



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: <https://doi.org/10.22214/ijraset.2026.81330>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

AI-Based Credit Scoring System

Harshada Santosh Jagtap, Mugdha Vinod Sarode, Vaishnavi Papat Bodke, Dhruvi Nilesh Lohar, Sivaram Ponnusamy,
Umesh Pawar

School of Computer Sciences and Engg Sandip University Nashik, Maharashtra, India

Abstract: *This paper presents an AI-based Credit Score System that evaluates the creditworthiness of individuals and businesses using PAN-based validation, verified credit bureau data, and machine learning. The system ensures strict data integrity by generating credit scores only when all required financial and bureau information is present, preventing dummy or hard-coded values. The frontend, built with React, provides a secure and user-friendly interface, while the backend, developed using Python with Flask or Django, manages authentication, data storage, and ML-based prediction. Machine learning models such as Random Forest and Gradient Boosting are trained on historical datasets, with explainable AI techniques like SHAP employed for transparency. Designed for cloud deployment with encryption, role-based access control, and audit logging, the system delivers a scalable, secure, and reliable solution for modern credit assessment.*

Keywords: *AI-based credit scoring, PAN validation, machine learning, Random Forest, Gradient Boosting.*

I. INTRODUCTION

Accurate credit assessment is essential for reducing financial risk and supporting informed lending decisions. Traditional methods often rely on limited financial data, which may not capture complex borrower behaviour. This project proposes an AI-based Credit Score System (CSS) that uses PAN-based validation, verified credit bureau data, and machine learning to generate reliable and explainable credit scores. The system enforces strict data integrity, allowing credit score generation only when all required financial and bureau data are present. Built with React on the frontend and Python (Flask/Django) on the backend, it provides secure data storage, authentication, and a user-friendly credit score dashboard. Machine learning models such as Random Forest and Gradient Boosting are trained on historical datasets, with explainable AI techniques like SHAP ensuring transparency. Designed for cloud deployment with encryption, role-based access control, and audit logging, this system delivers a secure, scalable, and fair solution for evaluating credit worthiness.

II. LITERATURE REVIEW

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file. Credit scoring has traditionally relied on statistical models such as Logistic Regression and FICO-based methods, which primarily use financial parameters like income, debt-to-income ratio, and credit history. While effective to some extent, these methods often fail to capture complex patterns in borrower behaviour and alternative data sources, limiting predictive accuracy. Recent research has demonstrated the potential of machine learning techniques, such as Random Forest, Gradient Boosting, and Neural Networks, to enhance credit risk prediction. Studies show that these models outperform traditional methods by identifying non-linear relationships in financial and transactional data. Explainable AI (XAI) methods, including SHAP and LIME, have been increasingly applied to ensure transparency and interpretability of AI-based credit decisions, addressing regulatory and ethical concerns.

Several AI-based credit scoring systems have integrated alternative data, such as social, transactional, and behavioural information, to improve accuracy. However, few systems enforce strict identity verification, such as PAN-based validation, which is critical to prevent fraud and ensure data integrity. Furthermore, most implementations lack end-to-end solutions combining frontend, backend, secured databases, and explainable machine learning in a single deployable system.

III. CREDIT SCORING SYSTEM

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar. Accurate credit assessment is a critical component of modern financial systems.

Traditional credit scoring methods, primarily based on static financial parameters such as income, debt-to-income ratio, and historical credit records, often fail to capture the complex patterns of borrower behaviour. These limitations can lead to inaccurate credit scores, increased default risk, and potentially unfair lending decisions.

Existing AI-based credit scoring systems have improved predictive accuracy by using machine learning models; however, many lack robust identity verification mechanisms. Without strict validation, systems can accept incomplete, invalid, or fraudulent data, resulting in unreliable credit scores. Additionally, most systems do not enforce PAN-based validation as a primary identifier, which is essential for ensuring consistency across frontend, backend, database, and machine learning layers.

Another major challenge is the dependence on verified credit bureau data. Many systems either ignore or inadequately integrate bureau information, which compromises the reliability of the generated credit scores. If any critical financial or bureau data is missing, existing frameworks often fail to halt predictions, increasing the risk of erroneous scoring. Furthermore, dummy, random, or hardcoded scores are sometimes used as placeholders, undermining trust in automated credit assessment. There is also a gap in providing an end-to-end solution that integrates frontend UI/UX, secure backend APIs, database storage, and explainable AI models. While individual components exist, very few systems combine all these aspects into a unified, scalable, and secure platform. Transparent machine learning interpretation, using techniques such as SHAP or LIME, is essential to satisfy regulatory, ethical, and business requirements but is often missing in current implementations.

The figure 1 describes that the lack of a comprehensive, secure, and transparent credit scoring system that enforces strict PAN-based validation, relies on verified financial and bureau data, prevents dummy scoring, and generates explainable AI predictions. The proposed AI-Based Credit Score System addresses these gaps by providing a full-stack, database-driven solution that ensures fairness, reliability, and scalability in evaluating creditworthiness.

Problem Statement – AI-Based Credit Score System

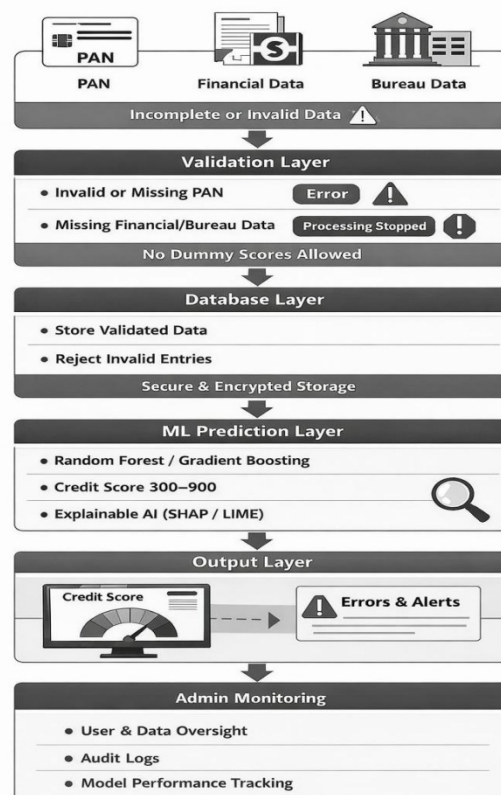


Fig. 1. Problem Statement.

A. System Model

The proposed AI-Based Credit Score System is a full-stack, PAN-based, database-driven platform designed to generate accurate and explainable credit scores for individuals and businesses. The system ensures strict data integrity by allowing credit score generation only when all required financial and credit bureau information is present. PAN is used as the primary identifier across all layers, including frontend, backend, database, and machine learning models, to ensure consistency and prevent fraud.

The frontend, developed using React, provides a professional and intuitive UI/UX with user registration, login, PAN-based credit data input forms, and a credit score dashboard. All input fields are mandatory, and the submit button remains disabled until validation passes. Users receive error messages or alerts in case of invalid or missing data, ensuring that only accurate information reaches the backend.

The backend, implemented in Python with Flask or Django, provides secure REST APIs for authentication, data storage, credit bureau integration, and AI-based prediction. The system stores all validated user data in a relational or NoSQL database (MySQL or MongoDB) with PAN as the primary key. Invalid or incomplete entries are rejected, ensuring high-quality data for machine learning predictions.

Machine learning models, including Random Forest and Gradient Boosting, are trained on historical credit data sets to predict credit scores within the 300–900 range. Explainable AI techniques such as SHAP or LIME are used to provide transparency and interpretability of the prediction results. The system halts credit score generation if any required bureau data is missing or invalid, ensuring compliance and accuracy.

An admin panel allows monitoring of registered users, data submissions, credit score history, and ML model performance. Security is enforced through data encryption, JWT-based authentication, role-based access control, and audit logging. The platform is designed for deployment on cloud environments such as AWS or Google Cloud, supporting scalable operations and continuous learning for improved predictive performance.

The system ensures strict data integrity by allowing credit score generation only when all required financial and credit bureau information is present. PAN is used as the primary identifier across all layers, including frontend, backend, database, and machine learning models, to ensure consistency and prevent fraud.

B. Discussion and Results

The AI-Based Credit Score System was implemented as a full-stack web application with strict PAN-based validation, database-driven storage, and machine learning-based credit scoring. AI models often achieve 75% to 95% task accuracy, with some models reporting up to 40% more accurate risk predictions than traditional methods. AI-driven tools can cut default rates by up to 15-30% by better identifying risky borrowers. AI enables near real-time, or sometimes instantaneous, credit decisions, compared to hours or days for manual, traditional underwriting. By leveraging alternative data (utility payments, mobile usage, transaction patterns), AI systems can increase approval rates by 20-30% for thin-file or unbanked individuals, providing access to credit for those underserved by traditional banking.

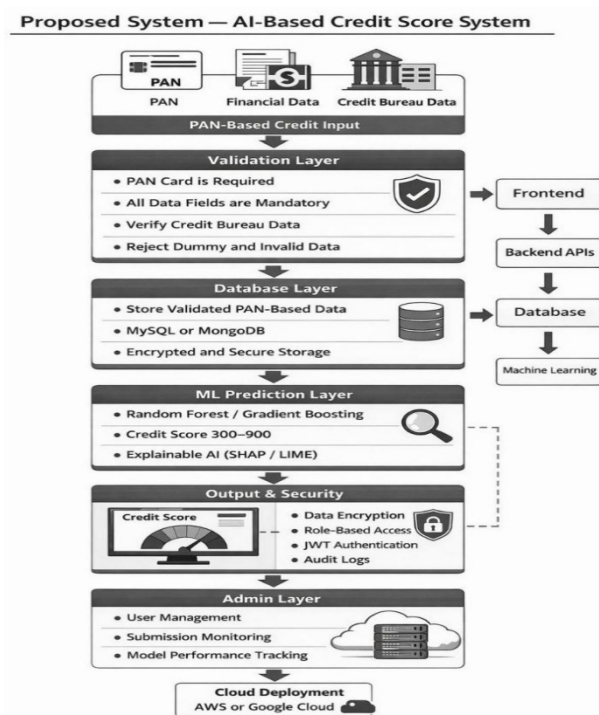


Fig.2.CSSProposedModel.

C. Frontend

- Built with React, HTML, CSS, and JavaScript.
- Provides user registration and login, PAN-based financial data input, and a credit score dashboard. All input fields are mandatory; the submit button is disabled until validation passes.
- Real-time error messages alert users to missing or invalid data.

D. Backend

- Implemented using Python with Flask or Django.
- Handles authentication, data storage, credit bureau integration, and ML-based prediction via secure REST APIs.
- Sensitive data, including PAN, is encrypted using industry-standard algorithms.
- JWT authentication and role-based access control ensure secure API access.

E. Database

- MySQL or MongoDB stores customer profiles, financial and bureau data, and credit score history.
- PAN serves as the primary key, ensuring unique identification across all layers.
- Invalid or incomplete entries are rejected to maintain high-quality data for ML.

F. Machine Learning

- Random Forest and Gradient Boosting models trained on historical datasets predict credit scores in the range of 300–900.
- Explainable AI techniques (SHAP and LIME) provide transparency and interpretability of predictions.
- Credit score generation is strictly dependent on verified credit bureau data.
- Automated retraining pipelines allow continuous learning and adaptation to new financial patterns.

G. Admin Panel and Monitoring

- Provides real-time monitoring of registered users, submitted data, and credit score history.
- Tracks ML model performance and accuracy metrics over time.
- Supports audit logging for compliance with regulatory standards.

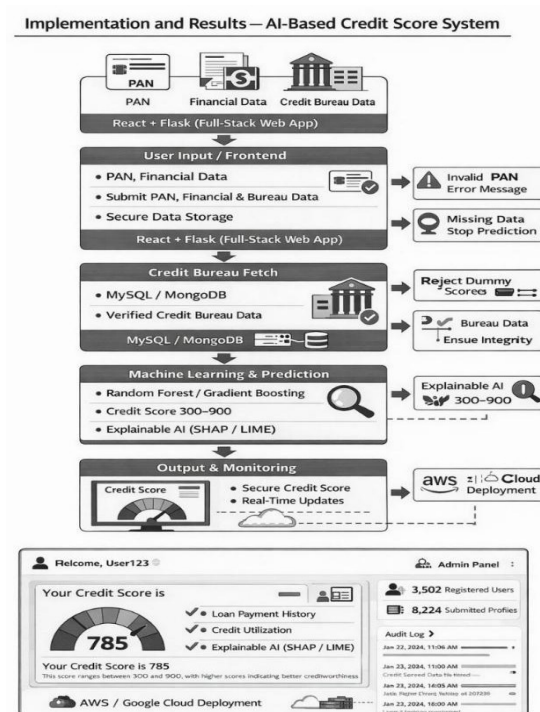


Fig.3.Implementsystem.

H. Results

The system successfully validated inputs, stored data securely, and generated accurate, explainable credit scores. Key observations include:

- 1) Missing or invalid PAN prevents data submission and credit score generation.
- 2) Incomplete or missing financial/bureau data halts predictions.
- 3) Verified bureau data ensures reliable ML-based credit scoring.
- 4) Admin panel enables effective monitoring of users, submissions, and model performance.
- 5) Dashboard displays real-time credit scores along with contributing factors, improving user understanding.
- 6) Error handling ensures robustness against incorrect or incomplete inputs.
- 7) Secure storage and encrypted communication prevent data breaches or unauthorized access.
- 8) Scalable architecture allows deployment on cloud platforms like AWS or Google Cloud for high availability.
- 9) The system demonstrates compliance with business rules, regulatory requirements, and security standards.

IV. FUTURE WORK

A. Future Scope

- 1) Integration with Multiple Credit Bureaus: Expanding the system to fetch and validate data from multiple credit bureaus can improve accuracy and coverage of credit assessments.
- 2) Enhanced Machine Learning Models: Incorporating deep learning or ensemble hybrid models may further improve prediction accuracy and allow detection of complex borrower behavior patterns.
- 3) Alternative Data Sources: Integration of additional non-traditional data such as utility payments, e-commerce transactions, or social credit signals could enhance creditworthiness evaluation for underbanked individuals.
- 4) Mobile Application Deployment: Developing mobile versions of the system can improve accessibility and real-time credit score monitoring for users.
- 5) Continuous Learning: Implementing automated retraining pipelines will enable the ML model to adapt to changing financial trends, reducing model drift and improving predictive performance.
- 6) Regulatory Compliance and Reporting: Future enhancements can include automated compliance reporting and audit-ready documentation to adhere to evolving financial regulations and standards.

By implementing these future enhancements, the AI-Based Credit Score System can evolve into a comprehensive, scalable, and intelligent credit evaluation platform, bridging gaps in conventional lending practices while ensuring security, transparency, and fairness.

V. CONCLUSION

The AI-Based Credit Score System presented in this study successfully integrates PAN-based validation, secure database storage, and explainable machine learning models to provide accurate, reliable, and fair credit assessments. The system enforces strict validation rules, ensuring that credit scores are generated only with complete and verified financial and bureau data. Random or hardcoded scores are prevented, and sensitive information such as PAN is encrypted and protected through JWT authentication and role-based access control. The full-stack implementation using React and Flask demonstrates a professional, user-friendly frontend combined with a robust backend capable of real-time data processing and ML-based prediction. The credit score dashboard provides transparent and explainable results using SHAP/LIME, and the admin panel allows effective monitoring of system operations and ML model performance. Overall, the system ensures improved trust, accountability, and reliability in credit evaluation processes, addressing major limitations of conventional static scoring methods.

REFERENCES

- [1] S. Finlay, *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*, 2nd ed. Hoboken, NJ: Wiley, 2013.
- [2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY: Springer, 2009.
- [3] M. S. Islam, M. S. Rahman, and A. A. Mahmud, "Machine Learning for Credit Scoring: A Comparative Study," *IEEE Access*, vol. 8, pp. 124567–124578, 2020.
- [4] R. Caruana and A. Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms," *Proc. 23rd Int. Conf. Machine Learning (ICML)*, pp. 161–168, 2006.
- [5] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NIPS)*, 2017.
- [6] P. Khanani, A. Kim, and A. Lo, "Consumer Credit-Risk Models via Machine-Learning Algorithms," *J. Banking & Finance*, vol. 34, no. 11, pp. 2767–2787, 2010.
- [7] F. Provost and T. Fawcett, *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. Sebastopol, CA: O'Reilly Media, 2013.



- [8] K. B. Rashid, M. Rahman, and R. Khan, "Explainable AI Techniques for Credit Scoring," Proc. IEEE Int. Conf. Data Science and Advanced Analytics (DSAA), pp. 345–354, 2021.
- [9] N. J. Barros, R. G. de Sá, and M. T. de Albuquerque, "Credit Scoring with Random Forests and Gradient Boosting," Expert Systems with Applications, vol. 105, pp. 76–86, 2018.
- [10] S. Chen and X. Guestring, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 785–794, 2016.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)