INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

ANTI-FRAUD MODEL FOR INTERNET LOAN USING DEEP LEARNING

[1]D.V.S.Deepak, Assistant professor [2]I.Durga Prasad [3]Ch.jagadeesh [4]G.jhansi

Miracle Educational Group of Institutions, Vizianagaram, A.P, India

ABSTRACT

Recently, Internet finance is increasingly popular. However, bad debt has become a serious threatto Internet financial company. The fraud detection models commonly used in conventionalfinancial companies is logistic regression. Although it is interpretable, the accuracy of the logisticregression still remains to be improved. This paper takes a large public loan dataset, e.g. Lendingclub, for example, to explore the potential of applying deep neural network for fraud detection. We first _ll the missing values by a random forest. Then, an XGBoost algorithm is employed toselect the most discriminate features. After that, we propose to use a synthetic minorityoversampling technique to deal with the sample imbalance. With the preprocessed data, we designa deep neural network for Internet loan fraud detection. Extensive experiments have beenconducted to demonstrate the outperformance of the deep neural network compared with thecommonly-used models. Such a simple yet effective model may brighten the application of deeplearning in anti-fraud for Internet loans, which would benefit the financial engineers in small andmedium Internet financial companies

1. INTRODUCTION

Internet fraud methods are increasing dramatically in recent years, together with the rapiddevelopment of Internet financial models and the Internet business used to be handled by traditionalfinancial institutions. In this regard, Internet lending companies face an unprecedented risk of onlinefraud. Luckily, the rapid development of computer technology, the accumulating data, and theemerging data analysis techniques bring new opportunities to financial risk management and analysison the big data in the financial industry.

Researchers have developed various anti-fraud measures and fraud prevention systemsover the years. Leonard [1] proposed a rule-based expert system for fraud detection. The rules of thismodel were manually constructed by the fraud experts from the bank. Sanchez et al. [2] proposed touse association rules to detect fraud and help risk analysts extract more fraud rules. Edge and Sampaio [3] proposed a set of a financial fraud modeling language (FFML) for better describing andcombining fraud rule sets to assist fraud analysis. However, the rule-based models require sufficientand accurate expertise knowledge and cannot be updated timely to new frauds.

To this end, machine learning models have been introduced for fraud detection. Ghosh andReilly [4] uses neural networks to detect credit card fraud.

Kokkinaki [5] proposed decision trees andBoolean logic functions to characterize normal transaction patterns to detect fraudulent transactions.Peng et al. [6] compared nine machine learning models for fraud detection. The results demonstratelinear logistic and IEEE Access and Transaction on Deep Learning,Volume:9,IssueDate:12.January.2021 Bayesian networks are more effective. Lei and Ghorbani [7] proposed a newclustering algorithm namely improved competitive learning network (ICLN) and supervised animproved competitive learning network (SICLN). Sahin et al. [8] designed a decision tree based oncost sensitivity. Halvaiee and Akbari [9] proposed to use an AIRS improved algorithm for fraud

detection. However, these traditional machine learning methods heavily rely on manual subjectiverules and easily lead to model risk. These methods also tend to over due to the imbalance trainingdataset with serious pollution by noises. Thus, ensemble learning methods have also been introducedto integrate different models for complicated fraud detection. Louzada and Ara [10] proposed abagging ensemble model that integrates k-dependence probabilistic networks. The results show thatthe proposed ensemble model has stronger modeling capabilities. Carminati et al. [11] proposed acombination of semi-supervised and unsupervised fraud and anomaly detection methods, mainlyusing a histogram-based outlier score (HBOS) algorithm to model the user's past behavior.

Recently, deep learning techniques have attracted a lot of academic and industrialattention that provides a new insight for financial data analysis. Fu et al. [12] used con volutionalneural networks to effectively reduce feature redundancy. Tu et al. [13] design a deep featurerepresentation technique for fraud detection. To incorporate with prior knowledge with the deepnetwork, Greiner and Wang [14] pointed out the borrower is likely to conceal information that is notbeneficial to him or even fictitious favorable information before obtaining the loan. After obtainingthe loan, the borrower is likely to default unilaterally. Pope and Sydnor [15] also found it difficult tojudge the risk of the personal information provided by the borrower unilaterally because theauthenticity of this information cannot be verified. Freedman and Jin [16] uncovered that theborrower may commit fraudulent behavior by reporting false information, which exacerbates theinformation asymmetry between the two parties. Herzenstein et al. [17] also found that theborrowers repayment ability and credit rating are the factors that have the greatest impact onpersonal credit risk. They concluded that economic strength is the determinant of judging theavailability of borrowing. At the same time, Herzen stein et al. [18] depicted the borrowersspendingpower can also directly affect the success rate of borrowing. These methods reveal the characteristicsof the borrowers would be helpful for fraud detection.

Motivated by such an idea, we propose a deep learning technique to mine the fraudin a public lending dataset with 200,000 records. We analyze the customer credit rating, which canhelp us to identify customers' actual situations. Intuitively, the lower a customer has a credit rating,such as the rating, the greater the likelihood of being a fraudulent user. Internet finance small loancompanies set different thresholds

ISSN: 2057-5688

on their customer credit rating data to build anti-fraud rules basedon the true information of their customers. This paper aims to provide small financial creditcompanies a simple yet effective model to improve their risk control and the level of anti-fraud. Suchcompanies often have a poor-risk control capacity with limited capacity for data engineering,modeling, and optimization. The main contribution of this paper is summarized as follows

First, we analyze the real-world Internet financial data for the missing data andsample imbalance. We propose to all the missing with a random forest and deal with the sampleimbalance with a synthetic minority oversampling technique.

We train a deep neural network by the preprocessed data. We make comprehensibleexperiments for the setting of the network architecture and hyper parameters.Extensive experiments have been conducted to demonstrate the outperformancecomparing with the commonly-used loan fraud detection models. The rest of this paper is organizedas follows. The second part is the methodology, and the third part is an empirical study from real-world data. The fourth part is the conclusion

2. LITERATURE SURVEY

[1] K. J. Leonard, ''The development of a rul based expert system model for fraud alert i consumer credit,'' Eur. J. Oper. Res., vol. 80 no. 2, pp. 350–356, Jan. 1995.

As credit loan products significantly increase i most financial institutions, the number o fraudulent transactions is also growing rapidly Therefore, to manage the financial risk successfully, the financial institutions shoul reinforce the qualifications for a loan an augment the ability to detect a credit loan fraud proactively. In the process of building a classification model to detect credit loan frauds, utility from classification results (i.e., benefits from correct prediction and costs from incorrect prediction) is more important than the accuracy rate of classification. The objective of this paper is to propose a new approach to building a classification model for detecting credit loan fraud based on an individual-level utility. Experimental results show that the model comes up with higher utility than the fraud detection models which do not take into account the individual-level utility concept. Also, it is shown that the individual-level utility computed by the model is more accurate than the mean-level utility computed by other models, in both opportunity utility and cash flow perspectives. We provide diverse views on the experimental results from both perspectives

[2] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, ''Association rules applied to credit card fraud detection,'' Expert Syst. Appl., vol. 36, no. 2, pp. 3630–3640, Mar. 2009.

Association rules are considered to be the best studied models for data mining. In this article, we propose their use in order to extract knowledge so that normal behavior patterns may be obtained in unlawful transactions from transactional credit card databases in order to detect and prevent fraud. The proposed methodology has been applied on data about credit card fraud in some of the most important retail companies in Chile

[3] M. E. Edge and P. R. F. Sampaio, ''The design of FFML: A rule-based policy modelling language for proactive fraud management in financial data streams,'' Expert Syst. Appl., vol. 39, no. 11, pp. 9966–9985, Sep. 2012.

ISSN: 2057-5688

Developing fraud management policies and fraud detection systems is a vital capability fo financial institutions towards minimising the effect of fraud upon customer service delivery bottom line financial losses and the adverse impact on the organisation's brand image reputation. Rapidly changing attacks in real-time financial service platforms continue to demonstrate fraudster's ability to actively re engineer their methods in response to ad hoc security protocol deployments, and highlights the distinct gap between the speed of transaction execution within streaming financial data and corresponding fraud technology frameworks tha safeguard the platform. This paper presents the design of FFML, a rule-based policy modelling language and encompassing architecture fo facilitating the conceptual level expression and implementation of proactive fraud controls within multi-channel financial service platforms It is demonstrated how a domain specific language can be used to abstract the financia platform into a data stream based information model to reduce policy modelling complexity and deployment latencies through an innovative policy mapping language usable by both exper and non-expert users. FFML is part of a comprehensive suite of assistive tools and knowledge-based systems developed to suppor fraud analysts' daily work of designing new high level fraud management policies, mapping into executable code of the underpinning application programming interface and deployment of active monitoring and compliance functionality within the financial platform

[4] S. Ghosh and D. L. Reilly, Credit Card Fraud Detection With a NeuralNetwork Wailea, HI, USA: IEEE, 1994.

Using data from a credit card issuer, a neura network based fraud detection system was trained on a large sample of labelled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20) over rulebased fraud detection procedures. We discuss the performance of the network on this data set in terms of detection accuracy and earliness of fraud detection. The system has been installed on an IBM 3090 at Mellon Bank and is currently in use for fraud detection on that bank's cmlit card portfolio.

[5] A. I. Kokkinaki, ''On atypical database transactions: Identification of probable frauds using machine learning for user profiling,'' in Proc. IEEE Knowl. Data Eng. Exchange Workshop, 1997, pp. 229–238.

The paper proposes a framework for deriving users' profiles of typical behaviour and detecting atypical transactions which may constitute fraudulent events or simply a change in user's behaviour. The anomaly detection problem is presented and previous attempts to address it are discussed. The proposed approach proves that individual user profiles can be constructed and provides an algorithm that derives user profiles and an algorithm to identify atypical transactions. Lower and upper bounds for the number of misclassifications are also provided. An evaluation of this approach is discussed and some issues for further research are outlined

[6] Y. Peng, G. Wang, G. Kou, and Y. Shi, ''An empirical study of classification algorithm evaluation for financial risk p

A wide range of classification methods have been used for the early detection of financial risks in recent years. How to select an adequate classifier (or set of classifiers) for a given dataset is an important task in financial risk prediction. Previous studies indicate that classifiers' performances in financial risk prediction may vary using different performance measures and under different circumstances. The main goal of this paper is to develop a two-step approach to evaluate classification algorithms for financial risk prediction. It constructs a performance score to measure the performance of classification algorithms and introduces three multiple criteria decision making (MCDM) methods (i.e., TOPSIS, PROMETHEE, and VIKOR) to provide a final ranking of classifiers. An empirical study is designed to assess various classification algorithms over seven real-life credit risk and fraud risk datasets from six countries. The results show that linear logistic, Bayesian Network, and ensemble methods are ranked as the top-three classifiers by TOPSIS, PROMETHEE, and VIKOR. In addition, this work discusses the construction of a knowledge-rich financial risk management process to increase the usefulness of classification results in financial risk detection.

3. PROBLEM STATEMENT

Ghosh and Reilly [4] uses neural networks to detect credit card fraud. Kokkinaki [5] proposed decision trees and Boolean logic functions to characterize normal transaction patterns to detect fraudulent transactions. Peng *et al.* [6] compared nine machine learning models for fraud detection. The results demonstrate linear logistic andBayesian networks are more effective.

Lei and Ghorbani [7] proposed a new clustering algorithm namely improved competitive learning network (ICLN) and supervised an improved competitive learning network (SICLN). Sahin*et al.* [8] designed a decision tree based on cost sensitivity. Halvaiee and Akbari [9] proposed to use an AIRS improved algorithm

for fraud detection. However, these traditional machine learning methods heavily rely on manual subjective rules and easily lead to model risk. These methods also tend to over fit due to the imbalance training dataset with serious pollution by noises. Thus, ensemble learning methods have also been introduced to integrate different models for complicated fraud detection.

Louzada and Ara [10] proposed a bagging ensemble model that integrates k-dependence probabilistic networks. The results showthat the proposed ensemble model has stronger modeling capabilities. Carminati *et al.* [11] proposed a combination of semi-supervised and unsupervised fraud and anomaly detection methods, mainly using a histogram-based outlier score (HBOS) algorithm to model the user's past behavior.

3.1 DISADVANTAGES OF EXISTING SYSTEM

The system doesn't analyze for large number of data sets due to lack of ml classifies. The system couldn't implement to detect the following (i) Level of Loan activity, (ii) Level of Loan Prediction, (iii) Loan Profile information

4. PROPOSED SYSTEM

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

This paper aims to provide small financial creditcompanies a simple yet effective model to improve their risk control and the level of anti-fraud. Such companies often have a poor-risk control capacity with limited capacity for data engineering, modeling, and optimization. The main contribution of this paper is summarized as follows.

_ First, we analyze the real-world Internet financial data for the missing data and sample imbalance. We propose to fill the missing with a random forest and deal with the sample imbalance with a synthetic minority oversampling technique.

_ We train a deep neural network by the preprocessed data. We make comprehensible experiments for the setting of the network architecture and hyper parameters.

_ Extensive experiments have been conducted to demonstrate the outperformance comparing with the commonly-used loan fraud detection models

4.1 ADVANTAGES OF PROPOSED SYSTEM

_ **Identity theft:** Criminals steal user's personal financial information in order to conduct fraudulent financial transaction activities or withdraw money from your account.

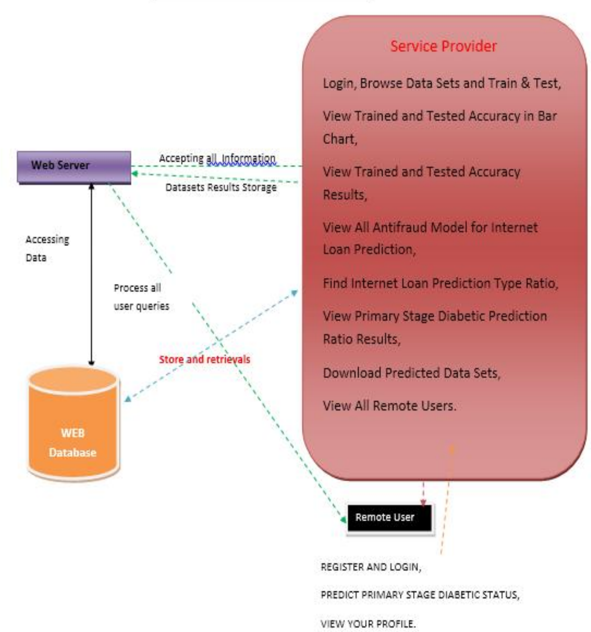
_ **Investment fraud:** Selling investment or securities with false, misleading, or fraudulent information.

_ **Mortgage and loan fraud:** The borrower uses false information to open a mortgage or loan, or the lender uses a high-pressure sales strategy to sell the mortgage or loan or predatory loan to users.

_ **Large-scale marketing fraud:** Criminals usually use a lot of mail, telephone, or spam to steal users' personal financial information or request donations and fees from fraudulent organizations, usually involving fake checks, charities, sweepstakes, lotteries, and exclusive clubs or honor society invites

5. SYSTEM ARCHITECTURE

Architecture Diagram



6. IMPLEMENTATION

6.1 Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download

ISSN: 2057-5688

Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

6.2 Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE

7. INTERNAL MODULES

7.1 Numpy

numPy is a Python library used for working witl arrays. It also has functions for working ii domain of linear algebra, fourier transform, anc matrices. NumPy was created in 2005 by Travi: Oliphant. It is an open source project and yoı can use it freely. NumPy stands for Numerica Python.

NumPy arrays are stored at one continuous placı in memory unlike lists, so processes can acces: and manipulate them very efficiently.

This behavior is called locality of reference iı computer science. This is the main reason why NumPy is faster than lists. Also it is optimized tı work with latest CPU architectures.

7.2 Pandas

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

Pandas allows us to analyze big data and make conclusions based on statistical theories. Pandas can clean messy data sets, and make them readable and relevant. Relevant data is very important in data science.

7.3 Matplotlib

Human minds are more adaptive for the visual representation of data rather than textual data. We can easily understand things when they are visualized. It is better to represent the data through the graph where we can analyze the data more efficiently and make the specific decision according to data analysis. Before learning the matplotlib, we need to understand data visualization and why data visualization is important.

Graphics provides an excellent approach for exploring the data, which is essential for presenting results. Data visualization is a new term. It expresses the idea that involves more than just representing data in the graphical form (instead of using textual form).

This can be very helpful when discovering and getting to know a dataset and can help with classifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts. The static does indeed focus on quantitative description and estimations of data. It provides

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

an important set of tools for gaining a qualitativ understanding.

7.4 Keras

Keras is an open-source high-level Neura Network library, which is written in Python i capable enough to run on Theano, TensorFlow or CNTK. It was developed by one of the Googl engineers, Francois Chollet. It is made user friendly, extensible, and modular for facilitatin faster experimentation with deep neural network It not only supports Convolutional Networks an Recurrent Networks individually but also thei combination.

It cannot handle low-level computations, so i makes use of the **Backend** library to resolve it The backend library act as a high-level AP wrapper for the low-level API, which lets it ru on TensorFlow, CNTK, or Theano.

Initially, it had over 4800 contributors during it launch, which now has gone up to 250,00 developers. It has a 2X growth ever since ever year it has grown. Big companies like Microsoft Google, NVIDIA, and Amazon have activel contributed to the development of Keras. It ha an amazing industry interaction, and it is used i the development of popular firms likes Netflix Uber, Google, Expedia, etc.

Focus on user experience has alway been a major part of Keras. Large adoption in th industry. It is a multi backend and support multi-platform, which helps all the encoder come together for coding. Research communit present for Keras works amazingly with th production community. Easy to grasp al concepts. 1It supports fast prototyping. I seamlessly runs on CPU as well as GPU. I provides the freedom to design any architecture which then later is utilized as an API for th project. 1It is really very simple to get started with. Easy production of models actually makes Keras special.

7.5 Tensorflow

TensorFlow is a software library or framework, designed by the Google team to implement machine learning and deep learning concepts in the easiest manner. It combines the computational algebra of optimization techniques for easy calculation of many mathematical expressions.

Let us now consider the following important features of TensorFlow −

It includes a feature of that defines, optimizes and calculates mathematical expressions easily with the help of multi-dimensional arrays called tensors.It includes a programming support of deep neural networks and machine learning techniques.It includes a high scalable feature of computation with various data sets.TensorFlow uses GPU computing, automating management. It also includes a unique feature of optimization of same memory and the data used.

TensorFlow is well-documented and includes plenty of machine learning libraries. It offers a few important functionalities and methods for the same.

TensorFlow is also called a “Google” product. It includes a variety of machine learning and deep learning algorithms. TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embedding and creation of various sequence models.

7.6 Scikit-learn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine

ISSN: 2057-5688

learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy SciPy and Matplotlib.

Supervised Learning algorithms − Almost all the popular supervised learning algorithms, like Linear Regression, Support Vector Machine (SVM), Decision Tree etc., are the part of scikit learn.

Unsupervised Learning algorithms − On the other hand, it also has all the popular unsupervised learning algorithms from clustering factor analysis, PCA (Principal Component Analysis) to unsupervised neural networks.

Clustering − This model is used for grouping unlabeled data.

Cross Validation − It is used to check the accuracy of supervised models on unseen data.

Dimensionality Reduction − It is used for reducing the number of attributes in data which can be further used for summarisation visualisation and feature selection.

Ensemble methods − As name suggest, it is used for combining the predictions of multiple supervised models.

Feature extraction − It is used to extract the features from data to define the attributes in image and text data.

Feature selection − It is used to identify useful attributes to create supervised models.

Open Source − It is open source library and also commercially usable under BSD license.

8. ALGORITHMS USED

8.1 Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

8.2 K-Nearest Neighbors (KNN)

Simple, but a very powerful classification algorithm. Classifies based on a similarity measure. Non-parametric. Lazy learning. Does not “learn” until the test example is given. Whenever we have a new data to classify, we find its K-nearest neighbors from the training data.

Example

Training dataset consists of k-closest examples in feature space. Feature space means, space with categorization variables (non-metric variables). Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

8.3 Logistic regression Classifiers

ISSN: 2057-5688

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

8.4 Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin KamHo[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochasticdiscrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

8.5 SVM

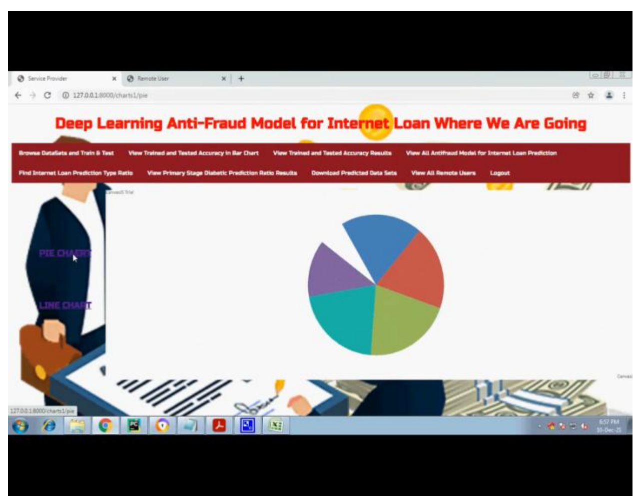
In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels fornewly acquired

ISSN: 2057-5688

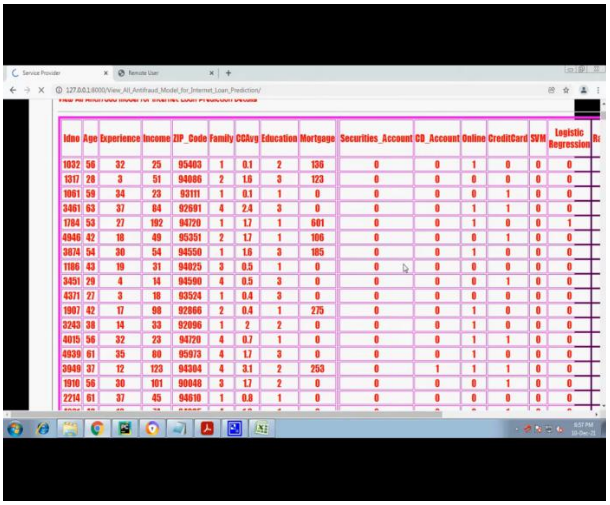
instances. Unlike generative machine learning approaches, which require computations ofconditional probability distributions, a discriminant classification function takes a data point x and assignsit to one of the different classes that are a part of the classification task. Less powerful than generativeapproaches, which are mostly used when prediction involves outlier detection, discriminant approachesrequire fewer computational resources and less training data, especially for a multidimensional featurespace and when only posterior probabilities are needed. From a geometric perspective, learning a classifieris equivalent to finding the equation for a multidimensional surface that best separates the different classesin the feature space.

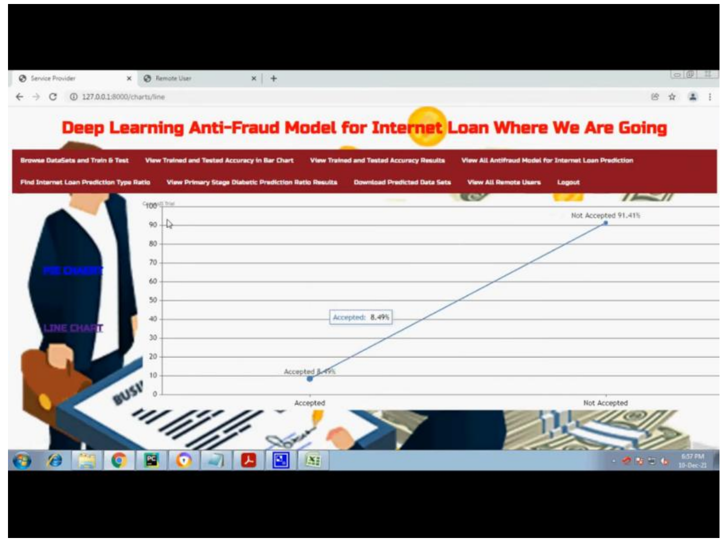
SVM is a discriminanttechnique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

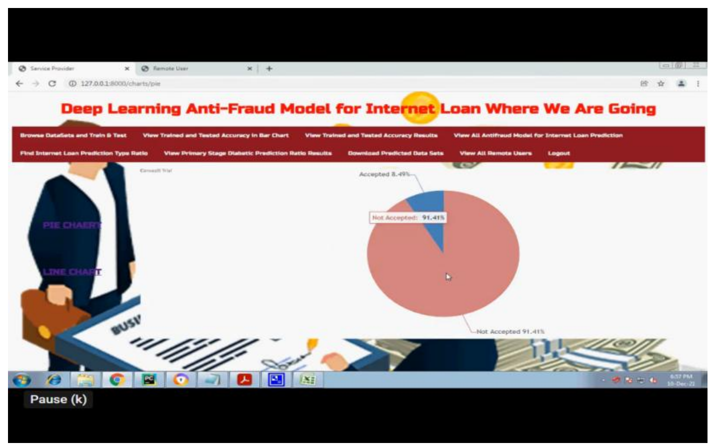
9. OUTPUT RESULTS



ISSN: 2057-5688

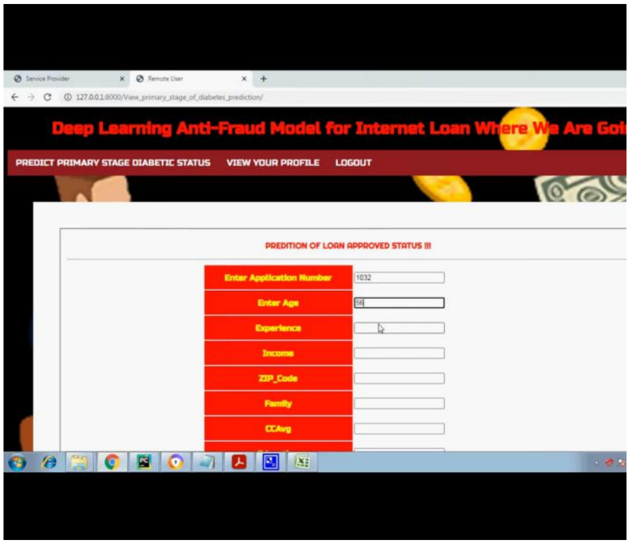






INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688



10. CONCLUSION

In this paper, we take the real customer information of the public loan data set of the lending club company as a sample. Then, we build a deep learning based Internet fraud detection model. We introduce the main parameters of the model and optimizes to find the optimal parameter combination of the model. Finally, the most popular logistic regression in the financial industry as well as other comparisons are used as a baseline to evaluate the performance of the proposed model. The results reveal the deep neural network achieves better Performance, which is promising to be used in the financial industry for Internet fraud detection. In the future, we plan to cooperate with mature Internet financial technology companies and banks in China for blacklists and white lists. The deep neural network combined with such blacklists and white lists and the expertise anti fraud rules is promising to increase fraud detection capability.

11. FUTURE SCOPE

In enhancement we will add some ML Algorithms to increase accuracy

12. REFERENCE

[1] K. J. Leonard, ``The development of a rule based expert system model for fraud alert in consumer credit," *Eur. J. Oper. Res.*, vol. 80, no. 2, pp. 350_356, Jan. 1995.
[2] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, ``Association rules applied to credit card fraud detection," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3630_3640, Mar. 2009.
[3] M. E. Edge and P. R. F. Sampaio, ``The design of FFML: A rule-based policy modelling language for proactive fraud management in _nancial data streams," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 9966_9985, Sep. 2012.
[4] S. Ghosh and D. L. Reilly, *Credit Card Fraud Detection With a Neural- Network.* Wailea, HI, USA: IEEE, 1994.
[5] A. I. Kokkinaki, ``On atypical database transactions: Identi_cation of probable frauds using machine learning for user pro_ling," in *Proc. IEEE Knowl. Data Eng. Exchange Workshop*, 997, pp. 229_238.
[6] Y. Peng, G.Wang, G.Kou, andY. Shi, ``An empirical study of classi_cationalgorithm evaluation for _nancial risk prediction," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2906_2915, Mar. 2011.
[7] J. Z. Lei and A. A. Ghorbani, ``Improved competitive learning neural networks for network intrusion and fraud detection," *Neurocomputing*, vol. 75, no. 1, pp. 135_145, Jan. 2012.
[8] Y. Sahin, S. Bulkan, and E. Duman, ``A cost-sensitive decision tree approach for fraud detection," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5916_5923, Nov. 2013.
[9] N. SoltaniHalvaiee and M. K. Akbari, ``A novel model for credit card fraud

ISSN: 2057-5688

detection using arti_cial immune systems,'' *Appl. Soft Comput.*, vol. 24, pp. 40_49, Nov. 2014.

[10] F. Louzada and A. Ara, ``Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool,'' *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11583_11592, Oct. 2012.

[11] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, ``BankSealer: A decision support system for online banking fraud analysis and investigation,'' *Comput. Secur.*, vol. 53, pp. 175_186, Sep. 2015.

[12] K. Fu, D. Cheng, Y. Tu, and L. Zhang, ``Credit card fraud detection using convolutional neural networks,'' in *Proc. Int. Conf. Neural Inf. Process.*, Oct. 2016, pp. 483_490.

[13] B. Tu, D. He, Y. Shang, C. Zhou, and W. Li, ``Deep feature representation for anti-fraud system,'' *J. Vis. Commun. Image Represent.*, vol. 59, pp. 253_256, Feb. 2019.

[14] M. E. Greiner and H. Wang, ``Building consumer-to-consumer trust in EFinance marketplaces: An empirical analysis,'' *Int. J. Electron. Commerce*, vol. 15, no. 2, pp. 105_136, Dec. 2010.

[15] D. G. Pope and J. R. Sydnor, ``What's in a picture? Evidence of discrimination from prosper.com,'' *J. Hum. Resour.*, vol. 46, no. 1, pp. 53_92, 2011.

[16] S. Freedman and G. Z. Jin, ``The information value of online social networks: Lessons from peer-to-peer lending,'' *Int. J. Ind. Org.*, vol. 51, pp. 185_222, Mar. 2017.

[17] M. Herzenstein, U. M. Dholakia, and R. L. Andrews, ``Strategic herding Behavior in peer-to-peer loan auctions,'' *J. Interact. Marketing*, vol. 25, no. 1, pp. 27_36, Feb. 2011.

[18] M. Herzenstein, R. L. Andrews, and U. M. Dholakia, ``The democratization of personal consumerloans? Determinants of successin online peer-topeer lending communities,'' *J. Marketing Res.*, vol. 15, pp. 274_277, 2008.

[19] N. Mykhalchenko and J. Wiegratz, ``Anti-fraud measures in Southern Africa,'' *Rev. Afr. Political Economy*, vol. 46, no. 161, pp. 496_514, Jul. 2019.

[20] B. Riley, ``Anti-fraud technologies: A business essential in the card industry,'' *Card Technol. Today*, vol. 19, no. 10, pp. 10_11, Oct. 2007.

[21] L. Breiman, ``Random forests,'' *Mach. Learn.*, vol. 45, no. 1, pp. 5_32, 2001.

[22] T. Chen and C. Guestrin, ``XGBoost: A scalable tree boosting system,'' in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785_794.

[23] T. Pan, J. Zhao,W.Wu, and J. Yang, ``Learning imbalanced datasets based on SMOTE and Gaussian distribution,'' *Inf. Sci.*, vol. 512, pp. 1214_1233, Feb. 2020.

[24] E. Micheu-Tzanakou, ``Arti_cial neural networks: An overview,'' *Netw.*, vol. 22, nos. 1_4, pp, pp. 208_230, 2011.

[25] T. Zhou, G. Han, X. Xu, Z. Lin, C. Han, Y. Huang, and J. Qin, ``_-agree AdaBoost stacked autoencoder for short-term traf_c _ow forecasting,'' *Neurocomputing*, vol. 247, pp. 31_38, Jul. 2017.

ISSN: 2057-5688