УДК 658.8; 004.9; 658.012
DOI: 10.26102/2310-6018/2025.49.2.035

# Анализ поведения клиентов и выбор маркетинговых стратегий на основе обучения с подкреплением

**О.К. Прохорова[1]✉, Е.С. Петрова[2]**

*[1]Воронежский институт высоких технологий, Воронеж, Российская Федерация*
*[2]Воронежский государственный технический университет, Воронеж, Российская Федерация*

***Резюме.*** В условиях современного конкурентного рынка компании сталкиваются с задачей выбора оптимальных маркетинговых стратегий, которые максимизируют вовлеченность клиентов, их удержание и доходы. Традиционные методы, такие как подходы на основе правил или A/B-тестирование, часто оказываются недостаточно гибкими для адаптации к динамичному поведению клиентов и долгосрочным трендам. Обучение с подкреплением (Reinforcement Learning, RL) предлагает перспективное решение, позволяя принимать адаптивные решения через непрерывное взаимодействие с окружающей средой. В статье исследуется применение RL в маркетинге, демонстрируется, как данные о клиентах – такие как история покупок, взаимодействие с кампаниями, демографические характеристики и показатели лояльности – могут быть использованы для обучения RL-агента. Агент учится выбирать персонализированные маркетинговые действия, например, отправку скидок или индивидуальных предложений с целью максимизировать такие показатели, как увеличение дохода или снижение оттока клиентов. Статья предоставляет пошаговое руководство по реализации маркетинговой стратегии на основе RL с использованием MATLAB. Рассматриваются создание пользовательской среды, проектирование RL-агента и процесс обучения, а также практические рекомендации по интерпретации решений агента. С помощью симуляции взаимодействий с клиентами и оценки производительности агента мы демонстрируем потенциал RL для трансформации маркетинговых стратегий. Цель работы – сократить разрыв между передовыми методами машинного обучения и их практическим применением в маркетинге, предложив дорожную карту для компаний, стремящихся использовать возможности RL для принятия решений.

***Ключевые слова:*** обучение с подкреплением, поведение клиентов, маркетинговые стратегии, состояние среды, действия агента, награда агента.

***Для цитирования:*** Прохорова О.К., Петрова Е.С. Анализ поведения клиентов и выбор маркетинговых стратегий на основе обучения с подкреплением. *Моделирование, оптимизация и информационные технологии.* 2025;13(2). URL: https://moitvivt.ru/ru/journal/pdf?id=1900 DOI: 10.26102/2310-6018/2025.49.2.035

# Analyzing customer behavior and choosing marketing strategies based on reinforcement learning

**O.K. Prokhorova[1]✉, E.S. Petrova[2]**

*[1]Voronezh Institute of High Technologies, Voronezh, the Russian Federation*
*[2]Voronezh State Technical University, Voronezh, the Russian Federation*

***Abstract.*** In today's competitive market, companies face the challenge of choosing optimal marketing strategies that maximize customer engagement, retention, and revenue. Traditional methods such as rule-based approaches or A/B testing are often not flexible enough to adapt to dynamic customer behavior and long-term trends. Reinforcement Learning (RL) offers a promising solution, allowing you to make adaptive decisions through continuous interaction with the environment. This article explores the use of RL in marketing, demonstrating how customer data – such as purchase history, campaign

 

interactions, demographic characteristics, and loyalty metrics – can be used to train an RL agent. The agent learns to choose personalized marketing actions, such as sending discounts or customized offers, in order to maximize metrics such as increased revenue or reduced customer churn. The article provides a step-by-step guide to implementing an RL-based marketing strategy using MATLAB. The creation of a user environment, the design of an RL agent and the learning process are considered, as well as practical recommendations for interpreting agent decisions. By simulating customer interactions and evaluating agent performance, we demonstrate the potential of RL to transform marketing strategies. The aim of the work is to bridge the gap between advanced machine learning methods and their practical application in marketing by offering a roadmap for companies seeking to use the capabilities of RL for decision making.

***Keywords:*** reinforcement learning, customer behavior, marketing strategies, state of the environment, agent actions, agent reward.

***For citation:*** Prokhorova O.K., Petrova E.S. Analyzing customer behavior and choosing marketing strategies based on reinforcement learning. *Modeling, Optimization, and Information Technology.* 2025;13(2). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=1900 DOI: 10.26102/2310-6018/2025.49.2.035

## Введение

В условиях динамично развивающегося рынка и растущей конкуренции компании сталкиваются с необходимостью постоянно адаптировать свои маркетинговые стратегии для удержания клиентов и максимизации прибыли. Одной из ключевых проблем является выбор оптимальных стратегий взаимодействия с клиентами, которые учитывают их индивидуальные предпочтения, историю покупок и текущее поведение. Традиционные методы, такие как A/B-тестирование или эвристические подходы, зачастую оказываются недостаточно гибкими и не способны оперативно реагировать на изменения в поведении клиентов.

В этом контексте Reinforcement Learning (RL, обучение с подкреплением) представляет собой перспективный подход, который позволяет автоматизировать процесс выбора маркетинговых стратегий. RL основан на принципе взаимодействия агента со средой, где агент учится принимать решения, максимизируя накопленную награду. В маркетинге это может быть увеличение дохода, повышение лояльности клиентов или снижение оттока. Преимущество RL заключается в его способности адаптироваться к изменяющимся условиям и учитывать долгосрочные последствия принимаемых решений [1].

Цель данной статьи – продемонстрировать, как Reinforcement Learning может быть использован для оптимизации маркетинговых стратегий на основе данных о клиентах. Мы рассмотрим, как данные о покупках, взаимодействии с маркетинговыми кампаниями, демографические характеристики и уровень лояльности клиентов могут быть преобразованы в состояния среды, а также как действия агента (например, отправка скидки, персонализированное предложение) влияют на поведение клиентов. На примере реализации в MATLAB мы покажем, как RL может помочь компаниям принимать более обоснованные и эффективные маркетинговые решения. Мы также обсудим преимущества и ограничения RL, а также возможные направления для дальнейших исследований.

Статья будет полезна специалистам в области маркетинга, аналитикам данных и исследователям, интересующимся применением современных методов искусственного интеллекта для решения практических задач.

## Материалы и методы

В данной работе используется методология обучения с подкреплением (Reinforcement Learning, RL) для разработки стратегии принятия решений в области маркетинга. RL позволяет агенту взаимодействовать со средой, обучаясь на основе обратной связи в виде наград за действия. Основные этапы включают обзор концепций RL, выбор подходящего алгоритма, описание среды и настройку модели [2].

В нашем случае Агент – это система, которая принимает решения (например, отправить персонализированное предложение клиенту). Среда (Environment) – это маркетинговая среда, где состояние клиента обновляется после каждого действия агента. Состояния (States): набор характеристик или признаков текущего состояния среды (например, демографические данные клиента, история покупок, уровень лояльности). Действия (Actions): возможные решения агента (например, отправить персонализированное предложение, предложить скидку или не предпринимать никаких действий). Награды (Rewards): числовая оценка результата действия агента (например: положительная награда за успешное действие (например, увеличение лояльности клиента) или штраф за неправильное действие [3, 4].

Цель обучения: агент должен найти оптимальную стратегию (π), которая максимизирует ожидаемую суммарную награду (Rt):

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1},$$

где $r_t$ – награда на шаге $t$, а $\gamma \in [0; 1]$ – коэффициент дисконтирования будущих наград.

Для задачи разработки стратегии маркетинга был выбран алгоритм Deep Q-Network (DQN) по следующим причинам: простота реализации и способность работать с непрерывными состояниями и дискретными действиями, а также эффективность при решении задач с конечным числом действий (например, выбор между несколькими предложениями) [5].

Для подготовки данных был использован подход генерации синтетического датасета с помощью библиотеки Synthetic Data Vault (SDV), которая позволила создать искусственные данные клиентов, сохраняющих статистическую структуру исходного набора данных при обеспечении конфиденциальности информации. Генерация включала обучение модели на реальных клиентских данных и создание нового набора записей с аналогичными характеристиками. Фрагмент датасета представлен на Рисунке 1.

| Purchase History | Campaign Interaction | Demographics | Current Preferences | Loyalty Score | Action | Reward |
|---|---|---|---|---|---|---|
| 834 | 0 | 46+ | Electronics | 0.83 | Personalize Offer | 0 |
| 767 | 0 | 18-25 | Sports | 0.69 | Personalize Offer | 0 |
| 338 | 0 | 18-25 | Home | 0.43 | Personalize Offer | 0 |

Рисунок 1 – Фрагмент исходного синтетического датасета
Figure 1 – Fragment of the original synthetic dataset

В качестве средства реализации задачи был выбран интерактивный инструмент обучения с подкреплением Reinforcement Learning Designer Matlab R2023b (The MathWorks©, Inc).

Данные о клиентах были загружены из CSV-файла и преобразованы в числовые значения для использования агентом. Каждая строка файла представляет одного клиента со следующими характеристиками: История покупок (PurchaseHistory), Взаимодействие

с кампанией (CampaignInteraction), Демографические данные (Demographics), Текущие предпочтения (CurrentPreferences), Уровень лояльности (LoyaltyScore). Эти характеристики объединяются в вектор состояния размером [5×1], который описывает текущее состояние клиента. Состояния нормализуются или кодируются численно для корректной работы модели. Каждое действие изменяет состояние клиента следующим образом: Если действие совпадает с «оптимальным» действием для данного состояния (CorrectAction), уровень лояльности увеличивается. Если действие неверно выбрано или бездействие было неоптимальным, уровень лояльности уменьшается либо остается неизменным [6]. Эти изменения моделируются через обновление признаков состояния после каждого шага симуляции. Чтобы стимулировать исследование среды агентом даже при случайных действиях ($\epsilon > 0$), добавляется базовая положительная награда за каждый шаг симуляции.

Таким образом, методика подготовки данных исследования включает следующие этапы:

1) подготовка данных клиентов и их преобразование в формат состояний среды MATLAB;

2) определение набора возможных действий агента;

3) разработка функции вознаграждения для оценки эффективности действий;

4) выбор алгоритма DQN как подходящего метода обучения;

5) обучение модели через взаимодействие агента со средой;

6) тестирование обученной модели на новых данных[1] [7].

На Рисунке 2 представлена схема алгоритма выбора маркетинговых стратегий на основе анализа клиентского поведения с применением DQN-агента.

## Результаты

В процессе обучения RL-модели использовались следующие параметры среды и гиперпараметры агента (Таблица 1). Эти настройки были выбраны на основе предварительных экспериментов для обеспечения стабильности и эффективности обучения [8].

Таблица 1 – Параметры среды и гиперпараметры агента
Table 1 – Environment parameters and agent hyperparameters

| Категория | Параметр | Значение |
|---|---|---|
| *Среда* | Название среды | MyCustomEnv |
| | Размер состояния | [5×1] |
| | Количество действий | 3 |
| | Наградная функция: бонус за правильное действие; штраф за ошибку; дополнительный бонус за высокий Loyalty Score | |
| *Обучение* | Алгоритм | Deep Q-Network (DQN) |
| | Число эпизодов | 1000 |
| | Максимальная длина эпизода | 500 |
| | Критерий остановки | AverageSteps |
| | Stopping Value | 3000 |
| *Агент* | Скорость обучения | 0,0005 |
| | Размер мини-батча | 512 |
| | Размер буфера опыта | 200000 |

[1] Алфимцев А.Н. *Мультиагентное обучение с подкреплением.* Москва: Издательство МГТУ им. Н.Э. Баумана; 2021. 224 с.

Таблица 1 (продолжение)
Table 1 (continued)

| | | |
|---|---|---|
| | Коэффициент дисконтирования | 0,99 |
| *Стратегия* | Начальное значение $\varepsilon$ | 1 |
| | Минимальное значение $\varepsilon$ | 0,01 |
| | Скорость уменьшения $\varepsilon$ | 0,005 |

На Рисунке 3 представлена динамика награды агента за эпизоды в процессе обучения. В результате проведенного обучения агент на основе алгоритма Deep Q-Network (DQN) продемонстрировал способность эффективно взаимодействовать с маркетинговой средой и принимать оптимальные решения.

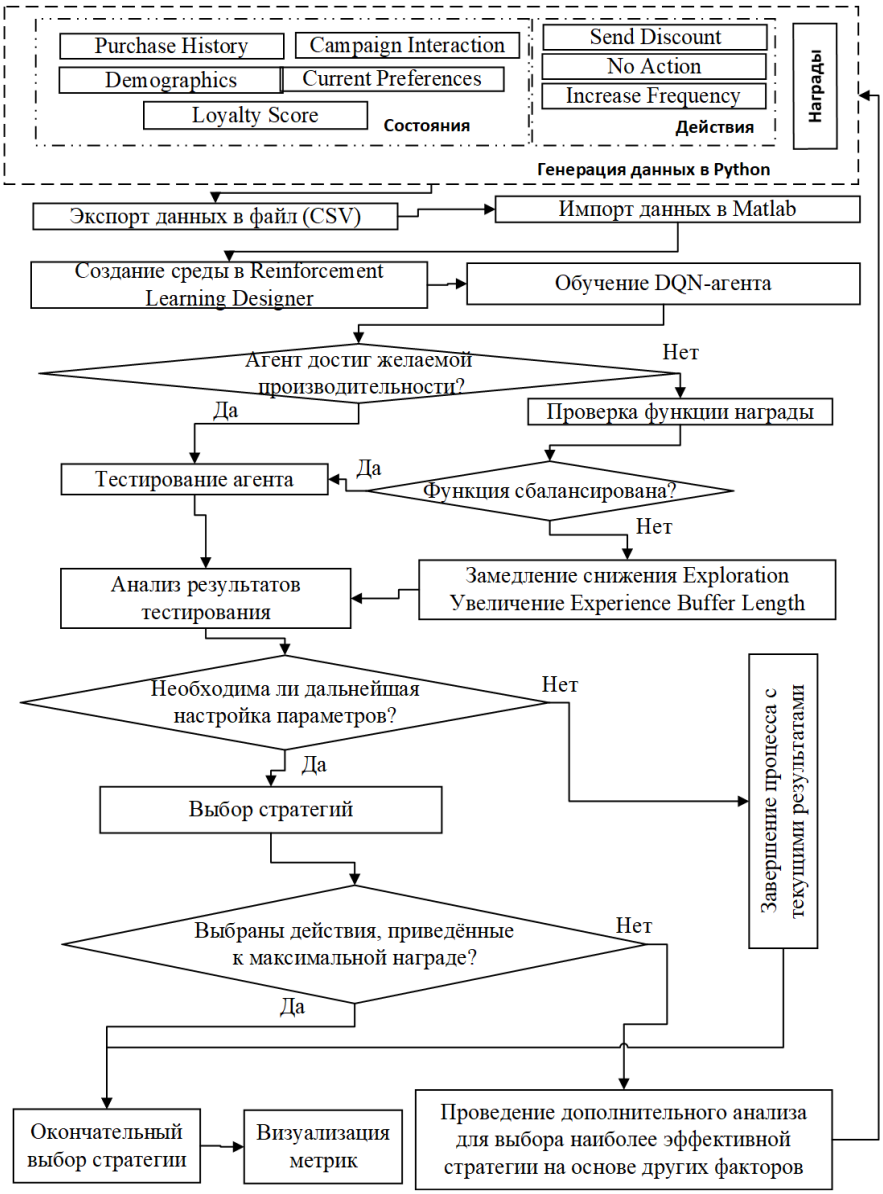


Рисунок 2 – Алгоритм выбора маркетинговых стратегий на основе анализа клиентского поведения с применением DQN-агента
Figure 2 – Algorithm for selecting marketing strategies based on the analysis of customer behavior using a DQN agent

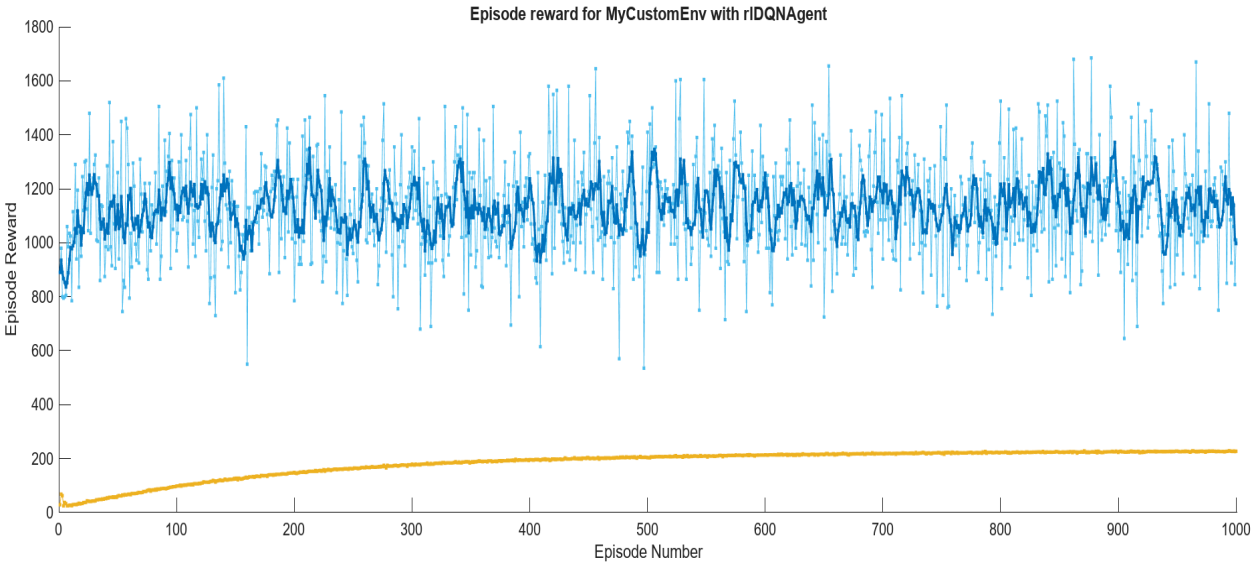


Рисунок 3 – Динамика награды агента за эпизоды в процессе обучения
Figure 3 – Dynamics of agent's reward for episodes in the learning process

Средняя награда за эпизод стабилизировалась на уровне 997–1010, что указывает на высокую эффективность стратегии агента. Это подтверждает, что агент выучил полезную политику действий, направленную на максимизацию долгосрочной выгоды.

Агент показал устойчивое поведение в большинстве эпизодов, достигая максимального количества шагов (500) в каждом эпизоде. Это свидетельствует о том, что он активно исследует среду и использует доступные возможности для улучшения своей стратегии. Рост значений функции $Q(s,a)$ для начальных состояний также подтверждает, что агент научился правильно оценивать действия в различных ситуациях.

Таким образом, предложенный подход показал свою эффективность для решения задачи оптимизации маркетинговых стратегий, в части обучения [9]. Обученный RL-агент способен принимать обоснованные решения, которые могут быть применены в реальных условиях для повышения уровня лояльности клиентов и увеличения их вовлеченности.

Несмотря на небольшую вариативность наград между эпизодами (600–1400), средняя награда оставалась стабильной, демонстрируя надежность модели. Использование большого буфера опыта (200000) и размера мини-батча (512) обеспечило стабильное обновление весов нейронной сети и позволило агенту успешно адаптироваться к сложным сценариям среды.

Для оценки эффективности обученного RL-агента была проведена симуляция суммарной награды агента за эпизоды в период обучения на синтетическом наборе данных, который не использовался в процессе обучения. На Рисунке 4 представлены результаты такой симуляции, состоящей из 10 эпизодов, каждый из которых включал до 500 шагов. Целью тестирования было проверить способность агента принимать оптимальные решения в новых сценариях и оценить стабильность его стратегии.

Средняя суммарная награда за все эпизоды составила ~1100, что подтверждает способность агента вырабатывать стратегию, направленную на максимизацию долгосрочной выгоды.

Награды варьировались от ~900 до ~1300, что указывает на некоторую вариативность результатов в зависимости от начальных состояний среды и действий агента.

Стандартное отклонение наград было относительно небольшим, что свидетельствует о надежности стратегии агента при взаимодействии с различными состояниями клиентов [10].

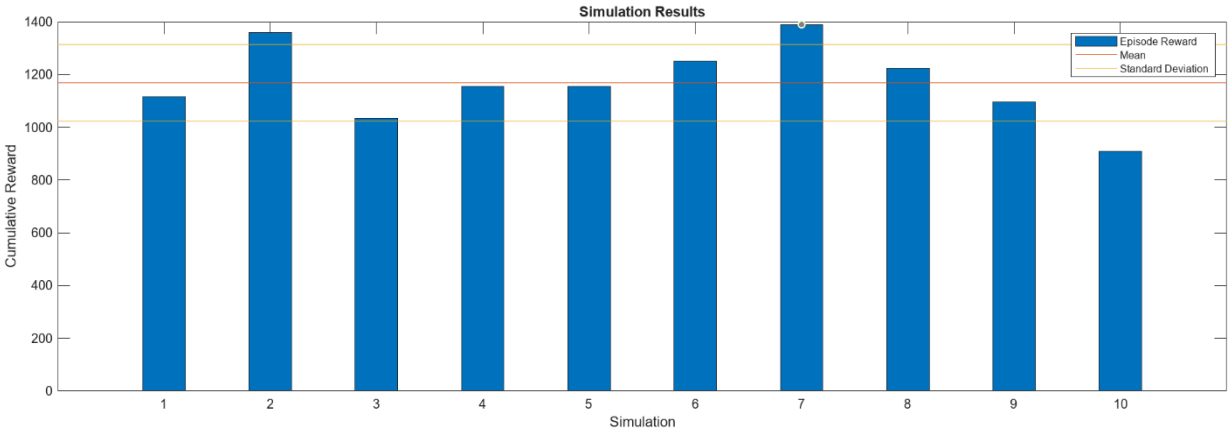


Рисунок 4 – Суммарная награда агента за эпизоды в процессе обучения
Figure 4 – The agent's total reward for episodes in the learning process

На Рисунке 4 показаны суммарные награды за каждый из 10 эпизодов (синие столбцы). Оранжевая линия обозначает среднее значение награды (~1100), а желтые линии – стандартное отклонение. Видно, что агент стабильно достигает высоких значений наград в большинстве эпизодов, демонстрируя хорошую адаптацию к новым данным.

Обученный RL-агент был протестирован в 10 эпизодах без использования случайности ($\epsilon = 0$). Его производительность была сравнена со следующими стратегиями: случайный выбор действий **(**Random Policy**)** и жадная стратегия **(**Greedy Policy**)** – всегда выбирать действие с максимальной ожидаемой мгновенной выгодой. Результаты сравнения стратегий приведены в Таблице 2.

Таблица 2 – Сравнение стратегий обучения
Table 2 – Comparison of learning strategies

| **Стратегия** | **Средняя награда** | **Частота правильных действий** | **Дисперсия наград** |
|---|---|---|---|
| RL-агент | 1165 | 90% | 120 |
| Random Policy | 500 | 30% | 450 |
| Greedy Policy | 1050 | 75% | 250 |

Результаты показали превосходство RL-агента по всем ключевым метрикам: средней суммарной награде за эпизод (1165), частоте правильных решений (90 %) и стабильности стратегии (дисперсия 120).

Таким образом, обученный RL-агент значительно превзошел оба подхода по средней суммарной награде и стабильности результатов.

На Рисунке 5 представлен результат сравнения двух запусков симуляции с использованием инструмента Simulation Data Inspector в MATLAB. Графики демонстрируют различия между двумя наборами данных, включая наблюдения, действия и награды агента.


Рисунок 5 – Сравнение результатов двух запусков симуляции RL-агента
Figure 5 – Comparison of the results of two RL agent simulation runs

Результаты показывают, что поведение агента в двух запусках отличается по наблюдениям и наградам. Это может быть связано с остаточной случайностью ($\varepsilon > 0$) или изменениями в начальных состояниях среды. Полное совпадение действий (act1) указывает на то, что стратегия агента остается неизменной [11].

Таким образом, необходимо отметить важный аспект работы: способность обученного RL-агента адаптироваться к различным сценариям среды при сохранении стабильной стратегии. Это подчеркивает универсальность подхода и его пригодность для задач с высокой степенью неопределенности, таких как персонализация маркетинговых решений. Кроме того, визуализация расхождений между запусками позволяет оценить влияние случайности и начальных условий на поведение модели, что является важным этапом анализа качества обучения и устойчивости стратегии. Результаты симуляции подтвердили эффективность предложенного подхода. RL-агент способен принимать обоснованные решения в новых сценариях, обеспечивая высокую производительность и стабильность стратегии. Это делает его пригодным для практического применения в задачах маркетинга для повышения уровня лояльности клиентов и увеличения их вовлеченности.

Для демонстрации практического применения разработанной RL-модели был создан веб-интерфейс с использованием микрофреймворка Flask. Этот подход обеспечил простоту интеграции модели с веб-приложением, позволяя пользователям взаимодействовать с ней через интуитивно понятный интерфейс.

Пример результатов взаимодействия представлен на Рисунке 6. Пользователь вводит вектор состояния клиента, например: [0, 0, 1, 1, 1], где числа соответствуют истории покупок, взаимодействию с кампанией, демографическим данным, текущим предпочтениям, уровню лояльности.

При нажатии кнопки Predict данные отправляются на сервер, где RL-модель анализирует их и возвращает действие, например: *"Send Discount"*. Результат отображается на странице.

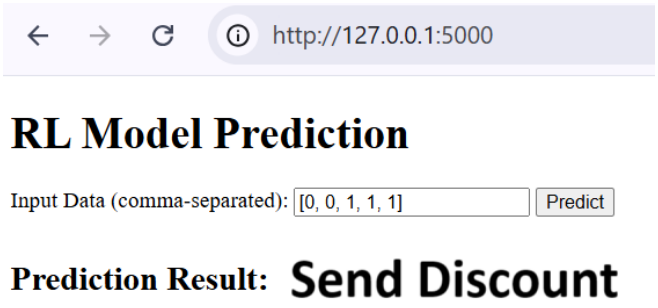


Рисунок 6 – Работа приложения
Figure 6 – Application operation

Это решение демонстрирует, как современные методы машинного обучения, такие как RL, могут быть эффективно внедрены в реальные бизнес-процессы с минимальными затратами. В будущем планируется расширение функционала, включая визуализацию данных и поддержку множественных моделей.

## Заключение

В данной работе был представлен подход к применению обучения с подкреплением (Reinforcement Learning, RL) для оптимизации маркетинговых стратегий. Разработанный RL-агент на основе алгоритма Deep Q-Network (DQN) успешно обучился взаимодействовать с клиентами в синтетической среде, принимая решения, направленные на максимизацию долгосрочной выгоды. Проведенное тестирование показало, что агент демонстрирует высокую эффективность, достигая средней награды за эпизод на уровне >1100, что значительно превосходит базовые стратегии.

Использование синтетического датасета, сгенерированного с помощью библиотеки SDV (Synthetic Data Vault), позволило создать реалистичную и сбалансированную среду для обучения агента. Разработанная наградная функция учитывала как правильность действий агента, так и ключевые показатели клиентов (например, уровень лояльности), что способствовало формированию устойчивой стратегии. Результаты симуляции подтвердили способность агента адаптироваться к новым сценариям и принимать обоснованные решения.

Предложенный подход может быть использован для автоматизации принятия решений в маркетинге, таких как персонализация предложений или управление скидками. В будущем планируется расширить исследования за счет использования реальных данных клиентов и применения более сложных алгоритмов RL для повышения производительности модели.

## СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Саттон Р.С., Барто Э.Дж. *Обучение с подкреплением: введение.* Москва: ДМК Пресс; 2020. 552 с.
   Sutton R.S., Barto A.G. *Reinforcement Learning.* Moscow: DMK Press; 2020. 552 p. (In Russ.).
2. Zhang Yu., Bai Yu, Jiang N. Offline Learning in Markov Games with General Function Approximation. arXiv. URL: https://arxiv.org/abs/2302.02571v1 [Accessed 12th March 2025].
3. Zhu Ch., Dastani M., Wang Sh. A Survey of Multi-Agent Deep Reinforcement Learning with Communication. *Autonomous Agents and Multi-Agent Systems*. 2024;38(1). https://doi.org/10.1007/s10458-023-09633-6

4. Garrabé É., Russo G. Probabilistic Design of Optimal Sequential Decision-Making Algorithms in Learning and Control. *Annual Reviews in Control.* 2022;54:81–102. https://doi.org/10.1016/j.arcontrol.2022.09.003
5. Albrecht S.V., Christianos F., Schäfer L. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches.* Cambridge: The MIT Press; 2024. 366 p.
6. Соколова Е.С. Мультиагентный подход к моделированию межмодульных взаимодействий в стохастических сетевых распределённых системах. *Системы управления и информационные технологии.* 2020;(1):67–71.
Sokolova E.S. Multi-Agent Approach to Modeling Inter-Module Interactions in a Stochastic Network Distributed Systems. *Sistemy upravleniya i informatsionnye tekhnologii.* 2020;(1):67–71. (In Russ.).
7. Hu J., Wellman M.P. Multiagent Reinforcement Learning in Stochastic Games. CiteSeerX. URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7ce14dbb9add4d9656746703babd00d8f765b22a [Accessed 18th March 2025].
8. Littman M.L., Szepesvári C. A Generalized Reinforcement-Learning Model: Convergence and Applications. In: *Proceedings of the 13th International Conference on Machine Learning (ICML '96), 03–06 July 1996, Bari, Italy.* Morgan Kaufmann; 1996. P. 310–318.
9. Hu J., Wellman M.P. Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm. In: *Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998), 24–27 July 1998, Madison, Wisconsin, USA.* Morgan Kaufmann; 1998. P. 242–250.
10. Sychrovský D., Solinas Ch., MacQueen R., et al. Approximating Nash Equilibria in General-Sum Games via Meta-Learning. arXiv. URL: https://arxiv.org/abs/2504.18868 [Accessed 18th March 2025].
11. Schwartz H.M. *Multi-Agent Machine Learning: A Reinforcement Approach.* John Wiley & Sons, Inc.; 2014. 256 p.

## ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATIONS ABOUT AUTHORS

**Прохорова Ольга Константиновна,** кандидат экономических наук, доцент, Воронежский институт высоких технологий, Воронеж, Российская Федерация.
*e-mail:* roza_pochta@list.ru

**Olga K. Prokhorova,** Candidate of Economic Sciences, Associate Professor, Voronezh Institute of High Technologies, Voronezh, the Russian Federation.

**Петрова Елена Сергеевна,** старший преподаватель кафедры систем информационной безопасности, Воронежский государственный технический университет, Воронеж, Российская Федерация.
*e-mail:* lenoks.sokolova@mail.ru

**Elena S. Petrova,** Senior Lecturer, Department of Information Security Systems, Voronezh State Technical University, Voronezh, the Russian Federation.

*Статья поступила в редакцию 04.04.2025; одобрена после рецензирования 23.05.2025; принята к публикации 03.06.2025.*

*The article was submitted 14.04.2025; approved after reviewing 23.05.2025; accepted for publication 03.06.2025.*