Credit Risk Assessment in Banking Using Machine Learning

Pratyush Kumar Tripathi
Master of Business Administration
Galgotias University, Greater Noida, Uttar Pradesh, India
Email: tripathipratyush84@gmail.com

ABSTRACT

In the modern digital economy, the banking sector is undergoing a transformation, driven by rapid technological advancements and evolving customer expectations. At the heart of this transformation lies the pressing need to assess and mitigate credit risk more accurately and efficiently. Traditional credit scoring systems—although functional—often struggle to capture the nuances of today's dynamic financial environments. With the increasing availability of alternative and behavioural data, Machine Learning (ML) has emerged as a revolutionary force in redefining credit risk assessment. This research paper explores the practical and theoretical dimensions of using ML in Indian banking, with a focus on improving accuracy, reducing Non- Performing Assets (NPAs), and ensuring financial inclusion. Through a detailed survey involving 52 professionals and a comprehensive review of secondary literature, this study provides a grounded perspective on the current status, challenges, and future potential of ML-driven credit assessment. It not only identifies key technical and regulatory bottlenecks but also presents human-centric insights into how professionals perceive this shift. The findings underline a critical balance between automation and ethical oversight, urging banks to integrate ML responsibly for long-term resilience and trust.

Keywords: NPA, Machine learning

I. INTRODUCTION

Setting the Context

Imagine a loan officer in a small-town Indian bank, making decisions based on a customer's income proof, credit bureau report, and a gut feeling. Now contrast this with a fintech app in Mumbai that uses a borrower's smartphone behaviour, digital payments, and social network to approve a loan in 30 seconds. This contrast illustrates the technological leap banking has taken—and also the disparity in adoption. Credit risk, one of the most fundamental aspects of lending, has now moved beyond traditional scorecards to the realm of algorithms, real-time data, and intelligent systems.

The Growing Complexity of Credit Risk

The rise of MSMEs, gig economy workers, and first-time borrowers—many of whom operate outside formal income structures—has made creditworthiness harder to assess through conventional means. Add to this the regulatory push for financial inclusion, and banks face a dilemma: how to lend more while risking less? This is where Machine Learning promises a way forward. By analysing vast and varied datasets, ML algorithms can predict default probabilities with greater accuracy and nuance.

Why This Research Matters

India, with its rapidly digitizing economy and diverse customer base, presents both an opportunity and a challenge for credit risk assessment. While private banks and fintech firms are actively exploring ML, public sector banks—responsible for financial outreach—often lag behind due to legacy systems and compliance concerns. This research aims to map this gap, understand the barriers, and recommend a path toward scalable, transparent, and inclusive ML adoption.

Personal Motivation

As a student of business and technology, I've always been fascinated by the intersection of data and decision-making. Working on this topic has deepened my understanding of how transformative ML can be—not just for corporate profits, but for common people seeking credit for education, homes, or business dreams. I hope this paper contributes meaningfully to ongoing discussions in academia and industry alike.

II. LITERATURE REVIEW

Traditional Credit Risk Models: Strengths and Shortcomings

Credit risk has historically been assessed using statistical models based on financial ratios, repayment history, and credit bureau scores. While these models are explainable and easy to implement, they assume linear relationships and often ignore contextual factors. For instance, a salaried professional with a poor repayment history might be penalized more heavily than a small business owner with irregular but steady income—something traditional models struggle to distinguish.

Machine Learning: A Game-Changer

ML has the ability to ingest thousands of features, spot non-linear patterns, and adjust itself over time. Algorithms such as Random Forest, Gradient Boosting, and Support Vector Machines are already being used in countries like the U.S. and China to underwrite loans in real-time. In India, firms like CASHe, ZestMoney, and KreditBee use ML to assess credit for new-to-credit borrowers based on behavioral data.

Behavioural and Alternative Data: A New Frontier

Mobile phone usage, spending behaviour, geolocation, and even typing speed are now used in psychometric models to evaluate trustworthiness. These alternative data points help include those excluded by traditional scoring, such as migrant workers, women entrepreneurs, and rural borrowers. ML enables this by processing vast, unstructured data and finding patterns that correlate with repayment behaviour.

Ethical and Regulatory Considerations

There are concerns. What if an algorithm is biased against a certain region or gender due to skewed training data? How do borrowers appeal a decision made by a black-box model? These questions have prompted regulators like the RBI to caution against over-reliance on opaque systems. Recent global discussions around Responsible AI underscore the importance of fairness, accountability, and transparency.

Gaps in Existing Research

Most academic studies focus on algorithm performance—AUC, precision, recall—without exploring how frontline staff and regulators interact with these models. There is limited Indian research on organizational readiness, training gaps, and real-world deployment experiences. This paper aims to bridge that gap.

Past research highlights the significance of branding, quality, and consumer trust in the FMCG industry. MahaboobBasha (2020) emphasized the role of customer satisfaction in improving sales. Ganesh (2019) discussed how brand loyalty is driven by consumer perception and quality. Thanisorn and Byaporn (2018) showed how packaging and promotions affect purchase behavior, particularly in the herbal cosmetics market.
Additionally, Sarfaraz and Pratik (2018) illustrated that even private label products can gain a foothold if marketed as quality alternatives. The literature reveals that buying behavior is shaped by multiple factors, including socio-demographics, lifestyle, awareness, and cultural alignment with the product.

III. RESEARCH METHODOLOGY

Purpose and Approach

The research aims to provide a human-centric understanding of ML in credit risk assessment. It uses a combination of descriptive and exploratory approaches. Descriptive elements help quantify adoption, while exploratory elements capture narratives, perceptions, and organizational behaviours.

Target Respondents

Banking professionals, fintech consultants, data scientists, credit analysts, and risk managers formed the core of the sample. Their diversity offered a 360-degree view of challenges and opportunities.

Sampling Strategy

A mix of purposive and snowball sampling was used. Respondents were selected from professional circles, LinkedIn groups, and industry referrals.

Data Collection Tools

A 13-question online survey via Google Forms was used. It included:

- Multiple choice questions
- Likert scale assessments
- One open-ended question

Analysis Techniques
Quantitative responses were coded, tabulated, and visualized using Excel. Qualitative responses were thematically analysed to extract deeper insights.

Validity and Ethics
All responses were anonymous. Participants were informed about the academic nature of the research. Bias was minimized by avoiding leading questions and allowing space for dissenting views.

IV. DATA ANALYSIS

Professional Roles
Out of 52 respondents, 33% were risk officers, 27% were credit analysts, 19% were data scientists, and the rest included consultants, auditors, and operations managers.

Experience with ML
58% had less than one year of experience applying ML in banking. This suggests that ML adoption is still at a learning stage across most institutions.

Methods Used for Credit Scoring
Rule-based scoring was used by 61%, while only 16% reported active use of ML tools. 23% were in transition or pilot phase.

Types of Data Used
33% used a mix of demographic, behavioral, and transactional data. This marks a shift from over-reliance on traditional variables like income and CIBIL scores.

Familiarity with ML Algorithms
Decision Trees (CART, C4.5) were most popular (74%), followed by Logistic Regression (68%), Random Forest (45%), and Neural Networks (18%).

Confidence in ML Models
Only 23% of respondents were very confident in using ML outputs for lending decisions. This emphasizes the need for transparent, explainable systems.

Barriers to Adoption
Key challenges include:

- Lack of skilled personnel (31%)
- Poor data quality (27%)
- Regulatory uncertainty (21%)
- Model explainability (14%)

Validation Techniques
Manual cross-verification is still common. 54% of banks validate ML decisions through traditional underwriting steps, indicating hybrid model usage.

Perceived Benefits of ML
Respondents highlighted:

- Faster credit scoring (46%)
- Reduced NPAs (41%)
- Improved borrower segmentation (39%)
- Better fraud detection (35%)

Training Exposure
Only 18% had formal training in ML applications for banking. Most learned through self-study or on-the-job exposure.

V. FINDINGS

- ML is primarily viewed as a supportive tool rather than a standalone risk engine.
- Indian banks show strong interest but weak implementation maturity.
- Public sector banks lag behind private banks and fintech firms.
- There's increasing focus on behavioral analytics and alternative data.
- Training, regulatory guidance, and model explainability are pivotal for adoption.
- ML is primarily viewed as a supportive tool rather than a standalone risk engine.
- Indian banks show strong interest but weak implementation maturity.
- Public sector banks lag behind private banks and fintech firms.

- There's increasing focus on behavioral analytics and alternative data.
- Training, regulatory guidance, and model explainability are pivotal for adoption.

VI. CONCLUSION

The adoption of Machine Learning in Indian banking, especially for credit risk assessment, holds transformational potential. While challenges around regulation, skills, and infrastructure persist, the direction is clear. Institutions that invest early in data readiness, human capital, and explainable models will gain competitive advantage. The findings suggest that ML is not a replacement but a reinforcement to traditional underwriting, especially when transparency and ethics are prioritized. Moving forward, banks must align their digital transformation with responsible AI principles to ensure inclusive and sustainable growth.

VII. RECOMMENDATIONS

- Structured Training Programs Banks must invest in ML upskilling through partnerships with institutions like NIBM, IITs, or online platforms. Specialized modules on model validation, compliance, and explainability are crucial.
- Model Governance Frameworks Develop internal policies to monitor model performance, bias, and transparency. Regulatory compliance should be central to ML deployment.
- Explainable AI Tools Implement SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) to demystify black-box models.
- Regulatory Sandboxes RBI and SEBI should expand sandbox environments to allow innovation with supervised oversight.
- Collaborative Ecosystems Create fintech-bank collaborations for mutual learning and faster ML implementation. Banks can leverage fintech agility while providing compliance frameworks.

VIII. REFRENCES

Reserve Bank of India. (2023). Financial Stability Report.

NASSCOM & BCG. (2023). AI in Indian Banking.

Reddy, A., & Rao, K. (2021). ML Algorithms in Loan Default Prediction.

Singh, R., & Kumar, P. (2020). Credit Scoring in Cooperative Banks.

Das, S., & Iyer, R. (2020). Bias in ML Models for Credit Risk.

Bajaj Finserv Case Study. (2023).

ICICI Bank. (2023). iLens Automation Report.

CRISIL. (2022). NPA Trends and Analytics.

NIT Trichy. (2019). ML Use in Retail Banking.