The Improving perdition of Credit card fraud detection on PSO optimized SVM

T. Pavithra [#1], Dr. K. Thangadurai [#2]
[#1] *Research Scholar, P.G & Research Department of Computer Science, Government Arts College(A), Karur*
[#2] *Assistant Professor & Head, P.G & Research Department of Computer Science, Government Arts College(A), Karur*
[1] pavithrakrishnaveni2011@gmail.com
[2] ktramprasad04@gmail.com

Abstract— **Credit card fraud is an ever-growing problem with far-reaching commercial consequences and several methods. Information mining is successfully applied to credit card databases to automatically examine enormous volumes of complex data. The main topic in data mining, in particular for large data sets, is the selection of features. Functional selection is a method that is widely employed in the machine education and utilizes an algorithm of features outlined from the data. The best subset consists of the least accurate quantity of measurements; the remaining non-important dimensions are removed. This is one of two ways to avoid the curse of dimension (the other is to remove characteristics). There are two advances in the choice of characteristics known as transmission and reverse selection. A thorough field of research was chosen for pattern recognition, statistics and mining of data. The elimination of features with small or no predictive information determines a division of input variables. The main idea is to choose features. Feature selection techniques can be decayed into three broad classes. First one is Filter techniques and the second one is Wrapper method and the last one is Embedded method. Thus, in this paper particle swarm optimization based feature selection algorithm is implemented on credit card fraud dataset to improve anomaly detection in credit cards transactions.**

Keywords— **Feature Selection, Particle Swarm Optimization, Support Vector Machine.**

I. INTRODUCTION

Credit card fraud is a still increasing threat with far attainment consequences in the finance industry, corporate organizations, and government. Credit card Fraud can be defined as a criminal trick with a target of acquiring financial grow. High confidence in internet technology has liked increased credit card transactions. As credit card transactions develop into the most current mode of payment for both online and offline transaction, credit card fraud speed also accelerates. Credit card fraud can approach in moreover inner card fraud or external card fraud. Though, financial institutions have alert consideration to new computational methodologies to handle the credit card fraud problem.

The technology of data mining is one remarkably useful method for solving the issue of credit fraud [1]. Data mining is a method used by businesses to make raw data helpful. Through the use of software in big batches of information to make model products available, companies are able to know more about their clients and develop effective marketing strategies. Efficient information collections and storage as well as computer processing are essential for information mining. Data mining in the digital globe, which is interdisciplinary and computational in the process of detecting trends through technologies such as artificial intelligence, machinery learning, statistics and database systems in big information sets. The objective of the data mining technique is to extort and convert data from a data set into a comprehensible framework. The mining of data includes six popular functions such as identification of abnormalities, the mining of association rules, clustering, classification and regression. The rapid growth and advancements in knowledge datasets and computer techniques motivated the data accumulation at high speed. The all over supposed tasks in data mining requires the knowledge datasets to be processed to get any sort of understandable structure.

In order to recognize appropriate and relevel characteristics to enhance predictive precision, the processing of collected information itself has been a large challenge for the scientists, and the amount of data reduction techniques has been suggested for this. For example, data reduction may reduce information size through aggregation, removal or clustering of superfluous characteristics. The choice of features is one of the significant and commonly used methods for information decrease or information mining preprocessing. There are several feature selection benefits including reducing feature size, deleting no relevant, redundant, or noisy data, reducing computer costs, speeding up an algorithm for information mining and improving classification precision.

Feature choice is a process where the initial feature subset is selected [2]. An assessment criterion calculates the optimality of a function subgroup. Four key steps are taken to select the function: subset assessment, subset generation, stop criterion and validation of the results [3]. Subset generation is a method of discovery [4] that generates assessment sub-sets of applicant features based on the search approach. All candidates are assessed according to a compelling assessment criterion and compared to the past best. To improve the fresh subset, it replaces the previous best subset. The subdivision creation and assessment procedure is repeatedly conducted waiting for a certain stop criterion. Finally, the best subset chosen for validation by domain specialists or any other test can be provided as an entry for any data mining assignment.

Filter model, the wrapper model and the integrated model are usually confidential into three categories. The filter model reliably calculates and selects the overall personality of the information without linking an algorithm. A mining algorithm is required for the wrapper model and the assessment criterion is its presentation. It is aimed at improving mining presentation by using enhanced functionality suitable for an extraction algorithm but is also more computationally costly than the filter model. The embedded model attempts to take advantage of the two models by exploiting their special evaluation principle in different search stages.

II. RELATED WORK

SVM Algorithm based on Particle Swarm Optimization (PSO) is suggested by Duan, Min [5]. By analysing the impact of the SVM parameters on their performance, the parameter ranges and optimization procedure is suggested and the algorithm is used to predict short traffic flow to produce excellent outcomes. The algorithm reduces the information dimensions and retains the traffic flow sequence features by using SVM, which has the features of structural risk decrease and rapid worldwide PSO optimisation.

Le Bris [6] This paper aims to compare several FS requirements, from easy classical FS scoring results (e.g. JM reparability measure) to wrapper classification accuracies using state-of - the-art classifiers (e.g. Random Forestry). In a stochastic optimization frame, all these criteria have been evaluated.

Visalakshi [7] A important element of water contamination is the suggested feature space. (1) The primary aim is to speed up the feature selection procedure so that the suggested filter-based pre-selection characteristic is implemented and ensures that helpful information is unlikely to be separated in the original phase that this article addressed briefly. (2) In order to facilitate the shorthand sub-set of functions with high accuracy, the resultant functions will be again filtered using the genetics algorithm coded with the Support Vector Machine method and reduce the expenses. Experimental results show that the proposed methods trim down redundant features effectively and achieved better classification accuracy.

The hybrid Fuzzy expert system and the Fogg behavioural analysis were put forward by R. HaratiNik [8] and other individuals. A good hybrid model must be developed in conjunction with a costly method, which requires a lot of time to practice but provides extremely precise and accurate outcomes with optimization techniques that reduce system costs and rapidly make the system train. The selection of methods for a hybrid will rely on the use of the fraud sensing scheme and its environment.

Sam Maes [9] Credit card fraud detection proposal using the Bayesian Network and the Network of Artificial Neural two methods for machine learning. The article addressed how the Bayesian networks produced excellent outcomes following a brief training course and how the use of ANN increased their velocity.

Y. Sahin and E. Duman [10] Proposed credit card fraud identification by using the Support Vector and Decision Trees mix. Decision Trees exceeded the SVMs when the size of the data set was small but the size of the data set increased and SVM reached the exactness of decision-making bodies.

PoojaChougule and others [11] Submitted a easy fraud detection K-means and Simple Genetic Algorithms. In this paper they showed that the transactions based on the separate values of attributes and genetic algorithms are grouped into a k-means algorithm for optimization because the input k-means produces a boundary with an increase in the input size. K-means generated clusters which were optimized with the genetic algorithm in principle.

ThurayaRazooqi [12] The fraud detection scheme suggested by the Network and the Fuzzy Logic. They discovered that ANN was 33% more precise than fluid logic. The current system information was used to decide and each information received an affiliate attribute using the faded logic, as opposed to the use of ANN for validation of outcomes.

III. METHODOLOGY

This document incorporates the ideas of PSO and Support Vector Machinery (SVM), which are developed to obtain the finest characteristics in the forecast of credit card fraud.

A. SVM Theory

Supposing there are l independent and identically distributed sample data points $\{x_i, y_i\}$. $x_i \in R^m$ is input and $y_i \in R^m$ is output, $i - 1,2, \ldots, l$. Support vector regression machine maps the sample data $\{x_i, y_i\}$ to high dimensional feature space $\{\phi(x_i), y_i\}$ by non-linear function ϕ(x) . Then linear regression is performed by feature space to construct the optimum liner function

$$\hat{y} = < \phi(x), w > +b >$$

Where w is a weight vector; b is the paranoia. Construct the following optimization problems and solve the equation as

$$\min\frac{1}{2}\ \|w\|^2 + C\sum_i^1(\xi_{i+}\xi_i)$$

$$s.t. < \phi(x_i), w > +b - y_i \leq \in\ +\xi_i$$

$$y_i - < \phi(x_i), w > -b \leq \in +\xi_i\ \ \xi_{i,}\xi_i \geq 0, i = 1,2, \dots, l$$

Since the actual problem is often non-linear, we must extend linear cases. Kernal function K(w,x) is introduced to solute the equation by the dual problem, that is

$$\min\frac{1}{2}\sum_{i=1}^1(a_i - b_i)(a_i - b_j)K(x_i, x_j) + \in\sum_i^1(a_i + b_i)(a_i + b_j) - \sum_{i=1}^1(a_i - b_i)$$

$$s.t.\sum_{i=1}^1(a_i - b_i) = 0$$

$$0 \leq a_i, b_i \leq \frac{C}{l}, i = 1,2 \dots, l$$

Where C is punishment factor and it is a parameter between model generalization ability and accuracy. By solution, we get the optimum solution

$(a = a_{1,}b_{1,}a_{2,}b_2, \dots, a_i, b_i)$ Choose positive component b of a get

$$b^* = y_i - \sum_{i=1}^l(a_i - b_i)K(x_i, x_j) - \in$$

Then the estimation function is

$$y = \sum_{i=1}^l(a_i - b_i)K(x_i, x_j) - b^*$$

It can be used for regression analysis of unknown samples.

B. PSO Theory

The position of the particle at the optimization space is the feasible solution to the optimization problem in the PSO algorithm. A particle has a velocity vector to determine its movement direction and speed so that every particle is able to search for the best part at the moment and draw on its movement experience to discover the best value in solution space. The position and speed in the solution space at the starting state of each particle and distributed randomly. The particle then dynamically shifts the particle's place and velocity, depending on the known ideal individual value and global optimal value. The position of the particles and the speed of the flight are updated by the following equations:

$$v_{id}(t + 1) = v_{id}(t) + c_1 r_1(P_{bid} - x_{id}(t)) + c_2 r_2(P_{gd} - x_{id}(t))$$

$$x_{id}(t + 1) = x_{id}(t) + v_{id}(t + 1)$$

Where x_{id} denotes the location of para ticle; v_{id} denotes the velocity of partica le; c_1 and c_2 denotes learning factor; P_{bid} denotes the optimum individual value; P_{gd} denotes global optimum value; r_1 and r_2 denotes random number whose value is among 0 and 1.

The following algorithm denotes proposed PSO algorithm

1. Input Dataset
2. For each particle do
   2.1. Use SVM classifier as an objective function
      2.1.1. first consider the equation of the hyperplane w·x+b=0.
         2.1.1.1. Separate two classes X and Y using the following
            a. f(x)=w.x+b is positive if and only if xϵX
            b. f(x)≤0 if and only if xϵY
            c. Distance from a point x to a hyperplane wx + d = 0 is: |w x + d |/||w||
            d. Let the width of the space between observations is 2/||w||
         2.1.1.2. To find the optimal hyperplane for non-separable case
            a. for every i, yi(a.xi+b)≥1-ξi
            b. 1/2 ||a||2 +δ∑iξi is minimal
      2.1.2. Discover a non-linear separable hyperplane that accurately splits class variables
   2.2. Calculate the fitness value and compare it with the class label
   2.3. If the fitness value is better than the p_best_{id} in history
      2.3.1. Set the current fitness value as the new p_best_{id}
3. Choose the particle having the best fitness value of all particles as g_best_d
4. For each particle do
   4.1. Calculate velocity according to as $v_{id}(t) + c_1 r_1 (P_{bid} - x_{id}(t)) + c_2 r_2 \left(P_{gd} - x_{id}(t)\right)$
   4.2. Update particle position as $x_{id}(t+1) = x_{id}(t) + v_{id}(t+1)$
5. Repeat steps 2 and 4 until maximum iterations reached or minimum error criterion is not attained
6. Input the modified dataset in to SVM classifier
   6.1. determine the hyperplane that accurately splits class variables
   6.2. calculate and evaluate accuracy of traditional SVM with PSO-SVM

The algorithm suggested illustrates the process of detecting anomalies in credit card transactions. The method starts with the preparation of the information for missing values for the credit card fraud. These values are removed and this method allows the search space to be constructed for a certain amount of moment or to remain stagnant. The present strategy provides information for the transaction. Part in this search space is allocated and motions are activated in accordance with the standard PSO guidelines. Following completion of the iteration, the best global values are crossed by the best currently available particle values and the values showing the minimum distance are considered to be the best global values today. This process is repeated until the best solution is acquired or time goes by. The ultimate best worldwide values are regarded as the ultimate alternatives. The modified dataset is inserted in the SVM classification to correctly split the hyper plane between variables of class. The precise calculation of the traditional PSO-SVM. The precision of the data set is determined by the Vector Support Machine (SVM), which also serves as the PSO fitness assessor. SVM fitness is achieved by dividing the data set into exercises and test sets. The dataset is divided into 70% instruction and 30% test sets. The precision acquired in classification is not specified as an Adaptive Functional Value (AFV) for which the highest AFV for each particle is called the most personal and the finest AFV for each individual is worldwide. Position and velocity are the best characteristics. For each particle the bit value' 1' is a selected feature, and' 0' a not selected feature, the binary PSO is used. The PSO is the result of this work.

IV. EXPERIMENTATION AND RESULT DISCUSSIONS

The data sets for German credit cards include 20 characteristics of 1000 credit card transactions in Germany used in contests for KDD99. Since some of its information are descriptive, a second version was provided in order to use them in various categories, showing the information in numerical form in 24 properties [13].

The cases are randomly divided up into two different sets: 70% and 30%. The PSO binary is used to enhance presentation with lower numbers of characteristics as a feature selection algorithm for classification problems. Every iteration optimizes a particle in the PSO with fitness admiration. The test demonstrates the loading with all 16 characteristics of German credit cards and the outcomes achieved by the non-PSO SVM classifier in Figure 1. The SVM classification with 15 characteristics achieves precision of 78.6957%. The cases properly categorized are 543.

Once the credit card dataset is trained for extracting the feature subsets, the proposed PSO algorithm is performed and then on testing the classification performance of the feature subset of the SVM learning algorithm is calculated. Inertia weight w=0,8, acceleration variables x= y= 0,546, velocity maximum= 5, population size maximum= 300 and population size= 50, all PSO parameter algorithms are as follows. With 617 correctly classified instances, the accuracy of classification achieved by the proposed PSO-SVM classifier is 89.4203 percent. The following depicts the loading of PSO-SVM feature subset followed by the accuracy gained by the sub set of PSO-SVM features in Figure 2.

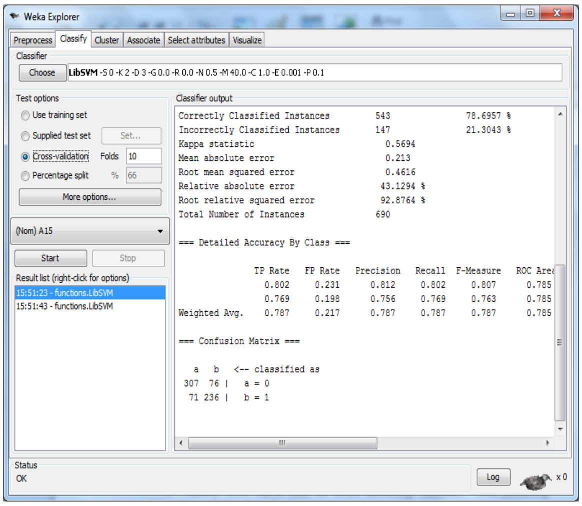

Figure 1 Prediction accuracy of German dataset (Original)

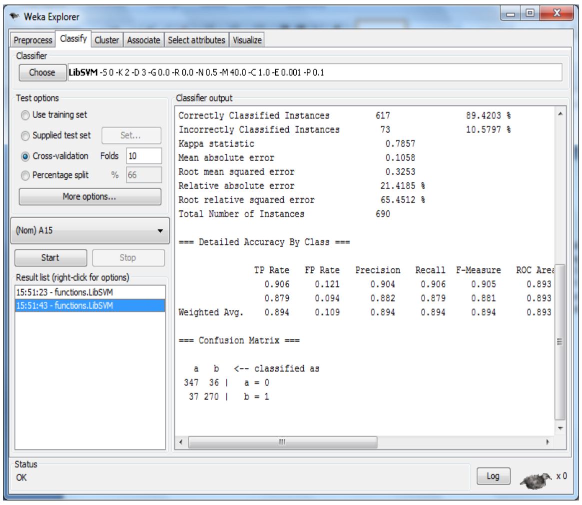

Figure 2 Prediction accuracy of Australian dataset (Original)

The test demonstrates that German credit card datasets with all 16 data characteristics are loaded and the outcomes of the SVM classification are achieved without PSO in Figure 1. The precision of the SVM classification with 16 characteristics is 78, 6957%. The cases properly categorized are 543.

Datasets from Australian credit cards comprise 14 characteristics of 690 credit card transactions in Australia. The information is categorized by 0 and 1, respectively [14] as fraudulent and normal.

Like the last chapter, Australian credit card data sets containing all 16 characteristics followed by the outcomes achieved by the SVM classification without the PSO in Figure 3 have been shown to demonstrate the performance and effectiveness of the suggested psOsvm algorithms. The precision achieved by 16 characteristics of the SVM classification is 78,8406%. The cases properly categorized are 544. Inertia w=0.8, acceleration variables x= y=0.546, speed maximum=5, population size= 300 and population size=50 the parameters for the PSO algorithm of function selection. The maximum velocity=5.The classification accuracy obtained by the proposed PSO-SVM classifier is 89.2754% with 616 correctly classified instances. With 616 correctly classified instances, the classification accuracy of the proposed classification PSO-SVM is 89. 2754 percent. This shows the charge of the PSO-SVM subset function followed by the precise results achieved by the sub-set of PSO-SVM characteristics in Figure 4.

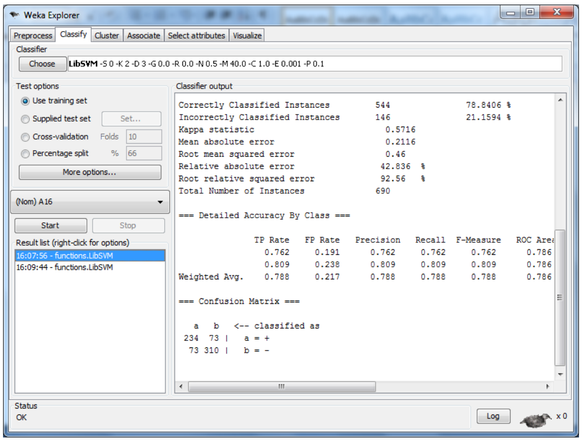

Figure 3 Prediction accuracy of PSO-SVM (German Credit card)

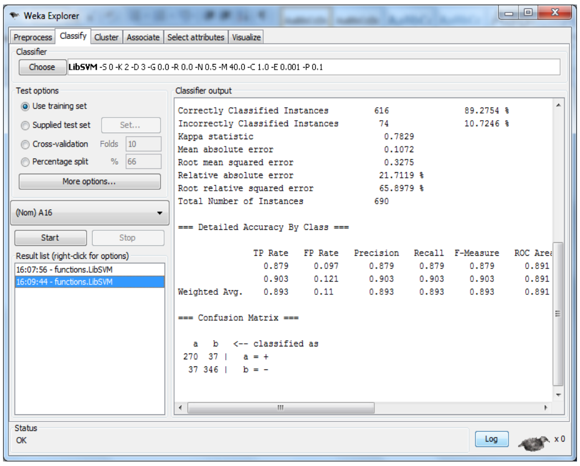

Figure 4 Prediction accuracy PSO-SVM (Australian dataset)

The suggested PSO-SVM algorithm has obtained 8 characteristics as the best characteristics for predicting two loan card bases with the greatest classification precision, showing a rise of 10% to 12% in forecast precision relative to the standard SVM classification. However, in the proposal PSO-SVM, which states that not all features need to get the highest accuracy, the number of features selected in the proposed method is slightly lower. In addition, the PSO focuses more on changing' w' and the acceleration variables ' x' and' y' inert weights. The efficiency of PSO is thus improved. The consumption of lengthy practice when dealing with large datasets is one of the primary problems with evolutionary based algorithms. Parallel development of the choice of the function subset in such a scenario. When the computational stress on the client system is important, the parallel architecture of PSO-SVM is important. Table 1. Presents the relative outcomes of PSO-SVM with the data set for credit card fraud.

Table I
PSO-SVM Comparative Analysis with Traditional Approaches

Algorithm	*Accuracy for German dataset*	*Accuracy for Australian Dataset*
Traditional Dataset with SVM	78.6957	78.8406
PSOSVM	89.4203	89.2754

The graphic demonstration of the PSO-SVM feature selection comparative results are presented in Figure 5.

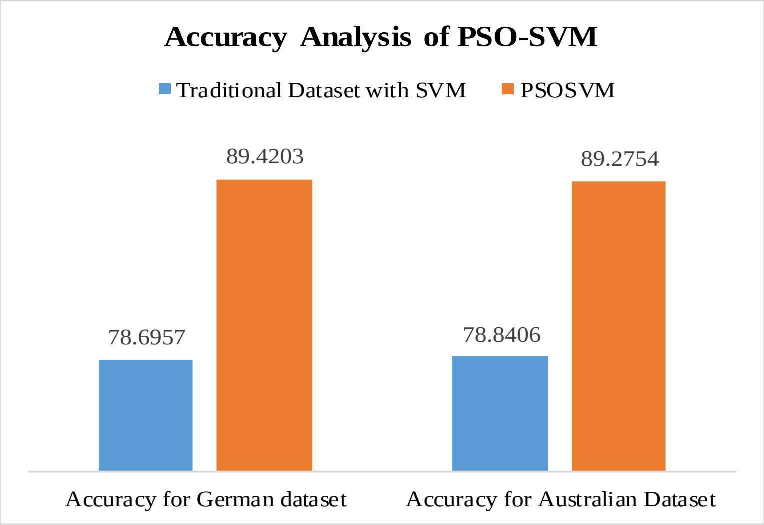

Figure 5 Accuracy Analysis of PSO-SVM

V. CONCLUSION

This document has a hybrid architecture involving the optimization of the particles swarm (PSO) and a feature selection algorithm based on Support Vector Machine (SVM) which increases the forecast precision of two credit card data sets. The most important tasks of the PSO-SVM algorithm are to select the finest characteristics and the classification model choice. The PSO algorithm is used in this job to select characteristics and the SVM algorithm is used for the iterative development of the feature selection. The findings of job have shown that a minimum of functionalities is extracted by the suggested PSO-SVM algorithm, giving 89 percent greater precision by misclassifying a single instance in the data set for heart disease. In addition, the suggested algorithm is implemented in conjunction with the developmental algorithm to minimize the computational cost involved. The PSO-SVM algorithm is thus an optimal preparatory instrument for enhancing feature selection optimisation.

REFERENCES

[1] Yee, Ong Shu, Saravanan Sagadevan, and Nurul Hashimah Ahamed Hassain Malim. "Credit card fraud detection using machine learning as data mining technique." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 10, no. 1-4 (2018): 23-27.

[2] Lee, Hojin, Dahee Choi, Habin Yim, Eunyoung Choi, Woong Go, Taejin Lee, Insuk Kim, and Kyungho Lee. "Feature selection practice for unsupervised learning of credit card fraud detection." *Journal of Theoretical and Applied Information Technology* 96, no. 2 (2018): 408-417.

[3] Zhao, Hong, Pengfei Zhu, Ping Wang, and Qinghua Hu. "Hierarchical Feature Selection with Recursive Regularization." In *IJCAI*, vol. 2017, pp. 3483-3489. 2017.

[4] Harish, B. S., and M. B. Revanasiddappa. "A comprehensive survey on various feature selection methods to categorize text documents." *International Journal of Computer Applications*164, no. 8 (2017): 1-7.

[5] Duan, Min. "Short-Time Prediction of Traffic Flow Based on PSO Optimized SVM." In *2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 41-45. IEEE, 2018.

[6] Le Bris, Arnaud, Nesrine Chehata, Xavier Briottet, and Nicolas Paparoditis. "A comparison of several feature selection scores applied to hyperspectral data." In *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1-4. IEEE, 2015.

[7] Visalakshi, S., and V. Radha. "A Hybrid Filter And Wrapper Feature Selection Approach For Detecting Contamination in Drinking Water Management System." Journal of Engineering Science and Technology 12, no. 7 (2017): 1819-1832.

[8] HaratiNik, Mohammad Reza, Mahdi Akrami, Shahram Khadivi, and Mahdi Shajari. "FUZZY: A hybrid model for credit card fraud detection." In *6th International Symposium on Telecommunications (IST)*, pp. 1088-1093. IEEE, 2012.

[9] Maes, Sam, Karl Tuyls, Bram Vanschoenwinkel, and Bernard Manderick. "Credit card fraud detection using Bayesian and neural networks." In *Proceedings of the 1st international naiso congress on neuro-fuzzy technologies*, pp. 261-270. 2002.

[10] E. D. Y. Sahin, "Detecting credit card fraud by decision trees," in *Proceedings of the international multiconference of engineers and computer science*, Hong Kong, 2011.

[11] Chougule, Pooja, A. D. Thakare, Prajakta Kale, Madhura Gole, and Priyanka Nanekar. "Genetic K-Means Algorithm for Credit Card Fraud Detection." *International Journal of Computer Science and Information Technologies (IJCSIT)* 6, no. 2 (2015): 1724-1727.

[12] [12]Razooqi, Thuraya, Pansy Khurana, Kaamran Raahemifar, and Abdolreza Abhari. "Credit card fraud detection using fuzzy logic and neural network." In Proceedings of the 19th Communications & Networking Symposium, p. 7. Society for Computer Simulation International, 2016.

[13] https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data).

[14] https://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval).
[15] Zhang, Q., Segall, R.S., Web Mining: A Survey of Current Research, Techniques and Software, International Journal of Information Technology & Decision Making. Vol.7, No. 4, pp. 683-720. World Scientific Publishing Company (2008).
[16] Q. Yang and X. Wu, 10 challenging problems in data mining research, International Journal Information Technology Decision Making 5(4) (2006).
[17] Andrei Broder. A taxonomy of web search. SIGIR Forum, 36(2):3–10, 2002.
[18] Jiawei Han, Kevin, Chen-Chuan Chang "Data Mining for Web Intelligence" IEEE International Conference on Data Mining, 2002.