

Prompting is All You Need: Predicting Credit Defaults Using Diverse Text Data

Abstract

Emerging phenomena of ubiquitous text data offer promising avenues for financial institutions to make breakthroughs in credit risk management. While previous studies on credit scoring have used various natural language processing (NLP) methods to extract credit-relevant information from text, these studies typically focus on single-source text data, potentially overlooking the complex factors that collectively influence financial risks across diverse text content. Our study addresses this gap by proposing a novel prompt-based approach that integrates and assesses the importance of diverse text data. We empirically demonstrate the efficacy of our method with a unique consumer loan dataset and experiment with two variant setups: off-the-shelf and fine-tuned prompting. The results indicate that our approach significantly enhances default prediction by leveraging diverse text content and exhibits the interpretability of each text type within the predictions. Moreover, our method outperforms standard NLP benchmarks across the board. The results further demonstrate that each design component of our method contributes to improved prediction performance. This study highlights the potential of leveraging diverse text data in managing credit risk and provides strategic insights for guiding user-generated loan assessments. Our prompt-based approach marks a significant advancement in the methodological landscape of interdisciplinary research in text mining, paving the way for more nuanced and effective decision analytics.

Authors & Affiliations

Zongxiao Wu¹, Yizhe Dong¹, Yaoyiran Li², Baofeng Shi³

¹University of Edinburgh, Edinburgh, United Kingdom. ²University of Cambridge, Cambridge, United Kingdom. ³Northwest A&F University, Yangling, China