International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN (Online): 2347-5552, Volume-14, Issue-6, Nov-Dec 2026
DOI: https://doi.org/10.55524/ijircst.2026.14.6.4
Article ID IRPC-1435, Pages 30-37
www.ijircst.org

An AI-Based Framework for Credit Risk Assessment of Thin-File Microenterprises Using Alternative Financial Behavioral Data

Shivangi Srivastava[1], and Raj Bihari Lal Srivastava[2]

[1] Assistant Professor, Department of Computer Application, C.M.P Degree College, Prayagraj, Uttar Pradesh, India
[2] Professor, Department of Commerce, C.M.P. Degree College, Prayagraj, Uttar Pradesh, India

Correspondence should be addressed to Shivangi Srivastava; shivangi srivastavashivangi361@gmail.com

Received: 5 Oct 2026; **1st Revision:** 18 Oct 2026; **2nd Revision:** 29 Oct 2026; **Accepted:** 8 Nov 2026; **Published Online:** 16 Nov 2026

Copyright © 2026 Shivangi Srivastava et al. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Microenterprises often have difficulty accessing formal finance due to limited credit history, incomplete financial records and lack of documented borrowing information. So, their financial situation and repayment ability may not be truly reflected by traditional credit evaluation methods. Opportunities in machine learning and artificial intelligence to improve credit risk assessment involve incorporating alternative financial and behavioral indicators. However, predictive models can make unreliable decisions when the information they have is incomplete, inconsistent, or insufficient. In this context, this paper proposes an uncertainty-aware AI framework for credit risk assessment of thin-file microenterprises with alternative financial behavioral indicators. The proposed framework consists of financial preprocessing of data, behavioral feature extraction, machine learning based risk prediction, uncertainty quantification, and human-in-the-loop decision support. The candidate prediction models considered are Logistic Regression, Random Forest and XGBoost. To benchmark the context, we consider a public dataset of credit-risk of SMEs from Bangladesh with 1,200 loan records and 27 structured features. The repository benchmark results using five-fold stratified cross-validation show that Logistic Regression has the highest ROC-AUC (0.900 ± 0.024) and PR-AUC (0.776 ± 0.045), while XGBoost has the lowest Brier score (0.174). Random Forest shows the relatively lower discrimination performance and the largest reported train-test gap. These results can serve as a comparative reference for candidate prediction model selection. The proposed mechanisms of probability calibration, conformal prediction and human-review have not yet been evaluated experimentally. The benchmark dataset is a representation of the general SMEs' population, which does not prove that all are thin file microenterprises. Thus, the reported results do not serve as direct evidence of the target population's performance. We put forth a framework which is a structure for developing more open, which includes and presents the uncertainty in results and responsible AI based credit assessment systems.

KEYWORDS— Artificial Intelligence, Credit Risk Assessment, Microenterprises, Alternative Financial Data, Machine Learning, Uncertainty Estimation, Financial Behavioural Indicators

I. INTRODUCTION

Jobs creation, local economic growth and financial inclusion are all at the center of microenterprises. Many micro firms contribute to the economy but struggle to access formal financing due to lack of accurate financial accounts, established credit history and recorded borrowing records. Most traditional credit scoring systems use historical repayment information, collateral, income documentation and financial ratios. This can be a problem for businesses that don't have much in the way of formal financial record. The advent of digital financial services has given access to alternative financial information. Transaction histories, regular cash flows, payment behavior, sales patterns, and consistent costs also provide further insights into a company's financial activity. Such data may include alternate credit information which lenders use to determine repayment ability in addition to traditional credit reports. In past research which looked at SME lending [1], [2] it was found that alternative transaction data is a useful tool for credit assessment.

Also, we have seen that machine learning techniques play a role in credit risk prediction as they are able to identify complex relationships between financial characteristics and repayment results. Logistic Regression is a simple interpretable baseline. Random forest and gradient boosting are able to capture non-linear relationships between input variables [3], [4]. However, the predictive accuracy does not tell us if a model's individual prediction is accurate or not. Credit risk model may return a high risk or low risk classification, even if the data on which it is based is incomplete or poorly calibrated in terms of prediction likelihood. . This ambiguity is very important in financial decision making, because a wrong assessment may result in a wrong loan approval or rejection. The credit risk uncertainty quantification has been studied to analyze the trustworthiness of the machine learning predictions [5]. To alleviate these problems, the paper proposes an uncertainty-aware AI approach for credit risk assessment of thin-file microenterprises. It includes alternative financial

behavior indicators and prediction and uncertainty estimation by means of machine learning . It also proposes a confidence-based decision process for referring uncertain cases to human evaluation.

The key contribution is a structured methodology that incorporates behavioral financial data, credit risk prediction, uncertainty estimation, and human review into one decision support process.

II. LITERATURE REVIEW

The relevance of SME credit information in eliminating information gaps and promoting financial inclusion is discussed by Nguyen and Sagara [1].

Caire and Fernandez Vidal [2] analyze micro and small enterprise lending transactions, and find that sales, expenses, orders, invoices and other operational records might give useful evidence of business activity. Their case studies with Indian fintech providers show that transaction information can have predictive value similar to credit-history information and that a combination of both sources can improve evaluation. This supports the use of observed company behavior as a complement, rather than automatic substitution, to traditional credit information.

Gu et al. [3] report on the application of machine learning to credit risk assessment of small and micro firms which they put forth as a case of the value of data driven methods in this field. Bitetto et al. [4] look at which among machine learning algorithms do a better job in credit risk rating of small enterprises when lenders do not have full information. They report that which models do best is very promising for fintech lending but also that it is too early to tell if we are to use model results for high stake lending decisions. It is also the case that they keep Logistic Regression as a simple base line which is interpretable and at the same time put Random Forest and XGBoost forward as more flexible options. The aim is not to say that complex models are in themselves better but to put forward a framework in which models are compared using the same data partitions, outcome definitions and evaluation criteria.

Nagl et al. [5] study uncertainty quantification in machine-learning approaches to loss-given-default modeling and show the importance of uncertainty analysis in credit risk modeling. Djeundje et al. [6] address the use of alternative information in credit scoring.The authors report on the use of non-traditional data which they put forth as a solution for evaluating students that have had little to no traditional credit history.

In the work by Bitetto et al. [9] they look at past use of random forest technique versus an ordered probit model for the case of Italian small and medium sized firms which they analyze with quarterly data. We see that the machine learning approach did in fact work better in this setting and also that they used Shapley values to study the input variables' contribution. What we find is that tree-based models should be put forth as a very well may be a very good option for the assessment of risk in SME's but also that there is a great need for study of the results put out by the models. This literature supports the distinction between the level of risk projected and the reliability of the prediction.

Di Maggio and Ratnadiwakara [13] further explore the fintech underwriting with alternative data, and explain how atypical information might be useful to capture thin-file applicants who are conservatively rated by standard score algorithms. Their findings suggest the possibility for alternative data to be useful for financial inclusion, but the environment and characteristics in consumer fintech lending cannot be assumed to apply directly to microenterprise lending.

Bahlool et al. [14] systematically evaluate AI-based credit rating, considering prediction performance, fairness, explainability and governance. In their review they underline the importance of considering these features together, and not anticipated accuracy alone, as the measure of model quality. This means that in designing a thin-file microenterprise architecture, choices of input variables, the handling of missing data, the monitoring of subgroup performance and the provision of human supervision must be taken into account. Therefore, in the proposed architecture, the model is considered as a decision-support component without the autonomy to approve or deny credit applications.

Ariza-Garzón et al. [15] study the profit-sensitive machine learning categorization with explanations in the credit risk of small enterprises in Peer-to-Peer Lending. Collectively, these studies imply that the predictive performance should be assessed in conjunction with the capacity to explain the reasons related to a risk assessment. Explainability can be used as a supporting diagnostic to assist reviewers understand the major drivers of a prediction in the proposed framework, but should not be taken as evidence of causation.

Xu et al [16] develop a powerful deep learning model for predicting high risk behaviors in financial traders by examining large volumes of transaction data. The model is first pre-trained in an unsupervised way to develop distributed representations that capture complex data relations on its own.

III. RESEARCH GAP

The reviewed studies looked at the issue of transactional and alternative data [2], [6], [13] machine learning models for SME credit risk [3], [4], [9] explainability and fairness [10], [11], [14], [15] and predictive uncertainty estimation [5], [7], [12]. But these are mostly looked at in separate studies or in what may not be relevant settings to thin file microenterprises. Also, what we see is that the availability and quality of business level behavioral indicators may greatly vary between sectors and between digital and non-digital firms.

This paper reports on what we see as a gap in the field which is the development of an integrated decision support workflow which includes

- business relevant alternative financial behavior indicators,
- comparative machine learning risk prediction,
- probability calibration and uncertainty assessment, and
- confidence based referral to human review.

We present the put forth integration and treatment of uncertainty which is a novel element in our work, we did not develop a new machine learning technique, report on a new behavioral signal, or claim to have achieved proven predictive performance. The framework is still a put forth which requires empirical validation in a dataset that is representative of thin file microenterprises and which has the appropriate variables.

IV. PROPOSED AI FRAMEWORK AND RESEARCH METHODOLOGY

A. Overview of the Proposed Framework

The proposed framework consists of five main stages: (1) collecting of alternative financial data, (2) preprocessing and feature engineering, (3) machine-learning based credit risk prediction, (4) uncertainty estimate and (5) confidence-based decision support. The framework is designed to employ financial behavioral information along with the conventional financial indicators that are accessible. In the last step of decision support, the system decides if a forecast can be made with adequate confidence or if it needs further study.

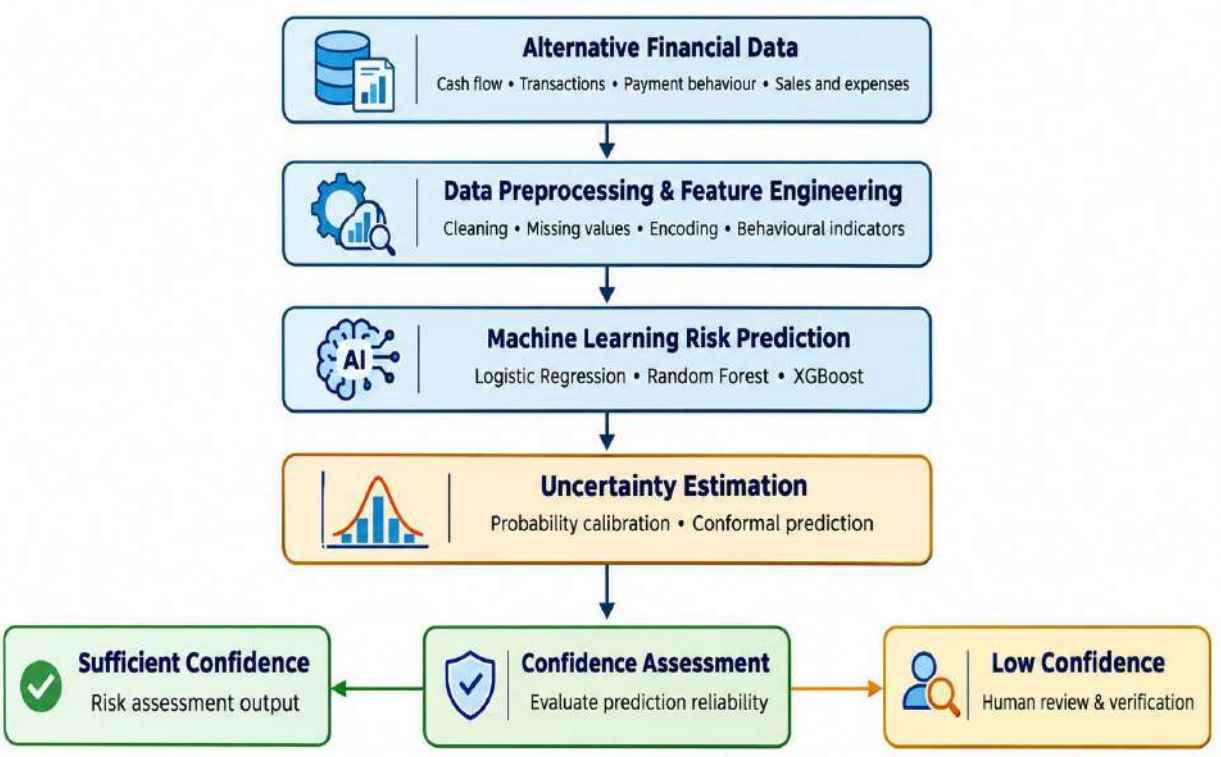


Figure 1: Proposed AI framework which is uncertainty aware for microenterprise credit risk assessment

We present in Figure 1 the full system workflow. We start out by collecting and processing financial and behavioral data which in turn produce relevant indicators. Then we use a machine learning model to assess the credit risk which is in turn evaluated for its reliability by the uncertainty estimation module.

B. Alternative Financial Behavioural Indicators

The proposed framework considers financial indicators that may provide information about the operational and repayment behaviour of microenterprises.

The behavioural indicators that have been suggested, as well as the potential relevance for credit risk assessment, are provided in Table 1.

These variables are conceptual inputs and not actual variables from a specific dataset. Availability and usefulness of their prediction should be considered in the process of data set evaluation and feature selection.

For instance, the coefficient of variation of periodic net cash flow may be used to calculate cash-flow stability, as follows: $CV_CF = \sigma_CF / \mu_CF$

Table 1: Proposed alternative financial behavioural indicators

Indicator	Description	Potential credit relevance
Cash-flow stability	Regularity of incoming and outgoing cash flows	Indicates financial consistency
Transaction frequency	Number of financial transactions over a defined period	Reflects business activity
Payment regularity	Consistency of payments to suppliers and lenders	Indicates repayment behaviour
Sales consistency	Variation in business sales over time	Provides information about revenue stability
Expense consistency	Regularity of recurring operating expenses	Supports assessment of cost management
Repayment history	Record of previous repayment activities, where available	Indicates past repayment behaviour
Cash balance trend	Changes in available business cash balance	Provides information about liquidity

This is because the standard deviation of the periodic net cash flow, represented as σ_CF, and the mean of the periodic net cash flow, represented as μ_CF, are used. A lower coefficient of variation could be a sign of more predictable cash flow, but this does not necessarily apply and depends on the nature of the business and the seasonality concerns.

The number of transactions could be shown as: TF = N_T / T

The number of recorded transactions (N_T) and the observation period (T). These indicators should be based on the same time period of observation. Lack of transactions or incomplete information in transactions should not automatically be seen as a sign of poor creditworthiness.

C. Data Collection and Preprocessing

The proposed methodology employs a public data set of SME loans as a prospective source for methodological validation. One resource we found is the Zenodo dataset "Structured Data Suffices: Machine Learning for SME Credit Risk Assessment in an Emerging Market. Its repository describes 1,200 anonymized SME loan applications, 27 characteristics and a binary target variable [8].

Its repository entry indicates that the dataset is available under a CC BY 4.0 data licence. However, it must first be evaluated to determine if its records specifically represent microenterprises, if it comprises thin-file borrowers, and if the necessary behavioral indications are available.

If the present dataset does not have the required variables, we will have to use a public dataset which will augment the suggested framework or which will present an alternate feature mapping which is properly specified. “You can’t just create variables out of thin air or treat some data as if it were a transaction which it is not. Preprocessing at this stage includes the removal or correction of invalid records, identification of missing values, encoding of categorical variables, scaling of numerical variables as required, identification of outliers, assessment of class distribution, and the separation of training, validation and test data. Preprocessing parameters should be derived only from the training data to avoid information leakage, and then applied to the validation and test data.

D. Machine Learning-Based Credit Risk Prediction

The framework considers three candidate machine learning models.

Table 2: Candidate machine learning models

Model	Main function	Reason for inclusion
Logistic Regression	Estimates default probability	Interpretable baseline model
Random Forest	Combines multiple decision trees	Captures nonlinear relationships
XGBoost	Uses gradient-boosted decision trees	Models complex feature interactions

Proposed models for credit risk prediction are in Table 2. We use Logistic Regression as a base model for comparison also put Random Forest and XGBoost forward as more flexible nonlinear options. Their true value will only be seen out in validation which is not to say at the start they are proven.

The prediction function we put forth is:

$\hat{p}_i = f_\theta(X_i)$

where X_i denotes the input features of enterprise i, f_θ is the trained prediction model, and $\hat{p}_i$ is the estimated probability of default. The output is a probability estimate, not an automatic lending decision. The relative costs of false approvals and false rejections must be selected, and a decision threshold must be selected based on validation data.

E. Uncertainty Estimation

The uncertainty estimation part is designed to identify predictions that may not be reliable enough for automated decision support. The methodology we propose combines two approaches in a complementary way.

- **Calibration of probabilities**

Calibration is the issue of the relationship between what is predicted and what we see in the real world. If the model is well calibrated, we should see that 20% of the cases which are given a default probability of around 0.20 will in fact default in the given time frame. Also, you can use the Brier score to see how good the probability forecasts are:

$BS = (1/N) \Sigma_{i=1}^{N} (\hat{p}_i - y_i)^2$

where N is the number of observations. For each observation $\hat{p}_i$ is the predicted default probability and y_i is the observed binary outcome.

- ***Inference of conformal***

Conformal prediction is used for generating prediction sets and in the process of uncertainty quantification which is based on a calibration sample [7]. In the field of classification this approach puts forth a set of possible classes for each observation. A prediction set which includes both default and non-default classes may indicate that the model is not able to tell the difference between the two outcomes at the chosen significance level. The conformal approach requires to be put through an appropriate calibration procedure and also certain assumptions which may break down in the case of temporal or distributional shifts.

F. Confidence-Based Decision Support

The final stage uses prediction confidence and uncertainty information to support risk assessment.

Table 3: Confidence-based credit risk decision mechanism

Stage	Process
1. Input	Receive enterprise financial and behavioural data.
2. Prediction	Generate a credit risk probability using the trained model.
3. Uncertainty	Estimate prediction reliability using calibration and/or conformal prediction.
4. Confidence assessment	Determine whether the prediction meets predefined validation-based criteria.
5A. Sufficient confidence	Present the risk assessment as decision-support information.
5B. Insufficient confidence	Refer the case for human review and additional information verification.

6. Reassessment	Reassess the case when additional verified information becomes available.

Table 3 reports the decision support process which comes after the uncertainty analysis. We present the risk assessment as a which is of sufficient reliability, and we put forward uncertain predictions for further verification. This approach is to avoid the system from treating all model outputs the same as very reliable. We did not require that loans be automated in terms of approval or denial. All loan decisions are subject to institutional policy, regulatory rules and authorized human review.

V. EXPERIMENTAL SETUP AND BENCHMARK RESULTS

A. Dataset and Experimental Protocol

Machine Learning for SME Credit Risk Assessment in an Emerging Market" [8]. We present a repository of 1,200 anonymized SME loan applications from Bangladesh which include 27 structured features and a binary missed payment outcome. We report a class distribution of 897 non-default observations (74.8% and 303 default observations (25.2%. The dataset is at a level of SMEs' benchmark; the repository description does not go into that all observations are microenterprises or that the borrowers are in fact thin file applicants. We use this dataset here only as a context for the performance of the put forth prediction models and not as a proof of performance on thin file microenterprises.

The dataset repository reports an analysis pipeline which we present and also which reports model performance using fivefold stratified cross validation. We report comparison of Logistic Regression, Random Forest and XGBoost. The metrics we present are ROC-AUC, PR-AUC, Brier score and the train test performance gap. The numbers which follow are as reported by the repository they have not been re calculated in this manuscript. In particular, the proposed probability-calibration, conformal-prediction and human-review stages have not been experimentally evaluated here; no result is claimed for those components.

B. Evaluation Metrics

The evaluation metrics proposed are described in Table 4. Predictive performance is assessed by classification measures and probability reliability is assessed by Brier score and calibration error . The uncertainty procedure is assessed using conformal coverage, and the operational effect of referring uncertain situations to a human for assessment is measured with review rate.

Table 4: Evaluation metrics for the proposed framework

Metric	Purpose
Accuracy	Measures overall classification correctness
Precision	Measures the proportion of predicted defaults that are actual defaults
Recall	Measures the proportion of actual defaults identified
F1-score	Provides a harmonic balance between precision and recall
ROC-AUC	Evaluates discrimination across classification thresholds
Brier Score	Measures the accuracy of predicted probabilities
Calibration Error	Evaluates the agreement between predicted and observed probabilities
Conformal Coverage	Measures the proportion of outcomes contained in prediction sets
Review Rate	Measures the proportion of cases referred for human assessment

The proposed evaluation metrics are grouped into three complementing aspects as shown in Figure 2. Classification performance indicators evaluate the quality of discrimination and classification, probability reliability metrics evaluate the quality and calibration of the predicted probabilities, and uncertainty and workflow metrics evaluate the coverage of the prediction set and the proportion of cases requiring human review. The picture is a conceptual grouping of metrics, and does not represent measured experimental data.

The following standard metrics can be used:

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1\text{-}score = 2 \times (Precision \times Recall) / (Precision + Recall)$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

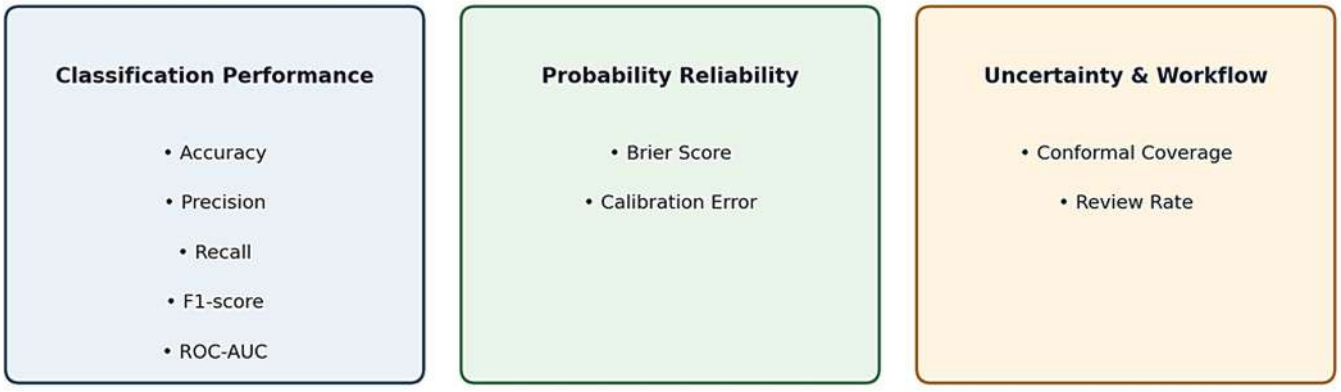


Figure 2: Grouping of evaluation metrics for the proposed framework (Conceptual group only)

C. Reported Benchmark Results

Table 5: Model performance reported with the public SME dataset [8]

Performance metric	Logistic Regression	Random Forest	XGBoost
ROC-AUC (mean ± SD)	0.900 ± 0.024	0.810 ± 0.019	0.875 ± 0.019
PR-AUC (mean ± SD)	0.776 ± 0.045	0.611 ± 0.038	0.717 ± 0.035
Brier score	0.194	0.208	0.174
Train–test gap	0.017	0.190	0.125

The benchmark metrics supplied by the dataset repository are presented in Table 5. Logistic Regression attains the greatest recorded ROC-AUC and PR-AUC , while XGBoost attains the lowest reported Brier score . These numbers are meant to be a guide for potential models, and should not be considered as independent experimental findings from this study. We provide the results on the repositories to clearly differentiate from the previously reported benchmark evidence and to validate the uncertainty-aware approach proposed in this paper.

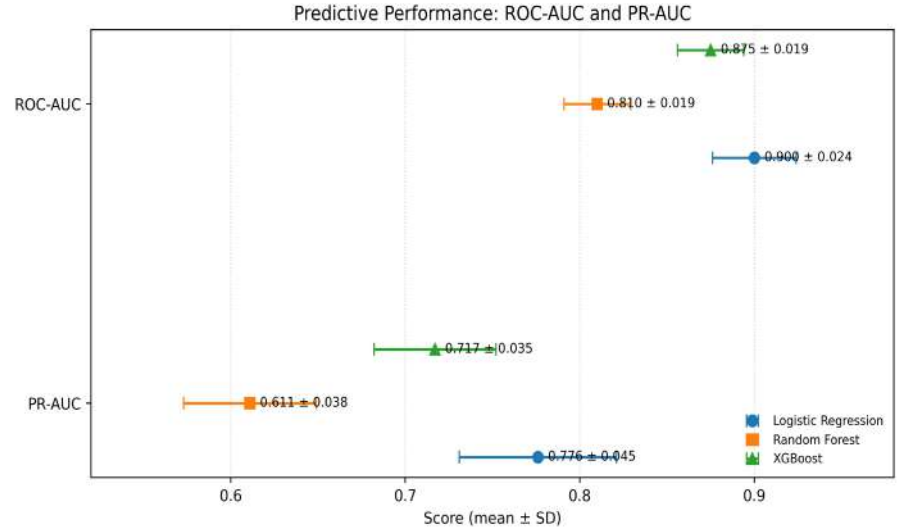


Figure 3: Comparative ROC-AUC and PR-AUC Performance of Benchmark Machine Learning Models

Figure 3 reports a comparison of Logistic Regression, Random Forest, and XGBoost based on ROC-AUC and PR-AUC scores. In this report we present the mean performance which we see to also include the standard deviation SD) which we have illustrated with error bars. Logistic Regression does best in terms of ROC-AUC (0.900) which XGBoost does not match in terms of that score, though XGBoost does out perform the other two models in terms of PR-AUC.

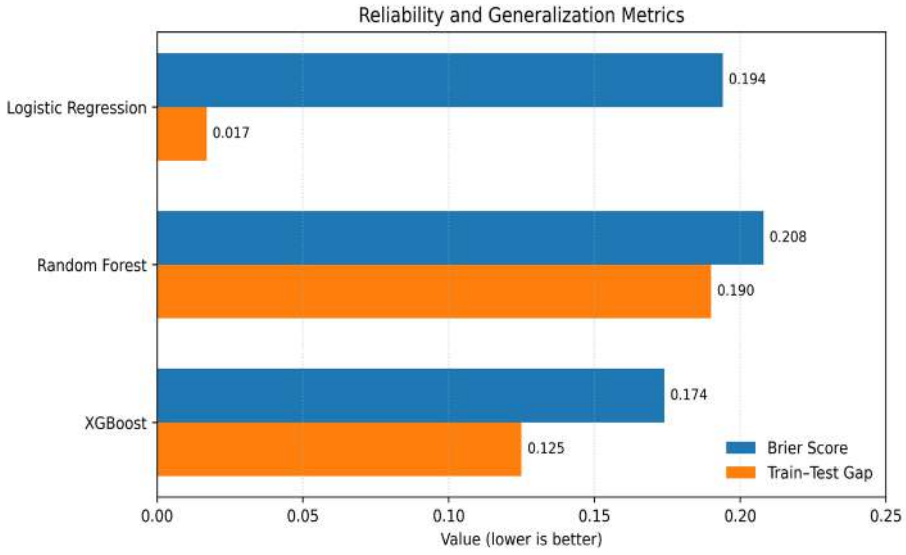


Figure 4. Comparative Evaluation of Brier Score and Train–Test Gap Across Benchmark Machine Learning Models

Figure 4 shows the brier score and train Test Gap. Lower values are better for probability reliability and generalization performance. XGBoost does best with a Brier Score of 0.174 which Logistic Regression does worst for at 0.017 for Train Test Gap. Thus we see that there is a variation in probability estimation reliability and model generalization between the three study algorithms.

C. Discussion

For the benchmark supplied by the repository, Logistic Regression demonstrates the best discrimination with ROC-AUC 0.900 ± 0.024 and PR-AUC 0.776 ± 0.045. Random Forest has the lowest discrimination (ROC-AUC 0.810 ± 0.019; PR-AUC 0.611 ± 0.038) and the highest observed train-test gap (0.190) while XGBoost obtains ROC-AUC of 0.875 ± 0.019 and PR-AUC of 0.717 ± 0.035. • XGBoost has the lowest Brier score (0.174) followed by Logistic Regression (0.194) and Random Forest (0.208). The Brier score is a combination of calibration and discrimination effects, and its value alone is not indicative of a well-calibrated model.

These benchmark results report that we see value in using Logistic Regression as an interpretable base model and putting forward nonlinear alternatives which isn't always the case with more complex models. Also, we do not see that the put forth conformal prediction technique is evaluated nor do we see coverage, prediction set size, or human review rate measured. These results we think should have a separate out their replicable study which reports on training, calibration and independent test partitions. They also need to evaluate whether the selected records are representative of the desired thin-file microenterprise population and contain the behavioral indications according to Table 1.

The proposed framework makes use of alternative financial behavioural information and uses machine learning for credit risk prediction and uncertainty estimation. The main methodological property is the differentiation between risk prediction and the reliability of prediction. However, it is crucial to remember that a model can also produce a probability when it does not have sufficient, consistent or previously unseen information, as the model may be making a prediction beyond its scope of training data. The proposed uncertainty module will help to spot those instances and guide further review. Using behavioral financial indicators could offer data not recorded in traditional credit histories. However, caution should be used in interpreting these indicators. For example, we see that at times cash flows may be irregular which is often a result of seasonal activity and not of financial distress. Also, it may be the business' character which causes low transaction frequency and not low repayment capacity. The put forth human review process is a extra layer of protection for gray areas which does not make automatic decisions. In practice, however, these must be set to effective levels, based on the quality of the data on which they are built, and the consistency of human evaluation. The framework is also somewhat practical. The public SME data identified may not be comprehensive enough to capture thin-file microenterprises, and may not include all behavioural indicators identified in Table 1. Moreover, the framework is not tested in this paper. Thus, it is not claimed to be more accurate in predicting credit or to provide greater financial inclusion or reduce default rates.

VI. CONCLUSION AND FUTURE SCOPE

This paper suggests an AI-based method for credit risk identification of thin-file microenterprises based on alternative financial behavioral data. The framework proposed is a combination of the following steps: preprocessing of financial data, extraction of behavioral features, prediction of risk using machine learning, estimation of uncertainty, and human-in-the-loop decision support. The following logistic Regression, Random Forest and XGBoost are regarded as potential prediction models for credit risk assessment. The study includes also the calibration of probabilities and the use of conformal prediction for the assessment of the expected credit risk which is based on the model's output quality. Also, any case that requires more attention should be put to human review which is in addition to what the computer models present. We looked at a public data set of 1,200 SME loan records which also included 27 structural variables for the purpose of contextual benchmarking. The performance of the top three candidate models using five-fold stratified cross validation in the repository is presented with the highest ROC-AUC of 0.900 ± 0.024 and PR-AUC of 0.776 ± 0.045 for Logistic Regression. Random Forest achieved ROC-AUC 0.810 ± 0.019 and PR-AUC 0.611 ± 0.038. XGBoost gets ROC-AUC of 0.875 +- 0.019 and PR-AUC of 0.717 +- 0.035 . Logistic Regression (0.194), Random Forest (0.208) and XGBoost (0.174) scored the lowest Brier scores. The reported discrimination performance of Logistic Regression seems to be better, while the Brier score of XGBoost is relatively small. The selection of models should be done based on the performance of the model in discrimination and probability, and not based on one evaluation metric. The benchmark results are a first comparison for the candidate prediction models, but do not constitute an experimental validation of the overall proposed framework. In particular, we have not independently implemented and evaluated probability calibration, conformal prediction, prediction-set coverage, prediction-set size and human-review processes in this study. Furthermore, the selected data set is representative of the SME loan applications overall, which does not necessarily imply that all the results are representative for thin-file microenterprises. The results of the benchmark are therefore to be seen in the light of the limitations of the used data set and not in isolation, but cannot be directly transferred to the intended target group. Future research should aim to establish and gather representative microenterprise data that contain validated alternative financial behavioral indicators like cash-flow stability, cash flow frequency, regularity of cash-flow, cash flow sales consistency, cash flow expense consistency, repayment history, cash flow balance trends, etc. The probability calibration and conformal prediction should be applied to and tested on various training, calibration and test data sets in future works. The proposed human review approach should also be evaluated with respect to the quality of the approach in terms of prediction-set coverage, average prediction-set size, review rate, and other appropriate metrics. Furthermore, fairness should be checked between categories of organizations, data privacy, drift of models must also be considered, as well as model explainability and applicable legal requirements before deployment. In that which we present, we put forth a framework which includes alternative financial behavioral data sources, machine

learning and uncertainty estimation and human supervision in the credit risk assessment. We aim to put forward a transparent and comprehensive set of elements for credit application which thin file microenterprises may use, also we add predictive credit performance and reliability of the model results. Also, we will report on the results of the use of representative data sets and reproducible experimental studies which will in turn validate the model's practical effectiveness and application in real world lending settings.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] L. H. Nguyen and M. Sagara, "Credit Risk Database for SME Financial Inclusion," Asian Development Bank Institute, ADBI Working Paper No. 1111, 2020. Available from: https://www.adb.org/publications/credit-risk-database-sme-financial-inclusion

[2] Caire and M. Fernandez Vidal, "Leveraging Transactional Data for Micro and Small Enterprise (MSE) Lending," Consultative Group to Assist the Poor (CGAP), Mar. 2024. Available from: https://www.cgap.org/research/publication/leveraging-transactional-data-for-micro-and-small-enterprise-lending

[3] Z. Gu, J. Lv, B. Wu, Z. Hu, and X. Yu, "Credit risk assessment of small and micro enterprise based on machine learning," *Heliyon*, vol. 10, no. 5, Art. no. e27096, Mar. 2024. Available from: https://doi.org/10.1016/j.heliyon.2024.e27096

[4] Bitetto, P. Cerchiello, S. Filomeni, A. Tanda, and B. Tarantino, "Can we trust machine learning to predict the credit risk of small businesses?" *Review of Quantitative Finance and Accounting*, vol. 63, pp. 925–954, 2024. Available from: https://doi.org/10.1007/s11156-024-01278-0

[5] M. Nagl, M. Nagl, and D. Rösch, "Quantifying uncertainty of machine learning methods for loss given default," *Frontiers in Applied Mathematics and Statistics*, vol. 8, Art. no. 1076083, Dec. 2022. Available from: https://doi.org/10.3389/fams.2022.1076083

[6] V. B. Djeundje, J. Crook, R. Calabrese, and M. Hamid, "Enhancing credit scoring with alternative data," *Expert Systems with Applications*, vol. 163, Art. no. 113766, 2021. Available from: https://doi.org/10.1016/j.eswa.2020.113766

[7] N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," *Foundations and Trends in Machine Learning*, vol. 16, no. 4, pp. 494–591, 2023. Available from: https://doi.org/10.1561/2200000101

[8] S. Chowdhury, A. Das, and S. Paul, "Structured Data Suffices: Machine Learning for SME Credit Risk Assessment in an Emerging Market—Dataset and Analysis Code," Zenodo, Feb. 2026. Available from: https://doi.org/10.5281/zenodo.18525278

[9] Bitetto, P. Cerchiello, S. Filomeni, A. Tanda, and B. Tarantino, "Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses," *Socio-Economic Planning Sciences*, vol. 90, Art. no. 101746, Dec. 2023. Available from: https://doi.org/10.1016/j.seps.2023.101746

[10] Babaei, P. Giudici, and E. Raffinetti, "Explainable FinTech lending," *Journal of Economics and Business*, vols. 125–126, Art. no. 106126, 2023. Available from: https://doi.org/10.1016/j.jeconbus.2023.106126

[11] Y. Chen, P. Giudici, K. Liu, and E. Raffinetti, "Measuring fairness in credit ratings," *Expert Systems with Applications*, vol. 258, Art. no. 125184, Dec. 2024. Available from: https://doi.org/10.1016/j.eswa.2024.125184

[12] Tyralis and G. Papacharalampous, "A review of predictive uncertainty estimation with machine learning," *Artificial Intelligence Review*, vol. 57, Art. no. 94, 2024. Available from: https://doi.org/10.1007/s10462-023-10698-8

[13] M. Di Maggio and D. Ratnadiwakara, "Invisible primes: Fintech lending with alternative data," *Management Science*, published online May 13, 2026. Available from: https://doi.org/10.1287/mnsc.2024.07854

[14] R. Bahlool, N. Hewahi, and W. Elmedany, "Performance, fairness, and explainability in AI-based credit scoring: A systematic literature review," *Journal of Risk and Financial Management*, vol. 19, no. 2, Art. no. 104, Feb. 2026. Available from: https://doi.org/10.3390/jrfm19020104

[15] M.-J. Ariza-Garzón, J. Arroyo, M.-J. Segovia-Vargas, and A. Caparrini, "Profit-sensitive machine learning classification with explanations in credit risk: The case of small businesses in peer-to-peer lending," *Electronic Commerce Research and Applications*, vol. 67, Art. no. 101428, 2024. Available from: https://doi.org/10.1016/j.elerap.2024.101428

[16] K. Xu, Y. Wu, Z. Li, R. Zhang, and Z. Feng, "Investigating Financial Risk Behavior Prediction Using Deep Learning and Big Data," *International Journal of Innovative Research in Engineering and Management (IJIREM)*, vol. 11, no. 3, pp. 77–81, 2024. Available from: https://doi.org/10.55524/ijirem.2024.11.3.12