International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

Credit Card Transactions Data Adversarial Augmentation in the Frequency Domain

Kattela Vineetha[1], Dr.Dhanraj Cheelu[2]
[1]*MCA Student, Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh, India*
[2]*Professor, Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh, India*

Abstract
With the rapid development of e-commerce, a huge amount of trade data has been generated, which may be used by criminals for fraudulent transactions. In actual application, not all of the funds steal incidents can be found by the anomaly detection system in the bank because the system is not sensitive enough. A large amount of capital was taken by offenders every year, this makes a bad effect on the whole society. Thus, it becomes important to identify such fraudulent transactions in the massive commerce data. The advance of science and technology brings about the increase of data volume, as well as reveal the new application of machine learning, which provides researchers a way to detect vulnerability. Machine learning methods are widely used in the field of credit card fraud detection, but the imbalance between different categories poses an obstacle in the learning tasks. To alleviate this kind of issue, in this paper a model based on frequency domain characteristics be proposed, which is combined with a generative adversarial network (GAN) to augment minority class. In contrast to the traditional adversarial network, which only generates adversarial samples from the training data itself, this approach uses the frequency domain amplitude features of the data to generate diverse training data that matches the trend of data changes. Experimental results demonstrate classification performance is considerably improved and that over-fitting is alleviated when applying to multiple original datasets. Moreover, outperform other existing state-of the-art approaches.
Keywords: credit, card, GAN, ML

Introduction

As an efficient and accurate means of data analysis, machine learning is widely used in the financial field for data analysis [1]–[6], especially to estimate the possibility of fraudulent credit loans in banking systems around the world. When the financial institutions make loan decisions, the main reference to analysis data comes from historical transactions, user personal credit evaluation, and market risk assessments. At the same time, through data processing, the final basis of the decision is considered as the second classification or scoring results [1], [2]. With such kind of data analysis methods, a very important problem emerges. Though the machine learning classification methods offer excellent classification, sometimes the dataset itself has defects that prevent the algorithm to work efficiently. In a large amount of transaction data, most records are legal, while a small amount of illegal or untrusted data is there with higher observation values. In order to solve the defects of the sample, we attempted to find other dimensional features of the data, which are rarely used in data augmentation, such as time dimension, frequency dimension, etc. However, the experimental model found that the time dimension could not represent the current transaction information very well. On the contrary, researchers have discovered the obvious characteristics of credit card data in the frequency domain [7] and have shown significant differences between the positive and negative samples of the amplitude of the spectrum. We make an attempt to reasonably use this difference and explain how it can help in data classification. Thus we propose that data can be transformed into the frequency domain by using each transaction as an equal interval sampling. Most of the data abnormality can be eliminated by data cleaning, it can be transformed into a perfect form, and then used in machine learning feature extraction [8]. However, in this approach, the impact of insufficient sample becomes more prominent. This problem is defined as a sample imbalance

DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

problem. Generally, for such a problem, the optimization direction can be divided into data augmentation and model optimization. The optimization approach proposed in this paper is based on data augmentation discrete Fourier transform (DFT) [9], which is used in the process of generating samples. We used fast Fourier transform (FFT) to transform data into the frequency domain and extract the features of the frequency domain because the amplitudes of the positive and negative samples have obvious stratification. In order to augment this feature, we used Wasserstein generative adversarial network (WGAN) [10] to generate diverse training dataset by fitting the spectral amplitude. Then with inverse Fourier transform, we obtained training dataset with augmented features, at the same time all operations are considered only for the minority class. For the sake of simplicity, hereinafter, we refer to our proposed approach as the frequency domain adversarial augmentation (FDAA). The proposed method is capable of solving the problem of credit card fraud data. At first, we considered a European real-world credit card dataset using different classifiers, where corresponding indicators were improved and compared with other data augmentation methods. Further, two more public credit card datasets with relatively best classifiers were used, which also yielded good results. It was found that the model has some level of universality. The results of this approach are analysed theoretically and the principles are summarized.

Literature Survey

A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring

AUTHORS: Y. Xia, C. Liu, Y. Li, and N. Liu

ABSTRACT: Credit scoring is an effective tool for banks to properly guide decision profitably on granting loans. Ensemble methods, which according to their structures can be divided into parallel and sequential ensembles, have been recently developed in the credit scoring domain. These methods have proven their superiority in discriminating borrowers accurately. However, among the ensemble models, little consideration has been provided to the following: (1) highlighting the hyper-parameter tuning of base learner despite being critical to well-performed ensemble models; (2) building sequential models (i.e., boosting, as most have focused on developing the same or different algorithms in parallel); and (3) focusing on the comprehensibility of models. This paper aims to propose a sequential ensemble credit scoring model based on a variant of gradient boosting machine (i.e., extreme gradient boosting (XGBoost)). The model mainly comprises three steps. First, data pre-processing is employed to scale the data and handle missing values. Second, a model-based feature selection system based on the relative feature importance scores is utilized to remove redundant variables. Third, the hyper-parameters of XGBoost are adaptively tuned with Bayesian hyper-parameter optimization and used to train the model with selected feature subset. Several hyper-parameter optimization methods and baseline classifiers are considered as reference points in the experiment. Results demonstrate that Bayesian hyper-parameter optimization performs better than random search, grid search, and manual search. Moreover, the proposed model outperforms baseline models on average over four evaluation measures: accuracy, error rate, the area under the curve (AUC) H measure (AUC-H measure), and Brier score. The proposed model also provides feature importance scores and decision chart, which enhance the interpretability of credit scoring model.

Credit Card Fraud Detection Using AdaBoost and Majority Voting

AUTHORS: K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi

ABSTRACT:

Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are first used. Then, hybrid methods which use AdaBoost and majority voting methods are applied. To evaluate the model efficacy, a publicly available credit card data set

DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

is used. Then, a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

Detecting financial restatements using data mining techniques

AUTHORS: I. Dutta, S. Dutta, and B. Raahemi

ABSTRACT: Both intentional and unintentional restatements may destroy shareholders value.Financial restatements (intentional/unintentional) detection models are developed.Performance of all widely used data mining techniques are compared.A reduced set of significant attributes are identified for predicting restatements.Class imbalance and cost imbalance issues are addressed. Financial restatements have been a major concern for the regulators, investors and market participants. Most of the previous studies focus only on fraudulent (or intentional) restatements and the literature has largely ignored unintentional restatements. Earlier studies have shown that large scale unintentional restatements can be equally detrimental and may erode investors confidence. Therefore it is important for us to pay a close to the significant unintentional restatements as well. A lack of focus on unintentional restatements could lead to a more relaxed internal control environment and lessen the efforts for curbing managerial oversights and instances of misreporting. In order to address this research gap, we focus on developing predictive models based on both intentional (fraudulent) and unintentional (erroneous) financial restatements using a comprehensive real dataset that includes 3,513 restatement cases over a period of 2001 to 2014. To the best of our knowledge it is the most comprehensive dataset used in the financial restatement predictive models. Our study also makes contributions to the datamining literature by (i) focussing on various datamining techniques and presenting a comparative analysis, (ii) ensuring the robustness of various predictive models over different time periods. We have employed all widely used data mining techniques in this area, namely, Decision Tree (DT), Artificial Neural Network (ANN), Nave Bayes (NB), Support Vector Machine (SVM), and Bayesian Belief Network (BBN) Classifier while developing the predictive models. We find that ANN outperforms other data mining algorithms in our empirical setup in terms of accuracy and area under the ROC curve. It is worth noting that our models remain consistent over the full sample period (2001-2014), pre-financial-crisis period (2001-2008), and post-financial-crisis period (2009-2014). We believe this study will benefit academics, regulators, policymakers and investors. In particular, regulators and policymakers can pay a close attention to the suspected firms and investors can take actions in advance to reduce their investment risks. The results can also help improving expert and intelligent systems by providing more insights on both intentional and unintentional financial restatements.

Some Experimental Issues in Financial Fraud Mining

AUTHORS: J. West and M. Bhattacharya

ABSTRACT: Financial fraud detection is an important problem with a number of design aspects to consider. Issues such as problem representation, choice of detection technique, feature selection, and performance analysis will all affect the perceived ability of solutions, so for auditors and researchers to be able to sufficiently detect financial fraud it is necessary that these issues be thoroughly explored. In this paper we will analyse some of the relevant experimental issues of fraud detection with a focus on credit card fraud. Observations will be made on issues that have been explored by prior researchers for general data mining problems but not yet thoroughly explored in the context of financial fraud detection, including problem representation, feature selection, and performance metrics. We further investigated some of these issues with controlled simulations, concentrating on detection algorithms, feature selection, and performance metrics for credit card fraud.

Stock market prediction and Portfolio selection models: A survey

AUTHORS: A. M. Rather, V. N. Sastry, and A. Agarwal

ABSTRACT: Stock data is known to be chaotic in nature and it is a challenging task to predict the non-linear patterns of such data. Forming an optimal portfolio of stocks is yet another challenging

International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

task and limitations do exist in every portfolio model in some form or the other. In order to resolve such problems, many artificial intelligence models have appeared in literature which are also known as intelligent models. Prediction of stocks as well as investing in appropriate stocks has remained in focus among investors, industrialists as well as among academicians. This paper surveys important published articles in the related area available in literature. This survey highlights traditional mathematical models available in articles which have appeared decades back till artificial intelligence based models available in recent articles.

Existing System

In actual application, not all of the funds steal incidents can be found by the anomaly detection system in the bank because the system is not sensitive enough. A large amount of capital was taken by offenders every year, this makes a bad effect on the whole society. Thus, it becomes important to identify such fraudulent transactions in the massive commerce data.

Disadvantages:

1.In actual application, not all of the funds steal incidents can be found by the anomaly detection system in the bank because the system is not sensitive enough.

2. it becomes important to identify such fraudulent transactions in the massive commerce data.

Proposed System

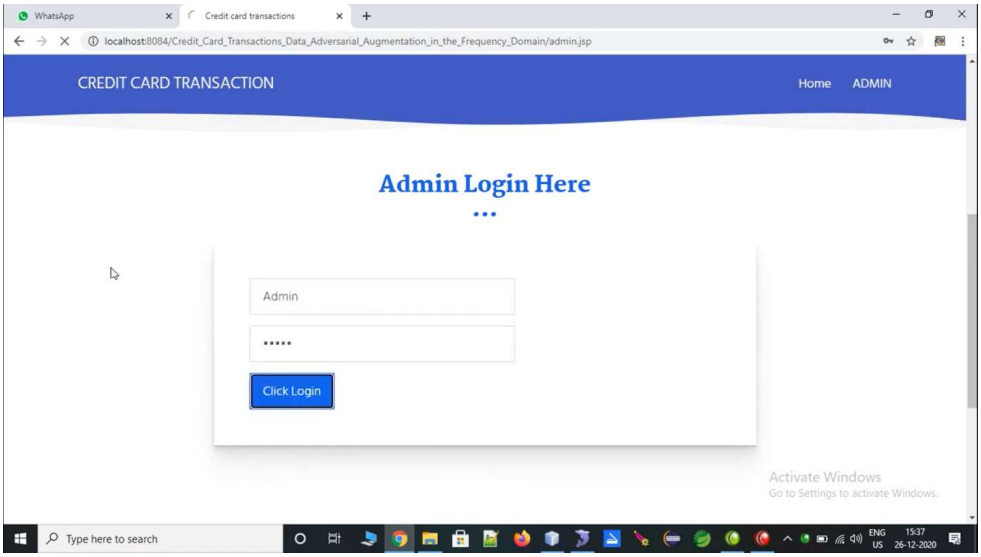
we propose a new unbalanced data augmentation method based on the transformation with frequency domain and generative model to obtain a high quality of data augmentation. Moreover, by performing targeted augmentation on the categories of samples lacking training data, the training obtains a model with higher classification AUC and F1-score, meanwhile over-fitting is mitigated. The experimental results on multiple credit card datasets show that our method is effective and can improve the performance of many common machine learning classification models. The constraint of frequency domain to adversarial network helps in data generation a lot. By generating more training data, the performance of the classifier can be improved for ensuring speed and sample diversity.

Advantages:

1.It is combined with a generative adversarial network (GAN) to augment minority class.

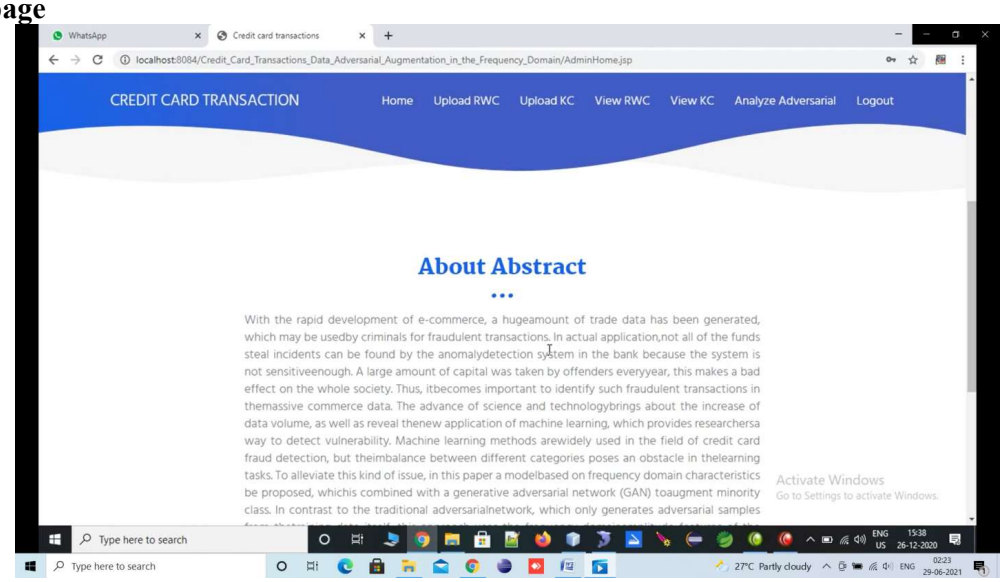
2. Experimental results demonstrate classification performance is considerably improved and that over-fitting is alleviated when applying to multiple original datasets.

Results

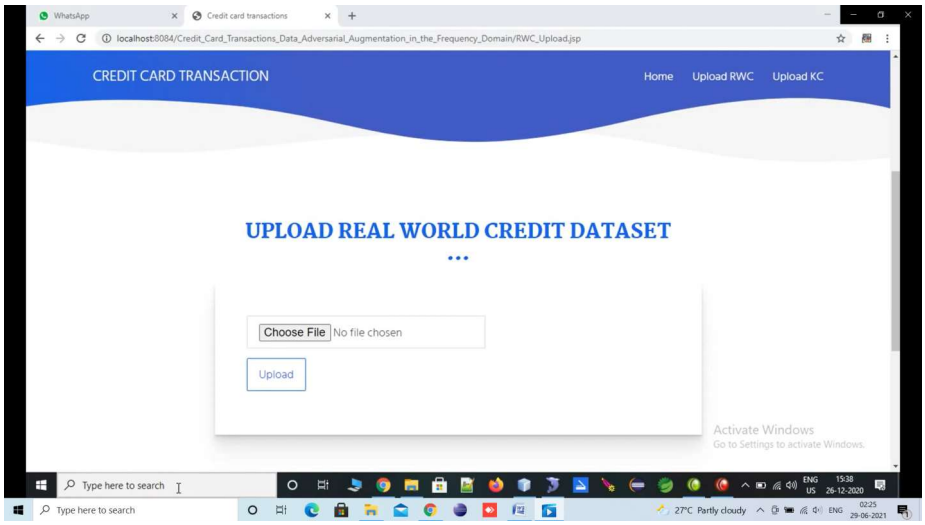


International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

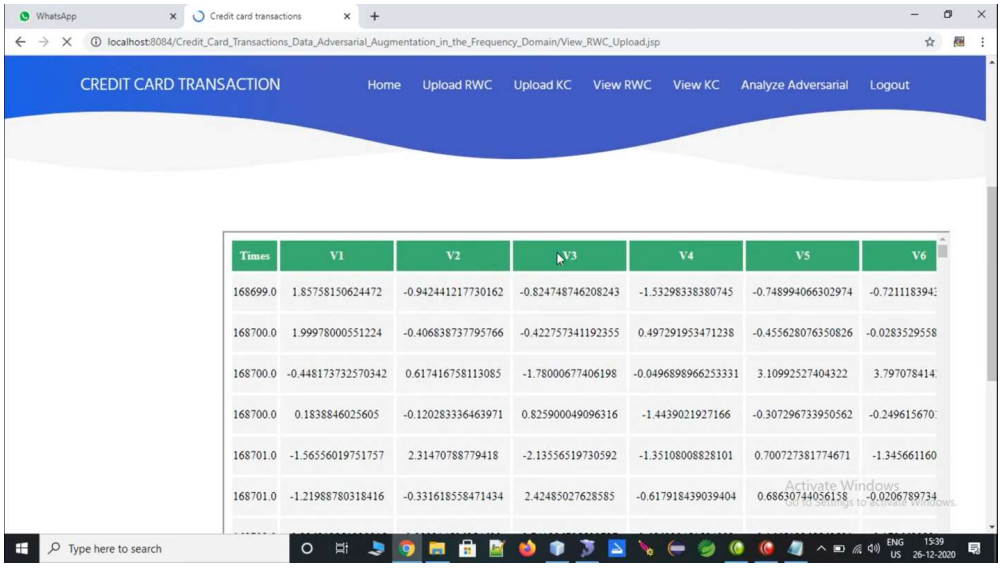
Home page



Upload RWC

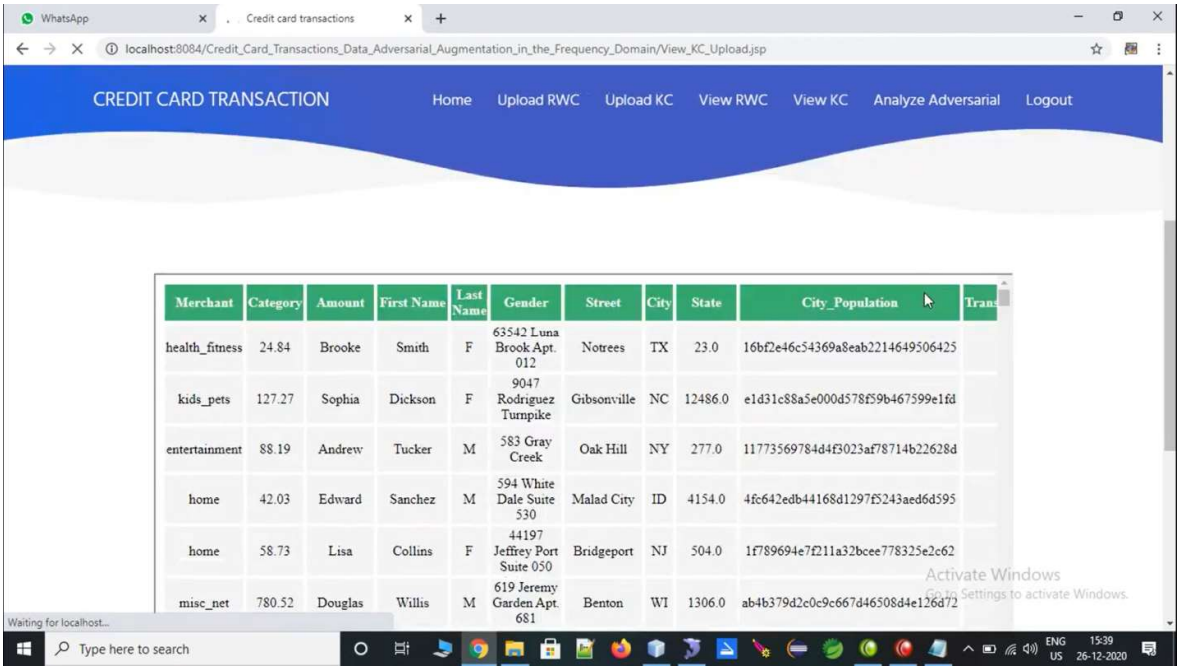


View RWC



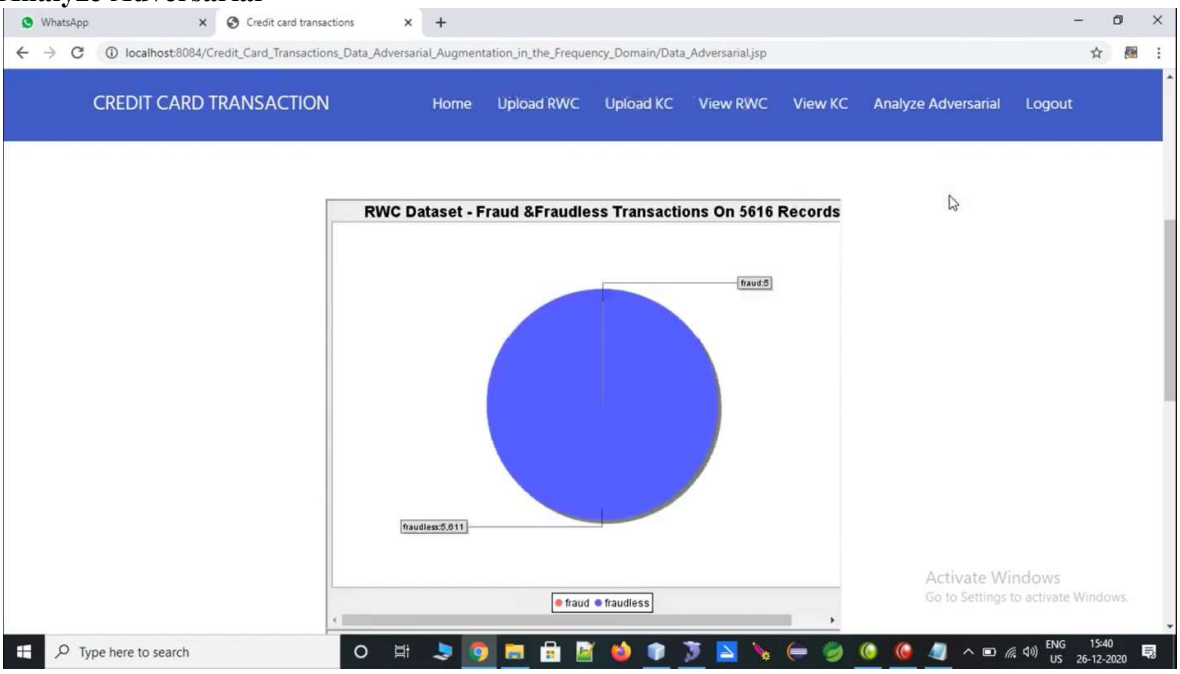
International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

View KC



Merchant	Category	Amount	First Name	Last Name	Gender	Street	City	State	City_Population	Trans
health_fitness	24.84	Brooke	Smith	F	63542 Luna Brook Apt. 012	Notrees	TX	23.0	16bf2e46c54369a8eab2214649506425	
kids_pets	127.27	Sophia	Dickson	F	9047 Rodriguez Turnpike	Gibsonville	NC	12486.0	e1d31c88a5e000d578f59b467599e1fd	
entertainment	88.19	Andrew	Tucker	M	583 Gray Creek	Oak Hill	NY	277.0	11773569784d4f3023af78714b22628d	
home	42.03	Edward	Sanchez	M	594 White Dale Suite 530	Malad City	ID	4154.0	4fc642edb44168d1297f5243aed6d595	
home	58.73	Lisa	Collins	F	44197 Jeffrey Port Suite 050	Bridgeport	NJ	504.0	1f789694e7f211a32bcee778325e2c62	
misc_net	780.52	Douglas	Willis	M	619 Jeremy Garden Apt. 681	Benton	WI	1306.0	ab4b379d2c0c9c667d46508d4e126d72	

Analyze Adversarial



Conclusion

In this paper, we propose a new unbalanced data augmentation method based on the transformation with frequency domain and generative model to obtain a high quality of data augmentation. Moreover, by performing targeted augmentation on the categories of samples lacking training data, the training obtains a model with higher classification AUC and F1-score, meanwhile over-fitting is mitigated. The experimental results on multiple credit card datasets show that our method is effective and can improve the performance of many common machine learning classification models. The constraint of frequency domain to adversarial network helps in data generation a lot.

DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

By generating more training data, the performance of the classifier can be improved for ensuring speed and sample diversity. For the European credit card benchmark dataset, the AUC of the classifier trained by the training set augmented by our method increased by 5.66% and the F1-score increased by 10.89%, this was better than other methods.

References

[1] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," Expert Systems with Applications, vol. 78, pp. 225–241, Jul. 2017.
[2] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," IEEE Access, vol. 6, pp. 14 277–14 284, 2018.
[3] I. Dutta, S. Dutta, and B. Raahemi, "Detecting financial restatements using data mining techniques," Expert Systems with Applications, vol. 90, pp. 374–393, Dec. 2017.
[4] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," Expert Systems with Applications, vol. 42, no. 1, pp. 259–268, Jan. 2015.
[5] J. West and M. Bhattacharya, "Some Experimental Issues in Financial Fraud Mining," Procedia Computer Science, vol. 80, pp. 1734–1744, 2016.
[6] A. M. Rather, V. N. Sastry, and A. Agarwal, "Stock market prediction and Portfolio selection models: A survey," OPSEARCH, vol. 54, no. 3, pp. 558–579, Sep. 2017.
[7] M. D. Godfrey, "An Exploratory Study of the Bi-Spectrum of Economic Time Series," Applied Statistics, vol. 14, no. 1, p. 48, 1965.
[8] R. Bradley, "Adaptive data cleaning," US Patent US20 060 238 919A1, Oct., 2006.
[9] R. Saia and S. Carta, "Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach:," in Proceedings of the 14th International Joint Conference on E-Business and Telecommunications. Madrid, Spain: SCITEPRESS - Science and Technology Publications, 2017, pp. 335–342. [10] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53–65, Jan. 2018.
[11] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," Expert Systems with Applications, vol. 39, no. 3, pp. 3446–3453, Feb. 2012.
[12] C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," Expert Systems with Applications, vol. 33, no. 4, pp. 847–856, Nov. 2007.
[13] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System. New York: Assoc Computing Machinery, 2016, wOS:000485529800092.
[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, Jun. 2002.
[15] D. M. McNeish and L. M. Stapleton, "The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration," Educational Psychology Review, vol. 28, no. 2, pp. 295–314, Jun. 2016.
[16] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," in 2008 Ieee International Joint Conference on Neural Networks, Vols 1-8. New York: Ieee, 2008, pp. 1322–1328, wOS:000263827200212.
[17] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Transactions on Audio and Electroacoustics, vol. 15, no. 2, pp. 70–73, 1967. [18] S. W. Smith, The Scientist & Engineer's Guide to Digital Signal Processing. California Technical Pub, 1997.

DOI:10.46647/ijetms.2025.v09i02.092 ISSN: 2581-4621

[19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in 2017 International Conference on Machine Learning, 2017, pp. 214–223.

[20] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging- , Boosting-, and Hybrid-Based Approaches," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 463–484, Jul. 2012.

[21] B. Pang, E. Nijkamp, and Y. N. Wu, "Deep Learning With TensorFlow: A Review," Journal of Educational and Behavioral Statistics, p. UNSP 1076998619872761, 2019, wOS:000487155100001.

[22] G. Daniel and A. Pekka, "A Web-based modeling tool for the SEMAT Essence theory of software engineering," Journal of Open Research Software, vol. 1, no. 1, p. e4, Sep. 2013. [23] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," SIGKDD Explor. Newsl., vol. 6, no. 1, pp. 20–29, Jun. 2004.