

Credit Risk Assessment in Digital Lending: Evaluating Alternative Data vs. Traditional Credit Scoring in the Indian Banking System

¹Vaibhav C. Bendewar, ²Dr. Ujwala Narkhede

¹Department of Management Studies, Wainganga College of Engineering and Management Nagpur, Maharashtra, India

²Associate Professor & Research Guide, Wainganga College of Engineering and Management Nagpur, Maharashtra, India.

Email: vaibhavbendewar123@gmail.com, ujwalanarkhede@wcemnagpur.in

Abstract:

The rapid growth of digital lending platforms and financial technology firms has transformed retail credit delivery across India. Traditional commercial banks and non-banking financial companies (NBFCs) have historically depended on centralized bureau scores, such as TransUnion CIBIL, along with formal salary slips and collateral security to evaluate borrower credit risk. While this conventional framework works well for salaried professionals in formal sectors, it leaves out millions of self-employed individuals, gig workers, small merchants, and first-time borrowers who lack formal credit records. In response, digital lending entities now evaluate alternative data sources, including Unified Payments Interface (UPI) transaction flows, utility bill settlements, goods and services tax (GST) filings, and Account Aggregator statements, using machine learning models to assess repayment capacity. This study compares traditional credit scoring against alternative data-driven scoring models within the Indian banking ecosystem. Analyzing loan application and repayment data from 12,500 retail borrowers, the study evaluates predictive performance, credit approval rates for thin-file applicants, and twelve-month default outcomes across traditional, alternative, and hybrid underwriting approaches. The results show that relying only on bureau scores yields an AUC-ROC of 0.734 with a thin-file approval rate of just 19.2%. When consent-based alternative cash flow data is combined with bureau records in a hybrid model, the AUC-ROC improves to 0.882, thin-file approvals reach 56.4%, and portfolio defaults drop to 3.2%. The paper concludes with recommendations for Indian lending institutions to adopt compliant hybrid underwriting aligned with Reserve Bank of India (RBI) guidelines.

Keywords— *Alternative Data, Credit Risk, Digital Lending, Indian Banking System, CIBIL, Financial Inclusion.*

I. INTRODUCTION

Access to timely and affordable credit is vital for economic development, job creation, and personal financial security in India. For decades, the Indian banking system—comprising scheduled commercial banks, regional rural banks, and non-banking financial companies (NBFCs)—has relied on conservative credit appraisal standards. Loan approvals have traditionally required documented income proof, three years of Income Tax Returns (ITR), salary slips, audited business balance sheets, and formal bureau scores from agencies like TransUnion CIBIL, Experian, or Equifax.

Although these criteria protect lenders against willful defaults, they create a major barrier for a large segment of the population. India's vast informal workforce, comprising street vendors, freelance professionals, gig-economy workers, and rural micro-entrepreneurs, earns reliable

incomes but rarely holds the formal paper documentation required by traditional bank branch managers. Consequently, these "thin-file" or "new-to-credit" individuals are frequently denied bank loans, forcing them to turn to unorganized moneylenders who charge exorbitant interest rates.

Over the last decade, India's digital public infrastructure—often referred to as the India Stack—has fundamentally reshaped financial services. The widespread rollout of Aadhaar-based identification, affordable 4G/5G mobile connectivity, Jan Dhan bank accounts, and the Unified Payments Interface (UPI) has shifted daily business and retail transactions into digital formats. Local kirana store owners and delivery partners now handle daily commercial turnover through QR code payments rather than cash.

Digital lending platforms and FinTech-partnered NBFCs recognized that these digital

footprints contain valuable behavioral information about a borrower's income consistency, cash flows, and bill payment discipline. Using predictive machine learning algorithms, digital lenders began issuing collateral-free micro-loans within minutes based on alternative data inputs. However, questions remain regarding how well these alternative models predict actual defaults compared to traditional bureau scores over an entire economic cycle. In addition, aggressive app permissions and unstructured data collection prompted the Reserve Bank of India (RBI) to establish formal Digital Lending Guidelines. This paper examines the practical performance, benefits, and regulatory considerations of using alternative data alongside traditional credit scoring in India.

II. LITERATURE REVIEW

The core problem of credit lending has long been framed by George Akerlof's study on information asymmetry and Joseph Stiglitz and Andrew Weiss's analysis of credit rationing. When lenders cannot tell creditworthy borrowers apart from risky ones due to missing financial information, they either charge high interest rates across the board or ration credit entirely. In India, public and private sector banks set up centralized credit bureaus to reduce this information gap. Studies confirm that bureau scores, notably CIBIL scores ranging between 300 and 900, significantly reduce default rates by penalizing late repayments and high credit utilization.

However, research on financial inclusion in India points out that credit bureaus are inherently backward-looking. A borrower must already hold a formal loan or credit card to build a bureau score. As a result, millions of enterprising individuals remain locked out of formal credit simply because they have never taken a bank loan before.

In recent years, researchers worldwide have examined whether non-traditional digital footprints can fill this gap. Studies analyzing digital transactions have found that continuous daily cash inflows, consistent telecom bill settlements, and stable operating habits provide dependable clues about an individual's repayment willingness and financial health. In India, the adoption of GST filings and merchant UPI transaction data has allowed lenders to evaluate micro-business revenue without waiting for annual audited statements.

At the same time, academic literature cautions against treating alternative data as an absolute replacement for traditional financial metrics. Standalone alternative algorithms can misjudge credit risk during economic disruptions, when non-financial behaviors may not match real debt repayment capacity. Furthermore, early digital lending apps relied on controversial data points, such as reading customer phone contacts, SMS messages, and device locations. This raised serious consumer privacy concerns and sparked regulatory interventions. Scholars now emphasize that the most sustainable lending framework combines permissioned financial data—facilitated by India's Account Aggregator ecosystem—with core bureau metrics to balance credit expansion with institutional safety.

III. RESEARCH METHODOLOGY

This study uses an empirical, comparative approach to evaluate credit risk assessment models using loan performance data from the Indian retail lending market.

A. Data Scope and Variables

The dataset consists of 12,500 retail unsecured personal and micro-business loan applications processed between January 2023 and December 2024 by an Indian digital lending partner. Borrowers were tracked over a 12-month post-disbursement window to monitor repayment status.

The features are divided into two primary input categories:

1) *Traditional Credit Variables*: TransUnion CIBIL credit score, past repayment history (delinquencies categorized into 30, 60, and 90+ days past due), total existing active loan lines, credit card utilization ratio, and length of formal credit history.

2) *Alternative Data Variables*: Monthly UPI transaction volume and inflow consistency, utility bill payment timeliness (electricity, piped gas, water), mobile recharge and postpaid billing records, average daily minimum balance derived via Account Aggregator (AA) banking feeds, and GST return filing regularity for small enterprises.

3) *Target Variable*: Loan default status, defined as non-repayment exceeding 90 days past due (90+ DPD) within 12 months from the loan disbursement date.

B. Model Formulations

Three practical credit evaluation models were developed and compared:

- **Model 1 (Traditional Bureau Approach):** A standard Logistic Regression scoring model based exclusively on bureau scores, past delinquency history, and declared monthly income.
- **Model 2 (Alternative Data Only):** An Extreme Gradient Boosting (XGBoost) model trained entirely on UPI transaction velocity, utility bill consistency, telecom payment regularity, and Account Aggregator cash-flow trends.
- **Model 3 (Hybrid Framework):** An ensemble model that combines traditional bureau history (where available) with consent-based Account Aggregator bank cash flows and alternative payment behaviors.

C. Evaluation Criteria

Each model's performance was measured using four standard metrics: Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Precision (accuracy in identifying actual defaulters), Recall (ability to catch true default cases), and the 12-month Non-Performing Asset (NPA) default rate observed across approved applicants. Additionally, the thin-file approval rate was calculated to evaluate each model's contribution toward financial inclusion.

IV. RESULTS AND DISCUSSION

The experimental evaluation shows clear differences in predictive accuracy and lending reach across the three approaches, as detailed in Table I.

TABLE I: MODEL PERFORMANCE IN INDIAN BANKING CONTEXT

Metric	Model 1 (Bureau)	Model 2 (Alt Data)	Model 3 (Hybrid)
AUC-ROC Score	0.734	0.819	0.882
Precision (Defaulters)	0.678	0.746	0.831
Recall (Detection Rate)	0.612	0.785	0.849
Thin-File Approval (%)	19.2%	62.5%	56.4%
12-Month NPA Rate (%)	6.1%	4.9%	3.2%

The traditional bureau model (Model 1) achieved an AUC-ROC score of 0.734. While it reliably assesses salaried borrowers with established banking relationships, it performs poorly with new-to-credit applicants. Because applicants without a CIBIL history receive a "-1" or "0" score, the traditional model rejected more than 80% of thin-file applicants, approving only 19.2%. This confirms that conventional scoring continues to exclude creditworthy informal workers.

In contrast, the alternative data model (Model 2) raised the AUC-ROC to 0.819 and approved 62.5% of thin-file applicants. Inflow frequency on UPI and consistent electricity bill payments proved to be dependable indicators of liquid cash availability and financial discipline. However, because Model 2 lacked visibility into existing non-digitized borrowings or overall debt burden, it experienced some false positives during irregular earning months, yielding an NPA rate of 4.9%.

The hybrid model (Model 3) delivered the most balanced and dependable outcome. It achieved the highest AUC-ROC at 0.882 and reduced the observed NPA default rate to 3.2%, while maintaining a strong 56.4% approval rate for thin-file applicants. For applicants with an existing bureau record, alternative data offered near-real-time updates on cash-flow health, eliminating the typical 30-day reporting delay seen in bureau updates. For first-time borrowers, cash-flow verification via Account Aggregator accounts confirmed repayment ability without requiring collateral.

From a regulatory perspective, the findings support the Reserve Bank of India's updated Digital Lending Guidelines. Early lending applications gathered invasive device details such as contacts, media files, and real-time location. Our analysis indicates that these invasive variables offer minimal predictive benefit compared to structured financial cash flows from UPI, utility payments, and Account Aggregator records. Using verified financial transaction trails protects consumer privacy while complying with the Digital Personal Data Protection (DPDP) Act.

V. CONCLUSION

This study evaluated the role of alternative data in credit risk assessment within the Indian banking and digital lending sector. Traditional credit bureau metrics remain valuable for prime

borrowers, but relying on them exclusively restricts credit access for self-employed individuals, micro-enterprises, and gig workers. Standalone alternative data models significantly improve financial inclusion, but they require sound cash-flow verification to manage default risk over economic cycles.

The most effective strategy for Indian banks and NBFCs is a Hybrid Credit Risk Assessment model. By combining baseline credit bureau records with consent-based digital cash-flow data through India's Account Aggregator framework, lenders can safely serve thin-file borrowers while lowering default rates. Adopting this structured, privacy-focused approach allows financial institutions to support responsible financial inclusion while safeguarding portfolio quality and complying with regulatory standards.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Management Studies, Wainganga College of Engineering and Management, Nagpur, for providing academic guidance, computational resources, and support during the preparation of this research paper.

REFERENCES

- [1] G. A. Akerlof, "The market for 'lemons': Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500, 1970.
- [2] A. Banerjee and E. Duflo, "Do firms want to borrow more? Testing credit constraints using a directed lending program," *The Review of Economic Studies*, vol. 81, no. 2, pp. 572–607, 2014.
- [3] T. Berg, V. Burg, A. Gombović, and M. Puri, "On the rise of FinTechs: Credit scoring using digital footprints," *The Review of Financial Studies*, vol. 33, no. 7, pp. 2845–2897, 2020.
- [4] J. Jagtiani and C. Lemieux, "The roles of alternative data and machine learning in FinTech lending: Evidence from the LendingClub consumer platform," *Financial Management*, vol. 48, no. 4, pp. 1009–1029, 2019.
- [5] L. J. Mester, "What's the point of credit scoring?," *Federal Reserve Bank of Philadelphia Business Review*, vol. 3, pp. 3–16, 1997.
- [6] M. Óskarsdóttir, C. Bravo, C. Sarraute, J. Vanthienen, and B. Baesens, "The value of big data for credit scoring: Improving financial inclusion with alternative data," *Decision Support Systems*, vol. 125, p. 113119, 2019.
- [7] Reserve Bank of India, "Guidelines on digital lending," *RBI/2022-23/111 DOR.CRE.REC.66/21.07.001/2022-23*, Reserve Bank of India, 2022.
- [8] R. Sahay, U. von Kleist, N. Ni, and M. Cihak, "The promise of FinTech: Financial inclusion in the post-COVID-19 era," *IMF Working Paper*, no. 20/190, 2020.
- [9] J. E. Stiglitz and A. Weiss, "Credit rationing in markets with imperfect information," *The American Economic Review*, vol. 71, no. 3, pp. 393–410, 1981.