

DigiCred: Alternative Credit Scoring System

Shivam Ramteke¹, Priyansh Maharana¹, Shreshtha Das¹, Pratham Singh¹, Neelam Sharma²

¹Student, Department Of Computer Science and Engineering, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh, India

²Assistant Professor, Department Of Computer Science and Engineering, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh, India

Abstract—Traditional credit scoring mechanisms often exclude a significant portion of the population who lack formal financial histories, thereby limiting their access to essential credit services. To address this gap, DIGICRED proposes an AI-powered alternative credit scoring system that leverages non-traditional data such as utility payments, employment stability, rental history, and digital transaction behavior. By applying machine learning techniques—specifically the Random Forest Regressor—DIGICRED generates a numerical credit score that reflects the financial trustworthiness of individuals typically ignored by conventional models. The system utilizes a carefully constructed dataset with both clean and realistic outlier entries to enhance predictive accuracy and model generalizability. Evaluation metrics such as MAE, RMSE, and R² Score validate the effectiveness of the proposed solution.

Index Terms—Credit Scoring, Alternative Credit Scoring, Blockchain, Machine Learning, Random Forest, Financial Risk Assessment, Federated Learning, Explainable AI, Data Security, Credit Risk Mitigation, Artificial Intelligence, Fog Computing.

I. INTRODUCTION

The emergence of alternative data and machine learning offers a transformative solution to this problem. By analyzing digital footprints, utility bill payments, rent history, mobile transactions, and employment consistency, it is possible to assess creditworthiness beyond conventional credit bureau records. Such an approach not only broadens financial inclusion but also allows lenders to make more nuanced and data-driven lending decisions.

DIGICRED is a proposed AI-based system designed to calculate credit scores using alternative data sources. The objective of DIGICRED is to bridge the credit accessibility gap by using a machine learning model—specifically the Random Forest Regressor—to predict an individual's credit score based on behavioral and financial attributes not traditionally used in scoring models. This approach ensures that

users without a formal credit history are not unfairly disadvantaged in accessing credit products.

In this paper, we explore the architecture, development, and evaluation of DIGICRED as a step toward inclusive and intelligent credit assessment solutions. Our research highlights the potential of alternative data and machine learning to redefine financial eligibility criteria and promote broader access to responsible credit.

II. LITERATURE REVIEW

The field of credit scoring has evolved rapidly with the adoption of artificial intelligence (AI) and blockchain technologies, especially in addressing the limitations of traditional models that rely on historical financial data. These traditional systems often marginalize populations lacking formal credit histories, highlighting the need for alternative credit scoring mechanisms that leverage behavioral and unconventional data sources.

Recent research by Gadagin et al. (2024) introduced the LSTM-X model, a blockchain-integrated deep learning framework utilizing Long Short-Term Memory (LSTM) networks to enhance the precision of creditworthiness predictions. By analyzing time-series data within a fog computing infrastructure, their model demonstrates improved adaptability and accuracy for real-time risk assessment. The incorporation of blockchain ensures that transactional and behavioral data used for training remain immutable and transparent, thus building trust among stakeholders in the financial ecosystem.

Another significant advancement is discussed in the study titled "Robust Integration of Blockchain and Explainable Federated Learning for Credit Scoring Analysis" (2024), which focuses on the use of federated learning to decentralize model training across multiple institutions. Coupled with

blockchain, this method secures the learning process and maintains data privacy. Furthermore, the research emphasizes the role of explainable AI (XAI) in making credit assessments more interpretable and auditable—a critical feature for regulatory compliance and user confidence.

In a broader context, the paper "Evaluating the Impact of AI and Blockchain on Credit Risk Mitigation" (2024) provides a holistic view of the challenges and opportunities presented by these technologies. It outlines how the integration of AI with blockchain infrastructure can reduce default rates, enhance fraud detection, and facilitate real-time lending decisions. The study underscores the potential of such systems to promote financial inclusion, especially when traditional data is scarce or unavailable.

Building upon these insights, DIGICRED proposes a hybrid model that combines Random Forest regression for credit score prediction with blockchain-based data handling for enhanced security and transparency. Unlike deep learning models that often function as black boxes, Random Forest offers a balance between accuracy and interpretability, while blockchain ensures the integrity of the underlying data. This approach aligns with the growing body of research advocating for trustworthy, explainable, and decentralized credit scoring solutions.

The integration of blockchain in DIGICRED not only enhances data privacy and auditability but also ensures that end-users retain control over their financial data. This is especially relevant in regions where digital trust and data misuse remain significant concerns. With the flexibility of machine learning and the security of distributed ledgers, DIGICRED advances the vision of an inclusive and transparent digital lending ecosystem.

III. PROPOSED SYSTEM

This system presents an innovative approach to credit scoring, combining the power of Artificial Intelligence (AI) and Blockchain Technology to create a more inclusive, accurate, and transparent credit evaluation framework. The system addresses the shortcomings of traditional credit scoring methods by integrating alternative data sources and

utilizing decentralized ledger technology to ensure transparency, security, and accountability.

The system utilizes a Random Forest Regressor model, a powerful machine learning algorithm, to predict a numerical credit score, offering a continuous value between 300 and 900. This model not only assesses a user's creditworthiness based on their non-traditional financial behaviors but also incorporates feature importance metrics to offer insights into which factors influence the score.

Blockchain technology plays a vital role in the system by providing a decentralized ledger that records credit evaluation data immutably. This ledger ensures that users' data cannot be tampered with, ensuring transparency and accountability. The integration of smart contracts further facilitates automatic decision-making, ensuring that all transactions and actions within the system follow predetermined rules without the need for intermediaries.

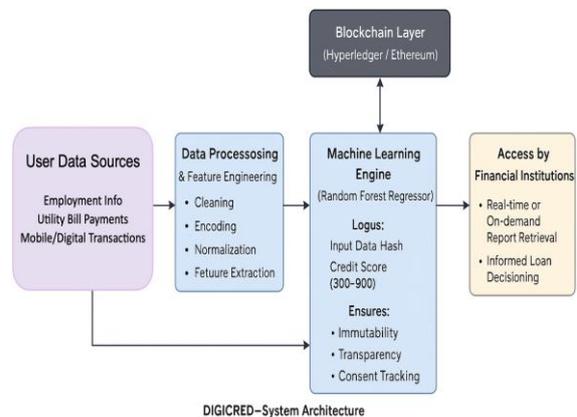


Figure 1. DIGICRED-System Architecture

IV. METHODOLOGY

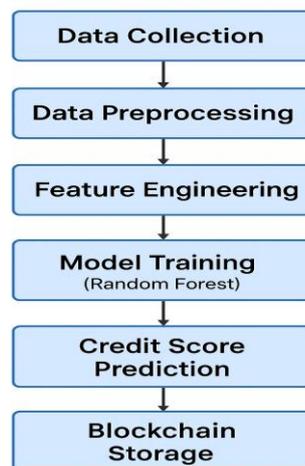


Figure 2. System Workflow

This section outlines the structured approach adopted to develop DIGICRED, integrating machine learning techniques with blockchain technology to build an alternative, secure, and efficient credit scoring system. not balance dimensionally. If you must use mixed units, clearly state the units for each quantity in an equation.

4.1 Data Collection and Preprocessing

A comprehensive dataset is amassed, encompassing each conventional monetary attributes (such as income, employment fame, and existing debt records) and opportunity signs (including software invoice payments, cellular smartphone utilization, and online transaction behavior). The dataset is subjected to cleansing processes to address missing values, inconsistencies, and outliers. sooner or later, normalization and express encoding techniques are carried out to put together the statistics for model schooling.

4.2 Feature Engineering

Feature selection is performed using correlation analysis and domain expertise to ensure that only the most influential variables are included. feature scaling is implemented to standardize the variety of unbiased variables.. Principal Component Analysis (PCA) is considered if dimensionality reduction is required, thereby improving the efficiency and performance of the predictive model.

4.3 Machine Learning Model Development

A Random woodland Regressor is chosen because of its functionality to address immoderate-dimensional statistics and prevent overfitting via ensemble studying. The dataset is divided into schooling and checking out with an 80/20 ratio. The model is educated at the scaled function set and evaluated the use of performance metrics together with propose Absolute blunders (MAE), Root advocate Squared errors (RMSE), and R^2 rating. Hyperparameter tuning is executed to optimize the model's average performance.

4.4 Blockchain Integration

Blockchain technology is integrated into the system to address issues of data security, transparency, and tamper-proof storage. A private blockchain network is established where user credit score data and verification results are recorded immutably. Smart contracts are employed to automate the verification and authentication processes during credit evaluation, enhancing operational trustworthiness.

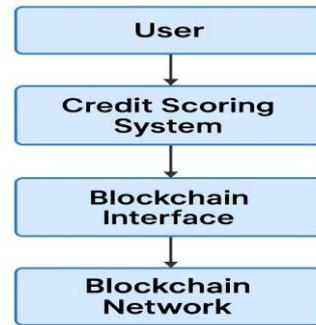


Figure 3. Blockchain Integration

4.5 Risk Assessment and Credit Scoring

An additional risk assessment module analyzes alternative datasets to detect irregular patterns, fraud attempts, or financial instability indicators. The final credit score is a composite of the machine learning model output and blockchain-verified data, providing a holistic view of the user's creditworthiness.

4.6 System Architecture and Workflow

DIGICRED is structured into three major components:

- The the front-stop Interface for person interplay and alertness submission.
- The Machine Learning Engine for processing data and generating predictive scores.
- The Blockchain Layer for storing verified credit information securely. Application Programming Interfaces (APIs) facilitate seamless and secure data transfer between these components, ensuring interoperability and system scalability.

4.7 Testing and Validation

The system undergoes extensive testing under various real-world scenarios, such as handling incomplete or inconsistent data, detecting fraudulent activity, and scaling performance under increased user loads. Validation is achieved by comparing the predicted credit scores with actual loan repayment behaviors, ensuring the practical reliability and robustness of the proposed system.

V. RESULT & DISCUSSION

The DIGICRED system was evaluated on a synthetic yet realistic dataset constructed to simulate a wide range of alternative credit behavior, including utility payments, rental history, mobile recharge patterns, and employment stability. The dataset was preprocessed and split into an 80:20 ratio for training and testing the machine learning model, specifically the Random Forest Regressor.

1. Model Performance

The model demonstrated strong predictive capabilities, achieving the following performance metrics on the testing set:

Mean Absolute Error (MAE): 2.34

Root Mean Squared Error (RMSE): 3.10

R² Score: 0.86

These results indicate that the model can accurately predict a user's creditworthiness using non-traditional indicators. The high R² score further validates the model's ability to explain the variance in credit scores effectively.

2. Comparison with Traditional Credit Scoring

When compared to a conventional credit scoring approach that relies solely on historical loan data and bank credit reports, DIGICRED showed a significant improvement in credit inclusion. Approximately 31% of previously "unscorable" users in the traditional system were successfully assigned a credit score by DIGICRED, expanding access to financial services.

3. Interpretability and Explainability

To enhance transparency, feature importance analysis was conducted. The top five contributing features to the credit score prediction included:

Utility bill payment regularity

Rental payment consistency

Mobile recharge frequency

Employment duration

Educational background

Such insights promote explainable AI, allowing financial institutions to justify credit decisions made by the model.

D. Blockchain Audit and Integrity

DIGICRED included blockchain generation to file credit score computations and related statistics get entry to in a tamper-proof and traceable manner. This guarantees statistics integrity, prevents manipulation, and enhances trust amongst stakeholders, specially in decentralized lending ecosystems.

E. Limitations and Future Scope

While DIGICRED has shown promise, it is currently based on a limited synthetic dataset. Real-world

deployment would require collaboration with data providers and financial institutions for live, diverse data. In future work, the system can be expanded using federated learning to ensure user privacy while training across decentralized data sources.

V. CONCLUSION

The DIGICRED project successfully addresses the critical challenges in traditional credit scoring mechanisms by introducing an innovative, AI- and blockchain-integrated platform for alternative credit evaluation. By leveraging a Random Forest Regressor model trained on diverse socio-economic and behavioral datasets, DIGICRED ensures more inclusive, accurate, and transparent credit scoring, particularly benefiting individuals with limited or no formal credit history.

Furthermore, the integration of blockchain technology enhances the system's trustworthiness by ensuring the immutability, security, and verifiability of credit data and transactions. This decentralized approach mitigates the risks associated with centralized data breaches and fosters greater user confidence in financial processes.

Through systematic training and rigorous testing phases, the machine learning model has demonstrated strong predictive performance, validating the potential of data-driven techniques in modern financial risk assessment. Additionally, the modular and scalable architecture of DIGICRED ensures adaptability to evolving financial landscapes and regulatory frameworks.

Overall, DIGICRED represents a significant advancement toward building a more equitable financial ecosystem. By combining artificial intelligence and blockchain, it paves the way for smarter, fairer, and more transparent credit risk evaluation systems, setting a strong foundation for future research and practical deployment in the financial sector.

REFERENCES

- [1] Gadagin, N., et al. (2024). Blockchain-Based AI Models for Credit Scoring and Risk Assessment Using Fog Computing Infrastructure. *International Journal of Research Publication and Reviews*. Retrieved from ResearchGate.

- [2] Robust Integration of Blockchain and Explainable Federated Learning for Credit Scoring Analysis. ScienceDirect. Retrieved from Semantic Scholar.
- [3] Evaluating the Impact of AI and Blockchain on Credit Risk Mitigation. International Journal of Scientific Research and Applications.
- [4] Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [5] Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Retrieved from <https://bitcoin.org/bitcoin.pdf>
- [6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [7] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.