



COMPARING THE ACCURACY IN CREDIT CARD FRAUD DETECTION USING XGBOOST COMPARING WITH NOVEL ADABOOST

B.Sri Sai Chowdary¹, J. Chennai Kumaran^{2*}

Article History: Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

Abstract

Aim: The main aim of the research is to detect Credit Card Fraud using XGBoost compared with the Novel AdaBoost.

Materials and Methods: When implementing an accurate prediction model it might not be sufficient to just consider one or two parameters. This analysis will be fed to the prediction model. Following XGBoost algorithm novel AdaBoost algorithm based on the previously collected datasets. Can predict the upcoming credit card fraud with calculations.

Result: Comparison is done by using SPSS Software. The XGBoost algorithm produces 93.89% whereas AdaBoost algorithm produces 94.38% accuracy while detecting credit card fraud on a data set ($p > 0.05$). Hence AdaBoost is better than XGBoost.

Conclusion: After using iterations to get that by using XGBoost algorithm get 93.89% (0.89) and novel AdaBoost algorithm get 94.38% (0.94). So can say that By using the novel AdaBoost algorithm to get more Accuracy than the XGBoost Algorithm.

Keywords: XGBoost, Novel AdaBoost, Credit Card Fraud Detection, Mean Accuracy, Predictions, Performance.

¹Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and technical Science, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

^{2*}Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

1. Introduction

Credit card fraud is concerned with the illegal use of credit card information for purchases. Credit card transactions can be accomplished either physically or digitally (Marabad 2021). Since it is installed in a transactional hub for operation distribution, and not on a card-issuing institution, it acts solely on the information of the operation to be rated and of its immediate previous history, and not on historic databases of past cardholder activities (Nandi et al. 2022). Credit Card Fraud Detection also has two other highly peculiar characteristics. The first one is obviously the very limited time span in which the acceptance or rejection decision has to be made (Dorransoro et al. 1997). Credit card fraud is a growing problem in the credit card industry (Ao et al. 2022). Data prevention is where the protection layer is created to prevent outsiders from attacking them. It first attempts to prevent fraud (Zhang, Bhandari, and Black 2020). Understanding Problem-solving data sets revealed many solutions over the years. Most general proposed answers fall into three broad collections: data layer, algo, & synthesis solutions (AI-Nuemat 2017).

In the last five years, Google scholar identified almost 459 research articles on Credit Card Fraud Detection using machine learning using XGBoost compared with AdaBoost. As credit card transactions become the most prevailing mode of payment for both online and offline transactions, credit card fraud rate also accelerates (Baesens, Verbeke, and Van Vlasselaer 2015). Check Fraud occurs when a person forges a check or pays for something with a check knowing that there is not enough money. Internet sales is fraud where fraudsters sell fake items or counterfeit items, or take payment without delivering the item (IEEE Staff 2017). Presents coupon personalized placement recommendations, which use XGBoost algo to calculate whether users are using coupons for specified age of time (Pumsirir: Yan 2018). As a result, the concept of classification using multiple classifiers has become one of the most significant methodologies to improve the classification performance. All classifiers provide their predictions of the class of an incoming data sample, and these predictions are analyzed and combined using some fusion strategy (Nandi et al. 2022). It not only endangers the property security of users, but also hinders the development of digital finance in the world (McCarthy et al. 2019). The SMOTE approach was employed to oversample the dataset because it was severely unbalanced (Marabad 2021). Our team has extensive knowledge and research experience that has translated into high quality publications (Pandiyan et al. 2022;

Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The research gap identified from the existing system shows poor accuracy. The study is to improve the accuracy of Classification (Webster and Ball 2018). by incorporating XGBoost and comparing performance with the novel AdaBoost. The proposed model improves classifiers to achieve more accuracy for Credit Card Fraud Detection.

2. Materials and Methods

This study setting was done in the Soft Computing Laboratory, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The number of required samples in research are two in which group 1 is Logical Regression compared with group 2 of SVM Algorithm. The samples were taken from the device and iterated 10 times to get desired accuracy with G power. A dataset consisting of a collection of stocks was downloaded from Kaggle repository (salihfurkansaglam 2021)

XGBoost

XGBoost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost implements gradient-boosted decision trees designed for speed and performance.

Algorithm for XGBoost

Stage 1: Trouble the instructive records
Stage 2: Import the dataset and eliminate the X variables and Y autonomously.
Stage 3: Separation of the dataset into two sections
Stage 4: Introducing the Information Testing in XGBoost model
 $xgb = XGBClassifier(max_depth = 4)$
 $xgb.fit(X_train, y_train)$
 $xgb_yhat = xgb.predict(X_test)$
Stage 5: Proceed with the means
Stage 6: Concocting expectations
Stage 7: Assessing model's performance

Novel AdaBoost

Novel AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances.

Algorithm for Novel AdaBoost

Stage 1: Burden the significant libraries
Stage 2: Import dataset and remove the X factors and Y independently.
Stage 3: Gap the dataset into train and test
Stage 4: Instating the AdaBoost model
For AdaBoost, Each instance in the training dataset is weighted. Initial weight is set to: $Weight(x_i) = (1/n)$
Where, x_i – ith training instance,
 n – Number of training instance
Stage 5: Fitting the AdaBoost model
Stage 6: Thinking of forecasts
Stage 7: Evaluating model's display

Recall that the testing setup includes both hardware and software configuration choices. The computational environment has an Intel Core i5 5th generation CPU with 8GB of RAM, an x86-based processor, a 64-bit operating system, and a hard drive and it runs on Windows X environment.

Statistical Analysis

SPSS is a software tool used for statistics analysis. The proposed system utilized 10 iterations for each group with predicted accuracy noted and analyzed. Independent samples t-test was done to obtain significance between two groups.

3. Results

Table 1 shows the accuracy value of the iteration of XGBoost and novel AdaBoost. Table 2 represents the Group statistics results which depicts XGBoost with mean accuracy of 93.89%, and standard deviation is 1.377. Novel AdaBoost has a mean accuracy of 94.389% and standard deviation is 1.55. Proposed novel AdaBoost algorithm provides better performance compared to the XGBoost algorithm. Table 3 shows the independent samples T-test value for XGBoost and novel AdaBoost with mean difference as -0.49, std. error difference as 0.656. Significance value is observed as 0.461 ($p > 0.05$).

Figure 1 shows the bar graph comparison of mean of accuracy on XGBoost and novel AdaBoost algorithm. Mean accuracy of XGBoost is 93.89% and novel AdaBoost is 94.38%.

4. Discussion

In this study the iteration based on the previous historical datasets are considered for getting the accurate value Credit Card Fraud Detection of Using XGBoost Algorithm and novel AdaBoost algorithm on the basis of anaconda navigator(Jupituer notebook) and SPSS.After Getting many iterations to compare all the Selective algorithms are verified and decide to get

The accurate values of Credit Card Fraud Detection.

The implementation of the machine became to learn about Credit Card Fraud Detection. The have a look at makes a speciality of the Credit card Frauds datasets obtained from diverse portals belonging to some districts of Karnataka inside the country. Datasets ordered in a properly based totally way. The XGBoost set of guidelines is used for the prediction model and its accuracy is received. The destiny is outstanding for the implementation of system studying algorithms inside the vicinity of Credit Card Fraud Detection and are hoping to put in force greater advanced algorithms in order that the system becomes more efficient. Hoping to make tool prediction extra sturdy and attain excessive accuracy with the assistance of greater datasets and advanced algorithms.

5. Conclusion

In this study, Credit Card Fraud Detection using the novel AdaBoost algorithm provides better accuracy than XGboost algorithm.

Declaration

Conflict of Interests

No conflict of interests in this manuscript

Authors Contribution

Author B.Sri Sai Chowdary was involved in data collection, data analysis, manuscript writing. Author J.Chenni Kumaran was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgement

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: Thanks for the following organizations for providing financial support that enabled us to complete the study.

1. Qbec Infosol
2. Saveetha University
3. Saveetha Institute of Medical and Technical sciences
4. Saveetha School of Engineering

6. Reference

Al-Nuemat, Ahmed Adnan. 2017. "The Legal System for Credit Cards in Jordan." *Beijing Law Review* 8 (2): 171–90.

- Ao, Guangyu, Jing Li, Yang Yuan, Yushu Wang, Basma Nasr, Mulong Bao, Ming Gao, and Xin Qi. 2022. "Intravenous Vitamin C Use and Risk of Severity and Mortality in COVID-19: A Systematic Review and Meta-Analysis." *Nutrition in Clinical Practice: Official Publication of the American Society for Parenteral and Enteral Nutrition*, February. <https://doi.org/10.1002/ncp.10832>.
- Baesens, Bart, Wouter Verbeke, and Veronique Van Vlasselaer. 2015. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. John Wiley & Sons.
- Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2022.134596>.
- Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresour Technol*. <https://doi.org/10.1016/j.biortech.2022.127234>.
- Dorronsoro, J. R., F. Ginel, C. Sgnchez, and C. S. Cruz. 1997. "Neural Fraud Detection in Credit Card Operations." *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council* 8 (4): 827–34.
- IEEE Staff. 2017. *2017 International Conference on Computing Networking and Informatics (ICCNi)*.
- Karpagam, M., R. Beulah Jeyavathana, Sathiy Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. <https://doi.org/10.1007/s00500-022-06931-1>.
- Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. <https://doi.org/10.1007/s10973-022-11298-4>.
- Marabad, Sanmati. 2021. "CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING." *Asian Journal For Convergence In Technology (AJCT) ISSN - 2350-1146* 7 (2): 121–27.
- McCarthy, Richard V., Mary M. McCarthy, Wendy Ceccucci, and Leila Halawi. 2019. *Applying Predictive Analytics: Finding Value in Data*. Springer.
- Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.
- Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. <https://doi.org/10.1007/s12633-022-01825-1>.
- Nandi, Asoke K., Kuldeep Kaur Randhawa, Hong Siang Chua, Manjeevan Seera, and Chee Peng Lim. 2022. "Credit Card Fraud Detection Using a Hierarchical Behavior-Knowledge Space Model." *PLoS One* 17 (1): e0260579.
- Pandiyani, P., R. Sitharthan, S. Saravanan, Natarajan Prabakaran, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. <https://doi.org/10.1016/j.seta.2022.102155>.
- Pumsirirat, Apapan, and Liu Yan. 2018. "Credit Card Fraud Detection Using Deep Learning Based on Auto-Encoder and Restricted Boltzmann Machine." *International Journal of Advanced Computer Science and Applications (ijacsa)* 9 (1). <https://doi.org/10.14569/IJACSA.2018.090103>.
- salihfurkansaglam. 2021. "FraudDetection-Accuracy%99 Recall%99 Precision%99." Kaggle. December 24, 2021. <https://kaggle.com/salihfurkansaglam/frauddetction-accuracy-99-recall-99-precision-99>.
- Venu, Harish, Ibbam Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*.

<https://doi.org/10.1016/j.energy.2022.123806>

Webster, William, and Kirstie Ball. 2018. *Surveillance and Democracy in Europe: Courting Controversy?* Routledge.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn

Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Zhang, Dongfang, Basu Bhandari, and Dennis Black. 2020. "Credit Card Fraud Detection Using Weighted Support Vector Machine." *Applied Mathematics* 11 (12): 1275–91.

Tables and Figures

Table 1. Accuracy Values for XGBoost and novel AdaBoost

S.NO	XGBoost	AdaBoost
1	96.00	97.00
2	94.20	93.02
3	92.10	94.80
4	95.10	96.03
5	92.20	92.88
6	93.89	93.60
7	92.80	93.00
8	94.00	94.77
9	95.65	96.00
10	93.05	92.79

Table 2. Group Statistics Results - novel AdaBoost has an mean accuracy (94.38%), std.deviation (1.55), whereas for XGBoost has mean accuracy (93.89%), std.deviation (1.37).

Group Statistics					
	Groups	N	Mean	Std deviation	Std. Error Mean
Accuracy	XGBoost	10	93.899	1.37740	.43557
	AdaBoost	10	94.389	1.55204	.49080

Table 3. Independent Samples T-test -novel Adaboost seems to be significantly better than XGboost (p=0.99)

Accuracy	Independent Samples Test								
	Levene's Test for Equality of Variances					T-test for Equality of Means			
	F	Sig	t	df	Sig(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.567	.461	-.747	18	.465	-.49000	.65621	-1.86864	.88864
Equal variances not assumed			-.747	17.749	.465	-.49000	.65621	-1.87003	.8900

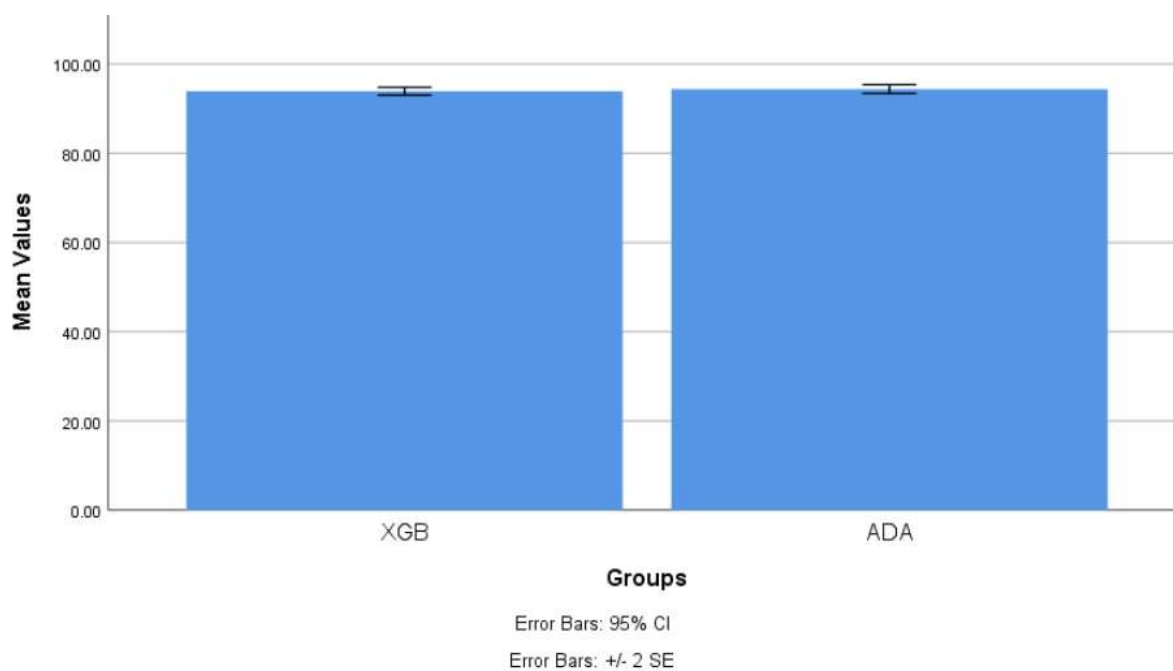


Fig. 1. Bar Graph Comparison on mean accuracy of XGboost (93.89%) and novel Adaboost (94.38%). X-axis: XGBoost,novel ADAboost, Y-axis: Mean Accuracy with ± 2 SE.