An Efficient Machine Learning Model for Location Aware Credit Fraud and Risk Classification and Detection

V. Muthulakshmi[1],C. Saravanakumar[2] and A. Tamizhselvi[3]
{ hoditlabaffairs@stjosephs.ac.in[1], mailofcsk@gmail.com[2], atamizhselvi@gmail.com[3]}

Associate Professor, Department of Information Technology
St. Joseph's College of Engineering, OMR, Chennai 600 119, India[1]
Associate Professor, Department of Information Technology,
St.Joseph's Institute of Technology, OMR, Chennai - 600119, India[2]
Associate Professor, Department of Information Technology,
St. Joseph's College of Engineering, OMR, Chennai 600 119, India[3]

Abstract. Online transaction grows in enormous rate because of the strength of usage by the user. User always use online mode to pay the amount to the respective merchant. Various method of payment is available in the market but credit card is so popular due to the pre credit is assigned to the customer by banker. Card user gets extra time for paying the payment which gives comfortable live to them. Security of the card suffers in various factors such as theft, fraud, illegal access, so it is protected by using modern algorithm with automated capability. Artificial Intelligent algorithms are applied to detect the fraud but that is not achieving enough accuracy. This type of problem is overcome by using location based risk identification model with multidimensional features for analysis. Three phases of processing is carried out namely feature management, risk management and Location awareness. The focus of the model is to protect the credit card frauds in multi level security by identifying the source and location of access. It achieves high level of security when compared to all exiting algorithms with reliable manner.

Keywords: Machine Learning, SVM, Deep Learning, Fraud Detection, Risk Analysis

1 Introduction

Nowadays a credit card usage grows in large because of the popularity among the humans life. Credit card is like a wallet which holds the money for using the purchase of any goods from the marker by user. This is provided by any bank to the customer and offers the usage of card. The user has to pay the amount what he can purchase along with additional charge with some time duration. The card user gets some rewards point for his purchase and also get loan amount for specific purpose. Cards are classified into various types namely business cards, secure cards, prepaid cards and digital based cards. This card eliminates the excess holding of money in hand and achieves cashless mode for payment. It provides the risk free transaction of buying any product. The customers are suffers credit card because of fraud occurs in the payment process. Some users get the card and do for illegal transaction which leads the fraud. This is identified and classified using machine learning based approach with maximum trust

ICASISET 2020, May 16-17, Chennai, India

DOI 10.4108/eai.16-5-2020.2304200

[1]. The hidden patterns are explored by using big data analytics methods for making business decision. These types of methods are consider to detect the fraud occurs in the credit card usage with the help of latest tools. Real time transactions are suffers a security problem. The detection process is also carried out by implementing classification algorithms [2]. Online payment methods are so popular in recent days and also increase the fraud propositionally. Artificial Intelligent type methods are used to detect the fraud automatically with high level of accuracy [3]. Most of the online frauds are occur due to the credit card sources. The information is stolen by the others for making their own usage and purpose. There is problem occur for finding illegal transaction from the large collection of transactions [4]. Electronic commerce application suffers the problem of online purchase such as theft of user identity, money loss illegal usage of the card and information. It leads the heavy crime in the human life and also affects the financial condition of the human [5]. Traditional fraud detection algorithms are not suitable detection completer fraud so it will be handled using automated method [6]. Card user treated as a defaulter due to the problem in the payment on time. This problem affects the user for getting loan and other kind of benefits from the users. The bankers block the account to this type of card holder [7]. There are no generic solution is available for detecting the credit fraud detection. Two types of problem exist the detection process namely balanced and unbalanced. Unbalanced type is difficult to detect so there is need for intelligent solution [8]. Security threats are high during the regular purchase through online mode. This problem is solved by using efficient technique with modern tools with innovative feature support [9]. Credit card fraud identification requires an efficient algorithm to minimize the loss in the customer and achieves high accuracy [10]. Various machine learning methods and related algorithms support the bank and customer for identifying and preventing from the fraud and achieve maximum reliability. The aim of the proposed work is to analyze the credit rate of the user and recommend for further bank process. The location of the fraud is also considered for imposing high level of security.

2 Proposed Fraud Analysis Model

Transaction and card information are collected from various users by the banker who identify and analyze whether user is genuine or not. The data are preprocessed from the collected raw data by removing useless information, generating relevant data and perform various operations. The features are selected from the cleaned data and stored into the database for subsequent analysis. Data are classified with various categories such as default user, moderate and genuine user based on the credit score of the particular user. Default user always like fraud who unable to the pay the amount on the particular time. The other categories are good for card usage. The risk identification phase analyzes the user with their current usage of the card which is maintained in the database. Various types of risk are available in the card management process, so it will be handled proper level and eliminate completely in order to achieve high retention and reliability. The classified parameters are divided as training and testing dataset which is suitable for further model evaluation process. The model is prepared for further prediction process and identify the loss occurs by the bankers due to fraud action performed by the unauthorized persons. The classification report is generated for accurate analysis. Figure 1 show that the flow model of the proposed algorithm. Work flow is modeled in figure 2. Algorithm 1 presents the identification of defaulter of the card user.

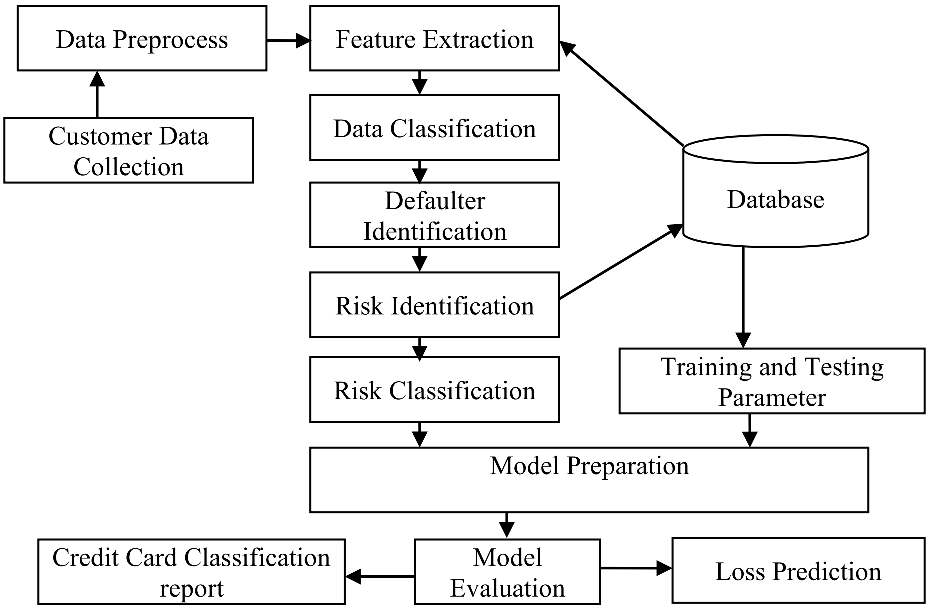

Fig. 1. Flow of the proposed model

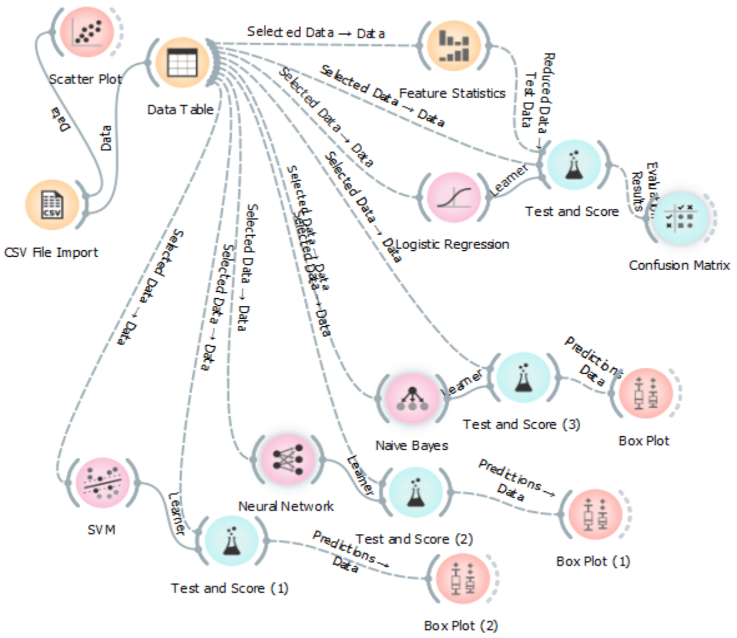

Fig. 2. Work flow of machine learning model

Algorithm 1 Defaulter_Identification()

```
Begin
Let Customer as CS;
Let data as D;
Let features as F;
For each customer Ɛ CS do
  Begin
Collect the credit card data;
Identify the irrelevant data;
Remote those data;
Generation of data using interpolation method;
Save the data into data store DS;
Classify the data based on the credit score;
If user Ɛ U and Category == " Defaulter" then
Set risk as high level;
Else if user Ɛ U and Category == " Ok" then
Set risk as Moderate level;
Else
No risk is set to the customers;
  End
Risk_Identification(Category);
End
```

3 Outer Analysis

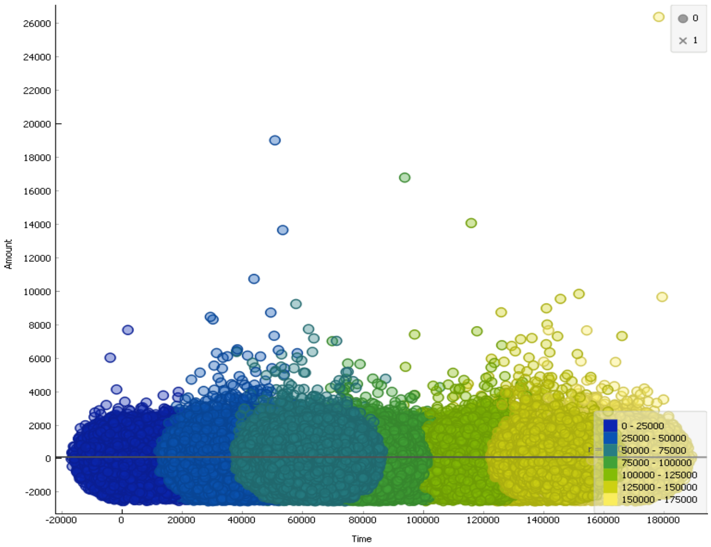

Fig. 3. Data set distributions

The insider attack in the network uses some new path and channel to access the company's data. The anomalies are detected by using multi stage method. Outlier factor is assessed using the indicators with mathematical models [11]. Various algorithms related to machine learning are not fulfilling the constraints in all level of activity. The dataset are goes to less than one percent leads the unbalance condition during the training of the model. The over fitting and under fitting problem is overcome by using required methods and evaluate the model very efficient level [12]. Real-time data are multi dimensional in nature which adds the complexity to the model goes to least accuracy. This problem is overcome by using the reduced dimensions. The relevant features are selected and removed irrelevant features called feature pruning [13]. Inventory management needs the internal components and technologies for maintaining the data in more reliable way. If the outlier's appearance is high then the performance degrades until removed from the data set. It determines the position of the components in correct place without any struggle [14]. Integrated version of machine learning and design approach which targets the regression problems. It also solves the problems like noise removal, parameter mismatch in the model, manual errors and so on [15]. Figure 3 and figure 4 represents the distribution of data and analysis of the outliers respectively.

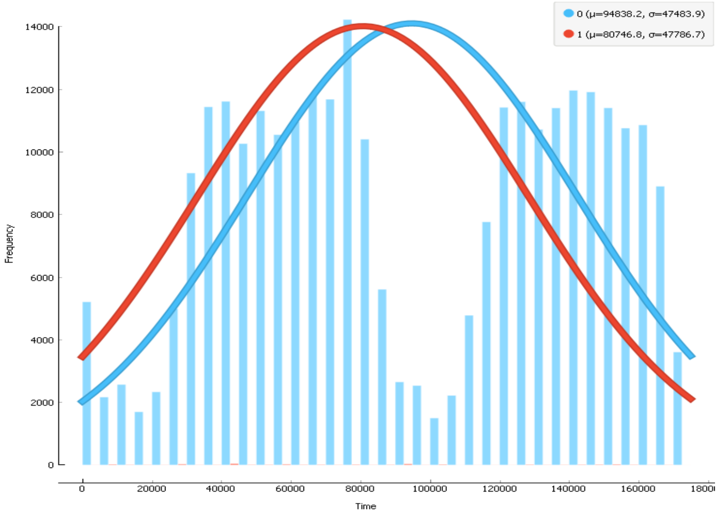

Fig. 4. Outlier Detection

4 Risk Identification and Classification

Statistical method of learning is used for identifying the risk which is present in the credit evaluation process. The aim is to achieve high precision with SVM approach and also uses the indexes as a parameter [16]. First and foremost step of any business is risk analysis because it leads the entire industry goes to loss stage. Hierarchy based approach are used with various factors and also maintaining the risk table in order to reduce in some level [17]. Risk management of the software application suffers in the debt occur due to various factors. The loss reduction and risk detection is a severe challenge in the corresponding industry. Automated model with necessary parameters are used for getting an accurate result [18].

Software lifecycle gets changes in negative side due to uncertainty present in the environment. RAF is a risk assessment framework which assigned the risk process in some priority manner [19]. Empirical model provides only the analysis of the risk but practical and real evaluation methods needed for reducing the risk. There are various AI problems are present so it will be selected based on the application and problem area [20]. Credit card risks are classified with the relevant category by maintaining various levels depending upon the users' capability. It decides the user behavior related to the genuine level while deals with card. Algorithm 2 provides the steps for risk categorization of the fraud assessment.

```
Algorithm 2 Risk_Classification(Category as RC)
Begin
If risk Ɛ RC then
Don't consider the customer for credit card;
Add to the defaulter list;
Else if risk Ɛ RC then
Identify the past record with time duration;
Calculate the duration of credit score S;
End
If score Ɛ S and duration Ɛ D == long then
Consider for credit score;
Else
Customer get new credit card and set genuine;
End
End
```

Location of the card usage is very important because the fraud is happen by un-trusted person in unidentified location. It will be handled by using the device ID based approach which is added with the tag. The tags are created in two ways namely mobile based or host based. Mobile based approach adds the tag with OTP for verification purpose before completing the transaction. Host based approach the corresponding IP address is added with the tag which identifies the location. The proposed algorithm is location aware model in order to track the card usage in very rapid manner. Location Awareness and identification is presented in Algorithm3. Figure 5 gives the frequency analysis.

```
 Algorithm Location_Identication()
 Begin
 Let customer as C;
 Let devices as D;
 For each user Ɛ C do
 Identify the source of transaction TD;
 If TD == "Mobile" then
 Identify the location and add it to the tag;
 Verify the OTP with automated Tag;
 Else
 Identify the location based on the IP address and add it to
the tag;
 Verify the OTP with IP based automated tag;
 End
End
```

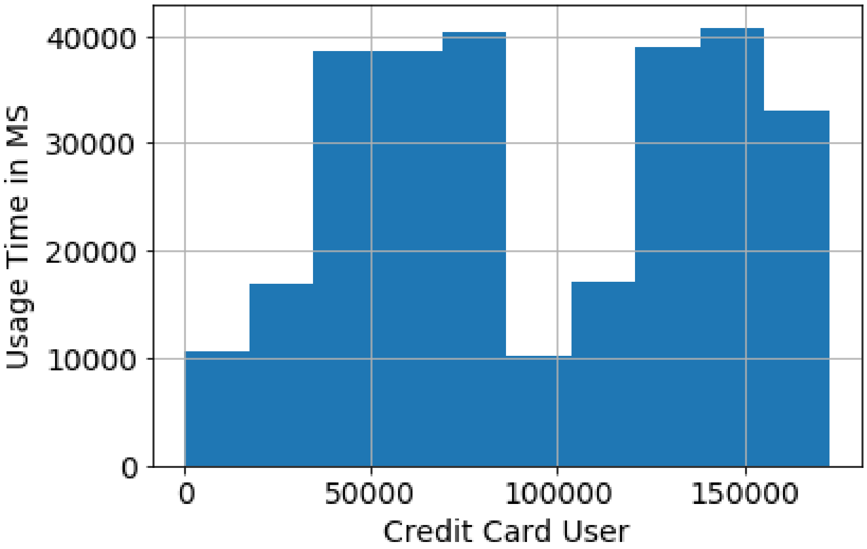

Fig. 5. Histogram analysis

5 Results and Discussion

Data science domain uses various kinds of attributes which applies for prediction and classification problems. Logistic regression model suffers to handle the raw data because it only performs categorical data. It is overcome by using estimator element with compression [21]. Classification of multimedia data needs preprocessing phase which performs complex task to identify the classes. Navie bayes algorithm provides the chances of occurrences of the particular data related to problem area, but it requires other algorithm support to improve accuracy [22]. Credit score are calculated with respective models performs the operation in two ways online and offline. Traditional schemes uses offline mode, so it is not update the score instantly. Neural network model eliminates this issue by performing both the type of operation with exact features [23]. SVM model handles transaction data with sparse type for detecting the fraudulent card usage.

Table 1. Classification Report

Algorithm	Precision	Recall	F1-Score	Support
Logistic Regression Model	1	0.76	0.31	0.44
Navie Bayes Model	0.99	0.99	0.12	0.34
SVM Model	0.98	0.85	0.42	0.39
Neural Networks Model	0.96	0.99	0.52	0.47
Proposed Model	1	1	0.65	0.75

It is a non-linear model follows grid model which combines more number of features in an feasible manner [24]. Table 1 represents the classification summary of the algorithms. Figure 6 shows that various algorithms comparison over proposed algorithm. Proposed model achieves improved result when compared to existing algorithm with high security with reliability.

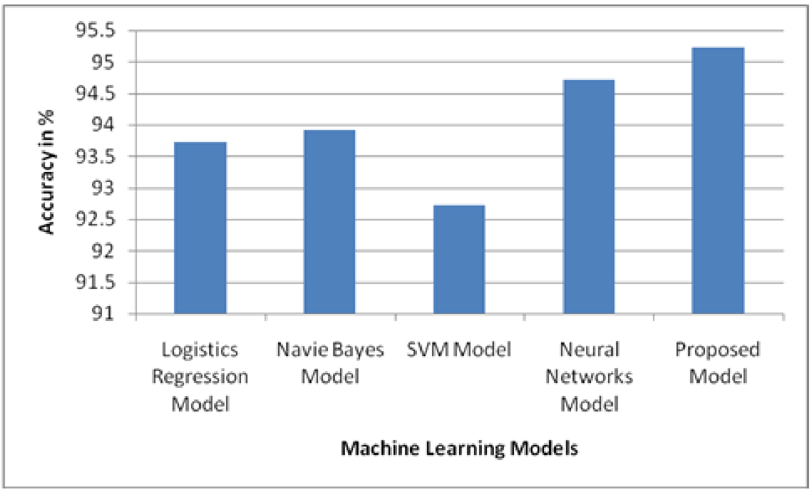

Fig. 6. Comparison of ML algorithms

6 Conclusion

Internet users are goes in high rate because everything is available from various sources. Traditionally the user goes to market for buying all kind of products with liquid cash is carried by them. It leads the security problem which handling large amount of cash. This problem is overcome by using the online buy of all products through digital currency. Credit card is a one type which is used for buy any protect in customer location. Credit card also suffers a problem of fraud happens by the third party. These problems are overcome by using integrated multilevel Machine learning algorithm with reduced risk. Customers have been classified based on their credit rating and corresponding score. Risks are identified from the classified users and identify them in order to achieved high reliability. Location based detection has been implemented for identifying the location of the card usage with multilevel authentication and verification method. This model achieves high accuracy with multidimensional parameter. This model can be applied to all kind of payment mode used for online transaction.

References

[1] SamidhaKhatri; AishwaryaArora; ArunPrakashAgrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison",10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, IEEE, Pp. 680-683.

[2] SahilDhankhad; Emad Mohammed; BehrouzFar, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study", International Conference on Information Reuse and Integration (IRI), IEEE, 2018, Pp.122-125.

[3] Vinod Jain; MayankAgrawal; AnujKumar,"Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection", 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, 2020, DOI: 10.1109/ICRITO48877.2020.9197762, Pp. 86-88.

[4] Sangeeta Mittal; ShivaniTyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection", 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2019, DOI: 10.1109/CONFLUENCE.2019.8776925, Pp. 320-324.

[5] OlawaleAdepoju; Julius Wosowei; Shiwanilawte; HemaintJaiman, "Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques",Global Conference for Advancement in Technology (GCAT), IEEE, 2019, DOI: 10.1109/GCAT47503.2019.8978372, Pp.1-6.

[6] Fatima Zohra El hlouli, Jamal Riffi, Mohamed AdnaneMahraz, Ali El Yahyaouy, Hamid Tairi, "Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures", International Conference on Intelligent Systems and Computer Vision (ISCV), IEEE, 2020, DOI: 10.1109/ISCV49265.2020.9204185, Pp.1-5.

[7] S. S. HarshiniPadmanabhuni, AdityaSaiKandukuri, DebachudamaniPrusti, Santanu Kumar Rath, "Detecting Default Payment Fraud in Credit Cards",IEEE International Conference on Intelligent Systems and Green Technology (ICISGT), IEEE, 2019, DOI: 10.1109/ICISGT44072.2019.00018, Pp. 15-18.

[8] Gokhan Goy, Cengiz Gezer, VehbiCagriGungor, "Credit Card Fraud Detection with Machine Learning Methods", 4th International Conference on Computer Science and Engineering (UBMK), IEEE, 2019, Pp. 350-354.

[9] Arun Kumar Rai; Rajendra Kumar Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme",International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, 2020, DOI: 10.1109/ICESC48915.2020.9155615,Pp. 421-426.

[10] S. Benson Edwin Raj, A. Annie Portia, "Analysis on credit card fraud detection methods",International Conference on Computer, Communication and Electrical Technology (ICCCET), IEEE, 2011, DOI: 10.1109/ICCCET.2011.5762457,Pp. 152-156.

[11] WangyanFeng, Wenfeng Yan, Shuning Wu, NingweiLiu, "Wavelet transform and unsupervised machine learning to detect insider threat on cloud file-sharing",International Conference on Intelligence and Security Informatics (ISI), IEEE, 2017, DOI: 10.1109/ISI.2017.8004896, Pp.155-157.

[12] NizarIslah, Jamie Koerner, Roman Genov, Taufik A. Valiante, Gerard O'Leary, "Machine Learning with Imbalanced EEG Datasets using Outlier-based Sampling",42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, DOI: 10.1109/EMBC44109.2020.9175401,Pp. 112-115.

[13] Yan-Wei Pang, Xin Lu, Yuan Yuan, Jing Pan, "An outlier-insensitive Linear Pursuit Embedding algorithm",International Conference on Machine Learning and Cybernetics, IEEE, 2009, Pp. 2792-2796.

[14] YimingQuan, Lawrence Lau, Faming Jing, QianNie, Alan Wen, Siu-YeungCho, "Analysis and machine-learning based detection of outlier measurements of ultra-wideband in an obstructed environment", 15th International Conference on Industrial Informatics (INDIN), IEEE, 2015, DOI: 10.1109/INDIN.2017.8104909,Pp. 997-1000.

[15] Xiaohua Li, JianZheng, "Joint machine learning and human learning design with sequential active learning and outlier detection for linear regression problems",Annual Conference on Information Science and Systems (CISS), IEEE, 2016, DOI: 10.1109/CISS.2016.7460537, Pp. 1-5.

[16] Chunsheng Zhu, Yuanrui Zhan, ShijunJia,"Credit Risk Identification of Bank Client Basing on Supporting Vector Machines",Third International Conference on Business Intelligence and Financial Engineering, IEEE, 2010, DOI: 10.1109/BIFE.2010.25,Pp.62-66.

[17] Liu Shidong, Zheng Bin, Li Teng, Li Gengzhen, ShenPingyi, "Research on Risk Classification Based on AHP in Automobile Insurance", International Conference on Smart Grid and Electrical Automation (ICSGEA), IEEE, 2016, DOI: 10.1109/ICSGEA.2016.18, Pp. 157-159.

[18] Harry Raymond Joseph, "Poster: Software Development Risk Management: Using Machine Learning for Generating Risk Prompts",IEEE/ACM 37th IEEE International Conference on Software Engineering, IEEE, 2015, DOI: 10.1109/ICSE.2015.271, Pp. 822-834.

[19] AnirbanGanguly, Mo Mansouri, Roshanak Nilchiani, "A Risk Assessment Framework for analyzing risks associated with a Systems Engineering Process",IEEE International Systems Conference, IEEE, 2010, DOI: 10.1109/SYSTEMS.2010.5482460, Pp. 1-6.

[20] Yong Hu, Xiangzhou Zhang, Xin Sun, Jing Zhang, Jianfeng Du, Junkai Zhao, "A Unified Intelligent Model for Software Project Risk Analysis and Planning",3rd International Conference on Information Management, Innovation Management and Industrial Engineering, IEEE, 2010, DOI: 10.1109/ICIII.2010.504, Pp.110-113.

[21] Ruibin Xi, Nan Lin, Yixin Chen, "Compression and Aggregation for Logistic Regression Analysis in Data Cubes", IEEE Transactions on Knowledge and Data Engineering, Volume: 21, Issue: 4, 2009,Pp. 479 - 492.

[22] Leonardo O. Iheme, ŞükrüOzan, "Multiclass Digital Audio Segmentation with MFCC Features using Naive Bayes and SVM Classifiers", Innovations in Intelligent Systems and Applications Conference (ASYU), 2019, IEEE, DOI: 10.1109/ASYU48272.2019.8946441, Pp. 1-5.

[23] Zaimei Zhang, Kun Niu, Yan Liu, "A Deep Learning Based Online Credit Scoring Model for P2P Lending",IEEE, 2020, IEEE Access, Volume: 8, Pp. 177307 - 177317.

[24] Wei Xu, Yuan Liu, "An Optimized SVM Model for Detection of Fraudulent Online Credit Card Transactions", International Conference on Management of e-Commerce and e-Government, IEEE, 2012, DOI: 10.1109/ICMeCG.2012.39,Pp. 14-17.