International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-2, July 2020

Credit Card Fraud Detection using Machine Learning and Deployment of Model in Public Cloud as a Web Service

S Kiruthika, Sowmyarani C N

Abstract: ***In recent times, usage of credit cards has increased exponentially which has given way to an increase in the number of cybercrimes related to transactions using credit cards. In this paper, the aim is to reduce the fraudulent credit card transactions happening around the world. Latest technologies like machine learning algorithms, cloud computing and web service implementation has been used in this paper. The model uses Local outlier factor algorithm and Isolation forest algorithm to develop the credit card fraud detection model using unsupervised learning techniques. The model has been implemented as a Web service to make the solution integratable with other applications and clients across the world. A third party prototype application is developed and integrated to the Fraud Detection Model using Web Services. The complete Fraud Detection System is deployed on the cloud. The Fraud Detection Model shows exceptionally high accuracy when compared to other models already existing.***

Keywords**:** ***Fraud detection model, machine learning, local outlier factor, isolation forest, web service, prototype application, public cloud, amazon web services, EC2 instance***

I. INTRODUCTION

Internet speed has expanded largely and the costs of portable mobile devices and data rates have diminished impressively in recent years, the data rates have gotten significantly more moderate to a large portion of the individuals. This has come about into the digitization of the greater part of the foundations as it is simple and advantageous for the individuals and institutes to keep up the records. Along these lines, it brought about the greater part of the banks and different establishments getting and moving cash through credit cards and online medium of transactions [1]. Be that as it may, with the programmers and other digital hoodlums around the world there are consistently odds of the fake exchanges in the market which results in huge financial losses [2].

Manuscript received on May 25, 2020.
Revised Manuscript received on June 29, 2020.
Manuscript published on July 30, 2020.
* Correspondence Author
S. Kiruthika*, Department of Computer Science and Engineering, R V College of Engineering, Bengaluru, India.
Email: skiruthika.scn18@rvce.edu.in
Dr. Sowmyarani C.N., Department of Computer Science and Engineering, R V College of Engineering, Bengaluru, India.
Email: sowmyaranicn@rvce.edu.in

Fraud Transactions are very less when compared to legitimate transactions, since they occur very few in number. The fraudulent transactions must always be avoided in order not affect the bank or institute's reputation. There are multiple techniques that has been come up by researchers so far to detect fraud. Outlier detection is a pivotal part as outliers also known as anomalies show abnormal running conditions of the transactions from the generic behavior expected from a genuine transaction. In classification problems the skewed data of the transaction class impacts the model accuracy more. The imbalance in this distribution of class is because generally the ratio of fraud is less than 1% of the complete set of overall transactions [3]. This has been the major reason for the skewed data [4].

The fraud detection system mostly uses this type of unbalanced and skewed transaction data. The common machine learning algorithms like logistic regression, random forest and decision trees showed better results and good prediction accuracy for the majority class [5]. Because of this, minority class gets disregarded as the algorithm calculation regards it as a noise. But, the problem in this type of research is to precisely detect and predict the minority class, i.e., the fraud transaction class.

This paper focuses on the development of Fraud Detection System, involving the Machine Learning Fraud Detection Model which is a combination of two algorithms, Isolation forest (IF) and Local Outlier Factor (LOF) algorithm. This paper, also shows how to implement this fraud detection model as a Web Service. A prototype application is developed through which the clients such as banks can access the fraud detection model. The complete Fraud Detection System is deployed in the amazon web services public cloud. The Fraud Detection Model shows exceptionally high accuracy when compared to other models already existing.

The rest of the paper is presented as follows. Section 2 gives a brief background to the related work in this field. Section 3 explains about the proposed work. Section 4 provides the experimental methodology including results. Finally, Section 5 is the conclusion of this work and Section 6 lists the references.

II. RELATED WORK

Credit card fraud detection can be categorized as a binary classification issue with a class of genuine transactions and another class of very small quantity of fraud transactions [1].

If the standard or common most used machine learning algorithms are applied to these datasets, it tends to incline towards the majority class [6], [7].

This limitation can be addressed by using 3 different approaches: data-level approach, algorithm-level approach or cost-sensitive approach [1], [6]. We focus on algorithm level approach.

Yusuf Sahin *et. al.,* have come up with a new perspective by considering a variable misclassification cost that helps the proposed algorithm to predict correct results. The inclusion of fraud priorities and a freshly proposed metric w.r.t. performance has helped in reducing the fraud [8].

Samaneh Sorournejad *et. al.,* have surveyed on numerous Fraud Detection techniques. This paper touches upon a wide range of datasets that is being used in researches. All the limitations and criterias that can be considered for the evaluation of proposed methods in this research area of fraud detection is listed in this paper [9].

Divya Iyer *et. al.,* have formulated a set of operations, which helps in detecting the occurrence of fraud by using the card holder's general activities. This behaviour of the card usage is fed into the model. Any high non-probabilistic activity of the card w.r.t. transactions is considered as a fraud [10].

Jon T.S Quah *et. al.,* have used a multi layered system and has taken advantage of the various qualities of Self Organizing maps. The output patterns are analysed for various cluster nodes and transactions [11].

L. U. Oghenekaro *et. al.,* have experimented on an UCI repository dataset. The algorithm forms three major formulations and based on the values gathered from the temporal and spatial pooler components, the learning columns and duration is adjusted for the results. Multiple cell size in terms of columns along with the run times is analyzed and accuracies are listed [12].

Suvasini Panigrahi *et. al.,* have come up with a algorithm that uses past and present transaction details, filters that are based on certain rules and various adders. The authors provide a three classification output as normal, abnormal or suspicious as results [13].

III. PROPOSED SYSTEM

In this proposed system, a machine learning model is developed using a combination of Local outlier factor algorithm and Isolation forest algorithm. The system is a fraud detection machine learning model which classifies a credit card transaction to be genuine or fraudulent. This machine learning model is implemented as a web service and a third party prototype application is developed as well. The complete credit card fraud detection system runs in the EC2 instances of Amazon Web Services in the Live Cloud environment. This provides access to the system across all geographical locations.

IV. EXPERIMENTAL METHODOLOGY

A. Data Description

The credit card transactions dataset is taken from kaggle data source. The dataset consists of 284807 real time credit card transaction details. This CSV file holds all the credit card transactions that occurred in real time in a bank during a specific time duration including the genuine and fraudulent transactions. For confidentiality reasons, the dataset has undergone Principal Component Analysis (PCA) to hide or encrypt all the customer sensitive information such as card holders' personal information, pin number, etc. The real time credit card transactions dataset has 31 columns of data namely Time, Amount, Class and v1-v28 columns that has undergone PCA. The dataset has 492 fraud transactions and 284315 genuine or legitimate credit card transactions recorded. The data present in the dataset is skewed, but this is not the case with other generic datasets of other research areas. This is caused since the number of fraud transactions are always far less when compared to genuine transactions. The attributes of the transactions dataset are mentioned as follows in Table I:

Table I: Columns in Credit Card Transactions Dataset

Column Name	Description
Time	Transaction Time
Amount	Transaction Amount
V1- V28	Unknown Fields (Encrypted with PCA)
Class	Genuine or Fraud

B. Modeling and Testing

In this proposed system, a machine learning Fraud detection model is developed using a combination of 2 algorithms, i.e., Local Outlier factor algorithm and Isolation Forest Algorithm. Since the dataset in skewed, anomaly detections algorithms are used to predict the outliers and arrive with correct classification of the transaction. Both the algorithms are tested separately and provide outcomes or outputs independent of the other. The fraud detection model finally classifies the real time credit card transaction based on table II.

Table II: Final Output of Fraud Detection Model

Output of LOF Algorithm	Output of IF Algorithm	Final Predicted Output of Model
Fraud	Fraud	Fraud
Fraud	Genuine	Fraud
Genuine	Fraud	Fraud
Genuine	Genuine	Genuine

These algorithms have been applied to the real time credit card dataset. It is found that these anomaly detection algorithms achieve exceptional high accuracy rates. The Accuracy of each of the algorithms when tested for the transaction class prediction is show in figure 1.

The LOF and IF algorithm are tested against all the 284807 number of transactions to predict the output class of the transaction as genuine or fraud. It is found that very few transactions are mispredicted by both the algorithms when compared to the actual output of the transaction class. The number of errors or wrongly predicted transaction count is shown in figure 2.

Local Outlier factor predicts the wrong class for 83 transactions and Isolation Forest algorithm predicts with a even lesser count of 71 transactions with the wrong result. While comparing the Wrong Output predictions against the total number of predictions with the total number of credit card transactions, these algorithms show exceptional results.

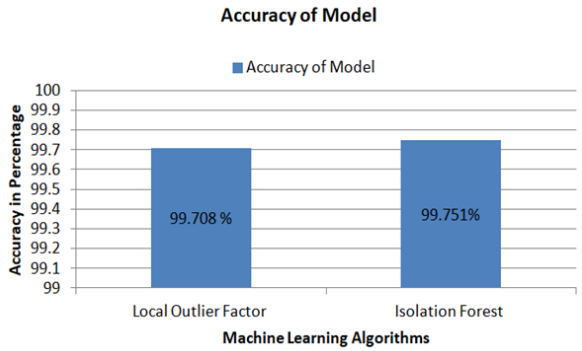

Figure 1: Accuracy results of Algorithms

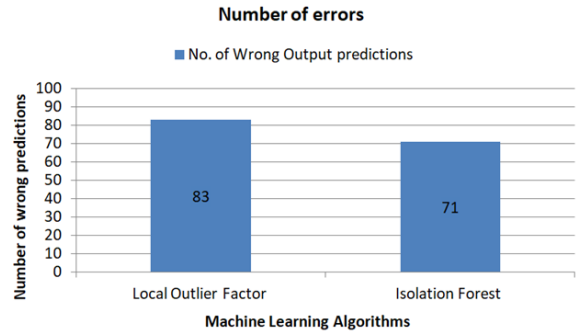

Figure 2: Number of Errors

C. Web Service Implementation

An Application Programming Interface (API) is implemented to make the Machine Learning Fraud Detection Python Model available to other applications and clients. In this paper, the Web service is implemented as Representational State Transfer (REST) API to provide interoperability between computer systems on the internet. It is implemented using java programs called Executor Controller, Executor Service and Constants. A major advantage of this paper is to provide this web service hosted in the AWS Cloud that provides access to the service all around the world. The flow of how the client or third party application calls the ML Fraud Detection model through these programs is shown in below figure 3.

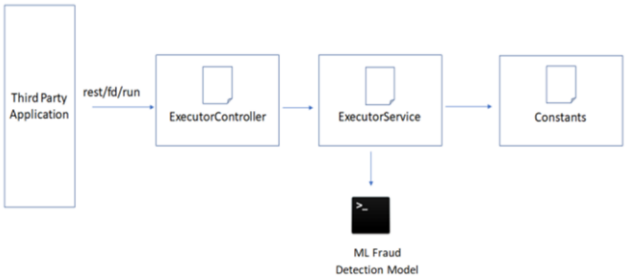

Figure 3: REST API Implementation

The third party application invokes the web service by using the Context Root. The context root hits the Executor Controller which acts as the first entry point to the web service. This Controller reads the real time credit card transaction input variables and passes it to the Executor Service. The Executor Service invokes the ML Python program in command line (cmd) and reads the output of the Fraud Detection model. The output is given back to Executor Controller and is in turn displayed in the third party application. Constants store the directory or location of the Model and name of the ML python model program to be invoked.

D. User Interface Design for Fraud Detection Model

The user can interact with the ML Fraud Detection Model using a prototype or third party application. The application user interface developed in this paper provides five major operations: Profile operations, End point Model Configuration of web service, Input Data specification, Model Execution and Result/Output Collector.

The user profile operations include creating a new account, logging in to the existing account, logging out, editing the profile, changing the password, and deleting the profile. This application is also deployed on the cloud server so that this can be accessed by anyone across the globe using the public IP address of this cloud server. The implementation is done using the Java 2 Platform Enterprise Edition (J2EE) architecture and for the database needs we have used SQLITE3.

The user can configure the end point model web service so that the prototype application will be capable of communicating with the model without having any issues. For configuring the machine learning fraud detection model through the prototype application, the user will have to provide the Host name, Port number, Application Name and the Context root of the Model Web services. The end point module configuration details in the application is shown in the below figure 4.

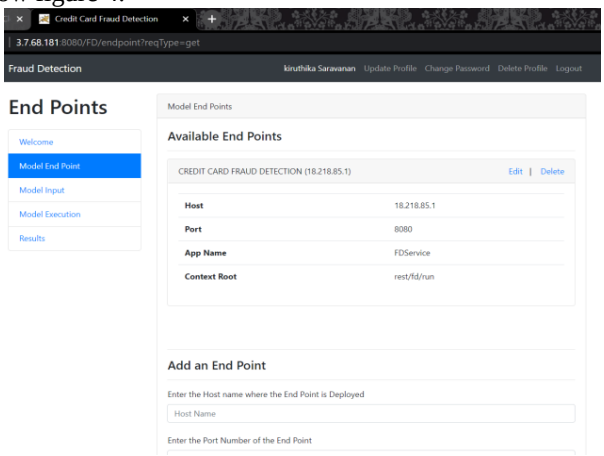

Figure 4: End Point Module Configuration

The user can provide input to the model through the application in either of these two ways: comma separated way or field-by-field way.

The system will validate the input before sending it to the model for execution. The Model Execution module provides a button to the user upon clicking it, the system will invoke the web services as per the configured model by sending the input supplied in the Model Input module. The Model input module is shown in the below figure 5.

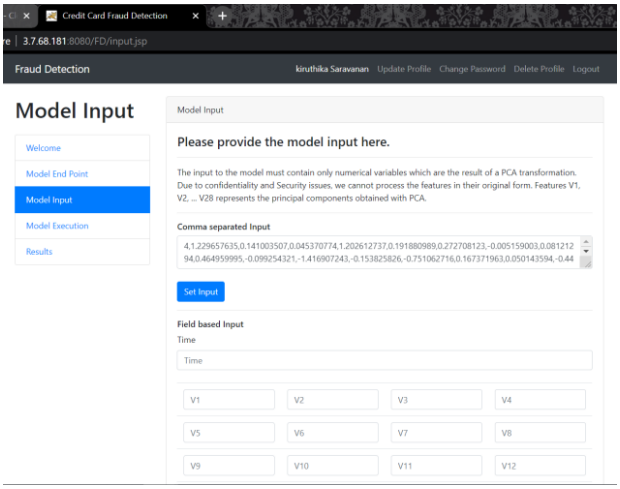

Figure 5: Prototype Application Model Input

The Results collector module will keep polling the web service system to check if the result of the model is available or not. Once the results are available, it displays the same to the user.

In this paper, we consider two types of transactions. The input data of the real time credit card transaction is given to the Fraud Detection Machine Learning Model through the third party application that is provided to the client. The web service helps in passing the input and output values between the Model and the application. If the ML Fraud Detection model, classifies the transaction as genuine and legitimate, figure 6 is displayed to the client application.

If the model predicts that the given real time credit card transaction is fraudulent, the prototype application displays the client with figure 7.

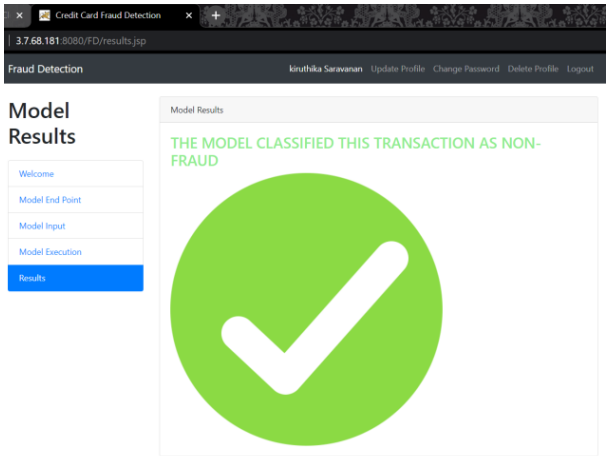

Figure 6: Results of Real time Genuine Credit Card Transaction

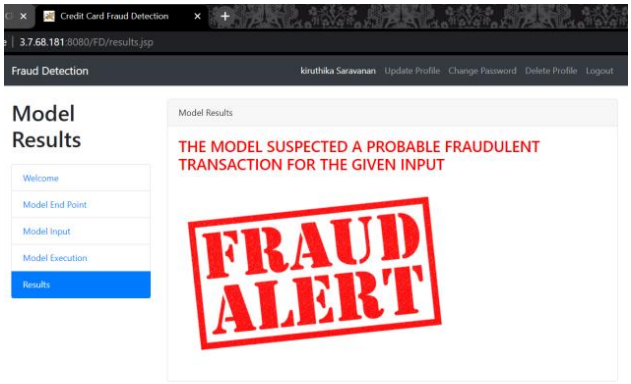

Figure 7: Results of a Real time Credit Card Fraudulent Transaction

Thus this Fraud Detection System helps the clients from detecting fake or fraud transactions and avoids it in prior stages.

E. Cloud Deployment

The Fraud Detection model is deployed on a cloud server to make the solution accessible across the geographical areas. For the cloud deployment process, we use Amazon web services. Two EC2 instances are created and the ML Model, REST API services and application files are all deployed in the cloud instances.

F. Fraud Detection System

One of the key objectives of this project is to detect the credit card fraud. This fraud detection system has three major modules: Fraud Detection Machine Learning Model, Web Services REST API Implementation Module and Development of Prototype Application for clients. Another advantage of this project is, the fraud detection system in completely deployed in the cloud to provide access to clients all over the world. The overall flow is shown for the Fraud Detection System in figure 8.

The user interacts with the Machine Learning model using the User Interface application and polls the output class of the transaction using the web service REST API. Any transaction that is fraudulent and not legitimate is identified by the fraud detection model and the result is given to the API and in turn to the Client application Result Collector.

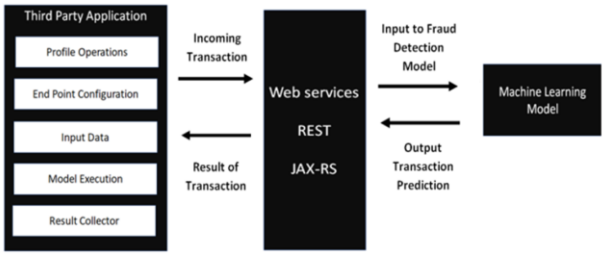

Figure 8: Fraud Detection System

V. CONCLUSION

Credit card fraud detection has been a very interesting and intriguing area of research for many researchers worldwide. In this research paper, we have presented a credit card fraud detection system by using a combination of two best suitable anomaly detection algorithms. The API and Prototype Application helps in indicating the End user or client the immediate minute when a fraudulent transaction takes place. This quick feedback can help the fraud investigation teams in banks and related clients to take the necessary action. The model has achieved exceptionally high accuracy rates for both the machine learning models used namely Local Outlier factor algorithm with 99.708% and Isolation Forest algorithm with 99.751%. We have also effectively addressed the issue of skewed data. Since only a single type of fraud has been looked into in this project, the future extensions will be aimed to focus on other type of frauds and fraud patterns and develop a fraud detection system for those.

REFERENCES

1. Andrea, Dal, Pazzolo; Oliver, Caelen; Reid, A, Johnson; Gianluca, Bontempi . (2015). Calibrating Probability with Undersampling for Unbalanced Classification. *IEEE Symposium Series on Computational Intelligence*.
2. Dal, P. A., & Johnson, A. R. (2014). Using HDDT to avoid instances propagation in unbalanced and evolving data streams. *Proceedings of international joint conference on neural networks.*
3. Datta, S., & Arputharaj, A. (2018). An Analysis of Several Machine Learning Algorithms for Imbalanced Classes. 5th International Conference on Soft Computing and Machine Intelligence.
4. Divya, I., Arti, M., Sneha, J., Dhanashree, R., & Amrutha, S. (2011). Credit Card Fraud Detection using Hidden Markvo Model. *World Congress on Information and Communication Technologies.*
5. Jon, T. S., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications , 35* (4), 1721-1732.
6. L.U. Orghenekaro, C. U. (2016). A Novel Machine Learning Approach to Credit Card Fraud Detection. *International Journal of Computer Applications , 140* (5).
7. Lin, W. C., Tsai, C. F., Hu, Y. H., & Jhang, J. S. (2017). Clustering-based undersampling in class-imbalanced data. *Information Sciences , 409-410*, 17-26
8. Roberston, D. (2016). Investments &amp, Acquisitions – September 2016 Top Card Issues in Asia – Pacific Card Fraud Losses reach $21.84 Billion.
9. Samaneh, S., Zahra, Z., Reza, E. A., & Amir, H. M. (2016). A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective. *IOT Security* .
10. Suvasini, P., Amlan, K., Shamik, S., & Manjumdar, A. (2009). Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning. *Information Fusion , 10*, 354-363.
11. Wang, H., Zhu, P., Zou, X., & Qin, S. (2018). An Ensemble Learning Framework for Credit Card Fraud Detection based on Training Set Partitioning and Clustering. *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovations*, (pp. 94-98).
12. Xuan, Z., Liu, G., Li, Z., Zheng, L., Wang, S., & Surname, G. (2018). Random Forest for Credit Card Fraud . *15th International Conference Networking Sensor Control*.
13. Yusuf, S., Serol, B., & Ekram, D. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications* , 5916-5923.

AUTHORS PROFILE

S. Kiruthika, pursuing M.Tech in Computer Network Engineering from the Department of Computer Science and Engineering, R V College of Engineering in Bengaluru, India. B.Tech in Electronics and Communication Engineering from Amrita School of Engineering, Bengaluru, India. Her areas of interests are Machine Learning, Web Development and Telecommunication. She has worked as a research intern for six months for the design of security protocol for 5G networks from Nanyang Technological University, Singapore.

Dr. Sowmyarani C.N., Associate Professor in Department of Computer Science and Engineering at R V College of Engineering, Bengaluru, India. She has guided 25 UG Projects and 10 PG Projects. She has paper publications in 14 International Journals and 12 International Conferences. She has a patent filed in Privacy preserving data publishing. She is also an associate editor in International Journal of Information Security and Privacy (IJISP) and Editorial Review board member of International Journal of Open Source Software and Processes (IJOSSP). She has delivered expert lectures, session chair and judge in multiple Project exhibitions.