

УДК 004.62

DOI: [10.26102/2310-6018/2026.54.3.010](https://doi.org/10.26102/2310-6018/2026.54.3.010)

## Автоматизированная система поддержки принятия решений для прогнозирования покупательского поведения пользователей интернет-магазинов

Р.С. Святлов✉

*Российский университет дружбы народов, Москва, Российская Федерация*

**Резюме.** Актуальность исследования обусловлена стремительным развитием электронной коммерции и необходимостью построения эффективных инструментов прогнозирования поведения пользователей интернет-магазинов. Проблема заключается в том, что существующие решения в этой области часто ограничены применением к конкретным наборам данных, не обладают достаточной масштабируемостью и редко поддерживают автоматизацию процесса прогнозирования в реальном времени. Целью работы является разработка системы поддержки принятия решений, позволяющей на основе анализа поведенческих данных пользователей формировать прогноз вероятности совершения покупки в будущем и предоставлять лицам, принимающим решения, готовые рекомендации для дальнейших маркетинговых действий. Методологическая основа исследования заключается в использовании системы веб-аналитики в качестве источника информации о действиях пользователей, предобработке и структурировании данных, а также применении градиентного бустинга в качестве алгоритма машинного обучения для прогнозирования вероятности совершения покупки. Для определения внутренних и внешних факторов, которые могут оказать положительное или отрицательное влияние на достижение поставленной цели, был проведен SWOT-анализ. Проведена экспериментальная апробация системы на данных четырех интернет-магазинов различной направленности. Полученные результаты показали, что общее значение показателя F-меры превышает 80 % во всех экспериментах. Материалы статьи представляют практическую ценность для специалистов в области электронной коммерции, аналитиков и маркетологов, а также лиц, принимающих решения, поскольку разработанная система позволяет автоматизировать процесс прогнозирования покупательского поведения, формировать интерпретируемые сегменты пользователей и использовать полученные результаты в задачах персонализации маркетинговых коммуникаций и оптимизации управленческих решений.

**Ключевые слова:** машинное обучение, система поддержки принятия решений, анализ поведения пользователей, электронная коммерция, прогнозирование покупательского поведения, интернет-магазины.

**Для цитирования:** Святлов Р.С. Автоматизированная система поддержки принятия решений для прогнозирования покупательского поведения пользователей интернет-магазинов. *Моделирование, оптимизация и информационные технологии.* 2026;14(3). URL: <https://moitvvt.ru/ru/journal/article?id=2230> DOI: 10.26102/2310-6018/2026.54.3.010

## Automated decision support system for predicting online shopping behavior of e-commerce users

R.S. Svyatov✉

*RUDN University, Moscow, the Russian Federation*

**Abstract.** The relevance of this study is caused by the rapid development of electronic commerce and the growing need for effective tools to predict user behavior in online retail environments. The main problem lies in the fact that existing solutions in this domain are often limited to specific datasets, lack sufficient scalability, and rarely support real-time automation of the forecasting process. The purpose of

this study is to develop a decision support system that enables the estimation of the probability of future purchase completion based on the analysis of user behavioral data and provides decision-makers with actionable recommendations for subsequent marketing activities. The methodological framework of the study is based on the use of a web analytics system as a source of information on user activities, data preprocessing and structuring procedures, and the application of gradient boosting as a machine learning algorithm for predicting the probability of purchase. To identify internal and external factors that could have a positive or negative impact on achieving the goal, a SWOT analysis was conducted. Experimental validation of the system was conducted using data from four online stores representing different business domains. The results demonstrate that the overall F-score exceeds 80 % across all experiments. The materials presented in this article have practical relevance for e-commerce professionals, data analysts, and marketing specialists, as well as for decision-makers, since the proposed system enables automated prediction of purchasing behavior, the formation of interpretable user segments, and the application of the obtained results to marketing personalization and optimization of managerial decision-making.

**Keywords:** machine learning, decision support system, user behavior analysis, e-commerce, consumer behavior prediction, online stores.

**For citation:** Svyatov R.S. Automated decision support system for predicting online shopping behavior of e-commerce users. *Modeling, Optimization and Information Technology*. 2026;14(3). (In Russ.). URL: <https://moitvivr.ru/ru/journal/article?id=2230> DOI: 10.26102/2310-6018/2026.54.3.010

## Введение

Стремительное развитие онлайн-торговли и рост количества платформ электронной коммерции обусловили необходимость разработки инструментов прогнозирования покупательского поведения пользователей. Оценка вероятности совершения покупки является ключевым элементом оптимизации маркетинговых коммуникаций и распределения рекламных бюджетов. Использование событийных данных систем веб-аналитики позволяет формировать количественные оценки будущих действий пользователей и интегрировать их в процессы поддержки управленческих решений.

В научной литературе представлен широкий спектр методов прогнозирования пользовательского поведения. В работах [1, 2] для оценки вероятности совершения покупки применяются алгоритмы машинного обучения и глубокого обучения, основанные на анализе данных пользовательских сессий. Аналогичные подходы к прогнозированию пользовательского поведения в электронной коммерции также рассматриваются в работе [3]. Исследования, посвященные формированию и отбору признакового пространства, подтверждают значимость данного этапа для повышения точности прогнозных моделей [4, 5]. Соответствующие результаты также представлены в работе [6].

В ряде работ используются открытые наборы данных для обучения моделей машинного обучения, что позволяет достигать высоких значений метрик качества [7, 8].

Несмотря на полученные положительные результаты, большинство существующих решений ограничено использованием данных одной платформы электронной коммерции, либо открытых датасетов, что существенно снижает их переносимость на другие интернет-магазины. Кроме того, значительная часть исследований сосредоточена на незначительных модификациях ранее предложенных подходов, основанных на одних и тех же открытых наборах данных.

В работах [9, 10] проводится сравнительный анализ прогностических моделей на основе открытых наборов данных, при этом достигается точность порядка 90 %. Однако данные результаты имеют ограниченную практическую значимость, поскольку применимость моделей определяется выбранной методологией и характеристиками конкретной выборки, а использование открытых данных ограничивает обобщаемость

решений и их использование в реальных условиях функционирования интернет-магазинов.

Таким образом, сохраняется потребность в разработке систем, обеспечивающих автоматизированное получение прогнозных значений для различных интернет-магазинов на основе единого источника событийных данных и наличия масштабируемости.

Целью настоящего исследования является разработка и апробация автоматизированной системы поддержки принятия решений (СППР), предназначенной для прогнозирования покупательского поведения пользователей интернет-магазинов на основе данных веб-аналитики и ориентированной на информационную поддержку лиц, принимающих решения (ЛПР), в задачах маркетингового и управленческого планирования. В рамках исследования проведены онтологический анализ, обучение модели машинного обучения и оценка ее эффективности на данных нескольких интернет-магазинов.

### Материалы и методы

В последние годы значительное количество исследований посвящено разработке моделей и систем прогнозирования покупательского поведения пользователей интернет-магазинов. В большинстве таких работ для оценки вероятности совершения покупки используется анализ поведенческих и транзакционных данных с применением методов машинного обучения [11, 12].

В Таблице 1 представлен обзор исследований, отражающих современные подходы к прогнозированию покупательского поведения в области электронной коммерции. Рассмотренные работы различаются по используемым источникам данных, применяемым алгоритмам, достигнутым значениям метрик качества, а также по характеристикам системного уровня.

Таблица 1 – Сравнительный анализ существующих исследований по прогнозированию покупательского поведения пользователей интернет-магазинов

Table 1 – Comparative analysis of existing studies on predicting online shoppers purchasing behavior

Исследование	Тип источника данных	Результаты качества	Используемые алгоритмы
[1]	Логи	AUC = 0,97	Дерево решений; Bagging; Градиентный бустинг
[13]	Открытый набор данных	Улучшение F-меры на 2,44 %	Улучшенный Deep Forest
[14]	Открытый набор данных	–	Графовые нейронные сети (GNN)
[15]	Открытый набор данных	Accuracy = 92,39 %; F-score = 0,924; AUC-ROC = 0,975	Случайный лес
[16]	Открытый набор данных	Accuracy = 90,65 %	Градиентный бустинг
[17]	Открытый набор данных	–	TPGN; LSTM
[18]	Логи	–	Fusion RF; CatBoost
[19]	Транзакционные данные	F-score = 0,65	Логистическая регрессия; Дерево решений; LightGBM
[20]	Данные кликов	–	LightGBM
[21]	Поведенческие логи	–	Искусственная нейронная сеть

Таблица 1 (продолжение)  
Table 1 (continued)

[22]	Демографические и транзакционные данные	Градиентный бустинг и RF – лучшие по F1 и accuracy	KNN; Градиентный бустинг; SVC; RF
[23]	Открытый набор данных	–	Глубокое обучение
[24]	Демографические и поведенческие данные	RF и GB превосходят по accuracy и ROC-AUC	LogReg; RF; GB; SVM; KNN; XGBoost
[25]	Пользовательские логи (Big Data)	Улучшение персонализации и пользовательского опыта	Глубокое обучение
[26]	Clickstream и данные покупок	Accuracy $\approx$ 72–75 %	RF; GBDT; XGBoost; RNN; LSTM

Значительная часть существующих исследований опирается на использование открытых наборов данных и логов пользовательских сессий, собранных в рамках конкретного интернет-магазина. В ряде работ для оценки разрабатываемых моделей машинного обучения используются открытые наборы данных [13, 14]. При этом применяются различные алгоритмические подходы, включая случайные леса и градиентный бустинг [15, 16], а также глубокие нейронные сети и их модификации [17]. В отдельных исследованиях для обучения и тестирования моделей используются логи пользовательских сессий, полученные из реальных интернет-магазинов [18].

В ряде работ сообщается о высоких показателях качества прогнозирования, при этом значения точности и показателя AUC превышают 90 %. Однако полученные результаты в значительной степени зависят от структуры и качества конкретных используемых наборов данных.

Еще одним существенным ограничением представленных подходов является недостаточная масштабируемость и адаптивность. Большинство предлагаемых решений разрабатывается под конкретный набор данных или веб-сайт и не ориентировано на интеграцию с различными интернет-магазинами. Применение обученных моделей для других интернет-магазинов, как правило, требует ручной предобработки данных, переработки признакового пространства и повторного обучения моделей, что существенно ограничивает их практическую применимость.

Ряд исследований затрагивает системные аспекты прогнозирования покупательского поведения, включая автоматизацию процессов обработки данных и стратегии формирования признаков [18, 19]. Тем не менее данные работы по-прежнему остаются ограниченными по практическому применению, специфичными для используемых наборов данных, и не предлагают единого и системного решения для применения в условиях нескольких интернет-магазинов.

В отличие от существующих подходов, предлагаемая СППР использует данные веб-аналитической платформы Яндекс.Метрика в качестве единого источника поведенческих событий пользователей. Это обеспечивает автоматизированный сбор данных, снижает зависимость от внутренних схем данных отдельных интернет-магазинов и позволяет реализовать масштабируемое развертывание системы в рамках различных направленностей интернет-магазинов.

*Теоретическое описание СППР.* С развитием производства и расширением функций человека выработка решений без применения компьютеров стала затруднительной. В связи с этим были созданы системы поддержки принятия решений. СППР определяют как интерактивные автоматизированные системы, используемые при принятии решений в ситуациях, когда невозможна или затруднительна полная автоматизация процесса решения вследствие слабой структурированности или неструктурированности задач. Целью разработки и внедрения СППР является информационная поддержка ЛПР, создание условий для принятия обоснованных решений на основе доставки данных и аналитической информации, а также формирование и эксплуатация моделей процессов. На Рисунке 1 представлена классификация разрабатываемой СППР.

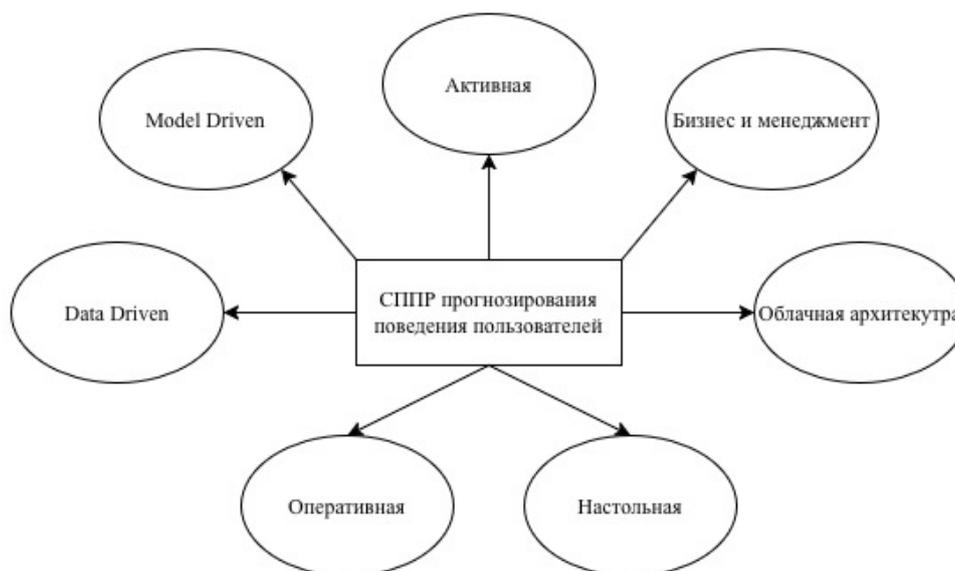


Рисунок 1 – Классификация представленной СППР  
Figure 1 – Classification of proposed DSS

По позиции пользователя СППР делят на:

- пассивные (поддерживают процесс принятия решения, но не предлагают конкретного варианта выбора);
- активные (формируют предложения о предпочтительном решении);
- кооперативные (позволяют ЛПР изменять и дополнять предложения системы с последующей проверкой и согласованием).

Разрабатываемая СППР прогнозирования поведения пользователей интернет-магазинов относится к активным СППР, поскольку не только предоставляет данные и аналитику, но и выдает рекомендации в виде прогнозных оценок и сегментов пользователей по готовности к покупке, на основе которых ЛПР принимает решения по маркетинговым действиям.

На концептуальном уровне выделяют СППР, управляемые:

- Communication-Driven DSS (СППР, управляемые сообщениями), поддерживают группу пользователей, работающих над выполнением общей задачи;
- Data-Driven DSS (СППР, управляемые данными), ориентированные на работу с данными и манипуляции с ними;
- Document-Driven DSS (СППР, управляемые документами), осуществляют поиск и манипулируют неструктурированной информацией, заданной в различных форматах;

– Knowledge-Driven DSS (СППР, управляемые знаниями), обеспечивают решение задач в виде фактов, правил, процедур;

– Model-Driven DSS (СППР, управляемые моделями), характеризуются в основном манипуляциями с математическими моделями (статистическими, финансовыми, оптимизационными, имитационными).

Разрабатываемая система опирается на исторические данные о поведении пользователей и на прогнозную модель машинного обучения, обученную по этим данным. Рекомендации формируются выходом обученной модели по сформированным признакам. По данному признаку система относится к гибриду типа, управляемого данными, и типа, управляемого моделями.

По характеру решаемых задач и используемых данных СППР разделяют на:

– оперативные (предназначены для быстрой реакции на текущее состояние процессов, опираются на детальные оперативные данные);

– стратегические (служат обоснованию долгосрочных решений на основе агрегированной ретроспективной информации).

Разрабатываемая система предназначена для поддержки оперативных маркетинговых решений на основе ежедневно обновляемых поведенческих данных и актуальных прогнозов, то есть относится к оперативным СППР.

По области применения СППР ориентируют на конкретный рынок, например, финансы, маркетинг, управление ресурсами. Разрабатываемая система применяется в области бизнеса и менеджмента и обеспечивает прогнозирование вероятности совершения покупки пользователями интернет-магазина для обоснования решений по распределению рекламного бюджета или иным маркетинговым действиям.

По архитектуре различают централизованные, децентрализованные, облачные и гибридные СППР. По охвату пользователей выделяют корпоративные системы и индивидуальные настольные СППР. Разрабатываемая система реализует облачную архитектуру с использованием инфраструктуры облачного провайдера для обеспечения масштабируемости и отказоустойчивости и ориентирована на использование одним или несколькими ЛПР в рамках интернет-магазина, что соответствует признакам настольной СППР в облачном исполнении.

Главной целью разработки СППР прогнозирования покупательского поведения является повышение качества результатов прогнозирования в сравнении с существующими решениями. Для определения внутренних и внешних факторов, которые могут оказать положительное или отрицательное влияние на достижение поставленной цели, был проведен SWOT-анализ. SWOT-анализ широко применяется при разработке технологических решений, как, например, в ряде исследований [27, 28]. Данный инструмент позволяет определять:

– сильные стороны системы, т. е. ее положительные характеристики, которые могут способствовать достижению намеченных целей;

– слабые стороны системы, т. е. отрицательные характеристики, которые могут препятствовать достижению поставленных целей;

– потенциальные возможности системы, т. е. внешние условия, которые могут способствовать достижению целей;

– возможные угрозы, т. е. внешние условия, которые могут препятствовать достижению намеченных целей.

В Таблице 2 представлена матрица проведенного SWOT-анализа, описывающая положительные и отрицательные влияния для внутренней и внешней среды рассматриваемой системы.

Таблица 2 – Матрица SWOT-анализа рассматриваемой системы  
Figure 2 – SWOT analysis matrix of the proposed system

	<b>Положительное влияние</b>	<b>Отрицательное влияние</b>
<b>Внутренняя среда</b>	<ol style="list-style-type: none"> <li>1. Гибкость и масштабируемость</li> <li>2. Ориентация на пользователей, недавно взаимодействующих с интернет-магазином</li> <li>3. Увеличение продаж</li> <li>4. Снижение расходов на проведение рекламных кампаний</li> <li>5. Автоматизированное формирование значений прогноза покупки</li> <li>6. Возможность применения результата в разных маркетинговых коммуникациях</li> </ol>	<ol style="list-style-type: none"> <li>1. Система не охватывает пользователей, которые не знакомы с интернет-магазином</li> <li>2. Зависимость точности модели от полноты и доступности исторических данных пользователей интернет-магазина</li> <li>3. Требуется частое обновление при изменении пользовательского поведения</li> <li>4. Ограниченная применимость для интернет-магазинов со средним сроком принятия решения менее двух дней и более 60 дней</li> </ol>
<b>Внешняя среда</b>	<ol style="list-style-type: none"> <li>1. Рост рынка электронной торговли</li> <li>2. Возможность интеграции с программами лояльности и CRM для комплексных коммуникационных стратегий</li> </ol>	<ol style="list-style-type: none"> <li>1. Ужесточение законодательства в области защиты персональных данных</li> <li>2. Угрозы кибербезопасности: риски утечки и несанкционированного доступа</li> <li>3. Снижение интереса к покупкам в интернет-магазинах</li> </ol>

Разрабатываемая СППР содержит множество положительных сторон. СППР прогнозирования поведения пользователей позволяет увеличить продажи и снизить расходы на проведение рекламных кампаний за счет проведения индивидуальных коммуникаций с пользователями интернет-магазина. Система направлена на получение значений прогноза для пользователей, которые недавно посещали интернет-магазин. Возникает ограничение, которое описывается невозможностью формирования значений прогноза для пользователей, еще не посещавших интернет-магазин, поскольку системы веб-аналитики, аккумулирующие данные о действиях пользователей в интернет-магазине, не содержат данных об несуществующих визитах и поведении пользователей вне интернет-магазина.

Благодаря наличию открытых API-интерфейсов, предоставляемых веб-аналитическими системами, становится возможной автоматизация процесса извлечения данных о посещениях пользователей интернет-магазина. Весь процесс формирования итоговых прогнозных значений может быть реализован в виде программного кода, что обеспечивает высокую степень автоматизации и минимизирует необходимость вмешательства человека. СППР прогнозирования покупательского поведения осуществляет обучение алгоритмов машинного обучения на основе исторических данных о посещаемости интернет-магазина. Полученные вероятностные оценки совершения покупки пользователем в течение заданного временного интервала могут применяться в различных маркетинговых коммуникациях.

Одним из ключевых инструментов маркетинговой коммуникации с пользователями является настройка рекламных кампаний, осуществляющих таргетинг с конкретными рекламными сообщениями. Дополнительно возможна реализация взаимодействия с определенным пользователем через сам интернет-магазин. Интегрировав СППР, становится возможным получение значений прогноза для

конкретного пользователя посредством API взаимодействия веб-сайта и СППР, что позволит динамически изменять контент веб-сайта или выводить отдельные баннеры с индивидуальной скидкой в рамках текущего визита пользователя.

Персонализированные коммуникации оказываются эффективными как для пользователей с низкой вероятностью совершения покупки, выявленной системой, так и для пользователей с высокой прогнозной вероятностью.

Из изложенного выше следует ограничивающий фактор системы, заключающийся в зависимости точности модели и эффективности всей системы от полноты и доступности исторических данных о пользователях интернет-магазина. Перед началом обучения система предпринимает попытку получить исторические данные о визитах пользователей. В случае отсутствия таких данных обучение системы становится невозможным.

Учитывая высокую изменчивость пользовательского поведения, система требует регулярного обновления для учета новых паттернов активности. Данная проблема может быть решена посредством регулярного дообучения алгоритмов машинного обучения на новых данных, которые ранее не использовались для обучения.

Система имеет важное ограничение, заключающееся в том, что ее эффективность напрямую зависит от продолжительности цикла принятия решения о покупке пользователем в интернет-магазине. Если срок принятия решения составляет менее двух дней, система сможет обучиться, однако практическая результативность ее применения будет ограниченной. Например, в случае интернет-магазинов, специализирующихся на продаже цветов, цикл принятия решения, как правило, измеряется часами или, в лучшем случае, несколькими днями.

Говоря о внешней среде, положительным влиянием на систему может стать рост рынка электронной торговли, что приведет к еще большему количеству интернет-магазинов и большей посещаемостью интернет-магазинов пользователями. Более того, система может быть интегрирована с программами лояльности или CRM-системами для дополнительных коммуникаций. Как правило, в CRM-системах содержится информация о контактных данных пользователей, а именно номера телефонов и электронные почты. Дополнительные контактные данные так же могут быть использованы при настройке рекламных кампаний или в SMS и Email коммуникациях.

Отрицательным влиянием на систему может быть ужесточение законодательства в области защиты персональных данных, угрозы кибербезопасности с рисками утечки данных и несанкционированным доступом, а также снижение интереса к совершению покупок в сети Интернет.

Таким образом, проведенная классификация и SWOT-анализ позволяют всесторонне охарактеризовать разрабатываемую СППР прогнозирования покупательского поведения с точки зрения функциональных, архитектурных и прикладных аспектов. Выявленные сильные стороны подтверждают практическую значимость и потенциал применения системы в задачах персонализированного маркетинга, тогда как обозначенные ограничения и внешние риски формируют требования к условиям ее эффективного использования и дальнейшего развития.

*Источник данных.* Существует множество способов извлечения данных о действиях пользователей: открытые и общедоступные наборы данных, системные логи веб-сайтов, данные систем веб-аналитики, а также собственные модули сбора данных. Наиболее распространенным инструментом накопления информации о пользователях являются системы веб-аналитики, которые сегодня установлены практически на каждом веб-сайте. Среди наиболее часто применяемых систем следует выделить Яндекс.Метрика и Google Analytics. Эти инструменты интегрируются на сайт в несколько простых шагов и обеспечивают автоматизированный сбор информации о

действиях пользователей с последующим хранением данных на специализированных серверных ресурсах.

В представляемой СППР предлагается использовать именно Яндекс.Метрику, поскольку данная система установлена более чем на 95 % российских веб-сайтов, что позволяет рассматривать ее в качестве единого источника данных для различных интернет-магазинов. Дополнительным преимуществом является наличие бесплатного API-интерфейса, который обеспечивает извлечение поведенческих данных пользователей за произвольный временной интервал. Таким образом, одним из ключевых условий функционирования СППР является наличие подключенного счетчика Яндекс.Метрики на веб-сайте, в противном случае система не сможет быть использована.

Между тем доступ к сырым пользовательским данным на уровне отдельных событий в Google Analytics требует активной интеграции с Google BigQuery [29]. Такая интеграция влечет за собой дополнительные инфраструктурные зависимости и ограничивает доступность данных периодом, следующим за моментом настройки интеграции, что ограничивает возможность использования исторических данных.

Следует также отметить, что в условиях санкционных ограничений использование Google Analytics в России стремительно сокращается, в том числе из-за риска штрафов со стороны Роскомнадзора, связанных с трансграничной передачей персональных данных.

Другие перечисленные методы сбора данных не позволяют выработать единый и приемлемый подход для различных интернет-магазинов. В частности, большинство компаний не осуществляют хранение системных логов и не располагают собственными модулями сбора данных, разработка которых требует значительных затрат и привлечения специалистов.

Дополнительно следует отметить технические ограничения веб-аналитической системы Яндекс.Метрика, заключающиеся в том, что данные о действиях пользователей доступны только по вчерашний день включительно. Это означает, что при формировании прогноза вероятности покупки на текущий день система будет использовать данные до вчерашнего дня. Таким образом, существует риск, что пользователь, которому система присвоит высокую вероятность совершения покупки, уже завершил транзакцию, что снижает рентабельность рекламных коммуникаций. С другой стороны, если цикл принятия решения превышает 60 дней, объем доступных данных о покупках может оказаться недостаточным для качественного обучения модели, что также снижает общую эффективность системы.

Указанное ограничение системы в наибольшей степени характерно для узкого класса интернет-магазинов, в которых процесс принятия решения о покупке существенно отличается от типичных сценариев электронной коммерции. К таким случаям относятся, в частности, онлайн-сервисы по продаже цветов, где решение принимается в течение короткого временного интервала, а также веб-ресурсы, связанные с рынком недвижимости, для которых характерен значительно более продолжительный цикл принятия решения.

Для извлечения данных о действиях пользователей используется Logs API, предоставляемый Яндекс.Метрика. Для каждого интернет-магазина извлекается один и тот же набор полей из API Яндекс.Метрика уровня визитов. В Таблице 3 представлен полный список извлекаемых полей из API Яндекс.Метрика [30].

Таблица 3 – Список выгружаемых полей из API Яндекс.Метрика  
Table 3 – 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 данных считается достаточным для формирования итогового признакового пространства, поскольку охватывает ключевые аспекты пользовательской активности в рамках визита пользователя. Для выявления наиболее подходящего признакового пространства с наибольшей точностью прогнозирования, были проведены серии экспериментов, описанные в работе [30].

В данной работе было сформировано 7 подмножеств, каждое из которых формировалось на основе определенного набора признаков или их комбинаций. В работе было описано формирование следующих наборов 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).

В ходе проведенного эксперимента было выявлено, что наиболее высокие показатели качества разработанной модели машинного обучения достигались за счет использования базовых признаков и событийных данных с их вариациями (п. 6). Значения показателя F-меры превышало значения 70 % по всем рассмотренным категориям интернет-магазинов, что свидетельствовало о практической применимости и эффективности предложенного подхода. В связи с этим, в рамках текущего исследования, разработанная СППР использует такое же признаковое пространство, что и в п. 6.

*Формирование признакового пространства.* В данном исследовании процесс формирования итогового признакового пространства будет описан более

детализировано, в отличие от [30]. Для анализа сложных технических систем и понимания той окружающей среды, в которой находится рассматриваемая система, необходимо использовать методы анализа, позволяющие рассматривать систему на уровне отдельных терминов в конкретной предметной области. Одним из таких методов является метод онтологического анализа, результатом которого является онтология.

Онтология представляет из себя иерархический законпорядок терминов между собой, которым можно формализовать набор определенных терминов рассматриваемой предметной области. Полученная онтология может послужить фундаментом для формирования признакового пространства для обучения моделей машинного обучения.

Предметной областью данного исследования является интернет-магазин, рассматриваемый как организационно-техническая система с тесно взаимосвязанными технологическими, экономическими и поведенческими процессами. Технологический уровень системы включает в себя веб-приложения, базы данных, системы веб-аналитики и интеграции с внешними системами. Организационный уровень охватывает логистические, маркетинговые и управленческие процессы, обеспечивающие функционирование интернет-магазина как единой системы.

На Рисунке 2 представлена концептуальная модель предметной области системы прогнозирования покупательского поведения пользователей интернет-магазинов [31]. Модель отражает ключевые сущности предметной области и связи между ними, обеспечивая формализованное описание процессов, происходящих в ходе взаимодействия пользователя с интернет-магазином.

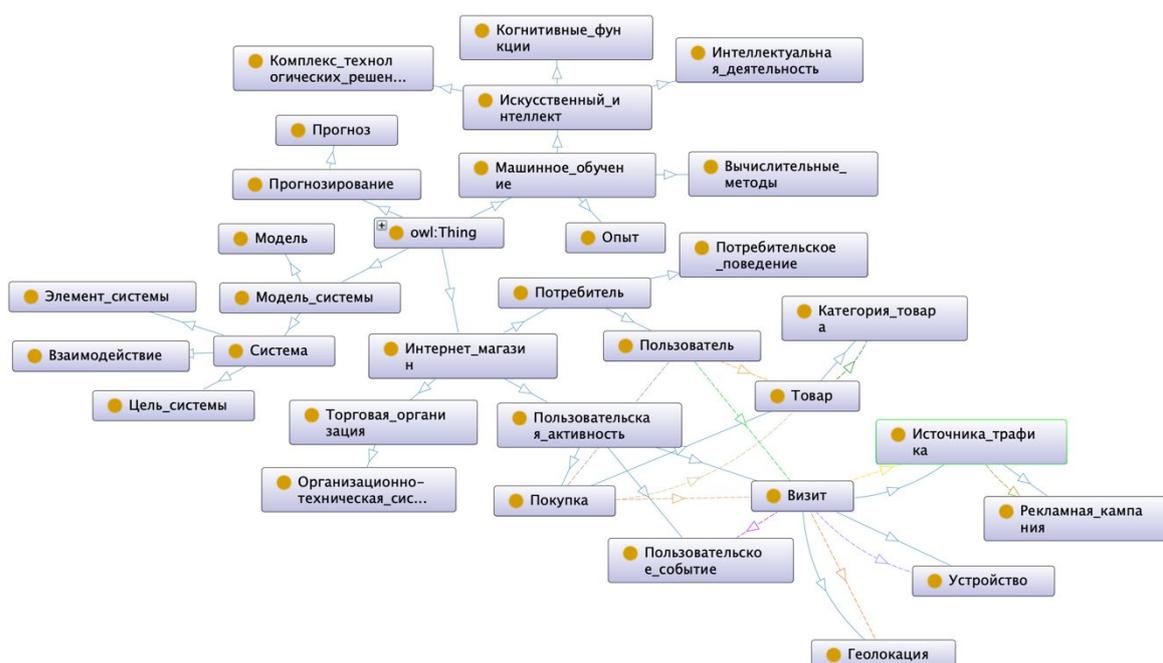


Рисунок 2 – Концептуальная модель предметной области системы прогнозирования покупательского поведения пользователей интернет-магазина

Figure 2 – Conceptual model of the subject domain of the system for predicting online shoppers purchasing behavior

Результаты современных исследований подтверждают применимость онтологического анализа в задачах в области электронной коммерции [32, 33]. Предлагаемая концептуальная модель интегрирует как структурные компоненты интернет-магазина, так и поведенческие аспекты пользователей, взаимодействующих с платформой. Онтология была разработана с использованием программного средства

Protégé, предоставляющего графический интерфейс для создания, визуализации и управления онтологическими структурами.

Онтологически обусловленная концептуальная модель использовалась для определения принципов отбора и структурирования поведенческих признаков, извлекаемых из данных Яндекс.Метрика. Такой подход позволяет достичь единой терминологии, повысить согласованность данных и обеспечить масштабируемость процесса формирования признакового пространства при интеграции различных интернет-магазинов с разными профилями направленности.

Итоговые признаки реализуются в виде проверок условий и операций над данными в программном коде на языке программирования Python. Итоговое количество условий является динамическим значением и зависит от количества доступных настроенных целей в интерфейсе Яндекс.Метрика конкретного интернет-магазина.

Яндекс.Метрика поддерживает функционал создания целей с настройкой по разным условиям. Настроенная цель инкрементально накапливает достижения определенного указанного условия, например, клик по кнопке, добавление товара в корзину, оформление заказа и многое другое. API Яндекс.Метрика также позволяет выгружать данные о текущих доступных настроенных целях конкретного интернет-магазина.

Для формирования признаков на основе настроенных целей в счетчике Яндекс.Метрика интернет-магазина используется набор из четырех условий. Поскольку количество доступных целей различается для каждого счетчика Яндекс.Метрика, то есть конкретного интернет-магазина, то итоговое количество условий масштабируется пропорционально их количеству. Одним из внутренних параметров СППР является процент используемых целей для анализа и по умолчанию данное значение равно 50 %. Так, при наличии 100 настроенных целей и значением в 50 % отбираемых целей из них сформируется 200 условий.

Выделяется 4 условия временных окон для расчета признаков со смещением в прошлое на несколько визитов пользователя для учета ретроспективных данных. Выделяется 2 условия агрегации покупок – расчет общей выручки по покупкам пользователя в рамках определенного визита и расчет общего количества покупок, совершенных за визит пользователя. Дополнительно реализованы 2 условия кодирования категориальных признаков, обеспечивающих корректное преобразование категориальных переменных в числовую форму для обучения модели.

Примерами таких условий могут быть:

1. Если признак является категориальным (номинальным) и участвует в обучении модели, то он преобразуется в числовой вид с помощью порядкового кодирования (OrdinalEncoder) по обучающей выборке.

2. Если для визита известен факт достижения цели в этом визите, то формируется бинарный признак «цель достигнута в текущем визите» для данной цели.

3. Если рассматривается визит пользователя, то для каждой отобранной цели формируются признаки «цель достигнута 2 визита назад» и «цель достигнута 3 визита назад».

*Описание предлагаемой системы.* В ходе исследования была разработана функциональная модель системы для отображения структуры и функций, а также потоков информации и материальных объектов, преобразуемых этими функциями. На Рисунке 3 изображена функциональная модель СППР прогнозирования покупательского поведения пользователей интернет-магазинов (диаграмма A-0).

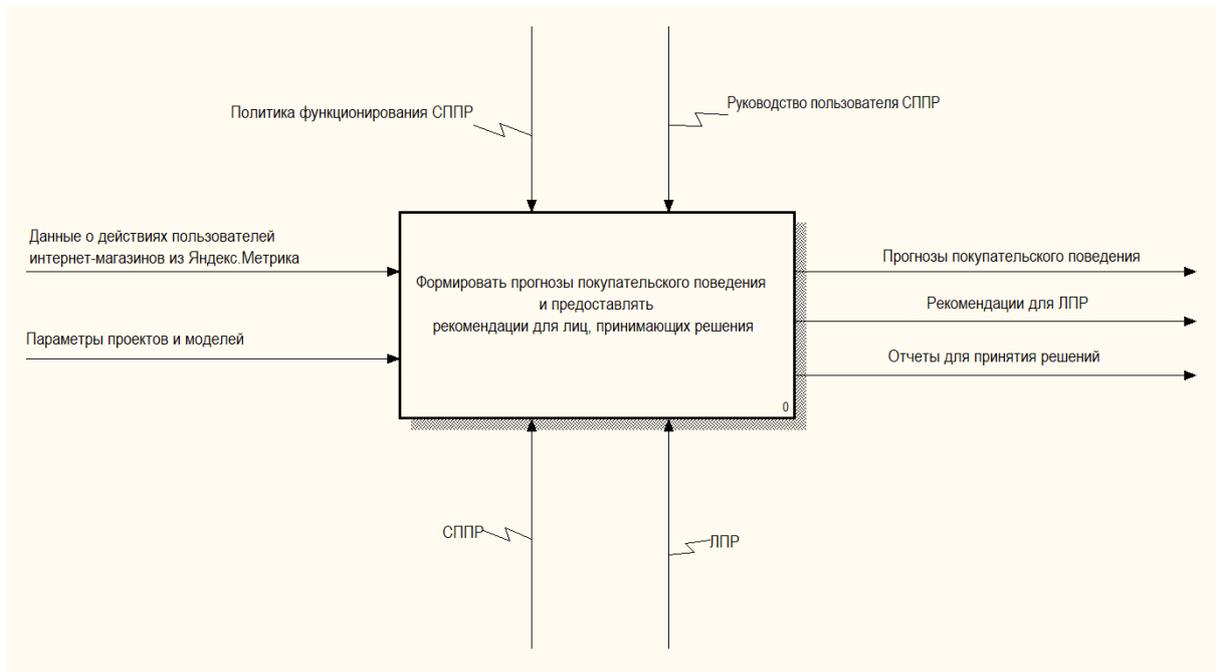


Рисунок 3 – Функциональная модель СППР прогнозирования покупательского поведения (диаграмма А-0)

Figure 3 – Functional model of the DSS for predicting consumer behavior (A-0 diagram)

Входным элементом системы являются данные о действиях пользователей интернет-магазинов, поступающие из сервиса веб-аналитики Яндекс.Метрика, а также параметры проекта и модели, задаваемые ЛПР в пользовательском интерфейсе системы. Управляющие воздействия формируются политикой функционирования СППР и руководством пользователя СППР.

Формирование прогнозов покупки для конкретного интернет-магазина описывается одним проектом и одной моделью, создаваемыми при запуске конкретного проекта через пользовательский интерфейс. Данные сущности содержат в себе базовую информацию по настройкам проекта и параметрам модели, используемым при обучении алгоритмов машинного обучения.

В нижней части диаграммы показаны механизмы, обеспечивающие выполнение функций: сама СППР и ЛПР. Выходами модели являются прогнозы покупательского поведения, рекомендации для ЛПР и отчеты, предназначенные для поддержки процесса принятия решений. Дополнительно была разработана функциональная модель СППР уровня А0, представленная на Рисунке 4.

На диаграмме представлена декомпозиция функции верхнего уровня, отражающая последовательность формирования прогнозов покупательского поведения и рекомендаций для ЛПР. Модель включает в себя четыре основных функции: сбор и структурирование данных о действиях пользователей, формирование обучающих и тестовых наборов данных, разработку и обучение модели прогнозирования, а также формирование прогнозов и рекомендаций.

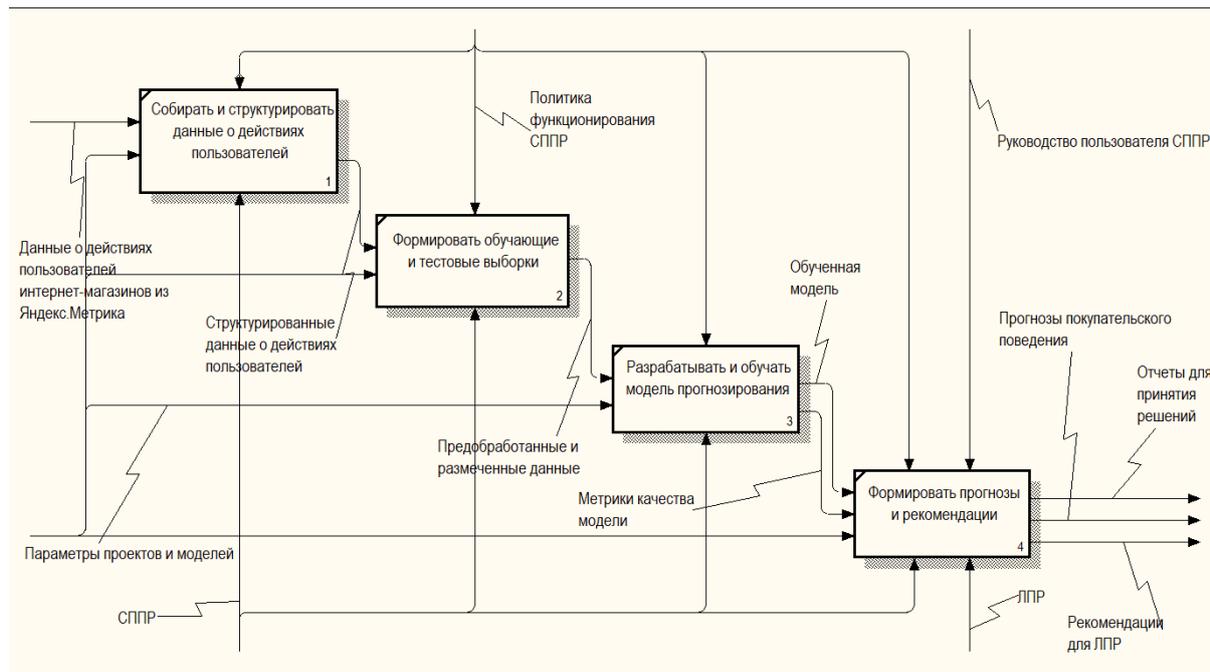


Рисунок 4 – Функциональная модель СППР по прогнозированию покупательского поведения (диаграмма A0)

Figure 4 – Functional model of the DSS for predicting consumer behavior (A0 diagram)

Предлагаемая система реализована на языке программирования Python версии 3.10 с использованием библиотек pandas, scikit-learn, numpy и инструментов оркестрации Apache Airflow версии 2.9.1 и Kubernetes версии 1.28. Хранение событийных данных о действиях пользователей организовано в СУБД ClickHouse, являющимся эффективным решением при анализе и хранении больших объемов данных [34]. Данные в ClickHouse загружаются посредством ETL-обработки (Extract-Transform-Load) через API интерфейс Яндекс.Метрика. Для хранения метаданных моделей и настроек проекта, метрик качества моделей и других вспомогательных данных используется СУБД PostgreSQL, демонстрирующая высокую эффективность как транзакционная реляционная СУБД с оперативной доступностью данных [35].

Описаны 2 условия формирования целевой переменной, которые описывают логику определения факта достижения цели пользователем в текущем и прогнозируемом периодах. Процесс обучения модели машинного обучения регулируется тремя условиями, определяющими разбиение данных, параметры обучения и выбор оптимального порога классификации. Для отбора данных и построения прогнозов применяются три условия фильтрации, а для разделения пользователей по степени готовности совершения покупки реализовано три условия сегментации. Контроль корректности и устойчивости вычислений обеспечивается тремя условиями валидации и обработки ошибок.

ЛПР получает результаты прогнозирования в табличном виде через интерфейс системы. В таблице представлен список пользователей интернет-магазина с оценкой вероятности совершения покупки в заданном горизонте. Пользователи отнесены к сегментам по степени готовности к покупке. Таким образом, ЛПР понимает, у кого вероятность покупки наиболее высокая, а у кого низкая в ближайший период. Полученные данные позволяют планировать коммуникации с клиентами через маркетинговые каналы, в том числе Яндекс.Директ, SMS и email-рассылки.

В качестве основного метода машинного обучения был выбран градиентный бустинг. Для реализации модели использовалась библиотека CatBoost, что обусловлено ее способностью эффективно работать с разнородными типами признаков, устойчивостью к масштабированию данных и встроенной поддержкой категориальных переменных. Обучение модели выполнялось с использованием стандартных механизмов оптимизации, предоставляемых библиотекой, с применением ранней остановки для предотвращения переобучения.

Система обладает высокой производительностью и масштабируемостью. Процессы обучения и обработки данных выполняются через инструменты Apache Airflow и Kubernetes. Одной из основных причин использования Kubernetes в предлагаемой системе является поддержка автоматического масштабирования вычислительных ресурсов в зависимости от входящей нагрузки на вычислительные узлы. Инфраструктурная часть системы реализована через сервис Яндекс.Облако, предоставляющий облачные вычислительные ресурсы и инструменты. Apache Airflow выступает в роли инструмента, позволяющего прописать единую инструкцию по выполнению определенных блоков программного кода, которые запускаются в определенной указанной последовательности в назначенное время.

Система функционирует в распределенной среде с одним вычислительным узлом конфигурации Kubernetes, поддерживающей горизонтальное масштабирование в диапазоне от 1 до 6 узлов. Каждый узел располагает вычислительными ресурсами в 96 CPU и 576 ГБ оперативной памяти. Ежедневно в ночное время осуществляется автоматический запуск задач на обновление прогнозов для активных проектов. Для обеспечения надежности предусмотрены механизмы повторного выполнения задач в Apache Airflow в случае ошибок или недоступности вычислительных узлов.

*Пользовательский интерфейс системы.* После прохождения регистрации пользователем открывается возможность создания проекта для обучения системы. Пользователю необходимо предоставить доступ к API к конкретному аккаунту Яндекса, который имеет доступ к нужному счетчику Яндекс.Метрики интернет-магазина. Это достигается прохождением OAuth авторизации, для получения нужных уровней доступа к аккаунту пользователя. Следующим этапом является указание названия создаваемого проекта и выбор счетчика Яндекс.Метрика, данные которого необходимо использовать для обучения системы. Дополнительно необходимо указать наименование цели, отвечающей за факт совершения покупки пользователем (Рисунок 5).



Рисунок 5 – Выбор счетчика Яндекс.Метрика и выбор цели для прогнозирования  
 Figure 5 – Selection of the Yandex.Metrica counter and the target goal for prediction

Следующим этапом является заполнение основных параметров для запуска процесса формирования прогнозов (Рисунок 6). Среди параметров выделяется указание горизонта прогнозирования в днях, описывающего значение, в течение которого ожидается факт совершения покупки пользователем.

При необходимости доступно указание количества формируемых сегментов, разделенных по степени готовности совершения покупки пользователями. По умолчанию количество сегментов равно 5, что позволяет ЛПР получать достаточно детализированные данные, избегая избыточной фрагментации при большом количестве сегментов, и потери информативности при их малом количестве.

Реализован механизм доступа к полученным прогнозам посредством API интерфейса, предоставляемого системой. Данный интерфейс создан с использованием фреймворка FastAPI на языке Python. Входными параметрами запроса на получение прогноза является уникальный идентификатор проекта и идентификатор пользователя, присваиваемый веб-аналитической системой Яндекс.Метрика.

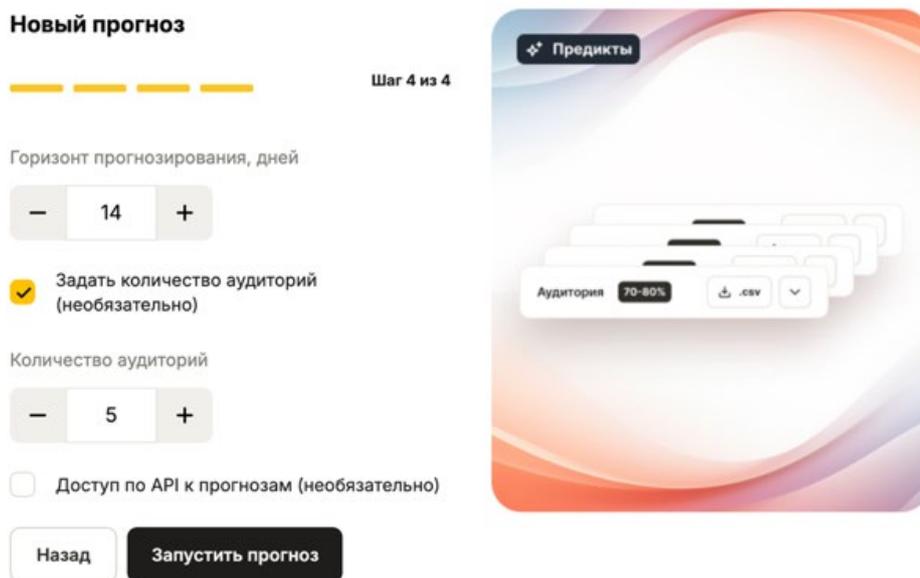


Рисунок 6 – Заполнение необходимых параметров для запуска процесса формирования прогнозов

Figure 6 – Configuration of the required parameters for initiating the forecast generation process

После запуска прогноза начинается процесс извлечения данных из Яндекс.Метрика, формирования обучающих выборок и обучения алгоритмов. После успешного обучения системы пользователю становится доступна страница проекта, которая содержит информацию о метриках качества обученной модели (Precision, Recall, F-score) и сформированных сегментов, разделенных по степени готовности совершения покупки пользователями (Рисунок 7). Пользователю доступна информация о размере сегментов, диапазоне полученных вероятностей формирования покупки, а также типе сегмента с характерным названием.

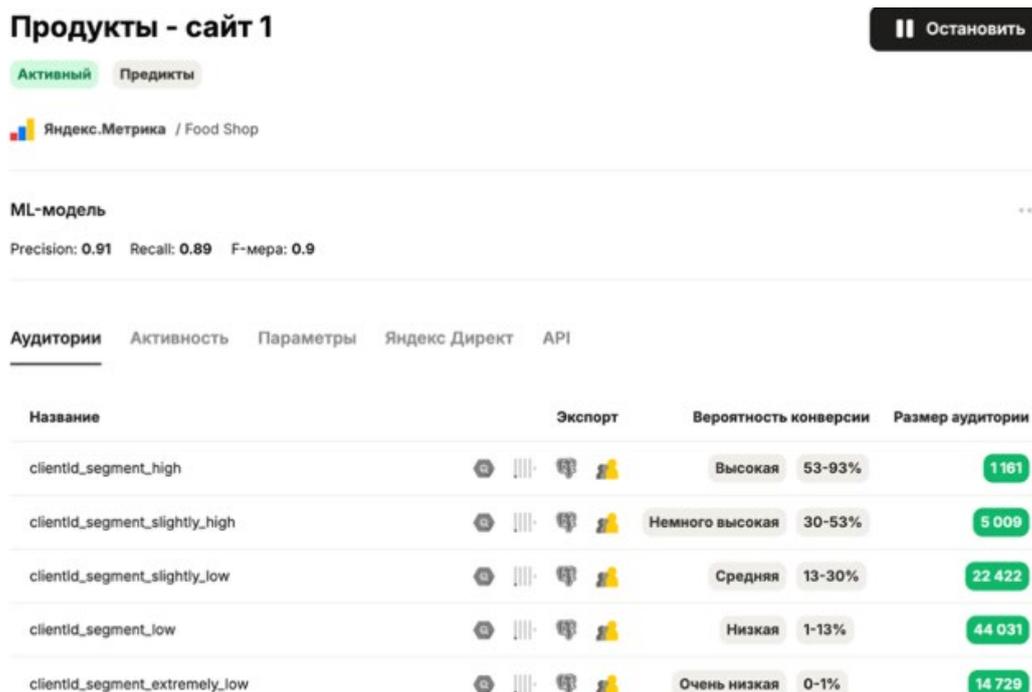


Рисунок 7 – Пользовательский интерфейс созданного проекта в системе  
Figure 7 – User interface of the created project in the system

ЛПР в роли пользователя интерфейса доступна возможность получения данных сегментов посредством настройки ETL процесса выгрузки данных в одну из поддерживаемых СУБД: PostgreSQL, ClickHouse, BigQuery. Результаты также могут быть экспортированы в виде CSV файла, содержащего информацию о значении прогноза от 0 до 1 по уникальному идентификатору пользователя (Рисунок 8). Прогноз формируется на основе последнего визита пользователя на веб-сайт.

client_id	visit_id	datetime	prediction	cluster
"1761674435326333153"	"6381433101519945920"	2025-10-29 00:00:35+03:00	0.6656257504922485	high
"176016720453895489"	"6381452074872471926"	2025-10-29 00:01:46+03:00	0.7110035671040339	high
"1761674525735932831"	"6381456353376600386"	2025-10-29 00:02:04+03:00	0.730568391275103	high
"1761646968503680938"	"6381465097137291407"	2025-10-29 00:02:37+03:00	0.7184528951752651	high
"1761584172920247747"	"6381477180982952262"	2025-10-29 00:03:23+03:00	0.5781990775617976	high
"1758403056101381870"	"6381479654990545148"	2025-10-29 00:03:32+03:00	0.8500937969103972	high
"1761674623685665996"	"6381482651157790757"	2025-10-29 00:03:44+03:00	0.729502758897375	high
"1761674698597477575"	"6381502164299677969"	2025-10-29 00:04:57+03:00	0.7108411497790182	high
"1761674711912094219"	"6381505560334041173"	2025-10-29 00:05:11+03:00	0.7097284944416918	high
"1761665964381786858"	"6381513396902953337"	2025-10-29 00:05:41+03:00	0.7239991422394437	high
"1757145969297093702"	"6381549205496004942"	2025-10-29 00:07:57+03:00	0.6455999684268973	high
"1761035978488915066"	"6381548991293685985"	2025-10-29 00:07:57+03:00	0.6364619827554616	high
"1761675020779323998"	"6381585857405517972"	2025-10-29 00:10:18+03:00	0.7114132000372998	high
"1761674498376500879"	"6381586707074777246"	2025-10-29 00:10:21+03:00	0.8975925054935155	high
"1724307735831828392"	"6381592716171018349"	2025-10-29 00:10:44+03:00	0.5826272286507254	high

Рисунок 8 – Пример формируемых данных сегмента по прогнозу совершения покупки  
Figure 8 – Example of generated segment data based on purchase probability prediction

Кроме того, предусмотрена интеграция с сервисом Яндекс.Аудитории, что позволяет передавать сформированные сегменты для их последующего использования в системе Яндекс.Директ. Это позволяет настраивать рекламные кампании на определенный сегмент пользователей интернет-магазина, что обеспечивает повышение эффективности маркетинговых коммуникаций с пользователями.

### Результаты и обсуждение

Предложенная СППР была апробирована в реальных условиях на данных нескольких интернет-магазинов. Для эксперимента были выбраны четыре интернет-магазина, компании-владельцы которых предоставили доступ к данным Яндекс.Метрики. Среди них:

- Интернет-магазин по продаже специальной одежды;
- Интернет-магазин по продаже автозапчастей;
- Два интернет-магазина, специализирующихся на доставке продуктов питания и готовых блюд.

В Таблице 4 представлены характеристики выборок данных, использованных для обучения, тестирования и валидации моделей. По умолчанию система извлекает данные сессионной активности пользователей за последние 3 месяца. В качестве модели машинного обучения использовался алгоритм градиентного бустинга с использованием библиотеки CatBoost.

Таблица 4 – Характеристики выборок данных для обучения и тестирования СППР  
 Table 4 – Characteristics of the datasets used for DSS training and testing

Интернет-магазин	Размер обучающей выборки	Размер тестовой выборки	Размер валидационной выборки	Лучшая итерация обучения модели	Время обучения (сек.)
Спецодежда	191 042	23 880	23 880	5 460	536
Автозапчасти	636 338	79 542	79 542	1 007	1 196
Продукты (сайт 1)	4 320 193	540 024	540 024	2 999	12 070
Продукты (сайт 2)	1 354 344	169 293	169 293	2 949	3 732

Из полученных данных можно сделать вывод, что исследуемые интернет-магазины существенно различаются по объему данных, характеризующих пользовательскую активность. Это выражается в размере обучающих выборок. Наименьший объем наблюдений зафиксирован у интернет-магазина специализированной одежды (191 тыс. записей), тогда как наибольший объем представлен в данных одного из магазинов по продаже продуктов питания (более 4,3 млн наблюдений). Такой разброс объясняется спецификой предметной области и различиями в пользовательском трафике.

Следует обратить внимание на время обучения моделей. Результаты показывают, что время обучения положительно коррелирует с объемом обрабатываемых данных. Наибольшая продолжительность составила 3,5 часа (12 070 секунд) при 2 999 итерациях обучения, что соответствует самому крупному набору данных. В то же время наименьшее время обучения составило порядка 9 минут (536 секунд) при 5 460 итерациях, что отражает меньший объем входных данных.

Полный набор данных для каждого интернет-магазина был разделен на три части: обучающую, тестовую и валидационную. Данные были распределены в пропорции 80 %, 10 %, 10 %.

10 % и 10 % соответственно. Тестовый набор применялся для мониторинга качества модели в процессе обучения, тогда как валидационный набор предназначался для финальной проверки эффективности модели на ранее не встречавшихся данных.

В Таблице 3 представлены результаты качества моделей после обучения для рассматриваемых интернет-магазинов.

Таблица 5 – Метрики качества прогнозных моделей для интернет-магазинов

Table 5 – Quality metrics of predictive models for online stores

Интернет-магазин	Precision (класс 0)	Recall (класс 0)	F-мера (класс 0)	Precision (класс 1)	Recall (класс 1)	F-мера (класс 1)	F-мера (общая)
Спецодежда	0,98	0,99	0,99	0,86	0,50	0,63	0,81
Автозапчасти	0,99	0,99	0,99	0,80	0,68	0,73	0,86
Продукты (сайт 1)	0,93	0,95	0,94	0,89	0,84	0,86	0,90
Продукты (сайт 2)	0,98	0,98	0,98	0,69	0,69	0,69	0,83

Полученные модели после обучения оцениваются как соответствующие требованиям качества при условии, что значение общей F-меры превышает 65 %. Согласно представленным результатам, все модели для исследуемых интернет-магазинов преодолели данный порог, что подтверждает эффективность разработанной СППР. Следует подчеркнуть, что значения общей F-меры выше 80 % могут быть рассмотрены как высокое качество прогноза не только в рамках решаемой задачи, но и в целом в практике построения моделей машинного обучения. Наилучшие результаты были достигнуты для одного из интернет-магазинов по продаже продуктов, модель которого обучалась в течение 3,5 часов.

Анализ эффективности по показателю Precision для положительного класса показывает, что по всем интернет-магазинам, за исключением одного продуктового сайта, данный показатель превышает 80 %. Несмотря на это, итоговое значение F-меры для данного магазина составило 83 %, что также соответствует уровню надежного прогнозирования.

Полученные экспериментальные результаты подтверждают устойчивость и универсальность разработанной СППР при применении к интернет-магазинам различной направленности и с существенно различающимися объемами данных. Достижение значений общей F-меры выше 80 % во всех экспериментах свидетельствует о корректности выбранного подхода к формированию признакового пространства и используемого алгоритма машинного обучения.

Следует отметить, что различия в значениях метрик качества между интернет-магазинами обусловлены спецификой предметной области, характером пользовательского трафика и особенностями цикла принятия решения о покупке. В частности, для интернет-магазинов с более длительным и осознанным процессом выбора товаров (например, автозапчасти) наблюдается более сбалансированное соотношение показателей Precision и Recall для положительного класса, тогда как в сегментах с импульсным поведением пользователей возможны смещения в сторону повышения точности при снижении полноты.

Результаты также демонстрируют, что увеличение объема обучающей выборки положительно влияет на итоговое качество прогнозирования, однако сопровождается ростом вычислительных затрат. Это подтверждает целесообразность использования масштабируемой облачной инфраструктуры и инструментов оркестрации, применяемых в разработанной СППР.

## Заключение

В ходе исследования была разработана автоматизированная СППР прогнозирования покупательского поведения пользователей интернет-магазинов, продемонстрировавшая высокую точность прогноза на примере рассмотренных интернет-магазинов. В отличие от существующих решений, представленный подход сочетает в себе высокую точность, возможность масштабирования и работу с интернет-магазинами различной направленности.

Проведен онтологический анализ для формирования концептуальной модели рассматриваемой предметной области, что послужило основой для формирования признакового пространства. Разработана функциональная модель системы, отражающая ключевые процессы сбора, обработки и анализа данных, а также формирование прогнозов и рекомендаций для ЛПР.

Разработанная система объединяет онтологический подход к описанию предметной области, функциональную модель и модуль машинного обучения, реализованный на языке Python с использованием библиотек Pandas, Scikit-learn и CatBoost.

Использование облачной инфраструктуры Яндекс.Облако, инструментов оркестрации Apache Airflow и Kubernetes обеспечили высокую производительность системы, ее устойчивость и гибкость при обработке больших объемов данных. Автоматическое масштабирование вычислительных узлов системы позволило выполнять регулярные прогнозные расчеты без вмешательства человека.

Дополнительно представлен пользовательский интерфейс системы, в котором ЛПР имеет возможность настройки нужного проекта по формированию прогнозов, а также интеграции с внешними сервисами и СУБД для применения прогнозов в маркетинговых коммуникациях с пользователями и принятия важных стратегических решений.

Экспериментальные исследования на реальных данных интернет-магазинов различных тематик подтвердили универсальность предложенного решения и стабильные показатели качества прогнозов, где значения F-меры во всех случаях превысили 80 %.

Разработанная СППР формирует сегменты пользователей по степени готовности совершения покупки и обеспечивает возможность экспорта во внешние маркетинговые системы для повышения эффективности рекламных кампаний и качества управленческих решений.

Полученные результаты демонстрируют достижение поставленной цели исследования и подтверждают практическую значимость предложенного подхода для задач прогнозирования и персонализации в электронной коммерции.

## 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*. 2020;31:697–715. <https://doi.org/10.1007/s12525-020-00448-x>
2. 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>
3. Ketipov R., Angelova V., Doukovska L., Schnalle R. Predicting User Behavior in e-Commerce Using Machine Learning. *Cybernetics and Information Technologies*. 2023;23(3):89–101. <https://doi.org/10.2478/cait-2023-0026>
4. 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>

5. Xu J., Wang J., Tian Y., et al. 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>
6. Abhichandani D., Vadrevu N.R.T., Doshi P., Shrivastava Sh. Predicting Online Purchases Using Six Machine Learning Models Based on Customer Demographics. In: *2025 6<sup>th</sup> International Conference on Data Intelligence and Cognitive Informatics (ICDICI), 09–11 July 2025, Tirunelveli, India*. IEEE; 2025. P. 1787–1792. <https://doi.org/10.1109/icdici66477.2025.11135228>
7. 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>
8. Shi X. The application of machine learning in online purchasing intention prediction. In: *ICBDC '21: Proceedings of the 6<sup>th</sup> International Conference on Big Data and Computing, 22–24 May 2021, Shenzhen, China*. New York: ACM; 2021. P. 21–29. <https://doi.org/10.1145/3469968.3469972>
9. Hamami F., Muzakki A. Machine learning pipeline for online shopper intention classification. *AIP Conference Proceedings*. 2021;2329(1). <https://doi.org/10.1063/5.0043452>
10. Liu Ch.-J., Huang T.-Sh., Ho P.-T., Huang J.-Ch., Hsieh Ch.-T. Correction: Machine learning-based e-commerce platform repurchase customer prediction model. *PLoS ONE*. 2024;19(12). <https://doi.org/10.1371/journal.pone.0315518>
11. Hesvindrati 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>
12. Prasad A.K., M D.K., Macedo V.D.J., Mohan B.R., N A.P. Machine Learning Approach for Prediction of the Online User Intention for a Product Purchase. *International Journal on Recent and Innovation Trends in Computing and Communication*. 2023;11(1s):43–51. <https://doi.org/10.17762/ijritcc.v11i1s.5992>
13. 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>
14. Zhou S., 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>
15. 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>
16. Tanvir A.-A., Khandokar I.A., Islam A.K.M.M., Islam S., Shatabda S. A gradient boosting classifier for purchase intention prediction of online shoppers. *Heliyon*. 2023;9(4). <https://doi.org/10.1016/j.heliyon.2023.e15163>
17. Liu Y., Tian Y., Xu Y., et al. TPGN: A time-preference gate network for e-commerce purchase intention recognition. *Knowledge-Based Systems*. 2021;220. <https://doi.org/10.1016/j.knosys.2021.106920>
18. Liu Zh., Zhang Y., Abedin M.Z., et al. Profit-driven fusion framework based on bagging and boosting classifiers for potential purchaser prediction. *Journal of Retailing and Consumer Services*. 2024;79. <https://doi.org/10.1016/j.jretconser.2024.103854>
19. 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>
20. Tokuç A.A., Dağ T. Customer Purchase Intent Prediction using Feature Aggregation on E-Commerce Clickstream Data. In: *2024 8<sup>th</sup> International Artificial Intelligence and Data Processing Symposium (IDAP), 21–22 September 2024, Malatya, Turkiye*. IEEE; 2024. P. 1–5. <https://doi.org/10.1109/idap64064.2024.10711144>
  21. Wang H., Wang L., Zhu F. E-Commerce User Behavior Analysis and Prediction Based on Artificial Neural Network and Data Mining. In: *2024 IEEE 7<sup>th</sup> Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 20–22 September 2024, Chongqing, China*. IEEE; 2024. P. 583–586. <https://doi.org/10.1109/itnec60942.2024.10733243>
  22. Kumari L., Bhattacharjee K., Sharma N., Kumar Sh., Kumari A. Machine Learning Models in Customer Behaviour Prediction: A Comparative Analysis. In: *2024 7<sup>th</sup> International Conference on Contemporary Computing and Informatics (IC3I), 18–20 September 2024, Greater Noida, India*. IEEE; 2024. P. 957–959. <https://doi.org/10.1109/ic3i61595.2024.10828637>
  23. Al-Otaibi Y.D. Enhancing e-Commerce Strategies: A Deep Learning Framework for Customer Behavior Prediction. *Engineering, Technology & Applied Science Research*. 2024;14(4):15656–15664.
  24. Deniz E., Çökekoğlu Bülbül S. Predicting Customer Purchase Behavior Using Machine Learning Models. *Information Technology in Economics and Business*. 2024;1(1):1–6. <https://doi.org/10.69882/adba.iteb.2024071>
  25. Lv Q. E-Commerce Big Data Analysis and User Behavior Prediction Algorithm Based on Deep Learning. In: *2024 3<sup>rd</sup> International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), 29–31 July 2024, Bristol, United Kingdom*. IEEE; 2024. P. 219–224. <https://doi.org/10.1109/aiars63200.2024.00046>
  26. 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>
  27. Fu Z., Han J. Research on Marketing Strategies of Pinduoduo based on SWOT Analysis. *SHS Web of Conferences*. 2023;154. <https://doi.org/10.1051/shsconf/202315402009>
  28. Budiman S., Ahidin U. Optimizing digital marketing strategies for Indonesian retail companies through SWOT analysis and strategic development. *Journal of Industrial and Logistics Management*. 2025;9(1):86–98. <https://doi.org/10.30988/jmil.v9i1.1612>
  29. Chauleva B., Capeska Bogatinoska D., Karadimce A. Optimizing Customer Journey through Advanced Analytics Techniques over Google Analytics 4 Data in Google BigQuery. *WSEAS Transactions on Computers*. 2024;23:336–346. <https://doi.org/10.37394/23205.2024.23.33>
  30. Святлов Р.С. Прогнозирование покупательского поведения пользователей интернет-магазинов на основе событийных данных. *Моделирование, оптимизация и информационные технологии*. 2025;13(4). <https://doi.org/10.26102/2310-6018/2025.51.4.064>  
Svyatov R.S. Forecasting e-commerce user purchase behavior based on event Data. *Modeling, Optimization and Information Technology*. 2025;13(4). (In Russ.). <https://doi.org/10.26102/2310-6018/2025.51.4.064>

31. Святлов Р.С. Онтологический подход к прогнозированию покупательского поведения пользователей в электронной коммерции. *Моделирование, оптимизация и информационные технологии*. 2026;14(2). <https://doi.org/10.26102/2310-6018/2026.53.2.018>  
Svyatov R.S. Ontology-based approach to predicting consumer purchasing behavior in e-commerce. *Modeling, Optimization and Information Technology*. 2026;14(2). (In Russ.). <https://doi.org/10.26102/2310-6018/2026.53.2.018>
32. Bhutani P., Baranwal Sh.K., Jain S. Semantic Framework for Facilitating Product Discovery. In: *ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021, 25–27 February 2021, Delhi, India*. 2021. P. 30–36.
33. Kim H. Developing a Product Knowledge Graph of Consumer Electronics to Manage Sustainable Product Information. *Sustainability*. 2021;13(4). <https://doi.org/10.3390/su13041722>
34. Schulze R., Schreiber T., Yatsishin I., Dahimene R., Milovidov A. ClickHouse – lightning fast analytics for everyone. *Proceedings of the VLDB Endowment*. 2024;17(12):3731–3744. <https://doi.org/10.14778/3685800.3685802>
35. Schneider M., Martínez D. A comparative benchmark analysis of transactional and analytical performance in PostgreSQL and MySQL. *International Journal of Modern Computer Science and IT Innovations*. 2025;2(10):51–63.

#### ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

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

*Статья поступила в редакцию 15.02.2026; одобрена после рецензирования 10.03.2026; принята к публикации 23.03.2026.*

*The article was submitted 15.02.2026; approved after reviewing 10.03.2026; accepted for publication 23.03.2026.*