Next-Gen Fraud Detection: Machine Learning and Encryption for Credit Card Security

Giriharan V., Dilip Raj K. and Karthiga R.
Department of Computer Science and Engineering, St. Joseph's Institute of Technology, Chennai, Tamil Nadu, India

Keywords: Credit Card Fraud Detection, Customer Activation Prediction, Machine Learning, Django Framework, Web Application Development.

Abstract: Today in the field of finance, it is critical to secure credit card transactions and the customer engagement initiatives. This paper studies how machine learning is used to predict fraudulent transactions and customer engagement, exemplifying the use of encryption algorithms in order to safeguard sensitive information. Using cutting edge algorithms, our models achieve great accuracy in fraud detection and consumer behavior predictions. Integrating these predictive models into the Django framework provides for an agile and scalable web architecture, ensuring easy integration and improved usability. This approach not only strengthens the protection of data with encryption, but also provides the foundation through which specific data sealed by the institution can be uncovered, and thus improves the efficiency on combating fraudulent activities and retain customers.

1 INTRODUCTION

As the digital revolution continues to accelerate, protecting financial transactions is more important than ever when card fraud threatens both individuals and enterprises. The ever-evolving nature of fraud has outpaced antifraud detection tools. To solve this problem, the state of the art in machine learning offers a solution. Using advanced algorithms and data-driven techniques, machine learning models can sift through large records of transactions to detect suspicious activity. This not only improves the accuracy of fraud detection, but also reduces false alerts so legitimate transactions can be completed without issue." Against the backdrop of this evolving threat, a significant innovation in protecting financial transactions is using encryption in fraud detection predictions supported by advanced machine learning.

In the financial services industry, credit card fouls pose a challenge owing to the rosy figures that come with considerable losses every year. In recent years, the usage of credit cards has increased significantly within the banking Sector. Nevertheless, this increased usage of credit cards hasalso resulted in an increase in default levels faced by financial institutions. Credit card frauds, as defined, involve the misuse of cardholder's information for financial gains, which become identifying factors for fraudulent credit transactions. The dataset consists of 284,807 transactions of European clients, which were scrutinized to measure the performance of the various algorithms. The results obtained 0.02% Neural Network technique and 100% accuracy rate, which demonstrates this technique has the highest predictive power and accuracy.

2 LITERATURE SURVEY

In paper, a fraud detection system for credit cards is presented. The overall system consists of three stages. Firstly, Luhn algorithm check for verification card numbers and thus to differentiate legitimate from fraudulent cards. Secondly, dynamically checking of the expiration date of the card. Finally, the validation of Card Verification Value (CVV) or Card Verification Code (CVC), where digit counts fall within an acceptable range. More, the solution incorporates a user-friendly GUI for easier application. With rigorous testing conducted on both valid and fraudulent cards, our system showed better efficiency than the distance sum model; it was also able to capture serial numbers not detected by the other methods relying on the Luhn algorithm.

V., G., K., D. R. and R., K.
Next-Gen Fraud Detection: Machine Learning and Encryption for Credit Card Security.
DOI: 10.5220/0013909300004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 138-145
ISBN: 978-989-758-777-1

In paper (T.P. Bhatla et al., 2013), valuable insights into the mapping processes and current trends in the yearly developments of machine learning applications have been gathered for future research. The study follows a three-phase approach. The first phase involved compiling descriptive statistical analysis from Scopus and Herzing's Publish or Perish, the second phase utilized Vos Viewer for quantitative citation analysis, and the third phase focused on interpreting the findings. This research categorized the studies into three clusters: machine learning related to data mining, fraud detection, and fraud classification. The second cluster addresses the complexities of credit card transaction data, aiming to enhance fraud detection systems, while the third cluster is dedicated to improving classification techniques for managing imbalanced data and limited records. This area has seen rapid growth in recent years, especially regarding the performance of fraud classification with unbalanced data. Future research will aim to understand how machine learning has evolved over time for fraud detection and conduct a comparative study across Scopus, Web of Science, and ScienceDirect.

Paper (Chan PK et al., 1999) is online transactions grow in number; credit card fraud is becoming one of the most frequent in today's world. Thereby it incurs substantial losses. Thus, a good strategy for initial stage of fraud detection to minimize loss that has already been imposed by this crime. Using of machine learning algorithms for fraud detection is a potential solution. This paper examines the recent developments in the application of machine learning to detect credit card fraud. Four machine learning models were compared in terms of accuracy. According to the results, the Catboost algorithm achieves the highest accuracy of 99.87%. The analyzed data was obtained from Kaggle. In paper (A.C. Bahnsen et al., 20213), Logistic Regression as well as Random Forest have been employed to enhance the performance of fraud detection systems and operational effectiveness.

After thorough examination of these algorithms in the context of credit card fraud, the study demonstrates that both models are used jointly within Fraud Fort. Thus, this confirms the fact that the integration between logistic regression and random forest will result in a more reliable anti-fraud system, thus contributing to a safer and more trustworthy economic environment. In paper (Dal Pozzolo et al., 2014), a model is proposed that utilizes the XG Boost classifier for detecting fraud in transactions. Contrasting from traditional methods that are defined using only one threshold value, our model calculates and compares a set of multiple threshold values to determine the optimum, thus maximizing effectiveness and efficiency.

Paper A.C. Bahnsen et al., proposes a technique based on machine learning for the detection of credit card fraud using labeled data to distinguish between fraudulent and genuine transactions. Supervised learning techniques were used in the experiment.

In paper (West J, Bhattacharya M., 2016) covers various credit card fraud detection techniques that offer advanced protection against multiple fraud types. We provide a comparison of these techniques in terms of their accuracy, time efficiency, and cost-effectiveness, while pointing out their strengths and weaknesses to help determine the most appropriate technique.

Paper (Zhang, S et al., 2015) evaluates and compares several algorithms, including Random Forest, Decision Tree, and Artificial Neural Networks (ANN). This research is crucial for advancing fraud detection technologies, as it not only examines the performance of various machine learning algorithms but also sheds light on feature engineering. The findings emphasize the importance of designing scalable, resilient, and real-time solutions to address the evolving nature of fraud strategies.

The paper centers on credit card fraud detection by using transaction analysis, and specifically aims to detect fraudulent items. Then the results are applied to classify the new transactions as fraudulent or legitimate. Detecting all fraudulent transactions with less misclassification of unfraudulent ones is of paramount importance. The analysis uses "neighbor outliers" and "isolation forest" approaches to the PCA-transformed credit card transaction data (Van Vlasselaer V et al., 2015).

The presented credit card fraud detection model addresses class imbalance by synthetic minority oversampling and the feature extraction and representation are enhanced by sequential deep learning methodologies. This model achieves accuracy, detection rate and F-measure that are greater than those of the current ones. Its utility may have a major impact not just on online markets but may also limit fraud and increase trust in online purchases. The advanced feature extraction methods employed by the model contribute to its superior performance compared to other models (R.J. Bolton, D.J. Hand, et al., 2001).

3 EXISTING SYSTEM

As the credit card industry is flourishing with the

advent of Internet technology, the number of fraudulent credit card transactions is also on the rise, more too in view of the complexity of transaction information. Many of the AI or machine learning models applied for the detection of fraud have lacked effectiveness due to redundant feature data and the imbalanced nature of transaction datasets. Truly speaking, better feature engineering and smart sampling methods could help boost detection accuracy. In this paper, we proposed thereby embodies a compound elimination strategy to remove correlation and redundant features, thus improving the quality of the data. Also, a multifactor synchronous embedding mechanism was implemented, making each feature contributing best towards the decision- making aspect of fraud detection. Furthermore, SOBT eases the imbalance in legitimate vs. fraudulent transactions, thereby improving the model's detection capacity for fraud. These detailed experiments shown with real-life datasets demonstrate that this method outperforms the existing ones in the face of limitations such as computing intensity, risk of overfitting, and concerns regarding scalability and implementation.

4 PROPOSED SYSTEM

A system is these, proposes to enhance the detection of credit card fraud and the prediction of customer activation over a platform of Django that integrates a couple of models in machine learning. Historical data about transactions and customer behaviors are used to develop robust predictive models to detect fraud and predict accurate customer activation. Industry-standard cryptographic algorithms are used for input and output encryption based on "lost data" and "security data" to help reduce the risk of sensitive data being compromised. This helps to ensure that the data remains secure and untampered with. Django forms the backbone of the framework and offers a scalable platform for deploying and managing the machine learning models. It interacts with the database, manages the user authentication workflow, and implements critical security practices. This has been built on top of Django and its powerful features for making prediction in realtime, process data at ease and communicate with web securely. Moreover, we framework-based application was developed to simplify redact, utilizing dataset balancing and algorithmic comparisons for accuracy and scalability. Django's modular structure enables straightforward updates and easier system maintenance.

5 METHODOLOGY

Employing the latest algorithms, it scans through all transaction records for the most suspicious activities with the highest precision. The algorithms, powered by deep learning techniques, assisted with supervised models, help the system learn and adjust to changing fraud patterns automatically, improving detection rates continuously. An intuitive and seamless interface that seamlessly connects to existing financial networks as well as real- time fraud warnings and in depth reports to help re-. make informed decisions. Designed with both security and performance in mind, our algorithm fights fraud without introducing too many false positives, creating a secure and effective financial protection solution. Figure 1 shows the system architecture.

5.1 System Architecture

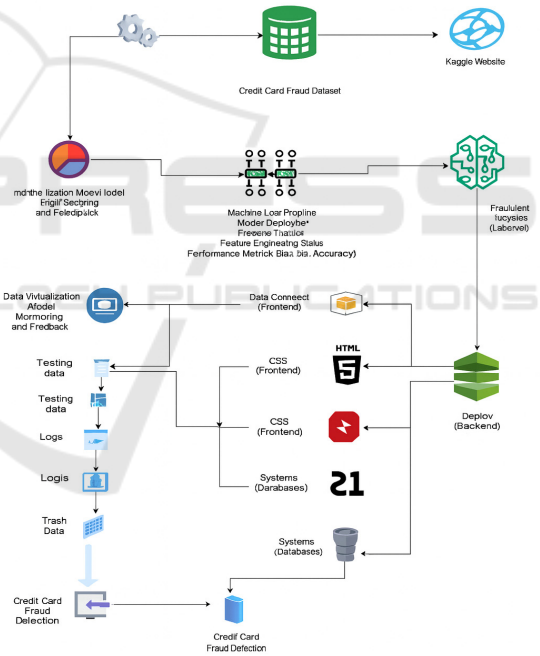

Figure 1: System architecture.

5.2 Module Description

5.2.1 Data Pre-Processing

In most practical applications, data samples do not fully capture population characteristics, making

validation indispensable. These methods help detect missing and duplicate values while determining whether data types are integers or floating-point variables. Validation datasets are used to impartially measure model performance during hyperparameter tuning. As model configurations adjust based on validation data, the risk of evaluation bias increases. This dataset is frequently utilized to refine machine learning models, allowing engineers to fine-tune hyperparameters. The tasks of data collection, preprocessing, and ensuring structural integrity require significant time. Gaining insights into data properties during identification aids in selecting the optimal algorithm for model development. Figure 2 shows the data processing steps.

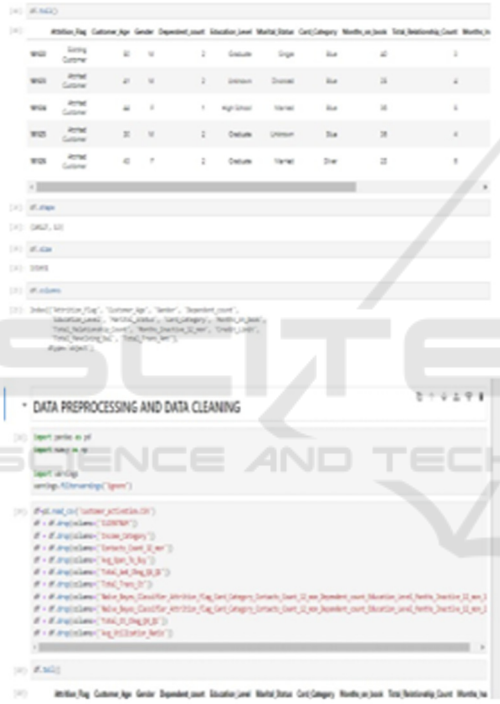

Figure 2: Data processing.

5.2.2 Data Validation

Import library packages and load the dataset; analyze variables, examine data structure, identify types, and detect non-found entries. A verified dataset is kept for an unbiased estimate of the model's skill and to optimize hyperparameter tuning with the best possible use of test and validation datasets. Data preparation involves renaming datasets, removing unnecessary columns, and doing variate analyses. The cleaning methods and techniques are specific to each dataset because every dataset is unique. The core purpose of cleaning the data is to correct the errors or irregularities, therefore enhancing the value for analytics and informed decisions (figure 3).

Figure 3: Data validation.

5.2.3 Exploration Data Analysis of Visualization

Visualizing visualizations is the natural way of working with datasets to discover patterns, identify errors/outliers, etc. Based on domain knowledge, visualization in the form of plots/graphs can signify important relationships meaningfully. Displaying data graphically using plots and graphs may be more stimulating than numerical figures.

In prediction outcomes, estimates are developed with algorithms of linear equations that contain independent variables, which can be practically infinite in the negative and the positive direction. In classification, the output must be mapped into

categories, and among them, logistic regression showed the best prediction accuracy than the other models tested.

Data visualization and exploratory data analysis are huge areas and research beyond that suggested in the books mentioned above. Visual representation of data often makes patterns and trends clear, so visualization is an important ingredient in applied statistics and machine learning. Python plotting mastery enables us to explore and understand datasets more effectively.

5.3 Algorithm

5.3.1 Algorithm Explanation

In the context of supervised learning where a system is trained on labeled data inputs and then makes use of the training to classify new data points. Datasets may involve binary classes-for example, identifying whether or not an email is spam, or determining whether an individual is male or female-or multiple classes. Some common examples of classification are voice recognition, identification of handwritten texts, biometric authentication, and categorizing documents. Supervised learning algorithms are used on labeled datasets for identifying patterns, relating the identified patterns to new inputs, and thereby assigning suitable labels to unseen data.

5.3.2 Adaboost Classifier

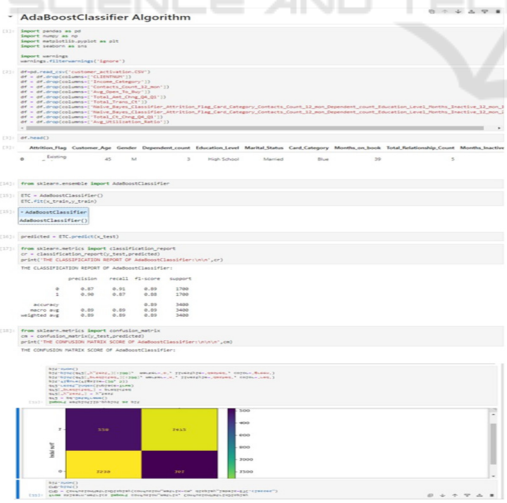

Figure 4: AdaBoost classifier.

The algorithm raises the weights of misclassified instances, directing the next weak learners to concentrate on these more difficult cases. Weak learners are trained sequentially, modifying the weights of misclassified instances. The model is constructed at the end by aggregating the learners, with each contributing proportionally to its accuracy, and the ultimate prediction being based on the weighted majority vote (figure 4).

5.3.3 Decision Tree Classifier

Decision Tree (figure 5) is a popular machine learning algorithm for classification and regression problems. It is depicted in the form of a tree, where every node is a test or choice on a feature. Decision Trees are most preferred in domains such as finance, healthcare, and marketing for their simplicity, easy interpretation, and capacity to solve complex decision- making problems.

Figure 5: Decision tree classifier.

5.3.4 K N Neighbors Algorithm

It is a simple yet powerful, the query point's class is found by the majority vote of its K nearest neighbours (figure 6). The choice of K is a critical parameter, as it can affect the model's overall performance is a non-parametric algorithm, i.e., it doesn't make any assumption about the underlying distribution of the data.

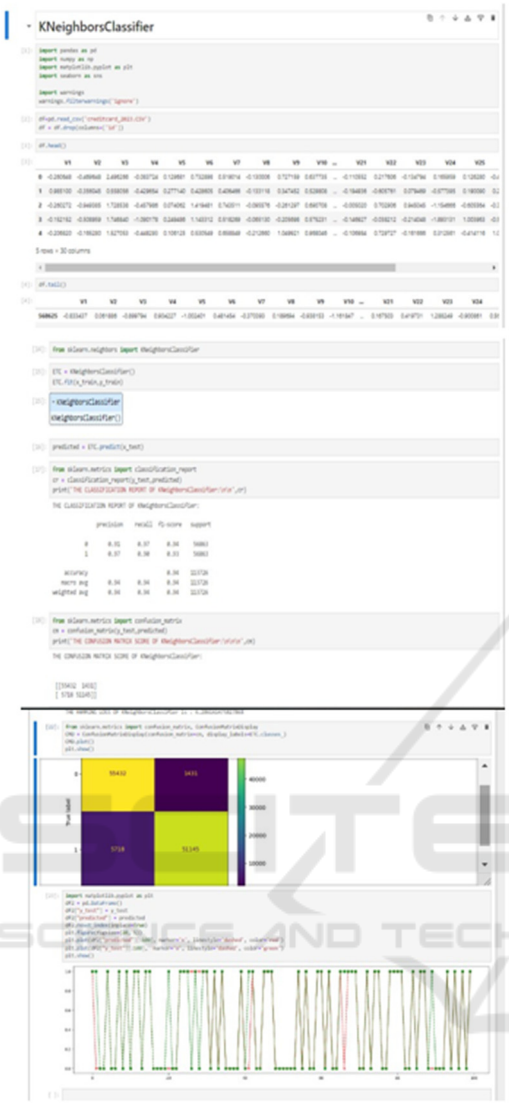

Figure 6: K N Neighbors algorithm.

The algorithm computes similarity in terms of distance metrics, commonly Euclidean distance, but other metrics like Manhattan or Minkowski distances can be used based on the dataset. Since KNN doesn't learn a model from the data, it is referred to as a lazy learning algorithm, where the whole dataset is retained and predictions are generated at the querying time. KNN's simplicity and ease of implementation are two of its strongest points. It can handle multi-class classification problems and learn complex decision boundaries. However, it has some drawbacks as well, such as its computationally costly prediction step, as it needs to compute distances to all the training data, and its sensitivity to the curse of dimensionality, where the performance of the model degrades with an increase in the number of features. By choosing an optimal K and distance metric and applying dimensionality reduction techniques, most of these problems can be resolved.

5.3.5 Extra Tree Classifier

It creates an ensemble of decision trees, which makes predictions and averages out their results for better overall performance. Extra Trees is a regular decision tree with extra randomness added to the occasion of building a tree. For instance, it randomizes every split by selecting subsets of features instead of developing the best split, which makes it more diverse between the trees (figure 7).

Figure 7: Extra tree classifier.

The algorithm can produce better accuracy and generalization than a single decision tree by averaging predictions from these trees. So this Extra Trees Classifier is very useful in cases of very large datasets with high-dimensional features. The algorithm reduces the complexity of determining the best split point a randomized manner due to this mechanism, thus accelerating training. Also it is less prone to overfitting compared to other ensemble learners like Random Forest, as Randomness used in its construction reduces the chances of model overfitting the training set and predicting accurately

on new instances. The Extra Trees Classifier is used in various practical applications for real-time classification that requires good accuracy and low latency. It was able to deal with more data and provide more accurate output making this technique useful in a variety of sectors such as finance, healthcare, and image processing. The algorithm is particularly useful in cases of datasets with a high dimensionality (many features), its randomization mechanisms allow to control the complexity of feature interactions while maximising accuracy and computational efficiency.

6 CONCLUSIONS

Overall, the use of encryption models adds a better layer of filtering for credit card fraud detection. We are able to enhance the accuracy and safeguard of anti-fraud technologies with latest algorithms and strong encryption protocols. Machine learning techniques such as deep learning and ensemble models are powerful tools available to identify suspicious activity by analyzing large volumes of transactional data and recognizing abnormal patterns. Encryption protects sensitive information at rest during the analysis window from intrusion. This activity further refines the accuracy of fraud detection while promoting user confidence through top-tier data protection practices.

6.1 Future Work

Thus, the approach is making the combination of predictive techniques and cloud infrastructure as effective as possible for IoT application where cloud platforms such as AWS or Azure or Google Cloud is thereby used to run well-trained ML models that process large amounts of IoT data. Hosting and training these models are available through cloud-based services like Google AI Platform and AWS Sage Maker. For smooth communication of IoT device to the cloud, messaging protocols like MQTT or HTTP can be used for exchanging data in real-time. In addition, edge computing could be employed to help reduce cloud- based processing dependence by processing data near IoT devices, allowing for rapid decision making and reducing the cloud computational load. IoT specific lightweight models can be rolled out on the edge for real- time predictions. To handle the dynamic volume of data traffic and keep the prediction models available all the time, auto- scaling can be used in the cloud infrastructure. Moreover both encryption and security protocols are necessary to ensure confidentiality as well as integrity in order to transport data across IoT devices and cloud infrastructure. This allows for better real-time decision-making with improved performance, scalability, and security.

REFERENCES

[Acessed:14- SEP-2018]

A.C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten,"Cost sensitive credit card fraud detection using bayes minimum risk, Machine Learning and Applications" ICMLA 12th International Conference on. Vol. 1, IEEE, 2013, pp. 333–338.

A.C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten,"Improving credit card fraud detection with calibrated probabilities", Proceedings of the, SIAM International Conference on Data Mining. Pennsylvania, USA2014, pp. 677–685.

Alex Shenfield, David Day, and Aladdin Ayesh, "Intelligent intrusion detection system using artificial neural networks," vol. 4, no.2, pp. 95-99, June 2018.

Antonia Nisioti, Alexios Mylonas, Paul D. Yoo, Vasilios Katos. "From Intrusion Detection to Attacker Attribution: A Comprehensive Survey of Unsupervised Methods", IEEE Communications Surveys & Tutorials, 2018.

Chan PK, Fan W, Prodromidis AL, Stolfo SJ. "Distributed data mining in credit card fraud detection". IEEE Intelligent Systems and Their Applications, Nov; 14(6): 1999, pp. 67-74.

D. Stiawan, A.H. Abdullah, and M.Y. Idris, "The trends of intrusion prevention system network, in 2010" 2nd International Conference on Education Technology and Computer, vol. 4, pp. 217-221, June 2010.

D.J.Weston, D.J. Hand, N.M. Adams, C.Whitrow, P. Juszczak," Plastic card fraud detection using peer group analysis", ADAC 2 (1) 2008, pp 45–62 ,

Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, G. Bontempi,"Learned lessons in credit card fraud detection from a practitioner" perspective, Expert Syst. Appl. 41 (10), 2014 , pp. 4915– 4928.

ECB, "Third Report on Fraud, European Central Bank", February 2014.

Gómez, J. A., Arévalo, J., Paredes, R., & Nin, J, "End-to-end neural network architecture for fraud scoring in card payments." Pattern Recognition Letters, 2017.

Gulshan Kumar. "The use of artificial intelligence based techniques for intrusion detection: a review", Artificial Intelligence Review, 09/04/2010.

Iman Sharafaldin, ArashHabibiLashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018.

J.T. Quah, M. Sriganesh,"Real-time credit card fraud detection using computational intelligence," Expert Syst. Appl. 35 (4), 2008, PP 1721– 1732.

L. Akoglu, H. Tong, D. Koutra, "Graph based anomaly detection and description: a survey", Data Mining and Knowledge Discovery Available Online", http://dx.doi.org/ 10.1007/s10618-014-0365-y. 2014.

Liao H.-J., Lin C.-H.R., Lin Y.-C., and Tung K.-Y. "Intrusion detection system: A comprehensive review" Network Computing. Appl., Rev., 36 (1), pp. 16-24 ,2013.[Online].Availablehttps://www.kdnuggets.com/ 2016/10/beginners-guide- neuralnetworks-python-scikit-learn.html.

Monowar H. Bhuyan, Dhruba K. Bhattacharyya, Jugal K. Kalita. "Network Traffic Anomaly Detection and Prevention", Springer Nature, 2017.

R.J. Bolton, D.J. Hand, et al.,"Unsupervised profiling methods for fraud detection," Proceedings of the VII Conference on Credit Scoring and Credit Control, 2001, pp. 235–255, (Edinburgh, United Kingdom).

Rosenblatt F. "The perceptron: A probabilistic model for information storage and organization in the brain" Psychol. Rev., 65 (6), pp. 386-408, 1958.

Singh R., Kumar H., Singla R.K., and Ketti R.R. "Internet attacks and intrusion detection system: A review of the literature" Online Inform. Rev., 41 (2), pp. 171-184, 2017.CrossRefView Record in ScopusGoogle Scholar.

T.P. Bhatla, V. Prabhu, A. Dua, "Understanding credit card frauds", Cards Bus. Rev. 1 (6), 2003, pp. 1–17.

V. Zaslavsky, A. Strizhak, "Credit card fraud detection using selforganizing maps" Inf. Secur. 18, 2006, PP 48.

Van Vlasselaer V, Bravo C, Caelen O, Eliassi-Rad T, Akoglu L, Snoeck M, Baesens B. "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions." Decision Support Systems. 2015, pp. 38-48.

West J, Bhattacharya M. "Payment Card Fraud Detection Using Neural Network Committee and Clustering. Computers & Security. 2016; 57:47- 66.

Wu J., Peng D., Li Z., Zhao L., and Ling H. "Network intrusion detection based on a general regression neural network optimized by an improved artificial immune algorithm."Rev.,10 (3), 2015. [Online] Available https://www.ncbi.nlm.nih.gov/pubmed/25807466 [Acessed:14-SEP-2018].

Zhang G.P. "Neural networks for classification: A survey" IEEE Trans. Syst. Man Cybern. C, Rev., 30 (4), pp. 451-462, 2000.

Zhang, S., Xiong, W., Ni, W., & Li, X. "Value of big data to finance: observations on an internet credit Service Company in China." Financial Innovation, 1(1), 2015.