iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VIII Month of publication: August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37696

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Identity Theft Detection Using Machine Learning

Vanshita Agarwal
UG Student, Dept. of I.T., TCET, University of Mumbai, India

Abstract: Online Identity Theft is known to be quite possibly the most genuine and developing dangers to people and organizations for the past decade because of the colossal financial harm caused by it. It is a critical form of cybercrime which uses information such a victim's name, bank details, email address, passwords, passport or identification details, and other valuable information to gain access to accounts. This aim of this research is to make use of Machine Learning algorithms for the detection of such crimes.

Keywords: identity theft; cybercrime; Machine Learning; detection; algorithms.

I. INTRODUCTION

The things people used to buy at shops years ago are now purchased online. Online users surf cyberspace to exchange information, to interact, to entertain, to buy and sell goods and services, to access and use public and private transport, to share their interests and to make financial transactions across a wide range of industries, both nationally and internationally.

When this confidential information is somehow obtained by a third party, it can be used to impersonate the victim for several reasons, such as opening a new account, gaining influence over the victim's credit account, gaining government benefits.

Identity Theft Detection becomes possible using Machine Learning due to the ability of ML algorithms to learn from historical fraud patterns and recognize them in future transactions. Machine Learning is the general term for analytical algorithms and systems that 'learn' patterns from samples and datasets to an extent where one can be largely free of human interference.

Machine Learning in ID theft detection is vital since it is a lot quicker cycle than manual investigation and is considerably more useful to scale when bigger associations and swarms of information are in question. Additionally, ML algorithms can find sophisticated fraud traits that a human might not be able to detect.

II. RELATED WORK

In [1] the author proposed a novel fraud detection method that has four stages they first utilize the historical transaction data to divide them into groups to form clusters of transactions having the same behaviour then thus they came up with a sliding window strategy to aggregate transactions. This algorithm is used to characterize the behavioural pattern of a cardholder then after aggregation, we use the new window formed the feature extraction is done. At last, the classification takes place and classifies behavioural patterns and assignments. As a result, their method of Logistic Regression with raw data (RawLR), Random Forest with aggregation data (AggRF), and Random Forest and feedback technique with aggregation data (AggRF +FB) are the best method with 80% accuracy as compared to other methods.

In [2] they applied supervised machine learning algorithms on the real-world data set and then used those algorithms to implement a super classifier using ensemble learning and then they compared the performance of supervised algorithms with their implementation of a super classifier. They used ten machine learning algorithms such as Random Forest, Stacking Classifier, XGB Classifier, Gradient Boosting, Logistic Regression, MLP Classifier, SVM, Decision Tree, KNN, Naïve Bayes. And compared the accuracy, Recall Precision, confusion matrix with the result of their super classifier. As a result, they found that the Logistic Regression is better for predicting fraud transactions.

In [3] They used twelve machine learning algorithms for credit card fraud detection in which their range standard from a neural network to deep learning. They are tracing the performance of benchmark and real-world datasets. In addition, the AdaBoost and majority voting methods are applied for forming the hybrid models. As their related study explains about single and hybrid models. For both the parameters (Benchmark and real-world datasets) they had given the results using there twelve selected algorithms that are Naïve Bayes, Random Forest, Decision Tree, Gradient Boosted Tree, Decision Stump, Random Tree, Neural Network, Linear Regression, Deep Learning, Logistic Regression, SVM, Multilayer Perceptron. As a result, when standard algorithms used with AdaBoost and majority voting methods under benchmark data the best accuracy and sensitivity acquired by Random Forest algorithm 95% and 91% respectively. When experimented with real-world data the accuracy rate is still above 90% even with 30% noise in the dataset. MCC (Mathews correlation coefficient) is standard to measure the performance of a model so in case of majority voting the best MCC score is 0.823 whereas 0.942 with 30% of noise added to the dataset.

In [4] the author proposed the various machine learning algorithms and analysed them concerning to credit card fraud detection methods. The various methods of machine learning are Logistic Regression, Naïve Bayes, Random Forest, Multilayer Perceptron. Here for multilayer perceptron (ANN) is used (Artificial neural network) which consist of 4 hidden layers and relu activation functioned is used that is to avoid negative values and optimizer used is Adam for its best performance. As a result for Logistic regression the accuracy score is 97.46% with the data set containing 56962 samples in which 98 fraud transactions. For the same dataset Naïve Bayes and Random Forest, accuracy score is 99.23% and 99.96% respectively. At last for ANN it was 99.93% of accuracy as we can observe that random forest gives the best result in case of credit card fraud detection.

III.PROPOSED ALGORITHM

A. *Using Neural Networks*

Neural Network is a concept inspired by the working of a human brain. Neural networks in Deep Learning uses different layers for computation. It uses cognitive computing that helps in building machines capable of using self-learning algorithms that involve the use of data mining, pattern recognition, and natural language processing. It is trained on a dataset passing it through different layers several times. It gives more accurate results than other models as it uses cognitive computing and it learns from the patterns of authorized behaviour and thus distinguishes between ‘fraud’ and ‘genuine’ transactions.

B. *Description of the Proposed Algorithm*

Aim of the proposed algorithm is to use the different layers in a neural network that focus on different parameters to decide whether a transaction is ‘fraud’ or ‘non-fraud.’

1) *Step 1: Training the Model*

The data is fed to the Neural Network:

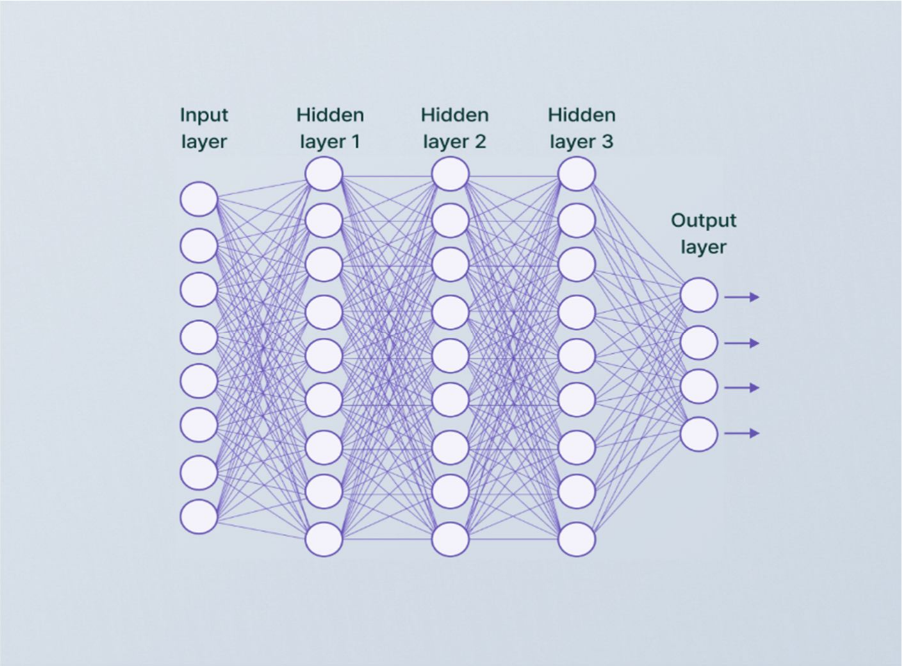

Fig 1. Hidden Layers in a Neural Network

2) *Step 2: Testing Data*

When a Transaction is initiated, the Hidden Layer 1 checks the amount of transaction.

Similarly, the other layers check for the location, identity, IP address of the location, the frequency of transaction, and the mode of payment.

There can be more business-specific parameters. These individual layers work on these parameters, and computation is done based on the models’ self-learning and experience to calculate the probabilities for detecting frauds.

Neural networks work on data and learn from it, and it improves the model’s performance over every iteration.

IV. PSEUDO CODE

A. Training Part

1) *Step 1:* Loading and observing the dataset.

```
pd.read.csv(.csv) # reads the dataset
resampling of data
StandardScaler() #scaling and normalization of data
```

2) *Step 2:* Data pre-processing

```
Train_test_split() #Splitting of data
```

3) *Step 3:* Training the model

```
Dense() #Adding data to activation function
```

4) *Step 4:* 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)

B. Testing Part

It is carried out similar way only difference is that the stored trained model is used to test the data and classify it.

V. EVALUATION OF RESULTS

The end result is evaluated based on the confusion matrix and precision, recall and accuracy is calculated.
It contains two classes: actual class and predicted class. The confusion metrics depends on these features:

1) *True Positive:* In which both the values positive that is 1.
2) *True Negative:* It is case where both values are negative that is 0.
3) *False Positive:* this is the case where true class is 0 and non-true class is 1.
4) *False Negative:* It is the case when actual class is 1 and non-true class is 0.

Precision defined as follows:
Precision = true positive / Actual result
Precision = true positive/(true positive + false positive)
Recall defined as follows:
Recall = true positive / predicted result
Recall = true positive/(true positive + false negative)
Accuracy defined as:
Accuracy = (true positive + true negative)/ total

VI. CONCLUSION AND FUTURE WORK

This research implements a neural network algorithm for identity theft detection. In this model, by using an artificial neural network (ANN) which gives accuracy approximately equal to 100% is best suited for identity theft detection. It gives accuracy more than that of the unsupervised learning algorithms. In this research work, data pre-processing, normalization and under-sampling carried out to overcome the problems faced by using an imbalanced dataset. Future work will investigate the improvement of this algorithm. Future work can also include combining this algorithm with different algorithms and applying new technologies.

REFERENCES

[1] Xuan Shivang ,"Random forest for credit card fraud detection.", 2018 IEEE 15th International Conference on Networking Sensing and Control (ICNSC). IEEE, 2018.

[2] Dhankhad, Sahil, Emad Mohammed, and Behrouz Far. "Supervised machine learning algorithms for credit card fraudulent transaction detection: : a comparative study." 2018 IEEE international conference on information reuse and integration (IRI). IEEE, 2018

[3] Randhawa, Kuldeep, et al. "Credit card fraud detection using AdaBoost and majority voting." IEEE access 6 (2018)): 14277-14284.

[4] Varmedja, Dejan, et al. "Credit card fraud detection-machine learning methods." 2019 18th International Symposium INFOTEH-JAHORINA(INFOTEH) IEEE 2019.

[5] Suresh K Shirgave, Chetan J. Awati , Rashmi More, Sonam S. Patil,A Review Review on Credit Card Fraud Detection Using Machine Learning, , Published in International journal of Science and Technology Research, vol.no.8, Issue no. 10, October 2019, pp. 1217-1220.

[6] Sadgali, N. Sael, F. Benabbou, Fraud detection in credit card transaction using neural networks, Proceedings of the 4th International Conference on Smart City Applications (SCA '19). Association for Computing Machinery, New York, NY, USA (2019), pp. 1-4

[7] D. Prusti, S.K. Rath,Web service based credit card fraud detection by applying machine learning techniques, Proceedings of the TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India (2019), pp. 492-497

[8] P. Kumar, F. Iqbal,Credit card fraud identification using machine learning approaches,Proceedings of the 1st International Conference on Innovations in Information and Communication Technology (ICIICT), CHENNAI, India (2019), pp. 1-4

[9] Mukkamala, S., G. Janoski, and A. Sung. Intrusion detection: support vector machines and neural networks. in proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO. 2002.

[10] Geeta, D.V., Online identity theft–an Indian perspective. Journal of Financial Crime, 2011. 18(3): p. 235-246.

[11] Allison, S.F.H., A.M. Schuck, and K.M. Lersch, Exploring the crime of identity theft: Prevalence, clearance rates, and victim/offender characteristics. Journal of Criminal Justice, 2017. 33(1):19- 29.

[12] Yucheol Cho and Sangjin Lee, “Detection and Response of Identity Theft within a Company Utilizing Location Information”, IEEE Xplore, 2016.