ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022

Comparison of Supervised Learning Algorithms for Credit Card Fraud Detection

[1*]B. V. Ramana, [2] G. Nageswara Rao, [3] T. Ravi Kumar, [4]B. R. Sarath Kumar
[1,2] Dept of IT, Aditya Institute of Technology and Management, Tekkali, AP, India
[3] Dept of CSE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[4]Dept of CSE, Lenora College of Engineering, Rampachodavaram, A.P, India
[*]Corresponding Author: ramana.bendi@gmail.com

Abstract
Fraud in credit card transactions is common today as most of us are using the credit card payment methods more frequently. This is due to the advancement of Technology and increase in online transaction resulting in frauds causing huge financial loss. Therefore, there is a need for effective methods to reduce the loss. In addition, fraudsters find ways to steal the credit card information of the user by sending fake SMS and calls, also through masquerading attack, phishing attack and so on. This paper aims in using the multiple algorithms of Machine learning such as Support Vector Machine, K -Nearest Neighbors (KNN), Random Forest algorithm in predicting the occurrence of the fraud. Further, we conduct a differentiation of the accomplished supervised machine learning techniques to differentiate between fraud and non-fraud transactions.

Keywords: Credit card, K-Nearest Neighbor, Random Forest, Support Vector Machine, Detection Techniques, Machine Learning, Fraud Detection.

1. Introduction

Credit cards are greater secure than ever, with regulators, card providers and banks taking massive time and effort to collaborate with investigators to make sure fraudsters are not successful. Cardholders' money is normally blanketed from scammers with regulations that make the card company and bank accountable. The generation and protection measures in the back of credit cards are becoming more and more sophisticated making it tougher for fraudsters to steel cash. While this is probably detected after thorough historical past assessments, if done, this could permit criminals to apply a valid credit card with a fake paper trail. A similar form of fraud entails taking up a legitimate credit card account by posing as the patron using a similar faux paper trail. The following are the types of credit card frauds.

1. Card-not-present (CNP) fraud
2. Counterfeit and skimming fraud
3. Lost and stolen card fraud
4. Card-by never arrived-fraud.
5. False software / application fraud

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022**

The credit card fraud [1-3] detection takes area as the person or the patron enters the important credentials with a purpose to make any transaction using credit card and the transaction need to get authorized best upon being checked for any fraud interest. For this to manifest, we first bypass the transaction details to the verification module where, it is categorized underneath fraud and non-fraud categories. Any transaction this is positioned underneath fraud category is rejected. Otherwise, the transaction gets authorized.

Main demanding situations involved in credit card fraud detection are as follows:

1. Enormous Data is processed each day and the version construct have to be fast enough to respond to the scam in time.
2. Imbalanced Data i.e. most of the transactions (99.8%) are not fraudulent which makes it in reality tough for detecting the fraudulent ones
3. Data availability because the data is basically private.
4. Misclassified Data may be any other predominant problem, as not each fraudulent transaction is stuck and reported.
5. Adaptive techniques used against the model by using the scammers.

2. Classifications of credit card frauds

1. Application fraud: When a fraudster acquires the control over the application, Steals the credentials of purchaser, and makes a faux account after which the transactions takes place.

2. Electronic or manual card imprints: In this form of fraud, the fraudster skims the data from the magnetic strip which is present on the card then uses the credentials and fraud transactions are performed.

3. Card not present: This is a type of credit card in which physical card is not present during transaction

4. Counterfeit card fraud: The fraud kind wherein the fraudster will copy all the information from magnetic strip and actual card looks as if authentic card and works as original card most effectively and this card used for fraud.

5. Lost/stolen card: This form of fraud is because of losing of the card by means of the cardholder or by the stealing the card from the cardholder.

6. Card id theft: The type of fraud wherein the id of the cardholder is stolen, and the fraud is done.

7. Mail non-received card fraud: While issuing the credit card there will be procedure of sending a mail to the recipient, here fraud can occur by defrauding the mail or phishing.

8. Account Takeover: Here the fraudster will take complete control of the account holder and fraud will be done.

9. Fake fraud in website: Fraudster will introduce a malicious code which does their work in the website.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022**

10. Merchant collision: In this fraud type, cardholder details are shared third party or the fraudster by merchants without cardholder authorization.

The fraud in credit card transaction takes place when the stealer makes use of the other man or woman card without authorization of the respective person via stealing the necessary information like PIN, password and other credentials with or without the physical card. Using fraud detection module related to device studying, we can discover whether the upcoming transaction is fraud or valid. Machine Learning [4-6] is the treading and maximum used technology due to its diverse applications and much less time consumption, more correct in result.

Machine learning is an era that offers with the set of rules, which provides the pc, a capability to look at and boost thru experience without being explicitly programmed. Machine learning [7-8] has applications in a couple of fields. For example, medical, diagnosis, regression etc. Machine

learning [9-10] involves the combination of algorithm and statically models which allow computer to perform the task without hard coding then a model is build through a training data and then it is tested on the trained model.

3. Methodology

This section explains about the implementation, which includes the algorithm used for implementation of proposed system. In this paper, Implementations start from loading the dataset. Then data pre-processing is carried out that includes data cleansing and normalizing the data. Dataset is split into two datasets as train data and test data and model is trained and tested. Finally, the system predicts whether transaction is fraud or non-fraud.

In the implementation of the proposed system, we used python as a programming language. Python is a beginner's language, which provides various applications. In recent years, python has set the new trend because it is easy to use, interpreted, object- oriented, high-level, scripting language. Python is one of the best languages for implementing machine learning. It provides rich packages and libraries that are used in machine learning. The following python libraries and packages used in this paper.

Numpy is a python library. Abbreviation of Numpy is numerical python library. Numpy package is used for multidimensional arrays and linear algebraic operations.

Pandas is a python library. Pandas is used for data analysis and data manipulation tool. It is used to read the dataset and load the dataset. It is fast, flexible when working with data.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022

Scikitlearn a python package which is suitable for statistical model and machine learning models. A best suited python package for machine learning modeling. Keras is advanced stage of neural network application programming interface (API). It is able of run on top of tensor flow. Keras is mainly used while implementing deep learning algorithms such as CNN, RNN because its user friendly, modularity, and easy to extensibility. It runs on both CPU and GPU. In the experiment of finding the fraud or non fraud credit card transaction we had used Keras along with backend running tensor flow. This Keras along with tenor flow backend makes excellent choice for training neural network architecture.

The Machine learning algorithms used in this project are:

1. K-Nearest Neighbour
2. Support Vector Machine
3. Neural Network Algorithm

3.1. K-Nearest Neighbour

A simple, clean-to-enforce supervised machine-gaining knowledge of technique that uses classified enter statistics to increase a feature that gives a suitable output while given extra unlabelled facts. Both classification and regression problems may be solved with the k-nearest neighbor (KNN) algorithm, that's quick and straightforward to use. Uses labeled facts to teach a feature that generates an acceptable overall performance for brand new facts. In the K-Nearest Neighbor set of rules, the resemblance among the new case and the instances which might be already categorized is calculated. Once the new case is positioned in a class this is most corresponding to the available ones, it is carried out to all remaining instances in that institution. In an analogous style, KNN organizes all on hand information and categorizes new factors relying on how similar they're. This describes each time new records emerges, it is just a remember of becoming a K-N category scheme to it. The algorithm is very truthful and simple to place into practice. If a version does now not want to be constructed, so a few parameters and expectations may be tuned, it's miles needless. The algorithm receives appreciably slower as predictors/independent variables increase. As shown in the figure below:

Step 1: START
Step 2: Loading of dataset pd.read.csv (csv file) # reads the csv file and loads
Step 3: Cleaning and normalization of data
Normal = 0
Fraud = 1 # resampling
Data is scaled and normalized
Train_test_split() # splitting of dataset into train and test data
Step 4: Train the model then fit the trained model
Trained the data using Knn classifier
KNeighborsClassifier() # knn classifier which does classification of transactions

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022

Step 5: Calculating the number of fraud, valid transactions and recall, precision and accuracy
calculated.
Step 6: STOP

3.2. Support Vector Machine

Support Vector Machine are a popular machine learning method for classification, regression, & other learning tasks. LIBSVM is a library for Support Vector Machines (SVMs). A typical use of LIBSVM involves two steps: first, training a data set to obtain a model & second, using the model to predict information of a testing data set. For SVC & SVR, LIBSVM can also output probability estimates. Many extensions of LIBSVM are available at libsvmtools. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane.

Step 1: START
Step 2: Loading and observing the dataset
pd.read.csv(.csv) # reads the dataset
resampling of data
StandardScaler() #scaling and normalization of data
Step 3: Data pre-processing
Train_test_split() #Splitting of data
Step 4: Training the model

Dense() #Adding data to activation function
Step 5: Analyzing the model
Prediction of fraud is made and this trained data is stored .it can used to test
(training the model takes longer time so it is stored)
Step 6: STOP

3.3. Artificial Neural Network

Artificial Neural Networks are a class of machine learning algorithms inspired by the structure and functioning of the human brain. They are widely used for various tasks, including image recognition, natural language processing etc. An Artificial Neural Network consists of interconnected nodes, often referred to as neurons or units. These nodes are organized into layers: an input layer, one or more hidden layers, and an output layer. Information flows through the network from the input layer, passes through the hidden layers, and results in an output at the output layer.

The ANN algorithm has two parts:
Training part and testing part.
Training part:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022**

Def ANN:
Step 1: START
Step 2: Loading and observing the dataset
- pd.read.csv(.csv) # reads the dataset
- resampling of data
- StandardScaler() #scaling and normalization of data

Step 3: Data pre-processing
- Train_test_split() #Splitting of data

Step 4: Training the model
- Dense() #Adding data to activation function

Step 5: Analyzing the model
- ·Prediction of fraud is made and this trained data is stored .it can used to test (training the model takes longer time so it is stored)

Step 6: STOP

4. Results

The experimentation was done with machine learning algorithms that are K-Nearest Neighbor, Support Vector Machine and Artificial Neural Network. The results indicate that The Artificial Neural Network algorithm gives highest accuracy than other algorithms.

Machine Learning Algorithms	**Accuracy**
K-Nearest Neighbor	0.94
Support Vector Machine	0.90
Artificial Neural Network	0.99

5. Conclusion

In this research, we have proposed a method to detect fraud in credit card transactions that is based on deep learning. We first compare it with machine learning algorithms such as K-Nearest Neighbor, Support vector machine etc. Finally, we have used the neural network, even though tough to train the model which would fit fine to model for detecting afraud in credit card Transactions. In our model, using Machine Learning Algorithms which gives accuracy approximately equal to 99% is best suited for credit card fraud detection. It gives accuracy more than that of the unsupervised learning algorithms. In this research work, data pre-processing, normalization and under-sampling was carried out to overcome the problems faced by using an imbalanced dataset. Further, by using Deep learning we can detect Credit Card Frauds and Calculate the Accuracy.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved,** UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, July 2022

References:

1. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," in IEEE Access, vol. 8, pp. 2557925587, 2020, doi: 10.1109/ACCESS.2020.2971354.
2. S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid and H. Zeineddine, "An Experimental Study with Imbalanced Classification Approaches for Credit Card Fraud Detection," in IEEE Access, vol. 7, pp. 93010-93022, 2019, doi: 10.1109/ACCESS.2019.2927266.
3. Jiang, J. Song, G. Liu, L. Zheng and W. Luan, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," in IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3637-3647, Oct. 2018, doi: 10.1109/JIOT.2018.2816007.
4. Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar. 2018. Ensemble learning for credit card fraud detection. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data Association for Computing Machinery, New York, NY, USA, 289–294. DOI: https://doi.org/10.1145/3152494.3156815
5. Phuong Hanh Tran, Kim Phuc Tran, Truong Thu Huong, Cédric Heuchenne, Phuong HienTran, and Thi Minh Huong Le. 2018. Real Time Data-Driven Approaches for Credit Card Fraud Detection. In Proceedings of the 2018 International Conference on EBusiness and Application. Association for Computing Machinery, New York, NY, USA, 6–9. DOI: https://doi.org/10.1145/3194188.3194196
6. Imane Sadgali, Nawal Sael, and Faouzia Benabbou. 2019. Fraud detection in credit card transaction using neural networks. In Proceedings of the 4th International Conference on Smart City Applications (SCA '19). Association for Computing Machinery, New York, NY, USA, Article 95, 1–4. DOI: https://doi.org/10.1145/3368756.3369082
7. Prusti and S. K. Rath, "Web service based credit card fraud detection by applying machine learning techniques," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 492-497, doi: 10.1109/TENCON.2019.8929372. 16
8. M. Zamini and G. Montazer, "Credit Card Fraud Detection using autoencoders based clustering," 2018 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp. 486-491, doi: 10.1109/ISTEL.2018.8661129.
9. S. Akila and U. S. Reddy, "Credit Card Fraud Detection Using Non-Overlapped Risk Based Bagging Ensemble (NRBE)," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, 2017, pp. 1-4, doi:10.1109/ICCIC.2017.8524418.
10. M. S. Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika and E. Aswini, "Credit Card Fraud Detection Using Random Forest Algorithm," 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2019, pp. 149-153, doi: 10.1109/ICCCT2.2019.8824930.