

Научная статья

Original article

УДК 330.43:004.85

doi: 10.55186/2413046X_2025_10_5_141

**ПРОГНОЗИРОВАНИЕ ОТТОКА ПОЛЬЗОВАТЕЛЕЙ
КОМПЬЮТЕРНОЙ ИГРЫ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ
МАШИННОГО ОБУЧЕНИЯ**

**USER CHURN PREDICTION IN COMPUTER GAMES USING MACHINE
LEARNING METHODS**



Лебедева Людмила Николаевна, к.ф.-м.н., доцент кафедры прикладной математики и информатики, Санкт-Петербургский государственный экономический университет, г. Санкт-Петербург, e-mail: l.n.lebedeva@yandex.ru

Корнев Даниил Владимирович, руководитель отдела анализа данных, ООО «Эксбо Север», г. Санкт-Петербург, e-mail: daniil.kornev.spb@gmail.com

Постольник Роман Денисович, Санкт-Петербургский государственный экономический университет, г. Санкт-Петербург, e-mail: rpd1303@gmail.com

Lebedeva Lyudmila Nikolaevna, PhD, Associate Professor of the Department of Applied Mathematics and Informatics, Saint Petersburg State University of Economics, Saint Petersburg, e-mail: l.n.lebedeva@yandex.ru

Kornev Daniil Vladimirovich, Head of Data Analysis Department, Exbo North LLC, Saint Petersburg, e-mail: daniil.kornev.spb@gmail.com

Postolnik Roman Denisovich, Saint Petersburg State University of Economics, Saint Petersburg, e-mail: rpd1303@gmail.com

Аннотация. В статье рассматривается задача прогнозирования оттока пользователей компьютерной игры на основе методов машинного обучения.

Проведён анализ поведения игроков с учётом их уровня опыта, что позволило выделить однородные кластеры аудитории и адаптировать подходы к построению моделей для каждой группы. Для решения задачи использовались современные алгоритмы ансамблевого обучения, включая XGBoost, LightGBM и CatBoost, а также методы балансировки классов и отбора информативных признаков. Особое внимание уделено интерпретации результатов: применён комплекс объяснимых моделей, что позволило выявить ключевые факторы риска и повысить прозрачность принимаемых решений. Разработанные модели прошли апробацию на реальных данных и были интегрированы в бизнес-процессы компании, обеспечив своевременное выявление до 80% склонных к уходу пользователей и способствуя оптимизации стратегий удержания.

Abstract. This article addresses the challenge of predicting user churn in computer games using machine learning methods. The analysis of player behavior, taking into account the level of user experience, allowed for the identification of homogeneous audience clusters and the adaptation of modeling approaches for each group. State-of-the-art ensemble learning algorithms, including XGBoost, LightGBM, and CatBoost, were employed alongside class balancing techniques and feature selection methods. Special emphasis was placed on the interpretability of the results: a suite of explainable models was applied, making it possible to identify key risk factors and enhance the transparency of decision-making. The developed models were tested on real data and integrated into the company's business processes, enabling the timely detection of up to 80% of users prone to churn and contributing to the optimization of retention strategies.

Ключевые слова: прогнозирование оттока пользователей, машинное обучение, анализ поведения игроков, интерпретируемый искусственный интеллект, ансамблевые алгоритмы

Keywords: user churn prediction, machine learning, player behavior analysis, explainable AI, ensemble algorithms

Современный рынок онлайн-игр отличается высокой конкуренцией, где удержание пользователей становится критически важной задачей для развития проектов. Потеря активных игроков негативно сказывается на динамике игрового сообщества и финансовых результатах, что требует внедрения эффективных инструментов для своевременного выявления риска оттока.

В условиях постоянного роста пользовательской базы и усложнения поведенческих сценариев традиционные методы анализа зачастую оказываются недостаточно чувствительными к скрытым закономерностям, определяющим решение игрока покинуть проект.

Методы машинного обучения открывают новые возможности для анализа сложных паттернов поведения и позволяют выявлять предикторы оттока на ранних этапах. Использование современных алгоритмов, способных работать с большими и разнородными массивами данных, делает возможным построение адаптивных моделей, учитывающих индивидуальные особенности различных сегментов аудитории. В данной статье рассматриваются подходы к построению и интерпретации моделей прогнозирования оттока на основе реальных игровых данных, а также анализируются ключевые факторы, влияющие на уход пользователей.

Вопрос прогнозирования оттока пользователей компьютерных игр активно изучается с начала 2010-х годов. Одна из наиболее известных работ – исследование Nadiji F. [3], где были протестированы различные алгоритмы машинного обучения на реальных игровых данных. В данной работе отмечено, что качество прогноза во многом определяется выбором признаков и корректным определением периода оттока.

Очень важен при проведении исследований выбор временных окон для анализа поведения игроков. В работе Lee E., Kim B. [5] предложен метод подбора оптимальных размеров временных промежутков с использованием

метода «локтя», что позволяет повысить точность прогноза и избежать избыточности.

Эффективность ансамблевых алгоритмов в задаче прогнозирования оттока, таких как XGBoost и градиентный бустинг, была подтверждена в исследовании Arik K., Gezer M., Tayali S.T. [1], где эти методы показали высокую устойчивость к разнородности и несбалансированности данных. Развитие методов прогнозирования оттока приведено в [6] и [7].

Влияние социальных факторов на отток подробно рассмотрено в работе Kawale J., Pal A., Srivastava J. [4], где анализ структуры социальных связей позволил повысить точность выявления склонных к уходу пользователей.

Современные подходы также акцентируют внимание на интерпретируемости моделей. Например, Bertens P., Guitart A., Periáñez Á. [2] предложили использовать survival-ансамбли для одновременного прогнозирования оттока и анализа факторов риска, что повышает прозрачность принимаемых решений.

В дальнейшем изложении под оттоком понимается полное прекращение пользователем игровой активности в течение фиксированного временного интервала, следующего за окном наблюдения. Такой подход позволяет чётко разграничить временную неактивность и фактический уход, что критически важно для корректного формирования целевой переменной.

На рисунке 1 представлена полученная авторами зависимость доли пользователей, классифицируемых как ушедшие, от продолжительности выбранного окна отсутствия активности.

Можно заметить, что увеличение длины интервала приводит к росту доли оттока, однако начиная с определённого значения темп прироста существенно замедляется. На необходимость учета этого фактора обращали внимание авторы работы [5].

Анализ темпов прироста позволяет обосновать выбор границы между временной неактивностью и фактическим уходом пользователя.

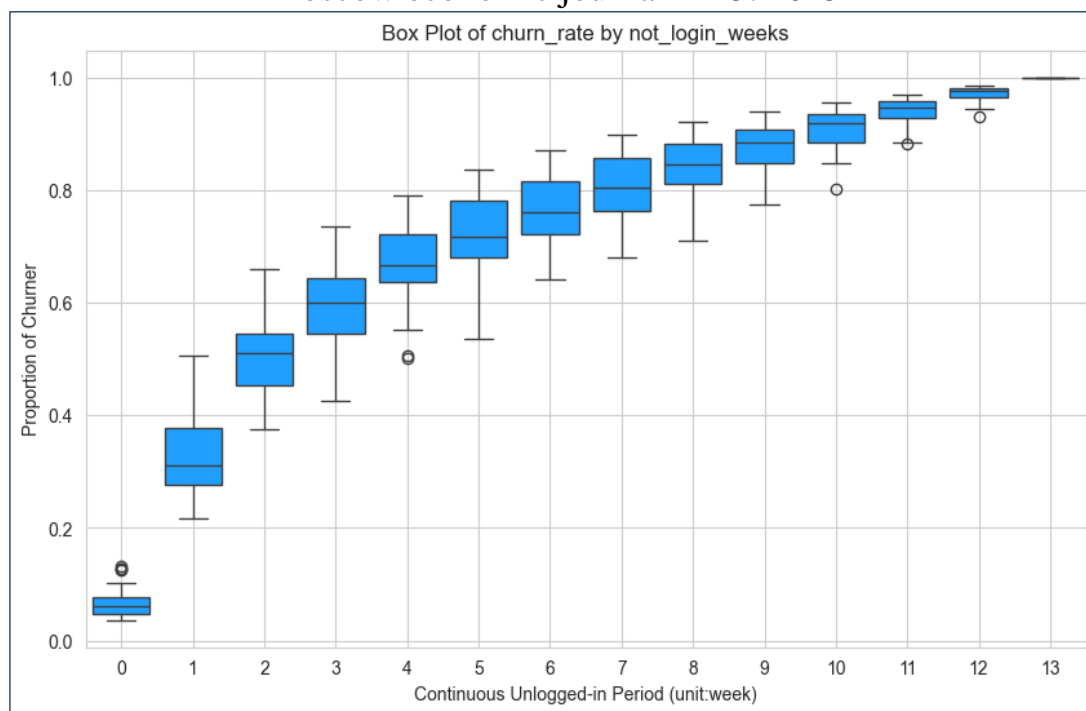


Рисунок 1. Зависимость вероятности ухода игрока от его периода неактивности

В проведенном авторами исследовании в качестве порогового значения выбрано 7 дней. Это позволяет обеспечить баланс между чувствительностью и специфичностью определения оттока.

В качестве исходного набора данных для проведения исследования использовались обезличенные данные об активности пользователей крупного MMORPG проекта. Система сбора данных охватывала широкий спектр характеристик: фиксировались игровые сессии, внутриигровые транзакции, взаимодействия с другими игроками, а также показатели прогресса и вовлечённости. Для каждого пользователя формировался агрегированный профиль, в котором ключевые метрики рассчитывались за ежемесячные интервалы. Такой формат хранения данных обеспечивает возможность их динамического обновления и регулярного переобучения моделей по мере поступления новой информации.

Для учёта неоднородности аудитории была проведена кластеризация пользователей. Для кластеризации использовались такие характеристики, как

суммарная продолжительность игрового времени, а также показатели вовлечённости в социальные и групповые активности.

Применение алгоритма кластеризации позволило выделить три основных сегмента: новичков, игроков со средним уровнем опыта и опытных пользователей.

На рисунке 2 представлено распределение пользователей по кластерам, что наглядно демонстрирует различия в структуре аудитории и обосновывает необходимость индивидуализированного подхода к моделированию оттока для каждой группы.

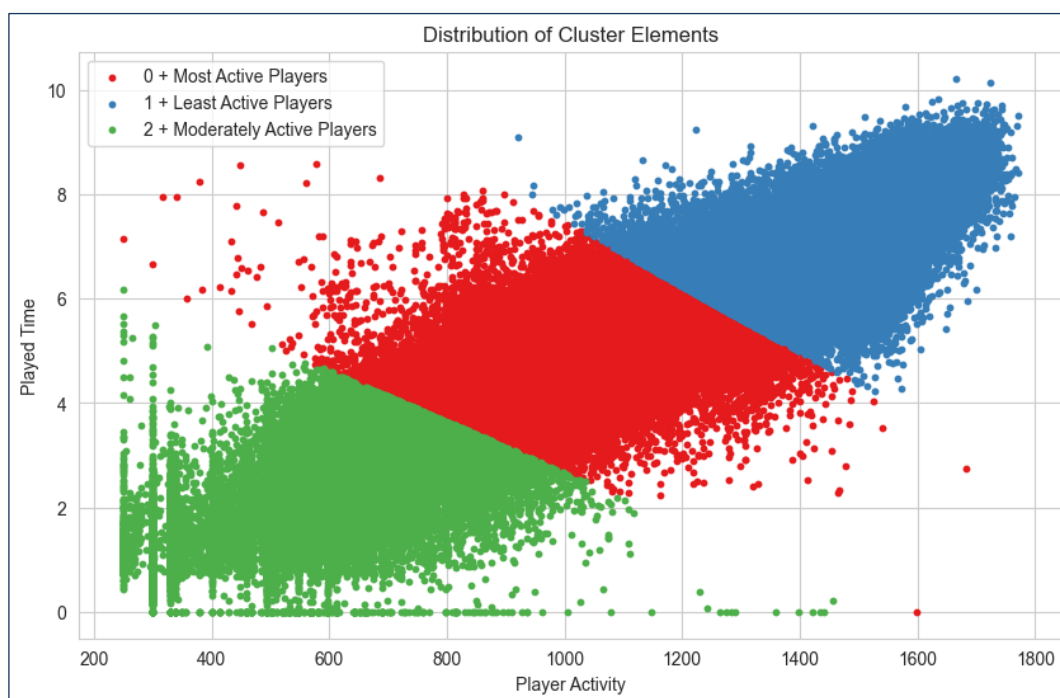


Рисунок 2. Распределение элементов кластеров

Анализ полученных кластеров показывает, что каждый сегмент характеризуется своим типичным уровнем активности и вовлечённости. Это подтверждает необходимость отдельного подхода к моделированию оттока для разных групп игроков.

Для построения моделей прогнозирования оттока использовались как классические алгоритмы, так и современные ансамблевые методы, что

позволило учесть специфику поведения различных пользовательских сегментов.

Обучение осуществлялось отдельно для каждого из выделенных кластеров, что обеспечило адаптацию моделей к особенностям динамики и структуры данных внутри группы.

На первом этапе были протестированы логистическая регрессия и случайный лес, однако лучшие результаты показали алгоритмы градиентного бустинга – XGBoost, LightGBM и CatBoost, благодаря способности выявлять сложные нелинейные зависимости и устойчивости к разнородности данных.

Качество прогнозирования оценивалось по метрикам recall и precision, что позволило объективно сравнить эффективность подходов.

В сегменте новичков XGBoost обеспечил recall 86% и precision 77%, что свидетельствует о высокой полноте и приемлемой точности выявления склонных к уходу пользователей.

Для игроков со средним стажем лучшей оказалась модель LightGBM (recall 70%, precision 40%), что отражает сложность прогнозирования оттока в этой неоднородной группе.

Среди опытных пользователей CatBoost достиг максимального значения recall – 80%, однако precision составил лишь 22%, что указывает на значительное количество ложноположительных срабатываний, характерных для сегмента с высокой вариативностью поведения.

На рисунке 3 представлено сравнение точности рассмотренных моделей для построения прогноза на кластере наименее опытных игроков.

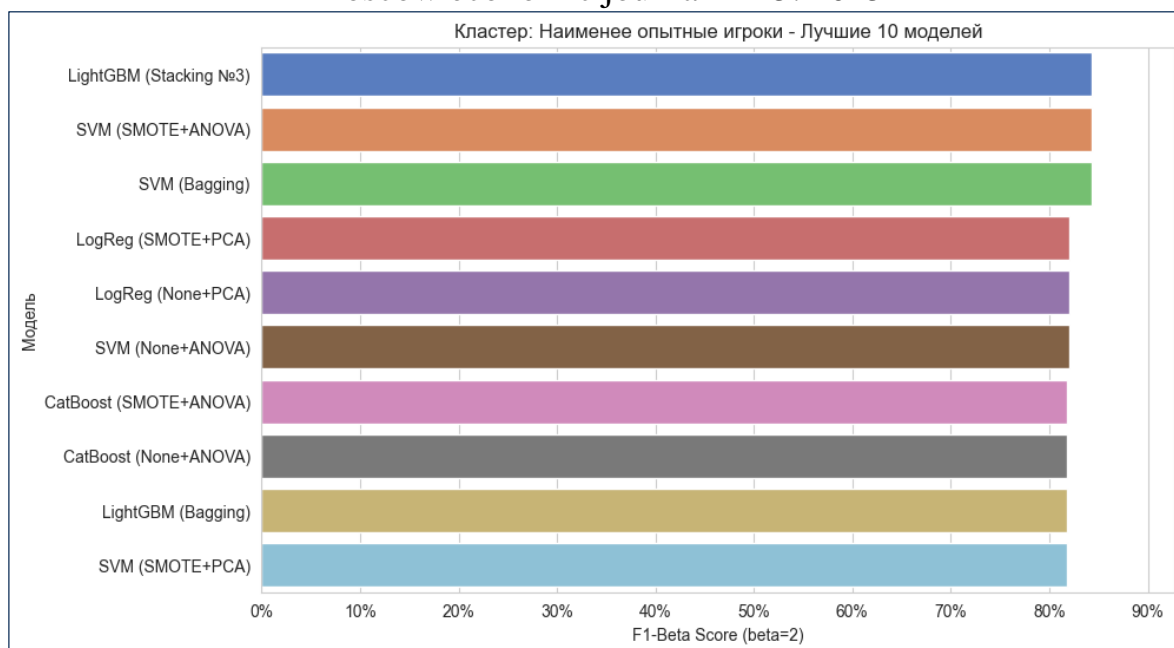


Рисунок 3. Лучшие модели для прогнозирования оттока на кластере наименее опытных игроков

Подбор отдельных моделей для каждого кластера позволил повысить точность прогнозирования и выявить различия в структуре факторов риска, что создаёт основу для дальнейшей персонализации стратегий удержания и оптимизации работы с аудиторией проекта.

Анализ факторов, влияющих на вероятность оттока, выявил существенные различия в структуре значимых признаков для каждого из выделенных кластеров игроков.

Для новичков решающую роль играют показатели ранней вовлечённости: частота входов в игру, количество завершённых обучающих миссий, а также первые внутриигровые достижения. Именно эти параметры с наибольшей вероятностью сигнализируют о формировании устойчивого интереса или, напротив, о риске быстрой потери мотивации. Высокая чувствительность моделей в этом сегменте объясняется относительной однородностью паттернов поведения и небольшим объёмом накопленной истории.

В группе игроков со средним стажем структура факторов риска становится более сложной. Здесь на первый план выходят показатели

регулярности игровой активности, участие в групповых событиях, а также динамика внутриигровых транзакций. Существенное значение приобретают изменения в привычных паттернах: резкое сокращение времени в игре или уменьшение числа взаимодействий с другими пользователями часто предшествуют уходу. Высокая вариативность поведения внутри этого сегмента приводит к снижению точности моделей, что подтверждается сравнительно низкими значениями precision.

Для опытных пользователей ключевым предиктором оттока становится изменение привычной интенсивности участия в высокоуровневых активностях и снижение вовлечённости в социальные взаимодействия. В этой группе модели демонстрируют высокую полноту, однако точность страдает из-за сложности однозначной интерпретации причин ухода. Значительная часть ложноположительных срабатываний связана с временными перерывами, характерными для игроков с длительной историей, что затрудняет дифференциацию между временной неактивностью и реальным уходом.

Результаты проведенного исследования способствуют повышению эффективности управления пользовательской базой и оптимизации маркетинговых кампаний.

Список источников

1. Arik K., Gezer M., Tayali S.T. The study of indicators affecting customer churn in MMORPG games // *Управленец*. – 2022. – Т. 13, № 6. – С. 70–85
2. Bertens P., Guitart A., Periañez Á. Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model // *Proceedings of 2017 IEEE Conference on Computational Intelligence and Games (CIG)*. – New York, NY: IEEE, 2017. – P. 33–36
3. Hadiji F., Sifa R., Drachen A. Predicting player churn in the wild // *2014 IEEE Conference on Computational Intelligence and Games (CIG)*. – Piscataway, NJ: IEEE, 2014. – P. 1–8

4. Kawale J., Pal A., Srivastava J. Churn Prediction in MMORPGs: A Social Influence Based Approach // Proceedings of 2009 International Conference on Computational Science and Engineering (CSE). – Vancouver, BC: IEEE, 2009. – P. 423–428
5. Lee E., Kim B. Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games // IEEE Transactions on Games. – 2020. – Vol. 12, No. 1. – P. 41–53. – DOI: 10.1109/TG.2018.2871215.
6. Lee M., Park S., Lee S. et al. AI-based early detection to prevent user churn in MMORPG // The Korean Journal of Applied Statistics. — 2024. — Vol. 37, No. 4. — P. 525
7. Varun P.N., Yaswanth V., Tapasvi Y.T., Kulkarni V., Padmasree N. Banking Churn Prediction with Random Forest and Discrete Trees in ML // Indiana Journal of Multidisciplinary Research. — 2024. — Vol. 4, No. 3

© Лебедева Л.Н., Корнев Д.В., Постольник Р.Д., 2025. Московский
экономический журнал, 2025, № 5.