CREDIT CARD FRAUD DETECTION SYSTEM

NASSCOM
REVIEW - 03

BY –
ROHAN ALLEN
18BCI0247

PROBLEM STATEMENT

- The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. The aim of the project here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

ABSTRACT

- The main purpose of this project is understand and implement the distinct approach of Isolation forest algorithm and LOF algorithm to identify fraudulent transactions in a database instead of using generic Random Forest approach.
- The model will be able to identify the transactions with greater accuracy and by comparing both the approaches we will be working towards a more optimal solution.

INTRODUCTION

The following are reasons why we need to develop a robust system to detect fraudulent transactions. The following challenges have to be overcome to make sure that the final product is resilient.

- The challenge is to recognize fraudulent credit card transactions so that the customers of credit card companies are not charged for items that they did not purchase. Main challenges involved in credit card fraud detection are:
- Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.
- Imbalanced Data i.e. most of the transactions (99.8%) are not fraudulent which makes it really hard for detecting the fraudulent ones Data availability as the data is mostly private.
- Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported.
- Adaptive techniques used against the model by the scammers.

OBJECTIVES

- To identify fraudulent transactions
- To reduce number of false positives and false negatives during detection process
- To maximize efficiency and throughput

PROPOSED METHODOLOGY - ISOLATION FOREST ALGORITHM

- Isolation forest is an unsupervised learning algorithm for anomaly detection that works on the principle of isolating anomalies, instead of the most common techniques of profiling normal points.
- In statistics, an anomaly (a.k.a. outlier) is an observation or event that deviates so much from other events to arouse suspicion it was generated by a different mean.
- Anomalies in a big dataset may follow very complicated patterns, which are difficult to detect “by eye” in the great majority of cases. This is the reason why the field of anomaly detection is well suited for the application of Machine Learning techniques.

PROPOSED METHDOLOGY(CONT.)

- The most common techniques employed for anomaly detection are based on the construction of a profile of what is “normal”: anomalies are reported as those instances in the dataset that do not conform to the normal profile.
- Isolation Forest uses a different approach: instead of trying to build a model of normal instances, it explicitly isolates anomalous points in the dataset. The main advantage of this approach is the possibility of exploiting sampling techniques to an extent that is not allowed to the profile-based methods, creating a very fast algorithm with a low memory demand.

PROPOSED METHODOLOGY - LOCAL OUTLIER FACTOR ALGORITM

- Similar to K-NN but not exactly KNN
- Branch of Unsupervised ML
- Calculates outlier density on the basis of existing points surrounding every point and hence more efficient.

LITERATURE SURVEY

- Chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). **Distributed data mining** in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, *14*(6), 67-74.
- Ghosh, S., & Reilly, D. L. (1994, January). Credit card fraud detection with a **neural-network**. In *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on* (Vol. 3, pp. 621-630). IEEE.
- Raj, S. B. E., & Portia, A. A. (2011, March). Analysis on credit card fraud detection methods. In *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)* (pp. 152-156). IEEE.
- Brause, R., Langsdorf, T., & Hepp, M. (1999, November). **Neural data mining** for credit card fraud detection. In *Proceedings 11th International Conference on Tools with Artificial Intelligence* (pp. 103-106). IEEE.
- Srivastava, A., Kundu, A., Sural, S., & Majumdar, A. (2008). Credit card fraud detection using **hidden Markov model**. *IEEE Transactions on dependable and secure computing*, *5*(1), 37-48.

LITERATURE SURVEY

- Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Credit card fraud detection using **Bayesian and neural networks**. In *Proceedings of the 1st international naiso congress on neuro fuzzy technologies* (pp. 261-270).
- Aleskerov, E., Freisleben, B., & Rao, B. (1997, March). Cardwatch: A neural network based **database mining system** for credit card fraud detection. In *Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr)* (pp. 220-226). IEEE.
- Quah, J. T., & Sriganesh, M. (2008). Real-time credit card fraud detection using **computational intelligence**. *Expert systems with applications*, *35*(4), 1721-1732.
- Ogwueleka, F. N. (2011). **Data mining** application in credit card fraud detection system. *Journal of Engineering Science and Technology*, *6*(3), 311-322.
- Panigrahi, S., Kundu, A., Sural, S., & Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using **Dempster-Shafer theory and Bayesian learning**. *Information Fusion*, *10*(4), 354-363.

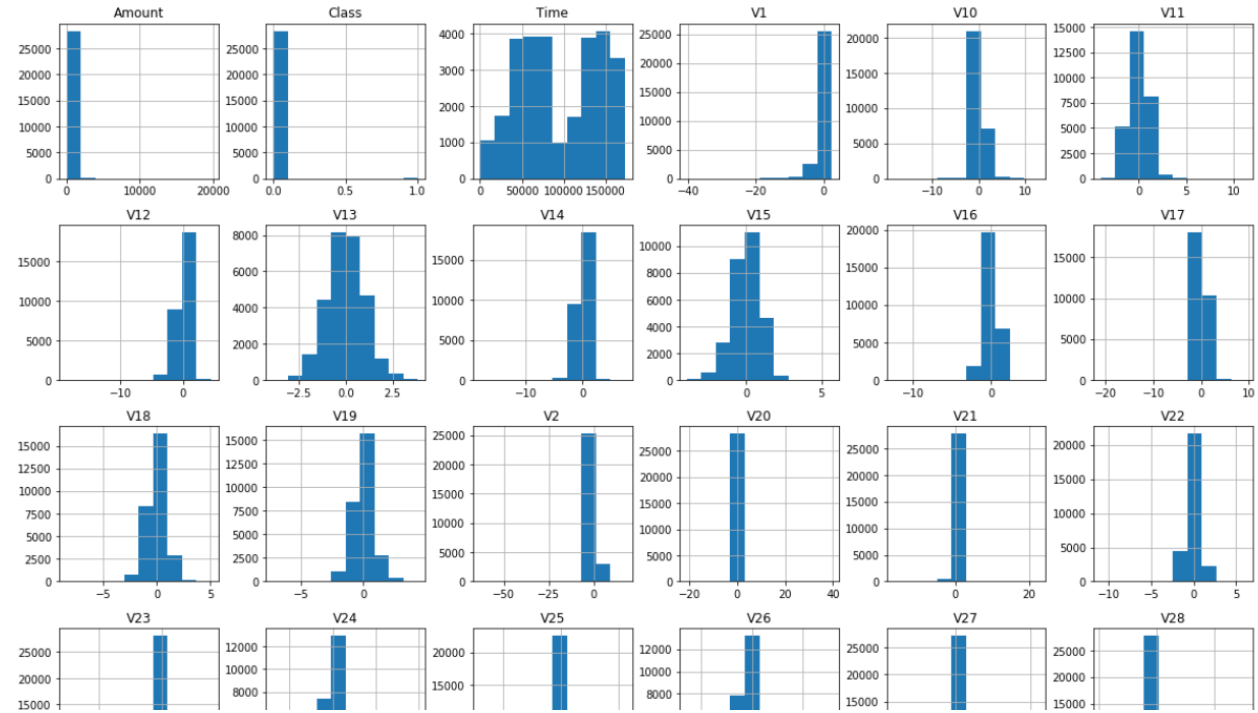
LITERATURE SURVEY

- Syeda, M., Zhang, Y. Q., & Pan, Y. (2002, May). **Parallel granular neural networks** for fast credit card fraud detection. In *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291)* (Vol. 1, pp. 572-577). IEEE.
- Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert systems with applications*, *41*(10), 4915-4928.
- Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997, July). Credit card fraud detection using meta-learning: Issues and initial results. In *AAAI-97 Workshop on Fraud Detection and Risk Management* (pp. 83-90).
- Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018**). Sequence classification** for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234-245.
- Carneiro, N., Figueira, G., & Costa, M. (2017). A **data mining** based system for credit-card fraud detection in e-tail. *Decision Support Systems*, *95*, 91-101.

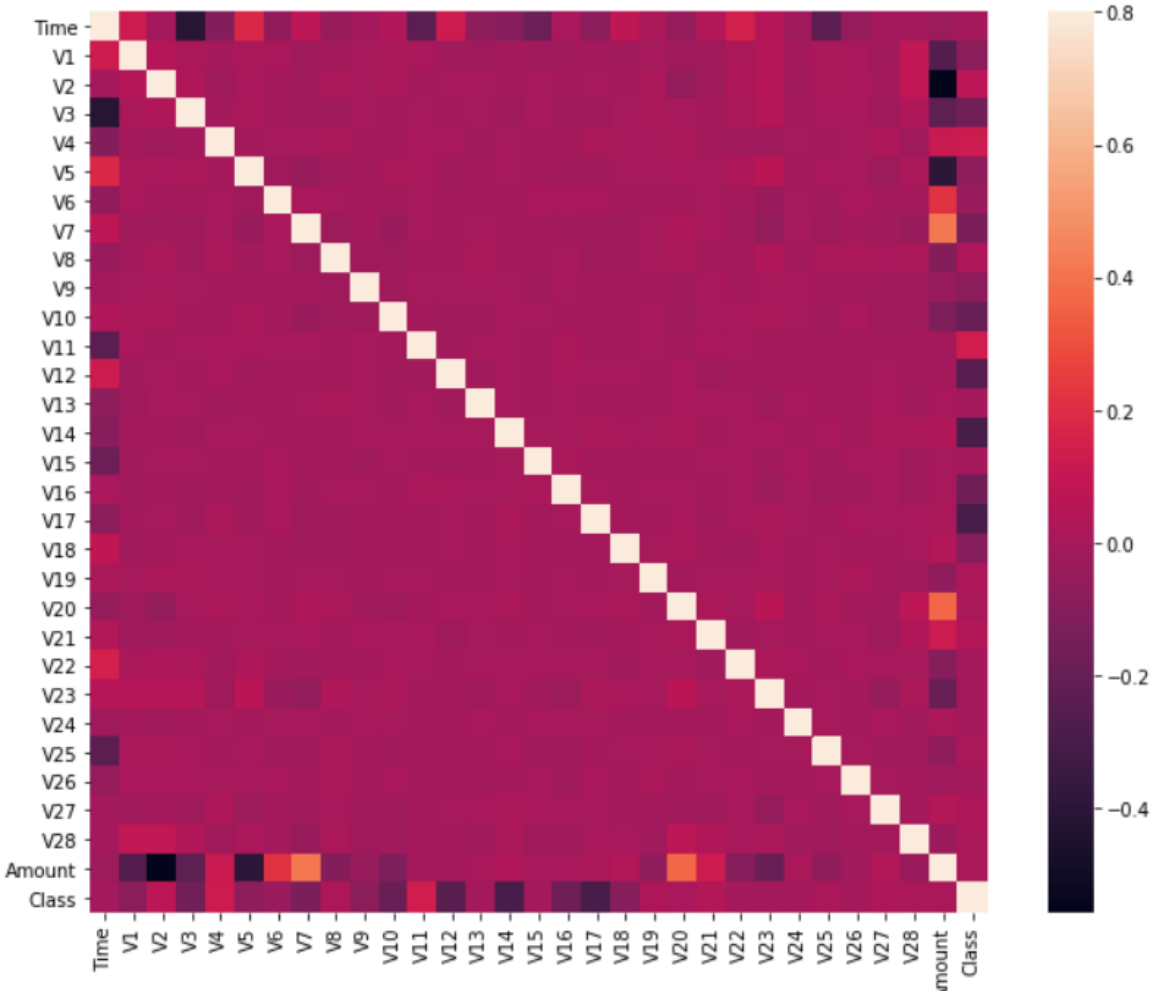
LITERATURE SURVEY

- Zhu, Kevin. "**Phone usage pattern** as credit card fraud detection trigger." U.S. Patent No. 8,386,386. 26 Feb. 2013.
- Shen, A., Tong, R., & Deng, Y. (2007, June). Application of classification models on credit card fraud detection. In *2007 International conference on service systems and service management* (pp. 1-4). IEEE.
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). **Transaction aggregation as a strategy** for credit card fraud detection. *Data mining and knowledge discovery*, *18*(1), 30-55.
- Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). **Feature engineering strategies** for credit card fraud detection. *Expert Systems with Applications*, *51*, 134-142.
- Patidar, R., & Sharma, L. (2011). Credit card fraud detection using neural network. *International Journal of Soft Computing and Engineering (IJSCE)*, *1*(32-38).

IMPLEMENETATION



IMPLEMENTATION



IMPLEMENTATION OUTCOME

```
Isolation Forest: 71
0.99750711000316
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28432
           1       0.28      0.29      0.28        49

    accuracy                           1.00     28481
   macro avg       0.64      0.64      0.64     28481
weighted avg       1.00      1.00      1.00     28481

Local Outlier Factor: 97
0.9965942207085425
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28432
           1       0.02      0.02      0.02        49

    accuracy                           1.00     28481
   macro avg       0.51      0.51      0.51     28481
weighted avg       1.00      1.00      1.00     28481


<Figure size 648x504 with 0 Axes>
```

PROJECT OUTCOME

- The model is simple and fast enough to detect the anomaly and classify it as a fraudulent transaction as quickly as possible.
- Imbalance is dealt by properly using some Isolation forest algorithm and Local Outlier factor (LOF).
- For protecting the privacy of the user the dimensionality of the data will be reduced. A more trustworthy source will be taken which double-checks the data, at least for training the model.
- The model is interpretable so that when the scammer adapts to it with just some tweaks, we can have a new model up and running to deploy.