FinBoost: Boosting Survival Modelling for Financial Transactions

Manjil Nepal
SRM University-AP
India
manjil_nepal@srmap.edu.in

Parth Kiran Hanchate
SRM University-AP
India
parthkiran_hanchate@srmap.edu.in

Gao Fei
A*STAR IHPC
Singapore
gaofei@a-star.edu.sg

Wei Qingsong
A*STAR IHPC
Singapore
wei_qingsong@a-star.edu.sg

Renuga Kanagavelu
A*STAR IHPC
Singapore
renuga_k@a-star.edu.sg

Abstract

We present FinBoost, a survival modelling method for financial transactions to predict time-to-event outcomes. Leveraging a dataset of over 21.8 million records with 90 engineered features, we evaluate classical statistical models, deep learning approaches, and tree-based ensemble methods. Our experiments show that XGBoost achieves the highest predictive performance, with a maximum concordance index of **0.8472178861**. The proposed approach enables accurate and timely risk assessment in financial systems. The open-source implementation is available at: GitHub Code.

CCS Concepts

• **Computing methodologies → Machine learning**.

Keywords

Decentralized Finance, Ethereum, Survival Analysis

ACM Reference Format:
Manjil Nepal, Parth Kiran Hanchate, Gao Fei, Wei Qingsong, and Renuga Kanagavelu. 2025. FinBoost: Boosting Survival Modelling for Financial Transactions. In *Proceedings of 6th ACM International Conference on AI in Finance (ICAIF), 2025 (Financial Survival Challenge at 'ICAIF'25).* ACM, Singapore, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Survival modelling is a statistical technique that is employed to approximate the duration of time before an event happens. Although widely applied in medical and technological fields, survival analysis is not often employed to predict financial failures [7]. Finance is a sensitive domain inherently defined by risk over time, making the prediction of event timing critical. Such foresight can help businesses take proactive measures to avoid financial distress or bankruptcy. More commonly, it allows firms to mitigate the costs associated with financial distress and business failure [3]. In finance, the event of interest includes a risk-related occurrence such as default, bankruptcy, or customer churn. The observation time starts on a specific point referred to as the index event and ends with the outcome event, also referred to as the event of interest. Those subjects who do not encounter the outcome event within the observation period are said to be censored, which represents incomplete observations. However, survival datasets in finance are often difficult to obtain due to their sensitive or confidential nature. Many deep learning–based survival analyses rely on proprietary economic data requiring costly subscriptions or on medical datasets with restricted access. [4].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Financial Survival Challenge at 'ICAIF'25, Singapore

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

2 Methodology

Algorithm 1 presents the FinBoost method, which combines multiple strategies using performance-weighted ensemble predictions.

Algorithm 1 FinBoost

procedure TRAIN(X_{train}, y_{train})
  Define three strategies: Conservative, Aggressive, Balanced
  Assign weights based on transition performance
  for each strategy **do**
    Transform target using time and status
    Train XGBoost with the strategy's parameters
  end for
end procedure
procedure PREDICT(X_{test})
  for each trained model **do**
    Compute predictions
  end for
  Return weighted average of all predictions
end procedure

2.1 Dataset Description

The starting dataset for this study predicts the time to the next financial event, covering 16 possible transitions among Borrow, Deposit, Repay, Withdraw, and Liquidated. It contains over 21.8 million records with 90 engineered features, supporting 16 prediction tasks, and model performance is evaluated using the C-index.

2.2 Models Used

We evaluated a range of survival models, including classical statistical approaches, parametric models, and modern machine learning techniques. The Cox Proportional Hazards model [2] served as a baseline, while the Weibull Accelerated Failure Time (AFT) model captured parametric survival patterns. We also experimented with deep learning via DeepSurv [5] and tree-based gradient boosting methods, including XGBoost [1] and LightGBM [6], which effectively model complex nonlinear relationships and feature interactions. Among all models, XGBoost achieved the highest predictive performance, attaining the best concordance index.

2.3 Feature Engineering

We enhanced the dataset with transition-specific features to improve predictive performance.

Base survival features included user risk scoring *(liquidation_risk, repayment_ratio, leverage_ratio)*, activity patterns *(activity_volatility, transaction_frequency)*, market interaction features *(market_borrow_ratio, market_deposit_ratio, market_volatility_impact)*, temporal indicators *(is_business_hours, is_month_end, temporal_interaction* and hazard rates *(borrow_hazard, deposit_hazard)* to support better C-index performance.

Transition-specific features were engineered for each outcome: `Liquidated` *(liquidation_risk_score, collateral_health, liquidation_proximity)*, `Repay` *(repay_urgency, debt_maturity, repay_capacity)*, `Withdraw` *(withdraw_opportunity, withdraw_timing, liquidity_preference)*, `Deposit` *(deposit_attractiveness, deposit_timing, deposit_momentum)*, and `Borrow` *(borrow_necessity, borrow_capacity, borrow_risk_tolerance)*.

Additional **cross-transition interaction** features captured the *user_activity_diversity* and *market_engagement*, **temporal features** included *weekend_risk* and *month_end_risk* indicators, and **volatility features** measured *amount_volatility* and *market_deviation*, resulting in a feature-enhanced dataset tailored for each event transition.

3 Results and Discussion

3.1 Comparative Analysis

Our experiments show that XGBoost achieves the highest predictive performance, outperforming classical and ensemble survival models.

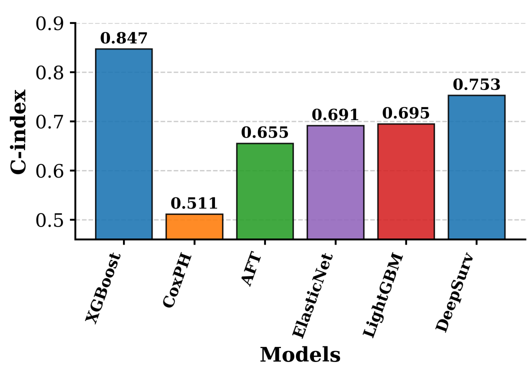

Figure 1: C-index Comparison Across Models

Figure 2 presents the detailed C-index values for all 16 event pairs across the evaluated models, offering a comprehensive comparison of model performance for each transition.

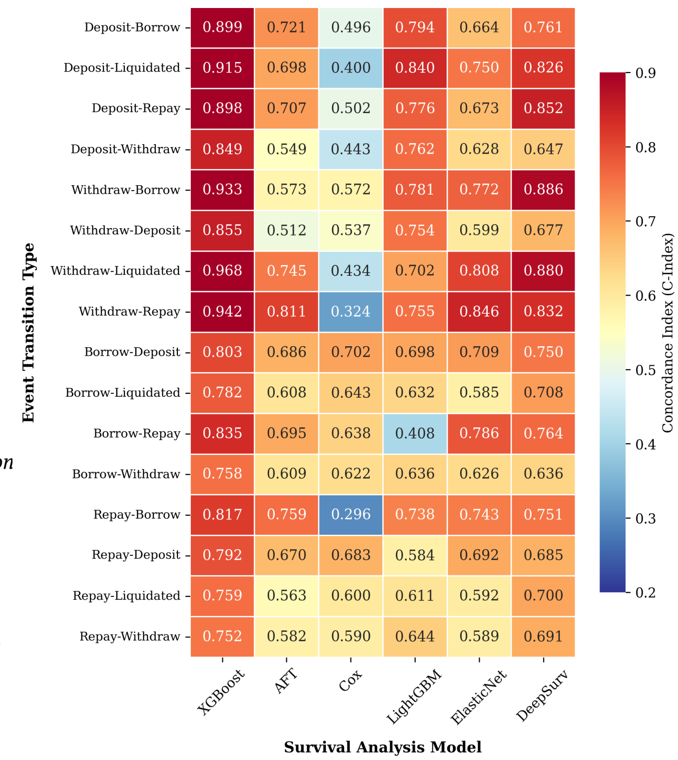

Figure 2: C-index Across Various Index-Outcome Event Pairs

4 Conclusion

In this work, we introduced FinBoost, a technique of predicting time-to-financial events based on survival modelling techniques. Our experiments show that XGBoost is effective compared to the other models, and the achieves the highest concordance index is **0.8472178861**. The results show that machine learning–based survival analysis enables accurate, timely risk assessments for proactive financial decision-making.

References

[1] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[2] David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.

[3] Adrian Gepp and Kuldeep Kumar. 2015. Predicting financial distress: A comparison of survival analysis and decision tree techniques. *Procedia Computer Science* 54 (2015), 396–404.

[4] Aaron Green, Zihan Nie, Hanzhen Qin, Oshani Seneviratne, and Kristin P Bennett. 2025. FinSurvival: A Suite of Large Scale Survival Modeling Tasks from Finance. *arXiv preprint arXiv:2507.14160* (2025).

[5] Jared Katzman, Uri Shaham, Jonathan Bates, Alexander Cloninger, Tingting Jiang, and Yuval Kluger. 2016. DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network. *Bmc Medical Research Methodology* 18, 1 (2016), 24.

[6] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017).

[7] Diego Vallarino. 2024. A Comparative Machine Learning Survival Models Analysis for Predicting Time to Bank Failure in the US (2001-2023). *Journal of Economic Analysis* 3, 1 (2024), 129–144.

Received 20 October 2025