



# CredSaathi

**Transforming data chaos into credit intelligence!**

*EXTRACT, CURATE, ANALYSE*



## **Team Members:**

Sahil Dhiman, Shubham Goel, Sandeep, Bhavya Jain, Yatharth Nehra, Akshat Gupta, Anant Jain, Lakshay Aggarwal

# The Lending Divide: **Two Perspectives**



**Ravi, Credit Manager at Mid-Sized Fintech NBFC**

- Manages 8-person underwriting team
- Reviews 40+ new loan files daily
- Battles corrupted documents, slow GST portals, and manual spreadsheet entries.
- Faces tough calls: thin-file and NTC applicants often rejected, despite promising data clues.



**Vinay, Online food delivery partner**

- Food delivery partner, migrated for work
- Wants to buy an EV, rejected by banks
- No formal loan history, zero credit score
- Earns 30k/month.
- Rs. 10,000 FD maturing in 3 years
- Active insurance policy, pays utility bills regularly

# Decoding **The LENDING BLINDSPOT**

## **Growth vs. Quality** Tradeoff

Fintech lending grew very fast—  
20–24% YoY growth.

**But asset quality got worse!**

- *Net 90+ Days Past Due (DPD) doubled for small-ticket PL.*
- *worsening slippage ratios (2.19% for PL, 4.7% for BL)*

 **Prioritise  
TopLine over Sustainability!**

## **\$530B MSME Credit Gap**

*22% false rejection rate (good borrowers denied due to incomplete data)*

## **70% WC = Informal**

Manual  
underwriting takes  
3-7 days

Only **30–35%** of Indian  
adults have a  
meaningful CIBIL score.

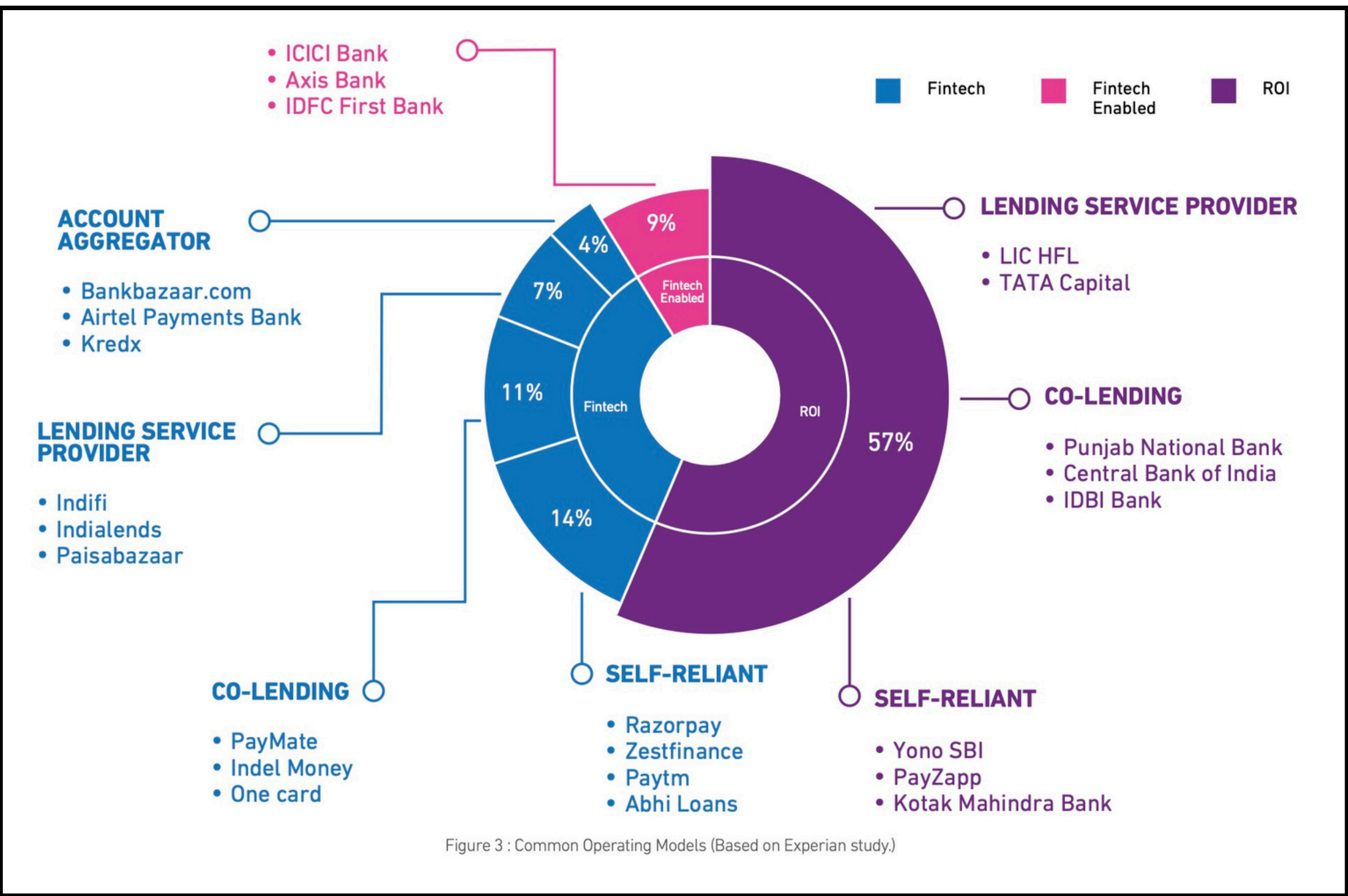
 **300-900**

## **Flat Pricing ≠ Real Risk; Margins <2%**

- Same rates for high-risk and safe borrowers.
- ECL 5.89% vs 3.41% industry average
- High costs + manual reviews squeeze profit on small loans

***Analyst Bottleneck!***  
***Each analyst processes 8-12 files/day manually. (₹40-50K/month per analyst)***

# Multiple Models, One Shared Problem



₹2.8 Lakh Cr Total **AUM**

**87%** of small-ticket loans (<₹1L) involve fintech

**62.5 Lakh** applications processed in FY24

## The **Hidden** Truth

Lending decisions fail not because of missing data, but because of **INVISIBLE DATA!**

**2.3X**

### Behavioral Signals

- Payment timing predicts default **2.3× better** than income.
- **Rent paid 1–3rd → 4.2% default vs 25–30th → 11.7%.**

**40%**

### Network Insights

- Borrowers with **≥3 verified supplier/vendor ties** default 40% less.

## The Core **Job-to-be-Done**

*“ **When** evaluating unsecured loan applications from thin-file borrowers, **I need** a solution that instantly consolidates, verifies, and extracts actionable creditworthiness signals from multiple non-traditional sources—**so** my lending team can confidently reduce false rejections of good borrowers, and ensure fair and consistent decisions.”*



# Idea-Mapping Canvas

## USERS AND Job-to-be-Done

### USERS Platform workers

*Enabling every person seeking a loan or a complete credit score to access a more complete and diverse metric to measure their reliability when they apply for loans, with a focus on gig-workers*

## AI-LEVERAGE

- Using **AI models** combined with **financial data, social media interactions** and **company reviews** to assign a **credit score**.
- First use a **sentiment analysis model** to assign scores to the interactions the borrowers have on **social media sites, feedback from employers etc.**
- Social media sites such as **Instagram, X, UPI information from Google pay/Paytm etc** and **assign weights** of importance on them to **aid in the calculation of the score**.
- **Forward the sentiment scores to a second model** which shall **perform regression** to **predict the credit score** for the individual.
- This is then **shared with the lenders** who ultimately **make the final decision**.

## PAIN POINTS

### Platform workers

- Process of loan approval feels forced and like a transaction, they feel like banks just wish to sell a loan as a 'product'.
- **Non-financial aspects of the borrower's history are completely ignored, which may lead to people still not paying back their loans despite appearing financially capable.**
- 60% of gig workers face barriers to formal credit due to irregular incomes and a lack of traditional credit histories.
- Some companies like Swiggy and Zomato do offer some emergency loans but only 36% report financial resilience
- **One of the biggest issues is also the lack of a proper credit score for gig-workers**

## DATA ADVANTAGE

- **Financial information:** Track financial information of the customer like prior loan history, liabilities, UPI transaction history, spending habits etc.
- **Social media data:** Monitor online activity of the consumer by getting permission to access posts, account information etc from sites such as Reddit, Instagram, X, YouTube etc.
- **Employer feedback:** Capture feedback and reviews from the employers of our consumers.
- All this is combined to first calculate a sentiment analysis score and then predict the credit score of the individual.

## CredSaathi — Solution Concept & Value Proposition

***A data-driven credit score prediction/generation and interactive dashboard platform aimed at creating an all-encompassing credit score which reinforces gig-worker profiles and allows them to avail formal credit card and loan approval services from banks.***

- **App:** Mobile/web app in order to gain approval of the workers and start inputting the financial and social media data into the AI models for credit score prediction.
- **AI Sentiment analysis:** Used to assign sentiment scores to the social media data, employer feedback and other data received through scraping methods (through O-Auth APIs).
- **Credit score prediction model:** Using an ML model to perform regression by assigning weights to the scores generated by the first model and then calculate the credit scores.

## Risks & Ethics (harm, bias, privacy)

- **Low Adoption:** Reluctance to adopt a more complete credit score and interactive dashboard may limit system effectiveness.
- **GDPR conformity:** We may not be able to scrape data without violating GDPR guidelines.
- **Lack of awareness:** Lack of awareness of the problem and solution will limit the adoptability of the solution.
- **Cost vs ROI:** High initial and maintenance costs for the AI models may not immediately yield proportional savings or efficiency.
- **Personalization vs Standardization:** ***AI-driven credit score predictions may produce varied or biased scores, making it hard to assign a general-purpose score that balances financial data with social/societal scores.***

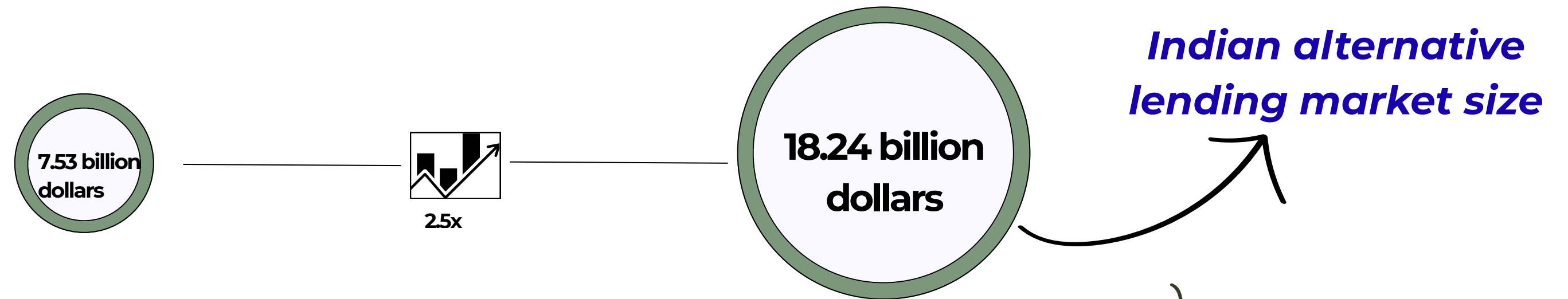
## Success Metrics/ KPI's

- **20-30% increase in rate of loan approvals** through credit score prediction.
- **Higher consumer satisfaction (+0.5 to +1 point in surveys/NPS)** from better loan application processes and credit card application processes at various banks.
- **Increased awareness about complete credit scores and formal loan and credit card approval practices.**

# Multi-Source Alternative Data

**“Entire process (sanction, approval, issuance, and disbursement) shall be digital, and digital footprints will form the basis of assessment.”**

**— Union Budget 2025–26 (Page 15)**

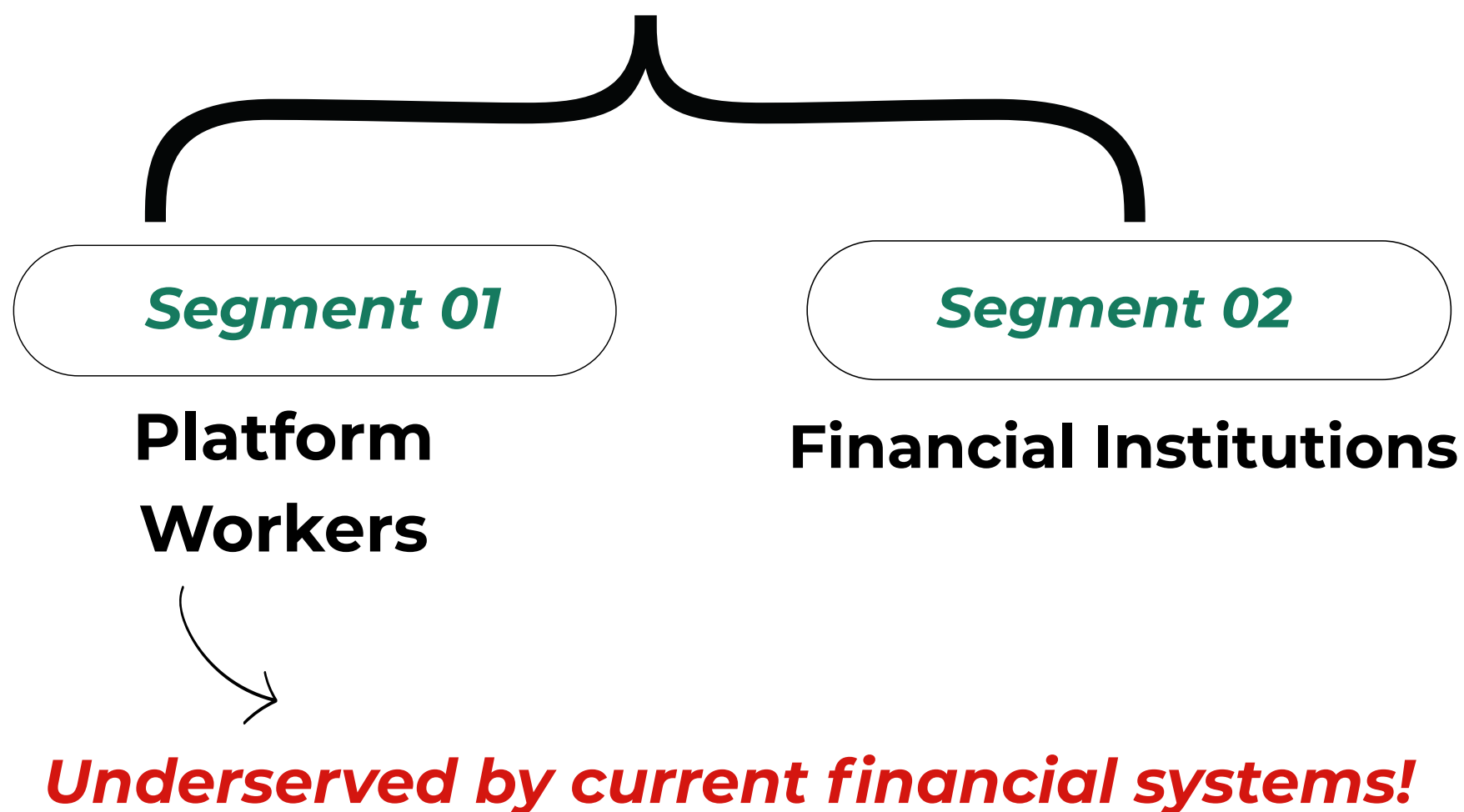


## Data Sources Used:

- **Financial Data:** Bank transactions, UPI history, GST records, and utility bill payments (Collected securely via Account Aggregator)
- **Social Data:** Social media activity & engagement patterns.
- **Employee/Work Data:** Earnings, work frequency, ratings/reviews, tenure, delivery volumes, reliability (Fetched via Gig Platform APIs or manual uploads)

***Proxies for  
Financial  
Credibility!***

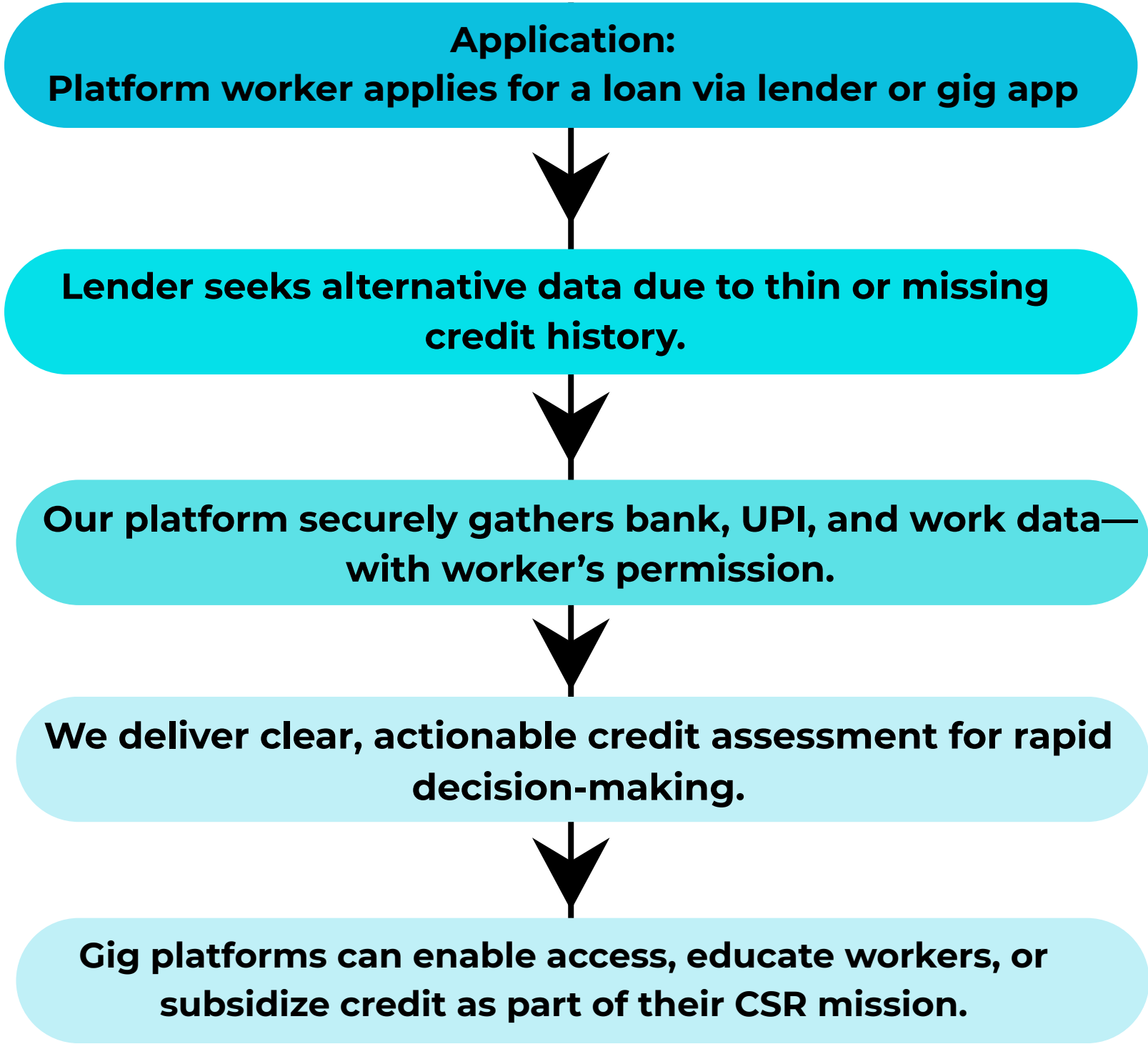
# Target Segment Definition



## MARKET OPPORTUNITY

**TAM: \$2.25B (platform lending)**  
**SAM: ₹98 Cr (350 lenders)**  
**SOM: ₹10 Cr Year 3 (10% share)**

# Our Solution





# Creating Our **Blue** Ocean!

## Crowded Digital Lending & Credit Scoring Landscape!

- *Exclusive Platform Partnerships- Direct Data Pipelines*
- *Proprietary Risk Models*
- *Lenders (and platforms) get a single integration*
- *Early Warning & Liquidity Flags*

- *Rely on outdated bureau data and income proxies*
- *Generic, low-resolution risk models*
- *not tailored for gig work volatility.*
- *Compete on price, race to the bottom with high NPAs, poor retention*

## A **B2B SaaS** Platform!

**“Unlike** traditional credit bureaus that rely only on old financial records, **we** combine financial, social, and work performance data to serve the “invisible” workforce.”

# Pricing Strategy to Cashflow

## Platform Pricing Tiers !



### 1. Essential Tier **INR 50K/month**

- Up to 500 applications/month
- Simple dashboard, 2 analyst seats
- Perfect for pilot programs at small NBFCs/Fintechs



### Growth Tier **INR 1.5L/month**

- Up to 2,500 applications/month, full feature suite, 10 analyst seats, API access.
- Target: Mid-sized NBFCs, digital lenders.



### Enterprise Tier **INR 4L+/month**

- Unlimited applications, white-label option, custom model training, dedicated support.
- Target: Large NBFCs, banks.

## Revenue Model

- **Monthly SaaS Subscription (By Tier):** Fixed monthly fee based on usage/features.
- **Performance-based Add-ons:** Optional fee per approved loan, alternative-data insight bonus, or API call volume.
- **Custom Projects:** Integration and analytics consulting for enterprise clients (one-time/annual fees).

# Targets & Costs

1	2	3	4
Month 3	Month 6	Month 12	Month 18
First paying customers beyond pilots	8 customers, INR 40L ARR	25 customers, INR 1.5 Cr ARR	50 customers, INR 4 Cr ARR, breakeven

**Traction Targets**  
**First 100 customers!**

## Projected Cost Structure

Cost Category	Type	Benchmark Amount	Description
Cloud & Tech Infrastructure	Fixed	₹1.5–5L/year	AWS, security, compliance, monitoring
Core Team/Product	Fixed	₹25–35L/year	Founders + 2 devs + 1 data scientist
G&A/Admin	Fixed	₹5–8L/year	Legal, HR, basic accounting
Data/API Usage	Variable	₹2–3K/customer/month	Account aggregator/API access, scaling
Customer Success/Support	Variable	₹1.5–2K/customer/month	Onboarding, L1 support
Sales/Marketing	Variable	₹1.5–2K/customer/month	CRM, outreach, digital campaigns
Partner Integrations	Variable	₹1–2K/customer/month	API/revenue share, joint promotions

# Metrics That Matter

## **North Star Metric** ✨

**Successful First-Time  
Loans Disbursed**



*Direct platform engagement,  
revenue driver*

## **Leading Indicator 1**

**Active Institutional Customers  
(Monthly/Quarterly)**

*Customer growth, recurring  
revenue*

## **Leading Indicator 2**

**Time to Disbursal (Efficiency &  
User Experience Metric)**

***Instant liquidity.*** A faster disbursal time  
*indicates operational efficiency*

# Responsibility

*Building trust is at the core of CredSaathi.*

## 1. Fairness & Bias Mitigation

*We audit models to detect and reduce demographic and contextual bias.*

## 2. Transparency & Explainability

*We provide clear, interpretable insights to support informed credit decisions.*

## 3. Privacy & Data Security by Design

*Borrower data is protected using strict privacy-by-design principles and collected only with user consent.*

## 4. Accountability & Human-in-the-Loop

*Final decisions remain human-controlled, supported by HITL validation and audit trails.*

# Sustainability

## 1. Right-Sized AI Models

*Optimized, energy-efficient models like XGBoost/CatBoost deliver accurate results without heavy compute.*

## 2. Intelligent Caching for Speed

*AI-based caching accelerates frequent queries and reduces redundant processing.*

## 3. Low-Bandwidth & Resilient Access

*Designed to work efficiently in low-bandwidth environments across India.*



# REIMAGINING LOANS

## AI Solution

Unified system using financial (RBI Account Aggregator), social, and gig-economy data to build holistic, real-time credit profiles with **XGBoost/CatBoost models**.

## Key Data Sources

1. **Financial:** Finvu, OneMoney (UPI, EMI, balance trends)
2. **Social:** Twitter, LinkedIn, Facebook (sentiment, reliability, work stability)
3. **Gig:** Swiggy, Zomato, Uber, Ola, Rapido (performance metrics)

## Technology Stack

- **Frontend:** React.js / Next.js + Chart.js
- **Backend:** FastAPI (async APIs, typed schemas)
- **Data Layer:** PostgreSQL, Pinecone, Redis, AWS S3
- **AI Engine:** XGBoost / CatBoost with continuous retraining

# AI Leverage, Not Just Automation

## RAG for Document Intelligence

Fine-tuned LLM reads PDFs in any format, retrieves relevant historical patterns, generates structured summaries with source citations. Handles 200+ document formats without manual templates.

## AI Agents for Multi-Step Verification

Autonomous agents for income verification, identity verification, and business stability. Agents collaborate and escalate conflicts to human analysts. Replaces 4 hours of manual work with 90-second workflow.

## Intelligent Routing

ML classifier categorises applications: Auto-approve (40%), Standard review (50%), Deep-dive (10%). Senior analysts only see high-risk cases. 6x throughput increase.

## Human-in-the-Loop Active Learning

System captures analyst corrections and retraines weekly. Confidence scores improve 12% every quarter. HITL interventions reduced from 30% 10% in 6 months with full audit trail.

# Before & After: The Transformation

## Current Manual Workflow

### Application Arrives (Day1)

- Analyst requests documents via email/Whatsapp

### Document Collection (Day1-2)

- Borrower uploads 8-15 PDFs

### Manual Data Entry(Day2-3)

- Analyst downloads, converts, enters data into spreadsheet

### Cross-Checking(Day3-4)

- Analyst verifies income, expenses, GST across multiple sheets

### Manager Review (Day4-5)

- Final decision made

***Total 4-5 days !***

## AI- Powered Workflow

### Application Arrives (Instant)

- Platform auto-fetches data via Account Aggregator API

### AI Extraction (30 seconds)

- AI extracts, structures, validates data from all sources

### ML Scoring (45 seconds)

- ML generates multi-dimensional risk score with explainability

### Dashboard Presentation

- Recommendation presented with supporting evidence

***Total 1-2 hours !***

# Roadmap

## Phase 1 - months 0 to 6

Establish connection with delivery partner corporations and launch MVP solution for employees

## Phase 2 - months 6 to 12

Scale revenue and expand data partnerships (social media, UPI, employer feedback)

## Phase 3 - year 1 to 2

Profitability through automation by rolling out auto-learning models to reduce update cycles with global and regional templates for company expansion

## Phase 4 - year 2 to 4

Expansion of services provided to MSME's in addition to the original consumer base

# Funding Ask

Round	Timeline	Source	Amount (INR Lakhs)
Pre-Seed	Q2 2025	Angels/Friends	25-40
Seed	Q1 2026	Seed VC/Angels	100-150
Series A	Q1 2027	Early VC	400-600
Series B	Q2 2028	VC/Growth Funds	1200-2000

# DE-RISKING THE RISK

## **CYBERSECURITY & DATA BREACH**

*Sensitive data + digital infra = prime attack targets*

- *Regular audits & monitoring*
- *Cloud-native, AI-driven threat detection*
- *Partner with cybersecurity experts*

- *In-house compliance function*
- *Engage with regulators/industry forums*
- *Agile, configurable product to adapt fast*

## **REGULATORY UNCERTAINTY (DATA/FINTECH LAW)**

*RBI, Data Privacy, AA rules change rapidly*




## **PLATFORM & API DEPENDENCY/CONCENTRATION**

*Over-reliance on 3rd-party APIs (e.g., Account Aggregator)*

- *Multi-partner integration (not single vendor)*
- *Build redundancy & fallback logic*



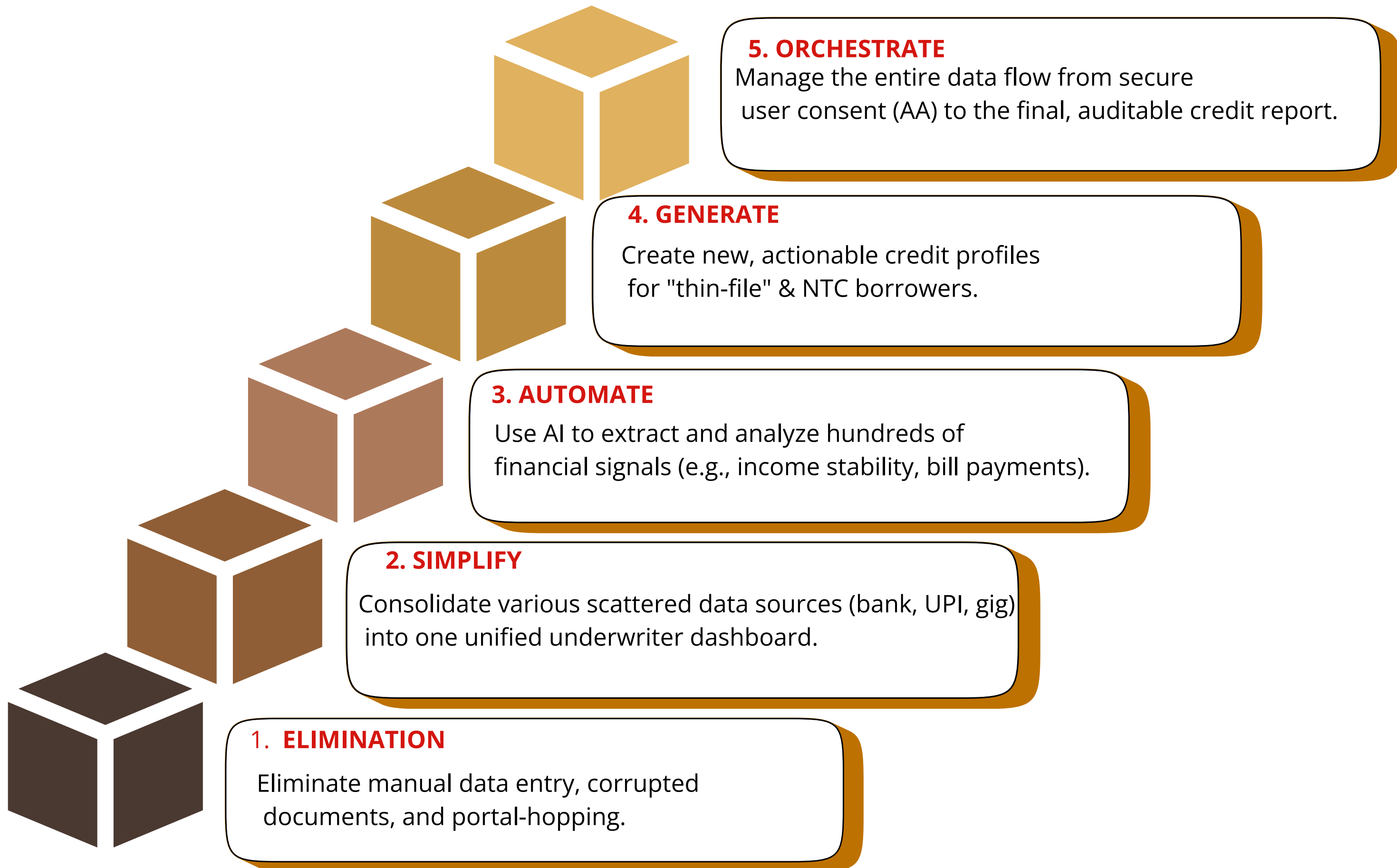
# Who are the Key Players ?

	Company	Valuation	Revenue
 CREDITVIDYA	CreditVidya	~\$30 million	~\$17.2 million
	Perfios	~\$1 billion	~\$68.8 million
 KreditBee	KreditBee	~\$700 million USD	~\$185 million USD

# FVR ScoreCard(out of 5)

SCORE	ASPECT OF SOLN	FEASIBILITY	VIABILITY	RESPONSIBILITY
1	ML model for sentiment score	3.5	4	5
2	ML model for credit score prediction	3.5	3.5	5
3	Web application for consumers	4	3.5	4

# Re-Imagination Ladder



# Proof and Evidence



## Before: The Broken Artifact

Using **anonymized public gig-worker data**, our pipeline extracted a **stable ₹28,000/month income** profile that a **partner bank instantly rejected** due to thin bureau history.

## Our AI-Generated Artifact

Using the same anonymized dataset, our pipeline extracted additional decision-grade signals:

- **98% positive customer-interaction sentiment** from verified feedback
- **Consistent weekly activity patterns** indicating stable earning capacity
- **2.3× lower default-risk score**

## PILOT INSIGHT !

*The simulator's output was clear:*

- **A borrower who would traditionally be tagged as "high-risk"**
- **Reclassified as low-risk and creditworthy under our alternative-data model.**

# Interview Guide

## For Platform Workers (Delivery, Ride Hailing, Freelance, etc.)

- Have you ever applied for a loan or credit card? If yes, what was your experience?
- Were you able to get the loan? If not, what challenges did you face?
- What documents or information did banks/lenders ask you for?
- Was any part of the process confusing or hard for you?
- Did you feel the bank or lender understood your real income and work patterns, or did they only look at formal documents?
- How long did it take to get an answer about your loan? Did the waiting time affect your decision or plans?
- If lenders could see your app earnings, work hours, or ratings, would you be comfortable sharing that—for a better chance of getting a loan? Why or why not?
- What kind of financial support (loan amount, repayment flexibility, emergencies) would be most helpful to you?
- Have you ever heard of sharing your financial data digitally for a loan (like through account aggregators, or linking bank apps)? Would you be open to it if it was secure and easy?
- What would make you trust a new lender or platform with your information and loan needs?

## For Finance Domain Professionals (NBFCs, Banks, Digital Lenders)

- How do you currently assess creditworthiness for applicants with little or no formal credit history?
- What data points do you rely on most?
- Can you walk me through your data collection and verification process for loan applications, especially for gig workers or new-to-credit borrowers?
- Which parts are most time-consuming or error-prone?
- How do you handle cases where applicants can't provide traditional documents (like salary slips or tax returns)?
- What alternatives have you tried or would like to try?
- What are the main reasons you end up rejecting gig workers or thin-file borrowers? Is it more about data, policy, or risk appetite?
- How confident are you in the consistency and fairness of your credit decisions across analysts or branches?
- Have you explored using alternative data (bank statements, GST, UPI, gig platform data) to supplement credit decisions?

