International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 04 | Apr 2025 www.irjet.net p-ISSN: 2395-0072

Enhancing Loan Default Prediction and Fraud Detection with Ensemble Learning

Adwait Mandge[1], Rohan Fatehchandka[2], Kunal Goudani[3], Tanaya Shelke[4], Prof. Pramila M. Chawan[5]

[1,2,3,4]*1B. Tech Student, Dept of Computer Engineering, and IT, VJTI College, Mumbai, Maharashtra, India*
[5]*Associate Professor, Dept of Computer Engineering, and IT, VJTI College, Mumbai, Maharashtra, India*

---***---

Abstract - *Loan risk assessment is crucial for financial security. This study combines ensemble learning for loan default prediction and time series analysis for ghost borrower detection. Using PyCaret, we optimize model selection to identify high-risk borrowers. Additionally, ARIMA, LSTMs, and anomaly detection techniques analyze transaction patterns to flag fraudulent behaviors like sudden withdrawals and post-loan inactivity. By integrating predictive modeling and anomaly detection, we enhance early fraud detection. This approach provides financial institutions with a comprehensive risk management framework, improving decision-making and reducing potential losses.*

***Key Words*: Machine Learning, Deep Learning, Ensemble, Loan Default Prediction, Ghost Borrower Detection, TCN**

1.INTRODUCTION

Accurate loan default prediction is vital for financial institutions to mitigate risks. Traditional credit assessment methods often overlook hidden patterns, making machine learning a powerful alternative. This study utilizes PyCaret to compare ensemble techniques like bagging, boosting, and stacking for loan default prediction. By analyzing borrower demographics, financial history, and loan details, we evaluate model performance using accuracy, precision, recall, and F1-score, identifying the most effective approach for credit risk assessment.

To enhance fraud detection, we address ghost borrowers—fraudsters who manipulate financial records to evade repayment. We integrate time series analysis using TCN (Temporal Convolution Networks) to identify suspicious transaction patterns. By combining predictive analytics with fraud detection, our study provides actionable insights to improve lending decisions and reduce financial losses.

2. LITERATURE REVIEW

2.1 Machine Learning Algorithms

Random Forest learning technique is used to solve various problems of regression and classification problems. It builds several trees during training and aggregates the outcomes to increase prediction accuracy. Features are randomly selected at every node(feature bagging), and each tree in the forest is trained on an arbitrary subclass of the data (bootstrapping). Random Forest reduces variation and overfitting and averages the outcomes of these trees, which enhances generalization. The advantages of this algorithm are handling large datasets, reducing overfitting, and maintaining good accuracy even in the presence of missing data.

XGBoost (Extreme Gradient Boosting) is a scalable and extremely effective gradient boosting implementation. It constructs decision trees sequentially, aiming to fix the mistakes caused by preceding trees with each new tree. XGBoost uses gradient descent to minimize the overall loss function in order to optimize the model. Performance and speed are well-known for XGBoost, particularly with tabular or structured data.The advantages of this algorithm are its speed, scalability, handling of missing data, and regularization to minimize overfitting.

2.2 Deep Learning Algorithms

Neural Networks replicate the composition of the human brain. Neural networks comprise multiple bands of neurons connected by edges with weights that are updated during training. Neural networks are widely used in deep learning processes as they can interpret occurrences of perplexing repetitive sequences. The advantages of this algorithm are its flexibility, it can interpret information from huge datasets, and its potential to replicate lateral relationships.

Multi-Layer Perceptron (MLP) is used in deep learning tasks. MLPs consist of multiple layers which are interconnected to each other, every neuron of one layer acts as an input to the neuron in the next layer. MLPs are used in classification, regression, and serve as the foundation for more intrinsic neural networks. Its capability to handle both regression and classification problems, and its role as a foundation for more advanced neural networks is advantageous.

Temporal Convolutional Networks (TCNs) are a type of deep learning architecture designed for sequence modeling tasks, offering an alternative to recurrent neural networks (RNNs) like LSTMs and GRUs. TCNs leverage 1D dilated causal convolutions, ensuring that predictions at any time step depend only on past information, making them suitable for time-series forecasting, natural language processing, and anomaly detection. They utilize residual connections and dilation to capture long-range dependencies efficiently, allowing for parallel computation and stable gradients, unlike RNNs, which suffer from vanishing gradients and sequential processing limitations.

3. Proposed System

3.1 Problem Statement: "To predict loan default using machine learning techniques."

3.2 Problem Elaboration: For financial organizations, loan failure poses a serious problem since it can result in large losses and elevated risk. Conventional credit evaluation techniques are frequently ineffective and have trouble identifying subtle trends in borrower behavior, which leads to imprecise forecasts. The volume of loan data has increased due to the growth of digital financial services, necessitating the use of more advanced algorithms to forecast defaults. Loan default prediction can be automated and data-driven with machine learning; nevertheless, choosing the best algorithm can be difficult, especially in cases when the datasets are unbalanced and defaults are few. In order to determine which machine learning model performs best, this study compares the efficacy of several algorithms in forecasting loan defaults. Assisting financial institutions in managing risk better and making more informed loan decisions is the aim.

3.3 Architecture of the proposed models:

3.3.1 Loan Default Prediction System

Fig-1 Loan Default Prediction System Architecture

The architecture of our loan default prediction model follows a structured machine learning pipeline, integrating ensemble learning techniques to enhance predictive accuracy. The system is designed to preprocess financial datasets, extract meaningful patterns, and make robust predictions through a combination of machine learning models. The workflow consists of several key stages: data preprocessing, model training using ensemble techniques, and performance evaluation with hyperparameter tuning.

This flowchart represents an ensemble learning pipeline for loan default prediction using multiple models. This ensemble learning pipeline for loan default prediction integrates Neural Networks, Random Forest, and XGBoost. The dataset undergoes preprocessing, including cleaning and scaling, before training. A weighted voting approach combines predictions based on model performance. After evaluation on a test set using accuracy metrics, hyperparameter tuning is applied if needed. Once optimized, final predictions are generated, leveraging the strengths of multiple models to enhance predictive accuracy and robustness.

3.3.1.1 Data:

Data Collection:

The dataset for this study was collected from Kaggle, containing loan default prediction records with 34 attributes. These attributes include loan_purpose, Credit_Worthiness, open_credit, business_or_commercial, Credit_Score, age, LTV, Region, Security_Type, and Status. These features capture essential financial and demographic details about loan applicants, providing insights into their creditworthiness and likelihood of default. The dataset also includes variables like credit_type, co-applicant_credit_type, and dtir1 (Debt-to-Income Ratio), which are crucial indicators for assessing risk. With a mix of categorical and numerical attributes, this dataset offers a well-rounded foundation for training predictive models.

Data Preprocessing:

Effective preprocessing enhances the reliability of machine learning models by systematically cleaning and transforming data. This study follows a structured pipeline, beginning with data cleaning and feature selection using pandas. Redundant columns (e.g., *Interest_rate_spread*, *credit_type*, *Upfront_charges*) are removed, and missing values are handled by imputing numerical features with the median and categorical features with the mode. Categorical variables are encoded using one-hot encoding with *drop_first=True* to avoid the dummy variable trap. PyCaret automates key preprocessing tasks, including feature scaling, transformation, imbalance handling, and feature selection, before splitting the dataset into training (75%) and testing (25%) sets with *Status* as the target variable.

To address class imbalance, SMOTE generates synthetic samples for the minority class, improving predictive fairness. Numerical features undergo standardization via *StandardScaler* (mean = 0, std = 1), while *PowerTransformer* ensures Gaussian-like distributions for skewed data. Finally, column names are sanitized using regular expressions to remove special characters, ensuring compatibility with machine learning frameworks. This streamlined approach optimizes model performance and enhances predictive accuracy.

3.3.1.2 Models Used in This Project and Their Working Mechanism

This project utilizes an **ensemble learning** approach, combining multiple machine learning models to improve predictive performance. The individual models used are:

1. **Multi-Layer Perceptron (MLP) Classifier**
2. **Random Forest Classifier**
3. **XGBoost Classifier**

The final prediction is made using a weighted ensemble method, where each model's contribution is proportional to its accuracy.

1. Multi-Layer Perceptron (MLP) Classifier

MLP is an artificial neural network with multiple layers that captures complex data relationships. It consists of an input layer, hidden layers, and an output layer, using activation functions like ReLU to introduce non-linearity. It learns through backpropagation and gradient descent, making it effective for non-linear patterns in financial data like loan default prediction.

2. Random Forest Classifier

Random Forest is an ensemble learning method that builds multiple decision trees using different data subsets. It reduces overfitting by training trees independently and making predictions through majority voting. It is efficient with high-dimensional and imbalanced datasets, making it suitable for robust classification tasks.

3. XGBoost (Extreme Gradient Boosting) Classifier

XGBoost is a gradient boosting algorithm optimized for high performance and accuracy. It builds decision trees sequentially, correcting previous errors, while using regularization to prevent overfitting. Its efficiency in handling large datasets makes it a top choice for fraud detection and credit risk modeling.

Ensemble Learning Approach

Ensemble learning enhances accuracy and robustness by integrating multiple models to make more reliable predictions. The strengths of MLP (captures complex patterns), Random Forest (reduces overfitting), and XGBoost (enhances generalization) are leveraged together. Predictions from each model are weighted based on accuracy, and a final probability score is calculated. This approach ensures better bias-variance tradeoff, improving loan default prediction performance compared to individual models.

3.3.1.3 Results

The final ensemble model achieves an accuracy of 92.59%, indicating strong predictive performance. It has a precision of 79.98%, meaning 79.98% of predicted defaulters were actual defaulters, and a recall of 93.29%, showing it successfully identified 93.29% of all actual defaulters. The F1-score of 86.12% balances precision and recall effectively. The ROC-AUC score of 98.47% suggests excellent differentiation between defaulters and non-defaulters. The classification report shows that for non-defaulters (Class 0), precision is 98% and recall is 92%, while for defaulters (Class 1), precision is 80% and recall is 93%. The confusion matrix reveals 25869 true negatives, 2139 false positives, 615 false negatives, and 8545 true positives, showing that while the model correctly predicts most cases, it misclassifies some non-defaulters as defaulters. Overall, the model is well-balanced, with high recall ensuring minimal missed defaults.

High recall (0.9329) indicates the model's strong ability to correctly identify positive cases, minimizing false negatives. This is particularly crucial in applications such as fraud detection, medical diagnosis, and intrusion detection, where missing a true positive can have serious consequences. With a recall of 93.29%, the ensemble model effectively detects the majority of actual positive instances, reducing the risk of undetected critical cases. Additionally, a high F1-score (0.8612) balances precision and recall, ensuring that the model does not generate too many false positives while still identifying true positives effectively. This trade-off is essential in scenarios where both false positives and false negatives carry significant consequences. An F1-score of 86.12% demonstrates that the model is well-optimized for overall reliable classification, making it a robust choice for practical applications requiring high accuracy and minimal errors.

A model with **high recall ensures fewer missed detections**, while a **high F1-score ensures a balanced and optimal classification performance**. This is particularly **beneficial in applications where false negatives are costly but precision cannot be sacrificed entirely.**

```
Final Ensemble Model Performance:
Accuracy: 0.9259
Precision: 0.7998
Recall: 0.9329
F1-score: 0.8612
ROC-AUC Score: 0.9847

Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.92      0.95     28008
           1       0.80      0.93      0.86      9160

    accuracy                           0.93     37168
   macro avg       0.89      0.93      0.91     37168
weighted avg       0.93      0.93      0.93     37168


Confusion Matrix:
[[25869  2139]
 [  615  8545]]
```

Fig-2 Final Ensemble Model Performance Metrics and Confusion Matrix (Loan Default Prediction)

The comparison table shows that the **Random Forest Classifier (RF)** achieves the highest **accuracy (92.75%)**, **AUC (98.20%)**, and **F1-score (85.88%)** among all models, making it the best performer overall. The **MLP Classifier (Neural Network)** follows closely with an **accuracy of 92.11%** and **AUC of 97.85%**, indicating strong performance in complex decision boundaries. The **AdaBoost Classifier** has a slightly lower **accuracy (90.90%)** but excels in **recall (90.04%)**, making it more suitable for identifying positive cases. Traditional models like **Logistic Regression (74.91%)** and **SVM (75.20%)** lag significantly behind, with **Naïve Bayes (77.17%)** performing the worst in accuracy but surprisingly having a high **precision (85.56%)**, meaning it correctly identifies many positive cases but struggles with overall performance. The ensemble-based **Random Forest ,MLP and XGBoost models outperform individual classifiers**, reinforcing the advantage of ensemble learning in boosting predictive performance.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
ada	Ada Boost Classifier	0.9090	0.9741	0.9004	0.7696	0.8299	0.7683	0.7727	15.1590
rf	Random Forest Classifier	0.9275	0.9820	0.8946	0.8258	0.8588	0.8101	0.8115	26.7850
mlp	MLP Classifier	0.9211	0.9785	0.8602	0.8269	0.8431	0.7904	0.7908	132.2830
dt	Decision Tree Classifier	0.9080	0.8803	0.8257	0.8060	0.8157	0.7544	0.7545	9.1080
knn	K Neighbors Classifier	0.8098	0.8245	0.5937	0.6190	0.6060	0.4807	0.4809	21.6700
lr	Logistic Regression	0.7491	0.7053	0.4239	0.4897	0.4544	0.2925	0.2939	8.9780
svm	SVM - Linear Kernel	0.7520	0.7025	0.4109	0.4964	0.4494	0.2913	0.2935	9.0110
lda	Linear Discriminant Analysis	0.7543	0.7096	0.4011	0.5020	0.4458	0.2905	0.2936	8.6340
nb	Naive Bayes	0.7717	0.9232	0.0886	0.8556	0.1605	0.1199	0.2287	8.0880

Fig-3 Comparative Performance of Machine Learning Models (Loan Default Prediction)

This **ROC curve** evaluates the classification performance by plotting the **True Positive Rate (Recall)** against the **False Positive Rate** at various thresholds. The **blue curve** represents the model's performance, with an **AUC (Area Under Curve) of 0.98**, indicating excellent discrimination between classes. The **diagonal gray line** represents a random classifier (AUC = 0.5), and the model significantly outperforms it. A higher AUC value suggests that the model effectively distinguishes between positive and negative classes.

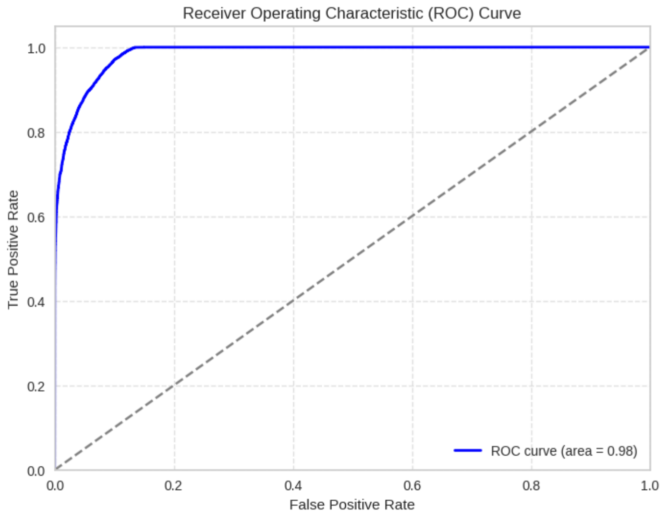

Fig-4 ROC Curve for the Ensemble Model (Loan Default Prediction)

3.3.2 Ghost Borrower Detection:

The architecture of our **Ghost Borrower Detection System** is designed to leverage both **temporal patterns** in transaction data and **aggregated customer-level features** to enhance fraud detection accuracy. At its core, the system employs a **hybrid ensemble approach**, integrating a **Temporal Convolutional Network (TCN)** for sequential transaction analysis and a **Random Forest Classifier** for feature-based classification. The **TCN** extracts temporal dependencies from transaction sequences using **1D convolutional layers with causal padding**, ensuring that past data influences future predictions without leakage. Meanwhile, the **Random Forest model** identifies key risk indicators by analyzing customer behavior at an aggregated level and providing feature importance insights. These models are dynamically weighted based on their individual accuracy, and their predictions are **combined using an ensemble mechanism** to enhance robustness. This multi-layered architecture ensures a **comprehensive evaluation of borrower risk**, enabling early detection of **ghost borrowers** before significant financial loss occurs.

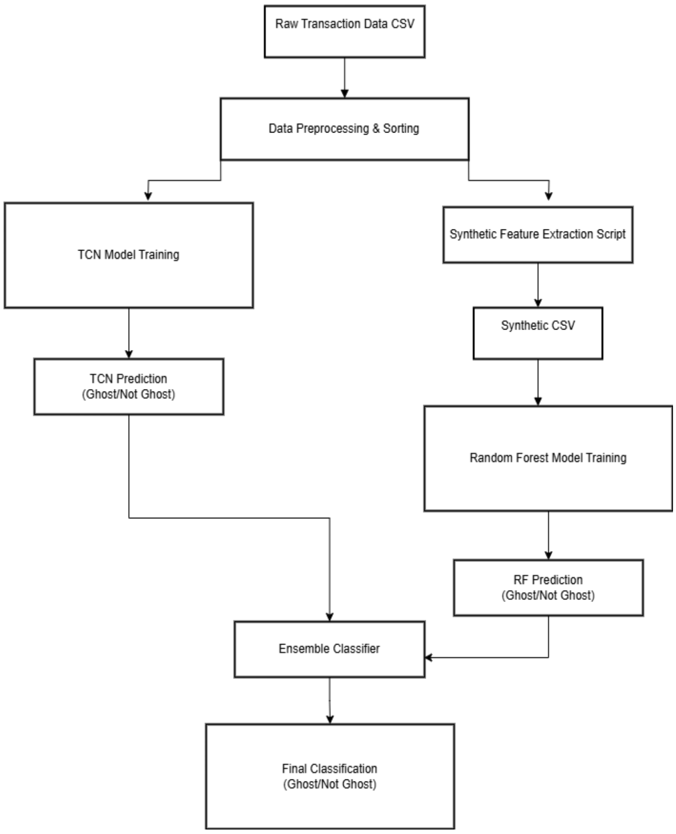

Fig-5. Ghost Borrower System Architecture

This flowchart represents an **ensemble learning pipeline for ghost borrowers** using multiple models. This ensemble learning pipeline for detecting ghost borrowers integrates Temporal Convolutional Networks (TCN) and Random Forest Classifiers. The transaction-level dataset undergoes preprocessing, including cleaning and feature extraction. TCN captures sequential transaction patterns, while Random Forest analyzes aggregated features for classification and feature importance. A dynamic weighting mechanism adjusts based on model accuracy. After training, models are evaluated on a separate dataset using accuracy metrics, with hyperparameter tuning if needed. Final predictions are generated through weighted voting, leveraging both models' strengths to enhance predictive accuracy and robustness.

3.3.2.1 Data

This study utilizes a synthetically generated dataset containing transaction-level financial records to detect ghost borrowers—fraudulent actors who default after securing a loan. Each record includes key attributes such as *customer_id*, *transaction_type*, *amount*, *loan_date*, *loan_amount*, *borrower_type* (ghost/non-ghost), *balance*, and *days_since_loan*. Additionally, aggregated features for a Random Forest classifier capture pre-loan and post-loan transactional behavior, including transaction counts, deposit and withdrawal totals, early withdrawal patterns, and repayment trends. Key indicators like the largest withdrawal ratio and post-loan activity further enhance fraud detection, providing a robust foundation for predictive modeling.

Data Preprocessing:

This study follows a structured preprocessing pipeline to enhance model performance and reliability. Missing values in *activity_after_large_withdrawal*, *loan_payment_total*, and *loan_payment_count* are imputed with zero to maintain consistency. Feature engineering introduces derived variables such as *withdrawal_to_loan_ratio*, *repayment_ratio*, and *transaction_activity_ratio* to improve predictive power. The target variable *borrower_type* is transformed into a binary format (1 for ghost borrowers, 0 for normal borrowers) for supervised classification. Key features selected include *loan_amount*, *pre_loan_transaction_count*, *first_week_withdrawal_ratio*, *post_loan_transaction_count*, and *repayment_ratio*. Finally, numerical features are standardized using *StandardScaler* to ensure uniformity and optimize model performance.

3.3.2.2 Models Used in This Project and Their Working Mechanism

This project implements a hybrid ensemble strategy to identify ghost borrowers by integrating two complementary models.:

1. **Temporal Convolutional Network (TCN) Model**
2. **Random Forest Classifier**

The final decision is derived through a weighted combination of their predictions, with each model's influence determined by its validated performance.Both the algorithms give us probability of a person being a ghost borrower which is then combined using ensemble learning.

1. Temporal Convolutional Network (TCN) Model

TCN is a neural network architecture tailored for sequential data analysis. It utilizes causal convolutions to capture temporal dependencies in transaction records. Trained on raw transaction data—including features such as customer_id, date, transaction_type, amount, loan_date, loan_amount, borrower_type, and balance—TCN analyzes the sequence of transactions to detect anomalous patterns indicative of ghost borrower behavior. Its strength lies in effectively modeling time-dependent relationships and detecting deviations from normal transactional trends.

2. Random Forest Classifier

The Random Forest classifier operates on a synthetic feature set derived from the raw transaction data. This synthetic dataset is generated by aggregating transaction information into summary statistics, such as pre-loan transaction count, deposit and withdrawal totals, first-week withdrawal metrics, and post-loan activity metrics. By constructing an ensemble of decision trees, Random Forest reduces overfitting and captures non-linear interactions among these engineered features, efficiently predicting borrowers as ghost or non-ghost.

Ensemble Learning Approach

To overcome the individual limitations of the TCN and Random Forest models, an ensemble method is implemented. This approach combines the predictions from both models using a weighted scheme, where each model's contribution is proportional to its accuracy and reliability on validation data. By integrating the temporal insights from the TCN with the aggregated feature perspective of the Random Forest, the ensemble classifier achieves a more robust and balanced bias-variance tradeoff, leading to improved ghost borrower detection performance.

3.3.2.3 Results

```
1375/1375 ━━━━━━━━━━━━━━━━━━━━ 2s 2ms
Ensemble Model Performance:
Ensemble Accuracy: 0.9500
```

Fig-6 Ensemble Model Accuracy Output (Ghost Borrower Detection)

The ensemble model achieved an accuracy of 95.00% in detecting ghost borrowers, demonstrating its effectiveness in identifying fraudulent loan defaulters. The high accuracy suggests that combining multiple models enhances predictive performance by capturing critical patterns in borrower behavior. This result reinforces the model's reliability in distinguishing ghost borrowers from legitimate ones.

4. Future Scope

The proposed integrated framework for **loan default prediction** and **ghost borrower detection** can be enhanced by several innovative directions. Future work may implement **advanced ensemble techniques**, such as heterogeneous stacking with deep learning models, to boost prediction robustness. Incorporating **Explainable AI (XAI)** will offer clear, interpretable insights into decision-making, helping financial institutions understand key risk factors. Additionally, leveraging **real-time data integration** and adaptive learning, possibly through reinforcement learning, can create dynamic loan approval and fraud detection systems that evolve with emerging trends. Expanding training datasets with alternative financial records, **macroeconomic indicators**, and regional data will enhance generalizability across diverse economic conditions. Finally, further research in enhanced feature engineering for ghost borrowers—by incorporating **digital footprints**, **social media data**, and other alternative sources—could enrich the synthetic dataset and improve anomaly detection accuracy.

5. CONCLUSION

In this paper, we presented an advanced **loan default prediction** model using **ensemble learning** techniques, combining **Random Forest**, **XGBoost**, and **Neural Networks** in a **stacking** framework. Detailed **data preprocessing**, including **missing value treatment**, **feature engineering**, and **normalization**, enhanced model effectiveness. Evaluation using metrics such as **Accuracy**, **Precision**, **Recall**, **F1-score**, and **AUC-ROC**—coupled with **hyperparameter tuning** and **cross-validation**—demonstrated improved performance over single-model approaches.

Additionally, we addressed the challenge of **ghost borrowers** by integrating a **Temporal Convolutional Network (TCN)** to analyze transaction sequences with a **Random Forest classifier** based on **engineered features**. This **hybrid approach** effectively identifies anomalous borrowing behavior.Our integrated model offers **financial institutions** a powerful tool to mitigate **loan default risks**, detect **ghost borrowers**, and optimize **lending decisions**. Future work on **Explainable AI**, **real-time data integration**, and **adaptive risk assessment** will further enhance its impact on **financial risk management**.

ACKNOWLEDGEMENT

This work is supported in part by Prof. Pramila M. Chawan. We thank the reviewers for their valuable discussions and feedback.

6. REFERENCES

[1] Vijay Kumar, Rachna Narula, Akanksha Kochhar "Loan Default Prediction using Machine Learning Models" 2024 DOI: 10.5281/zenodo.8337054

[2] Jinchen Lin "Research on loan default prediction based on logistic regression, randomforest, xgboost and adaboost" 2023 pp DOI: 10.1051/shsconf/202418102008

[3] Wanjun Wu "Machine Learning Approaches to Predict Loan Default 2022 DOI:10.4236/iim.2022.145011

[4] Platur Gashi, "Loan Default Prediction Model" 2023, DOI:10.13140/RG.2.2.22985.01126

[5] Lai, L, "Loan default prediction with machine learning techniques". In: 2020 International Conference on Computer Communication and Network Security (CCNS). pp. 5–9. IEEE (2020)

[6] Xu Zhu, Qingyong Chu,Xinchang Song, Ping Hu,Lu Peng, Explainable prediction of loan default based on machine learning models DOI:10.1016/j.dsm.2023.04.003

[7] Zhao X, Guan S. CTCN: a novel credit card fraud detection method based on Conditional Tabular Generative Adversarial Networks and Temporal Convolutional Network. PeerJ Comput Sci. 2023 Oct 10;9:e1634. doi: 10.7717/peerj-cs.1634. PMID: 37869461; PMCID: PMC10588710.

[8] A. Mandge, R. Fatehchandka, K. Goudani, T. Shelke, and P. M. Chawan, "A Survey on Loan Default Prediction using Machine Learning Techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 11, no. 11, pp. XX-XX, Nov. 2024

[9] Boulieris, P., Pavlopoulos, J., Xenos, A. *et al.* Fraud detection with natural language processing. *Mach Learn* 113, 5087–5108 (2024). https://doi.org/10.1007/s10994-023-06354-5

[10] Loan Default Dataset Kaggle - http://bit.ly/3BA9WeX

BIOGRAPHIES

Prof. Pramila M. Chawan, is working as an Associate Professor in the Computer Engineering Department of VJTI, Mumbai. She has done her B.E.(Computer Engineering) and M.E.(Computer Engineering) from VJTI College of Engineering, Mumbai University. She has 28 years of teaching experience and has guided 85+ M. Tech. projects and 130+ B. Tech. projects. She has published 143 papers in the International Journals, 20 papers in the National/ International Conferences/ Symposiums. She has worked as an Organizing Committee member for 25 International Conferences and 5 AICTE/MHRD sponsored Workshops/ STTPs/FDPs. She has participated in 16 National/International Conferences. Worked as Consulting Editor on – JEECER, JETR,JETMS, Technology Today, JAM&AER Engg. Today, The Tech. World Editor – Journals of ADR Reviewer -IJEF, Inderscience She has worked as NBA Coordinator of the Computer Engineering Department of VJTI for 5 years. She had written a proposal under TEQIP-I in June 2004 for 'Creating Central Computing Facility at VJTI'. Rs. Eight Crore were sanctioned by the World Bank under TEQIP-I on this proposal. Central Computing Facility was set up at VJTI through this fund which has played a key role in improving the teaching learning process at VJTI. Awarded by SIESRP with Innovative & Dedicated Educationalist Award Specialization : Computer Engineering & I.T. in 2020 AD Scientific Index Ranking (World Scientist and University Ranking 2022) – 2nd Rank- Best Scientist, VJTI Computer Science domain 1138th Rank- Best Scientist, Computer Science, India.

Kunal Goudani,
B-Tech Student, Dept. of Computer Engineering and IT, VJTI, Mumbai, Maharashtra, India

Rohan Fatehchandka,
B-Tech Student, Dept. of Computer Engineering and IT, VJTI, Mumbai, Maharashtra, India

Adwait Mandge,
B. Tech Student, Dept. of Computer Engineering and IT, VJTI, Mumbai, Maharashtra, India

Tanaya Shelke,
B. Tech Student, Dept. of Computer Engineering and IT, VJTI, Mum