IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comparative Study on Different Fraud Detection Methods in Credit Card-Based Transactions

M.s.s.Teja Sriram[1], Nishant Agnihotri[2], K.Devika[3], D.Sirisha[4], K.Vivek[5], B.Rajasekhar[6]

[1]Student, Lovely Professional University

[2]**Assistant professor**, Lovely Professional University

[3]Student, Lovely Professional University

[4]Student, Lovely Professional University

[5]Student, Lovely Professional University

[6]Student, Lovely Professional University

Abstract

As communication technology and e-commerce have improved, credit cards have become the most common mode of payment for both online and offline purchases. As a result, security in this system is expected to be very good to avoid fraudulent transactions. Each year, the number of fraudulent credit card data transfers rises. Researchers are also experimenting with unique approaches to detect and prevent such frauds in this direction. However, some strategies that can accurately and efficiently detect these scams are always needed. This study proposes a method for detecting credit card fraud using supervised and unsupervised learning approaches. The suggested neural network method outperforms existing methods such as Logistic Regression, Local Outlier Factor, Isolation Forest, and K-means clustering when compared. As the accuracy values of each model are nearly negligible because the dataset is imbalanced, we will compare the false-negative rate, recall, precision, and recall parameters of each model.

Keywords

Machine learning, Logistic regression, Isolation Forest, Local Outlier Factor, K means clustering, Neural Networks.

1. Introduction

Fraud detection (In a range of industries, including banking and insurance, fraud detection is used. Banking fraud also includes things like check forgery and the use of stolen credit cards.) [7]. The purpose of fraud detection is to correctly identify legitimate and fraudulent credit card transactions, which is a data mining classification task. only a little research has been published in this field due to the lack of real-world data on which researchers may conduct experiments. The bank's unwillingness to provide sensitive consumer transaction data for privacy reasons. Shopping through internet-based applications and paying bills online has become commonplace, thus a physical card is no longer required to make transactions.[1][10][21]

2. Detection methods

2.1 Logistic regression

It is a statistical model which uses a logistic function to classify the binary dependent variable.it understands the relationship between the dependent variable and one or more independent variables by estimating the probabilities using a logistic regression equation [12][13]. this type of model is useful in predicting the chances of the choice being made.it is a supervised machine learning model.[12] [20]

2.2 Isolation Forest

Isolation forest tree machine learning algorithm is an anomaly detection method. It can work with either supervised or unsupervised learning methods. For outlier detection, the isolation forest tree technique is different from other types of distance or density-based methods, and the algorithm tried tree to build an incredibly randomized decision tree for separating outliers and return the anomaly score of each sample using the Isolation Forest algorithm.[2][19]

2.3 Local Outlier Factor

The LOF (Local outlier factor) approach is an unsupervised anomaly detection methodology that can be used to detect the outlier. The local density deviation of a data point is calculated by comparing it to its neighbors. Because of its lesser density than the rest of the neighborhood, it is considered an outlier. Outliers are samples that have a density that is much lower than that of their neighbors. Local density is determined by calculating distances between neighboring data points (K-nearest neighbors). As a result, each data point's local density may be calculated.[2][19]

2.4 K-means clustering

The basic purpose of the K-means algorithm is to lower the sum of distances between points and the cluster centroid that corresponds to them. It allows us to cluster data into different groups and is a simple and rapid technique for determining the categories of groups in an unlabeled dataset without any prior training. A centroid-based algorithm is assigned to each cluster. The value of k should be known ahead of time in this algorithm.[3][18]

2.5 Neural networks:

A neural network is a set of algorithms that recognizes hidden correlations in a set of data by simulating how the human brain functions.[22] It is a mathematical model that is implemented neurally. To perform all functions, it contains a large number of interconnected processing components known as neurons. A weighted connection of neurons stores information in the neurons. The neural network approach is mostly used for image classification and detection.[6][16][17]

3. Dataset:

The data set includes credit card transactions from European cardholders in September 2013. We have 492 frauds out of 284,807 transactions in our dataset, which is two days' data. The positive class (frauds) is 0.172 percent of all transactions. It only has numerical input variables that have been transformed using PCA. We are unable to get the original features of the data or additional information due to confidentiality concerns. V1; V2; V3; V4; V5; V6; V7... The only features not modified by PCA are 'Time' and ‘Amount’. These are classified as fraud and valid in the class column.[8][9]

4. Proposed approach

Here, we used supervised and unsupervised techniques like logistic regression, k-means clustering, isolation forest, local outlier factor, and neural networks to identify the frauds. In figure 1 we can see the flowchart of the proposed work. Data preprocessing is done to extract the features and clean the data and then the data is split into training and testing. we build the models and train them using training data if the results are variable then the parameters are tuned and the training process is repeated. if the results are constant, we proceed to the test phase where the model detects the fraud transactions and gives the performance metrics of the model.

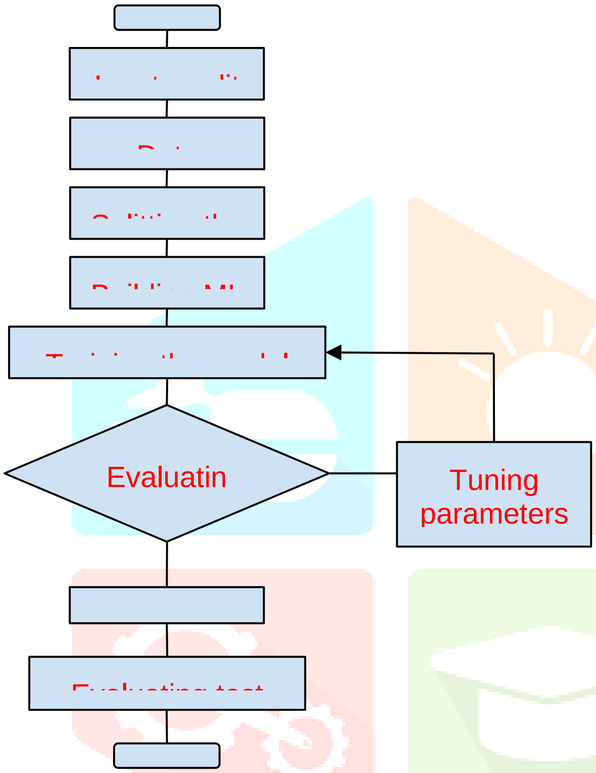

figure 1: flowchart of the fraud detection model

5. Performance evaluation

5.1 Confusion metrics

A confusion matrix is a table that displays a machine learning model's performance on a set of test data for which the true values are known.[11][14][10]

Based on the observations the confusion matrics for Logistic regression, K-means clustering, Local outlier factor, Isolation Forest, Autoencoders, and Neural network is presented in table 1. from this table we can observe that the neural network model is giving better true positive values compared to respective models

Table 1: Observations of confusion matrics for dataset 1

ML models	True Negative	False Positive	False Negative	True Positive
Logistic regression	85278	23	50	92
K-means	83817	1484	137	5
Local outlier factor	85140	161	140	2
Isolation forest	85194	107	104	38
Neural networks	85272	29	36	106

Table 2: Observations of confusion matrics for dataset 2

ML models	True Negative	False Positive	False Negative	True Positive
Logistic regression	5634	215	854	2046
K-means	2919	2930	1605	1295
Local outlier factor	3782	2067	1934	966
Isolation forest	3816	2033	1861	1039
Neural networks	5767	82	769	2131

5.3 Results

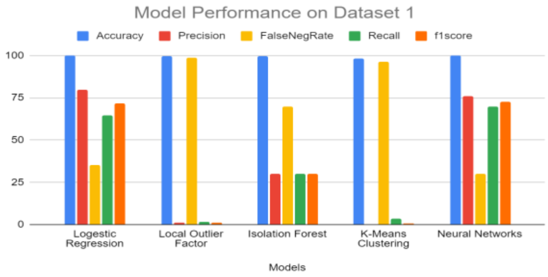

figure 2: Model performance on dataset 1

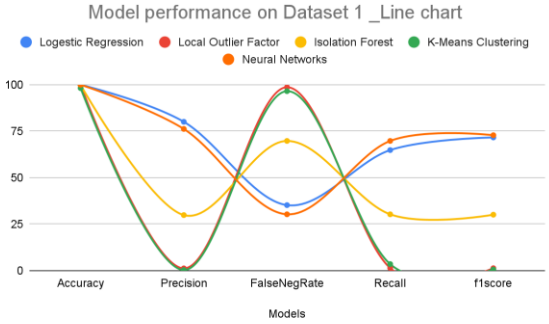

figure 3: Model performance on dataset 1

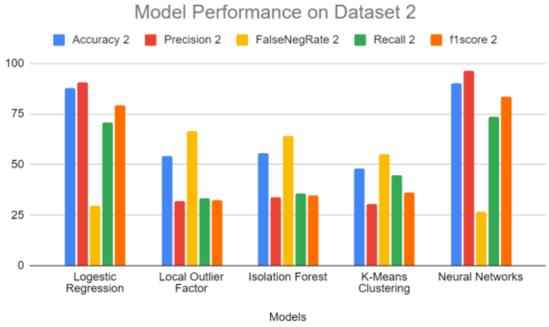

figure 4: Model performance on dataset 2

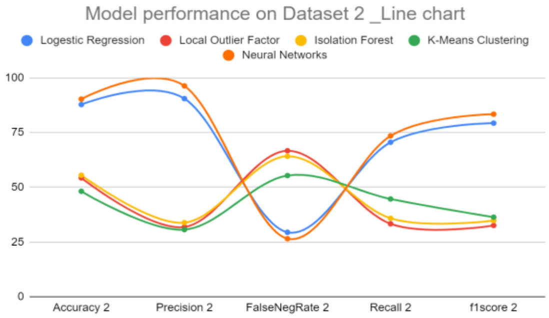

figure 5: Model performance on dataset 2

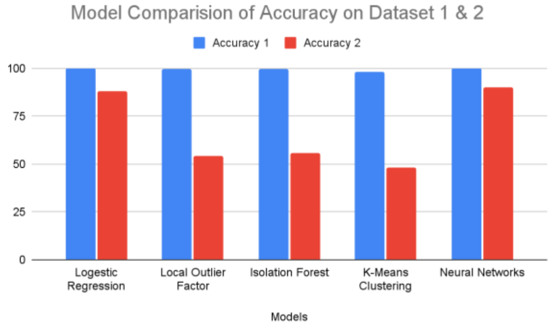

figure 6: Accuracy vs models

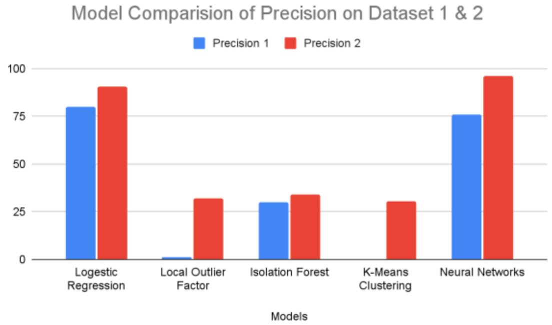

figure 7: precision vs models

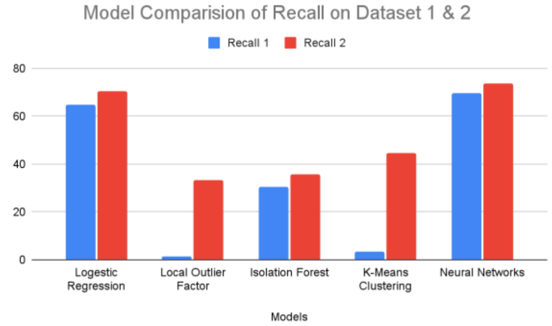

figure 8: recall vs models

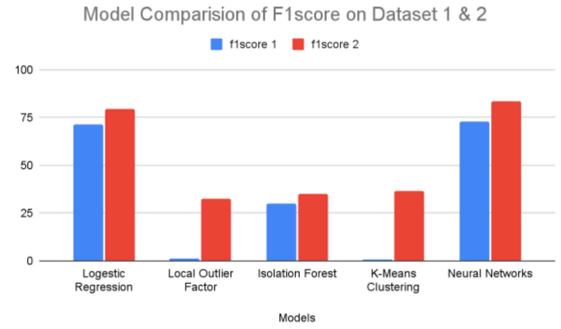

figure 9: f1score vs models

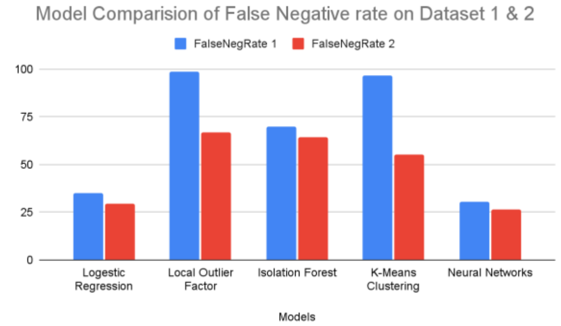

figure 10: false-negative rate vs models

6. Conclusion and Future Directions

we can conclude from above figure 3, figure 5 that Neural-network based machine learning model performs better than other models i.e, local outlier factor, k-means clustering, isolation forest, and logistic regression. from figure 8 we can observe that the neural network detects fewer false-negative rates which are 25.36% and 26.51%

In the future we can improve the neural network model by lowering the false-negative rate and improving the performance metrics and by building a more advanced neural system it can be integrated with mobile and web applications to detect the fraud in real-time and prevent from happening.

7. References

[1] Analysis of Credit Card Fraud Detection Techniques. (2016). *International Journal of Science and Research (IJSR)*, *5*(3), 1302–1307. https://doi.org/10.21275/v5i3.nov162099

[2]Early Prediction of Credit Card Fraud Detection using Isolation Forest Tree and Local Outlier Factor Machine Learning Algorithms. (2020). *Journal of Xidian University*, *14*(6). https://doi.org/10.37896/jxu14.6/077

[3] Vankayalapati, R., Ghutugade, K. B., Vannapuram, R., & Prasanna, B. P. S. (2021). K-Means Algorithm for Clustering of Learners Performance Levels Using Machine Learning Techniques. *Revue d'Intelligence Artificielle*, *35*(1), 99–104. https://doi.org/10.18280/ria.350112

[4] Rajaratne, M. (2018, November 22). *Credit Card Fraud Detection using Autoencoders* https://towardsdatascience.com/credit-card-fraud-detection-using-autoencoders-in-h2o-399cbb7ae4f1

[5] Dzakiyullah, N. (2021). Semi-Supervised Classification on Credit Card Fraud Detection using AutoEncoders. *Journal of Applied Data Sciences*, *2*(1), 1–7. https://doi.org/10.47738/jads.v2i1.16

[6] L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, & S. Karishma. (2020). Credit Card Fraud Detection using Classification, Unsupervised, Neural Networks Models. *International Journal of Engineering Research And*, *V9*(04). https://doi.org/10.17577/ijertv9is040749

*[7]*Gillis, A. S. (2021, September 30). fraud detection. SearchSecurity. https://www.techtarget.com/searchsecurity/definition/fraud-detection

[8] Credit Card Fraud Detection. (2018, March 23).Kaggle.https://www.kaggle.com/mlg-ulb/creditcardfraud

[9]*OpenML*. (n.d.). Open Ml. Retrieved April 7, 2022, from https://new.openml.org/search?type=data&sort=runs&id=1597&status=active

[10]*Credit card fraud detection using artificial neural network*. (2021, June 1). ScienceDirect. Retrieved April 7, 2022, from https://www.sciencedirect.com/science/article/pii/S2666285X21000066

[11]*Confusion Matrix - an overview | ScienceDirect Topics*. (n.d.). Science Direct. Retrieved April 7, 2022,

from https://www.sciencedirect.com/topics/engineering/confusion-matrix

[12]Analysis of Credit Card Fraud Detection Techniques. (2016). *International Journal of Science and Research (IJSR)*, *5*(3), 1302–1307. https://doi.org/10.21275/v5i3.nov162099

[13]*What is Logistic regression? | IBM*. (n.d.). Ibm. Retrieved March 10, 2022, from https://www.ibm.com/topics/logistic-regression

[14] sklearn.metrics.confusion_matrix. (n.d.). Scikit-Learn. Retrieved April 7, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

[15] Lin, T. H., & Jiang, J. R. (2021). Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest. Mathematics, 9(21), 2683. https://doi.org/10.3390/math9212683

[16] 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), 3373–3375. https://doi.org/10.1166/jctn.2018.7628

[17] L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, & S. Karishma. (2020). Credit Card Fraud Detection using Classification, Unsupervised, Neural Networks Models. International Journal of Engineering Research And, V9(04). https://doi.org/10.17577/ijertv9is040749

[18] Vaishali, V. (2014). Fraud Detection in Credit Card by Clustering Approach. International Journal of Computer Applications, 98(3), 29–32. https://doi.org/10.5120/17164-7225

[19] John, H., & Naaz, S. (2019). Credit Card Fraud Detection using Local Outlier Factor and Isolation Forest. International Journal of Computer Sciences and Engineering, 7(4), 1060–1064. https://doi.org/10.26438/ijcse/v7i4.10601064

[20] Kumar, T. (2021). Comparison of Logistic Regression and Decision Tree method for Credit Card Fraud Detection. International Journal for Research in Applied Science and Engineering Technology, 9(5), 680–683. https://doi.org/10.22214/ijraset.2021.34241

[21] Survey of Data-mining Techniques used in Fraud Detection and Prevention. (n.d.). Science Alert. Retrieved March 16, 2022, from https://scialert.net/fulltext/?doi=itj.2011.710.716

[22]*Neural Networks*. (2021, August 4). Neural Network. Retrieved March 18, 2022, from https://www.ibm.com/in-en/cloud/learn/neural-networks