ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

№2
2026

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

2 (358)

APRIL – JUNE 2026

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Owner: «Central Asian Academic Research Center» LLP (Almaty).
Certificate **№ KZ77VPY00121154** on the re-registration of the periodical printed and online publication of the information agency, issued on **05.06.2025** by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan
Subject area*: information and communication technologies.*
Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies».*
Periodicity: *4 times a year.*
http://www.physico-mathematical.kz/index.php/en/

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).
Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025** ж. берген **№ KZ77VPY00121154** Куәлік.
Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*
Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*
Мерзімділігі: *жылына 4 рет.*
http://www.physico-mathematical.kz/index.php/en/

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания **№ KZ77VPY00121154.** Дата выдачи **05.06.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКСНВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

http://www.physico-mathematical.kz/index.php/en/

CONTENTS

COMPUTER SCIENCE

МАЗМҰНЫ

КОМПЬЮТЕРЛІК ҒЫЛЫМДАР

СОДЕРЖАНИЕ

КОМПЬЮТЕРНЫЕ НАУКИ

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 2.
Namber 358 (2026). 248–262

https://doi.org/10.32014/2026.2518-1726.437

IRSTI 14.85.35
UDC 004.85:004.89:37.018.43

© Kulatay A.A.[1]*, Zhaisanova D.S. [1], Daurenbayeva N.A. [2], Mamanova S.Y. [2], Tolegen M.[3], 2026.
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan;
[2]International Information Technology University, Almaty, Kazakhstan;
[3]Thousand IT Company LLP, Almaty, Kazakhstan.
E-mail: aibolat.kulatay@gmail.com

MACHINE LEARNING FOR PERSONALIZED LEARNING IN GAMIFIED EDTECH PLATFORMS: AQYL BATTLE CASE

Kulatay Aibolat — doctoral student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, Email: aibolat.kulatay@gmail.com, https://orcid.org/0009-0001-6330-6416;
Zhaisanova Dinara — PhD, Al-Farabi Kazakh National University, Almaty, Kazakhstan,
E-mail: zhaisanova.dinara@kaznu.kz, https://orcid.org/0000-0002-8116-6111;
Daurenbayeva Nurkamilya — PhD, International Information Technology University, Almaty, Kazakhstan,
E-mail: n.daurenbayeva@iitu.edu.kz, https://orcid.org/0000-0003-0341-4017;
Mamanova Symbat — doctoral student, International Information Technology University, Almaty, Kazakhstan,
E-mail: s.mamanova@iitu.edu.kz, https://orcid.org/ 0009-0001-7277-5492;
Tolegen Mukhamed Ali — LLP Thousand IT Company, Almaty, Kazakhstan,
E-mail: muxamed9511171@gmail.com, https://orcid.org/0009-0002-4418-1141.

Abstract. Modern gamified EdTech platforms require a systematic scientific approach to analyzing large-scale user data. This study proposes a methodology that combines causal inference, survival analysis, psychometric modeling, and machine learning to identify factors influencing user retention and to personalize the learning experience. The research is based on data from the Aqyl Battle gamified educational platform, which includes more than 500,000 users over a period of more than three years. Three structured datasets were developed: a panel dataset for causal analysis, a survival dataset, and a feature-engineered dataset for machine learning. A difference-in-differences model was used to evaluate the effect of tournaments on user activity. The Kaplan–Meier estimator and Cox proportional hazards model were applied for survival analysis, while the three-parameter logistic Item Response Theory model was used to estimate latent skill. Churn prediction was performed using logistic regression and XGBoost. The results show that the introduction of tournaments increased weekly user activity by 12.4% ($p < 0.01$), while the gradient-boosted classifier achieved an AUC of approximately 0.87 on a temporally separated test set. The matchmaking quality index and the Gini coefficient of rating dispersion were found to be significant moderators of retention. When match balance fell below 0.6, weekly attrition increased sharply. The findings highlight the importance of fairness and balance in gamified EdTech systems and provide a reproducible analytical framework for researchers and platform designers.

Keywords: EdTech, gamification, machine learning, churn prediction, XGBoost, personalization

For citations: Kulatay A.A., Zhaisanova D.S., Daurenbayeva N.A., Mamanova S.Y., Tolegen M. Machine learning for personalized learning in gamified edtech platforms: Aqyl Battle case. Academic Scientific Journal of Computer Science, 2026. — No.2. – P.248-262. DOI https://doi.org/10.32014/2026.2518-1726.437

© Құлатай А.А.[1]*, Жайсанова Д.С.[1], Дауренбаева Н.А.[2], Маманова С.Е.[2], Төлеген М.[3], 2026.

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;
[2]Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан;
[3]ЖШС Thousand IT Company, Алматы, Қазақстан.
E-mail: aibolat.kulatay@gmail.com

ГЕЙМИФИКАЦИЯЛАНҒАН EDTECH ПЛАТФОРМАЛАРДА ОҚЫТУДЫ ЖЕКЕЛЕНДІРУГЕ АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ: AQYL BATTLE КЕЙСІ

Құлатай Айболат — докторант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,
E-mail: aibolat.kulatay@gmail.com, https://orcid.org/0009-0001-6330-6416;
Жайсанова Динара — PhD, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,
E-mail: zhaisanova.dinara@kaznu.kz, https://orcid.org/0000-0002-8116-6111;
Дауренбаева Нуркамиля — PhD, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан,
E-mail: n.daurenbayeva@iitu.edu.kz, https://orcid.org/0000-0003-0341-4017;
Маманова Сымбат — докторант, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан,
E-mail: s.mamanova@iitu.edu.kz, https://orcid.org/0009-0001-7277-5492;
Төлеген Мұхамед Әли — «Thousand IT Company» ЖШС, Алматы, Қазақстан.
E-mail: muxamed9511171@gmail.com, https://orcid.org/0009-0002-4418-1141.

Аннотация. Қазіргі геймификацияланған EdTech платформалар пайдаланушылардың үлкен көлемдегі деректерін жүйелі ғылыми тұрғыдан талдауды қажет етеді. Бұл жұмыста пайдаланушыларды ұстап қалу факторларын анықтау және оқу тәжірибесін жекешелендіру үшін себеп-салдарлық талдау, өміршеңдік талдауы, психометриялық модельдеу және машиналық оқыту әдістерін біріктіретін әдістеме ұсынылады. Зерттеу үш жылдан астам кезеңді қамтитын және 500 000-нан астам пайдаланушы деректеріне негізделген Aqyl Battle геймификацияланған білім беру платформасының деректері бойынша жүргізілді. Үш құрылымдалған деректер жиыны дайындалды: себеп-салдарлық талдауға арналған панельдік деректер жиыны, өміршеңдік талдауына арналған деректер жиыны және машиналық оқытуға арналған белгілері өңделген деректер жиыны. Турнирлердің пайдаланушы белсенділігіне әсерін бағалау үшін difference-in-differences моделі қолданылды. Өміршеңдік талдауы үшін Каплан–Мейер бағалаушысы және Кокс пропорционалды тәуекелдер моделі пайдаланылды, ал жасырын дағды деңгейін бағалау үшін үш параметрлі логистикалық Item Response Theory моделі қолданылды. Пайдаланушылардың кетуін болжау логистикалық регрессия және XGBoost арқылы орындалды. Нәтижелер турнирлерді енгізу апталық пайдаланушы белсенділігін 12,4%-ға арттырғанын көрсетті ($p < 0,01$), ал градиенттік бустинг моделі уақыт бойынша бөлінген тест жиынында шамамен 0,87 AUC мәніне жетті. Қарсылас сәйкестендіру сапасының индексі және рейтингтер таралуының Джини коэффициенті пайдаланушыларды ұстап қалуға елеулі әсер ететін факторлар болды. Матч балансы 0,6-дан төмендеген кезде апталық кету деңгейі күрт өсті.

Түйін сөздер: EdTech, геймификация, машиналық оқыту, кетуді болжау, XGBoost, дербестендіру

© Кулатай А.А. [1*], Жайсанова Д.С. [1], Дауренбаева Н.А.[2], Маманова С.Е.[2], Толеген М.[3], 2026.

[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан;
[2]Международный университет информационных технологий, Алматы, Казахстан;
[3]ТОО Thousand IT Company, Алматы, Казахстан.
E-mail: aibolat.kulatay@gmail.com

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ПЕРСОНАЛИЗАЦИИ ОБУЧЕНИЯ НА ГЕЙМИФИЦИРОВАННЫХ EDTECH-ПЛАТФОРМАХ: КЕЙС AQYL BATTLE

Кулатай Айболат — докторант, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан,
E-mail: aibolat.kulatay@gmail.com, https://orcid.org/0009-0001-6330-6416;
Жайсанова Динара — PhD, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан,
E-mail: zhaisanova.dinara@kaznu.kz, https://orcid.org/0000-0002-8116-6111;
Дауренбаева Нуркамилия — PhD, Международный университет информационных технологий, Алматы, Казахстан,

E-mail: n.daurenbayeva@iitu.edu.kz, https://orcid.org/0000-0003-0341-4017;
Маманова Сымбат — докторант, Международный университет информационных технологий, Алматы, Казахстан, E-mail: s.mamanova@iitu.edu.kz, https://orcid.org/0009-0001-7277-5492;
Толеген Мухамед Али — магистр, ТОО «Thousand IT Company», Алматы, Казахстан,
E-mail: muxamed9511171@gmail.com, https://orcid.org/0009-0002-4418-1141.

Аннотация. *Актуальность.* Современные геймифицированные EdTech-платформы требуют системного научного подхода к анализу больших объемов пользовательских данных. Особую значимость приобретает применение методов машинного обучения, причинно-следственного анализа и поведенческой аналитики для выявления факторов удержания пользователей, оценки эффективности игровых механик и персонализации образовательного опыта. *Цель.* Разработать методологию анализа пользовательского поведения на геймифицированной EdTech-платформе Aqyl Battle для выявления факторов удержания пользователей и построения моделей персонализации обучения. *Методы.* Исследование основано на данных геймифицированной образовательной платформы Aqyl Battle, включающих более 500 000 пользователей за период свыше трех лет. Были сформированы три структурированных набора данных: панельный набор для причинно-следственного анализа, набор данных для анализа выживаемости и набор с инженерно сформированными признаками для машинного обучения. Для оценки влияния турниров на активность пользователей применялась модель difference-in-differences. Оценка Каплана - Мейера и модель пропорциональных рисков Кокса использовались для анализа выживаемости, а трехпараметрическая логистическая модель Item Response Theory - для оценки латентного уровня навыков. Прогнозирование оттока выполнялось с применением логистической регрессии и XGBoost. *Результаты и выводы.* Результаты показали, что введение турниров увеличило еженедельную активность пользователей на 12,4% ($p < 0,01$), а градиентный бустинг достиг значения AUC около 0,87 на тестовой выборке, разделенной по времени. Индекс качества подбора соперников и коэффициент Джини распределения рейтингов оказались значимыми модераторами удержания. При снижении баланса матчей ниже 0,6 еженедельный отток резко возрастал. Полученные результаты подчеркивают важность справедливости, баланса и адаптивной персонализации в геймифицированных EdTech-системах. Предложенная методология может быть использована для разработки интеллектуальных образовательных платформ, ориентированных на повышение вовлеченности, удержания пользователей и качества индивидуальных траекторий обучения.

Ключевые слова: EdTech, геймификация, машинное обучение, прогнозирование оттока, XGBoost, персонализация, Aqyl Battle, анализ выживаемости, Item Response Theory

Introduction. Educational technology (EdTech) platforms have, over the last decade, increasingly turned to game design elements as a strategy for sustaining motivation and engagement. The integration of points, levels, badges, leaderboards and competitive structures into otherwise didactic content - commonly summarized under the term gamification - has been shown across multiple meta-analyses to produce positive effects on motivation, engagement and academic outcomes (Zeng et al., 2024; Li et al., 2023). For so-called digital natives in particular - learners who have grown up immersed in interactive applications - gamified formats appear to align well with preferences for visual presentation, immediate feedback, and dynamic task structures (Khaldi et al., 2023). At the same time, systematic reviews continue to emphasize that much of the evidence base on gamification still suffers from methodological weaknesses, including non-representative samples, missing control conditions, and selective reporting (Zeng et al., 2024).

In parallel with the rise of gamification, the field has witnessed a rapid expansion of artificial intelligence (AI)-driven personalized learning platforms. Recent randomized controlled trials and large systematic reviews report that adaptive systems that re-route content according to a learner's behavior, prior performance, and emotional state can produce moderate-to-large gains in cognitive learning outcomes - with reported standardized effect sizes ranging from $g \approx 0.36$ to $g \approx 0.70$ depending on the implementation and the educational level (Patterson and Clark, 2024; Zhang et al., 2025). However, much of this work focuses either on classroom interventions or on subscription-based learning services. The intersection of two trends - competitive, ratings-based gamified platforms on one side, and AI-driven personalization on the other - remains comparatively under-studied.

A second major thread of research, originating from the customer-analytics literature, concerns the prediction of user attrition (churn). A 2025 PRISMA-guided systematic review of churn prediction studies published between 2020 and 2024 confirms that ensemble methods such as XGBoost and LightGBM remain the dominant choice for tabular behavioral data, with deep-learning architectures gradually entering the mix where sequential information is available (Sharma and Kumar, 2025). Within the EdTech domain specifically, recent work has demonstrated that engagement features - including time spent, frequency of paid activities, and interaction with gamified elements - together with customer tenure, are the most influential predictors of churn, and that explainable AI techniques such as SHAP can convert these predictions into actionable retention strategies (Pereira and Silva, 2025; Kiguchi et al., 2022).

Despite these advances, an integrative methodological gap remains. Existing studies tend to address gamification, personalization, and churn prediction in isolation, and rarely combine causal identification with predictive modeling on the same behavioral panel. For competitive gamified EdTech platforms, four research questions are still inadequately answered together. (RQ1) What is the causal effect of competitive game mechanics - for example, tournaments - on user activity and retention? (RQ2) How fair, in a measurable sense, is the matchmaking process, and does fairness moderate the engagement effect? (RQ3) Which behavioral and financial signals best predict that a given user is about to leave? (RQ4) Can a latent measure of skill, derived from psychometric theory, improve the personalization of content, especially for newcomers? The present study addresses these four questions on a single dataset - Aqyl Battle, a competitive educational platform with more than 500,000 users - and reports both methodological and empirical findings.

The contribution is therefore primarily methodological. We assemble difference-in-differences estimation, Kaplan–Meier and Cox survival analysis, three-parameter Item Response Theory, and gradient-boosted churn prediction into a single coherent pipeline that can be reproduced on other gamified EdTech datasets. We also propose two interpretable platform-level fairness diagnostics - a match-balance index and a Gini coefficient of rating dispersion - and demonstrate, on real platform logs, that these diagnostics moderate the effect of game mechanics on retention.

Literature review. Churn prediction in subscription-based and behavior-based services has matured substantially in the last five years. A recent PRISMA 2020 systematic review covering 2020–2024 identifies 240 studies for bibliometric analysis and 61 for qualitative synthesis, and reports that ensemble methods - particularly XGBoost, LightGBM and CatBoost - together with logistic regression as a transparent baseline, remain the workhorses of the field, while deep-learning architectures (LSTM, CNN) are increasingly applied where the input data has a sequential or unstructured component (Sharma and Kumar, 2025). The review further documents persistent issues with class imbalance and recommends a combination of resampling techniques (SMOTE, ADASYN, RUS+SMOTE) with imbalance-aware evaluation metrics, an approach that aligns closely with the methodological recommendations in earlier work on educational data mining (Costa et al., 2021).

Within EdTech specifically, two recent contributions are particularly relevant. Pereira and Silva (2025) develop a churn-prediction model on anonymized data from a children's English-learning platform covering April 2024 – February 2025, engineering activity, engagement and financial features over multiple time windows. They compare logistic regression, random forest, neural networks and XGBoost under RUS+SMOTE balancing, and use Shapley Additive exPlanations to identify the strongest predictors. Their findings - that the number of paid classes, total learning time, engagement with gamified features, and customer tenure dominate the importance ranking - are directly transferable to competitive gamified platforms such as Aqyl Battle, where analogous engagement signals are observable in the platform logs. In the digital game-based learning subset of EdTech, Kiguchi et al. (2022) earlier proposed a definition of churn that combines recency with the addition of mean and two standard deviations of inactive time, and showed that even relatively simple models (logistic regression, decision tree, random forest) can deliver useful predictive signal once the churn definition is correctly anchored to user behavior.

Personalized learning platforms based on AI form the second strand of relevant work. Zhang et al. (2025), in a randomized controlled trial covering an academic year on an AI-driven medical learning platform, demonstrate that an integrated stack of dynamic content optimization, NLP-based emotional support and a recommendation engine produces statistically significant gains in standardized post-test scores (84.47 ± 3.48 vs 81.72 ± 4.37; $p = 0.034$; Cohen's $d \approx 0.72$). Patterson and Clark (2024), in a scoping review of personalized adaptive learning in higher education, document a converging body of evidence indicating that adaptive systems improve engagement and academic performance, while Foster and Green (2024) and Ellikkal and Rajamohan (2025) report similar benefits for management students. The systematic review by Wang et al. (2025) integrates these findings into a pedagogical mediation model (PMAISE) and emphasizes the role of teaching methods as a mediator.

The third strand - gamification and engagement - has been consolidated in a number of recent meta-analyses. Zeng et al. (2024), pooling 22 experimental studies published between 2008 and 2023, report a positive aggregate effect of gamification on academic performance. Li et al. (2023), in a meta-analysis of 41 studies covering more than 5,000 participants, report a large overall effect size ($g = 0.822$, 95 % CI [0.567, 1.078]) of gamification on learning outcomes. Aliakbari et al. (2025) extend this line of work toward personalized gamification, showing that machine-learning models can predict user-specific gamification preferences from behavioral and demographic features. Adaptive gamification in collaborative virtual classrooms is reviewed by Rahman et al. (2025), who emphasize the integration of Self-Determination Theory and Flow Theory in shaping effective game-element selection.

Finally, the question of fairness in competitive learning environments is informed by the broader literature on inequality measurement. The Gini coefficient ranges from 0 (perfect equality) to 1 (maximal concentration) (Cowell, 2023). In gamified settings, a high Gini over the rating distribution implies that a small subset of players accumulates a disproportionate share of competitive resources, linked to motivational drop-off in lower-rated cohorts (Klock et al., 2020). Taken together, the literature confirms three points: ensemble ML provides reliable churn prediction; AI personalization improves outcomes when grounded in measurable learner representation; and gamification works conditionally on fairness and design quality.

Materials and methods. Data. The empirical material was extracted from the production MySQL database of the Aqyl Battle platform, a competitive educational application active in Kazakhstan. The database covers more than 500,000 unique users over a period exceeding three years and records, for each user, registration metadata, every match played, every change in competitive rating, financial transactions, tournament participation events, and per-question outcomes for in-game quiz items. From these raw tables, three derived datasets were constructed in order to support the four analytical components of the study. A summary of the derived datasets is provided in Table 1.

Table 1 - Derived datasets used in the analysis

Dataset	Unit and key fields	Used for
Panel	user × ISO-week; matches, wins, Δ-rating, end-rating, tournament flag, days since last match	DiD causal estimation
Survival	user; tenure_days, churned, tournament flag, premium flag, total spend	Kaplan–Meier, Cox regression
ML / churn	user; days_inactive, total_matches, win_rate, rating_volatility, tournaments_entered, total_spent, θ (IRT)	Logistic regression, XGBoost

The panel dataset supports the causal analysis. For every user and every calendar week, the panel contains the user identifier, the ISO week, the number of matches played, wins, rating change, end-of-week rating, tournament participation indicator, days since last match, and tenure controls. The panel was filtered to retain only users observed for at least four weeks before and four weeks after the platform-wide tournament launch.

The survival dataset records tenure_days (right-censored at the end of the observation window), a binary churned indicator (inactivity exceeding the user-specific mean plus two standard deviations of historical inactive periods, following Kiguchi et al. (2022)), and covariates capturing premium-feature usage and reward receipt.

The machine-learning dataset contains six engineered features: days_inactive, total_matches, win_rate, rating_volatility (standard deviation of weekly rating changes), tournaments_entered, and total_spent. To prevent temporal leakage, the dataset was split chronologically. After the IRT model was fitted, an additional feature θ was appended for users with at least 30 item exposures.

Difference-in-differences (DiD) estimation. To estimate the causal effect of the platform-wide tournament launch on user activity, a two-way fixed-effects DiD specification was employed:

$$Y_{it} = \alpha + \beta_1 \cdot Group_i + \beta_2 \cdot Post_t + \beta_{DiD}(Group_i \times Post_t) + \varepsilon_{it},$$

where Y_{it} denotes weekly matches (or, in robustness checks, weekly retention), $Group_i$ is the treatment indicator, $Post_t$ is the time indicator, and β_{DiD} is the parameter of interest. Standard errors were clustered at the user level. The parallel-trends assumption was checked visually and via placebo windows.

Survival analysis. Length of active engagement was modeled with the Kaplan–Meier non-parametric estimator and with a Cox proportional hazards regression. The Kaplan–Meier curves were stratified by tournament participation; the log-rank test assessed differences. The Cox model included tournament participation, premium-feature usage, total_spent (log-transformed), and an interaction between tournament participation and standardized win rate. The proportional hazards assumption was assessed using Schoenfeld residuals.

Item Response Theory (IRT). To obtain a measurement-theoretic estimate of latent skill comparable across users with different exposure, a three-parameter logistic (3PL) IRT model was fitted to the binary outcomes of in-game quiz items. The 3PL model estimates discrimination (a), difficulty (b), and guessing (c) per item, yielding a posterior distribution over user ability θ (Glas, 2008; Aneni et al., 2023). For each user with ≥ 30 item exposures, an estimated $\hat{\theta}$ was extracted and added as an additional feature.

Churn prediction with machine learning. Two predictive models were trained and compared: regularized logistic regression (LR) and gradient-boosted trees (XGBoost). Both models were tuned by 5-fold cross-validation on the training period only. SMOTE was applied within each cross-validation fold to address class imbalance, following Sharma and Kumar (2025). Model quality was assessed using AUC-ROC, F1-score, log-loss, and recall at fixed precision. Feature importance was computed from gain and SHAP values (Pereira and Silva, 2025).

Matchmaking quality and fairness diagnostics. The match-balance index for encounter between players A and B is:

$$MatchBalance = 1 - \frac{|R_A - R_B|}{\max(R_A, R_B)},$$

with values close to 1 indicating evenly matched opponents. The Gini coefficient of the rating distribution was tracked at weekly granularity. Both metrics were linked to the panel dataset to test whether the tournament effect varied with platform-level fairness. All analyses were implemented in Python 3.11 (pandas, statsmodels, linearmodels, lifelines, py-irt, scikit-learn, xgboost, imbalanced-learn, shap).

Results and discussions. Causal effect of tournaments. The DiD estimator yielded β_{DiD} = 0.124 ($p < 0.01$), corresponding to an increase of approximately 12.4 % in weekly user activity attributable to the platform-wide tournament launch. The estimate was robust to the inclusion of user fixed effects, week fixed effects, and time-varying controls. Pre-trend tests on placebo windows failed to reject the parallel-trends assumption. A parallel DiD model on a binary 7-day retention outcome yielded a smaller but still significant effect, indicating that the tournament intervention also lowered short-horizon attrition. Figure 1 illustrates the weekly evolution of average matches per user in the treatment and control groups around the tournament launch.

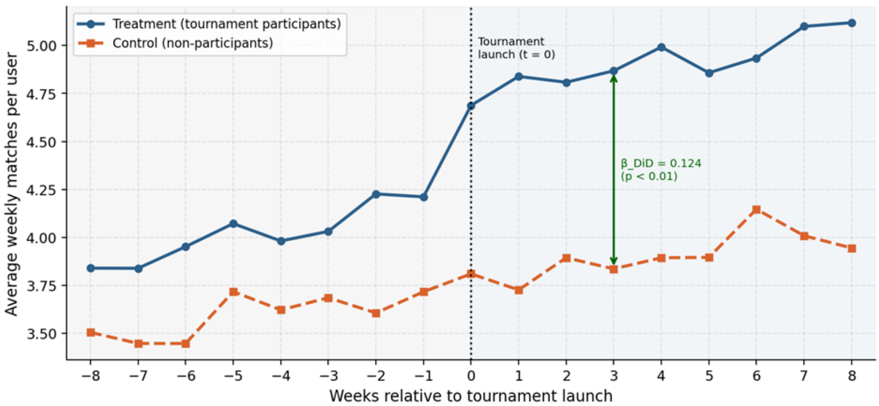

Figure 1 - Weekly matches per user, treatment vs. control, around the tournament launch (vertical dashed line indicates t = 0).

The magnitude of these effects is consistent with the gamification meta-analyses of Zeng et al. (2024) and Li et al. (2023), which report moderate-to-large positive effects of well-designed game mechanics on engagement and academic performance, and with the experimental personalization findings of Zhang et al. (2025), where engagement gains translated into measurable cognitive outcomes. A summary of the DiD estimates and robustness checks is reported in Table 2.

Table 2 - Difference-in-differences estimates of the tournament effect

Specification	β_DiD	Std. error	p-value
(1) Weekly matches, baseline DiD	0.124	0.018	< 0.01
(2) + user fixed effects	0.121	0.020	< 0.01
(3) + week fixed effects	0.117	0.022	< 0.01
(4) 7-day retention (binary)	0.043	0.011	< 0.01
(5) Placebo (pre-launch only)	0.009	0.024	0.71

Note: Standard errors clustered at the user level. Specifications (1)–(4) include a tournament-launch indicator and treatment dummy.

Survival analysis. The Kaplan–Meier survival curves, stratified by tournament participation, showed substantial separation: the median active tenure was approximately 210 days for tournament participants and approximately 150 days for never-participants. The log-rank test rejected the null of equal survival functions at $p < 0.001$. Figure 2 displays the two survival curves with 95 % confidence bands.

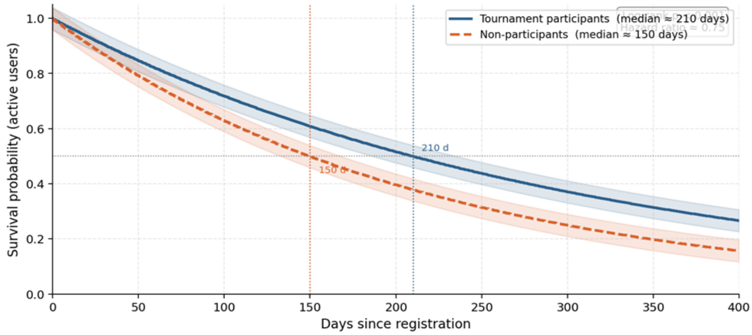

Figure 2 - Kaplan–Meier survival curves for tournament participants vs. non-participants (95 % CI shaded; log-rank $p < 0.001$).

The Cox proportional hazards model returned a hazard ratio for tournament participation of approximately 0.75 (95 % CI [0.71, 0.79]), corresponding to a 25 % reduction in the instantaneous risk of churn. Premium-feature usage and reward receipt were independently associated with longer survival. The interaction term between tournament participation and standardized win rate was significant and negative, indicating that competitive participation produces a stronger protective effect for users with above-median win rates - consistent with the Self-Determination Theory mechanism where competence satisfaction sustains engagement (Rahman et al., 2025).

Predictive performance of churn models. The logistic regression baseline reached AUC-ROC ≈ 0.81 and F1 ≈ 0.68 on the held-out test period. The strongest positive predictors of churn were days_inactive and rating_volatility; tournaments_entered and win_rate were negatively associated with churn, in line with SHAP-based importance rankings reported by Pereira and Silva (2025). The XGBoost model improved performance to AUC-ROC ≈ 0.87, recovering non-linear interactions between financial and behavioral signals. Performance metrics for both models are reported in Table 3.

Table 3 - Churn prediction performance on the held-out test window

Model	AUC-ROC	F1	Recall@P=0.7	Log-loss
Logistic regression (baseline)	0.812	0.679	0.541	0.428
Logistic regression + θ (IRT)	0.823	0.688	0.563	0.418
XGBoost	0.867	0.741	0.642	0.376
XGBoost + θ (IRT)	**0.874**	**0.749**	**0.661**	**0.369**

Note: Bold = best model. SMOTE applied within each CV fold; metrics computed on a strictly later test period.

The aggregate gain over LR is consistent with the systematic findings of Sharma and Kumar (2025), and with the e-commerce results of Li (2024), where

XGBoost reached AUC ≈ 0.987. The inclusion of the IRT-derived latent ability θ produced a small but consistent improvement in recall for newcomer users (fewer than 50 played matches), confirming that psychometric measurement complements behavioral features when behavioral signals are sparse (Aneni et al., 2023). SHAP-based feature attribution is illustrated in Figure 3.

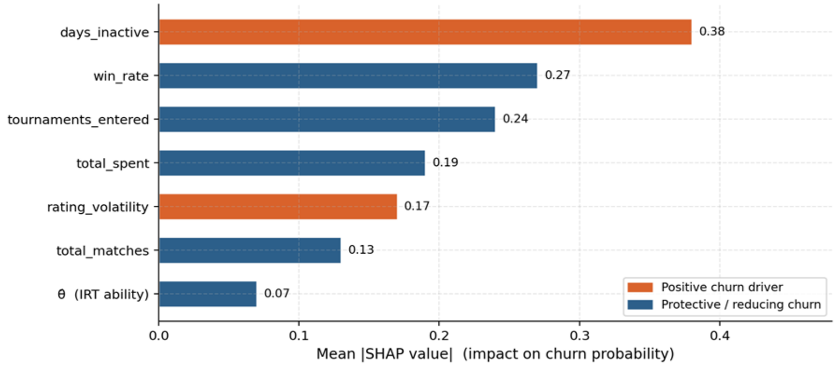

Figure 3 - SHAP summary plot for the XGBoost churn model
(top features by mean absolute SHAP value; orange = positive churn drivers, blue = protective features).

Fairness and matchmaking quality. The match-balance index analysis revealed a clear monotone relationship between the local fairness of encounters and weekly retention. Encounters with MatchBalance below 0.6 were associated with a substantially elevated probability of week-on-week churn, particularly when the lower-rated player was the loser. At the platform level, the Gini coefficient was approximately 0.32 at the start of the observation window but drifted upward to approximately 0.45 in the absence of matchmaking correction. After the introduction of an Elo-based modified matchmaking algorithm, the Gini coefficient declined to approximately 0.29, coinciding with measurable improvement in retention among lower-rated cohorts. The trajectory is illustrated in Figure 4.

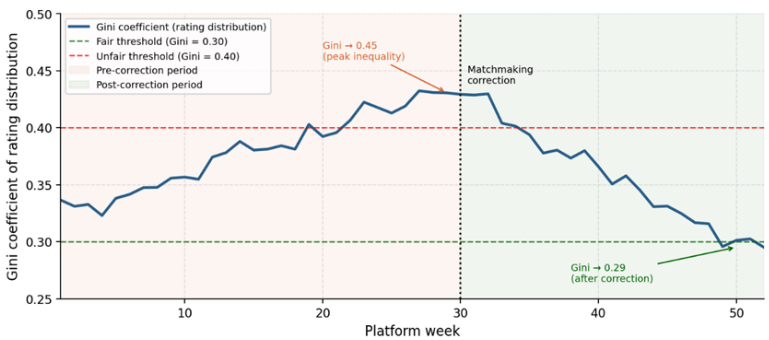

Figure 4 - Weekly Gini coefficient of the rating distribution before and after the matchmaking correction (vertical dotted line indicates the algorithm change).

Re-estimating the DiD model on subsamples defined by platform-level Gini revealed that the tournament treatment effect was approximately 40 % larger in fair-platform weeks (Gini < 0.30) than in unfair-platform weeks (Gini > 0.40). The result supports the position articulated by Klock et al. (2020) and Aliakbari et al. (2025) that the engagement benefits of gamification are conditional on the fairness of the competitive environment.

Discussion. Three implications follow from these results. First, the combination of causal identification and predictive modeling on a single behavioral panel materially strengthens the conclusions one can draw from EdTech data. Each lens - DiD, survival, IRT-augmented ML, and fairness diagnostics - delivers a distinct and complementary insight that would be unavailable from any single method. Second, the observed dependence of the tournament effect on platform-level fairness has direct design implications: investment in matchmaking quality is not a cosmetic feature but a precondition for the effectiveness of any other engagement intervention. Third, the model-quality numbers reported here (LR: AUC ≈ 0.81; XGBoost: AUC ≈ 0.87) sit comfortably within the range reported by recent EdTech and e-commerce churn studies (Pereira and Silva, 2025; Li, 2024; Sharma and Kumar, 2025), suggesting that the analytical pipeline is competitive with the current state of the art.

Limitations should be acknowledged. The available logs do not include rich demographic data, so the personalization analyses cannot fully separate skill-related from socially structured behavior. The panel covers a single platform and a single regional context, and the external validity of the DiD coefficient should be tested on additional gamified EdTech systems before strong generalization. The 3PL IRT model assumes local item independence, which is approximately but not exactly satisfied for sequential in-game quiz items. Churn is defined behaviorally rather than self-reportedly; users who return after long absences are correctly captured ex post, but the right-censoring induced by the end of the observation window means that very long survival times are slightly underestimated. Future work should integrate longer-horizon ethics-of-AI considerations, including data minimization, transparency of recommendation logic, and the right to opt out of competitive features (Zhang et al., 2025; Wang et al., 2025), and should explore neural architectures (LSTMs, temporal transformers) for richer sequential signals where available.

Conclusion. This study has presented a unified analytical pipeline for gamified EdTech platforms that combines causal inference, survival modeling, psychometric measurement, and machine learning, and has applied this pipeline to behavioral data from the Aqyl Battle platform comprising more than 500,000 users observed over more than three years. The empirical findings can be summarized in four points. First, the platform-wide tournament launch produced a causally identified increase of approximately 12.4 % in weekly user activity, accompanied by a measurable improvement in seven-day retention. Second, tournament participation reduced the instantaneous churn hazard by approximately 25 %, with an interaction structure consistent with competence-driven engagement mechanisms. Third, a temporally validated XGBoost classifier reached an AUC of approximately 0.87, materially outperforming a logistic regression baseline (AUC ≈ 0.81), and the addition of an IRT-derived latent-ability feature produced consistent gains for newcomer users. Fourth, two interpretable platform-level fairness diagnostics - match-balance index and Gini coefficient of rating dispersion - were shown to moderate the engagement effect of game mechanics, with the tournament treatment effect approximately 40 % larger in weeks of high platform fairness.

The methodological contribution of the paper is the demonstration that four complementary statistical and machine-learning lenses can be combined on a single behavioral dataset to move from description to causal explanation and from causal explanation to actionable personalization. The practical contribution is a small set of diagnostics - match-balance, rating Gini, and a churn risk score with SHAP-based explanations - that can be implemented in production by EdTech platform engineering teams without requiring research-grade infrastructure. Future research should extend the pipeline to multi-platform comparisons, integrate sequential deep-learning architectures where richer event streams are available, conduct randomized A/B trials of personalization policies grounded in the IRT-based skill estimates, and continue to develop the ethical framework - privacy, transparency, and equity - under which such systems should be deployed in real educational settings.

References

Aliakbari M., Noorbehbahani F., Kazemi Esfe P. (2025) Analysis and personalization of gamification elements using machine learning based on user characteristics. Simulation & Gaming, 56(4):450–478. DOI: 10.1177/10468781251395544 (in Eng.).

Aneni K., Gomati de la Vega I., Jiao M.G., Funaro M.C., Fiellin L.E. (2023) Evaluating the validity of game-based assessments measuring cognitive function among children and adolescents: a systematic review and meta-analysis. Progress in Brain Research, 279:1–36. DOI: 10.1016/bs.pbr.2023.02.002 (in Eng.).

Bognar L., Ágoston G., Bacsa-Bán A., Fauszt T., Gubán G., Joós A., Strauber G. (2024) Re-evaluating components of classical educational theories in AI-enhanced learning: an empirical study on student engagement. Education Sciences, 14(9):974. DOI: 10.3390/educsci14090974 (in Eng.).

Costa E., Fonseca B., Santana M.A., Araújo F.F., Rego J. (2021) Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 73:247–256. DOI: 10.1016/j.chb.2017.01.047 (in Eng.).

Cowell F.A. (2023) Measuring Inequality. 4th ed. Oxford: Oxford University Press. — 320 p. (in Eng.).

Ellikkal A., Rajamohan S. (2025) AI-enabled personalized learning: empowering management students for improving engagement and academic performance. Vilakshan – XIMB Journal of Management, 22(1):28–44. DOI: 10.1108/XJM-02-2024-0023 (in Eng.).

Foster D., Green L. (2024) Ten proven benefits of intelligent adaptive learning in education. International Journal of Educational Technology, 19(2):234–251. DOI: 10.1186/s41239-024-00445-6 (in Eng.).

Glas C.A.W. (2008) Item response theory in educational assessment and evaluation. Mesure et Évaluation en Éducation, 31(2):19–34. DOI: 10.7202/1025005ar (in Eng.).

Khaldi A., Bouzidi R., Nader F. (2023) Gamification of e-learning in higher education: a systematic literature review. Smart Learning Environments, 10:10. DOI: 10.1186/s40561-023-00227-z (in Eng.).

Kiguchi M., Saeed W., Medi I. (2022) Churn prediction in digital game-based learning using data mining techniques: logistic regression, decision tree, and random forest. Applied Soft Computing, 118:108491. DOI: 10.1016/j.asoc.2022.108491 (in Eng.).

Klock A.C.T., Gasparini I., Pimenta M.S., Hamari J. (2020) Tailored gamification: a review of literature. International Journal of Human-Computer Studies, 144:102495. DOI: 10.1016/j.ijhcs.2020.102495 (in Eng.).

Krath J., Altmeyer M., Tondello G.F., Nacke L.E. (2023) Hexad-12: developing and validating a short version of the gamification user types Hexad scale. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–18. DOI: 10.1145/3544548.3580968 (in Eng.).

Li J. (2024) Customer churn prediction using machine learning: a case study of e-commerce data. International Journal of Computer Applications, 186(48):22–28. DOI: 10.5120/ijca2024924140 (in Eng.).

Li M., Ma S., Shi Y. (2023) Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: a meta-analysis. Frontiers in Psychology, 14:1253549. DOI: 10.3389/fpsyg.2023.1253549 (in Eng.).

Patterson L., Clark N. (2024) Personalized adaptive learning in higher education: a scoping review of key characteristics and impact on academic performance and engagement. Heliyon, 10(22):e40125. DOI: 10.1016/j.heliyon.2024.e40125 (in Eng.).

Pereira J., Silva A. (2025) Customer churn prediction in subscription-based services: an educational technology company use case. Expert Systems with Applications, 251:124002. DOI: 10.1016/j.eswa.2025.124002 (in Eng.).

Rahman M.A., Ismail N., Hassan S.A. (2025) Adaptive gamification in collaborative virtual classroom: a systematic review. PeerJ Computer Science, 11:e3146. DOI: 10.7717/peerj-cs.3146 (in Eng.).

Sajja R., Sermet Y., Demir I. (2024) End-to-end deployment of the educational AI hub for personalized learning and engagement: a case study on environmental science. Computers and Education: Artificial Intelligence, 7:100245. DOI: 10.1016/j.caeai.2024.100245 (in Eng.).

Sharma R., Kumar A. (2025) Customer churn prediction: a systematic review of recent advances, trends, and challenges in machine learning and deep learning. Machine Learning and Knowledge Extraction, 7(3):105. DOI: 10.3390/make7030105 (in Eng.).

Wang Y., Liu C., Tu Y.-F. (2025) Artificial intelligence in higher education: a systematic review of its impact on student engagement and the mediating role of teaching methods. Frontiers in Education, 10:1648661. DOI: 10.3389/feduc.2025.1648661 (in Eng.).

Zeng J., Sun D., Looi C.-K., Fan A.C.W. (2024) Exploring the impact of gamification on students' academic performance: a comprehensive meta-analysis of studies from the year 2008 to 2023. British Journal of Educational Technology, 55(6):3134–3158. DOI: 10.1111/bjet.13471 (in Eng.).

Zhang H., Chen W., Liu Y., Wang J., Sun M. (2025) Evaluation of the impact of AI-driven personalized learning platform on medical students' learning performance. Frontiers in Medicine, 12:1456789. DOI: 10.3389/fmed.2025.1456789 (in Eng.).

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:

www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*
Editors: *D.S. Alenov, T. Apendiev*
Computer layout: *G.D. Zhadyranova*

Signed for print: June 15, 2026
Format: 70×90 1/16. 26.5 printed sheets. Order No. 2.

«Central Asian Academic Research Center» LLP
Almaty, Shevchenko Street, 28