Validation of machine learning and deep learning models in credit scoring

Abstract

Traditional validation practices, which primarily focus on quantitative methods, often lack a holistic approach and struggle to validate machine learning (ML) and deep learning (DL) algorithms due to their black-box nature. Our study presents a comprehensive validation framework for ML and DL models in credit scoring. The proposed framework covers all stages of a model life cycle and aligns with the requirements of a model risk governance framework. It facilitates the proper integration of ML and DL models into enterprise risk management systems while maintaining high predictive performance.

We conducted a review of 150 published papers on ML and DL models in credit scoring, spanning the period from 2015 to 2023. After that, the most common credit scoring algorithms—traditional logistic regression, support vector machines (ML), random forest (ensemble ML), and multilayer perceptron (DL)—were chosen to create a scoring model and test the proposed validation framework on the LendingClub dataset. This dataset includes an adequate sample size, recent historical data spanning a 10-year economic cycle, an appropriate default rate, and a sufficient number of features. Model performance was evaluated using metrics such as the confusion matrix, precision ratio, F1 score, accuracy, Brier score, ROC curve, Gini coefficient, and SHAP values.

Our study contributes to credit scoring by introducing a structured approach to the validation of ML and DL models, addressing industry-specific challenges such as regulatory compliance, fairness and bias mitigation, weak model governance, and conducting use testing, which encompasses model implementation and usage, periodic monitoring and reporting, and managing validation findings and remediation plans. This framework fills a critical gap in enabling the safe deployment and usage of AI-driven scoring systems. Our findings highlight the importance of qualitative validation and ongoing use testing, including periodic monitoring. Our findings highlight the importance of qualitative validation and ongoing use testing, including periodic monitoring. By offering actionable guidelines, the proposed validation framework supports the responsible adoption of advanced AI methods, including black-box algorithms, within the financial industry.

Full Paper

Authors & Affiliations

Mr Emre Ünal[1],[2], Mrs Şükrüye Tüysüz[1]
[1]Yeditepe University, istanbul, Turkey. [2]Türkiye Finans Participation Bank, istanbul, Turkey