



ПРИМЕНЕНИЕ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОГНОЗИРОВАНИЯ ОТТОКА КЛИЕНТОВ: ПРЕИМУЩЕСТВА ДЛЯ БИЗНЕСА

Адилов Рустам Тимурович

Студент компьютерных наук, Университет ИНХА в Ташкенте

E-mail: rstmadylov@gmail.com

ORCID: 0009-0004-7250-4627

<https://doi.org/10.5281/zenodo.21873033>

ARTICLE INFO

Qabul qilindi: 06-avgust 2026 yil

Ma'qullandi: 08-avgust 2026 yil

Nashr qilindi: 10-avgust 2026 yil

KEYWORDS

искусственный интеллект;
машинное обучение; отток
клиентов; удержание
клиентов; Random Forest;
XGBoost; CatBoost; SHAP;
пожизненная ценность
клиента (CLV)..

ABSTRACT

Статья посвящена применению моделей искусственного интеллекта и машинного обучения для прогнозирования оттока клиентов (customer churn) в телекоммуникационной отрасли и смежных сферах обслуживания. Согласно проведённому анализу, телекоммуникационный сектор занимает первое место среди крупных отраслей по ежегодному уровню оттока, который превышает 30%. Различные авторы демонстрируют, что модели Random Forest, XGBoost, CatBoost и их ансамблевые комбинации способны достигать точности (accuracy) от 88% до 99%. В работе разбираются конкретные бизнес-преимущества внедрения подобных моделей: сокращение затрат на удержание клиентов,кратно более высокая - в 3-14 раз - вероятность успешной продажи действующему клиенту по сравнению с новым, а также итоговый рост прибыли компании. Приводится сравнение показателей качества рассмотренных моделей и предлагается обобщённая архитектура системы прогнозирования оттока. Заключительная часть посвящена вопросам интерпретируемости моделей (SHAP), дисбаланса данных и практических сложностей внедрения подобных решений в бизнес-процессы.

Введение

Под оттоком клиентов (customer churn) понимается переход клиента к конкуренту либо полный отказ от услуг компании - и эта проблема остаётся одной из наиболее болезненных для отраслей, работающих по подписной модели. Телекоммуникационный сектор среди всех крупных отраслей отличается самым высоким годовым уровнем оттока свыше 30%, поэтому удержание абонентской базы напрямую определяет финансовую устойчивость операторов связи [1]. Поскольку каждый процент оттока оборачивается прямой потерей выручки, а привлечение новых абонентов обходится компаниям крайне дорого,

бизнес вынужден искать инструменты для заблаговременного выявления клиентов, готовых уйти.

Приоритет удержания клиентов над их привлечением подкреплён устойчивыми эмпирическими данными. Ставшее классическим исследование Bain & Company совместно с Гарвардской школой бизнеса установило, что рост удержания клиентов всего на 5% способен увеличить прибыль компании на 25-95% [2]. Известно также, что шанс продать что-то существующему клиенту составляет 60-70% против лишь 5-20% для нового потенциального покупателя; при этом привлечение нового клиента обходится в 5-25 раз дороже его удержания, а постоянные клиенты в среднем тратят на 31% больше, чем новые [3]. Эти цифры объясняют, почему операторы связи всё чаще рассматривают прогнозирование оттока на основе ИИ не как вспомогательный аналитический инструмент, а как одно из ключевых звеньев стратегии управления доходами.

Ранее анализ оттока строился на описательной статистике и простых правилах. Например, фиксации падения активности абонента ниже заданного порога, - что позволяло выявлять отток лишь постфактум, когда клиент уже фактически ушёл. С развитием методов машинного обучения, от логистической регрессии и деревьев решений до ансамблевых алгоритмов (Random Forest, XGBoost, CatBoost) и объяснимого ИИ (explainable AI), стал возможен проактивный подход: модели заранее выявляют клиентов с повышенным риском ухода, опираясь на историю использования услуг, параметры тарифного плана, платёжную дисциплину и обращения в поддержку, и у компании появляется время для целевого удерживающего воздействия.

Цель данной статьи - систематизировать актуальные подходы к применению моделей искусственного интеллекта в задаче прогнозирования оттока клиентов, сопоставить их эффективность по показателям accuracy, precision, recall, F1 и AUC-ROC, приведённым в опубликованных исследованиях, а также разобрать конкретные бизнес-преимущества, которые подобные модели дают компаниям.

Анализ литературы и методология

Работа представляет собой аналитический обзор: были изучены научные публикации за 2019-2026 гг. по применению моделей машинного обучения для прогнозирования оттока клиентов в телекоммуникационной отрасли, после чего рассмотренные подходы систематизированы по используемым алгоритмам, наборам данных и заявленным показателям качества. Источники подбирались по запросам «customer churn prediction», «telecom churn», «machine learning retention», «explainable churn model», «Random Forest churn», «XGBoost churn». Дополнительно были изучены отраслевые данные об экономическом эффекте удержания клиентов.

Среди первых широко цитируемых работ по данной теме выделяется исследование Ullah et al.: авторы предложили модель прогнозирования оттока на базе Random Forest с отбором признаков методами information gain и корреляционного ранжирования, которая верно классифицировала 88,63% случаев, после чего клиентов дополнительно группировали по мере схожести (cosine similarity) для формирования коллективных удерживающих предложений [4]. Развитие этого направления продемонстрировано в работе Wagh et al.: сочетание случайного леса с техниками балансировки данных

(передискретизация, ENN) позволило поднять точность классификации до 99% при сопоставимых значениях precision и recall [5].

В работе Bhatnagar и Srivastava сравниваются девять алгоритмов машинного обучения - Random Forest, KNN, дерево решений, SVM, логистическая регрессия, бэггинг, XGBoost, стохастический градиентный спуск и наивный байесовский классификатор, - протестированных с применением техник балансировки классов (SMOTE, ENN, под и передискретизация) и кросс-валидации. Лучший результат продемонстрировал именно Random Forest: 90,30% точности на полном наборе признаков и 90,90% на сокращённом, что говорит о возможности упростить модель без потери качества предсказаний [6].

В более поздних исследованиях заметное место занимает вопрос объяснимости моделей (explainable AI). Так, Yuan et al. построили интерпретируемую модель на алгоритме CatBoost с подбором гиперпараметров и получили accuracy 95,54%, precision 90,10%, recall 80,72%, F1-меру 85,03% и AUC-ROC 98,23% на открытом наборе данных иранского оператора связи; стековый ансамбль (CatBoost, Random Forest и LightGBM с логистической регрессией в роли метамодели) показал чуть более высокую полноту обнаружения - recall 82,02%, но уступил по точности (accuracy 94,56%). Для объяснения решений модели авторы использовали метод SHAP (SHapley Additive exPlanations), который выделил длительность пользования услугами, число неудачных вызовов и частоту SMS-сообщений как ключевые факторы, влияющие на отток [7].

Krishna et al. также опирались на классификатор Random Forest, сопоставляя его с SVM, градиентным бустингом, XGBoost и LightGBM, и получили точность 98,1% на реальных данных телекоммуникационной компании [8]. Sharma Poudel, Pokharel и Timilsina сосредоточились на табличных моделях машинного обучения со встроенными и постфактумными методами объяснения (explainer models), что даёт возможность не только прогнозировать отток, но и формировать локальные и глобальные объяснения решений модели для бизнес-пользователей [9]. Логическим продолжением этого направления стал мультимодельный ансамблевый подход El Attar и El-Hajj: семь моделей машинного обучения (в том числе XGBoost, Random Forest и многослойный перцептрон) объединены с методом SMOTE для балансировки классов и SHAP-анализом признаков на наборе данных из 7043 записей [10].

Базовым алгоритмом почти во всех рассмотренных работах служит случайный лес (Random Forest) - ансамблевый метод, объединяющий множество деревьев решений и определяющий итоговый класс большинством голосов, что обеспечивает устойчивость к переобучению даже на сравнительно небольших выборках [11]. Для объяснения решений сложных ансамблевых и градиентных моделей исследователи всё активнее применяют метод SHAP, опирающийся на теоретико-игровой подход к распределению вклада каждого признака в итоговый прогноз [12]. С методологической точки зрения статья опирается на сопоставление показателей accuracy, precision, recall, F1-score и AUC-ROC, приводимых непосредственно в первоисточниках; собственный эксперимент не проводился, поскольку работа носит аналитический характер.

Результаты

Обобщение изученных источников позволяет описать типовую архитектуру ИИ-системы прогнозирования оттока клиентов, состоящую из четырёх последовательных уровней.

1. Уровень сбора данных - охватывает сведения о тарифном плане и контракте, историю платежей и задолженностей, интенсивность пользования услугами (минуты разговоров, интернет-трафик, SMS), число и характер обращений в службу поддержки, а также демографические данные клиента.

2. Уровень предобработки и формирования признаков - на этом этапе устраняется дисбаланс классов (SMOTE, ENN, взвешивание), кодируются категориальные признаки, нормализуются числовые показатели и строятся агрегированные индикаторы поведения, например динамика интенсивности пользования услугами за последние месяцы.

3. Уровень моделирования используются как отдельные алгоритмы (логистическая регрессия, дерево решений, Random Forest, XGBoost, CatBoost), так и ансамблевые либо стековые схемы, объединяющие несколько базовых моделей для повышения устойчивости итогового прогноза.

4. Уровень интерпретации и принятия решения, каждому клиенту присваивается вероятность оттока, а для клиентов с высоким риском строится объяснение решения модели (например, методом SHAP), позволяющее отделу удержания подобрать персонализированное предложение: специальный тариф, скидку, дополнительную услугу либо прямой контакт с клиентом.

Показатели качества моделей прогнозирования оттока, заявленные в проанализированных исследованиях, сведены в таблице 1.

Таблица 1. Сравнение показателей качества моделей прогнозирования оттока клиентов

Модель / подход	Accuracy	Precision	Recall	Источник
Random Forest (с отбором признаков)	88,63%	-	-	[4]
Random Forest + балансировка данных (SMOTE, ENN)	99%	99%	99%	[5]
Random Forest (полный набор признаков)	90,30%	-	-	[6]
Random Forest (сокращённый набор признаков)	90,90%	-	-	[6]
CatBoost (с настройкой гиперпараметров)	95,54%	90,10%	80,72%	[7]
Стековый ансамбль (CatBoost + RF + LightGBM)	94,56%	82,95%	82,02%	[7]
Random Forest (реальные данные оператора)	98,1%	-	-	[8]

Таблица 1 показывает, что независимо от конкретной реализации - Random Forest или градиентные ансамбли (CatBoost, XGBoost, LightGBM) - все рассмотренные модели

демонстрируют высокую точность классификации, от 88,6% до 99%, заметно превосходя простые правило-ориентированные методы и одиночные линейные модели. При этом самая высокая заявленная точность (99%) получена именно за счёт применения техник балансировки классов, что подчёркивает значимость этапа предобработки данных для итогового качества прогноза.

В таблице 2 собраны количественные оценки экономического эффекта от удержания клиентов по данным отраслевых источников, они наглядно показывают, что точное прогнозирование оттока представляет прямую ценность для бизнеса, а не только техническую задачу классификации.

Таблица 2. Экономический эффект удержания клиентов по данным отраслевых исследований

Показатель	Значение	Источник
Ежегодный уровень оттока клиентов в телекоме	> 30% (максимум среди крупных отраслей)	[1]
Рост прибыли при увеличении удержания на 5%	25-95%	[2]
Стоимость привлечения нового клиента vs удержания	в 5-25 раз дороже	[3]
Вероятность продажи существующему клиенту	60-70% (против 5-20% новому)	[3]
Дополнительные траты постоянных клиентов	на 31% больше, чем новые клиенты	[3]
Средняя доля бизнеса, приходящаяся на текущих клиентов	около 65%	[3]

Если сопоставить таблицы 1 и 2, видно, что даже умеренное повышение точности модели прогнозирования оттока скажем, с 88,6% до 95-99% - способно дать компании ощутимый финансовый результат: при годовом оттоке свыше 30% абонентской базы и стоимости привлечения нового клиента, в 5-25 раз превышающей затраты на его удержание, каждый вовремя выявленный «клиент риска» напрямую превращается в сохранённую выручку. Кроме того, поскольку шанс успешной допродажи (up-sell) или кросс-продажи (cross-sell) действующему клиенту в 3-14 раз выше, чем новому, своевременное вмешательство отдела удержания, запущенное моделью, способно не только удержать клиента, но и повысить средний доход на клиента (ARPU).

Обсуждение

Внедрение моделей прогнозирования оттока даёт компаниям сразу несколько взаимосвязанных преимуществ. Прежде всего это экономия на удержании клиентов за счёт таргетирования: вместо того чтобы равномерно распределять удерживающие ресурсы (скидки, бонусы, персональные предложения) на всю базу, компания направляет их только на клиентов с высокой предсказанной вероятностью ухода, что повышает отдачу от программ лояльности. Не менее важно и раннее обнаружение

факторов риска, например, роста числа неудачных вызовов или снижения интенсивности пользования услугами, выявленных методом SHAP [7], которое позволяет не просто удержать конкретного клиента, но и устранить системную причину недовольства, скажем, проблемы качества связи в отдельном регионе.

Точное прогнозирование оттока также напрямую улучшает финансовое планирование: выручка становится более предсказуемой, а инвестиции в привлечение новых клиентов можно корректировать исходя из реального оттока действующей базы, а не усреднённых исторических цифр. Наконец, персонализация удерживающих предложений на основе объяснимых моделей повышает удовлетворённость клиентов, поскольку предложение строится не на статичном сегменте, а на индивидуальном профиле риска и конкретных факторах, выявленных моделью для каждого отдельного клиента.

При этом внедрение подобных моделей не лишено ограничений. Как и в задачах обнаружения мошенничества, здесь сохраняется дисбаланс классов: доля ушедших клиентов обычно заметно меньше доли оставшихся, поэтому уже на этапе предобработки данных требуются техники балансировки (SMOTE, ENN, взвешивание ошибок) [5, 6]. Кроме того, модели, обученные на исторических данных, со временем теряют актуальность из-за изменений рыночной конъюнктуры, тарифной политики или появления новых конкурентов, что требует регулярного переобучения и отслеживания дрейфа данных.

Особого внимания требует вопрос интерпретируемости моделей: сложные ансамблевые архитектуры (CatBoost, стековые ансамбли) дают более высокую точность, но без методов объяснимого ИИ, таких как SHAP, остаются «чёрным ящиком» для бизнес-пользователей и специалистов по работе с клиентами [7, 12]. Применение SHAP-анализа не только повышает доверие бизнеса к модели, но и даёт возможность обосновать перед руководством конкретные удерживающие мероприятия, а это критически важно, если модель должна применяться на практике, а не оставаться исследовательским прототипом.

Заключение

Проведённый обзор подтверждает, что модели искусственного интеллекта и машинного обучения в первую очередь Random Forest, XGBoost, CatBoost и их ансамблевые комбинации обеспечивают точность прогнозирования оттока клиентов от 88,6% до 99% по данным различных исследований, что заметно превосходит возможности традиционных описательных методов. Учитывая, что телекоммуникационная отрасль ежегодно теряет более 30% абонентской базы [1], а рост удержания клиентов всего на 5% способен увеличить прибыль компании на 25-95% [2], практическая значимость подобных моделей для бизнеса выходит далеко за рамки чисто технической задачи классификации.

Среди приоритетных направлений дальнейшего развития можно выделить: более широкое использование методов объяснимого ИИ (SHAP, LIME) для повышения доверия бизнес-пользователей к прогнозам моделей; связывание прогнозирования оттока с расчётом пожизненной ценности клиента (customer lifetime value, CLV), чтобы приоритизировать удерживающие усилия применительно к наиболее ценным клиентам; переход от разовых batch-прогнозов к потоковой обработке данных, близкой

к реальному времени, что позволит реагировать на признаки оттока непосредственно по мере их появления; а также применение методов причинно-следственного вывода (causal inference) и uplift-моделирования, которые оценивают не просто вероятность оттока, а ожидаемый эффект конкретного удерживающего воздействия для каждого клиента.

Список литературы:

1. Chang V., Hall K., Xu Q. A., Amao F. O., Ganatra M. A., Benson V. Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models // Algorithms. 2024. Vol. 17, No. 6. P. 231. DOI: 10.3390/a17060231
2. Reichheld F. F., Sasser W. E. Jr. Zero Defections: Quality Comes to Services // Harvard Business Review. 1990. Vol. 68, No. 5. P. 105-111.
3. Yotpo. Customer Acquisition Vs. Retention: Which Drives Profit? [Электронный ресурс]. Yotpo, 2025. URL: <https://www.yotpo.com/blog/cost-of-customer-acquisition-vs-retention/>
4. Ullah I., Raza B., Malik A. K., Imran M., Islam S. U., Kim S. W. A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector // IEEE Access. 2019. Vol. 7. P. 60134-60149. DOI: 10.1109/access.2019.2914999
5. Wagh S. K., Andhale A. A., Wagh K. S., Pansare J. R., Ambadekar S. P., Gawande S. Customer Churn Prediction in Telecom Sector Using Machine Learning Techniques // Results in Control and Optimization. 2023. Vol. 14. Art. 100342.
6. Bhatnagar A., Srivastava S. Customer Churn Prediction: A Machine Learning Approach with Data Balancing for Telecom Industry // International Journal of Computing. 2025. Vol. 24, No. 1. P. 9-18.
7. Yuan J. et al. Telecom Customer Churn Prediction with Explainable Machine Learning // Proceedings of the 2025 2nd International Conference on Modeling, Natural Language Processing and Machine Learning (CMNM 2025). ACM, 2025. DOI: 10.1145/3757110.3757152
8. Krishna R., Jayanthi D., Shylu Sam D. S., Kavitha K., Maurya N. K., Benil T. Application of Machine Learning Techniques for Churn Prediction in the Telecom Business // Results in Engineering. 2024. Vol. 24. Art. 103165.
9. Sharma Poudel S., Pokharel S., Timilsina M. Explaining Customer Churn Prediction in Telecom Industry Using Tabular Machine Learning Models // Machine Learning with Applications. 2024. Vol. 17. Art. 100567. DOI: 10.1016/j.mlwa.2024.100567
10. El Attar A., El-Hajj M. Explainable AI-driven Customer Churn Prediction: A Multi-model Ensemble Approach with SHAP-based Feature Analysis // Frontiers in Artificial Intelligence. 2026. DOI: 10.3389/frai.2026.1748799
11. Breiman L. Random Forests // Machine Learning. 2001. Vol. 45, No. 1. P. 5-32. DOI: 10.1023/A:1010933404324
12. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4765-4774.