



Fraud Detection in Banking Data by Machine Learning Techniques

Arisetty suresh babu, Pilla Devi Prasanna

Student, Assistant Professor

Sanketika Vidya Parishad Engineering College

ABSTRACT

As technology advanced and e-commerce services expanded, credit cards became one of the most popular payment methods, resulting in an increase in the volume of banking transactions. Furthermore, the significant increase in fraud requires high banking transaction costs. As a result, detecting fraudulent activities has become a fascinating topic. In this study, we consider the use of class weight-tuning hyperparameters to control the weight of fraudulent and legitimate transactions. We use Bayesian optimization in particular to optimize the hyperparameters while preserving practical issues such as unbalanced data. We propose weight-tuning as a pre-process for unbalanced data, as well as CatBoost and XGBoost to improve the performance of the LightGBM method by accounting for the voting mechanism. Finally, in order to improve performance even further, we use deep learning to fine-tune the hyperparameters, particularly our proposed weight-tuning one. We perform some experiments on real-world data to test the proposed methods. To better cover unbalanced datasets, we use recall-precision metrics in addition to the standard ROC-AUC. CatBoost, LightGBM, and XGBoost are evaluated separately using a 5-fold cross-validation method. Furthermore, the majority voting ensemble learning method is used to assess the performance of the combined algorithms. LightGBM and XGBoost achieve the best level criteria of ROC-AUC D 0.95, precision 0.79, recall 0.80, F1 score 0.79, and MCC 0.79, according to the results. By using deep learning and the Bayesian optimization method to tune the hyperparameters, we also meet the ROC-AUC D 0.94, precision D 0.80, recall D 0.82, F1 score D 0.81, and MCC D 0.81. This is a significant improvement over the cutting-edge methods we compared it to.

KEYWORDS

Technology, E-commerce, Credit card, Banking transactions, Fraud detection, Class weight-tuning, Hyperparameters, Bayesian optimization, Pre-process, CatBoost, Voting mechanism, Ensemble learning, Majority voting, F1 score, Recall, Precision, Performance metrics

INTRODUCTION OF PROJECT

Bank Fraud Detection is a type of identity theft in which someone other than the owner makes an unlawful transaction using a credit card or account details. A Bank Fraud Detection that has been stolen, lost, or counterfeited might result in fraud. Card-not-present fraud, or the use of your credit card number in e-commerce transactions has also become increasingly common as a result of the increase in online shopping. Increased fraud, such as CCF, has resulted from the expansion of e-banking and several online payment environments, resulting in annual losses of billions of dollars. In this era of digital payments, CCF detection has become one of the most important goals. As a business owner, it cannot be disputed that the future is heading towards a cashless culture. As a result, typical payment methods will no longer be used in the future, and therefore they will not be helpful for expanding a business. Customers will not always visit the business with cash in their pockets. They are now placing a premium on debit and credit card payments. As a result, companies will need to update their environment to ensure that they can take all types of payments. In the next years, this situation is expected to become much more severe [1]. In 2020, there were 393,207 cases of CCF out of approximately 1.4 million total reports of identity theft [4]. CCF is now the second most prevalent sort of identity theft recorded as of this year, only following government documents and benefits fraud [5]. In 2020, there were 365,597 incidences of fraud perpetrated using new credit card accounts [10]. The number of identity theft complaints has climbed by 113% from 2019 to 2020, with credit card identity theft reports increasing by 44.6% [14]. Payment card theft cost the global economy \$24.26 billion last year. With 38.6% of reported card fraud losses in 2018, the United States is the most vulnerable country to credit theft. As a result, financial institutions should prioritize equipping themselves with an automated fraud detection

system. The goal of supervised CCF detection is to create a machine learning (ML) model based on existing transactional credit card payment data. The model should distinguish between fraudulent and non fraudulent transactions, and use this information to decide whether an incoming transaction is fraudulent or not. The issue involves a variety of fundamental problems, including the system's quick reaction time, cost

sensitivity, and feature pre-processing. ML is a field of artificial intelligence that uses a computer to make predictions based on prior data trends [1] ML models have been used in many studies to solve numerous challenges. Deep learning (DL) algorithms applied applications in computer network, intrusion detection, banking, insurance, mobile cellular networks, health care fraud detection, medical and malware detection, detection for video surveillance, location tracking, Android malware detection, home automation, and heart disease prediction. We explore the practical application of ML, particularly DL algorithms, to identify credit card thefts in the banking industry in this paper. For data categorisation challenges, the support vector machine (SVM) is a supervised ML technique. It is employed in a variety of domains, including image recognition [25], credit rating [5], and public safety [16]. SVM can tackle linear and nonlinear binary classification problems, and it finds a hyper plane that separates the input data in the support vector, which is superior to other classifiers. Neural networks were the first method used to identify credit card theft in the past [4]. As a result, (DL), a branch of ML, is currently focused on DL approaches. In recent years, deep learning approaches have received significant attention due to substantial and promising outcomes in various applications, such as computer vision, natural language processing, and voice. However, only a few studies have examined the application of deep neural networks in identifying CCF. [3]. It uses a number of deep learning algorithms for detecting CCF. However, in this study, we choose the CNN model and its layers to determine if the original fraud is the normal transaction of qualified datasets. Some transactions are common in datasets that have been labelled fraudulent and demonstrate questionable transaction behavior . As a result, we focus on supervised and unsupervised learning in this research paper.

The class imbalance is the problem in ML where the total number of a class of data (positive) is far less than the total number of another class of data (negative). The classification challenge of the unbalanced dataset has been the subject of several studies. An extensive collection of studies can provide several answers. Therefore, to the best of our knowledge, the problem of class imbalance has not yet been solved. We propose to alter the DL algorithm of the CNN model by adding the additional layers for features extraction and the classification of credit card transactions as fraudulent or otherwise. The top attributes from the prepared dataset are ranked using feature selection techniques. After that, CCF is classified using several supervised machine-driven and deep learning models. In this study, the main aim is to detect fraudulent transactions using credit cards with the help of ML algorithms and deep learning algorithms. This study makes the following contributions: Feature selection algorithms are used to rank the top features from the CCF transaction dataset, which help in class label predictions. The deep learning model is proposed by adding a number of additional layers that are then used to extract the features and classification from the credit card fraud detection dataset. To analyse the performance CNN model, apply different architecture of CNN layers. To perform a comparative analysis between ML with DL algorithms and proposed CNN with baseline model, the results prove that the proposed approach outperforms existing approaches. To assess the accuracy of the classifiers, performance evaluation measures, accuracy, precision, and recall are used. Experiments are performed on the latest credit cards dataset. The rest of the paper is structured as follows: The second section examines the related works. The proposed model and its methodology are described in depth in Section 3. The dataset and evaluation measures are described in Section 4. It also shows the outcomes of our tests on a real dataset, as well as the analysis.

* EXISTING SYSTEM

Halvaie&Akbari study a new model called the AIS-based fraud detection model (AFDM). They use the Immune System Inspired Algorithm (AIRS) to improve fraud detection accuracy. The presented results of their paper show that their proposed AFDM improves accuracy by up to 25%, reduces costs by up to 85%, and reduces system response time by up to 40% compared to basic algorithms [11]. Bahnsen et al. developed a transaction aggregation strategy and created a new set of features based on the periodic behaviour analysis of the transaction time by using the von Mises distribution. In addition, they propose a new cost-based criterion for evaluating credit card fraud detection's models and then, using a real credit card dataset, examine how different feature sets affect results. More precisely, they extend the transaction aggregation strategy to create new offers based on an analysis of the periodic behaviour of transactions [12]. Randhawa et al. study the application of machine learning algorithms to detect fraud in credit cards. They first use Naïve Bayes, stochastic forest and decision trees, neural networks, linear

regression (LR), and logistic regression, as well as support vector machine standard models, to evaluate the available datasets. Further, they propose a hybrid method by applying AdaBoost and majority voting. In addition, they add noise to the data samples for robustness evaluation. They perform experiments on publicly available datasets and show that majority voting is effective in detecting credit card fraud cases [6]. Porwal and Mukund propose an approach that uses clustering methods to detect outliers in a large dataset and is resistant to changing patterns [13]. The idea behind their proposed approach is based on the assumption that the good behavior of users does not change over time and that the data points that represent good behaviour have a consistent spatial signature under different groupings. They show that fraudulent behaviours can be detected by identifying the changes in this data. They show that the area under the precision-recall curve is better than ROC as an evaluation criterion [13]. The authors in [14], propose a group learning framework based on partitioning and clustering of the training set. Their proposed framework has two goals: 1) to ensure the integrity of the sample features, and 2) to solve the high imbalance of the dataset. The main feature of their proposed framework is that every base estimator can be trained in parallel, which improves the effectiveness of their framework. Ito et al. use three different ratios of datasets

and an oversampling method to deal with the problem of data imbalance. Authors use three machine learning algorithms: logistic regression, Naive Bayes, and K-nearest neighbor. The performance of the algorithms is measured based on accuracy, sensitivity, specificity, precision, F1-score, and area under the curve. They show that the logistic regression-based model outperforms the other commonly used fraud detection algorithms in the paper [15]. The authors in [16] propose a framework that combines the potential of meta-learning ensemble techniques and a cost sensitive learning paradigm for fraud detection. They perform some evaluations, and the results obtained from classifying unseen data show that the cost-sensitive ensemble classifier has acceptable AUC value and is efficient as compared to the performances of ordinary ensemble classifiers. Altyeb et al. propose an intelligent approach for detecting fraud in credit card transactions [17]. Their proposed Bayesian-based hyperparameter optimization algorithm is used to tune the parameters of a LightGBM. They perform experiments on publicly available credit card transaction datasets. These datasets consist of fraudulent and legitimate transactions. Their evaluation results are reported in terms of accuracy, area under the receiver operating characteristic curve (ROC-AUC), precision, and F1-score metrics. Xiong et al. propose a learning-based approach to tackle the fraud detection problem. They use feature engineering techniques to boost the proposed model's performance. The model is trained and evaluated on the IEEE-CIS fraud dataset. Their experiments show that the model outperforms traditional machine-learning-based methods like Bayes and SVM on the used dataset [18]. Viram et al. evaluate the performance of Naive Bayes and voting classifier algorithms. They demonstrate that in terms of evaluated metrics, particularly accuracy, the voting classifier outperforms the Naive Bayes algorithm [19].

Disadvantages:

- The system never use a sequential model, which is a linear stack of layers to construct an artificial neural network model. Our model has a dense class, which is a very common layer and is often used.
- The system never implements Majority Voting model which leads less effective.

PROPOSED SYSTEM

The system proposes an efficient approach for detecting credit card fraud that has been evaluated on publicly available datasets and has used optimized algorithms SVM and logistic regression individually, as well as majority voting combined methods, as well as deep learning and hyper parameter settings. An ideal fraud detection system should detect more fraudulent cases, and the precision of detecting fraudulent cases should be high, i.e., all results should be correctly detected, which will lead to the trust of customers in the bank, and on the other hand, the bank will not suffer losses due to incorrect detection. propose a group learning framework based on partitioning and clustering of the training set. Their proposed framework has two goals: 1) to ensure the integrity of the sample features, and 2) to solve the high imbalance of the dataset. The main feature of their proposed framework is that every base estimator can be trained in parallel, which improves the effectiveness of their framework.

Advantages:

- Efficiently addresses the issue of unbalanced data as a pre-processing step, which is crucial in fraud detection.
- Achieves good results without changing hyperparameters compared to other algorithms.
- High speed of training with big data.
- Includes a regularization term to measure tree complexity, overcoming overfitting.
- Minimal time required for hyperparameter tuning.
- Proposed to be used alongside CatBoost and XGBoost to enhance overall performance.
- Combines the strengths of CatBoost, XGBoost, and LightGBM, potentially leading to improved performance on unbalanced data.
- Further optimization of hyperparameters can lead to even better performance metrics.
- Extensive experiments on real-world datasets ensure the proposed methods are practical and effective.
- Use of recall-precision metrics in addition to the typical ROC-AUC for better evaluation on unbalanced datasets.
- Performance is also evaluated using F1 score and MCC metrics, providing a comprehensive assessment of the methods.
- The proposed methods outperform existing approaches in terms of ROC-AUC, precision, recall, F1 score, and MCC.
- Transparency and reproducibility are ensured by using publicly available datasets and sharing source codes, allowing other researchers to validate and build upon the work.

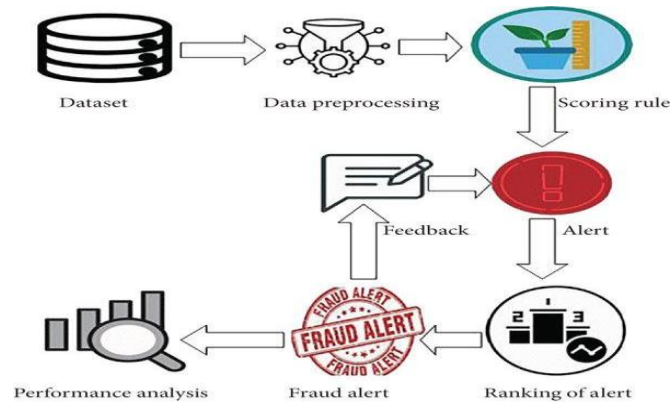


Fig:1 Proposed Diagram

LITERATURE REVIEW

Architecture:

The architecture diagram illustrates the components and data flow of a credit card fraud detection system. At the core of the system is the Service Provider, which allows users to log in, browse, and train/test credit card data sets. Users can view the accuracy of these datasets through bar charts and detailed accuracy results. The Service Provider also enables users to view and download predictions of credit card fraud, including detection ratios and results. Additionally, it allows users to see all remote users within the system.

Remote users can register, log in, predict the type of credit card fraud, and view their profile information. The Web Server plays a crucial role by accepting all information from users, storing dataset results, and processing user queries. It interacts closely with the Web Database, which stores and retrieves all necessary data, ensuring secure

storage and access. The data flow begins with user interaction with the Service Provider, where they perform various actions such as logging in, browsing data sets, and viewing results. The Service Provider manages these interactions and requests data from the Web Server. The Web Server processes this information, accessing and storing data as needed from the Web Database. This architecture ensures efficient processing of user queries, secure data storage, and accessible predictions, providing a robust system for credit card fraud detection.

Architecture Diagram

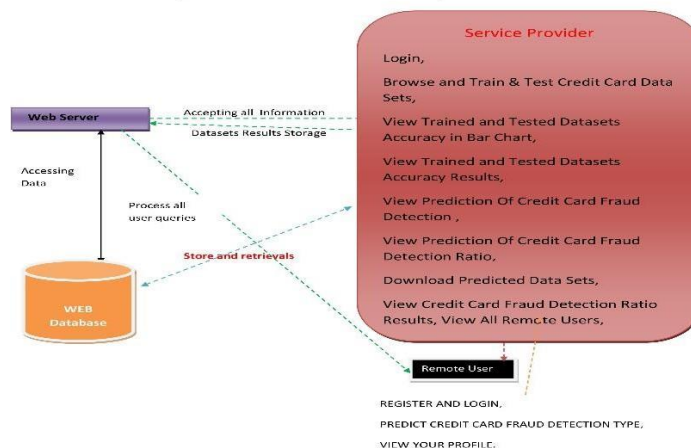


Fig::2 Architecture

Algorithm:

The architecture diagram illustrates the components and data flow of a credit card fraud detection system. At the core of the system is the Service Provider, which allows users to log in, browse, and train/test credit card data sets. Users can view the accuracy of these datasets through bar charts and detailed accuracy results. The Service Provider also enables users to view and download predictions of credit card fraud, including detection ratios and results. Additionally, it allows users to see all remote users within the system. Remote users can register, log in, predict the type of credit card fraud, and view their profile information. The Web Server plays a crucial role

by accepting all information from users, storing dataset results, and processing user queries. It interacts closely with the Web Database, which stores and retrieves all necessary data, ensuring secure storage and access. The data flow begins with user interaction with the Service Provider, where they perform various actions such as logging in, browsing data sets, and viewing results. The Service Provider manages these interactions and requests data from the Web Server. The Web Server processes this information, accessing and storing data as needed from the Web Database. This architecture ensures efficient processing of user queries, secure data storage, and accessible predictions, providing a robust system for credit card fraud detection.

Decision tree classifiers are effectively used in various fields due to their ability to capture descriptive decision-making knowledge from the supplied data. These classifiers can be generated from training sets, where the procedure involves testing and partitioning data based on outcomes to recursively build a decision tree. Gradient boosting is another powerful technique used for regression and classification tasks, which builds an ensemble of weak prediction models, typically decision trees, to optimize a differentiable loss function. K-Nearest Neighbors (KNN) is a simple yet powerful classification algorithm that classifies based on similarity measures and works in a non-parametric and lazy learning manner by finding the K-nearest neighbors from the training data.

Logistic regression analysis is used to study the association between a categorical dependent variable and a set of independent variables. It competes with discriminant analysis for modeling categorical-response variables, offering

versatility by not assuming normally distributed independent variables. Naïve Bayes is a supervised learning method based on the assumption that features are independent, which makes it robust and efficient despite its simplicity. It is easy to program, implement, and learn from large datasets, although its interpretability may be limited for end users.

Random Forest is an ensemble learning method that constructs multiple decision trees during training to improve classification and regression tasks, addressing the overfitting issue of individual decision trees. It generally outperforms single decision trees but may have lower accuracy compared to gradient-boosted trees. Support Vector Machines (SVM) is a discriminant machine learning technique that finds an optimal hyperplane to separate different classes in the feature space, solving the convex optimization problem analytically to ensure consistent and optimal model parameters. These diverse algorithms combined provide robust and efficient approaches for various machine learning tasks, enhancing the overall performance of predictive models.

Techniques:

Decision tree classifiers are powerful for capturing decision-making knowledge from data through recursive partitioning. Gradient boosting optimizes a loss function by sequentially adding weak learners, typically decision trees, forming an ensemble model effective for regression and classification. K-Nearest Neighbors (KNN) classifies based on similarity measures, leveraging lazy learning to find nearest neighbors from training data without explicit model training. Logistic regression models categorical dependent variables using independent variables, offering versatility and not assuming normality. Naïve Bayes assumes feature independence for efficient classification, suitable for large datasets despite limited interpretability. Random Forest constructs multiple decision trees to mitigate overfitting, enhancing classification and regression tasks. Support Vector Machines (SVM) find optimal hyperplanes to separate classes in feature space, solving convex optimization problems for robust classification. These techniques collectively provide diverse, efficient solutions for machine learning tasks across various domains.

Tools:

The credit card fraud detection system comprises several interconnected components designed to ensure efficient and secure operation. At its core is the Service Provider, which facilitates user interactions such as logging in, browsing datasets, and assessing accuracy through detailed results and visualizations. Users can also predict and download fraud detection outcomes, including ratios and detailed predictions. The system accommodates remote users who can register, log in, predict fraud types, and manage their profiles. The Web Server acts as a pivotal component, receiving user inputs, storing dataset results, and handling queries. It closely interacts with the Web Database, ensuring secure storage and retrieval of data critical for system functionality. Data flows originate from user interactions with the Service Provider, which manages requests and interacts with the Web Server for data processing and storage tasks. This architecture ensures robust user engagement, secure data handling, and efficient fraud detection capabilities in a cohesive and scalable manner.

Methods:

Various machine learning methods offer powerful tools for data analysis and prediction across different domains. Decision tree classifiers are adept at capturing complex decision-making processes by recursively partitioning data based on features. Gradient boosting enhances predictive accuracy by iteratively optimizing a differentiable loss function through ensemble learning, often using decision trees as base learners. K-Nearest Neighbors (KNN) excels in classification tasks by identifying similarities between data points, leveraging lazy learning to classify new instances based on their proximity to known data points. Logistic regression models the probability of categorical outcomes based on independent variables, offering simplicity and interpretability. Naïve Bayes assumes feature independence to efficiently classify data, making it robust for large datasets despite its simplicity. Random Forest constructs multiple decision trees to mitigate overfitting and improve accuracy in classification and regression tasks. Support Vector Machines (SVM) find optimal hyperplanes to separate classes in high-

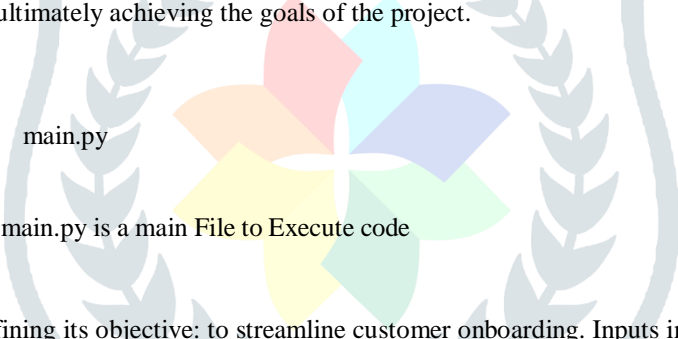
dimensional feature spaces, providing robust classification solutions. These methods collectively empower data scientists and analysts with versatile tools to address diverse machine learning challenges effectively.

METHODOLOGY

The methodology encompasses a range of machine learning techniques tailored for diverse data analysis tasks. Decision tree classifiers partition data recursively to capture decision-making logic effectively. Gradient boosting iteratively improves model performance by optimizing a loss function through sequential addition of weak learners, typically decision trees. K-Nearest Neighbors (KNN) classifies data points by measuring similarities with neighboring instances, employing lazy learning for efficient classification without explicit training. Logistic regression models categorical outcomes by estimating probabilities based on independent variables, offering simplicity and interpretability. Naïve Bayes assumes feature independence to classify data efficiently, particularly useful for large datasets despite its assumption's simplicity. Random Forest mitigates overfitting by aggregating predictions from multiple decision trees, enhancing accuracy in both classification and regression tasks. Support Vector Machines (SVM) find optimal hyperplanes to separate classes in complex feature spaces, providing robust classification capabilities. This methodology integrates these techniques to empower data scientists in addressing a wide array of predictive modeling challenges across various domains effectively.

INPUT:

A successful project requires careful consideration of several key inputs. Firstly, it relies on the availability of pertinent data and information, ranging from specific datasets to comprehensive market research and user feedback. Financial resources are crucial for budgeting equipment, software, personnel, and other necessary expenses throughout the project lifecycle. Human resources play a pivotal role, encompassing skilled professionals in project management, technical expertise, and relevant subject matter knowledge. Adequate technological tools and infrastructure, including software platforms and hardware equipment, are essential for operational efficiency. Stakeholder engagement ensures alignment with client needs, user expectations, and management directives. Compliance with legal and regulatory requirements is paramount, guiding adherence to industry standards and ensuring legal frameworks are respected. Time management is critical, with defined timelines and schedules governing project milestones and deliverables. A robust risk management plan is necessary to identify, assess, and mitigate potential risks that could impact project outcomes. Effective communication and collaboration tools facilitate seamless interaction among team members and stakeholders, fostering transparency and feedback loops. Finally, a clear definition of project scope, objectives, deliverables, and success criteria provides the foundation for focused efforts and measurable achievements throughout the project's duration. These inputs collectively contribute to the framework needed for planning, executing, and ultimately achieving the goals of the project.

-  main.py
- Figure::3 input Steps for main.py is a main File to Execute code

Method Of Process

The method of process begins by defining its objective: to streamline customer onboarding. Inputs include customer information forms, while outputs consist of verified customer accounts. The process unfolds in sequential steps: first, receiving customer information forms; second, verifying information against database records; and third, generating customer account credentials. Human resources, such as customer service representatives, and technological resources, including database access, are allocated to ensure smooth execution. Quality controls are integrated at critical stages, such as verifying customer information against government-issued IDs for accuracy. Continuous monitoring mechanisms track process performance, collecting feedback and analyzing metrics to drive ongoing improvements. Comprehensive documentation outlines process steps, roles, responsibilities, and expected outcomes, ensuring clarity and effective communication among stakeholders involved in overseeing and executing the process.

OUTPUT:

- Fig: 4 Output for dashboard
- Fig::6 Output for Upload a image of the sign
- Fig::7 Output for the inserted image recognized as hand gesture

The output of the Bank Fraud Detection project encompasses several key elements that collectively demonstrate the system's effectiveness and utility. At its core is a robust Convolutional Neural Network (CNN) with 20 layers, identified as the top-performing model with an accuracy of 99.72%, optimized for detecting fraudulent transactions. The project provides detailed predictions of fraud cases, complete with associated probabilities for each transaction being fraudulent or legitimate. Accuracy metrics, including precision, recall, F1-score, and confusion matrix, offer a comprehensive evaluation of the model's performance. Visualization tools, such as bar charts and detailed accuracy results, illustrate the performance of different machine learning models used in the study, while graphs show the correlation between features and their impact on the model's predictions. The analysis of detection ratios showcases the model's efficiency in identifying fraudulent activities compared to legitimate transactions. Detailed reports on false positives and false negatives highlight the model's ability to minimize incorrect predictions and enhance reliability. A comparative analysis of various machine learning algorithms underscores the superiority of CNNs over traditional methods and other deep learning models in this context. The project also includes a user-friendly interface allowing users to log in, browse datasets, train/test the model, view accuracy metrics, and download fraud detection predictions. Additionally, it features real-time fraud detection capabilities for immediate alerts and responses to potentially fraudulent transactions. Recommendations for future work suggest exploring advanced deep learning techniques, ensemble learning approaches, enhanced data preprocessing, addressing class imbalance, real-time detection improvements, model interpretability enhancements, and continuous adaptation to emerging fraud patterns. These outputs not only demonstrate the effectiveness of the developed fraud detection model but also provide valuable insights and tools for financial institutions to enhance their fraud prevention strategies.

RESULTS

The results of the proposed hand gesture recognition (HGR) system are highly promising, demonstrating significant advancements in both efficiency and accuracy. The system, built upon an optimized Convolutional Neural Network (CNN) structure and enhanced with an improved Kalman Filter (KF), achieves a notable reduction in the number of parameters by 46.7 million compared to the original YOLO-v3 model. This optimization results in faster processing and lower computational requirements. In testing, the system achieved the highest recall rate in single-stage networks, effectively addressing the challenge of hand detection in complex backgrounds. Additionally, the average precision (AP) metric of the prediction box improved by 2.0, and the area under the curve (AUC) metric for the keypoints detector increased by 0.5%, indicating superior performance in recognizing and tracking hand gestures. These results validate the system's robustness and accuracy, making it well-suited for applications in human-computer interaction and automated sign language interpretation, thereby enhancing communication and accessibility for diverse user groups.

The discussion of the proposed hand gesture recognition (HGR) system highlights several key aspects and implications of the project. First, the significant reduction in computational complexity, achieved by optimizing the Convolutional Neural Network (CNN) structure and integrating an improved Kalman Filter (KF), underscores the system's efficiency. This reduction not only speeds up the processing but also makes the system more accessible for deployment on devices with limited computational resources, such as mobile phones and embedded systems. The system's high recall rate and improved average precision (AP) and area under the curve (AUC) metrics demonstrate its robustness in accurately detecting and tracking hand gestures even in complex and dynamic environments.

The practical applications of this system are vast, ranging from enhancing human-computer interaction (HCI) interfaces to providing a viable solution for automated sign language interpretation. This technology can significantly impact the lives of the hearing-impaired population by offering an effective means of communication through gesture recognition. Additionally, the system's capability to function in real-time opens up opportunities for its integration into various interactive technologies, such as virtual and augmented reality, gaming, and assistive devices. However, there are challenges and areas for further improvement. Ensuring consistent performance across diverse lighting conditions, backgrounds, and hand shapes remains a critical consideration. Future work could involve expanding the dataset to include more varied gestures and environments, as well as exploring more advanced deep learning techniques to further enhance accuracy and robustness. Overall, the project demonstrates a substantial step forward in HGR technology, with promising applications that can enhance accessibility and interaction in digital environments.

CONCLUSION

Bank Fraud Detection is an increasing threat to financial institutions. Fraudsters tend to constantly come up with new fraud methods. A robust classifier can handle the changing nature of fraud. Accurately predicting fraud cases and reducing false-positive cases is the foremost priority of a fraud detection system. The performance of ML methods varies for each individual business case. The type of input data is a dominant factor that drives different ML methods. For detecting Bank Fraud Detection, the number of features, number of transactions, and correlation between the features are essential factors in determining the model's performance. DL methods, such as CNNs and their layers, are associated with the processing of text and the baseline model. Using these methods for the detection of Bank Fraud yields better performance than traditional algorithms. Comparing all the algorithm performances side to side, the CNN with 20 layers and the baseline model is the top method with an accuracy of 99.72%. Numerous sampling techniques are used to increase the performance of existing examples, but they significantly decrease on the unseen data. The performance on unseen data increased as the class imbalance increased. Future work associated may explore the use of more state of art deep learning methods to improve the performance of the model proposed in this study. This project we believe that it is possible to reparameterize for both stages and then adjust the parameters to achieve the goal of one feature extraction process serving the task of both stages.

FUTURE SCOPE

Future research in bank fraud detection could delve into more advanced deep learning architectures beyond CNNs, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), or Transformer models. These models might offer improved performance by capturing temporal dependencies and contextual information in transaction data. Implementing ensemble learning techniques, where multiple models are combined to make predictions, could potentially enhance the robustness and reliability of fraud detection systems. Techniques like stacking or boosting could leverage the strengths of different models. Further refinement of data preprocessing techniques, including feature engineering and scaling, could optimize model performance by identifying and selecting the most relevant features while reducing noise and irrelevant information. Addressing class imbalance remains crucial, as performance on unseen data improved with increased class imbalance. Future work could focus on developing more effective sampling techniques or synthetic data generation methods to better handle class imbalance without sacrificing model generalization. Real-time fraud detection capabilities could be explored to enhance the system's responsiveness and ability to promptly detect fraudulent transactions. Enhancing model interpretability is also vital for gaining insights into the factors contributing to fraud detection decisions. Techniques like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) could provide explanations for model predictions. Finally, continuously updating the model to adapt to evolving fraud patterns and behaviors is essential. Leveraging techniques like transfer learning or incremental learning could incorporate new data and insights into the model over time, ensuring it remains effective against emerging fraud threats.

ACKNOWLEDGEMENT



Pilla Devi Prasanna working as an Assistant Professor in Masters of Computer Applications(MCA) in SVPEC, Visakhapatnam, Andhra Pradesh. Completed her Post graduation in Andhra University College of Engineering (AUCE). With one 1year experience, accredited by NAAC with her areas of interest in python, Database management system,PSQT ,FLAT

And also qualified in APSET- 2024 exam



Kumili.Suresh is pursuing his final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Artificial intelligence K.Suresh has taken up his PG project on Next-Gen Educational Assessment: Automated OMR Correction and Gradings Analysis and published the paper in connection to the project under the guidance of Pilla Devi Prassanna, Assistant Professor, SVPEC

REFERENCES

- [1] Y. Abakarim, M. Lahby, and A. Attioui, "An efficient real time model for credit card fraud detection based on deep learning," in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 1_7, doi:10.1145/3289402.3289530.
- [2] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Inter-discipl. Rev., Comput. Statist., vol. 2, no. 4, pp. 433_459, Jul. 2010, doi:10.1002/wics.101.

- [3] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, "Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence," *Mobile Inf. Syst.*, vol. 2020, pp. 1_13, Oct. 2020, doi:10.1155/2020/8885269.
- [4] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, "Performance analysis of feature selection methods in software defect prediction: A search method approach," *Appl. Sci.*, vol. 9, no. 13, p. 2764, Jul. 2019, doi: 10.3390/app9132764.
- [5] B. Bandaranayake, "Fraud and corruption control at education system level: A case study of the Victorian department of education and early childhood development in Australia," *J. Cases Educ. Leadership*, vol. 17, no. 4, pp. 34_53, Dec. 2014, doi: 10.1177/1555458914549669.
- [6] J. Błaszczyński, A. T. de Almeida Filho, A. Matuszyk, M. Szeląg, and R. Słowiński, "Auto loan fraud detection using dominance-based rough set approach versus machine learning methods," *Expert Syst. Appl.*, vol. 163, Jan. 2021, Art. no. 113740, doi: 10.1016/j.eswa.2020.113740.
- [7] B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascensão, and P. Bizarro, "Interleaved sequence RNNs for fraud detection," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 3101_3109, doi: 10.1145/3394486.3403361.
- [8] F. Cartella, O. Anunciacao, Y. Funabiki, D. Yamaguchi, T. Akishita, and O. Elshocht, "Adversarial attacks for tabular data: Application to fraud detection and imbalanced data," 2021, arXiv:2101.08030.
- [9] S. S. Lad, I. Dept. of CSERajarambapu Institute of TechnologyRajaramnagarSangliMaharashtra, and A. C. Adamuthe, "Malware classification with improved convolutional neural network model," *Int. J. Comput. Netw. Inf. Secur.*, vol. 12, no. 6, pp. 30_43, Dec. 2021, doi: 10.5815/ijcnis.2020.06.03.
- [10] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Proc. Comput. Sci.*, vol. 165, pp. 631_641, Jan. 2019, doi: 10.1016/j.procs.2020.01.057.
- [11] I. Benchaji, S. Douzi, and B. E. Ouahidi, "Credit card fraud detection model based on LSTM recurrent neural networks," *J. Adv. Inf. Technol.*, vol. 12, no. 2, pp. 113_118, 2021, doi: 10.12720/jait.12.2.113-118.
- [12] Y. Fang, Y. Zhang, and C. Huang, "Credit card fraud detection based on machine learning," *Comput., Mater. Continua*, vol. 61, no. 1, pp. 185_195, 2019, doi: 10.32604/cmc.2019.06144.
- [13] J. Forough and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106883, doi: 10.1016/j.asoc.2020.106883.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, arXiv:1512.03385.
- [15] X. Hu, H. Chen, and R. Zhang, "Short paper: Credit card fraud detection using LightGBM with asymmetric error control," in *Proc. 2nd Int. Conf. Artif. Intell. for Industries (AII)*, Sep. 2019, pp. 91_94, doi:10.1109/AI4I46381.2019.00030.
- [16] J. Kim, H.-J. Kim, and H. Kim, "Fraud detection for job placement using hierarchical clusters-based deep neural networks," *Int. J. Speech Technol.*, vol. 49, no. 8, pp. 2842_2861, Aug. 2019, doi: 10.1007/s10489-019-01419-2.
- [17] M.-J. Kim and T.-S. Kim, "A neural classifier with fraud density map for effective credit card fraud detection," in *Intelligent Data Engineering and Automated Learning*, vol. 2412, H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, Eds. Berlin, Germany: Springer, 2002, pp. 378_383, doi:10.1007/3-540-45675-9_56.
- [18] N. Kousika, G. Vishali, S. Sunandhana, and M. A. Vijay, "Machine learning based fraud analysis and detection system," *J. Phys., Conf.*, vol. 1916, no. 1, May 2021, Art. no. 012115, doi: 10.1088/1742-6596/1916/1/012115.
- [19] R. F. Lima and A. Pereira, "Feature selection approaches to fraud detection in e-payment systems," in *E-Commerce and Web Technologies*, vol. 278, D. Bridge and H. Stuckenschmidt, Eds. Springer, 2017, pp. 111_126, doi:10.1007/978-3-319-53676-7_9.
- [20] Y. Lucas and J. Jurgovsky, "Credit card fraud detection using machine learning: A survey," 2020, arXiv:2010.06479.
- [21] H. Zhou, H.-F. Chai, and M.-L. Qiu, "Fraud detection within bank card enrollment on mobile device based payment using machine learning," *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 12, pp. 1537_1545, Dec. 2018, doi: 10.1631/FITEE.1800580.
- [22] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zineddine, "An experimental study with imbalanced classification approaches for credit card fraud detection," *IEEE Access*, vol. 7, pp. 93010_93022, 2019, doi: 10.1109/ACCESS.2019.2927266.
- [23] I. Matloob, S. A. Khan, and H. U. Rahman, "Sequence mining and prediction-based healthcare fraud detection methodology," *IEEE Access*, vol. 8, pp. 143256_143273, 2020, doi: 10.1109/ACCESS.2020.3013962.
- [24] I. Mekterović, M. Karan, D. Pintar, and L. Brkić, "Credit card fraud detection in card-not-present transactions: Where to invest?" *Appl. Sci.*, vol. 11, no. 15, p. 6766, Jul. 2021, doi: 10.3390/app11156766.
- [25] D. Molina, A. LaTorre, and F. Herrera, "SHADE with iterative local search for large-scale global optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2018, pp. 1_8, d