IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

Probabilistic Support Vector Regression Classification Model for Credit Card Fraud Detection

Nikita Sawhney[1*], B. Kaur[2], H. Kaur[3]

[1]Dept. of Information Technology, Chandigarh Engineering College, Mohali, India
[2]Dept. of Information Technology, Chandigarh Engineering College, Mohali, India
[3]SGGS collegiate public school, Mohali, India

Available online at: www.ijcseonline.org

Accepted: 24/Jun/2018, Published: 30/Sept./2018

Abstract—The researchers have already worked with many supervised and unsupervised methods for the purpose of credit card fraud detection. The supervised models have been found more efficient for the purpose of credit card fraud detection. The major goal of the credit card fraud detection research is to improve the accuracy while decreasing the elapsed time. The proposed credit card fraud detection models purposes the use of feature extraction and selection of the credit card data with linear regression algorithm for the credit card fraud detection. The feature engineering and analysis would be performed over the given transactional data and then final classification of the anomalies or outliers is done using linear regression classifier. The proposed model has been tested under the various experiments from the various groups of test cases. The test case groups have been obtained after applying the various levels of the feature elimination and feature selection over the collection of credit card transaction data. The proposed credit card fraud classification model is based upon two different models, which includes the Naïve Bayes and Support Vector Regression. The main aim of the research is to achieve the higher credit card fraud recognition accuracy, with the minimum classification complexity.

Keywords— Credit card fraud detection, Early Fraud Detection, Regressive analysis, Linear regression models.

I. INTRODUCTION

The credit cards are one of the popular payment transaction methods now-a-days with worldwide popularity. The credit card is a kind of debt issued to the individual to increase their purchasing power, which allows them to buy the things, they can't exactly afford with their income. There are several cases of credit card frauds come to the fore every year causing the loss of several millions to billions of dollars to the national exchequers. The historical credit card user data is used to determine the frauds on the basis of spending pattern based case matching. A number of data mining techniques are employed individually or in variety of combinations in order to determine the credit card fraud cases.

Data mining and outlier discovery is the process to find the valuable information and related stuff from the messages, threads, discussion forums and other sources attached to the data source, which may be used to classify the abnormalities in the database. The descriptive and predictive analytics is the branch of data science specialization, where the data is analyzed by using the various data processing methods. The data mining methods require the 'High Quality' combinations for the various techniques altogether for the discovery of data using connections between the data rows and phrases, estimating their novelty and dynamic methodology. [1] The statistical methods based upon the feature description, dimensionality reduction and pattern discovery, which has been described in the various computing methods, which involves the classification methods such as artificial neural networks (ANN), support vector machine (SVM), Naïve Bayes (NB), Random Forest (RF), Co-Forest (CF), Simple Regression, Linear Regression, Logistic Regression, etc. [3] Outlier analytical methods have been utilized in the versatile applications, ranging from the data retrieval to the data processing applications. The data mining application requires the multiple steps to be executed in the particular arrangement, which is shown in the following steps:

1. **Data Retrieval** systems establish the documents in a very assortment that match a user's question. The foremost acknowledge IR systems are search engines like google, that establish those documents on the globe wide net that are relevant to a collection of given entrie.

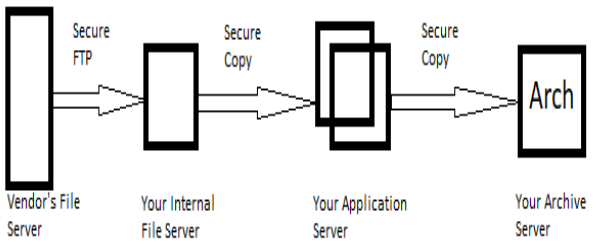

Figure 1.1: Data Retrieval Process from the target servers [5]

2. **Data Methoding** (DM) is that the process of characteristic patterns in massive sets of knowledge. The aim is to uncover antecedently unknown, helpful information. Once employed in pattern discovery, DM is applied to the facts generated by the data extraction section and places the results of our DM method into another information which will be queried by the end-user via an acceptable graphical interface. The info generated by such queries may be delineated visually.

3. **Data Extraction** is that the method of mechanically getting structured knowledge from an unstructured language document.

II. LITERATURE SURVEY

Li, Jinyan et. al. [1] have proposed the hierarchical classification in pattern discovery for outlier analysis of credit card database. In this paper, the authors have evaluated several popular classification algorithms, along with three filtering schemes.

Prollochs, Nicolas et. al. [2] has worked on the enhancement of outlier analysis of financial news by detecting negation scopes.To predict the corresponding negation scope, related literature commonly utilizes two approaches, namely, rule-based algorithms and machine learning.

Cui, Limeng et. al. [3] has developed a hierarchy method based on lda and svm for credit card fraud detection. In this paper the authors have focused on the data classification, which is meaningful for information provider to organize and display the data rows but also for the users to reach the valuable information easily.

Ouyang, Yuanxin et. al. [4] has proposed the outlier classification with support from auxiliary large databases. In this paper, the authors have targetted on the problem of the data classification which is an essential and typical membership and propose an approach which employs external information from larger databases to address the problem the sparseness.

Ouyang, Yuanxin, Yao Huangfu, Hao Sheng, and Zhang Xiong [9] proposed the Outlier Classification with Support from larger sized datasets. Denial I.Morariu, Lucian N.
Vintan, and Volker Tresp [10] Investigated three approaches to build an efficient meta-classifier. In this select 8 different SVM Classifiers. For each of the classifier modified the kernel, the degree of the kernel and input data representation based on the selected classifier calculate the upper limit of the proposed meta- classifier that is 94.21 %.

Hyeran Byun 1 and Seong-Whan Lee2 [11] present brief introduction is presented on SVMs and several applications of SVMs in pattern recognition problems.

D. Morariu, R. Cre,Tulescu and L.,Vin,tan [12] says that building up on the meta- classifier presented based on 8 SVM components, we add to these a navie bayes based classifier which leads to a significant improvement of the upper that the meta classifier can reach.

A. Literature Table

Index	Authors	Problem Addressed	Technologies Used	Classification Model
1	D. Morariu et. al. [12]	Meta-classifier for dynamic data categorization	Multi-component and multvaraiate classification	Naïve Bayes, SVM
2	Hyeran Byun et. a. [11]	Pattern Recognition	Image text and color features.	SVM
3	Vintan et. al. [10].	Built multiple models for meta-classification among multi category data	Applied support vectors with RBF, Linear, Polynomial and other kernels	SVM
4	Ouyang, Yuanxin et. al. [4]	Membership based data classification	**Sparse signatures in the large datasets**	SVM,KNN

III. EXPERIMENTAL DESIGN

Credit card fraud detection database is the source to distribute the information about the financial transactions completed from the credit cards to buy certain items from the offline or online markets. The financial transactional control patterns plays thevital role in distributing such eventful and outlier data in the given database. Data provide useful information about the daily events, which covers the purchasing behaviour, outliers, medium, high and low spending possibilities, etc. In the proposed model, the combination of the linear regression classifier along with outlier analysis model has been implemented in order to improvise the overall performance of the proposed model. Data classification system involves various steps like feature extraction, feature selection, feature

vector generation, training and classification. The flow chart in figure 4.2 demonstrates all the phases system goes through. **Data preparation phase:** The data preparation phase involves the gathering the credit card transaction dataset from available sources. The actual design of the system starts with gathering data for experimentation. The collected transaction sets are then transformed into an appropriate format suitable for system design,

Preprocessing phase: Preprocessing involves following steps:

Read data to usable vector:- Reading the credit card transactional data into selective feature vector is the very first step involved in data classification based credit card transactional database. The size of feature vector is proportional to size of data set. Each cell of string vector stores the corresponding feature value of database.

Feature extraction: The selective features are further extracted from the string vector, which primarily contains the clean and fraud transactions in our case. These features are the major categories for the fraud classification models.

Training Model:

Algorithm 1: Probabilistic Support Vector regression Algorithm

Read the source credit card transactional data, and Extract the features from the credit card transactional data source. Feature descriptor will be the set of selective features, and will describe smaller details than the original dataset.

1. Apply the preprocessing methods over the input data to fill the missing values, treat the outliers and create a smooth and validated training data
2. Create the group vector mentioning all of the rows in the training data
3. Split the data in testing and training matrices
4. Build the support vector machine (SVM) classifier model by making a call to the library
5. Configure the classifier with configuration parameters including learning rate & support vector ratio
6. Train the classifier with training data and labels in the group vector
7. Test the classifier with testing data and observe the predictions
8. Evaluate the predictions and original labels of the testing data to calculate the performance of classification model

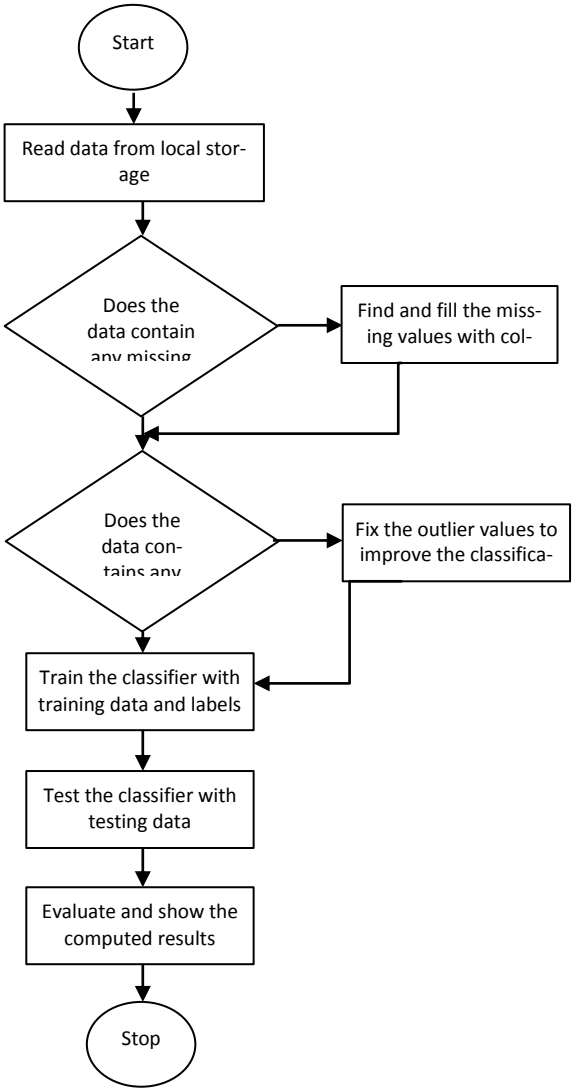

Figure 1: Flow diagram of the supervised classification model for credit card fraud evaluation

Testing Model:

- **Classification**:- Feature vector of training dataset is fed to learnt model to assign a class label to given data samples in the form of a unclassified dataset.

IV. CONCLUSION

The proposed model has been entirely based upon the guided feature extraction using the smart feature selection method, which utilizes the frequency based outlier analysis model along with the linear regression classifier. The proposed model has been designed to work over the multiple types of credit card transactional entries. The proposed model has been tested with the locally stored credit card fraud detection database. A number of experiments have been conducted over the proposed model by using the various forms of the input data gen-

erated after various levels of pre-processing. The proposed model has been designed by using the outlier analysis with linear regression model to predict the category of the clean or fraud transaction. The further improvement can be further improved by using the convolution neural network, deep learning and other related classification methods for the automatic fraud detection.

REFERENCES

[1]. Li, Jinyan, Simon Fong, Yan Zhuang, and Richard Khoury. "Hierarchical classification in pattern discovery for outlier analysis of credit card database." Soft Computing (2015): 1-10.

[2]. Prollochs, Nicolas, Stefan Feuerriegel, and Dirk Neumann. "Enhancing Outlier analysis of Financial News by Detecting Negation Scopes." InSystem Sciences (HICSS), 2015 48th Hawaii International Conference on, pp. 959-968. IEEE, 2015.

[3]. Cui, Limeng, Fan Meng, Yong Shi, Minqiang Li, and An Liu. "A Hierarchy Method Based on LDA and SVM for Credit card fraud detection." In Data Mining Workshop (ICDMW), 2014 IEEE International Conference on, pp. 60-64. IEEE, 2014.

[4]. Ouyang, Yuanxin, Yao Huangfu, Hao Sheng, and Zhang Xiong. " target on the problem of news title classification which is an essential and typical member in short text family." In Neural Information Processing, pp. 581-588. Springer International Publishing, 2014.

[5]. Krishnlal G, S Babu Rengarajan, K G Srinivasagan, " A new pattern discovery approach based on HMM-SVM for web credit card fraud detection" International Journal of Computer Applications (0975-8887) Volumn 1- No.19,2010.

[6]. Vandana Korde, C namrata Mahender, "Text classification and classifier a survey," International Journal of Artificial Intelligence and Application (IJAIA), vol.3, No.2, March2012.

[7]. Mita K. Dalal, Mukesh A.Zaveri," Automatic text Classification," International Journal of Computer Applications (0975-8887) Volumn 28- No.2, August 2011.

[8]. Denial I.Morariu, Lucian N. Vintan, and Volker Tresp, "Meta-Clalssification using SVM classifiers for text documents ", "World Academy of science engineering and technology" 21,2006.

[9]. Hyeran Byun 1 and Seong-Whan Lee2, " Application of Support Vector machines for pattern recognition: A Survey," SVM 2002, LNCS 2388,pp.213-236,2002.

[10]. D. Morariu, R. Cre,Tulescu and L.,Vin,tan, " improving the SVM Meta Clssifier for text document by using Naïve bayes,"Int. J. of Computers, communication and control, ISSN 1841-9844.

[11]. Lie Lu, Stan Z. Li and Hong –Jiang Zhang, " Content based Audion Segmentation using Support vector machine."

[12]. CREŢULESCU, Radu George, and N. VINŢAN Lucian. "Contributions to Document Classification System Design." Vol. 1, Issue 1, SIBIU, 2011.

[13]. Hyeran Byun 1 and Seong-Whan Lee2, " Application of Support Vector machines for pattern recognition: A Survey," SVM 2002, LNCS 2388,pp.213-236,2002.

[14]. Krishnlal G, S Babu Rengarajan, K G Srinivasagan, " A new pattern discovery approach based on HMM-SVM for web credit card fraud detection" International Journal of Computer Applications (0975-8887) Volumn 1- No.19,2010.

[15]. Zhang, Yulei, Yan Dang, Hsinchun Chen, Mark Thurmond, and Cathy Larson. "Automatic credit card database monitoring and classification for syndromic surveillance." Decision Support Systems 47, no. 4 (2009): 508-517.

Authors Profile

Nikita Sawhney received the B.TECH degree in information technology from Rayat Bahra University mohali, Punjab, in 2015. She is currently pursuing the MTECH in information technology with the Chandigarh engineering college mohali, Punjab. Her research interests include machine learning.

Dr.Bikram Pal Kaur is an Professor in the Deptt. of Computer Science & Engineeringand was also HOD of Deptt. Of Computer Application in Chandigarh Engineering College, Landran,Mohali. She has contributed more than 28 articles in various national/ international conferences and 43 papers in research Journals. Her areas of interest are Information System.

Harliv Haur is a Research Student at Sri Guru Gobind Singh Collegiate Public School, Chandigarh.. She has already published three research papers in various reputed journals. Her keen areas are of interest are IOT, Machine Learning and ANN.

