World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Logistic regression on banking fraud

Ethan Brooks * and Daniel Mercer

University of Florida USA.

World Journal of Advanced Engineering Technology and Sciences, 2022, 07(02), 334-348

Publication history: Received on 12 October 2022; revised on 22 November 2022; accepted on 24 November 2022

Article DOI: https://doi.org/10.30574/wjaets.2022.7.2.0132

Abstract

Bank fraud has been an increasing concern for banks as fraudsters use more advanced methods to take advantage of weaknesses in online transactions. Banks use machine learning algorithms to detect fraud, and logistic regression has been among the most popular methods used to detect fraud. This paper examines the use of logistic regression for the detection of banking fraud and its benefits, use, and limitations.

The article begins with a background of banking fraud, listing common types such as credit card fraud, identity fraud, loan fraud, and insider fraud. It then goes on to logistic regression, explaining why it is suitable for fraud detection and how it compares to other classification models. Data gathering, data preprocessing, and principal features that affect fraud classification are treated in the article.

Moreover, the paper discusses logistic regression model building and assessment using performance metrics such as accuracy, precision, recall, and F1-score. Some of the issues such as imbalanced data, false positives, and privacy concerns are taken into consideration, and ethical and legal concerns informing fraud detection systems are discussed. How banks optimize fraud detection by integrating logistic regression with cutting-edge methods such as deep learning and blockchain technology is also explored in the paper.

Finally, the paper discusses the future of banking fraud detection with an emphasis on AI innovation and emerging technologies that will shape the future of financial security. Through the adoption of machine learning and new fraud prevention strategies, financial institutions can mitigate fraud risks while providing a secure and seamless banking experience.

Keywords: Banking fraud detection; Logistic regression; Machine learning in finance; Fraud prevention strategies; AI in financial security

1. Introduction

1.1. Overview of Banking Fraud

Although the threat is always present in the financial industry, the prospect of banking fraud still has a long way to go. Hackers are becoming much smarter in their stealing techniques as a large number of financial corporations are moving to digital platforms. Sometimes, fraud occurs, credit card details can be used fraudulently, and others become the victim of identity theft, get caught in a phishing scam, insider fraud, or fraudulent loan applications. Such crimes lead to financial loss but also to the public loss of trust in banking systems. Banks are required to deploy ever more complex detection means in the fight against fraud, which becomes more elaborate, but the funds of their customers must not be endangered and the reliability of their financial proceedings be threatened.

* Corresponding author: Ethan Brooks.

1.2. Importance of Fraud Detection in Financial Institutions

Yes, it is correct that the cost of false positives is sometimes higher than the cost of false negatives, however, it should not deprive us from focusing on specific areas of the problem to minimize financial losses and to protect customers as well as to fulfill the regulatory requirements. Banks handle vast amounts of money daily, making them prime targets for fraudulent activities. If fraudulent transactions are not noticed by financial institutions, the results can mean substantial monetary loss for them, which can result in the instability and loss of profits of that particular segment of the industry. It is very important for the security of customer accounts for banking services to be trusted and confident. In this respect, customers take banks on faith to protect their financial data; therefore, any security violation may bring innumerable negative ramifications for losing your reputation in the long run.

What we know for sure is that fraud is strictly prohibited, and regulatory authorities mandate and impose strict laws so as not to attract fraud; as such, complacency can earn you heavy penalties, legal action, and even loss of operating license. These regulations are a must for banks to follow, but they must have a great degree of operational efficiency. Banks are assisted in this process by fraud detection mechanisms. The prevention of fraud is also important in beefing up security in financial ecosystems, other than for financial and legal purposes. They implement advanced fraud detection models to put banks into proactive action instead of being reactive toward risks.

1.3. The Role of Machine Learning in Fraud Detection

The modern changes in fraud schemes, which continue to evolve, usually cannot be detected effectively by traditional methods of fraud detection like rule-based systems. We have a powerful solution as machine learning uses data-driven models that can detect fraudulent activities with a very high accuracy. However, unlike manual systems, machine learning algorithms can process a large amount of data to recognize patterns, and find out anomalies in real time. This however allows us to train the models such that their output will be more and more prone to correctly predict faults in the new dataset while minimizing false positives and false negatives.

The time required for response is accelerated and the workload volume in the case of human analysts is reduced as well, so that financial institutions are capable of automating the fraud detection process as required. Logistic regression, decision trees and neural networks, and ensemble methods to name a few, are the methods used for fraud detection. Among these problems, logistic regression is one of the popularly used methods, which is very simple, interpretable, and compares well with these problems on binary classification problems like separating legitimate transactions from fraudulent transactions.

2. Understanding logistic regression

2.1. Definition and Purpose

Therefore, Logistic regression is a kind of statistical method used for binary classification, so it is very useful for fraud detection tasks. The difference between logistic regression and linear regression is the former predicts the probability of an event occurring (shown on 0 and 1) while the latter predicts continuous values. Logistic regression is used for transaction classification as fraud or not in the scenario of banking fraud.

Input features are passed to a sigmoid function to produce a probability between 0 and 1 in our model. The decision is that, if the probability is greater than a predetermined threshold, the transaction is categorized as fraudulent, otherwise it is legitimate. The simplicity of logistic regression makes it a go-to for the detection of fraud in situations where the interpretability of the model is essential.

This blog deals with two main questions related to why Logistic Regression is suitable (sometimes, the only solution) when it comes to fraud detection.

Since logistic regression is an efficient tool to solve classification problems, it is very fit for fraud detection. The method of fraud detection uses analyzing transaction patterns to identify if such transaction patterns indicate fraudulent activity. Because fraudulent transactions are usually much more infrequent compared to legitimate ones, logistic regression can be trained to detect such key risk factors and the probability estimation per transaction.

Logistic regression has the advantage of being interpretable. Logistic regression does not offer smooth insights as complex machine learning models do (i.e., neural networks), but it gives us clear idea on how different feature affects the prediction. Financial institutions require explainability for regulatory compliance and making decisions, this is

essential due to transparency in this industry. Moreover, logistic regression is computationally efficient and has good performance in the case of small to moderately large datasets, useful for practical fraud detection purposes.

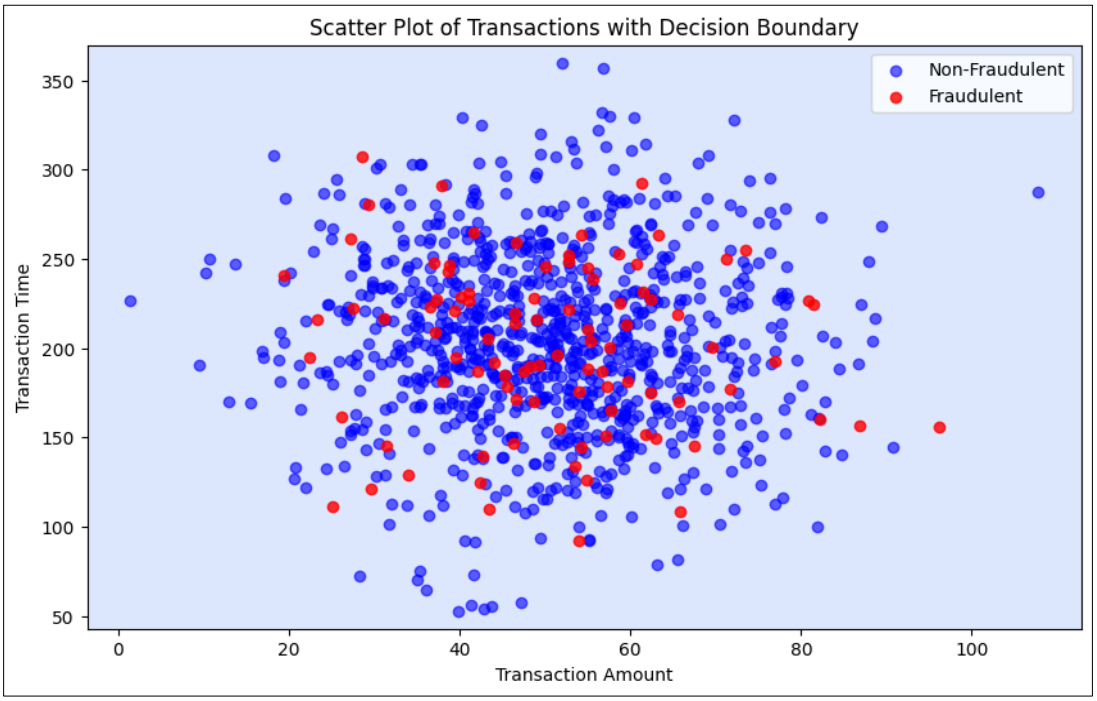

Figure 1 Logistic Regression Decision Boundary

2.2. Comparison with Other Classification Algorithms

It turns out that logistic regression as we know it is not the only classification algorithm for fraud detection although it is popularly used. However, even more complex models like decision trees, random forests, and neural networks provide higher predictive accuracy in some cases. Random Forest and gradient-boosting ensemble methods together with Decision Trees can model non-linear relations between the features and they are both often used to detect complex fraud patterns. With the ability to learn hierarchical representations from large datasets, neural networks, and especially deep learning models, become able to detect intricate fraud schemes.

However, those advanced models have a price. Considering that neural networks and ensemble methods are black box models and are computationally expensive, they cannot be used practically with the given data set. However, logistic regression is simple to interpret, easy to implement, and lower in computational power, thus, making for the practical choice of fraud detection in financial institutions. Logistic regression is often used as a benchmark, before using a more complex algorithm.

Affiliated with various feature engineering and anomaly detection techniques, logistic regression can perform great in fraud detection. This makes possible the probabilistic outputs that allow banks to fix different risk threshold levels according to their fraud tolerance. Logistic regression can be enhanced with other machine learning models to create a foolproof fraud detection mechanism for financial institutions that promotes accuracy, efficiency, and the ability to understand the reasons for an organization's choices.

3. Types of Banking Fraud

3.1. Credit Card Fraud

The most common type of banking fraud is credit card fraud in which someone makes unauthorized transactions using stolen card details. Phishing, skimming devices or data breaches let fraudsters get the card information and go online for purchasing or to ATM withdrawals or to sell it on the dark web. Fraudsters are reluctant to use cards, and there are some also whose only crime is to clone cards to which they do not have any authorization for making unauthorized transactions.

Banks use machine learning models to monitor transaction patterns reduce false positives, identify anomalies, and flag suspicious things. Some of the features one can consider are the server processed transactions' amount, frequency, location, and the device type that was used to perform the transactions to try and estimate if the transaction is a legitimate one or a fake. Multi-factor authentication as well as real-time detection mechanisms serve to lessen the likelihood of false transactions.

3.2. Identity Theft

Identity theft refers to the crime in which the fraudsters get hold of personal info like Social Security Numbers, bank account information, or login credentials to avail of financial services. This information is used by the criminals to open new bank accounts, apply for credit cards or loans in that person's name or engage in any unauthorized transactions of that person. The victims of identity theft usually incur considerable financial and reputational loss.

Thus, banks possess their fraud detection systems that take into account public behavior, login patterns and activities on an account to prevent identity theft. The features like log in at an unusual location, several unsuccessful authentication attempts or sudden huge transactions can be indicative of identity fraud. As for logistics regression, machine learning models are used to classify fraudulent versus legitimate user activities.

3.3. Loan Fraud

Loan fraud refers to the act of providing false information about someone to enable him or her to obtain loan that the person has no intention or capacity to repay. Fraudsters can submit fake income documents, stolen ID's or they may even create Synthetic IDs for the loans. However, some fraud schemes include organized groups that submit several fraudulent loan applications for bank systems exploitation.

Banks calculate credit histories, employment details, and patterns of financial transactions of a prospective borrower to ensure that he or she is a genuine applicant. The advanced fraud detection models look at the risk factors and flag any inconsistencies in the applications. The loan applicant is classified as a low risk or high risk based on specific characteristics like credit score, employment history, past financial transactions, and so on, using logistic regression.

3.4. Insider Fraud

Insider fraud refers to the circumstances when employees of a financial institution misuse access to transactions and funds or to divulge sensitive data to fraudsters outside of the company. Implementing methods to detect fraud is difficult since insiders who are authorized to have access to banking systems interact with these banking systems. Unauthorized account modifications, embezzlement of funds, and data breaches are some of the common insider fraud schemes.

Fraud detection models are used by financial institutions to watch the activity of employees, see unusual activity, and issue flags on odd transactions. A combination of behavioral analysis with access control mechanisms reduces the risk of such fraud. To pick out high-risk employees from employee logs, transaction approvals, and past fraud incidents, you could use logistic regression for example.

Table 1 Summary of Common Types of Banking Fraud

Type of Fraud	Description	Example
Credit Card Fraud	Unauthorized use of credit card information to make purchases or withdraw funds.	A fraudster steals card details to buy expensive electronics online.
Identity Theft	Illegally obtaining personal information to impersonate someone else.	Using stolen personal details to open a bank account in the victim's name.
Loan Fraud	Providing false information to obtain loans or credit.	Submitting fake documents to secure a mortgage.
Insider Fraud	Fraudulent activities conducted by employees within the organization.	An employee manipulates account balances for personal gain.

4. Data Collection and Preprocessing

4.1. Sources of Fraud Detection Data

Vast amounts of data are used in a fraud detection model from multiple sources. From transaction logs, customer profiles, account histories and other external fraud reports we can gain a lot of gudets to train machine learning models. In addition, banks work with the financial regulator and fraud prevention organizations to exchange fraud intelligence to make more accurate detection.

The data sources for fraud detection include online transaction, usage of credit cards, devices, IP addresses, geolocation data and user authentication logs. Machine learning models, by pull together several data points, can identify anomalies and related patterns with instances of fraud.

4.2. Missing values handling and Data Cleaning

Typically, this involves cleaning raw financial data and removing data inconsistencies, missing values, and duplicate records such that they can be used in training fraud detection models. Data cleaning includes removing the errors, standardizing the format and dealing with missing values in the data. Statistical methods, such as mean imputation, predictive modeling, or simply removing all the records in case there is a lack of data, can be used to solve the problem.

Outliers are also critical to handle in fraud detection because the fraudulent transactions usually vary greatly with the normal ones. To detect actual anomalies (and not fraudulent behavior), machine learning models need to be trained. Taking into account the above warning, logistic regression is much better if it is fed with clean and properly processed data, which becomes crucial step in fraud detection.

4.3. Feature Selection and Engineering

Improving the accuracy of the fraud detection model depends on feature selection. The dataset contains useful features that include transaction amount, transaction frequency, customer location, and payment method, which separates the legitimate transactions as compared to fraudulent transactions. Excess or unimportant features can either increase the computational complexity of the model and can have a negative impact on the performance of the model.

Feature engineering means crafting new features from existing data in order to increase accuracy of model. For example, one can calculate the average transaction amount over a certain time interval, identify sudden changes pattern of spending or to track the history of login location. The most important variables of interest for fraud classification are identified using feature selection techniques like Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA).

5. Logistic Regression for Banking Fraud Detection

5.1. How Logistic Regression Works

Fraud detection is a binary classification problem, and logistic regression is a machine learning algorithm that can be used and is widely implemented. It estimates the probability that an event will happen and then uses this result to determine either a classification label above some threshold or a probability. Logistic Regression predicts if a given transaction is fraud (1) or not (0) Fraud Detection.

To accomplish this, the model maps input features to a probability score, i.e. 0 to 1, using a logistic (sigmoid) function. The transaction is classified as a fraudulent one if the predicted probability is greater than a predefined threshold (let us say 0.5), and a legitimate one otherwise. Using the historical transactional data along with the fraud cases labeled, a logistic regression model is trained to learn the patterns associated with the fraudulent behavior.

5.2. Implement a Logistic Regression Algorithm in Fraud Detection

The following steps are followed to implement logistic regression for fraud detection

- Data Collection – Gather transaction records, user profiles, and fraud labels.
- Data Preprocessing includes cleaning the data, handling the missing values, and normalizing the features.
- Feature Selection – selecting the important variable that can impact fraud detection.
- Secondly, Model Training is to train the logistic regression model with historical fraud data.

- Prediction and Classification – Classify new transactions as fraud or not fraud using the trained model.
- Model Evaluation – Evaluate the performance of the model with accuracy, precision, recall, and F1 score.

Other techniques, such as anomaly detection and ensemble learning, are often used with logistic regression to increase the performance of fraud detection. Logistic regression can be used as a baseline model by advanced fraud detection systems, which base further enhancement either on deep learning or an ensemble approach.

5.3. The Key Metrics Used in Model Evaluation

The performance of a fraud detection model must be evaluated to check whether it is effective. Logistic regression can be used by having common metrics:

- Accuracy: It measures the proportion of correctly classified transactions. Nevertheless, just get accuracy can be deceptive because in fraud detection problems, there is an issue of class imbalance (i.e there are much fewer fraudulent transactions compared with normal transactions).
- Precision (also called positive predictive value): Tells how many of the transactions that are classified as being fraudulent truly are. High precision helps in reducing false positives and therefore reduces the impact on the legitimate users.
- Sensitivity or Recall: The Proportion of actual fraud cases correctly identified by the model. Fewer fraud cases are missed due to high recall.
- F1-Score: A balanced metric that is a single performance measure with precision and recall. It is very useful for imbalanced datasets.
- AUC-ROC (Area Under the Curve – Receiver Operating Characteristic) – where the model's performance to discriminate between fraudulent and non-fraudulent transactions is projected over different probability thresholds. Better Model Performance is indicated by a higher AUC score.

Optimizing these metrics will help banks to improve the fraud detection accuracy, reduce false alerts, and provide a smooth customer experience. In spite of the fact, it is an old technique (automated machine learning not yet so old), logistic regression still represents a useful tool in fraud detection, as it is quite simple, interpretable and efficient in financial security applications.

6. Model Training and Performance Evaluation

6.1. The Dataset is divided into Training and Testing.

But, before training a logistic regression model for fraud detection, it is important to be divide the dataset into two parts – training and testing. Often a dataset is divided into about 70–80% for training set and 20–30% for the testing set. The training set is used for making the model learn patterns from historical data transactions and testing set is used to evaluate the model's capability of the model in generalizing to new datasets.

However, k-fold cross-validation techniques can be used to improve the model reliability. With this, different subsets of the data tend to be used for training and evaluation, and hence, the performance of the model is more balanced. When it comes to fraud detection, the class distributions of the numbers of fraudulent and nonfraudulent transactions in the two sets must be kept the same or they can bias model learning.

6.2. Performance Metrics: Accuracy, Precision, Recall, F1-Score

To evaluate the effectiveness of a logistic regression model, multiple performance metrics are needed since accuracy alone is not accurate enough for fraud detection, as classes might be highly unbalanced. Accuracy indicates the percentage of correct predictions, however, in a highly imbalanced dataset, a model can obtain high accuracy only by classifying all the transactions as genuine.

The precision value has great relevance on fraud detection as the higher the precision value, the less flagged fraudulent transactions are fraudulent. A more precise model means that it classifies fewer legitimate transactions as fraud, and thus a smaller number of customers are hindered unnecessarily. Another key metric is recall which indicates the proportion of actual fraudulent cases that have been correctly identified by the model. The score for recall is high enough to minimize the number of undetected fraudulent transactions. F1-score is a measure of precision and recall, and it is more appropriate when there is a class imbalance. The best fraud detection model should have a high recall without compromising its precision so that it will catch fraudulent transactions without raising false alerts.

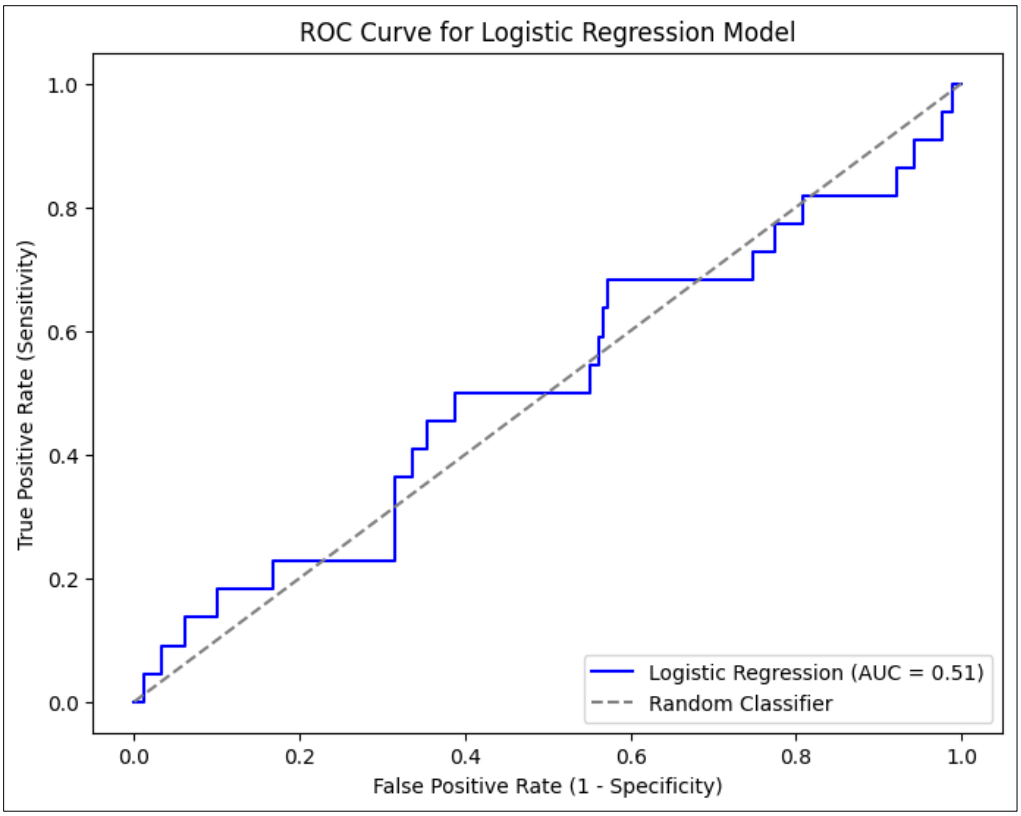

Figure 2 ROC Curve for Logic Regression Model

6.3. Handling Class Imbalance in Fraud Detection

The nature of the problem and class imbalance difficulty makes fraud detection a challenging one since only small portion of all banking transactions is the fraudulent one. If not handled properly, a model starts to predict that all transactions are legitimate and thus we have a bad fraud detection rate.

A way to tackle this issue is to carry out resampling of the dataset by sampling the minority class or reducing the majority class. Oversampling works by duplicating fraudulent transactions or in a way creating synthetic fraud transactions & SMOTE; undersampling, on the other hand, removes some of the legitimate transactions to make the dataset more balanced. One way is to change the class weight within the logistic regression algorithm by giving more importance to fraudulent transactions to make more effort in finding them.

In addition, the techniques of anomaly detection can be incorporated to treat the fraud as an outlier detection problem. This approach finds those transactions that certainly retreat from normal behavior and helps to uncover fraud tactics never seen before. Class imbalance in financial datasets can be handled properly by financial institutions, thus improving fraud detection accuracy while maintaining a good user experience for their customers.

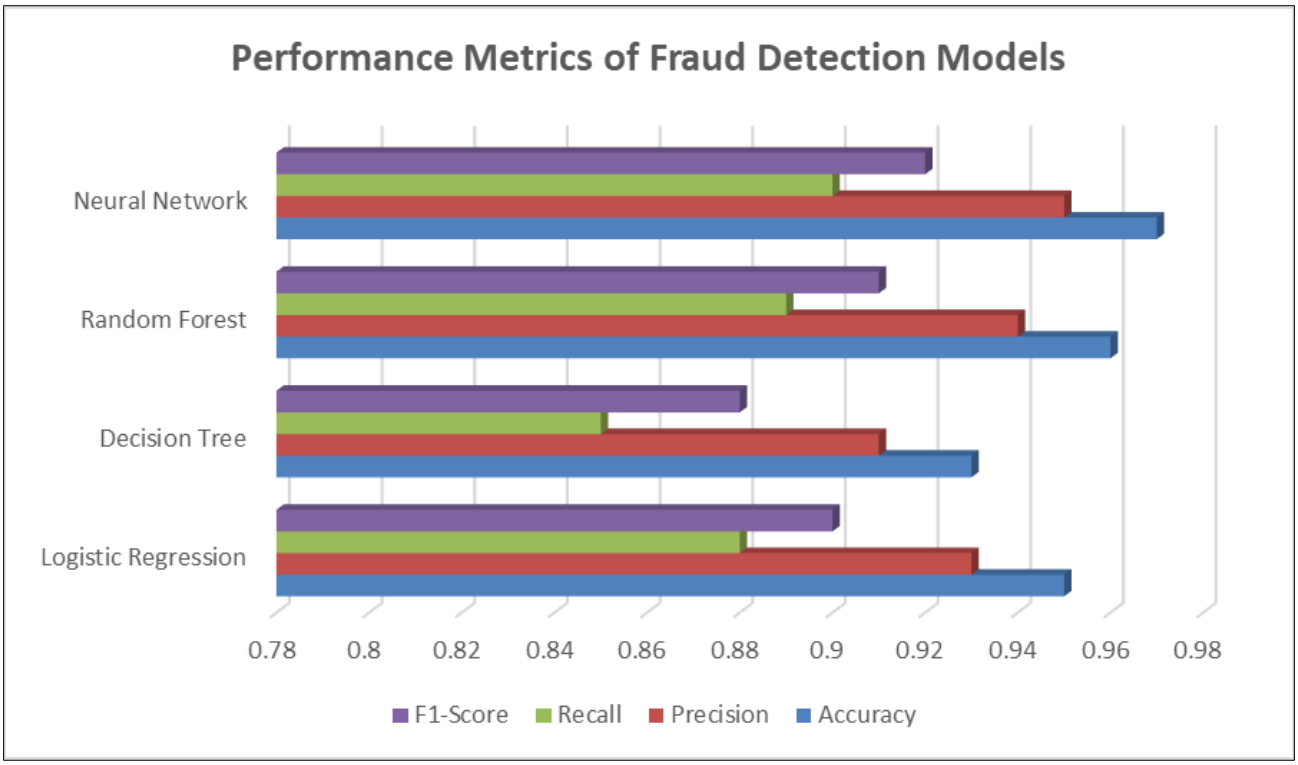

Figure 3 Performance Metrics of Fraud Detection Models

7. Feature Importance in Fraud Detection

7.1. Key Features Influencing Fraud Classification

To improve the performance of the fraud detection model, feature selection is critical. Features that are some of the most influential in fraud classification are transaction amount, transaction frequency, transaction time, location discrepancies, and changes in device or IP address. The most common indicators of possible fraud are large, unexpected transactions or transactions happening in a short time and at unusual hours. Additionally, purchases made from a country that the buyer has never visited before can also raise a red flag for a sudden change in transaction location. They are also advised if any new device or unknown IP address accesses the account to initiate a transaction such might suggest unauthorized access to the account.

Another important factor is the behavioral history of an account. A change in customer spending pattern can lead to classification of the transaction as fraud, such as the high value purchase when the spending pattern shows the typically low value spending. This allows logistic regression models to identify key features on which they can successfully separate between legitimate and fraudulent transactions.

7.2. Methods for Feature Importance Analysis

Different feature importance analysis techniques can be used to find out the most critical features in fraud detection. The feature that contributes the most to final prediction resides in the coefficient provided by logistic regression. A greater coefficient value increases influence on fraud classification. Nevertheless, raw coefficients may not explain what is going on as there are correlations between features.

The other option is Recursive Feature Elimination (RFE) which gets rid of the least important features one by one, retraining the model until the important ones remain. This way helps to improve the model efficiency since it reduces the noise in irrelevant data. Another widely used technique is Principal Component Analysis (PCA) which transforms correlated features into a smaller set of uncorrelated variables which are the most important and reduce dimensionality.

Moreover, in the feature importance analysis, the model performance improves and it also increases the interpretability. Having a mastery of why a transaction is classified as fraudulent is important to gain customer trust and for legal

requirements as in the case of banking fraud detection, where regulatory compliance demands transparency in decision making.

8. Comparison with Other Machine Learning Algorithms

8.1. Decision Trees

Another good choice for fraud detection also is decision trees, as they may capture complex decision boundaries. Whereas logistic regression assumes a linear relationship between the features and the target variable, decision trees can model non-linear relationships as a team of its leaves are made by recursively splitting data based on feature value in data. As a result, they are useful to spot out sophisticated fraudulent transaction patterns. However, decision trees suffer from overfitting the training data, whereby the representation performance on the training data is generally decreased.

8.2. Random Forests

The overfitting issue is addressed by averaging predictions from bunch of different decision trees; Random forests, the ensemble learning method based on multiple decision trees. It makes the fraud detection more accurate and robust. The random forest produces results by majority voting from multiple models while logistic regression outputs anyway in probabilities. Although better predictive performance is gained by random forests, they are computationally more intensive and less interpretable than logistic regression.

8.3. Neural Networks

Fraud detection has gotten interested in neural networks because of their potential to serve and learn very complicated, high-dimensional patterns in very large data sets. Because deep learning models can automatically extract features from the raw data, they tend to be extremely powerful at detecting schemes that are sophisticated enough to escape the attention of humans. Nevertheless, due to the high level of computational resources and requirement of large amounts of labeled data for training neural networks, they are considered unfit for computationally restrictive and low-labeled data application domains such as IoT. They are also not interpretable, so a financial institution cannot explain why a specific transaction was categorized as fraudulent. However, this limitation does not help regulatory compliance and customer trust.

8.4. Pros and Cons of Logistic Regression versus Other Algorithms

Despite being simple, efficient and interpretable, logistic regression is a strong contender for fraud detection. While decision trees and neural networks can be quite complex, logistic regression gives clear probability estimates which financial institutions can then use to set their fraud risk tolerance. It is computationally efficient as well, which would make it practical to deploy in real time for fraud detection purposes.

However, one of the problems with such models is that logistic regression assumes a linear relationship between features and fraud probability. On the contrary, models such as random forests and neural networks can formulate intricate relationships at the expense of being complicated and costly computationally. Ultimately, the selection of the algorithm is based on the needs of the financial institution. When interpretability and efficiency are important, logistic regression is a good practical choice. In such cases, either ensemble methods or deep learning models are more suited for more advanced fraud detection with larger data sets and complex patterns.

From learning the strengths and constraints of each algorithm we must use; financial institutions can develop hybrid fraud detection processes using numerous models.

9. Challenges in Banking Fraud Detection

9.1. Imbalanced Datasets

Dealing with imbalanced datasets is considered to be one of the biggest challenges in banking fraud detection. In most cases, when we talk about transactional data, fraudulent transactions make up only a small part of total transactions and in some cases, even below 1% of the whole dataset. An imbalance case can distort machine learning models (it can happen, for instance, with logistic regression), making them predictable to believe all transactions are legit. Simply put a model trained on imbalanced data with high accuracy maybe because it is just predicting every transaction as non–fraudulent, which in turn gives a poor recall score and thus many fraudulent cases undetected.

To tackle this challenge, we can employ oversampling on the fraudulent transactions, undersampling on the legitimate transactions, or using some synthetic data generation techniques such as SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset. Furthermore, class weights are adjusted in the logistic regression algorithm to make sure that fraudulent transactions are weighted more important during this algorithm's training.

9.2. False Positives and False Negatives

One more dimension of another significant challenge in fraud detection is the false positives and false negatives. In a false positive situation, a genuine transaction is labeled as fraud. As a result, customers experience inconvenience, transactions may be declined and the bank's reputation compromised. However, a false negative will get the fraudulent transaction mistaken for a genuine one, and this would be a financial loss in customers as well as institutions.

The fraud detection models need to balance precision and recall, to minimize false positives, and at the same time not miss fraudulent transactions. Finding this balance is possible by setting an appropriate decision threshold based on the institution's risk tolerance. Moreover, it can also benefit by combining logistic regression with other advanced techniques (e.g. anomaly detection or deep learning) to increase its accuracy in fraud detection.

9.3. Data Privacy Concerns

The fraud detection models are designed to analyze the indelibly large amounts of sensitive financial data such as transaction history, location specifics, device information, and individual identification details. Banks and financial institutions have a huge challenge guaranteeing data privacy and complying with regulations such as GDPR and PCI DSS.

Institutions implement encryption, anonymization, and secure data storage techniques to address privacy concerns and protect customers' data. An emerging approach called federated learning enables the training of machine learning models across multiple institutions with privacy maintained and allows fraud detection to be improved. Fraud detection models have to adhere to strict data protection policies to ensure that customers trust the company, and also to avoid legal problems.

10. Real-world Applications and Case Studies

10.1. Success Stories of Logistic Regression in Fraud Detection

Logistic regression can be implemented by many financial institutions and has been very successful in detecting fraudulent activities. Logistic regression is used as a first fraud detection tool for example by major banks for its efficiency, interpretability and its ability to compute probability-based risk assessments. By analyzing transaction amount, when transaction occurred, user location and user behavior patterns, logistic regression models can detect any suspicious activities that are happening in real time.

In this particular real-world case, a bank worldwide implemented a logistic regression model to identify credit card fraud. The final model was trained on historical transaction data and the fraud probability score was generated for each new transaction. This was achieved by defining a risk threshold, thereby reducing fraudulent activities to a great extent whilst minimizing the disruptions experienced by legitimate customers. The logistic regression model became more adaptable to new fraud patterning and better at detecting patterns that potentially deem transactions to be fraud than traditional rule-based systems.

10.2. How Banks are integrating Machine Learning Models to enhance their capabilities.

Logistic regression as a machine learning model is being used by banks and financial institutions, growingly, to strengthen their fraud detection powers. An architecture of real-time transaction monitoring system that integrates with these models is defined and where each transaction is scored with fraud risk based on historical data. Legitimate transactions continue without delay, whereas high risk transactions are passed on the further review.

Ensemble approaches, where logistic regression is combined with others more advanced models constituting decision trees, random forests, and deep learning networks are also used by the financial institutions to increase the accuracy of fraud detection. Most of the banks commonly use cloud hosted fraud detection platform to allow real time analysis over enormous amount of transaction data to immediately detect suspicious activities. Besides, machine learning models are continually re-educated with new fraud data to anticipate emerging frauds.

Leveraging machine learning, banks not only raise their fraud detection, but also help create more enhanced customer experience by reducing false positives and thus minimize unnecessary transaction declines.

11. Improving the Accuracy of Fraud by Applying New Techniques

Logistic regression can also be combined with other techniques, even if such use may not be immediately evident.

Logistic regression is a very powerful fraud detection tool, with the clear potential of being combined with other techniques. Logistic regressions are integrated with anomaly detection (e.g. bottom and up lift models), rule-based systems and other machine learning algorithms to form the hybrid fraud detection models that would produce better accuracy and fewer false positives.

Logistic regression can be used as a screening model (an initial model) where high risk transactions are found and passed on to a second, more complex model, such as a decision tree and a neural network. Such an approach makes it easy to detect sophisticated fraud schemes that might be difficult for one algorithm to detect.

An ensemble learning approach is also very efficient which involves multiple models being combined in order to make a final prediction. Bagging and boosting are techniques that use the strengths of multiple different models to improve overall fraud detection performance. The hybrid approach, these, enable financial institutions to get a more accurate and adaptable fraud detection.

In recent years, deep learning has become one of the powerful tools in the fight against fraud by being able to identify complex patterns in large datasets. To analyze the sequence of transactions and detect anomalies in user behaviors, we use the neural networks specialized to work at the sequences of data, both recurrent neural networks (RNNs) and convolutional neural networks (CNNs).

Deep learning model takes transaction data in raw form instead of manually selected features as is the case in logistic regression. This makes it possible for them to find minute fraud patterns which are out of observation of normal tactics. The deep learning models, together with the logistic regression, are used by the financial institutions to complement fraud detection.

Although brings many advantages, it presents challenges, for instance, extremely high computational requirements and little interpretability. Unlike logistic regression, deep learning models work as black boxes that provide probability scores for each transaction, it is hard to answer the question of why a transaction was flagged as fraud. Deep learning models are said to be lacking transparency and interpretability; to overcome this limitation of deep learning models, explainable AI (XAI) techniques are introduced to make deep learning models more transparent and interpretable.

The banks can use deep learning combined with logistic regression and other advanced techniques to develop better fraud detection systems that will respond to emerging threats but also be economical, accurate, and transparent.

While fraud detection remains a pivotal aspect of banking security, logistic regression remains one of the basic models for identifying cases of fraudulent transactions. However, since such fraud techniques keep evolving, financial institutions must move to hybrid approaches that involve logistic regression and more sophisticated forms of machine learning to be a step ahead of the cybercriminals. Banks can take advantage of real-time analytics, secure data practices, and AI-based fraud detection systems in enhancing security as well as giving seamless experience to customers.

12. Ethical and Legal Considerations

12.1. Regulations Governing Fraud Detection Models

Fraud detection models implemented by financial institutions must adhere to strict regulations so that the processes are transparent and accountable, and aim at consumer protection or least harm to consumers. There are various global regulations associated with how banks can use artificial intelligence (AI) and machine learning (ML) to detect fraud. The General Data Protection Regulation (GDPR) in the European Union dictates that customer data is to be handled with more responsibility and more specific regulations when it comes to data privacy, consent, and user rights. So too with the Payment Card Industry Data Security Standard (PCI DSS) which enforces rigorous security requirements for dealing with credit card transactions, so that the rules around fraud detection models do not expose sensitive financial information.

Financial institution's fraud detection practices are regulated in the US by the Fair Credit Reporting Act (FCRA) and the Bank Secrecy Act (BSA) which compels them to report suspicious activities while protecting consumers' rights. Further, real-time monitoring of transactions is mandated by anti-money laundering (AML) laws from banks to prevent money laundering activities. Accordingly, in the UK, regulatory agencies like the Financial Conduct Authority (FCA) are responsible for the ethical use of AI-driven fraud detection models, while in the US, this function is provided by the Office of the Comptroller of the Currency (OCC).

Fake detection systems are growing more robust and regulators keep developing the rules to ensure that the AI and ML models operate equally and don't discriminate any customer groups based on any bias. To comply with such evolving regulations, financial institutions must, based on regular auditing of their fraud detection systems, adjust the systems.

12.2. Ethical Concerns in Using AI for Fraud Detection

It is true that while AI and machine learning have managed to improve fraud detection, they also bring forward the ethical questions companies need to answer. Fraud detection models have one of the biggest problems in terms of bias. If biased data is used to train an AI model, it may end up unfairly singling out targets to different demographics, thereby discriminating financially against them. For instance, if historical fraud data has identified that transactions coming from particular regions and profiles have been frauded more often than others, the model can develop biases leading to higher false positive rates for these groups. Unwanted blocking of legitimate transactions is caused, generating inconvenience for the customer and reputation damage to the FI."

The second ethical issue is that AI fraud detection models usually operate with limited transparency. Advanced models such as Deep Learning models and many others are 'black boxes', it is not easy to explain when why a transaction was flagged as Fraudulent. If the bank is unable to give a clear explanation the customer wrongly accused of fraud may find it difficult to contest the decision. This is a problem that can be solved only if institutions install explainable AI (XAI) techniques and make fraud detection decisions interpretable and justifiable.

Moreover, the AI use in fraud detection poses a concern for mass surveillance and customer privacy. Independent of that point, it is appropriate that monitoring transactions is crucial to detect fraud, unfettered use of which could impede upon customers' rights. Ethical AI policies are something upon which financial institutions should focus to avoid misuse of customer data. To be precise, the use of ethical AI helps in maintaining a balance between security and privacy.

Fairness, transparency, and accountability of AI driven fraud detection helps to create customer trust and comply with ethical and legal parameters related to it.

13. Future of Banking Fraud Detection

13.1. AI and ML Advancements

Continuous development in artificial intelligence and machine learning is going to be the key driver of fraud detection in banking in the future. With every passing day's AI driven fraud detection models become smarter and hence have the power of analyzing very large amount of transactions data in real time and along with this the subtlest of patterns for fraud detection. With the improving deep learning, fraud detection systems can now detect complex associations between transactions to deliver a better fraud detection on the emerging fraud techniques.

The biggest advancement is the interjection of reinforcement learning which means that fraud detection models keep learning new fraud patterns from real-time feedback about what was, and is correctly, flagged. Instead of being purely dependent on historical data, these models learn dynamically and get better accuracy as more and more fraud tactics come into play. Fraud is additionally being increasingly discovered proactively as you compare it to old-fashioned methods with AI-driven fraud detection that swiftly picks out suspicious activities and prevents fraud from being carried out. Banks can assess a customer's risk profile and even know, detect advance, and flag potential fraud before a transaction takes place.

The other important development is the use of Natural Language Processing (NLP) in fraud detection. With NLP–powered AI models, it is possible to use the voice of customers (complaints, chat, social mentions) and analyze unstructured information to track fraudulent activities before they blow up. With the help of AI-driven behavioral analytics, financial institutions can spot anomalies in customer behaviors, which is one extra defending approach to prevent fraud.

13.2. The Role of Blockchain in Fraud Prevention

Blockchain is becoming a serious fraud prevention tool. Being decentralized, and immutable, blockchain offers transparency to transactions ensuring there are no tampered or unauthorized changes in transactions. Traditional banking systems involve acts of sending a transaction by one party through various intermediaries, who have ample opportunities to manipulate financial data without the trader's knowledge and a blockchain eliminates this opportunity by creating a direct and secure way for transactions to take place.

Smart contracts, one of the most promising applications of blockchain, are also the most promising application of blockchain in fraud prevention. Smart contracts run automatically while the predefined conditions are met with no requirement of human intervention. Such a thing would cut down the risk of malware executing fraudulent activities such as identifying unauthorized fund transfers or identity theft.

Decentralized identity management is also a part of blockchain that helps in identity verification. Blockchain eliminates the need to trust in undermined centralized databases when it comes to digital identities, instead, providing a secure and verifiable digital identity. A blockchain based system allows customers to store identity information instead of the government in order to reduce the risk of identity theft and fraud.

Combining blockchain with AI based fraud detection models helps financial institutions to raise the bar of security of the banking ecosystem. Blockchain eradicates transparency and security, while AI ensures that all of this can take place intelligently thereby helping in preventing financial crimes.

14. Conclusion

Modern banking is fraught with the risk of banking, and fraud detection plays a crucial role in providing the much-needed financial security and protection of customers from fraudulent financial transactions. However, logistic regression is still a cornerstone of fraud detection because it is effective, easy to use, and easy to interpret. However, with the tendency to use different fraud tactics by criminals, financial institutions shall seek to employ more expensive machine learning models which include combining logistic regression, deep learning, anomaly detection, and blockchain technology to improve fraud prevention.

However, fraud detection is one of the biggest challenges, which includes imbalanced datasets, false positives, and ensuring the privacy of users. Techniques like resampling, anomaly detection, and ethical AI frameworks are to be used by financial institutions to avert biases and for further fairness in fraud detection. Compliance is also very important as the bank has to adhere to a strict policy that helps protect customer data from falling into the wrong hands and stops financial discrimination.

Better still, the future of fraud detection has some help from AI advancements, real-time analytics, and blockchain technology, among others. The AI-powered fraud detection models will become increasingly more accurate in identifying new fraud patterns and more accurately assess the risk of customers. About blockchain, technology is also going to serve a crucial role in enabling the control of transaction transparency, and security and eliminating risks of fraud related to traditional banking systems.

Banks have to keep on updating their fraud detection strategies, they have to invest in AI-driven solutions and be even while keeping them secure and at the same time, making it convenient for customers. The solution for fraud detection is not once and done, but continues onward as continually evolving, ethical, and compliant. The use of cutting-edge technology and proactive fraud prevention measures that financial institutions can employ will cultivate a safer and more trustworthy banking environment for their customers.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Kumar, Y., Saini, S., & Payal, R. (2020). Comparative Analysis for Fraud Detection Using Logistic Regression, Random Forest and Support Vector Machine. SSRN Electronic Journal.

[2] Höppner, S., Baesens, B., Verbeke, W., & Verdonck, T. (2020). Instance-Dependent Cost-Sensitive Learning for Detecting Transfer Fraud. arXiv preprint arXiv:2005.02488.

[3] Niu, X., Wang, L., & Yang, X. (2019). A Comparison Study of Credit Card Fraud Detection: Supervised versus Unsupervised. arXiv preprint arXiv:1904.10604.

[4] Bhat, N. (2019). Fraud detection: Feature selection-over sampling. Kaggle. Retrieved from https://www.kaggle.com/code/nareshbhat/fraud-detection-feature-selection-over-sampling

[5] InsiderFinance Wire. (2021). Logistic regression: A simple powerhouse in fraud detection. Medium. Retrieved from https://wire.insiderfinance.io/logistic-regression-a-simple-powerhouse-in-fraud-detection-15ab984b2102

[6] Olaitan, V. O. (2020). Feature-based selection technique for credit card fraud detection. Master's Thesis, National College of Ireland. Retrieved from https://norma.ncirl.ie/5122/1/olaitanvictoriaolanlokun.pdf

[7] Raymaekers, J., Verbeke, W., & Verdonck, T. (2021). Weight-of-evidence 2.0 with shrinkage and spline-binning. arXiv preprint arXiv:2101.01494. Retrieved from https://arxiv.org/abs/2101.01494

[8] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2017). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643

[9] Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., Mazzer, Y., & Bontempi, G. (2019). Scarff: A scalable framework for streaming credit card fraud detection with spark. Information Fusion, 41, 182–194. https://doi.org/10.1016/j.inffus.2017.09.005

[10] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47–66. https://doi.org/10.1016/j.cose.2015.09.005

[11] Jain, M., & Shah, A. (2022). Comparison of machine learning models for stress detection from sensor data using long short-term memory (LSTM) networks and convolutional neural networks (CNNs). International Journal of Scientific Research and Management (IJSRM), 12(12), 1775-1792.

[12] Zareapoor, M., & Shamsolmoali, P. (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Computer Science, 48, 679–685. https://doi.org/10.1016/j.procs.2015.04.201

[13] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613. https://doi.org/10.1016/j.dss.2010.08.008

[14] Patel, H., & Zaveri, M. (2011). Credit card fraud detection using neural network. International Journal of Innovative Research in Computer and Communication Engineering, 1(2), 1–6. https://www.ijircce.com/upload/2011/october/1_Credit.pdf

[15] Duman, E., & Ozcelik, M. H. (2011). Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications, 38(10), 13057–13063. https://doi.org/10.1016/j.eswa.2011.04.102

[16] Jain, M., & Shah, A. (2022, February 28). Machine learning with convolutional neural networks (CNNs) in seismology for earthquake prediction. IRE Journals, 5(8).

[17] Jain, M., & Srihari, A. (2021). Comparison of CAD detection of mammogram with SVM and CNN. IRE Journals, 8(6), 63-75.

[18] Kaushik, P., Jain, M., & Jain, A. (2018). A pixel-based digital medical images protection using genetic algorithm. International Journal of Electronics and Communication Engineering, 31-37.

[19] Kaushik, P., Jain, M., & Shah, A. (2018). A Low Power Low Voltage CMOS Based Operational Transconductance Amplifier for Biomedical Application.

[20] Jain, M., & Shah, A. (2022). Machine Learning with Convolutional Neural Networks (CNNs) in Seismology for Earthquake Prediction. Iconic Research and Engineering Journals, 5(8), 389–398. https://www.irejournals.com/paper-details/1707057

[21] Kaushik, P., & Jain, M. (2018). Design of low power CMOS low pass filter for biomedical application. International Journal of Electrical Engineering & Technology (IJEET), 9(5).