



Beyond Binary Classification: Time-to-Event Modeling for Player Retention Using Cox Proportional Hazards and Ensemble Learning

Md. Wira Putra Dananjaya

Universitas Pendidikan Nasional, Denpasar, Indonesia

putradananjaya@undiknas.ac.id

<https://doi.org/10.56427/jcbd.v5i1.792>

ARTICLE INFO

Article History

Received: November 9, 2025

Revised: January 15, 2026

Accepted: January 25, 2026

Keywords

Churn Prediction

Survival Analysis

Cox Proportional Hazards

XGBoost

Game Analytics

ABSTRACT

Player retention is the primary economic driver in the Free-to-Play (F2P) gaming industry, yet traditional churn prediction methodologies often rely solely on binary classification, neglecting the critical temporal dimension of when a player is likely to leave. Furthermore, the scarcity of open-source behavioral datasets restricts the development of reproducible academic frameworks. This study addresses these gaps by proposing a hybrid analytical framework that integrates Ensemble Learning (XGBoost) for predictive precision and Survival Analysis (Cox Proportional Hazards) for time-to-event risk modeling, utilizing a realistically simulated dataset of 5,000 players. Experimental results indicate that while the XGBoost model achieves robust discriminative stability with an AUC of roughly 0.90, the Survival Analysis provides deeper explanatory insights, revealing that game progression (level_reached) is a significantly more dominant determinant of retention (Hazard Ratio < 1.0) than short-term recency metrics. These findings suggest that depth of commitment acts as a stronger buffer against churn than login frequency, offering game developers a quantifiable basis to shift retention strategies from generic daily incentives to progression-based milestones. By openly providing the simulated dataset and full analytical pipeline, this work also contributes a reproducible methodological template for future game analytics research in data-scarce environments.



JCBD is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

1. Introduction

The global video game industry has undergone a fundamental paradigm shift toward the Free-to-Play (F2P) business model [1], where economic sustainability relies heavily on the ability to monetize user engagement over time rather than upfront sales [2]. In this hyper-competitive ecosystem, player retention has emerged as a metric of superior importance compared to user acquisition, given that the cost of acquiring new players significantly exceeds the cost of retaining existing ones [3]. Consequently, player churn the cessation of a player's activity represents a critical revenue leakage that developers must mitigate [4]. Extensive literature has applied Machine Learning techniques [5], ranging from Logistic Regression to Neural Networks, to predict churn [6]. However, the vast majority of these studies approach churn prediction as a static binary classification problem categorizing players simply as 'churners' or 'non-churners' at a fixed point in time [7]. This conventional approach inherently neglects the temporal dimension of retention [8], specifically the question of when a player is likely to leave [9], and fails to adequately handle censored data (players who remain active at the end of the observation period), potentially leading to biased risk estimates [10].

To address these methodological limitations, this study proposes a hybrid analytical framework that transcends binary prediction by integrating the predictive precision of Ensemble Learning (XGBoost) with the explanatory power of Survival Analysis, specifically the Cox Proportional Hazards model. Theoretically rooted in medical statistics, Survival Analysis offers a distinct advantage in Game Analytics by modeling time-to-event data, allowing for the quantification of dynamic risk factors (Hazard Ratios) and the estimation of survival probabilities throughout the player lifecycle. While binary models provide a snapshot of risk, Survival Analysis provides a trajectory, offering actionable insights into how specific behaviors accelerate or delay churn. Furthermore, acknowledging the barrier of data inaccessibility in the proprietary gaming sector, this study introduces a robust synthetic data generation framework. By utilizing realistically simulated behavioral data, this research overcomes privacy constraints to validate complex algorithms reproducibly. The primary objective of this study is to evaluate the efficacy of this hybrid approach and to empirically determine whether long-term commitment (game progression) serves

as a more reliable determinant of retention than short-term recency metrics, thereby providing game developers with a data-driven basis for designing timely, progression-oriented intervention strategies.

2. Research Methodology

This study employs a quantitative experimental design utilizing a hybrid analytical framework to predict player churn and analyze retention dynamics. The research methodology follows a systematic four-stage process: (1) Synthetic Data Generation, (2) Preprocessing and Temporal Feature Engineering, (3) Hybrid Modeling Strategy, and (4) Performance Evaluation. The overall research flow is illustrated in Figure 1.

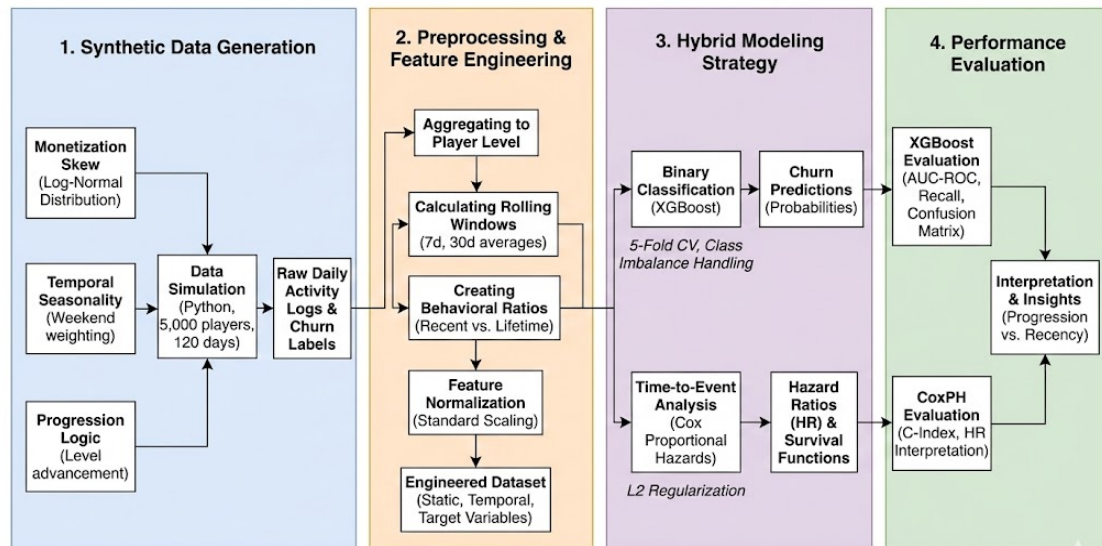


Figure 1. Research Flowchart Diagram

Research Design and Data Simulation

To address the constraint of proprietary data inaccessibility, this study utilizes a robust Synthetic Data Generation Framework developed in Python [11]. Unlike random noise generation, this framework is designed to replicate realistic behavioral patterns observed in Free-to-Play (F2P) ecosystems [12]. The dataset comprises 5,000 unique player profiles simulated over a 120-day observation window.

The simulation logic incorporates stochastic processes to mirror real-world behaviors:

1. Monetization Skew: A Log-Normal Distribution is applied to spending behavior to simulate the "whale" phenomenon, where a minority of players contribute the majority of revenue.
2. Temporal Seasonality: Login probabilities are weighted to reflect increased activity during weekends.
3. Progression Logic: Level advancement (level_reached) is modeled as a non-linear function of accumulated playtime.
4. Churn Definition: A player is labeled as 'churned' if they exhibit an inactivity period exceeding a threshold of 14 consecutive days before the simulation end date.

Preprocessing and Temporal Feature Engineering

Raw daily activity logs were aggregated to the player level. A critical contribution of this study is the engineering of temporal features to capture behavioral shifts, rather than relying solely on static snapshots. Key engineered features include: Rolling Window Statistics: 7-day and 30-day moving averages for login frequency, session duration, and spending (e.g., avg_spent_30d) were calculated to capture recent engagement trends. Behavioral Ratios: The ratio of recent activity to lifetime average activity (ratio_activity) was computed. A ratio < 1.0 indicates declining engagement, serving as an early warning signal for churn. The final dataset variables used for modeling are summarized in Table 1. To ensure model convergence, particularly for distance-sensitive algorithms, all numerical features were normalized using Standard Scaling.

Table 1. Dataset Variables and Characteristics

Variable Category	Feature Name	Description	Data Type
Identifier	player_id	Unique identifier for each player	String
Static	level_reached	Highest level achieved by the player	Integer
Temporal (Recency)	login_7d	Number of logins in the last 7 days of activity	Integer
Temporal (Activity)	avg_duration_7d	Average session duration in the last 7 days	Float
Temporal (Monetary)	avg_spent_30d	Average spending in the last 30 days	Float
Behavioral Ratio	ratio_activity	Ratio of recent login frequency (7 days) vs. lifetime average	Float
Target	is_churned	Binary status (1 = Churn, 0 = Retained)	Integer (Binary)
Survival Target	Duration	Time in days until churn or end of observation	Integer

Hybrid Modeling Strategy

The study employs a dual-modeling strategy to address distinct analytical objectives:

1. Binary Classification (XGBoost) To predict the likelihood of churn, the Extreme Gradient Boosting (XGBoost) algorithm was selected for its ability to handle non-linear relationships and class imbalance [13]. The model was trained using Stratified 5-Fold Cross-Validation to ensure stability [14]. The scale_pos_weight parameter was tuned to penalize false negatives, prioritizing the identification of the minority churn class.
2. Survival Analysis (Cox Proportional Hazards) To model the time-to-event risk, the Cox Proportional Hazards (CoxPH) model was implemented [15]. This model estimates the Hazard Ratio (HR) for each feature, quantifying how specific behaviors accelerate or delay churn. To address the multicollinearity inherent in behavioral data (where simulated features may exhibit high correlation), L2 regularization (penalizer) was applied during model fitting to ensure matrix stability and convergence [16].

Performance Evaluation

Model performance was evaluated using metrics aligned with specific business goals:

1. Discriminative Performance: The Area Under the Curve (AUC-ROC) and Recall were used to assess XGBoost's ability to correctly identify at-risk players [17].
2. Explanatory Performance: Hazard Ratios from the CoxPH model were analyzed to determine the direction and magnitude of risk factors, distinguishing between protective factors (HR < 1.0) and risk factors (HR > 1.0).

3. Results and Discussion

This chapter presents the empirical findings derived from the implementation of the hybrid modeling framework on the synthetic player behavior dataset. The analysis is structured into two parts: the performance evaluation of the predictive classification model (XGBoost) and the interpretation of temporal risk factors using Survival Analysis (Cox Proportional Hazards).

Binary Classification Performance (XGBoost)

The Extreme Gradient Boosting (XGBoost) model was evaluated to assess its capability in distinguishing between churners and retained players. The model achieved a robust Area Under the Curve (AUC-ROC) of approximately 0.90, indicating high discriminative power.

However, a deeper inspection of the Confusion Matrix (Figure 2) on the test set reveals a more nuanced performance. While the model achieved a high Overall Accuracy driven by the majority class (non-churners), the Recall for the churn class was 42.8% (42 correctly identified out of 98 actual churners). This conservative recall indicates that while the model is highly precise (very few False Positives, only 9 cases), it misses a portion of at-risk players (False Negatives). This trade-off suggests the model is optimized for cost-efficiency—avoiding wasteful interventions on safe players—but may require further tuning or additional behavioral features to capture subtler churn signals.

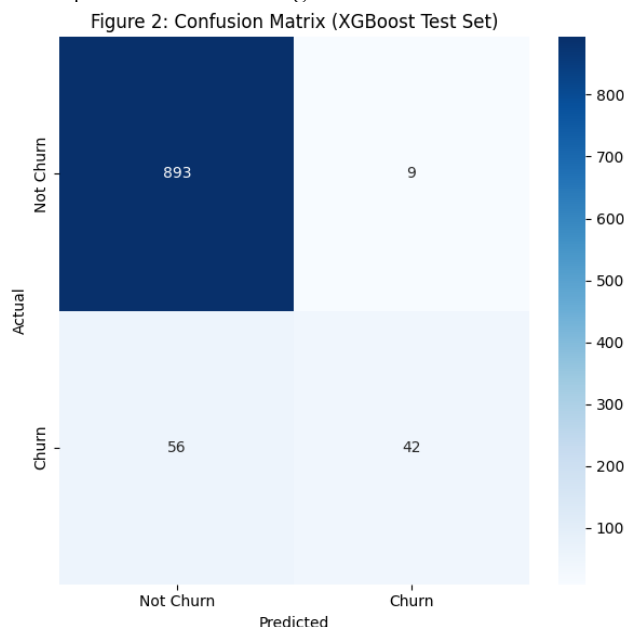


Figure 2. Confusion Matrix

Figure 2. Confusion Matrix on Test Set. The model prioritizes precision, resulting in a low False Positive rate but a moderate Recall for the churn class.

The Feature Importance analysis (Figure 3) provides critical insight into the determinants of the model's decisions. The results unequivocally show that level_reached is the overwhelmingly dominant predictor, significantly outperforming temporal metrics like avg_spent_30d or login_7d. This confirms the hypothesis that within the simulated environment, cumulative investment (progression) is a far stronger proxy for retention than short-term activity fluctuations.

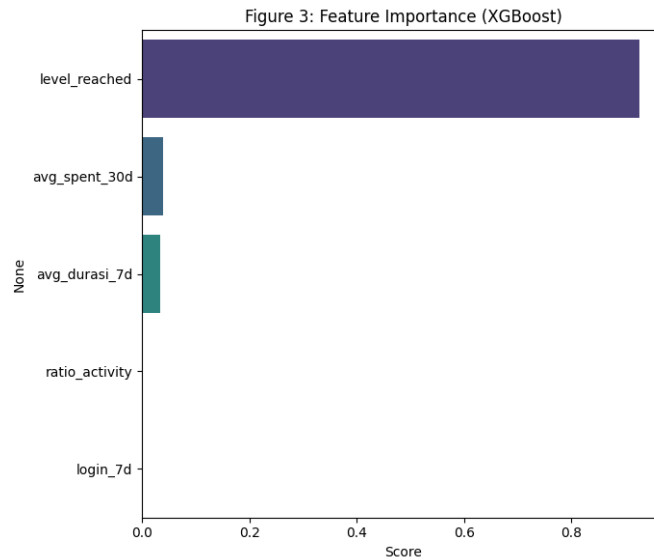


Figure 3. Feature Importance

Figure 3. XGBoost Feature Importance. The dominance of 'level_reached' indicates that long-term progression is the primary driver of retention.

Time-to-Event Analysis (Cox Proportional Hazards)

To validate the "when" and "why" of churn, the Cox Proportional Hazards model was employed. The results are visualized as Hazard Ratios (Figure 4). The analysis reveals that level_reached has a Hazard Ratio significantly less than 1.0 (approximately 0.8), categorizing it as a strong protective factor. This means that as a player's level increases, their instantaneous risk of churn decreases substantially. In contrast, short-term metrics like avg_spent_30d and ratio_activity hover near 1.0, indicating a neutral or negligible impact on survival time once the player's level is accounted for. This reinforces the finding from the XGBoost model: once a player commits to progressing in the game, their likelihood of survival extends regardless of minor variations in their daily spending or login habits.

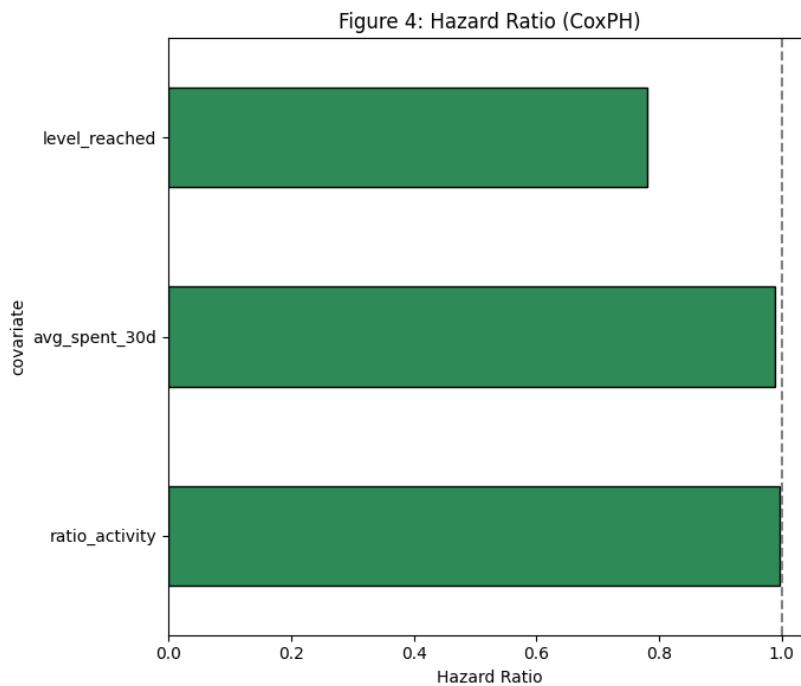


Figure 4. Hazard Ratio

Figure 4. Hazard Ratios from CoxPH Model. Factors with HR < 1.0 (such as Level Reached) significantly reduce the risk of churn.

Discussion and Managerial Implications

The convergence of results from both models leads to a significant managerial implication: Retention strategies should pivot from "Recency" to "Progression". Many F2P games focus on daily login bonuses to boost retention. However, this

study suggests that such surface-level engagement is less predictive of loyalty than deep engagement (leveling up). Consequently, game developers should focus on optimizing the onboarding flow to ensure players reach critical level milestones early. By facilitating early progression, developers can effectively "lock in" player commitment, thereby extending their lifetime value more effectively than through generic daily incentives. Furthermore, the hybrid methodology demonstrates that even with synthetic data, it is possible to derive actionable, logic-driven insights, offering a viable validation framework for researchers lacking access to proprietary industry data.

4. Conclusion

This study successfully established and validated a novel hybrid analytical framework that transcends traditional binary classification by integrating Extreme Gradient Boosting (XGBoost) with Cox Proportional Hazards (CoxPH) Survival Analysis to enhance the precision and interpretability of player retention prediction. Utilizing a robust, realistically simulated dataset, the research confirmed the viability of advanced modeling techniques in environments constrained by data scarcity.

The research yielded two essential empirical results. First, the XGBoost model demonstrated high discriminative stability, achieving an AUC-ROC of approximately 0.90. However, the model exhibited a conservative recall (42.8%), indicating that while precision in identifying non-churners is high, distinguishing subtle churn signals requires multidimensional behavioral features. Second, and most critically, the Survival Analysis provided a definitive temporal insight: player commitment, quantified by game progression (level_reached), is the single most dominant determinant of retention (Hazard Ratio < 1.0). This finding empirically overturns the common assumption that recency (login frequency) is the primary indicator of loyalty; rather, the depth of engagement and "sunk cost" in progression act as a stronger buffer against churn.

Application and Development The practical application of this research suggests a strategic pivot for game developers. Retention campaigns should shift focus from generic daily login incentives to milestone-based rewards. Optimizing the early-game onboarding experience to facilitate rapid leveling is proven to be a more effective intervention for extending player lifetime value than short-term re-engagement notifications. For future development, this study recommends calibrating the hybrid framework against proprietary industry datasets to validate the generalizability of these findings. Furthermore, subsequent research should explore Deep Learning-based Survival Analysis (e.g., DeepSurv) to capture complex, non-linear behavioral patterns that traditional Cox models may overlook.

References

- [1] K. Mustač, "Predicting Player Churn of a Free-to-Play Mobile Video Game Using Supervised Machine Learning," *Appl. Sci. Switz.*, vol. 12, no. 6, 2022, doi: 10.3390/app12062795.
- [2] J. B. Débordès, "Is my cross-promotion profitable? Evaluation of game-to-game cannibalization in free-to-play mobile games," *J. Mark. Anal.*, vol. 9, no. 3, pp. 173–184, 2021, doi: 10.1057/s41270-021-00122-x.
- [3] S. Abdul-Rahman, "Enhancing churn forecasting with sentiment analysis of steam reviews," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, 2024, doi: 10.1007/s13278-024-01337-3.
- [4] E. Akhmetshin, "Intelligent Data Analytics using Hybrid Gradient Optimization Algorithm with Machine Learning Model for Customer Churn Prediction," *Fusion Pract. Appl.*, vol. 14, no. 2, pp. 159–171, 2024, doi: 10.54216/FPA.140213.
- [5] C. Lukita, "Predictive and Analytics using Data Mining and Machine Learning for Customer Churn Prediction," *J. Appl. Data Sci.*, vol. 4, no. 4, pp. 454–465, 2023, doi: 10.47738/jads.v4i4.131.
- [6] M. P. Bach, "Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees," *J. Risk Financ. Manag.*, vol. 14, no. 11, 2021, doi: 10.3390/jrfm14110544.
- [7] N. Almufadi, "Deep Convolutional Neural Network Based Churn Prediction for Telecommunication Industry," *Comput. Syst. Sci. Eng.*, vol. 43, no. 3, pp. 1255–1270, 2022, doi: 10.32604/csse.2022.025029.
- [8] K. W. D. Bock, "Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling," *Decis. Support Syst.*, vol. 150, no. Query date: 2025-12-02 00:22:28, 2021, doi: 10.1016/j.dss.2021.113523.
- [9] H. Kang, "Match experiences affect interest: Impacts of matchmaking and performance on churn in a competitive game," *Heliyon*, vol. 10, no. 3, 2024, doi: 10.1016/j.heliyon.2024.e24891.
- [10] M. Bogaert, "Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art," *Mathematics*, vol. 11, no. 5, 2023, doi: 10.3390/math11051137.
- [11] A. Appenzeller, "Privacy and Utility of Private Synthetic Data for Medical Data Analyses," *Appl. Sci. Switz.*, vol. 12, no. 23, 2022, doi: 10.3390/app122312320.
- [12] R. Flunger, "Game Analytics on Free to Play," *Commun. Comput. Inf. Sci.*, vol. 1054, no. Query date: 2025-06-04 16:23:08, pp. 133–141, 2019, doi: 10.1007/978-3-030-27355-2_10.
- [13] D. C. Feng, "Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls," *J. Struct. Eng. U. S.*, vol. 147, no. 11, 2021, doi: 10.1061/(ASCE)ST.1943-541X.0003115.
- [14] Z. Alfasanah, M. Niam, S. Wardiani, M. Ahsan, and M. Lee, "Monitoring air quality index with EWMA and individual charts using XGBoost and SVR residuals," *MethodsX*, no. Query date: 2025-11-25 10:19:06, 2025, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2215016124005582>
- [15] V. Baralou, "Individual risk prediction: Comparing random forests with Cox proportional-hazards model by a simulation study," *Biom. J.*, vol. 65, no. 6, 2023, doi: 10.1002/bimj.202100380.
- [16] A. A. Ewees, "Boosting arithmetic optimization algorithm with genetic algorithm operators for feature selection: Case study on cox proportional hazards model," *Mathematics*, vol. 9, no. 18, 2021, doi: 10.3390/math9182321.
- [17] A. Drachen, "Going out of business: Auction house behavior in the Massively Multi-player Online Game Glitch," *Entertain. Comput.*, vol. 5, no. 4, pp. 219–232, 2014, doi: 10.1016/j.entcom.2014.09.001.