CreditIQ Analytics: AI-Powered Credit Risk & Alternative Data Platform for Indian NBFCs

Team: Risk Visionaries
Girish Basavaraj Hiremath (24MDA030)
Poojitha C. G. (24MDA060)
Nikhita (24MDA082)
Vanithaa Ranga (24MDA092)

August 2025

Abstract

This report details *CreditIQ Analytics*—a FinTech platform purpose-built for Indian Non-Banking Financial Companies (NBFCs). CreditIQ fuses traditional bureau metrics with alternative data sources (UPI, GST, utilities, mobile metadata) to underwrite thin-file or credit-invisible borrowers, provide early-warning default signals, and automate compliance aligned with the Reserve Bank of India (RBI) Core Financial Services Solution (CFSS) mandate. We present the problem context, regulatory landscape, system architecture, machine-learning (ML) methodology, implementation roadmap, market sizing, competitive positioning, risks, and an evidence-based business case.

Keywords: NBFC, CFSS, Digital Lending, Alternative Data, Credit Scoring, PD/LGD, Explainability, Account Aggregator, India Stack.

Contents

1. Area of Focus

Domain: Lending (NBFC Sector).
Mandate Context: RBI CFSS adoption is mandatory for NBFC-ML and NBFC-UL by 30 September 2025. CreditIQ targets these institutions with cloud-native analytics and compliance tooling, delivered as APIs and dashboards.

2. Problem Statement and Idea Description

Indian NBFCs face: (i) limited visibility into thin-file borrowers, (ii) manual and siloed underwriting, (iii) high compliance overheads, and (iv) lagging risk surveillance. **CreditIQ** addresses these via:

- **Composite CreditIQ Score:** Real-time credit-risk scoring that fuses bureau data with alternative signals (UPI payment behaviour, GST returns, utility bill history, mobile usage patterns).

- **Early-Warning & Stress Testing:** Portfolio and borrower-level anomaly detection, macro and sectoral stress scenarios, and challenger-model monitoring.

- **Compliance Automation:** MIS and audit trails aligned to CFSS, Digital Lending Guidelines (2022), and DPDP Act (2023).

Quantified value (indicative):

- Up to **25%** lower default rates through richer data signals.

- **60%** faster loan approval cycles (hours → minutes), with **40%** lower credit-operations cost.

- **15%** expansion in addressable borrower base among thin-file segments.

3. Domain Understanding and Value-Chain Fit

3.1 Current Lending Value Chain

Borrower onboarding → KYC → Credit appraisal → Sanction & disbursement → Servicing → Collections → Reporting. Pain points include manual underwriting, fragmented data, and evolving regulatory requirements.

3.2 CreditIQ Insertion Points

1. Post-KYC, generate the **CreditIQ Score** (PD/LGD-ready) and explainability artefacts.

2. Pre-sanction, feed the decisioning engine with PD/LGD, policy rules, and affordability checks.

3. Post-disbursement monitoring: behavioural drift, transaction anomalies, delinquency flags, and collection prioritisation.

4. Geography and Regulation

Geography: India.
Regulators: Reserve Bank of India (RBI), Ministry of Electronics & IT (MeitY), Goods & Services Tax Network (GSTN).
Key Instruments:

- RBI **Scale-Based Regulation** and **CFSS Circular** (23 Feb 2022)—mandatory CFSS by 30 Sep 2025 for NBFC-ML/UL.
- **Digital Lending Guidelines** (2022): consented data use, FLDG caps, disclosure norms.
- **Digital Personal Data Protection (DPDP) Act** (2023): data minimisation, purpose limitation, storage and consent obligations.
- **Enablers:** RBI regulatory sandbox, Account Aggregator (AA) framework, OCEN APIs.

5. Technology Architecture

5.1 High-Level Components

- **Data Ingestion Layer:** APIs for credit bureaus, AA, UPI/IMPS rails, GSTN, and utility billers.
- **Feature Store:** Scalable store (vector/feature DB) with 500+ engineered features and lineage.
- **ML Model Hub:** Ensemble Gradient Boosting + DNN stack; monthly retraining; bias/drift monitoring.
- **Decision Engine:** Real-time PD/LGD, policy rules, affordability, and explanations.
- **Compliance & MIS:** Auto-populated RBI reports, audit trails, and evidence logs.
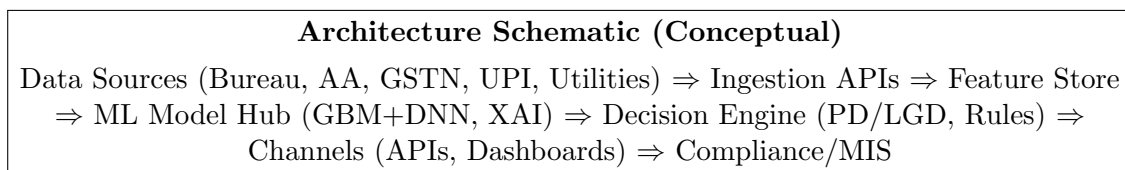

Architecture Schematic (Conceptual)

Data Sources (Bureau, AA, GSTN, UPI, Utilities) $\Rightarrow$ Ingestion APIs $\Rightarrow$ Feature Store $\Rightarrow$ ML Model Hub (GBM+DNN, XAI) $\Rightarrow$ Decision Engine (PD/LGD, Rules) $\Rightarrow$ Channels (APIs, Dashboards) $\Rightarrow$ Compliance/MIS

Figure 1: CreditIQ system dataflow and components (conceptual view).

5.2 Data Sources

Core banking/CFSS, bureau pulls, GST returns, phone metadata, utility bills, social and location metrics. Sample CSVs and a Colab notebook accompany this submission for demonstration on synthetic data.

5.3 ML and Governance

- **Feature Engineering:** Payment regularity indices, income proxies from GST, volatility of UPI flows, bill-payment punctuality, device-risk scores.
- **Modelling:** Gradient boosting trees for tabular signals; feed-forward DNNs for interaction terms; calibrated outputs (Platt/Isotonic) for PD.
- **Explainability:** SHAP/feature attributions, counterfactual reason codes, policy override logs.
- **MLOps:** Model registry, CI/CD, data/version lineage, drift alarms, challenger models.

6. Customer and User Behaviour

Primary users: Credit risk analysts, underwriting teams, CROs.
Buyers: NBFCs (SaaS licence), co-lending fintechs (per-call API pricing).

6.1 Pricing and ROI

- Platform: **12 lakh/year**

- Per-application scoring: **20** (volume-tiered)

- Portfolio monitoring (optional): **3 lakh/year**

Projected ROI: save ∼40 lakh/year in manual credit-ops and reduce credit cost by ∼30 bps on AUM via better risk selection and faster cycles.

7. Market and Strategic Analysis

7.1 Market Sizing

Table 1: Market Metrics and Forecasts (illustrative)

Metric	2024	2025F	2026F	2027F
NBFC Credit Stock (tn)	52	60	70	82
Digital Lending Share (%)	60	68	75	80
TAM for Credit-Analytics SaaS (cr)	650	850	1100	1400

Market share capture goal: 0.9% (8 cr) in 2025; 2% (22 cr) in 2026; 4% (56 cr) in 2027.

7.2 Competitive Landscape

Key players: Perfios, Finezza, Credolabs, Experian (Alt-Data).
Differentiators: India-Stack integrations (AA/OCEN), NBFC-first compliance, rapid time-to-value, explainability-first approach.

7.3 Porter's Five Forces (2025 vs 2027)

Table 2: Five Forces Trajectory

Force	2025 Intensity	2027 Intensity	Trend
Threat of New Entrants	Medium (tech barriers)	High (API commoditisation)	↑
Supplier Power (data providers)	High	Medium	↓ (AA opens access)
Buyer Power (NBFCs)	Medium	Medium-High	↑
Substitutes (in-house models)	Low	Medium	↑
Rivalry	Medium-High	High	↑

7.4 SWOT (2025)

Table 3: SWOT Summary

Strengths	Weaknesses
Proprietary alt-data models; CFSS-ready; fast deployment.	Reliance on external data licences; integration dependencies.
Opportunities	**Threats**
ECL transition; rural NBFC growth; co-lending expansion.	Data-privacy risks; intensifying fintech competition.

8. Implementation Plan

8.1 12-Month Build (Seed Stage)

- v1 platform with ingestion, feature store, baseline models, and decision engine.

- Secure 3 pilot NBFCs; deliver PD/LGD, monitoring dashboards, and CFSS-aligned MIS.

- Governance: privacy-by-design, DPIA, security controls (KMS, tokenisation, VPC peering).

8.2 Scale-Up (Series A, 2026)

- Expand to 40 NBFCs; regional clusters; API marketplace distribution.

- Partnerships: GSTN, NPCI, AA TSPs, cloud providers, co-lenders.

9. Business Model and Financials

Pricing as per Section 6.2.
Revenue trajectory: focus on platform ARR + usage-based scoring; optional monitoring module for stable ARR; services for onboarding and model customisation.

10. Risk Management

Strategic and operational risks with mitigations:

- **Regulatory Change:** Maintain compliance library; automated impact assessment; sandbox participation.

- **Data Privacy/Security:** Consent management, encryption at rest/in transit, strict access controls, regular audits.

- **Model Risk:** Backtesting, calibration, drift detection, human-in-the-loop overrides.

- **Vendor Risk:** Multi-source data providers; resiliency engineering; SLAs.

11. Conclusion

CreditIQ enables NBFCs to responsibly expand credit access while tightening risk controls and streamlining compliance. With India-Stack integrations, explainable ML, and CFSS-ready MIS, the platform is positioned to deliver measurable uplift in approval speed, loss mitigation, and operational efficiency.

Appendix A: LLM Usage

ChatGPT was used for strategic option evaluation, synthetic data generation guidelines, and regulatory summarisation. Generated artefacts include sample CSVs and a demonstration notebook for model training and validation on synthetic data.

Appendix B: References (Selected Web Sources)

The following references informed the regulatory and market framing of this report (URLs provided for convenience):

1. https://www.nelito.com/blog/understanding-the-rbis-cfss-mandate-for-nbfcs.html

2. https://fidcindia.org.in/wp-content/uploads/2022/02/RBI-CORE-FIN-SERV-SOLUTIONS-23-02-22.pdf

3. https://www.bankingfinance.in/the-role-of-credit-scoring-and-alternative-lending-in-indias-financial-inclusion.html

4. https://precisa.in/alternative-credit-scoring-using-bank-statement-analysis/

5. https://finezza.in/blog/alternative-credit-scoring-in-india/

6. https://newgensoft.com/in/blog/simplifying-rbis-mandate-for-nbfcs-to-implement-core-financial-services-solution-cfss/

7. https://rvks.in/blogs/implementation-of-core-financial-services-solution-by-non-banking-financial-companies-nbfcs/

8. https://faceofindia.org/wp-content/uploads/2025/04/State-of-Alternative-Data-in-India-A-Lending-Use-Case.pdf

9. https://www.linkedin.com/pulse/12-powerful-insights-alternative-credit-scoring-india-driving-oa5bc

10. https://www.gnani.ai/resources/blogs/digital-lending-in-nbfcs-projected-to-grow-at-a-cagr-of-25-through-2025/

11. https://www.rbi.org.in/scripts/NotificationUser.aspx?Id=12247Mode=0

Appendix C: Colab Notebook Reference

The complete model training, statistical analysis, and explainability workflow for CreditIQ Analytics is implemented in the following Google Colab notebook:
Notebook Link: CreditIQ_Model_Training.ipynb
This notebook includes:

- Simulated datasets for borrower, alternative data, and loan records.

- Shapiro–Wilk normality testing and ANOVA across borrower segments.

- Gradient Boosting model training and evaluation.

- SHAP-based explainability visualizations.

- Simulated RBI-compliant dashboard outputs.