

**СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS,
MANAGEMENT AND PROCESSING OF INFORMATION**

DOI: <https://doi.org/10.60797/itech.2025.8.1>

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ФАКТОРОВ, ВЛИЯЮЩИХ НА СТАТУС ЗАКАЗА КЛИЕНТА

Научная статья

Баранова А.П.^{1,*}, Миролюбова А.А.²

²ORCID : 0000-0003-3785-0538;

^{1,2} Ивановский государственный химико-технологический университет, Иваново, Российская Федерация

* Корреспондирующий автор (ann.brnv24[at]gmail.com)

Аннотация

В данной работе проводится интеллектуальный анализ данных интернет-магазина техники с целью выявления ключевых факторов, оказывающих влияние на статус заказа клиента. Исследование включает анализ различных характеристик клиентов и их заказов, таких как пол, возраст, стоимость заказа, участие в программе лояльности и другие параметры. Для оценки взаимосвязей между этими характеристиками и статусом заказа проведен расчет коэффициентов корреляции, что позволило выявить наиболее значимые факторы, влияющие на процесс оформления и выполнения заказов. Результаты данного анализа могут быть использованы для оптимизации бизнес-процессов интернет-магазина, повышения эффективности логистики, улучшения прогнозирования спроса и минимизации рисков, связанных с отменой или задержкой заказов.

Ключевые слова: интеллектуальный анализ данных, статус заказа, корреляция, машинное обучение, прогнозирование заказов.

INTELLIGENT ANALYSIS OF FACTORS AFFECTING THE STATUS OF A CUSTOMER'S ORDER

Research article

Baranova A.P.^{1,*}, Mirolubova A.A.²

²ORCID : 0000-0003-3785-0538;

^{1,2} Ivanovo State University of Chemistry and Technology, Ivanovo, Russian Federation

* Corresponding author (ann.brnv24[at]gmail.com)

Abstract

This work conducts an intelligent analysis of data from an online technology store in order to identify key factors that influence the status of a customer's order. The study includes an analysis of various customer characteristics and their orders, such as gender, age, order value, participation in loyalty programmes, and other parameters. To evaluate the relationships between these characteristics and order status, correlation coefficients were calculated, which made it possible to identify the most significant factors influencing the order placement and fulfilment process. The results of this analysis can be used to optimise the business processes of the online store, increase the efficiency of logistics, improve demand forecasting, and minimise the risks associated with order cancellations or delays.

Keywords: intelligent data analysis, order status, correlation, machine learning, order prediction.

Введение

В условиях стремительного развития электронной коммерции и цифровизации бизнес-процессов особое значение приобретает задача эффективного управления заказами клиентов. Одним из ключевых аспектов в этой области является анализ факторов, влияющих на статус заказа, что позволяет оптимизировать бизнес-процессы, повысить удовлетворённость клиентов и снизить издержки. Современные методы интеллектуального анализа данных, в частности машинное обучение и глубокое обучение, предоставляют широкие возможности для решения подобных задач.

Ранее для анализа факторов, влияющих на заказы, применялись преимущественно методы регрессионного анализа и традиционные статистические подходы. Однако современные исследования демонстрируют эффективность использования автоматизированных моделей машинного обучения, которые позволяют выявлять сложные нелинейные зависимости между переменными и обеспечивать более высокую точность прогнозирования.

Целью данной работы является построение и обоснование моделей машинного обучения для анализа влияния различных факторов на статус заказа клиента, а также сравнение эффективности различных подходов. В работе особое внимание уделяется корректному выбору методов, этапам построения модели и интерпретации полученных результатов.

Методы и принципы исследования

Описание данных

В исследовании использовался набор данных, содержащий информацию о заказах в интернет-магазине техники за 2023–2024 гг. Для каждого заказа были доступны следующие признаки:

1. Пол клиента (мужской/женский).
2. Возраст клиента.
3. Участие в программе лояльности (да/нет).

4. Способ оплаты (онлайн/наличными).
5. Сумма заказа.
6. Количество товаров в заказе.
7. География доставки (город/регион).
8. Статус заказа (успешно выполнен/отменён/возвращён).

Постановка задачи

Задача формулируется как задача классификации: на основе имеющихся признаков необходимо предсказать статус заказа. Статус заказа рассматривается как целевая переменная с тремя классами.

Предварительная обработка данных

Пропущенные значения обработаны методом медианного заполнения.

Категориальные признаки преобразованы методом однократного кодирования.

Данные разделены на обучающую (70%) и тестовую (30%) выборки.

Выбор и обоснование методов

Для решения задачи классификации были выбраны следующие модели:

Логистическая регрессия — это метод классификации, который использует логистическую функцию для прогнозирования вероятности принадлежности наблюдения к определённому классу [5, С. 77].

Случайный лес — ансамблевый метаалгоритм, предназначенный для улучшения стабильности и точности алгоритмов машинного обучения, используемых в задачах классификации и регрессии [3], а также позволяющий выявлять сложные взаимосвязи и обладающий высокой обобщающей способностью.

Градиентный бустинг — современный ансамблевый алгоритм, который широко применяется для задач классификации и регрессии, поскольку обеспечивает высокую точность, устойчивость к переобучению и хорошую работу с данными, содержащими сложные зависимости [1, С. 264].

К-ближайших соседей — выбран для сравнения, так как часто используется для задач классификации с небольшим количеством признаков.

Глубокая нейронная сеть — использовалась для оценки возможности выявления сложных нелинейных зависимостей.

Выбор моделей обусловлен необходимостью сравнения как простых, так и сложных алгоритмов, их интерпретируемости, устойчивости к переобучению и способности работать с разными типами данных.

Оценка качества моделей

Для оценки качества моделей использовались следующие метрики:

1. Ассигасу (точность).
2. Оценка F1 для каждого класса.
3. Матрица ошибок (матрица смещения).
4. ROC-AUC (для бинарных классификаций).

Для предотвращения переобучения использовалась кросс-валидация (5-кратная).

Основные результаты

Корреляционный анализ

В ходе исследования был проведён комплексный анализ факторов, влияющих на статус заказа клиента интернет-магазина техники. На первом этапе выполнен корреляционный анализ с использованием коэффициента Крамера, который показал, что наиболее существенные связи наблюдаются между статусом заказа и такими признаками, как участие в программе лояльности (0,32), способ оплаты (0,28) и сумма заказа (0,27). Согласно шкале Чеддока, эти значения соответствуют слабой или умеренной связи [4], что указывает на ограниченность простого корреляционного анализа для выявления всех закономерностей.

Сравнение моделей

Для более глубокого анализа были построены и сравнены различные модели машинного обучения: логистическая регрессия, случайный лес, градиентный бустинг, метод k-ближайших соседей и глубокое обучение. Сравнение проводилось по метрикам точности (ассигасу) и F1-оценки для каждого класса (успешно выполнен, отменён, возвращён).

Таблица 1 - Сравнение эффективности моделей машинного обучения для классификации статуса заказа

DOI: <https://doi.org/10.60797/itech.2025.8.1.1>

Модель	Точность	Оценка F1 (успешно)	Оценка F1 (отмена)	Оценка F1 (возврат)	Примечания
Логистическая регрессия	0,71	0,78	0,65	0,60	Хорошая интерпретируемость, базовый уровень
Случайный лес	0,77	0,82	0,71	0,68	Высокая устойчивость к переобучению
Градиентный бустинг	0,79	0,84	0,74	0,70	Лучшая точность, хорошо работает с данными
К-ближайших соседей	0,69	0,76	0,62	0,58	Чувствителен к масштабу признаков

Модель	Точность	Оценка F1 (успешно)	Оценка F1 (отмена)	Оценка F1 (возврат)	Примечания
Глубокое обучение	0,78	0,83	0,72	0,69	Требуется больше данных и вычислений

Лучшие результаты продемонстрировал градиентный бустинг с общей точностью 0,79 и наивысшими F1-оценками по всем классам, что объясняется его способностью выявлять сложные нелинейные зависимости и устойчивостью к переобучению. Глубокое обучение показало сопоставимые результаты (точность 0,78), однако требует большего объёма данных и вычислительных ресурсов. Модели случайного леса и логистической регрессии также показали достойные результаты, но уступили по точности ансамблевым методам.

Важность признаков

Анализ важности признаков (по модели случайного леса и градиентного бустинга) показал, что наибольшее влияние на статус заказа оказывают:

1. Сумма заказа.
2. Участие в программе лояльности.
3. Способ оплаты.
4. География доставки.

Обсуждение

Полученные результаты исследования подтверждают, что для анализа факторов, влияющих на статус заказа клиента, целесообразно использовать современные ансамблевые методы машинного обучения и глубокое обучение, поскольку они способны выявлять значимые закономерности и тенденции [2], сложные и нелинейные взаимосвязи между признаками, которые не обнаруживаются простыми статистическими методами или корреляционным анализом.

Выбор моделей обусловлен их способностью работать с разнородными и частично коррелированными признаками, а также устойчивостью к переобучению, что особенно важно при наличии большого количества параметров и ограниченного объёма данных. Кросс-валидация позволила объективно оценить качество моделей и снизить риск переобучения.

Важно отметить, что коэффициенты корреляции показали лишь слабую или умеренную связь между отдельными признаками и статусом заказа, что свидетельствует о необходимости комплексного подхода к анализу, включающего построение сложных моделей и анализ их интерпретируемости. Кроме того, выявленные значимые факторы (сумма заказа, программа лояльности, способ оплаты, география) могут быть использованы для разработки персонализированных стратегий взаимодействия с клиентами и оптимизации бизнес-процессов.

Заключение

В работе проведён всесторонний интеллектуальный анализ факторов, влияющих на статус заказа клиента в интернет-магазине техники. Использование ансамблевых методов машинного обучения и глубоких нейронных сетей позволило достичь наилучших результатов в задаче классификации статуса заказа, что подтверждается высокими значениями точности и F1-оценки.

Наиболее значимыми факторами, влияющими на статус заказа, были определены сумма заказа, участие в программе лояльности, способ оплаты и география доставки. Полученные результаты могут быть практически применены для повышения эффективности логистики, оптимизации бизнес-процессов, улучшения прогнозирования спроса и минимизации рисков отмены или возврата заказов. Кроме того, они создают основу для дальнейших исследований и внедрения персонализированных сервисов в электронной коммерции.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Арланова А.А. Интеллектуальный анализ данных: виды и методы / А.А. Арланова, А.М. Нобатов // Вестник науки. — 2023. — Т. 2. — № 1 (58). — С. 7–10. — URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-vidy-i-metody> (дата обращения: 15.02.25).
2. Лыгина Н.И. Разработка информационной системы принятия решения интернет-магазина на основе ситуационного анализа / Н.И. Лыгина, Т.Н. Окшин // Молодой ученый. — 2020. — № 21 (311). — С. 56–63. — URL: <https://moluch.ru/archive/311/70586/> (дата обращения: 13.02.25).
3. Мищенко А.А. Обзор прикладных решений в области анализа данных на примере ювелирной отрасли / А.А. Мищенко, Т.Е. Тимашкова // Российская наука, инновации, образование (РОСНИО-II-2023) : сборник научных статей по материалам II Всероссийской (национальной) научной конференции с международным участием. — 2023. — С. 8–23. — DOI: 10.47813/rosnio-II.2023.8.8-23. — EDN DJNKEU.
4. Тесленко И.Б. Big Data = Большие данные : учебное пособие / И.Б. Тесленко, С.А. Кузнецов, Н.Н. Ползунова [и др.]. — Владимир : Издательство ВлГУ, 2021. — 123 с.
5. Трегуб А.В. Методика построения модели ARIMA для прогнозирования динамики временных рядов / А.В. Трегуб, И.В. Трегуб // Лесной вестник. — 2011. — № 5. — С. 179–183. — URL:

<https://cyberleninka.ru/article/n/metodika-postroeniya-modeli-arima-dlya-prognozirovaniya-dinamiki-vremennyh-ryadov> (дата обращения: 15.02.25).

Список литературы на английском языке / References in English

1. Arlanova A.A. Intellektualnii analiz danih: vidi i metodi [Intelligent data analysis: types and methods] / A.A. Arlanova, A.M. Nobatov // Vestnik nauki [Bulletin of Science]. — 2023. — Vol. 2. — № 1 (58). — P. 7–10. — URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-vidy-i-metody> (accessed: 15.02.25). [in Russian]
2. Lygina N.I. Razrabotka informatsionnoi sistemi prinyatiya resheniya internet-magazina na osnove situatsionnogo analiza [Development of information system decision internet store based on a situational analysis] / N.I. Lygina, T.N. Okshin // Molodoj uchenyj [Young scientist]. — 2020. — № 21 (311). — P. 56–63. — URL: <https://moluch.ru/archive/311/70586/> (accessed: 13.02.25). [in Russian]
3. Mishchenko A.A. Obzor prikladnyh reshenij v oblasti analiza dannyh na primere yuvelirnoj otrasli [Overview of applied solutions in the field of data analysis on the example of the jewelry industry] / A.A. Mishchenko, T.E. Timashkova // Rossijskaya nauka, innovacii, obrazovanie (ROSNIO-II-2023) [Russian Science, Innovation, education (ROSNIO-II-2023)] : collection of scientific articles based on the materials of the II All-Russian (National) Scientific Conference with international participation. — 2023. — P. 8–23. — DOI: 10.47813/rosnio-II.2023.8.8-23. — EDN DJNKEU. [in Russian]
4. Teslenko I.B. Big Data = Bolshie dannie [Big Data = Big Data] : study guide / I.B. Teslenko, S.A. Kuznetsov, N.N. Polzunova [et al.]. — Vladimir : Vladimir State University Publishing House, 2021. — 123 p. [in Russian]
5. Tregub A.V. Metodika postroeniya modeli ARIMA dlya prognostirovaniya dinamiki vremennikh ryadov [The procedure of construction the ARIMA model for forecasting the time series dynamics] / A.V. Tregub, I.V. Tregub // Lesnoj vestnik [Forestry Bulletin]. — 2011. — № 5. — P. 179–183. — URL: <https://cyberleninka.ru/article/n/metodika-postroeniya-modeli-arima-dlya-prognostirovaniya-dinamiki-vremennyh-ryadov> (accessed: 15.02.25) [in Russian]