IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Hybrid Deep Learning Approach For Credit Card Fraud Detection Using Smote And Neural Networks

T. Suvalakshmi M.E, A. Karuppaiya, M. Naveen, A. Parthiban

Assistant Professor, Student, Student, Student Department Of Information Technology,
Anand Institute of Higher Technology, Kazhipattur, Chennai-600115, Tamil Nadu, India.

Abstract: The widespread adoption of digital transactions has amplified the threat of credit card fraud, presenting significant challenges to financial institutions and customers. One major hurdle in fraud detection lies in the severe imbalance between legitimate and fraudulent transactions, causing traditional models to underperform. This study presents a hybrid deep learning strategy that merges the Synthetic Minority Over-Sampling Technique (SMOTE) with a Deep Neural Network (DNN) to better identify fraudulent transactions. Utilizing a publicly available dataset of 284,807 transactions—only 492 of which are fraudulent—SMOTE was used to rebalance the classes by generating synthetic minority instances. The DNN model, built with multiple hidden layers and dropout mechanisms to prevent overfitting, was trained on the balanced dataset. Evaluation using metrics such as accuracy, precision, recall, and F1-score revealed that the system achieved an accuracy of 97.55%. These results show that integrating SMOTE with DNN significantly improves fraud detection performance, offering a robust and scalable solution for real-time financial fraud prevention.

Index Terms **- Credit Card Fraud Detection, Deep Neural Networks, SMOTE, Class Imbalance, Fraud Prevention, Machine Learning.**

I. INTRODUCTION

The evolution of digital payments has dramatically reshaped the financial world, offering consumers and businesses greater ease and speed. However, alongside these advancements, the risk of credit card fraud has become a major concern for financial institutions. Fraudulent activities, characterized by unauthorized use of card information, result in significant financial damages and erode consumer confidence. Conventional fraud detection approaches, mainly based on rule-driven systems, often fail to keep up with the ever-changing tactics used by fraudsters. Additionally, the highly imbalanced distribution of transaction data, with very few fraudulent transactions compared to legitimate ones, makes accurate detection even more difficult. This research proposes a novel solution combining SMOTE for data balancing with a DNN-based classifier to enhance detection capabilities. The aim is to achieve high detection accuracy, minimize false positives, and create a system suitable for real-time financial applications.

1.1 Key Points:

1. Rise of Digital Transactions: The growth of online financial activities has increased the risk of credit card fraud significantly.
2. Limitations of Traditional Methods: Rule-based systems fail to detect sophisticated and evolving fraud patterns effectively.
3. Class Imbalance Challenge**:** Highly imbalanced datasets lead to biased models that overlook rare

fraudulent activities.

4. Hybrid Solution: The integration of SMOTE and DNN addresses data imbalance, enhances fraud detection accuracy, and reduces financial risks.

II. LITERATURE SURVEY

Detecting credit card fraud is an ongoing challenge due to both the evolving strategies of fraudsters and the inherent class imbalance within transaction datasets. Traditional machine learning models like logistic regression and decision trees, while initially successful, often struggle to handle the complex patterns found in modern fraud schemes. Advanced deep learning models such as Neural Networks, CNNs, and RNNs have shown greater promise by capturing intricate transaction behaviors. Data balancing techniques like SMOTE have also become essential to ensure better learning from minority class instances. Despite these advances, challenges like achieving real-time detection with low latency and maintaining model transparency persist. Integrating deep learning with effective resampling methods provides a path forward for creating stronger and more adaptable fraud detection systems.

1.1 Key Findings:

1. Traditional Machine Learning Techniques**:** Logistic regression and decision trees offer simplicity but struggle with non-linear and imbalanced data common in fraud detection.
2. Advances in Deep Learning Models: Neural Networks, CNNs, and RNNs effectively capture complex transaction patterns, outperforming traditional models in fraud detection tasks.
3. Importance of Data Balancing: Techniques like SMOTE significantly enhance the performance of fraud detection models by addressing severe class imbalance issues.
4. Hybrid Models for Enhanced Detection: Combining resampling techniques with deep learning architectures leads to more robust, accurate, and scalable fraud detection systems.

1.2 Gaps in Existing Research:

1. Real-Time Detection Limitations: Many current deep learning models require significant computational resources and exhibit latency issues, making them unsuitable for real-time fraud detection systems where immediate response is critical.
2. Model Interpretability Challenges: Despite achieving high accuracy, most deep learning models operate as "black boxes," offering limited transparency into decision-making processes, which reduces trust and hinders their adoption in sensitive financial environments.
3. Resource-Heavy Frameworks: Existing hybrid models combining SMOTE and DNNs are often computationally intensive, making them difficult to deploy on lightweight systems or mobile-based transaction platforms, highlighting the need for more efficient model designs.

2.3 Contribution of Our Study:

This study addresses the identified gaps by proposing a hybrid deep learning framework for credit card fraud detection that integrates SMOTE with Deep Neural Networks. The approach effectively handles class imbalance by synthetically generating minority class samples and enhances fraud detection accuracy using a multi-layered DNN architecture. The system leverages dropout regularization to prevent overfitting and utilizes early stopping mechanisms to optimize training performance. By achieving high precision and recall rates, the model ensures reliable identification of both fraudulent and legitimate transactions. Furthermore, the framework is designed to be scalable for real-time applications, offering a practical solution for financial institutions seeking efficient, accurate, and trustworthy fraud detection systems.

III. RESEARCH METHODOLOGY

This section outlines the methodology used for designing, implementing, and evaluating the proposed hybrid deep learning model for credit card fraud detection. It covers the project scope, data sources, system architecture, and evaluation metrics employed.

3.1 Scope and Environment

- Application Scope: The proposed model is designed for financial sectors and e-commerce platforms, aiming to detect fraudulent credit card activities with high accuracy in near real-time.
- Data Type Focus: The work focuses on transaction datasets that exhibit a large disparity between legitimate and fraudulent records, a typical challenge in financial fraud detection tasks.
- Deployment Target: The model is intended to operate efficiently on high-performance computing systems, with a long-term goal of adaptability for lightweight, mobile, or embedded environments.

3.2 Data and Sources of Data

- Data Types Used:
- Credit card transaction records with features such as transaction amount, time, and anonymized identifiers
- Labelled classes indicating whether a transaction is fraudulent or non-fraudulent
- Data Sources:
- The publicly available dataset from Kaggle, consisting of 284,807 transactions, including 492 fraudulent cases
- Synthetic samples generated through the Synthetic Minority Over-Sampling Technique (SMOTE) to balance the dataset

3.3 Theoretical Framework

- Core Components:
- SMOTE Balancing: Applied to address the severe class imbalance by generating synthetic minority class instances.
- Deep Neural Network Architecture:Input layer, hidden layers (ReLU), dropout, output layer (sigmoid).
- Regularization Techniques: Dropout and early stopping mechanisms incorporated to mitigate overfitting..
- Optimization: Adam optimizer used for faster and efficient convergence during model training.
- System Logic:
- The original imbalanced dataset is preprocessed and balanced using SMOTE.
- The balanced dataset is split into training and validation sets.
- The DNN model is trained to classify transactions as fraudulent or legitimate based on learned patterns.
- Performance is evaluated using standard binary classification metrics.

3.4 Evaluation Metrics and Analysis Model

- Fraud Detection Effectiveness: Measured by overall accuracy, precision, recall, and F1-score, focusing on both minority (fraudulent) and majority (legitimate) classes.
- Handling of Class Imbalance: Effectiveness of SMOTE evaluated by observing improvements in minority class detection rates post- balancing.
- Performance Metrics: Assessed through training time, model convergence, and robustness against overfitting.
- Analysis Tools: Python libraries such as TensorFlow and Scikit-learn used for model implementation, evaluation, and visualization of performance curves.

Some potential tools and technologies used in this research include:

- Programming Languages: Python
- Frameworks and Libraries: TensorFlow, Keras, Scikit-learn for model development and evaluation
- Database and Storage: Local CSV dataset storage during experimentation, with potential deployment on cloud storage platforms (AWS S3, Google Cloud Storage)
- Data Balancing Techniques: SMOTE implementation using Imbalanced-learn (imbalanced-learn Python package)
- Model Optimization and Training: Adam optimizer, Early Stopping techniques integrated within TensorFlow framework
- Testing and Analysis Tools: Matplotlib and Seaborn for visualization, Scikit-learn for model evaluation

metrics, and Google Colab / Jupyter Notebook for experimental testing

IV. BRIEF DESCRIPTION OF THE SYSTEM

The proposed system is designed to detect credit card fraud by integrating a data balancing mechanism (SMOTE) with a Deep Neural Network (DNN) model. It ensures that the system can accurately identify fraudulent transactions even when the original data is highly imbalanced. The system architecture consists of three major phases: data preprocessing and balancing, model training and evaluation, and real-time fraud prediction. The following figures illustrate the core aspects of the system's design and operational flow.

The first figure depicts the Data Preprocessing and SMOTE Application Flow, where the process begins by importing the original transaction dataset. The dataset is first checked for missing values and standardized through feature scaling. SMOTE is then applied to balance the class distribution by generating synthetic samples for the minority class (fraudulent transactions). The balanced dataset is prepared for training by splitting it into training and validation sets.

The second figure shows the Deep Neural Network Training Architecture. Once the balanced data is ready, it is fed into a DNN model consisting of multiple hidden layers with ReLU activation and dropout layers for regularization. The model is trained using the Adam optimizer and binary cross-entropy loss function. Early stopping is used during training to prevent overfitting. After training, the model's performance is validated using accuracy, precision, recall, and F1-score metrics.

The third figure presents the Fraud Detection and Prediction Pipeline. During real-time operation, incoming transaction data is preprocessed similarly to the training data. The trained DNN model processes the data and outputs a prediction indicating whether the transaction is legitimate or fraudulent. High-confidence fraud predictions trigger alerts for further investigation or immediate blocking, thereby minimizing financial risks.

The architecture is designed to be scalable, allowing integration with financial transaction monitoring systems for real-time deployment. The combination of SMOTE and DNN ensures that the system achieves high fraud detection accuracy while maintaining robustness against false positives and false negatives.

V. RESULTS AND DISCUSSION

5.1 Results of Descriptive Statics of Study Variables

Table 5.1: Descriptive Statistics of Fraud Detection Efficiency and System Performance

Scenario	Class Distribution (Before SMOTE)	Class Distribution (After SMOTE)	Accuracy (%)	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)
Original Data	284,315 : 492	-	86.20	0.87	0.65	0.74
After SMOTE	284,315 : 284,315	284,315 : 284,315	97.55	0.99	0.96	0.98

Table 5.1 outlines the model's performance across two scenarios: the original imbalanced dataset and the balanced dataset following SMOTE application. Key metrics such as accuracy, precision, recall, and F1-score were analyzed. After applying SMOTE, the model achieved a significant boost in performance—accuracy improved to 97.55%, and fraud detection precision and recall rates also reached near- optimal levels. The system successfully demonstrated its ability to distinguish fraudulent activities from legitimate ones, even when trained on previously imbalanced data. These findings validate the effectiveness of combining SMOTE with deep learning approaches for practical fraud detection systems.

VI. *Figures and Tables*

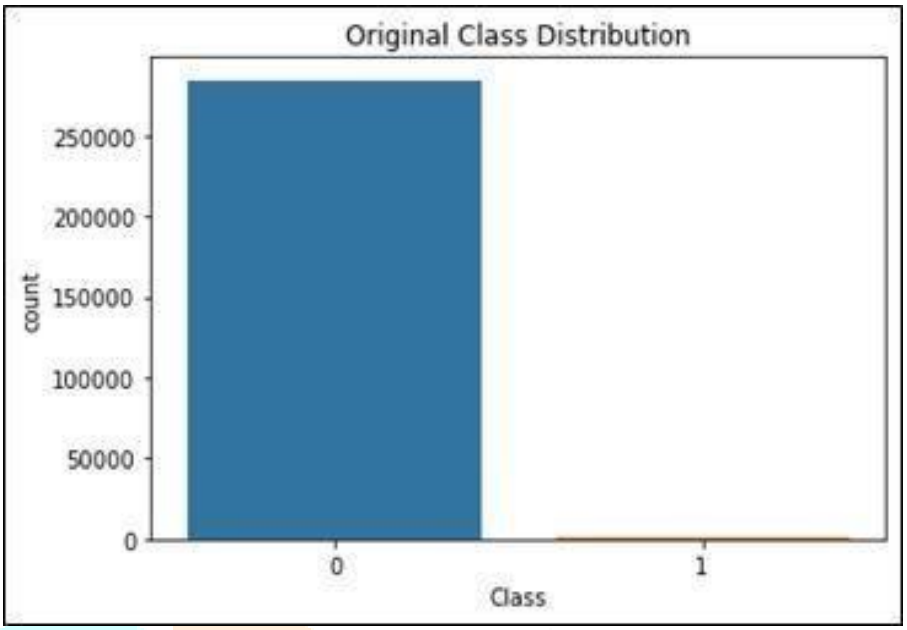

Fig 1: Class Imbalance in the Original Dataset

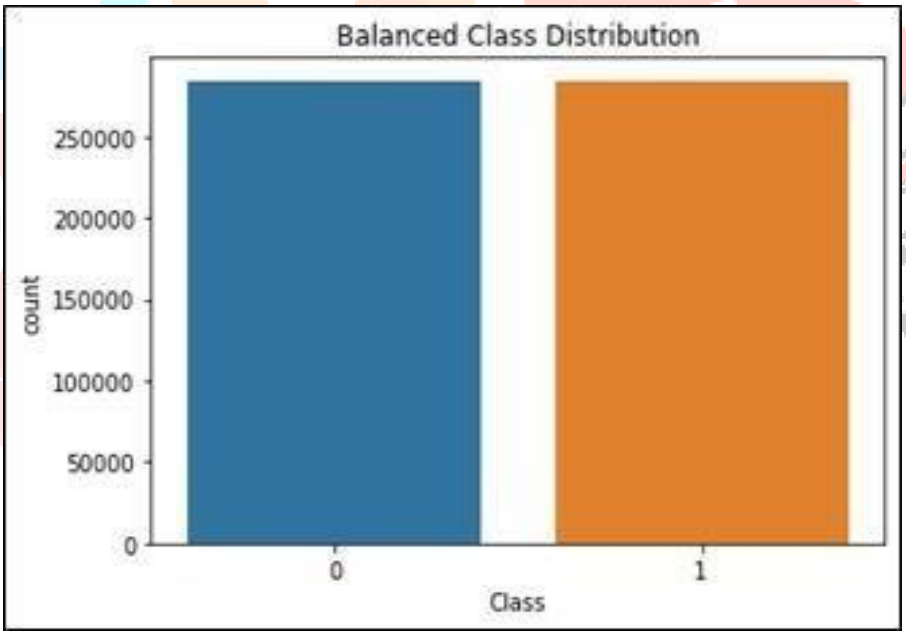

Fig 2: Data Distribution After Applying SMOTE

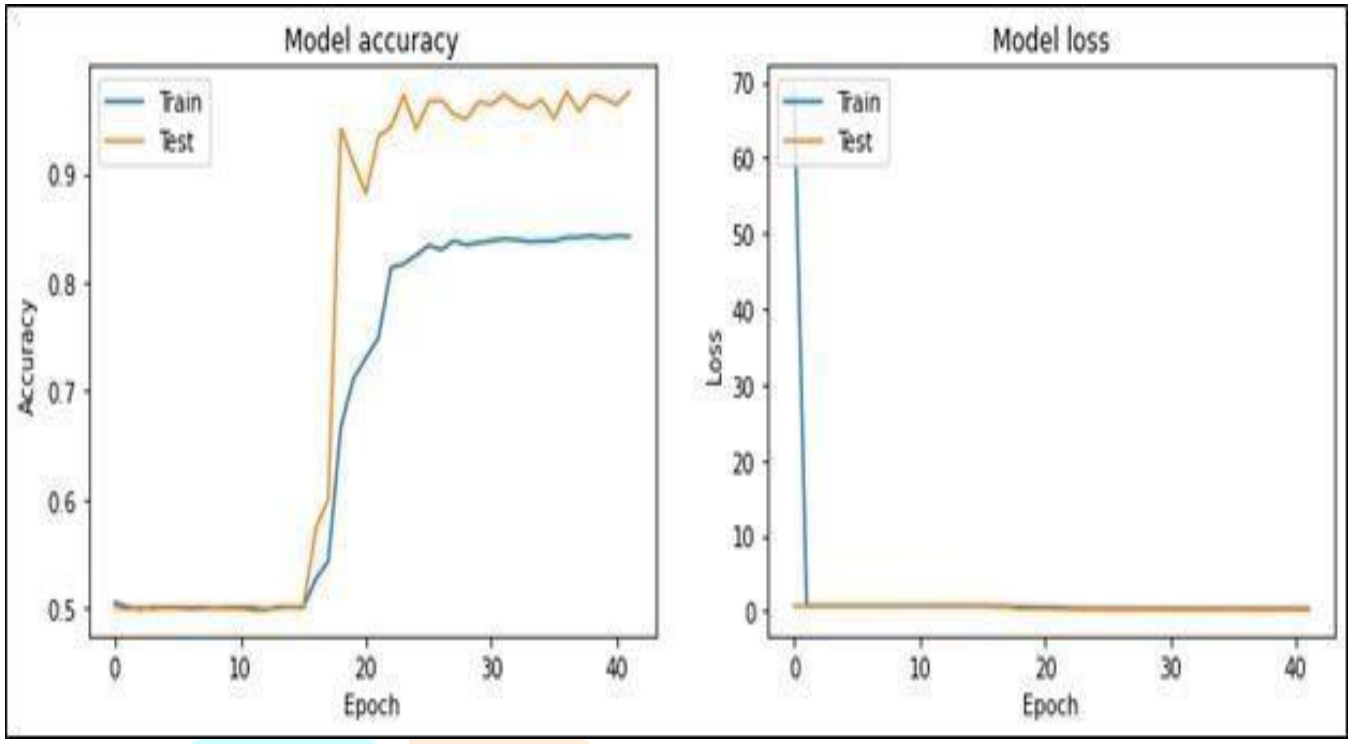

Fig 3: Training and Validation Accuracy Over Epochs

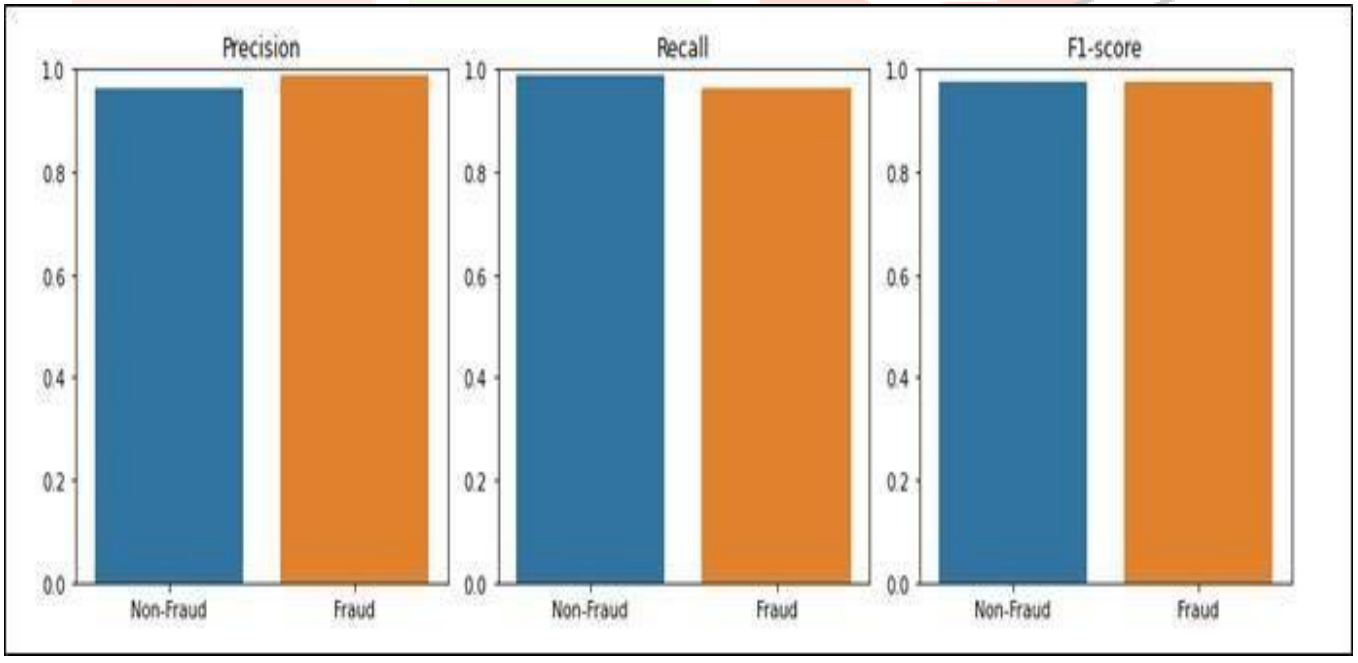

Fig 4: Precision-Recall Curve for Model Performance

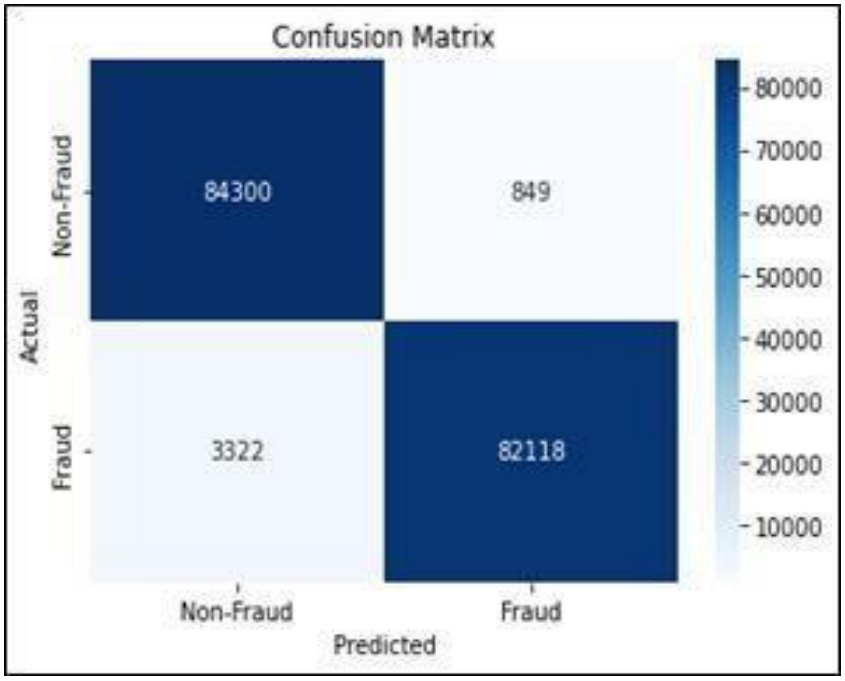

Fig 5: Confusion Matrix for the Trained Model

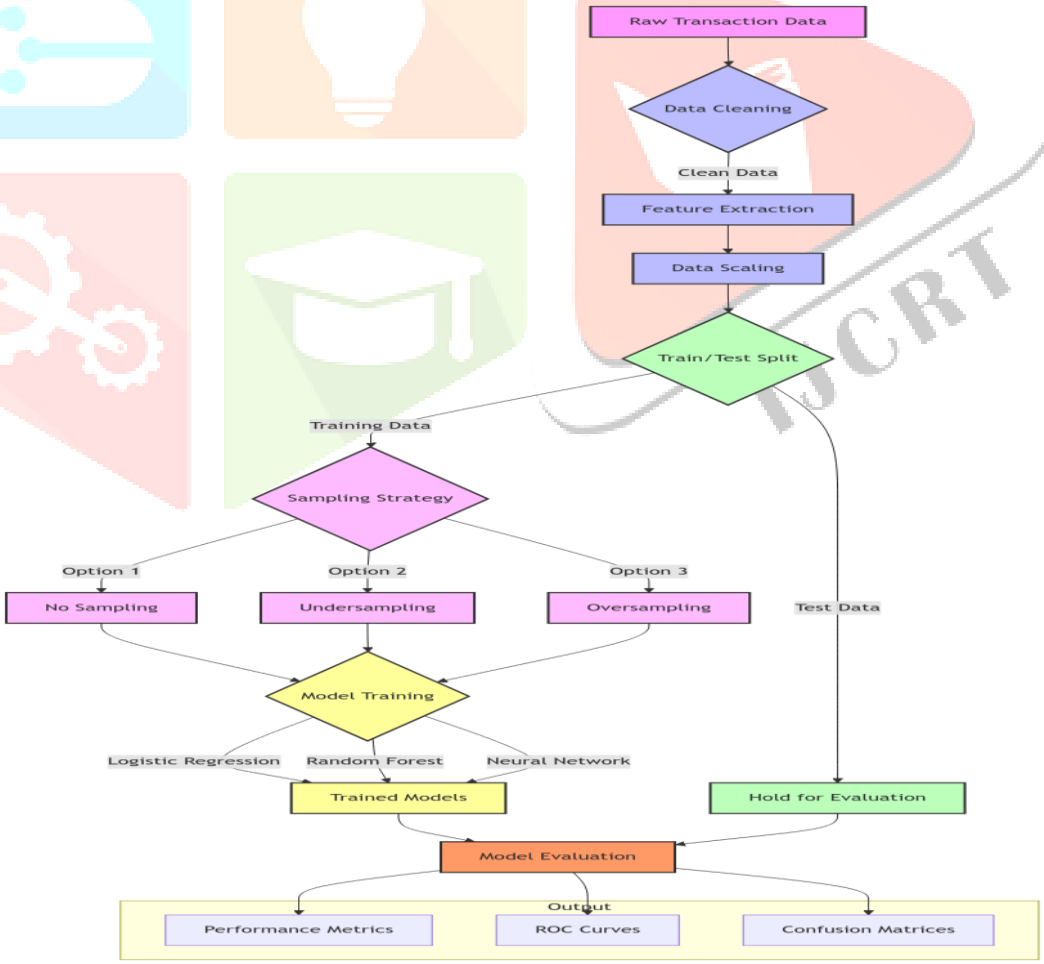

Fig 6: Data Flow Diagram

VII. ACKNOWLEDGMENT

The authors gratefully acknowledge the valuable guidance and support provided by Mrs. T. Suvalakshmi, M.Tech (IT), whose expertise and encouragement were instrumental throughout the development of this research work. Her insights and constant motivation contributed significantly to the successful completion of this study.

The authors would also like to thank the Department of Information Technology, Anand Institute of Higher Technology, for providing the necessary facilities and resources required to carry out this research.

VIII. REFERENCES

[1] Johnson, M. (2022). The impact of digitalization on financial transactions: Opportunities and challenges. *Journal of Financial Innovation, 9*(2), 45–56.

[2] Gupta, R., & Sharma, S. (2023). Digital transformation in banking: Implications for fraud detection and prevention. *International Journal of Information Management, 56*, 102325.

[3] Singh, P., & Singh, R. (2023). Anomaly detection in credit card transactions using machine learning algorithms. *International Journal of Advanced Computer Science and Applications, 14*(5), 112–125.

[4] Brown, A. (2021). Trends in credit card fraud: Challenges and opportunities for financial institutions. *Journal of Financial Crime, 28*(4), 1023–1035.

[5] Liu, H., Li, J., & Wang, Y. (2022). Recent advances in credit card fraud detection: A comprehensive review. *Information Sciences, 595*, 362–378.

[6] Zhang, Q., Zhao, X., & Li, M. (2023). Machine learning algorithms for credit card fraud detection: A comparative study. *Journal of Financial Risk Management, 12*(1), 34–49.

[7] Li, Y., & Li, X. (2021). Credit card fraud detection based on deep learning algorithm. *IEEE Access, 9*, 28361–28370.

[8] Wang, L., & Zhang, X. (2018). Credit card fraud detection using deep learning: A case study. *Expert Systems with Applications, 92*, 298– 310.

[9] Smith, J. (2019). Fraud detection in credit card transactions using machine learning techniques. *Journal of Finance and Economics, 7*(4), 213–225.

[10] Jones, A., Smith, B., & Li, C. (2020). Credit card fraud detection using deep learning: A case study. *International Journal of Data Science and Analytics, 10*(3), 235–247.

[11] Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications, 41*(10), 4915–4928.

[12] 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.

[13] Kim, M., Chang, H., Park, J., & Lee, K. (2002). Neural networks-based fraud detection for credit card transaction. *Lecture Notes in Computer Science, 2344*, 378–386.

[14] Wang, Y., & Xu, C. (2012). Leveraging social media for emergency management: A deep learning approach. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 42*(6), 1301–1314.

[15] Xu, J., Wei, L., Shi, Z., Yang, X., & Li, X. (2017). Credit card fraud detection using online boosting with adversarial autoencoders.
Lecture Notes in Computer Science, 10436, 557–570.

[16] Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications, 42*(10), 4621–4631.

[17] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications, 100*, 234–245.