

Comparative Analysis of Random Forest Classification Over SVM Classifier to Detect Cyber Thefts in Credit Card to Reduce False Rate

K.R. Ruthvik¹; Dr.G. Charlyn^{2*}

¹Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.

¹ruthvik.work@gmail.com

^{2*}Project Guide, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.

^{2*}charlynpushpalathag.sse@saveetha.com

Abstract

Aim: To reduce the false rate of cyber thefts in credit card attacks based on binary selection Random Forest classifier and SVM classifier. **Materials and Methods:** Classification is performed by Random forest classifier (N=28) over SVM classifier (N=28) is for false rate detection. **Results and Discussion:** The values obtained in terms of accuracy is identified by random state in Random forest (94.4%) over SVM (91.4%) **Conclusion:** The reduction of false rate with sigma value 0.126 appears to be better in Random Forest classifier than SVM classifier.

Key-words: False Rate Reduction, Efficient Approach, Machine Learning, Novel Detection for Cyber Threats, Unique Approach for False Rate Reduction.

1. Introduction

Credit card fraud detection is to identify the unauthorised or fraudulent transactions to acquire money or money equivalent commodity this is one of the advanced methods of money laundering which is also known as Undisclosed Aggregation Factoring (UAF) (Awoyemi, Adetunmbi, and Oluwadare 2017). The importance of this research is to find suspicious transactions and intimate to the investigator simultaneously allowing non fraudulent transactions to be carried out without any delay in the system. It is also important to prevent fraudsters from getting a chance of performing the similar frauds and make multiple transactions on counterfeit cards (Sherly and Nedunchezian 2015).

This research can be further implicated on debit cards (Borzykowski 2013). It can also be further implemented on E - wallets (Levitin, n.d.).

In the era of the internet, there is a drastic increase in online shopping as it provides convenience to human life, this also led to drastic increase in use of the credit cards online. (Awoyemi, Adetunmbi, and Oluwadare 2017) As much as it provides convince it also led to increase of credit card frauds (Thennakoon et al. 2019) in past few years and it has caused huge problems to payment gateways, banks and sometimes even customers (Benchaji, Douzi, and ElOuahidi 2018). This can be stopped using machine learning techniques and incorporated in payment gateways like UNO pay, pay pal, linkway (Shabad, Rahim Shabad, and Kavitha 2018). Using these detection methods the credit card frauds can be detected and stopped (Wang et al. 2018).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Sathish and Karthick 2020; Varghese, Ramesh, and Veeraiyan 2019; S. R. Samuel, Acharya, and Rao 2020; Venu, Raju, and Subramani 2019; M. S. Samuel et al. 2019; Venu, Subramani, and Raju 2019; Mehta et al. 2019; Sharma et al. 2019; Malli Sureshbabu et al. 2019; Krishnaswamy et al. 2020; Muthukrishnan et al. 2020; Gheena and Ezhilarasan 2019; Vignesh et al. 2019; Ke et al. 2019; Vijayakumar Jain et al. 2019; Jose, Ajitha, and Subbaiyan 2020). Now the growing trend in this area motivated us to pursue this project.

The existing research uses highly imbalanced data as its input data set. The positive classes are highly oversampled. The number of tuples obtained in the dataset after cleaning of the dataset is very low. Hence the size of the input dataset is very small. The accuracy obtained by the existing research for some techniques are very low. The authors are expertise in the field of Machine Learning to conduct present research with peer reviewed publication as supporting documents. This research aims to increase the size of the input dataset and also to improve the accuracy of the algorithms.

2. Materials and Methods

The study and study setting was done in university the implementation, the execution, the error rectification was done in university. This study does not require any human samples or humanly data. Also this research does not require any ethical permission. The number of groups mentioned in this research are 2 where the first group is considered to be Random forest and the second group to be considered as SVM. The sample size of this work is considered to be 28. The study set taken for this research is from Experia.com (Awoyemi, Adetunmbi, and Oluwadare 2017).

Random Forest is supervised machine learning which can be used for both Classifier and Regression. But it is mainly used for classification in the majority of cases. This algorithm makes decision trees on the input data given and then delivers its prediction. In the final phase it chooses the best pattern in the means of voting. (V 2019; Benchaji, Douzi, and ElOuahidi 2018). **Support Vector Machines (SVM)** is a supervised and powerful machine learning algorithm yet a flexible algorithm. SVM is basically a multiple dimensional representation of various classes present in a hyperplane. This hyperplane will be created on the basis of iterative manner. This procedure is done by the SVM algorithm to minimise the errors. The datasets are divided into classes to find maximum marginal hyperplane. (Zhang, Bhandari, and Black 2020).

The independent data are the attributes like id, name, url, status_count and others and the dependent variables are update, transaction_class which consists of binary data. This study uses t test as its testing element. A t-test is a type of inferential static used to determine if there is a significant difference between the means of two groups, which may be related in certain features. The t-test is one of many tests used for the purpose of hypothesis testing in statistics. The statistical tool used for this study is IBM SPSS version 21. The independent data are the attributes like id, name, url, status_count and others, Attributes like update, transaction_class are considered as dependent variables for this study. In SPSS, the datasets are prepared using 28 as sample size for both the classifiers Random Forest and SVM. Groupid is given as 1 for Random Forest and 2 for SVM, the Groupid is given as a grouping variable and accuracy is given as a testing variable.

3. Results

Table 1- Group, Accuracy and Loss Value Calculation for Random Forest and SVM Classifiers

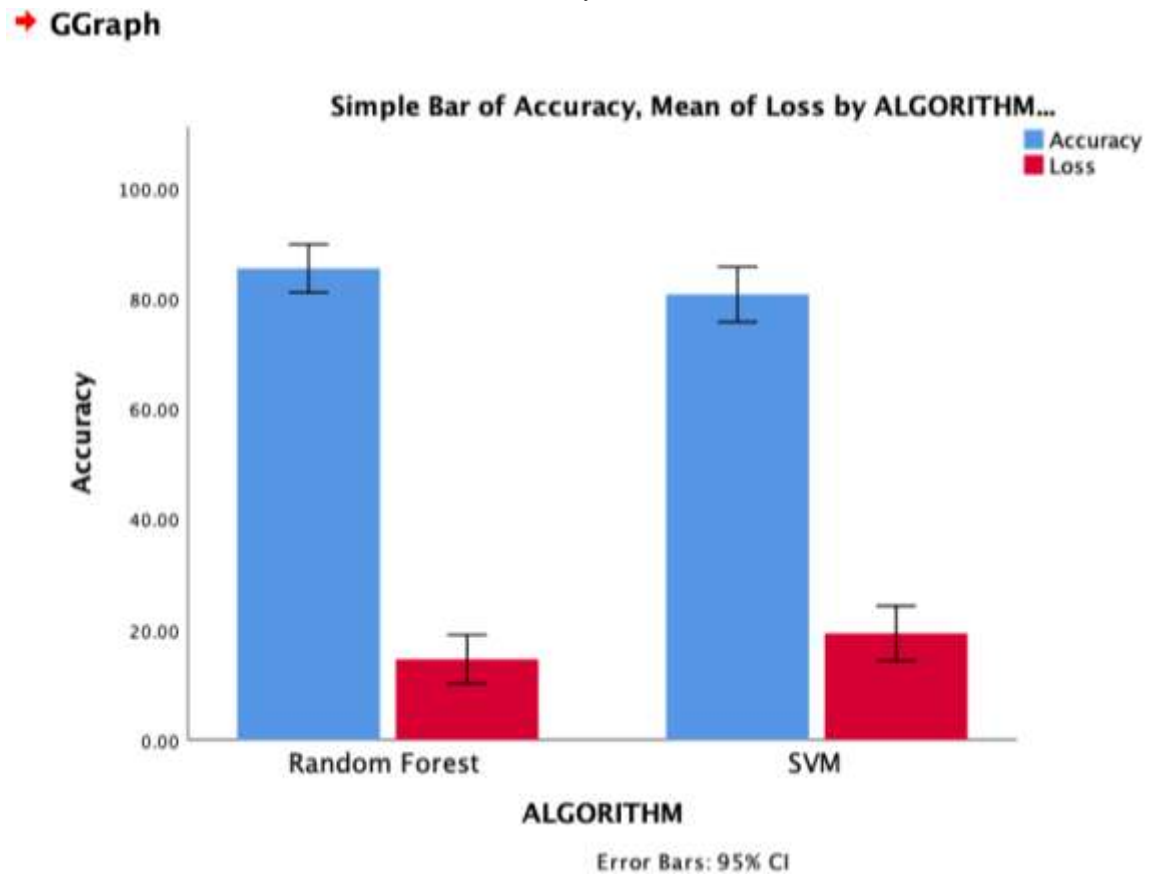
	Name	Type	Width	Decimals	Columns	Measure	Role
1	Group	Numeric	8	0	30	Nominal	Input
2	Accuracy	Numeric	8	4	30	Scale	Input
3	Loss	Numeric	8	2	30	Scale	Input

In that statistical tool the sample size used is 28. This data is used for analysis of Random forest and SVM. These 28 data samples used for each algorithm along with their loss are also used to calculate statistical values that can be used for comparison. From Table 1, it is inferred that the group, accuracy and the loss values for the two algorithms Random Forest classifier and SVM classifier are denoted. The group statistics table shows the number of samples that are collected and the mean and standard deviation obtained for the accuracies are entered.

Table 2- Mean, Standard Deviation and Standard Error Mean Calculation

	Group	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	1	Random Forest	28	85.4240	6.12781	1.93778
	2	SVM	28	80.7070	6.9692	2.20630
Loss	1	Random Forest	28	14.5860	6.14436	1.94302
	2	SVM	28	19.2930	6.97692	2.20630

Fig. 1- Comparison of Random Forest and SVM Classifier in Terms of Mean Accuracy. The Mean Accuracy of Random Forest is better than SVM the Standard Deviation of Random Forest is Slightly better than SVM X Axis: Random Forest vs SVM Classifier Y Axis: Mean Accuracy of Detection \pm 1 SD



From Table 2, the group statistics values along with the mean, standard deviation and the standard error mean for the two algorithms are also specified. The Independent sample T test is applied for the data set fixing confidence interval as 95%. The following image represents the graph that depicts the comparison between Random Forest and SVM based on their accuracy. Table 3 shows the independent t sample test for the algorithms. The comparative accuracy analysis, mean of loss between the two algorithms are specified. Fig. 1 shows the comparison of mean of accuracy and mean loss between Random Forest Classifier and SVM Classifier.

4. Discussion

In the given study, the accuracy analysis of Random forest classifier is 94.4% when compared to the accuracy of SVM classifier which is 91.4%. For the given sample size 28, statistical data analysis is done for both the prescribed algorithms namely Random Forest Classifier and SVM classifier. The group and the accuracy values are being calculated for the given classifiers. The mean, standard deviation and the standard mean values for the Random Forest classifier algorithms are 85.4240, 6.12781 and 1.93778 respectively. (Sherly and Nedunchezian 2015) Similarly for SVM classifier, the mean, standard deviation and the standard mean values are 80.7070, 6.9692 and 2.206030 respectively. (Thennakoon et al. 2019).

Table 3- Independent t Test Sample Calculation

Levene's Test for Equality of variances			T-test for Equality of means				95% confidence of the Difference			
		F	Sig.	t	df	Sig.(2-tailed)	Mean difference	Std. Error Diffencener	Lower	Upper
Accuracy	Equal variances assumed	.091	.840	1.606	18	.126	4.7170	2.93645	-1.45226	10.88626
	Equal variances not assumed			1.606	17.705	.126	4.7170	2.93645	-1.45963	10.89363
Loss	Equal variances assumed	.086	.840	-1.601	18	.127	-4.70700	2.93991	-10.88352	1.46952
	Equal variances not assumed			-1.601	17.717	.127	-4.70700	2.93991	-10.89060	1.47660

The independent sample t test values applied for the data set fixing confidence interval as 95% is tabulated in Table 3. Figure 1 shows the comparative accuracy analysis between Random Forest and SVM classifier in which Random forest shows the accuracy of 94.4% and SVM classifier shows the accuracy of 91.4%. (Xuan et al. 2018).

Compared to the previous analysis of Random Forest classifier and SVM classifier for credit card detection, peers have obtained the classification accuracy of 90.2% for Random Forest Classifier

(Awoyemi, Adetunmbi, and Oluwadare 2017) and the accuracy for SVM is 89%(Wang et al. 2018) Our analysis shows higher accuracy than our peers.

The training process for this data takes a long duration of time. The execution time is also very high and depending on the random state the larger the data the more time of execution is required. The training data change can be made but it's a very time consuming process and with low low training data the accuracy is reduced.

Our institution is passionate about high quality evidence based research and has excelled in various fields ((Vijayashree Priyadharsini 2019; Ezhilarasan, Apoorva, and Ashok Vardhan 2019; Ramesh et al. 2018; Mathew et al. 2020; Sridharan et al. 2019; Pc, Marimuthu, and Devadoss 2018; Ramadurai et al. 2019). We hope this study adds to this rich legacy.

The application for this study is built on the basis of the training and testing process of both the algorithms Random forest and SVM. The structured based classification can be done for testing and training for both the Machine learning Techniques Random Forest and SVM. The data cleaning process can be further improved and time of execution can be decreased. The process of time consumption in training dataset can be decreased. Fraud detection can be used on debit cards and also E- wallets. Hence this novel detection for cyber threats is very much helpful in identifying an unique approach for false rate reduction.

5. Conclusion

The data for this work has been acquired from various resources for the usage in credit card fraud detection system. The Random Forest classifier and SVM classifier are used as the classifier for this whole study. The algorithm has exhibited accuracies that are mentioned as follows. The accuracy of Random Forest classifier with accuracy 94.4% and sigma value (0.126) appears to be better when compared to the accuracy of SVM with an accuracy of 91.4%

Declarations

Conflict of Interests

No conflict of interest in this manuscript.

Authors Contribution

Author KRR was involved in data collection, data analysis and manuscript writing. Author CPL was involved in conceptualization, data validation and critical reviews of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Infysec Solutions, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

References

- Awoyemi, J.O., Adetunmbi, A.O., & Oluwadare, S.A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *In International Conference on Computing Networking and Informatics (ICCNi)*, 1-9. <https://doi.org/10.1109/iccni.2017.8123782>
- Benchaji, I., Douzi, S., & El Ouahidi, B. (2018). Using genetic algorithm to improve classification of imbalanced datasets for credit card fraud detection. *In International Conference on Advanced Information Technology, Services and Systems*, 220-229. <https://doi.org/10.1109/csnet.2018.8602972>
- Borzykowski, B. (2013). *Debit Card Fraud Detection for Dummies (Custom)*. John Wiley & Sons.
- Ezhilarasan, D., Apoorva, V.S., & Ashok Vardhan, N. (2019). Syzygium cumini extract induced reactive oxygen species-mediated apoptosis in human oral squamous carcinoma cells. *Journal of Oral Pathology & Medicine, Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*, 48(2), 115-121.
- Gheena, S., & Ezhilarasan, D. (2019). Syringic acid triggers reactive oxygen species-mediated cytotoxicity in HepG2 cells. *Human & experimental toxicology*, 38(6), 694-702.

- Jose, J., & Subbaiyan, H. (2020). Different treatment modalities followed by dental practitioners for Ellis class 2 fracture—A questionnaire-based survey. *The Open Dentistry Journal*, 14(1), 59-65.
- Ke, Y., Al Aboody, M.S., Alturaiki, W., Alsagaby, S.A., Alfaiz, F.A., Veeraraghavan, V.P., & Mickymaray, S. (2019). Photosynthesized gold nanoparticles from *Catharanthus roseus* induces caspase-mediated apoptosis in cervical cancer cells (HeLa). *Artificial cells, nanomedicine, and biotechnology*, 47(1), 1938-1946.
- Krishnaswamy, H., Muthukrishnan, S., Thanikodi, S., Antony, G.A., & Venkatraman, V. (2020). Investigation of air conditioning temperature variation by modifying the structure of passenger car using computational fluid dynamics. *Thermal Science*, 24(1 Part B), 495-498.
- Levitin, A.J. (2016). Pandora's Digital Box: Digital Wallets and the Honor All Wallets Rules. Available at SSRN 2819146.
- Malli Sureshbabu, N., Selvarasu, K., Nandakumar, M., & Selvam, D. (2019). Concentrated growth factors as an ingenious biomaterial in regeneration of bony defects after periapical surgery: A report of two cases. *Case reports in dentistry*.
- Mathew, M.G., Samuel, S.R., Soni, A.J., & Roopa, K.B. (2020). Evaluation of adhesion of *Streptococcus mutans*, plaque accumulation on zirconia and stainless steel crowns, and surrounding gingival inflammation in primary molars: Randomized controlled trial. *Clinical oral investigations*, 24(9), 3275-3280. <https://link.springer.com/article/10.1007/s00784-020-03204-9>
- Mehta, M., Tewari, D., Gupta, G., Awasthi, R., Singh, H., Pandey, P., & Satija, S. (2019). Oligonucleotide therapy: an emerging focus area for drug delivery in chronic inflammatory respiratory diseases. *Chemico-biological interactions*, 308, 206-215.
- Muthukrishnan, S., Krishnaswamy, H., Thanikodi, S., Sundaresan, D., & Venkatraman, V. (2020). Support vector machine for modelling and simulation of Heat exchangers. *Thermal Science*, 24 (1 Part B), 499-503.
- PC, J., Marimuthu, T., Devadoss, P., & Kumar, S.M. (2018). Prevalence and measurement of anterior loop of the mandibular canal using CBCT: A cross sectional study. *Clinical implant dentistry and related research*, 20(4), 531-534. <https://europepmc.org/article/med/29624863>
- Ramadurai, N., Gurunathan, D., Samuel, A.V., Subramanian, E., & Rodrigues, S.J. (2019). Effectiveness of 2% Articaine as an anesthetic agent in children: randomized controlled trial. *Clinical oral investigations*, 23(9), 3543-3550.
- Ramesh, A., Varghese, S., Jayakumar, N.D., & Malaiappan, S. (2018). Comparative estimation of sulfiredoxin levels between chronic periodontitis and healthy patients—A case-control study. *Journal of periodontology*, 89(10), 1241-1248.
- Samuel, M.S., Bhattacharya, J., Raj, S., Santhanam, N., Singh, H., & Singh, N.P. (2019). Efficient removal of Chromium (VI) from aqueous solution using chitosan grafted graphene oxide (CS-GO) nanocomposite. *International journal of biological macromolecules*, 121, 285-292.
- Samuel, S.R., Acharya, S., & Rao, J.C. (2020). School Interventions—based Prevention of Early-Childhood Caries among 3–5-year-old children from very low socioeconomic status: Two-year randomized trial. *Journal of public health dentistry*, 80(1), 51-60.
- Sathish, T., & Karthick, S. (2020). Wear behaviour analysis on aluminium alloy 7050 with reinforced SiC through taguchi approach. *Journal of Materials Research and Technology*, 9(3), 3481-3487.

- Shabad, M.A.R., & Kavitha, M. (2018). Credit Card Fraud Detection Using Neural Networks at Merchant Side. *Journal of Computational and Theoretical Nanoscience*, 15(11-12), 3373-3375. <https://doi.org/10.1166/jctn.2018.7628>
- Sharma, P., Mehta, M., Dhanjal, D.S., Kaur, S., Gupta, G., Singh, H., & Satija, S. (2019). Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. *Chemico-biological interactions*, 309, 108720.
- Sherly, K.K., & Nedunchezian, R. (2015). A improved incremental and interactive frequent pattern mining techniques for market basket analysis and fraud detection in distributed and parallel systems. *Indian Journal of Science and Technology*, 8(18), 1-12. <https://doi.org/10.17485/ijst/2015/v8i18/55109>.
- Sridharan, G., Ramani, P., Patankar, S., & Vijayaraghavan, R. (2019). Evaluation of salivary metabolomics in oral leukoplakia and oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine*, 48(4), 299-306.
- Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019, January). Real-time credit card fraud detection using machine learning. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 488-493. <https://doi.org/10.1109/confluence.2019.8776942>
- Varghese, S.S., Ramesh, A., & Veeraiyan, D.N. (2019). Blended Module-Based Teaching in Biostatistics and Research Methodology: A Retrospective Study with Postgraduate Dental Students. *Journal of dental education*, 83(4), 445-450.
- Venu, H., Raju, V.D., & Subramani, L. (2019). Combined effect of influence of nano additives, combustion chamber geometry and injection timing in a DI diesel engine fuelled with ternary (diesel-biodiesel-ethanol) blends. *Energy*, 174, 386-406.
- Venu, H., Subramani, L., & Raju, V.D. (2019). Emission reduction in a DI diesel engine using exhaust gas recirculation (EGR) of palm biodiesel blended with TiO₂ nano additives. *Renewable Energy*, 140, 245-263.
- Kumar, M.S., Soundarya, V., Kavitha, S., Keerthika, E.S., & Aswini, E. (2019). Credit card fraud detection using random forest algorithm. In *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*, 149-153. <https://doi.org/10.22214/ijraset.2019.3215>
- Vignesh, R., Ditto Sharmin, C., Annamalai, S., & Baghkomeh, P.N. (2019). Management of complicated crown-root fracture by extra-oral fragment reattachment and intentional reimplantation with 2 years review. *Contemporary clinical dentistry*, 10(2), 397-401.
- Jain, S.V., Muthusekhar, M.R., Baig, M.F., Senthilnathan, P., Loganathan, S., Wahab, P.A., & Vohra, Y. (2019). Evaluation of three-dimensional changes in pharyngeal airway following isolated lefort one osteotomy for the correction of vertical maxillary excess: a prospective study. *Journal of maxillofacial and oral surgery*, 18(1), 139-146.
- Vijayashree Priyadharsini, J. (2019). In silico validation of the non-antibiotic drugs acetaminophen and ibuprofen as antibacterial agents against red complex pathogens. *Journal of periodontology*, 90(12), 1441-1448.
- Wang, C., Wang, Y., Ye, Z., Yan, L., Cai, W., & Pan, S. (2018). Credit card fraud detection based on whale algorithm optimized BP neural network. In *13th International Conference on Computer Science & Education (ICCSE)*, 1-4. <https://doi.org/10.1109/iccse.2018.8468855>

Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018). Random forest for credit card fraud detection. *In IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 1-6. <https://doi.org/10.1109/icnsc.2018.8361343>

Zhang, D., Bhandari, B., & Black, D. (2020). Credit Card Fraud Detection Using Weighted Support Vector Machine. *Applied Mathematics*, *11*(12), 1275-1291. <https://doi.org/10.4236/am.2020.1112087>