ISSN: 0974-5823 Vol. 7 (Special Issue, Jan.-Mar. 2022)

International Journal of Mechanical Engineering

An Approach to Reduce False Positive Rate in Analysing Redflag Gap for Anti Money Laundering System using Recurrent Neural Network in Comparison with Support Vector Machine Algorithm

Lokesh R[1] , Dr. Rashmita Khilar[2]

[1]Research Scholar, Department of Information Technology,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Saveetha University,Chennai,TamilNadu. India Pincode: 602105.

[2]Project Guide,Corresponding Author, Department of Information Technology,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Saveetha University,Chennai,TamilNadu.India. Pincode: 602105.

ABSTRACT

Aim: To reduce false positive rate in analysing Anti Money Laundering systems using Neural Network Algorithm and Support Vector Machine Algorithm. **Materials and Methods**: Classification is performed by Recurrent Neural Network Algorithm of sample size (N=10) with accuracy 87.75% over Support vector machine of sample size (N=10) algorithm with accuracy 79.22% for analysing the Anti Money Laundering. Sample size is calculated using GPower with pretest power as 0.8 and alpha 0.05. **Results:** Mean accuracy identified by the Recurrent Neural network (87.75%) is high compared to Support Vector Machines (79.22%). Significance value for accuracy and loss is 0.365 ($p>0.05$). **Conclusion:** The mean accuracy of the Anti Money Laundering system in Neural Networks is better than the Support Vector Machine.

Keywords: Recurrent Neural Network Algorithm, Support Vector Machine, Machine Learning, Transparent Transaction, Anti Money laundering, Novel Detection

INTRODUCTION

Money laundering is the interaction by which a lot of wrongfully obtained cash from drug dealing, fear monger movement, or other genuine wrongdoings, is given the presence of having started from an authentic source. Suspicious interest reporting is part of the anti-cash laundering statutes and guidelines and its role is very critical inside the fight towards AML [1]. Anti-money laundering refers to the legal guidelines, rules and processes meant to prevent criminals from disguising illegally obtained budgets as legitimate earnings [2]. Its broader risk even though is to the economy itself and the civil society it serves. Financial crime is growing every day [3]. Therefore AML desires to be diagnosed and tracked that allows you to avoid unethical and illegal interest. Detecting AML additionally avoids human beings from funding terrorists and consequently reduces terrorist attacks [2]. AML procedures also operate as gatekeepers against general fraudulent activity [4].

In this research, the anti money laundering system, the databases are collected from banking sectors. This research work is carried out by many researchers and there are a number of articles published. There are 30 papers in the last 5 years, 18 articles on IEEE Explore and 12 on Researchgate [5]. The usage of secret key and security key is useful as it changes every time the user runs the program [6]. Hence the account is safe and the key is only generated during every run of the program hence giving the user account an impenetrable security. The use of a secret key and security key is done as nowadays the OTP (One Time Password) is hacked and is got by the third person easily hence it secures our account and ensures the user the safety of their bank [7] Collecting the bank database is a huge and challenging process and is a difficult process as well, because there are a lot of banks in India [8]. Also we have taken a lot of data into consideration such as Operating expenses, Slow processes, High commissions, Low stimulus to savings, ATM permanent network, Online or virtual banking, etc. Anti money laundering includes 5 steps to combat money laundering such as: Improve search with technology, Regular cross communication, Leverage data analytics to detect patterns, System standardization, Training. Previously our team has a rich experience in working on various research projects across multiple disciplines[9]–[19]

The existing system has a drawback which is not suitable for large datasets due to high training time. It takes more time to complete the process so the other transactions get delayed. The aim of this research is to explore and extend working skills towards money laundering systems to reduce its time for training data and decrease the false positive rate in analyzing the redflag gap.

MATERIALS AND METHODS

The study setting was done at Data Analytics Lab, Saveetha School of Engineering. The implementation, the execution, the error rectification was done in Saveetha School of Engineering. This study does not require human samples or human data. 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 Recurrent Neural Network and the second group to be considered as support vector machine [20]. In each group 10 sets of samples are used in total 20 samples are taken for this research, where the accuracy for Recurrent Neural Network is 87.75 and Support vector machine is 79.22. The sample size of this work is considered to be 10 for each group using Gpower 3.1 software with pretest power as 0.8 and alpha 0.05. The study set taken for this research is from GitHub [3], [21].

The dataset that can be used for real time prediction is collected from GitHub.com. The dataset has a total no. of 8 columns and 1575 rows with important variables such as minimum balance, maximum balance, bank ID, Business type, count. The data collected from here is put into a separate folder named anti money laundering [1] The obtained dataset is well qualified for evaluating the machine learning algorithm. Then the dataset is splitted into two parts with 25% of the data used for testing and called anti money laundering and the rest 75% will be used as training dataset. The Novel Recurrent Neural network algorithm will be implemented and the dataset will be processed for testing and training [22] followed by calculation of accuracy percentage. The pseudocode for neural networks is described in Table 1. The comparative algorithm which is the support vector machine algorithm will be implemented and the dataset will be going through the process of testing and training, after which the accuracy percentage for the algorithm is obtained [23] . The pseudocode Support vector machine is described in Table 2.

The proposed work is designed and implemented with the help of Python OpenCV software. The platform to assess deep learning was Windows 10 OS. Hardware configuration was an Intel core i5 processor with a RAM size of 4GB. System sort used was 64-bit. For implementation of code, java programming language was used. As for code execution, the dataset is worked behind to perform an output process for accuracy.

STATISTICAL ANALYSIS

The independent datas are the attributes like minimum balance, maximum balance, bank ID and the dependent variables are Business type, count. This study uses the T-test as its testing elements. The T-test is a type of inferential static used to determine if there is a significant difference between the mean of two groups, which may be related in certain features. The Statistical T-test is one of many tests used for hypothesis testing in statistics. The statistical tool used for the study is IBM SPSS version 21.

RESULTS

In the statistical tool the total sample size used is 10.This data is used for analysis of Recurrent Neural Network and Support Vector Machine.These 10 data samples used for each algorithm along with their loss are also used to calculate statistical values that can be used for comparison. From Table1, it is inferred that the group, accuracy and the loss values for the two algorithms Recurrent Neural Network and Support Vector Machine algorithm 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.

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%. Table 3 consists of criteria of the independent sample t test.Table4 shows the values of the independent t-test sample 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 Recurrent Neural Network algorithm and Support Vector Machine algorithm. This study indicates that the Neural Network algorithm is a better algorithm with 87.75% accuracy percent when compared to the Support Vector Machine algorithm with 79.22% accuracy percent.

DISCUSSION

In the given study, the significance value obtained is (P=0.365) which implies that the Recurrent Neural Network algorithm appears to be better than the Support Vector Machine algorithm. Due to insufficient dataset the g power value is less than specified. The accuracy analysis of the Neural Network algorithm is 87.75% when compared to the accuracy of the Support Vector Machine algorithm which is 79.22%. For the given sample size 25. The mean, standard deviation and the standard mean values for Neural Network algorithms algorithm are 87.7500, 6.43657, 2.03542. This clearly indicates that Neural Network is a better algorithm compared to Support Vector Machine algorithm. Support Vector Machine algorithms are applied and the mean, the standard deviation and the standard mean are 79.2200, 5.10486, 1.61430 [24].

This paper shows the comparative accuracy analysis between Neural Network and Support Vector Machine algorithm in which Recurrent Neural Network shows the accuracy of 87.75% and Support Vector Machine algorithm shows the accuracy of 79.22% [25]. Compared to the previous analysis of detection than other algorithms and Support Vector Machine Identification, peers have obtained the classification accuracy of 87.75% for Neural Networks algorithm for anti money laundering of Neural Network algorithms. Recurrent Neural Network technique shows better performance algorithms and the accuracy for Support Vector Machine algorithm is 79.22%. Support Vector Machine is used for identification of Black money and other finding techniques. Performance is improved in the Neural Network algorithm [26]. Our analysis shows higher accuracy than our peers. Accuray difference is shown and performance is calculated for fraud detection using many other techniques.

Support vector machine have a drawback which is difficult to understand and interpret the final model, it does not perform well for large dataset leads to overlap. It requires more time for training. The anti money laundering system can be applied to software used in the finance and legal industries to meet the legal requirements for financial institutions and other regulated entities to prevent or report abnormalities in money laundering activities.

CONCLUSION

From this study of Anti money laundering , the accuracy of support vector machines is 79.22% whereas the Neural network has a higher accuracy of 87.75%. Hence it is inferred that Neural networks appear to be better in accuracy when compared to support vector machines.

DECLARATION

Conflict of interests

No conflict of interest in this manuscript.

Authors Contribution

Author RL was involved in data collection, data analysis and manuscript writing. Author RK 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

REFERENCE

[1] Z. Luo, "A new SVM algorithm and selected application in sequence analysis," *Future Information Engineering*. 2014. doi: 10.2495/icie130892.

[2] M. Ziemski, T. Wisanwanichthan, N. A. Bokulich, and B. D. Kaehler, "Beating Naive Bayes at Taxonomic Classification of 16S rRNA Gene Sequences," *Front. Microbiol.*, vol. 12, p. 644487, Jun. 2021.

[3] C.-W. Zhang and Y.-B. Wang, "Research on application of distributed data mining in anti-money laundering monitoring system," *2010 2nd International Conference on Advanced Computer Control*. 2010. doi: 10.1109/icacc.2010.5487272.

[4] D. Chau and M. van Dijck Nemcsik, *Anti-Money Laundering Transaction Monitoring Systems Implementation: Finding Anomalies*. John Wiley & Sons, 2020.

[5] Y. Hou, "A Novel SVM Algorithm and Experiment," *2012 International Conference on Computer Science and Electronics Engineering*. 2012. doi: 10.1109/iccsee.2012.121.

[6] A. U. Bello, *Improving Anti-Money Laundering Compliance: Self-Protecting Theory and Money Laundering Reporting Officers*. Springer, 2017.

[7] J. Tsiligaridis, "Classification with neural network and SVM via decision tree algorithm." 2018. doi: 10.1063/1.5045446.

[8] M. Levi, "Making sense of professional enablers' involvement in laundering organized crime proceeds and of their regulation," *Trends Organ Crime*, pp. 1–15, Nov. 2020.

[9] D. Ezhilarasan, T. Lakshmi, M. Subha, V. Deepak Nallasamy, and S. Raghunandhakumar, "The ambiguous role of sirtuins in head and neck squamous cell carcinoma," *Oral Dis.*, Feb. 2021, doi: 10.1111/odi.13798.

[10] R. Balachandar *et al.*, "Enriched pressmud vermicompost production with green manure plants using Eudrilus eugeniae," *Bioresour. Technol.*, vol. 299, p. 122578, Mar. 2020.

[11] S. Muthukrishnan, H. Krishnaswamy, S. Thanikodi, D. Sundaresan, and V. Venkatraman, "Support vector machine for modelling and simulation of heat exchangers," *Therm. Sci.*, vol. 24, no. 1 Part B, pp. 499–503, 2020.

[12] A. Kavarthapu and K. Gurumoorthy, "Linking chronic periodontitis and oral cancer: A review," *Oral Oncol.*, p. 105375, Jun. 2021.

[13] S. C. Sarode, S. Gondivkar, G. S. Sarode, A. Gadbail, and M. Yuwanati, "Hybrid oral potentially malignant disorder: A neglected fact in oral submucous fibrosis," *Oral Oncol.*, p. 105390, Jun. 2021.

[14] Hannah R, P. Ramani, WM Tilakaratne, G. Sukumaran, A. Ramasubramanian, and R. P. Krishnan, "Author response for 'Critical appraisal of different triggering pathways for the pathobiology of pemphigus vulgaris—A review.'" Wiley, May 07, 2021. doi: 10.1111/odi.13937/v2/response1.

[15] D. Sekar, D. Nallaswamy, and G. Lakshmanan, "Decoding the functional role of long noncoding RNAs (lncRNAs) in hypertension progression," *Hypertension research: official journal of the Japanese Society of Hypertension*, vol. 43, no. 7. pp. 724–725, Jul. 2020.

[16] P. Appavu, V. Ramanan M, J. Jayaraman, and H. Venu, "NOx emission reduction techniques in biodiesel-fuelled CI engine: a review," *Australian Journal of Mechanical Engineering*, vol. 19, no. 2, pp. 210–220, Mar. 2021.

[17] S. Menon, H. Agarwal, S. Rajeshkumar, P. Jacquline Rosy, and V. K. Shanmugam, "Investigating the Antimicrobial Activities of the Biosynthesized Selenium Nanoparticles and Its Statistical Analysis," *Bionanoscience*, vol. 10, no. 1, pp. 122–135, Mar. 2020.

[18] R. Gopalakrishnan, V. M. Sounthararajan, A. Mohan, and M. Tholkapiyan, "The strength and durability of fly ash and quarry dust light weight foam concrete," *Materials Today: Proceedings*, vol. 22, pp. 1117–1124, Jan. 2020.

[19] V. R. Arun Prakash, J. F. Xavier, G. Ramesh, T. Maridurai, K. S. Kumar, and R. B. S. Raj, "Mechanical, thermal and fatigue behaviour of surface-treated novel Caryota urens fibre–reinforced epoxy composite," *Biomass Conversion and Biorefinery*, Aug. 2020, doi: 10.1007/s13399-020-00938-0.

[20] M. Turki, A. Hamdan, R. T. Cummings, A. Sarea, M. Karolak, and M. Anasweh, “The regulatory technology ‘RegTech’ and money laundering prevention in Islamic and conventional banking industry,” *Heliyon*, vol. 6, no. 10, p. e04949, Oct. 2020.

[21] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*. Springer, 2018.

[22] N. A. L. Khac, N. A. Le Khac, and M.-T. Kechadi, “Application of Data Mining for Anti-money Laundering Detection: A Case Study,” *2010 IEEE International Conference on Data Mining Workshops*. 2010. doi: 10.1109/icdmw.2010.66.

[23] K. Gurney, *An Introduction to Neural Networks*. CRC Press, 2018.

[24] A. I. Canhoto, “Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective,” *J. Bus. Res.*, Oct. 2020, doi: 10.1016/j.jbusres.2020.10.012.

[25] H. D. Lodge, “Blackstone’s Guide to The Sanctions and Anti-Money Laundering Act.” 2020. doi: 10.1093/oso/9780198844778.001.0001.

[26] K. Benson, “Money laundering and the anti-money laundering regime,” *Lawyers and the Proceeds of Crime*. pp. 15–31, 2020. doi: 10.4324/9781315179735-2.

TABLES AND FIGURES

Table 1. Pseudocode for Anti Money Laundering System based on Neural Network

INPUT: Training dataset for Anti Money Laundering
Step 1: Read the test data for Anti Money Laundering. Step 2: Extract the Anti Money Laundering attributes. Step 3: Extract the Anti Money Laundering data. Step 4: Comparison of Secret Key and Security Key. Step 5: Apply Neural Network Algorithm. Step 6: Formula for Neural Network Algorithm $f(b + \sum_{i=1}^{n} x_i w_i)$ Step 7: Forecast fraud detection. Step 8: Return accuracy.
OUTPUT: Classified as Genuine or Fraudulent

Table 2. Pseudocode for Anti Money Laundering System based on SVM classifier

INPUT: Training dataset for Anti Money Laundering
Step 1: Read the test data for Anti Money Laundering. Step 2: Extract the Anti Money Laundering attributes. Step 3: Extract the Anti Money Laundering data. Step 4: Comparison of Secret Key and Security Key. Step 5: Apply SVM classifier Algorithm. Step 6: Formula for SVM classifier Algorithm $\sum_{i=1}^{n} \quad c_i - \frac{1}{2}\sum_{i=1}^{n} \quad \sum_{j=1}^{n} \quad y_i c_i k(x_i, x_j) y_j c_j$ Step 7:To find optimal boundaries between the possible outputs. Step 8: Return accuracy.
OUTPUT: Classified as Genuine or Fraudulent

Table 3. Group, Accuracy and Loss value uses 31 columns with 8 width data for Anti Money Laundering System

S.Nos.	**Name**	**Type**	**Width**	**Decimal**	**Columns**	**Measure**	**Role**
1	Group	Numeric	8	0	27	Nominal	Input
2	Accuracy	Numeric	8	4	27	Scale	Input
3	Loss	Numeric	8	2	27	Scale	Input

Table 4. Group Statistical analysis for neural network and support vector machine Mean, Standard Deviation and standard error mean are determined

	Group	**Algorithm**	**N**	**Mean**	**Std Deviation**	**Std.Error Mean**
Accuracy	1	Neural Network	10	87.7500	6.43657	2.03542
	2	Support Vector Machine	10	79.2200	5.10486	1.61430
Loss	1	Neural Network	10	12.2500	6.43657	2.03542
	2	Support Vector Machine	10	20.7800	5.10486	1.61430

Table 5. Independent sample T-test t is performed on two groups for significance and standard error determination. p value is greater than 0.05 (0.365) and it is considered to be statistically insignificant with 95% confidence interval

		Levene's test for Equality of variance		T-Test for equality of mean						
		F	Sig	t	df	Sig(2 - tailed)	Mean difference	Std.Error Difference	95% confidence of Difference: Lower	95% confidence of Difference: Upper
Accuracy	Equal variances assumed	0.863	0.365	3.283	18	0.004	8.53000	2.59786	3.07209	13.98791
	Equal Variances not assumed			3.283	17.112	0.004	8.53000	2.59786	3.05173	14.00827
Loss	Equal variances assumed	0.863	0.365	-3.283	18	0.004	-8.53000	2.59786	-13.98791	-3.07209
	Equal Variances not assumed			-3.283	17.112	0.004	-8.53000	2.59786	-14.00827	-3.05173

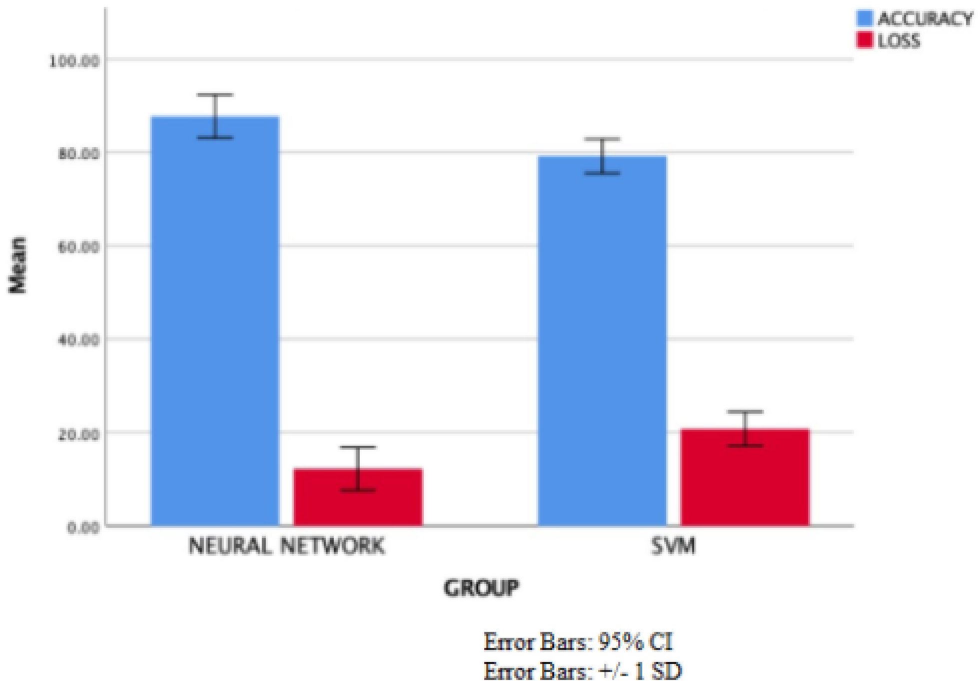

Fig. 1 Comparison of Neural Network and SVM in terms of mean accuracy and mean loss. The mean accuracy of the Neural Network is better than SVM. The standard deviation of Neural Network is slightly better than SVM X Axis: Backpropagation neural network vs SVM Y Axis: Mean accuracy of detection ± 1 SD.