International Journal of Engineering and Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 14 Issue 4 April 2025 || PP 66-72

Forcasting Credit Card Fraud Detection Using Machine Learning

[1] P. Veeresh, [2] M. Thaher Basha, [3] Vajrala Varshith, [4] Beldar Taher Basha, [5] Kadapala Giribabu

[1] Associate professor, [2, 3, 4, 5,] BTECH

[1, 2, 3, 4, 5,] Dept of CSE, St. Johns College of Engineering and Technology, Yerrakota, Yemmiganur, Kurnool, AP, Affiliated by JNTUA, INDIA

ABSTRACT

*Identification of bank cards fraud remains a serious worry for financial companies because of how sophisticated criminal behavior is becoming.. This study addresses this challenge by leveraging models increase the efficiency of detection systems. Utilizing a dataset from Kaggle, we implement and compare five distinct algorithms: LSTM networks, neural networks based on CNN, Decision Trees, Random Forests, and a Stacking Classifier. CNNs are employed to capture intricate patterns in transaction data, while LSTM address sequential dependencies and temporal dynamics. Decision Trees and Random Forests provide robust classification through hierarchical decision-making and ensemble learning. Additionally, a Stacking Classifier integrates the strengths of these algorithms to potentially improve overall performance. The comparative analysis of these methods tries to determine the best method for real-time identification of fraud. The results are expected to contribute significantly to the development of more secure credit card transaction systems, thereby mitigating financial losses and enhancing consumer trust.***

I. INTRODUCTION

Overview

Credit card fraud detection is a vital aspect of the financial industry, designed to protect both institutions and consumers from financial losses. As fraudulent activities become more sophisticated, traditional detection methods are struggling to keep pace with the evolving tactics used by cybercriminals. Fraudulent practices like identity theft, account takeover, and transaction manipulation are now being carried out using advanced technologies and algorithms, making detection increasingly difficult.

Fraudsters employ complex techniques, including artificial intelligence, to bypass conventional detection systems, highlighting the need for more advanced, adaptable fraud detection mechanisms. These sophisticated threats demand real-time, dynamic detection systems to identify and combat fraud effectively.

The economic consequences of credit card fraud are severe, leading to substantial losses for financial institutions and damage to their reputations. Consumers are also impacted by financial losses and emotional distress. Consequently, there is a pressing need for more effective and advanced fraud detection systems that can evolve alongside these new and complex threats.

Problem Statement

Fraud detection remains A critical concern for financial organazations, as traditional methods struggle to keep pace with the rapid evolution of fraudulent tactics. Conventional methods for detecting fraud frequently depend on rule-based techniques, which involve predefined criteria and patterns for identifying suspicious activities. While these systems can be effective against known types of fraud, they are inherently limited by their static nature. They often fail to detect new, sophisticated fraud schemes that do not match the established rules or patterns. This limitation leaves significant gaps in fraud prevention, allowing innovative fraudulent activities to go undetected. Another significant challenge is the dynamic Fraudsters are constantly evolving their techniques and strategies to carry out fraudulent activities. They consistently modify and refine their methods to stay ahead of detection systems.

Objective of the project

Main goal of this investigation is to leverage cutting-edge ML techniques to enhance the accuracy and efficiency of detection systems. Fraud remains a major challenge. For financial institutions because of their increasing sophistication of fraudulent activities. Traditional techniques for detecting cheating often fall

.CNN: are employed to capture intricate patterns and anomalies within transaction data. CNNs excel in identifying complex features through their hierarchical layers, making them suitable for detecting subtle signs of fraud.

LSTM : are utilized to address sequential dependencies and temporal dynamics inherent in transaction data. LSTM are particularly effective in analyzing time-series data, which is crucial for detecting fraud patterns that evolve over time.

Decision Trees: provide a transparent, hierarchical approach to classification, making decisions based on feature values. They offer interpretability and are useful for understanding decision rules.

Random Forests: build upon Decision Trees by employing an ensemble learning approach. This method aggregates using choice trees to enhance classification accuracy. Accuracy and reduce overfitting, enhancing the overall robustness of the fraud detection system

Limitations of the Project

This study employs a Kaggle dataset to Create and assess ML models for the detection of credit card fraud using the dataset provided. a critical component of this research, comprises transaction records with labeled instances of fraudulent and legitimate activities. The dataset encompasses multiple attributes, including the transaction amount and timestamp. and user-specific details, which are essential for training and testing the algorithms.

II. LITERATURE SURVEY

1. R. Sailusha, V. Gnaneswar, R. Ramesh, and G. R. Rao (2022) Title: Credit Card Fraud Detection Using Machine Learning

This study compares two popular machine learning algorithms, Random Forest and Adaboost, for credit card fraud detection. The paper evaluates both algorithms based on several metrics, including accuracy, precision, recall, F1-score, and the ROC curve. Through this comparison, the authors aim to determine which method is the most effective for detecting fraud in credit card transactions.

2. A. Thennakoon, C. Bhagvani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarac (2019)

Title: Real-time Credit Card Fraud Detection Using Machine Learning

This paper explores the use of machine learning models for real-time credit card fraud detection. The authors evaluate optimal algorithms for detecting various types of fraud and implement a system for real-time fraud detection through predictive analytics. An API is also used to facilitate the real-time application of the system, providing efficient fraud detection solutions.

*****3 1. J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare (2017)* Title: Credit card fraud detection using machine learning techniques: A comparative analysis

In this paper, the authors highlight the critical role of data mining in detecting credit card fraud. The study emphasizes the challenges posed by adaptive detection mechanisms and handling imbalanced datasets, particularly when working with real-time transaction data. Through a comparative analysis of various machine learning techniques, the paper offers insights into how data mining can enhance the accuracy and efficiency of fraud detection systems.

4. F. K. Alarfaj, L. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Aluned (2022) Title: Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms

This paper introduces a hybrid system combining an autoencoder with a Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) for credit card fraud detection. The proposed model aims to address the challenges of detecting financial fraud with high accuracy, particularly in the context of imbalanced datasets. The authors use oversampling techniques to balance the dataset, which significantly improves both recall and precision in fraud detection, making it a highly effective solution in real-world applications.

5 D. Varmedia, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla (2019) Title: Credit Card Fraud Detection Using Machine Learning Methods

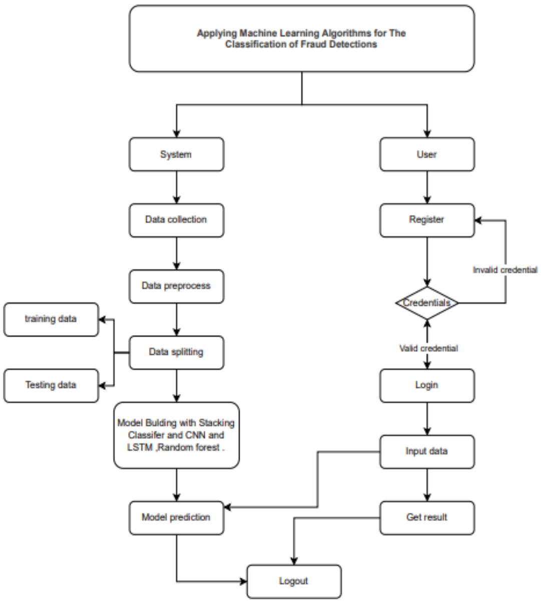
In this research, the authors employ various machine learning techniques to detect credit card fraud. They use SMOTE (Synthetic Minority Over-sampling Technique) for balancing the dataset, which helps in dealing with the common problem of imbalanced classes in fraud detection tasks. The paper also applies feature selection to enhance the performance of machine learning models, including Logistic Regression, Random Forest, Naive Bayes, and Multilayer Perceptron (MLP). Their approach demonstrates high accuracy in detecting fraud despite the challenges of working with imbalanced datasets.

III. SYSTEM ANALYSIS

Overview of Existing Model: Current credit card fraud detection systems primarily rely on rule-based methods, statistical models, and traditional machine learning algorithms such as Logistic Regression and Support Vector Machines (SVM). These systems often face limitations in detecting complex and evolving fraud patterns due to their reliance on static rules and the inability to capture temporal dependencies in transaction data. As fraudsters develop more sophisticated techniques, these systems struggle to maintain high accuracy and timely detection, leading to false positives and false negatives, which can result in either customer dissatisfaction or financial losses for institutions

Overview of Proposed Model: The proposed method for enhancing credit card fraud detection involves the implementation and comparison of five machine learning algorithms: Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), Decision Trees, Random Forests, and a Stacking Classifier. Each algorithm is chosen for its unique strengths in handling complex transactional data. CNNs are utilized to identify intricate patterns within the data, while LSTMs are employed to capture sequential dependencies and temporal relationships. Decision Trees offer straightforward, interpretable decision-making, and Random Forests enhance this by aggregating multiple trees to improve classification accuracy. The Stacking Classifier integrates predictions from the CNN, LSTM, Decision Trees, and Random Forests to create a more robust and comprehensive model. This method aims to optimize detection accuracy and efficiency, thereby reducing false positives and improving real-time detection capabilities in credit card fraud prevention systems.

Workflow of Proposed Model



IV. METHODOLOGY

Dataset Description

1Unnamed: 0 ==> This feature represents Likely an index or row identifier from the dataset. This column might not hold meaningful data and can often be dropped.

2trans_date_trans_time ==> This Componet indicates The timestamp of the transaction, typically in a date-time notation (e.g., HH:MM:SS or YYYY-MM-DD).

3cc_num ==> This feature indicates The credit card number used for the transaction. This is sensitive information and should be handled with care 4.merchant ==> This aspect indicates The name of the merchant or store where the transaction occurred.

5.category ==> This attribute records the kind or category of products or services that were bought (e.g., groceries, electronics).

6.amt ==> it's represents The amount of money spent in the transaction.

7. first ==> This section indicates The first name of the cardholder.

8. last ==> This feature captures The last name of the cardholder.

9. gender ==> it shows The gender of the cardholder (e.g., Male, Female).
10. street ==> The place of residence connected to the hyping information of the cardholder.
11.city ==> This aspect captures The merchant billing address city.
12.state ==> This feature indicates The status of the advertising address.
13.zip ==> This col represents The postal ZIP code associated with the cardholder's address.
14.lat ==> This component indicates The azimuth of the merchant's latitude coordinate.
15.long ==> This feature represents The geographic location reflected in the merchant's longitude coordinate.
16.city_pop ==> This variable represents The number of people residing in the city where the transaction took place.
17.job ==> This Columns indicates This feature representsThe occupation of the cardholder.
18.dob ==> it tells Record the cardholder's date of birth.
19. trans_num ==> This feature donates The unique transaction number or ID.
20.unix_time ==> it represents The timestamp of the transaction in Unix epoch time format.
21.merch_lat ==> This columns shows the latitude organize of where the merchant is located
22.merch_long ==> Column indicates the longitude coordinate of the holder location.
23.is_fraud ==> its represents A binary indicator (usually 0 or 1) showing whether the transaction was fraudulent (1) or not (0).

Data Collection

In this study, we utilize a dataset sourced from Kaggle, which is widely recognized for its extensive collection of data suitable for machine intelligence applications. The dataset is designed specifically for fraud detection and includes a diverse range of features essential for identifying fraudulent transactions. It comprises transaction details such as transaction amount, timestamp, and anonymized variables derived from credit card transactions, which capture various aspects of user behavior and transaction patterns.

Data Preprocessing

Effective data training is crucial for building robust subset of AI models, especially in the context of identificaton fraudwhere data quality can significantly impact model performance. The preprocessing stage involves several key steps:

Handling Missing Values:

Null values in the dataset are Located to ensure that the machine learning algorithms operate on complete data. Common methods include inference, in which statistical measurements are used to replace missing valuesSuch as the standard deviation, the median, or mode of operation, or more complex approaches. like k-Nearest Neighbors imputation. Alternatively, rows or columns with excessive missing values may be removed if they do not provide critical information.

Normalization:

To guarantee that different features contribute evenly to the model's performance., normalization is applied. This involves scaling feature values to a standard range, typically between 0 and 1 or -1 and 1. Methods such as Z-score Uniformity and Min-Max Scaling are commonly used. Normalization helps in stabilizing the training process and improving the convergence rate of algorithms like The networks made up of Long Short-Term Memory LSTM and Convolutional neural networks, more commonly are sensitive to the size of input features.

Feature Selection:

Selecting relevant features is essential to enhance model efficiency and accuracy. Feature selection methods, such as statistical tests, correlation analysis, and techniques like Recurring Structures Feature Elimination, or RFE or Principal Component Analysis (PCA), are used to identify and retain the most informative features while discarding redundant or irrelevant ones. This step helps in reducing dimensionality and mitigating the risk of overfitting.

Data Splitting:

The dataset is separated into training, validation, and test sets to evaluate model performance effectively. Typically, the data is split into Ten to fifteen percent for testing, seventy to eighty percent for training. The test set is used to gauge how well the final model can generalize, the validation set is used to adjust hyper parameters and monitor performance throughout training, and the training set is used to train the models.

V. Algorithm Implementation and Result

The implementation of algorithms for credit Detecting card theft requires a methodical approach to implementing CNN, LSTM, Decision Trees, and Random Forests. Each algorithm is tailored to address specific aspects of fraud detection and enhance the overall system's accuracy and efficiency.

CNN: The CNN architecture is structured to achieve and detect complex patterns in transaction data. The implementation begins with defining the network's layers, including feature extraction using convolutional layers that pooling layers for dimensionality reduction, and fully connected layers for classification. The convolutional layers use filters to scan through the transaction data, identifying intricate patterns that might indicate fraudulent activity. Pooling layers help in reducing the computational load by summarizing the features extracted. In order to minimize the loss function, the network is trained using gradient descent and backpropagation. and optimize performance.

LSTM:

LSTM networks are configured to handle sequential data, capturing temporal dependencies crucial for detecting fraudulent patterns over time. The implementation involves setting up LSTM cells that maintain and update internal states through forget, input, and output gates. This structure allows the LSTM to remember long-term dependencies and recognize patterns in transaction sequences. Training involves using sequences of transaction data, with the LSTM adjusting its parameters to learn the temporal relationships between events.

Stacking Classifier

A Stacking Classifier is an ensemble learning method that blends several machine learning models rise predictive performance. Unlike traditional ensemble methods like bagging or boosting, which use a single type of model, stacking integrates different types of models to leverage their unique strengths.

Decision Trees and Random Forests:

To create decision forests, split data at each node to classify transactions, and build a hierarchical model that makes judgments based on feature values. By training an ensemble of decision trees using a randomized a portion of the characteristics and data, Random Forests expand on this concept. The forest aggregates The different trees' forecasts show increased robustness. and accuracy. Both algorithms are trained and evaluated using metrics such as accuracy, precision, and recall, to ensure effective fraud detection.

LIME Implementation

LIME (Local Interpretable Model-agnostic Explanations) is a technique that explains predictions of machine learning models. It works by approximating a complex model with a simpler, interpretable model locally around the prediction.

Model Training and Tuning

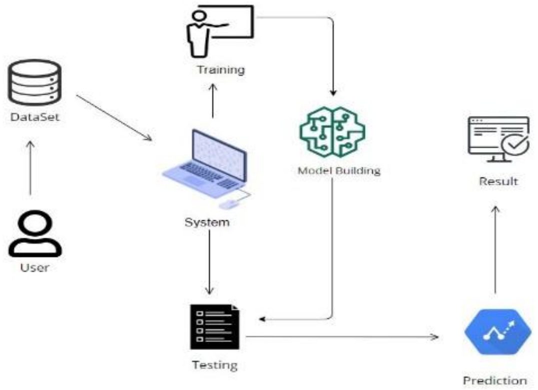
In this study, the model training process is designed to optimize the performance of five ML algorithms specifically for fraud detection. Each model undergoes extensive training using a preprocessed dataset from Kaggle, ensuring that the data is properly balanced to resolve the class disparity that exists in fraud detection activities.

For the CNN model, a grid search is conducted to fine-tune hyper parameters such as the number of convolutional layers, filter sizes, and learning rates. The LSTM model, given its ability to capture temporal dependencies, is optimized by adjusting the number of LSTM units, dropout rates, and sequence lengths. Decision Trees are pruned and tuned for depth and split criteria to avoid overfitting. Random Forests, being an ensemble of decision trees, are tuned by varying To find a Establish a balance between variation and bias, then calculate the maximum depth with all tree.

The Stacking Classifier combines the forecasts made by each separate model to generate a more reliable forecast. The base models for stacking are trained with optimal hyper parameters obtained from prior tuning processes. A meta-learner, typically a logistic regression or another tree-based model, is then trained on the outputs of these base models.

Cross-validation techniques are employed throughout the tuning process to validate the models' generalization performance, ensuring that the final models are not only accurate but also capable of handling real-world credit card transaction data efficiently.

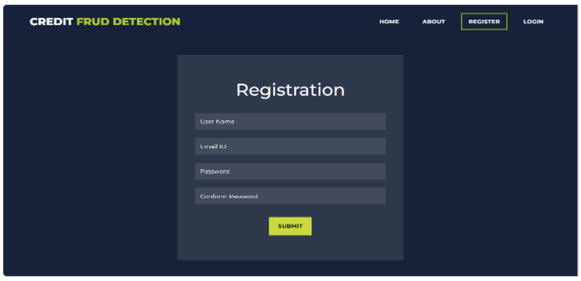
Optimization techniques such as grid search or random search are employed to systematically explore different hyper parameter combinations and identify the optimal settings. Cross-validation is used to evaluate model performance and ensure that it generalizes well to unseen data. This comprehensive training and tuning process target to boost the overall performance of the fraud detection system, resulting in a more precise and trustworthy identification of fraudulent transactions.



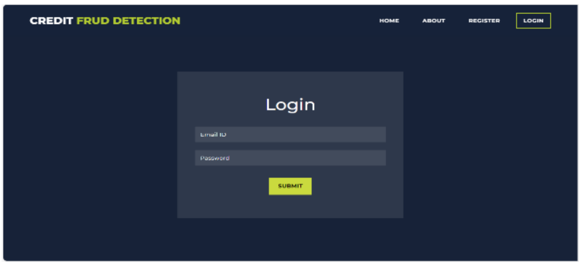
Architecture diagram

OUTPUT SCREEN:

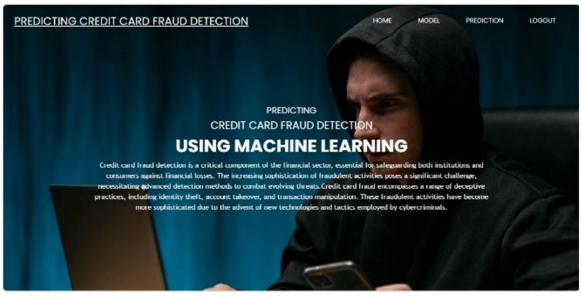
Registration Page:



Login Page:



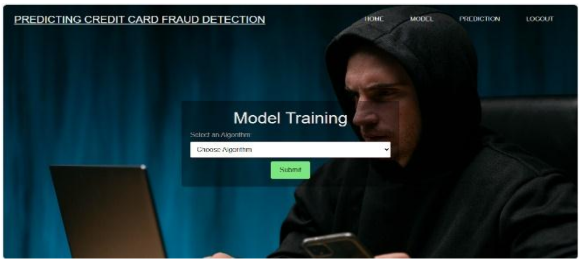
Model Home Page:



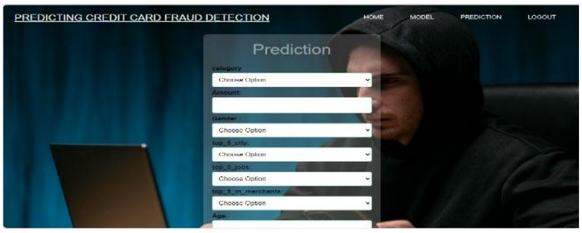
PREDICTING
CREDIT CARD FRAUD DETECTION
USING MACHINE LEARNING

Credit card fraud detection is a critical component of the financial sector, essential for safeguarding both institutions and consumers against financial losses. The increasing sophistication of fraudulent activities poses a significant challenge, necessitating advanced detection methods to combat evolving threats.Credit card fraud encompasses a range of deceptive practices, including identity theft, account takeover, and transaction manipulation. These fraudulent activities have become more sophisticated due to the advent of new technologies and tactics employed by cybercriminals.

Model Selection page:



Prediction page:



Prediction

category
Choose Option

Amount:

Gender :
Choose Option

top_5_city:
Choose Option

top_5_jobs:
Choose Option

top_5_in_merchants:
Choose Option

Age:

VI. CONCLUSION

The intention of this study was to compare four ml algorithms for fraud detection. Findings indicate that each of the four algorithms, namely CNN, LSTM, Decision Trees, and Random Forests, demonstrated promising performance in identifying fraudulent transactions.

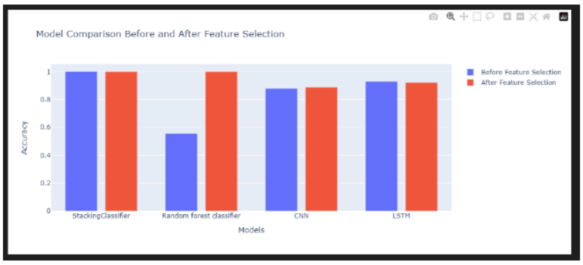


Figure Accuracy comparison

Key Results:
CNNs achieved an accuracy of 88.6% in detecting fraud, outperforming the other algorithms. This suggests that CNNs are effective in capturing intricate patterns in transaction data.
LSTM showed an accuracy of 92.2%, indicating their ability to model sequential dependencies and temporal dynamics in transaction data.

REFERENCES

[1]. Adi Saputra1, Suharjito2L: Fraud Detection using Machine Learning in e-Commerce, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.

[2]. Dart Consulting, Growth of Internet Users in India and Impact on Country's Economy: https://www.dartconsulting.co.in/marketnews/growth-of-internet-users-in-india-and-impact-on-countryseconomy/

[3]. Ganga Rama Koteswara Rao and R.Satya Prasad, " -Shielding The Networks Depending On Linux Servers Against Arp Spoofing, International Journal of Engineering and Technology(UAE),Vol. 7, PP.75-79, May 2018, ISSN No: 2227-524X, DOI - 10.14419/ijet.v7i2.32.13531.

[4]. Heta Naik , Prashasti Kanikar: Detctionsbased on Machine Learning Algorithms,International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.

[5]. Navanshu Khare ,Saad Yunus Sait: DetctionsUsing Machine Learning Models and Collating Machine Learning Models, International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.

[6]. Randula Koralage, , Faculty of Information Technology, University of Moratuwa,Data Mining Techniques for Credit Card Fraud Detection.

[7]. Roy, Abhimanyu, et al:Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.

[8]. Sahayasakila.V, D. Kavya Monisha, Aishwarya, Sikhakolli VenkatavisalakshiseshsaiYasaswi: DetctionsSystem using Smote Technique and Whale Optimization Algorithm,International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019.

[9]. Statista.com. retail e-commerce revenue forecast from 2017 to 2023 (in billion U.S. dollars). Retrieved April 2020, from India : https://www.statista.com/statistics/280925/e-commercerevenueforecast-in-india/

[10]. Yashvi Jain, NamrataTiwari, Shripri yaDubey, Sarika Jain:A Comparative Analysis of Various DetctionsTechniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019