Loan Approval Estimation Deploying Deep Learning and Blockchain Technologies

Jaya Sravani Janapareddy
Student, Computer Science and Engineering
Anurag University
Hyderabad, Telangana, India
jayasravani21@gmail.com

Aitha Sai Vishwanth
Student, Computer Science and Engineering
Anurag University
Hyderabad, Telangana, India
vishwanthreddyaitha@gmail.com

Manikanta Tagore
Student, Computer Science and Engineering
Anurag University
Hyderabad, Telangana, India
manikantatagore9@gmail.com

Varthiya Chandana
Student, Computer Science and Engineering
Anurag University
Hyderabad, Telangana, India
varthiyachandana2406@gmail.com

Dr. G. Balakrishna
Assistant Professor, Computer Science and Engineering
Anurag University
Hyderabad, Telangana, India
balakrishnacse@anurag.edu.in

Abstract*— Bank profitability relies heavily on credit lines and the interest revenue from loans, making it vital to predict loan defaulters to minimize Non-Performing Assets (NPAs) and manage risk. This study explores different approaches to loan default prediction, emphasizing the deep learning and machine learning models used in predictive analytics.The research underscores the importance of incorporating a wide range of customer characteristics into loan evaluation to efficiently detect potential defaulters and optimize lending approaches. Financial institutions are urged to adopt comprehensive strategies that transcend wealth indicators for assessing creditworthiness and reducing the likelihood of loan default.

Keywords— ***Loan approval, Deep learning, Blockchain technology, Financial services, Immutability***

I. Introduction

Cutting-edge technology like Deep Learning and Blockchain is revolutionizing loan approval processes. The collaborative project aims to enhance effectiveness and security by leveraging Deep Learning algorithms to sort through vast datasets and identify complex patterns, potentially transforming the banking sector through innovation.

Blockchain technology and machine learning to address economic challenges, notably impacting SMEs in India. It explores integrating these technologies to establish a peer-to-peer lending framework for micro-loans, with a focus on mitigating the impact of frequent lockdowns.[1]

Loans play a pivotal role in a bank's revenue, emphasizing the challenge of accurately assessing loan requests and managing associated risks. Machine learning techniques like K-Nearest Neighbours, Decision Trees, and Support Vector Classifier. It discusses credit risk, reliance on credit scores, and the need for effective machine learning methods. Human decision-making complexity in banking adds further challenge, setting the stage for exploring diverse algorithms to revolutionize loan approval. [2]

Machine learning approach to intelligently approve loans, leveraging extensive data to automate and refine the selection process. [3]

Amidst the financial crisis, loan approval practices are being reevaluated globally. This work advocates for using advanced machine learning techniques to predict approvals, considering factors like occupation and existing EMIs. The best models, Gradient Boosting and Random Forest, are integrated for optimal performance. [4]

Study predicts loan defaulters, stressing the importance of credit lines for banks. It highlights the Naïve Bayes model's effectiveness in predictive analytics. The introduction underscores the need for automated loan eligibility processes due to banking challenges.[5]

In today's banking landscape, primary revenue often relies on credit lines and loan interest. Explores predicting loan defaulters to mitigate Non-Performing Assets (NPAs). Utilizing advanced predictive analytics, especially Logistic Regression, aims to enhance precision in identifying potential defaulters and strengthen banking institutions' financial health.[6]

Blockchain technology offers revolutionary opportunities in financial inclusion by enabling cheaper prices, faster settlement times, and improved customer experiences for payments.[7]

Researchers utilize advanced computer technology and data analytics to develop robust anti-fraud measures, including sophisticated algorithms like XGBoost and deep neural networks. Their model demonstrates high accuracy in detecting online fraud, fortifying the financial industry against evolving cyber threats.[8]

II. Literature Survey

Deep learning has significantly advanced credit risk assessment in the financial industry. A novel model, utilizing an Adaptive Binarized Spiking Marine Predators Neural Network, demonstrates exceptional accuracy, reinforcing the effectiveness of sophisticated neural networks in financial decision-making. Integration with blockchain technology ensures data security.[9]

In parallel, a research highlights the usefulness of machine learning methods like Random Forest, Naive Bayes, Decision Tree, and KNN, offering an insightful viewpoint on the models' suitability for use in banking systems.[10]

The study reflects the extensive research and potential of decision trees to streamline financial procedures, particularly in anticipating loan approvals, offering a user-friendly application. [11]

Study aims to forecast loan acceptance using SVM, logistic regression, and other machine learning techniques, highlighting the importance of user-friendly interfaces for practical application ease and wider adoption. [12]

The methodologies and practical implementation, aim in to understand the effectiveness of these tools in real-

world banking scenarios where model can anticipate outcomes and is quickly adaptable to a wide range of inputs using Logistic regression and random forest.[13]

The primary objective is to enhance the loan approval process by utilizing historical data for predicting outcomes, employing Deep learning methods like ANN and data mining. Their focus on accurate predictions and prudent fund management contributes to optimizing the loan approval process discourse.[14]

A real-time loan approval classification method utilizing an auto-encoder in deep learning outperforms traditional binary classifiers like SVM in F1 score. This framework holds promising potential for selecting creditworthy candidates in real-time, with future research aiming to enhance performance using advanced deep learning techniques.[15]

The methodology of the paper involves a hybrid approach combining Convolutional Neural Network (CNN) with Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) classifiers for credit risk prediction.[16]

This research explores machine learning in finance, emphasizing ensemble learning techniques like Random Forest, Decision Tree, and XGBoost. It provides insights into challenges, limitations, recent advancements, and the current state of research in loan prediction and machine learning applications in finance.[17]

The literature survey reviews studies on credit forecasting systems, emphasizing data processing, customer trait consideration, and model comparisons to enhance loan approval processes using machine learning and deep learning techniques along with models containing blockchain.

III. Problem Statement

To tackle inefficiencies in manual loan approval processes, this project propose integrating Blockchain and Deep Learning technologies for automation and security. This approach aims to streamline evaluations, reduce processing times, and enhance overall efficiency and security in loan approval procedures.

IV. Proposed Architechture

The proposed architecture, depicted in Fig (1), starts with a React.js-based user interface where a bank employee inputs various data. When the "predict" icon is clicked, the request is sent to the Python-based Flask infrastructure.

In the backend, a deep learning model, contained within a pickle file, evaluates the input data and generates a loan approval prediction ('yes' or 'no'). This prediction is then returned to the frontend for display.

Interaction between frontend and backend occurs when the bank employee selects the "save data" icon. The backend, linked to JavaScript, integrates with the MetaMask extension on the webpage to authenticate requests for accessing and engaging with the Ethereum blockchain.

Upon authorization via MetaMask, the request is sent to the Solidity code, specifically the Ethereum blockchain-implemented data storage contract. The furnished data is securely stored on the blockchain.

Finally, the interface presents the blockchain-preserved data in a tabular representation, concluding the process and establishing a secure and efficient framework for loan estimation and data storage.

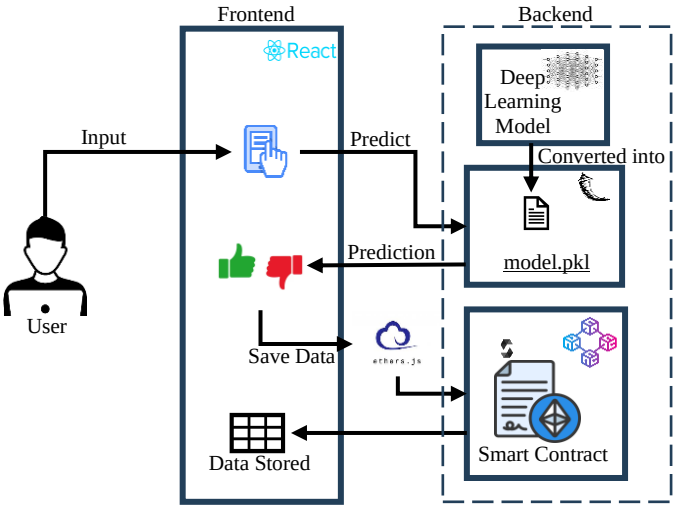


Fig 2. Architecture for Loan Approval Estimation

V. Deep Learning Model

The proposed credit risk assessment model employs a neural network framework with multiple layers, analyzing input data. By minimizing a designated loss function, the model is trained to forecast loan acceptance or denial, with its performance evaluated by performance matrix offering valuable insights for decision-makers in the lending industry to efficiently control credit risk.

A. *Data Collection*

The dataset used for building and evaluating the credit risk assessment model is sourced from Kaggle, and it is named the "Credit Risk" dataset.

TABLE I. Columns in Dataset

person_age	person_income
person_home_ownership	person_emp_length
loan_intent	loan_grade
loan_percent_income	loan_int_rate
loan_status	loan_amnt
cb_person_default_on_file	cb_person_cred_hist_length

B. *Pre- processing*

a. Handling Missing Values: Identify missing values and impute missing values by replacing them with the mean values of their respective columns.

b. Drop Columns: Remove unnecessary columns. This step helps reduce dimensionality and focuses on relevant features for credit risk assessment.

c. Analysis and Handling Outliers: Fig(2) represents outlines in data. Remove outliers by filtering the data within the interquartile range (IQR).

d. Convert categorical data to numerical data

e. Fig(3) represents the co-relation matrix

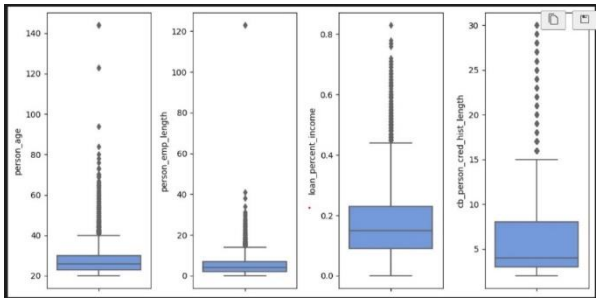


Fig 2. Analysis of Outliners

Fig 3. correlation matrix

f. Feature Selection and Finalization: Feature selection process was undertaken to identify the most relevant attributes for credit risk assessment.

TABLE II. ANALYSIS OF FEATURE SELECTION

S.no	Number of features	Feature names	Accuracy
1.	5 features	loan_percent_income loan_int_rate loan_grade loan_amnt person_emp_length	0.8576
2.	6 features	loan_percent_income loan_int_rate loan_grade loan_amnt person_emp_length cb_person_cred_hist_length	0.8569
3.	7 features	loan_percent_income loan_int_rate loan_grade loan_amnt person_emp_length person_age cb_person_cred_hist_length	0.8523
4.	8 features	loan_percent_income loan_int_rate loan_grade loan_amnt person_emp_length person_age cb_person_cred_hist_length person_income	0.8537

The model with eight features achieves a commendable accuracy of 0.8537, while the one with five features reaches 0.8576. The model with eight features is chosen for its balanced accuracy and inclusion of necessary features for comprehensive analysis.

g. Pie chart of loan status:

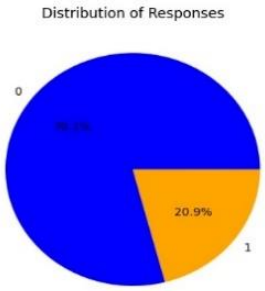


Fig 4. loan status

C. *Model Selection*

A Sequential neural network model was selected for credit risk prediction due to its capacity to capture complex relationships within the data. The model includes multiple layers with varying numbers of neurons and employs the ReLU activation function. Different experiments are performed to bring the best model and output.

TABLE III. EXPERIMENT WITH CHANGE IN NEURON LAYERS

Number of Layers	Number of neurons in each layer	Accuracy
2 layers	256,256	0.8500
3 layers	128, 256, 256	0.8519
3 layers	265, 256, 256	0.8500
4 layers	**128, 265, 256, 256**	**0.8553**
4 layers	256, 256, 256, 256	0.8533
5 layers	128, 256,256,256,256	0.8499

TABLE IV. EXPERIMENT WITH CHANGING EPOCHES

Number of Epoches	**50**	100	200
Accuracy	**0.8557**	0.8495	0.8460
Precision	**0.6757**	0.6675	0.6512
F1	**0.6012**	0.5637	0.5736
recall	**0.6363**	0.6112	0.6099

TABLE V. EXPERIMENT WITH CHANGING LEARNING AND EPOCHES

Learning rate	Epoch=50	Epoch=100
0.1	**0.8551**	0.8519
0.01	0.8515	0.8537
0.001	0.8539	0.8499

TABLE VI. EXPERIMENT USING DIFFERENT OPTIMIZERS

Optimizers	Accuracy
Stochastic Gradient Descent (SGD):	0.7901
Adams: (Adaptive Moment Estimation):	**0.8520**
RMSprop (Root Mean Square Propagation):	0.8501
Adagrad (Adaptive Gradient Algorithm):	0.7901
Nadam	0.7901

This comprehensive approach aimed to develop a deep learning model with four hidden layers using ReLU activation functions, each with neuron counts of 128, 256, 256, and 256. The output layer utilized a linear activation function. Training included historical data of approved and denied cases, and hyperparameter fine-tuning involved thorough experimentation to optimize accuracy.

D. Model Evaluation

A Sequential neural network model was selected for credit risk prediction due to its capacity to capture complex relationships within the data. The model includes multiple

a. Confusion Matrix:

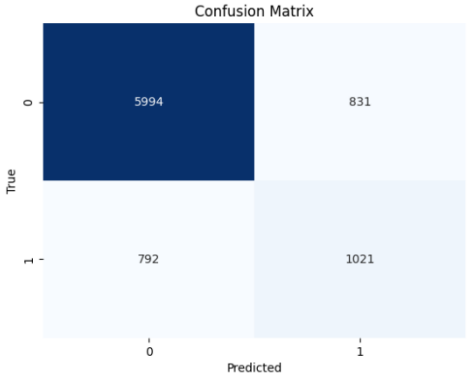


Fig 5. Confusion matrix

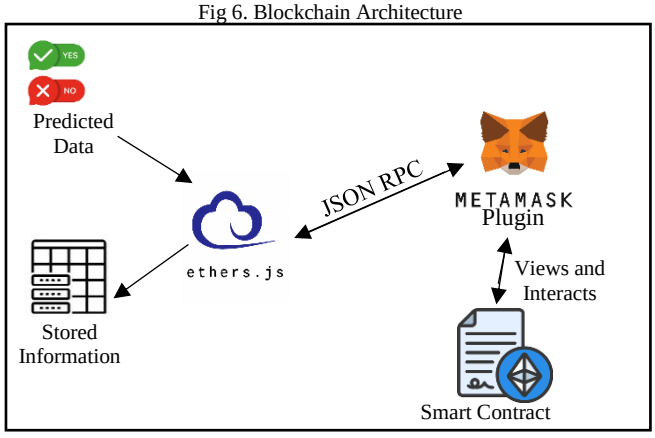
b. Performance Matrix:

- Accuracy - 0.8512387126649688
- Precision - 0.6694480102695763
- Recall -0.5752895752895753
- F1 score -0.61880747552655

VI. Blockchain Based Storeage System

Blockchain technology offers a robust and transparent solution for recording transactions, leveraging decentralization and cryptography for security. It ensures the integrity of various data types, including financial records and digital assets, by dispersing them among network nodes. The inclusion of smart contracts and distributed consensus mechanisms further bolsters trust and reliability in blockchain-based storage systems, making them highly desirable for sectors in need of efficient and secure data management solutions.

Fig 6. Blockchain Architecture



The architecture depicted in Fig (6) includes:

Predicted data: Generated by the deep learning model in response to input predictions.

Ether.js: A JavaScript library for Ethereum blockchain communication, enabling decentralized app development and smart contract interfacing.

MetaMask Plugin: A browser extension and mobile app serving as a digital wallet for Ethereum assets and dApp interaction.

Smart contract: Self-executing contracts on the blockchain, enforcing terms transparently and immutably.

Saved information: Data saved in blocks after smart contract execution, displayed on the front end as a table

A. Flow of Execution:

Upon user interaction with the "Save Data" button, input data is routed to Ether.js' addInformation() function, facilitating frontend-to-blockchain connectivity. Utilizing JSON-RPC, data is formatted and MetaMask is prompted for user validation, ensuring transaction security. MetaMask then initiates a transaction to the Ethereum network, where a predefined smart contract governs data storage with transparency and integrity. After network confirmation, the smart contract securely stores the data. Subsequently, getAllInformation() retrieves and formats stored data for seamless display on the frontend, enhancing user accessibility and application usability.

VII. Results and Discussion

The model achieved an accuracy of 85%, with precision at 67%, recall at 58%, and an F1 score of 62%, demonstrating a balanced performance in predicting positive instances while minimizing false positives.

The application is designed to provide users with detailed information retrieved from a blockchain. It offers a user-friendly front end where users can input specific details and predict the loan status the blockchain integration helps in storage. The retrieved information is then displayed in a structured format, primarily in the form of a table, providing users with clear and organized results.

Key Features:

User-Friendly Front End: The application offers an intuitive and easy-to-use front end, enabling users to input their desired details or criteria effortlessly.

Blockchain Integration: Utilizing blockchain technology, the application securely accesses and retrieves information stored within the blockchain network.

Dynamic Output Table: Retrieved information is dynamically presented in a structured table format for user clarity and ease of understanding. Each row represents a distinct record, with columns delineating various attributes. The table is designed to be informative, displaying all relevant details concisely and organized manner upon query initiation.

VIII. Conclusion

To estimate loan acceptance utilizing a Deep Learning model and blockchain technology for safe data storage, the study offers a decentralized application. The program guarantees data immutability by using the model's accurate predictions and the permanence and tamper-proof

characteristics of blockchain storage. The project's future includes for expanding the program to include encryption features to better protect sensitive data and highlighting the possibility of improving the model to suit other datasets. This new method provides a solid answer to the urgent demand for trustworthy loan approval systems, opening the door to improved model accuracy and safer data storage in the future.

REFERENCES

[1] Gogia, S., and U. Sharma. "Blockchain and machine learning based peer-to-peer lending for the post-pandemic economy." (2021).

[2] Deborah, R. Nancy, et al. "An Efficient Loan Approval Status Prediction Using Machine Learning." *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*. IEEE, 2023.

[3] Arun, Kumar, Garg Ishan, and Kaur Sanmeet. "Loan approval prediction based on machine learning approach." *IOSR J. Comput. Eng* 18.3 (2016): 18-21.

[4] Gupta, Kanishk, et al. "Loanification-loan approval classification using machine learning algorithms." *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*. 2021.

[5] Kadam, Ashwini S., et al. "Prediction for loan approval using machine learning algorithm." *International Research Journal of Engineering and Technology (IRJET)* 8.04 (2021).

[6] Pramod, Ms Kathe Rutika, et al. "Prediction of Loan Approval using Machine Learning Algorithm: A Review Paper." (2021).

[7] Chowdhury, Minhaj Uddin, et al. "Blockchain application in banking system." *Journal of Software Engineering and Applications* 14.7 (2021): 298-311.

[8] Fang, Weiwei, et al. "Deep learning anti-fraud model for internet loan: Where we are going." *IEEE Access* 9 (2021): 9777-9784

[9] Amarnadh, Vadipina, and Nageswara Rao Moparthi. "Prediction and assessment of credit risk using an adaptive Binarized spiking marine predators' neural network in financial sector." *Multimedia Tools and Applications* (2023): 1-37.

[10] AC, Ramachandra, and Vishwas KN. "Prediction of Loan Approval in Banks Using Machine Learning Approach." (2023).

[11] Anurima Majumdar, Romik Banerjee, Rounak Ghosh, Sagar Ghosh, Rudradip Bhattacharjee5, Sayak Saha6, Subhrajit Pallob7, Antara Ghosal8, Palasri Dhar9, Nabaneeta Banerjee1. "Loan Prediction by using Machine Learning". INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY June 2022 | IJIRT | Volume 9 Issue 1 | ISSN: 2349-6002

[12] Mamatha, Madhura, Mahalakshmi Mukri. "PREDICTING LOAN APPROVAL USING MACHINE LEARNING". International Research Journal of Modernization in Engineering Technology and Science(Peer-Reviewed, Open Access, Fully Refereed International Journal) Volume:05/Issue:05/May-2023.

[13] Loan Prediction System Using Machine Learning Anant Shinde1 , Yash Patil2 , Ishan Kotian3 , Abhinav Shinde4 and Reshma Gulwani5 1,2,3,4Department of Information Technology, RAIT, Nerul, India 5 D.Y. Patil Deemed to be University, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India.

[14] Jena, Soumya Ranjan, and S. Vasantha. "PREDICTION OF MODERNIZED LOAN APPROVAL SYSTEM BASED ON MACHINE LEARNING APPROACH."

[15] Abakarim, Youness, Mohamed Lahby, and Abdelbaki Attioui. "Towards an efficient real-time approach to loan credit approval using deep learning." 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC). IEEE, 2018.

[16] Melese, Tamiru, et al. "Credit-Risk Prediction Model Using Hybrid Deep—Machine-Learning Based Algorithms." Scientific Programming 2023 (2023).

[17] Vishal Singh "Prediction of Modernized Loan Approval System Based on Machine Learning Approach",2021