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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue V May 2026- Available at www.ijraset.com*

# AI-Based Credit Scoring System

Harshada Santosh Jagtap, Mugdha Vinod Sarode, Vaishnavi Popat Bodke, Dhruti 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 financialandbureauinformationispresent,preventingdummy orhardcodedvalues.Thefrontend,builtwithReact,providesa secureanduser-friendlyinterface,whilethebackend,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 auditlogging,thesystemdeliversascalable,secure,andreliable solution for modern credit assessment.*
***Keywords:** AI-basedcreditscoring,PANvalidation,machine learning, Random Forest, Gradient Boosting.*

## I. INTRODUCTION

Accurate credit assessment is essential for reducing financial risk and supporting informed lending decisions. Traditionalmethodsoftenrelyonlimitedfinancialdata,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 bureaudataarepresent.BuiltwithReactonthefrontendand Python(Flask/Django) onthebackend,I tprovidessecuredata storage, authentication, and a user-friendly credit score dashboard.MachinelearningmodelssuchasRandomForest andGradientBoostingaretrainedonhistoricaldatasets,with explainableAItechniqueslikeSHAPensuringtransparency. 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,confirmthatyouhavethecorrecttemplateforyour paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close thisfile anddownload the MicrosoftWord, Letterfile. Credit scoring has traditionally relied on statistical models suchasLogisticRegressionandFICO-basedmethods,which primarily use financial parameters like income, debt-to- income ratio, and credit history. While effective to some extent,thesemethodsoftenfailtocapturecomplexpatternsin borrower behaviour and alternative data sources, limiting predictive accuracy. Recent research has demonstrated the potential of machine learning techniques, such as Random Forest,GradientBoosting,andNeuralNetworks,toenhance 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 alternativedata,suchassocial,transactional,andbehavioural 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 solutionscombiningfrontend,backend,securedatabases,and explainablemachine learning in a single deployable system.

## III. CREDIT SCORING SYSTEM

Beforeyoubegintoformatyourpaper,firstwriteandsave the content as a separate text file. Complete all content and organizationaleditingbeforeformatting.Pleasenotesections A-D below for more information on proofreading, spelling and grammar Accurate credit assessment is a critical component of modern financial systems.

©IJRASET: All Rights are Reserved 

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 patternsofborrowerbehaviour.Theselimitationscanleadto inaccuratecreditscores, increaseddefaultrisk,andpotentially unfair lending decisions.

Existing AI-based credit scoring systems have improved predictive accuracy by using machine learning models; however,manylackrobustidentityverificationmechanisms. Without strict validation, systems can accept incomplete, invalid, or fraudulent data, resulting in unreliable credit scores.Additionally,mostsystemsdonotenforcePAN-based validation as a primary identifier, which is essential for ensuringconsistencyacrossfrontend,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 frameworksoftenfailtohaltpredictions,increasingtherisk of erroneous scoring. Furthermore, dummy, random, or hardcoded scores are sometimes used as placeholders, undermining trust in automated credit assessment. There is alsoagapinprovidinganend-to-endsolutionthatintegrates frontendUI/UX,securebackendAPIs,databasestorage,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 requirementsbutisoftenmissingincurrentimplementations.

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 ScoreSystemaddressesthesegapsbyprovidingafull-stack, database-drivensolutionthatensuresfairness, reliability,and scalability in evaluating creditworthiness.

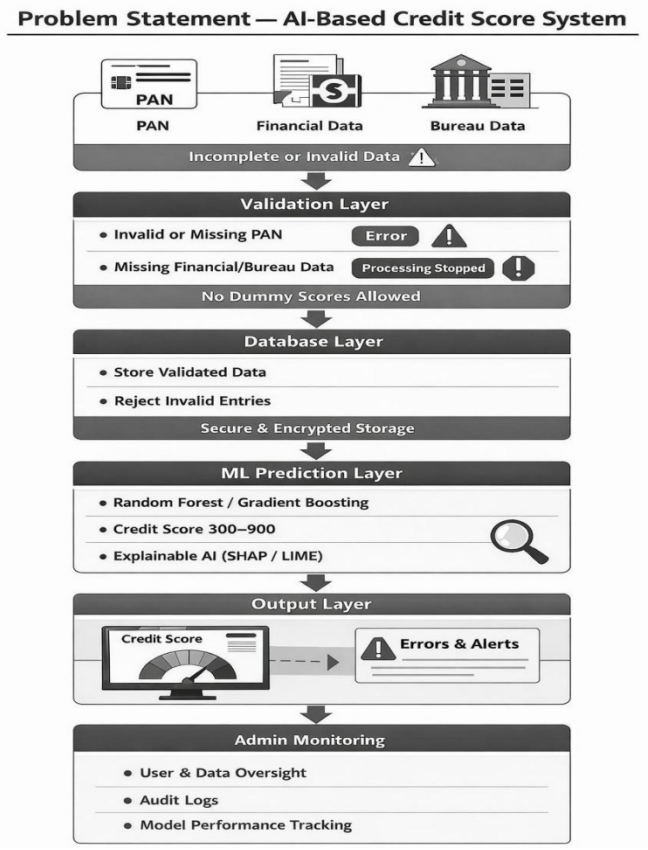



Fig.1.ProblemStatement.

*A. System Model*

The proposed AI-Based Credit Score System is a full- stack, PAN- based, database-driven platform designed to generateaccurateandexplainablecreditscoresforindividuals 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,andmachinelearningmodels,to ensure consistency and prevent fraud.

The frontend, developed using React, provides a professionalandintuitiveUI/UXwithuserregistration,login, PAN-based credit data input forms, and a credit score dashboard. All input fields are mandatory, and the submit buttonremainsdisableduntilvalidationpasses.Usersreceive error messages or alerts in case of invalid or missing data, ensuringthatonlyaccurateinformationreachesthebackend.

The backend, implemented in Python with Flask or Django, provides secure REST APIsfor 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 GradientBoosting,aretrainedonhistoricalcreditdatasetsto 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 dataismissingorinvalid,ensuringcomplianceandaccuracy

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

Thesystemensuresstrictdataintegritybyallowingcredit 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,andmachinelearningmodels,toensureconsistency and prevent fraud.

### *B. Discussion and Results*

TheAI-BasedCreditScoreSystemwasimplementedasa full-stackwebapplicationwithstrictPAN-basedvalidation, database- driven storage, and machine learning-based credit scoring.AImodelsoftenachieve75%to95%taskaccuracy, 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,creditdecisions,comparedtohoursordaysfor manual, traditional underwriting. By leveraging alternative data(utilitypayments,mobileusage,transactionpatterns),AI systemscanincreaseapprovalratesby20-30%forthin-fileor unbanked individuals, providing access to credit for those underserved by traditional banking.

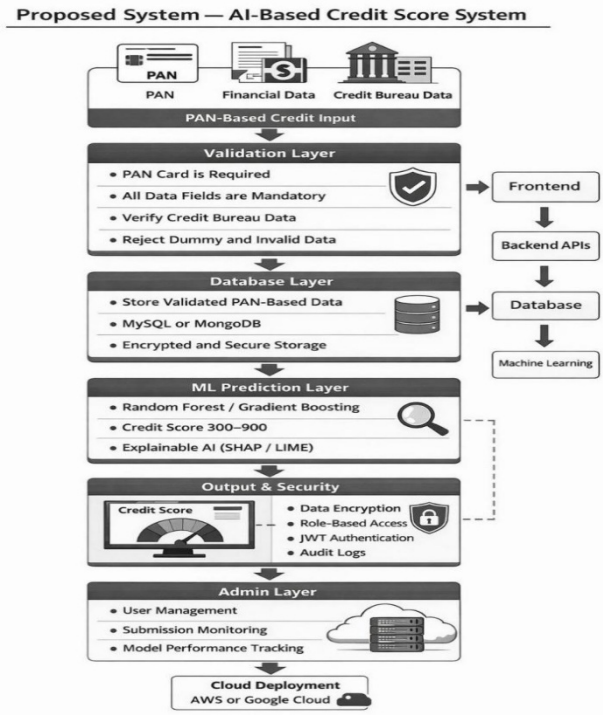



Fig.2.CSSProposedModel.

*C. Frontend*

- BuiltwithReact,HTML,CSS,andJavaScript.
- Provides user registration and login, PAN-based financialdatainput,andacreditscoredashboard.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*

- ImplementedusingPythonwithFlaskorDjango.
- 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*

- MySQLorMongoDBstorescustomerprofiles,financial and bureau data, and credit score history.
- PANservesastheprimarykey,ensuringuniqueidentification across all layers.
- Invalidorincompleteentriesarerejectedtomaintain 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.
- Creditscoregenerationisstrictlydependentonverified credit bureau data.
- Automated retraining pipelines allow continuous learning and adaptation to new financial patterns.

*G. AdminPanelandMonitoring*

- Providesreal-timemonitoringofregisteredusers, submitted data, and credit score history.
- TracksMLmodelperformanceandaccuracymetrics over time.
- Supports audit logging for compliance with regulatory standards.

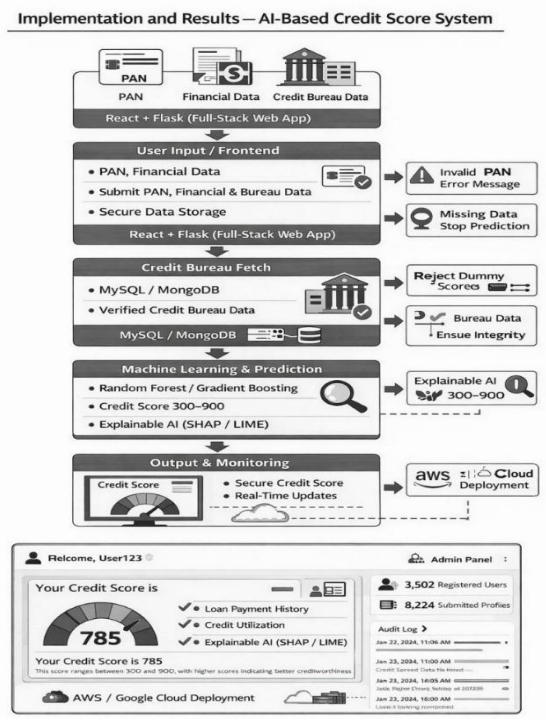


Fig.3.Implementsystem.

### 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) Incompleteormissingfinancial/bureaudatahalts predictions.
3) Verified bureau data ensures reliable ML-based credit scoring.
4) Adminpanelenableseffectivemonitoringofusers, 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. FutureScope

1) Integration with Multiple Credit Bureaus: Expanding thesystemtofetchandvalidatedatafrommultiplecredit 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) AlternativeDataSources:Integrationofadditionalnon- traditional data such as utility payments, e-commerce transactions, or social credit signals could enhance creditworthinessevaluationforunderbankedindividuals.
4) Mobile Application Deployment: Developing mobile versionsofthesystemcanimproveaccessibilityandreal- time credit score monitoring for users.
5) Continuous Learning: Implementing automated retrainingpipelineswillenabletheMLmodelstoadaptto 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- BasedCreditScoreSystemcanevolveintoacomprehensive, scalable, and intelligent credit evaluation platform, bridging gaps in conventional lending practices while ensuring security, transparency, and fairness.

## V. CONCLUSION

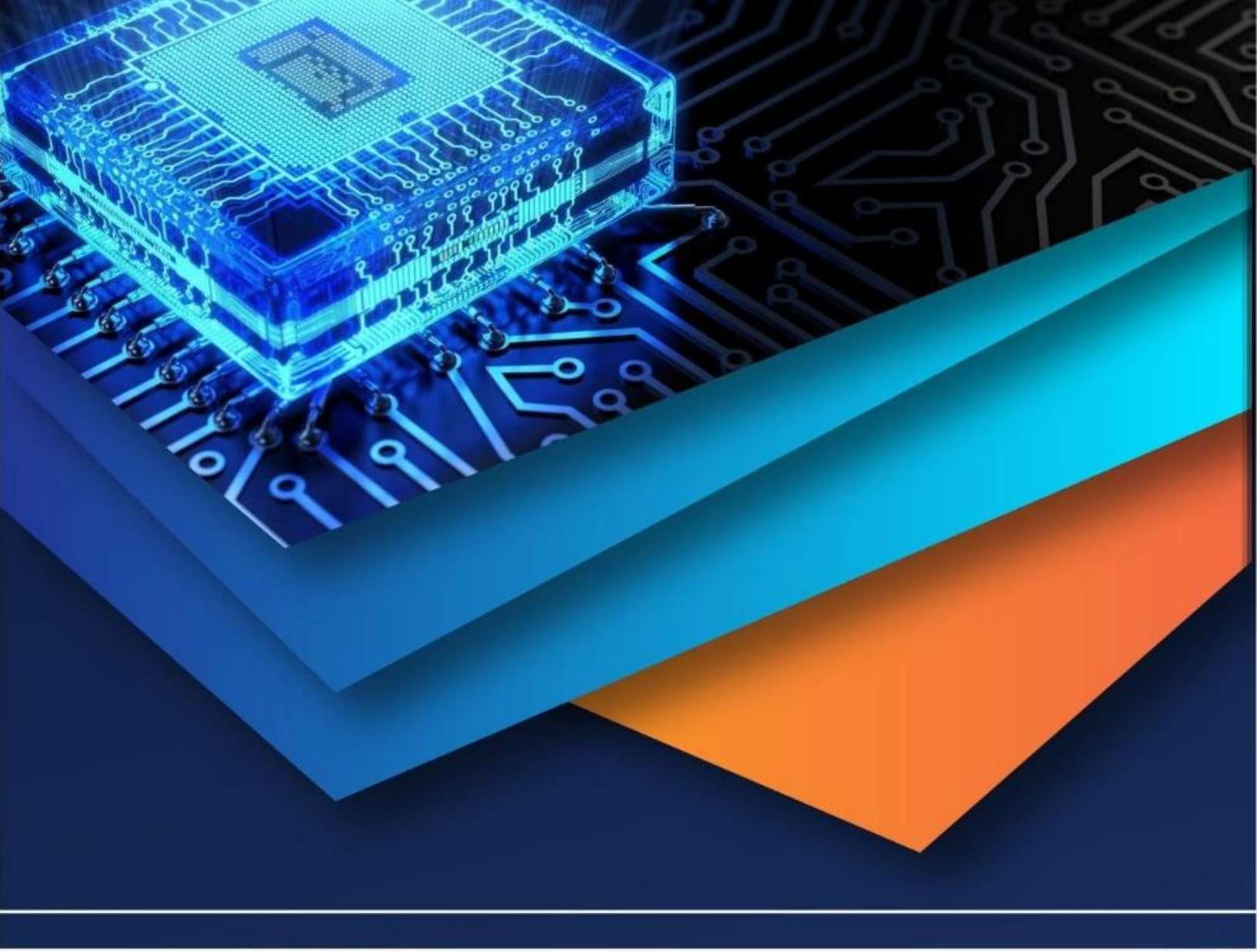
TheAI-BasedCreditScoreSystempresentedinthisstudy successfullyintegratesPAN-basedvalidation,secure database storage, and explainable machine learning models toprovideaccurate,reliable,andfaircreditassessments.The system enforces strict validation rules, ensuring that credit scores are generated only with 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-friendlyfrontendcombinedwitharobustbackendcapableof 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 StatisticalLearning: DataMining,Inference,and Prediction,2nd ed.NewYork,NY: Springer, 2009.
[3] M. S. Islam, M. S. Rahman, and A. A. Mahmud, "Machine Learningfor Credit Scoring: A Comparative Study," IEEE Access, vol. 8, pp.124567–124578,2020.
[4] R. Caruana and A. Niculescu-Mizil, "An Empirical Comparison ofSupervised Learning Algorithms," Proc. 23rd Int. Conf. MachineLearning (ICML), pp. 161–168, 2006.
[5] S. Lundberg and S. Lee, "A Unified Approach to Interpreting ModelPredictions," Advances in Neural Information Processing Systems(NIPS), 2017
[6] P. Khanani, A. Kim, and A. Lo, "Consumer Credit-Risk Models viaMachine-LearningAlgorithms,"J.Banking&Finance,vol.34,no.11,pp.2767–2787,2010.
[7] F.ProvostandT.Fawcett,DataScienceforBusiness:WhatYouNeedtoKnowAboutDataMiningandData-AnalyticThinking.Sebastopol,CA: O'Reilly Media, 2013.

[8] K. B. Rashid, M. Rahman, and R. Khan, "Explainable AI TechniquesforCreditScoring,"Proc.IEEEInt.Conf.DataScienceandAdvancedAnalytics (DSAA), pp. 345–354, 2021.
[9] N. J. Barros, R. G. de Sá, and M. T. de Albuquerque, "Credit Scoringwith Random Forests and Gradient Boosting," Expert Systems withApplications, vol. 105, pp. 76–86, 2018.
[10] S. Chen and X. Guesting, "XGBoost: A Scalable Tree BoostingSystem,"Proc.22nd ACMSIGKDDInt.Conf.KnowledgeDiscoveryand Data Mining, pp. 785–794, 2016.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)