IOSR Journal of Engineering (IOSRJEN) www.iosrjen.org
ISSN (e): 2250-3021, ISSN (p): 2278-8719 Special Issue || September 2021 || PP 16-20
7th National Conference On Advancements in Communication, Computing and Electronics Technology-ACCET 2021

Credit Card Fraud Detection: AComparative Study

Siddharth Shinde
Electronics and Telecommunication Engineering
Savitribai Phule Pune University, Pune.

Pravin Chavan
Electronics and Telecommunication Engineering
Savitribai Phule Pune University, Pune.

Dr. Mrs. K. S. Tiwari
Faculty of Electronics and Telecommunicationdept., Savitribai Phule Pune University, Pune.

Abstract-*The goal of data analytics is to delineate hidden patterns and use them to support informed decisions in a variety of situations. Credit card fraud is escalating significantly with the advancement of modernized technology and became an easy target forfrauds. Credit card fraud has highly imbalanced publicly available datasets. In this paper, we apply many supervised machine learning algorithms to detect credit card fraudulent transactions using a real-world dataset. Furthermore, we employ these algorithms to implement a super classifier using ensemble learning methods. We identify the most important variables that may lead to higher accuracy in credit card fraudulent transaction detection.*
Additionally, we compare and discuss the performance of various supervised machine learningalgorithms that exist in literature against the super classifier that we implemented in this paper.
Keywords: *CreditCard,Fraud detection, Supervised machine learning, Classification, Imbalanced dataset, Sampling.*

I. Introduction

Today, all around the world data is available very easily, fromsmall to big organizations are storing information that has high volume, variety, speed and worth. This information comes fromtons of sources like social media followers, likes and comments,user's purchase behaviours. All this information

pattern. Early analysis of big data was centred primarily on datavolume, for example, general public database, biometrics, financial analysis. For frauds, the credit card is an easy and friendly target because without any risk a significant amount ofmoney is obtained within a short period. To commit credit card fraud, fraudsters try to steal sensitive information such as creditcard number, bank account and social security number.

Fraudsters try to make every fraudulent transaction legitimate which makes fraud detection a challenging problem. Increased credit card transactions show that approximately 70% of thepeople in the US can fall into the trap of these fraudsters.

Credit card dataset is highly imbalanced dataset because it carries more legitimate transactions as compared to the fraudulentone. That means prediction will get very high accuracy score without detecting a fraud transaction. To handle this kind of problem one better way is to class distribution, i.e., sampling minority classes. In sampling minority, class training example can be increased in proportion to the majority class to raise the chance of correct prediction by the algorithm. In this paper, we use ten machine learning models and compare their Accuracy, TPR, FPR, G-mean, Recall, Precision, Specificity and F1-Score. All machine learning algorithm is evaluated using a real-world credit card transaction to identity fraud or nonfraud transaction. The main motive of this paper to apply supervised learning method on the real-world dataset.

II. Related Studies

Logistic regression and artificial neural network give flags whenever fraudulent and legitimate transaction happens based upon their transaction score. The performance of all the machine learning models decreases because of the skewness of the trainingdataset.

To make the unbalanced dataset balanced two different methods are used namely, intrinsic features and network-based features. Intrinsic features compare customer's past transactions looks for any suspiciousness score for each network object. These two methods lead to a very high accuracy score in Random Forest getting a 1% false positive making the perfect model obtaining fraudulent transaction. Comparisons are made between different modelling and algorithm techniques on a real dataset. Some of the algorithms underperform because of the unbalanced dataset. To learn from (non-stream credit card and data stream) unbalanced dataset has three different methods used (static, update and DataStream). They also used two methods of under sampling SMOTE and Easy Ensemble to make their dataset balanced from an unbalanced dataset. While in RF &SVM there is a decrement to see in AUC and increment in F-measure. The neural network architecture used upon an unsupervised method ofusing real-time transaction entry. Self-organizing map of the neural network by using optical classification it can solve the problem for each associated with an associated group. With 95% detection of fraud with ROC curve without causing any falsealarm.

Data Mining reports the development & implementation of a fraud detection system in a large e-tail merchant. Using a cost- based performance to train the algorithm to get the business outcomes take longer time. A bank seller decision support system that used in outline banking fraud analysis and investigation, that automatically find the fraud give them ranks and understand the user spending habits using their past transaction (based on mathematically and statistical technique).

III. Material and Methods

A. Supervised learning and unsupervised learning

Using supervised method helps to find out the label on past transaction, they tend to not recognized fraud pattern that has occurred in the past. While unsupervised technique helps to findout the class of transactions.

B. Unbalanced data

It is quite challenging to learn from an unbalanced dataset and for balancing it, the sampling method used.
A publicly available dataset that contains 284,807 transactions made in Sep. 2013 by European cardholders. The dataset includes 492 fraud transactions, which is highly imbalanced. Hence, under-sampling was applied.

C. **Fraud Detection Classifier**

Logistic Regression can handle the data with theoretical and statistical characteristics. Decision Tree is a supervised learning method that widely uses models for classification and regression tasks. Random Forest method used for classification and regression using tThaebcloell1e.cPtieornformance evaluation of different classifiers. of the decision tree, each one is slightly different from each other.

With first introduction in 1995 Navies Bayes using Bayes theorem for independence hypothesis.

K-Nearest Neighbourhood (KNN) is a necessary calculation which stores every single accessible occurrence. The Gradient Boosted Tree Classifier (GBT) is a collection of classification and regression models. Boosting supports improve the tree accuracy. XGB (XG boost Classifier) is the most refined classifier that works with all type of dataset.

The support vector machines (SVM) are initially presented in 1995, and they have been observed to be extremely fruitful in an assortment of exemplary classification tasks. The MLP organize comprises of no less than three layers of hubs, i.e., input, covered up, and yield.

Ensemble learning (also known as meta-classifier) helps to improve the results by combining multiple machine learning classifier to improve the predictive outcomes. Accuracy is one important method to compare the performance of classification models we also look at the other factors like F1-Score, Precision, TPR, FPR, Recall, G-mean and Specificity. All these evaluations measure adequately reveal validation of the study very well.

IV. Results and Discussion

We used 70% of the data is used for training and 30% used for the testing set. Data was balanced by using an under-samplingtechnique. So, we used Accuracy, F1-Score, Recall, Precision, G-Mean, FPR, TRP and specificity are used to compare the models. Table 1 shows all classifier results and comparisons. In table 1, stacking classifier (0.9527 accuracies) is leading the other classifiers, followed by the random forest (0.94594 accuracies) and XGB classifier (0.94594 accuracies) is helpful only when we have a symmetric dataset. Having a high precision is related to the low false rate. In Figure Random Forest, stacking and XGB classifier all have the same precision score of 0.95 followed by the Gradient boosting and logistic regression with the precision score of 0.94. We find out recall also developed the same ranking of precision in Figure. The F1-score is the weighted median of precision and recall, and its score take false positive and false negative into account F1-score. F1-score also followed the same ranking of Precision and Recall in Figure. SVM has the highest ranking with 0.5360 FPR, and stacking classifier has the lowest ranking with 0.0335 in Figure. TPR of the logistic regression has the highest ranking followed by the MLP and stacking classifier. We find out the top five features in table 2. Features 14 is the essential features and features and got selected by all algorithms. And V4 is decided by four features.

V. Conclusion

- Under-sampling is done for balancing the unbalanced dataset.
- The learning model's evaluation is based on their accuracy, recall, precision, TPR, FPR, specificity and G-mean.
- The result of all the purposed models were superior in overall performance.
- Overall results show that stacking classifier which is used LR as meta classifier is most promising for predicting fraud transaction in the dataset, followed by the random forest and XGB classifier.

Table-1

Model	Accuracy	Precision	Recall	TPR	FPR	F1- Score	G-Mean	Specificity
SC	0.95270	0.95	0.95	0.9387	0.0335	0.95	0.9524	0.9664
RF	0.94594	0.95	0.95	0.9251	0.0335	0.95	0.9455	0.9664
XGB Classifier	0.94594	0.95	0.95	0.93197	0.0402	0.95	0.9457	0.9597
KNN	0.94256	0.91	0.91	0.9183	0.0335	0.91	0.942	0.9664
LR	0.93918	0.94	0.94	0.93877	0.0604	0.94	0.9391	0.9395
GB	0.93581	0.94	0.94	0.9183	0.0335	0.94	0.942	0.9664
MLP Classifier	0.93243	0.93	0.93	0.9387	0.0738	0.93	0.9323	0.9261
SVM	0.93243	0.93	0.93	0.9183	0.536	0.93	0.9321	0.9463
Decision Tree	0.90878	0.91	0.91	0.9047	0.0872	0.91	0.9086	0.9127
Navies Bayes	0.90540	0.91	0.91	0.85714	0.04697	0.91	0.9037	0.953

Model	Precision
Random Forest	0.95000
Stacking Classifer	0.95000
XGB Classifier	0.95000
Gradient Boosting	0.94000
Logistic Regression	0.94000
MLP classifer	0.93000
SVM	0.93000
Decision Tree	0.91000
KNN	0.91000
Navies Bayes	0.91000

Precision: 0.91000 0.95000

Figure 1. Classifier ranking based on precision score

Model	Recall
Random Forest	0.95000
Stacking Classifer	0.95000
XGB Classifier	0.95000
Gradient Boosting	0.94000
Logistic Regression	0.94000
MLP classifer	0.93000
SVM	0.93000
Decision Tree	0.91000
KNN	0.91000
Navies Bayes	0.91000

Recall: 0.91000 0.95000

Figure 2. Classifier ranking based on Recall score

Future implications

- Future work will be conducting the using the voting classifier an[d6] check the performance with other ML learning methods, increase the size of training and testing dataset.
- We can work on using the all the machine learning algorithm to find out the feature's importance.
- We can work on top ten features and find-out the accuracy, Recall,Precision, Confusion matrix and compare it with our old result.

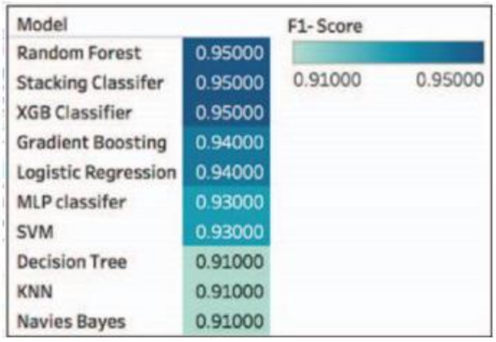
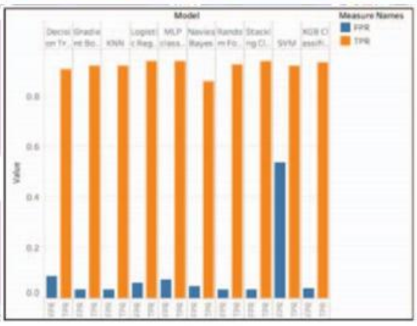


Figure 3. Classifier ranking based on F1-score

Table 2. Feature rankin s of dataset

Ranking	Random Forest	Decision Tree	Gradient Boosting	XGB	Logistic Regression
1	V10	V14	V14	V4	V4
2	V11	V17	V10	V14	V12
3	V12	V20	V4	V8	V14
4	V13	V10	V12	V11	V11
5	V14	V4	V19	V13	V8



4. TPR and FPR performance of all the classifier

Reference

[1]. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data mining for credit card fraud: A comparative study,” *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011.

[2]. K. Chaudhary, J. Yadav, and B. Mallick, “A review of Fraud Detection Techniques: Credit Card,” *Int. J. Comput. Appl.*, vol.45, no. 1, pp. 975–8887, 2012.

[3]. “Mining of Massive Datasets Second Edition.”

[4]. F. N. Ogwueleka, “Data Mining Application in Credit Card FraudDetection System,” vol.6, no. 3, pp. 311–322, 2011.

[5]. H. Nordberg, K. Bhatia, K. Wang, and Z. Wang, “BioPig: a Hadoop-based analytic toolkit for large-scale sequence data,” *Bioinformatics*, vol. 29, no.23, pp. 3014– 3019, Dec. 2013.

[6]. M. Hegazy, A. Madian, and M. Ragaie, “Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques,” Egypt. Comput. Sci., no. 03, pp. 72–81, 2016.

[7]. M. Zareapoor and P. Shamsolmoali, “Application of credit card fraud detection: Based on bagging ensemble classifier,” Procedia Comput. Sci., vol. 48, no. C, pp. 679–686, 2015.

[8]. O. S. Yee, S. Sagadevan, N. Hashimah, and A. Hassain, “Credit Card Fraud Detection Using Machine Learning As Data Mining Technique,” vol. 10, no. 1, pp. 23–27.

[9]. 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. 14277– 14284, 2018.

[10]. N. Mahmoudi and E. Duman, “Detecting credit card fraud by Modified Fisher Discriminant Analysis,” Expert Syst. Appl., vol. 42, no. 5, pp. 2510–2516, 2015.

[11]. A. Dal Pozzolo, O. Caelen, Y. A. Le Borgne, S. Waterschoot, and G. Bontempi, “Learned lessons in credit card fraud detection from a practitioner perspective,” Expert Syst. Appl., vol. 41, no. 10, pp. 4915–4928, 2014.

[12]. M. A. Scholar, M. Ali, and P. Fellow, “Investigating the Performance of Smote for Class Imbalanced Learning : A Case Study of Credit Scoring Datasets,” vol. 13, no. 33, pp. 340–353, 2017.

[13]. H. He, W. Zhang, and S. Zhang, “A novel ensemble method for credit scoring: Adaption of different imbalance ratios,” Expert Syst. Appl., vol. 98, pp. 105–117, May 2018.

[14]. A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, “Calibrating Probability with Undersampling for Unbalanced Classification.”

[15]. Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines,” Int. Multiconference Eng. Comput. Sci., vol. I, pp. 442–447, 2011.

[16]. V. Van Vlasselaer et al., “APATE: A novel approach for automated credit card transaction fraud detection using networkbased extensions,” Decis. Support Syst., vol. 75, pp. 38–48, 2015.

[17]. C. Phua, D. Alahakoon, and V. Lee, “Minority report in fraud detection,” ACM SIGKDD Explor. Newsl., vol. 6, no. 1, p. 50, 2004.

[18]. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literatu[2re6,]” Decis. Support Syst., vol. 50, no. 3, pp. 559–569, 2011.

[19]. C. C. Lin, A. A. Chiu, S. Y. Huang, and D. C. Yen, “Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts’ judgments,” Knowledge-Based Syst., vol. 89, pp. 459–470, 2015.

[20]. N. Carneiro, G. Figueira, and M. Costa, “A data mining based system for credit-card fraud detection in e-tail,” *Decis. Support Syst.*, vol. 95, pp. 91–101, 2017.

[21]. A. C. Bahnsen, D. Aouada, and B. Ottersten, “Example- dependent cost-sensitive logistic regression for credit scoring,” Proc. - 2014 13th Int. Conf. Mach. Learn. Appl. ICMLA 2014[,2p8p]. 263–269, 2014.

[22]. M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, “BankSealer: A decision support system for online banking fraud analysis and investigation,” Comput. Secur., vol. 53, pp. 175– 186, 2015.

[23]. C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, “Transaction aggregation as a strategy for credit card fraud detection,” Data Min. Knowl. Discov., vol. 18, no. 1, pp. 30–55, Feb. 2009.

[24]. G. Rushin, C. Stancil, M. Sun, S. Adams, and P. Beling, “Horse race analysis in credit card fraud - Deep learning, logistic regression, and Gradient Boosted Tree,” 2017 Syst. Inf. Eng. Des. Symp. SIEDS 2017, pp. 117–121, 2017.

[25]. R. J. Bolton, D. J. Hand, F. Provost, L. Breiman, R. J. Bolton, and D. J. Hand, “Statistical Fraud Detection: A ReviewCommentCommentRejoinder,” Stat. Sci., vol. 17, no. 3, pp. 235–255, 2002. [26] X.-Y. Liu, J. Wu, and Z.-H. Zhou, “Exploratory Undersampling for Class- Imbalance Learning,” vol. 39, no. 2, 2009.

[26]. E. A. Mohammed, M. M. A. Mohamed, C. Naugler, and B. H. Far, “Toward leveraging big value from data: chronic lymphocytic leukemia cell classification,” *Netw. Model. Anal.Heal. Informatics Bioinforma.*, vol. 6, no. 1, p. 6, Dec. 2017. G. H. John and P. Langley, “Estimating Continuous Distributionsin Bayesian Classifiers,” Feb. 2013.