesian Network is one of theclassifier types that have been generally applied to recognizemisrepresentationinthechargecardindustry.Maesetalinspected the genuine positive and misleading positive createdby Bayesian Belief Network and Artificial Neural Network onorderingMastercardmisrepresentationoccurrences.TheinvestigationdiscoveredthatBayesianorganizationperformedaround 8% higher than Artificial Neural Network and assertedthat the previous' classifier handling time is more limited thanthelast[12].Insteadofexaminingutilizingcustomaryarrangementstrategies,theexaminationby[13]startedtoperform cost delicate Visa misrepresentation recognition inlightof BayesMinimumRiskstrategy.

The review estimated the exhibitions of Logistic Regression(LR), C4.5 and Random Forest (RF). The review showed thatchanging the probabilities of Bayes Minimum Risk classifieron RF order yielded reliably improved results than LR andC4.5.Allthroughourperceptionandexaminationofpastinvestigations, Bayesian Network classifiers have become oneof the well known classifier types that are generally used tocharacterize Visa misrepresentation information. Hence, thisreview endeavored to research the characterization by a fewBayesianclassifiers,forexample,K2,TreeAugmentedNaïve Bayes (TAN), and Naïve Bayes. In addition, this concentratelikewise estimated the exhibitions of Logistics Regression andJ48 in light of the proposed philosophy. A short conversationabout Bayesian Network Classifier and proposed classifiersareexpressedunderneath.

Essentially,theobjectiveofmisrepresentationrecognitionoughttobematchedtoaninformationminingtechnique.Information, as a rule, mining procedures can be partitionedintotwokindsasfaraswhetherthefakeoccasionisrecognized in the past information: managed andsolo [3].Ngaietal.[4]haveshownthatgroupingasa managedstrategyisthemostofteninvolvedinformationminingapplicationinmonetarymisrepresentationidentification.Regardless, a classifier ought to characterize every client intooneofthetwoclassesoftypicalorfalseclients.Withacomplete view,we observe thatweareconfrontedwith aspecific kind of characterization issue. Taking into account abank data set with a large number of exchanges in a day, justexactly couple of exchanges might be dubious in a month. Allinall,weareconfrontedwithasuperimbalanceddataset.Theissue with an awkwardness informational index is the slanteddisseminationoftheinformationthatmakesthelearningcalculationsineffectual,particularlyinforeseeingtheminorityclasses. In this segment, we audit the writing in which issueswithimbalancedinformationarrangementandchargecardextortion discovery methods are. Albeit the absence of freelyaccessibledatasetshasrestrictedthedistributionsonmonetary extortion identification, in this part we will survey aportionoftheaccessibleones.

Olivier Caelen,2014, [22] Detectionissuesarecommonlytended to in two distinct ways. In the static getting the hang ofsetting,anidentificationmodelisoccasionallyrelearntwithout any preparation (for example when a year or month).Inthewebbasedpickingupsetting,theidentificationmodelis refreshed when new information shows up. However thistechniqueisthemostsufficienttomanageissuesofnonstationarity (for example because of the advancement of thespending conduct of the normal card holder or the fraudster),littleconsiderationhasbeengiveninthewritingto theunequal issue in evolving climate. One more dangerous issuein Visa identification is the shortage of accessible informationbecause of classification gives that allow little opportunity tothe local area to share genuine datasets and survey existingstrategies. This paper targets making a trial correlation of afew cutting edge calculations and demonstrating procedureson one genuine dataset, zeroing in specifically on a few openinquiries like: Which AI calculation ought to be utilized? Is itenough to gain proficiency with a model one time each monthor it is important to refresh the model regular? What numberof exchanges are adequate to prepare the model? Should theinformation be broke down in their unique lopsided structure?If not, which is the most ideal way to rebalance them? Whichexecution measure is the most satisfactory to asses results? Inthis paper we address these inquiries fully intent on evaluatingtheirsignificanceongenuineinformationandaccordingtoan expertpointofview.Thesearenonlysomeofpotentialinquiriesthatc ouldraiseduringtheplanofarecognitionframework. We don't really have the option to offer a positiveresponse to the issue, however we desire to that our work fillsinasruleforothersinthefield.

Rencheng Ton, 2007, [23] Credit card misrepresentation canbeseparatedinto2kinds:inwardcardextortionandoutercardmi srepresentation. Inward card extortion plans to cheat themoney.Generallyitistheintrigueamongtradersandcardholder s, utilizing bogus exchanges to cheat banks cash.Outsidecardextortionisfundamentallyencapsulatedatutiliz ingthetaken,phonyorfakeVisatoconsume,orutilizingcardstoget cashinmaskedstructures,forexample,purchasingthecostly, littlevolume waresortheitemsthatcanundoubtedly be changed into cash. This paper is principallycommittedtotheexaminationoftheoutsidecardmisrep resentation,whichrepresentsmostofVisacheats.IdentifyingMast ercardextortionisatroublesome errandwhile utilizing typical methods, so the improvement of theVisa misrepresentation identification model has happened toimportance, whether in the scholar or business local area as oflate.Thesemodelsaregenerallymeasurementsdrivenorcounter feitcleverbased,whichenjoythehypotheticalbenefitsinnotmonu mentalerraticsuppositionsontheinfofactors.

Khyati Chaudhary, 2012, [24] In present situation when theterm misrepresentation comes into a conversation, charge cardextortionsnapstomindupuntilthispoint.Withtheextraordina ryexpansioninMastercardexchanges,Visaextortionhasexpandin gunnecessarilyasoflate.Misrepresentation identification incorporates checking of thespending conduct of clients/clients to assurance, discovery, oraversion of bothersome way of behaving. As Visa turns intothe most overall method of installment for both online as wellas ordinary buy, extortion relate with it are likewise speedingup. Extortion identification is worried about catching the fakeoccasions,yetadditionallycatchingofsuchexercisesasfastas could beexpected. Theutilizationofchargecardsisnormalincurrentsociety.Misrepre sentationisamillionsdollarbusinessanditisrisingconsistently.Mi srepresentationpresentstremendousexpenseforoureconomyaro undtheworld. Present day methods in view of Data mining, Machinelearning,SequenceAlignment,FuzzyLogic,GeneticPro gramming,ArtificialIntelligenceandsoon,hasbeenpresented for distinguishing Mastercard false exchanges. Thispaper demonstrates the way that information mining strategiescan be consolidated effectively to get a high misrepresentationinclusionjoinedwith aloworhighphonyproblem rate.

3. METHODOLOGIES

- **DecisionTree(DT)**

Thepresentationofdatainformofatreestructureisusefulforeaseof interpretationbyusers.TheDecisionTree(DT)isacollectionofnod esthatcreatesdecisiononfeatures connected to certain classes. Every node represents a splittingruleforafeature.Newnodesareestablished untilthestoppingcriterion is met. The class label is determined based on themajorityofsamplesthatbelongtoaparticularleaf.TheRandom Tree(RT)operatesasaDToperator,withtheexceptionthatineachs plit,onlyarandomsubsetoffeaturesis available. It learns from both nominal and numerical datasamples.Thesubsetsizeisdefinedusingasubsetratioparamete r.
TheRandomForest(RF)createsanensembleofrandomtrees.Theu sersetsthenumberoftrees.Theresultingmodelemploysvotingofal lcreatedtreestodeterminethefinalclassification outcome. The Gradient Boosted Tree (GBT) isan ensemble of classification or regression models. It usesforward-learning ensemble models, which obtain predictiveresults using gradually improved estimations. Boosting helpsimprovethe treeaccuracy.

- **NaïveBayes(NB)**

Naïve Bayes (NB) uses the Bayes' theorem with strong ornaïveindependenceassumptionsforclassification.Certainfeat uresofaclassareassumedtobenotcorrelatedtoothers.It requires only a small training data set for estimating themeansand variancesisneededfor classification.

- **TheRandomForest(RF)**

TheRandomForest(RF)createsanensembleofrandomtrees.Theu sersetsthenumberoftrees.Theresultingmodelemploysvotingofal lcreatedtreestodeterminethefinalclassification outcome. The Gradient Boosted Tree (GBT) isan ensemble of classification or regression models. It usesforward-learning ensemble models, which obtain predictiveresults using gradually improved estimations. Boosting helpsimprovethetreeaccuracy.TheDecisionStump(DS)generat es adecisiontreewithasinglesplit only. It can beusedinclassifyingunevendatasets.

- **AdaBoostandMajorityVoting**

Adaptive Boosting or Ada Boost is used in conjunction withdifferent types of algorithms to improve their performance.The outputs are combined by using a weighted sum, whichrepresentsthecombinedoutputoftheboostedclassifier.Ada Boost tweaks weak learners in favor of misclassified datasamples.It is, however,sensitivetonoise and outliers. Aslong as the classifier performance is not random, AdaBoost isabletoimprovetheindividualresultsfromdifferentalgorithms.
AdaBoost helps improve the fraud detection rates,with a noticeable difference for NB, DT, RT, which produce aperfect accuracy rate. The most significant improvement isachieved by LIR. Majority voting is frequently used in dataclassification, which involves a combined model with at leasttwo algorithms. Each algorithm makes its own prediction forevery test sample. The final output is for the one that receivesthemajorityofthevotes.Themajorityvoting methodachieves good accuracy rates in detecting fraud cases in creditcards.

- **MachineLearningAlgorithms**

Machine learning is the science of designing and applyingalgorithms that are able to learn things from past cases. It usescomplexalgorithmsthatiterateoverlargedatasetsandanalyze thepatternsindata.Thealgorithmfacilitatesthemachinestorespon dtodifferentsituationsforwhichtheyhavenotbeenexplicitlyprogr ammed.Itisusedinspamdetection,imagerecognition,productrec ommendation,predictive analytics etc. Significant reduction of human effortis the main aim of data scientists in implementing ML. Evenwith modern analytics tools, it takes a lot of time for humansto read, collect, categorize and analyze the data. ML teachesmachines to identify and gauge the importance of patterns inplaceofhumans.Particularlyforusecaseswheredatamustbeanal yzed and acted upon in a short amount of time, having thesupport of machines allows humans to be more efficient andactwithconfidence.

4. SYSTEMARCHITECTURE

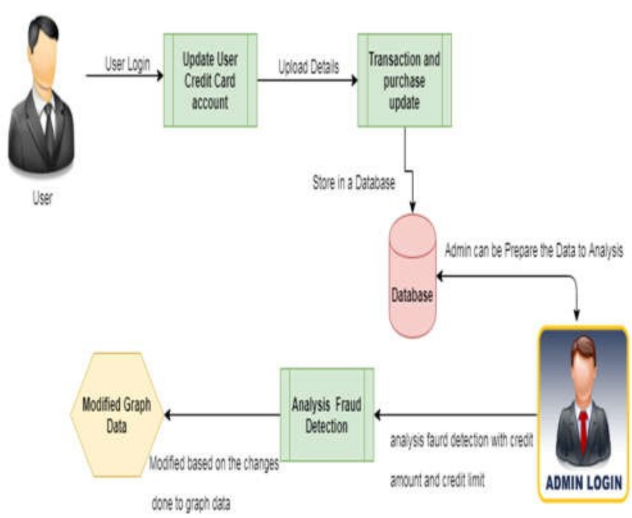

Figure1 SystemArchitecture

5. RESULTS

In this paper outcome part, we step up and study themachinelearning algorithms are used for detecting credit card fraud.The algorithms range from standard neural networks to deeplearning models. They are evaluated using both benchmarkand real-world credit card data sets. In addition, the AdaBoostand majority voting methods are applied for forming hybridmodels. To further evaluate the robustness and reliability ofthe models, noise is added to the real-world data set. The keycontribution of this paper is the evaluation of a variety ofmachine learning models with a real-world credit card data setfor fraud detection. While other researchers have used variousmethods on publicly available data sets, the data set used inthis paper are extracted from actual credit card transactioninformationoverthree months.

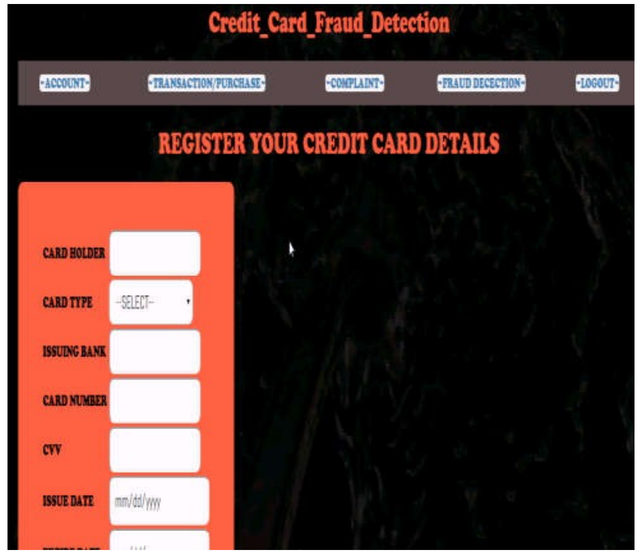

Figure2:Registercreditcardsdetails

In figure 2, user can register for credit card, by providing theirpersonaldetailsbankname,citynameandotherrelateddetails which is required. After completing this process user details isstoredinserverwithreferencetocreditcardregistrationdetails.

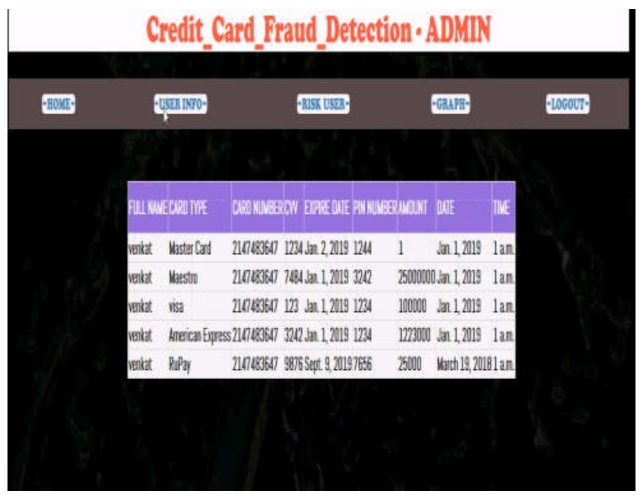

Figure3:creditcardsdetails

In figure 3, it shows the credit card user details, like issuingdate, name, card number and address etc. Admin can view thecreditcarddetailsofanyuser.

Figure4:creditcardsfraudalertmessage

In figure 4, display the credit cards fraud alert messages, sothat user is aware about any possible fraud. These fraud alertmessages are very help full to prevent and detection of anyfraudpossibilities.

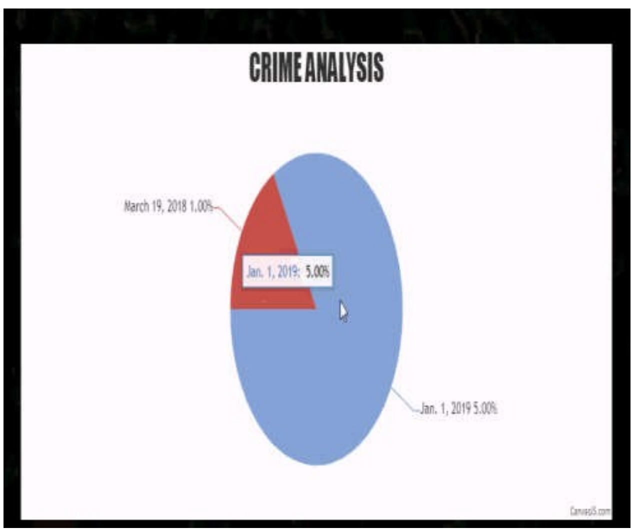

Figure5:Crimeanalysispiechart

In figure 5, display the pie chart graph of the crime analysis, itishelpfulltounderstandthenumberfraudcrimeyearbyyear.

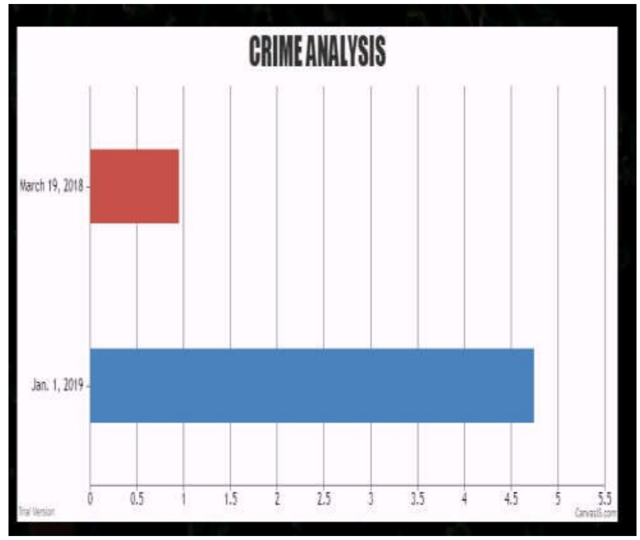

Figure6:Crimeanalysisbarchart

In figure 6, display the bar chart graph of the crime analysis, itishelpfulltounderstandthenumberfraudcrimeyearbyyear.

6. CONCLUSION

In this research paper, we have considered the novel conceptof study on credit card fraud detection using machine learningalgorithms has beenpresentedin this paper.A number ofstandard models which include NB, SVM, and DL have beenused in the empirical evaluation. A publicly available creditcard data set has been used for evaluation using individual(standard)modelsandhybridmodelsusingAdaBoosta ndmajority voting combination methods. The MCC metric hasbeenadoptedasaperformancemeasure,asittakesintoaccountt hetrueandfalsepositiveandnegativepredictedoutcomes.Thebest MCCscoreis0.823,achievedusingmajority voting. A real credit card data set from a financialinstitutionhasalsobeenusedforevaluation.Thesameindi vidual and hybrid models have been employed. A perfectMCCscoreof1hasbeenachievedusingAdaBoostandmajo rityvotingmethods.Tofurtherevaluatethehybridmodels, noise from 10% to 30% has been added into the datasamples. The majorityvoting method has yielded thebestMCC score of 0.942 for 30% noise added to the data set. Thisshowsthatthemajorityvotingmethodoffersrobustperforman ce inthepresenceofnoise.

REFERENCE

[1] WorldPay.(2015,Nov).Globalpaymentsreportpreview:your definitiveguidetotheworldofonlinepayments.Retrieved September 28, 2016, fromhttp://offers.worldpayglobal.com/rs/850-JOA856/images/GlobalPaymentsReportNov2015.pdf.

[2] FederalTradeCommision.(2008).consumersentinelnetwork - data book for January - December 2008. RetrievedOct20, 2016.Fromhttps://www.ftc.gov/.

[3] Bhatla,T.P.,Prabhu,V.,andDua,A.(2003).understandingcreditcardfrauds.CradsBusinessReview#2003-1,TataConsultancyServices.

[4] TheNilsonReport. (2015). Global fraud losses reach $16.31Billion.Edition:July2015,Issue1068.

[5] Y. Sahin and E. Duman, "Detecting credit card fraud bydecision trees and support vector machines", *Proceedings oftheInternationalMulti-ConferenceofEngineersandComputerScientists2011VolI,IMECS2011,March2011.*

[6] Elkan, C. (2001). Magical thinking in data mining: lessonsfromCOILchallenge2000.Proc.ofSIGKDD01,426-431.

[7] Mohammed, J. Zaki., & Wagner, Meira Jr. (2014). Datamining and analysis: fundamental concepts and algorithms.CambridgeUniversityPress.ISBN978-0-521-76633-3.

[8] F.N.Ogwueleka.(2011).Dataminingapplicationincredit card fraud detection system. *Journal of EngineeringScienceandTechnology,Vol.6,No.3 (2011)311-322.*

[9] V.Bhusari&S.Patil.(2011).Applicationofhiddenmarkov model in credit card fraud detection. *InternationalJournal of Distributed and Parallel Systems (IJDPS) Vol.2,No.6.*

[10] S.J.Stolfo,D.W.Fan,W.Lee,A.L.Prodromidis,and P.K. Chan. (1998). Credit card fraud detection using meta-learning: issues and initial results, *Proc. AAAI Workshop AIMethodsinFraudandRiskManagement,pp.83-90.*

[11] Sen,SanjayKumar.,&Dash,Sujatha.(2013).Metalearningalgorithmsforcreditcardfrauddetection.*InternationalJournalofEngineeringResearchandDevelopmentVolume6,Issue6,pp.16-20.*

[12] Maes,Sam,TuylsKarl,VanschoenwinkelBram&Manderick, Bernard. (2002). Credit card fraud detection usingbayesian and neural networks. *Proc. of 1st NAISO CongressonNeuroFuzzyTechnologies. Hawana.*

[13] A.C. Bahnsen, Aleksandar, Stojanovic., D. Aouada&Bjorn,Ottersten.(2013).Costsensitivecreditcardfraud detectionusingbayesminimumrisk.*12thInternationalConferenceonMachineLearningandApplications*.

[14] AmlanKundu,SuvasiniPanigrahi,ShamikSuralandArunK. Majumdar. (2009). Credit cardfraud detection: afusion approach using dempster–shafer theory and bayesianlearning. *Special Issue on Information Fusion in ComputerSecurity, Vol.10, Issue No.4,pp.354-363*.

[15] Lam,Bacchus(1994).Learningbayesianbeliefnetworks:anapproachbasedontheMDLprinciple.*ComputationalIntelligence,Vol.10,IssueNo.3,pp.269–293*.

[16] M. Mehdi, S. Zair, A. Anou and M. Bensebti (2007). Abayesiannetworksinintrusiondetectionsystems.*International JournalofComputationalIntelligenceResearch,Issue No.1,pp.0973-1873Vol.3*.

[17] R.Najafi&Afsharchi,Mohsen.(2012).Networkintrusiondetectionusingtreeaugmentednaive-bayes.*TheThird International Conference on Contemporary Issues inComputerandInformationSciences(CICI)2012.*

[18] G. Cooper, E. Herskovits (1992). A bayesian method forthe induction of probabilistic networks from data. MachineLearning.9(4):309-347.

[19] Quinlan,J.R.(1993).C4.5:ProgramsforMachineLearning. MorganKaufmannPublishers.

[20] Friedman,N.andGoldszmidt,M.(1996).Buildingclassifiersusingbayesiannetworks.Proc.13thNationalConferenceonArtificialIntelligence.Vol.2,pp1277-1284.

[21] Friedman, N., Geiger, D. and Goldszmidt, M. (1997).Bayesian network classifiers. machinelearning,Vol. 29, pp131-163.KluwerAcademicPublishers, Boston.

[22] Andrea Dal Pozzolo a, Olivier Caelenb ,Yann-Aël LeBorgne a , Serge Waterschoot, "Learned lessons in credit cardfrauddetectionfromapractitionerperspective",A.DalPozzolo et al. / Expert Systems with Applications 41 (2014)4915–4928.

[23] RenchengTong,AihuaShen,"ApplicationofClassification Models on Credit Card Fraud Detection", 1-4244-0885-7/07/2007IEEE.

[24] KhyatiChaudhary,JyotiYadav,"AreviewofFraudDetection Techniques: Credit Card", International Journal ofComputerApplications(0975–8887)Volume45–No.1,May2012.