Evaluating AutoML Performance: Insights from Financial Predictive Analytics Tasks

Mr.Nicolas Leyh

Technical University of Munich, TUM School of Management, Germany

Abstract

Automated Machine Learning (AutoML) frameworks offer great promise for financial predictive analytics, yet limited comparative guidance exists on their domain-specific suitability. This study systematically benchmarks nine open-source AutoML frameworks - AutoGluon, Auto-Sklearn, FLAML, GAMA, Hyperopt-Sklearn, LAMA, MLJAR, ML-Plan, and TPOT - alongside the transformer-based model TabPFN. Evaluation spans 25 financial datasets, leveraging the AMLB benchmark, and is structured along three dimensions: data segment variation (e.g., class imbalance, feature count), runtime constraints (10 vs. 30 minutes), and robustness across repeated executions. The findings confirm that no single framework dominates across all conditions. While TabPFN achieved state-of-the-art results in multiclass classification on large, complex datasets, it struggled with small, binary tasks. In contrast, frameworks like Hyperopt-Sklearn and MLJAR delivered more stable results in binary classification but underperformed in more complex multiclass settings. Notably, domain-optimized frameworks like LAMA did not consistently outperform general-purpose alternatives, highlighting the context-dependence of performance. Additionally, runtime extensions yielded negligible gains, suggesting short, fixed-time budgets are often sufficient for practical deployments. This study contributes theoretically by demonstrating that cross-domain performance assumptions do not hold under financial constraints and validates the AMLB framework in finance-specific contexts. Practically, it offers actionable insights for framework selection based on task complexity, dataset characteristics, and resource limitations. The inclusion of TabPFN also marks the first empirical comparison between transformer-based and traditional AutoML approaches in finance. Future research should explore the integration of hybrid techniques, enhanced categorical handling, and benchmarking beyond classification to capture the full scope of AutoML's potential in financial services.

Keywords: AutoML frameworks, performance benchmarking, financial predictive analytics, transformer-based models