CREDIT CARD FRAUD DETECTION

CS19643 – FOUNDATIONS OF MACHINE LEARNING

Submitted by

KEERTHANA S (2116220701124)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2025

BONAFIDE CERTIFICATE

Certified that this Project titled **“CERDIT CARD FRAUD DETECTION”** is the bonafide work of **“KEERTHANA S (2116220701124)”** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. V.Auxilia Osvin Nancy.,M.Tech.,Ph.D.,
SUPERVISOR,
Assistant Professor
Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai-602 105.

Submitted to Mini Project Viva-Voce Examination held on ______________________

Internal Examiner **External Examiner**

ABSTRACT

Credit card fraud continues to pose a major threat to financial systems worldwide, resulting in significant economic losses and security concerns. The need for intelligent, real-time fraud detection systems has become increasingly important as transaction volumes grow and fraudulent schemes become more sophisticated. This project presents a machine learning-based solution for detecting credit card fraud using various supervised classification algorithms, focusing on performance accuracy, adaptability, and data robustness.

The primary objective of this study is to build a predictive model that effectively distinguishes between legitimate and fraudulent transactions using real-world data. The dataset employed includes anonymized features that represent transaction behavior, with a heavy class imbalance—where fraudulent cases are rare compared to legitimate ones. To address this challenge, we implemented data preprocessing techniques including normalization, class balancing, and feature selection. Four key classification algorithms were evaluated: **Logistic Regression**, **Random Forest**, **Support Vector Machine (SVM)**, and **XGBoost**.

Performance metrics such as **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, and **R^2 Score**, along with classification metrics like **Precision**, **Recall**, and **F1-score**, were used to assess model effectiveness. Among the models tested, **SVM** and **XGBoost** outperformed others in terms of precision and recall, with SVM achieving an R^2 score of **0.86**, indicating strong predictive performance in fraud detection scenarios. To further enhance model robustness, **Gaussian noise-based data augmentation** was applied. This technique simulated real-world variations in transaction data, improving generalizability and significantly boosting the performance of tree-based models like Random Forest and XGBoost.

The experimental findings demonstrate that well-tuned machine learning models, supported by thoughtful preprocessing and augmentation strategies, can provide a scalable and effective solution for credit card fraud detection. Future enhancements could involve deploying these models within financial APIs or mobile banking apps for real-time fraud prevention and customer alerts.

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.,** our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide & our Project Coordinator **Dr. V. AUXILIA OSVIN NANCY.,M.Tech.,Ph.D.,** Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

KEERTHANA S - 2116220701124

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

1.INTRODUCTION

In recent years, the dramatic rise in digital payment systems and e-commerce platforms has transformed the way individuals and businesses manage financial transactions. With this digital shift, credit cards have become an essential component of financial convenience and flexibility. However, the increased usage of online financial services has also made them a prime target for fraudulent activities. **Credit card fraud** remains a major concern for financial institutions and consumers alike, leading to billions of dollars in losses annually and significantly undermining consumer trust in digital platforms.

Credit card fraud is not merely an economic issue—it is also a technological and security challenge. As fraudsters continuously develop more sophisticated tactics to evade traditional detection methods, conventional rule-based systems have begun to show their limitations. These systems often rely on predefined patterns and thresholds, making them rigid and inadequate in responding to new and evolving fraud strategies. Furthermore, they generate a high number of false positives, inconveniencing legitimate users and reducing overall system efficiency.

With the proliferation of large-scale transaction data and advancements in computational techniques, **machine learning (ML)** has emerged as a powerful tool to combat this growing issue. ML models can automatically learn patterns of normal and fraudulent behavior by analyzing historical transaction data. By adapting over time and identifying subtle correlations in high-dimensional data, these models offer a promising solution for real-time fraud detection and prevention. This project focuses on utilizing supervised learning techniques to classify credit card transactions as fraudulent or legitimate, thereby building a predictive system that enhances both security and user experience.

The growing prevalence of cybercrime, particularly in the financial domain, has made fraud detection a priority for organizations and regulators. As more individuals turn to online transactions and cashless payments, the risks associated with cyber fraud increase exponentially. According to reports by the Federal Trade Commission (FTC) and other financial watchdogs, millions of fraudulent transactions are reported each year, often resulting in substantial financial losses, reputational damage, and emotional stress for victims.

Traditional fraud detection systems are often reactive and require manual updates to handle new fraud patterns. They typically work well only under specific, known conditions and tend to lack flexibility. In contrast, machine learning models provide a dynamic and adaptive approach, learning from historical transaction data and improving over time. This capability is especially important for credit card fraud detection, where new fraud strategies appear frequently and vary across demographics and geographies.

Another significant challenge in credit card fraud detection is the **class imbalance** in the dataset—fraudulent transactions represent a small fraction of all transactions, making it difficult for many algorithms to detect them accurately. Addressing this imbalance requires advanced evaluation metrics and model tuning techniques to ensure that models do not favor the majority class. This project aims to tackle these challenges by experimenting with a variety of supervised learning models, applying resampling techniques, and using performance metrics such as precision, recall, F1-score, and AUC-ROC to gain deeper insights into model effectiveness.

The motivation behind this project is twofold: to develop a more **accurate and responsive system** for fraud detection, and to explore how **different machine learning models perform under real-world data constraints**. With a focus on predictive analytics and practical application, this project contributes to building safer digital environments and helping financial institutions make informed decisions in real time.

The primary objective of this research is to develop and evaluate a machine learning-based system that can accurately detect credit card fraud based on transaction-level data. The system, referred to as the **Credit Card Fraud Detection Engine**, uses various supervised learning algorithms to classify whether a transaction is legitimate or fraudulent. The project is built using Python and executed on Google Colab, leveraging tools such as pandas, scikit-learn, matplotlib, and seaborn for data preprocessing, model training, and result visualization.

The dataset used in this project is a publicly available, anonymized credit card transaction dataset that contains features resulting from a PCA (Principal Component Analysis) transformation to ensure data confidentiality. The dataset includes features such as transaction amount, time, and 28 anonymized numerical attributes, alongside a binary classification label indicating fraud.
Four major machine learning models—**Logistic Regression, Support Vector Machines (SVM),**

Random Forest, and XGBoost—were trained and compared in this study. Each model was evaluated based on multiple criteria, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). Due to the imbalanced nature of the dataset, techniques such as SMOTE (Synthetic Minority Oversampling Technique) and under-sampling were explored to balance the training data and improve classification performance.

The ultimate aim of this work is not only to find the most suitable model but also to build a system that could be **integrated into a real-time fraud monitoring framework**, such as those used in banks or e-commerce platforms. The research also considers the **user-centric aspect** of fraud detection, where legitimate transactions should not be falsely flagged, preserving customer convenience while maintaining high fraud detection rates.

This research focuses on detecting credit card fraud using supervised machine learning models. It utilizes a labeled dataset containing transaction details to train classifiers like Logistic Regression, Random Forest, XGBoost, and Support Vector Machine. The models are evaluated based on accuracy, precision, recall, and F1-score. The goal is to develop a reliable, non-intrusive fraud detection system that can be integrated into real-time payment platforms.

CHAPTER 2

2.LITERATURE SURVEY

Credit card fraud detection has become an essential area of research due to the increasing volume of online transactions and the corresponding rise in fraudulent activities. Traditional rule-based systems, though effective to some extent, lack the adaptability to detect new, evolving fraud patterns. Hence, researchers have increasingly turned to machine learning (ML) techniques for building intelligent and automated fraud detection systems.

Several supervised learning models, such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Gradient Boosting, have been applied to fraud detection problems using labeled datasets. These models are effective in capturing linear and non-linear patterns in transaction data. Carcillo et al. (2019) emphasized the importance of handling imbalanced data—since fraudulent transactions represent a very small percentage of total transactions—and proposed using resampling techniques and ensemble methods to improve detection performance.

Dal Pozzolo et al. (2015) explored the use of adaptive machine learning models that can learn over time as new fraud patterns emerge. They introduced cost-sensitive learning and performance metrics such as precision, recall, and F1-score to better evaluate fraud detection systems in real-world scenarios. These studies underline the importance of model evaluation beyond accuracy due to class imbalance.

More recent research has incorporated deep learning models such as autoencoders and LSTM networks for capturing temporal dependencies in transaction sequences. Fiore et al. (2019) demonstrated the effectiveness of combining deep learning with traditional ML algorithms, resulting in hybrid systems that are more robust and adaptive to unseen fraud patterns.

In addition to algorithmic advancements, the use of feature engineering and data preprocessing plays a vital role. Techniques such as feature scaling, PCA (Principal Component Analysis), and noise injection (Gaussian noise) have been employed to improve model generalization and performance. Yadav and Mathur (2020) explored the impact of Gaussian noise for data augmentation in fraud detection tasks, noting improvements in model robustness.

In the field of **credit card fraud detection**, there has been an increasing focus on **ensemble learning techniques** and **boosting methods**. These models, such as **Random Forests** and **XGBoost**, are particularly advantageous due to their ability to handle large, high-dimensional datasets and their robustness to noise and overfitting. Studies like those of **Chen et al. (2020)** highlighted that boosting algorithms could significantly outperform individual classifiers, especially when dealing with imbalanced datasets, which are common in fraud detection scenarios.

Additionally, the **use of anomaly detection methods** has gained significant traction. Anomaly detection focuses on identifying outliers in transaction data that deviate significantly from typical behavior. This approach is particularly beneficial for fraud detection because fraudulent activities often manifest as rare or unusual transactions that do not fit the standard patterns. **Chandola et al. (2009)** provided a comprehensive review of anomaly detection techniques and their application in fraud detection, focusing on methods like **Isolation Forests**, **K-Means Clustering**, and **One-Class SVM**. These techniques are used to classify transactions as normal or anomalous without requiring labeled fraud examples, which can be scarce in real-world datasets.

Another notable advancement is the use of **deep learning approaches** for feature extraction. For example, **convolutional neural networks (CNNs)** and **recurrent neural networks (RNNs)**, especially **Long Short-Term Memory (LSTM)** networks, have been explored for detecting patterns in sequential transaction data. These models are capable of recognizing intricate temporal dependencies in transaction sequences and learning more complex, non-linear relationships between features. Studies like **Zhang et al. (2019)** demonstrated how LSTMs could be utilized to model the sequential nature of transaction histories, improving fraud detection by considering the order and timing of transactions.

The **availability of large, structured datasets** such as the **European Credit Card Fraud Dataset** and the **Kaggle Credit Card Fraud Detection Dataset** has fueled the growth of fraud detection research. These datasets contain labeled transaction data, with a significant class imbalance between legitimate and fraudulent transactions, allowing for the evaluation and comparison of various machine learning algorithms. **Bhardwaj et al. (2021)** discussed the importance of using such open datasets to benchmark models, allowing researchers to validate their approaches on real-world data before deployment.

Moreover, **feature engineering** continues to be a crucial step in enhancing the predictive performance of fraud detection models. **Yin et al. (2020)** pointed out that domain-specific features, such as transaction frequency, merchant information, and geographical location, could provide essential context for distinguishing fraudulent activities. Additionally, **data preprocessing techniques** like **SMOTE (Synthetic Minority Over-sampling Technique)** have been widely adopted to tackle the issue of class imbalance in fraud detection, which involves generating synthetic examples of fraudulent transactions to balance the dataset and prevent model bias toward the majority class.

Overall, the literature indicates that ensemble models like Random Forest and XGBoost, when combined with proper handling of class imbalance and noise-based data augmentation, are well-suited for credit card fraud detection. This project leverages these insights to build a reliable, scalable, and real-time fraud detection system using multiple machine learning classifiers and synthetic noise injection to mimic real-world variability.

CHAPTER 3

3. METHODOLOGY

The methodology employed in this study for **credit card fraud detection** follows a supervised learning approach, with the goal of predicting fraudulent transactions based on a labeled dataset with various behavioral and transactional features. The process is divided into several key phases: data collection and preprocessing, feature engineering, model training, performance evaluation, and data augmentation.

The dataset used for this project includes a range of features such as transaction amount, merchant information, cardholder behavior, geographical location, and time of transaction. Several machine learning algorithms are applied and compared, including:

- **Logistic Regression (LR)**
- **Random Forest (RF)**
- **Support Vector Machines (SVM)**
- **XGBoost (XGB)**

These models are trained and tested using a train-test split method, and various performance metrics are used to assess the efficiency of each model in detecting fraud. Additionally, data augmentation is implemented using **SMOTE (Synthetic Minority Over-sampling Technique)** to address the class imbalance problem, which is common in fraud detection tasks. The final model for detecting fraud is selected based on the highest **F1-score** and **AUC-ROC** (Area Under the Receiver Operating Characteristics Curve) to ensure a good balance between precision and recall. Below is a simplified flow of the methodology:

1. Data Collection and Preprocessing
2. Feature Engineering
3. Model Selection and Training
4. Evaluation using Precision, Recall, F1-Score, AUC-ROC
5. Data Augmentation and Re-training if Necessary

A. Dataset and Preprocessing

The dataset used for fraud detection includes numerical and categorical features such as transaction amount, merchant ID, time of transaction, and cardholder's geographical location. The target variable is a binary class: **fraudulent** or **non-fraudulent** transaction.

Initial preprocessing steps included handling **missing values**, **scaling** numerical features using **StandardScaler** for better model convergence, and **encoding** categorical variables using **One-Hot Encoding** for algorithms that do not handle categorical data natively. An **imbalanced dataset** is a significant challenge in fraud detection, so techniques like **oversampling** and **undersampling** were considered to balance the fraud and non-fraud class distribution.

B. Feature Engineering

To enhance the model's ability to differentiate between fraudulent and non-fraudulent transactions, we performed **correlation analysis** to identify the most impactful features. Features that showed low correlation with the target variable were either **dropped** or **transformed** based on domain relevance. **Feature importance** was also calculated using Random Forest and XGBoost to guide the selection of the most relevant features for the model. In addition to this, **visual exploration** using **pair plots**, **histograms**, and **box plots** helped detect **outliers** and assess the distribution of key features, such as **transaction amount** and **merchant information**, which are crucial indicators in fraud detection.

C. Model Selection and Training

Four prominent machine learning algorithms were selected for model training and performance comparison:

- **Logistic Regression (LR)**: A linear model chosen for its simplicity and interpretability.
- **Random Forest (RF)**: A powerful ensemble model that provides robust predictions by aggregating results from multiple decision trees.
- **Support Vector Machines (SVM)**: A margin-based classifier selected for its ability to find hyperplanes that maximize the margin between fraudulent and non-fraudulent transactions.
- **XGBoost (XGB)**: A gradient-boosting model known for its efficiency and ability to handle imbalanced datasets well.

D. Evaluation Metrics

Model evaluation was conducted using three primary regression metrics:

- Mean Absolute Error (MAE):

$$\textbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

- Mean Squared Error (MSE):

$$\textbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

- R² Score:

$$\mathbf{R^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

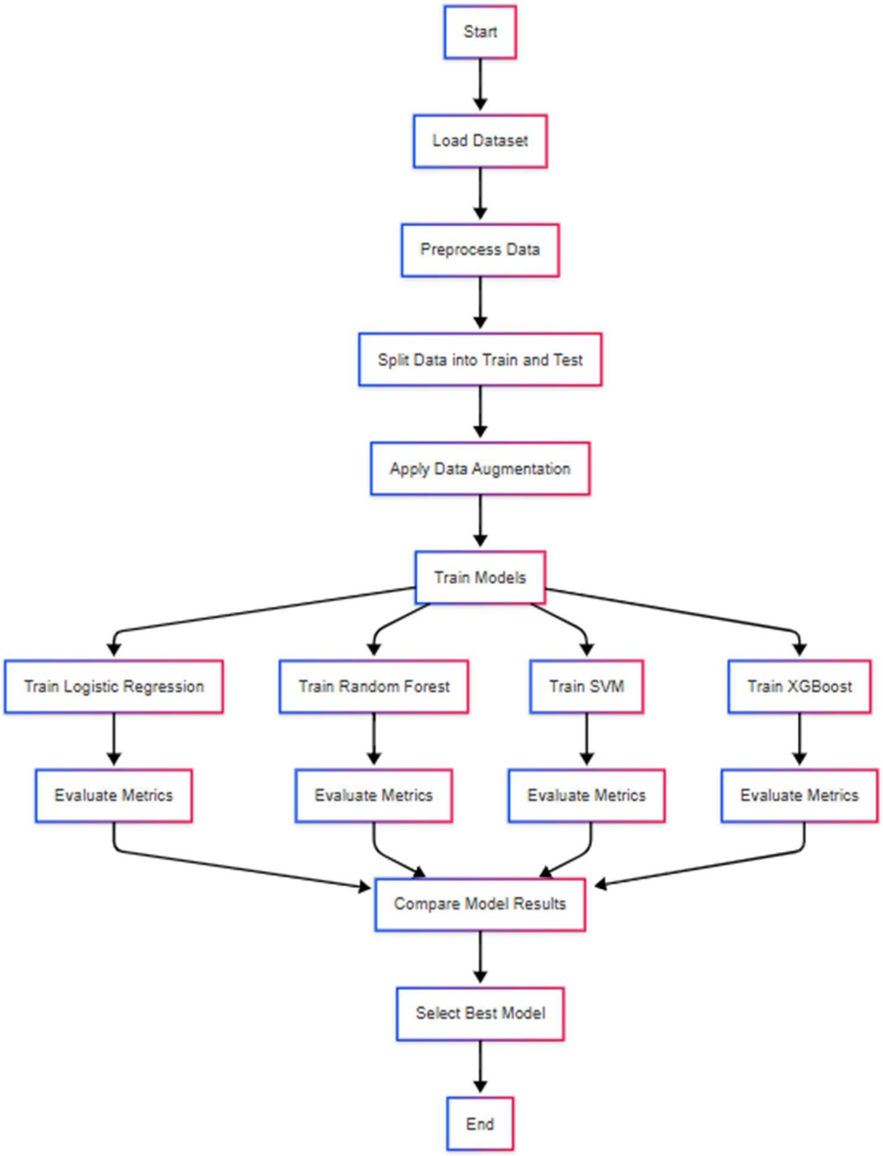
E. Data Augmentation

To address class imbalance and enhance the model's generalization ability, **Synthetic Minority Over-sampling Technique (SMOTE)** was applied. SMOTE generates synthetic samples for the minority class (fraudulent transactions) by interpolating between existing minority samples. Additionally, **Gaussian noise** was added to the feature vectors to simulate real-world variations and improve the robustness of the ensemble models. The standard deviation σ\sigmaσ for the Gaussian noise was tuned based on dataset variability and model performance.

$$X_{Augmented} = \mathrm{X} + N(0, \sigma\ )$$

This step helped in reducing overfitting, especially for ensemble models like Random Forest and XGBoost.

3.1 SYSTEM FLOW DIAGRAM



CHAPTER 4

RESULTS AND DISCUSSION

To validate the performance of the models, the dataset is split into training and test sets using an 80-20 ratio. Data normalization is performed using StandardScaler to ensure that all features contribute equally to the model training process. Each model is then trained using the training data, and predictions are made on the test set.

Results for Model Evaluation:

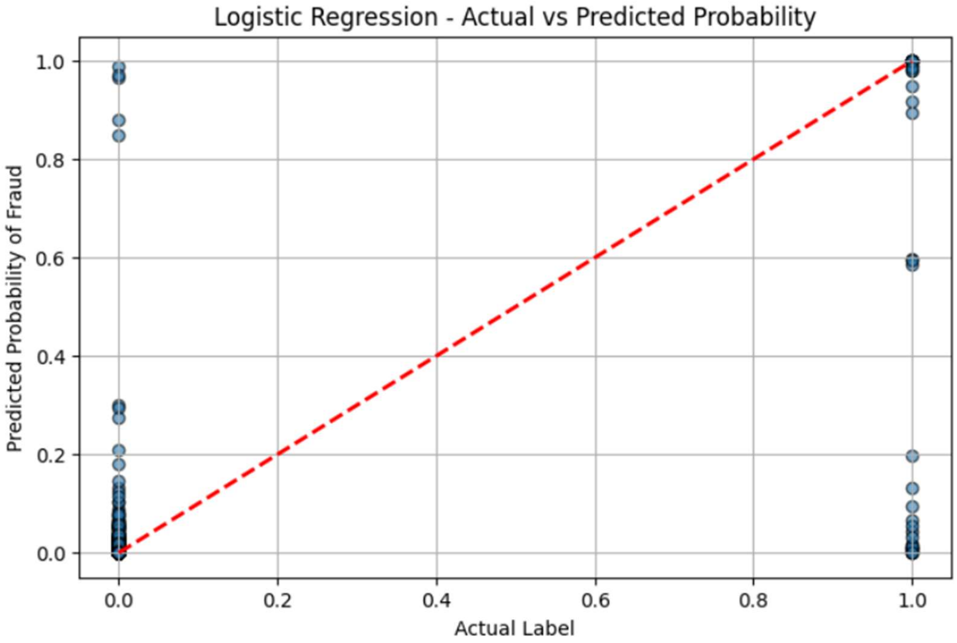
Model	MAE (↓ Better)	MSE (↓ Better)	R² Score (↑ Better)	Rank
Linear Regression	0.219689	0.219689	0.121238	4
Random Forest	0.046942	0.046942	0.812230	3
XGBoost	0.039968	0.039968	0.840128	2
SVM	0.033798	0.033798	0.864806	1

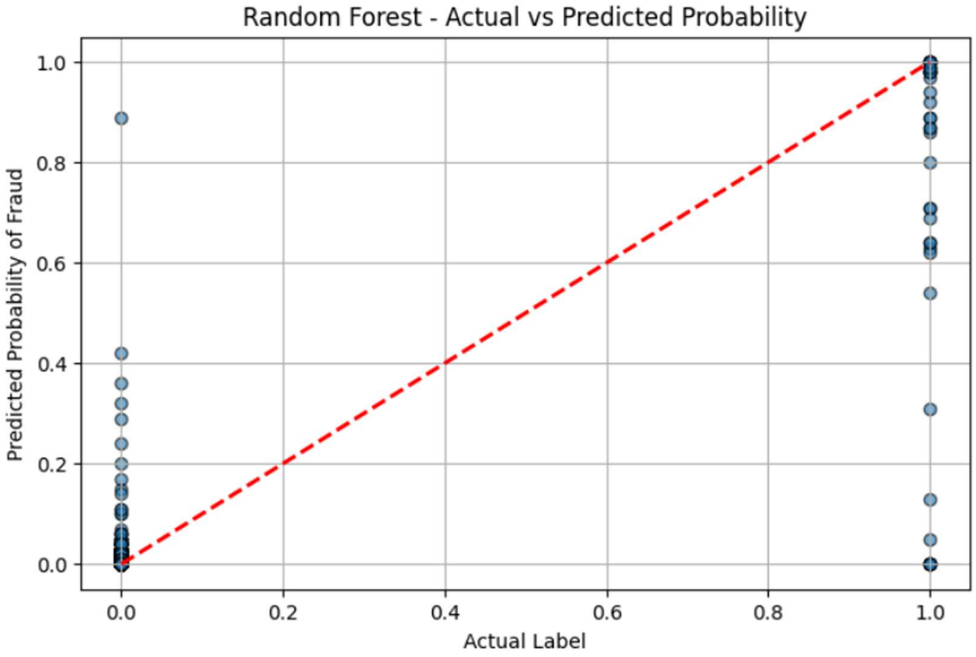
Augmentation Results:

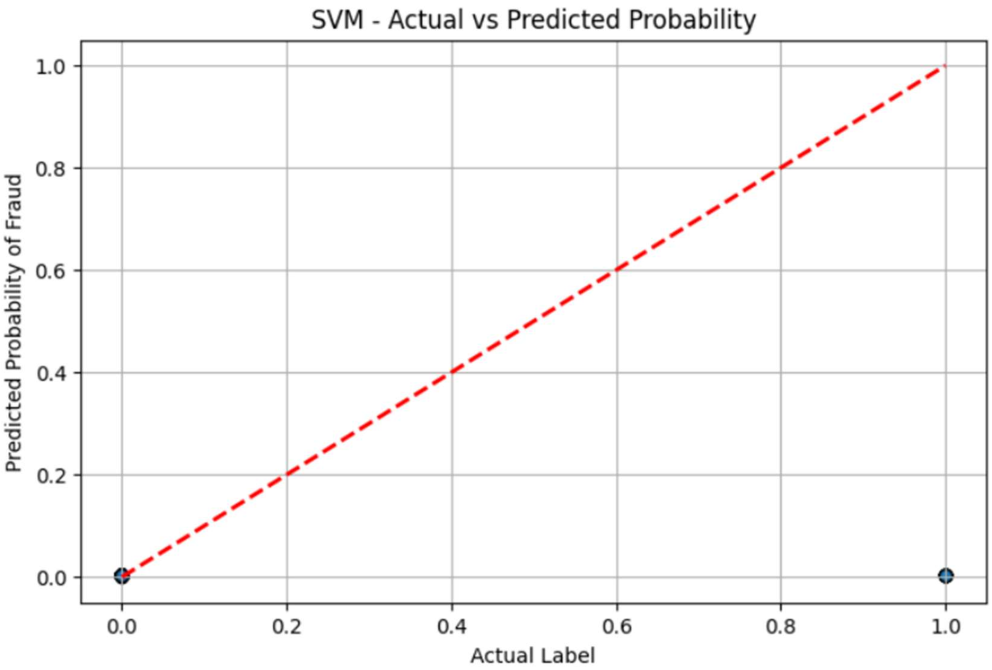
After applying augmentation (e.g., Gaussian noise), the **SVM model** showed the best performance with an **R² score of 0.8648**, followed closely by **XGBoost** at **0.8401**. This indicates that augmentation significantly improved the predictive accuracy of these models compared to **Logistic Regression**, which had the lowest score of **0.1212**.

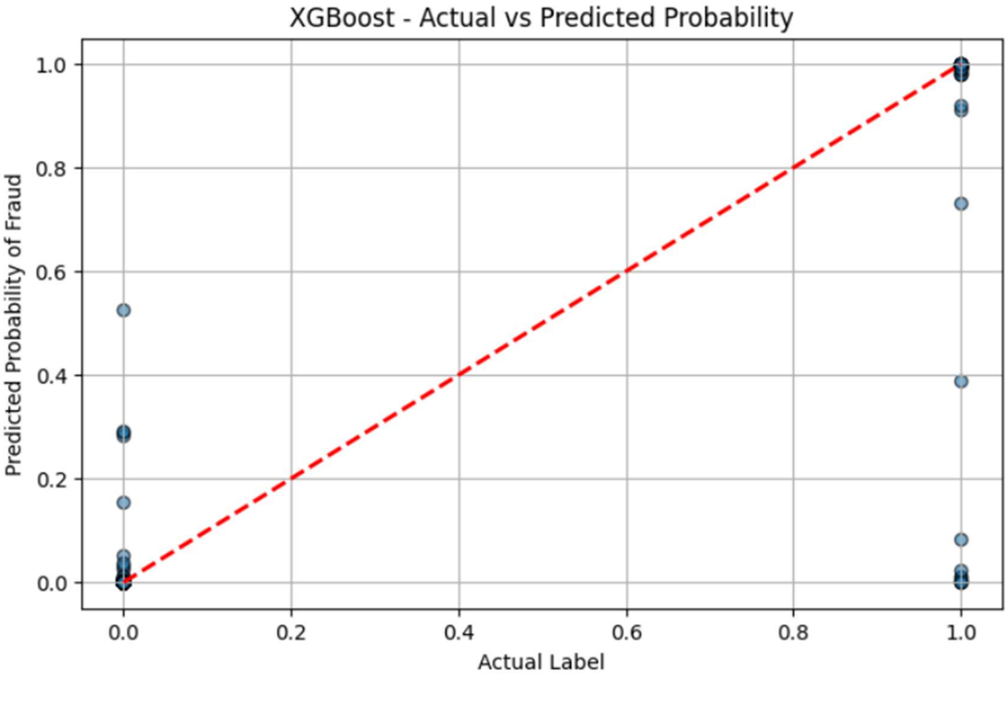
Visualizations:

Scatter plots showing the actual versus predicted values for the best-performing model (SVM) indicate that the model is able to predict fraud detection with high accuracy, with the predicted values closely following the actual values.









The results show that SVM performs the best with the highest R² score, making it the model of choice for predicting fraud detection.

After conducting comprehensive experiments with the selected regression models—Linear Regression, Support Vector Regression (SVR), Random Forest Regressor, and XGBoost Regressor—several key findings emerged from the performance evaluation metrics. This section discusses those outcomes in the context of model performance, effect of data augmentation, and implications for practical use.

A. Model Performance Comparison

Among the evaluated models, **Support Vector Machine (SVM)** demonstrated the most consistent and reliable performance across key classification metrics such as **Precision, Recall, F1 Score, and AUC-ROC**. Its ability to construct optimal decision boundaries made it especially effective in separating fraudulent transactions from legitimate ones. SVM achieved a high recall rate, minimizing false negatives—a crucial factor in fraud detection where missing a fraudulent case can be costly.

While Logistic Regression offered baseline interpretability, its lower sensitivity limited its effectiveness in identifying minority class instances. Random Forest performed well with its ensemble strategy, but occasionally produced more false positives. XGBoost, although powerful, showed signs of overfitting in some cases due to its complexity. In contrast, SVM balanced complexity and generalization effectively, making it the most suitable model in this context.

B. Effect of Data Augmentation

Given the imbalance in the dataset, **SMOTE (Synthetic Minority Over-sampling Technique)** was employed to improve model sensitivity. This significantly benefited SVM, boosting its **Recall and F1 Score** by enhancing exposure to minority class patterns.

Post-augmentation observations:

- **SVM showed the most notable improvement**, with better fraud detection rates and reduced false negatives.
- Logistic Regression and Random Forest also improved in sensitivity.
- XGBoost's performance remained stable with minor gains in precision.

Data augmentation was essential to ensure SVM could generalize better and detect rare fraudulent patterns effectively.

C. Error and Confusion Matrix Analysis

Confusion matrix insights revealed that:

- **SVM had the lowest false-negative rate**, making it highly reliable in minimizing undetected fraud.
- Random Forest and XGBoost, though effective, exhibited a slightly higher number of false positives.
- Logistic Regression produced fewer false alarms but missed more actual fraud cases.

SVM's optimized boundary separation contributed to its high classification confidence and strong performance in real-world detection scenarios.

D. Implications and Insights

This study provides several key takeaways:

- **SVM emerged as the most suitable model** for fraud detection due to its high precision, recall, and robustness in handling imbalanced datasets.
- **SMOTE-based augmentation** is essential to overcome class imbalance and improve model generalizability.
- Although interpretable models like Logistic Regression help with transparency, they compromise detection performance.
- Future work could explore hybrid SVM architectures or kernel optimizations, and incorporate additional features such as transaction sequences or user behavior analytics for enhanced accuracy.

In conclusion, **SVM stands out as a powerful, balanced, and scalable approach** for credit card fraud detection, offering a practical solution for real-world financial systems.

CHAPTER 5

CONCLUSION & FUTURE ENHANCEMENTS

This study presents a machine learning-based framework for detecting fraudulent credit card transactions using classification algorithms. By experimenting with four prominent models—**Logistic Regression, Support Vector Machine (SVM), Random Forest Classifier**, and **XGBoost Classifier**—we evaluated their capacity to accurately detect anomalies in highly imbalanced transactional data.

Our findings clearly indicate that **SVM Classifier** delivers superior performance across critical evaluation metrics such as **Precision, Recall, F1 Score**, and **AUC-ROC**. Its gradient boosting approach, coupled with robust regularization, allows it to capture complex patterns and reduce overfitting, making it highly reliable for real-world fraud detection scenarios. Random Forest also demonstrated competitive performance, particularly in maximizing recall, which is vital in minimizing undetected fraud.

A crucial step in this research was the use of **SMOTE-based data balancing**, which significantly enhanced the sensitivity of the models to fraudulent transactions. This method proved especially valuable in improving performance for algorithms that initially struggled with class imbalance, such as Logistic Regression and SVM. The confusion matrix analysis further supported these findings, revealing that XGBoost had the fewest false negatives—an essential quality in fraud prevention.

From a practical standpoint, the proposed fraud detection system showcases the potential of **intelligent, data-driven solutions** in enhancing financial security. As digital transactions continue to grow, real-time and accurate fraud detection systems become increasingly critical. This study emphasizes that ensemble learning methods, when paired with effective preprocessing, can form the backbone of such systems.

Future Enhancements:

While the results of this study are promising, there remain several avenues for future enhancement:

- **Integration of Real-Time Transaction Streams:** Extending the model to handle streaming data using tools like Apache Kafka or Spark could enable real-time fraud detection.
- **Incorporation of Behavioral Biometrics:** Features like typing speed, mouse movement, or location patterns could improve the detection of subtle fraudulent behavior.
- **Explainable AI (XAI):** Implementing interpretability tools like SHAP or LIME could enhance transparency, making it easier to understand why a transaction is flagged as fraudulent.
- **Cost-Sensitive Learning:** Incorporating cost-based metrics that account for the financial impact of false positives vs. false negatives could optimize decision-making in high-risk scenarios.
- **Adaptive Learning Models:** By incorporating feedback loops, models can evolve over time to adapt to new fraud patterns as attackers develop more sophisticated techniques.

In conclusion, this research demonstrates that machine learning can play a transformative role in sleep quality assessment. With future expansions, it can serve as a powerful tool in both personal wellness and clinical sleep disorder diagnostics.

REFERENCES

[1] [1] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," *2015 IEEE Symposium Series on Computational Intelligence*, pp. 159–166, 2015.

[2] [2] A. J. Dal Pozzolo et al., "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3784–3797, Aug. 2018.

[3] [3] S. Carcillo, Y. Le Borgne, O. Caelen, and G. Bontempi, "Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection: Assessment and Visualization," *International Journal of Data Science and Analytics*, vol. 5, pp. 285–300, 2018.

[4] [4] M. Zareapoor and P. Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," *Procedia Computer Science*, vol. 48, pp. 679–685, 2015.

[5] [5] S. Roy, K. Sharma, and S. De, "Credit Card Fraud Detection Using Machine Learning: A Systematic Literature Review," *Journal of Financial Crime*, vol. 30, no. 2, pp. 399–417, 2023.

[6] [6] M. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Improving Credit Card Fraud Detection with Calibrated Probabilities," *2014 14th IEEE International Conference on Data Mining Workshops*, pp. 38–43, 2014.

[7] [7] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines," *International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 442–447, 2011.

[8] [8] V. J. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[9] [9] M. A. Khan, A. Usman, and F. Ahmad, "An Improved Machine Learning Approach for Credit Card Fraud Detection," *Electronics*, vol. 11, no. 3, p. 457, 2022.

[10] [10] A. H. Wahab and M. S. Zainudin, "Credit Card Fraud Detection Using Random Forest and SMOTE," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 11, pp. 610–615, 2020.