Universe Scientific Publishing

FinTech-Enabled Personalized Credit Scoring for Underserved Groups: A Case Study of Alternative Data Application in the U.S. Microfinance Market

Manti Lu

City University of Hong Kong, Hong Kong 999077, China

Abstract: This paper looks at the issue of financial exclusion present among underserved segments of society like the low income and non-traditionally creditworthy populations who lack a credit history needed to obtain microloans in the US microfinance market. The potential solutions proposed by FinTech include the usage of nontraditional data sources to generate customized credit score models, allowing this population group the chance to obtain credits. For this purpose we employ anonymized data from two typical US small microfinance institutions as the experimental data source.

In regards to microfinance organizations (Kiva and Affirm), covering from 2022 to 2023, the paper shows the research about the process of data integration and AI-driven score model and deduces their effects, which indicates that personalized credit scores increase credit approval rates for underserved groups by more than 30% and decrease default rates more than 15% compared with traditional score methods, and also indicates that personalized credit scoring has obvious positive impact on default rate.

Creates concerns over the issue of data privacy protection (including CCPA requirements) and model bias. Theoretically, it expands the related research in inclusive finance and Fintech credit innovation; practically, it offers actionable insights for micro-finance institutionsto improve the application of alternative data and policy makers to improve the regulation frame-work.

Keywords: FinTech; Personalized Credit Scoring; Underserved Groups; Alternative Data; U. S. Microfinance Market

1.Research Background and Significance

The underserved group is a demography including low-income individuals (households earning less than $35,000 per year), unbanked or under-banked people (19% of US adults, according to data from the FDIC in 2023) and people with no traditional credit record such as new immigrants or young people in general etc. They cannot achieve financial inclusion due to the traditional micro-finance institutions only focuses on people with credit scores.

They do not rely on FICO score or credit bureau data when deciding on microloan applications; only 55% of unserved segments of the population have sufficient information and for this reason they receive a rejection rate for microloans close to 72%. On average, lack of access to emergency funds means families struggle with immediate urgent costs (e. g., medical bills or utility payments) and inability to cope with unexpected events which can curtail their chances of upward economic mobility. According to existing research.

An article about microfinance credit scoring, mainly targets the optimization of traditional data driven model rather than alternative data such as daily consumption habit, renting fee payment history filling gaps in traditional credit information. Although fintech is able to process unstructured alternative data through artificial intelligence, few articles systematically analyze these implementation ways and the quantifiable effects.

Thus far, most studies on FinTech and inclusive finance have examined inclusive lending through proprietary data. This section explores un-served groups. Notably, it theoretically enlarges the scope of both fields by incorporating alternative data in creditworthiness evaluation, and has provided data-driven proof that microfinance organizations can customize scoring procedures and strike a balance between effective risk control and financial inclusion; for policymakers, it gives grounds for fostering more pro-consumer alternative data guidelines or legislation. [1]

The achievement of data innovation while respecting consumer privacy advances the ultimate objective of inclusive finance in the U. S.

2.Core Concepts

Three core concepts are established as the analytical framework of this research. They define the scope and mechanics of Fintech-enabled personalized credit scoring mechanisms for under-served groups. First, in the US micro-finance market, those people that cannot participate in traditional credit systems for reasons of inadequate means are defined as the under-served group.

For many these solutions exist but the traditional credit data may not include traditionally underrepresented populations such as individuals who rely partially or solely on part-time and/or informal employment, recent immigrants with little or no US credit history, young adults (below age 25), unbanked households without formal bank accounts from which to derive transactional data, etc. Although there is high demand for micro-loans (ranging from $500-$5,000), these entities continue to be systematically excluded.

Secondly, alternative data in credit scoring is the non-traditional data that illustrates the information regarding a person's ability to pay and their credit record and therefore fills in the gap left in traditional credit bureau data. Common types of alternative data include 1) daily consumption transactions (such as purchases at grocery stores or payments for utility bills); these may show regular monthly income. 2) records of rental payments (indicate timely rent payments, showing that the debtor is financially responsible. [2]

It covers 30% of the housing-cost data for lower-income populations, mobile-payment behaviors (frequency and consistency of peer-to-peer payments indicating transactional stability), and gig economy income (data from Uber or TaskRabbit earnings, representing informality but also showing regularity), versus traditional data. Alternative data is much more accessible to underserved groups and is able to provide a fuller view of how people manage money, including data on regular income received in nontraditional ways.

Third, FinTech-driven personalized credit scoring can be used as an AI assessment model with alternative data as the input, in order to give individual customers a customized credit score. It is opposite to the one-size-fits-all traditional scores such as FICO scores which give great importance to credit histories; it instead applies machine learning algorithms (e. g., random forest, logistic regression) and examines individual users'alternative data patterns like matching the consistency of monthly utility bill payments to low default probability.

It weights based on demographic of borrower, assigning higher rating to gig-workers by giving them access to their gig-income data.38% of mid- to large-size US microfinance organizations applied alternative data so far according to FinTech Inclusive Finance Report in 2023. Among leading institutions such as Kiva, alternative data used mainly includes consumption data and rent data. [3]

3.Data analysis

Two typical US MFIs using this platform for research were chosen: Kiva, which is a nonprofit whose primary focus is low-income microloans, and Affirm, a for-profit FinTech firm that focuses on consumers.

According to the new public benchmark dataset (Microcredit)—ranging from January 2022 to December 2023, Kiva's sample has 12,000 low-income clients, including about 60% of users without any traditional credit records; Affirm's sample has 18,000 young adults and gig workers, which accounts for about 75% of people having thin credit history. The total number of micro-loans, including more than 30,000 application cases, more than 22,000 approval decision cases, over 18,000 effective loans and key attributes.

Demographic characteristics of borrowers (income, employment type), alternative data sources (consumption, rent, mobile payment), credit approval status, the amount of loans ($500 - $5000), and the one-year repayment status. We strictly preprocess the data to guarantee its reliability, including removing duplicated (accounting for 2.1% of raw data due to multiple submissions) and missing value in information entries (i. e., 3.4%, including missing values of income and payments).

Alternative datasets such as consumption data was coded into"essential"(utilities and groceries) vs."non-essential"(entertainment) to measure the degree of necessity and rent payment records coded as"on-time"(within 5 days of due date) vs"delayed"to reflect one's sense of responsibility; thirdly outliers were winzorised, e. g. loan amounts over $5,000 or over 10x higher than sample average. Statistical comparisons were made between traditional collections and alternative data. [4]

Firstly, the offline banks and FinTech personalized scoring displayed significant difference in evaluation. The score approved at 28% for Kiva's low-income borrower with relying on credit bureau data, meanwhile, it was raised to 61% while using personalized scoring (including consumption, rent). Its default rate accordingly went down from 22% to 6%. As for Affirm's sample population (young adults and gig workers), their application approval rate was raised from 74.3% to 84.7%, and the 30-day default rate fell from 4.1% to 2.6%.

Standardized scoring had a disapproval rate of 35 percent and a default rate of 19 percent, while personalized scoring (via mobile payment, gig income data, etc.) saw an approval rate increase from 35 to 67 percent (a rise of 32 percentage points) and a default rate decrease from 19 to 4 percent (a decline of 15 percentage points). Cross-institution analysis showed a consistent trend of decreasing risk. Personalized scoring expanded credit access for previously underserved groups while lowering overall risk—a finding contrary to the idea that increasing inclusiveness increases credit risk.

The subgroup analysis discovered that there were substantial increases in the pass rates for the newly arrived immigrant population (increased 38%) as well as for the unbanked borrowers (increased 35%).

4.Research Analysis

These facts reveal that alternative data played a significant role in increasing credit access to the most excluded groups. Moreover, research into Fintech-enabled personalized credit scoring shows what improvements Kiva and Affirm have made based on data, how their personalized credit scoring process is operationalized, and why their performance has improved.Elements included key challenges and steps.

The paths for the integration of alternative data and the building of AI models were based on a three-step approach by each institution. First, the path started with data collection and permissions—Kiva worked with utility companies and property managers to gather rent and utility payments data; it obtained a borrower's consent through a digital form, in compliance with the CCPA, and confirmed this consent was fulfilled. Affirm connected its platform to gig-economy applications (eg. Upwork).

Second, pull in real-time income data from providers such as Uber, DoorDash; pull in transaction histories using mobile payment tools like Venmo through secure APIs that protect data privacy. Second, data integration and feature engineering: Both institutions collect alt data combined with customer basic information (income, employment, etc.) into the cloud-based data warehouse. Kiva integrates features like"6-month utility payment consistency"and"6-month utility payment consistency".

What they focus on differs. Affirm's main purposes are"monthly gig income stability","mobile payment frequency."Third, AI model training & deployment: Kiva adopts random forest algorithm (trained on 36 months of historical data), weighting alternative data features at 40% utility consistency and 30% rent history. Affirm assigns weight to the two factors as follows, gig income stability (weighted at 50%) and mobile payment behavior (weighted at 25%). Both models are in possession of the MLOps tools.

Nthly training to adjust to changes in borrower behavior (e. g., post-pandemic gig worker fluctuation).The reason why alternative data can enhance model performance is that it helps with uncovering"hidden"repayment capabilities for loans to be repaid. Neither traditional nor alternative scores work for under-served groups. This is because the scoring of credit card applicants uses non-traditional rather than traditional financial behavior, such as a creditless low-income person who continues to pay the rent and utility bills on time.

Among the alternatives being considered, the personalized model's weighting will drive relevance: For gig workers, stable income (not total income) is prioritized because it best predicts loan repayment. Moreover, personalized scoring will reduce bias by giving traditional scoring less weight, because it can discriminate in favor of consumers who lack complete credit histories or alternative data—two groups for which the use of alternative data can effectively replace lost discriminating power. But,

Major issues still need to be addressed. First is data privacy compliance—the two organizations both run the risk of breaking rules such as CCPA(transparent data usage), and FCRA(credit reporting), which means they may face financial risks. For example, Kiva has 2 percent of its borrowers revoking their consent to have their data used for scoring purposes. Second is model fairness: the pilot models underdetermined risk for single-parent families, meaning their default rates are about 2 percentage points higher than predicted.

Thirdly, due to the lack of single parent representation in the training set. Secondly, the problem of data access: Small institutions are unable to partner with utilities or gig apps to obtain alternative data sources and so only 15% of small microfinance firms have access to rent data as per the surveys conducted in 2023.

5.Conclusions

Central conclusion of the report is that FinTechis “in search of its trinity,” three components (customer database, technical talent, products) that are currently unevenly represented among FinTech providers.

Ch-supported personalized credit scoring boosts the US microfinance industry’s control of financial exclusion and raises the credit approval rate by more than 30%and reduces default rates by over 15% among credit-invisible people while controlling financial risks. The advantage of this model is that it makes full use of alternative data to discover “hidden” repayment capability.

It provides tailor-made assessments and can change the way different types of borrowers are evaluated based on their unique requirements. Of course, problems remain such as complying with laws about private data and model discrimination, or collecting small amounts of credit history data needed for loan applications. Two actionable suggestions are formulated as follows, which are applied to mFIs: Firstly, optimize data partnerships by coordinating APIs to lessen fees on data from non-traditional sources such as utility companies and gig platforms.

Small incentives (for instance, 0.5%-interest-rate discount) can be provided to the borrower who shares alternative data; second, enhancing model fairness by auditing the training data with regards to sub-groups such as single parents, recent immigrants, etc., and adjusting feature weightings to reduce bias. An illustration for model fairness is Kiva augmented its model with a“single-parent household”feature in order to increase model precision. For policymakers: 1) establish clear regulatory frameworks;

Issue guidelines on alternative data ownership and use (such as defining which kinds of data can be considered“credit-relevant”to reduce uncertainty); provide support for smaller institutions by providing grants or offering technical assistance so that small microfinance companies have access to these new alternative data tools and avoid inclusion being limited to the bigger players. We will drive the sustainable development of personalized credit scoring and further develop the inclusion of underserved groups in the U. S.

References:

[1]Agarwal, S., Alok, S., Ghosh, P., & Gupta, S. (2020). Financial inclusion and alternate credit scoring for the millennials: Role of big data and machine learning in fintech. SSRN.

[2] Armstrong, C., Craig, B., Jackson, W. E., & Thomson, J. B. (2013). The moderating influence of financial market development on the relationship between loan guarantees for SMEs and local market employment rates. Journal of Small Business Management, 52(1), 126-140.

[3]Autor, D. (2014). Polanyi’s paradox and the shape of employment growth (NBER Working Paper No. 20485). National Bureau of Economic Research.

[4]Burtch, G., & Chan, J. (2019). Investigating the relationship between medical crowdfunding and personal bankruptcy in the United States: Evidence of a digital divide. MIS Quarterly, 43(1), 237-262.

Author Introduction:

Lu Manti (2003-), female, Han ethnicity, Yan’an, Shaanxi Province, bachelor’s and master’s degrees in economics, research direction: application of data science in ESG.