

УДК 004.62

DOI: [10.26102/2310-6018/2025.51.4.064](https://doi.org/10.26102/2310-6018/2025.51.4.064)

Прогнозирование покупательского поведения пользователей интернет-магазинов на основе событийных данных

Р.С. Святов

Российский университет дружбы народов, Москва, Российская Федерация

Резюме. Актуальность исследования определяется стремительным развитием рынка электронной коммерции, где все больше покупок происходит онлайн. Это обуславливает необходимость прогнозирования покупательского поведения для повышения эффективности маркетинговых стратегий. Проблема заключается в ограниченной применимости существующих подходов. Они, как правило, основаны на открытых наборах данных, не отражающих специфику реальных пользовательских сценариев. В связи с этим работа направлена на разработку подхода к прогнозированию покупательского поведения на основе событийных данных, формируемых в системах веб-аналитики. Основным методом исследования выступает экспериментальное моделирование с применением алгоритмов машинного обучения. Они реализованы в вычислительном стенде, включающем интеграцию с API Яндекс.Метрики и использование градиентного бустинга. Проведены эксперименты на данных шести интернет-магазинов, различающихся профилем деятельности и объемами пользовательской активности. В результате установлено, что использование событийных данных и их производных существенно повышает качество прогнозирования: значения F-меры, Precision, Recall и AUC-ROC увеличиваются на 10–20 процентных пунктов по сравнению с базовыми признаками. Таким образом, предложенный подход позволяет формировать интерпретируемые и масштабируемые модели прогнозирования покупательского поведения, применимые к интернет-магазинам различного уровня. Полученные результаты обладают практической ценностью для специалистов в области аналитики электронной коммерции и проектирования систем персонализации.

Ключевые слова: поведение пользователей, интернет-магазин, машинное обучение, анализ поведения пользователей, идентификация пользователей, электронная коммерция, прогноз покупки, событийные данные.

Для цитирования: Святов Р.С. Прогнозирование покупательского поведения пользователей интернет-магазинов на основе событийных данных. *Моделирование, оптимизация и информационные технологии*. 2025;13(4). URL: <https://moitvivt.ru/ru/journal/pdf?id=2129> DOI: 10.26102/2310-6018/2025.51.4.064

Forecasting e-commerce user purchase behavior based on event data

R.S. Svyatov

RUDN University, Moscow, the Russian Federation

Abstract. The relevance of this study is determined by the rapid growth of the e-commerce market, in which the share of online purchases continues to increase. This trend highlights the need to predict consumer behavior to enhance the effectiveness of marketing strategies. The problem lies in the limited applicability of existing approaches, which are mainly based on open datasets that do not reflect the specific features of real user scenarios. Therefore, this research aims to develop an approach for predicting consumer behavior based on event data collected from web analytics systems. The primary research method is experimental modeling using machine learning algorithms. The computational framework integrates with the Yandex.Metrica API and employs gradient boosting. Experiments were conducted on data from six online stores with different profiles and levels of user activity. The results

demonstrate that the use of event data and their derived features significantly improves prediction quality: F-measure, Precision, Recall, and AUC-ROC values increase by 10–20 percentage points compared to baseline features. Thus, the proposed approach enables the creation of interpretable and scalable models for predicting consumer behavior, applicable to online stores of different sizes. The findings have practical value for professionals in e-commerce analytics and the development of personalization systems.

Keywords: user behavior, online store, machine learning, user behavior analysis, user identification, e-commerce, purchase prediction, user events.

For citation: Svyatov R.S. Forecasting e-commerce user purchase behavior based on event data. *Modeling, Optimization and Information Technology*. 2025;13(4). (In Russ.). URL: <https://moitvivt.ru/ru/journal/pdf?id=2129> DOI: 10.26102/2310-6018/2025.51.4.064

Введение

За последние десятилетия электронная коммерция демонстрирует устойчивый рост, вследствие чего интернет-магазины стали основным каналом взаимодействия компаний с клиентами. В этих условиях возможность прогнозирования покупательского поведения приобретает стратегическое значение, поскольку позволяет оптимизировать маркетинговые расходы, повысить персонализацию предложений и увеличить конверсию продаж.

Актуальность задачи подтверждается и практическими соображениями. Разработка собственной системы прогнозирования требует значительных временных и технологических ресурсов, а также компетенций в области обработки больших данных и машинного обучения, что не всегда доступно компаниям малого и среднего бизнеса.

Вопрос прогнозирования покупательского поведения подробно рассматривается как отечественными, так и зарубежными исследователями. Основное внимание уделяется выбору алгоритмов машинного обучения, включая методы глубокого обучения, и сравнению их точности для различных постановок задачи. Однако не менее важным аспектом является формирование признакового пространства, отражающего реальные паттерны поведения пользователей. Именно качество данных и корректность их интерпретации определяют возможности моделей при анализе пользовательских действий.

Источники данных для подобных исследований могут быть разделены на несколько групп – открытые наборы данных, системные логи, встроенные модули сбора данных и системы веб-аналитики. Открытые наборы данных, как правило, представляют собой искусственно подготовленные выборки или ограниченные данные конкретного интернет-магазина, публикуемые в исследовательских целях. Они обеспечивают воспроизводимость экспериментов, однако отражают лишь усредненные сценарии и редко учитывают специфику конкретного бизнеса [1, 2].

Системные логи и встроенные модули сбора данных позволяют фиксировать события на уровне конкретного веб-сайта, однако требуют разработки и поддержки собственных решений по сбору и хранению данных [3, 4]. Такой подход используется ограниченно и в основном характерен для крупных интернет-магазинов, располагающих собственными командами аналитиков.

Наиболее распространенными инструментами для отслеживания активности пользователей являются системы веб-аналитики. Google Analytics активно применяется интернет-магазинами для получения сведений о посещаемости и поведенческих характеристиках аудитории [5, 6]. Сопоставимый функционал реализован в Яндекс.Метрике, обеспечивающей детализированный сбор событийных данных и описание структуры пользовательских взаимодействий [7, 8]. Оба инструмента

предоставляют не только базовые метрики визитов, но и информацию о последовательности действий пользователей, что расширяет возможности анализа.

Следует отметить, что значительная часть исследований в области прогнозирования покупательского поведения опирается преимущественно на открытые наборы данных. Это ограничивает их практическую ценность и презентабельность. Между тем в реальной практике существует множество интернет-магазинов, работающих в различных бизнес-сферах и обладающих собственными, зачастую уникальными данными о пользователях. Эти различия проявляются не только в ассортименте товаров, но и во временных горизонтах принятия решений о покупке, уровне вовлеченности пользователей, а также в структуре воронки продаж.

В то время как большинство существующих работ фокусируется на применении алгоритмов машинного обучения к стандартным наборам признаков сессионного уровня, остается недостаточно исследован потенциал событийных данных, позволяющих учитывать специфику поведения пользователей конкретного интернет-магазина. Использование подобных данных открывает возможности для построения более точных и интерпретируемых моделей прогнозирования.

Цель настоящего исследования – разработка и экспериментальная проверка подхода к прогнозированию покупательского поведения пользователей интернет-магазинов на основе событийных данных. Для достижения цели формируются различные конфигурации признаков, включающие стандартные характеристики визитов и событийные метрики пользователей. Это позволяет оценить вклад каждого типа данных в итоговое качество прогноза.

Рассматривая существующие исследования, можно отметить, что в большинстве работ вопросам использования событийных данных уделено недостаточно внимания. Как отмечалось ранее, значительная часть публикаций ориентирована на анализ открытых наборов данных, предоставляемых компаниями или организациями, организующими соревнования по машинному обучению. Как правило, в таких наборах содержатся транзакционные данные, сведения о товарах, посещаемости веб-сайтов, а также дополнительная информация о клиенте – возраст, пол, уровень дохода и другие характеристики.

В проведенных исследованиях применяются разнообразные алгоритмы и их комбинации с целью получения наиболее точных результатов. Однако полученные показатели, как правило, представляют собой единичные значения метрик точности, рассчитанные для конкретного набора данных. Каждое последующее исследование стремится лишь незначительно повысить точность, используя те же открытые источники. Подобный подход не позволяет в полной мере оценить практическую применимость и устойчивость предложенных решений в реальной среде.

Ограниченная доступность данных, приближенных к реальным, остается одной из ключевых проблем в данной области. Этот недостаток существенно снижает возможность адаптации предложенных моделей к условиям конкретных интернет-магазинов. В предлагаемом исследовании данный пробел устраняется: разработанный алгоритм и набор признаков могут быть применены к различным типам интернет-магазинов, что будет продемонстрировано на примере нескольких функционирующих площадок.

Прежде чем представить предлагаемое решение, рассмотрим ряд существующих исследований, посвященных прогнозированию покупательского поведения пользователей, и выделим их ключевые особенности и ограничения. В Таблице 1 приведен обзор таких работ, включающий тип исходных данных, достигнутые показатели точности и использованные алгоритмы.

Таблица 1 – Результаты исследований и используемые алгоритмы прогнозирования
Table 1 – Research results and applied forecasting algorithms

Статья	Тип исходных данных	Результаты точности	Используемые алгоритмы
[9]	Открытый набор данных	Accuracy $\approx 0,91$, F $\approx 0,90$	RF, GBDT, LR, DT
[10]	Открытый набор данных	Accuracy 93,2 %	RNN + Naive Bayes (RNN-NB)
[11]	Открытый набор данных	Accuracy 97,6 % F = 0,9763	XGBoost
[12]	Открытый набор данных	Accuracy = 98 %	TF-IDF + Logistic Regression
[13]	Открытый набор данных	AUC 0,6681	LR, KNN, RF, XGBoost
[14]	Открытый набор данных	Accuracy 72–75 %	RF, GBDT, XGBoost, RNN
[15]	Данные веб-аналитики	–	RNN
[16]	Поведенческие логи	Accuracy $\approx 0,9$	Naive Bayes
[17]	Открытый набор данных	F = 0,83	DF
[18]	Открытый набор данных	Accuracy $\sim 95,2$ %	C4.5, J48, CS-MC4, MLR, Apriori
[19]	Открытый набор данных	AUC = 0,85	GNN + Time Attention + Meta-Learning
[7]	Транзакционные данные	F = 0,65 (LGBM) RMSE = 3947,9	LR, KNN, Decision Tree, LGBM
[8]	Данные веб-аналитики	Accuracy 0,87 AUC 0,91	CatBoost, LightGBM, Logistic Regression

Анализ публикаций показывает, что подавляющее большинство из них оперирует открытыми наборами данных, тогда как данные веб-аналитики используются крайне редко. Между тем именно такие данные представляют собой наиболее перспективную основу для обучения алгоритмов и построения универсальных моделей, применимых к различным интернет-магазинам.

Рассматривая рынок электронной коммерции в Российской Федерации, можно отметить, что большинство интернет-магазинов, независимо от размера бизнеса используют Яндекс.Метрику как инструмент для сбора и анализа пользовательских данных. Следовательно, данные этого сервиса обладают высокой степенью унификации и представляют собой ценный источник событийной информации для последующего обучения моделей прогнозирования.

Что касается открытых наборов данных, они, как правило, публикуются однократно в рамках определенных мероприятий и предназначены для решения узкоспециализированных задач. Значительная часть исследований направлена исключительно на достижение максимальных показателей точности в рамках предоставленного набора, что ограничивает возможность обобщения результатов. Более того, подобные данные зачастую недоступны большинству интернет-магазинов. Так, например, в работах [11, 12] анализируемые наборы включают признаки, характеризующие конкретных клиентов, такие как возраст, пол, семейное положение, доход и др. Подобные сведения могут собираться лишь в отдельных отраслях, например, в банковской сфере, где клиенты предоставляют чувствительную персональную информацию. В то же время для интернет-магазинов подобные данные чаще всего отсутствуют.

Во всех упомянутых исследованиях продемонстрированы высокие показатели точности обученных алгоритмов. С одной стороны, это свидетельствует о корректности выбора методов и подходов, однако с другой стороны полученные результаты имеют ограниченную практическую ценность, поскольку основаны на искусственно

сформированных наборах данных. Для объективной оценки качества моделей необходимо проведение серий экспериментов на различных выборках и с разными наборами исходных признаков, что позволит подтвердить стабильность результатов.

Особого внимания заслуживает работа [8], в которой рассматривается прогнозирование покупательского поведения пользователей на основе данных Яндекс.Метрики. В ней представлен веб-сервис, принимающий параметры пользователя и возвращающий рекомендации о целесообразности предоставления скидки. Авторам удалось достичь значения точности AUC в 0,91, применяя ассоциативные правила, анализ временных рядов и скоринговую модель.

Сервис реализован на платформе Loginom, предоставляющей набор готовых компонентов машинного обучения для построения аналитических цепочек. Данное решение близко к предлагаемому в настоящем исследовании, однако имеет ряд существенных ограничений:

- небольшой объем выборки, на что указывают сами авторы;
- ограниченный набор признаков, не позволяющий выявлять более сложные поведенческие паттерны;
- ориентация на определение необходимости скидки, тогда как прогнозирование готовности к совершению покупки дает более широкие возможности для последующей персонализированной коммуникации с клиентом;
- отсутствие учета событийных данных пользователей;
- высокая точность при малом объеме выборки, что не позволяет объективно оценить прогностическую способность модели.

Таким образом, анализ существующих решений показывает, что, несмотря на достигнутые успехи, исследования в области прогнозирования покупательского поведения нуждаются в более глубоком использовании событийных данных, позволяющем повысить адаптивность и применимость моделей к условиям реальных интернет-магазинов.

Материалы и методы

Ключевые определения. Типология событийных данных. Интернет-магазин доступен для посещения любому пользователю сети Интернет. Каждый пользователь может взаимодействовать с сайтом, совершая различные действия в течение определенного промежутка времени. События, совершаемые пользователем, можно определить как фиксируемые системой веб-аналитики действия, подлежащие дальнейшему анализу. В качестве таких действий могут выступать, например, клик по элементу интерфейса, добавление товара в корзину, оформление заказа и другие формы активности.

Совокупность действий, совершенных пользователем в рамках непрерывного периода взаимодействия с сайтом, образует визит. Иными словами, в пределах одного визита пользователь осуществляет последовательность действий, после чего либо покидает сайт, либо визит автоматически помечается как завершенный по истечении заданного промежутка времени (как правило, 30 минут). Данный параметр может изменяться в зависимости от настроек веб-аналитической системы (по умолчанию, в Яндекс.Метрике он равен 30 минутам).

Поскольку пользователи могут посещать интернет-магазин неоднократно, каждому пользователю соответствует множество визитов, а внутри каждого визита фиксируется набор событий.

На Рисунке 1 представлена обобщенная схема иерархии данных интернет-магазина, отражающая отношения между пользователями, их визитами и совершамыми в рамках этих визитов событиями.

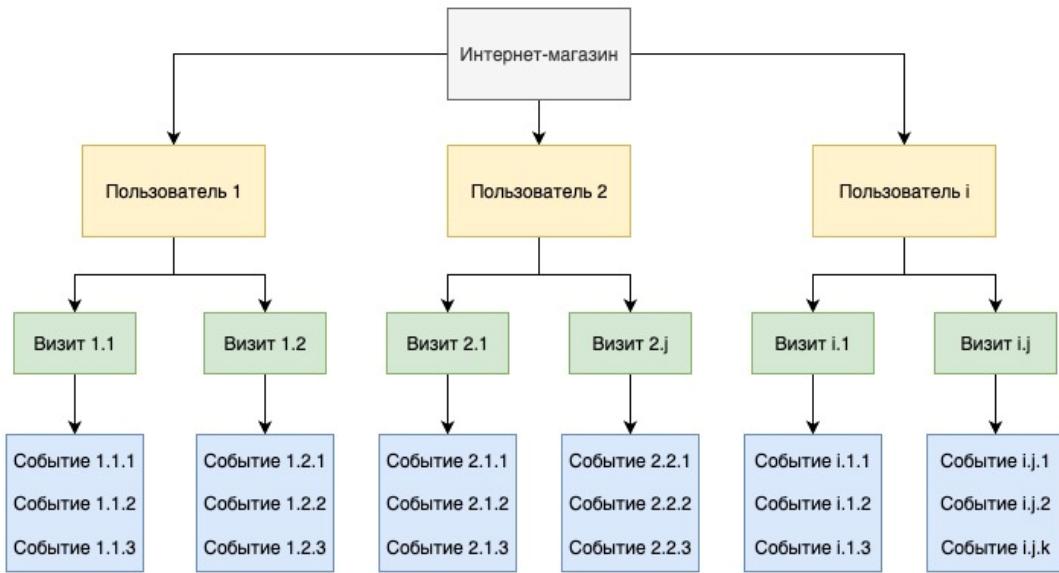


Рисунок 1 – Обобщенная схема иерархии данных интернет-магазина

Figure 1 – Generalized schema of the data hierarchy of an online store

Пусть множество пользователей определенного интернет-магазина описывается как:

$$U = \{u_1, u_2, \dots, u_n\}, \quad (1)$$

где n – общее количество пользователей.

Каждому пользователю u_i соответствует множество его визитов:

$$V_i = \{v_{i1}, v_{i2}, \dots, v_{iM_i}\}, \quad (2)$$

где M_i – количество визитов пользователя u_i .

Каждый визит v_{ij} представлен последовательностью событий:

$$E = \{(t_{ijk}, a_{ijk}), (t_{ijk}, a_{ijk}), \dots, (t_{ijk}, a_{ijk})\}, \quad (3)$$

где t_{ijk} – временная метка события, а $a_{ijk} \in A$ – тип действия, выбранного из множества доступных событий $A = \{a_1, a_2, \dots, a_M\}$, например, просмотр определенной страницы, удаление товара из корзины, чтение блока комментариев в карточке товара на сайте. Таким образом, каждый визит v_{ij} описывает временную последовательность событий, фиксирующих поведение пользователя u_i при взаимодействии с интернет-магазином.

Раскрывая типологию событийных данных, прежде всего следует отметить, что практически любое действие пользователя на веб-сайте может быть зафиксировано и передано в систему веб-аналитики. Возможности отслеживания событий, как правило, ограничиваются лишь технической реализацией логики срабатывания события и возможностями программной интеграции – например, посредством внедрения соответствующих скриптов на языках JavaScript или Python.

В общем случае все события, фиксируемые на сайте интернет-магазина, можно разделить на несколько категорий в зависимости от их назначения и роли в анализе поведения пользователей.

Во-первых, к базовой категории относятся *события электронной коммерции*. Данные события направлены на отслеживание ключевых этапов воронки взаимодействия пользователя с сайтом – от первого просмотра товаров до совершения покупки. Как правило, именно этот набор событий внедряется в первую очередь при разработке интернет-магазина и настройке веб-аналитической системы.

К событиям данной группы может относиться:

- просмотр списка товаров;
- переход к карточке товара;
- просмотр карточки товара;
- добавление товара в корзину;
- удаление товара из корзины;
- переход в корзину;
- начало оформления заказа;
- завершение покупки.

Следует подчеркнуть, что конкретный перечень событий может варьироваться в зависимости от специфики бизнеса и особенностей структуры сайта. Однако во всех случаях такие события отражают основную последовательность действий, приводящих к конверсии.

Во-вторых, выделяются так называемые *поведенческие события*, которые позволяют зафиксировать дополнительную активность пользователей, не связанную напрямую с покупкой, но оказывающую влияние на их вовлеченность и готовность к целевому действию. К данной группе относятся, например:

- чтение отзывов или комментариев к товару;
- просмотр фотографий в галерее;
- переходы по внутренним ссылкам (например, в раздел новостей или акций);
- подписка на рассылку или уведомления;
- использование фильтров и сортировок;
- добавление товара в избранное или сравнение.

Такие события отражают глубину взаимодействия пользователя с контентом сайта и позволяют выявлять скрытые закономерности в покупательском поведении, которые неочевидны при анализе только транзакционных данных.

В-третьих, в ряде случаев выделяют вспомогательную категорию – навигационные или технические события, регистрирующие общее взаимодействие пользователя с сайтом: загрузку страницы, прокрутку контента, переход между разделами и т. п.

Несмотря на их второстепенный характер, такие события позволяют оценивать качество пользовательского пути и выявлять возможные узкие места в интерфейсе сайта. Таким образом, типология событийных данных охватывает как *основные коммерческие действия*, непосредственно связанные с продажами, так и *поведенческие события*, характеризующие глубину интереса и вовлеченность пользователя. Комплексное использование всех типов событий позволяет сформировать более полное представление о поведении клиентов и повысить точность прогнозных моделей, основанных на событийной аналитике.

Используемые наборы данных. Для оценки эффективности использования событийных данных при прогнозировании покупательского поведения в интернет-магазине были сформированы наборы данных, полученные посредством API-интеграции с системой веб-аналитики Яндекс.Метрика. Данный сервис позволяет извлекать данные конкретного интернет-магазина в виде событийной структуры, либо данных на уровне визитов, где каждая строка данных является отдельным визитом определенного пользователя.

Для проведения дальнейшего эксперимента определимся с теми наборами признаков, по которым будут проводиться замеры точности алгоритма машинного обучения. Для каждого интернет-магазина извлекается один и тот же набор полей из API Яндекс.Метрика уровня визитов. В Таблице 2 представлен список выгружаемых полей из API Яндекс.Метрика.

Таблица 2 – Список выгружаемых полей из API Яндекс.Метрика
Table 2 – List of exported fields from the Yandex.Metrica API

Наименование поля	Тип данных поля	Описание поля
ym:s:visitID	UInt64	Идентификатор визита, уникален в рамках одного года
ym:s:clientID	UInt64	Анонимный идентификатор пользователя в браузере
ym:s:dateTime	DateTime	Дата и время визита
ym:s:pageViews	Int32	Глубина просмотра
ym:s:visitDuration	UInt32	Время на сайте
ym:s:bounce	UInt8	Отказность
ym:s:isNewUser	UInt8	Первый визит посетителя
ym:s:goalsID	Array(UInt32)	Номера целей, достигнутых за данный визит
ym:s:<attribution>TrafficSource	String	Источник трафика
ym:s:purchaseID	Array(String)	Идентификатор покупки. Список всех идентификаторов транзакций для визита
ym:s:purchaseRevenue	Array(Float64)	Общий доход или суммарная ценность транзакции
ym:s:watchIDs	Array(UInt64)	Просмотры, которые были в данном визите
ym:s:deviceCategory	String	Тип устройства. Возможные значения: 1 – десктоп, 2 – мобильные телефоны, 3 – планшеты, 4 – TV

Такой предлагаемый набор считается достаточным, для формирования различных наборов признаков для проведения экспериментов, поскольку охватывает ключевые аспекты пользовательской активности в рамках визита пользователя. Говоря об указанных данных, хочется отметить, что набор состоит из некоторых вспомогательных полей, которые не будут участвовать в процессе обучения алгоритма. Среди таких это дата и время визита, идентификатор визита и уникальный идентификатор пользователя. Все эти поля используются для формирования промежуточных срезов и дополнительной фильтрации данных при необходимости, а также понимания какому пользователю или какому визиту принадлежит то или иное значение получившегося прогноза совершения покупки.

Отдельное внимание стоит уделить полю, включающему в себя номера целей, которые были достигнуты за определенный визит – это и есть те событийные данные, которые пользователь совершают в течение визита. То есть каждый отдельный визит хранит в себе перечень из достижений определенной цели, то есть события.

Остальные поля являются либо категориальными, которые будут так же присутствовать в итоговых наборах данных, либо же поля с числовыми значениями определенных сущностей по каждому пользователю, например, сколько страниц посетил, сколько совершил покупок и на какую сумму.

Извлеченный набор данных был разделен на семь подмножеств, каждое из которых формировалось на основе определенного набора признаков или их комбинаций. Такой подход позволяет оценить влияние различных групп признаков на качество прогноза. Наборы данных:

- 1) только базовые признаки;
- 2) только событийные данные;
- 3) только накопительные признаки по событийным данным;
- 4) только событийные данные со смещением на уровне визитов (1, 2, 3 уровней);
- 5) только событийные данные и их вариации (п. 2, п. 3, п. 4);

- 6) базовые признаки и событийные данные с их вариациями (п. 1 и п. 5);
- 7) базовые признаки и событийные данные (п. 1 и п. 2).

Базовыми признаками в данном случае называют признаки, относящиеся к конкретному визиту пользователя и его историческое посещение интернет-магазина, при этом событийные сюда не входят. Некоторые из базовых признаков были извлечены из исходных данных посредством проектирования признаков, что позволило сформировать набор базовых признаков, описывающих пользователя, его устройство, источник трафика и характеристики поведения при взаимодействии с интернет-магазином.

В базовые признаки входят следующие данные.

1. Характеристики пользователя и источника визита:

- новый пользователь (`is_new_user`) – бинарный признак, указывающий, является ли данный пользователь новым посетителем сайта;
- категория устройства (`device_category`) – тип устройства, с которого был совершен визит (например, десктоп, смартфон, планшет);
- источник трафика (`traffic_source_encoded`) – закодированная категория источника перехода (поисковый трафик, реклама, прямой переход и т. д.);
- признак выходного дня (`is_weekend`) – бинарный индикатор, показывающий, пришел ли визит в выходной день.

2. Метрики текущего визита:

- длительность визита (`visit_duration_in_minutes`) – общее время визита пользователя в минутах;
- количество действий в текущей сессии (`current_session_hits_count`) – число зафиксированных событий за визит;
- количество покупок в текущей сессии (`current_session_purchase_count`) – число завершенных покупок в рамках визита;
- выручка текущей сессии (`current_session_purchase_revenue`) – суммарная стоимость покупок, совершенных пользователем в текущем визите.

3. Исторические показатели активности:

- количество предыдущих визитов (`previous_sessions_count`) – общее число визитов пользователя до текущего визита;
- количество визитов с ошибками или отказами (`previous_forbidden_sessions_count`)
- число сессий, завершившихся отказом, технической ошибкой или минимальной активностью;
- общее количество покупок (`total_purchase_count`) – совокупное количество завершенных транзакций пользователя за весь период наблюдения;
- среднее количество покупок (`average_purchase_count`) – усредненное значение количества покупок на визит;
- дни с момента последней покупки (`days_since_last_purchase`) – временной промежуток между текущим визитом и последней покупкой;
- коэффициент отказов (`bounce_rate`) – доля визитов, завершившихся без взаимодействия с контентом (односторонние сессии);
- дни с момента предыдущего визита (`days_since_previous_visit`) – промежуток времени между текущей и предыдущей сессией пользователя.

Данные из п. 2 содержат первые 50 % наиболее встречающихся по количеству событий. Конкретные события по каждомуциальному интернет-магазине будут отличаться, поскольку компании могут продавать как одежду, электронные книги, так и лекарства. Соответственно каждая компания сама принимает решение о том, какие события она будет отслеживать для аналитики, поэтому данные содержат те события, которые были настроены за время функционирования конкретного интернет-магазина.

Первые 50 % наиболее частых событий были выбраны для того, чтобы отсечь те события, которые срабатывают редко, что не будет являться ценным для обучения будущего алгоритма.

Данные из п. 3 представляют собой суммарные значения количества срабатываний отдельных событий за то количество визитов, которое совершил определенный пользователь за установленное количество дней в прошлом. Данное значение по умолчанию равно 30 дням, назначенное в рамках текущего исследования.

Говоря о данных из п. 4, они представляют собой событийные данные, для которых были посчитаны значения со смещением на одну, две и три позиции в прошлое на уровне визитов. То есть определенный визит пользователя будет описываться значением количества совершенных событий за прошлый визит и два ранних визита после, при условии, что визиты были, иначе значение равно 0.

Все остальные наборы данных являются комбинацией из нескольких отдельных наборов данных, применяемые для того, чтобы изучить вопрос совместного использования разных наборов признаков для выявления наиболее точного из предложенных экспериментов.

Экспериментальный стенд. В рамках данного исследования был разработан экспериментальный стенд, для проведения экспериментов с разными наборами данных. Данный стенд был реализован посредством языка программирования Python версии 3.10, использованием СУБД PostgreSQL версии 15.14, для хранения промежуточных данных по каждому эксперименту, объектное хранилище S3 для хранения артефактов, необходимых для дальнейшего обучения, программное обеспечение (ПО) для оркестрации процессов Apache Airflow версии 2.9.1, для формирования единого поэтапного процесса обучения алгоритма, а также ПО Kubernetes версии 1.28 для оркестрации и управления элементами процесса. На Рисунке 2 показана диаграмма IDEF0 уровня A0, описывающая процесс формирования прогноза совершения покупки пользователем.

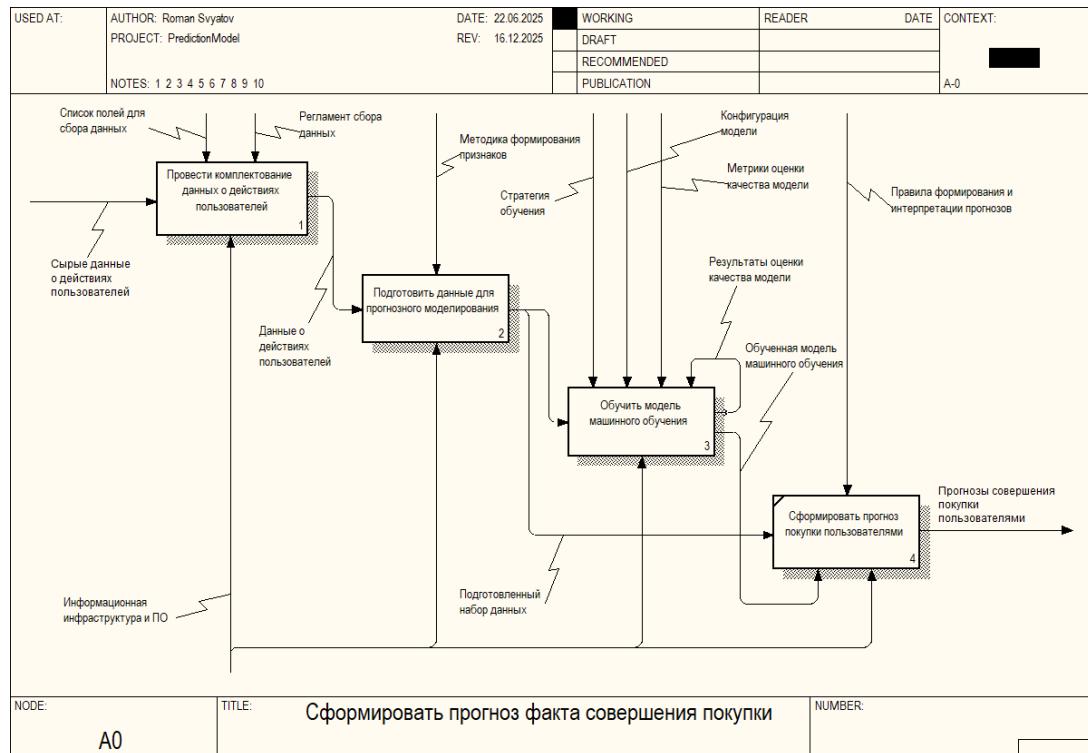


Рисунок 2 – Диаграмма IDEF0 уровня A0 формирования прогноза факта совершения покупки
 Figure 2 – IDEF0 Level A0 diagram of the purchase prediction process

Входными параметрами для запуска процесса обучения алгоритма машинного обучения является идентификатор проекта и идентификатор модели. Эти две сущности описываются набором полей, которые заполняются для определенного эксперимента, что позволит извлекать данные по определенному полю для эксперимента из СУБД PostgreSQL. На Рисунке 3 представлена структура базы данных, используемой для хранения информации о проектах, моделях и проводимых экспериментах.

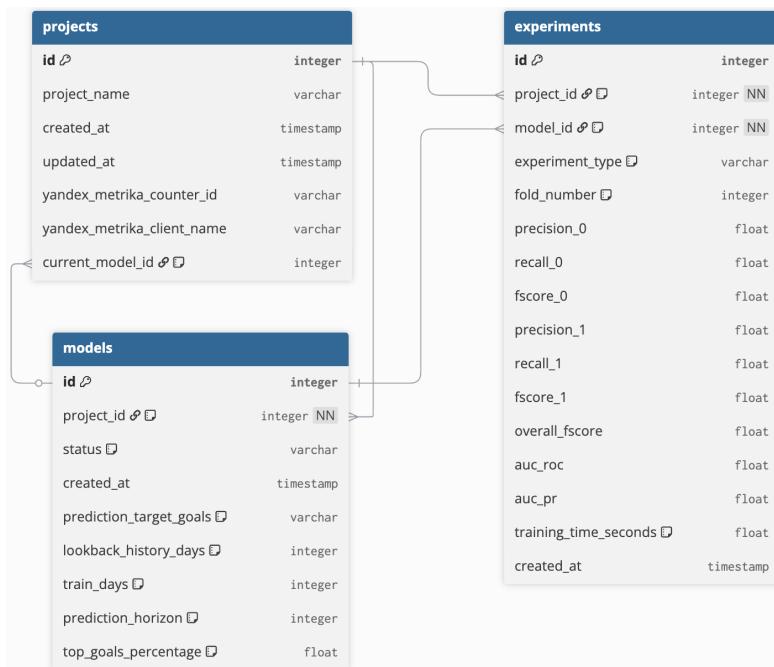


Рисунок 3 – Структура базы данных, используемой для хранения информации о проектах, моделях и экспериментах

Figure 3 – The database structure used for storing information about projects, models, and experiments

Сущность проекта включает в себя уникальный идентификатор проекта, наименование проекта, дату создания и обновления проекта, идентификатор счетчика Яндекс.Метрика, данные которого будут использоваться.

Сущность модели включает в себя уникальный идентификатор модели, идентификатор проекта к которому она относится, статус, дату создания, и список из идентификаторов цели, отвечающих за совершение покупки, что и будет прогнозироваться в дальнейшем. Также указывается параметр, отвечающий за промежуток в днях для расчета накопительных данных (значение по умолчанию в 30 дней). Количество дней, используемое для обучения алгоритма и горизонт прогнозирования в днях, то есть с какой долей вероятности пользователь совершил покупку в течение заданного промежутка в днях в будущем. Значение для отбора наиболее встречающихся событий в процентах (по умолчанию 25 %).

Дополнительно к двум упомянутым сущностям, существует третья, которая хранит в себе информацию о проведенных экспериментах. Она содержит информацию о том, для какого проекта и какой модели были проведены эксперименты, и каких показателей точности удалось достичь для того или иного набора данных.

Получив входные идентификаторы проекта и модели, программный код в ПО Apache Airflow начинает выполнение на основе инструкции, прописанной в DAG (Directed Acyclic Graph) – направленного ациклического графа, описывающего порядок выполнения задач (tasks). Каждая задача в графе представляет собой отдельный шаг

обработки данных или вычислений, реализованный в виде программного скрипта на языке программирования Python.

В рамках данного исследования DAG обеспечивает полный цикл обработки данных и обучения модели прогнозирования покупательского поведения пользователей интернет-магазинов. Все задачи выполняются последовательно или параллельно, в зависимости от их логических зависимостей, что позволяет оптимизировать процесс и минимизировать время выполнения.

Процесс начинается с загрузки параметров проекта, где из СУБД извлекаются настройки, содержащие идентификаторы счетчика Яндекс.Метрики, период анализа, целевые события и другие параметры, необходимые для работы.

Затем выполняется загрузка данных из Яндекс.Метрики, включая информацию о визитах пользователей, событиях и достижениях целей. На следующем этапе происходит извлечение перечня целей и определение наиболее часто встречающихся событий, что позволяет сузить фокус анализа на значимых поведенческих паттернах.

После этого выполняется обработка и кодирование категориальных признаков (например, источников трафика или типов устройств), что подготавливает данные для последующего обучения алгоритмов машинного обучения. Далее данные разделяются на отдельные батчи (пакеты), что дает возможность параллельно обрабатывать большие объемы информации.

На каждом батче выполняется построение базовых признаков (например, длительность визита, число просмотренных страниц, количество покупок) и признаков, основанных на событийных данных, отражающих взаимодействие пользователя с ключевыми элементами веб-сайта. Полученные признаки объединяются в единый набор данных, который служит входом для экспериментов с обучением алгоритмов.

Заключительный этап DAG включает модуль экспериментального обучения, в котором выполняется серия экспериментов с различными конфигурациями модели и методами валидации (например, по типам задач и фолдам по кросс-валидации). Для каждой комбинации параметров рассчитываются показатели качества, такие как точность (Precision), полнота (Recall), F-мера и площадь под кривыми ROC и PR (AUC). Результаты экспериментов сохраняются в СУБД и могут быть использованы для выбора наилучшей модели, которая в дальнейшем применяется для прогнозирования поведения пользователей.

Вычислительная инфраструктура, использованная в ходе исследования, была организована с применением программного обеспечения Kubernetes, обеспечивающего управление ресурсами и оркестрацию контейнеризированных вычислений.

Ключевым преимуществом применения данной системы является возможность автоматического масштабирования числа доступных вычислительных узлов в зависимости от текущей нагрузки, что позволяет эффективно использовать ресурсы и поддерживать стабильность вычислительного процесса.

В рамках проведенного исследования была сформирована группа узлов, включающая шесть вычислительных машин, каждая из которых оснащена 96 ядрами CPU, 576 ГБ оперативной памяти и 64 ГБ SSD-накопителя. Разворачивание Kubernetes осуществлялось в среде Яндекс.Облако, что обеспечило гибкость конфигурации и высокую отказоустойчивость инфраструктуры.

Следует отметить, что представленный объем вычислительных ресурсов является избыточным для целей текущего исследования, направленного на проведение серии экспериментальных запусков моделей. Тем не менее данная инфраструктура используется в рамках более широкого проекта, в частности – для построения и эксплуатации моделей прогнозирования покупательского поведения пользователей в различных интернет-магазинах.

Для выполнения экспериментальной части работы был задействован один вычислительный узел, ресурсы которого оказались достаточными для обеспечения требуемого уровня производительности и повторяемости результатов.

Проведение экспериментов. Для проведения экспериментов были задействованы данные 6 интернет-магазинов, разных по направленности и объему пользователей. В Таблице 3 приведен перечень интернет-магазинов, отобранных в качестве кандидатов для проведения экспериментов.

Таблица 3 – Параметры интернет-магазинов, участвующих в эксперименте
Table 3 – Parameters of the online stores participating in the experiment

Порядковый номер	Тематика	Уникальные пользователи (мес.)	Количество визитов (мес.)
1	Специальная одежда	61100	99800
2	Украшения и аксессуары	249500	366200
3	Одежда	72800	98300
4	Игровая периферия	111000	175600
5	Ювелирные изделия	526000	824000
6	Настольные игры	495800	852300

Исходя из данных таблицы, видно, что для экспериментов используются данные интернет-магазинов из разных сфер, с разным объемом пользователей и их визитов на веб-сайт. Это позволит протестировать предложенные наборы данных на разных объемах и паттернах поведения пользователей.

Помимо использования разных интернет-магазинов и различных наборов данных, дополнительно для сокращения случайности в полученных результатах, также применена кросс-валидация на данных конкретного эксперимента. В качестве количества фолдов было выбрано значение 5, что является стандартом для оценки устойчивости моделей и обеспечивает баланс между скоростью и надежностью валидации. Кросс-валидация позволяет сравнить между собой несколько моделей и выбрать оптимальную для конкретной задачи.

В данном исследовании дополнительно применяется стратификация, что позволяет сохранить соотношение классов, представленное в исходных наборах данных. Итоговое количество моделей для обучения равно 210. В качестве алгоритма для обучения прогнозирования покупательского поведения пользователей был выбран градиентный бустинг, являющийся мощным алгоритмом, показывающим высокие результаты в большинстве задач машинного обучения. В качестве реализации градиентного бустинга был использован CatBoost от компании Яндекс.

Входные параметры модели для всех проектов:

- 1) количество дней для расчета кумулятивных значений – 30 дней;
- 2) количество дней пользовательской активности для обучения модели – 60 дней;
- 3) горизонт прогнозирования – 14 дней;
- 4) процент отбора наиболее встречающихся событий для использования – 25 %.

Входные параметры для всех алгоритмов градиентного бустинга:

- 1) Количество итераций обучения – 2000;
- 2) Количество итераций перед преждевременной остановкой обучения – 450;
- 3) Скорость обучения – 0,01.

В рамках обучения моделей использовалась пользовательская метрика, реализующая процедуру поиска оптимального порога классификации t , при котором достигается максимальное значение F-меры. Для этого предсказанные вероятности преобразуются в бинарные метки при порогах $t \in [0,1; 1,0)$ с шагом 0,1, после чего

вычисляется значение F-меры для каждого варианта. Итоговое значение метрики соответствует наибольшему из полученных значений F-меры. Данный подход необходим для выбора оптимального порога определения факта покупки пользователем и расчета итоговых метрик качества алгоритмов.

Результаты и обсуждение

Для анализа оценки вклада каждого вида данных в итоговое значение прогноза были проведены следующие типы экспериментов:

- 1) только базовые признаки;
- 2) только событийные данные;
- 3) только накопительные признаки по событийным данным;
- 4) только событийные данные со смещением на уровне визитов (1, 2, 3 уровней);
- 5) только событийные данные и их вариации (п. 2, п. 3, п. 4);
- 6) базовые признаки и событийные данные с их вариациями (п. 1 и п. 5);
- 7) базовые признаки и событийные данные (п. 1 и п. 2).

На Рисунке 4 изображена зависимость среднего значения F-меры от типа эксперимента для различных интернет-магазинов.

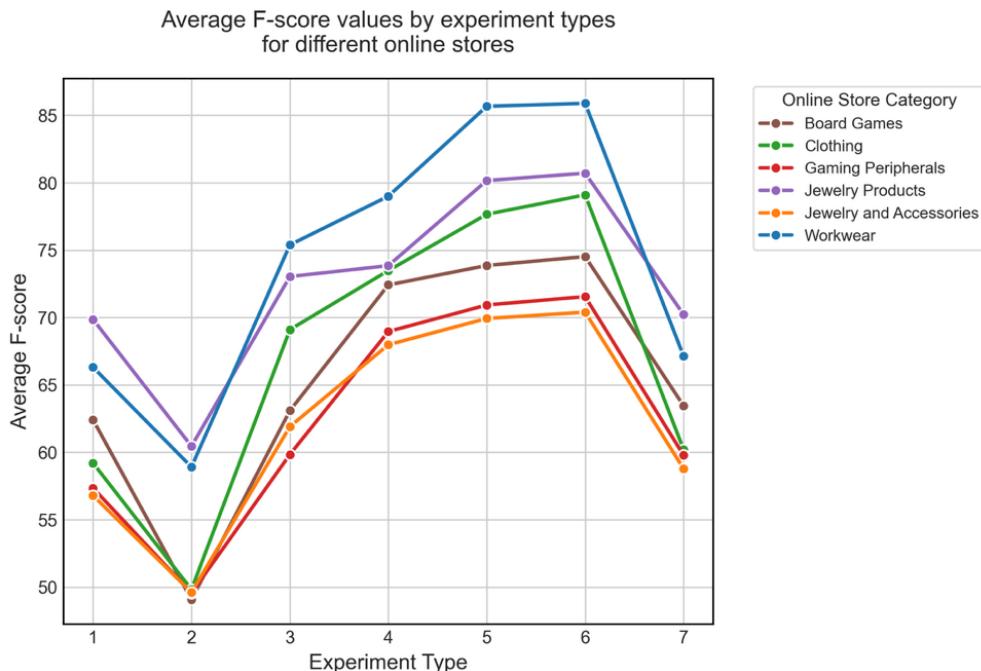


Рисунок 4 – Зависимость средней F-меры от типа эксперимента для различных интернет-магазинов

Figure 4 – The relationship between average F-score values and the type of experiment across different online stores

Каждый из семи экспериментов отражал определенную комбинацию признаков, используемых при обучении модели прогнозирования покупательского поведения. Основная цель заключалась в оценке влияния событийных данных и их производных на итоговое качество модели, измеряемое показателем F-меры.

Анализ полученных значений позволяет выявить несколько устойчивых закономерностей. Во-первых, во всех тематических направлениях интернет-магазинов наихудшие результаты продемонстрировали модели, обученные исключительно на базовых признаках (эксперимент 1) и только на исходных событийных данных

(эксперимент 2). Это указывает на ограниченность каждой группы признаков при их изолированном применении.

Так, для большинства тематик, F-мера в данных экспериментах не превышала 60–67 %, что свидетельствует о недостаточной информативности данных без учета взаимодействий и временной структуры пользовательских действий.

Во-вторых, при переходе к использованию накопительных событийных признаков (эксперимент 3) и событий со смещением на уровне визитов (эксперимент 4), наблюдается существенный рост показателя F-меры. Средние значения достигают 70–79 % в зависимости от тематики, что подтверждает эффективность подхода, учитывавшего последовательность и контекст действий пользователей.

Наиболее высокие результаты были получены в экспериментах 5 и 6, где объединялись различные вариации событийных признаков, а также их комбинация с базовыми характеристиками пользователей. Для всех интернет-магазинов именно эти эксперименты продемонстрировали максимальные значения F-меры (в диапазоне от 70 % до 86 %).

Наибольший эффект наблюдается в категориях «Спецодежда» (85,9 %) и «Ювелирные изделия» (80,7 %), что может быть связано с более выраженной структурой пользовательских сценариев и повторяемостью действий в данных сегментах.

В то же время в эксперименте 7, где использовались только базовые и исходные событийные признаки без их производных, наблюдается снижение точности. Это подтверждает гипотезу о том, что наибольший вклад в качество прогноза вносят именно производные событийные признаки, формируемые на основе агрегирования и временных сдвигов.

Продолжая анализ полученных результатов, следует рассмотреть временные характеристики обучения моделей. На Рисунке 5 представлено распределение среднего времени обучения по типам экспериментов для различных тематик интернет-магазинов.

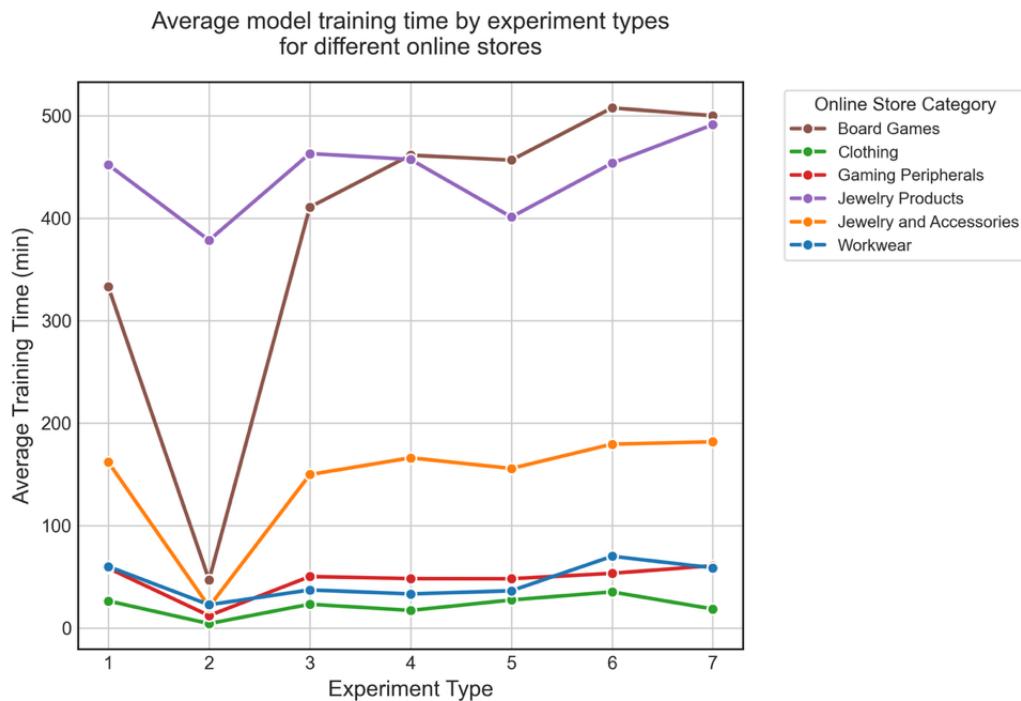


Рисунок 5 – Зависимость времени обучения (в мин.) от типа эксперимента для различных интернет-магазинов

Figure 5 – The relationship between training time (in min.) and the type of experiment across different online stores

Отмечается, что временные затраты на обучение варьируются весьма значительно – от нескольких десятков минут до нескольких часов в зависимости от категории и типа эксперимента.

Наибольшее время обучения наблюдается в тематиках «Ювелирные изделия» и «Настольные игры», что напрямую коррелирует с объемом исходных данных (Таблица 3), где количество уникальных пользователей достигает 500–520 тыс. в месяц.

Соответственно, увеличение объема данных приводит к пропорциональному росту количества событий, что усложняет процесс формирования признаков и увеличивает вычислительную нагрузку.

Также заметно, что во всех категориях пиковые значения времени обучения приходятся на эксперименты 5 и 6 – именно те, где применялись наиболее сложные комбинации базовых и событийных признаков, включая их вариации и агрегированные формы.

Это закономерно, поскольку данные эксперименты предполагают формирование значительно большего числа признаков и проведение более трудоемких процедур подготовки данных.

В то же время полученные результаты демонстрируют рациональный баланс между затратами на обучение и качеством моделей, поскольку рост вычислительной сложности сопровождается существенным улучшением показателя F-меры.

Таким образом, увеличение времени обучения моделей при использовании событийных данных является оправданным и подтверждает практическую ценность их применения в задачах прогнозирования пользовательского поведения.

Дополнительно рассмотрим эффективность разработанных моделей по разным интернет-магазинам по метрикам Precision, Recall и AUC-ROC. Метрики были получены путем усреднения получившихся значений по 5 фолдам по положительному и отрицательному классу (Precision_0, Precision_1, Recall_0, Recall_1) и дополнительному усреднению значений по каждому из классов. На Рисунках 6–8 изображены графики зависимости показателей Precision, Recall и AUC-ROC от типа эксперимента для различных интернет-магазинов.

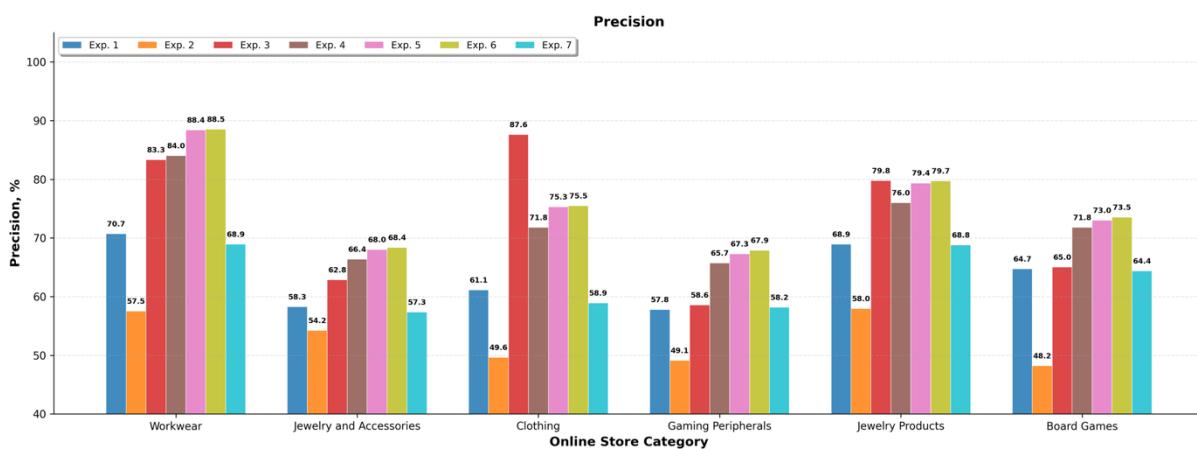


Рисунок 6 – Зависимость показателя Precision от типа эксперимента для различных интернет-магазинов

Figure 6 – The relationship between the Precision metric and the type of experiment across different online stores

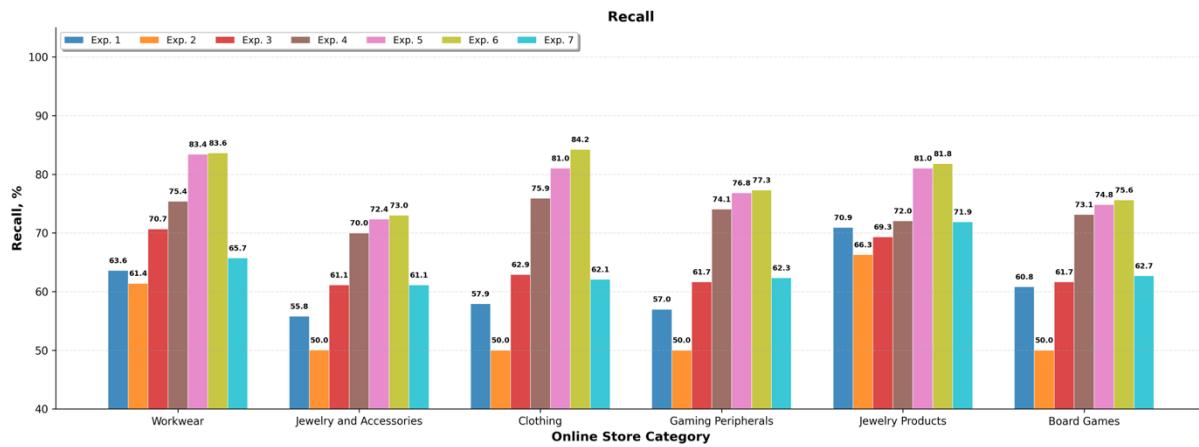


Рисунок 7 – Зависимость показателя Recall от типа эксперимента для различных интернет-магазинов

Figure 7 – The relationship between the Recall metric and the type of experiment across different online stores

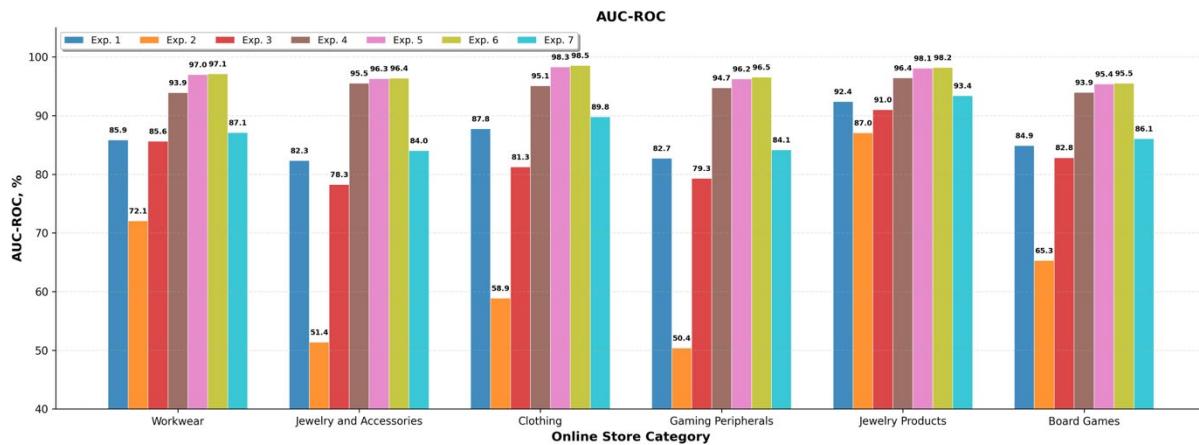


Рисунок 8 – Зависимость показателя AUC-ROC от типа эксперимента для различных интернет-магазинов

Figure 8 – The relationship between the AUC-ROC metric and the type of experiment across different online stores

Анализируя полученные результаты по метрикам Precision, Recall и AUC-ROC, прослеживается четкая динамика увеличения точности метрик по мере перехода от более ранних к более поздним экспериментам. Начиная с четвертого эксперимента, для всех категорий интернет-магазинов наблюдается существенный рост всех трех показателей.

Как и в случае с метрикой F-меры, лидером по-прежнему являются эксперименты под номером 5 и 6, где эксперимент 6 опережает эксперимент 5 на несколько процентов по всем показателям. Среди всех типов интернет-магазинов стоит выделить продажу специальной одежды и ювелирных изделий, где в среднем показатели по всем метрикам больше, чем в других типах интернет-магазинов.

В совокупности анализа всех метрик можно заключить, что данные, используемые в экспериментах 5 и 6, оказались наиболее эффективными с точки зрения рассматриваемых метрик качества. Отдельное внимание стоит уделить результатам эксперимента 6, который включал в себя как базовые признаки, так и все вариации признаков событийных данных пользователей, что подтверждает важность применения событийных данных при прогнозировании покупательского поведения пользователей

интернет-магазинов. Дополнительно не стоит забывать про базовые признаки, которые так же вносят свой вклад в качество прогнозирования.

Обсуждение

Полученные результаты подтверждают гипотезу о том, что использование событийных данных позволяет значительно повысить качество прогнозирования покупательского поведения пользователей интернет-магазинов. При этом наиболее высокая эффективность достигается при объединении событийных данных с базовыми признаками, описывающими характеристики визита и историю активности пользователя. Данный эффект прослеживается во всех тематических категориях интернет-магазинов, что свидетельствует об универсальности предложенного подхода.

Особое значение имеет тот факт, что использование производных и агрегированных событийных признаков (эксперименты 5 и 6) привело к устойчивому росту показателей F-меры, Precision, Recall и AUC-ROC при сохранении стабильных результатов на различных выборках. Это демонстрирует, что модели, основанные на динамических аспектах поведения пользователей, способны улавливать временные закономерности и выявлять паттерны, недоступные при анализе только статических признаков.

Несмотря на высокие показатели точности рассмотренных конкурирующих исследований, текущее исследование показало свою стабильность на всевозможных вариациях входных признаков и разных срезов, используемых при обучении моделей. Текущие исследования в области прогнозирования покупательского поведения ограничиваются своей практической применимостью и стабильностью получаемых метрик качества моделей. Значения показателя F-меры в текущем исследовании превышало значение в 70 % по всем категориям интернет-магазинов (эксперимент 6), что свидетельствует о практической применимости и эффективности предложенного подхода.

Дополнительно в ходе экспериментов было отмечено, что время обучения моделей напрямую зависит от объема исходных данных. Так, для интернет-магазинов с наибольшим числом пользователей и визитов, например «Ювелирные изделия» и «Настольные игры», время обучения оказалось значительно выше, по сравнению с другими тематиками. Данный эффект объясняется увеличением числа событий и усложнением признакового пространства при росте выборки. Тем не менее даже при увеличении вычислительной нагрузки качество моделей оставалось стабильно высоким, что подтверждает практическую устойчивость предложенного подхода при масштабировании на более крупные данные.

Следует отметить, что предложенный подход сохраняет масштабируемость и может быть применен для различных интернет-магазинов. Архитектура экспериментального стенда в сочетании с используемыми технологиями обеспечивает воспроизводимость и переносимость решений без необходимости ручной настройки. Это, в свою очередь, открывает перспективы практического внедрения разработанной методики в коммерческих аналитических системах.

Таким образом, проведенное исследование подтверждает значимость событийных данных в задачах прогнозирования покупательского поведения и демонстрирует возможность их эффективного интегрирования в существующие процессы анализа пользовательских данных.

Заключение

В работе представлено исследование, направленное на повышение эффективности прогнозирования покупательского поведения пользователей интернет-магазинов за счет применения событийных данных.

Проведен анализ существующих решений, показавший, что большинство из них основано на открытых наборах данных и ограниченных признаках, что снижает их применимость в реальных бизнес-сценариях.

Предложенный подход использует события, формируемые в реальной среде интернет-магазинов, что позволяет учитывать специфику пользовательской активности и особенности конкретных бизнес-доменов.

Методика включает формирование базовых, событийных и производных признаков, а также их комбинаций, что дало возможность оценить влияние каждой группы данных на качество прогноза.

Результаты экспериментов показали, что наилучшие значения F-меры, Precision, Recall и AUC-ROC достигаются при использовании комплексных наборов данных, включающих событийные признаки и их вариации. Это подтверждает, что динамические аспекты поведения пользователей являются ключевыми для построения точных моделей прогнозирования.

Кроме того, предложенный подход продемонстрировал устойчивость и масштабируемость. Используемая архитектура вычислительного стенда обеспечивает воспроизводимость экспериментов и возможность адаптации методики к интернет-магазинам различного масштаба без существенных изменений.

В дальнейшем развитие подхода может быть связано с расширением набора событийных признаков, применением нейросетевых моделей для анализа последовательностей действий и интеграцией разработанной методики в существующие аналитические системы интернет-торговли.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Torres D., Cepeda L.K. Machine Learning for Predicting Online Shoppers' Purchase Intentions. [Preprint]. ResearchGate. URL: <https://doi.org/10.13140/RG.2.2.15297.36968> [Accessed 3rd November 2025].
2. Satu M.Sh., Islam S.F. Modeling Online Customer Purchase Intention Behavior Applying Different Feature Engineering and Classification Techniques. *Discover Artificial Intelligence*. 2023;3(1). <https://doi.org/10.1007/s44163-023-00086-0>
3. Bigon L., Cassani G., Greco C., et al. Prediction Is Very Hard, Especially about Conversion. Predicting User Purchases from Clickstream Data in Fashion E-Commerce. arXiv. URL: <https://doi.org/10.48550/arXiv.1907.00400> [Accessed 3rd November 2025].
4. Hendriksen M., Kuiper E., Nauts P., Schelter S., de Rijke M. Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers. arXiv. URL: <https://doi.org/10.48550/arXiv.2012.08777> [Accessed 3rd November 2025].
5. Gkikas D.C., Theodoridis P.K. Predicting Online Shopping Behavior: Using Machine Learning and Google Analytics to Classify User Engagement. *Applied Sciences*. 2024;14(23). <https://doi.org/10.3390/app142311403>
6. Pîrvu M.C., Anghel A. Predicting Next Shopping Stage Using Google Analytics Data for E-Commerce Applications. arXiv. URL: <https://doi.org/10.48550/arXiv.1905.12595> [Accessed 4th November 2025].
7. Лагерев Д.Г., Савостин И.А., Герасимчук В.Ю., Полякова М.С. Исследование склонности пользователя интернет-магазина к покупке на основе технических данных о визитах посетителей интернет-магазина. *Современные информационные*

технологии и ИТ-образование. 2018;14(4):911–922. <https://doi.org/10.25559/SITITO.14.201804.911-922>

Lagerev D.G., Savostin I.A., Gerasimchuk V.Yu., Polyakova M.S. The Study of the Purchase Propensity of the User of the Online Store on the Basis of Technical Data on Visits of Visitors to the Online Store. *Modern Information Technology and IT-Education*. 2018;14(4):911–922. (In Russ.). <https://doi.org/10.25559/SITITO.14.201804.911-922>

8. Мамиев О.А., Финогенов Н.А., Сологуб Г.Б. Использование методов машинного обучения для решения задач прогнозирования суммы и вероятности покупки на основе данных электронной коммерции. *Моделирование и анализ данных*. 2020;10(4):31–40. <https://doi.org/10.17759/mda.2020100403>

Mamiev O.A., Finogenov N.A., Sologub G.B. Using Machine Learning Methods to Solve Problems of Forecasting the Amount and Probability of Purchase Based on E-Commerce Data. *Modelling and Data Analysis*. 2020;10(4):31–40. (In Russ.). <https://doi.org/10.17759/mda.2020100403>

9. Hes vindrati N., Aminuddin A., Mahadhni J., Pambudi A., Sudaryatno B. Behavior-Based Purchase Intent Prediction in E-Commerce: A Machine Learning Approach. *International Journal of Current Science Research and Review*. 2025;8(8):3970–3980. <https://doi.org/10.47191/ijcsrr/v8-i8-03>

10. Zhang Ch., Liu J., Zhang Sh. Online Purchase Behavior Prediction Model Based on Recurrent Neural Network and Naive Bayes. *Journal of Theoretical and Applied Electronic Commerce Research*. 2024;19(4):3461–3476. <https://doi.org/10.3390/jtaer19040168>

11. Wang W., Xiong W., Wang J., et al. A User Purchase Behavior Prediction Method Based on XGBoost. *Electronics*. 2023;12(9). <https://doi.org/10.3390/electronics12092047>

12. Xiao Sh., Tong W. Prediction of User Consumption Behavior Data Based on the Combined Model of TF-IDF and Logistic Regression. In: *Journal of Physics: Conference Series: Volume 1757, International Conference on Computer Big Data and Artificial Intelligence (ICCBDAI 2020), 24–25 October 2020, Changsha, China*. IOP Publishing; 2021. <https://doi.org/10.1088/1742-6596/1757/1/012089>

13. Yuwei J. Research on Prediction of E-Commerce Repurchase Behavior Based on Multiple Fusion Models. *Applied and Computational Engineering*. 2023;2:90–104. <https://doi.org/10.54254/2755-2721/2/20220555>

14. Liu D., Huang H., Zhang H., Luo X., Fan Zh. Enhancing Customer Behavior Prediction in E-Commerce: A Comparative Analysis of Machine Learning and Deep Learning Models. *Applied and Computational Engineering*. 2024;55:181–195. <https://doi.org/10.54254/2755-2721/55/20241475>

15. Liu Z. Prediction Model of E-commerce Users' Purchase Behavior Based on Deep Learning. *Frontiers in Business, Economics and Management*. 2024;15(2):147–149. <https://doi.org/10.54097/p22ags78>

16. Li H. Research on Consumer Behavior Prediction Based on E-commerce Data Analysis. *BCP Business and Management*. 2023;49:106–110. <https://doi.org/10.54691/bcpbm.v49i.5411>

17. Zhang W., Wang M. An Improved Deep Forest Model for Prediction of E-Commerce Consumers' Repurchase Behavior. *PLoS ONE*. 2021;16(9). <https://doi.org/10.1371/journal.pone.0255906>

18. 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). <https://doi.org/10.14569/IJACSA.2022.0130249>

19. Zhou Sh., Hudin N.S. Advancing E-Commerce User Purchase Prediction: Integration of Time-Series Attention with Event-Based Timestamp Encoding and Graph Neural Network-Enhanced User Profiling. *PLoS ONE*. 2024;19(4). <https://doi.org/10.1371/journal.pone.0299087>

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Святов Роман Сергеевич, аспирант, **Roman S. Svyatov**, Postgraduate, RUDN Российской университет дружбы народов, University, Moscow, the Russian Federation. Москва, Российская Федерация.

e-mail: romasvyatov@yandex.ru

ORCID: [0009-0009-0322-1443](https://orcid.org/0009-0009-0322-1443)

*Статья поступила в редакцию 08.11.2025; одобрена после рецензирования 22.12.2025;
принята к публикации 26.12.2025.*

*The article was submitted 08.11.2025; approved after reviewing 22.12.2025;
accepted for publication 26.12.2025.*