iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40799

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Credit Card Fraud Detection System Using CNN

P. Shanmugapriya[1], R. Shupraja[2], V. Madhumitha[3]
[1]*Associate Professor, Department of Computer Science and Engineering, SCSVMV, Kanchipuram*
[2, 3]*B.Tech Graduate (IV year), Department of Computer Science and Engineering, SCSVMV, Kanchipuram*

Abstract: ***This Project is focused on credit card fraud detection in real-world scenarios. Fraud is one of the major ethical issues in the credit card industry. In this era we can see many innovative financial services like ATMs, online banking, etc. Besides, along with the rapid advances of e-commerce, the use of credit cards has become a convenient and necessary part of financial life. A credit card is a payment card supplied to customers as a system of payment. Using a third-person credit card or its information without the knowledge of that person is referred to as credit card fraud. Application and Behavioral fraud are major types of fraud where people are easily tricked and lose their money. The same user may submit multiple applications which may lead to identical fraud. Application fraud takes place when fraudsters apply for new cards from a bank or issuing companies using false or others' information. In this project, we will be applying some supervised and unsupervised algorithms and will classify the credit card dataset. We will use CNN and correlate the data train and get the model's accuracy.***
Keywords: CCFD, CNN, Logistic Regression, Feature extraction.

I. INTRODUCTION

The main aim of the project is to create a fraud detection system. Our vision is to create a detection system that detects fraudulent transactions to help people and create awareness about this deceit.

In this current period, purchasing items from e-commerce sites has become common. Many companies developed an online transaction mode for payment via credit cards to identify the different types of credit card fraud happening and to review alternative techniques used in fraud detection. We are going to apply machine learning algorithms to past data and predict the possibility of a transaction being a fraud transaction. Purchasing items from e-commerce sites and paying via credit card has become common. To solve this problem, we will be using some machine learning methods like CNN the dataset will be divided into a training set and test set. The algorithms used in the experiment are Logistic Regression and CNN. To detect fraud in a transaction we need to perform and use complex algorithms and functions. In this project, we will try to train our dataset to get max accuracy to detect fraud transactions. Credit card fraud is mainly divided into two forms: and talks to them. To shape our thoughts into an application, we created a personal assistant with emotion recognition.

1) *Non-Internet-based Fraud:* It happens without permission n of the actual card owner. In most cases, the card-issuing can lock it previously it's used in a fraudulent procedure.
2) *Internet-based Fraud:* It happens through the web just hacker traces the card details and login details, and he/she can access the account and transact money. In this project, we will be designing a model to detect fraudulent activity in credit card transactions. Our model is trained to provide the essential features required to detect fraudulent transactions and we will be using machine learning algorithms to automate this process and to detect fraud transactions. The statistics alternatively focus on data inference and probability, ML contains further issues about the efficiency and feasibility of data processing algorithms and architectures, compounding several learning tasks with one compactor performance measure.

II. LITERATURE REVIEW

The literature survey includes all the research done by various researchers on the respective topic. After a lot of contemplation on the core idea of our project, its pros and cons, and the probability of completing the project successfully, To increase our knowledge on the topic we started our work by doing a lot of study on the projects and research papers that are similar to our core idea and project.

©IJRASET: All Rights are Reserved

These research papers and surveys enabled us to understand what is necessary for us to venture into a project of this sort.

Author	Topic	Remarks
A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi	A paper based four main fraud occasions in real-world transactions	Focused on better prediction level and an API module to decide if a particular transaction is genuine.
Sapna Sahu, Shikha Agarwal, Raju Baraskar.	Comparing various ML algorithms	Clustering and CNN perform well in unsupervised and supervised learning algorithm respectively
Aayushi Agarwal, Baldivya, Mitra, Md Iqbal	A Survey of various techniques used for CCFD	Random forest and K-NN can be the best techniques to get better accuracy

Using these research papers, we understood how and which algorithm can be more exact. Also, we used and surveyed a lot of fraud detection programs and algorithms to detect fraudulent transaction. We got inspired by a lot of ideas that arose after reading below mentioned papers. There are numerous fraud detection methods but we can't rely on or state one specific algorithm method that distinguishes the fraud totally [5].

III. PROPOSED METHOD

We propose a Machine learning model to detect fraudulent credit card activities in online financial transactions. Analyzing fake transactions manually is impracticable due to the vast amounts of data and its complexity. To classify fraudulent and legitimate credit card transactions by using supervised and unsupervised algorithms. To help us in getting awareness about the fraudulent activities without any financial loss. In our project we will be getting raw data and then performing data cleaning and transformation then we will be performing feature correlation and feature selection and applying ML algorithms like Logistic regression and predicting accuracy

A. Model Architecture

After importing the dataset and analyzing the dataset and performing data exploration and statistical analysis we further balance the data and separate the dataset to test and train and build a CNN model. So after importing the dataset, we have performed cleaning and then we have split the dataset into fraud and non-fraud data and created a balance data for checking fraud transactions, and plotted a bar graph to analyze the frequency of fraudulent transaction.

B. Machine Learning Algorithm Used

Convolutional neural network (CNN) –It is a neural network that has one or more convolutional layers used for classifying and segmenting images processing other correlated-data.
Logistic Regression **is** a classification algorithm i.e. used to predict the probability of a categorical dependent variable
Prediction- if we want a machine to make predictions for us, we should train it with some data.
Decision tree classification-
Statistical techniques: average, quantiles, probability distribution, association rules e data.
Supervised ML algorithm used: Logistic regression.

C. Algorithm

1) Step 1: Dataset is imported
2) Step 2: The imported dataset is read
3) Step 3: For balancing the dataset Random Sampling is done
4) Step 4: The dataset is divided into two parts, i.e., the Train dataset and Test dataset.
5) Step 5: Feature selection is applied for the proposed models.
6) Step 6: Accuracy and performance metrics are calculated to know the efficiency of the algorithm.
7) Step7: We build a Convolutional Neural Network
8) Step8: We predict from the test dataset.
9) Step 9: Perform logistic regression to find accuracy precision and recall value
10) Step 10: Then will predict the loss and accuracy and create a confusion matrix for our model

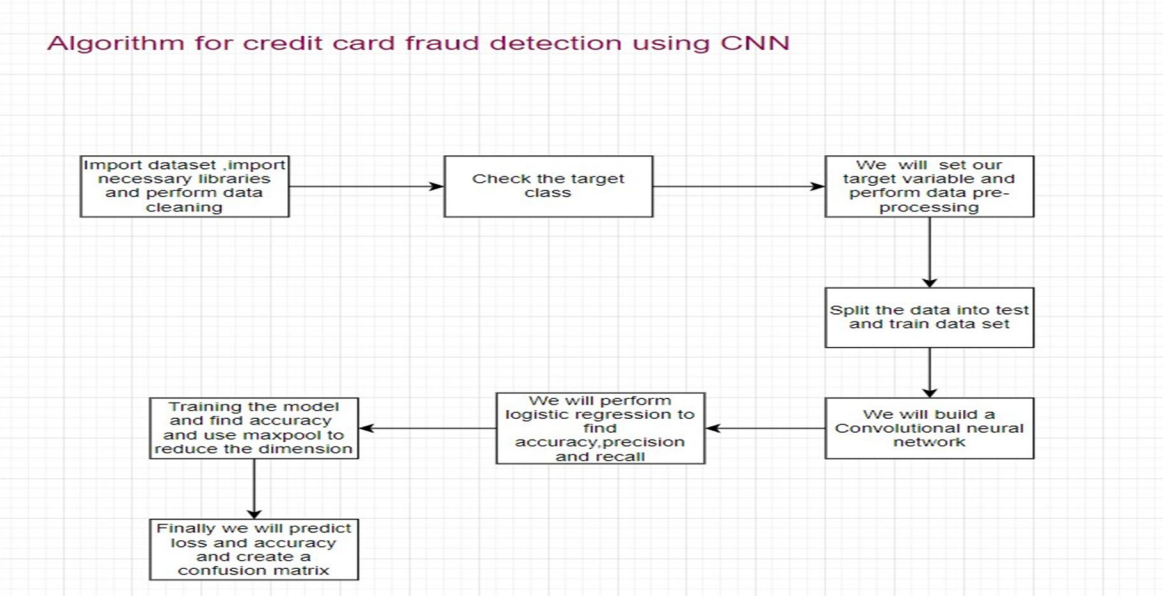


Fig 1. Algorithm For CCFD

D. Methodology

We will be using classical ML methods and training our dataset and performing oversampling to reduce the imbalance data also we will be using CNN and after balancing the data and attain precise accuracy.

The following are the modules of the project, which are planned aid to in completing the project concerning the proposed system while overcoming the existing system and providing support for enhancing in future.

1) Data Collection
2) Data Pre-processing
3) Feature Extraction
4) Evaluation model

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   Time    284807 non-null  float64
 1   V1      284807 non-null  float64
 2   V2      284807 non-null  float64
 3   V3      284807 non-null  float64
 4   V4      284807 non-null  float64
 5   V5      284807 non-null  float64
 6   V6      284807 non-null  float64
 7   V7      284807 non-null  float64
 8   V8      284807 non-null  float64
 9   V9      284807 non-null  float64
 10  V10     284807 non-null  float64
 11  V11     284807 non-null  float64
 12  V12     284807 non-null  float64
 13  V13     284807 non-null  float64
 14  V14     284807 non-null  float64
 15  V15     284807 non-null  float64
 16  V16     284807 non-null  float64
 17  V17     284807 non-null  float64
 18  V18     284807 non-null  float64
 19  V19     284807 non-null  float64
 20  V20     284807 non-null  float64
 21  V21     284807 non-null  float64
 22  V22     284807 non-null  float64
 23  V23     284807 non-null  float64
 24  V24     284807 non-null  float64
 25  V25     284807 non-null  float64
 26  V26     284807 non-null  float64
 27  V27     284807 non-null  float64
 28  V28     284807 non-null  float64
 29  Amount  284807 non-null  float64
 30  Class   284807 non-null  int64  
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Fig 2. Data Set Summary

After importing the dataset and analyzing the dataset and performing data exploration and statistical analysis we further balance the data and separate the dataset to test and train and build a CNN model. So After importing the dataset, we have performed cleaning and then we have split the dataset into fraud and non-fraud data and created a balance data for checking fraud transactions, and plotted a bar graph to analyze the frequency of fraud.

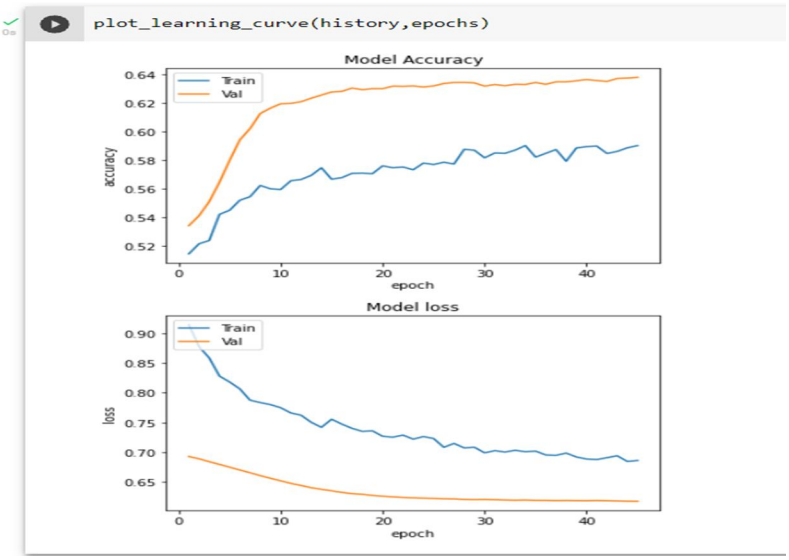


Fig 3. The learning Curve or the model

E. Software Requirement

1) Google Collab
2) Windows 64bit
3) Python Language

IV. RESULT AND DISCUSSION

The training dataset is passed on to the model and training of the model is done. The accuracy of the model is predicted and printed. Hence, we have acquired the result of an accurate value of credit card fraud detection i.e. 0.9539 (95.3%) by building and training a Convolutional Neural Network model with new enhancements.

In comparison to existing model, this proposed model can be applied for the larger dataset for more accurate result

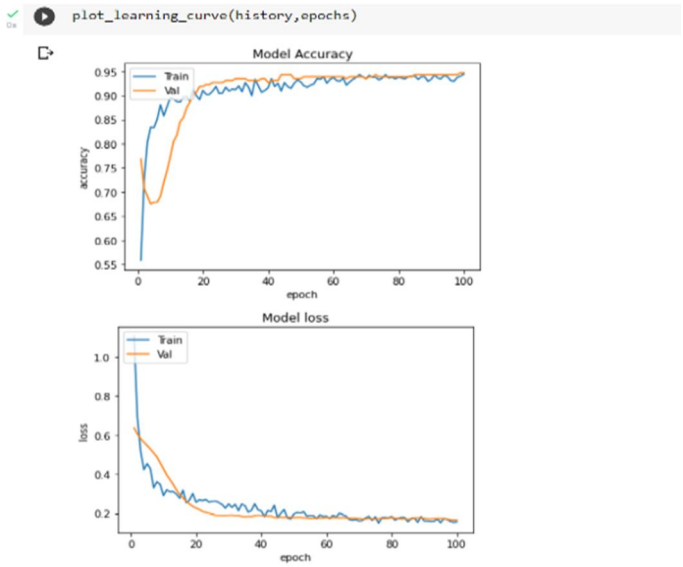


Fig 4. Accuracy and loss for the model

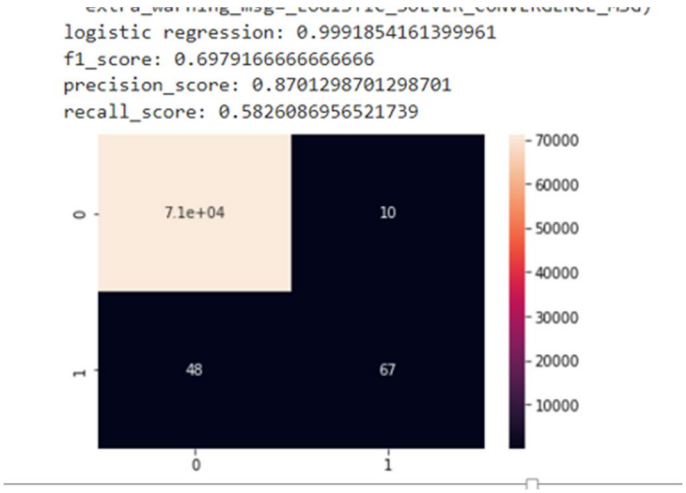


Fig 5. Predicting precision and plotting Confusion Matrix for the target value

V. CONCLUSIONS

In this Credit Card Fraud Detection System project, machine learning algorithms were briefly compared in fraud detection, also a detailed comparative analysis is provided. It is seen that various algorithms give different accuracy in performance. Even though there are numerous fraud discovery approaches we can't state that this specific method distinguishes the fraud. At the point when we look at the precision, recall, and F1-score are 0.8701, 0.5826, and 0.679166 using Logistic Regression accuracy is 0.999185. Accuracy of 0.9539 and loss of-0.1108 using CNN are obtained. Hence we will analyze the possibility of a transaction being a fraud transaction. The CNN algorithm will provide a better performance, but the speed during testing and application will still be an issue.

VI. FUTURE ENHANCEMENT

Nowadays, credit card has become the modern tool for, online payments, and many e-transactions have a credit card transaction option the credential from the credit card is used by the third-party to deduct money. Due to this, we need to analyze and find the best solution to detect the maximum number of frauds in online systems. We will try to implement deep learning algorithms and will further compare which algorithm works best and will test this with real time dataset.

.

REFERENCES

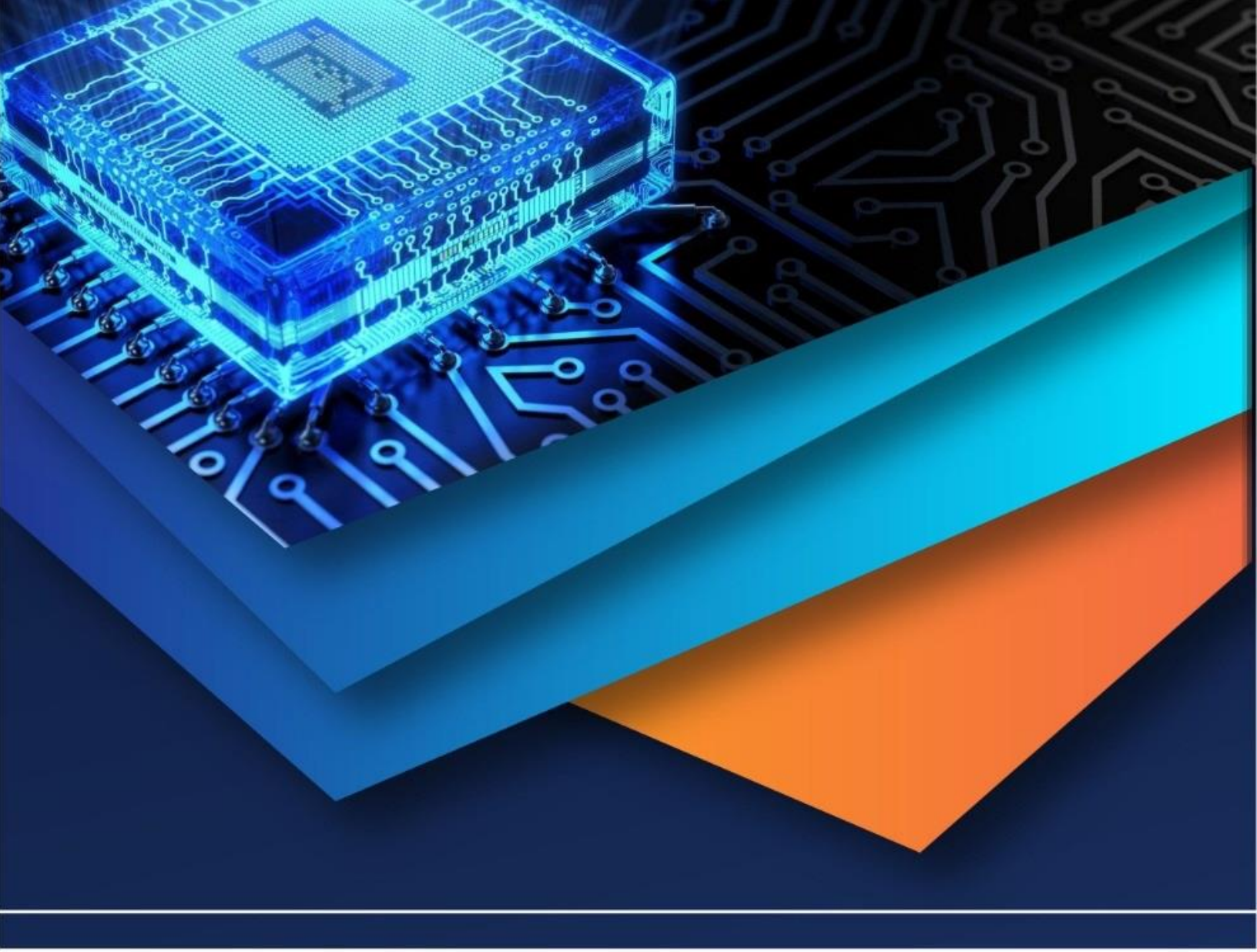
[1] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, published in the year 2019.

[2] A Brief review on Machine Learning for credit card frauddetection Sapan sahu[1], Shikha Agrawal[2], Raju Baraskar[3] – Volume-8 Issue-12, October Published in the year 2019.

[3] Survey of Various Techniques used for Credit Card Fraud Detection Aayushi Agarwal1 Md Iqbal2 , Baldivya Mitra3 1, 2, 3Department of Computer Science & Engineering, MIET, Meerut, U.P, India published on, 8th july 2020

[4] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, published on 22 March 2019

[5] Hema Gonaboina Appala Srinuvasu Muttipati, MACHINE LEARNING METHODS FOR DISCOVERING CREDIT CARD FRAUD published on 12, January 2021



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)