Foundation Models for Credit Risk Prediction: Game Changer or False Hope?

Abstract

Predictive models play a pivotal role in credit risk management, guiding critical financial decisions through accurate estimation of default probabilities and losses. Their performance influences the stability of the financial system and the profitability of lending operations. The relevance of predictive modeling has generated extensive interest in credit risk research, with numerous studies introducing new modeling techniques. Large-scale benchmarking studies complemented these endeavors by periodically consolidating the status quo and systematically uncovering the merits and demerits of methodological advancements. Today, quasi-standards such as a gradient-boosting model for prediction, paired with a SHAP explainer, have emerged for specific contexts. However, the continuous improvement of risk models and modeling practices remains a top priority.

Concurrently, the rapid advancements in AI, most notably through large language models, have disrupted predictive modeling paradigms in many fields. Foundation models, pretrained on extensive datasets from diverse domains, have demonstrated remarkable performance by leveraging prior knowledge acquired during pretraining. While prevalent in natural language processing and computer vision, foundation models specifically designed for tabular data have recently emerged and may hold great potential for credit risk management. We conjecture that pretraining on out-of-domain data is particularly beneficial in small-data settings, such as SME lending or managing specialized corporate portfolios. More technically, foundation models may help address longstanding credit-scoring challenges, including low default portfolios or class imbalance. However, the actual value of foundation models for credit risk prediction remains an open question.

This paper focuses on TabPFN, a pretrained, tabular data foundation model. We benchmark TabPFN against a broad set of competitors, including established and advanced (deep) machine learning techniques in two core prediction tasks: PD and LGD modeling. Our evaluation follows previous benchmarking studies, encompassing various datasets, performance indicators, and experimental conditions to clarify pretraining benefits across relevant risk modeling challenges. We also examine the profitability impact of pretraining and its consequences for model interpretability.

Preliminary results on smaller datasets suggest that TabPFN performs significantly better than state-of-the-art alternatives. Unlike the latter, TabPFN does not require hyperparameter tuning, ensuring ease of use and mitigating the computational costs associated with employing foundation models. Ongoing experiments extend the empirical study, verify the generality and robustness of results, and uncover the boundary conditions of foundation models. We anticipate that our findings will be relevant to academics interested in advancing machine learning methodologies and practitioners seeking robust, efficient, and scalable solutions for credit risk management.

Authors & Affiliations

Simon De Vos[1], Bart Baesens[1], Wouter Verbeke[1], Victor Medina-Olivares[2], Stefan Lessmann[3]
[1]KU Leuven, Leuven, Belgium. [2]University of Edinburgh Business School, Edinburgh, United Kingdom.
[3]School of Business and Economics, Humboldt-University of Berlin, Berlin, Germany