ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

Fast and Effective Credit Card Fraud Detection in Imbalanced Data using Parallel Hybrid PSO

Sivakumar Nadarajan, Dr.Balasubramanian Ramanujam

Research Scholar, PG and Research Department of Computer Science, J.J. College of Arts and Science, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

Professor, Research Supervisor, PG and Research Department of Computer Science, J.J. College of Arts and Science, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

ABSTRACT: Credit card fraud detection has been one of the major necessities of the current e-commerce based world. The ease of use provided by e-commerce transactions is hindered by the threat caused by fraudsters. Several models have been proposed for identifying fraudulent transaction in a credit card system. However, the threats still do tend to exist. This paper discusses and analyzes the major reasons for credit card models to fail and also proposes a metaheuristic based detection technique to overcome the problems and provide effective detections. This work proposes a parallel PSO based hybrid technique that incorporates Simulated Annealing to provide time effective predictions that are also more accurate compared to conventional techniques.

KEYWORDS: Credit card fraud detection, PSO, Simulated Annealing, Parallelization, Imbalance.

I. INTRODUCTION

E-Commerce has become one of the major utilities due to the flexibility of cross border purchases offered by them. Any purchase is made just a few clicks away, providing ease of purchase to the customers. However, any technological advancement is not without downsides. The improvement in payment techniques acts as the most luring prospect for fraudsters, as all the information is made available online. This scenario has also resulted in a huge increase in fraudulent cases concerning credit cards. Though several techniques such as CVV and PIN based authentications have been used, POS based frauds are still in the increase. This leads to a huge requirement for fraud detection techniques. The general knowledge of fraud systems and could be learnt from [1, 2]. A survey of standalone and ensemble fraud detection techniques was presented in [3].

Credit card frauds can be committed in several ways [4], this creates huge complexities in designing the detection systems. However, the categorization of frauds in terms of research is divided into application and behavioral [5]. The major difficulty arises with the fact that, as the fraud detection techniques evolve, so does the fraud committing patterns. The behavioural adaptations are quite rapid due to the evolution and availability of advanced technologies to both researchers and fraudsters [6]. Another major drawback in creating an effective fraud detection system is the requirement for confidentiality and the requirement for anonymization of records prior to the actual pattern mining [7, 8]. This leads to the loss of several behavioural patterns contained in the credit card transactions. The implicit imbalance existing in credit card transactions also provide a huge downside in determining an efficient algorithm for fraud detection [9]. This paper presents a solution for most of the above discussed problems by providing a PSO based parallel fraud detection algorithm that can be used in a real-time environment for fraud detection.

II. RELATED WORKS

Credit card fraud detection technique has several publications to its credit, however due to the technological advancements, the techniques tend to get outdated very frequently. This section describes some of the most recent and prominent techniques in credit card fraud detection.

An online store based credit card fraud detection technique was presented by Vlasselaer et al. in [10]. This technique is based on the intrinsic features obtained from the transaction history of the customers. This technique also proposes a network based analysis from the perspective of both the customers and the merchants to provide a suspiciousness score. Similar usage pattern based fraud detection techniques such as [11] are currently on the raise due to the failure of conventional techniques in identifying certain fraud patterns. A decision tree based fraud detection approach was presented by Sahin et al. in [12]. This technique minimizes the misclassification cost by identifying the best splitting

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

attribute for each terminal node in a decision tree. A similar decision tree based approach utilizing rough sets was presented by Jain et al. in [13] and Breiman et al. in [18]. Other cost sensitive fraud detection approaches include [20]. A meta-classifier that combines decision tree, naïve Bayesian and k-nearest neighbor algorithms was presented by Pun et al. in [14]. This technique was employed on an already analyzed dataset from bank to provide higher prediction efficiency. An ensemble score based credit card fraud detection technique was presented by Duman et al. in [15]. This technique uses an ensemble of Genetic Algorithm and Scatter search to identify frauds in credit card transactions. This is a cost based analysis method and the weightage of a misclassification cost is set to the maximum cash in the account. Hence this mode of selection has been observed to improve the detection process. A consumer buying behavior based credit card fraud detection system was presented by Jha et al. in [16]. A statistical rule based fraud detection system was presented by Bolton et al. in [17]. An association rule based fraud detection technique has been proposed by Sanchez et al. in [19]. This technique identifies fraudulent transactions on the basis of customer's buying patterns. Neural networks based fraud detection techniques [21, 22] have also been prominently used. However, most of these techniques fail to concentrate on imbalance, which is the prominent focus of this work.

III. CREDIT CARD FRAUD DETECTION IN IMBALANCED DATE USING PARALLEL HYBRID PSO

The credit card fraud detection system utilizes the past transactions for training, but in the actual operational scenario, it performs predictions on real time data generated by the customers carrying out credit card transactions. The training time is not particularly considered, but the prediction time plays a vital role in determining the efficiency of the system. It was observed from the previous contributions [24, 25] by the authors that even though the accuracy provided by PSO is excellent, time taken for providing the result does not comply with the time requirement of real time specifications. Hence this contribution provides a parallel variant of the hybrid PSO (or parallel PSO SA) to reduce the processing time.

The major motivation behind this contribution is that PSO [23], being embarrassingly parallel in nature can be parallelized to bring out its maximum efficiency. The actual working nature of particles is parallel, hence performing the actual process in a parallel environment can ultimately bring out the maximum efficiency of PSO.

The operations of parallel PSO begins with initializing the search space using the input data. The input data is analyzed and the boundaries are identified. These boundaries are set as the boundaries for the search space used by the particles in PSO. The input data is segregated to training and test data. Particle distribution is carried out on a single test data. All the other nodes in the search space are constructed using the training data. Initial *pbest* and *gbest* values are set to 0 and the initial velocity is identified in random using eq 1.

$$V_i \sim U(-|b_{up} - b_{lo}|, |b_{up} - b_{lo}|) \qquad (1)$$

where b_{up} and b_{lo} are the upper and lower bounds of the search space. This stage marks the beginning of the parallel processing. Each particle is accelerated in-parallel using their corresponding velocity values. The number of parallel executions are determined by the particle count *p*. However, the most optimal technique is to use the maximum available cores as the value for *p*. Each parallel thread identifies the values for r_p and r_g and the new velocity for the particle is obtained using eq 2.

$$V_{i,d} \leftarrow \omega V_{i,d} + \varphi_p r_p (P_{i,d} - X_{i,d}) + \varphi_g r_g (g_d - X_{i,d}) \qquad (2)$$

where r_p and r_g are the random numbers, $P_{i,d}$ and g_d are the parameter best and the global best values, $x_{i,d}$ is the value current particle position, and the parameters ω, φ_p, and φ_g are selected by the practitioner.

The calculated velocity is added to the particle's position and the particle is moved to a new location. Movement is PSO is continuous, however, the current application demands discrete particle movement to identify the appropriate fitness, hence the movement of the particle is discretized and the particle is moved to the closest node. Fitness of the node is compared with the current *pbest* and the lowest is taken as the new *pbest.* The *pbest* values here correspond to the prediction (normal or anomaly) for the current credit card transaction (test data). At this stage, several predictions are available for a single test data. This process is carried out in parallel for all the particles.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

Algorithm for Parallel Hybrid PSO:

1. *Search space boundary identification using input data*
2. *For each particle i=1…p*
    a. *Initialize particle p using uniform distribution function*
    b. *Initialize pbest and gbest values*
    c. *Initial velocity identification based on search space boundaries*
3. *For each particle i=1…p perform in parallel*
    a. *Generate r_p and r_g using normal distribution*
    b. *Identify the particle velocity using eq. (2)*
    c. *Update the particle's position to the nearest node*
    d. *If pbest< current fitness*
        i. *Assign current fitness to be the pbest*
4. *End Parallel*
5. *gbest<- Simulated Annealing (gbest,pbest,p)*
6. *Perform steps 3, 4 and 5 until stagnation condition is reached*
7. *gbest contains the best found solution*

Simulated Annealing(gbest,pbest,p)

1. *Let s = gbest*
2. *For k = 1 through p :*
    a. $T \leftarrow pbest_k$
    b. *Pick a random pbest (pb),* $s_{new} \leftarrow pb$
    c. *If* $P(E(s), E(s_{new}), T) \geq random(0, 1)$, *move to the new state:*
        - $s \leftarrow s_{new}$
3. *Output: the final state s*

P(e,e',T) was defined as 1 if e' < e and exp(-(e'-e)/T) otherwise.

All the *pbest* values and the *gbest* value are passed to Simulated Annealing module for processing. The Simulated Annealing module identifies the best solution from the set of solutions. This solution corresponds to the final prediction for the test data. If the node corresponding to the prediction contains normal data, the test data is also classified as normal data and if the prediction corresponds to an anomalous data, the test data is classified to be anomalous.

IV. RESULTS AND DISCUSSION

Hybrid Parallel PSO was implemented using C#.Net and the experiments were conducted on a DELL Precision T7600 Workstation with Intel Xeon E5 2680 CPU with 32 CPUs of 2.7 GHz each and 32 GB RAM. The results obtained by parallel hybrid PSO was compared with PSO SA to exhibit the efficiency of parallelization towards accuracy and time.
Accuracy obtained from PSO SA and parallel PSO SA are compared in Figure 1. It could be observed that the accuracy obtained by parallel PSO SA is much higher than the accuracy exhibited by PSO SA. The base algorithm used is PSO SA, however, it is to be noted that PSO is a metaheuristic technique. Metaheuristics in nature have the ability to provide improved results, as the time of processing increases. Since the proposed approach uses parallelization, the CPU time consumed is much higher than the actual time lapse, hence improved results were obtained.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 9 , September 2016

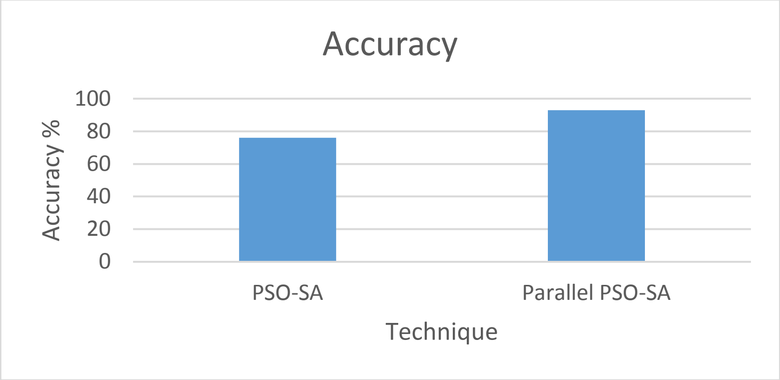

Fig. 1.Accuracy

A comparison between correct and incorrect predictions obtained from PSO SA and parallel PSO SA are shown in figures 3 and 4. It could be observed that while PSO SA exhibits 76% correct predictions, parallel PSO SA is observed to exhibit 93% correct predictions exhibiting improved performance. However, analysis in terms of correct and incorrect prediction levels alone are not sufficient. Analysis should also be conducted in terms of dominant and submissive entries due to the implicit imbalanced nature of the credit card transaction data.

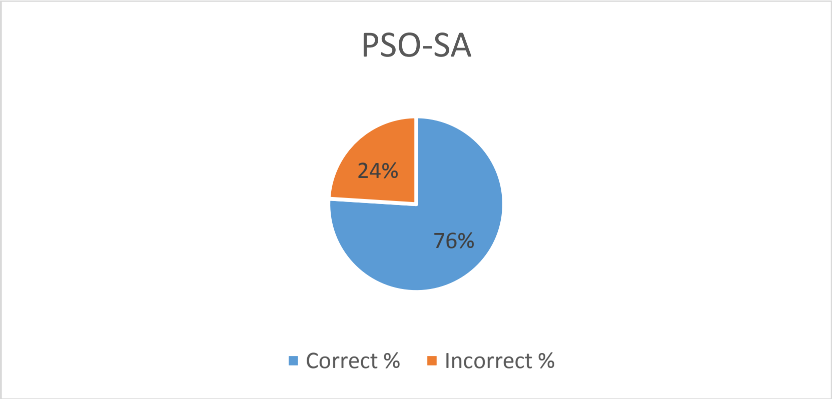

Fig. 2. PSO-SA

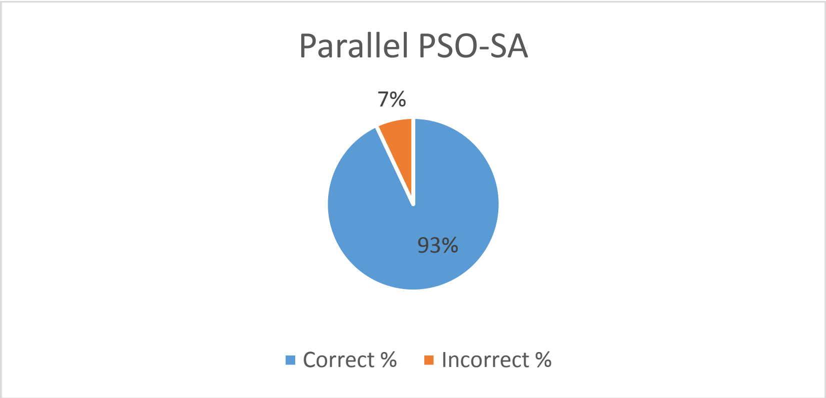

Fig. 3. Parallel PSO-SA

Dominant and submission prediction levels of PSO SA and parallel PSO SA are presented in figures 5-8. PSO SA exhibits 90% accuracy in predicting the dominant data, however it exhibits only 50% accuracy in predicting the

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

submissive data. This makes the prediction efficiency of PSO SA low, due to the highly imbalanced nature of the credit card transactions.

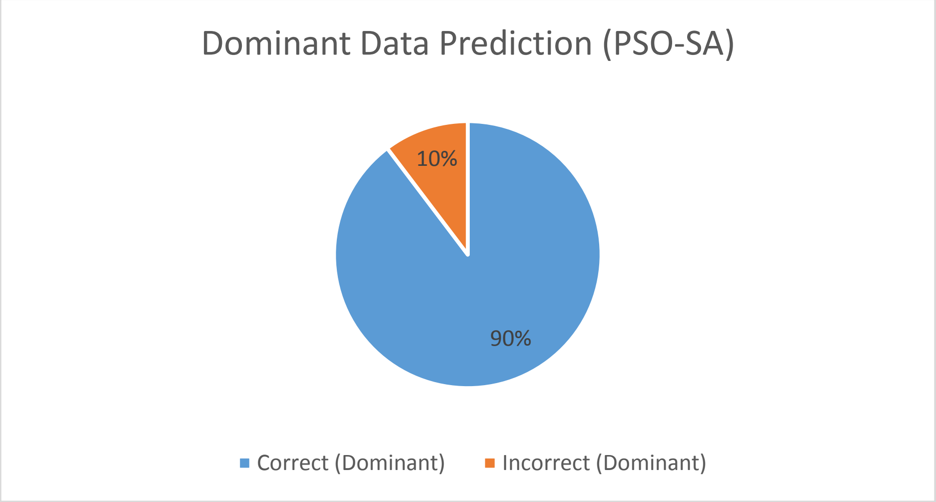

Fig. 4. Dominant Data Prediction (PSO-SA)

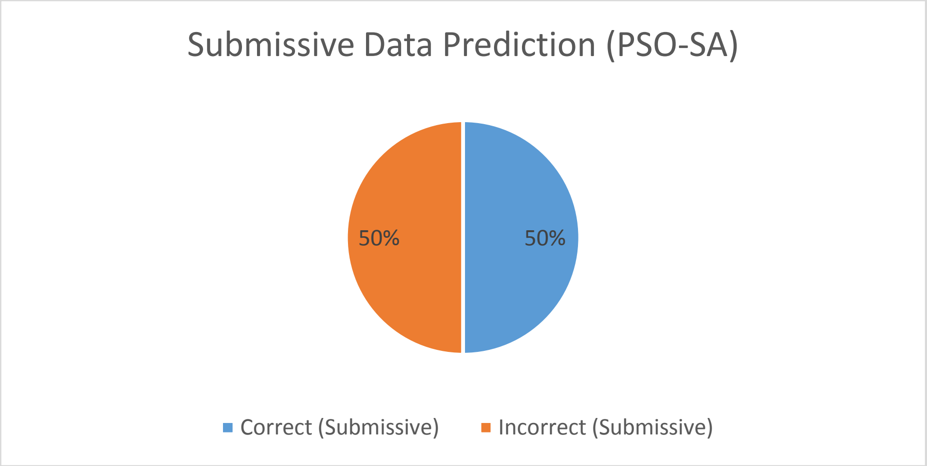

Fig 5: Submissive Data Prediction (PSO-SA)

However, when observing the dominant and submissive prediction levels of parallel PSO SA, it was observed that it exhibits accurate prediction levels both in terms of dominant and submissive data, making it the preferred choice for usage in fraud detection in credit card transactions.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

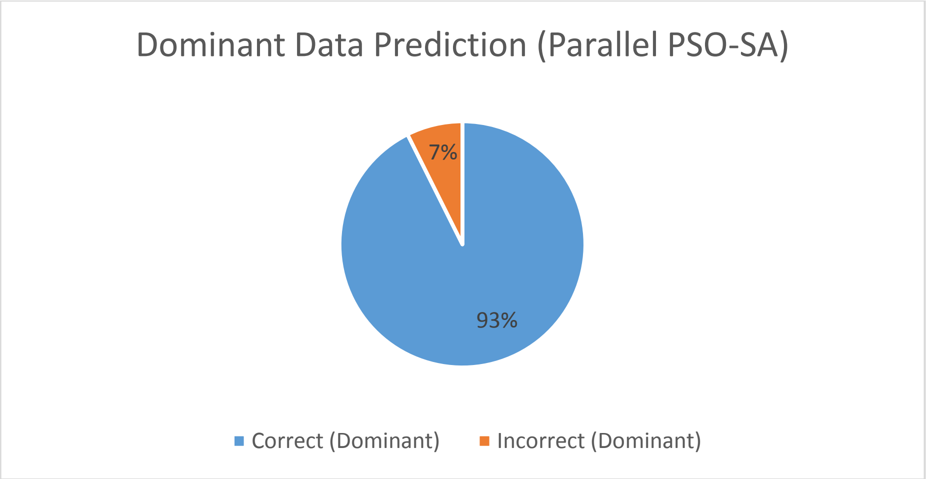

Fig 6: Dominant Data Prediction (Parallel PSO-SA)

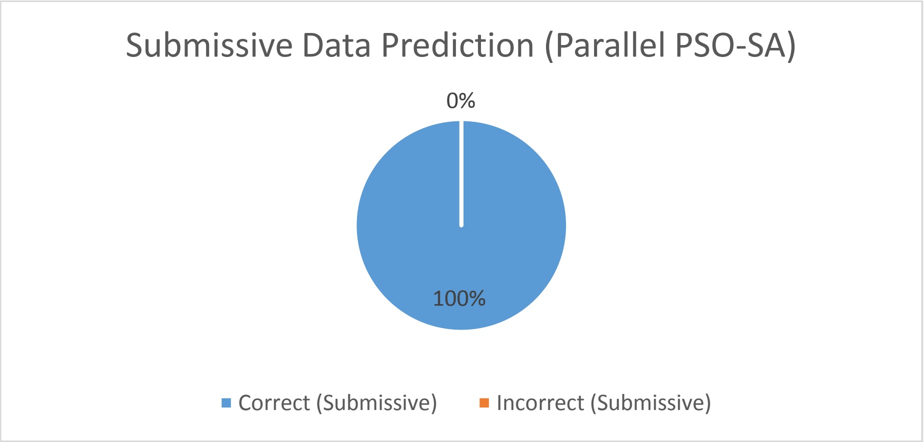

Fig 7: Submissive Data Prediction (Parallel PSO-SA)

V. CONCLUSION

PSO being a swarm based algorithm, lends itself to parallelization at ease. The nature of operation of swarm based algorithms is based on their parallel working model. Parallel PSO operates by parallelizing the operation of each of the particles available in the swarm. Hence irrespective of the number of particles contained in the swarm, the results can be obtained within the working time of a single particle. The experimental results show that a very considerable improvement was observed in the speed, in the order of one-third of the actual processing time was observed.

REFERENCES

[1] V. Hanagandi, A. Dhar, & K. Buescher, "Density-based clustering and radial basis function modeling to generate credit card fraud scores." In Proceedings of the IEEE/IAFE conference (pp. 247–251), 1996.
[2] D.J. Hand, G. Blunt, " Prospecting gems in credit card data." IMA Journal of Management Mathematics, 12, 173–200, 2001.
[3] Z. Masoumeh, K.R. Seeja, and M.A. Afshar, "Article: Analysis on Credit Card Fraud Detection Techniques: Based on Certain Design Criteria." International Journal of Computer Applications 52(3):35-42, 2012.
[4] G. Blunt, &D. J.Hand, "The UK Credit Card Market." Technical Report, Department of mathematics, Imperial College, London, 2000.
[5] R. J.Bolton, &D.J. Hand, "Unsupervised profiling methods for fraud detection." In Conference on credit scoring and credit control, Edinburgh, 2001.
[6] R. Bose, "Intelligent technologies for managing fraud and identity theft." In Proceedings of the third international conference on information technology, New Generations, 2006.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

[7] A. Dal Pozzolo, O.Caelen, Y.A. Le Borgne, S.Waterschoot, &G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective." Expert systems with applications, 41(10), 4915-4928, 2014.
[8] L. Delamaire, H. Abdou, &J. Pointon, "Credit card fraud and detection techniques: A review." Banks and Bank Systems, 4, 57–68, 2009.
[9] N. Japkowicz, & S. Stephen, "The class imbalance problem: A systematic study." Intelligent Data Analysis, 6, 429–449, 2002.
[10] V. Van Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, &B. Baesens, "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions." Decision Support Systems, 75, 38-48, 2015.
[11] R. Devaki, and V. Kathiresan, "Article: Credit Card Fraud Detection using Time Series Analysis." IJCA Proceedings on International Conference on Simulations in Computing Nexus ICSCN(3):8-10, 2014.
[12] Y. Sahin, S. Bulkan, &E. Duman, "A cost-sensitive decision tree approach for fraud detection." Expert Systems with Applications, 40(15), 5916-5923, 2013.
[13] R. Jain, B. Gour, and S. Dubey, "Article: A Hybrid Approach for Credit Card Fraud Detection using Rough Set and Decision Tree Technique." International Journal of Computer Applications 139(10):1-6, 2016.
[14] J. Pun, and Y. Lawryshyn, "Article: Improving Credit Card Fraud Detection using a Meta-Classification Strategy." International Journal of Computer Applications 56(10):41-46, 2012.
[15] E. Duman, &M.H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search." Expert Systems with Applications, 38(10), 13057-13063, 2011.
[16] S. Jha, M. Guillen, &J.C. Westland, "Employing transaction aggregation strategy to detect credit card fraud." Expert systems with applications, 39(16), 12650-12657, 2012.
[17] R.J. Bolton, &D.J. Hand, "Statistical fraud detection: A review." Statistical Science, 17(3), 235–249, 2002.
[18] L. Breiman, J.H. Friedman, R.A. Olshen, &C.J. Stone, "Classification and regression trees." CA, Wadsworth, Belmont, 1984.
[19] D. Sánchez, M.A. Vila, L. Cerda, &J.M. Serrano, J. M. "Association rules applied to credit card fraud detection." Expert Systems with Applications, 36(2), 3630-3640, 2009.
[20] A.C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten, "Cost sensitive credit card fraud detection using bayes minimum risk," Machine Learning and Applications (ICMLA), 2013 12th International Conference on. Vol. 1, IEEE 2013, pp. 333–338, 2013.
[21] E. Aleskerov, B. Freisleben, B. Rao, "Cardwatch: a neural network based databasemining system for credit card fraud detection," Computational Intelligence for Financial Engineering (CIFEr), 1997., Proceedings of the IEEE/IAFE 1997, IEEE 1997, pp. 220–226, 1997.
[22] B. Baesens, R. Setiono, C. Mues, J. Vanthienen, "Using neural network rule extraction and decision tables for credit-risk evaluation," Manag. Sci. 49 (3) (2003) 312–329, 2003.
[23] Y. Shi, and R. Eberhart, "A modified particle swarm optimizer." InEvolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on (pp. 69-73), 1998.
[24] Sivakumar Nadarajan & Dr. Balasubramanian Ramanujam, "Encountering Imbalance in Credit Card Fraud Detection with Metaheuristics" Advances in Natural and Applied Sciences. 10(8); Pages: 33-41,2016.
[25] N Sivakumar and R Balasubramanian. "Article: Enhanced Anomaly Detection in Imbalanced Credit Card Transactions using Hybrid PSO". International Journal of Computer Applications 135(10):28-32, 2016.