

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Кафедра Прикладная математика и информатика
(наименование)

01.03.02 Прикладная математика и информатика
(код и наименование направления подготовки / специальности)

Компьютерные технологии и математическое моделирование
(направленность (профиль) / специализация)

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(БАКАЛАВРСКАЯ РАБОТА)**

на тему Разработка алгоритма для анализа и прогнозирования поведения клиентов в
онлайн-банкинге

Обучающийся В.А. Нарицин _____
(Инициалы Фамилия) (личная подпись)

Руководитель канд. пед. наук, доцент, Т.А. Агошкова
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Консультант канд. филол. наук, М.В. Дайнеко
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2025

Аннотация

Тема данной выпускной квалификационной работы: «Разработка алгоритма для анализа и прогнозирования поведения клиентов в онлайн-банкинге».

Целью работы является разработка алгоритма для анализа и прогнозирования поведения клиентов в онлайн-банкинге с использованием методов машинного обучения.

Данная выпускная квалификационная работа состоит из: введения, трёх основных разделов, заключения, а также списка используемой литературы и задействованных источников.

Введение раскрывает актуальность темы, область применения, поставленную цель и задачи для реализации алгоритмов машинного обучения.

Первый раздел содержит обзор предметной области, обзор преимуществ машинного обучения для анализа данных и описание алгоритмов машинного обучения для определения состояния клиента.

Второй раздел содержит определение функциональных требований к разрабатываемой системе и создание концепта интерфейса.

Третий раздел представляет собой программную реализацию интерфейса и тестирование приложения для предсказания данных.

Заключение содержит выводы, которые были сделаны в результате проведенной работы.

Результатом работы является программа для предсказания поведения клиентов в онлайн-банкинге по заданным.

В работе было использовано 21 рисунок и 25 ссылок на внешние ресурсы.

Общий объем выпускной квалификационной работы 42 страницы.

Abstract

The title of the graduation work is *Developing an algorithm for analyzing and predicting the customer behaviour in online banking*.

The graduation work consists of an introduction, 3 parts, 21 figures, a conclusion, and a list of 25 references including foreign sources.

The aim of this research is to develop and evaluate algorithms for analyzing and predicting the customer behaviour in online banking using the machine learning methods.

The object of the graduation work is the processes of analyzing and predicting customer behaviour in the online banking system.

The subject of the graduation work is the machine learning methods used for analyzing the transactions, clustering customers, predicting their loyalty (churn), and preparing personalized offers for users.

The key issue of the graduation work is the application of machine learning to improve the customer behaviour analysis and prediction in online banking as well as to improve the decision-making and the service quality.

The graduation work may be divided into three logically connected parts.

The first part studies the subject area and the advantages of machine learning for a data analysis. The second part defines the functional requirements for the system being developed and presents the concept of the user interface. The third part deals with the software implementation of the interface and testing the application for data prediction.

In conclusion, it should be noted that the developed algorithm and its software implementation provide a practical tool for predicting the customer behaviour based on the input data.

The work is of interest for a narrow circle of readers interested in applying machine learning in banking sphere, customer behavior analysis, churn prediction, and personalized marketing in the financial sector.

Содержание

Введение	5
1 Анализ алгоритма прогнозирования поведения клиентов в онлайн-банкинге	8
1.1 Обзор предметной области	8
1.2 Преимущества машинного обучения для анализа данных	12
1.3 Описание алгоритма машинного обучения для определения состояния клиента	15
2 Проектирование приложения для анализа данных	22
2.1 Определение функциональных требований к разрабатываемой системе	22
2.2 Создание концепта интерфейса	24
3 Реализация приложения для прогнозирования поведения клиентов в онлайн-банкинге	27
3.1 Разработка и реализация интерфейса	27
3.2 Тестирование приложения для предсказания данных	33
Заключение	39
Список используемой литературы	41

Введение

В условиях стремительного развития цифровых технологий и увеличения объема данных, финансовые компании сталкиваются с необходимостью эффективного анализа и прогнозирования поведения клиентов. Онлайн-банкинг, как одна из ключевых областей финансового сектора, требует особого внимания к пониманию предпочтений и потребностей клиентов для улучшения качества обслуживания и повышения лояльности. Во-первых осуществляется персонализация клиентского опыта. Машинное обучение позволяет адаптировать услуги под индивидуальные потребности с помощью:

- рекомендательных систем: анализ транзакций и цифрового следа помогает предлагать клиентам персонализированные продукты;
- чат-ботов и голосовых помощников: ИИ-ассистенты обрабатывают 90% запросов, освобождая операторов для сложных задач;
- гиперперсонализации: алгоритмы учитывают геолокацию, эмоциональный тон и историю операций, чтобы предугадывать действия клиентов.

Во-вторых проводится прогностическая аналитика и управление рисками. Машинное обучение используется для прогнозирования рыночных трендов и минимизации рисков через:

- прогнозирование доходов: модели анализируют исторические данные и внешние факторы для составления финансовых прогнозов;
- управление кредитными рисками: алгоритмы оценивают вероятность дефолта, снижая убытки банков на 15–20%;
- мониторинг регуляторных требований: ИИ автоматизирует контроль за соблюдением нормативов, предотвращая штрафы.

В-третьих осуществляется работы с большим объёмом данных. Количеств обрабатываемых данных в онлайн-банкинге активно возрастает, а машинное обучение позволяет их эффективно использовать. Алгоритмы

машинного обучения способны ежедневно обрабатывать миллионы операций, обнаруживая определённые закономерности, что позволяет улучшить обслуживание клиентов.

Таким образом, методы машинного обучения способны решать различные задачи анализа данных и предсказания событий.

Объектом исследования являются процессы анализа и прогнозирования поведения клиентов в системе онлайн-банкинга.

Предмет исследования - это методы машинного обучения, используемые для анализа транзакционных операций, группировки клиентов, предсказания лояльности и подготовки персональных предложений пользователям.

Целью работы является разработка и экспериментальная оценка алгоритмов для анализа и прогнозирования поведения клиентов в онлайн-банкинге, направленных на повышение эффективности управленческих решений финансовых компаний.

Для достижения поставленной цели необходимо решить следующие задачи:

- провести анализ современных методов машинного обучения, применяемых в задачах анализа клиентского поведения;
- выявить ключевые особенности данных онлайн-банкинга, влияющие на выбор алгоритмов машинного обучения;
- разработать алгоритмы для решения конкретных задач: сегментации клиентов, прогнозирования оттока, анализа транзакций и персонализации предложений;
- разработать интерфейс для пользователей с целью упрощения работы с алгоритмами;
- написать руководство пользователя для приложения;
- провести эксперименты с использованием реальных данных для оценки эффективности предложенных алгоритмов;
- оценить точность моделей при помощи графиков и формул;
- сделать выводы по проделанной работе.

Актуальность исследовательской работы оправдана возрастающей потребностью финансовых организаций в программном обеспечении, позволяющем быстро и точно анализировать большие объемы данных для принятия рациональных решений в сфере управления. Использование алгоритмов машинного обучения в бизнес-процессах позволяет не только улучшить качество обслуживания клиентов, но и способствует улучшению организационной структуры предприятия, ведёт к снижению рисков и повышает конкурентоспособность на рынке.

Использование реалистичных данных при тестировании приложения, позволило оценить эффективность используемых алгоритмов и выбрать подходящие методы для решения задач. Полученное в результате работы приложение может быть использовано финансовыми компаниями для улучшения своих продуктов, услуг и концепций взаимодействия с клиентами.

Таким образом, данная бакалаврская работа направлена на решение актуальных проблем финансовых компаний с использованием современных методов машинного обучения, что позволит значительно повысить эффективность и качество обслуживания клиентов в онлайн-банкинге.

Первый раздел работы посвящен изучению предметной области, сравнительному анализу различных алгоритмов машинного обучения и приведено описание выбранных алгоритмов. Во втором разделе определены требования к разрабатываемой системе и пользовательское взаимодействие с ней, также разрабатывается концептуальная модель интерфейса приложения. Третий раздел посвящен разработке пользовательского интерфейса и программной реализации алгоритма машинного обучения, тестированию разрабатываемого продукта и анализу полученных результатов.

Результатом работы является анализ ключевых проблем финансового сектора, проведённый в области прогнозирования данных, программная реализация алгоритмов машинного обучения с пользовательским интерфейсом для повышения эффективности и качества обслуживания клиентов.

1 Анализ алгоритма прогнозирования поведения клиентов в онлайн-банкинге

1.1 Обзор предметной области

В современном банковском секторе цифровизация и переход к онлайн-сервисам стали ключевыми факторами успеха. Традиционные банки и компании, развивающие финансовые технологии, активно используют данные для улучшения качества обслуживания клиентов, разработки новых продуктов и прогнозирования поведения пользователей. Анализ данных в онлайн-банкинге включает в себя обработку и исследование большого объема информации, связанной с финансовыми операциями и пользовательской активностью в приложениях.

В условиях высокой конкуренции между банками данные о транзакциях и поведении пользователей играют стратегическую роль. Онлайн-банкинг генерирует огромные массивы данных, начиная от информации о платежах и переводах, до подробностей о том, как клиенты взаимодействуют с мобильными приложениями и веб-сайтами. Эти данные дают возможность банкам не только улучшать обслуживание, но и эффективно управлять рисками, обеспечивать безопасность, персонализировать предложения для клиентов, а также предсказывать их будущие действия.

Основные цели анализа данных для повышения качества обслуживания клиентов:

- персонализация услуг: банки используют данные для создания индивидуальных предложений и услуг, которые соответствуют потребностям конкретных клиентов. Проанализировав транзакционную активность, можно предлагать пользователям кредиты, инвестиционные продукты или скидки на определенные категории расходов;
- предсказание поведения клиентов: «методы анализа временных

рядов, включая рекуррентные нейронные сети, применяются для предсказания будущих действий клиентов, таких как запросы на кредиты или отток» [13];

- управление рисками: «системы обнаружения аномалий, основанные на алгоритмах изолированного леса или методах опорных векторов, выявляют мошеннические операции, снижая финансовые потери» [6];
- оптимизация внутренних процессов: анализ данных используется для оптимизации процессов внутри организации (для распределения ресурсов, улучшения технической поддержки клиентов и повышения эффективности рекламных кампаний).

Анализ данных в онлайн-банкинге имеет несколько направлений, каждое из которых требует индивидуального подхода и применения различных методов машинного обучения.

Транзакционный анализ – основное направление изучения данных в онлайн-банкинге, связанное с транзакциями (платежами, переводами и другими операциями клиентов). Транзакционные данные включают информацию о пользователях (получателях и отправителях средств), дате, времени операций, суммах, категориях расходов, местонахождении и т.п. Данная информация позволяет выявлять закономерности в поведении клиентов, и, используя их предпочтения и привычки, предсказывать возможные действия, для подготовки персональных предложений.

Основные задачи анализа транзакционной активности:

- выявление моделей поведения с учётом типов транзакций, частоты, времени операций, сумм, клиентов и т.п.;
- сегментация клиентов по активности путём разделения на группы в зависимости от их транзакционной активности (редкие и активные пользователи, с крупными или мелкими транзакциями и т.п.);
- оценка финансового поведения проводится по результатам расходных и доходных операций, с целью создания профиля клиента,

определения его финансового состояния и вероятности использования различных банковских продуктов;

- выявление подозрительных операций для анализа возможных мошеннических действий.

Анализ поведения пользователей в онлайн-приложениях изучает поведение клиентов в онлайн-приложениях (мобильные приложения и веб-интерфейсы). Этот вид данных включает информацию о том, как клиенты взаимодействуют с банком через цифровые каналы: как часто они входят в систему, какие сервисы используют, сколько времени проводят в приложении, какие действия выполняют.

Основные аспекты анализа поведения в приложениях:

- частота использования приложений: определение, как часто клиенты входят в онлайн-банкинг, когда происходит пик их активности (например, в выходные или рабочие дни);
 - используемые сервисы: какие сервисы и функции онлайн-банкинга наиболее популярны среди клиентов – переводы, оплата счетов, инвестиции, запросы по кредитам и др.;
 - анализ маршрутов пользователей: какие шаги предпринимают клиенты после входа в приложение, сколько времени они тратят на каждое действие, что помогает улучшить UX (пользовательский опыт). Например, «методы анализа кликов (clickstream analysis) и A/B-тестирование интерфейсов для улучшения UX» [12];
 - анализ откликов на уведомления: эффективность push-уведомлений и рассылок через приложение – как часто клиенты реагируют на сообщения от банка, какие предложения их больше всего интересуют
- В числе прочего, «использование методов NLP для оптимизации текстов рассылок и повышения конверсии» [9].

Изучение данных позволяет финансовым организациям решать всевозможные задачи, связанные с оптимизацией услуг и улучшением взаимодействия с клиентами. В зависимости от задач и типов данных

применяются различные методы анализа – от простых статистических моделей до сложных алгоритмов машинного обучения.

Персонализация и группировка клиентов осуществляется с помощью методов кластеризации и классификации. Финансовые организации делят клиентов на группы, основываясь на их транзакционной активности и поведении в приложениях. Таким образом, можно выделить группы клиентов, которые часто пользуются услугами кредитования, и предложить им персональные условия.

Прогнозирование действий клиентов производится методами машинного обучения. Анализ действий, которые будут совершены клиентами, позволяет не только предлагать своевременные услуги, но и предотвращать отток клиентов. В случае, если алгоритм определяет, что клиент, вероятно, прекратит пользоваться услугами банка в ближайшее время, можно заранее предложить ему улучшенные условия или бонусы.

Аномалии и мошеннические действия обнаруживаются системами машинного обучения, которые могут анализировать данные в реальном времени и выявлять подозрительные транзакции. Например, если клиент внезапно совершает необычно крупный платеж в другой стране, система может автоматически заблокировать операцию до подтверждения клиентом. Такие методы помогают банкам существенно снизить риски.

Оптимизация внутренних процессов за счёт анализа данных позволяет улучшать внутренние процессы банка.

Задачей анализа данных является улучшение пользовательского опыта, создание персонализированных интерфейсов и оптимизация функционала приложений. Изучение поведения клиентов в онлайн-банкинге помогает определить зависимость между активностью и лояльностью пользователя к организации, что используется для разработки методик по удержанию клиентов и приводит к снижению нагрузки на системы технической поддержки.

Следовательно, данные в онлайн-банкинге являются важным

инструментом для анализа и прогнозирования поведения клиентов, а применение методов машинного обучения значительно расширяет возможности их обработки, обеспечивая высокую точность и эффективность принимаемых финансовыми организациями решений. Грамотное использование полученной информации позволяет не только улучшить качество обслуживания, но и эффективно управлять рисками и разрабатывать персонализированные предложения для своих клиентов.

1.2 Преимущества машинного обучения для анализа данных

В настоящее время финансовые учреждения сталкиваются с огромным объемом данных, которые необходимо не только собирать и хранить, но и эффективно анализировать для принятия управленческих решений. Стандартные методы анализа информации (регрессионные модели и статистические подходы) не всегда могут справиться с задачами обработки высокоразмерных и нелинейных данных, характерных для банковской сферы. Для решения этой проблемы финансовые организации стали широко применять методы искусственного интеллекта (ИИ), позволяющие не только автоматизировать процесс обработки данных, но и повысить точность и скорость предсказаний.

Машинное обучение опережает стандартные методы анализа информации по нескольким важным параметрам. Во-первых, «в онлайн-банкинге ИИ обеспечивает обработку миллионов транзакций в реальном времени, что критически важно для оперативности решений» [5]. Во-вторых, «нейронные сети обнаруживают паттерны, недоступные классическим методам» [2].

Главные преимущества искусственного интеллекта для анализа данных в онлайн-банкинге:

- автоматизация процессов анализа данных: ИИ может самостоятельно собирать, обрабатывать и анализировать данные.

Это позволяет существенно сократить время, необходимое на выполнение анализа, и повысить его точность. В банках, где ежедневный объем транзакций может измеряться миллионами, автоматизация анализа данных становится критически важной для оперативного принятия решений. Благодаря применению ИИ, финансовые учреждения могут: выявлять подозрительные транзакции, сегментировать клиентов по уровню активности, прогнозировать кредитоспособность клиентов и многое другое. Например, «системы на основе градиентного бустинга сокращают время обработки данных» [14];

- повышение точности и адаптивность моделей: алгоритмы ИИ, такие как нейронные сети, случайный лес, градиентный бустинг и рекуррентные модели, способны анализировать данные с высокой степенью детализации и учитывать множество факторов, влияющих на поведение клиентов. Использование рекуррентных нейронных сетей и моделей временных рядов, таких как ARIMA, позволяет с высокой точностью предсказывать изменения активности клиентов в течение определённого времени и их потребность в конкретных продуктах или услугах. Алгоритмы ИИ также имеют свойство адаптироваться к изменениям входной информации. То есть, система может обучаться на новых данных и корректировать свои прогнозы в реальном времени. Для онлайн-банкинга это представляет особый интерес, так как поведение клиентов может меняться в зависимости от экономической ситуации, появления новых продуктов или услуг и других факторов. Также «рекуррентные сети и ARIMA-модели адаптируются к изменениям данных, прогнозируя активность клиентов даже в нестабильных условиях» [17];
- персонализация банковских услуг: традиционные методы сегментации клиентов, основанные на демографических данных или объемах транзакций, часто не дают полной картины. Искусственный

интеллект позволяет более глубоко анализировать поведение пользователей, их транзакционные и поведенческие данные, что позволяет создавать персонализированные предложения. Например, с помощью методов кластеризации, таких как К-средние или DBSCAN, можно сегментировать клиентов на основе их финансового поведения и предложить им подходящие кредитные продукты, услуги по управлению инвестициями или персонализированные предложения по возврату средств. Кроме того, ИИ может анализировать не только текущие действия пользователя, но и его потенциальное поведение в будущем, предсказывая, какие услуги или продукты могут быть ему интересны. Кроме того, «кластеризация (DBSCAN) анализирует финансовое поведение, создавая основу для индивидуальных предложений» [7];

- выявление мошенничества и аномалий: общепринятые методы обнаружения мошеннических действий часто основаны на фиксированных правилах, что делает их менее гибкими и эффективными. В отличие от этих методов, машинное обучение позволяет выявлять подозрительные транзакции на основе анализа аномалий в поведении клиентов. Алгоритмы, такие как деревья решений, случайный лес или глубокие нейронные сети, могут исследовать транзакции в реальном времени, сопоставляя текущие действия клиента с его предыдущей активностью. Например, если человек обычно совершает покупки в одном регионе, но вдруг происходит крупный перевод средств в другую страну, система может заподозрить мошенническую активность и временно заблокировать операцию до подтверждения. Это позволяет существенно повысить уровень безопасности и защитить клиентов от мошенников. «Алгоритмы изолированного леса выявляют аномалии в режиме реального времени, снижая риски» [8];
- оптимизация внутренних процессов: алгоритмы машинного

обучения могут анализировать эффективность маркетинговых кампаний, оптимизировать работу службы поддержки или прогнозировать потребность в увеличении штата сотрудников в зависимости от нагрузки на банковскую систему. Кроме того, ИИ позволяет прогнозировать финансовые потоки, оптимизировать управление активами и оценивать кредитные риски. Например, «ИИ прогнозирует нагрузку на службу поддержки, сокращая затраты на 20–30%» [16].

Таким образом, искусственный интеллект значительно превосходит традиционные методы анализа данных благодаря своей способности обрабатывать большие объемы информации, автоматически адаптироваться к изменениям, повышать точность прогнозов и персонализировать банковские услуги. Он не только улучшает качество обслуживания клиентов, но и помогает банкам оставаться конкурентоспособными в условиях стремительного развития цифровых технологий.

1.3 Описание алгоритма машинного обучения для определения состояния клиента

В данном пункте рассмотрена математическая модель, используемая для анализа и прогнозирования поведения клиентов в онлайн-банкинге. На основе данных о транзакциях и активности клиента можно предположить собирается ли клиент отказываться от услуг банка или нет. Для решения этой задачи используем метод логистической регрессии и метод кластеризации. Логистическую регрессию будем использовать, если в обучающей выборке будет произведена предварительная классификация (задача с учителем). Метод кластеризации, если в обучающей выборке будет отсутствовать предварительная классификация (задача без учителя).

Общий алгоритм решения задачи с использованием приведённых методов представлен на рисунке 1 и включает следующие этапы:

- постановка задачи: бинарная классификация или кластеризация;
- создание выборки: 7 признаков (1 признак бинарный);
- предобработка данных: масштабирование, перевод бинарного признака в значения 0 и 1;
- разделение на обучающую и тестовую выборки;
- выбор модели: логистическая регрессия с добавлением L2-регуляризатора и DBSCAN;
- обучение модели;
- оценка модели;
- интерпретация результатов: анализ и объяснение результатов;
- визуализация: использование графиков и диаграмм для наглядного представления результатов.

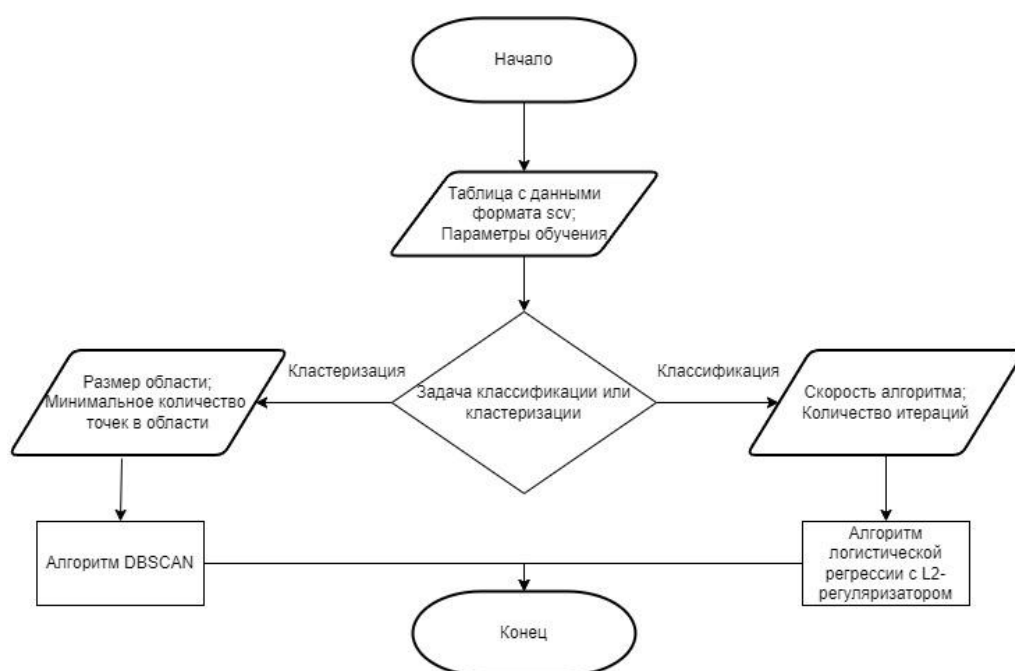


Рисунок 1 - Общий алгоритм решения задачи

Задача бинарной классификации заключается в предсказании одного из двух возможных классов для каждого клиента: отказ от услуг (класс 1) или продолжение использования услуг (класс 0). Входными данными для модели

являются признаки, описывающие транзакции и активность клиента, такие как: количество транзакций, сумма транзакций, частота входа в систему и другие.

Детально рассмотрим используемые методы.

Логистическая регрессия. Функция потерь(L) – бинарная кросс-энтропия будет измерять разницу между предсказанными вероятностями и истинными метками классов. «Формула бинарной кросс-энтропии (1) выглядит следующим образом.

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)] \quad (1)$$

где:

y_i - истинная метка класса для i-го образа,

$\hat{y}_i$ - предсказанная вероятность принадлежности к классу 1 для i-го образца,

N - количество образцов» [23].

В момент обучения модели мы можем столкнуться с различными проблемами в используемом датасете, которые могут привести к переобучению модели: линейно зависимые признаки, число наблюдений меньше числа признаков и прочее. Для решения проблемы переобучения используются регуляризация, которая добавляет к функции потерь штраф за большие значения параметров модели. Мы выберем L2-регуляризацию, потому что L2-регуляризация менее чувствительна к выбросам в данных по сравнению с L1-регуляризацией. Это связано с тем, что квадратичный штраф L2-регуляризации меньше влияет на коэффициенты модели. В задачах анализа поведения пользователей это важно, так как некоторые данные могут быть некорректными, из-за аномального поведения клиентов. «L2-регуляризация (2) имеет следующий вид:

$$R(w) = \frac{\lambda}{2} \|w\|^2 \quad (2)$$

где:

λ - коэффициент регуляризации,

w - вектор весов» [22].

«Формула функции потерь с L2-регуляризацией (3) выглядит следующим образом:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)] + \frac{\lambda}{2} \|w\|^2 \quad (3)$$

где:

y_i - истинная метка класса для i -го образа,

$\hat{y}_i$ - предсказанная вероятность принадлежности к классу 1 для i -го образца,

N - количество образцов,

λ - коэффициент регуляризации,

w - вектор весов» [4].

Для минимизации функции потерь используем градиентный спуск. Градиентный спуск - вектор частных производных функции потерь по каждому параметру модели. Градиент указывает направление наибольшего увеличения функции потерь. В данном методе шаг обучения - это гиперпараметр, который определяет, насколько большой шаг мы делаем в направлении, противоположном градиенту. Слишком большой шаг может привести к пропуску минимума, а слишком маленький - к медленной сходимости. «Градиент функции потерь относительно весов w (4) вычисляется как:

$$\nabla_w L(w) = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) x_i \quad (4)$$

где:

x_i – используемый объект.

Обновление весов происходит по формуле (5):

$$w = w - \eta \nabla_w L(w) \quad (5)$$

где:

η – шаг обучения» [10].

Алгоритм обучения модели включает следующие шаги:

- инициализация параметров модели;
- вычисление градиентов функции потерь с использованием L2-регуляризации;
- обновление параметров модели с использованием метода градиентного спуска;
- повторение шага 3 до достижения критерия остановки (например, достижение заданного числа эпох или стабилизация функции потерь).

Метод кластеризации. При решении задач без учителя используются методы кластеризации такие как: k-средних и DBSCAN.

«Алгоритм k-средних (k-means) – метод кластеризации, который разбивает набор данных на k кластеров, минимизируя суммарное квадратичное отклонение точек от центров кластеров. На каждой итерации точки перераспределяются по ближайшим центрам, после чего центры пересчитываются как средние значения координат точек внутри кластеров. Процесс повторяется до достижения сходимости (когда центры перестают изменяться)» [24].

«DBSCAN (Density-Based Spatial Clustering of Applications with Noise) - это алгоритм кластеризации на основе плотности, который идентифицирует кластеры как непрерывные области с высокой плотностью, разделенные областями с низкой плотностью. Кластер определяется как максимальный

набор точек, связанных по плотности, в то время как точки шума - это те, которые не принадлежат ни к одному кластеру» [20].

Мы используем DBSCAN, так как по сравнению с методом k -средних он быстрее работает с большими выборками, имеет возможность задавать кластеры произвольной формы и разделяет объекты на типы (корневые, граничные и шумовые), что позволяет определять некорректные данные.

Блок-схема алгоритма представлена на рисунке (рисунок 2).

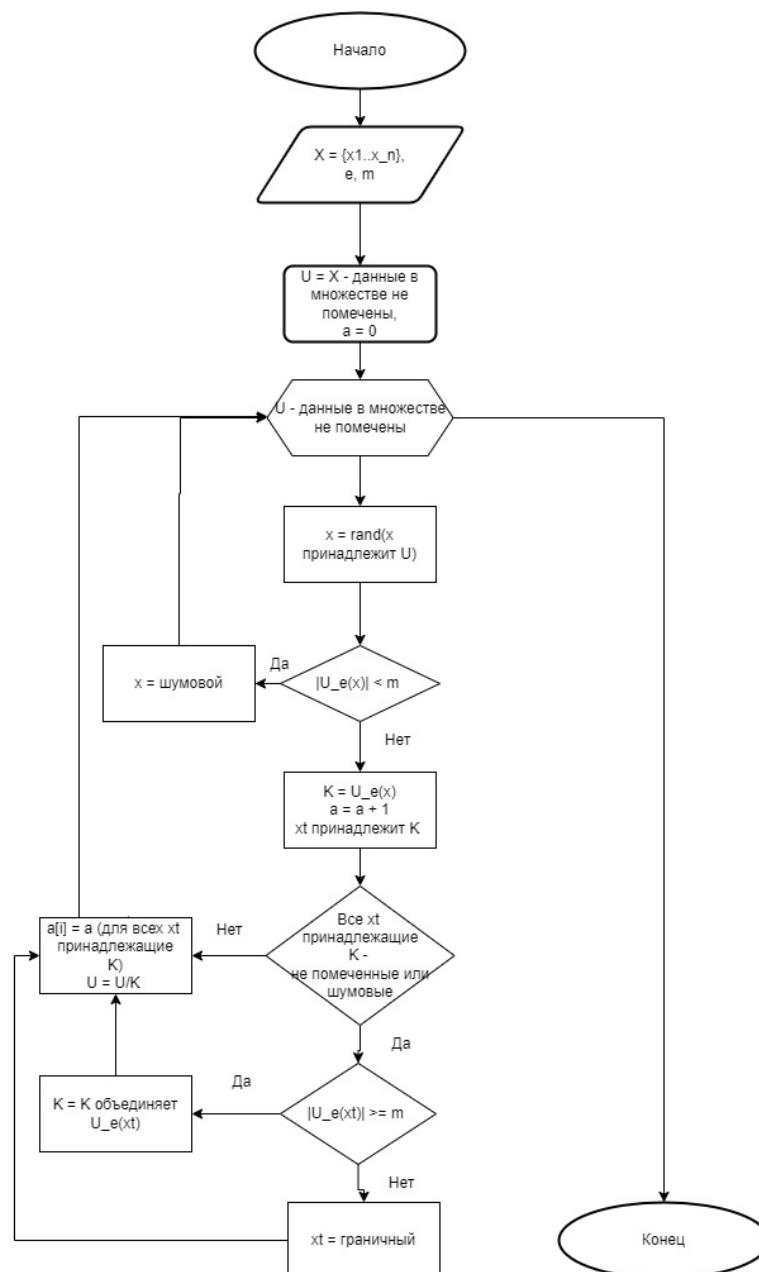


Рисунок 2 - Блок-схема DBSCAN

Алгоритм обучения модели с заданной областью включает следующие шаги:

- берётся случайная точка, и создаётся область заданного размера вокруг неё;
- измеряется количество точек, попавших в эту область;
- если количество этих точек меньше заданного значения, то этой точке присваивается тип – возможно шумовая точка;
- если количество этих точек больше или равно заданному значению присваивается тип – корневая точка;
- дальше идёт оценка точек в области корневой точки;
- если в её области точек меньше заданного значения присваивается тип – граничная точка;
- если в её области точек больше или равно заданному значению присваивается тип – корневая точка;
- шаги 3-7 выполняются до тех пор, пока на все точки в области корневых не будут определены;
- затем возвращаемся к шагу 1, пока точки в выборке не закончатся.

Описанные в данном разделе алгоритмы могут быть применимы для анализа и прогнозирования поведения клиентов в онлайн-банкинге. Далее в работе опишем модель бинарной классификации с использованием критерия качества бинарной кросс-энтропии и метода кластеризации.

2 Проектирование приложения для анализа данных

2.1 Определение функциональных требований к разрабатываемой системе

Цель работы - дать возможность пользователю удобно пользоваться созданными алгоритмами. Для достижения этой цели мы будем использовать:

- интуитивно-понятный интерфейс;
- систему подсказок и советов;
- возможность удобного добавления датасетов.

Основным этапом работы является создание удобного и эффективного инструмента для анализа и прогнозирования данных. Разрабатываемая система должна предоставлять пользователям интуитивно понятный интерфейс, позволяющий легко взаимодействовать с инструментами анализа данных и получать достоверные прогнозы. «Эффективность таких систем зависит от сочетания точности алгоритмов и удобства интерпретации результатов» [19].

При проектировании системы необходимо учитывать следующие ключевые аспекты:

- интуитивно понятный интерфейс: «логичная структура и предсказуемость элементов, что соответствует принципам UX/UI-дизайна» [1];
- точность прогнозов: «использование алгоритмов с высокой воспроизводимостью результатов, таких как логистическая регрессия и DBSCAN» [3];
- обратная связь и рекомендации: система должна предоставлять пользователю интерпретируемые результаты, а также рекомендации на основе анализа транзакционной активности и прогнозирования будущего поведения. «Генерация интерпретируемых отчетов с выделением ключевых факторов риска» [18];

- возможность корректировки входных данных: пользователь должен иметь возможность загружать обновленные данные и изменять параметры анализа.

Исходя из перечисленных аспектов, можно выделить основные функциональные требования для алгоритма кластеризации:

- загрузка и обработка данных об активности пользователей в онлайн-банкинге (входы в систему, использование различных сервисов и предсказанные данные первого алгоритма);
- отображение таблицы данных;
- кластеризация пользователей на основе их активности в системе с использованием метода DBSCAN;
- определение пользователей с высоким риском отказа от услуг банка;
- прогнозирование вероятности отказа клиента от услуг с использованием логистической регрессии;
- визуализация результатов в удобном формате (Отображение разгруппированных данных, ROC-кривая);
- формирование рекомендаций для банка по удержанию клиентов с высокой вероятностью отказа.

После определения функциональных требований к системе необходимо выбрать методологию для дальнейшего проектирования. В данном случае целесообразно использовать UML-моделирование, которое позволяет визуализировать ключевые компоненты системы и их взаимодействие. Это является стандартным языком проектирования, применяемым для описания структуры и поведения программных систем. Использование диаграмм поможет четко структурировать проект, определить основные модули и механизмы работы системы, что в дальнейшем облегчит её реализацию и тестирование.

2.2 Создание концепта интерфейса

Для разработки концептуальной модели интерфейса необходимо в первую очередь определить возможные взаимодействия пользователя с системой. Наиболее подходящим способом визуализации этих взаимодействий является диаграмма вариантов использования.

«Диаграмма вариантов использования или диаграмма прецедентов (англ. use case diagram) – это графический инструмент универсального языка моделирования (UML), который используется для описания функциональных требований к системе, ее возможных сценариев использования и взаимодействия системы с внешними сущностями (актерами). Диаграмма прецедентов представляет собой графическое изображение вариантов использования системы, акторов и их взаимодействия в виде эллипсов и прямоугольников» [21]. Создание такой диаграммы требует учета функциональных требований, заранее определенных для системы. Диаграмма вариантов использования представлена на рисунке 3.



Рисунок 3 - Диаграмма вариантов использования

На основе представленной диаграммы можно легко понять, какими функциями должна обладать будущая система, что позволит определить, из каких страниц, кнопок и функций будет состоять программные модули для алгоритмов анализа данных. Для отображения набора функций, будем использовать карту навигации, чтобы помочь пользователю ориентироваться в системе. Данная карта представлена на рисунке 4.



Рисунок 4 - Карта навигации приложения

Составим карту навигации для подготовки данных. Карта представлена на рисунке 5.

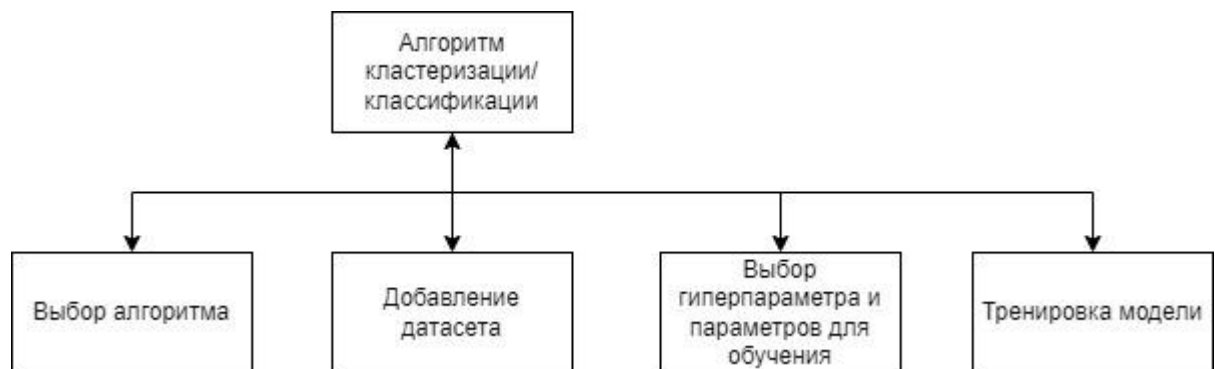


Рисунок 5 - Карта навигации для модуля подготовки данных

Карты навигации для настройки каждого алгоритма представлены на рисунках 6 и 7 соответственно.



Рисунок 6 - Карта навигации для модуля тренировки модели DBSCAN



Рисунок 7 - Карта навигации для модуля тренировки модели классификации

Таким образом, в данном разделе подготовлен проект приложения для анализа и предсказания данных на основе алгоритма кластеризации с использованием заданного датасета; определены функциональные требования к разрабатываемой системе с целью создания концепции приложения, которая будет интуитивно понятной и легко осваиваемой для любого пользователя; создана диаграмма вариантов использования и карта навигации для определения взаимодействий предполагаемого пользователя с программой.

3 Реализация приложения для прогнозирования поведения клиентов в онлайн-банкинге

3.1 Разработка и реализация интерфейса

Разработка пользовательского интерфейса (UI) является ключевым этапом создания приложения. «UI - это система правил и средств, регламентирующая и обеспечивающая взаимодействие программы с пользователем» [15]. Интерфейс выступает связующим звеном между алгоритмами машинного обучения (ML) и конечным пользователем, обеспечивая удобство взаимодействия, визуализацию данных и управление прогнозами. Интерфейс системы должен соответствовать следующим критериям:

- выполнять заложенные функциональные требования;
- визуализация результатов в виде графиков;
- интуитивная навигация, минималистичный дизайн;
- адаптивность под разные разрешения экранов;
- наличие руководства пользователя или справочной информации.

Для реализации интерфейса выбрана библиотека Tkinter, входящая в стандартную поставку Python, по следующим преимуществам:

- не требует установки дополнительных пакетов, что упрощает развертывание;
- поддерживает стандартные элементы управления (кнопки, поля ввода, метки) и кастомизации через ttk-модули;
- бесшовная интеграция с pandas, matplotlib для отображения данных.

Начальная страница приложения, расположенная на главном окне, включает в себя такие виджеты, как Label, для вывода статического текста и Button для отображения кнопок. Реализация начальной страницы приложения представлена на рисунке 8.

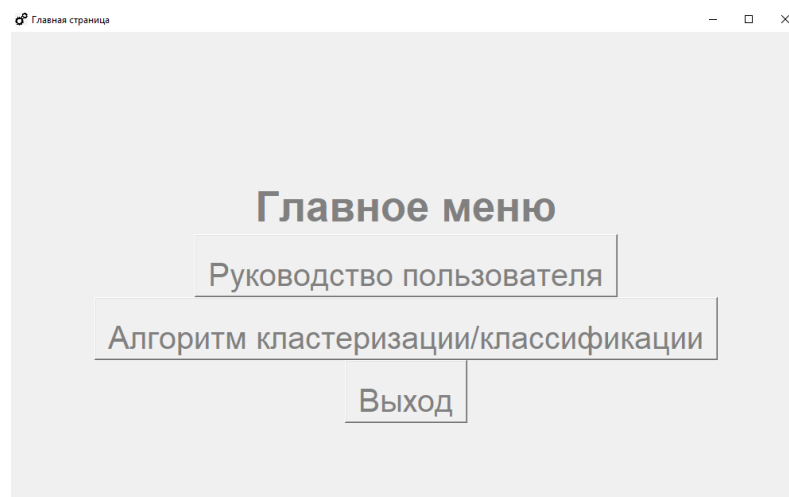


Рисунок 8 - Начальное меню приложения

После нажатия пользователем на кнопку «Алгоритм кластеризации/классификации» открывается страница для добавления датасета (рисунок 9).

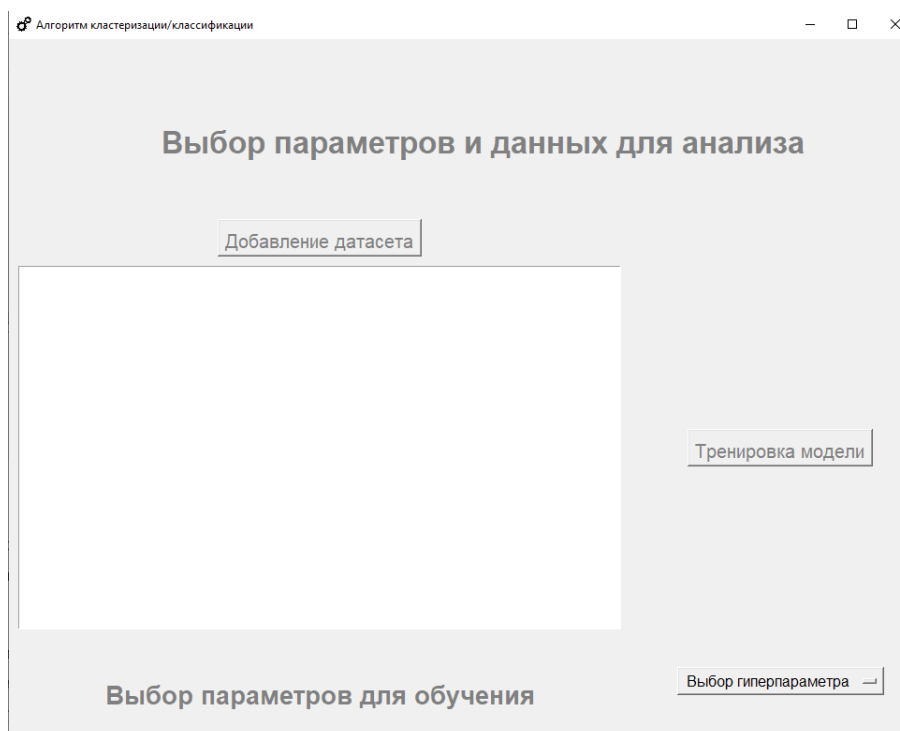


Рисунок 9 - Окно добавления датасета и выбора параметров

Когда пользователь добавит датасет, отобразится перечень данных, выбранного списка и автоматически появятся параметры для обучения алгоритма и целевой параметр. Страница для добавления данных и выбора параметров представлена на рисунке 10.

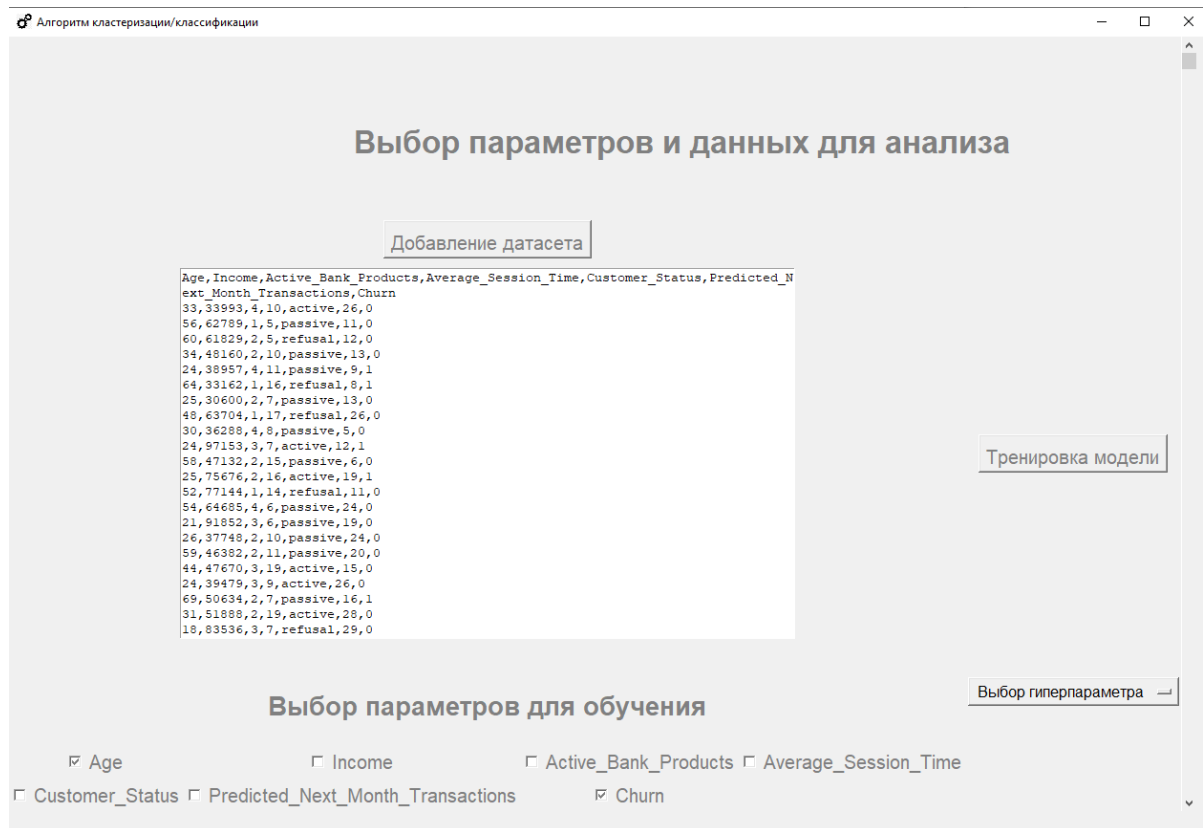


Рисунок 10 - Окно после добавления датасета и выбора параметров

Выбор алгоритма обучения происходит в выпадающем окне, реализованном с помощью элемента StringVar. Здесь пользователю даётся выбор, либо выбрать конкретный целевой параметр, либо выбрать строку «Без параметра(кластеризация)», что приведёт к выбору алгоритма классификации или кластеризации соответственно. Примерный список отображен на рисунке 11.

После нажатия кнопки «Тренировка модели», откроется окно, соответствующее выбранной модели.

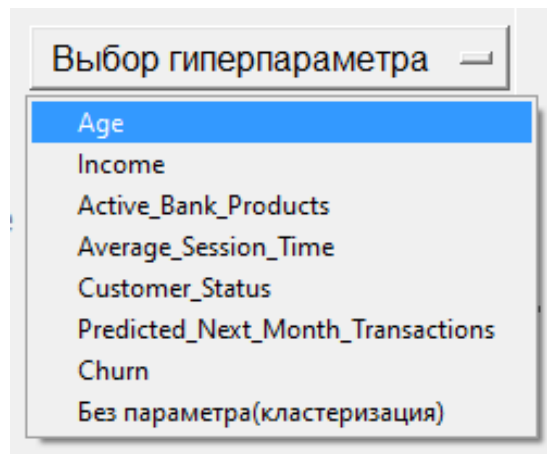


Рисунок 11 - Список для выбора гиперпараметра

В окне алгоритма классификации, при помощи объекта Entry, можно задать скорость обучения и количество итераций (рисунок 12).

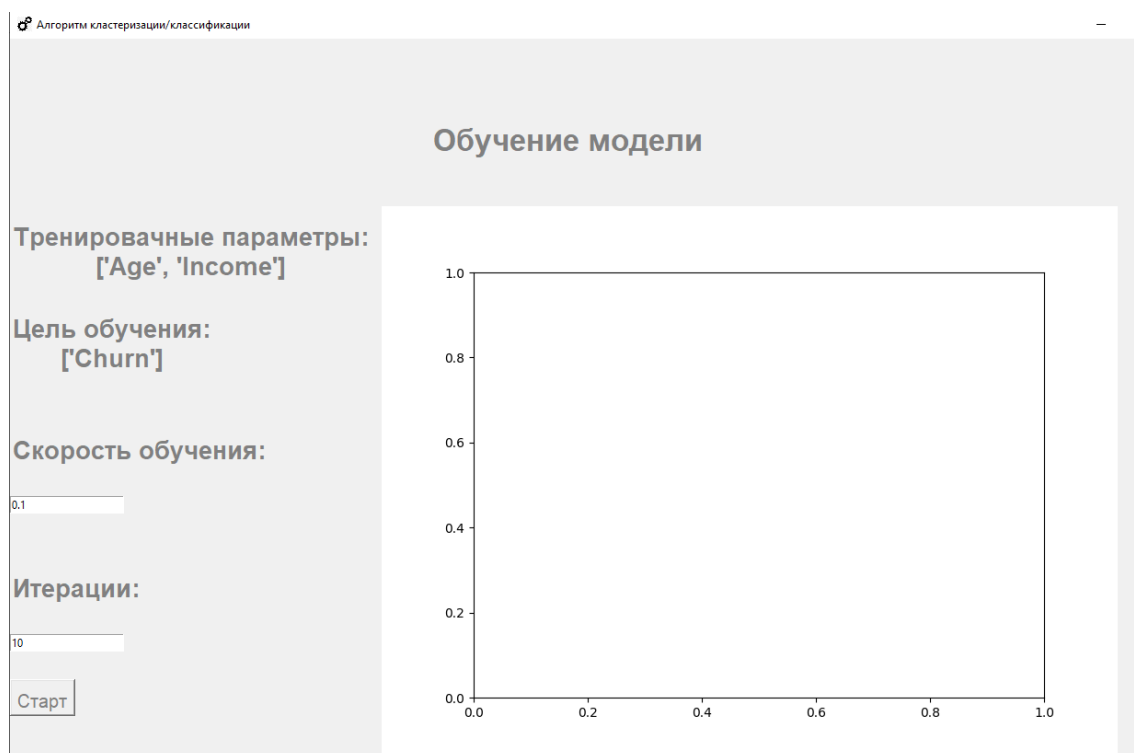


Рисунок 12 - Окно настройки алгоритма классификации

В окне алгоритма кластеризации, можно задать размер области и минимальное количество точек в области (рисунок 13).

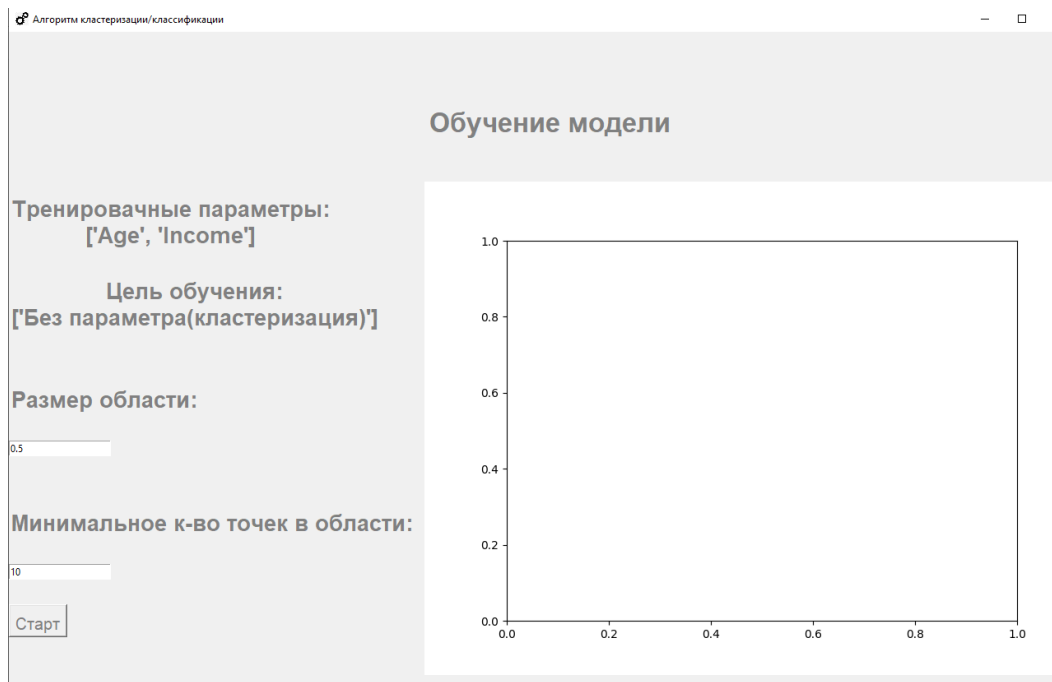


Рисунок 13 - Окно настройки алгоритма кластеризации

При нажатии на кнопку **старт** будет запущен алгоритм обучения выбранной модели и отображено итоговое предсказание данных на графике.

Если выбрано недостаточное количество параметров, то программа выдаст сообщение об ошибке (рисунок 14).

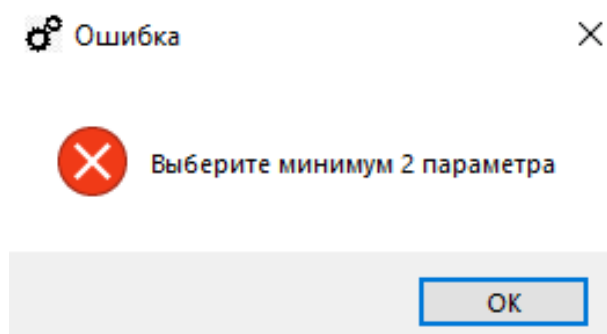


Рисунок 14 - Окно ошибки выбора параметров

При попытке ввести некорректные значения в поля программа выдаст сообщение об ошибке (рисунок 15).

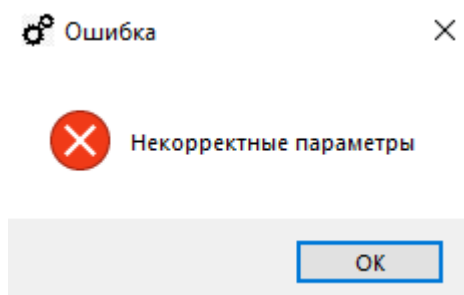


Рисунок 15 - Окно ошибки некорректных параметров

В данном пункте представлен интерфейс работы с алгоритмами машинного обучения для анализа данных. Для удобства пользования интерфейсом разработано руководство пользователя, ознакомиться с которым можно на главной странице приложения, нажав кнопку «Руководство пользователя» (рисунок 16).

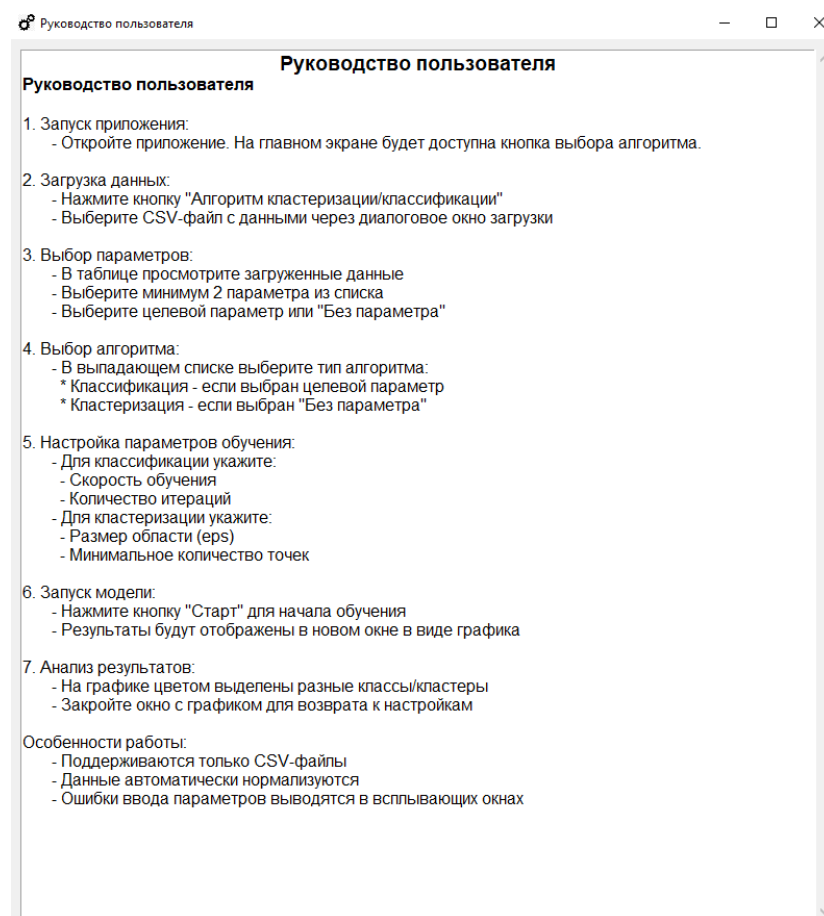


Рисунок 16 - Руководство пользователя

Таким образом в данном пункте представлена реализация интерфейса приложения для прогнозирования поведения клиентов с кратким руководством пользователя.

3.2 Тестирование приложения для предсказания данных

Тестирование алгоритма прогнозирования – критически важный этап создания приложения для анализа поведения клиентов в онлайн-банкинге. В финансовой сфере, где точность прогнозов напрямую влияет на принятие решений (например, оценка риска оттока клиентов или персонализация услуг), компании инвестируют значительные ресурсы в валидацию моделей – как временные, так и финансовые. Например, банки могут выделять до 40% общего бюджета проекта на проверку корректности работы алгоритмов, их устойчивости к аномалиям и согласованности с бизнес-логикой.

Пренебрежение этапом тестирования в таком контексте способно привести не только к финансовым потерям из-за ошибочных прогнозов, но и к репутационным рискам: сбои в работе системы, некорректные рекомендации клиентам или нарушение целостности персональных данных подрывают доверие пользователей. В условиях высокой конкуренции на рынке финтех-услуг это ставит под вопрос саму целесообразность внедрения подобного решения – даже инновационный алгоритм теряет ценность, если его точность не соответствует реальным сценариям поведения клиентов.

Проведём тестирование алгоритмов и интерфейса по-отдельности. Для примера выберем датасет для классификации со следующими параметрами:

- transaction_freq;
- login_attempts;
- account_age;
- session_duration;
- label.

Из перечисленных параметров выбираем 2 параметра обучения:

transaction_freq, login_attempts и 1 целевой параметр label (рисунок 17).

Алгоритм кластеризации/классификации

Выбор параметров и данных для анализа

Добавление датасета

```
transaction_freq,login_attempts,account_age,session_duration,label
5.745071229516849,1.792603548243223,41.0,1.1755130015021997,0.0
5.971532807151039,4.284544784612038,42.0,1.6528449014295743,0.0
4.648769937914996,1.6487945645762292,49.0,0.01992135219070268,0.0
7.368819223261087,3.151152093729363,22.0,7.187396178535303,0.0
4.295788421097572,2.813840065378947,46.0,7.324010490583791,0.0
4.304873460781307,1.3014053696446148,19.0,5.149333499132594,0.0
5.362943407349051,-0.8699203669866966,14.0,1.613628010553837,0.0
2.4126232512304506,1.156568706138541,25.0,0.8355433751068975,0.0
3.480753319498364,2.471370998892911,24.0,0.1912896643728701,0.0
3.6379638867181834,-0.11845555200293711,56.0,1.657481769652207,0.0
7.198473153382331,1.6613355492701964,43.0,8.908555349831467,0.0
5.101292307031886,-0.1371222793201854,45.0,2.4153684170168175,0.0
4.183425913212226,2.166383884564799,56.0,3.542713272172157,0.0
3.273509638665457,2.563547027518508,30.0,1.0534548448133318,0.0
4.099041965121793,1.5624593753100848,49.0,2.223613047481703,0.0
4.097440081655905,4.778417276763406,27.0,5.192245707802051,0.0
4.9797541628893099,0.41343360656614947,13.0,6.077526830956526,0.0
6.233817368154783,0.16873452504346664,51.0,2.4535231416621106,0.0
5.313295392507133,-0.9395051858196632,28.0,0.5773215922511721,0.0
3.0077209266523544,2.2952918538036853,45.0,3.9103286912739863,0.0
6.107699869993116,2.2570524217849557,53.0,2.342033709263341,0.0
4.826527576417639,1.548344456616067,27.0,2.195539232298019,0.0
2.782217014448859,0.9202336874079371,49.0,9.598628027134552,0.0
```

Тренировка модели

Выбор параметров для обучения

☒ transaction_freq ☒ login_attempts ☐ account_age ☐ session_duration

☐ label

label

Рисунок 17 - Датасет классификации

Для кластеризации выберем датасет со следующими параметрами:

- account_age;
- session_duration;
- cluster.

Из перечисленных параметров выбираем 2 параметра обучения: account_age, session_duration (рисунок 18).

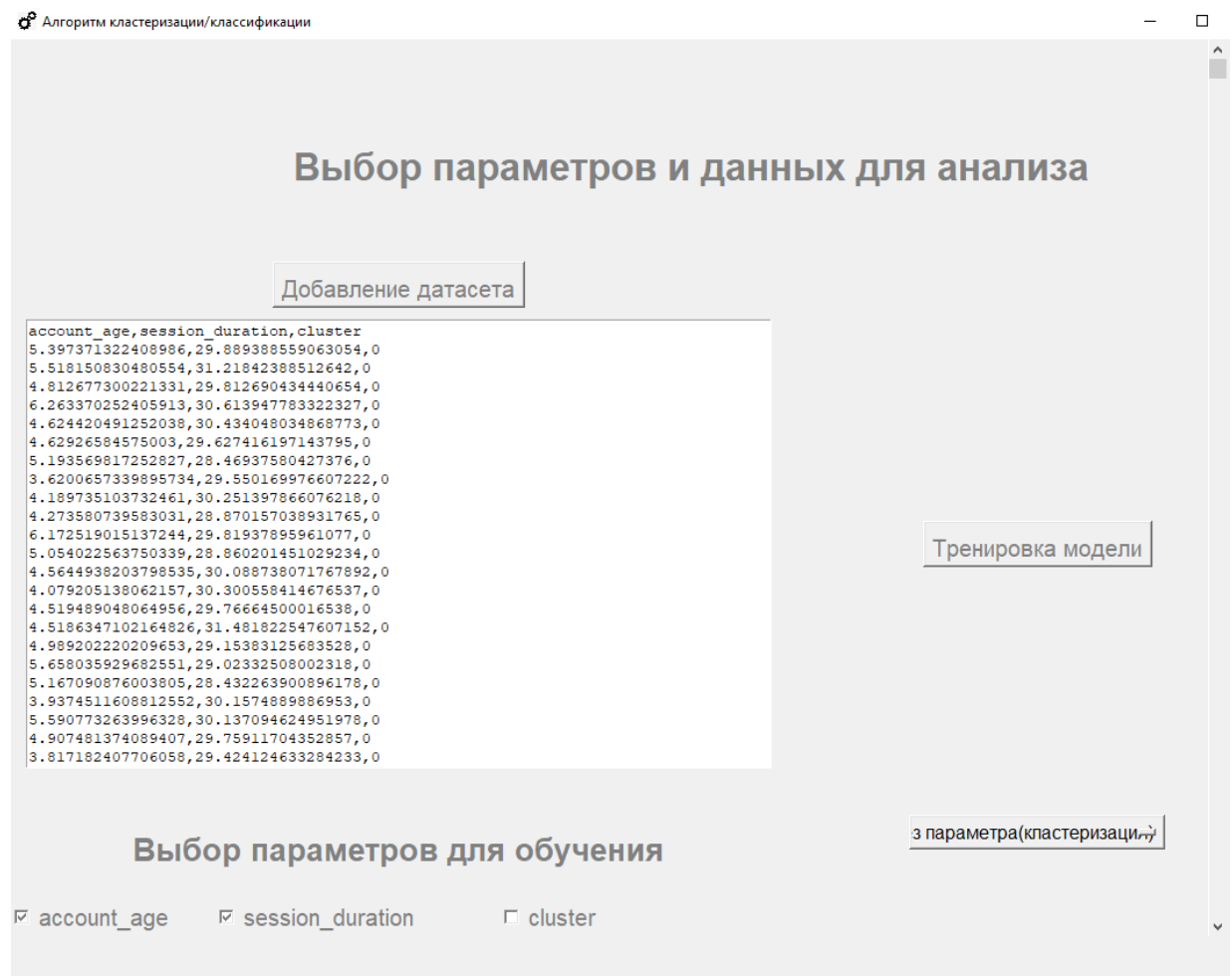


Рисунок 18 - Датасет кластеризации

Затем начинаем обучение моделей классификации и кластеризации. Для этого зададим следующие параметры:

- для классификации: скорость обучения - 0.1, количество итераций – 100;
- для кластеризации (DBSCAN): размер области – 0.3, минимальное количество точек в области – 15.

Для проверки точности работы модели запустим алгоритм в тестовом приложении. В результате тестирования алгоритма классификации получили график потерь (рисунок 19).

Из графика видно, что модель хорошо обучается, что приводит к снижению ошибок при увеличении итераций.

Проверим модель на точность. «Ассигасу (Точность) – метрика качества модели для задач классификации, которая определяет долю правильных ответов, которые модель дает для всех классов» [11].

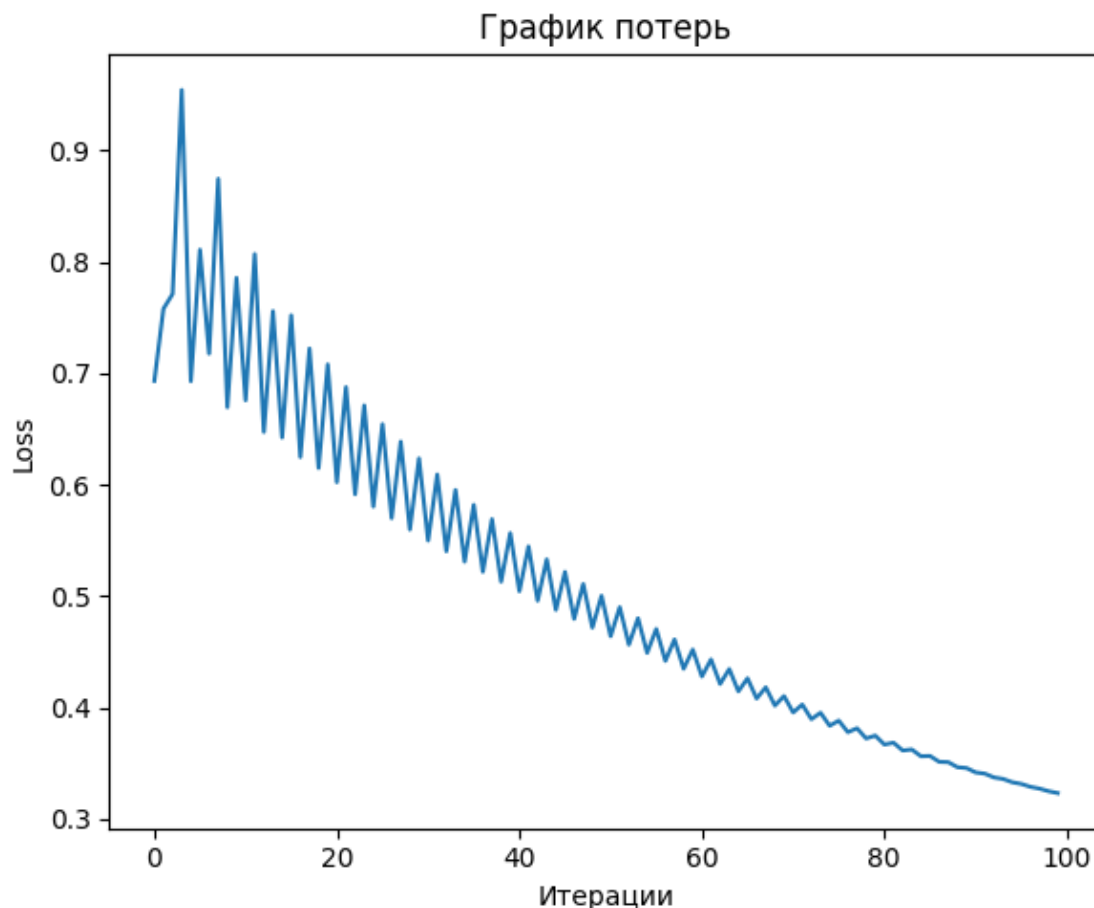


Рисунок 19 - График потерь

Проверка модели на тестовых значениях показала точность (accuracy) равную 0,9875, что говорит о хорошей работоспособности алгоритма.

Результат работы алгоритма в приложении представлен на рисунке 20.

Таким образом, по рисунку видно, что модель разделила клиентов на 2 категории: в верхнем правом углу лояльные пользователи услуг, в нижнем левом – нелояльные.

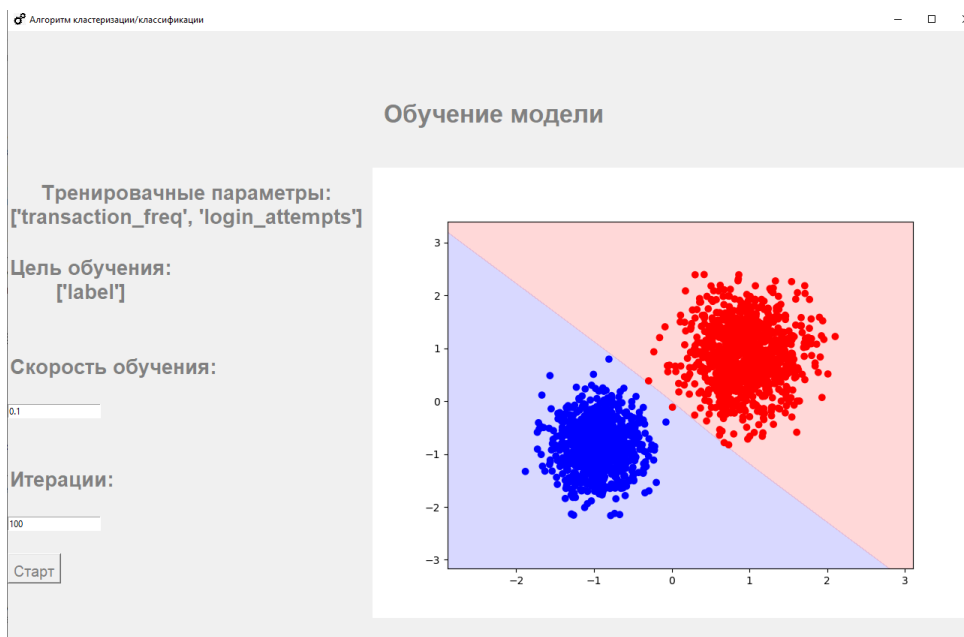


Рисунок 20 - Результат предсказания

Проведём тестирование на точность. «Коэффициент силуэта (silhouette score) – метрика оценки качества кластеризации, вычисляемая для каждого объекта как (6).

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}, \quad (6)$$

где

$a(i)$ - среднее расстояние от объекта i до других точек в его кластере,
 $b(i)$ - минимальное среднее расстояние до объектов соседнего кластера.

Значения коэффициента варьируются от -1 до 1, где 1 указывает на идеальное соответствие кластеру, 0 – на неопределённость, а -1 – на ошибочную кластеризацию» [25].

По результатам тестирования алгоритма кластеризации получили точность модели (silhouette_score) равную 0,85, что говорит о высокой схожести объектов внутри кластеров друг с другом и их от объектов в других кластерах. Значение близкое к 1 указывает на хорошую кластеризацию.

Результат работы алгоритма в приложении представлен на рисунке 21.

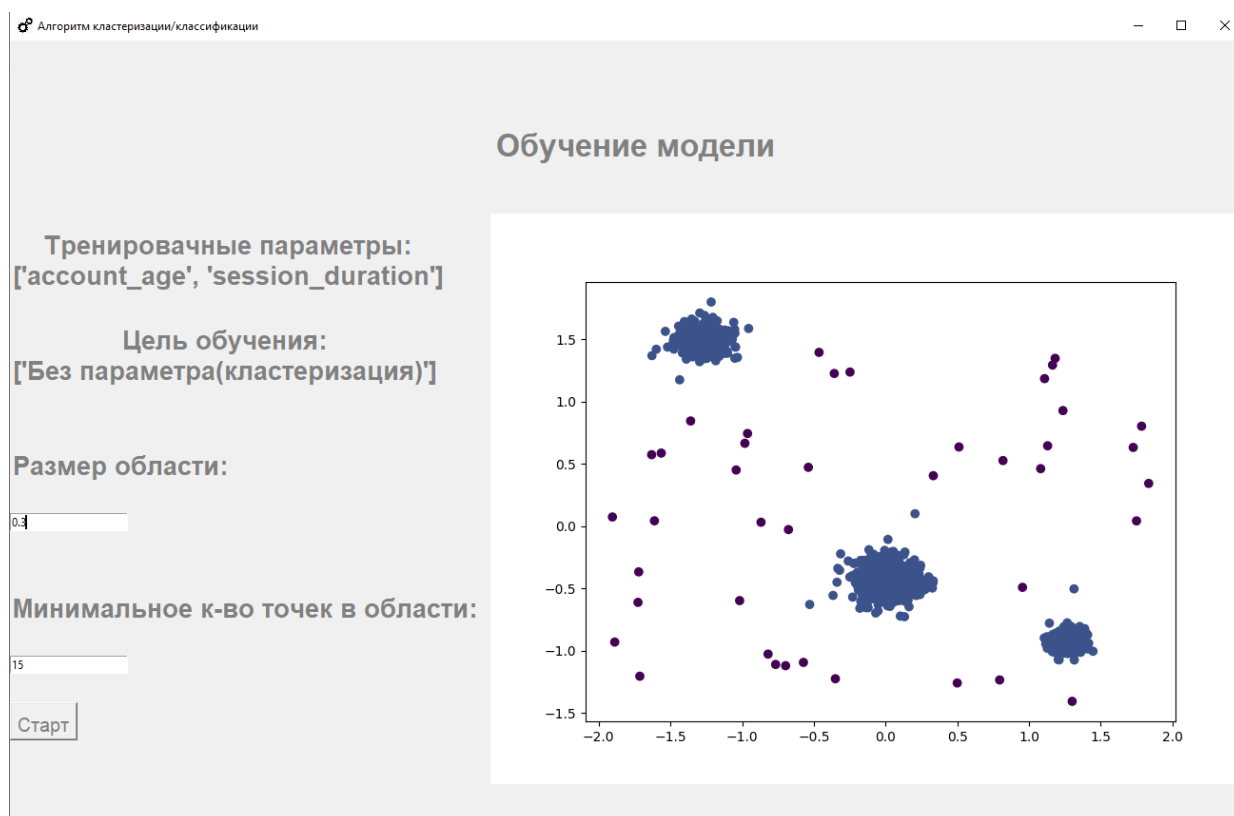


Рисунок 21 - Результат предсказания

В результате работы алгоритма кластеризации нам удалось разделить пользователей на 3 кластера и определить шумовые данные, что позволяет предсказать поведение клиентов.

Таким образом, в данном разделе проведен анализ обучения модели, протестированы алгоритмы и интерфейс. Полученные в результате тестирования оценки подтверждают хорошую обучаемость модели и высокий уровень прогнозирования. Для удобства пользования программой разработан интерфейс, позволяющий эффективно использовать алгоритмы анализа данных.

Заключение

В ходе выполнения бакалаврской работы была разработана и протестирована система алгоритмов машинного обучения для анализа и прогнозирования поведения клиентов в онлайн-банкинге. Исследование показало, что применение методов машинного обучения позволяет значительно повысить качество обслуживания клиентов в банковской сфере, оптимизировать внутренние бизнес-процессы, улучшить организационную структуру предприятия, что снижает финансовые риски и повышает конкурентоспособность на рынке.

Таким образом, в ходе представленной работы:

- проведено изучение и сравнение различных алгоритмов машинного обучения. В рамках исследования рассмотрены и протестированы методы регрессии, классификации и кластеризации. Эксперименты показали, что каждый из этих алгоритмов имеет свои преимущества и область применения, что позволяет выбрать наиболее подходящий метод машинного обучения для решения конкретной задачи;
- выполнен анализ транзакционных данных. В качестве исходных данных были использованы основные параметры транзакций такие как, `transaction_freq`, `login_attempts` и 1 целевой параметр `label` для алгоритма классификации и `account_age`, `session_duration` для алгоритма кластеризации. Алгоритмы машинного обучения разбили клиентов на группы с учётом скрытых закономерностей. Полученные данные использованы для прогнозирования дальнейших действий пользователей;
- выполнен анализ поведения клиентов в онлайн-банкинге. С помощью методов кластеризации транзакционные данные пользователей были разбиты на группы на основе их поведения и предпочтений. Точность метода составила 0,85. Использование алгоритма классификации позволило спрогнозировать поведения

клиентов в отношении их лояльности к банку. Точность прогноза составила 0,9875. Построенный график потерь говорит о высокой обучаемости модели. Использование полученной информации позволяет своевременно реагировать на потенциальные угрозы, принимать меры для удержания клиентов с помощью персонализированных предложений и улучшать взаимодействие с ними;

- разработан удобный интерфейс, позволяющий эффективно использовать алгоритмы машинного обучения, что является основным преимуществом перед аналогичными реализациями приложений.

Таким образом, проведенное исследование подтверждает эффективность использования алгоритмов машинного обучения в процессах онлайн-банкинга. Финансовые компании могут использовать предлагаемый в данной работе интерфейс для решения широкого спектра задач, связанных с анализом и прогнозированием поведения клиентов. Благодаря возможности оперативно анализировать большие объемы данных, финансовые организации будут принимать эффективные управленческие решения, разрабатывать стратегии, направленные на улучшение взаимодействия с клиентами, что обеспечит их устойчивое развитие и конкурентоспособность на рынке.

Список используемой литературы

1. Андреев С. В. Проектирование пользовательских интерфейсов. – М.: ДМК Пресс, 2022. – 198 с.
2. Беляев М. К. Глубинное обучение в аналитике. – Новосибирск: Наука, 2019. – 192 с.
3. Борисова М. А. Интеллектуальные системы поддержки решений. – СПб.: Лань, 2021. – 245 с.
4. Викиконспекты. Регуляризация. – НИУ ВШЭ, 2023 [Электронный ресурс]. – URL: <https://neerc.ifmo.ru/wiki/index.php?title=Регуляризация>
5. Громов А. С. Большие данные и ИИ в финансовых сервисах. – М.: Эксмо, 2020. – 410 с.
6. Иванов С. М. Анализ данных для финансовых институтов. – М.: Инфра-М, 2021. – 367 с.
7. Ильина О. В. Персонализация в цифровом банкинге. – Екатеринбург: УрФУ, 2021. – 180 с.
8. Карпов Р. Д. Безопасность в FinTech. – М.: Гелиос АРВ, 2020. – 305 с.
9. Лукашин А. Н. Машинное обучение в финансах: методы и практика. – СПб.: БХВ-Петербург, 2020. – 432 с.
10. Максимов Ю. Д. Введение в Data Science и машинное обучение. – М.: Издательство «Питер», 2018.
11. Михнин А. Оценка качества моделей машинного обучения: выбор, интерпретация и применение метрик [Электронный ресурс]. URL: <https://www.litres.ru/book/aleksey-mihnin/ocenka-kachestva-modeley-mashinnogo-obucheniya-vybor-interp-69618313/>.
12. Петров К. В. Цифровая аналитика в банковской сфере. – М.: Альпина Паблишер, 2022. – 214 с.
13. Рязанов В. А. Большие данные в банковской аналитике. – М.: ДМК Пресс, 2019. – 298 с.

14. Соколова Е. А. Машинное обучение для финансовых институтов. – М.: Инфра-Инженерия, 2023. – 355 с.
15. Соловьёв И. О., Цой А. Н., Гринкруг Л. С. Технология разработки прикладного программного обеспечения. – Санкт-Петербург, 2021. – 300 с.
16. Тихонова Л. М. Оптимизация банковских процессов с помощью ИИ. – СПб.: Питер, 2023. – 265 с.
17. Ткачев П. И. Проектирование аналитических систем. – Новосибирск: Наука, 2020. – 176 с.
18. Яковлев А. Р. Визуализация данных в банковской аналитике. – Казань: КФУ, 2021. – 154 с.
19. Dixon, M. F., Halperin, I., Bilokon, P. Machine Learning in Finance: From Theory to Practice. – Springer, 2020. – 500 p.
20. Ester, M., Kriegel, H.-P., Sander, J., Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). – AAAI Press, 1996. – P. 226–231.
21. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. – MIT Press, 2016. – 800 p.
22. ITOnboard. Диаграмма вариантов использования (Use Case Diagram) [Электронный ресурс]. URL: https://itonboard.ru/analysis/629-diagramma_variantov_ispolzovanija_use_case_diagram/ (дата обращения: 15.04.2025).
23. James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning. – Springer, 2013. – 426 p.
24. Lloyd, S. Least square quantization in PCM. – Bell Telephone Laboratories Paper, 1957.
25. Rousseeuw P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis [Электронный ресурс]. URL: [https://ru.wikibrief.org/wiki/Silhouette_\(clustering\)](https://ru.wikibrief.org/wiki/Silhouette_(clustering)) (дата обращения: 15.04.2025).