International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

Credit Card Fraud Detection Using Machine Learning and an Authenticator App

Adhil Muhammad Nazeer [1], Anjali Vijayan [2], Merin Ann Mathew [3], Roshni R [4], Sreelekshmi K.R [5], Vinod P.R [6]

U.G. Student, Department of Computer Engineering, College of Engineering, Chengannur, Kerala, India[1,2,3,4]

Assistant Professor, Department of Computer Engineering, College of Engineering, Chengannur, Kerala, India[5,6]

ABSTRACT: Financial fraud is a highly growing threat with far reaching consequences in the finance industry. The high dependence on internet technology has shown an increase in number of online credit card transactions. As credit card transactions become the most prevailing mode of payment for both online and offline transaction, credit card fraud rate also accelerates. Detecting fraudulent transactions using traditional methods of manual detection is time consuming and inefficient, thus the advent of Machine Learning methods like Logistic regression is used. Machine Learning algorithms has been applied for identifying fraudulent transactions. Logistic regression is one among the several machine algorithms reported in literature and the same has been employed for various applications such as predicting fault-prone code across software projects, predicting fault proneness in software models using three different logistic regression models etc. Another major key area that we address in our project is real-time credit card fraud detection. For this, we take the use the analytics done by the implemented machine learning model and an Authenticator App to decide if a particular transaction is genuine or fraudulent. It is the application based on two-factor Authentication that helps for identifying user identity and the confirmation on what a user claims to be and whether he actually is. The Authenticator App uses the Firebase Database to store the user's data.

KEYWORDS: Machine Learning, Logistic Regression, Firebase Database, Two-factor authentication, Unix timestamp

I. INTRODUCTION

The Credit Card is a small plastic card issued to users as a system of payment. It allows its cardholder to buy goods and services. A rapid growth in the number of credit card transactions has led to a substantial rise in fraudulent activities. Credit Card Fraudsters employ a large number of techniques to commit fraud. The research presented in this paper presents a machine learning approach to the credit card fraud detection. We apply a logistic regression based algorithm to training and testing the model. For our initial experiments we used a dataset consisting of 10,00,000 lakhs of transaction data which uses 6 fields such as mode of payment, cash-in, cash-out, debit, transfer. The logistic regression method is an effective supervised learning algorithm that can be applied for multiple classification tasks.
The intention of this study is to detect the exact fraud of the credit card transaction and to differentiate anomalous from legitimate transactions. With this, various learning algorithms like logistic regression will be evaluated by measuring their effectiveness in predicting the correct classification of the input dataset.

Paper is organized as follows. Section II describes the various machine techniques available and the existing system in credit card fraud detection. Section III describes how the model works and how the authenticator code can be generated. Section IV presents the experimental results of performing fraud detection using the dataset. Finally, Section V presents the conclusion.

II. RELATED WORK

Credit card fraud detection is the process of identifying those transactions that are fraudulent into two classes of genuine and fraudulent transactions [1]. Credit card fraud detection is based on analysis of a card's spending behaviour. Many techniques have been applied to credit card fraud detection but a comparative analysis of logistic regression and naive bayes is carried out in [2]. The paper [3] evaluates two advanced data mining approaches, support vector machines and random forests, together with logistic regression, as part of an attempt to better detect credit card fraud while neural network and logistic regression is applied on credit card fraud detection problem [4]. A number of challenges are associated with credit card detection, namely fraudulent behaviour profile are dynamic, that is fraudulent transactions tend to look like legitimate ones; credit card transaction datasets are rarely available and highly skewed. [5]

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

A two-step logistic regression (LR) based algorithm is proposed in this study for identifying the fraudulent transactions. A set of fraudulent transactions is first generated in step 1, followed by a second round of LR-based algorithm conducted on this smaller dataset for identifying the valid and non valid ones.[6] Logistic regression is a probabilistic linear classifier, parameterized by a weight matrix w and a bias b. It enables the system to estimate categorical results with the help of a group of independent variables. The classifier equation for logistic regression model is given as

$$y = \text{sgn}(wT\ x + b) \quad (1)$$

where y is an element of {-1,1} denoting the output class recognized for the input x fed to the system.[7]

The work in this paper is divided in two stages. 1) Model training using Logistic Regression 2) Authentication Code Generation using Unix Time Stamp. Machine learning model training and testing is done by using the LR(Logistic Regression). Thereafter, it is load and dumped in a pickle file. Then, creating the web application. For the secret code generation an Android Application is created. Finally, if the model detect the transaction as a fraudulent one, an authenticator code is generated in the app(by the use of the unix time stamp. After entering the code, the user have to confirm whether it is by them or not. If the user verifies, the transaction will be successful.

III. METHODOLOGY

The first approach in the proposed system was to collect the necessary datasets for the purpose of training and testing the model by using logistic regression. Developed the model using Python, where we have imported the necessary downloaded packages and modules. Thereby successfully trained the model.

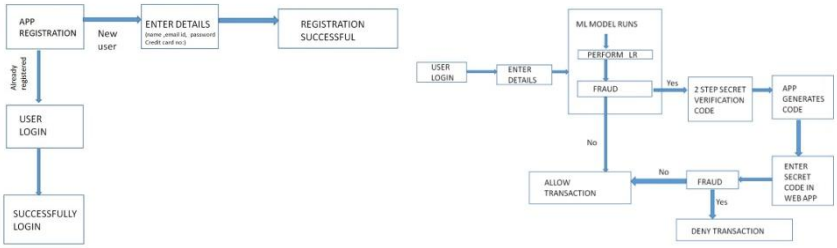
In the proposed system, the Authentication App and the Web application gives an extra layer of security for the users account. It is done by enabling 2-Factor Authentication.(2FA). The code can be entered manually and as a result, the Android Application will generate the secret code. The main difference is in expiry time: the OTP is valid until it hasn't been used or some predetermined period of time - usually 30 or 60 seconds. For this algorithm , time is also important as if the time on your phone is incorrect generated code will be different and you won't get an access to your account(Hence considering the Unix Time Stamp). The authenticator code improves the existing system's pitfall and preserves its essential features thus making the system more efficient and secure. To store the customer details, there should be a Firebase database.

If a person opens the app for the first time, then he/she should register into the app by submitting his details and then the corresponding details is stored in the Firebase authentication and Real-Time database and he gets registered. If the person already has an account, then he/she can login to the app by using their login-id and password.

When the person uses their card, our Web Application (which runs on Flask using an HTML template) which is also connected to the Firebase database , asks for email-id and password and verify the user. After verifying, it redirects to another page where the user needs to enter a number of six fields which are mode of payment, destination id, old balance, amount, new balance, receiver id. This page is connected to our machine learning model. After submitting the details the model will run and a fraud probability is calculated. If the transaction is found out as a fraudulent it shows a label of 1 else if it is valid it shows a label of 0.

If it shows the transaction as a fraudulent one, then an authenticator code will be generated on the app and the user needs to enter that secret code on to the next web page and submit it. Then the user needs to enter the secret code in perfect and then he can continue with his transaction otherwise it is identified as a fraudulent one and the transaction will be cancelled.

Fig. 1 (a) Flowchart of the Android Application. (b) Working of the whole proposed system.



(a) (b)

IV. EXPERIMENTAL RESULTS

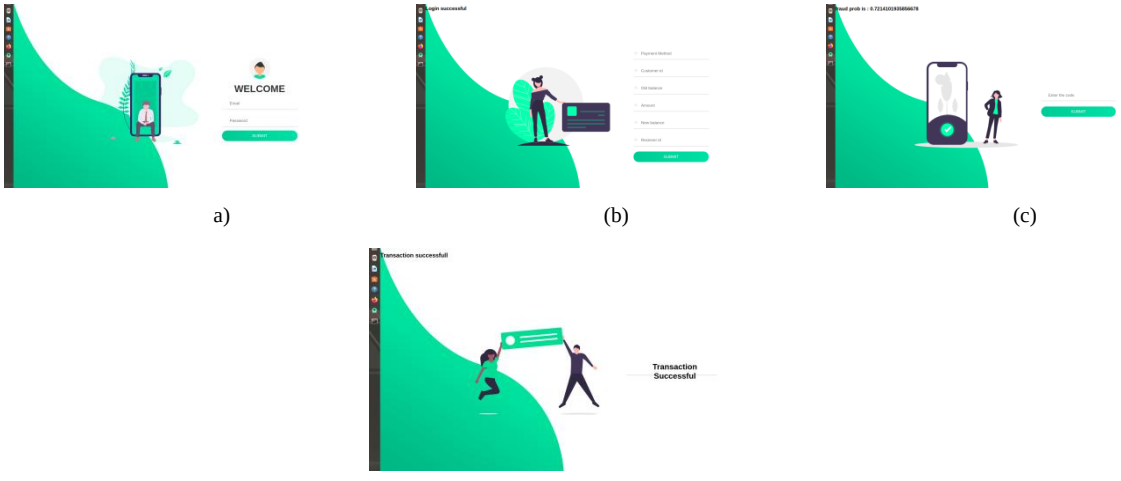
Figures show the result of Credit Card Fraud detection using dataset which consists of 10,00,000 lakhs transactions.

Fig. 2.Authenticator Application a) shows the registration phase in the app. If the person opens the app for the first time, his data needs to be collected and stored in the database. Fig b) shows the login phase which corresponds to a customer who is about to make the transaction. Fig c) shows the secret code generated .

(a) (b) (c)

Fig. 3. Web application (a) Login page where the user needs to enter his email-id and password (b) The page where the details need to be entered and the machine learning model runs after its submission. (c) The web page where the user needs to enter the secret code, if the transaction detected as a fraudulent one. Fig d) shows the page which will be displayed to the customer upon successful transaction.

Fig. 3. Web application (a) Login page (b) Details need to be submitted. (c) If fraudulent ,the user needs to enter the secret code. d) page displayed to the customer upon successful transaction.



a) (b) (c)

(d)

The results after performing the fraud detection on the dataset are found correct and accurate.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

VI. CONCLUSION

We have implemented a Credit Card Fraud Detection System using Machine Learning. Our LR algorithm successfully detects the fraudulent transactions from the datasets which consists of mixed datas (fraud and valid). We have applied our algorithm on different values and found that it successfully detected the fraudulent one. The use of Logistic Regression(LR) in the model has helped us achieve greater efficiency. With the introduction of the Android App we provided an extra layer of security in the transaction by implementing the 2-Factor Authentication.

REFERENCES

[1] John.O.Awoyemi , Adebayo O. Adetunmbi and Samuel A. Oluwadare ”Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis” International Conference on IEEE (2017)
[2] Ng, A. Y., and Jordan, M. I., (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems, 2, 841-848
[3] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602-613.
[4] Sahin, Y. and Duman, E., (2011). Detecting credit card fraud by ANN and logistic regression. In Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on (pp. 315-319). IEEE.
[5] The Nilson Report. (2015). U.S. Credit & Debit Cards 2015. David Robertson.
[6] Bolin Chen, Xuequn Shang, Min Lit , Jianxin Wangt and Fang-Xiang Wut” A two-step logistic regression algorithm for identifying individual-cancer-related genes” IEEE International Conference on Bioinformatics and Biomedicine (BTBM) (2015)
[7] Pranav Rao, and Manikandan JA “Design and Evaluation of Logistic Regression Model for Pattern Recognition Systems”
[8] Z. Wu, Q. Wang, A. Plaza, J. Li, L. Sun and Z. Wei, “Real-Time Implementation of the Sparse Multinomial Logistic Regression for Hyperspectral Image Classification on GPUs”, in IEEE Geoscience and Remote Sensing Letters, Vol. 12, No. 7, July 2015, pp. 1456-1460.
[9] O. Birkenes, T. Matsui, K. Tanabe, S. M. Siniscalchi, T. A. Myrvoll and M. H. Johnsen, “Penalized Logistic Regression With HMM Log Likelihood Regressors for Speech Recognition”, in IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 6, Aug. 2010, pp. 1440-1454.