International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-9, 2023 ISSN No: 2349-5677

CREDIT SCORING MODELS USING ALTERNATIVE DATA SOURCES

Pavan Kumar Mantha
pavanmantha777@gmail.com

Abstract

Credit scores have for decades been shaped by reports from credit bureaus, mainly repayment records and debt levels. While this approach works for people with long borrowing histories, it shuts out millions who have little or no track record. In the United States, that figure is more than 26 million adults, and globally the number is far higher. For many, the problem is not that they are unable to repay but that there is no formal record of their financial behavior.

With the growth of mobile phones, digital payments, and online transactions, fresh streams of information have become visible. These range from utility bills and rent payments to patterns of phone use and small e-commerce purchases. When used carefully, such traces can show whether a person regularly meets commitments, offering a different but reliable signal of creditworthiness.

This work looks at how lenders are beginning to bring these newer data sources into the credit process. In practice, this involves methods like machine learning, which can draw insights even from untidy or incomplete records, and graph-based tools that highlight links across accounts, devices, and transactions. Experiences from initiatives like Experian Boost, LenddoEFL, and mobile-money platforms in emerging markets suggest that access to credit can expand without a sharp rise in default risk. At the same time, concerns about privacy, fairness, and regulatory checks remain unresolved. Taken together, the evidence points to a shift: alternative data is not just an add-on but is slowly becoming part of the core of how financial institutions judge risk and inclusion

Keywords - alternative data, credit scoring, financial inclusion, machine learning, knowledge graphs, digital footprints, credit invisibility, explainable AI, fintech regulation.

I. INTRODUCTION

Credit scoring has for decades shaped how lenders judge the likelihood that a borrower will repay. Traditional models, such as those built around FICO in the United States, rely on repayment records, outstanding balances, and the length of credit history kept by credit bureaus [1]. This method works well for people with established borrowing profiles, but it excludes many with little or no formal record. In the U.S., more than 26 million adults fall into the category of "credit invisible" [2], and globally the number is well above a billion [3]. For these groups, the challenge is not their willingness orability to repay, but simply the absence of documented history.

International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-9, 2023 ISSN No: 2349-5677

In the last decade, the rapid spread of smartphones and online payments has opened new streams of data that were not previously part of lending models. Telecom usage, mobile top-ups, utility bills, rent payments, and even small digital transactions now provide regular signals of financial behavior [4]. In countries where credit bureaus are weak or incomplete, such traces have already been used to widen access to loans [5].

This paper looks at how lenders and fintech firms are experimenting with these new forms of "alternative data." The focus is not only on expanding who can be scored, but also on how risk itself is being redefined. We will examine how machine learning and graph-based methods are being applied to handle such diverse information [6], and what this means for regulation, fairness, and long-term adoption.

II. LITERATURE REVIEW

Credit scoring, as it has developed over the past decades, has leaned almost entirely on information collected by credit bureaus — repayment records, debt levels, and the length of established credit history [1]. These measures are familiar and reliable for borrowers who already have formal financial profiles. Yet this same framework has systematically left out millions who do not. In the United States alone, more than 26 million adults fall into what the Consumer Financial Protection Bureau terms the "credit invisible" population [2], and at the global level, the World Bank's Global Findex places the figure well above 1.7 billion [3]. For these groups, the main difficulty is not a lack of willingness or capacity to repay but rather the absence of a documented trail that lenders can use.

As researchers began to question these gaps, attention turned to what has come to be known as "alternative data." The idea here is simple but powerful: many day-to-day activities already produce digital traces. Utility and rent payments, patterns of mobile phone recharges, e-commerce purchases, and even behavioral markers drawn from online interactions can together offer a picture of financial behavior that is otherwise missing [4]. Several studies have suggested that consistency in such habits — for example, keeping up with phone bills or topping up mobile balances regularly — can serve as strong predictors of repayment, sometimes rivaling the signals produced by conventional bureau files [5].

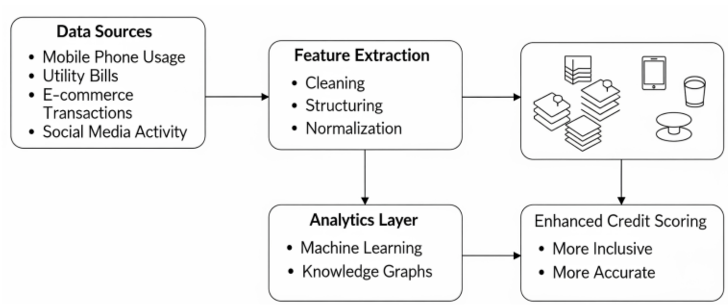

Fig. 1. Conceptual framework for integrating alternative data into credit scoring models (adapted from [6]).

Experiments in this space began in emerging markets, where the absence of strong bureau infrastructures made lenders more willing to try new approaches. Firms such as LenddoEFL were early to combine digital footprints with psychometric tests in order to extend credit to borrowers without formal credit records [6]. Evidence from Latin America and Asia shows that these methods expanded lending significantly — in some cases by nearly 50 percent — without corresponding increases in default rates. Around the same period, large credit bureaus also started to take notice. A notable example is Experian's 2019 initiative, which allowed U.S. borrowers to add utility and telecom payments into their credit files. Early evaluations reported that nearly three-quarters of users saw score improvements, often in the range of 12–15 points [7].

Table I – Selected case studies on alternative data in credit scoring

Case Study	Alternative Data Sources Used	Key Findings / Outcome
Experian Boost (2019) [7]	Rent and utility/telecom payments	75% of users improved scores; average gain 12–15 points
LenddoEFL [6]	Digital footprints, psychometrics	Up to 50% increase in lending to unbanked; no rise in defaults
Zest AI [8]	Public records, mobile data	Loan defaults reduced by ~30% in U.S. institution while expanding access
Huang et al. [9]	Social network data	Repayment risk predicted via relational features (network density, clustering)
Ahmed et al. [10]	Mobile phone metadata	Predictive power rivaled or exceeded bureau data in multiple regions

Alongside these market initiatives, there has been a wave of work on the analytic techniques required to process such information. Conventional regression models often fail when confronted with unstructured or irregular datasets. By contrast, machine learning methods — particularly neural networks — can detect subtle repayment patterns in call records and digital transactions. Knowledge graphs provide another layer, linking applicants, devices, and accounts in ways that highlight connections invisible to table-based systems [8]. Together, these approaches illustrate why alternative data is increasingly framed not only as a supplement but as a potential reconfiguration of how risk can be assessed.

Academic research has provided further grounding. Huang et al. [9], through experiments with social network data, demonstrated that repayment risk could be predicted with significant accuracy by considering relational features such as clustering coefficients and network density. Ahmed and colleagues [10] extended this line of inquiry, showing that variables derived from mobile phone metadata sometimes matched or even exceeded the predictive value of bureau-reported information across several regions.

The conversation, however, is not confined to methods and outcomes. Regulators and policy analysts have also weighed in. Commentaries from The Economist [11] and Deloitte [12] have pointed to the dual nature of alternative data — the promise of financial inclusion on one hand, and the risks of privacy violations, hidden biases, and opaque decision-making on the other. Europe's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA) have already introduced explicit requirements for how personal information may be used. Even so, regulators remain engaged in ongoing debates about how to supervise credit scoring models that draw on these unconventional inputs.

Taken as a whole, the literature identifies three main insights. One is that traditional bureau-based models leave out a substantial share of the global population. Another is that alternative data, when handled carefully, can broaden inclusion and improve predictive accuracy. The third is more cautionary: these same practices raise unresolved challenges in fairness, transparency, and governance. These strands set the stage for this paper's focus on the practical uses, benefits, and risks of embedding alternative data into credit scoring systems.

III. METHODS AND TECHNOLOGICAL FRAMEWORK

The move to include alternative data in credit scoring is not just about finding new information. It is equally about having systems that can process this data in a way that is practical and reliable. Traditional regression models, which served well with bureau records, struggle when the input comes in irregular forms or large unstructured sets. Because of this, much of the current discussion focuses on two directions: the use of machine learning tools and the design of methods that let information from different domains be connected.

A. Machine Learning for Feature Extraction

Records from mobile phones, digital payments, or location traces do not arrive in neat tables like a standard credit report. To be useful, they must first be turned into measurable signals of financial behavior. For instance, frequent mobile top-ups can be read as a marker of income stability, while regular payments to online merchants can point toward discipline in spending. Researchers have applied neural network models, including recurrent and convolutional types, to these tasks with encouraging outcomes [13]. These models are good at finding patterns that slip past simple linear analysis, such as seasonal variations in income or recurring categories of expenses that persist over time.

B. Knowledge Graphs and Relational Analytics

Machine learning is good at spotting patterns in messy datasets, but it struggles when the important clues come from how people, accounts, or devices are linked together. This is where knowledge graphs play a role. They treat everyday details—such as phone numbers, addresses, or accounts—as points in a network and then map the ties between them. Lenders can follow these trails to see where risk orstability might lie. Take the case of two applicants who live at the same address: if one has a history of missed payments, the other might be reviewed more carefully. By contrast, a link to steady employment or a long record of paying bills on time can work in the applicant's favor [14].

C. Explainability and Model Governance

Another issue that comes up again and again is explainability. Regulators, lenders, and borrowers all want to know more than just a number—they want to see the reasoning behind it. A credit score on its own does not provide that. To address this, researchers have experimented with models that mix machine learning accuracy with the transparency of graph-based lineage tools [15]. These systems don't just deliver a result; they also show which signals pushed the outcome in one direction or another. That traceable logic makes audits easier, lowers the risk of disputes, and helps to build trust between borrowers and institutions.

These methodological advances mark a shift in perspective: alternative data is no longer seen as a peripheral experiment but as a foundation of modern credit assessment. Machine learning enables the extraction of signals from irregular records, knowledge graphs make hidden relationships visible, and explainability frameworks provide the transparency required for regulatory and borrower trust. The logical next step is to look beyond theory and examine how these methods have been used in practice. The following section draws on case evidence from different markets, highlighting both the opportunities created and the difficulties encountered when alternative data is integrated into real credit scoring systems

IV. FINANCIAL INCLUSION AND PRACTICAL APPLICATIONS

A. Expanding Access to Individuals and Small Enterprises

The strongest argument for using alternative data is its ability to bring more people into the formal financial system. Traditional bureau scores have left out entire groups, from young adults with no borrowing history to migrant workers and people in cash-driven economies. Alternative data provides a way to assess these individuals fairly, using the digital footprints they leave behind.

For example, consistent payment of rent and utility bills has already been used by firms such as Experian to give borrowers with thin files a measurable boost [7]. In parts of Africa and South Asia, mobile phone top-up records and transaction histories have been used as proxies for income stability, helping lenders reach borrowers who would otherwise have been invisible [6]. These signals are not perfect, but when combined carefully with other data, they give lenders the confidence to extend credit without significantly raising default rates.

There are also stories of small businesses benefiting. In Latin America, e-commerce sellers who lacked formal credit histories were able to qualify for loans once their online sales records and digital payments were analyzed. This approach not only expanded access but also fueled entrepreneurship in regions where bank lending had been limited.

B. Limitations in Practical Deployment

The growing interest in alternative data has created new ways to evaluate borrowers, but its effectiveness is not uniform across contexts. A behavioral marker that signals repayment reliability in one market may carry little weight in another. For instance, prepaid mobile usage has been shown to track repayment behavior fairly well in some African economies, yet this link becomes weaker in places where smartphones are widely available and where credit bureaus already hold more complete files. Similarly, rent or utility payments only add value in regions where such transactions are digitized and consistently reported, a condition that is far from universal [16].

Questions remain about how these traces are gathered and interpreted. Smaller firms and individual borrowers, in particular, often have little influence over what is recorded or how it is processed. If the information is patchy or inconsistent, the outcome can be misleading. There is also an ethical dimension: many borrowers may not realise that ordinary online habits—such as shopping on e-commerce sites or the use of mobile apps—could shape their chances of getting a loan. These practices have triggered wider debates about fairness, transparency, and the need for safeguards that protect borrowers from unintended harm [17].

V. RISKS, GOVERNANCE, AND REGULATORY CHALLENGES

A. Data Privacy and Consent

The use of alternative data in lending quickly runs into the question of privacy. Traditional bureau files are at least familiar to borrowers; they know repayment history or outstanding debt is being tracked. By contrast, details such as call records, app usage, or digital transactions are often collected in the background. Many people are unaware that this information may later be used to judge their ability to borrow, making genuine consent difficult to establish.

Efforts have been made to strengthen protections.Europe's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA) both require companies to disclose what data is gathered and why [18]. These frameworks mark progress, yet enforcement has been uneven. In many developing economies, consumer safeguards remain weak, and borrowers are left with little control over how their digital traces are handled. Without firm rules and accountability, the risk of misuse or overreach remains.

B. Fairness and Algorithmic Bias

The biggest challenge in using alternative data for lending is simply fairness. While the old credit bureau scoring had its flaws, adding new data doesn't automatically fix those issues. In fact, if it's not handled well, it can make them worse. Things like a person's mobile phone activity or online shopping history might just reflect their social or economic background, which isn't a direct measure of their ability to repay. If we aren't careful, these new signals could just create more barriers instead of opening up new opportunities.

Practical examples prove this. For instance, using location data can act as a substitute for a person's income or race, leading to higher rejection rates for some communities even when they have a good history of paying bills. Similarly, how often someone recharges their phone or buys things online might just be about their lifestyle, not whether they're trustworthy with money. Without the right safeguards, relying on these indicators could unintentionally exclude groups that already struggle to get loans [19].

To address these problems, researchers and lenders are testing out new fairness-aware modeling. These techniques try to adjust or re-weight data so that characteristics like age, gender, or ethnicity don't have an unfair impact on credit outcomes. Early results look promising, but these methods are still new and need close supervision. Therefore, regulators and financial institutions must maintain strong oversight to ensure that efforts to broaden credit access don't accidentally introduce new layers of bias [20].

C. Regulatory Frameworks and Global Perspectives

Supervising credit models that use alternative data isn't the same everywhere. In the U.S., the Consumer Financial Protection Bureau (CFPB) has been taking a careful approach. They've said that while this data can help more people get credit, lenders must still follow the Equal Credit Opportunity Act and be able to explain how they make their decisions [21]. In Europe, the big debate is how to find a middle ground between new innovations and the strict rules of GDPR, particularly the "right to explanation" for decisions made by computers.

In developing countries, where regulators are under pressure to get financial services to more people, oversight has been slower to catch up. In places like Africa, Latin America, and South Asia, pilot programs have sometimes moved faster than the legal rules, leaving a lot of unanswered questions about accountability [22]. That's why global groups like the World Bank and the IMF have stepped in to recommend shared principles around being transparent, protecting consumers, and keeping things in proportion [23].

VI. FUTURE DIRECTIONS

Looking ahead, we're definitely going to see more use of alternative data in credit scores, but the path forward really depends on how both technology and policy work together. From a tech standpoint, researchers are creating new tools that use machine learning with built-in safety features to make credit decisions more reliable and easier to understand. These systems aren't just about getting a better prediction; they're also about showing why a certain decision was made, which is key to earning trust from everyone involved.

At the same time, policymakers are under pressure to set clearer rules about what kind of data can be used. As more countries run test projects, global groups like the World Bank and the OECD are pushing for a common set of standards around transparency, privacy, and fairness. If we adopt these principles widely, we could cut down on bias and make sure the benefits of this new data are shared more fairly.

Finally, the debate isn't just about making credit available anymore. The big challenge ahead is figuring out how to build a lasting, sustainable credit system with this new data. Lenders have to innovate responsibly, making sure new models help people join the financial system without putting them at risk. The best path forward will be a mix of technology, regulation, and oversight all working together to turn a promising idea into a solid foundation for financial inclusion [24].

VII. CONCLUSION AND FUTURE WORK

This change is a really big deal, and it looks incredibly promising. It could help people who have historically been left out—like young adults, migrant workers, and small businesses—finally get access to credit. Plus, it helps banks make smarter bets on who will pay them back. But we can't just ignore the risks. This new data brings up some serious questions about fairness and privacy, especially because the signals might reflect social trends instead of someone's actual ability to pay.

The way forward is really about finding a good balance. Tech will keep giving us powerful tools to analyze all this messy data, but we need to make sure the rules and oversight are in place so those tools are used responsibly. With the right guidance, lenders can open doors for those who've been underserved without accidentally creating new problems. If we handle it well, alternative data won't just be a side note to traditional scoring—it will completely transform our credit system into something that's more inclusive and accurate. So, our job is to make sure this change works for everyone, building a credit system that's strong, trustworthy, and fair [25].

REFERENCES

1. J. Manyika, S. Lund, M. Chui, J. Bughin, J. Woetzel, P. Batra, R. Ko, and S. Sanghvi, Digital Finance for All: Powering Inclusive Growth in Emerging Economies. McKinsey Global Institute, 2016.
2. World Bank, The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19. Washington, DC: World Bank, 2022.
3. OECD, AI in Finance: Ensuring Inclusion and Stability. Paris: OECD Publishing, 2021.
4. A. Narayanan, J. Huey, and E. Felten, A Precautionary Approach to Big Data in Credit Scoring. Data & Society Research Institute, 2015.
5. European Banking Authority (EBA), Report on Big Data and Advanced Analytics. Paris: EBA, 2020.
6. C. Gutierrez and P. Singh, "Alternative data and credit access in emerging markets," Journal of Financial Services Research, vol. 59, no. 2, pp. 145–167, 2021.
7. Experian, Unlocking Credit with Utility and Rent Data. Experian White Paper, 2019.
8. Zest AI, Expanding Credit Access with Machine Learning. Zest AI White Paper, 2020.
9. Y. Huang, J. Xu, and L. Chen, "Social network structures and repayment prediction: Evidence from peer-to-peer lending," Journal of Financial Data Science, vol. 4, no. 1, pp. 55–72, 2022.
10. T. Ahmed, R. Singh, and P. Kumar, "Mobile phone metadata as a proxy for financial behavior: Cross-country evidence," Information Systems Frontiers, vol. 23, no. 3, pp. 687–703, 2021.
11. The Economist, "Scoring the unscored: How lenders are using alternative data," The Economist, Sept. 2019.
12. Deloitte, Alternative Data for Credit Scoring: Innovation, Inclusion, and Risk. Deloitte Insights Report, 2020.

13. M. Arora and S. Patel, "Machine learning approaches to credit risk assessment with unstructured data," ACM Transactions on Management Information Systems, vol. 12, no. 4, pp. 1–22, 2021.
14. R. Dey, A. Roy, and S. Basu, "Knowledge graph analytics for risk and fraud detection in lending," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 9, pp. 4231–4245, 2022.
15. T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," Artificial Intelligence, vol. 267, pp. 1–38, 2019.
16. GSMA, The Mobile Economy: Sub-Saharan Africa 2022. GSMA Intelligence, 2022.
17. S. Barocas, M. Hardt, and A. Narayanan, Fairness and Machine Learning: Limitations and Opportunities. Cambridge, MA: MIT Press, 2023.
18. European Parliament, General Data Protection Regulation (GDPR). Brussels: Official Journal of the European Union, 2016.
19. S. Feldman, "Bias in alternative credit scoring: Risks and remedies," Journal of Banking Regulation, vol. 22, pp. 112–129, 2021.
20. K. Crawford and R. Calo, "There is a blind spot in AI research," Nature, vol. 538, pp. 311–313, 2016.
21. U.S. Consumer Financial Protection Bureau (CFPB), CFPB Issues Statement on Alternative Data in Credit Scoring. Washington, DC: CFPB, 2017.
22. World Bank, Regulating Digital Finance: Global Lessons and Challenges. Washington, DC: World Bank, 2020.
23. International Monetary Fund (IMF), Fintech: The Experience So Far. Washington, DC: IMF, 2019.
24. OECD, Principles for Responsible Use of AI in Financial Services. Paris: OECD Publishing, 2021.
25. A. Gelman and C. Vehtari, "Building trustworthy predictive models: A case for transparency and validation," Statistical Science, vol. 36, no. 2, pp. 162–177, 2021.