

REVIEW ARTICLE

Machine Learning Methods for Credit Risk Assessment in Microfinance: A Review

Parmeshwar Madhukar Patil^{a,*}, Rohini Rahul Anyapanawar^b^a Assistant Professor, Department of Statistics, Sangameshwar College (Autonomous), Solapur, Maharashtra, India^b Assistant Professor, Department of Management and Computer Science, Sangameshwar College (Autonomous), Solapur, Maharashtra, India

* Corresponding author

OPEN ACCESS

This article is licensed under CC BY-NC 4.0.

**ARTICLE HISTORY**

Received: 18 Sep 2026

Accepted: 01 Oct 2026

Published: 14 Oct 2026

CORRESPONDING AUTHOR

Parmeshwar Madhukar Patil, Department of Statistics, Sangameshwar College (Autonomous), Solapur, Maharashtra, India.

HOW TO CITEPatil, P. M., & Anyapanawar, R. R. (2026). Machine Learning Methods for Credit Risk Assessment in Microfinance: A Review. *Global Journal of Advanced Science and Technology*, 1(1), 1–6.

© 2026 The Author(s).

ABSTRACT

Microfinance institutions lend small amounts to borrowers who usually have no formal credit history, and they have long relied on group guarantees and the judgement of loan officers to manage the risk of default. As loan books grow and more data is captured digitally, statistical and machine learning models are increasingly used to support these decisions. This review examines the methods that have been applied to credit risk assessment, from logistic regression and discriminant analysis to decision trees, random forests, gradient boosting and neural networks, and considers how well they suit the conditions of microfinance. It also discusses three practical issues that often decide whether a model is useful: the imbalance between repaid and defaulted loans, the choice of evaluation measures, and the need to explain decisions to borrowers, loan officers and regulators. The literature shows that ensemble methods usually rank borrowers more accurately than single models, but the gains are often modest, and they come at a cost in transparency. For Indian microfinance and self-help group lending, the review argues that a well-specified logistic regression should remain the baseline, that tree ensembles are worth adopting where data volumes allow, and that explanation tools and careful validation matter as much as the choice of algorithm. Gaps include the scarcity of public microfinance datasets, limited use of group-level information, and few studies from India.

Keywords: Credit scoring; Microfinance; Machine learning; Random forest; Logistic regression; Self-help groups

1. Introduction

Microfinance grew out of a simple observation: poor households can be reliable borrowers even when they have no collateral and no credit record. Group lending, regular repayment meetings and close contact between loan officers and borrowers allowed early microfinance programmes to reach people whom

banks would not serve (Morduch, 1999). In India the same idea took the form of self-help groups (SHGs) linked to banks and of joint liability groups served by microfinance institutions.

These arrangements work well when lending is small and local, but they are expensive to scale. A loan officer can know a few hundred borrowers; a lender with several lakh accounts cannot rely on personal knowledge alone. Lenders therefore turn to credit scoring, the use of a statistical model to estimate the probability that a borrower will default from information available at the time of application or during the life of the loan (Thomas, 2000).

Credit scoring is well established in consumer and card lending, and the methods used there have been reviewed many times (Hand & Henley, 1997; Thomas, 2000). Its use in microfinance is newer and raises different questions. Data are thinner, borrowers are more homogeneous, repayment depends on group dynamics as well as on the individual, and the cost of a wrong rejection falls on households with few alternatives. Machine learning methods promise better predictions, but they also make decisions harder to explain.

This review brings together the main strands of this literature. It has three aims: to describe the statistical and machine learning methods that have been used for credit risk assessment; to examine the practical issues of data imbalance, evaluation and interpretability; and to identify what this means for microfinance and SHG lending in India.

2. Scope and approach of the review

The review is narrative rather than systematic. We concentrated on three kinds of work: benchmark studies that compare many classifiers on credit data, studies that build scoring models specifically for microfinance institutions, and methodological papers on techniques that are now standard in the field, such as random forests, gradient boosting, synthetic oversampling and model explanation. Studies were identified through Google Scholar, Scopus and the reference lists of the key papers. We give more weight to studies that report how models were validated, because accuracy figures without a clear validation design are difficult to compare.

3. Credit risk in microfinance

In conventional lending, a scorecard draws on a rich set of variables: income, employment history, past borrowing and repayment behaviour recorded by a credit bureau. Microfinance borrowers often have none of these on paper. The information available to the lender usually comprises basic demographics, the type and size of the household enterprise, the loan amount and purpose, membership of a group, and, for repeat borrowers, their repayment record with the same institution.

Two features of microfinance make risk assessment different. First, repayment is often a group outcome. Under joint liability, members cover for one another, so an individual default may be absorbed by the group and never appear in the data, while a shock to the group can lead several members to default together. Second, the lender's own staff hold a great deal of soft information about the borrower that never enters a database. Van Gool et al. (2012) examined whether credit scoring adds value in this setting and concluded that it can, provided the model complements rather than replaces the loan officer's judgement.

4. Methods for credit risk assessment

4.1 Statistical models

Linear discriminant analysis and logistic regression were the first methods used for credit scoring and remain the most common in practice (Hand & Henley, 1997). Logistic regression models the log-odds of default as a linear function of the borrower's characteristics. Its coefficients can be read directly, it produces probabilities that can be calibrated to observed default rates, and it can be converted into a

points-based scorecard that a branch manager can use without a computer. Its main limitation is that it captures non-linear effects and interactions only if the analyst builds them in, for example by grouping continuous variables into bands.

4.2 Decision trees and ensembles

A decision tree splits borrowers into groups using simple rules, such as loan amount above a threshold or fewer than two previous loans. Trees handle non-linearity and missing values naturally and are easy to explain, but a single tree is unstable: small changes in the data can produce a very different tree. Ensemble methods address this by combining many trees. A random forest grows each tree on a bootstrap sample of the data and a random subset of variables, and averages their predictions (Breiman, 2001). Gradient boosting builds trees one after another, each correcting the errors of the previous ones; XGBoost is an efficient and widely used implementation (Chen & Guestrin, 2016).

4.3 Neural networks and support vector machines

Neural networks and support vector machines can approximate complex relationships between borrower characteristics and default. In microfinance, Blanco et al. (2013) built multilayer perceptron models for a Peruvian microfinance institution and found that they classified borrowers more accurately, and at a lower cost of misclassification, than linear and quadratic discriminant analysis and logistic regression. Such models need more data and more careful tuning than logistic regression, and their predictions are harder to explain.

4.4 What benchmark studies show

Two large benchmark studies give the clearest picture of how these methods compare. Baesens et al. (2003) tested a wide range of classifiers on eight credit datasets and found that support vector machines and neural networks performed well, but that simple methods such as logistic regression and linear discriminant analysis were often close behind. More than a decade later, Lessmann et al. (2015) repeated the exercise with 41 classifiers and newer methods. They found that ensembles, especially heterogeneous ensembles that combine different types of model, gave the best results, and they recommended that random forest replace logistic regression as the standard benchmark for new studies.

Two cautions apply. The differences between the best methods are often small in absolute terms, and most benchmark datasets come from consumer lending in Europe and North America. Whether the same ranking holds for small, homogeneous microfinance portfolios is still largely untested. Table 1 summarises the strengths and limitations of the main methods.

Table 1. Main methods for credit risk assessment and their suitability for microfinance.

Method	Strengths	Limitations	Interpretability
Logistic regression	Stable; calibrated probabilities; easy scorecard	Misses interactions unless built in	High
Decision tree	Simple rules; handles missing data	Unstable; tends to overfit	High
Random forest	Accurate; robust to noise; little tuning	Many trees; less transparent	Medium (with SHAP)
Gradient boosting (XGBoost)	Often most accurate on tabular data	Needs careful tuning; can overfit small data	Medium (with SHAP)
Neural network	Captures complex patterns	Needs large data; hard to explain	Low
Support vector machine	Good with many variables	No direct probabilities; slow on large data	Low

Source: compiled by the authors from Hand and Henley (1997), Baesens et al. (2003), Lessmann et al. (2015) and Blanco et al. (2013).

5. Data, imbalance and evaluation

5.1 Class imbalance

Well-run microfinance portfolios have low default rates, often below five per cent. A model trained on such data can reach high accuracy simply by predicting that everyone repays. Common remedies include undersampling the repaid loans, oversampling the defaulted ones, adjusting class weights, or generating synthetic minority cases with SMOTE (Chawla et al., 2002). These techniques should be applied only to the training data; applying them before splitting the data inflates the reported performance.

5.2 Evaluation measures

Because accuracy is misleading on imbalanced data, credit models are judged mainly by how well they rank borrowers and how well their probabilities match observed default rates. The area under the ROC curve (AUC) and the Kolmogorov–Smirnov statistic measure ranking; the Brier score measures the quality of the predicted probabilities. Where the costs of approving a defaulter and rejecting a good borrower are known, a cost-based or profit-based measure is more informative than any statistical score (Lessmann et al., 2015). Table 2 lists the measures most often reported.

Table 2. Evaluation measures commonly used for credit scoring models.

Measure	What it tells the lender	Note
AUC (ROC curve)	How well the model ranks defaulters above good borrowers	0.5 = random; 1 = perfect
Kolmogorov–Smirnov	Maximum separation between good and bad score distributions	Common in industry practice
Brier score	Accuracy of predicted default probabilities	Lower is better
Sensitivity and specificity	Share of defaulters and good borrowers correctly identified	Depend on the cut-off chosen
Expected cost or profit	Financial effect of using the model at a given cut-off	Needs cost estimates

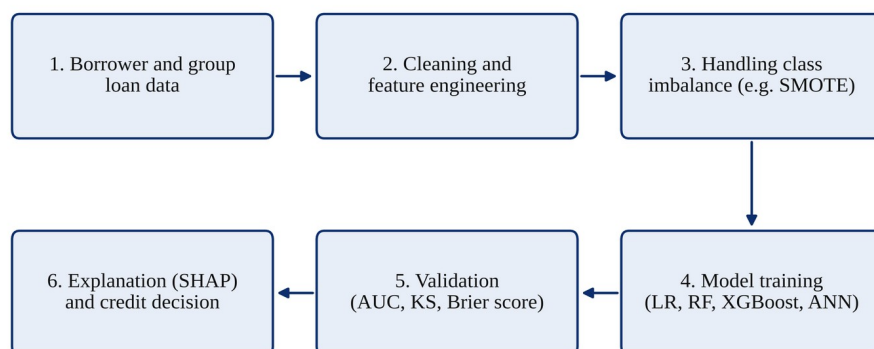


Figure 1. Typical workflow for building a credit scoring model for a microfinance portfolio.

Figure 1 sets these steps in order. In practice the validation step is where many published studies are weakest. A model should be tested on loans that were not used to build it, preferably loans issued later in time, because borrower behaviour and lending policy change. A model that performs well on a random hold-out sample may still fail when applied to next year's applicants.

6. Interpretability and fairness

A borrower who is refused a loan is entitled to know why, and a loan officer will not trust a score that contradicts what they know about a client without some explanation. This has made interpretability a central issue. Logistic regression and short decision trees explain themselves. For ensembles and neural networks, post-hoc methods are needed; SHAP values, which attribute each prediction to the contribution of each variable, are now the most widely used (Lundberg & Lee, 2017). SHAP can show, for example, that a particular applicant's score was lowered mainly by a large loan relative to the group's past borrowing.

Fairness is a related concern. Variables that predict default, such as location or household type, may also act as proxies for caste, religion or gender. A model that is accurate on average can still treat some groups unfairly. Checking performance and approval rates across groups, and excluding variables that cannot be justified, should be part of model development rather than an afterthought.

7. Implications for Indian microfinance and SHG lending

Four points follow from the literature. First, logistic regression remains a sound starting point. It is transparent, it performs close to more complex models in many settings, and it gives a benchmark against which any more complex model must justify itself. Second, where lenders have tens of thousands of loans and a reasonable number of defaults, random forests and gradient boosting are worth testing, provided they are validated on later loans and accompanied by explanations. Third, group-level variables, such as the group's age, its repayment history, attendance at meetings and the share of members with outstanding loans elsewhere, are likely to carry much of the predictive information in SHG and joint-liability lending, yet they are rarely used in published models. Fourth, scoring should support the loan officer rather than replace them: the most useful systems flag cases for closer review rather than issue automatic rejections.

8. Research gaps

- There are very few public microfinance datasets, which makes it hard to compare methods or reproduce results.
- Studies from India, and especially from SHG–bank linkage lending, are scarce compared with studies from Latin America and Europe.
- Group dynamics and joint liability are seldom modelled explicitly, for example with multilevel or network models.
- Most studies evaluate models statistically rather than by their financial effect on the lender or the borrower.
- Fairness across social groups is rarely tested in microfinance scoring models.

9. Conclusion

Machine learning has improved the ability of lenders to rank borrowers by risk, and ensemble methods such as random forests and gradient boosting now set the standard in benchmark studies. In microfinance, however, the gains from complex models are often small, data are limited, and the need for explanation is strong. The method matters less than the care with which a model is built, validated and used. For Indian microfinance institutions and banks lending to SHGs, the most productive path is likely to combine a transparent baseline model, well-validated ensembles where data allow, explanation tools such as SHAP, and the knowledge of field staff. Building shared, anonymised datasets would do more for research in this area than any new algorithm.

Author contributions: P. M. Patil: conceptualisation, methodology of the review, writing – review and editing. R. R. Anyapanawar: literature search, writing – original draft, preparation of tables.

Conflict of interest: The authors declare no conflict of interest.

Funding: This research received no external funding.

Declaration of generative AI use: During the preparation of this work the authors used an AI writing assistant (Claude, Anthropic) to help structure and draft the text. The authors reviewed, edited and verified the content and take full responsibility for this article.

References

- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627–635. <https://doi.org/10.1057/palgrave.jors.2601545>
- Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*, 40(1), 356–364. <https://doi.org/10.1016/j.eswa.2012.07.051>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A*, 160(3), 523–541. <https://doi.org/10.1111/j.1467-985X.1997.00078.x>
- Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136. <https://doi.org/10.1016/j.ejor.2015.05.013>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* 30 (pp. 4765–4774).
- Morduch, J. (1999). The microfinance promise. *Journal of Economic Literature*, 37(4), 1569–1614. <https://doi.org/10.1257/jel.37.4.1569>
- Thomas, L. C. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2), 149–172. [https://doi.org/10.1016/S0169-2070\(00\)00034-0](https://doi.org/10.1016/S0169-2070(00)00034-0)
- Van Gool, J., Verbeke, W., Sercu, P., & Baesens, B. (2012). Credit scoring for microfinance: Is it worth it? *International Journal of Finance & Economics*, 17(2), 103–123. <https://doi.org/10.1002/ijfe.444>