

Федеральное государственное автономное образовательное учреждение  
высшего образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Высшая школа бизнеса

### **КУРСОВОЙ ПРОЕКТ**

Построение рекомендательной системы для продвижения партнерских  
услуг физическим лицам коммерческого банка на основе данных их  
транзакций

образовательная программа «Бизнес-информатика»

Выполнили:

Хан А. -,

Поликарпов К.Н.,

Шрамко В.А.,

Озбетелашвили Г.О.

Руководитель:

приглашенный преподаватель,

Косарев И.М.

## Содержание

Введение.....	4
Основная часть.....	9
1. Этап исследования предметной области .....	9
1.1. Описание объекта исследования.....	9
1.2. Описание стейкхолдеров .....	10
1.3. Текущая архитектура системы .....	11
1.4. Монетизация .....	12
1.5. Исследование данных.....	12
1.6. Информационные процессы .....	15
2. Описание требований к ИТ-решению.....	16
2.1. Целевой принцип работы рекомендательной системы.....	16
2.2. Функциональные требования .....	17
2.2.1. Требования к рекомендательной модели .....	17
2.2.2. Требования к интерфейсу сервиса ИТ-решения.....	18
2.2.3. Требования к предоставлению информации о заказе пользователю .....	18
2.3. Нефункциональные требования .....	19
2.3.1. Требования к информационной модели рекомендательной системы .....	19
2.3.2. Требования к API рекомендательной системы.....	19
3. Описание ИТ-решения.....	21
3.1. Критерии выбора ИТ-решения .....	21
3.2. Сравнительный анализ ИТ-решения.....	21
3.3. Обоснование выбранного ИТ-решения .....	23
3.4. Описание сервиса ИТ-решения.....	25
3.5. Описание модели рекомендательной системы.....	32
3.5.1. Используемые инструменты.....	32
3.5.2. Методология ведения и построения модели рекомендательной системы .....	33
3.5.3. Постановка формальной задачи машинного обучения.....	33
3.5.4. Работа с данными.....	35
3.5.4.1. Понимание данных .....	35
3.5.4.2. Подготовка данных.....	36
3.5.5. Выбор модели рекомендательной системы .....	37
3.5.5.1. Модель по популярности.....	38
3.5.5.2. Коллаборативная фильтрация .....	38

3.5.5.2.1. Memory-based model.....	38
3.5.5.2.2. Модель со скрытыми переменными .....	39
3.5.5.3. Контентная модель .....	40
3.5.5.4. Оценка.....	42
3.5.6. Развертывание модели рекомендательной системы.....	44
3.6. Перспективы развития .....	46
3.7. Выгоды от реализации проекта .....	46
Заключение .....	47
Список использованных источников .....	48
Приложение.....	51

## Введение

Проблема быстрого роста влияния информации особенно затрагивает сферу бизнеса. Принятие эффективного решения зачастую тесно связано с владением актуальной информации о компании и тенденциях на рынке. При этом информация, которую необходимо учитывать для обоснования решений, неуклонно растет. Такие условия приводят к ситуации, когда невозможно управлять компанией без использования информационных технологий и средств автоматизации в бизнесе. Для того, чтобы привлекать как можно больше клиентов в банк и увеличивать их лояльность, необходимо сотрудничать со сторонними сервисами компаний. Взаимовыгодное сотрудничество с ними позволяет в дальнейшем создать экосистему вокруг банка.

Одним из возможных путей сотрудничества с компаниями-партнерами является рекламирование их услуг клиентам банка таким образом, чтобы реклама была наиболее эффективной и персонализированной.

Для решения данного **актуального** вопроса существуют технологии, которые позволяют из накапливаемых данных получать необходимые знания для эффективного управления компанией. Направление “Бизнес-аналитика” зачастую связано с внедрением предиктивных BI-систем, которые автоматизируют бизнес-процессы предприятия и упрощают работу с данными. С целью получения дополнительных знаний и практического опыта в сфере бизнес-аналитики, в рамках проектного семинара “ИТ-консалтинг” нами было выбрано именно данное направление. В последнее время банки стараются стать цифровой платформой в рамках экосистемы, поэтому в работе будет описание проекта, связанного с внедрением предиктивной системы для компании АО “Тинькофф Банк”, которая будет

предлагать релевантные услуги партнеров клиентам банка. Будут рассмотрены используемые методы и технологии машинного обучения и анализа данных.

**Объект исследования:** АО “Тинькофф Банк”.

**Предмет исследования:** рекомендательная система.

### **Бизнес-цель проекта**

Основная бизнес-цель - внедрение сервиса рекомендательной системы партнерских услуг клиентам банка. Рекомендации должны быть актуальны и индивидуально подобраны для каждого пользователя так, чтобы содержание рекламы было привлекательным для клиентов банка.

### **Задачи проекта**

1. Определить организационные и содержательные рамки проекта;
2. Провести анализ предметной области;
3. Выбрать и обосновать методы и инструменты для реализации проекта;
4. Сформулировать функциональные требования к ИТ-решению;
5. Описать архитектуру ИТ-решения;
6. Исследовать современные методы построения рекомендательных систем;
7. Разработать и описать алгоритм машинного обучения;
8. Выбор платформы для реализации проекта;
9. Описать функциональность решения и ожидаемые бизнес-эффекты;
10. Внедрение системы.

**Методы:** коллаборативная фильтрация, контентная модель.

**Характеристика структуры работы:** курсовая работа поделена на 3 блока : введение, основная часть, заключение. Основная часть имеет 3 смысловых главы:

- Этап исследования предметной области
  - В данном блоке прописана информация о Заказчике, проектной команде, допущения и ограничения проекта, план реализации проекта. Раздел носит информационный характер.
- Описание требований к ИТ-решению
  - В данном разделе обосновывается выбор методов и инструментов для разработки ИТ-решения. Кроме этого, формулируются функциональные требования к ИТ-решению и анализируется предметная область, описывается целевая ИТ-архитектура решения.
- Описание ИТ-решение
  - В данном разделе приводится описание и анализ предлагаемого ИТ-решения. Описывается итоговая ИТ-архитектура комплекса ИТ-решений и рекомендательной модели.

### **Допущения и ограничения проекта**

- В банке используется система устаревшая рекомендательная система с ручной настройкой;
- В банке используются платформы хранения данных ClickHouse, PostgreSQL, MongoDB, SQLite;
- Разработка системы производится доступными для студентов средствами и инструментами;
- Система не внедряется непосредственно в ИТ-архитектуру Тинькофф;

### **В рамках проекта:**

- Построение работающего ИТ-решения с описанием его архитектуры и функциональности;
- Описание ИТ-архитектуры Тинькофф по имеющейся информации и знаниям;
- Использование методов интеллектуального анализа данных;
- В рамках проекта данные для верификации решения были взяты из открытого источника Kaggle.

### **Ожидаемые результаты**

- Исследована предметная область
- Разработаны требования к ИТ-решению
- Разработана и внедрена рекомендательная система
- Разработан интерфейс ИТ-решения

Реализация проекта будет считаться успешной, если будут достигнуты следующие цели (в рамках курсового проекта):

- Спроектирована рекомендательная система;
- Разработана и внедрена рекомендательная модель;
- Будет достигнуто качество работы модели не ниже 80%

### **Роли в команде**

- Поликарпов Кирилл, руководитель команды
- Шрамко Влад, бизнес-архитектор
- Озбетелашвили Георгий, бизнес-аналитик
- Хан Александр, разработчик

Участник проекта	Роль в проекте	Описание роли
Поликарпов Кирилл	Руководитель команды, разработчик	<ol style="list-style-type: none"> <li>1. Провел анализ моделей рекомендательной системы.</li> <li>2. Выбрал и внедрил модель рекомендательной системы.</li> <li>3. Разработал схему взаимодействия модели с компонентами системы.</li> </ol>
Шрамко Влад	Бизнес-архитектор	<ol style="list-style-type: none"> <li>1. Разработал схему целевой ИТ-архитектуры.</li> <li>2. Выбрал бизнес-модель рекомендательной системы.</li> <li>3. Разработал модели монетизации и целевых метрик.</li> </ol>
Озбетелашвили Георгий	Бизнес-аналитик	<ol style="list-style-type: none"> <li>1. Описал функциональные и нефункциональные требования к ИТ-решению.</li> <li>2. Техническое описание и составление документации проекта.</li> <li>3. Сформировал гипотезы для возможного развития ИТ-решения.</li> </ol>
Хан Александр	Разработчик	<ol style="list-style-type: none"> <li>1. Провел исследование данных.</li> <li>2. Разработал сервис ИТ-решения и API.</li> <li>3. Внедрил СУБД PostgreSQL на сервер сайта.</li> <li>4. Внедрил ИТ-решение на сервер сайта.</li> </ol>



## **Основная часть**

### **1. Этап исследования предметной области**

#### **1.1. Описание объекта исследования**

«Тинькофф Банк» — российский коммерческий банк, основанный в 1994 году и сфокусированный полностью на дистанционном обслуживании, не имеющий розничных отделений. Штаб-квартира банка расположена в Москве. Является третьим крупнейшим банком по количеству клиентов на текущий момент. Штат банка на октябрь 2020 года составляет 13000 сотрудников. Капитализация банка равна 474 млрд. рублей, чистая прибыль - 44 млрд. рублей (2020 г.).

Основной продукт для физических лиц — кредитные и дебетовые карты, а также вклады. Также банк предоставляет клиентам дополнительные услуги, такие как кэшбэк за покупки в общем и на определенные категории товаров, процентная ставка по вкладам, курс обмена валюты, страховые услуги, бесплатный доступ в бизнес-залы Lounge Key и т. д.

Основным доходом банка является предоставление услуг, но также Тинькофф зарабатывает с помощью размещения рекламы партнеров в своих сервисах, например в мобильном приложении.

## 1.2. Описание стейкхолдеров

Стейкхолдеры - лица, оказывающие влияние на политику организации или заинтересованные в результатах ее деятельности<sup>1</sup>. Они могут быть заинтересованы: в прибыльности проекта, в непосредственном участии или же в конечном продукте, который будет удовлетворять их потребности.

Стейкхолдеров делят по различным группам:

1. Внешние и внутренние;
2. По степени влияния и заинтересованности.

К внутренним стейкхолдерам относятся лица, которые находятся внутри границ проекта и участвует в его непосредственной реализации. Например, руководители, инвесторы, собственники, сотрудники, спонсоры.

К внешним стейкхолдерам относятся лица, которые не принимают участие в реализации проекта, но они заинтересованы в его существовании и конечных продуктах. К внешним стейкхолдерам можем отнести потребителей, партнеров, поставщиков, госорганы,

Рассмотрим степень заинтересованности и влияния стейкхолдеров по матрице Менделоу<sup>2</sup>, которая их разделяет по степени влияния и по степени заинтересованности. Чем выше степень влияния у стейкхолдера, тем больше у него силы и власти. Чем выше степень заинтересованности у стейкхолдера, тем больше он хочет повлиять на реализацию проекта.

---

<sup>1</sup> А.Г. Харин, Т.Р. Гареев. "Стейкхолдерский подход в управлении организациями: перспективы применения теоретико-игровых моделей" URL: [https://te.sfedu.ru/evjur/data/2014/journal12\\_4.pdf](https://te.sfedu.ru/evjur/data/2014/journal12_4.pdf) (дата обращения: 05.03.2021)

<sup>2</sup> П.С.Щербоченко. "Управление взаимодействием со стейкхолдерами в российских компаниях" URL: <https://vestnik.guu.ru/jour/article/viewFile/1041/464> (дата обращения: 05.03.2021)

Для данного проекта составим матрицу внутренних стейкхолдеров, где Н (high) - высокий уровень, М (middle) - средний уровень, L (low) - низкий уровень (см. Приложение 1).

### 1.3. Текущая архитектура системы

Процесс предоставления услуг партнеров клиентам банка на текущий момент устроен следующим образом: клиентам, которые подписаны на рассылку по электронной почте от банка ежемесячно приходят рекомендации товаров и услуг от партнеров (см. Приложение 2).

Процесс формирования представлен следующим образом: рекомендации формируются путем ручного формирования правил по тому, в каких случаях и при каких условиях предложить какие услуги клиенту.

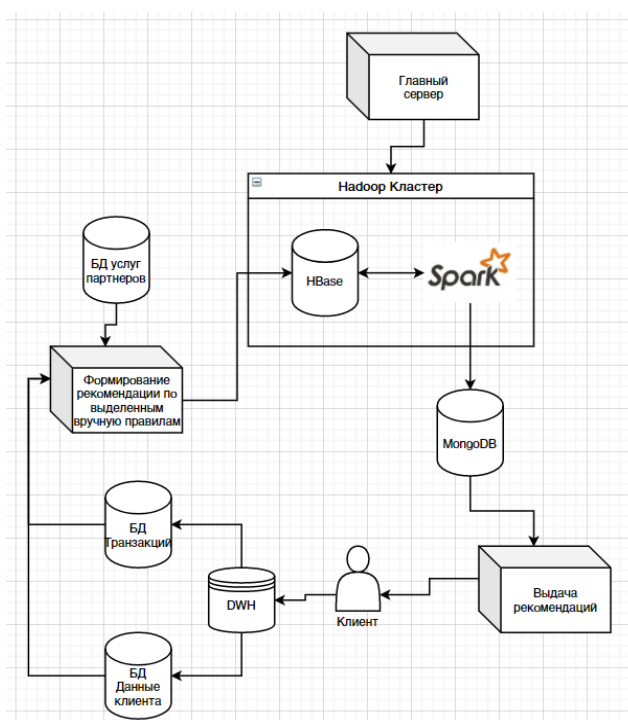


Рисунок 1.3.1. Архитектура текущего состояния системы рекомендации услуг партнеров

## 1.4. Монетизация

В данном кейсе можно выделить несколько возможностей монетизации рекомендательной системы:

**1. Получение % от стоимости чека клиента.**

Если клиент воспользовался рекомендацией от нашего сервиса и оплатил услуги картой Тинькофф Банк, то Тинькофф получает % от стоимости чека клиента (% прописывается в договоре с партнером).

**2. Использование метода “Cost per click”.**

Получения фиксированной выплаты от партнера за каждый клик по баннеру их услуг.

**3. Использование метода “Cost per view”.**

Получения фиксированной выплаты от партнера за каждый просмотр баннера их услуг.

## 1.5. Исследование данных

В результате проведения интервью с заказчиком было выявлено, что данные банка “Тинькофф Банк” хранятся в базе данных в Oracle. Для работы над системой команде проекта был открыт доступ к необходимой информации, а именно: данные о транзакциях клиентов, данные о самих клиентах, данные о партнерах и их дополнительных услугах.

Для нашей рекомендательной системы нам необходимы данные о транзакциях, которые имели категоризацию по типу продукта. К сожалению на Kaggle не нашлось подходящих по этому критерию банковских датасетов. Но мы нашли подходящие нам искусственно сгенерированные датасеты. В итоге, было принято решение, что больше

всего                      нам                      подходит                      данный                      датасет:

<https://www.kaggle.com/radistaleks/synthetic-bank-transactions>.

Рассмотрим таблицы данного датасета и соответствующие им признаки:

1. clients: данная таблица содержит базовые данные о клиентах банка (см. **Приложение 3**);

*Таблица 1.5.1. Таблица описания атрибутов сущности клиент*

Атрибут	Тип данных	Описание
id	int	Идентификатор клиента
registration_date	date	Дата регистрации клиента
full_name	string	Имя, фамилия, отчество клиента
phone_number	string	Номер телефона клиента
email	string	Электронная почта клиента
birthdate	date	Дата рождения клиента
gender	string	Пол клиента
workplace	string	Адрес места работы клиент
address	string	Адрес проживания клиента
income	int	Средний доход клиента за последние 12 месяцев
expenses	int	Средний объем расходов клиента за последние 12 месяцев
credit	boolean	Поле указывающее имеет ли клиент в данный момент открытый кредит
deposit	boolean	Поле, указывающее имеет ли клиент в данный момент открытый вклад

2. categories: данная таблица содержит данные о категориях. Каждая транзакция относится к какой-либо категории товаров/услуг (см. **Приложение 4**);

*Таблица 1.5.2. Таблица описания атрибутов сущности категория*

Атрибут	Тип данных	Описание
id	int	Идентификатор категории
name	string	Название категории
description	string	Описание категории
mcc-code	string	Коды категории продавцов

3. transactions: данная таблица содержит данные о транзакциях клиентов банка (см. **Приложение 5**);

*Таблица 1.5.3. Таблица описания атрибутов сущности транзакция*

Атрибут	Тип данных	Описание
id	int	Идентификатор транзакции
client_id	int	Идентификатор клиента, которому принадлежит транзакция
product_category	int	Идентификатор категории продукта
subtype	string	Подтип транзакции: - Покупка - Подписка - Автоплатеж - Перевод - Пополнение
product_company	string	Название компании, с которой произошла транзакция
amount	float	Объем транзакции
date	datetime	Дата и время транзакции
transaction_type	string	Тип транзакции: - Positive - Negative

4. subscriptions: данная таблица содержит данные об автоматических транзакциях клиента, иными словами, о подписках и автоплатежах (см. **Приложение 6**).

*Таблица 1.5.4. Таблица описания атрибутов сущности подписка*

Атрибут	Тип данных	Описание
id	int	Идентификатор подписки
client_id	int	Идентификатор клиента, которому принадлежит транзакция
product_category	int	Идентификатор категории продукта
product_company	string	название компании, с которой произошла транзакция
amount	float	Объем транзакции
date_start	date	Дата начала подписки\автоплатежа
date_end	date	Дата окончания подписки\автоплатеж

## 1.6. Информационные процессы

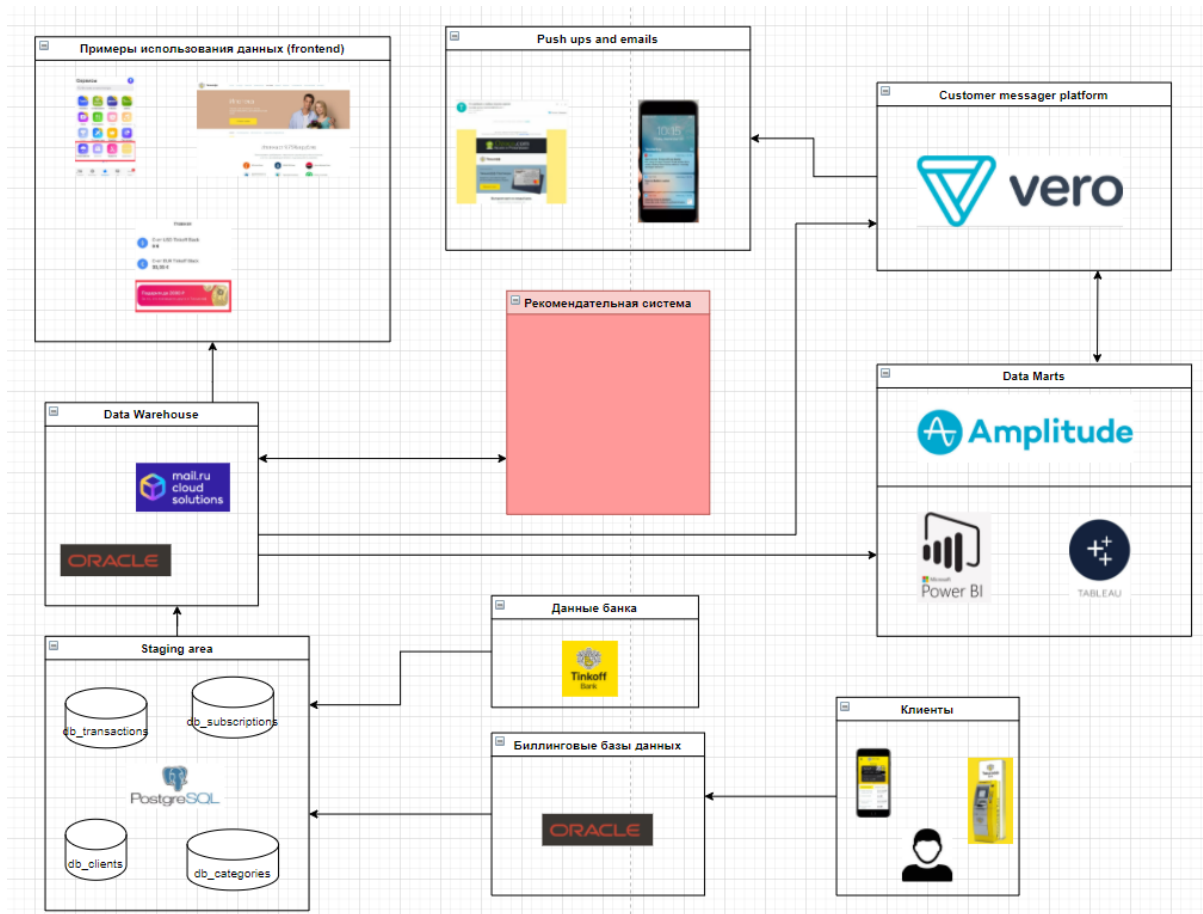


Рисунок 1.6.1. Диаграмма информационных процессов

- Рекомендательная система записывает данные в DWH, откуда Vero подтягивает информация и отправляет push-up уведомления. Очень важна аналитика открытия рассылок и последующие активации пользователей, для этого можно использовать программу Amplitude (Web, IOS, Android);
- CRM-маркетолог может добавлять правила рассылок в Vero, делать ручные рассылки;
- Frontend будет просто брать данные с DWH и визуализировать у себя.

## 2. Описание требований к ИТ-решению

### 2.1. Целевой принцип работы рекомендательной системы

На данный момент на главном экране приложения имеются различные баннеры, которые показывают персонализированную информацию о сумме возможных одобренных кредитах, новых акциях компании и выгодных предложениях (см. Рисунок 2.2.1).

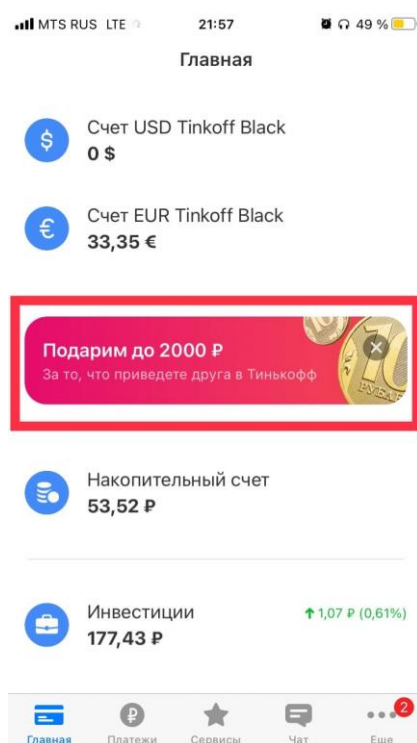


Рисунок 2.1.1. Пример баннера в главном меню

Для создания более персонализированной рекламы услуг партнеров должен производиться анализ транзакций клиентов и категорий сервисов партнеров. Рекомендательная модель, основываясь на вышеупомянутых данных, должен выдавать пул рекомендаций, в которой для каждого клиента соотносится предпочтительная для показа категория сервиса партнера.

Для заключения сделки по рекламной кампании потенциальный партнер должен перейти на сервис разрабатываемого ИТ-решения. В



разделе “Создать заказ” партнером создается заявка, в которой указывается необходимая информация (название компании, электронная почта, категория сервиса, желаемые даты начала и окончания рекламы). После создания заявки, используя информацию о категории сервиса и пула рекомендаций, который ИТ-решение создает и обновляет каждый день, модель определяет, каким пользователям приложения банка следует показывать данную рекламу. После формирования списка таких пользователей потенциальному партнеру предлагается краткая аналитика по заказанной рекламе, а заявка идет на рассмотрение специалисту банка.

## **2.2. Функциональные требования**

### **2.2.1. Требования к рекомендательной модели**

1. Рекомендательная модель должна создавать рекомендации, основываясь на транзакционных данных клиентов и категориях услуг партнеров;
2. Рекомендательная модель должна быть основана на методах машинного обучения;
3. Рекомендательная модель должна инициировать процессы создания рекомендаций и обновления модели после 1-го заказа в рамках 1-го дня;
4. Рекомендательная модель должна выдавать пул клиентов банка с категориями сервисов партнеров, рекламу которых им рекомендуется показать.

### **2.2.2. Требования к интерфейсу сервиса ИТ-решения**

1. Чтобы создать заявку на покупку рекламы, пользователь должен перейти по вкладке “Создать заявку”;
2. Чтобы получить демоверсию рекламы, пользователь должен перейти на главную страницу и нажать на кнопку “Показ демоверсии рекламы”;
3. Чтобы получить информацию о категориях сервисов, с которыми работает ИТ-решение, пользователь должен перейти по вкладке “Категории”;
4. Чтобы получить информацию об API сервиса, пользователь должен перейти по вкладке “API”;
5. Чтобы связаться с сотрудниками сервиса, пользователь должен перейти по вкладке “Контакты”;
6. Чтобы войти в систему сервиса со стороны администратора, пользователь должен перейти по вкладке “Управление” и пройти авторизацию.

### **2.2.3. Требования к предоставлению информации о заказе пользователю**

1. ИТ-решение должно выдавать пользователю следующую общую информацию:
  - Код заказа;
  - Название компании пользователя;
  - Категорию сервиса пользователя;
  - Электронную почту пользователя;
  - Количество клиентов, которое будет охвачено;
  - Даты начала и окончания рекламы;
  - Ожидаемая цена рекламной услуги;
  - СРС.

- СРА;
  - Статус заказа.
2. ИТ-решение должно выдавать краткую аналитика в виде диаграмм по заказу:
    - количество охваченных клиентов в зависимости от пола;
    - количество охваченных клиентов в зависимости от возраста;
    - ожидаемые показатели конверсии и кликабельности рекламного баннера.
  3. Чтобы скачать данные о заказе, пользователь должен нажать на кнопку “Скачать данные о заказе”.

### **2.3. Нефункциональные требования**

#### **2.3.1. Требования к информационной модели рекомендательной системы**

1. Должны быть настроены ETL-процессы данных из баз данных (клиентов, транзакций, услуг партнеров, подписок у клиентов);
2. ИТ-решение должно быть подключено к СУБД PostgreSQL;
3. Система должна проверять целостность и корректность полученных данных;
4. Система должна быть представлена в виде web-ресурса.

#### **2.3.2. Требования к API рекомендательной системы**

1. Проектирование API сервиса должно быть основано на стандарте REST API;
2. Весь функционал сервиса рекомендательной системы должен быть доступен в API;
3. Документация API должна лежать в открытом доступе.

### **2.3.3 Требования к рекомендательной модели**

1. Модель должна быть реализована на языке программирования Python;
2. Качество работы модели должно оцениваться метрикой F-score@k;
3. Значение метрики качества модели должно быть больше 80%.

### 3. Описание ИТ-решения

#### 3.1. Критерии выбора ИТ-решения

В качестве критериев для описания ИТ-решения были выбраны следующие:

- Денежные затраты на реализацию;
- Скорость адаптации системы;
- Сроки реализации;
- Стэк технологий;
- Вычислительная сложность;
- Скорость адаптации под новых партнеров;
- Уровень ожидаемой конверсии.

#### 3.2. Сравнительный анализ ИТ-решения

Критерии	Кластеризация и сегментация с помощью Microsoft Azure Machine Learning Studio	Формирование рекомендации по выделенным вручную правилам	Создание рекомендательной системы на основе применения машинного обучения
Денежные затраты на реализацию	Ежемесячно – 700 тыс. руб.: <ul style="list-style-type: none"><li>● Использование облачной службы - 300 тыс. руб.;</li></ul>	Ежемесячно - 800 тыс. руб.: <ul style="list-style-type: none"><li>● Команда, отвечающая за формирование</li></ul>	Ежемесячно – 800 тыс. руб.: <ul style="list-style-type: none"><li>● Команда ML-разработки - 2 * 300 тыс. руб.;</li></ul>

	<ul style="list-style-type: none"> <li>● Фонд оплаты труда (ФОТ) – 200 тыс. руб. * 2.</li> </ul>	последовательность правил для рекомендаций.	<ul style="list-style-type: none"> <li>● Data Engineer - 200 тыс. руб.;</li> </ul>
Скорость адаптации системы	<p>Высокая:</p> <ul style="list-style-type: none"> <li>● Подключение к PowerBI;</li> <li>● Несовместимость баз данных Oracle и MSSQL.</li> </ul>	<p>Низкая:</p> <p>уже интегрирована</p>	Средняя
Сроки реализации	5 мес.	Уже функционирует	6 мес.
Стэк технологий	Python, PowerBI	Excel, Tableau	Python, C++, PostgreSQL
Вычислительная сложность	Высокий	Уже закуплено	Средний
Скорость адаптации под новых партнеров	Средняя	Низкая	Высокая
Уровень ожидаемой конверсии	Средняя	Низкая	Высокая

### 3.3. Обоснование выбранного ИТ-решения

Для реализации проекта был выбран вариант ИТ-решения “Создание рекомендательной системы на основе применения машинного обучения” и разработана схема целевой архитектуры комплекса ИТ-решения.

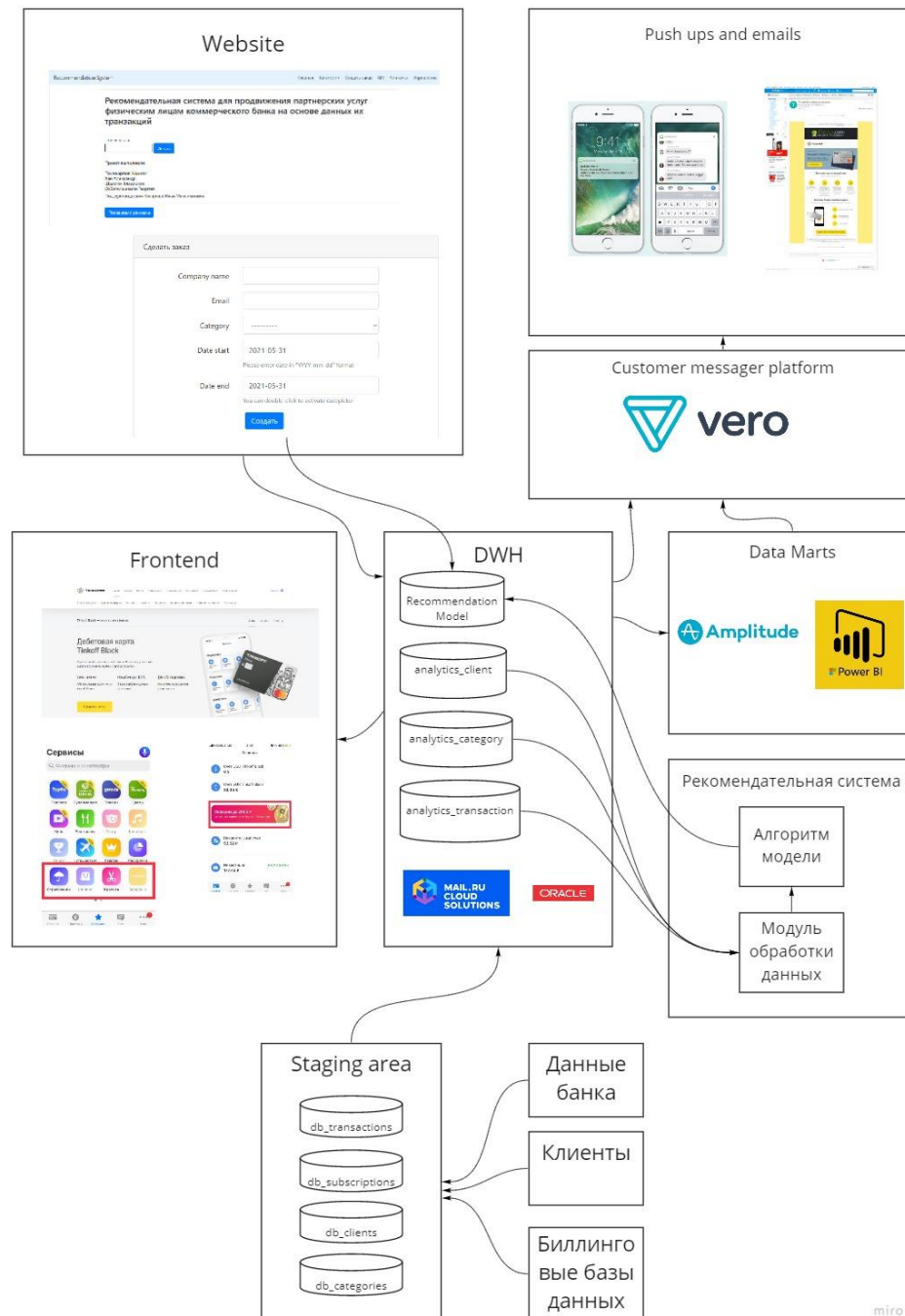


Рисунок 3.3.1. Целевая архитектура комплекса ИТ-решений

**Вычислительная сложность:** так как компания “Тинькофф Банк” сфокусирована на дистанционном обслуживании в формате онлайн, она

имеет достаточные крупные вычислительные мощности и необходимое аппаратное обеспечение. При нашем ИТ-решении уровень вычислительной сложности будет средним, так как система разрабатывается под ИТ-инфраструктуру и используемый стек технологий банка с целью автоматизации и оптимизации протекающих процессов в работе рекомендательной системы.

**Денежные результаты на реализацию:** в ходе улучшения системы рекомендаций уровень затрат на реализацию модели останется почти на прежнем уровне.

**Сроки реализации:** реализация ИТ-решения осуществляется ненамного дольше альтернативных вариантов.

**Уровень ожидаемой конверсии:** у выбранного ИТ-решения высокий ожидаемый уровень конверсии по сравнению с другими ИТ-решениями.

**Сложность адаптации системы:** Так как проект разрабатывается для ИТ-инфраструктуры банка, то она будет быстрее всего интегрирована в систему.

**Стек технологий:** для реализации выбранного ИТ-решения используются наиболее развитые и распространенные языки программирования. Специалисты по данным языкам есть как и внутри компании Тинькофф, так и в избытке на рынке труда.

**Скорость адаптации под новых партнеров:** данное ИТ-решение позволяет быстро подстроиться под изменение услуг партнеров и самих партнеров, так как модель рекомендательной системы не привязана к самим услугам напрямую, а только к категориям услуг, которые предсказывает модель.

**Оценка ожидаемой конверсии:** выбранное ИТ-решение имеет более высокую ожидаемую конверсию перехода клиентов по рекомендованным услугам по сравнению с другими вариантами.



### 3.4. Описание сервиса ИТ-решения

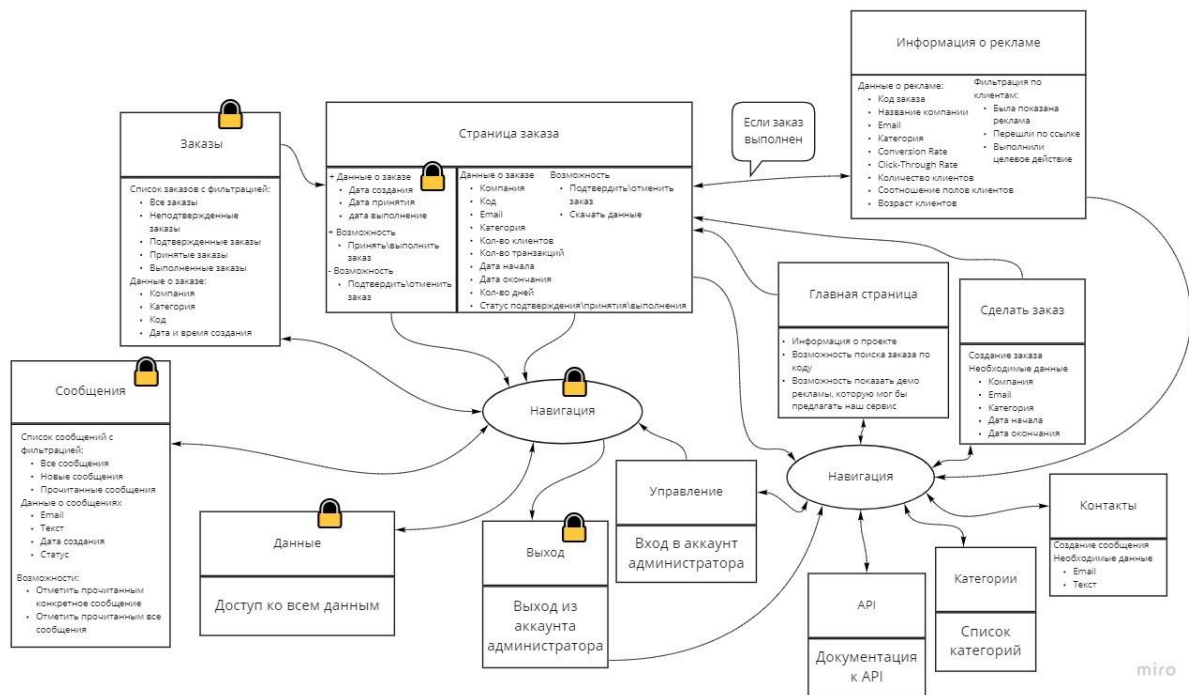


Рисунок 3.4.1. Диаграмма архитектуры сайта сервиса

Сайт сервиса разделен на две составляющие: пользовательскую и административную. На диаграмме административные страницы отмечены замком.

**Пользовательская составляющая** - это те страницы сайта, которые доступны всем клиентам, без авторизации. В ней вам доступны страницы:

- “Категории” - содержит информацию о доступных категориях для рекламы.

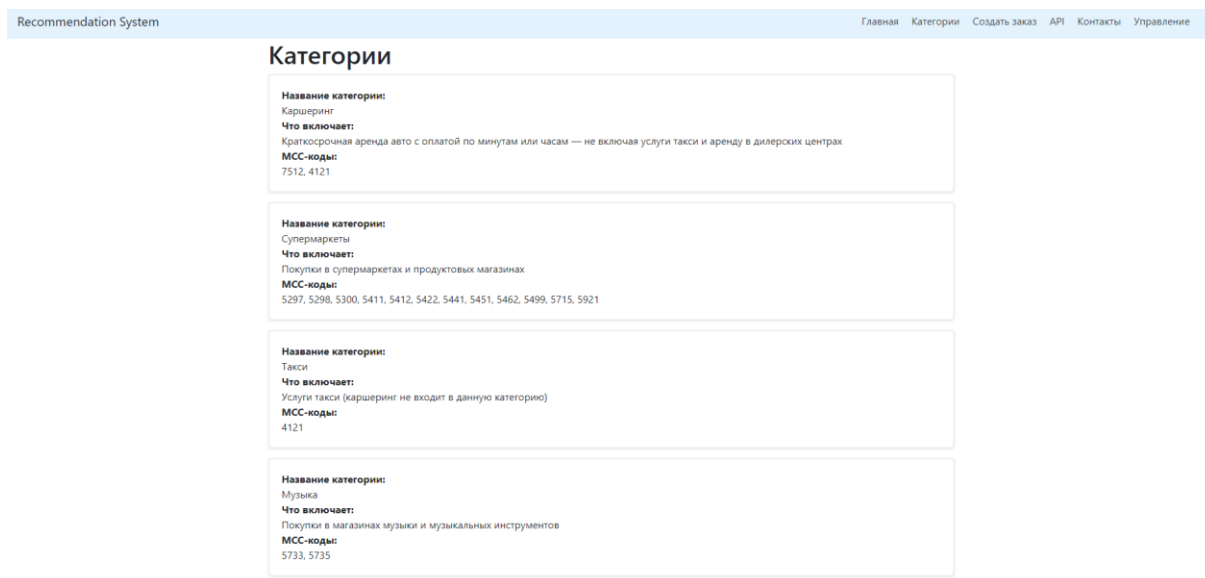


Рисунок 3.4.2. Страница “Категории”

- “API” - содержит документацию к API сервиса, для предоставления доступа к нашей системе другим сервисам и приложением банка.

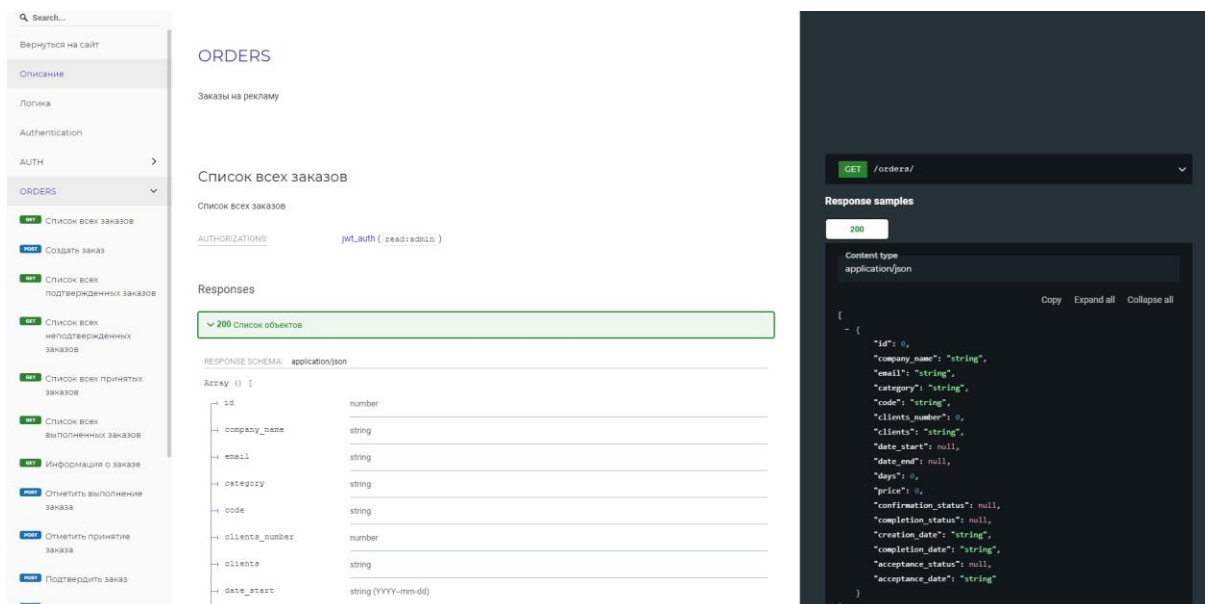


Рисунок 3.4.3. Страница “API”

- “Создать Заказ” - дает возможность клиентам создавать новые заказы. Пользователю необходимо ввести название компании, адрес электронной почты, даты начала и окончания действия рекламы и категорию, по которой должна проходить реклама. При каждом создании заказа, происходит подбор клиентов, которых выбирает рекомендательная система. После создания заказа, вас перекинет на

страницу “Заказ”, где будет представлена дополнительная информация, помимо ранее введенных данных: код заказа, количество клиентов, которым будет предоставлена реклама и небольшая статистика о них (соотношение полов клиентов, распределение возрастов клиентов), ожидаемая цена услуги, а также текущий статус заказа. У пользователей есть возможность подтвердить или отменить заказ. Однако если заказ уже принят сотрудниками, то отменить его нельзя. Также у клиентов есть возможность скачать данные о заказе в формате текстового файла.

Рисунок 3.4.4. Страница “Создать заказ”

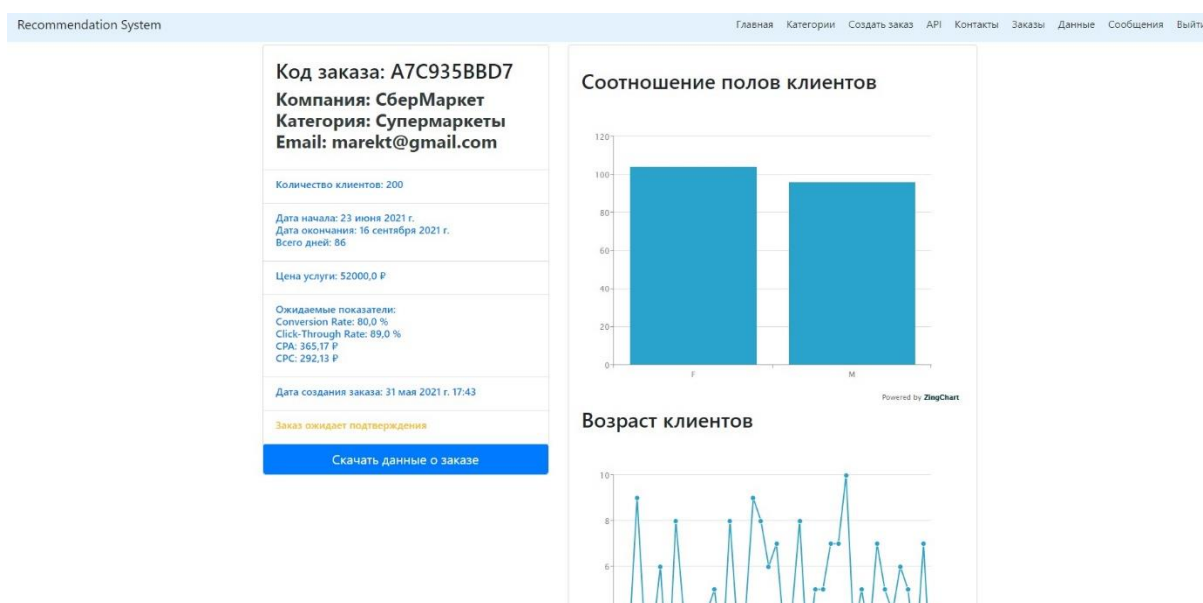


Рисунок 3.4.5. Страница “Заказ”

- “Главная” - содержит систему поиска заказа по коду, после ввода которого вас перекинет на страницу соответствующего заказа. Также доступна кнопка “Показ демоверсии рекламы”, которая покажет пример баннера рекламы, которую наш сервис мог бы предлагать.

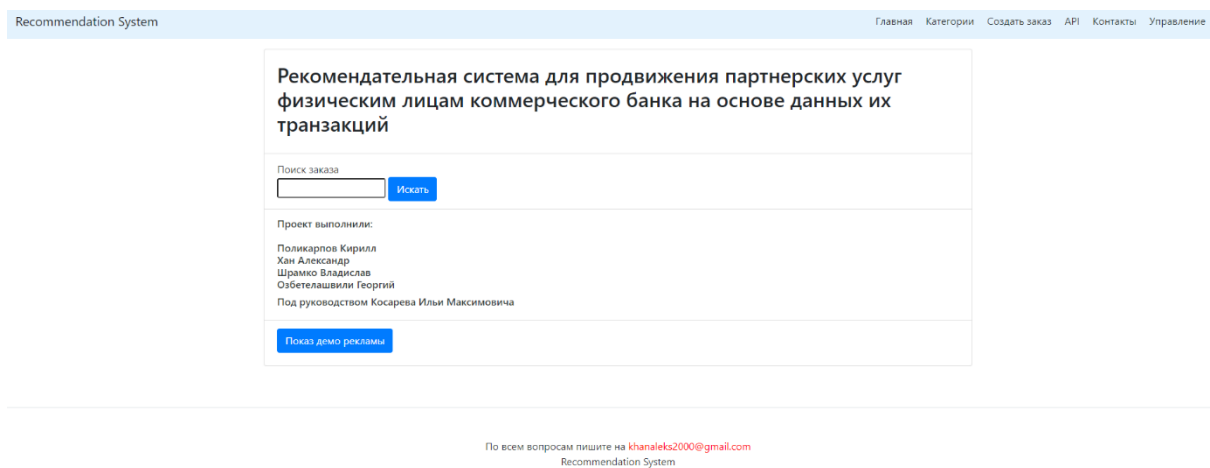


Рисунок 3.4.6. Страница “Главная”

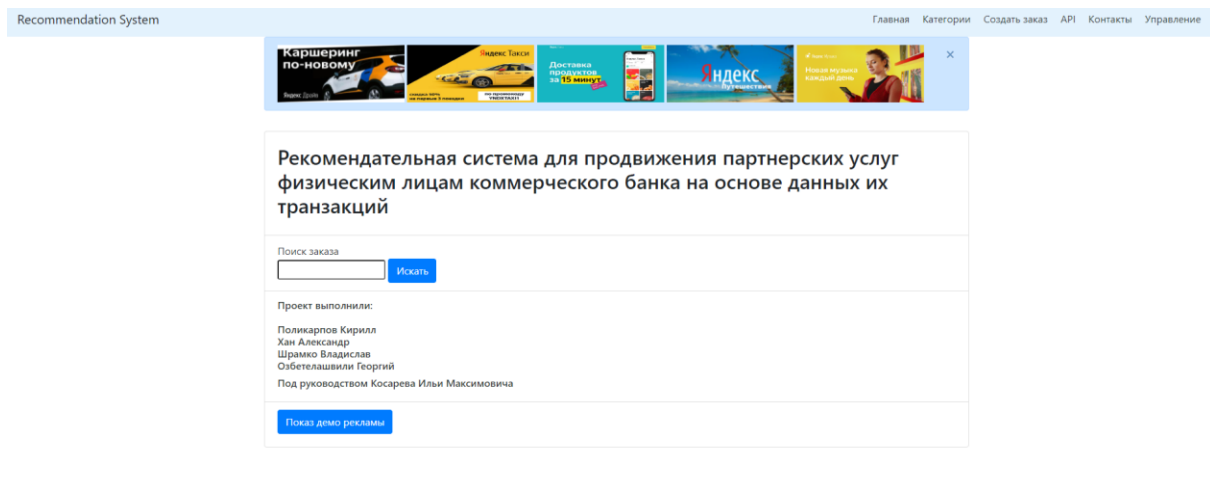


Рисунок 3.4.7. Страница “Главная” с показом демоверсии рекламы

- “Контакты” - дает возможность пользователям связаться с сотрудниками сервиса.

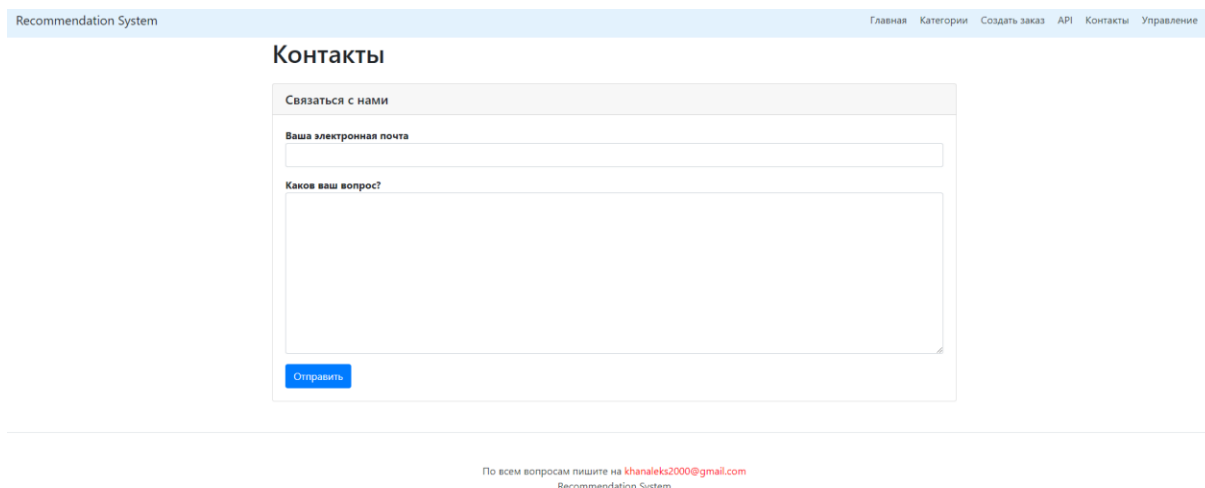


Рисунок 3.4.8. Страница “Контакты”

- “Управление” - дает возможность пользователю авторизоваться в качестве администратора.

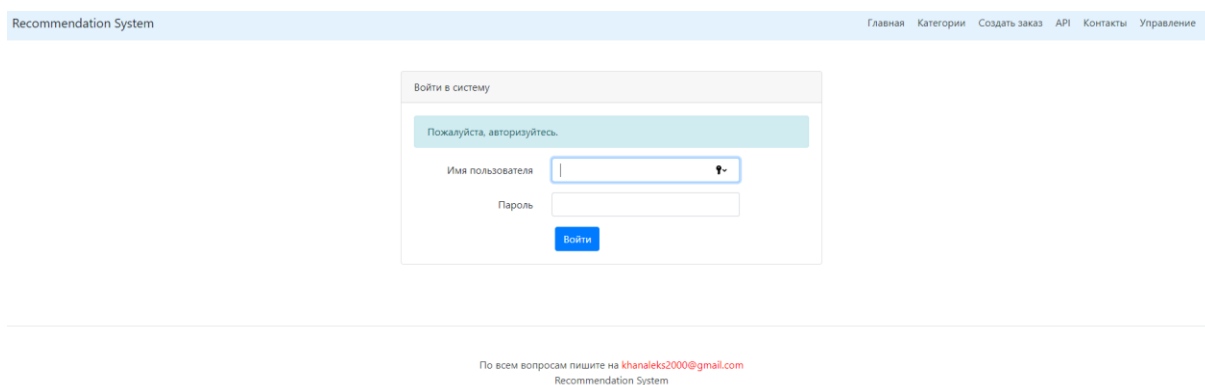


Рисунок 3.4.9. Страница “Управление”

Когда заказ только создан, то он ожидает подтверждения со стороны клиента. После подтверждения заказа, теперь уже со стороны сотрудников, заказ ожидает принятия. После принятия заказа, сервис начнет его выполнять. После выполнения заказа, сотрудники могут соответственно обновить статус заказа.

Если заказ выполнен, то на его странице появляется доступ к новой странице “Информация о рекламе”. Это страница будет содержать данные о текущем состоянии рекламы: число клиентов которым была показана реклама, число клиентов которые перешли по ссылке и число клиентов, которые выполнили целевое действие. Будут доступны показатели conversion rate и click-through rate, а также статистика о каждой группе

клиентов в форме данных о соотношении полов и возрастов клиентов. На данный момент показатели являются случайными данными, так как мы в действительности не проводим рекламу, и соответственно не имеем возможность получить такие данные.

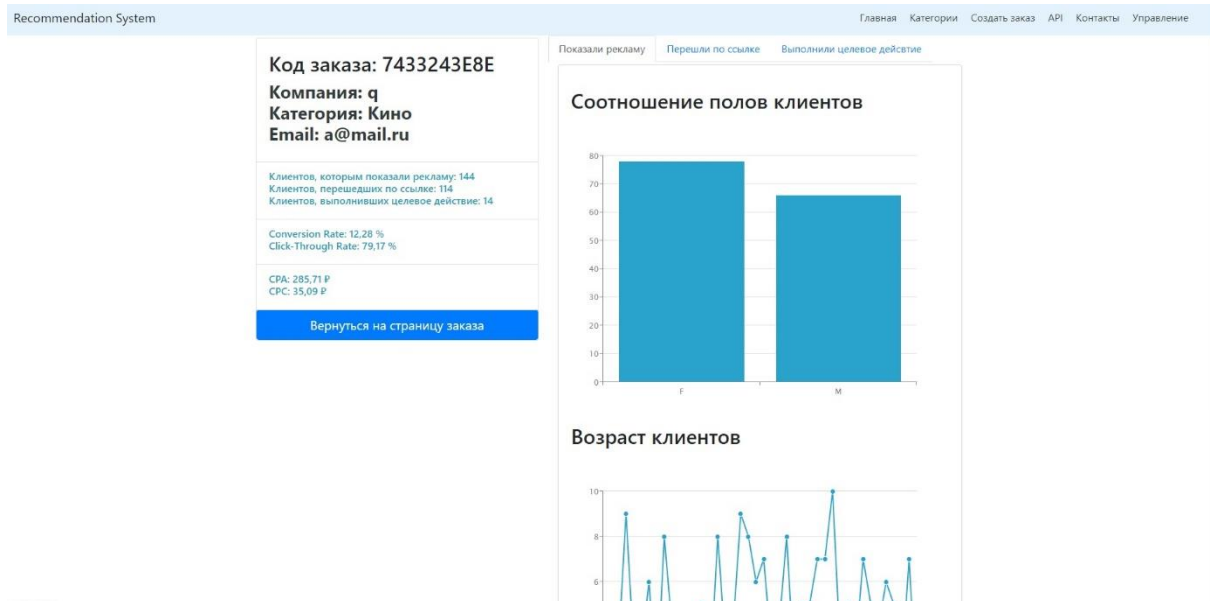
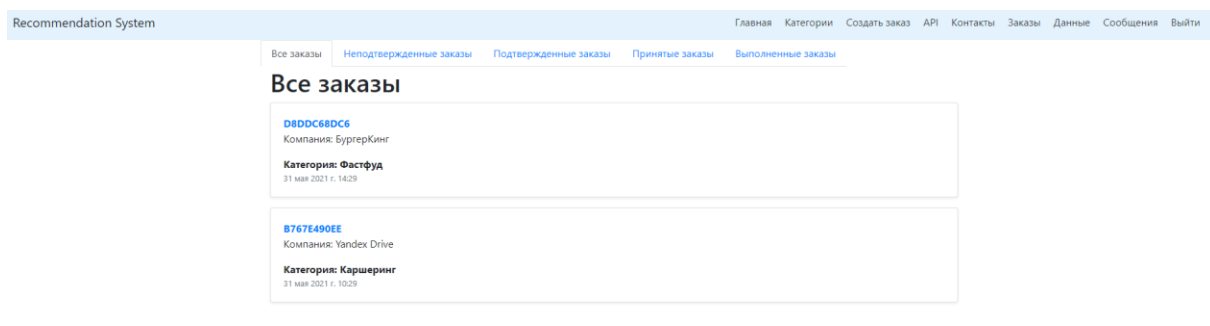


Рисунок 3.4.10. Страница “Информация о рекламе”

Административная составляющая сайта, помимо пользовательской части, дополнительно содержит:

- “Заказы” - содержит список заказов, с фильтрацией: все заказы, неподтвержденные, подтвержденные, принятые и выполненные заказы.



По всем вопросам пишите на [khanaleks2000@gmail.com](mailto:khanaleks2000@gmail.com)  
Recommendation System

Рисунок 3.4.11. Страница “Заказы”

- “Сообщения” - содержит список сообщений, которые были отправлены пользователями, с фильтрацией: все сообщения, новые и прочитанные сообщения.

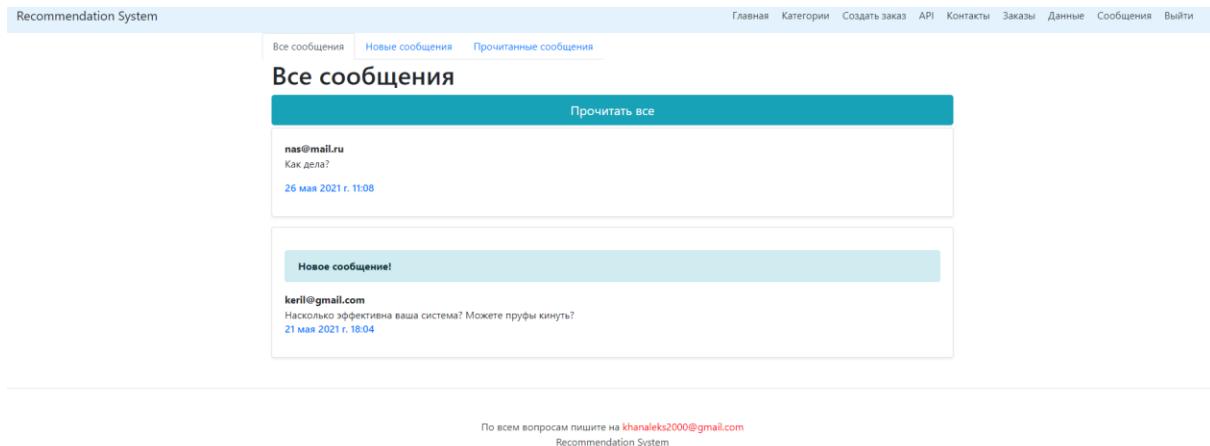


Рисунок 3.4.12. Страница “Сообщения”

- “Данные” - страница администрирования, которая дает доступ на чтение и изменение всех данных сайта, то есть доступ к базе данных.

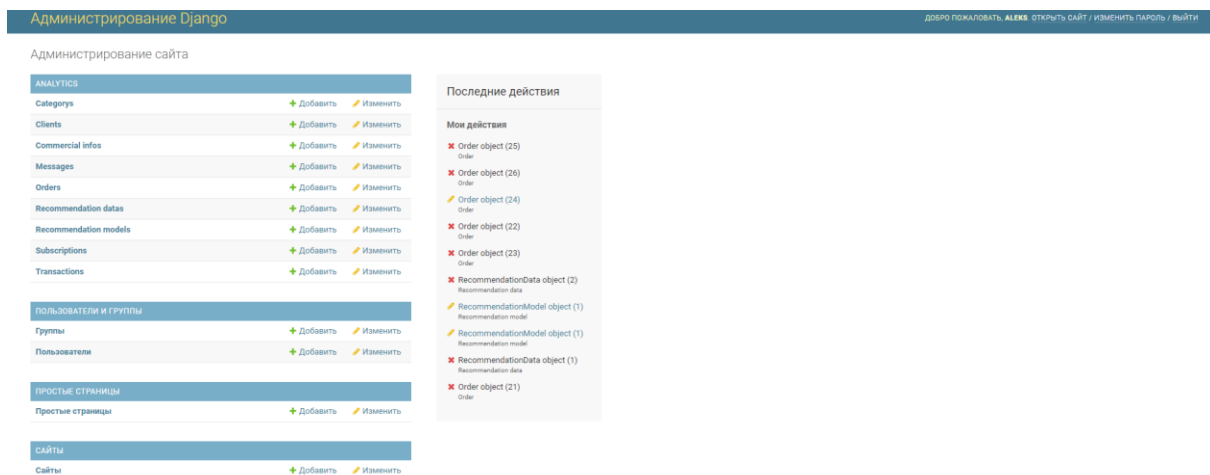


Рисунок 3.4.13. Страница “Данные”

- “Выйти” - дает возможность деавторизоваться пользователю.

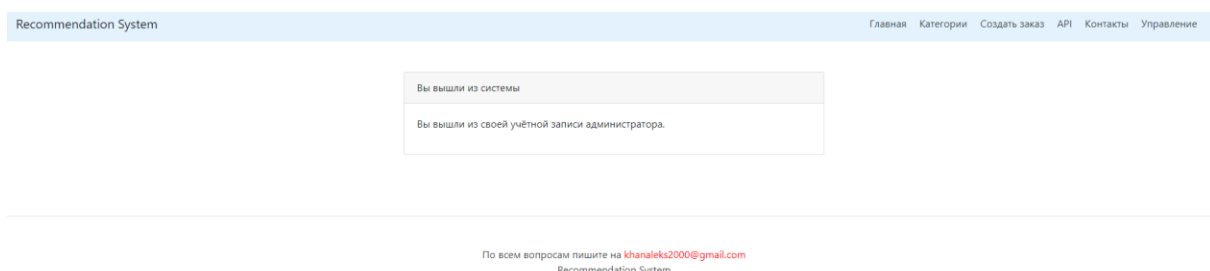


Рисунок 3.4.14. Страница “Выйти”

Хранение данных рекомендательной модели происходит при помощи сущности RecommendationModel. Она содержит поля для хранения даты и времени последнего обновления, ведь модель необходимо обновлять ежедневно. И так данные о категориях и клиентах модели хранятся в формате словаря, то необходимо было добавить дополнительную сущность RecommendationData с ManyToMany связью.

Для хранения дополнительных данных о заказах и модели, в базу данных было необходимо добавить новые сущности (см. **Приложение 7**).

### **3.5. Описание модели рекомендательной системы**

#### **3.5.1. Используемые инструменты**

Этапы работы с данными и исследования моделей рекомендательной системы производились с помощью языка Python (3.7.4) в среде Jupyter Notebook (Anaconda3) (см. **Приложение 8**). Для анализа и обработки данных использовались библиотеки NumPy (1.16.5) и Pandas (0.25.1). С помощью библиотеки SciPy (1.4.1) строилась модель со скрытыми переменными, а с помощью библиотек Scikit-learn (0.23.1) и CatBoost (0.25.1) – контентные модели случайного леса и градиентного бустинга, соответственно. Для визуализации данных использовались библиотеки Matplotlib и Seaborn.

На этапе развертывания скрипт итоговой модели model.py был реализован на сервере с помощью языка Python (3.8.5). Соединение между хранилищем данных PostgreSQL и моделью производилось с помощью библиотеки psycopg2-binary (2.8.6). Версии дополнительных используемых библиотек:

- NumPy: 1.20.3;
- Pandas: 1.2.4;



- SciPy: 1.6.3.

### **3.5.2. Методология ведения и построения модели рекомендательной системы**

Реализация любого проекта складывается из определенных этапов, соответствующих специфике проекта, поэтому важно выбрать подходящую методологию для его ведения, так как от этого во многом зависит результат работы. В качестве подхода построения модели рекомендательной системы была использована методология CRISP-DM. Данный фреймворк разделяет процесс построения модели машинного обучения на несколько этапов, которые могут корректироваться в зависимости от результатов каждого: понимание бизнеса (business understanding), подготовка данных (data understanding), моделирование (modeling), оценка (evaluating) и развертывание (deployment).

### **3.5.3. Постановка формальной задачи машинного обучения**

На данном этапе рассмотрим понимание поставленной задачи не только с точки зрения бизнеса, но и с точки зрения формальной постановки задачи машинного обучения. Ключевой задачей рекомендательной системы является предсказание наиболее релевантных категорий товаров или услуг клиентам банка по истории их транзакций. Зададим множество категорий товаров и услуг, которые банк может идентифицировать при совершении клиентом какой-либо транзакции как:

$$C = \{c_1, \dots, c_m\},$$

где  $c_j$  – наименование  $j$ -ой категории продукта.

Всех клиентов банка обозначим множеством пользователей следующим образом:

$$U = \{u_1, \dots, u_n\},$$

где  $u_i$  –  $i$ -ый клиент банка.

Также введем матрицу  $P$ , которая позволит отразить значения всех имеющихся пар  $(u, c)$ :

	$c_1$	$\dots$	$c_m$
$u_1$	$(u_1, c_1)$	$\dots$	$(u_1, c_m)$
$\dots$	$\dots$	$\dots$	$\dots$
$u_n$	$(u_n, c_1)$	$\dots$	$(u_n, c_m)$

Особенностью поставленной задачи является то, что предсказание того, какую категорию товара или услуги стоит вероятнее всего порекомендовать приобрести пользователю, идет не из множества тех категорий, с которыми он еще не взаимодействовал, а из всего множества категорий, что отличается от классического варианта работы рекомендательной системы. Это связано с тем, что пользователь время от времени может повторно приобретать товары или услуги одной и той же категории, поэтому одним из ограничений реализации рекомендательной системы является рекомендация пользователю категорий товаров и услуг как из того, что он уже приобретал, так и из того, что еще нет. На итоговую реализацию модели рекомендательной системы также влияет условие того, что для каждого пользователя будет строиться предсказание только одной категории товара или услуги в связи тем, что реклама только одного партнера банка может быть показана пользователю в приложении в течение определенного промежутка времени.

Процесс реализации модели рекомендательной системы в общем виде можно разделить на следующие этапы:

1. Формирование признаков для каждой пары  $(u, c)$ :  $x_1^{uc}, \dots, x_d^{uc}$ , где  $d$  – количество выделенных признаков, которым описывается каждая из пар  $(u, c)$ ;
2. Обучение алгоритма модели на каждой паре  $(u, c)$  из матрицы  $P$ :  $a(u, c) = a(x_1^{uc}, \dots, x_d^{uc})$ ;
3. По окончании обучения получаем модель  $a(u, c)$ , которую можно выводить на этап применения, где для каждого  $i$ -ого пользователя  $u_i$  получаем распределение категорий товаров и услуг, которые ранжируются по убыванию релевантности для него, и отбирается  $k$  категорий (в рамках данной задачи  $k = 1$ ), которые и будут в итоге порекомендованы пользователю:

$$u_i \rightarrow \begin{matrix} (u_i, c_1) & a(u_i, c_1) \\ \dots & \dots \\ (u_i, c_m) & a(u_i, c_m) \end{matrix} \rightarrow [c_{(1)}, \dots, c_{(k)}], \dots, c_{(m)}.$$

### 3.5.4. Работа с данными

#### 3.5.4.1. Понимание данных

Данная стадия уже частично была рассмотрена в данной работе по тому, каким образом собираются данные и какие данные нужны для построения модели рекомендательной системы. Помимо этого, на данном этапе была проведена первоначальная оценка качества полученных данных, в результате которой были подкорректированы типы данных, заполнены пропуски, переименованы столбцы в таблицах и удалены неинформативные данные. С точки зрения поставленной задачи были удалены из анализа транзакции, в которых содержатся категории «Переводы» и «Другое», поскольку данные категории нельзя связать со спецификой решаемой задачи. Также из анализа были удалены транзакции, которые имеют тип «Positive» в связи с тем, что для решения данной задачи рассматриваются

только те транзакции, которые связаны с покупкой пользователем какой-либо категории товара или услуги.

#### 3.5.4.2. Подготовка данных

Целью данного этапа является с помощью инструментов и методов предобработки данных получить из сырых данных итоговый набор данных, который в дальнейшем будет использоваться на этапе моделирования, поэтому важно качественно провести подготовку данных, так как от этого во многом зависит качество моделей машинного обучения.

В ходе предобработки данных было обнаружено то, что в данных транзакций у некоторых пользователей встречаются повторяющиеся покупки каких-либо категорий товаров или услуг, однако для построения модели рекомендательной системы необходимо работать только с одной величиной, характеризующей все взаимодействия пользователя с определенной категорией товара или услуги. Для решения данной проблемы для каждого пользователя посчитаем количество того, сколько раз он взаимодействовал с каждой категорией товара или услуги по данным его транзакций. Однако, после группировки пар  $(u_i, c_j)$  по количеству взаимодействий сталкиваемся с особенностью транзакционных данных покупок того, что пользователь может покупать одни и те же услуги с разной частотой, то есть отсутствует верхний предел на то, сколько раз пользователь может взаимодействовать всего с определенным товаром или услугой. Для того, чтобы решить данную проблему проведем масштабирование данных, но не по каждому клиенту отдельно, а по всем клиентам сразу. В качестве подхода масштабирования данных был использован `minmaxscaler`.

Помимо этого, одной из проблем рекомендательных систем является проблема холодного старта, поэтому для того, чтобы модели

рекомендательных систем наиболее качественно смогли провести обучение, оставим только тех пользователей, которые взаимодействовали больше, чем с четырьмя категориями товаров и услуг.

По окончании предобработки данных последним шагом является разбиение данных на тренировочную и тестовые выборки. С целью учета трендов и сезонности в покупках категорий товаров и услуг разобьем данные по времени, но перед этим добавим к преобразованным данным последнее время взаимодействия для каждой пары user-category, чтобы была возможность разделения их по времени. Для разделения выборки на обучение и контроль в отношении 80:20 был выбран следующий временной порог - 2020-12-30.

### **3.5.5. Выбор модели рекомендательной системы**

Данный этап включает себя рассмотрение и применение определенных алгоритмов машинного обучения, соответствующих специфике решаемой задачи, на данных, полученных на этапе их подготовки. Также одной из задач этапа моделирования является оптимизация используемых моделей с помощью подбора гиперпараметров, обеспечивающих наибольшее значение поставленной метрики качества для каждой из рассмотренных моделей.

В процессе моделирования были рассмотрены следующие модели рекомендательной системы:

- Модель по популярности (popularity model);
- Модели коллаборативной фильтрации (collaborative filtering):
  - Memory-based;
  - Модель со скрытыми переменными (latent factor model);
- Контентная модель (content-based model).

Рассмотрим данные модели рекомендательной системы с теоретической точки зрения в рамках поставленной задачи предсказания для каждого пользователя наиболее релевантной категории товара или услуги с учетом введенных ограничений, влияющих на их реализацию.

#### **3.5.5.1. Модель по популярности**

Данная модель является самой простой моделью рекомендаций, которая рекомендует наиболее популярные объекты. В связи с этим возьмем данную модель в качестве основной, относительно которой будет происходить сравнение с остальными.

Для определения наиболее популярных категорий товаров и услуг посчитаем популярность каждой категории, как сумму всех количеств взаимодействия с ней, и отсортируем категории по их популярности в порядке убывания. Далее модель рекомендует наиболее популярные товары и услуги каждому пользователю.

#### **3.5.5.2. Коллаборативная фильтрация**

Особенностью метода коллаборативной фильтрации является то, что данный метод учитывает схожесть пользователей и товаров между собой. Для работы с моделями коллаборативной фильтрации необходимо привести данные к матрице  $P$ , которая отражает все пары взаимодействий между пользователями и категориями, а в качестве значений данных пар хранит нормализованные количества взаимодействий.

##### **3.5.5.2.1. Memory-based model**

Данный подход коллаборативной фильтрации разделяется на два алгоритма: user-based и item-based. Идея алгоритма user-based заключается

в том, что он сравнивает схожесть двух пользователей, рассматривая только те категории, с которыми оба пользователя  $u$  и  $v$  взаимодействовали:

$$C_{uv} = \{c \in C \mid \exists p_{uc} \ \& \ \exists p_{vc}\}.$$

Для оценки схожести двух пользователей используется формула корреляции Пирсона, которая рассчитывается для каждой пары пользователей следующим образом:

$$w_{uv} = \frac{\sum_{c \in C_{uv}} (p_{uc} - \bar{p}_u)(p_{vc} - \bar{p}_v)}{\sqrt{\sum_{c \in C_{uv}} (p_{uc} - \bar{p}_u)^2} \sqrt{\sum_{c \in C_{uv}} (p_{vc} - \bar{p}_v)^2}}.$$

После формирования оценок схожести пользователей друг с другом алгоритм для каждого пользователя  $u_0$  оставляет только тех пользователей, которые похожи на него больше всего по определенному порогу значения схожести  $\alpha$ , который может задаваться самостоятельно:

$$U(u_0) = \{v \in U \mid w_{u_0v} > \alpha\}.$$

Далее алгоритм для каждого пользователя  $u_0$  рекомендует те категории  $c \in C$ , у которых наибольшая доля пользователей среди пользователей из множества  $U(u_0)$ .

Принцип работы алгоритма item-based строится таким же образом как user-based, где отличие только в том, что идет оценка схожестей не пользователей, а категорий.

### 3.5.5.2.2. Модель со скрытыми переменными

Модель со скрытыми переменными является одним из видов коллаборативной фильтрации, суть которой в том, что для каждого пользователя и каждой категории строятся вектора таким образом, чтобы скалярное произведение этих векторов приближалось к нормализованному количеству взаимодействий для каждой пары user-category:

$$l_u \in R^d$$

$$q_c \in R^d$$

$$\langle l_u, q_c \rangle \approx p_{uc},$$

где  $l_u$  – вектор пользователя;  $q_c$  – вектор категории;  $d$  – это размерность векторов, которую можно задавать самостоятельно.

Функционал качества модели со скрытыми переменными устроен следующим образом:

$$Q = \sum_{(u,c,p_{uc}) \in R} (p_{uc} - b_u - b_c - \langle l_u, q_c \rangle)^2 \rightarrow \min_{b_u, b_c, l_u, q_c}.$$

В данном функционале качества обучаются оба аргумента скалярного произведения  $l_u$  и  $q_c$ . В связи с тем, что данные вектора скалярно перемножаются, то возникает проблема сложности выпуклости функционала качества и, соответственно, проблема возникновения локальных экстремумов, поэтому вместо решения задачи нахождения глобального минимума в общем виде используются подходы приближения с помощью стохастического градиентного спуска (stochastic gradient descent, SGD), чередующихся наименьших квадратов (alternating least squares, ALS) или сингулярного разложения матрицы (singular value decomposition, SVD). В качестве способа приближения было использовано сингулярное разложение матрицы количества нормализованных взаимодействий.

### 3.5.5.3. Контентная модель

Контентная модель является альтернативным подходом к построению рекомендательной системы, идея которой заключается в том, что операции с данными происходят не с матрицей количества нормализованных взаимодействий, а с классической матрицей объекты-признаки, где каждый объект характеризуется парой user-category и содержит признаки, описывающие как пользователя, так и категорию. Кроме этого, признаки могут также описывать и саму пару целиком.

Данная модель рекомендательной системы позволяет поставить задачу классической классификации. В качестве целевой переменной будет



выступать само взаимодействие пар  $(u, c)$ , где отсутствующие взаимодействия отмечаются как отрицательные и выбираются случайным образом. Важно отметить, что так как модель оценивает каждую пару потенциального взаимодействия, то поэтому необходимо подготовить выборку из всех возможных пар из пользователей и категорий.

При построении контентной модели были выделены следующие признаки для каждой пары user-category:

- msc-code – количественная переменная, которая показывает количество msc-кодов, которые существуют для категории;
- gender – бинарная переменная, которая показывает пол пользователя;
- income – количественная переменная, которая показывает средний доход пользователя за последние 12 месяцев;
- expenses – количественная переменная, которая показывает средний объем расходов пользователя за последние 12 месяцев;
- credit – бинарная переменная, которая показывает взял ли пользователем кредит;
- deposit – бинарная переменная, которая показывает взял ли пользователем депозит;
- sum\_amount – количественная переменная, которая показывает то, на какую сумму было приобретено пользователем товаров или услуг определенной категории;
- subtype – категориальная переменная, которая показывает для пользователя то, какой был последний подтип транзакции для определенной категории;
- product\_company – категориальная переменная, которая показывает для пользователя, то у какой компании была совершена последняя транзакция для определенной категории;

- `days_from_registration` – количественная переменная, которая показывает то, сколько прошло дней с момента регистрации клиента в банке и совершением последней транзакции по определенной категории товара или услуги;
- `age` – количественная переменная, которая показывает то, какой возраст был у клиента на момент совершения последней транзакции по определенной категории товара или услуги.

Перед построением моделей необходимо убедиться в том, что в данных отсутствует мультиколлинеарность. По построенной матрице корреляций можно сделать вывод о том, что в тренировочных данных отсутствует данная проблема (см. **Приложение 9**).

В качестве моделей классификации были рассмотрены две наиболее сильные композиционные модели: модель случайного леса и модель градиентного бустинга. Модель случайного леса использует подход независимого обучения базовых моделей (в данном случае деревьев) на случайных подвыборках, образованных с помощью бэггинга и/или случайных подпространств, а подход модели градиентного бустинга состоит в последовательном обучении базовых моделей.

#### **3.5.5.4. Оценка**

В качестве метрик качеств для оценки рассмотренных моделей рекомендательной системы воспользуемся такими как `precision@1`, `recall@1` и `F-score@1`, так как по условию  $k = 1$ . В связи с тем, что строится рекомендация только одной категории для каждого пользователя, то в расчет не принималась метрика качества `nDCG`, поскольку она предназначена для учета порядка сделанных рекомендаций, а в рекомендации из одной категории отсутствует смысл учета какого-либо порядка.

Выбранные метрики качества рассчитываются следующим образом:

$$precision@1 = \frac{|R_u(1) \cap L_u|}{|R_u(1)|}$$

$$recall@1 = \frac{|R_u(1) \cap L_u|}{|L_u|}$$

$$F - score@1 = (1 + \beta^2) \frac{precision@1 * recall@1}{\beta^2 * precision@1 + recall@1},$$

где  $R_u(1)$  – список одной лучшей рекомендации алгоритма;  $L_u$  – список категорий, с которыми пользователь реально взаимодействовал.

При прогнозировании нам важнее то, чтобы модель рекомендательной системы не рекомендовала пользователям то, что им не интересно, поэтому строя и оценивая наши модели, будем использовать измененную метрику качества F-score, в которой дадим больший вес точности ( $\beta = 0.5$ ). При сравнении моделей возьмем метрику качества F-score@1 за основную, так как она позволяет объединить precision@1 и recall@1 с помощью гармонического среднего.

На тестовых данных модели показали следующие значения метрики качества F-score@1:

Модель	F-score@1
Модель по популярности	0.524
Memory-based (user-based)	0.855
Модель со скрытыми переменными	0.855
Модель случайного леса	0.077
Модель градиентного бустинга	0.029

Заметим, что лучшие значения метрики качества F-score@1 относительно модели по популярности получились у моделей коллаборативной фильтрации – 0.855. Поскольку на большем количестве рекомендаций категорий модель со скрытыми переменными показала лучшие значения метрики качества F-score, чем memory-based, поэтому возьмем ее за итоговую модель, которая будет использована в рекомендательной системе (см **Приложение 10**).

### **3.5.6. Развертывание модели рекомендательной системы**

Развертывание модели рекомендательной системы является заключительным этапом, на котором производится интеграция модели с другими компонентами системы и ее запуск в реальном секторе. Помимо этого, данный этап может включать в себя мониторинг и поддержку разработанной системы.

Итоговая архитектура взаимодействия модели рекомендательной системы с другими компонентами системы представлена следующим образом:

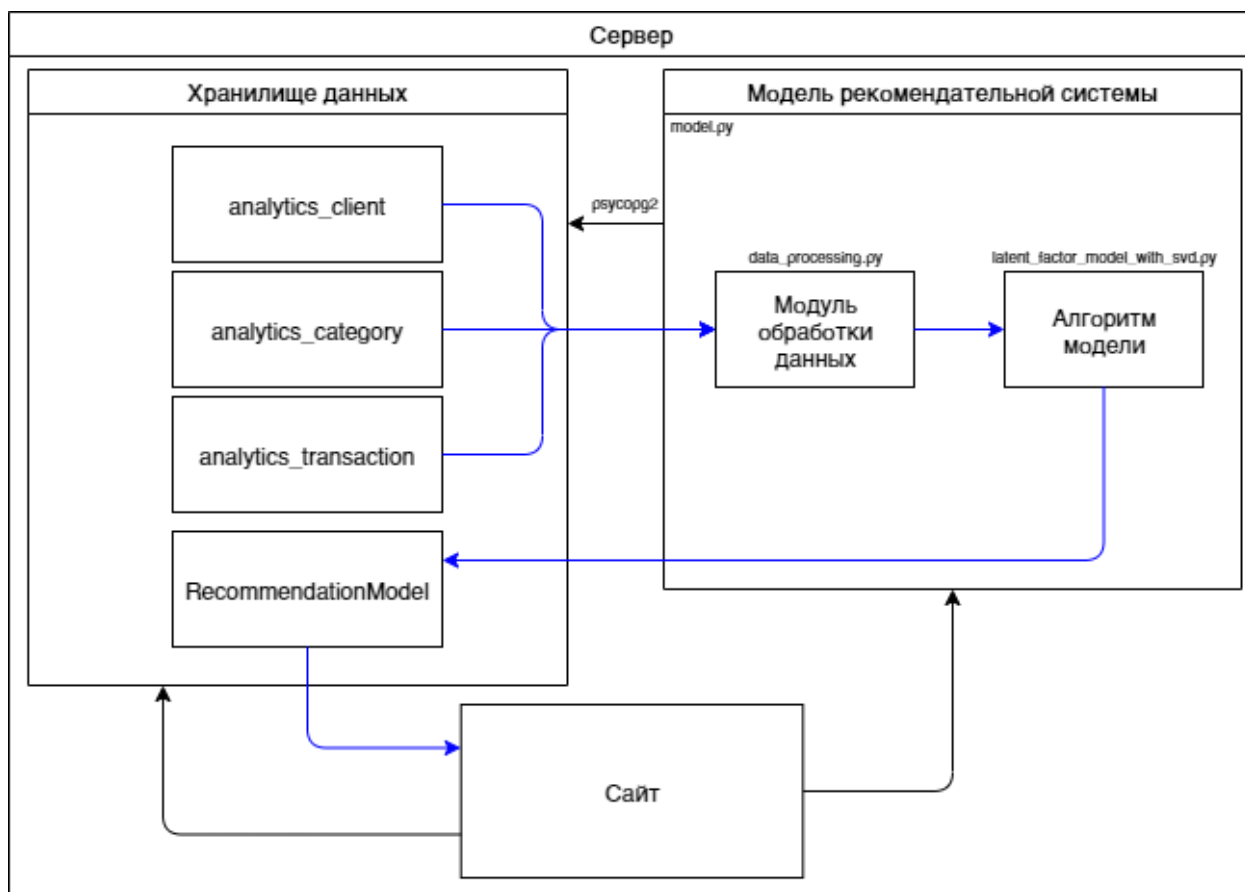


Рисунок 3.5.6.1. Диаграмма взаимодействия модели рекомендательной системы с ключевыми компонентами системы

Данный этап подразумевает под собой интеграцию итоговой модели рекомендательной системы в архитектуру рекомендательной системы. Развертывание модели происходило на удаленном сервере, на котором также расположены хранилище данных и сайт. Сначала был реализован скрипт `model.py`, состоящий из двух модулей: модуль обработки данных (`data_processing.py`) и сам алгоритм модели (`latent_factor_model_with.py`). Задача модуля обработки данных состоит в предобработке поступающих данных (`analytics_client`, `analytics_category` и `analytics_transaction`) из хранилища данных и преобразовании их в необходимый формат для модели. После того, как данные предобработаны, они поступают в модель, которая проводит обучение на них и возвращает рекомендации, где каждому клиенту соответствует одна из категорий товара или услуги. Выданные моделью рекомендации после сохраняются в хранилище данных

в таблицу RecommendationModel. Поскольку целью работы является реализация рекомендательной системы, был выбран упрощенный вариант вызова модели – модель рекомендательной системы вызывается через сервис сайта на момент создания первого заказа за день, иначе сайт забирает уже готовые рекомендации из хранилища данных.

### **3.6. Перспективы развития**

В будущем для клиентов банка можно будет добавить функцию выбора 5 категорий услуг, по которым они хотят получать предложения со стороны партнеров банка. Такой подход позволит повысить лояльность клиентов, использовать информацию о предпочтениях клиентов банка с целью улучшения алгоритма рекомендательной системы. Кроме этого, возможно внедрения онлайн-мониторинга и аналитики по каждой рекламной кампании.

Для улучшения качества рекомендательной системы можно внедрить технологии составления портрета каждого клиента. Такой подход позволит более детально понимать предпочтения клиентов, тем самым показывать более актуальную рекламу.

Для подсчета ожидаемых показателей conversion rate и CTR рекламного баннера, можно внедрить модели машинного обучения, предсказывающие успешность рекламы, основываясь на данных предыдущих рекламных кампаний.

### **3.7. Выгоды от реализации проекта**

Реализация проекта позволит проекту увеличить лояльность своих клиентов. Особенно среди тех пользователей, которые активно пользуются кэшбэком и следят за специальными предложениями. Кроме этого, банк

увеличит свою прибыль благодаря заказам рекламы со стороны компаний-партнеров.

### **Заключение**

Таким образом, в ходе разработки рекомендательной системы удалось построить такую архитектуру системы, в которую успешно была интегрирована модель рекомендательной системы, показавшая на тестовых данных метрику качества больше 0.8. Для выбора модели рекомендательной системы были исследованы основные модели рекомендательной системы.

Также был разработан макет сервиса в формате сайта и API. При помощи сайта клиенты (компании) могли бы делать заказы на рекламу. При помощи API другие сервисы банка могли бы обращаться к нашей системе, чтобы импортировать наш сервис в другие приложения.

Также это положительно повлияет на развитие отношений и программ с B2B партнерами. В дальнейшем возможны более сложные интеграции и обмен данными, технологиями. Затраты на реализацию и содержание проекта невелики, поэтому проект довольно быстро окупится.

Для реализации проекта были решены следующие задачи:

- Определены организационные и содержательные рамки проекта;
- Проведен анализ предметной области;
- Выбраны и обоснованы методы и инструменты для реализации проекта;
- Сформулированы функциональные требования к ИТ-решению;
- Описана архитектура ИТ-решения;
- Исследованы современные методы построения рекомендательных систем;
- Разработан и описан алгоритм машинного обучения;

- Выбрана платформа для реализации проекта;
- Описаны функциональность решения и ожидаемые бизнес-эффекты;
- Внедрена рекомендательная система.

### **Список использованных источников**

1. Nisbet, R., Elder, J., and Miner, G. Handbook of Statistical Analysis and Data Mining Applications. – Amsterdam: Elsevier, 2009.
2. Haiyun Lu. Recommendations Based on Purchase Patterns [Электронный ресурс] / International Journal of Machine Learning and Computing. – 2014. Режим доступа: <http://ijmlc.org/papers/462-C015.pdf>, свободный (30.05.2021).
3. Shivani Saini, Sunil Saumya и Jyoti Prakash Singh. Sequential Purchase Recommendation System for E-Commerce Sites [Электронный ресурс] / IFIP International Conference on Computer Information Systems and Industrial Management. – 2017. Режим доступа: [https://www.researchgate.net/publication/317232363\\_Sequential\\_Purchase\\_Recommendation\\_System\\_for\\_E-Commerce\\_Sites](https://www.researchgate.net/publication/317232363_Sequential_Purchase_Recommendation_System_for_E-Commerce_Sites), по запросу (30.05.2021).
4. Yeongwook Yang, Hong-Jun Jang и Byoungwook Kim. A Hybrid Recommender System for Sequential Recommendation: Combining Similarity Models With Markov Chains [Электронный ресурс] / IEEE. – 2020. Режим доступа: [https://www.researchgate.net/publication/347454718\\_A\\_Hybrid\\_Recommender\\_System\\_for\\_Sequential\\_Recommendation\\_Combining\\_Similarity\\_Models\\_With\\_Markov\\_Chains](https://www.researchgate.net/publication/347454718_A_Hybrid_Recommender_System_for_Sequential_Recommendation_Combining_Similarity_Models_With_Markov_Chains), по запросу (30.05.2021).



5. Mehdi Naseri. E-commerce Recommendation by an Ensemble of Purchase Matrices with Sequential Patterns [Электронный доступ]. – 2019. Режим доступа:  
<https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=8972&context=etd>, открытый (30.05.2021).
6. Anand V Bodapati. Recommendation Systems with Purchase Data [Электронный доступ] / Journal of Marketing Research. – 2008. Режим доступа:  
[https://www.researchgate.net/publication/247355167\\_Recommendation\\_Systems\\_with\\_Purchase\\_Data](https://www.researchgate.net/publication/247355167_Recommendation_Systems_with_Purchase_Data), по запросу (30.05.2021).
7. Moorissa Tjokro. How to Build a Recommendation System for Purchase Data (Step-by-Step) [Электронный доступ]. – 2018. Режим доступа:  
<https://medium.datadriveninvestor.com/how-to-build-a-recommendation-system-for-purchase-data-step-by-step-d6d7a78800b6>, свободный (30.05.2021).
8. Ashutosh Singh. Recommendation System for E-Commerce using Collaborative Filtering [Электронный доступ]. – 2019. Режим доступа:  
<https://medium.com/@ashutoshsingh93/recommendation-system-for-e-commerce-using-collaborative-filtering-fa04d6ab1fd8>, свободный (30.05.2021).
9. Taufik Azri. Recommendation System for Retail Customer [Электронный доступ]. – 2020. Режим доступа: <https://taufik-azri.medium.com/recommendation-system-for-retail-customer-3f0f80b84221>, свободный (30.05.2021).
10. Bruno Pradel, Savaneary Sean, Julien Delporte, Sebastien Guerif, Céline Rouveirol, Nicolas Usunier, Francoise Soulie Fogelman и Frédéric Dufau-Joël. A Case study in a Recommender System Based on Purchase Data [Электронный доступ] / Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. –

2011. Режим доступа:  
[https://www.researchgate.net/publication/221654162\\_A\\_case\\_study\\_in\\_a\\_recommender\\_system\\_based\\_on\\_purchase\\_data](https://www.researchgate.net/publication/221654162_A_case_study_in_a_recommender_system_based_on_purchase_data), свободный  
(30.05.2021).
11. Oladapo Oyeboode и Rita Orji. A Hybrid Recommender System for Product Sales in a Banking Environment [Электронный доступ] / Journal of Banking and Financial Technology. – 2020. Режим доступа:  
[https://www.researchgate.net/publication/338614945\\_A\\_hybrid\\_recommender\\_system\\_for\\_product\\_sales\\_in\\_a\\_banking\\_environment](https://www.researchgate.net/publication/338614945_A_hybrid_recommender_system_for_product_sales_in_a_banking_environment), свободный  
(30.05.2021).
12. Тимофеев М.В., Малькова К.М. и Лядова Л.Н. Архитектура Рекомендательной Системы, Настраиваемой на Предметные Области [Электронный доступ] / VIII Международная научно-техническая конференция "Технологии разработки информационных систем". – 2017. Режим доступа:  
[https://www.researchgate.net/publication/330564487\\_ARHITEKTURA\\_REKOMENDATELNOJ\\_SISTEMY\\_NASTRAIVAEMOJ\\_NA\\_PREDMETNYE\\_OBLASTI](https://www.researchgate.net/publication/330564487_ARHITEKTURA_REKOMENDATELNOJ_SISTEMY_NASTRAIVAEMOJ_NA_PREDMETNYE_OBLASTI), свободный (30.05.2021).

## **Приложение**

**Приложение 1**  
**Реестр стейкхолдеров**

<b>Категория</b>	<b>Наименование</b>	<b>Уровень влияния</b>	<b>Уровень заинтересованности</b>	<b>Ожидания</b>
Команда проекта	Команда № 318	Н	Н	Успешно внедрить систему, получить прибыль за проделанную услугу.
Председатель совета директоров	Близнюк С.Ю.	Н	М	Автоматизация бизнес-процессов с анализом данных, повышение эффективности бизнеса.
Директор по информационным технологиям	Цыганов В.В.	Н	М	Улучшение эффективности в сфере ИТ.
Руководитель проектного отдела	Филимонов В.И.	Н	Н	Успешное окончание проекта.
Руководитель отдела маркетинга	Волынка А.Н.	Н	Н	Развитие маркетинговой кампании, повышение эффективности и прибыльности отдела маркетинга
Разработчики приложения Тинькофф Банк	-	L	L	Стабильная и бесперебойная работа приложения.
Рядовые сотрудники Тинькофф Банк	-	L	L	Будущие бизнес-пользователи приложения.

## Приложение 2

### Реестр категорий предоставляемых услуг и сервисов партнеров

Категория	Сервис
Каршеринг	Яндекс.Драйв
Супермаркет	Яндекс.Лавка
Такси	Яндекс.Такси
Музыка	Яндекс.Музыка
Фастфуд	Яндекс.Еда
Аптеки	-
Кино	Кинопоиск
Книги	ЛитРес
Красота	Л'Этуаль
Образование	Яндекс.Практикум
Топливо	Яндекс.Заправка
Рестораны	Чайхона №1
Авиабилеты	Аэрофлот
Отели	-
Транспорт	-
Развлечения	-
Одежда и обувь	UNIQLO
Животные	-
Дом и ремонт	Moidom

Спорттовары	Спортмастер
Сувениры	-
Фото и видео	-
Цветы	-
Аренда авто	-
Автоуслуги	-
Дьюти-фри	-
Ж/д билеты	РЖД
Искусство	Пушкинский музей

## Приложение 3

### Пример данных о клиента

A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	fullName	address	phone_number	email	workplace	birthdate	registration_gender	income	expenses	credit	deposit
2	1	Носова Юрьевна Ковалева	г. Петрозаводск, ул. 50 лет СССР, д. 673 стр.6 (133) 969-49-57		mih_1991@hotmail.com	п. Гарно-Сале, бул. Раинка, д. 2, 877538	1978-01-17	2013-01-26 F		73107	ЛОЖЬ	ЛОЖЬ
3	2	Демидов Вячеслав Глебович	к. Новый Оскол, пр. Строителей, д. 8 (631) 768-4725		visnjakovfr@mail.ru	п. Бавло, пр. Революционный, д. 531 стр. 2/	1971-03-11	2012-07-28 M		168585	ЛОЖЬ	ИСТИНА
4	3	Волков Николай Владимирович	с. Октябрьское (Чепецк), пер. Ворон-17 (020) 257-2544		belovnikola@gmail.com	ст. Палана, бул. Астраханский, д. 79 стр. 6, 55	1963-06-23	2013-11-01 M		89208	ЛОЖЬ	ЛОЖЬ
5	4	Иселева Людмила Викентьевна	п. Холмогоры, а/л. Крестинская, д. 5 ст.17 (663) 425-5442		portiri66@yahoo.com	п. Иттемнери, пер. Уралский, д. 1/1 стр. 1/1	1980-09-01	2019-09-25 F		89713	ЛОЖЬ	ЛОЖЬ
6	5	И-И Лебедева Еронислава Артурович	к. Истра, наб. 50 лет Победы, д. 53 к/5 137 055 6648		pantelemon_1985@rambler.ru	к.л. Междуреченский, а/л. Шевченко, д. 66	1973-08-20	2015-11-16 M		82521	ЛОЖЬ	ИСТИНА
7	6	Денисова Клавдия Геннадьевна	п. Баргузин, пер. Моховой, д. 3 стр. -17 478 168 47 28		kondrat19@hotmail.com	к. Семьяники, наб. Пархоменко, д. 714, 5610	1962-09-18	2015-10-31 F		68593	ЛОЖЬ	ИСТИНА
8	7	Лукина Андреевна Алафоновна	ст. Галича, наб. Крылова, д. 30 стр. -17 (194) 665-80-00		lapinaoktjabrina@yandex.ru	к. Улан-Удэ, пр. Садовий, д. 222 стр. 2, 58928	1964-03-19	2014-08-28 F		76904	ЛОЖЬ	ЛОЖЬ
9	8	Леон Владимирович Архипов	д. Светлогорск (Калин.), ш. Центральная 5 563 161 0331		samolovgrigor@hotmail.com	ст. Манчигорск, бул. Главный, д. 79 стр. 59	1985-02-23	2014-01-16 M		62636	ЛОЖЬ	ИСТИНА
10	9	Алла Ариадьевна Давыдова	с. Усть-Ордынский, бул. Строителей 5 824 001 6894		kuznetsouparten@yandex.ru	п. Малоярославцы, пер. Грибоедова, д. 30 к.	1981-10-03	2013-11-08 F		80061	ЛОЖЬ	ИСТИНА
11	10	Заварова Фаина Идальевна	к.л. Одиноково, ш. Боксера, д. 694, 1 8 (266) 949-46-51		kulaginleonid@mail.ru	ст. Усть-Джукуты, пр. Ижевского, д.	1952-06-31	2012-06-04 M	282358	85952	ЛОЖЬ	ЛОЖЬ
12	11	Харин Сергей Александрович	к.л. Талов, пр. Флотский, д. 7 стр. 60, 17 (218) 343-405-64		vorobeynir@hotmail.com	ст. Тарпей, ш. Вулфганг, д. 29 стр. 35, 719808	1985-08-03	2015-05-21 F		76009	ЛОЖЬ	ЛОЖЬ
13	12	Виктор Гаврилович Никитин	ст. Москва, пр. Троицкий, д. 300 к. 3, 8 (733) 108-03-44		nikolagusev@gmail.com	п. Серафимовск, наб. Срековская, д. 785 к. 1, С	1986-05-22	2016-09-07 M		98752	ЛОЖЬ	ИСТИНА
14	13	Горбанев Михаил Харлампиевич	д. Шелково, бул. Советской Армии, 59949247754		pestovklavdi@hotmail.com	п. Юров (Ветка), ш. Советов, д. 5/6 стр. 2, 55	1960-09-21	2012-03-13 M		76892	ЛОЖЬ	ИСТИНА
15	14	Ахип Андрей Александрович Петузов	к.л. Барск, бул. Центральный, д. 6/5 17 951 650 87 20		marfa_52@rambler.ru	п. Дарвинск, а/л. Черемушкова, д. 416 стр.	2002-05-27	2012-08-06 M	232854	86568	ИСТИНА	ЛОЖЬ
16	15	Семенов Алексей Тарасович	г. Ербоганич, пр. Мелюратинский, д. 8 (104) 864-16-83		veniamin_1995@yandex.ru	с. Каргалы, ул. Тракторная, д. 2 к. 439, 509734	1991-09-01	2018-06-20 F		727719	97025	ИСТИНА
17	16	Шукина Феола Владимировна	к.л. Златог, ул. Ленина, д. 6, 393911 17 138 694 1533		vladimirov@hotmail.com	ст. Наркиса, а/л. Студенческая, д. 139, 2983	1996-06-12	2017-05-24 F		177280	74065	ИСТИНА
18	17	Уваров Тарас Ярославович	к. Ленинск, пр. Олимпийский, д. 17 к. 17 (083) 725-6047		artem2020@yandex.ru	ст. Саранск, пер. Владимирский, д. 5 к. 85, 74	1977-05-01	2015-03-16 M		90447	ЛОЖЬ	ИСТИНА
19	18	Вышеслав Герсентьевич Галин	с. Славогород, ш. Вершинное, д. 7/9 к 17 060 118 80 71		martin_1985@rambler.ru	к.л. Тверь, ш. Мичкина, д. 883 к. 717, 137661	1980-06-17	2016-11-17 M		101428	ЛОЖЬ	ИСТИНА
20	19	Артёмий Харитонович Ситыкин	к.л. Горняцкий, а/л. Гончарова, д. 85 17 399 611 84 41		demjan1975@rambler.ru	г. Флора, наб. Бакеева, д. 6, 922483	1952-05-14	2015-04-12 M		70591	ЛОЖЬ	ЛОЖЬ
21	20	Ульянов Русланович Сердюков	п. Нахичевань, а/л. Дроздовская, д. 8/1 17 (243) 455-49-14		gorfin@hotmail.com	г. Ленинград, бул. Сосновский, д. 17, 4306	1960-07-26	2016-10-26 F		86603	ЛОЖЬ	ЛОЖЬ
22	21	Шарапов Игорь Владимирович	п. Барнауль, ул. Верхова, д. 330, 958673440705056		makarovapollinari@mail.ru	с. Тверь, пер. Степана Разина, д. 2/7 к. 15, 82	1960-11-02	2019-05-02 M		85710	ЛОЖЬ	ИСТИНА
23	22	Полына Александровна Колдобова	к.л. Солная (Ханг.), ш. Пономарева, д. 1 (177) 933-43-77		prokl_79@yahoo.com	к. Якутск, ш. Крайняя, д. 42 стр. 9/5, 625580	1966-03-31	2014-09-23 F		87064	ЛОЖЬ	ИСТИНА
24	23	Анжела Владиславовна Пуляева	г. Магдогорск, бул. Кавказский, д. 24/52795160218		ntilipov@gmail.com	г. Ленинград, ул. Сосновская, д. 37 стр. 4, 5615	1969-03-07	2018-01-03 F	191891	92393	ИСТИНА	ЛОЖЬ
25	24	Галина Богдановна Иванова	г. Ангарск, ул. Кочубев, д. 46, 058518 (238) 388-5682		lapinagga@gmail.com	с. Шахты, наб. Шолохова, д. 4 к. 9/4, 083357	1988-09-09	2012-02-29 F	292102	81643	ИСТИНА	ЛОЖЬ
26	25	Харитонов Алексей Николаевич	к.л. Талов, пр. Королева, д. 7/5 стр.8 (408) 265-6399		arhilpovamaja@gmail.com	к. Хатанга, бул. Тракторный, д. 17 к. 2/7, 2611	1996-04-01	2019-07-08 F		85408	ЛОЖЬ	ИСТИНА
27	26	Юлия Александровна Лазарева	с. Север-Юрильск, пер. Поларный 8 332 786 69 73		grigor135@yandex.ru	с. Бузулу, ш. Свободная, д. 756, 997037	1984-02-27	2013-01-03 F	187105	91962	ИСТИНА	ЛОЖЬ
28	27	Исаев Юрий Гаврилович	с. Леринск, пер. Урицкого, д. 9/6, 21 17 (612) 165-98-56		mihail_85@yandex.ru	г. Краснодарт, пр. Маяковского, д. 3/5 к. 80, 2	1969-12-14	2017-07-17 M		20354	ЛОЖЬ	ЛОЖЬ
29	28	Заварова Фаина Идальевна	ст. Дмитрия, ш. Трудовое, д. 664, 48 17 (276) 524-8518		avoskija_91@yahoo.com	г. Багайск, пер. Новостройка, д. 186, 657631	1986-06-02	2018-06-13 F		88195	ЛОЖЬ	ИСТИНА
30	29	Туркина Валентина Ринатовна	к.л. Ока, а/л. Троицкая, д. 8/1 стр. 4/1 17 158 980 7783		lacom6@gmail.com	ст. Каргалы, пер. Запрудный, д. 48 к. 9, 63534	1951-03-15	2015-06-20 F	131667	87583	ИСТИНА	ЛОЖЬ
31	30	Молчанова Ульяна Николаевна	с. Нижний Тагил, ш. Северное, д. 7/1 17 (597) 115-8580		stodorovalev@gmail.ru	к.л. Шаховская, а/л. Крайняя, д. 831 стр. 2, 0	1962-02-13	2019-06-29 F		133531	98845	ИСТИНА
32	31	Сергеев Борислав Вячеславович	к.л. Болотный (Амур.), наб. Коллизия 7 685 246 7514		ladislav10@yahoo.com	п. Урус-Мартан, а/л. Крылова, д. 7/6, 937467	1985-02-02	2019-06-08 M		186068	87112	ИСТИНА
33	32	Лора Игоревна Кузнецова	г. Тирнекауш, ш. Радиново, д. 683, 4 17 (481) 603-1281		danilovlenita@hotmail.com	с. Петроавдарт, пер. Революционный, д. 8/	1969-08-14	2015-06-17 M	164839	79197	ИСТИНА	ЛОЖЬ
34	33	Полынов Александр Владимирович	п. Вологда, бул. Братский, д. 3 стр. 9 17 (478) 874-3778		mihalovaregina@mail.ru	к. Нижний Тагил, пер. Радищева, д. 91 к. 777,	1984-02-26	2015-12-17 M		82185	ЛОЖЬ	ИСТИНА
35	34	Барыкова Валентина Наумовна	г. Ижевск, ш. Интернациональный, 8 216 104 2473		nalina_32@gmail.com	ст. Тейрош, пр. Рабочий, д. 2 к. 8, 589070	1983-03-07	2013-11-06 F		57843	ЛОЖЬ	ЛОЖЬ
36	35	Карпов Глеб Игоревич	г. Уварово, наб. Садова, д. 22 к. 51, 6 17 (290) 880-19-12		ekomarov@gmail.com	с. Волынки, наб. Ленина, д. 9 стр. 215, 164872	1974-07-23	2014-07-25 M		71312	ЛОЖЬ	ЛОЖЬ
37	36	Илья Игоревич Рабов	к. Красноуральск, пр. Соколовый, д. 1 17 233 585 77 37		uljan57@yandex.ru	ст. Таштагол, а/л. Гвардейская, д. 1 стр. 8/2,	1987-07-13	2018-05-31 M	187224	75280	ИСТИНА	ЛОЖЬ
38	37	Михаил Дмитриевич Бобильев	с. Кабанск, а/л. Сибирская, д. 1/2, 4/5 418 584 6138		ladimir25@mail.com	с. Кудымкар, ул. Некрасова, д. 121 стр. 3, 247	1974-03-24	2012-06-13 M	112392	78890	ИСТИНА	ИСТИНА

## Приложение 4

### Пример данных о категориях

	A	B	C	D
1	id	name	description	mcc-code
2	1	Каршеринг	Краткосрочная аренда авто с оплатой по минутам или часам	7512, 4121
3	2	Супермаркеты	Покупки в супермаркетах и продуктовых магазинах	5297, 5298, 5300, 5411, 5412, 5422, 5441, 5451, 5462, 5499, 5715, 5921
4	3	Такси	Услуги такси (каршеринг не входит в данную категорию)	4121
5	4	Музыка	Покупки в магазинах музыки и музыкальных инструментов	5733, 5735
6	5	Фастфуд	Покупки в ресторанах быстрого питания	5814
7	6	Транспорт	Оплата городского и пригородного транспорта. Взносы за по	4111, 4121, 4131, 4457, 4468, 4784, 4789, 5013, 5271, 5551, 5561, 5592, 5598, 5599, 7511, 7523
8	7	Аптеки	Покупки в аптеках	5122, 5292, 5295, 5912
9	8	Кино	Покупки в кинотеатрах и магазинах кино	7829, 7832, 7841
10	9	Книги	Покупки в книжных магазинах и газетных киосках	2741, 5111, 5192, 5942, 5994
11	10	Развлечения	Оплата развлечений, (например, в магазинах видеоигр, боу	7911, 7922, 7929, 7932, 7933, 7941, 7991, 7992, 7993, 7994, 7996, 7997, 7998, 7999, 8664
12	11	Красота	Покупки в массажных и косметических салонах, SPA и салонах	5977, 7230, 7297, 7298
13	12	Образование	Оплата в государственных и частных школах, вузах и училищ	8211, 8220, 8241, 8244, 8249, 8299, 8493, 8494
14	13	Одежда и обувь	Покупки в магазинах одежды и обуви	5094, 5137, 5139, 5611, 5621, 5631, 5641, 5651, 5661, 5681, 5691, 5697, 5698, 5699, 5931, 5944, 5949,
15	14	Рестораны	Оплата счета в барах, ресторанах, кафе и клубах (кроме фаст	5811, 5812, 5813
16	15	Топливо	Покупка газа и бензина на заправках	5172, 5541, 5542, 5983
17	16	Животные	Покупки в зоомагазинах	5995, 0742
18	17	Дом и ремонт	Покупки в магазинах мебели, товаров для дома и ремонта	1520, 1711, 1731, 1740, 1750, 1761, 1771, 1799, 2791, 2842, 5021, 5039, 5046, 5051, 5065, 5072, 5074,
19	18	Спорттовары	Покупки в магазинах спортивных товаров	5655, 5940, 5941
20	19	Сувениры	Покупки в магазинах сувениров: открыток, упаковочной бум	5947
21	20	Фото и видео	Покупки в магазинах фото- и видеотехники	5044, 5045, 5946, 7332, 7333, 7338, 7339, 7395
22	21	Цветы	Покупки в цветочных магазинах и салонах флористики	5193, 5992
23	22	Аренда авто	Оплата аренды авто в дилерских центрах (каршеринг не вхо	3351–3398, 3400–3410, 3412–3423, 3425–3439, 3441, 7512, 7513, 7519
24	23	Автослужбы	Покупка автозапчастей, услуги в автосервисах и на автомойк	5511, 5521, 5531, 5532, 5533, 5571, 7012, 7531, 7534, 7535, 7538, 7542, 7549
25	24	Авиабилеты	Оплата авиабилетов	3000–3299; 4304; 4415; 4418; 4511; 4582
26	25	Дьюти-фри	Покупки в магазинах дьюти-фри	5309
27	26	Железнодорожные билеты	Оплата билетов на поезд	4011, 4112
28	27	Искусство	Покупки в галереях, антикварных магазинах или у арт-дилер	5932, 5937, 5970, 5971, 5972, 5973
29	28	Переводы	Переводы, пополнения	
30	29	Довгое	Интернет. коммунальные услуги	



## Приложение 5

### Пример данных о транзакциях

	A	B	C	D	E	F	G	H
1	id	client_id	product_cat	product_company	subtype	amount	date	transaction_type
2	48303	53	4	Яндекс.Музыка	Подписка	199	2020-01-01 00:00:00	Negative
3	54700	59	29		Автоплатеж	7505	2020-01-01 00:00:00	Negative
4	57477	62	29		Автоплатеж	2517	2020-01-01 00:00:00	Negative
5	14959	17	6	Стрелка	Покупка	52	2020-01-01 00:02:20	Negative
6	53210	58	6	Московский Метроп	Покупка	72	2020-01-01 00:16:19	Negative
7	80315	87	2	Ашан	Покупка	2290	2020-01-01 00:26:14	Negative
8	28995	31	28		Перевод	1620	2020-01-01 00:35:17	Negative
9	42012	45	28		Перевод	2755	2020-01-01 00:45:53	Negative
10	76139	82	28		Пополнение	211	2020-01-01 00:45:59	Positive
11	83740	90	15	АЗС Лукойл	Покупка	1515	2020-01-01 00:46:59	Negative
12	86869	94	6		Покупка	79	2020-01-01 01:01:26	Negative
13	80877	87	8	Каро Фильм	Покупка	354	2020-01-01 01:03:32	Negative
14	16076	18	6	Московский Метроп	Покупка	83	2020-01-01 01:04:07	Negative
15	54162	59	6	Стрелка	Покупка	65	2020-01-01 01:12:54	Negative
16	51027	56	3	UBER	Покупка	372	2020-01-01 01:29:36	Negative
17	68883	74	28		Пополнение	1688	2020-01-01 01:38:57	Positive
18	5822	7	6	Стрелка	Покупка	67	2020-01-01 01:40:36	Negative
19	89284	96	28		Перевод	2319	2020-01-01 01:42:37	Negative
20	77735	84	6	Московский Метроп	Покупка	98	2020-01-01 01:45:59	Negative
21	3767	5	5		Покупка	302	2020-01-01 01:46:50	Negative
22	45507	50	2	Утконос Онлайн	Покупка	1732	2020-01-01 01:48:16	Negative
23	92729	100	11	Косметик центр	Покупка	2006	2020-01-01 01:49:43	Negative
24	62274	68	3	Kiwi Taxi	Покупка	521	2020-01-01 02:02:50	Negative
25	68131	74	5	PizzaHut	Покупка	513	2020-01-01 02:03:47	Negative
26	78643	85	6	Московский Метроп	Покупка	59	2020-01-01 02:07:26	Negative
27	11416	13	6	Стрелка	Покупка	82	2020-01-01 02:08:38	Negative
28	73146	79	6	Стрелка	Покупка	91	2020-01-01 02:20:06	Negative
29	31440	34	6	Московский Метроп	Покупка	62	2020-01-01 02:21:57	Negative
30	64731	70	6	Московский Метроп	Покупка	72	2020-01-01 02:25:11	Negative
31	62085	67	8	Okko	Покупка	498	2020-01-01 02:27:28	Negative
32	91109	99	3	ЯндексGO	Покупка	532	2020-01-01 02:30:13	Negative
33	61782	67	6	Стрелка	Покупка	95	2020-01-01 02:30:54	Negative
34	80219	87	1	Яндекс.Драйв	Покупка	1213	2020-01-01 02:30:56	Negative
35	63729	69	6		Покупка	66	2020-01-01 02:34:35	Negative
36	70924	77	5	Макдоналдс	Покупка	448	2020-01-01 02:35:47	Negative
37	4208	5	14	Solo Burger Bar	Покупка	2767	2020-01-01 02:36:13	Negative
38	27256	30	3	ЯндексGO	Покупка	235	2020-01-01 02:43:16	Negative

## Приложение 6

### Пример данных о подписках

	A	B	C	D	E	F	G
1	id	client_id	product_category	product_company	amount	date_start	date_end
2	59399	64	29		4595	2012-08-06	
3	7188	9	4	Google Play Музыка	159	2013-02-13	
4	71811	78	4	Яндекс.Музыка	199	2013-03-07	2020-06-07
5	39410	43	4	Apple Music	169	2013-04-10	
6	21418	24	4	Яндекс.Музыка	199	2014-02-28	
7	42017	45	29		5095	2014-03-25	2020-07-25
8	78194	84	29		7731	2014-07-13	
9	41203	45	4	YouTube Music	169	2014-09-17	
10	31782	34	29		1162	2014-10-17	
11	48303	53	4	Яндекс.Музыка	199	2015-02-01	
12	4656	6	4	Boom	149	2015-02-14	
13	54839	60	4	YouTube Music	169	2015-02-21	
14	72654	78	29		9353	2015-03-06	
15	33855	37	4	Apple Music	169	2015-03-11	
16	57613	63	4	Boom	149	2015-03-27	
17	26358	29	4	Apple Music	169	2015-04-02	2020-09-02
18	38302	41	29		4291	2015-06-13	
19	64174	69	29		5812	2015-06-17	
20	27177	29	29		1590	2015-07-03	
21	7134	8	29		7999	2015-08-19	
22	64232	70	4	Apple Music	169	2015-08-25	
23	946	2	4	Spotify	169	2015-10-14	
24	58571	64	4	Boom	149	2015-10-18	2020-05-18
25	84833	92	4	Spotify	169	2016-02-07	
26	85687	92	29		1654	2016-02-11	
27	75320	81	29		4489	2016-02-14	
28	55672	60	29		1900	2016-03-10	
29	40208	43	29		6697	2016-04-11	
30	42950	47	4	YouTube Music	169	2016-05-13	
31	11959	13	29		3711	2016-08-10	
32	77252	84	4	Boom	149	2016-09-20	
33	5538	7	4	Boom	149	2016-11-10	
34	84720	91	29		3822	2017-01-06	2020-11-06
35	54700	59	29		7505	2017-02-01	
36	59456	65	4	YouTube Music	169	2017-02-12	2020-03-12
37	35755	39	4	Яндекс.Музыка	199	2017-02-23	

## Приложение 7

Таблица описания атрибутов сущностей заказа, сообщения и информации о рекламе

<b>Orders</b>	id	int	Идентификатор заказа
	company_name	string	Название компании
	email	string	Адрес электронной почты
	code	string	Код заказа
	date_start	date	Дата начала рекламы
	date_end	date	Дата окончания рекламы
	price	float	Цена
	category	string	Категория транзакции, по которой планируется пройти реклама
	clients	string	Список id клиентов, которым будет показываться реклама
	clients_number	int	Число клиентов, которым будет показываться реклама
	days	int	Число дней, сколько будет идти реклама
	confirmation_status	boolean	Статус подтверждения заказа
	acceptance_status	boolean	Статус принятия заказа
	acceptance_date	datetime	Дата и время принятия заказа
	completion_status	boolean	Статус выполнения заказа

	completion_date	datetime	Дата и время выполнения заказа
	creation_date	datetime	Дата и время создания заказа
<b>Commercial_infos</b>	id	int	Идентификатор информации о рекламе
	shown_number	int	Число клиентов, которым была показана реклама
	shown_clients	string	Список id клиентов, которым была показан реклама
	clicked_number	int	Число клиентов, которые перешли по ссылке
	clicked_clients	string	Список id клиентов, которые перешли по ссылке
	performed_action_number	int	Число клиентов, которые выполнили целевое действие
	performed_action_clients	string	Список id клиентов, которые выполнили целевое действие
<b>Messages</b>	id	int	Идентификатор сообщения
	text	string	Текст сообщения
	email	string	Адрес электронной почты автора сообщения
	creation_date	datetime	Дата и время создания сообщения
<b>RecommnedationData</b>	id	int	Идентификатор данных
	category	FK	Внешняя связь с сущностью Category

	clients	string	Список id клиентов, которым следует рекламировать соответствующую категорию
<b>RecommendationModel</b>	id	int	Идентификатор модели
	name	string	Название модели\версии
	f_score	float	Гармоническое среднее между точностью и полнотой дающее более точное понимание о качестве модели
	data	FK	Внешняя связь с сущностью RecommendationData
	last_update	datetime	Дата и время последнего обновления модели

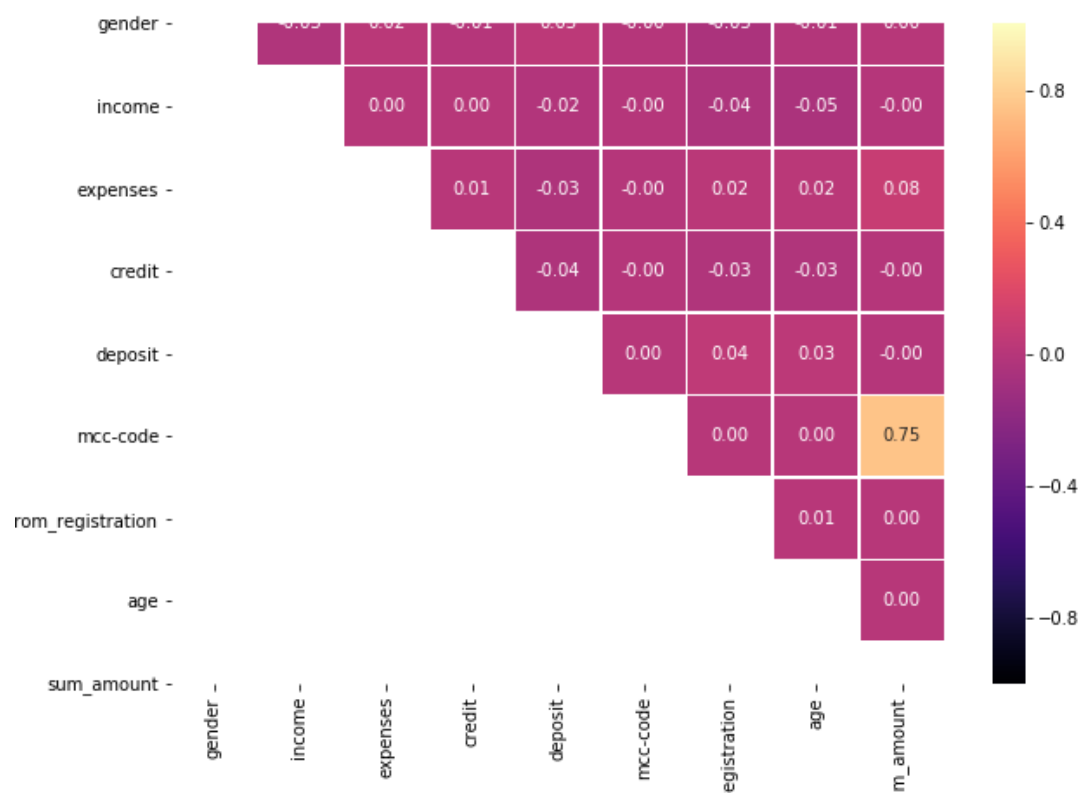
## Приложение 8

Ссылка на код исследования моделей рекомендательной системы

[https://github.com/AleksandrRadist/RecommendationSystem/tree/master/model/  
research\\_recommendation\\_models](https://github.com/AleksandrRadist/RecommendationSystem/tree/master/model/research_recommendation_models)

## Приложение 9

Матрица корреляций тренировочной выборки при построении контентных моделей



## Приложение 10

### Сравнение memory-based и модели со скрытыми переменными

