

A Credit Scoring Model Proposal Based on Social Media Data to Enhance Financial Assessment

Aufa Azhari Hafidh
School of Economic and Business
Telkom University
Bandung, Indonesia
aufaazharihafidh@gmail.com

Andry Alamsyah
School of Economic and Business
Telkom University
Bandung, Indonesia
andrya@telkomuniversity.ac.id

Abstract— The evolving landscape of credit scoring presents increasingly complex challenges in accurately assessing financial risk in the digital era. The complexity in credit scoring has escalated, driven by evolving industry needs and shifting digital demographics. Traditional credit scoring methods are challenged by the growing diversity in financial behaviors and the increasing number of individuals with limited or no credit history. Our model addresses these complexities by applying the '3Cs' of credit scoring: Capacity, Character, and Conditions. Capacity is assessed using demographic data, offering insights into financial stability. Character is evaluated through user activity, providing a deeper understanding of individual behaviors. Conditions are discerned from network data, reflecting broader economic influences. Our approach aids financial institutions in exploring new markets by offering insights into individuals with limited credit histories. Our model addresses the evolving landscape of credit risk assessment, enhances financial inclusivity, and facilitates more accurate risk evaluations and informed credit decisions.

Keywords— credit scoring, social media data, financial inclusivity, big data, digital finance.

I. INTRODUCTION

As an integral component in the financial sector, credit scoring plays a vital role in determining consumers' creditworthiness and subsequently influencing lending decisions. Traditionally, the credit scoring process has relied on historical financial data and tangible collateral, yet the credit scoring approach often overlooks individuals with minimal or no credit history. Credit scoring is fundamentally a statistical technique pivotal in categorizing credit applicants into 'good' and 'poor' categories [1]. The increasing importance of credit scoring, paralleling the dramatic growth in consumer credit, is well-documented [2].

The reliance on conventional credit scoring models presents a significant barrier, as highlighted by recent studies, which point out the dilemma faced by millions of people rendered invisible in credit assessments due to insufficient credit records [3]. These challenges underscore the urgent need to rethink and remodel the credit scoring system, particularly for the younger demographic, which often lacks physical assets but possesses valuable intangible assets like education and digital footprints [4], [5], [6], [7]. In addressing this gap, big data emerges as a potent solution with its expansive range of data sources, such as social media activities, online transactions, and telecommunication data [8], [9], [10]. Integrating social media activity data enhances the predictive accuracy of credit scoring models and marks a significant step toward financial inclusivity [11].

Building upon the context of addressing the limitations of conventional credit scoring systems, our research explores the integration of LinkedIn profile data into credit scoring models. As a professional networking platform, LinkedIn offers a rich dataset of an individual's professional attributes [12]. It provides detailed demographic information, including gender, age, education, and occupation, essential for understanding an individual's financial capacity [13]. In addition, LinkedIn's user activity data, such as interactions and language usage, offer insights into personal character traits and integrity [14]. Furthermore, Social Network Analysis of LinkedIn connections reveals an individual's societal influence, which is crucial for evaluating their credit conditions [15].

Prior research on modeling personal creditworthiness using social media data has focused on leveraging demographic and personality traits from platforms like LinkedIn [13]. Building on this foundation, our study enhances behavioral insights by introducing a 'social network category' that includes occupation and user engagement metrics to capture an individual's social interactions and professional standing more effectively. This approach complements Guo et al.'s work, which emphasizes the potential of user-generated content from broader social media platforms for credit scoring [16]. Our research aligns with these findings by incorporating professional networking data from LinkedIn, offering a nuanced perspective on financial capacity, character, and conditions. Together, these methodologies highlight the significance of combining professional and general social media activities to assess creditworthiness, advocating for a more inclusive credit evaluation system that utilizes the digital footprints of individuals with limited financial histories.

Utilizing LinkedIn's data, our model effectively translates demographics, user activities, and social networks into the '3Cs' of credit scoring: Capacity, Character, and Conditions [16]. Capacity is inferred from demographics like Age and Education, reflecting financial reliability [16], [17]. Character emerges from user activities, highlighting analytical thinking and authenticity, among others, as indicators of responsible behavior [16], [18], [19]. Conditions, drawn from social network analysis, reveal social engagement and connectivity's role. [16], [20], [21], [22].

Our research offers an alternative approach to traditional credit scoring practices by integrating a model that utilizes social media data analytics, particularly LinkedIn data. Our research demonstrates the potential of social media activity data in enhancing the precision and efficiency of credit scoring models and significantly contributes to financial inclusivity. By incorporating diverse data sources that reflect the modern consumer's digital footprint, we aim to create a more inclusive credit

evaluation system that recognizes the value of non-traditional assets and extends fair credit opportunities to a broader spectrum of individuals.

II. LITERATUR REVIEW

Credit scoring has evolved significantly in the digital era [9]. The evolution of credit scoring has been characterized by the integration of diverse data sources, the adoption of machine learning techniques, and the application of big data and advanced computational models.

Initially, credit scoring was heavily reliant on traditional financial data. However, with the rapid growth of social media and online platforms, researchers began exploring alternative data sources. A notable shift was observed with studies indicating the utility of social media activities, particularly from platforms like Weibo, in assessing creditworthiness [16]. The trend of this shift extended to using social network information from mobile phones, demonstrating improved loan default predictions [11].

Further exploring non-traditional data, researchers investigated the prediction of call data and social media activities from platforms like LinkedIn [8], [10], [13]. Exploring psychometric variables and email usage characteristics further substantiated the relevance of alternative data sources in credit scoring, particularly in regions with many unbanked individuals [3]. These studies highlighted the potential of digital footprints in providing valuable insights into an individual's creditworthiness, especially for those with limited traditional credit histories.

The integration of machine learning techniques marked a significant advancement in credit scoring. Employing models like Random Forest and AdaBoost, studies demonstrated enhanced accuracy using social network data [11]. The development of hybrid models, which combined genetic algorithms, decision trees, and logistic regression, represented a move toward more sophisticated credit-scoring methods [17]. Applying deep learning techniques, particularly in P2P lending, underscored the benefits of combining advanced methodologies with traditional credit scoring approaches [1].

The age of big data brought a paradigm shift, with studies emphasizing the increasing importance of diverse data sources and advanced analytical techniques [9]. The application of algorithms like neural networks in big data contexts showed promising results in improving the accuracy of credit risk assessments [23], [24]. These advancements highlighted the transformative potential of big data and computational models in credit scoring.

In conclusion, the evolution of credit scoring reflects the industry's adaptive response to the challenges and opportunities presented by the digital era. The journey of credit scoring from traditional financial data-based methods to models incorporating digital footprints, social media data, and big data underscores the continuous need for innovation in credit risk assessment methodologies, balancing technological advancements with ethical data use and regulatory compliance.