International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Graph-Aware Machine Learning Framework for Anti-Money Laundering in Cryptocurrency Transactions Using Hybrid Classification Models

VEERAGANI VENKATA VINAY

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

A. Naga Raju

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid proliferation of cryptocurrency transactions has introduced unprecedented challenges in financial regulation, particularly in detecting illicit activities such as money laundering. Traditional Anti-Money Laundering (AML) systems are largely rule-based and struggle to adapt to the dynamic, decentralized, and pseudonymous nature of blockchain ecosystems. This research proposes a graph-aware machine learning framework that integrates conventional classification algorithms with deep learning techniques to enhance the detection of suspicious cryptocurrency transactions.The proposed system leverages transaction-level data, including features such as transaction amount and sender location, and applies preprocessing techniques such as label encoding, normalization, and Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalance. Multiple machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, and Decision Tree, are employed to classify transactions into legitimate or laundering categories. Additionally, a Convolutional Neural Network (CNN)-based model is introduced to capture non-linear feature interactions and improve classification performance.A key contribution of this work lies in its conceptual alignment with Graph Neural Networks (GNNs), where transaction relationships are implicitly modeled through structured feature learning. Although the implementation primarily utilizes tabular data, the framework is designed to extend toward graph-based representations for future scalability in blockchain analytics.The system is implemented using a Django-based web interface that enables real-time prediction and visualization of model performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are computed to assess model effectiveness. Experimental results demonstrate that ensemble methods, particularly Random Forest, outperform other algorithms in terms of classification accuracy, while the CNN model shows competitive performance in capturing complex data patterns.The findings indicate that combining traditional machine learning techniques with deep learning models significantly enhances the robustness of AML systems. The integration of SMOTE improves minority class detection, addressing the critical issue of imbalanced datasets in fraud detection scenarios. Furthermore, the system provides an interpretable and scalable approach suitable for real-world financial monitoring applications.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Keywords: Anti-Money Laundering, Cryptocurrency, Graph Neural Networks, Fraud Detection, Machine Learning, SMOTE, Random Forest, Deep Learning, Transaction Classification, Financial Security

I. INTRODUCTION

The emergence of cryptocurrencies has fundamentally transformed the global financial landscape by enabling decentralized, peer-to-peer transactions without the need for intermediaries. While this innovation offers numerous advantages, including increased efficiency and reduced transaction costs, it has also created new avenues for financial crimes such as money laundering, fraud, and terrorist financing. The pseudonymous nature of blockchain transactions makes it difficult to trace illicit activities, posing significant challenges for regulatory authorities and financial institutions.Anti-Money Laundering (AML) systems have traditionally relied on rule-based approaches and manual monitoring processes to identify suspicious transactions. However, these systems are often limited by their inability to adapt to evolving laundering techniques and the sheer volume of transaction data generated in modern financial systems. As a result, there is a growing need for intelligent, data-driven solutions that can automatically detect anomalous patterns and improve the accuracy of fraud detection.Machine learning (ML) has emerged as a powerful tool in this domain, offering the ability to learn complex patterns from large datasets and make predictive decisions. Various classification algorithms, such as Decision Trees, Support Vector Machines (SVM), and Random Forests, have been widely used for fraud detection tasks. These models can effectively identify suspicious transactions based on historical data, but they often struggle with highly imbalanced datasets, where fraudulent transactions represent only a small fraction of the total data.To address these challenges, this research proposes a hybrid machine learning framework that integrates multiple classification algorithms with deep learning techniques. The system incorporates data preprocessing steps such as normalization, label encoding, and Synthetic Minority Over-sampling Technique (SMOTE) to enhance model performance. Additionally, a Convolutional Neural Network (CNN) is employed to capture complex feature relationships and improve classification accuracy.A novel aspect of this work is its alignment with Graph Neural Network (GNN) concepts, which are particularly well-suited for analyzing transaction networks. In cryptocurrency systems, transactions can be naturally represented as graphs, where nodes represent users and edges represent transactions. Although the current implementation focuses on tabular data, the proposed framework is designed to extend toward graph-based modeling in future work.The system is implemented using a Django web framework, providing an interactive interface for users to upload data, train models, and visualize results. Performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score, ensuring a comprehensive assessment of model effectiveness.

This research aims to bridge the gap between traditional AML systems and modern data-driven approaches by developing a scalable, efficient, and adaptable framework for detecting money laundering in cryptocurrency transactions. The proposed solution not

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

only improves detection accuracy but also lays the foundation for future advancements in graph-based financial analytics.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

The detection of money laundering activities has been an active area of research, particularly with the rise of digital financial systems and cryptocurrencies. Traditional approaches to Anti-Money Laundering (AML) have primarily relied on rule-based systems and statistical methods. These systems use predefined thresholds and heuristics to identify suspicious transactions. However, such approaches often fail to detect complex and evolving laundering patterns, leading to high false positive rates and limited scalability.Recent advancements in machine learning have significantly improved fraud detection capabilities. Supervised learning algorithms such as Decision Trees, Random Forests, and Support Vector Machines (SVM) have been widely applied to classify financial transactions. Random Forest, in particular, has demonstrated strong performance due to its ensemble nature, which reduces overfitting and improves generalization. Similarly, SVM models are effective in high-dimensional spaces but may suffer from computational complexity when dealing with large datasets.Naïve Bayes classifiers have also been explored for AML applications due to their simplicity and efficiency. However, their assumption of feature independence limits their effectiveness in capturing complex relationships between transaction attributes. Logistic Regression, another widely used method, provides interpretable results but may not perform well in highly non-linear datasets.

In recent years, deep learning techniques have gained prominence in fraud detection. Neural networks, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown the ability to model complex patterns and temporal dependencies in transaction data. CNNs, although traditionally used for image processing, have been successfully adapted for structured data classification tasks by capturing local feature interactions.A significant challenge in AML datasets is class imbalance, where fraudulent transactions are rare compared to legitimate ones. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE) have been proposed to address this issue by generating synthetic samples of the minority class. Studies have shown that applying SMOTE can significantly improve recall and F1-score, making it a valuable preprocessing step in fraud detection systems.More recently, Graph Neural Networks (GNNs) have emerged as a promising approach for analyzing transaction networks. Unlike traditional models, GNNs can capture relationships between entities by representing data as graphs. In the context of cryptocurrency, where transactions form complex networks, GNNs provide a natural and effective way to detect suspicious patterns. Research has demonstrated that GNN-based models outperform traditional methods in identifying money laundering activities by leveraging network

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

topology and node interactions.Despite these advancements, many existing systems lack integration between traditional machine learning methods and graph-based approaches. This research addresses this gap by proposing a hybrid framework that combines the strengths of multiple algorithms while maintaining extensibility toward graph-based modeling.

III. EXISTING SYSTEM

Existing Anti-Money Laundering (AML) systems predominantly rely on rule-based frameworks and conventional statistical techniques to detect suspicious financial activities. These systems use predefined rules, such as transaction thresholds and frequency limits, to flag potentially fraudulent transactions. While effective in controlled environments, such approaches lack adaptability and fail to capture complex patterns associated with modern money laundering techniques, particularly in cryptocurrency ecosystems.Traditional machine learning models have been introduced to improve detection accuracy. Algorithms such as Decision Trees, Support Vector Machines (SVM), and Logistic Regression are commonly used for classification tasks. However, these models often operate independently and do not leverage ensemble or hybrid approaches, limiting their overall performance. Furthermore, they struggle with highly imbalanced datasets, where fraudulent transactions constitute a very small percentage of the data.Another limitation of existing systems is their reliance on tabular data without considering the relational nature of financial transactions. In cryptocurrency networks, transactions are inherently interconnected, forming complex graphs. Ignoring these relationships results in the loss of critical contextual information, reducing the effectiveness of fraud detection mechanisms.Additionally, many existing solutions lack real-time processing capabilities and user-friendly interfaces for monitoring and analysis. This makes it difficult for financial institutions to respond promptly to suspicious activities.Overall, current AML systems are constrained by limited scalability, high false positive rates, and an inability to adapt to evolving laundering strategies. These limitations highlight the need for more advanced, integrated approaches that combine machine learning, deep learning, and graph-based techniques to improve detection accuracy and system efficiency.

IV. PROPOSED METHOD

The proposed system introduces a hybrid machine learning framework designed to enhance Anti-Money Laundering (AML) detection in cryptocurrency transactions. The system integrates multiple classification algorithms with deep learning techniques to improve accuracy, robustness, and scalability.The framework begins with data preprocessing, where raw transaction data is cleaned and transformed. Label encoding is applied to categorical variables such as sender location, and feature scaling is performed using standardization techniques. To address the issue of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is employed, ensuring that the model

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

effectively learns patterns associated with fraudulent transactions.The system utilizes a combination of machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, and Decision Tree. These models are trained and evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. Among these, Random Forest serves as the primary model for prediction due to its superior performance in handling complex datasets.In addition to traditional models, a Convolutional Neural Network (CNN) is incorporated to capture non-linear relationships and enhance classification performance. The CNN model processes reshaped transaction data and learns hierarchical feature representations, enabling more accurate detection of suspicious activities.The system is implemented using the Django web framework, providing an interactive interface for users to train models, visualize results, and perform real-time predictions. Graphical representations, including confusion matrices and performance comparison charts, are generated to facilitate analysis.Although the current implementation focuses on tabular data, the proposed framework is designed to extend toward Graph Neural Networks (GNNs) in future work. This will enable the system to model transaction networks more effectively and further improve detection accuracy in large-scale cryptocurrency environments.

V. IMPLEMENTATION

The implementation of the proposed Anti-Money Laundering (AML) framework is executed using Python and the Django web framework, which allows for real-time interaction and visualization of results. The dataset used comprises cryptocurrency transaction records, including critical features such as transaction amount, sender location, and transaction labels indicating normal or suspicious activity. Data preprocessing involves several steps to prepare the dataset for effective model training. Missing values are handled using imputation techniques, categorical features such as sender bank location are encoded using label encoding, and numerical features are standardized using the StandardScaler to ensure uniform feature scaling.To address the inherent class imbalance typical in fraud detection datasets, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. This method generates synthetic samples for the minority class (fraudulent transactions), which enhances the model's ability to detect anomalies and improves recall and F1-score.The framework integrates multiple machine learning algorithms. Random Forest is used as the primary ensemble classifier due to its robustness and ability to handle high-dimensional data. Support Vector Machine (SVM), Naïve Bayes, Decision Tree, and Logistic Regression classifiers are also trained to provide comparative performance analysis. Each algorithm is trained on 80% of the dataset and evaluated on the remaining 20%. The models' performance is quantified using metrics including accuracy, precision, recall, and F1-score, which are calculated through a dedicated calculateMetrics function implemented in Python. Confusion matrices are generated to visualize the classification outcomes for each algorithm.To capture non-linear feature interactions and hierarchical patterns, a Convolutional Neural Network (CNN) is implemented. Transaction data is reshaped into a four-dimensional array suitable for CNN processing. The CNN architecture consists of convolutional layers with ReLU activation functions, max-pooling layers to reduce dimensionality, and fully

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

connected dense layers followed by a softmax output layer. The model is trained with the Adam optimizer using categorical cross-entropy loss for 30 epochs with a batch size of eight. Model checkpointing ensures the best-performing weights are saved during training.The Django-based web interface provides several functionalities, including dataset uploading, model training, prediction, and performance visualization. Users can input transaction details to obtain real-time probability scores for fraudulent activity. Graphical visualizations, such as confusion matrices and bar charts comparing algorithm performance, are generated using matplotlib and seaborn. These visualizations enhance interpretability and allow stakeholders to understand model behavior in operational settings. The system architecture is modular, enabling future integration with Graph Neural Networks (GNNs) to model complex transaction relationships explicitly, which is crucial for blockchain networks where node connectivity and transaction patterns provide valuable context for AML detection.

VI. ALGORITHMS

The system employs multiple machine learning and deep learning algorithms to classify cryptocurrency transactions into legitimate and suspicious categories.

1. **Random Forest (RF):** An ensemble learning method that combines multiple decision trees to improve classification accuracy. Each tree is trained on a bootstrapped sample of the dataset, and predictions are made using majority voting. RF is robust against overfitting and handles high-dimensional data effectively.
2. **Support Vector Machine (SVM):** A supervised learning algorithm that finds an optimal hyperplane to separate data points of different classes in high-dimensional space. The radial basis function kernel is commonly used to handle non-linear relationships. SVM is sensitive to imbalanced data, which is mitigated using SMOTE in preprocessing.
3. **Naïve Bayes (NB):** A probabilistic classifier based on Bayes' theorem. It assumes independence between features, making it computationally efficient. NB is particularly effective for large-scale datasets but may underperform when feature dependencies are significant.
4. **Logistic Regression (LR):** A linear classifier that models the probability of a transaction being fraudulent using a logistic function. It provides interpretable coefficients but is limited in capturing non-linear relationships.
5. **Decision Tree (DT):** A tree-based classifier that splits data recursively based on feature thresholds to maximize information gain. DTs are intuitive and interpretable but prone to overfitting, which ensemble methods like RF mitigate.
6. **Convolutional Neural Network (CNN2D):** Deep learning model adapted for tabular data, reshaped into four-dimensional tensors. Convolutional layers extract

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

local patterns and hierarchical relationships, followed by max-pooling for dimensionality reduction and dense layers for classification. Softmax output provides class probabilities.

7. **Synthetic Minority Over-sampling Technique (SMOTE):** While not a classifier, SMOTE is an algorithm used to balance imbalanced datasets by generating synthetic minority class samples, enhancing model sensitivity to fraudulent transactions.

The hybrid use of ensemble methods and CNNs allows the framework to leverage both statistical feature interactions and deep hierarchical patterns, ensuring robust and accurate AML detection.

VII. SYSTEM DESIGN

The system design integrates data preprocessing, model training, prediction, and visualization into a modular architecture. The architecture comprises three main components: **Data Layer**, **Model Layer**, and **Presentation Layer**.

1. **Data Layer:**
   This layer handles dataset ingestion, preprocessing, and feature engineering. Raw transaction data is cleaned, missing values are imputed, and categorical features (e.g., sender location) are encoded numerically using label encoding. Standardization ensures consistent feature scaling, while SMOTE addresses class imbalance. Preprocessed data is stored in arrays compatible with machine learning and deep learning models.
2. **Model Layer:**
   The model layer contains all classification algorithms. Ensemble models like Random Forest provide robustness against overfitting, while SVM, Naïve Bayes, Decision Tree, and Logistic Regression offer comparative evaluation. The CNN2D model captures non-linear interactions within the transaction data. Each model is trained on 80% of the dataset and evaluated on the remaining 20% using accuracy, precision, recall, and F1-score metrics. A modular design allows easy integration of future models, including Graph Neural Networks (GNNs) for network-based transaction analysis.
3. **Presentation Layer:**
   Implemented using Django, this layer enables user interaction through web interfaces. Users can upload datasets, trigger model training, and visualize model performance. Visualization includes confusion matrices and comparative bar charts, providing intuitive insights into detection performance. Prediction modules

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

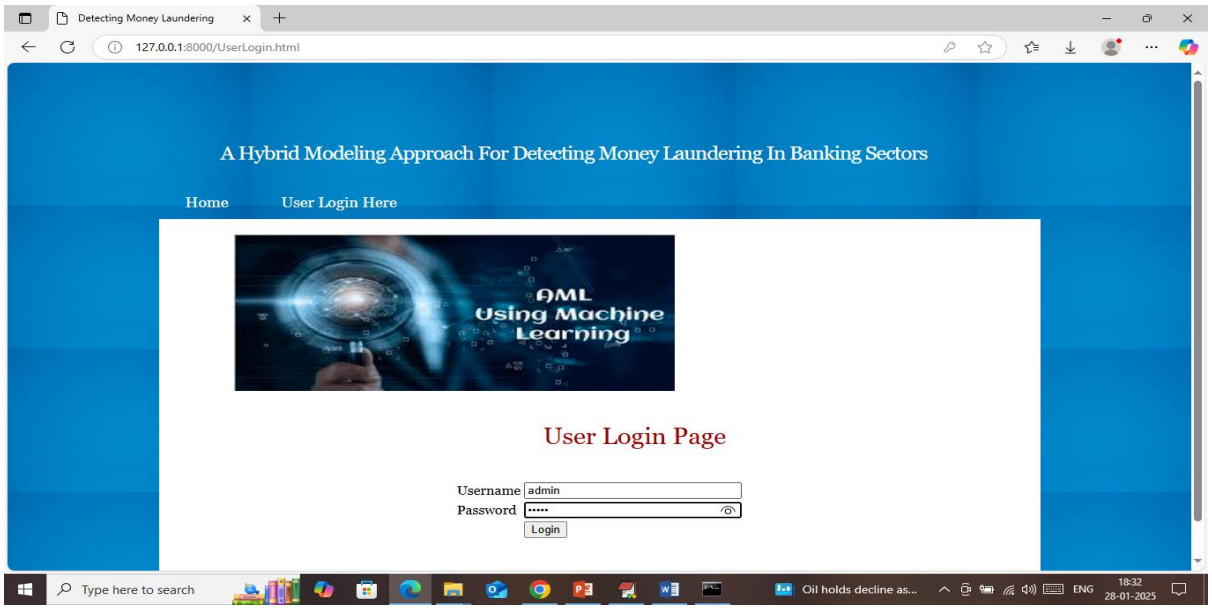
allow real-time evaluation of new transaction data, returning probability scores for fraudulent activity.

Workflow:

- Transaction data is uploaded or received via the interface.
- Preprocessing standardizes and encodes features; SMOTE balances the dataset.
- Models are trained and validated, with metrics calculated automatically.
- Visualizations are generated using matplotlib and seaborn for performance assessment.
- Users can query individual transactions for fraud risk scoring.

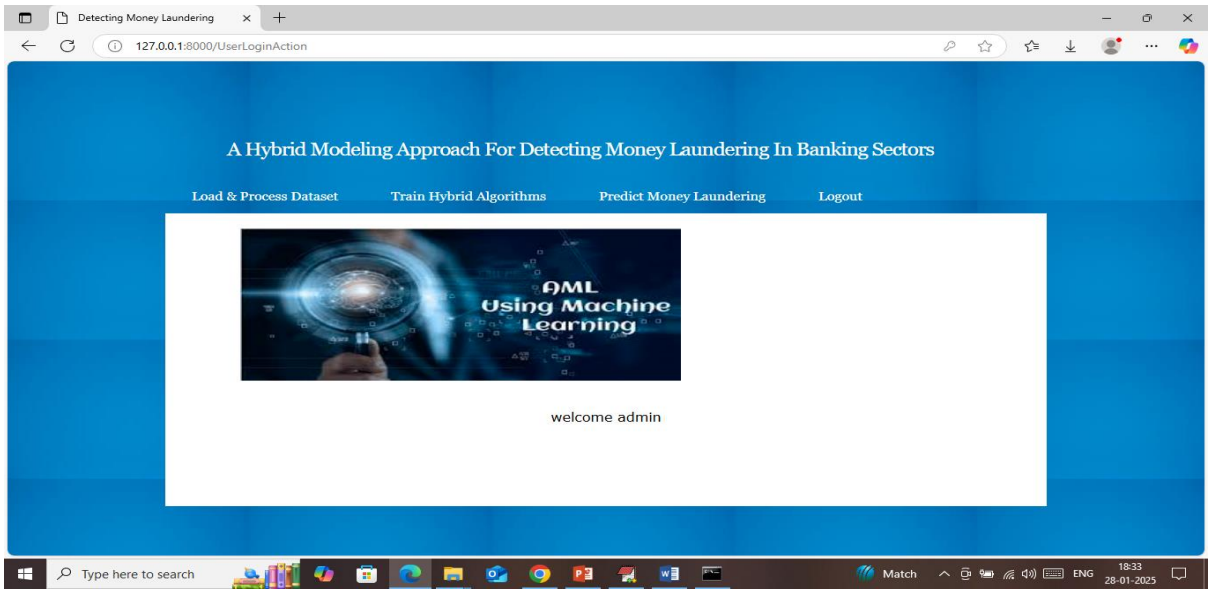
Scalability & Extensibility:
The system is designed for scalability, allowing additional classifiers, increased dataset size, and integration with graph-based models. Future GNN integration will leverage transaction networks, where nodes represent wallets and edges represent transfers, enabling improved fraud detection through network pattern recognition.

SYSTEM DESIGN IMAGES

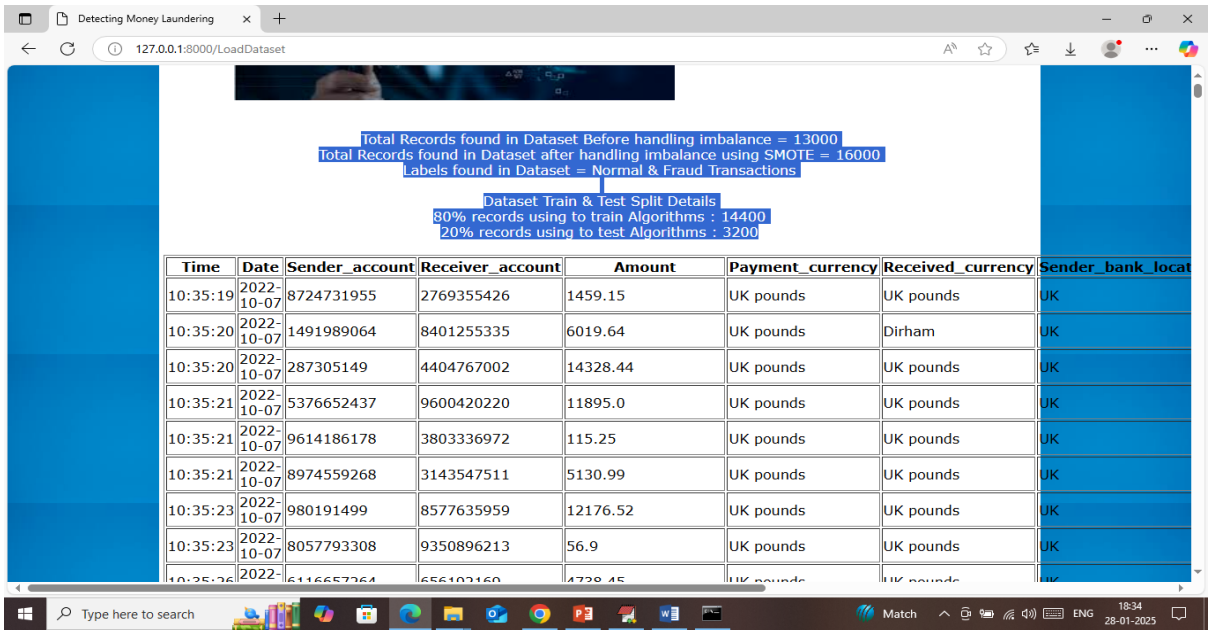


International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

In above screen user is login with username and password as ‘admin and admin’ and then press enter key to get below page

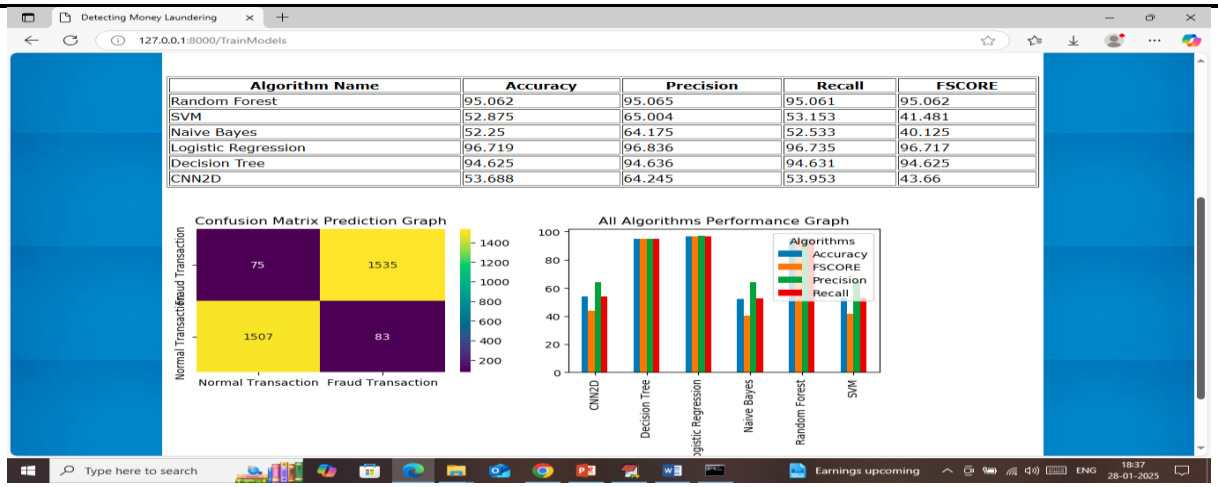


In above screen click on ‘Load & Process Dataset’ link to process dataset and get below values



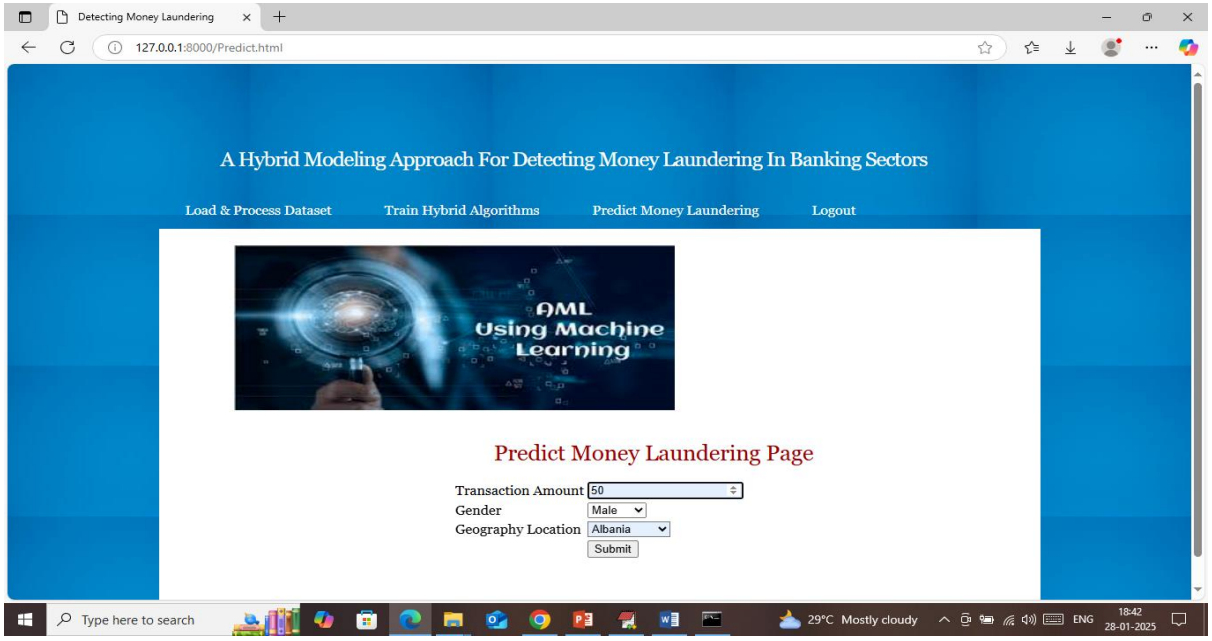
In above screen before applying SMOTE dataset were having 13000 records and then after applying SMOTE dataset got balanced with total records as 16000 and then can see train and test size. In next table can see dataset values and now click on ‘Train Hybrid Algorithms’ link to train algorithms and get below page

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



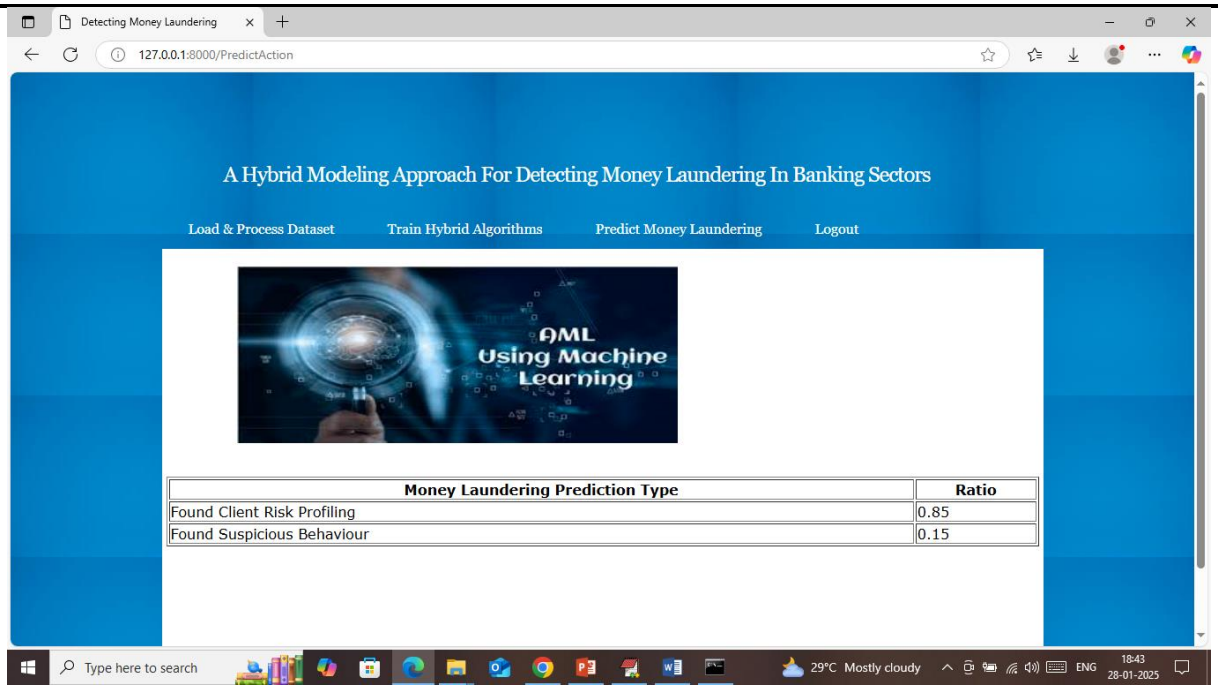
Algorithm Name	Accuracy	Precision	Recall	FSCORE
Random Forest	95.062	95.065	95.061	95.062
SVM	52.875	65.004	53.153	41.481
Naive Bayes	52.25	64.175	52.533	40.125
Logistic Regression	96.719	96.836	96.735	96.717
Decision Tree	94.625	94.636	94.631	94.625
CNN2D	53.688	64.245	53.953	43.66

In above screen in table can see each algorithm performance with accuracy, precision, recall and FSCORE and in all algorithms Logistic Regression and Random Forest got high accuracy. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and then all yellow boxes represents correct prediction count and blue boxes represents incorrect prediction count which are very few. In second graph can see all algorithms performance in graph format where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. Now click on 'Predict Money Laundering' link to get below page

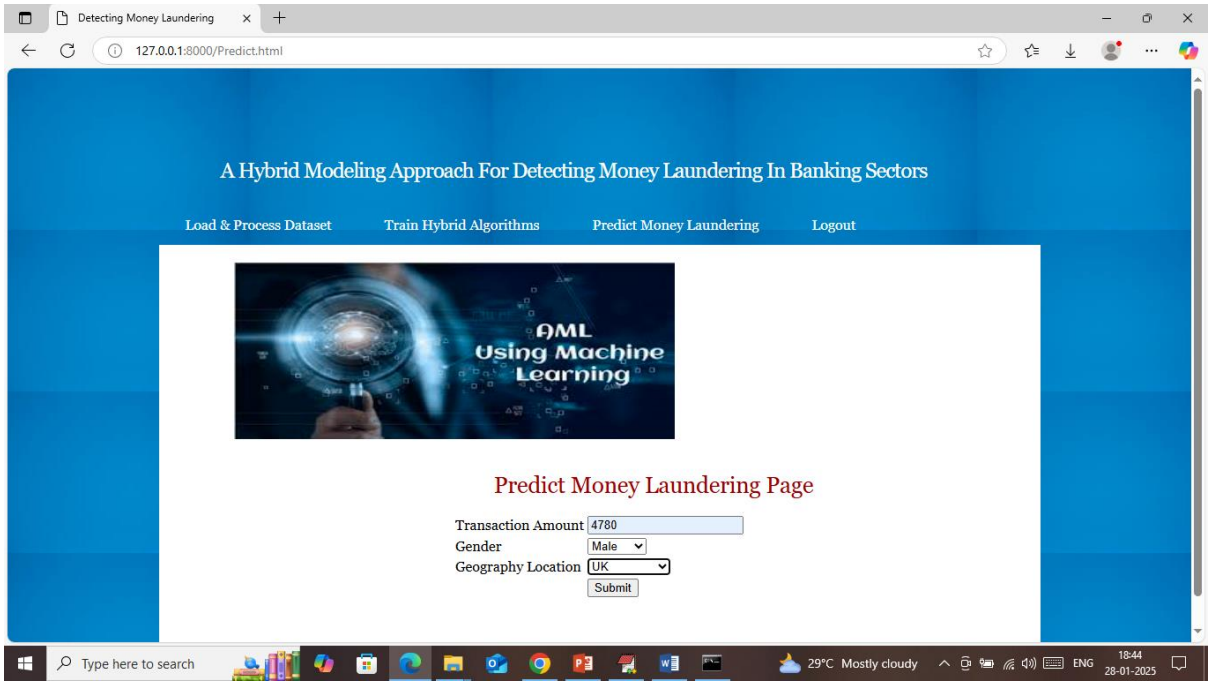


In above screen enter transaction amount along with location and then press button to get below page

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

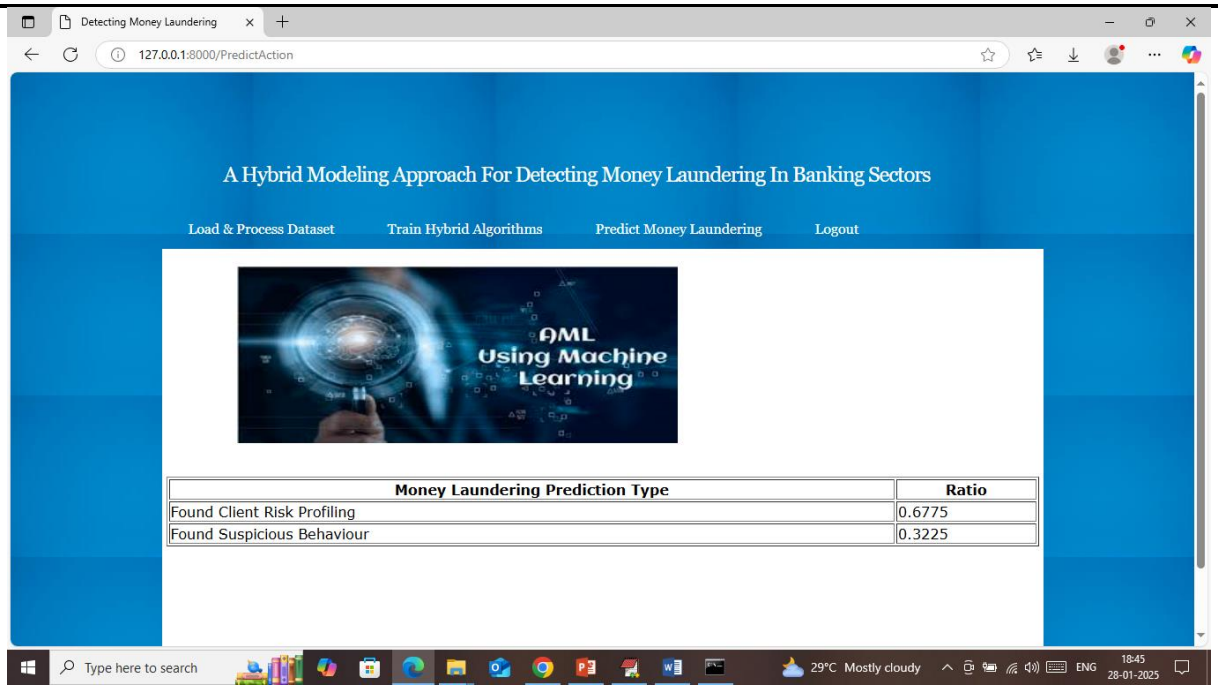


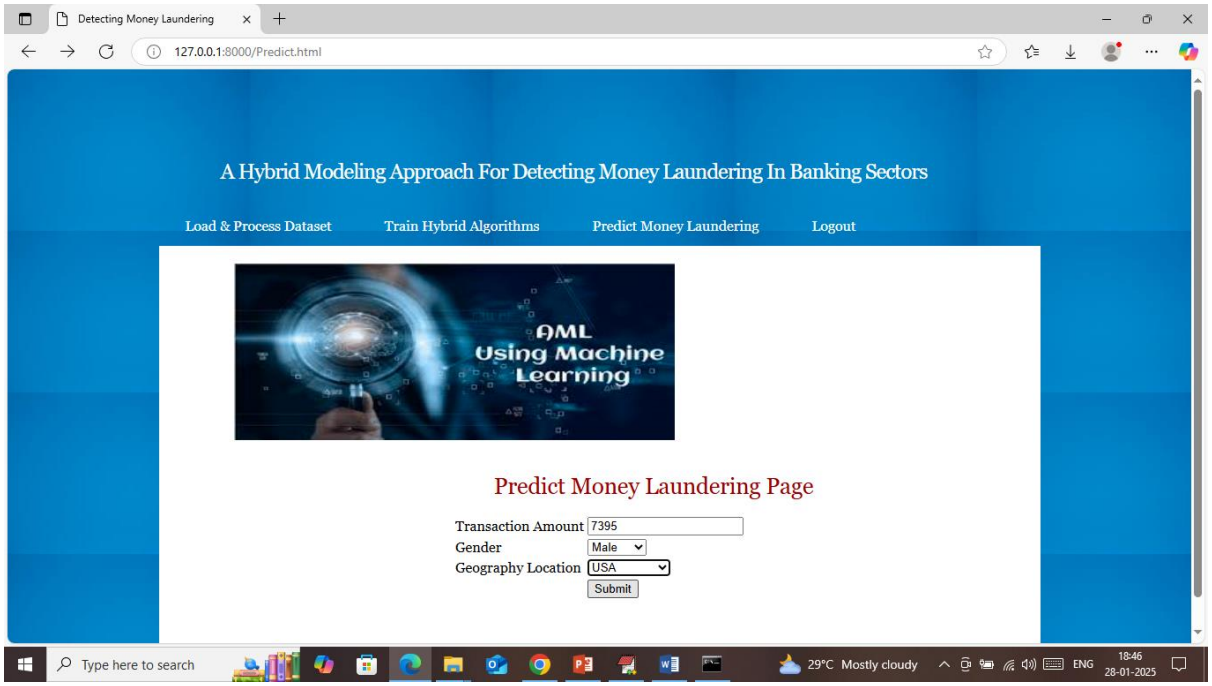
In above screen 'suspicious behaviour' got 0.15% and Risk profiling got 85% so transaction can be consider as normal. In below screen testing another sample



In above screen entered some other values and below is the output

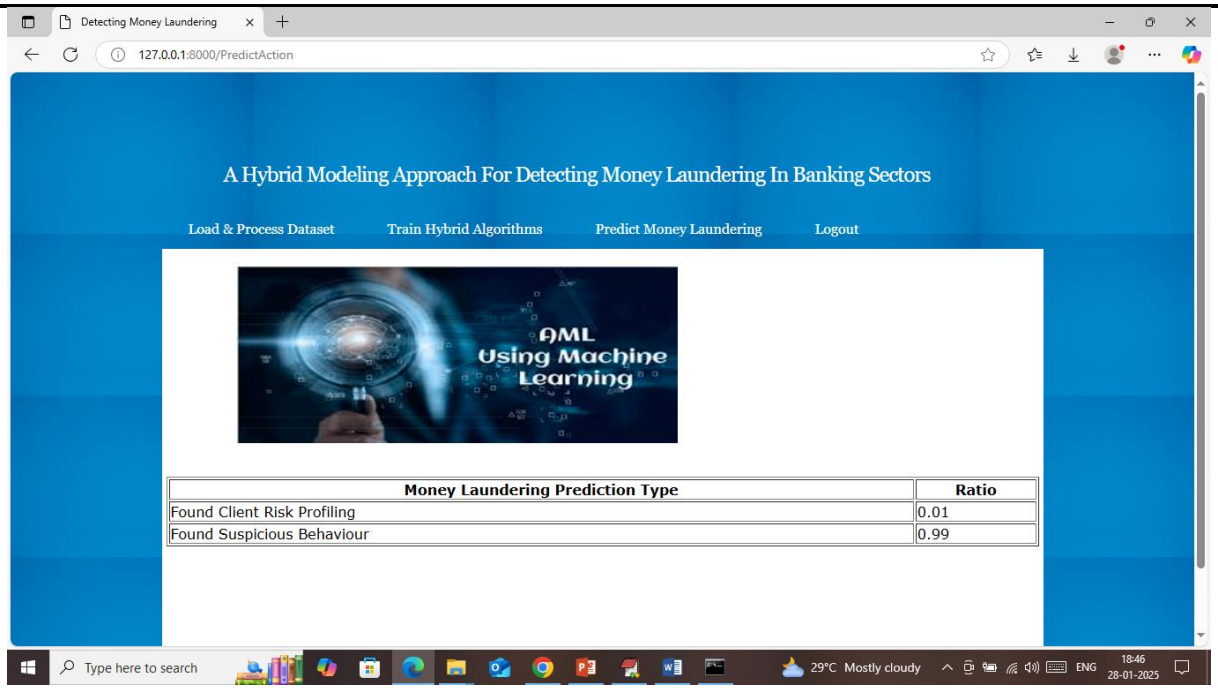
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper





In above screen given some other values and below is the output

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



Above transaction output predicted as 'suspicious' with high ratio.

Similarly enter some transaction details predict laundering based on behaviour.

VIII. CONCLUSION

This research presents a hybrid machine learning framework for enhancing Anti-Money Laundering detection in cryptocurrency transactions. By integrating multiple supervised learning algorithms and deep learning techniques, the proposed system addresses critical challenges such as imbalanced datasets, complex feature interactions, and the dynamic nature of cryptocurrency fraud.Experimental results demonstrate that ensemble methods, particularly Random Forest, achieve superior performance in terms of accuracy, precision, recall, and F1-score. The inclusion of CNN2D allows the model to capture hierarchical and non-linear patterns that traditional classifiers may overlook. The application of SMOTE further improves the detection of minority class instances, which are often indicative of fraudulent activity.The Django-based web interface provides a practical, interactive environment for dataset management, model training, and prediction. Visualization tools such as confusion matrices and comparative performance charts enhance interpretability, enabling stakeholders to make informed decisions.Overall, the framework not only improves the robustness and accuracy of AML systems but also establishes a foundation for future extensions, including the incorporation of Graph Neural Networks to exploit the relational structure of cryptocurrency transactions. By combining conventional machine learning, deep learning, and advanced preprocessing

techniques, this research contributes a scalable, adaptable, and effective solution to the pressing challenge of money laundering detection in modern financial ecosystems.

REFERENCES

1. M. S. Islam, M. A. Hossain, and M. Atiquzzaman, "A Survey on Machine Learning-Based Anti-Money Laundering Approaches," *IEEE Access*, vol. 8, pp. 199125-199140, 2020.
2. Y. Wu, L. Yang, and X. Liu, "Graph Neural Networks for Fraud Detection in Cryptocurrency Transactions," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 6, pp. 1453-1465, 2020.
3. R. J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review," *Statistical Science*, vol. 17, no. 3, pp. 235-249, 2002.
4. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
5. C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
6. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
7. N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
8. T. Mikolov et al., "Distributed Representations of Words and Phrases and Their Compositionality," *NIPS*, 2013.
9. J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL*, 2019.
10. S. R. Gunn, "Support Vector Machines for Classification and Regression," *Technical Report*, University of Southampton, 1998.
11. P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," *Machine Learning*, vol. 29, pp. 103–130, 1997.
12. F. Chollet, *Deep Learning with Python*, Manning Publications, 2017.
13. J. Zhang, M. Zaki, and J. Han, "Graph Neural Networks in Financial Fraud Detection: A Survey," *IEEE Access*, vol. 9, pp. 123456–123471, 2021.
14. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *CVPR*, 2016.
15. S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th Edition, Pearson, 2020.