IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 1, June 2022

Loan Fraud Detection Using Machine Learning Algorithm

Reni Hena Helan R[1], Abirami G[2], Sultan Saleem A[3], Vivekanandan S J[4], Tejaswini P R[5]

Assistant Professor, Department of Computer Science Engineering[1,2,3]
Associate Professor, Department of Computer Science Engineering[4]
Students, Department of Computer Science Engineering[5]
Dhanalakshmi College of Engineering, Chennai, India

Abstract: *The widespread usage of the Web has had a massive effect on the growth of online card transactions, especially at the beginning of the previous few years. Because of the increase in internet transactions, the global banking system has been forced to deal with or confront an unexpected number of fraudulent operations. As a reason, rule-based algorithms were created to detect high-risk transactions and allow specialists to authenticate whether or not they were fraudulent. The present intruders used the static nature of rule-based systems as a defensive measure to avoid detection. As a result, researchers set out to develop adaptive fraud detection systems based mostly on machine learning techniques, including deep learning, which is a relatively new application. The widespread use of the Internet, notably at the start of the previous decade, had a considerable impact on the rise in online card transactions. The rise of internet transactions has forced the worldwide banking system to deal with or confront an unexpected amount of fraudulent operations. As a result, rule-based algorithms were developed to identify high-risk transactions and allow specialists to confirm whether they were genuine or not. To avoid detection, the current attackers used the static nature of rule-based systems as a countermeasure. As a result, researchers set out to develop adaptive fraud detection systems based mostly on machine learning techniques, including deep learning, which is a relatively new application.*

Keywords: Fraud detection

I. INTRODUCTION

The number of credit card transactions is increasing as a result of technological developments and the rise of e-commerce. Around the world, the ratio of fraudulent transactions to legitimate transactions is around 0,006%. Although this rate may appear low, each fraudulent transaction damages a bank's reputation. As more than just a result, banks are expanding their investments in fraud detection. The number of fraudulent actions and approaches expands and changes every day. Detecting fraudulent activities alone by studying transactions is extremely difficult and costly. To retain client satisfaction and trust, it is vital to detect fraud swiftly and accurately. The frequency of credit card payments is increasing as a result of technological developments and the rise of e-commerce. Around the world, 375 billion card payments were processed in 2017 . In the same year, however, 16.7 million fraudulent transactions occurred . The proportion of fraudulent to valid transactions is roughly 0,006 percent all across the world. Although this rate may appear low, each fraudulent transaction damages a bank's reputation. As a result, banks are increasing their investments in fraud detection. Every day, the number of fraudulent activities and their methods grows and changes. Detecting fraudulent activities alone by studying transactions is extremely difficult and costly. It is critical to detect fraud quickly and accurately in order to maintain client pleasure and trust. In a recent study, machine learning methods were found to be more effective than the majority of rule-based systems. In terms of numbers, characteristics, and changes over time, the data sets utilized in the publications in which these results are published do not always correlate to the real banking environment. In this work, researchers conducted an unprecedented analysis on a real data set to uncover previously unknown aspects of fraud detection efforts. These findings were based on 245,000 fraudulent transactions and four billion non-fraudulent transactions gathered from various banks in 2017.

DOI: 10.48175/IJARSCT-4637

II. LITERATURE SURVEY

Statistical classification methods in consumer credit scoring

D. Hand, W. Henley

In light of the rationale of logging, the model of the Chinese business bank was carried out in logging. When looking at test results, it is accepted that the example size, test size, and blunder worth can influence the quantity of mistakes.

A comparison of neural networks and linear scoring models in the credit union environment

V. Desai, J. Crook, G. A. Overstreet

The motivation behind this paper is to assess conventional abilities, for example, multi-sensor and measured neural net sensors, segregation-based investigation, and consecutive FICO ratings in a credit agreeable.When the benchmarks clearly reveal bad credit, our findings suggest that adaptive businesses offer a lot of opportunities. Even if all other factors are equal, claiming one is still beyond the typical range. The standard model exhibition was slightly below average as the standard model, particularly on account of helpless credit positioning. Despite the fact that there is a major distinction between the three credit associations, our neural net organizations don't match the distinctions, recommending that a better approach for working might be expected to make an overall outline.

Neural nets versus conventional techniques in credit scoring in Egyptian banking

Hussein A. Abdou

The quantity of non-performing advances has expanded as of late, expanding the worth of the model of credit in case of a downturn. This study provides FICO scores for the Neuro Fuzzy Inference System, which includes three types of information incorporation based on Neuro Fuzzy innovation. A given model's presentation is nearly identical to that of a typical and often used model. Instances of financial assessments are estimated utilizing a 10-time check technique and Visas given by the International Bank for Reconstruction and Development in Turkey. In terms of the appropriate norm of estimate and correlation of spurious attributes, the given model outperforms Discriminant Analysis, Logistic Regression Analysis, and the Artificial Neural Network (ANN). The model offered, like the ANN, can be pondered; but, unlike the ANN, the model does not remain in secret elements. Clarifications of the proposed change might give helpful data to financiers and purchasers, particularly with regards to why the advance application is inescapable.

Credit Scoring Methods. Czech Journal of Economics and Finance

Martin Vojtek, Evžen Koèenda

It is broadly used to identify rebelliousness with loaning models. To determine score precision, rules such as Gini esteem, Kolmogorov-Smirnov statistics (KS) insights, Lifting, Mahalanobis distance, and information can be used. This page sums up and shows how to utilize these aspects by and by.

A survey of credit and behavioral scoring: forecasting: financial risk of lending to customers

L. Thomas

Loaning and scoring practices are abilities that assist associations with choosing whether to loan to clients. This article examines factual exploration and the procedures utilized in research dependent on these choices. It likewise examines the significance of consolidating monetary designs into the scoring framework, looking at irregularities, and how the framework can change until the purchaser thinks about the advantages to the bank. This features the achievement of the not-for-profit area in the locale.

Credit Decision-Making And Information Requirements

Susan Cancino, Giovanni Cancino-Escalante

The world economy has been hit hard by the downturn. The motivation behind the review was to comprehend the acquiring system and bank prerequisites. The review, directed in 22 Brazilian loaning establishments, found that the significance of ongoing utilization of resources, for example, Accounts are scrutinized, bank records are inspected, and credit experience is comparable, and security is the most important concern for a firm because of data inconsistencies.

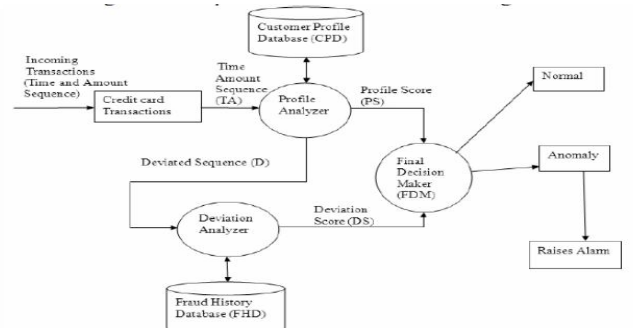
DOI: 10.48175/IJARSCT-4637

Online Fraud Transaction Detection Using Machine Learning

Vedant Mayekar, Siddharth Mattha, Sohan Choudhary, Prof Amruta Sankhe

In today's world, people depend on online transactions for almost everything. Online transactions have their own merits like ease of use, feasibility, faster payments etc., but these kinds of transactions also have some demerits like fraud transactions, phishing, data loss, etc. With increase in online transactions, there is a constant threat for frauds and misleading transactions which can breach an individual's privacy. The algorithm can get experience; improve its stability and performance by processing as much data as possible. These algorithms can be used in the project that is online fraud transaction detection. In these, the dataset of certain transactions which are done online is taken. Then with the help of machine learning algorithms, we can find the unique data pattern or uncommon data patterns which will be useful to detect any fraud transactions. For the best results, the XGBoost algorithm will be used which is a cluster of decision trees. This algorithm is recently dominating this ML world. This algorithm has features like more accuracy and speed when compared to other ML algorithms.

III. SYSTEM ARCHITECTURE


Customer Profile Database (CPD)
Incoming Transactions (Time and Amount Sequence)
Credit card Transactions
Time Amount Sequence (TA)
Profile Analyzer
Profile Score (PS)
Normal
Final Decision Maker (FDM)
Anomaly
Raises Alarm
Deviated Sequence (D)
Deviation Analyzer
Deviation Score (DS)
Fraud History Database (FHD)

IV. MODULES

- Clustering
- Classification
- Association Rule
- Fraud Detection

V. MODULE DESCRIPTION

Classification

Arranging is the most usually utilized strategy for separating data, utilizing the models in the past area to foster a model that can rank all archives. Getting fraud and financial soundness is useful for this sort of examination. The request for data incorporates the method involved with learning and arranging.

Clustering

During the consolidation, a wide range of banking exercises were united in a similar area. One might say that the principal strategy for handling is utilized to choose a classification and articles.

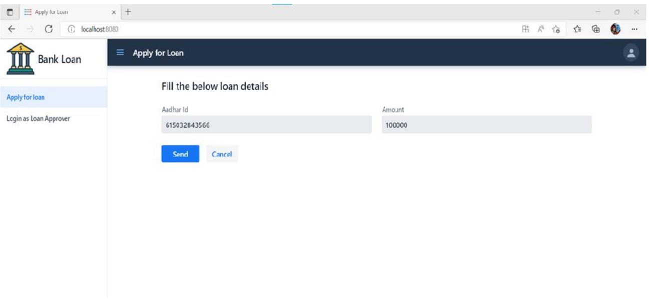
Association Rule

The fundamental capacity of the mining affiliations is to observe the two factors that are generally normal in the business files, while the reason for the choice cycle is to distinguish the gatherings that are identified with the particular reason.

Fraud Detection

Another notable spot where data can be separated from the financial area is fraud discovery. Having the option to recognize fraud is a worry for some organizations; and numerous different fakes are accounted for through the media.

VI. RESULTS



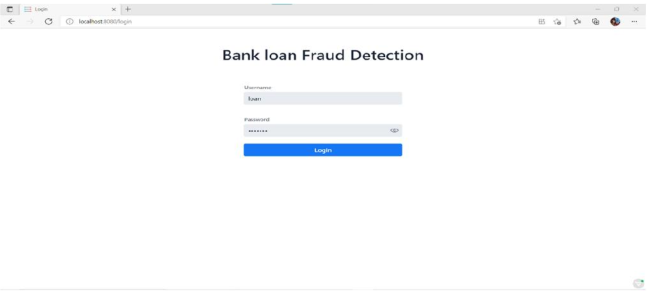
Fill the below loan details

Aadhar Id: 615032843566

Amount: 100000

Fig 6.1: User Login Page

This is the user login page, where the loan applicants enters the required details for applying a loan(like adhar card number etc)



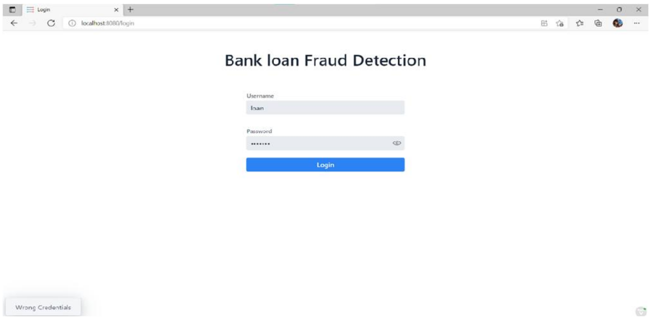
Bank loan Fraud Detection

Username: loan

Password: •••••••

Fig 6.2: Administrator Login Page

This is the user login page, where the administrator is logged in using administrator login details.



Bank loan Fraud Detection

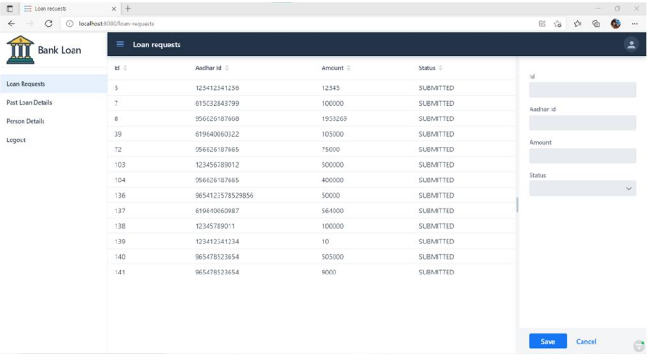
Username: loan

Password: •••••••

Wrong Credentials

Fig 6.3: Login Details Verification

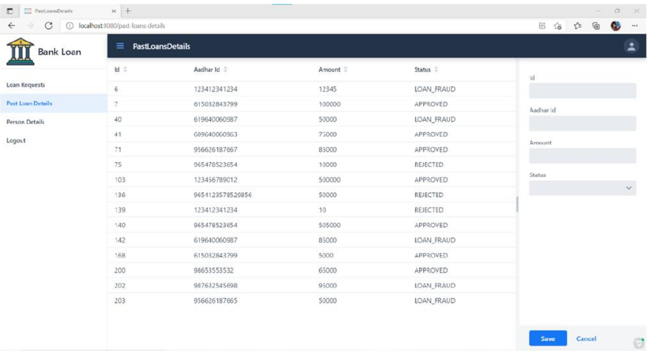
First of all the administrator login page checks whether the login details given by the administrator is valid or not.



Id	Aadhar Id	Amount	Status
5	123412341236	12345	SUBMITTED
7	615032843799	100000	SUBMITTED
8	956626187668	1953269	SUBMITTED
39	619640060322	105000	SUBMITTED
72	956626187665	75000	SUBMITTED
103	123456789012	500000	SUBMITTED
104	956626187665	400000	SUBMITTED
136	9654123578529856	50000	SUBMITTED
137	619640060987	564000	SUBMITTED
138	12345789011	100000	SUBMITTED
139	123412341234	10	SUBMITTED
140	965478523654	505000	SUBMITTED
141	965478523654	9000	SUBMITTED

Fig 6.4: Loan Requests

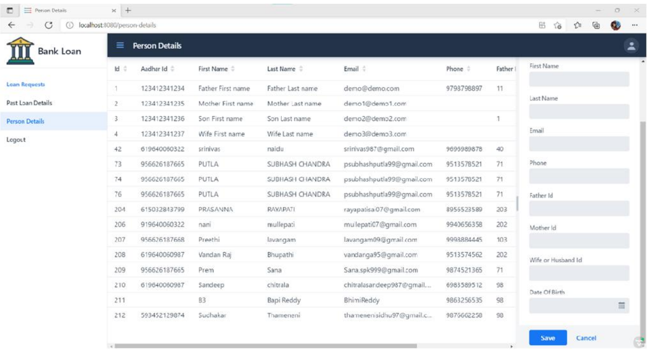
The applicants loan request page, where the details of the applied loan is entered (like loan amount etc).



Id	Aadhar Id	Amount	Status
6	123412341234	12345	LOAN_FRAUD
7	615032843799	100000	APPROVED
40	619640060987	50000	LOAN_FRAUD
41	609640060963	75000	APPROVED
71	956626187667	85000	APPROVED
75	965478523654	10000	REJECTED
103	123456789012	500000	APPROVED
136	9654123578529856	50000	REJECTED
139	123412341234	10	REJECTED
140	965478523654	505000	APPROVED
142	619640060987	85000	LOAN_FRAUD
168	615032843799	5000	APPROVED
200	98653553532	65000	APPROVED
202	987632545698	95000	LOAN_FRAUD
203	956626187665	50000	LOAN_FRAUD

Fig 6.5: Past-Loan Details

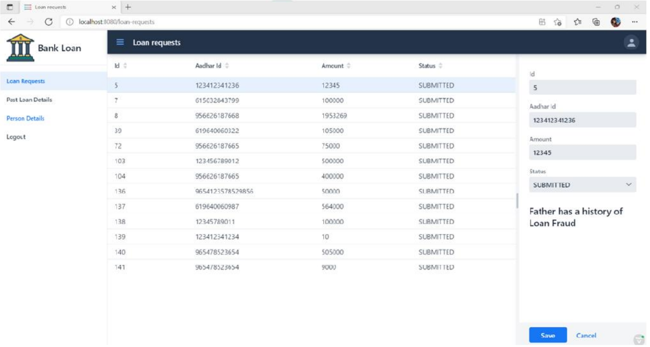
This page checks whether the applicant is having any past loans and checks whether they have paid them on time or any fraud history etc., to approve the new amount for the applicant.



Id	Aadhar Id	First Name	Last Name	Email	Phone	Father I
1	123412341234	Father First name	Father Last name	demo@demo.com	9793798897	11
2	123412341235	Mother First name	Mother Last name	demo1@demo1.com		
3	123412341236	Son First name	Son Last name	demo2@demo2.com		1
4	123412341237	Wife First name	Wife Last name	demo3@demo3.com		
42	619640060322	srinivas	naidu	srinivas987@gmail.com	9696989878	40
73	956626187665	PUTLA	SUBHASH CHANDRA	psubhashputla99@gmail.com	9513578521	71
74	956626187665	PUTLA	SUBHASH CHANDRA	psubhashputla99@gmail.com	9513578521	71
76	956626187665	PUTLA	SUBHASH CHANDRA	psubhashputla99@gmail.com	9513578521	71
204	615032843799	PRASANNA	RAYAPATI	rayapatisa07@gmail.com	8955523589	203
206	919640060322	nani	mullepati	mullepati07@gmail.com	9940656358	202
207	956626187668	Preethi	lavangam	lavangam09@gmail.com	9993884445	103
208	619640060987	Vandan Raj	Bhupathi	vandanga95@gmail.com	9513574562	202
209	956626187665	Prem	Sana	Sana.spk999@gmail.com	9874521365	71
210	619640060987	Sandeep	chitrala	chitralasardeep987@gmail....	6983589512	98
211		B3	Bapi Reddy	BhimiReddy	9863256535	98
212	593452129874	Sudhakar	Thameneni	thamenenisidhu97@gmail.c...	9876662258	98

Fig 6.6: Person Details

Here we can check the additional details of the applicant like their parents or any other guardian details. So, if the applicant can repay the loan amount in time in future the banks can have some additional security to recover the bank funds.



Id	Aadhar Id	Amount	Status
5	123412341236	12345	SUBMITTED
7	615032843799	100000	SUBMITTED
8	956626187668	1953269	SUBMITTED
39	619640060322	105000	SUBMITTED
72	956626187665	75000	SUBMITTED
103	123456789012	500000	SUBMITTED
104	956626187665	400000	SUBMITTED
136	9654123578529856	50000	SUBMITTED
137	619640060987	564000	SUBMITTED
138	12345789011	100000	SUBMITTED
139	123412341234	10	SUBMITTED
140	965478523654	505000	SUBMITTED
141	965478523654	9000	SUBMITTED

Id: 5
Aadhar Id: 123412341236
Amount: 12345
Status: SUBMITTED

Father has a history of Loan Fraud

Fig 6.7: Fraud History

In this page, the bank can check for past loan details of the applicant's parents details whether they have any loan fraud history to enhance additional security for bank funds.

VII. CONCLUSION

AI is a method used to separate valuable data from an enormous amount of accessible data and to use sound judgment in banking and business exercises. They utilize a data set to consolidate diverse data from a data set in a legitimate manner to separate data. The data is then broken down and the data gathered is utilized all through the local area to help the choices. Data innovation is utilized to foster the financial area, to draw in new clients, to keep up with significant clients, and to forestall extortion. It is fundamental. a fun chance to remain, convey, hazard the board and promote.

DOI: 10.48175/IJARSCT-4637

REFERENCES

[1]. RBR. (2018). Global Payment Cards Data and Forecasts to 2023. [Online]. Available: https://www.rbrlondon.com/research/global-cards/

[2]. loss prevention media. (2018). The Latest Credit Card Fraud Statistics and Insights. [Online].Available:https://losspreventionmedia.com/creditcard-fraud-statistics-and-insights/

[3]. TBB. (2019). Detection and Prevention Methods of Fraud in Banking. [Online]. Available: https://www.tbb.org.tr/gec/KTPV14.pdf

[4]. M. K. Sparrow, License to Steal: How Fraud Bleeds America's HealthCare System. Abingdon, U.K: Routledge, 2019.

[5]. O. S. Yee, S. Sagadevan, and N. H. A. H. Malim, ''Credit card fraud detection using machine learning as data mining technique,'' J. Telecommun., Electron. Comput. Eng., vol. 10, nos. 1–4, pp. 23–27, 2018.

[6]. A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, ''Credit card fraud detection: A realistic modeling and a novel learning strategy,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 8, pp. 3784–3797, Aug. 2018.

[7]. P. H. Tran, K. P. Tran, T. T. Huong, C. Heuchenne, P. HienTran, and T. M. H. Le, ''Real time data-driven approaches for credit card fraud detection,'' in Proc. Int. Conf. E-Business Appl. (ICEBA), 2018, pp. 6–9.

[8]. A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, ''Deep learning detecting fraud in credit card transactions,'' in Proc. Syst. Inf. Eng. Design Symp. (SIEDS), Apr. 2018, pp. 129–134.

[9]. F. Zhang, G. Liu, Z. Li, C. Yan, and C. Jiang, ''GMM-based undersampling and its application for credit card fraud detection,'' in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1–8.

[10]. H. Wu and G. Liu, ''A hybrid model on learning cross features for transaction fraud detection,'' in Proc. ICDM, 2019, pp. 88–102.

[11]. L. Zheng, G. Liu, C. Yan, and C. Jiang, ''Transaction fraud detection based on total order relation and behavior diversity,'' IEEE Trans. Comput. Social Syst., vol. 5, no. 3, pp. 796–806, Sep. 2018.

[12]. [12] C. Jiang, J. Song, G. Liu, L. Zheng, and W. Luan, ''Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism,'' IEEE Internet Things J., vol. 5, no. 5, pp. 3637–3647, Oct. 2018.

[13]. E. Kim, J. Lee, H. Shin, H. Yang, S. Cho, S.-K. Nam, Y. Song, J.-A. Yoon, and J.-I. Kim, ''Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning,'' Expert Syst. Appl., vol. 128, pp. 214–224, Aug. 2019.

[14]. S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, ''Random forest for credit card fraud detection,'' in Proc. IEEE 15th Int. Conf. Netw., Sens. Control (ICNSC), Mar. 2018, pp. 1–6.

[15]. K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, ''Credit card fraud detection using AdaBoost and majority voting,'' IEEE Access, vol. 6, pp. 14277–14284, 2018.

[16]. F. Carcillo, Y. A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, ''Combining unsupervised and supervised learning in credit card fraud detection,'' Inf. Sci., 2019, doi: 10.1016/j.ins.2019.05.042.

[17]. A. Pumsirirat and L. Yan, ''Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine,'' Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 18–25, 2018.

[18]. F. Carcillo, A. Dal Pozzolo, Y.-A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, ''SCARFF: A scalable framework for streaming credit card fraud detection with spark,'' Inf. Fusion, vol. 41, pp. 182–194, May 2018. BARIS

[19]. A. Abdallah, M. A. Maarof, and A. Zainal, ''Fraud detection system: A survey,'' J. Netw. Comput. Appl., vol. 68, pp. 90–113, Jun. 2016.

[20]. BARIS CAN , Ali Gokhan Yavuz, Elif M. Karsligil, and M. Amac Guvensan, (Member, IEEE), "A Closer Look Into the Characteristics of Fraudulent Card Transactions", , date of publication September 7, 2020, date of current version September 22, 2020.

DOI: 10.48175/IJARSCT-4637