International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com
doi : https://doi.org/10.32628/IJSRCSEIT

Detection of Personal Loan Fraud Based on Supervised Learning

N. Supriya[1], Dr. R. Viswanathan[2]
Master of Computer Applications [1], Assistant Professor [2]
Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India

ARTICLE INFO

Article History:

Accepted: 01 Aug 2023
Published: 09 Aug 2023

Publication Issue
Volume 9, Issue 4
July-August-2023

Page Number
258-263

ABSTRACT

Because it can result in significant financial losses, personal loan fraud is a major concern for financial institutions. Accurate and effective fraud detection systems are essential to reducing this risk. The goal of this study is to create a supervised learning-based fraud detection model that uses historical data to spot fraudulent loan applications. The proposed model uses a dataset containing named cases of both certifiable and deceitful credit applications. Relevant data, such as applicant demographics, credit history, and loan details, are extracted from the data using feature-engineering methods. After that, support vector machines, logistic regression, decision trees, random forests, and other supervised learning algorithms all take these features as inputs. The dataset is parted into preparing and testing sets, permitting the model to gain from the marked examples and assess its presentation on concealed information. The model is trained on a wide variety of legitimate and fraudulent loan applications, allowing it to identify patterns and correlations that indicate fraud. Cross-validation and grid search are used to fine-tune the model's parameters in order to improve its performance.

To survey the model's adequacy, different assessment measurements, like exactness, accuracy, review, and F1-score, are used. Measurements of the model's capacity to strike a balance between false positives and false negatives are also made using tools like area under the curve (AUC) analysis and receiver operating characteristic (ROC) curve analysis.

Keywords: Financial transactions, Fraud, Patterns etc.

I. INTRODUCTION

Personal loans are widely used financial products that provide individuals with the flexibility to meet various personal expenses such as debt consolidation, home renovations, or funding a vacation. However, the popularity of personal loans has also attracted fraudulent activities, posing significant risks to both borrowers and lending institutions. Detecting and preventing personal loan fraud has become a critical task for financial institutions to ensure the integrity of their lending operations. Supervised learning techniques offer a promising approach to combat personal loan fraud by leveraging historical data to train accurate and efficient fraud detection models. These models learn from labeled data, where instances are classified as fraudulent or legitimate, and can then predict the likelihood of new loan applications being fraudulent based on the learned patterns. The aim of this research is to develop a robust and reliable fraud detection system using supervised learning algorithms.

This study aims to explore various supervised learning algorithms, such as logistic regression, decision trees, random forests, support vector machines, and neural networks, to detect personal loan fraud. By employing these algorithms, the research aims to identify patterns and features that distinguish fraudulent loan applications from legitimate ones. The ultimate goal is to build a model that can accurately classify new loan applications as either fraudulent or genuine, thereby minimizing the risk of fraudulent activities and protecting the interests of lenders and borrowers alike. To achieve this, a comprehensive dataset comprising historical loan application records, including both fraudulent and legitimate instances, will be utilized. The dataset will be preprocessed to handle missing values, outliers, and feature scaling to ensure the data is suitable for training the supervised learning models. Various performance metrics, such as accuracy, precision, recall, and F1-score will be employed to evaluate and compare the effectiveness of different algorithms in detecting personal loan fraud, the outcomes of this research have the potential to significantly enhance the fraud detection capabilities of financial institutions, leading to improved risk management and enhanced customer trust. By developing accurate fraud detection models, lenders can mitigate financial losses caused by fraudulent loan applications, reduce the chances of identity theft, and protect their customers' financial well-being.

II. RELATED WORKS

N. Carneiro, G. Figueira, and M. Costa, "An information digging based framework for Mastercard misrepresentation recognition in e-tail," Choice Emotionally supportive networks, vol. 95, 2017.[1] Online retailers suffer losses totaling billions of dollars as a result of credit card fraud. Researchers have discovered increasingly sophisticated methods for detecting fraud thanks to the development of machine learning algorithms; however, reports of actual implementations are rare. We talk about how a large e-commerce company built and used a fraud detection system. The paper compares various machine-learning techniques, provides insights into the entire development process, and investigates the combination of manual and automatic classification. The paper can hence help specialists and experts to plan and execute information digging based frameworks for misrepresentation discovery or comparable issues. In addition to providing an automated system, this project has provided fraud analysts with insights for enhancing their manual revision procedures, resulting in improved performance overall.[2] C. Phuaacbd, "On the communal analysis suspicion scoring for identity crime in streaming credit applications," vol. 195, no. 2, pp. 595–612, 2009.

A quick method is described in this paper: Using implicit links to one another across time and space, communal analysis suspicion scoring (CASS) generates numerical suspicion scores for streaming credit applications. Pair-wise communal scoring of application identifier attributes, the definition of suspiciousness categories for application pairs, the inclusion of temporal and spatial weights, and smoothed k-wise scoring of multiple linked application pairs are all features of CASS. The mining of several hundred thousand real credit applications revealed that CASS maintains reasonable hit rates while lowering false alarm rates. CASS can quickly identify the early signs of identity theft and is scalable for this large data set. Additionally, new insights have been discovered from application-to-application relationships. [3] D. J. Hand and R. J. Bolton, "c, vol. 17, no. 3, pp. 235–249, 2002. With the rise of global superhighways of communication and modern technology, fraud is on the rise, resulting in annual losses of billions of dollars worldwide. Fraudsters are adaptable and, given time, will typically find ways to circumvent such measures, despite the fact that prevention technologies are the best way to reduce fraud. Strategies for the identification of misrepresentation are fundamental on the off chance that we are to get fraudsters once extortion anticipation has fizzled. Money laundering, e-commerce credit card fraud, telecommunications fraud, and computer intrusion, to name a few, have all been successfully detected using statistical and machine learning technologies for fraud detection. We talk about the most common applications of fraud detection technologies and the statistical fraud detection tools that are available. [4] "Hoba:" by X. Zhang, Y. Han, W. Xu, and Q. Wang A novel feature engineering approach with a deep learning architecture for the detection of credit card fraud," Information Sciences, May 2019.

Card issuers incur annual losses of billions of dollars due to credit card transaction fraud. An advanced misrepresentation recognition framework with an innovative extortion discovery model is viewed as crucial for diminishing extortion misfortunes. The creation of a fraud detection system that makes use of a cutting-edge feature engineering procedure based on homogeneity-oriented behavior analysis (HOBA) and a deep learning architecture is our primary contribution. We conduct a comparative evaluation of the proposed framework using a real-world dataset from one of China's largest commercial banks. The findings of the experiments show that the method we have proposed works and can be used to detect credit card fraud. With a reasonable false positive rate, our proposed method is able to identify a greater number of fraudulent transactions than the standard approaches. Our research has managerial implications because it suggests that credit card issuers can use the proposed method to quickly and effectively identify fraudulent transactions in order to safeguard the interests of their customers, minimize losses due to fraud, and save money on regulatory fees.

III. METHODOLOGY

Proposed system:

We suggest using this strategy, which can be advantageous because it helps to lessen the restrictions imposed by traditional and other methods. The goal of this project is to create a reliable, efficient approach for quickly recognizing fraudulent transactions. We created this system using some powerful algorithms in a Python-based framework.

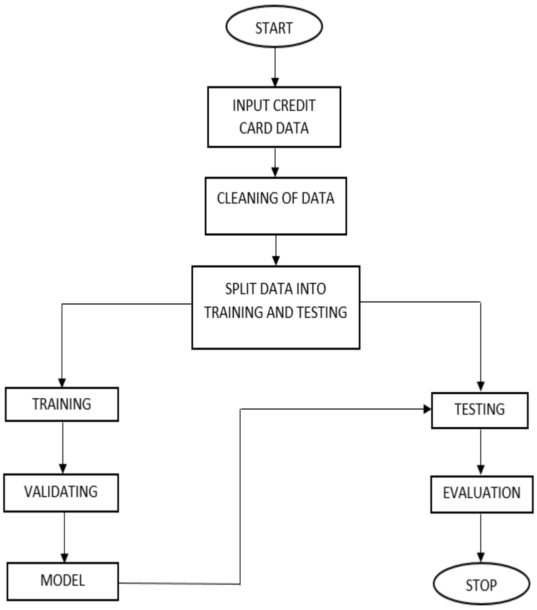


Figure 1: Block diagram

IV. IMPLEMENTATION

The project has implemented by using below listed algorithm.

XGBoost:

XGBoost is a regularizing gradient boosting framework for C++, Java, Python, R, Julia, Perl, and Scala that is available as an open-source software library. Linux, Windows, and macOS are all supported. A "Scalable, Portable and Distributed Gradient Boosting (GBM, GBRT, GBDT) Library" is the project's stated objective, according to the description. Along with the distributed processing frameworks Apache Hadoop, Apache Spark, and Apache Flink, it runs on a single machine. It has recently garnered a lot of attention and popularity as the algorithm of choice for numerous machine learning competition winners. Tianqi Chen worked on XGBoost as a research project for the Distributed (Deep) Machine Learning Community (DMLC) group. It started out as a terminal application that could be set up with the help of a libsvm configuration file. After its use in the Higgs Machine Learning Challenge is winning solution, it gained notoriety in the ML competition community. Before long, the Python and R bundles were constructed, and XGBoost currently has bundle executions for Java, Scala, Julia, Perl, and different dialects. As a result, the library gained access to a larger number of developers and gained popularity within the Kaggle community, where it has been used in a large number of competitions. It was quickly incorporated into a number of other packages, making its use in each community simpler. For Python users, it is now integrated with scikit-learn, and for R users, it is integrated with the caret package. Using the abstracted Rabit and XGBoost4J, it can also be integrated into Data Flow frameworks like Apache Spark, Apache Hadoop, and Apache Flink. XGBoost is also available for OpenCL for FPGAs. Tianqi Chen and Carlos Guestrin have distributed a proficient, adaptable execution of XGBoost.

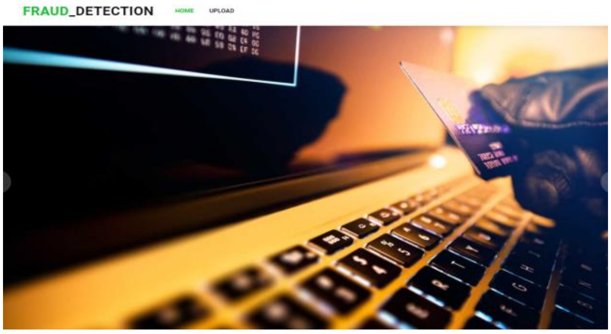
Stacking Classification:

Bagging, boosting, and stacking are just a few of the machine learning methods that can be used to group models together. Stacking is one of the most famous troupe AI procedures used to foresee numerous hubs to fabricate another model and work on model execution. Stacking empowers us to prepare numerous models to tackle comparable issues, and in view of their joined result, it constructs another model with further developed execution. The general architecture of stacking, important key points for implementing stacking, and the distinctions between bagging and boosting in machine learning will all be discussed in this section, "Stacking in Machine Learning." Prior to beginning this point, first, comprehend the ideas of the troupe in AI. So, in machine learning, let's start with the definition of ensemble learning.

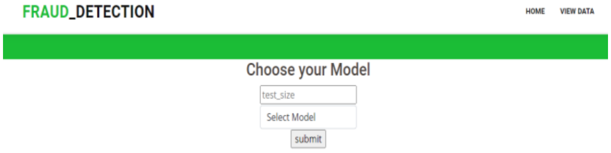
In machine learning, stacking is one of the most common ensemble modeling methods. We are able to make better forecasts for the future by combining various weak learners with Meta learners, who are ensembles in parallel. In order to construct a superior output prediction model, this ensemble method makes use of the predictions made by Meta learners and multiple weak learners when combined. In stacking, a calculation takes the results of sub-models as info and endeavors to figure out how to best join the info forecasts to make an improved result expectation. Stacking, which is also known as a stacked generalization, is an extended version of the Model Averaging Ensemble technique in which each sub-model contributes equally based on their performance weights to create a new model that makes better predictions. This new model is piled up on top of the others; this is the motivation behind why it is named stacking.

V. RESULTS

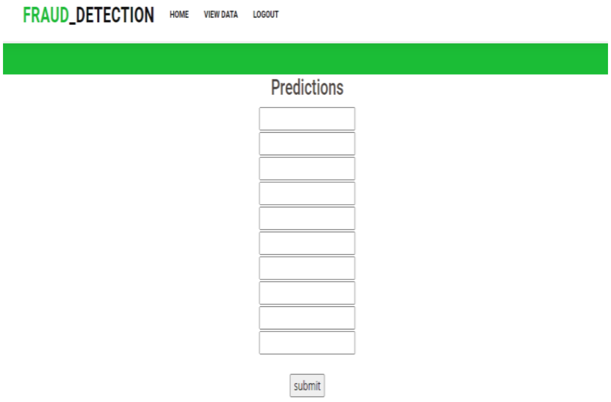
Home page:



Model Training Page:



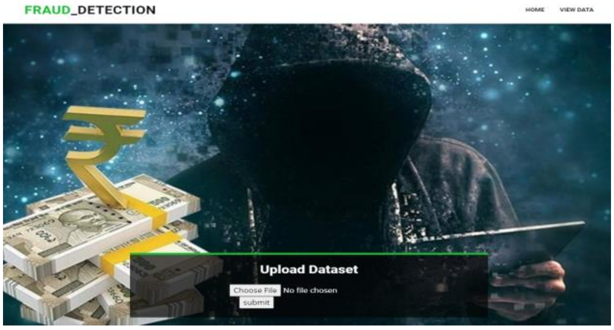
Model Prediction Page:



VI.CONCLUSION

In this application, we have successfully created a process to generate the classification for fraud detection. The system is likely to collect information from the user to predict the requirements.

Upload Page:



VII. REFERENCES

[1]. N. Carneiro, G. Figueira, and M. Costa, “A data mining based system for credit-card fraud detection in e-tail,” Decision Support Systems, vol. 95, 2017.

[2]. C. Phuaacbd, “On the communal analysis suspicion scoring for identity crime in streaming credit applications,” European Journal of Operational Research, vol. 195, no. 2, pp. 595–612, 2009.

[3]. R. J. Bolton and D. J. Hand, “Statistical fraud detection: A review,” Statistical Science, vol. 17, no. 3, pp. 235–249, 2002.

[4]. X. Zhang, Y. Han, W. Xu, and Q. Wang, “Hoba: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture,” Information Sciences, 05 2019.

[5]. E. Aleskerov, B. Freisleben, and B. Rao, “Cardwatch: a neural network based database mining system for credit card fraud detection,” in Computational Intelligence for Financial Engineering, 1997.

[6]. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection: A fusion approach using dempster–shafer theory and bayesian learning,” Information Fusion, vol. 10, no. 4, pp. 354–363, 2009.

[7]. Saputra, Adi & Suharjito, Suharjito. (2019). Fraud Detection using Machine Learning in e-Commerce. 10.14569/IJACSA.2019.0100943.

[8]. A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 421-426, doi: 10.1109/ICESC48915.2020.9155615.

[9]. John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare et al., “Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis”, IEEE, 2017.

[10]. Rajendra Kumar Dwivedi, Sonali Pandey, Rakesh Kumar “A study on Machine Learning Approaches for Outlier Detection in Wireless Sensor Network” IEEE International Conference Confluence, (2018).

Cite this article as :

N. Supriya, Dr. R. Viswanathan, "Detection of Personal Loan Fraud Based on Supervised Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp.258-263, July-August-2023