

УДК 004.62

DOI: [10.26102/2310-6018/2024.47.4.037](https://doi.org/10.26102/2310-6018/2024.47.4.037)

Роль идентификации пользователей в прогнозировании совершения целевых действий на сайте

Р.С. Святков

Российский университет дружбы народов, Москва, Российская Федерация

Резюме. Актуальность исследования обусловлена необходимостью повышения точности прогнозирования целевых действий пользователей на веб-сайтах, что является ключевым аспектом в оптимизации маркетинговых стратегий и персонализации пользовательского опыта. Сложность задачи усугубляется отсутствием стабильных идентификаторов, что приводит к разрыву данных и снижению точности предсказаний. Настоящая работа направлена на анализ влияния методов идентификации пользователей и разработку подходов к сегментации, что позволит устранить существующие пробелы в данной области. Ведущим методом исследования является применение алгоритмов машинного обучения, что позволило оценить влияние различных идентификаторов, таких как `client_id` и `user_id`, на точность прогнозирования. Проведена сегментация пользователей на основе метода градиентного бустинга, а также анализ эффективности ретаргетинговых кампаний в системе Яндекс.Директ по показателям конверсии, стоимости привлечения клиентов и доли рекламных расходов на примере клиента, специализирующемся на продаже электронных книг. В результате исследования установлено, что использование идентификатора `user_id` повышает точность предсказания покупок на 8 %, полноту на 6 % и F1-меру на 7 %. Кластеризовав пользователей на несколько сегментов, удалось показать снижение стоимости привлечения клиента на 67 %, уменьшение показателя доли рекламных расходов до 5,87 % по сравнению с автостратегиями Яндекса и увеличение процента конверсии до 34 %. Материалы статьи представляют значимость для специалистов в области электронной коммерции и маркетинга, обеспечивая научную основу для внедрения персонализированных рекламных кампаний. Предложенные методы также могут быть использованы для дальнейшего улучшения аналитики и интеграции данных в мультиканальной среде.

Ключевые слова: машинное обучение, анализ поведения пользователей, идентификация пользователей, сегментация пользователей, электронная коммерция, прогноз целевого действия.

Для цитирования: Святков Р.С. Роль идентификации пользователей в прогнозировании совершения целевых действий на сайте. *Моделирование, оптимизация и информационные технологии*. 2024;12(4). URL: <https://moitvivr.ru/ru/journal/pdf?id=1767> DOI: 10.26102/2310-6018/2024.47.4.037

The role of user identification in predicting target actions on a website

R.S. Svyatov

RUDN University, Moscow, the Russian Federation

Abstract. The relevance of study lies in the need to improve the accuracy of predicting users' target actions on websites, which is a key aspect of optimizing marketing strategies and personalizing user experiences. The complexity of the task is exacerbated by the lack of stable identifiers, leading to data fragmentation and reduced prediction accuracy. This paper aims to analyze the impact of user identification methods and develop approaches to segmentation, which will help to eliminate existing gaps in this area. The primary research method involves applying machine learning algorithms to evaluate the influence of different identifiers, such as `client_id` and `user_id`, on prediction accuracy.

Segmentation of users was carried out based on the gradient boosting method, as well as an analysis of the effectiveness of retargeting campaigns in the Yandex.Direct system based on conversion rates, customer acquisition costs, and the share of advertising expenses using the example of a client specializing in the sale of e-books. The findings reveal that utilizing the user_id identifier improves purchase prediction accuracy by 8 %, recall by 6 %, and the F1-score by 7 %. Segmenting users into targeted groups demonstrated a 67 % reduction in customer acquisition cost, a decrease in advertising expense share to 5.87 % compared to Yandex auto-strategies, and an increase in conversion rate to 34 %. The article's materials are of significance for specialists in the field of e-commerce and marketing, providing a scientific basis for the implementation of personalized advertising campaigns. The proposed methods also offer potential for further enhancement of analytics and data integration in multichannel environments.

Keywords: machine learning, user behavior analysis, user identification, user segmentation, e-commerce, target action prediction.

For citation: Svyatov R.S. The role of user identification in predicting target actions on a website. *Modeling, Optimization and Information Technology*. 2024;12(4). (In Russ.). URL: <https://moitvvt.ru/journal/pdf?id=1767> DOI: 10.26102/2310-6018/2024.47.4.037

Введение

В последние годы предсказание целевых действий пользователей на сайтах стало ключевым направлением исследований в области электронной коммерции, маркетинга и пользовательской аналитики. Эти задачи включают в себя прогнозирование таких действий, как регистрация, подписка на рассылку или совершение покупки.

Особенно важное место занимает предсказание вероятности покупки, поскольку оно позволяет повысить конверсию, оптимизировать маркетинговые стратегии и персонализировать взаимодействие с пользователями. Прогнозирование целевых действий пользователей на веб-сайтах требует точного анализа данных, направленного на выявление закономерностей и поведенческих паттернов каждого пользователя.

Традиционные методы сбора данных, такие как опросы и интервью, не обеспечивают достаточного охвата всей аудитории сайта. Более того, с увеличением объемов и динамики данных в онлайн-среде эти методы становятся менее эффективными. Современные подходы, основанные на анализе больших данных и использовании алгоритмов машинного и глубокого обучения, открывают новые возможности для прогнозирования действий пользователей. Эти технологии позволяют выявлять скрытые взаимосвязи и закономерности, которые сложно определить с помощью традиционных методов анализа.

Точность прогнозирующих моделей зависит от множества факторов, включая качество данных и набор признаков, описывающих пользовательскую активность. Одним из ключевых аспектов является использование данных о сессионной активности, что включает в себя последовательность действий пользователя за определенный временной промежуток. Для построения надежных моделей необходимо идентифицировать каждого пользователя, что позволяет создавать непрерывные цепочки его действий и визитов на сайте. Это, в свою очередь, способствует более глубокому пониманию поведения пользователей.

В работе [1] рассматривается проблема предсказания намерений покупки на раннем этапе с использованием передовых моделей машинного обучения, позволяя выявлять потенциальные покупки во время активной сессии. Это особенно полезно для анонимных пользователей, так как позволяет применять персонализированные предложения и скидки в реальном времени, что улучшает взаимодействие с пользователем.

В статье [2] предлагается развитие данной идеи в виде фреймворка для раннего предсказания покупок как для зарегистрированных, так и для незарегистрированных пользователей, что улучшает коэффициент конверсии. Это исследование выявляет значимость контекстных характеристик, которые могут использоваться до начала взаимодействия с товарами, достигая при этом высокой точности.

Использование технологий глубокого обучения для анализа поведенческих данных также продемонстрировало свою эффективность. Так, исследование [3] сравнивает различные алгоритмы, включая глубокие нейронные сети, для предсказания покупок, где глубокое обучение оказалось превосходящим традиционные методы. Данное исследование подчеркивает, что вовлечение пользователей на платформе может быть стимулировано через точное предсказание их поведения.

Исследование [4] акцентирует внимание на важности выбора признаков и методов ансамблирования для улучшения производительности моделей. Авторы предлагают SE-stacking модель на основе информации и ансамблевого обучения, достигающую высокой точности. Это подчеркивает важность качественного подбора факторов, связанных с покупательским поведением, и использования ансамблевых подходов для оптимизации результатов предсказания.

Другие методы, такие как использование Word2Vec для анализа схожести товаров, были рассмотрены в [5], где выявлено, что добавление метрик сходства между выбранными товарами в сессии может улучшить предсказательную точность модели. Эти подходы оказывают положительное влияние на F1-оценку, что подтверждает значимость комплексного анализа поведения пользователя при взаимодействии с различными товарами.

Исследования [6] и [7] анализируют оптимальные методы машинного обучения и предварительной обработки данных, применяя XGBoost и k-кратную кросс-валидацию для повышения точности предсказания. Применение таких подходов позволяет достичь высокой точности и надежности предсказания.

Для повышения эффективности предсказания некоторые исследования, например, в статье [8], предлагают интеграцию алгоритмов кластеризации для сокращения объема данных и ускорения вычислений. Они также применяют ассоциативные правила для выявления частых наборов товаров, что улучшает точность рекомендаций.

Также значительным вкладом являются исследования, ориентированные на использование данных о визитах пользователей, как в [9], где сравниваются методы классификации на основе вручную создаваемых признаков и глубокого обучения для раннего предсказания намерений. Такой подход демонстрирует надежность предсказаний даже при краткосрочных наблюдениях за поведением пользователей.

Помимо этого, исследование [10] доказывает эффективность применения RNN для анализа последовательных данных действий пользователей, предсказывая поведение в реальном времени. Показано, что такие модели способны эффективно поддерживать маркетинговые кампании, повышая доходность и снижая расходы на таргетирование. Использование ансамблевых моделей, комбинирующих RNN и традиционные классификаторы, представлено как наиболее эффективное решение для прогнозирования покупок.

Несмотря на активное развитие методов прогнозирования поведения пользователей, важной проблемой остается необходимость точной идентификации пользователей на сайте. Это обусловлено разнородностью данных и сложностью построения последовательных цепочек действий для каждого пользователя. Отсутствие стабильного и однозначного идентификатора приводит к разрыву данных, что снижает точность прогнозирующих моделей и эффективность персонализации.

Целью данной работы является исследование влияния идентификации пользователей на точность прогнозирования целевых действий на сайте. Особое внимание уделяется сравнению эффективности применения различных идентификаторов пользователей (`client_id` и `user_id`), а также сегментации пользователей на основе разработанной модели машинного обучения.

Материалы и методы

Для начала разберем структуру предлагаемой системы прогнозирования поведения пользователей. На Рисунке 1 представлена схема прогнозирования поведения пользователей.

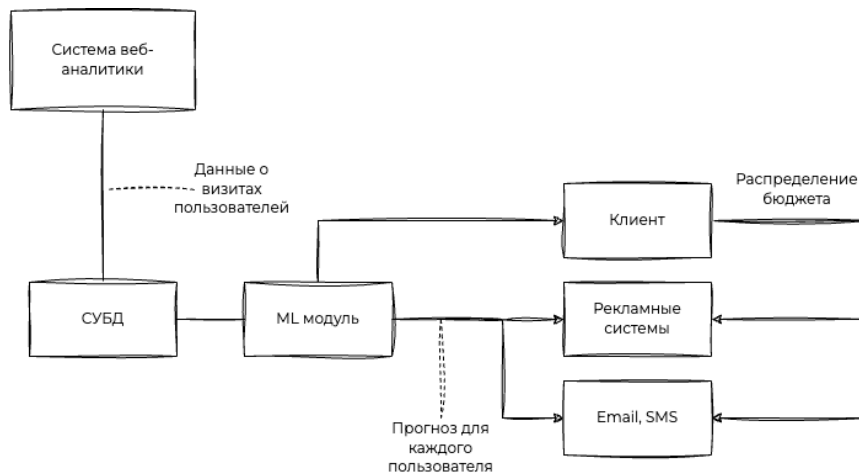


Рисунок 1 – Схема прогнозирования поведения пользователей
 Figure 1 – User behavior prediction scheme

Процесс начинается с получения данных из систем веб-аналитики. На сегодняшний день наиболее популярными системами являются Яндекс.Метрика и Google Analytics. Для анализа поведения пользователей необходимо работать с историей их визитов на сайт и анализировать последовательность действий, совершенных пользователями в процессе посещения. Благодаря интеграции с открытым API этих систем появляется возможность выгружать данные пользователей за заданный период времени.

После успешной выгрузки данных их необходимо загрузить в систему управления базами данных (СУБД) для дальнейшей обработки и формирования нужных данных. Это можно выполнить с использованием кода на языке Python. В качестве СУБД может быть использована любая реляционная база данных. Через SQL-запросы появляется возможность работать с данными и формировать набор признаков, который будет использован для обучения модели машинного обучения. Сформированный набор данных передается в ML-модуль, где происходит обучение модели. В зависимости от задачи, может быть выбрана любая подходящая модель машинного обучения. В рассматриваемой задаче используется бинарная классификация, где целевым признаком является факт совершения определенного действия пользователем в будущем в течение установленного периода времени. После обучения модели данные по каждому пользователю преобразуются в вероятность совершения целевого действия. Полученные данные затем сегментируются по степени готовности пользователя совершить целевое действие. Обычно пользователи делятся на пять сегментов, от низкоконтверсионных до высококонтверсионных.

Выходной результат системы на данном этапе представляет собой таблицу, содержащую следующую информацию:

- Client_id или user_id пользователя;
- вероятность совершения целевого действия;
- назначенный сегмент в зависимости от готовности совершить целевое действие;
- номер телефона (при наличии);
- электронная почта (при наличии).

Для получения контактных данных пользователей необходимо настроить интеграцию с CRM-системой клиента, где хранится связь между client_id или user_id и контактной информацией.

Выделены несколько способов применения полученной таблицы с данными о каждом пользователе:

1. Передача таблицы клиенту. Эти данные представляют большую ценность для бизнеса и клиента, так как позволяют принимать решения о дальнейшем использовании информации. Например, клиент может внедрить на сайте модуль, который будет адаптировать контент в зависимости от вероятности совершения целевого действия пользователем, показывая более дорогие товары или предлагая специальные скидки. Клиент может также самостоятельно управлять рекламным бюджетом, используя полученные сегменты в рекламных системах, а также для рассылок по Email и SMS.

2. Рекламные системы. Существует множество рекламных платформ, которые позволяют создавать рекламные кампании с таргетингом на конкретных пользователей. Например, в Яндекс.Директ возможно настроить рекламную кампанию на основе контактных данных или client_id из Яндекс.Метрики. Перед показом рекламы пользователю проводится аукцион. Аукцион является закрытым, то есть все участники делают ставки вне зависимости друг от друга, не зная ставок конкурентов. Объект аукциона достается участнику, предложившему наиболее высокую цену [11]. Для высококонверсионных сегментов можно увеличить ставку, чтобы гарантированно показать рекламу пользователю, так как вероятность его покупки высока. Для низкоконверсионных сегментов ставки могут быть снижены, чтобы исключить неэффективных пользователей из рекламной кампании.

3. Email и SMS коммуникации. Этот способ использования сегментов направлен на взаимодействие с пользователями через электронную почту или SMS. Сообщения могут содержать предложения по скидкам или акционным товарам, которые пользуются повышенным спросом.

Эту схему можно усовершенствовать и автоматизировать с помощью программного обеспечения Apache Airflow, которое является оркестратором процессов на основе Python. Основной сущностью Apache Airflow является DAG (Directed Acyclic Graph), где вершинами графа являются конкретные задачи. DAG может работать по заранее установленному расписанию. Автоматизацию можно настроить для передачи таблицы с прогнозами по каждому пользователю через API или для загрузки в СУБД. Одним из наиболее распространенных сценариев автоматизации является обновление сегментов в Яндекс.Директ, которое выполняется ночью, чтобы сегменты успели обработаться на стороне сервиса Яндекс.Аудитории. При наличии нескольких клиентов систему можно масштабировать с помощью программного обеспечения Kubernetes, что позволяет эффективно управлять ресурсами и обеспечивать автомасштабирование.

Все прогнозы и построение цепочек действий пользователей строятся на основе идентификаторов пользователей. Без корректной идентификации пользователь не сможет быть идентифицирован и его дальнейший анализ станет некорректным.

Идентификация пользователей является ключевым элементом в анализе их активности на сайте. Этот процесс заключается в сопоставлении действий конкретного пользователя с уникальным идентификатором, что позволяет отслеживать поведение на протяжении времени и через различные каналы взаимодействия.

На практике наиболее часто используются анонимные идентификаторы, которые присваиваются системами веб-аналитики, такими как Google Analytics или Яндекс.Метрика. Например, в Яндекс.Метрике каждому уникальному посетителю сайта автоматически назначается `client_id`, который формируется на основе случайного числа и даты первого посещения в формате Unix и сохраняется в cookies браузера. Однако у `client_id` есть недостатки: если пользователь посещает сайт с разных устройств или браузеров, ему присваиваются разные идентификаторы, что затрудняет анализ цепочек его действий. Кроме того, часть пользователей может отключать сохранение cookies, что также негативно влияет на точность идентификации.

Для устранения этих проблем возможно использование `user_id`, который назначается пользователю, например, во время регистрации или авторизации на сайте. Интеграция `user_id` требует доработки системы, чтобы идентификатор генерировался на основе данных из CRM или CMS сайта. Это позволяет связать различные идентификаторы одного пользователя даже при посещении с разных устройств или браузеров, если он авторизован.

Результаты

Для подтверждения значимости идентификации пользователей был проведен эксперимент, направленный на сравнение точности модели при использовании двух различных типов идентификаторов: `client_id` и `user_id`. В качестве исходных данных использовалась сессионная активность пользователей крупного российского онлайн-магазина, специализирующегося на продаже электронных книг.

Данные в виде визитов пользователей были получены при помощи интеграции с API Яндекс Метрика и загружены в СУБД Yandex ClickHouse при помощи собственного кода на Python. Набор данных включал в себя 15,5 миллионов строк, охватывающих четырехмесячный период. После предварительной обработки данные были преобразованы при помощи SQL запросов в формат, содержащий 34 признака, описывающих поведение каждого пользователя на сайте за последний месяц. Среди таких признаков были средняя длительность визита, количество часов с момента предыдущего визита, среднее число просмотренных страниц, а также метрики, связанные с покупками: общее число покупок, их суммарная стоимость, время, прошедшее с последней покупки, и другие.

Стоит отметить, что данные аналитических систем вбирают в себя куда больший спектр полезных и ценных данных. Чем больше признаков получится сформировать, тем проще будет определять паттерны поведения пользователей, ведущих к конверсии на сайте [11]. Несмотря на это, обилие всевозможных признаков не гарантирует высокую предсказательную способность модели.

Целевой переменной стала метка, отражающая факт совершения покупки пользователем в течение 14 дней после конкретного визита. Это позволило сформулировать задачу как проблему бинарной классификации, в которой модель предсказывала вероятность совершения покупки на основе предоставленных данных.

Обучение модели проводилось с использованием алгоритма градиентного бустинга, реализованного в библиотеке LightGBM на Python. Настройка гиперпараметров не производилась, использовались значения по умолчанию. Модель обучалась в течение 10,000 эпох с реализацией механизма преждевременной остановки,

установленного на 1,000 раундов. Для обучения модели было выделено 80 % исходных данных, тогда как оставшиеся 20 % были разделены поровну на тестовую и валидационную выборки.

Для формирования идентификаторов `user_id` на сайте разработчики внедрили процесс передачи данного идентификатора в момент авторизации или регистрации пользователя. Идентификатор был доступен из CRM клиента, после чего отправлялся на уровень визита конкретного пользователя на сайте. Этот идентификатор стал доступен при выгрузке данных из API Яндекс.Метрика.

Дополнительно, пользователи, идентифицированные с помощью `user_id`, были сегментированы на пять групп по степени вероятности совершения покупки в течение ближайших 14 дней. Эта сегментация использовалась для настройки рекламных кампаний в системе Яндекс.Директ с нулевыми корректировками ставок. Целью данного этапа была проверка точности модели и ее способности корректно прогнозировать поведение пользователей. Сегментация была выполнена с применением метода k-ближайших соседей, что позволило эффективно разделить пользователей на группы на основе схожести их поведения.

В Таблице 1 приведены оценки качества обученных моделей для `client_id` и `user_id` на валидационном наборе данных. В качестве метрик качества выступали Precision, Recall и F1-Score.

Таблица 1 – Метрики качества модели на валидационном наборе данных

Table 1 – Model quality metrics on the validation dataset

Тип	Класс	Точность	Полнота	F1-мера
client_id	0	0,967	0,960	0,964
client_id	1	0,486	0,584	0,535
user_id	0	0,966	0,953	0,959
user_id	1	0,564	0,644	0,602

Для класса 0 показатели модели практически идентичны при использовании обоих идентификаторов. Точность для `client_id` составила 0,967, а для `user_id` – 0,966. Полнота также близка: 0,960 для `client_id` и 0,953 для `user_id`. F1-мера, которая объединяет эти два показателя, составила 0,964 для `client_id` и 0,959 для `user_id`. Эти результаты демонстрируют стабильность модели в прогнозировании отсутствия покупки, независимо от типа идентификатора, что связано с высокой представленностью этого класса в данных.

Исходя из результатов для класса 1, можно сделать вывод, что использование идентификатора `user_id` приводит к заметному увеличению всех метрик. Точность увеличилась с 48 % до 56 %, полнота увеличилась с 58 % до 64 % и F1-мера увеличилась с 53 % до 60 %.

Сравнительный анализ показал, что идентификация пользователей через `user_id` позволяет достичь большей точности определения готовности совершения покупки в течение 14 дней.

Дополнительно на протяжении одного месяца сегментированные пользователи присутствовали в качестве ретаргетинг сегментов в нескольких рекламных кампаниях в Яндекс.Директ с нулевыми корректировками. Всего было 5 сегментов:

- пользователи с очень высокой вероятностью покупки;
- пользователи с высокой вероятностью покупки;
- пользователи со средней вероятностью покупки;
- пользователи с низкой вероятностью покупки;

– пользователи с очень низкой вероятностью покупки.

Сегменты обновлялись каждый день рано утром при помощи программного обеспечения Apache Airflow. Эти сегменты сравнивались по нескольким маркетинговым метрикам с автостратегиями Яндекса. В метрики вошли Conversion Rate, Cost Per Order и доля рекламных расходов (ДРР). На Рисунках 1, 2 и 3 изображено сравнение созданных сегментов и автостратегий Яндекса по CR, CPO и ДРР.

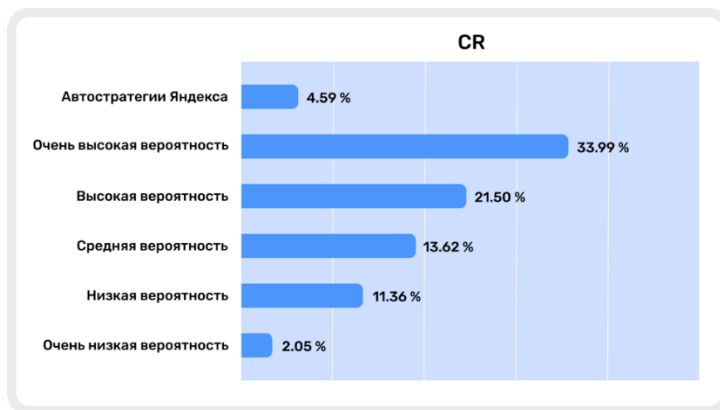


Рисунок 2 – Сравнение показателя CR в разрезе ML сегментов и автостратегий Яндекса
 Figure 2 – Comparison of the CR indicator in the context of ML segments and Yandex auto strategies

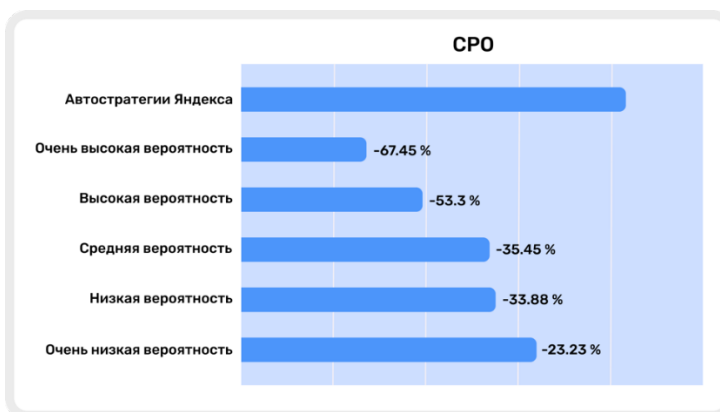


Рисунок 3 – Сравнение показателя CPO в разрезе ML сегментов и автостратегий Яндекса
 Figure 3 – Comparison of CPO indicator in terms of ML segments and Yandex auto strategies

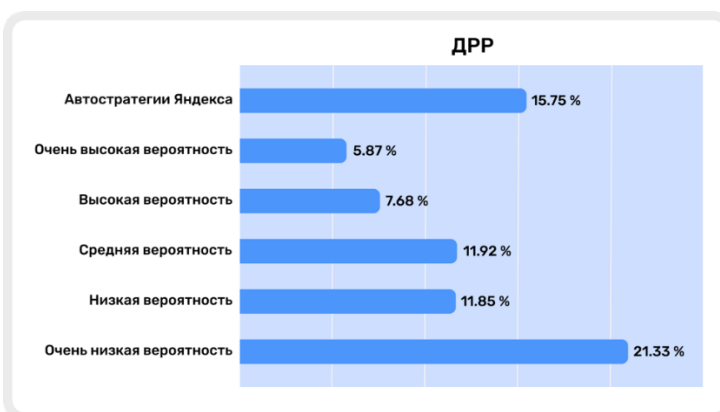


Рисунок 4 – Сравнение показателя ДРР в разрезе ML сегментов и автостратегий Яндекса
 Figure 4 – Comparison of the Advertising Cost Ratio indicator in the context of ML segments and Yandex auto strategies

Результаты проведенного эксперимента демонстрируют явное преимущество сегментированного подхода на основе ML сегментов в сравнении с использованием стандартных автостратегий Яндекса. Показатель CR наглядно подтвердил точность и качество прогнозирования поведения пользователей. Так, самый высококонверсионный сегмент обеспечил максимальное значение CR, равное 34 %. По мере снижения конверсионности сегментов наблюдается закономерное уменьшение CR.

При анализе стоимости привлечения клиента было выявлено значительное снижение затрат в сегментах, сформированных на основе ML, по сравнению с автостратегиями. В наиболее конверсионном сегменте затраты на привлечение клиента оказались ниже на 67 %, что подчеркивает экономическую эффективность применения модели для оптимизации маркетинговых бюджетов.

Показатель DPP также указал на эффективность ML сегментов. Для автостратегий Яндекса он составил 15,75 %, тогда как наиболее конверсионные сегменты продемонстрировали гораздо более низкие значения, достигая минимального уровня 5,87 % в сегменте с высокой вероятностью целевого действия. Напротив, самый низкоконверсионный сегмент показал максимальное значение DPP, что свидетельствует о низкой рентабельности вложений в данный сегмент. Этот результат подчеркивает целесообразность применения понижающих корректировок для таких сегментов с целью исключения их из рекламных кампаний и минимизации расходов на малоперспективную аудиторию.

Заключение

Исследование подтвердило, что использование идентификатора `user_id` и сегментации пользователей на основе модели машинного обучения значительно улучшает точность прогнозирования и эффективность рекламных кампаний. Для класса 0 (отсутствие покупки) модель показала стабильные результаты вне зависимости от идентификатора, однако для класса 1 (совершение покупки) применение `user_id` обеспечило увеличение точности на 8 %, полноты на 6 % и F1-меры на 7 %, что демонстрирует преимущества этого подхода. В ретаргетинговых кампаниях сегментация на основе ML позволила наиболее конверсионным сегментам достигнуть максимального значения CR в 34 % с последовательным уменьшением показателя для менее конверсионных групп. Стоимость привлечения клиента (CPO) в наиболее эффективных сегментах снизилась на 67 % по сравнению с автостратегиями, а показатель DPP оказался минимальным, равным 5,87 % против 15,75 % у автостратегий Яндекса, что подчеркивает экономическую эффективность такого подхода.

Полученные результаты имеют значимую практическую ценность, поскольку позволяют бизнесу оптимизировать распределение маркетинговых бюджетов. Высококонверсионные сегменты способствуют увеличению доходов, тогда как отказ от инвестиций в низкоконверсионные группы позволяет минимизировать затраты. Для дальнейшего развития методов предлагается интеграция дополнительных поведенческих признаков, использование более сложных алгоритмов, таких как глубокие нейронные сети, а также то, как данное решение можно автоматизировать под всех клиентов с электронной торговлей.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Esmeli R., Bader-El-Den M., Abdullahi H. Towards early purchase intention prediction in online session based retailing systems. *Electronic Markets*. 2021;31:697–715. <https://doi.org/10.1007/s12525-020-00448-x>

2. Esmeli R., Bader-El-Den M., Abdullahi H. An analyses of the effect of using contextual and loyalty features on early purchase prediction of shoppers in e-commerce domain. *Journal of Business Research*. 2022;147:420–434. <https://doi.org/10.1016/j.jbusres.2022.04.012>
3. Chaudhuri N., Gupta G., Vamsi V., Bose I. On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*. 2021;149. <https://doi.org/10.1016/j.dss.2021.113622>
4. Xu J., Wang J., Tian Y., Yan J., Li X., Gao X. SE-stacking: Improving user purchase behavior prediction by information fusion and ensemble learning. *PLoS ONE*. 2020;15(11). <https://doi.org/10.1371/journal.pone.0242629>
5. Esmeli R., Bader-El-Den M., Abdullahi H. Using Word2Vec Recommendation for Improved Purchase Prediction. In: *2020 International Joint Conference on Neural Networks (IJCNN), 19–24 July 2020, Glasgow, UK*. IEEE; 2020. pp.1–8. <https://doi.org/10.1109/IJCNN48605.2020.9206871>
6. Lee J., Jung O., Lee Y., Kim O., Park C. A Comparison and Interpretation of Machine Learning Algorithm for the Prediction of Online Purchase Conversion. *Journal of Theoretical and Applied Electronic Commerce Research*. 2021;16(5):1472–1491. <https://doi.org/10.3390/jtaer16050083>
7. Song P., Liu Y. An XGBoost Algorithm for Predicting Purchasing Behaviour on E-Commerce Platforms. *Technical gazette*. 2020;27(5):1467–1471. <https://doi.org/10.17559/tv-20200808113807>
8. Abu Alghanam O., Al-Khatib S.N., Hiari M.O. Data Mining Model for Predicting Customer Purchase Behavior in E-Commerce Context. *International Journal of Advanced Computer Science and Applications*. 2022;13(2):421–428. <https://doi.org/10.14569/IJACSA.2022.0130249>
9. Requena B., Cassani G., Tagliabue J., Greco C., Lacasa L. Shopper intent prediction from clickstream e-commerce data with minimal browsing information. *Scientific Reports*. 2020;10. <https://doi.org/10.1038/s41598-020-73622-y>
10. Koehn D., Lessmann S., Schaal M. Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*. 2020;150. <https://doi.org/10.1016/j.eswa.2020.113342>
11. Бабенкова Ю.А. Использование теории аукционов при построении оптимальной стратегии рекламодателя на Яндекс.Директ. *Труды Московского физико-технического института*. 2010;2(1):67–72.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Святлов Роман Сергеевич, аспирант, **Roman S. Svyatov**, postgraduate student, Российский университет дружбы народов, RUDN University, Moscow, the Russian Москва, Российская Федерация. Federation.
e-mail: romasvyatov@yandex.ru
ORCID: [0009-0009-0322-1443](https://orcid.org/0009-0009-0322-1443)

Статья поступила в редакцию 08.12.2024; одобрена после рецензирования 20.12.2024; принята к публикации 24.12.2024.

The article was submitted 08.12.2024; approved after reviewing 20.12.2024; accepted for publication 24.12.2024.