RESEARCH ARTICLE OPEN ACCESS

Sequence Classification of Credit Card Fraud Detection

Mr. J. David Sukeerthi Kumar[1], B. Nayab Rasool[2], V. Karthik Kumar Reddy[3], P. Vishnuvardhan Achary[4], P. Abraham[5], P. Hemanth Kumar[6]

[1]Assistant Professor, Department of Computer Science and Engineering,
Santhiram Engineering College, Nandyal
Email: david.cse@srecnandyal.edu.in
[2,3,4,5,6] Students at Department of Computer Science and Engineering,
Santhiram EngineeringCollege, Nandyal
Email:
19x51a0569@srecnandyal.edu.in19x51a0565@srecnandyal.edu.in19x51a05a1@srecnandyal.edu.in19x51a05a6@srecnandyal.edu.in19x51a05a3@srecnandyal.edu.in

--************************----------------------------------

Abstract:

The financial burden of credit-card fraud is becoming a significant concern for financial institutions and service providers as a result of the rising number of electronic payments, compelling them to continuously enhance their fraud detection systems. While data-driven and learning-based approaches are all the rage in certain fields, they are just now beginning to make their way into others. commercial software To include transaction sequences, we frame the fraud detection issue as a sequence classification job and use Long Short-Term Memory (LSTM) networks. Moreover, we use cutting-edge feature aggregation techniques and publish our findings using conventional retrieval measures. In tests comparing the LSTM to a baseline Random Forest (RF) classifier, it was shown that the LSTM provides better identification accuracy for in-store, offline transactions when the cardholder is physically present at the business. Manual feature aggregation procedures are very beneficial to both sequential and non-sequential learning approaches. Subsequent study of positive results showed that the two methods often pick up on distinct types of fraud, suggesting they be used together. Lastly, we explore some of the scientific and practical open questions that remain after completing this research.

--************************----------------------------------

I. INTRODUCTION

As more and more financial transactions take place online and companies move to a cashless banking sector, spotting fraudulent activity is more important than ever. The goal is to make sure that genuine consumers are not harmed by automated and human evaluations, and to reduce the direct losses caused by fraudulent transactions. Detecting fraudulent transactions is a vital task for payment processors due to the prevalence of card fraud. Card fraud happens when someone uses stolen card information to make purchases without the cardholder's knowledge or consent. A fraud detection system often combines automated software with human reviewers. The automated system uses predefined criteria to identify potential cases of fraud. It examines all freshly received transactions and rates their potential for fraud. To do this step by hand, fraud investigators are employed. They provide just binary feedback (fraud or legitimate) on all transactions they assess, focusing on those with a high fraudulent score. Expert-driven rules, data-driven rules, or a hybrid of the two are all possible foundations for a fraud detection system. Using their findings, fraud investigators strive to pinpoint particular fraud situations that can be identified using guidelines developed by experts in the field. A fraudulent scenario may include "a cardholder doing a transaction in a given nation, and then, in the following two weeks, (s)he conducts another transaction for a given amount in another given country." The stream of transactions will be analyzed for this pattern, and an alert will be generated if it is found. Rules that are "data driven" are developed using machine learning techniques. The system is trained to recognize fraudulent behavior in a continuous flow of transaction data. Logistic regression, support vector machines, and random forests are some of the most frequently used machine learning algorithms detecting fraud is a difficult machine learning task. The fraud detection issue is inherently a sequential classification work, because I the data distribution changes over time due to seasonality and new attack techniques, (ii) fraudulent transactions only account for a tiny proportion of total daily transactions, and (iii) there are relatively few of them. Here, we suggest adopting LSTM networks as machine learning techniques in fraud detection systems to largely solve the latter problem. We compile current advances in machine learning and credit card fraud detection, and demonstrate via an in-depth analysis how these improvements may be incorporated into a data-driven fraud detection system. We give empirical findings on a real-world credit-card transaction dataset and analyze the accuracy of categorization for both online (e-commerce) and offline (point-of-sale or face-to-face) purchases. Further, we address common problems with using a sequence learner, such an LSTM, in this setting.

Credit card fraud detection. Our specific contributions are as follows:

We evaluate a sequence learner (Long Short-Term Memory) and a static learner (Random Forest) on a real-world fraud detection dataset.

We demonstrate that using either a sequence learner or feature engineering to infer information about past transactions significantly increases the accuracy with which offline transactions may be detected as fraudulent. There is no advantage to using a sequence learner rather than a static learner for online purchases.

We demonstrate that LSTMs reliably identify different frauds than Random Forests do. This is true for both online and physical purchases.

II. RELATED WORK

“Methods for detecting fraud detection systems: a review.”

Most monetary transactions are now conducted through electronic commerce systems like the credit card system, the telephone system, the healthcare insurance system, etc., thanks to the rise in the usage of computers and the persistence of businesses. Unfortunately, both honest people and con artists use these techniques. Furthermore, fraudsters used a variety of techniques to get into e-commerce platforms. The current state of fraud prevention systems (FPSs) cannot guarantee the safety of online marketplaces. However, if FDSs and FPSs work together, it might help strengthen the safety of online transactions. However, FDSs have obstacles and constraints that limit their effectiveness, including idea drift, supporting real-time detection, skewed distribution, a vast volume of data, and so on. The purpose of this study is to present a high-level, all-encompassing summary of the problems that hinder FDS efficiency. We have chosen to focus on the credit card, telecommunication, healthcare insurance, auto insurance, and online auction sectors of e-commerce. The most common forms of fraud in these online marketplaces are presented in detail.

Moreover, the most up-to-date methods for FDSs are carefully explained in the context of a few examples of E-commerce platforms. After that, we provide our last thoughts and a short outlook on future directions for the field of study.

“Techniques based on feature engineering for monitoring credit card transactions for signs of fraudulent activity.”

Global credit card fraud costs businesses and consumers billions of euros annually. As a result, banks are under constant pressure to enhance their fraud prevention tools. Machine learning and data mining have been advocated in recent years as a means of solving this issue by a number of different research. To compare the effectiveness of the various approaches, most studies have relied on misclassification measures rather than accounting for the real costs of fraud detection. Also crucial is knowing what information to pull from a credit card transaction when building a fraud detection model.

financial records or data. Typically, this is accomplished by the aggregation of transactions so that client spending trends may be studied. In this study, we offer a new set of characteristics based on examining the periodic behaviour of the time of a transaction using the von Mises distribution, which represents a generalization of the transaction aggregation technique. We next compare state-of-the-art credit card fraud detection algorithms and assess the effects of various feature sets using an actual credit card fraud dataset given by a major European card processing operator. The findings suggest that savings may be increased by an average of 13% when the recommended periodic characteristics are included into the approaches.

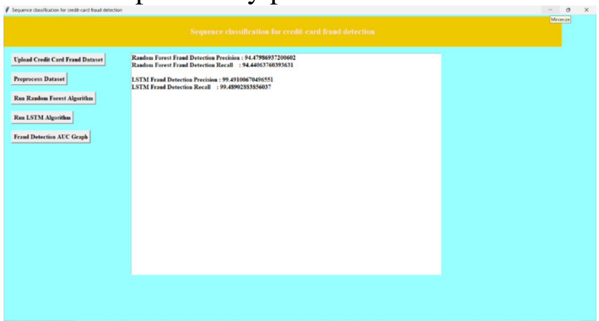
III. METHODOLOGY

One way to look at fraudulent transactions is as outliers in consumers' buying patterns; another is as a distinct class of transactions that stands in contrast to the legit ones. For two reasons, fraudulent and legitimate transactions may easily coexist in the feature space. To start, there is a wide range of diversity in the actual purchasing behaviours of millions of clients. And secondly, fraudsters use a wide variety of sophisticated, but undetectable, methods to carry out fraudulent operations that span several, unrelated client accounts and time periods, yet ultimately manifest themselves as isolated transactions in a dataset.

As a result, the same set of purchasing behaviours might simultaneously represent perfectly normal behaviour in the context of certain clients and glaring abnormalities in the context of others. We found two methods that allow us to summarize customers' transaction histories and then use this summary during the categorization of individual transactions, which aids in the discrimination of transactions that are otherwise difficult to tell apart. The first technique employs a Recurrent Neural Network to describe the transition dynamics between transactions in order to recover the sequential structure of a customer's transaction history In contrast, the second approach is a tried-and-true standard in the field of credit card fraud detection, and it relies on human feature engineering. There is a subset of recurrent neural networks called a Long Short-Term Memory Network (LSTM) (RNN). To better represent time series data, recurrent neural networks were created in the 1980s [Williams and Hinton, 1986, Werbos, 1988, Elman, 1990]. While recurrent neural networks (RNNs) are structurally comparable to traditional multilayer perceptrons, RNNs also permit connections between hidden units that are "in the same layer." with quantifiable, separate chunks of time. Every input sequence has its own index, which is the passage of time. The model is able to detect temporal correlations between events that may seem far apart in the input sequence since it can remember the information from previous inputs thanks to the connections across time steps.In time series, where the occurrence of one event may rely on the occurrence of numerous other occurrences in the past, this is an essential trait for accurate learning.

IV. RESULT AND DISCUSSION

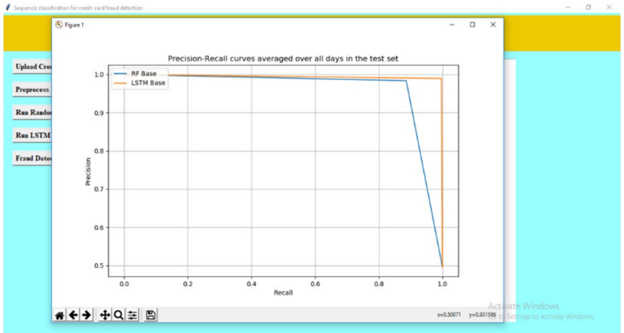
In this study, the author employs the LSTM (long short term memory) neural network technique to identify sequence patterns in a credit card dataset, therefore facilitating the detection of fraudulent activity. Existing algorithms don't take time series data into account for fraud detection, even though fraudsters' behaviour and methods change over time. LSTM is the only algorithm that treats data as a sequence of patterns, using those patterns to forecast whether a given transaction is fraudulent or not. An LSTM trained on time series data will be able to detect fraudulent activity in either newly processed or previously processed transactions.


Sequence classification for credit-card fraud detection
Minimize
Sequence classification for credit-card fraud detection
Upload Credit Card Fraud Dataset
Preprocess Dataset
Run Random Forest Algorithm
Run LSTM Algorithm
Fraud Detection AUC Graph

```
Random Forest Fraud Detection Precision : 94.47986937200602
Random Forest Fraud Detection Recall   : 94.44063760393631

LSTM Fraud Detection Precision : 99.49100670496551
LSTM Fraud Detection Recall   : 99.48902883856037
```

The training phase of Random Forest is complete; the resulting model achieved a 94/93 accuracy rate in fraud detection using the LSTM and a 99% accuracy rate using the LSTM.



Above, we observe that LSTM performs around 1, but random forest performs below 1. The x-axis indicates recall, and the y-axis represents accuracy; the blue line represents random forest, and the orange line represents LSTM. Based on the data shown above, it is clear that LSTM is superior than Random Forest at detecting fraud.

V. CONCLUSIONS

The In this article, we used long short-term memory networks to pool previous credit card purchases into a single profile for better fraud detection. An historical context was not used in the comparison to a baseline classifier. According to the results of our research, traditional There are significant differences in the characteristics of in-person and online transactions with regard to the sequential nature of subsequent transactions. For improved identification of latent sequential patterns in offline transactions, an LSTM is a suitable model. Manually aggregating the transaction history by means of extra characteristics is an alternative to the sequence learner that enhances detection for both offline and online transactions. Nonetheless, across all feature sets, the true positives produced by LSTM modeling of the transaction history are distinguishable from the frauds discovered by Random Forest, giving birth to a combination method.

REFERENCES

[1] Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90–113.

[2] David Sukeerthi Kumar, J., M. V. Subramanyam, and A. P. Siva Kumar. "A Hybrid Spotted Hyena and Whale Optimization Algorithm-Based Load-Balanced Clustering Technique in WSNs." Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022. Singapore: Springer Nature Singapore, 2023.

[3] https://www.scopus.com/authid/detail.uri?authorId=57202806468

[4] Bahnsen, A. C., Aouada, D., & Ottersten, B. (2015). Example-dependent cost-sensitive decision trees. Expert Systems with Applications, 42 (19), 6609–6619.

[5] 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. Bayer, J. S. (2015). Learning sequence representations. München, Technische Universität München, Diss. Ph.D. thesis. Bengio, Y., Simard, P., &Frasconi, P. (1994).

[6] KUMAR, J. DAVID SUKEERTHI. "Implementing the Effective File System of Secondary Memory Management in Wireless Sensor Networks."

[7] https://scholar.google.co.in/citations?user=LCUp-PsAAAAJ&hl=en

[8] Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5 (2), 157–166.

[9] Bhattacharyya, S., Jha, S0., Tharakunnel, K. ,& Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50 (3), 602–613.

[10] Breunig, M. M., Kriegel, H.-P., Ng, R. T. ,& Sander, J. (20 0 0). Lof: identifying densi- ty-based local outliers. In ACM sigmod record: 29 (pp. 93–104).

[11] ACM. Carneiro, N., Figueira, G., & Costa, M. (2017).

[12] https://www.scopus.com/authid/detail.uri?authorId=34975973700

[13] Rao, SCV Ramana, S. Naga Mallik Raj, S. Neeraja, P. Prathusha, and J. David Sukeerthi Kumar. "Flow Controlling of Access at Edge Routers." International Journal of Advanced Computer Science and Applications 1, no. 4 (2010).

[14] A data mining-based system for credit- card fraud detection in e-tail. Decision Support Systems, 95, 91–101. Chawla, N. V., Bowyer, K. W. , Hall, L. O. , &Kegelmeyer, W. P. (2002).

[15] Smote: syn- thetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16 , 321–357 . Cieslak, D. A. ,Hoens, T. R. , Chawla, N. V. , &Kegelmeyer, W. P. (2012).

[16] Maneesha, K., M. Sravani, S. Ayeesha, T. Kavyasree, S. Raziya Begum, and Mr J. David Sukeerthi Kumar. "TRAFFIC SIGN RECOGNITION USING CNN FOR DRIVERLESS CARS."

[17] Hellinger dis- tance decision trees are robust and skew-insensitive. Data Mining and Knowledge Discovery, 24 (1), 136–158.

[18] Collobert, R., Weston, J., Bottou, L. ,Karlen, M. , Kavukcuoglu, K. , &Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12 (Aug), 2493–2537.

[19] Dal Pozzolo, A., &Bontempi, G. (2015). Adaptive machine learning for credit card fraud detection. Universitélibre de Bruxelles . Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., &Bontempi, G. (2018).

[20] Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, PP (99), 1–14. doi: 10.1109/ TNNLS.2017.2736643. Davis, J., &Goadrich, M. (2006).

[21] The relationship between precision-recall and ROC curves. (pp. 233–240). ACM. Dietterich, T. (2002).

[22] Machine learning for sequential data: A review. Structural, Syn- tactic, and Statistical Pattern Recognition, 227–246. Doya, K. (1993).

[23] Bifurcations of recurrent neural networks in gradient descent learn- ing. IEEE Transactions on Neural Networks, 1, 75–80.