МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**
Кафедра информатики и программирования

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ ЗНАКОМСТВ С РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМОЙ

БАКАЛАВРСКАЯ РАБОТА

студента 4 курса 441 группы
направления 02.03.03 — Математическое обеспечение и администрирование
информационных систем
профиль «Технологии программирования»
факультета КНиИТ
Уталиева Султана Едильбаевича

Научный руководитель
ст.преп. кафедры ИиП _____ Казачкова А. А.

Заведующий кафедрой
к. ф.-м. н., доцент _____ Огнева М. В.

Саратов 2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В последние годы наблюдается устойчивый рост интереса к персонализированным цифровым сервисам, что обусловлено стремлением пользователей получать релевантный контент и улучшенный пользовательский опыт. Одним из ключевых инструментов персонализации являются рекомендательные системы — интеллектуальные алгоритмы, позволяющие адаптировать предложения под индивидуальные предпочтения. Их значение особенно велико в приложениях для знакомств, где качество рекомендаций напрямую влияет на успешность социальных взаимодействий и удовлетворённость пользователей [1, 2].

Современные приложения, такие как Tinder, OkCupid и Hinge, активно используют алгоритмы рекомендаций, включая коллаборативную фильтрацию, обучение представлений и гибридные методы, зачастую основанные на больших объёмах пользовательских данных и методах машинного обучения [3–5]. Однако многие существующие решения сталкиваются с проблемами в условиях дефицита данных — например, на ранних этапах использования приложения [6].

В данной работе рассматривается разработка мобильного приложения для знакомств, включающего в себя встроенную рекомендательную систему. В отличие от большинства существующих решений, предлагаемый подход предусматривает использование не только анкетных данных, но и результатов тестов-опросов, которые пользователи могут проходить по собственному желанию. Каждая карточка с вопросом позволяет выбрать один из трёх вариантов ответа — «да», «нет» или «пропустить», что позволяет формировать тернарные признаки $(-1, 0, 1)$, лежащие в основе профиля предпочтений пользователя.

Целью дипломной работы является разработка приложения для знакомств с рекомендательной системой, обеспечивающей релевантные и разнообразные рекомендации потенциальных партнёров на основе результатов тестирования. Для достижения этой цели решаются следующие задачи:

- анализ существующих подходов к построению рекомендательных систем в контексте мобильных приложений для знакомств;
- формализация задачи рекомендаций с учётом специфики представления пользовательских признаков;
- проектирование архитектуры приложения и рекомендательной системы;
- реализация рекомендательной подсистемы на основе методов контентной фильтрации и эвристических правил;

— разработка подхода к оценке качества рекомендаций с использованием оффлайн-метрик.

Практическая значимость работы заключается в создании масштабируемого мобильного решения, сочетающего в себе функции приложения для знакомств и интерпретируемой рекомендательной системы, адаптированной к условиям ограниченных вычислительных ресурсов и высокой динамики пользовательских предпочтений.

1 Анализ предметной области и существующих решений

1.1 Особенности рекомендательных систем

Рекомендательные системы являются неотъемлемой частью современных цифровых платформ, предоставляя пользователям персонализированные предложения продуктов, услуг или контента. Они находят широкое применение в различных областях, включая электронную коммерцию, стриминговые сервисы, социальные сети и онлайн-знакомства.

Современные рекомендательные сталкиваются с рядом проблем:

- отсутствие информации о новых пользователях и объектах (проблема холодного старта) [6];
- высокая разреженность пользовательско-объектных матриц [1];
- необходимость обеспечения справедливости и отсутствия предвзятости в рекомендациях [2].

Современные исследования направлены на преодоление этих ограничений, в том числе с использованием больших языковых моделей, методов объяснимого машинного обучения и расширения пользовательского контекста.

1.2 Обзор рекомендательных систем в сфере онлайн-знакомств

Онлайн-сервисы знакомств предъявляют особые требования к алгоритмам рекомендаций. Поскольку конечной целью является установление реального или виртуального контакта между людьми, системы должны максимально точно учитывать совместимость по широкому спектру признаков, при этом оставаясь достаточно лёгкими в вычислении и объяснимыми.

1.2.1 OkCupid

Платформа OkCupid применяет нетривиальный подход к построению рекомендаций, фокусируясь на глубоком анализе анкет пользователей и их ответов на вопросы. В отличие от многих систем, которые ориентируются лишь на поведение пользователей, OkCupid делает упор на содержательные признаки совместимости.

Пользователи проходят опросы, отвечая на вопросы о ценностях, привычках, интересах и взглядах. Для каждого вопроса пользователь:

- указывает свой собственный ответ;
- выбирает приемлемые для него ответы потенциального партнёра;
- определяет важность соответствия по этому вопросу.

Таким образом, для каждого пользователя можно получить богатый профиль предпочтений, включающий не только их мнения, но и ожидания от других.

Основой рекомендательной системы OkCupid служит собственная метрика совместимости, вычисляемая по формуле, напоминающей обобщённую версию взвешенного совпадения. Пусть пользователь A ответил на n вопросов, по которым можно сопоставить ответы с пользователем B. Тогда их совместимость вычисляется по следующей схеме:

1. Для каждого вопроса i определяется s_i^A — совпадает ли ответ B с приемлемыми ответами A; w_i^A — важность вопроса по мнению A.
2. Вычисляется доля удовлетворения B ожиданий A:

$$S_{AB} = \frac{\sum_{i=1}^{n} s_i^A \cdot w_i^A}{\sum_{i=1}^{n} w_i^A}$$

3. Аналогично считается S_{BA}.
4. Итоговая совместимость:

$$C(A, B) = \sqrt{S_{AB} \cdot S_{BA}}$$

Такой симметричный подход позволяет учитывать желания обеих сторон, что особенно важно в сфере знакомств [7].

Кроме ответов на вопросы, система также может учитывать:

— географическое положение пользователей;
— возраст и пол;
— поведенческие метрики (например, активность и отклики);
— интересы, указанные в профиле;
— алгоритмы коллаборативной фильтрации на основе лайков.

Однако основным источником данных остаются именно опросы, что выгодно отличает OkCupid от других платформ.

Рекомендательная система OkCupid строится на принципах симметричной оценки, учитывающей как личные предпочтения, так и взаимные ожидания. Модель сочетает элементы экспертных систем и идеи коллаборативной фильтрации, что позволяет выдавать персонализированные рекомендации, основанные на содержательных признаках [3].

Тем не менее, использование опросов как основной основы для рекомендаций имеет ряд ограничений:

— Самоотчетность. Пользователи могут отвечать неискренне, выбирая социально одобряемые ответы или предполагаемые предпочтения других, а не собственные.
— Ограниченность охвата. Даже при большом числе вопросов многие аспекты личности, поведения и совместимости остаются неохваченными.
— Неполные профили. Пользователи часто не отвечают на все доступные вопросы, что затрудняет построение точных рекомендаций.
— Фиксированные веса. Веса важности выставляются вручную, и пользователь может переоценить или недооценить значимость отдельных тем.
— Изменчивость во времени. Ответы могут устаревать, но система не всегда способна учитывать динамику изменения взглядов и предпочтений.

Таким образом, хотя подход OkCupid обладает значительной выразительной силой, его эффективность может снижаться при недостаточной мотивации пользователей честно и подробно заполнять анкету.

1.2.2 Tinder

Платформа Tinder делает акцент на скорости взаимодействия и минимализме в пользовательском интерфейсе. Это определяет и архитектуру рекомендательной системы: она почти полностью основана на поведенческих данных, а не на заранее структурированных опросах [8].

Рекомендации Tinder опираются на наблюдение за действиями пользователя в реальном времени, применяя идеи из области ранжирования, коллаборативной фильтрации и машинного обучения.

— Пользователь не заполняет анкету или тесты – основным источником сигналов становится поведение (свайпы, матчи, отклики).
— Цель системы – ранжировать потенциальных партнёров по вероятности положительного отклика.
— Взаимная симпатия (мэтч) служит основной меткой релевантности.

Ранее Tinder применял модификацию рейтинговой системы Elo, аналогичной той, что используется в шахматах. Каждому пользователю приписывался скрытый рейтинг R, который обновлялся при каждой паре действий:

$$R_A^{\text{new}} = R_A + K \cdot (S - E)$$

где:

— S — фактический результат (1, если A получил лайк от B, 0 — если дизлайк);

— E — ожидаемая вероятность положительного отклика, например:

$$E = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

— K — коэффициент обучения.

Со временем эта система была признана слишком статичной и неспособной учитывать сложные паттерны предпочтений.

В настоящее время Tinder использует более гибкий и масштабируемый подход на основе моделей обучения ранжированию и нейронных сетей. Ключевые особенности [5]:

1. Использование исторических свайпов как обучающего датасета.
2. Обогащение признаков за счёт:
   — времени суток, геолокации, возраста и пола;
   — информации о взаимодействиях (ответы в чатах, продолжительность общения);
   — изображений (модели компьютерного зрения извлекают визуальные эмбеддинги).
3. Применение моделей типа learning-to-rank, включая градиентный бустинг и глубокие нейросети.
4. Возможное использование sequence-based моделей (например, RNN или трансформеров), учитывающих порядок свайпов.

Хотя такой подход даёт хорошее качество персонализации, он не лишён ограничений:

— Холодный старт. Новым пользователям сложно получить релевантные рекомендации до накопления истории свайпов.

— Смещение данных. Пользователи чаще свайпают по привлекательности, а не по глубинной совместимости.

— Неустойчивость. Поведение может быть ситуативным, но алгоритм воспринимает его как предпочтение.

— Мало объяснимости. Модель сложно интерпретировать или объяснить пользователю, почему показан тот или иной профиль.

Модель рекомендаций Tinder эволюционировала от простой системы рейтингов к сложной системе поведенческого ранжирования. Она эффективно мас-

штабируется и адаптируется под предпочтения пользователя, но при этом страдает от отсутствия прозрачности и возможной поверхностности критериев совместимости [9].

1.2.3 Hinge

Платформа Hinge позиционирует себя как сервис для «удаления» приложения после нахождения подходящего партнёра. Это отражается и в её подходе к построению рекомендаций: модель ориентирована не на максимум свайпов, а на вероятность качественного взаимодействия. Рекомендательная система Hinge учитывает как поведенческие сигналы, так и контекстные данные, стремясь построить эффективные персонализированные предложения.

В основе лежат несколько ключевых принципов:

— Система стремится обучаться на успешных взаимодействиях — прежде всего на лайках, приводящих к продолжительным диалогам.
— Учитывается не только сам факт лайка, но и качество последующего общения.
— Рекомендации строятся с использованием ранжирования, основанного на моделях типа learning-to-rank.

Ключевым источником вдохновения послужил алгоритм Gale–Shapley, использующийся в задаче стабильного брака. В оригинальной постановке каждый участник ранжирует партнёров, и цель — найти устойчивое соответствие. В адаптации Hinge этот принцип реализуется эвристически, через поиск так называемого «самого совместимого» партнёра дня, при этом система моделирует предпочтения обеих сторон [10].

Если обозначить:

— $P(u, v)$ — вероятность успешного взаимодействия между пользователями u и v;
— R_u — рейтинг, отражающий склонность u к взаимодействию с разными типами партнёров;
— C_v — контекстные характеристики пользователя v;

то задача рекомендации может быть сведена к оценке:

$$\hat{y}_{uv} = f(R_u, C_v)$$

где f — обученная модель, приближающая вероятность успешного взаи-

модействия. Под успешностью понимается не просто лайк, а наличие значимого чата или повторного контакта.

Кроме основных моделей, в системе используются дополнительные эвристики:

— подавление повторяющихся шаблонов (например, слишком частые лайки одному типажу);
— учёт предпочтений по контенту профиля (фото, ответы на подсказки);
— отслеживание реакций на предложенные анкеты и их отложенное влияние.

Важной частью рекомендаций Hinge является система подбора пары дня. Она основывается на анализе двусторонних предпочтений, активности и недавнего поведения. Рекомендуемый партнёр имеет высокий прогнозируемый шанс на взаимную симпатию и заинтересованный диалог.

Несмотря на успех такой модели, она имеет определённые ограничения:

— зависимость от истории пользователя, что создаёт проблему холодного старта;
— возможные локальные оптимумы: пользователь может застревать в узком профиле предложений;
— невысокая прозрачность — пользователю сложно понять, почему предложен тот или иной контакт.

В отличие от Tinder, Hinge делает ставку не на частоту свайпов, а на глубину взаимодействий. Это требует от системы учёта более сложных поведенческих метрик, включая динамику общения после совпадения. Такой подход требует более тонкой настройки, но лучше отвечает цели платформы — формированию устойчивых связей [4].

1.2.4 eHarmony

Рекомендательная система eHarmony изначально создавалась как экспертная модель совместимости, основанная на глубоком психологическом тестировании. Платформа ориентирована на долгосрочные отношения, а не на быстрые знакомства. Это определяет как архитектуру системы, так и методологию сбора и обработки данных.

В отличие от более поведенчески-ориентированных платформ, eHarmony делает ставку на анкеты, основанные на психологической типологии. При регистрации каждый пользователь заполняет обширную анкету, содержащую от 100 до 150 вопросов, касающихся:

— личностных черт (экстраверсия, добросовестность, невротизм и др.);
— ценностей и жизненных установок;
— отношения к конфликтам, компромиссам, религии и карьере;
— предпочтений в партнёрстве и стиле общения.

Из ответов формируется вектор признаков $x \in \mathbb{R}^d$, где d — число выделенных скрытых характеристик. Далее используется модель оценки совместимости между двумя пользователями u и v на основе расстояния между их признаковыми векторами:

$$S(u, v) = 1 - \frac{\|x_u - x_v\|}{D}$$

где D — нормирующий коэффициент (максимально возможное расстояние), а $S(u, v)$ — мера совместимости, принимающая значения от 0 до 1.

Алгоритм может включать дополнительные поправки:
— усиление совпадений по наиболее значимым признакам;
— штрафы за критические несовпадения (например, по отношению к детям или религии);
— предпочтение партнёров, близких по возрасту, географии или культурному фону.

Система реализует фильтрацию на основе заранее определённой модели совместимости, а не обучения на пользовательском поведении. Это означает, что:

1. рекомендации стабильны во времени и не зависят от текущей активности;
2. пользователи получают небольшой, но тщательно отобранный список совпадений;
3. основной целью алгоритма является структурная совместимость, а не привлекательность.

Несмотря на высокую психологическую обоснованность, такой подход имеет ряд ограничений:
— высокая когнитивная нагрузка при регистрации, что отпугивает часть пользователей;
— отсутствие адаптации под поведение — система не обучается на реальных откликах;
— возможная переориентация на типаж, а не на реальное разнообразие предпочтений.

Тем не менее, eHarmony демонстрирует устойчивую эффективность в своей нише, благодаря глубокой проработке модели совместимости и ориентации на фундаментальные ценности и личностные черты. Такой подход хорошо подходит для пользователей, ищущих стабильные и продолжительные отношения [11].

1.3 Особенности сбора и представления пользовательских признаков

В рекомендательных системах, применяемых в онлайн-знакомствах, центральную роль играет сбор и интерпретация информации о предпочтениях пользователей. Эти данные служат основой для построения персонализированных профилей и определения потенциально совместимых кандидатов. Анализ существующих решений в данной предметной области выявляет два крайних подхода к сбору информации:

— длинные анкеты с детализированными вопросами, характерные для платформ вроде eHarmony, обеспечивают богатое представление о пользователе, но требуют значительных усилий при заполнении и приводят к высокой доле отказов;

— минималистичные интерфейсы типа Tinder предлагают быструю оценку по принципу свайпа пользователей, обеспечивая высокую конверсию, но теряя глубину предпочтений.

В качестве компромиссного решения рассматриваются тернарные опросы, в которых каждый вопрос допускает три варианта реакции: положительную, отрицательную и нейтральную. Это позволяет выразить отношение к различным характеристикам без излишней нагрузки. Каждый ответ кодируется значением $q_i \in \{-1, 0, 1\}$, где i — номер вопроса. Такая шкала обеспечивает:

— компактность векторного представления профиля;

— возможность учитывать отсутствие чёткой позиции;

— поддержку различных методов машинного обучения и заполнения пропусков.

Результаты сбора формируются в разреженную матрицу $R = (q_{u,i})$, где u — пользователь, i — вопрос, а $q_{u,i}$ — реакция. Пропущенные значения соответствуют отсутствию взаимодействия. На практике возможно использование стратегий заполнения: нулями, средним значением по вопросу, либо построение модели, устойчивой к разреженности.

В дополнение к ответам могут быть включены и другие признаки:

— демографические данные — возраст (нормированный), пол (бинарная переменная);
— описания «о себе», конвертируемые в эмбеддинги с помощью языковых моделей;
— временные характеристики — длительность прохождения, количество пропусков, уверенность в ответах.

Особенность предметной области знакомств заключается в том, что каждый пользователь является одновременно субъектом и объектом рекомендаций. Это накладывает дополнительные требования к симметричности представлений и способности учитывать двустороннюю заинтересованность. Более того, успешность рекомендации здесь не сводится к лайку, а может быть оценена через цепочку взаимодействий: переписка, ответный интерес, встречи.

На этапе анализа задач были также выявлены следующие характеристики, которые следует учитывать при проектировании системы:

— необходимость работать с разреженными признаковыми матрицами и отсутствием части информации;
— поддержка холодного старта для новых пользователей и вопросов;
— ориентация на сравнение пользователей между собой, а не на ранжирование объектов фиксированной природы.

Таким образом, представление предпочтений в тернарной форме, дополненное демографией и эмбеддингами, формирует универсальную основу для построения профилей. Эти профили могут использоваться в различных методах рекомендации — от коллаборативной фильтрации до кластеризации и нейросетевых моделей.

2 Методы построения рекомендательной системы

2.1 Коллаборативная фильтрация

Коллаборативная фильтрация является одним из наиболее популярных подходов в рекомендательных системах, особенно в условиях, когда отсутствует явное описание объектов или пользователей. Основная идея заключается в том, что предпочтения пользователей могут быть предсказаны на основе поведения других пользователей с похожими вкусами.

Существует два основных типа коллаборативной фильтрации: на основе памяти (memory-based) и на основе модели (model-based). Первый подход использует метрики сходства между пользователями или объектами (например, косинусное расстояние, корреляцию Пирсона) и агрегирует оценки соседей. Второй — строит параметризованную модель на основе данных о взаимодействиях, чаще всего через факторизацию матрицы.

Пусть имеется матрица взаимодействий $R \in \mathbb{R}^{m \times n}$, где m — количество пользователей, n — количество объектов (например, анкет потенциальных партнёров). Элемент $r_{u,i}$ может обозначать бинарную оценку (лайк/не лайк), числовой рейтинг или иной сигнал предпочтения. Коллаборативная фильтрация предполагает, что в матрице присутствует скрытая структура, отражающая закономерности во вкусах пользователей [12].

Одним из распространённых методов является сингулярное разложение (SVD) или его модификации (например, Funk-SVD, ALS), при которых R аппроксимируется как

$$R \approx UV^\top,$$

где $U \in \mathbb{R}^{m \times k}$ и $V \in \mathbb{R}^{n \times k}$ содержат латентные векторы пользователей и объектов соответственно, а k — число латентных признаков. Значение $\hat{r}_{u,i} = U_u \cdot V_i^\top$ интерпретируется как предсказанная степень интереса пользователя u к объекту i.

Интересным обобщением является применение коллаборативной фильтрации к данным не только о лайках, но и о признаках, таких как ответы пользователей на опросы. В этом случае каждый пользователь представлен тернарным или категориальным вектором, а матрица $Q \in \{-1, 0, 1\}^{m \times p}$ (где p — число вопросов) также может быть факторизована аналогичным способом:

$$Q \approx U'Z^\top.$$

Полученные векторы могут использоваться для оценки схожести между пользователями, для восстановления пропущенных ответов или как источник признаков для гибридных моделей.

Коллаборативная фильтрация демонстрирует высокую эффективность при наличии большого количества пользовательских взаимодействий, однако страдает от проблемы холодного старта и может усиливать популярность одних и тех же объектов, снижая разнообразие рекомендаций [13].

2.2 Контентная фильтрация

Контентная фильтрация представляет собой один из классических подходов к построению рекомендательных систем. Основная идея заключается в том, чтобы рекомендовать объекты, схожие с теми, которые пользователь оценил положительно ранее, основываясь на характеристиках самих объектов. В отличие от коллаборативной фильтрации, здесь не учитываются предпочтения других пользователей.

Каждый объект описывается вектором признаков, которые могут быть бинарными, числовыми или категориальными. Пусть объект j представлен вектором признаков $x_j \in \mathbb{R}^d$. Модель пользователя строится как агрегированное представление объектов, с которыми у него были положительные взаимодействия. Например, если пользователь i взаимодействовал с объектами $j_1, \ldots, j_k$, то вектор предпочтений можно получить как среднее:

$$p_i = \frac{1}{k} \sum_{s=1}^{k} x_{j_s}.$$

Рекомендации формируются на основе сходства между вектором предпочтений пользователя и векторами новых объектов. Чаще всего используется косинусное расстояние:

$$\mathrm{sim}(p_i, x_j) = \frac{p_i^\top x_j}{\|p_i\| \cdot \|x_j\|}.$$

Объекты с наибольшим значением sim включаются в топ рекомендаций [14].

Данный подход имеет ряд достоинств:

— высокая интерпретируемость: можно объяснить, почему был рекомендован тот или иной объект;

— независимость от количества других пользователей;
— устойчивость к проблеме холодного старта для объектов (если их описание доступно).

Однако контентная фильтрация имеет и ограничения:

— рекомендации ограничиваются областью уже проявленных интересов;
— трудно учитывать сложные зависимости между признаками;
— качество зависит от полноты и выразительности признакового описания.

Для повышения гибкости могут применяться методы машинного обучения. Например, обучающая выборка может включать пары (x_j, y_{ij}), где y_{ij} — бинарная переменная, указывающая наличие положительного отклика со стороны пользователя i на объект j. На этой основе можно обучить логистическую регрессию, SVM или градиентный бустинг, предсказывающий вероятность интереса к новому объекту.

Также возможны гибридные модели, объединяющие контентную и коллаборативную фильтрацию. Например, контентные признаки могут использоваться для регуляризации матричной факторизации или служить входом для нейронных сетей. Такие подходы позволяют улучшить обобщающую способность модели и преодолеть узость интересов, характерную для чисто контентной фильтрации [15].

Контентные методы особенно полезны в системах, где объекты имеют чётко выраженные признаки: тексты, категории, изображения, ответы на тесты или анкеты. При наличии информативного описания они позволяют получать рекомендации уже на самых ранних этапах использования системы, что делает их важным компонентом гибридных решений.

2.3 Эвристики: совпадения по ответам, популярность, фильтры

Эвристические методы в рекомендательных системах основаны на наборе простых правил и предположений, позволяющих быстро и эффективно формировать рекомендации. Несмотря на относительную простоту, такие подходы остаются актуальными, особенно в условиях ограниченности данных или требований к объяснимости.

Одним из базовых эвристических подходов является сравнение пользователей по их ответам на вопросы анкет или опросов. Если ответы представлены в тернарной шкале (например, -1 — несогласие, 0 — нейтрально, 1 — согласие), то схожесть между пользователями можно оценить по доле совпадающих

ответов:

$$\mathrm{sim}(u, v) = \frac{1}{|P|} \sum_{i \in P} (q_{u,i} = q_{v,i}),$$

где P — множество вопросов, на которые оба пользователя дали ответ. Этот подход применим как в системах знакомств, так и в других областях, где важна совместимость взглядов, предпочтений и интересов.

Другим эвристическим приёмом является использование показателя популярности. Объекты (например, профили или товары), получившие наибольшее количество положительных оценок, могут предлагаться новым пользователям в качестве стартовых рекомендаций. Популярность может быть нормирована, например:

$$\mathrm{pop}(i) = \frac{\text{число лайков объекта } i}{\text{максимальное число лайков среди всех объектов}}.$$

Популярные рекомендации часто дополняются фильтрацией по демографическим и другим признакам, таким как возраст, пол, географическое местоположение, язык, наличие общих интересов и т. д [16].

Такие фильтры применяются до или после основного ранжирования и позволяют исключить очевидно нерелевантные варианты. Например, пользователь, заинтересованный только в кандидатах определённого возраста или пола, должен получать только соответствующие предложения.

Также может применяться эвристика совпадения по ключевым признакам. Если в профиле пользователя указаны предпочтения (например, любимые фильмы, занятия, взгляды), система может искать совпадения с другими профилями и ранжировать их по количеству совпавших интересов.

Комбинирование эвристик позволяет построить гибкую систему, способную адаптироваться к условиям отсутствия данных, начальной загрузки, а также повысить обоснованность рекомендаций. При этом эвристические методы легко интерпретируемы, что важно в чувствительных сферах, таких как онлайн-знакомства или подбор персонала.

2.4 Применение кластеризация (k-means, DBSCAN) в рекомендациях

Кластеризация — это метод обучения без учителя, направленный на группировку объектов в кластеры таким образом, чтобы элементы одного кластера были похожи друг на друга и отличались от элементов других кластеров. В

контексте рекомендательных систем кластеризация применяется для:

— сегментирования пользователей (или объектов) по интересам, поведению или признакам;
— повышения масштабируемости рекомендаций за счёт ограничения поиска релевантных кандидатов внутри кластера;
— выявления нишевых предпочтений и персонализированных паттернов.

Одним из наиболее популярных алгоритмов является k-means. Он принимает на вход число кластеров k и итеративно минимизирует внутрикластерную дисперсию:

$$\sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2,$$

где C_j — кластер, а μ_j — его центр. Алгоритм эффективен при компактных и сферических кластерах, но чувствителен к выбору k и неустойчив к выбросам.

Для более гибкой кластеризации используется алгоритм DBSCAN. Он группирует точки по плотности: кластером считается связная по плотности область, где каждая точка имеет хотя бы $minPts$ соседей в пределах радиуса ε. DBSCAN способен находить кластеры произвольной формы и автоматически игнорирует шум (выбросы), что делает его особенно полезным в разнородных данных.

В рекомендательных системах кластеризация применяется по-разному:

— На пространстве пользователей: сегментация по поведенческим или опросным признакам позволяет формировать кластеры пользователей с похожими предпочтениями. Рекомендации для нового пользователя можно извлекать из наиболее близкого кластера.
— На пространстве объектов: группировка товаров, фильмов, анкет и пр. по тематике, стилю или целевой аудитории позволяет адаптировать рекомендации к интересам пользователя.
— В латентных пространствах: кластеризация векторов после факторизации (например, в SVD или autoencoder-подходах) даёт более сжатое и семантически значимое представление.

Кластеризация также применяется для визуализации и анализа структуры пользовательской базы, выявления целевых групп и построения тематических подборок. Её эффективность во многом зависит от выбора признаков и масштабов данных, поэтому нередко она используется в комбинации с другими

методами [17].

2.5 Стратегии холодного старта

Проблема холодного старта возникает, когда система не располагает достаточной информацией о пользователях или объектах, чтобы формировать персонализированные рекомендации. Выделяют два основных сценария: появление нового пользователя и добавление нового объекта (например, анкеты).

Для новых пользователей могут применяться следующие подходы:

- заполнение вступительных тестов или анкет — позволяет собрать первичные признаки и использовать их при формировании рекомендаций;
- использование демографических данных — рекомендации подбираются на основе поведения пользователей с аналогичными характеристиками (возраст, пол, география и т.д.);
- показ популярных объектов — временная стратегия, при которой пользователю демонстрируются анкеты с высокой оборачиваемостью, что помогает быстрее сформировать профиль предпочтений.

При появлении новых объектов, которые ещё не получили откликов, возможны такие меры:

- временное повышение приоритета в выдаче — например, показ новым или активным пользователям для ускоренного накопления статистики;
- использование признаков схожести — на основе анкетных данных, внешности (в случае CV), текста описания (в случае NLP) или ответов на вопросы;
- размещение в релевантных сегментах — объект может быть временно включён в выдачу по кластерам, в которые он потенциально попадает по признаковому пространству.

Кроме того, существуют универсальные стратегии, применимые и к новым пользователям, и к новым объектам:

- инициализация признаков с помощью доступных внешних данных — анкет, биографий, метаинформации;
- использование гибридных моделей, сочетающих элементы content-based и коллаборативной фильтрации — это снижает чувствительность к отсутствию истории;
- активное обучение — выбор контента, максимально полезного для уточнения предпочтений, что позволяет за минимальное число взаимодействий

улучшить качество рекомендаций.

Проблема холодного старта наиболее критична для систем, основанных на коллаборативной фильтрации, поскольку они требуют исторических данных о взаимодействии пользователей с объектами. Поэтому многие современные решения строятся с использованием дополнительных эвристик и предварительной инициализации признаков, что позволяет обеспечить устойчивость системы на ранних этапах использования [18].

2.6 Методы глубокого обучения в рекомендательных системах

Развитие нейросетевых архитектур оказало существенное влияние на область рекомендательных систем. Благодаря способности извлекать сложные скрытые зависимости из разнородных данных, модели глубокого обучения применяются для повышения качества рекомендаций как в традиционных задачах (например, предсказание рейтингов), так и в более сложных сценариях, включая мультимодальные рекомендации, учет временной динамики и персонализацию на основе контекста.

Одним из первых направлений стало расширение классической матричной факторизации с помощью нейронных сетей. Вместо простой линейной факторизации матрицы взаимодействий $R \in \mathbb{R}^{m \times n}$, где m — количество пользователей, n — количество объектов, и R_{ij} — факт взаимодействия, используются обучаемые эмбеддинги и нелинейные функции активации. Примером такой модели является Neural Collaborative Filtering (NCF) [19]. В NCF пары эмбеддингов (u_i, v_j) передаются через многослойный перцептрон:

$$\hat{r}_{ij} = \mathrm{MLP}([u_i, v_j]),$$

где $[\cdot, \cdot]$ обозначает конкатенацию векторов. Модель обучается по функции потерь, например, бинарной кросс-энтропии в задаче предсказания лайка/дизлайка.

Другое направление связано с использованием рекуррентных и трансформерных архитектур для моделирования последовательности взаимодействий. Так называемые sequence-based recommenders учитывают порядок взаимодействий пользователя с объектами. Например, модель GRU4Rec применяет Gated Recurrent Unit (GRU) [20] для обработки истории действий пользователя. Базовая идея заключается в следующем: пусть $x_1, x_2, \ldots, x_T$ — последовательность

взаимодействий, тогда на каждом шаге рассчитывается скрытое состояние h_t:

$$h_t = \mathrm{GRU}(x_t, h_{t-1}),$$

и предсказывается следующий элемент x_{t+1} с помощью softmax-слоя.

Для учёта более длинных зависимостей и параллельной обработки была предложена модель SASRec, основанная на механизме внимания. В отличие от рекуррентных моделей, здесь используется позиционно-кодированная последовательность эмбеддингов, проходящая через слои трансформера. Это позволяет учитывать контекст всех предыдущих действий при выборе следующей рекомендации [21].

Особое место занимают графовые нейронные сети, применяемые для моделирования взаимодействий в виде графов. В Graph Convolutional Matrix Completion [22], каждый пользователь и объект представляются вершинами, соединёнными ребром при наличии взаимодействия. Представления вершин обновляются по правилу:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ij}} W^{(l)} h_j^{(l)}\right),$$

где $\mathcal{N}(i)$ — соседи вершины i, $W^{(l)}$ — матрица весов на l-м слое, c_{ij} — коэффициент нормализации. Такой подход позволяет учитывать структуру взаимодействий и взаимосвязь объектов, что особенно актуально для задач, где важны не только пользовательские предпочтения, но и социальные или контекстные связи.

Ещё одно активно развивающееся направление — мультимодальные рекомендации. Здесь помимо взаимодействий учитываются дополнительные признаки, такие как текст описания, изображения, аудио. Для обработки таких данных применяются CNN, BERT и другие специализированные архитектуры. В модели VBPR визуальные признаки изображений используются для расширения латентного пространства:

$$\hat{r}_{ij} = u_i^\top v_j + u_i^\top E x_j,$$

где x_j — визуальный вектор, полученный из изображения объекта, E — обучаемая матрица проекции [23].

Основные преимущества методов глубокого обучения:

— способность моделировать сложные нелинейные зависимости между пользователями и объектами;
— возможность использовать разнородные источники информации;
— высокая гибкость и расширяемость архитектур.

Тем не менее, существуют и ограничения:

— высокая требовательность к вычислительным ресурсам;
— потребность в большом объёме размеченных данных;
— трудность интерпретации и объяснимости результатов.

Таким образом, методы глубокого обучения открывают широкие перспективы для построения более точных и персонализированных рекомендательных систем, особенно в условиях, когда доступны дополнительные источники информации и достаточно ресурсов для обучения. Однако выбор таких подходов должен быть обоснован задачами проекта, размером аудитории и доступной инфраструктурой.

2.7 Гибридные подходы

Гибридные рекомендательные системы объединяют преимущества различных методов, включая коллаборативную фильтрацию, контентную фильтрацию и эвристические алгоритмы. Такая интеграция позволяет компенсировать слабые стороны отдельных подходов и достичь более высокой точности, устойчивости к холодному старту и разнообразия рекомендаций.

Существуют разные стратегии гибридизации:

— объединение выходов нескольких моделей (late fusion);
— комбинирование признаков на входе одной модели (early fusion);
— использование одного подхода в качестве фильтра, а другого — для ранжирования.

Один из классических примеров — модель, в которой одновременно учитываются схожесть пользователей (коллаборативная составляющая) и сходство объектов (контентная составляющая). Пусть r_{ui} — предсказанная оценка пользователя u для объекта i. Тогда комбинированная формула может иметь следующий вид:

$$r_{ui} = \alpha \cdot r_{ui}^{\text{collab}} + (1 - \alpha) \cdot r_{ui}^{\text{content}},$$

где r_{ui}^{collab} — предсказание по коллаборативной модели, r_{ui}^{content} — результат контентного ранжирования, а $\alpha \in [0, 1]$ — параметр, регулирующий вклад каждой части.

Другой подход основан на поэтапной фильтрации. Например, можно сначала применить контентную фильтрацию для предварительного отбора релевантных объектов, а затем выполнить коллаборативное ранжирование по пользовательским лайкам или оценкам. Такой двухшаговый процесс снижает вычислительную нагрузку и повышает релевантность [24].

Гибридные системы особенно полезны в следующих сценариях:

— наличие разреженной матрицы пользовательских взаимодействий, при этом имеются структурированные признаки объектов;

— необходимость рекомендаций для новых пользователей или новых объектов;

— желание учитывать не только историю взаимодействий, но и семантическое содержание;

— ориентация на объяснимость модели и возможность интерактивной настройки предпочтений.

В последние годы широкое распространение получили модели, встраивающие оба подхода в общую латентную структуру. Например, в модели Factorization Machines (FM) и её нейронных расширениях (Neural FM, DeepFM) признаки пользователей и объектов подаются в общий обучаемый слой, позволяющий учитывать как контентную, так и взаимодействующую информацию. Обозначим входной вектор признаков как x, тогда предсказание в FM-модели имеет вид:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j,$$

где v_i — латентный вектор i-го признака. Такая модель способна улавливать взаимодействия между признаками без явного задания структуры [25].

Гибридные методы на практике показывают высокую гибкость и адаптивность, особенно в условиях изменяющихся предпочтений и ограниченных пользовательских данных. Они успешно применяются в рекомендательных системах крупных платформ (Netflix, YouTube, LinkedIn), где важно сочетать поведенческие паттерны с контекстной и персонализированной информацией.

Совокупность перечисленных подходов делает гибридные методы универсальным инструментом, который можно адаптировать под особенности конкретной предметной области, в том числе в сфере онлайн-знакомств.

2.8 Методы оценки рекомендательной системы

Оценка качества рекомендательной системы в приложении знакомств играет ключевую роль для обеспечения релевантности и привлекательности предложений, предоставляемых пользователю. Цель такой оценки — убедиться, что система способствует установлению взаимного интереса и активному взаимодействию между пользователями.

В рамках мобильного приложения знакомств эффективность рекомендаций оценивается с двух сторон: на основе исторических данных (оффлайн) и в реальном времени (онлайн).

Оффлайн-оценка проводится до запуска системы на основе известных пользовательских взаимодействий. При этом выбирается контрольная выборка, в которой определённые события (например, взаимные лайки) скрываются, а алгоритм должен их предсказать. Основное внимание в данном случае уделяется метрике HitRate@K — доле случаев, в которых хотя бы один релевантный объект оказался среди топ-K рекомендаций. Эта метрика отражает способность системы «попадать» в интересы пользователя и используется в качестве основной целевой метрики.

Также применяются следующие метрики:

— Precision@K — доля релевантных пользователей среди K рекомендованных;

— Recall@K — доля релевантных пользователей, которые удалось рекомендовать;

— MRR (Mean Reciprocal Rank) — учитывает позицию первого релевантного объекта в списке;

— NDCG — нормированная кумулятивная дисконтированная полезность, учитывающая ранжирование;

— Coverage — доля пользователей, которым система способна выдать хотя бы одну осмысленную рекомендацию. Высокий coverage свидетельствует о способности системы охватывать широкую аудиторию.

Оффлайн-оценка обычно проводится по схеме leave-one-out, когда одна интеракция пользователя исключается из тренировочного множества и исполь-

зуется для тестирования [26].

После внедрения системы важно отслеживать эффективность рекомендаций по поведенческим метрикам в режиме онлайн. Наиболее информативные из них:

1. Доля взаимных лайков среди выданных рекомендаций;
2. Конверсия в диалог — отношение числа начатых чатов к количеству рекомендованных профилей;
3. Среднее время до первого взаимодействия (лайка или сообщения);
4. Удержание пользователей — как долго и регулярно пользователь взаимодействует с рекомендованными контактами;
5. Повторные взаимодействия — оценивается, продолжается ли активность с рекомендованным пользователем спустя время.

Помимо точности, важно учитывать качественные характеристики системы, которые напрямую влияют на пользовательское восприятие и вовлечённость:

- Разнообразие — рекомендации не должны быть однотипными;
- Новизна — включение ранее не встречавшихся кандидатов повышает интерес;
- Персонализация — учёт индивидуальных особенностей конкретного пользователя;
- Равномерность — отсутствие систематического перекоса в сторону ограниченной группы пользователей.

Для оценки рекомендательной системы в приложении знакомств важно сочетать количественные метрики точности и охвата с показателями пользовательского поведения. Такой подход позволяет обеспечить релевантность, персонализацию и устойчивую вовлечённость аудитории.

3 Теоретические основы разработки мобильных приложений

3.1 Общие особенности мобильной разработки

Мобильные приложения функционируют в условиях, отличающихся от настольной или серверной среды. Эти отличия накладывают определённые ограничения и формируют особые требования к проектированию, реализации и тестированию.

Во-первых, мобильные устройства имеют ограниченные ресурсы. Смартфоны и планшеты уступают настольным системам по вычислительной мощности, объёму оперативной памяти и возможностям хранения данных. Кроме того, приложения должны быть чувствительны к расходу батареи, особенно при активном использовании мультимедиа, анимации и фоновых процессов.

Во-вторых, в экосистеме мобильных устройств наблюдается значительное разнообразие. Существует множество моделей устройств с различными размерами экранов, плотностью пикселей, аппаратными возможностями и версиями операционных систем. Это требует особого внимания к адаптивности интерфейса и совместимости приложения с широким спектром устройств.

Помимо технических ограничений, важным аспектом является пользовательский опыт. Пользователи ожидают от приложения высокой скорости отклика, визуальной плавности и интуитивной структуры. Низкое качество интерфейса или перегруженность функциональностью могут привести к отказу от использования, независимо от пользы приложения.

Среди ключевых особенностей мобильной разработки можно выделить:

- ограниченные вычислительные ресурсы устройства;
- требования к экономии энергии и управлению фоновыми задачами;
- разнообразие устройств, экранов и версий операционных систем;
- необходимость обеспечения высокого уровня интерактивности и отзывчивости интерфейса;
- ориентация на кратковременные, но частые сценарии использования.

Эти особенности определяют специфику проектирования мобильных решений. Для успешной реализации приложения разработчику необходимо соблюдать ряд принципов, направленных на обеспечение стабильности, производительности и удобства использования [27].

К таким принципам относятся:

1. адаптация пользовательского интерфейса под различные размеры и ори-

ентации экранов;

2. минимизация использования ресурсов устройства;
3. соблюдение рекомендаций по дизайну, принятых в целевой платформе;
4. обеспечение плавной и предсказуемой навигации по приложению;
5. проведение тестирования на разных устройствах и версиях операционной системы.

Таким образом, мобильная разработка представляет собой область, требующую сочетания инженерной дисциплины, внимания к деталям и ориентации на пользовательский опыт. Эти аспекты определяют основу архитектурных и технологических решений в мобильных системах.

3.2 Общие подходы к проектированию мобильных систем

Проектирование мобильных приложений требует системного подхода, включающего выбор архитектурной модели, структурирование компонентов и определение принципов взаимодействия между ними. Эти решения оказывают значительное влияние на надёжность, масштабируемость и удобство сопровождения системы.

К числу ключевых факторов, влияющих на архитектуру мобильного приложения, относятся:

— целевая платформа (Android, iOS или обе);
— требования к производительности и времени отклика;
— предполагаемый объём пользовательского трафика и сценарии нагрузки;
— организационные ограничения (время, бюджет, ресурсы);
— ожидаемая эволюция проекта и возможность масштабирования.

Современные мобильные приложения обычно строятся как распределённые системы, в которых клиент и сервер выполняют разные роли. Распространённые архитектурные подходы включают:

— монолитную архитектуру с полной локальной логикой (редко применяется в современных системах);
— клиент-серверную архитектуру с разделением обязанностей между устройством пользователя и серверной частью;
— модульную или микросервисную архитектуру, обеспечивающую масштабируемость и гибкость.

При этом важную роль играет концепция разделения ответственности: пользовательский интерфейс, бизнес-логика и работа с данными разносится

по разным слоям приложения. Это способствует улучшению тестируемости, повторному использованию компонентов и упрощению сопровождения [28].

Независимо от конкретной реализации, архитектура мобильной системы должна обеспечивать:

1. слабую связанность между модулями;
2. чёткие границы между слоями;
3. стандартизированное взаимодействие между компонентами (например, через REST API);
4. возможность независимого обновления и масштабирования подсистем.

Такой подход обеспечивает устойчивую основу для развития и поддержки мобильного приложения в условиях изменяющихся требований и роста нагрузки.

3.3 Клиент-серверная архитектура

Клиент-серверная архитектура является одним из наиболее устойчивых и широко применяемых подходов при построении современных мобильных приложений. Её основным принципом является логическое разделение системы на два взаимосвязанных компонента: клиентскую часть, работающую на пользовательском устройстве, и серверную часть, выполняющую обработку запросов, управление данными и реализацию бизнес-логики.

Данный архитектурный подход обеспечивает целый ряд преимуществ:

— централизованное управление и консистентность данных;
— разгрузка клиентской части за счёт переноса вычислений на сервер;
— возможность переиспользования серверной логики в различных клиентских интерфейсах (веб, мобильных и пр.);
— упрощение обновления клиентских приложений без необходимости модификации серверного кода;
— гибкость масштабирования и балансировки нагрузки на стороне сервера.

Роль клиентской части заключается в следующем:

— предоставление пользовательского интерфейса и взаимодействие с пользователем;
— сбор и первичная обработка пользовательских данных;
— формирование сетевых запросов и обработка полученных ответов;
— управление локальным состоянием приложения и навигацией.

Серверная часть, в свою очередь, обеспечивает:

— реализацию бизнес-правил и логики приложения;
— аутентификацию и авторизацию пользователей;
— централизованное хранение и обработку данных;
— взаимодействие с внешними сервисами;
— логирование, мониторинг и обеспечение безопасности.

Взаимодействие между клиентом и сервером, как правило, осуществляется через стандартизированные протоколы (например, HTTP) и форматы обмена данными (чаще всего JSON или XML). Такой подход обеспечивает платформонезависимость и простоту интеграции.

Для реализации клиентской части всё более широкое распространение получают кроссплатформенные фреймворки, такие как Flutter. Он позволяет разрабатывать приложения с единой кодовой базой, сохраняя при этом высокую производительность и выразительность интерфейса.

Серверная логика зачастую реализуется с использованием зрелых и гибких платформ, таких как Spring, обеспечивающих поддержку REST API, управление транзакциями, безопасность и масштабируемость [29].

Таким образом, клиент-серверная архитектура остаётся актуальной и надёжной основой для построения мобильных систем, сочетающих в себе адаптивность, расширяемость и удобство сопровождения.

3.4 Безопасность и конфиденциальность

В современных мобильных приложениях вопросы безопасности и конфиденциальности данных занимают центральное место. Пользователи ожидают, что их личная информация будет защищена от несанкционированного доступа, утечки или подмены. Это особенно актуально для приложений, обрабатывающих чувствительные персональные данные, включая профили пользователей, переписку, предпочтения и другую информацию личного характера.

Обеспечение безопасности требует комплексного подхода, охватывающего как клиентскую, так и серверную часть приложения. В числе ключевых задач:
— защита канала передачи данных от перехвата и модификации;
— надёжная аутентификация и авторизация пользователей;
— контроль доступа к защищённым ресурсам;
— предотвращение распространённых уязвимостей (включая XSS, CSRF, SQL-инъекции и другие);
— обеспечение целостности и актуальности пользовательской сессии;

— соблюдение требований к обработке и хранению персональных данных.

Одной из стандартных практик является реализация авторизации на основе токенов. Наиболее распространённым решением в этом направлении являются компактные самодостаточные токены, которые позволяют хранить информацию о пользователе и его правах доступа в зашифрованном или подписанном виде. Использование токенов удобно в распределённых системах и облегчает реализацию масштабируемой безсессионной архитектуры [30].

Для защиты пользовательских данных от перехвата при передаче, общепринятым стандартом является использование защищённых транспортных протоколов. В частности, применяется HTTPS, основанный на TLS, который обеспечивает шифрование и аутентификацию соединения между клиентом и сервером.

На серверной стороне безопасность реализуется посредством встроенных или внешних фреймворков, позволяющих централизованно управлять доступом, разграничивать привилегии и применять политики безопасности. Такие решения также обеспечивают:

— предварительную фильтрацию входящих запросов;
— управление сессиями или токенами;
— обработку исключений, связанных с неавторизованным доступом;
— регистрацию действий пользователей для целей аудита.

С точки зрения пользовательского опыта и доверия, защита конфиденциальности является неотъемлемой частью проектирования. Приложение должно обеспечивать безопасное хранение пользовательских данных, возможность их удаления, а также отказоустойчивость в случае попыток несанкционированного доступа [31].

На практике к задачам обеспечения безопасности подходят с учётом принципов минимизации данных, разделения ответственности и поэтапного усиления контроля. Это позволяет проектировать надёжные и устойчивые к угрозам системы, соответствующие современным ожиданиям и требованиям.

4 Проектирование архитектуры приложения для знакомств

4.1 Проектирование архитектуры мобильного приложения

Проектирование архитектуры мобильного приложения было выполнено на основе системного анализа предполагаемой функциональности, пользовательских сценариев и требований к масштабируемости. Приложение реализуется как клиентская часть, взаимодействующая с сервером через REST API, и играет роль интерфейса между конечным пользователем и рекомендательной системой.

Архитектура приложения строится вокруг концепции модулярности и разделения ответственности. Это означает, что каждый экран и компонент реализует строго определённую роль, что упрощает тестирование, расширение и поддержку.

В качестве основного навигационного каркаса был выбран стек-навигации с возможностью модального перехода, поскольку пользователи часто выполняют действия в глубину (например, редактирование профиля, просмотр чужих профилей, переход в чат) с последующим возвратом к предыдущему контексту.

Начальной точкой проектирования стало обеспечение контроля доступа и персонализации взаимодействия. Приложение начинается с экрана входа, где пользователь может авторизоваться или зарегистрироваться. Выбор между этими ветками предопределяет дальнейший маршрут в навигации.

На рисунке 1 представлена схема, иллюстрирующая данный процесс. При регистрации пользователь предоставляет базовую информацию (имя пользователя, пол, краткое описание), которая сохраняется на сервере и может быть использована как часть признаков в системе рекомендаций.

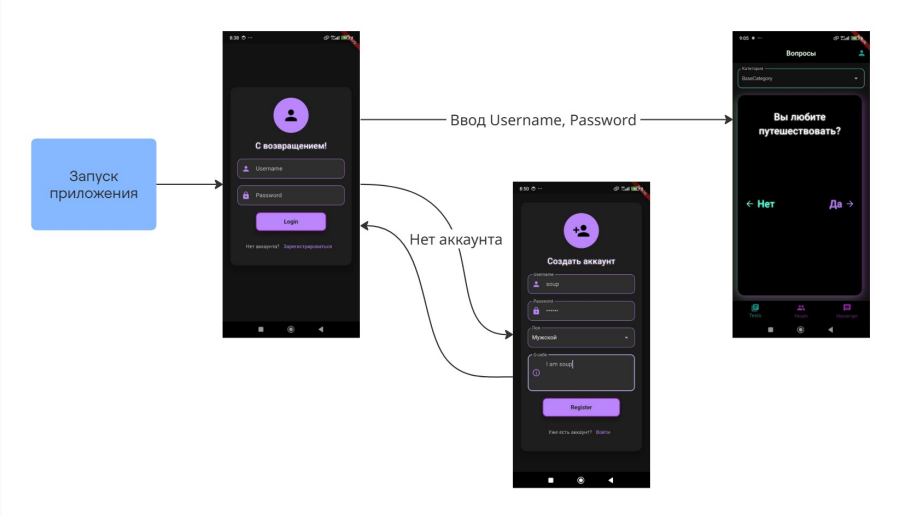

Рисунок 1 – Процесс входа и регистрации

Каждому пользователю доступен экран профиля, где отображается личная информация. В процессе проектирования было определено, что взаимодействие с профилем должно быть двухуровневым: просмотр и редактирование. Таким образом, экран "Профиль"реализован как статическое представление данных, а редактирование (аватара и описания) выполняется через отдельные маршруты. Это позволяет минимизировать когнитивную нагрузку и избежать случайных изменений.

На рисунке 2 показаны возможные действия пользователя с личным профилем. Архитектурно, экран реализован как обособленный компонент, который получает и обновляет данные через API, сохраняя локальное состояние до момента отправки изменений.

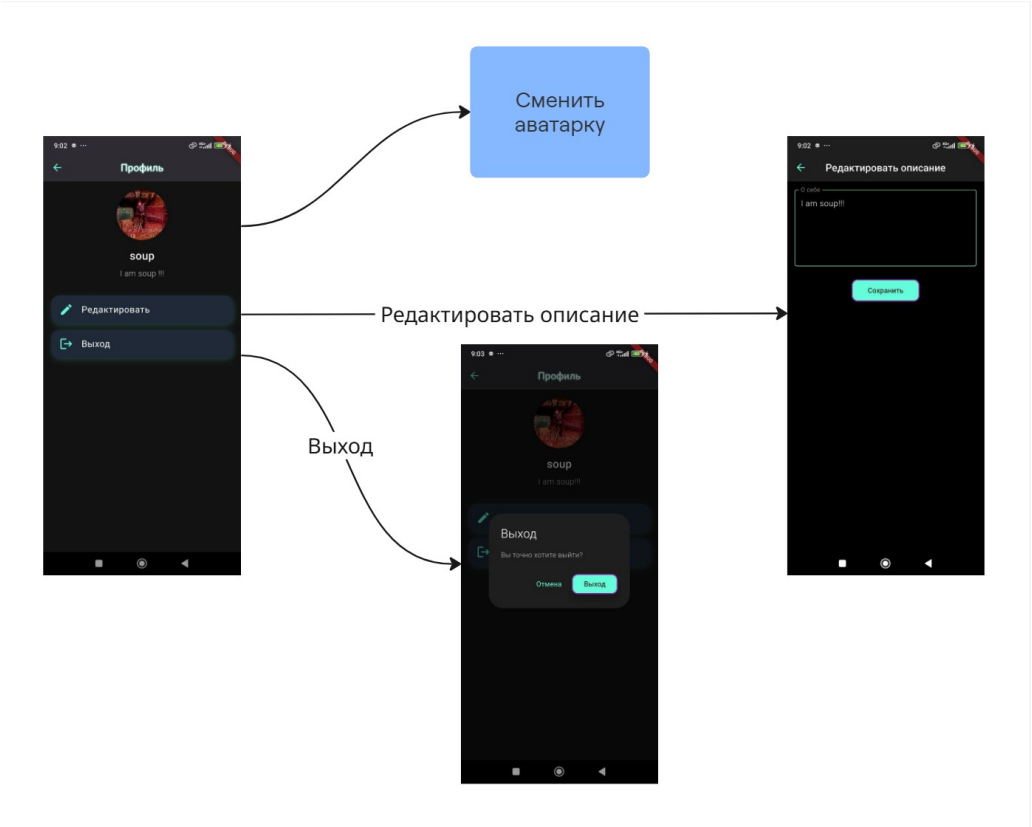

Рисунок 2 – Редактирование профиля и выход

Сердцем рекомендательной системы является вектор признаков, формируемый из ответов пользователя на серию вопросов. Было принято решение реализовать интерфейс прохождения теста через свайпы, поскольку это сочетает простоту использования и скорость. Такой формат также является привычным для пользователей в контексте приложений знакомств.

Как показано на рисунке 3, пользователь выбирает категорию, после чего отображается последовательность карточек с вопросами. Ответ осуществляется

свайпом влево (отрицательный ответ), вправо (положительный) или вверх (нейтральный). При достижении конца набора карточек приложение сообщает об этом. В архитектуре эта часть реализована как отдельный модуль с локальным буфером вопросов, загружаемых с сервера по категориям.

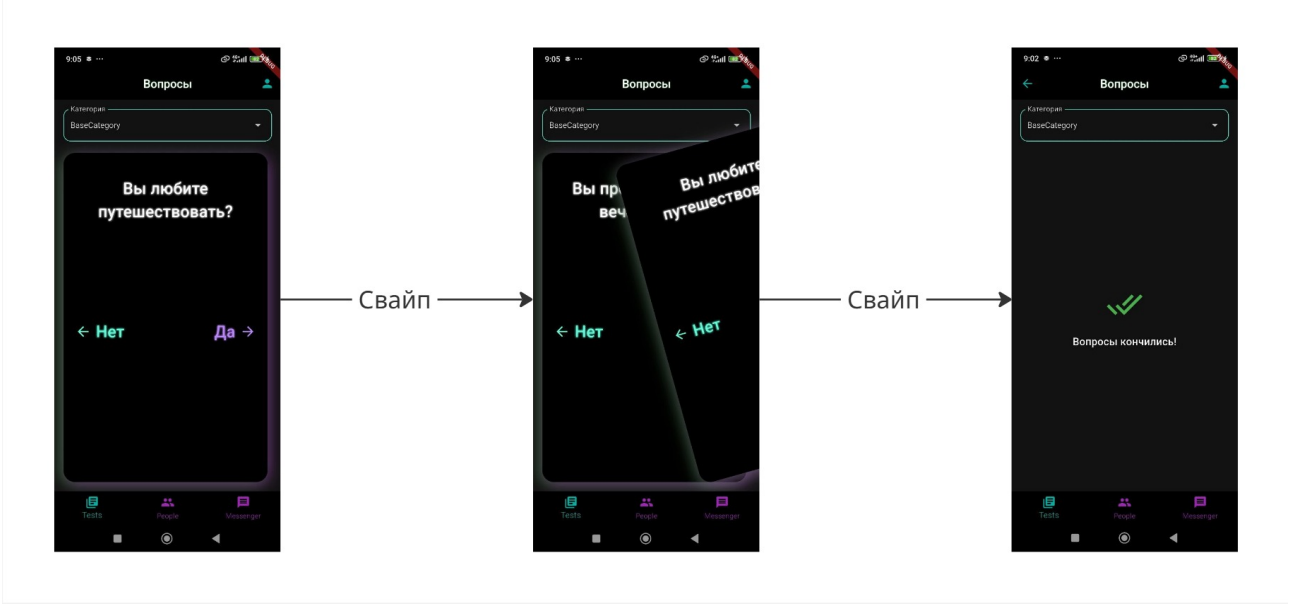

Рисунок 3 – Процесс ответов на вопросы

На основе сформированных векторов признаков сервер предоставляет список пользователей, наиболее близких по сходству. Эта информация отображается на экране рекомендаций . Данный экран проектировался как динамическая таблица, получающая данные из внешнего источника с возможностью фильтрации и сортировки.

Как показано на рисунке 4, каждая карточка рекомендации предоставляет краткую информацию: имя, сходство и кнопку для запроса на чат. Нажатие на карточку открывает модальное окно с расширенным профилем. Эта структура проектировалась для поддержки масштабируемости — при появлении новых фильтров или параметров сортировки можно расширить интерфейс без изменения существующего кода отображения.

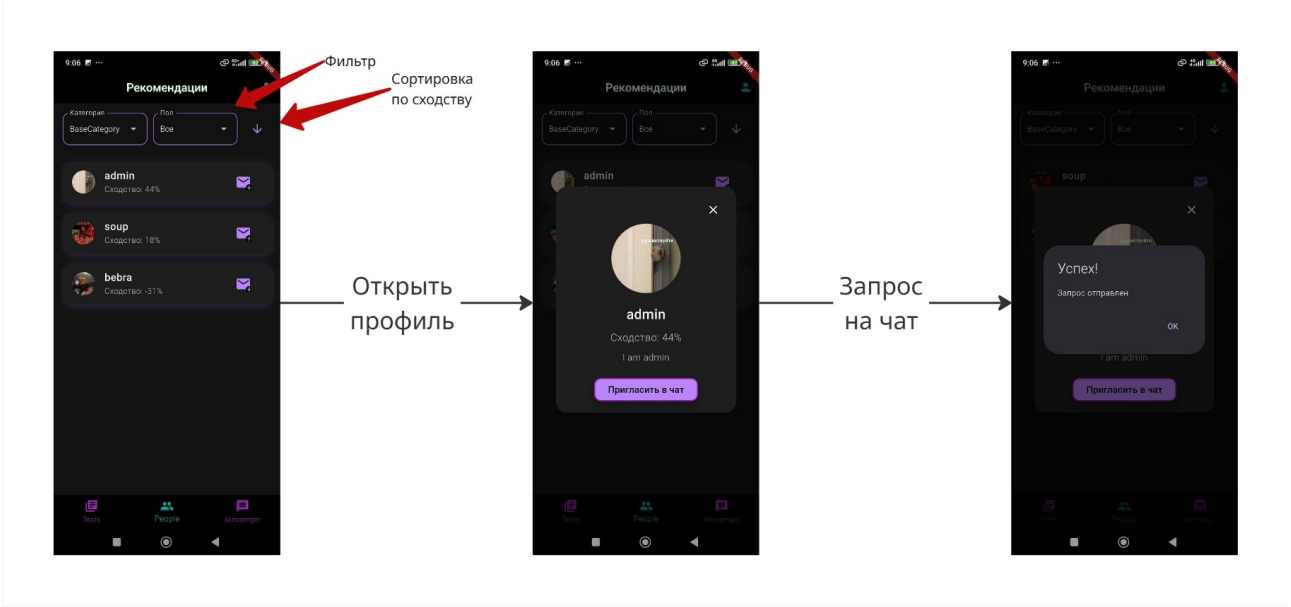

Рисунок 4 – Рекомендации, просмотр профиля и отправка запроса

Переход от рекомендаций к активному взаимодействию осуществляется через систему запросов на чат. С точки зрения архитектуры, чат требует наличия согласия двух сторон. Для этого реализован модуль управления заявками (Рисунок 5), в котором разделены входящие запросы и активные чаты.

Входящие запросы представлены в виде списка с возможностью принятия или отклонения. Принятые запросы автоматически перемещаются в раздел активных чатов. Таким образом, логика работы с чатами разделена на два уровня: согласование и непосредственное общение.

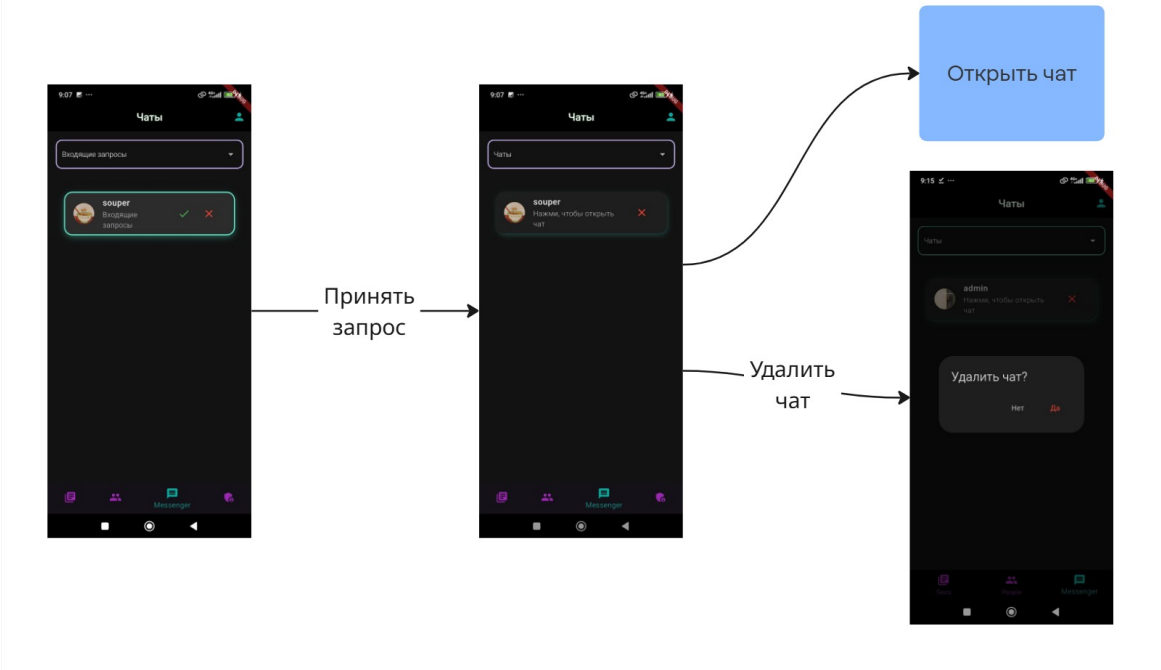

Рисунок 5 – Управление чатами и запросами на чат

Чат реализован как асинхронный поток сообщений. Было принято архитектурное решение о поддержке мультимедиа, что потребовало внедрения системы предварительного просмотра и управления прикреплёнными файлами. На рисунке 6 представлена логика интерфейса: отправка текста и вложений.

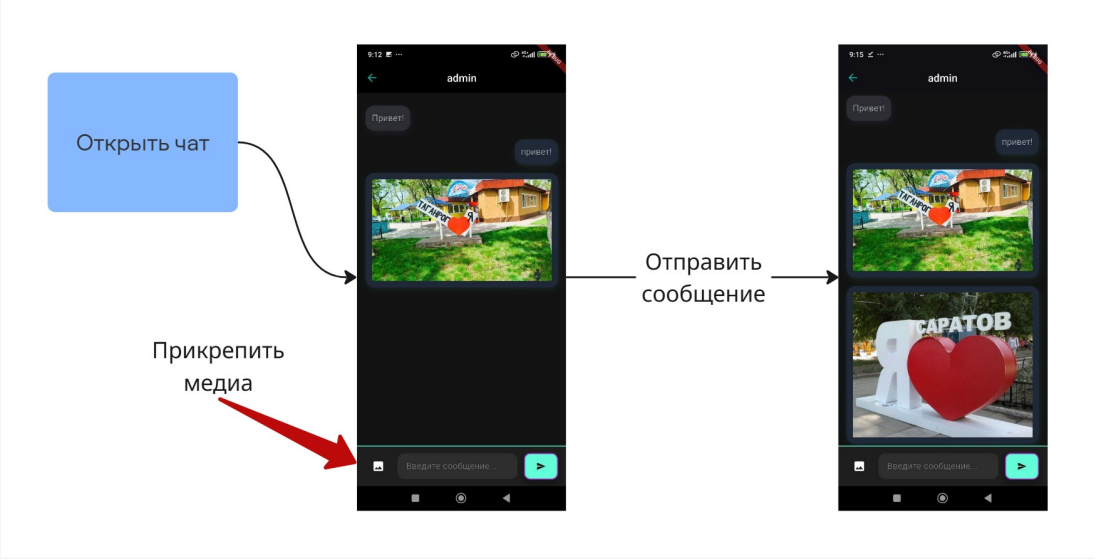

Рисунок 6 – Функционал чата: отправка сообщений и медиа

На этапе проектирования была предусмотрена возможность наполнения базы вопросов через интерфейс администратора. Как показано на рисунке 7, архитектура реализует проверку прав пользователя при запуске. При наличии административных прав активируется специальный экран, где можно вводить новые карточки. Таким образом, архитектура поддерживает разграничение ролей и безопасную изоляцию функционала.

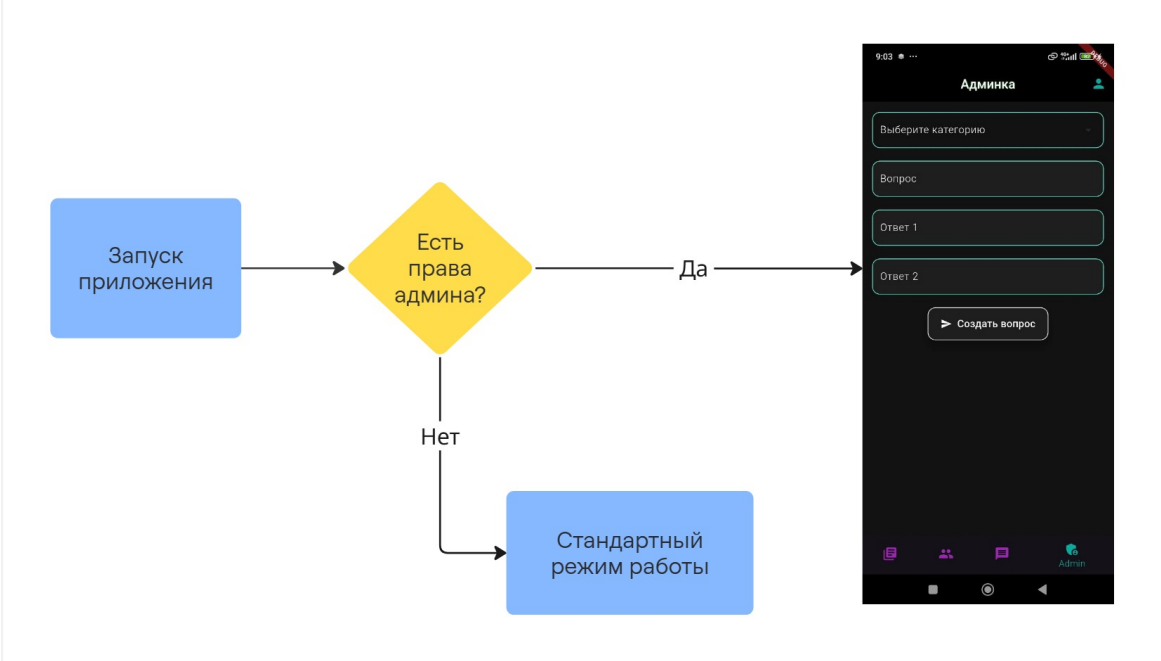

Рисунок 7 – Логика доступа к административной панели

Итоговая архитектура мобильного приложения представляет собой иерархическую структуру, основанную на пользовательских сценариях:

— экраны входа и регистрации — проверка подлинности;
— экран профиля — просмотр и редактирование данных;
— интерфейс вопросов — сбор данных для рекомендаций;
— рекомендации — выдача и фильтрация результатов;
— запросы на чат и чаты — коммуникация между пользователями;
— панель администратора — внутренняя поддержка наполнения базы.

Каждый из компонентов был спроектирован с учётом независимости, повторного использования и расширяемости. Такая архитектура обеспечивает устойчивость к росту функциональности и позволяет легко интегрировать рекомендательные алгоритмы на стороне сервера без необходимости внесения значительных изменений в клиентское приложение.

4.2 Проектирование архитектуры серверной части

Проектирование серверной части приложения для знакомств начинается с определения ключевых требований к системе. В первую очередь, необходимо обеспечить надёжное взаимодействие с мобильным клиентом, реализацию бизнес-логики приложения, управление пользовательскими данными, хранение медиафайлов, поддержку масштабируемой подсистемы рекомендаций, а также гибкость при развёртывании и сопровождении.

Для обмена данными между клиентским и серверным слоями требуется простой, расширяемый и широко поддерживаемый протокол. Учитывая эти факторы, архитектура взаимодействия строится на основе REST. Это решение обеспечивает стандартизованный обмен информацией через HTTP, совместимость с мобильными платформами и широкую поддержку в инструментах автоматизации.

При проектировании интерфейсов важно предусмотреть их документирование и проверку. Для этих целей подходит спецификация OpenAPI, поддерживающая автоматическую генерацию описания REST-эндпоинтов. Интеграция со Swagger позволяет предоставить наглядную, интерактивную документацию, упрощающую как разработку, так и тестирование API.

Серверная часть должна хранить структурированные пользовательские данные, такие как анкеты, результаты тестов, предпочтения, заявки на чат. Все эти данные обладают чёткой схемой и требуют транзакционной целостности.

Поэтому в качестве основной системы управления базами данных обоснован выбор реляционной СУБД. Среди доступных решений наиболее оптимальным по сочетанию стабильности, расширяемости и соответствию промышленным стандартам является PostgreSQL. Эта СУБД поддерживает богатый набор функций, включая работу с JSON, индексацию, расширения, и хорошо интегрируется с инструментами миграций.

Поддержка непрерывного развития и возможность безопасного обновления схемы базы данных требует внедрения системы управления миграциями. Это позволяет отслеживать изменения, выполнять откаты и обеспечивать согласованность структуры данных во всех окружениях. Исходя из популярности, удобства и хорошей интеграции со Spring Boot, в качестве системы миграций выбирается Liquibase.

Поскольку приложение позволяет пользователям загружать изображения профиля и отправлять вложения в чатах, необходимо предусмотреть механизм хранения бинарных данных. Хранение таких файлов непосредственно в базе данных было бы неэффективным и плохо масштабируемым. Вместо этого архитектура должна предусматривать использование объектного хранилища. Оно обеспечивает отдельное хранение медиафайлов с возможностью доступа через HTTP-интерфейс. В качестве подходящего решения выбирается MiniO — объектное хранилище с S3-совместимым API, которое легко разворачивается локально или в облачной среде и обеспечивает высокую отказоустойчивость.

Особое внимание в архитектуре уделяется подсистеме рекомендаций. Она требует специализированной обработки данных, может использовать иные языки программирования, библиотеки машинного обучения, и нуждается в независимом масштабировании. В этой связи было принято решение спроектировать рекомендательную систему как отдельный сервис. Это соответствует принципам микросервисной архитектуры и обеспечивает логическую и техническую изоляцию. Взаимодействие между основной серверной частью и рекомендательным сервисом происходит через HTTP-клиент, реализованный в модуле client.

Общая структура серверного приложения организована на основе модульного подхода с логическим разделением по уровням ответственности. Это упрощает поддержку кода, повышает читаемость и способствует соблюдению принципов SOLID. Каждый модуль отвечает за строго определённую функцию в рамках архитектуры.

Общая структура серверного приложения организована на основе модульного подхода, обеспечивающего логическое разделение по уровням ответственности. Это облегчает поддержку, расширение и тестирование системы.

- controller — REST-контроллеры, принимающие HTTP-запросы, валидирующие входные данные и передающие их в бизнес-логику;
- service — реализация бизнес-правил, координация взаимодействия между слоями и внешними сервисами;
- repository — интерфейсы доступа к базе данных на основе Spring Data JPA;
- entity — JPA-сущности, отражающие структуру таблиц и связи между ними;
- dto — модели передачи данных между слоями и при взаимодействии с клиентом;
- client — модуль HTTP-взаимодействия с внешней рекомендательной системой;
- config — конфигурационные классы, включая настройки безопасности, CORS, Swagger и подключение к внешним сервисам;
- exception — иерархия пользовательских исключений и глобальная обработка ошибок;
- mapper — преобразование между сущностями и DTO;
- utils — вспомогательные методы и классы общего назначения.

Такое разделение позволяет обеспечить читаемость, переиспользуемость и модульность, а также упрощает интеграцию дополнительных компонентов в архитектуру.

Подобная структура способствует высокой модульности, улучшает читаемость проекта и упрощает тестирование. Чёткое разграничение ответственности между слоями позволяет удобно масштабировать приложение, добавлять новые функции и интегрировать внешние сервисы без нарушения архитектурной целостности.

Для обеспечения воспроизводимости окружения и удобства развертывания на различных машинах проектируется контейнеризированная инфраструктура. Это особенно важно при наличии нескольких сервисов (основного API, рекомендательной системы, базы данных, хранилища медиафайлов). В этих условиях использование Docker становится естественным выбором. С его помо-

щью можно описать каждый компонент как независимый контейнер, настроить их взаимодействие в рамках одной сети и обеспечить надёжное воспроизводимое развёртывание, как в процессе локальной разработки, так и в продуктивной среде.

Таким образом, архитектура серверной части проектируется на основе требований надёжности, масштабируемости и гибкости. В ходе анализа выбираются технологии и подходы, соответствующие поставленным задачам. Итоговая архитектура включает REST-интерфейс, документацию через OpenAPI, использование PostgreSQL с миграциями Liquibase, объектное хранилище MiniO для медиафайлов, отдельный микросервис рекомендаций и контейнеризацию с помощью Docker. Эти решения образуют устойчивую и расширяемую платформу, готовую к дальнейшему развитию.

5 Реализация приложения для знакомств

5.1 Реализация кроссплатформенной клиентской части

Клиентская часть мобильного приложения для знакомств была разработана с использованием фреймворка Flutter и языка программирования Dart. Выбор данной технологии обусловлен ее способностью обеспечивать кроссплатформенную разработку с единой кодовой базой для операционных систем iOS и Android, а также высокой производительностью и гибкостью в создании пользовательских интерфейсов. Основной задачей при разработке клиентской части являлось создание интуитивно понятного, отзывчивого и функционального интерфейса, обеспечивающего комфортное взаимодействие пользователя с системой. Архитектура приложения ориентирована на модульность и переиспользование компонентов, что упрощает поддержку и дальнейшее развитие проекта.

Точкой входа в приложение является файл main.dart, который инициализирует корневой виджет и определяет глобальные настройки, такие как тема оформления и система навигации. Этот процесс инициализации и определения маршрутов можно увидеть в следующем фрагменте кода. В приложении реализована единая темная тема для консистентного визуального восприятия на всех экранах. Маршрутизация между экранами осуществляется с помощью именованных маршрутов, что позволяет структурировать навигацию и упрощает переходы.

```
// Фрагмент main.dart
class MyApp extends StatelessWidget {
    const MyApp({super.key});
    @override
    Widget build(BuildContext context) {
        return MaterialApp(
        title: 'Bim Bim App',
        theme: ThemeData(
        brightness: Brightness.dark,
        scaffoldBackgroundColor: //...
        ),
        initialRoute: '/login',
        routes: {
            '/login': (context) => const LoginScreen(),
            '/register': (context) => const RegisterScreen(),
            '/home': (context) => const MainScreen(),
            '/messenger': (context) => const MessengerPage(),
```

```
            '/editProfile': (context) => const EditProfilePage(),
        },
        );
    }
}
```

Взаимодействие с серверной частью осуществляется через специально разработанный сервис ApiClient, расположенный в services/api_client.dart. Этот класс инкапсулирует логику отправки HTTP-запросов (GET, POST, PUT, DELETE), а также загрузки файлов на сервер. Базовый URL для всех API-запросов вынесен в файл констант constants/constants.dart.

Ключевым аспектом безопасности при взаимодействии с API является использование JWT-токенов. После успешной аутентификации пользователя сервер возвращает JWT-токен, который сохраняется в локальном хранилище устройства с помощью пакета shared_preferences. ApiClient автоматически извлекает этот токен и добавляет его в заголовки всех последующих защищенных запросов в виде «Authorization: Bearer <token> ». Пример реализации этого механизма представлен ниже.

```
// Фрагмент services/api_client.dart
Future<Map<String, String>> _getHeaders({Map<String, String>? extraHeaders}) async {
    final prefs = await SharedPreferences.getInstance();
    final token = prefs.getString('jwt_token');
    final headers = <String, String>{
        'Content-Type': 'application/json',
        if (token != null) 'Authorization': 'Bearer ' + token,
        ...?extraHeaders,
    };
    return headers;
}

Future<http.Response> post(String endpoint,
    {Object? body, Map<String, String>? headers}) async {
    final fullHeaders = await _getHeaders(extraHeaders: headers);
    final url = Uri.parse(endpoint);
    return http.post(url, headers: fullHeaders, body: body);
}
```

Модуль аутентификации включает экраны входа (screens/login_screen.dart) и регистрации (screens/register_screen.dart). Экран входа позволяет пользовате-

лю ввести свои учетные данные, которые затем отправляются на сервер через ApiClient на эндпоинт /auth/login. В случае успеха, полученный JWT-токен сохраняется, и пользователь перенаправляется на главный экран приложения. Логика этого процесса показана в следующем фрагменте кода.

```
// Фрагмент screens/login_screen.dart
Future<void> _login() async {
    final String username = _usernameController.text;
    final String password = _passwordController.text;
    // ...
    try {
        final response = await _apiClient.post(
        '$baseUrl/auth/login',
        body: jsonEncode({'username': username, 'password': password})
        );

        if (response.statusCode == 200) {
            final Map<String, dynamic> responseData = jsonDecode(response.body);
            if (responseData.containsKey('token')) {
                final String token = responseData['token'];
                final prefs = await SharedPreferences.getInstance();
                await prefs.setString('jwt_token', token);
                Navigator.pushReplacementNamed(context, '/home');
            } // ...
        } // ...
    } catch (e) { /* ... */ }
}
```

Главный экран приложения (screens/main_screen.dart) служит центральным узлом навигации после аутентификации. Он содержит BottomNavigationBar для переключения между основными разделами: «Вопросы», «Рекомендации», «Чаты» и «Админка» (для администраторов).

Управление профилем пользователя реализовано на странице screens/pages/-profile_page.dart. Здесь отображается информация о пользователе, предоставляется возможность загрузки нового аватара и перехода к редактированию профиля. Загрузка аватара, механизм которой приведен ниже, осуществляется методом uploadMultipart класса ApiClient.

```
// Фрагмент screens/pages/profile_page.dart
Future<void> _uploadAvatarToBackend(File imageFile) async {
```

```
    try {
        final file = await http.MultipartFile.fromPath('image', imageFile.path);
        final response = await _apiClient.uploadMultipart(
          endpoint: '$baseUrl/user/updateAvatar',
          files: [file],
          fields: {'type': 'image'}
        );
        if (response.statusCode == 200) {
            _fetchUserData();
        } // ...
    } catch (e) { /* ... */ }
}
```

Важной частью функционала является система подбора партнеров, основанная на ответах пользователей на вопросы в разделе «Тесты» (screens/pages/-tests_page.dart). Пользователи отвечают на вопросы, свайпая карточки. Использование виджета SwipeCards для этой цели демонстрируется в коде ниже. Эти ответы отправляются на сервер и используются для формирования профиля интересов.

```
// Фрагмент screens/pages/tests_page.dart
void _initializeSwipeItems() {
    if (_questions.isNotEmpty) {
        _swipeItems = _questions.map((question) {
            QuestionItem questionItem = QuestionItem(
              id: question['id'].toString(),
              content: question['content'],
              answerLeft: question['answerLeft'],
              answerRight: question['answerRight'],
            );
            return SwipeItem(
              content: questionItem,
              likeAction: () => _onAnswer(questionItem.id, 1),
              nopeAction: () => _onAnswer(questionItem.id, -1),
              superlikeAction: () => _onAnswer(questionItem.id, 0),
            );
        }).toList();
        setState(() {
            _matchEngine = MatchEngine(swipeItems: _swipeItems);
        });
    }
```

На странице «Рекомендации» (screens/pages/people_page.dart) отображаются профили других пользователей с указанием процента «сходства». Загрузка таких рекомендаций и их последующее отображение в виде списка проиллюстрированы в следующем фрагменте.

```
// Фрагмент screens/pages/people_page.dart
Future<void> _fetchPeople(int? categoryId) async {
    // ...
    final response = await _apiClient.get('$baseUrl/matching/$categoryId');
    if (response.statusCode == 200) {
        final data = json.decode(response.body) as List<dynamic>;
        setState(() {
            _people = data.map((e) => e as Map<String, dynamic>).toList();
            // ...
        });
    }
    // ...
}
ListView.builder(
    itemCount: sortedPeople.length,
    itemBuilder: (context, index) {
        final person = sortedPeople[index];
        return ListTile(
          leading: CircleAvatar(backgroundImage: NetworkImage(person['avatar'])),
          title: Text(person['username']),
          subtitle: Text('Сходство: ${person['similarity']}%'),
           // ...
        );
    },
)
```

Модуль обмена сообщениями состоит из списка чатов (screens/pages/-messenger_page.dart) и экрана самого чата (screens/pages/chat_page.dart). Экран чата отображает историю переписки и позволяет отправлять текстовые сообщения и изображения. Для обработки реального времени и получения новых сообщений используется периодический опрос сервера (Timer.periodic). Реализация этого механизма обновления показана ниже.

```
// Фрагмент screens/pages/chat_page.dart
@override
void initState() {
```

```
        super.initState();
        _loadMessages();
        _updateTimer = Timer.periodic(const Duration(seconds: 3), (timer) {
            if (mounted) {
                _loadMessages();
            }
        });
}
Future<void> _loadMessages() async {
        // ...
        final response = await _apiClient.get('$baseUrl/chat/${widget.chatId}/messages');
        // ...
}
```

Таким образом, клиентская часть мобильного приложения для знакомств реализована с использованием современных подходов и инструментов фреймворка Flutter. Особое внимание уделено структурированию кода, обеспечению безопасности передачи данных через JWT-токены и созданию удобного пользовательского интерфейса для всех ключевых функций приложения.

5.2 Реализация масштабируемой серверной части

Серверная часть мобильного приложения для знакомств разработана с использованием фреймворка Spring Boot, который был выбран благодаря его способности обеспечивать быструю разработку, встроенной поддержке множества технологий и обширному сообществу. Spring Boot упрощает создание автономных производственных приложений на основе Spring, минимизируя необходимость в сложной конфигурации.

Архитектура серверной части следует классической многоуровневой модели, включающей уровень контроллеров (Controller), сервисов (Service) и репозиториев (Repository), что обеспечивает четкое разделение ответственности и повышает тестируемость и поддерживаемость кода.

Контроллеры отвечают за обработку входящих HTTP-запросов, их валидацию и передачу данных на уровень сервисов. Они определяют API эндпоинты приложения. Например, контроллер AuthController обрабатывает запросы, связанные с аутентификацией и регистрацией пользователей, как показано в следующем фрагменте кода, где определяется эндпоинт для входа пользователя.

```
@RestController
```

```
@RequiredArgsConstructor
@RequestMapping("/api/auth")
public class AuthController {
    private final UserServiceImpl userService;
    // ... другие методы ...
    @PostMapping("/login")
    public JwtDto loginUser(@RequestBody UserLoginRequest userLoginRequest) {
        return userService.loginUser(userLoginRequest);
    }
}
```

Уровень сервисов инкапсулирует основную бизнес-логику приложения. Сервисы координируют взаимодействие между контроллерами и репозиториями, выполняют операции над данными и реализуют специфические для домена правила. Пример реализации метода loginUser в классе UserServiceImpl, который проверяет учетные данные пользователя и генерирует JWT-токен, представлен ниже.

```
@Service
@RequiredArgsConstructor
public class UserServiceImpl implements UserService {
    private final UserRepository userRepository;
    private final PasswordEncoder passwordEncoder;
    private final JwtUtils jwtUtils;
    // ... другие зависимости ...
    @Override
    public JwtDto loginUser(UserLoginRequest userLoginRequest) {
        Optional<User> user = userRepository
            .findByUsername(userLoginRequest.username());
        if (user.isEmpty()) {
            throw new UnauthorizedException("Username not found");
        }
        if (!passwordEncoder.matches(userLoginRequest.password(),
            user.get().getPassword())) {
            throw new UnauthorizedException("Wrong password");
        }
        return jwtUtils.generateToken(user.get().getUsername(),
            user.get().getId(), user.get().getRoles());
    }
    // ... другие методы ...
}
```

Для взаимодействия с базой данных используется Spring Data JPA, который значительно упрощает создание уровня доступа к данным. Интерфейсы репозиториев, такие как UserRepository, наследуются от JpaRepository, что автоматически предоставляет стандартные CRUD-операции и возможность определения кастомных запросов.

```
@Repository
public interface UserRepository extends JpaRepository<User, Long> {
    Optional<User> findByUsername(String username);
}
```

Модель данных представлена JPA-сущностями (Entities), такими как User, Category, Question, Chat, Message. Эти классы аннотированы для маппинга на таблицы реляционной базы данных. Сущность User, например, содержит информацию о пользователе, его учетные данные и связи с другими сущностями.

```
@Getter
@Setter
@Entity
@Table(name="users")
public class User extends AbstractEntity {
    @Column(nullable = false, unique = true)
    private String username;
    @Column(nullable = false)
    private String password;
    // ... другие поля и связи ...
}
```

Для обмена данными между клиентом и сервером, а также между различными слоями приложения, активно используются объекты передачи данных (DTO). Они представляют собой простые Java-классы (часто реализуемые как records), которые определяют структуру данных для запросов и ответов. Пример DTO для запроса на вход пользователя UserLoginRequest:

```
public record UserLoginRequest(String username, String password) {
}
```

Безопасность приложения обеспечивается с помощью Spring Security и механизма аутентификации на основе JSON Web Tokens (JWT). JWT-токены используются для аутентификации пользователей после успешного входа в систему, позволяя им получать доступ к защищенным ресурсам. Центральным

элементом системы аутентификации является фильтр JwtAuthenticationFilter. При каждом запросе он извлекает JWT-токен из заголовка Authorization. Если токен присутствует и валиден, из него извлекаются данные пользователя (имя, идентификатор, роли), на основе которых создается объект аутентификации и помещается в SecurityContextHolder, делая пользователя доступным для последующих компонентов системы. Логика работы фильтра показана в следующем фрагменте кода.

```
@Override
protected void doFilterInternal(HttpServletRequest request,
HttpServletResponse response,
        FilterChain filterChain) throws ServletException, IOException {
        String authHeader = request.getHeader(HttpHeaders.AUTHORIZATION);
        if (authHeader != null && authHeader.startsWith("Bearer ")) {
                JwtDto jwtDto = // ... извлечение токена;
                if (jwtUtils.validateToken(jwtDto)) {
                        // ... извлечение необходимых параметров из токена
                        UsernamePasswordAuthenticationToken authentication =
                                new UsernamePasswordAuthenticationToken(
                                        userDetails, null, authorities
                                );
                        SecurityContextHolder.getContext().setAuthentication(authentication);
                }
        }
        filterChain.doFilter(request, response);
}
```

Генерация и валидация токенов осуществляется классом JwtUtils, который использует секретный ключ и заданное время жизни токена, настраиваемые в конфигурационном файле application.yaml. Пример метода генерации токена:

```
public class JwtUtils {
        private String secretKey;
        private Long expirationTime;
        public JwtDto generateToken(String username, Long id, String roles) {
                return new JwtDto(Jwts.builder()
                        .setSubject(username)
                        .claim("id", id)
                        .claim("roles", roles)
                        .setIssuedAt(new Date())
                        .setExpiration(new Date(System.currentTimeMillis() + expirationTime))
```

```
                .signWith(SignatureAlgorithm.HS512, secretKey)
                .compact());
    }
    // ... другие методы валидации и извлечения данных ...
}
```

Конфигурация безопасности Spring Security определяется в классе SecurityConfig. В фрагменте кода ниже показано, как настраиваются правила доступа к различным эндпоинтам, отключаются стандартные механизмы (например, CSRF, form login), и JwtAuthenticationFilter встраивается в цепочку фильтров.

```
@Bean
public SecurityFilterChain securityFilterChain(
    JwtAuthenticationFilter jwtAuthenticationFilter,
    HttpSecurity http) throws Exception {
    http
        .csrf(AbstractHttpConfigurer::disable)
        .formLogin(AbstractHttpConfigurer::disable)
        .sessionManagement(config -> config.sessionCreationPolicy(
            SessionCreationPolicy.STATELESS))
        .authorizeHttpRequests(auth -> auth
            .requestMatchers("/api/auth/**").permitAll()
            .requestMatchers("/api-docs/**").permitAll()
            // ... другие разрешенные пути ...
            .anyRequest().authenticated()
        )
        .addFilterBefore(jwtAuthenticationFilter,
            UsernamePasswordAuthenticationFilter.class);
    return http.build();
}
```

Разграничение доступа на уровне методов контроллеров или сервисов осуществляется с использованием аннотаций, таких как @PreAuthorize("hasRole('ADMIN')"), что позволяет гибко управлять правами доступа к отдельным операциям.

Для хранения пользовательских изображений (аватары, изображения в чатах) используется интеграция с S3-совместимым хранилищем MinIO. Сервис ImageServiceImpl отвечает за загрузку файлов в MinIO, генерацию уникальных имен файлов и предоставление URL для доступа к ним. Загрузка происходит через MinioClient, настроенный в MinioConfig.

Для реализации функции подбора подходящих пользователей (мэтчинга) серверная часть приложения интегрируется с внешним специализированным сервисом рекомендаций. Взаимодействие с этим сервисом осуществляется посредством HTTP-клиента, реализованного с использованием декларативного подхода Spring. Интерфейс MatchingClient определяет контракт взаимодействия, используя аннотацию @PostExchange для указания эндпоинта и типа запроса. Пример объявления клиента представлен ниже.

```
public interface MatchingClient {
    @PostExchange("/api/matching")
    List<MatchingResponse> getMatching(
    @RequestBody MatchingRequest request);
}
```

Конфигурация данного клиента, включая базовый URL внешнего сервиса, вынесена в отдельный класс свойств MatchingClientProperties, значения для которого загружаются из основного конфигурационного файла приложения application.yaml. Это обеспечивает гибкость настройки адреса сервиса рекомендаций без необходимости изменения кода.

```
@Data
@Configuration
@ConfigurationProperties(prefix = "rest.client.matching",
ignoreUnknownFields = false)
public class MatchingClientProperties {
    private String baseUrl;
}
```

Создание и настройка экземпляра MatchingClient происходит в конфигурационном классе RestClientConfig. В коде ниже показано, как используется RestClient.Builder для базовой настройки HTTP-клиента (например, указания baseUrl из MatchingClientProperties) и HttpServiceProxyFactory для создания прокси-объекта, реализующего интерфейс MatchingClient. Такой подход позволяет абстрагироваться от низкоуровневых деталей выполнения HTTP-запросов, сосредотачиваясь на бизнес-логике.

```
@Bean
public MatchingClient matchingClient(RestClient.Builder builder,
MatchingClientProperties properties) {
```

```
        RestClient restClient = builder
                .baseUrl(properties.getBaseUrl())
                .build();
        RestClientAdapter adapter = RestClientAdapter
                .create(restClient);
        HttpServiceProxyFactory factory = HttpServiceProxyFactory
                .builderFor(adapter)
                .build();
        return factory.createClient(MatchingClient.class);
}
```

При необходимости получения рекомендаций, соответствующий сервис серверной части (MatchingServiceImpl) подготавливает объект MatchingRequest, содержащий данные о текущем пользователе, других пользователях и их ответах на вопросы. Этот объект передается методу getMatching клиента MatchingClient, который выполняет HTTP-запрос к внешнему сервису и возвращает список подходящих кандидатов в виде объектов MatchingResponse.

Управление схемой базы данных и ее миграциями осуществляется с помощью Liquibase. Изменения схемы описываются в XML или YAML файлах, что обеспечивает версионирование и контролируемое обновление структуры БД. Пример изменения схемы для создания таблицы категорий:

```
databaseChangeLog:
  - changeSet:
        id: add_base_category_and_questions
          author: soup
          changes:
                - insert:
                  tableName: category
                  columns:
                        - column:
                          name: name
                          value: 'BaseCategory'
                        - column:
                          name: question_count
                          valueNumeric: 0
```

Все основные конфигурационные параметры приложения, такие как настройки подключения к БД, параметры JWT, адреса внешних сервисов и MinIO,

вынесены в файл application.yaml, что позволяет гибко настраивать приложение для различных окружений.

Таким образом, серверная часть приложения представляет собой структурированную систему, использующую современные подходы и технологии для обеспечения функциональности, безопасности и масштабируемости мобильного приложения для знакомств.

6 Разработка рекомендательной системы приложения для знакомств

6.1 Анализ предметной области и выбор данных для исследования

Разработка эффективной рекомендательной системы является ключевым аспектом современных мобильных приложений для знакомств. Целью данной работы является создание такой системы, основной особенностью которой является использование опросов с тернарными ответами (да, нет, пропустить) для формирования профилей пользователей и последующего подбора потенциальных партнеров. Данный подход предоставляет пользователям простой и интуитивно понятный способ выражения своих предпочтений и интересов, одновременно позволяя системе собирать структурированные данные для анализа.

Для моделирования и первичного тестирования предлагаемой рекомендательной системы был выбран публично доступный набор данных «Speed Dating Experiment» [32]. Этот датасет был собран профессорами Колумбийской бизнес-школы Рэйем Фисманом и Шиной Айенгар в ходе экспериментальных мероприятий по быстрым знакомствам, проводившихся с 2002 по 2004 год. Выбор данного набора данных обусловлен его высокой релевантностью поставленной задаче. Во-первых, он содержит информацию о демографических характеристиках участников, их интересах, самооценках по ключевым атрибутам (привлекательность, искренность, интеллект, веселье, амбициозность, общность интересов) и предпочтениях в потенциальном партнере. Во-вторых, что наиболее важно, датасет включает реальные исходы четырехминутных «первых свиданий» – решение участников о желании продолжить общение (переменные dec – решение участника, match – обоюдное согласие). Наличие как анкетных данных, так и результатов реальных взаимодействий позволяет оценить, насколько предпочтения, выраженные в опросах, коррелируют с фактическим выбором.

Ключевой особенностью датасета, имеющей значение для данного исследования, является его структура по «волнам» (переменная wave). Каждая «волна» представляет собой отдельное мероприятие по быстрым знакомствам, и участники взаимодействовали только с ограниченным подмножеством партнеров противоположного пола внутри своей волны. Это означает, что данные о взаимодействии каждой возможной пары пользователей в рамках всего датасета отсутствуют. Число участников в волнах варьировалось, что также вносит определенную разнородность в данные. Данная особенность структуры накладывает

ограничения на интерпретацию метрик качества рекомендаций, поскольку система может предложить объективно подходящего партнера, но если они не пересекались в рамках одной «волны», это взаимодействие не будет зафиксировано как совпадение, что потенциально занижает показатели точности и полноты.

Для адаптации данных к тернарному формату, используемому в разрабатываемом приложении, были выбраны признаки, отражающие интересы участников (например, sports, tvsports, exercise, dining и т.д.) и их самооценки/предпочтения по шести ключевым атрибутам (например, attr1_1 – важность привлекательности для себя, attr3_1 – самооценка привлекательности). Эти признаки, представленные в датасете оценками по 10-балльной шкале, были преобразованы в тернарный формат: значения от 1 до 3 интерпретировались как «нет» (-1), от 4 до 6 – как «пропустить» (0), а от 7 до 10 – как «да» (1). Такой подход позволяет симулировать механизм сбора предпочтений, предполагаемый в мобильном приложении.

Несмотря на упомянутые ограничения, датасет «Speed Dating Experiment» предоставляет ценную основу для первоначального исследования и обоснования базовых механизмов предлагаемой рекомендательной системы. Он позволяет проверить гипотезу о том, что сходство пользователей, выраженное через их ответы на тернарные опросы, может служить основой для формирования релевантных рекомендаций.

6.2 Разработка и сравнительный анализ моделей рекомендаций

Основной задачей данного этапа исследования являлась оценка эффективности построения рекомендаций, базирующихся на сходстве профилей пользователей, сформированных на основе их ответов на тернарные опросы. Для этого был проведен сравнительный анализ нескольких подходов к получению векторных представлений пользователей и последующему расчету их схожести.

Первоначальным шагом являлась загрузка и предварительная обработка данных из датасета «Speed Dating Experiment» [32]. Данные были загружены из CSV-файла с использованием библиотеки pandas. Одной из задач предварительной обработки было восстановление идентификаторов партнеров (pid) для некоторых записей, где они отсутствовали, на основе сопоставления номера волны (wave) и идентификатора партнера внутри волны (partner). Этот процесс показан в следующем фрагменте кода:

```
import pandas as pd

# ... загрузка df из CSV ...
df = pd.read_csv(csv_path, encoding='latin1')

id_lookup = df[['wave', 'id', 'iid']].drop_duplicates().set_index(['wave', 'id'])['iid']

df['pid'] = df.apply(lambda row: id_lookup.loc[(row['wave'], row['partner'])]
    if pd.isna(row['pid']) else row['pid'], axis=1)
```

Методология исследования включала несколько ключевых шагов. Во-первых, данные из датасета, касающиеся интересов участников и их оценок различных атрибутов, были преобразованы в тернарный формат. Как упоминалось ранее, значения от 1 до 3 были отображены в -1 («нет»), от 4 до 6 – в 0 («пропустить»), а от 7 до 10 – в 1 («да»). Пример функции для такого преобразования и ее применение к выбранному набору признаков (ternary_features, включающему интересы и самооценки атрибутов) приведен ниже:

```
def ternarize(value):
    if pd.isna(value) or (4 <= value <= 6):
        return 0
    elif value >= 7:
        return 1
    elif value <= 3:
        return -1
    return 0

    ternary_features = ["sports", "tvsports", ..., "amb3_1"]
    df_ternary = df["iid"] + ternary_features].copy()

    for feature in ternary_features:
    df_ternary[feature] = df_ternary[feature].apply(ternarize)

    df_ternary = df_ternary.drop_duplicates(subset=['iid'], keep='first')
```

Таким образом, для каждого уникального пользователя (идентифицируемого по iid) был сформирован вектор тернарных ответов. Далее, для уменьшения размерности и извлечения скрытых признаков из этих векторов, были применены и сравнены различные методы: сингулярное разложение (SVD), метод главных компонент (PCA), нелинейное снижение размерности с помощью

UMAP, а также автоэнкодер (Autoencoder) и вариационный автоэнкодер (VAE). Сходство между пользователями противоположного пола затем рассчитывалось с использованием косинусного расстояния между их полученными эмбеддингами. Функция для построения рекомендательной системы на основе PCA, например, имела следующую структуру:

```
def build_pca_recommender(df_ternary: pd.DataFrame, df_profiles: pd.DataFrame,
        n_components: int = 10):
        # df_profiles содержит уникальные iid и соответствующий пол (gender)
        features = df_ternary.drop(columns=['iid'])
        pca = PCA(n_components=n_components, random_state=42)
        reduced = pca.fit_transform(features)
        df_reduced = pd.DataFrame(reduced, index=df_ternary['iid'])

        genders = df_profiles.drop_duplicates('iid').set_index('iid')['gender']
        df_reduced['gender'] = genders

        def recommend(user_id: int, k: int = 10):
                if user_id not in df_reduced.index:
                        return []
                user_row = df_reduced.loc[user_id]
                user_vec = user_row.drop('gender').values.reshape(1, -1)
                user_gender = user_row['gender']
                opposite_gender = 1 if user_gender == 0 else 0

                candidates = df_reduced[df_reduced['gender'] ==
                        opposite_gender].drop(columns='gender')
                candidate_ids = candidates.index
                candidate_vectors = candidates.values

                similarities = cosine_similarity(user_vec, candidate_vectors)[0]
                top_indices = np.argsort(similarities)[::-1][:k]
                top_user_ids = candidate_ids[top_indices]
                return list(top_user_ids)
        return recommend
```

Для оценки качества полученных рекомендаций использовались стандартные метрики: точность на K позиции (Precision@K), полнота на K позиции (Recall@K), коэффициент попадания (HitRate@K) и покрытие (Coverage). В качестве данных о взаимодействиях (df_interactions) использовались записи из

исходного датасета, где было зафиксировано решение пользователя (dec) и фактическое совпадение (match).

Результаты сравнительного анализа моделей, использующих исключительно тернарные данные, показали, что более простые линейные методы, такие как PCA и SVD, а также нелинейный UMAP, продемонстрировали сопоставимую и в некоторых случаях лучшую производительность по сравнению с нейросетевыми подходами (Autoencoder и VAE) на данном конкретном датасете и объеме данных. Например, для K=20, значения Precision@K для PCA, SVD и UMAP находились в диапазоне 0.058-0.060, Recall@K – 0.073-0.077, в то время как для Autoencoder и VAE эти показатели были несколько ниже (P@20 0.056, R@20 0.071-0.072). Снижение производительности нейросетевых моделей может быть обусловлено относительно небольшим размером датасета (551 уникальный пользователь для обучения эмбеддингов) и разреженностью тернарных векторов, что затрудняет обучение сложных нелинейных зависимостей без переобучения. График зависимости метрик от значения K для модели PCA, показавшей одни из лучших результатов среди рассмотренных, приведен на рисунке 8.

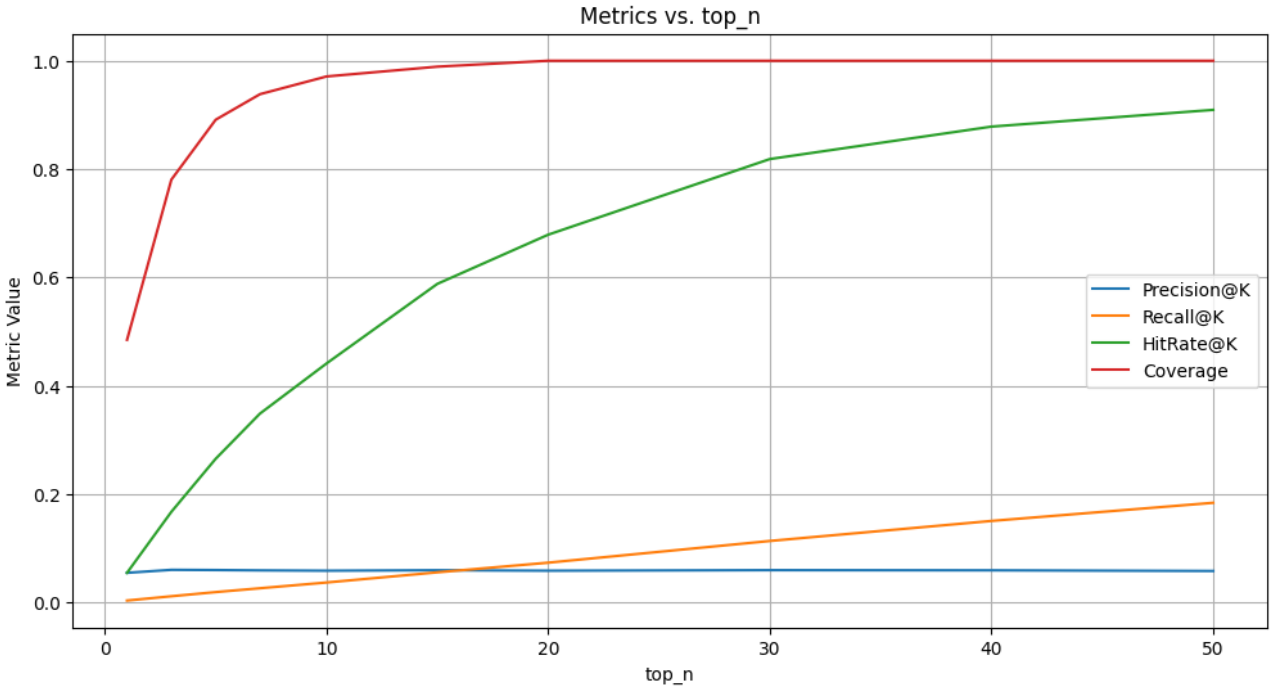

Рисунок 8 – Зависимость метрик Precision@K, Recall@K, HitRate@K и Coverage от числа рекомендаций K для модели PCA

При интерпретации этих значений важно учитывать упомянутую ранее «волновую» структуру датасета. Поскольку участники взаимодействовали лишь

с ограниченным числом партнеров в рамках своей волны, многие потенциально релевантные рекомендации не могли быть подтверждены как фактические совпадения, что искусственно занижает показатели Precision@K и Recall@K. В этом контексте более показательным является HitRate@K, который для K=20 составил порядка 0.67-0.68 для PCA, SVD и UMAP. Это означает, что примерно для 67-68% пользователей система смогла найти хотя бы одно реальное совпадение (из тех, кто действительно встретился на мероприятии) среди топ-20 предложенных кандидатов. Показатель Coverage для всех моделей был близок к единице.

Полученные умеренные значения метрик на одних лишь тернарных данных, а также тот факт, что более сложные нейросетевые модели не продемонстрировали явного преимущества на данном этапе, указывают на то, что, хотя опросы и несут полезную информацию, их одних может быть недостаточно для формирования высокоточных рекомендаций. Это создает предпосылки для разработки комбинированной (гибридной) рекомендательной системы.

В рамках данной дипломной работы предлагается расширить базовую систему, основанную на тернарных векторах, путем включения анализа текстовых описаний профилей пользователей. Предполагается, что текстовые описания могут содержать нюансы и детали, которые сложно выразить через ограниченный набор тернарных ответов. Для этого текстовые описания преобразуются в векторные представления. Итоговое сходство между пользователями рассчитывается как взвешенная сумма сходства их тернарных профилей и сходства их текстовых эмбеддингов: $Sim(u, v) = \alpha \cdot sim_{ternary}(u, v) + (1 - \alpha) \cdot sim_{text}(u, v)$, где $sim_{ternary}(u, v)$ – косинусное сходство тернарных векторов пользователей u и v, $sim_{text}(u, v)$ – косинусное сходство их текстовых эмбеддингов, а α – весовой коэффициент. Такой гибридный подход позволяет обогатить информацию о пользователях и потенциально повысить качество и релевантность предлагаемых кандидатур.

6.3 Реализация микросервиса рекомендательной системы

Для практической реализации предложенной гибридной рекомендательной системы был разработан прототип в виде микросервиса на языке Python. Выбор Python обусловлен его богатой экосистемой библиотек для обработки данных, машинного обучения и веб-разработки, а также простотой и скоростью разработки. В качестве основного фреймворка для создания API был выбран

FastAPI, известный своей высокой производительностью, асинхронной природой и удобной системой валидации данных на основе Pydantic. Для запуска приложения используется ASGI-сервер Uvicorn.

Архитектура сервиса спроектирована с учетом ключевой особенности – динамичности системы, то есть способности работать с произвольным набором вопросов для формирования тернарных профилей. Сервис не привязан к заранее определенной структуре опросов и вычисляет сходство на лету на основе данных, передаваемых в запросе.

Для определения структуры входящих запросов и исходящих ответов используются модели Pydantic, описанные в файле models.py. Модель MatchingRequest инкапсулирует данные о текущем пользователе, для которого запрашиваются рекомендации, списке всех доступных пользователей для подбора и перечне вопросов (QuestionMatchingRequest) с их идентификаторами и содержанием. Модель UserMatchingRequest содержит информацию о конкретном пользователе, включая его идентификатор, текстовое описание и словарь его тернарных ответов на вопросы. На выходе сервис возвращает список объектов MatchingResponse, каждый из которых содержит профильную информацию рекомендованного пользователя и вычисленный коэффициент сходства. Структура модели MatchingResponse представлена ниже:

```
from pydantic import BaseModel
from typing import Optional, List

class MatchingResponse(BaseModel):
    id: int
    avatar: Optional[str] = None
    gender: str
    username: str
    description: str
    similarity: float
```

Центральным элементом системы является класс DynamicRecommendationSystem, реализованный в файле recommender.py. При инициализации этого класса загружается предобученная модель «all-MiniLM-L6-v2» из библиотеки Sentence Transformers для преобразования текстовых описаний профилей в векторные представления (эмбеддинги). Выбор модели «all-MiniLM-L6-v2» обусловлен ее оптимальным балансом между качеством получаемых эмбеддингов и вы-

числительной эффективностью. Данная модель хорошо зарекомендовала себя в задачах семантического сходства текстов, при этом являясь относительно компактной, что важно для производительности микросервиса. Имя модели задается в конфигурационном файле config.py. Также инициализируется весовой коэффициент α, по умолчанию равный 0.8. Этот коэффициент определяет относительный вклад тернарного сходства и текстового сходства в итоговую оценку. Значение 0.8 было выбрано эмпирически как начальная точка, предполагающая больший вес для явных предпочтений, выраженных через тернарные опросы, но при этом позволяющая текстовым описаниям вносить коррективы. В дальнейшем, на основе анализа реальных данных и A/B тестирования, этот коэффициент может быть более точно настроен.

Процесс получения текстового эмбеддинга для описания пользователя реализован в методе get_text_embedding. Если описание отсутствует, возвращается нулевой вектор соответствующей размерности, что обеспечивает корректную работу системы даже при неполных данных.

```
class DynamicRecommendationSystem:
    def __init__(self, alpha: float = 0.8):
    self.text_embedder = SentenceTransformer(TEXT_EMBEDDING_MODEL)
    self.text_embedding_dim = self.text_embedder.get_sentence_embedding_dimension()
    self.alpha = alpha

def _get_text_embedding(self, description: str) -> np.ndarray:
    if not description:
        return np.zeros(self.text_embedding_dim)
    return self.text_embedder.encode(description)
```

Для обработки тернарных ответов используется вспомогательная функция get_ternary_vector. Она принимает словарь ответов пользователя и упорядоченный список идентификаторов вопросов, на основе которых формирует numpy-вектор тернарных ответов. Важно, что порядок вопросов в этом векторе строго задан, что обеспечивает корректное сопоставление профилей разных пользователей.

```
def _get_ternary_vector(answers: dict, question_ids: List[int]) -> np.ndarray:
    return np.array([answers.get(qid, 0) for qid in question_ids], dtype=float)
```

Расчет косинусного сходства между двумя векторами (тернарными или

текстовыми) выполняется функцией compute_similarity. В ней предусмотрена проверка на наличие нулевых векторов, чтобы избежать ошибок деления на ноль; в таком случае сходство полагается равным нулю.

Основная логика формирования рекомендаций заключена в методе get_recommendations класса DynamicRecommendationSystem. Этот метод последовательно обрабатывает запрос: сначала извлекаются упорядоченные идентификаторы вопросов и данные основного пользователя. Затем для основного пользователя и каждого кандидата вычисляются тернарные и текстовые векторы. На основе этих векторов рассчитываются два типа сходства: тернарное и текстовое. Итоговое комбинированное сходство определяется по формуле $Sim(u, v) = \alpha \cdot sim_{ternary}(u, v) + (1 - \alpha) \cdot sim_{text}(u, v)$. После расчета сходства для всех кандидатов (кроме самого пользователя) формируется список объектов MatchingResponse, который сортируется по убыванию коэффициента сходства. Фрагмент, иллюстрирующий расчет комбинированного сходства и формирование ответа, показан ниже:

```
    # ... получение main_ternary, main_text ...
for candidate in request.users:
    if candidate.id == main_user.id:
        continue
    candidate_ternary = _get_ternary_vector(candidate.answers, ordered_qids)
    candidate_text = self._get_text_embedding(candidate.description)
    ternary_sim = _compute_similarity(main_ternary, candidate_ternary)
    text_sim = _compute_similarity(main_text, candidate_text)
    combined_sim = self.alpha * ternary_sim + (1 - self.alpha) * text_sim
    recommendations.append(MatchingResponse(
        id=candidate.id,
        username=candidate.username,
        avatar=candidate.avatar,
        gender=candidate.gender,
        description=candidate.description,
        similarity=combined_sim * 100
    ))
    # ... сортировка recommendations ...
```

API сервиса реализован с использованием FastAPI в файле main.py. Определен единственный эндпоинт /api/matching, который принимает POST-запросы с телом в формате MatchingRequest. Эндпоинт выполняет базовую валидацию входных данных: проверяет наличие списка пользователей и, если предостав-

лены ответы на вопросы, то и наличие самих вопросов. Затем задача формирования рекомендаций делегируется объекту класса DynamicRecommendationSystem. В сервисе предусмотрена обработка стандартных исключений FastAPI и общих исключений для возврата корректных HTTP-ответов клиенту, что повышает надежность его работы. Код эндпоинта представлен следующим образом:

```
@app.post("/api/matching", response_model=List[MatchingResponse])
async def get_matching_users(request: MatchingRequest):
    try:
        if not request.users:
            return []
        if not request.questions and any(u.answers for u in request.users):
            raise HTTPException(status_code=400,
                detail="Answers provided but no questions defined to interpret")
        return dynamic_recommendation_system.get_recommendations(request)
    # ... обработка исключений ...
    except Exception as e:
    # ... логирование ошибки ...
    raise HTTPException(status_code=500, detail=f"Internal server error: {str(e)}")
```

Запуск микросервиса осуществляется с помощью Uvicorn, который обеспечивает асинхронное выполнение и способен обрабатывать большое количество одновременных запросов. Таким образом, разработанный прототип представляет собой функциональный и производительный сервис, способный предоставлять динамические рекомендации на основе комбинирования тернарных профилей и текстовых описаний пользователей. Модульная структура и использование современных фреймворков обеспечивают простоту его дальнейшего сопровождения и масштабирования.

6.4 Преимущества и перспективы развития

Разработанная гибридная рекомендательная система, сочетающая анализ тернарных опросов и текстовых описаний профилей, обладает рядом преимуществ и открывает перспективы для дальнейшего развития и совершенствования.

Ключевым преимуществом предложенного подхода является его динамичность и гибкость. Система не привязана к фиксированному набору вопросов, что позволяет администраторам приложения легко изменять, добавлять или удалять вопросы в опросах без необходимости переобучения основной модели

сходства. Обработка тернарных векторов и текстовых эмбеддингов происходит на лету на основе актуального набора вопросов и описаний, передаваемых в запросе. Это обеспечивает высокую адаптивность системы к изменяющимся потребностям пользователей и эволюции самого приложения.

Другим важным преимуществом является простота и интерпретируемость взаимодействия для пользователя. Тернарные ответы (да, нет, пропустить) интуитивно понятны и не требуют от пользователя сложных оценок или размышлений, что снижает когнитивную нагрузку и повышает вероятность заполнения опросов. При этом пользователь имеет явный контроль над предоставляемыми данными. Относительно небольшое количество вопросов, необходимых для формирования первичного тернарного профиля, и простота их заполнения также минимизируют проблему «холодного старта» для новых пользователей, позволяя системе достаточно быстро начать генерировать осмысленные рекомендации.

Несмотря на продемонстрированные в ходе исследования на датасете «Speed Dating Experiment» умеренные, но осмысленные результаты базовых моделей на тернарных данных, важно отметить ограничения текущего исследования и прототипа. Основное ограничение связано с самим датасетом: его «волновая» структура не позволяет оценить взаимодействие всех возможных пар, что, как обсуждалось ранее, занижает метрики точности и полноты. Кроме того, текущая реализация гибридной системы использует простую линейную комбинацию для агрегации сходств с фиксированным коэффициентом α.

Тем не менее, разработанная система представляет собой прочный фундамент и открывает широкие перспективы для дальнейшего развития и исследований. Одним из главных направлений является сбор и использование исторических данных о реальных взаимодействиях пользователей внутри разработанного мобильного приложения, таких как лайки, мэтчи и характер общения после мэтча. Эти данные позволят обучать более сложные и персонализированные модели, например, на основе коллаборативной фильтрации или нейросетевых подходов, способных улавливать скрытые предпочтения.

Далее, текущий весовой коэффициент α в гибридной модели может быть оптимизирован с помощью A/B тестирования или методов машинного обучения для более точного взвешивания вклада тернарных и текстовых данных. Возможна также разработка более сложных нелинейных функций для агрега-

ции различных типов сходства.

Перспективным направлением является и динамическая адаптация самих опросов. Можно реализовать механизмы, которые бы предлагали пользователям наиболее релевантные или информативные вопросы на основе их предыдущих ответов или активности, что позволит более эффективно собирать данные. Кроме того, система может быть расширена за счет учета контекстуальных факторов, таких как время, геолокация или недавняя активность пользователя, а также анализа дополнительных аспектов профиля, например, фотографий или музыкальных предпочтений, с использованием соответствующих технологий. Внедрение механизмов обратной связи от пользователей на предложенные рекомендации также будет способствовать непрерывному улучшению модели.

Предложенная и реализованная в виде прототипа микросервиса рекомендательная система, основанная на динамическом анализе тернарных опросов и текстовых описаний, демонстрирует свою жизнеспособность и является перспективной отправной точкой. Ее гибкость и возможность учета различных аспектов пользовательского профиля, в сочетании с потенциалом для интеграции более сложных алгоритмов на основе собираемых данных, делают ее ценным компонентом для разрабатываемого мобильного приложения знакомств.

ЗАКЛЮЧЕНИЕ

В рамках работы были успешно достигнуты все поставленные цели по разработке мобильного приложения для знакомств с гибридной рекомендательной системой. Основная задача состояла в создании системы, способной формировать релевантные рекомендации на основе как структурированных ответов пользователей на опросы, так и анализа текстовых описаний их профилей, эта задача была полностью решена.

Было проведено исследование предметной области, спроектирована и реализована многокомпонентная архитектура, включающая мобильное приложение, основную серверную часть и отдельный микросервис рекомендаций. Ключевым результатом является создание гибкой рекомендательной системы, способной динамически адаптироваться к изменениям в опросниках и пользовательских данных, что обеспечивает хороших релевантность рекомендаций и решает проблему «холодного старта». Экспериментальная оценка, проведенная на публичном датасете «Speed Dating Experiment», подтвердила эффективность выбранных подходов.

Разработанное решение демонстрирует практическую значимость, предлагая масштабируемую платформу для онлайн-знакомств. Несмотря на ограничения, связанные со структурой использованного для первоначальной оценки датасета и текущей линейной комбинацией признаков в гибридной модели, успешное выполнение поставленных задач заложило прочный фундамент для дальнейшего развития. Перспективы включают интеграцию более сложных алгоритмов персонализации, сбор и анализ реальных пользовательских данных для уточнения модели, а также расширение функциональности. Таким образом, дипломная работа представляет собой завершенное исследование и разработку, полностью соответствующую первоначальным целям.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Wang, X. Data scarcity in recommendation systems: A survey / X. Wang, Y. Liu, M. Chen at al. // arXiv preprint arXiv:2312.10073. — 2023.

2 Jin, D. A survey on fairness-aware recommender systems / D. Jin, L. Wang, H. Zhang, Y. Zheng, W. Ding, F. Xia, S. Pan // arXiv preprint arXiv:2306.00403. — 2023.

3 Jablons, Z. Large-scale collaborative filtering to predict who on okcupid will like you, with jax [Электронный ресурс] / Z. Jablons // OkCupid Tech Blog. — 2021. — URL: https://tech.okcupid.com/large-scale-collaborative-filtering-to-predict-who-on-okcupid-will-like-you-with-jax-8 (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

4 Carman, A. Finding love on a first data: Matching algorithms in online dating [Электронный ресурс] / A. Carman // Harvard Data Science Review. — 2021. — URL: https://hdsr.mitpress.mit.edu/pub/i4eb4e8b (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

5 Tinder Engineering Team,. Personalized user recommendations at tinder [Электронный ресурс] // Proceedings of the Machine Learning Conference). — San Francisco, USA: 2017. — URL: https://mlconf.com/sessions/personalized-user-recommendations-at-tinder-the-t/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

6 Zhao, Z. Weizhi zhang and yuanchen bei and liangwei yang and henry peng zou / Z. Zhao, W. Fan, J. Li at al. // arXiv preprint arXiv:2501.01945. — 2025.

7 Rudder, C. Okcupid: The math behind online dating [Электронный ресурс] / C. Rudder // AMS Graduate Student Blog. — 2013. — URL: https://blogs.ams.org/mathgradblog/2016/06/08/okcupid-math-online-dating/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

8 Tinder,. Powering tinder — the method behind our matching [Электронный ресурс]. — 2022. — URL: https://www.help.tinder.com/hc/en-us/articles/7606685697037-Powering-Tinder-The-Method-Behind-Our-Matching4 (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

9 Resnick, B. The tinder algorithm, explained [Электронный ресурс] / B. Resnick // Vox. — 2019. — URL: https://www.vox.com/2019/2/7/

18210998/tinder-algorithm-swiping-tips-dating-app-science (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

10 Wells, G. Hinge founder justin mcleod explains how the algorithm finds your match [Электронный ресурс] / G. Wells // Fortune. — 2024. — URL: https://fortune.com/2024/01/18/hinge-ceo-justin-mcleod-interview-attractiveness-score-algorithm-rose-jail/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

11 Relationstips,. How eharmony matches are made: Inside the algorithm [Электронный ресурс] / Relationstips. — 2025. — URL: https://www.relationstips.com/how-eharmony-matches-are-made-inside-the-algorithm/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

12 Sugahara, K. Hierarchical matrix factorization for interpretable collaborative filtering / K. Sugahara, K. Okamoto // arXiv preprint arXiv:2311.13277. — 2023.

13 Zhou, Z. Contrastive collaborative filtering for cold-start item recommendation / Z. Zhou, L. Zhang, N. Yang // arXiv preprint arXiv:2302.02151. — 2023.

14 Lops, P. Content-based recommender systems: State of the art and trends / P. Lops, M. d. Gemmis, G. Semeraro // Recommender Systems Handbook. — 2011. — Pp. 73–105.

15 Zhang, S. Deep learning based recommender system: A survey and new perspectives / S. Zhang, L. Yao, A. Sun, Y. Tay // ACM Computing Surveys (CSUR). — 2019. — Vol. 52, no. 1. — Pp. 1–38.

16 Nabil, S. Demographic information combined with collaborative filtering for an efficient recommendation system / S. Nabil, M. Y. Chkouri, J. El Bouhdidi // International Journal of Electrical and Computer Engineering (IJECE). — 2024. — Vol. 14, no. 5. — Pp. 5916–5925.

17 Beregovskaya, I. Review of clustering-based recommender systems / I. Beregovskaya, M. Koroteev // arXiv preprint arXiv:2109.12839. — 2021.

18 Nadimi-Shahraki, M. H. Cold-start problem in collaborative recommender systems: Efficient methods based on ask-to-rate technique / M. H. Nadimi-Shahraki, M. Bahadorpour // Journal of Computing and Information Technology. — 2014. — Vol. 22, no. 2. — Pp. 105–113.

19 Xiangnan, H. Neural collaborative filtering / H. Xiangnan, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua // arXiv preprint arXiv:1708.05031. — 2017.

20 Hidasi, B. Session-based recommendations with recurrent neural networks / B. Hidasi, A. Karatzoglou, L. Baltrunas, D. Tik // arXiv preprint arXiv:1511.06939. — 2016.

21 Kang, W. Self-attentive sequential recommendation / W. Kang, J. J. McAuley // arXiv preprint arXiv:1808.09781. — 2018.

22 van den Berg, R. Graph convolutional matrix completion / R. van den Berg, T. Kipf, M. Welling // arXiv preprint arXiv:1706.02263. — 2017.

23 Ruining, H. Visual bayesian personalized ranking from implicit feedback / H. Ruining, J. McAuley // arXiv preprint arXiv:1510.01784. — 2015.

24 Cano, E. Hybrid recommender systems: A systematic literature review / E. Cano, M. Morisio // Intelligent Data Analysis. — 2017. — Vol. 21, no. 6. — 1487–1524 p.

25 Guo, H. Deepfm: A factorization-machine based neural network for ctr prediction / H. Guo, R. Tang, Y. Ye, Z. Li, X. He // arXiv preprint arXiv:1703.04247. — 2017.

26 Recommender model evaluation: Offline vs. online [Электронный ресурс] // Shaped Blog. — 2023. — URL: https://www.shaped.ai/blog/evaluating-recommender-models-offline-vs-online-evaluation (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

27 Verma, R. Mobile app performance optimization - how it works and more? [Электронный ресурс] / R. Verma // Bacancy Technology. — 2024. — URL: https://www.bacancytechnology.com/blog/mobile-app-performance (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

28 Dhaduk, H. Mobile application architecture: Layers, types, principles, factors [Электронный ресурс] / H. Dhaduk // Simform Blog. — 2024. — URL: https://www.simform.com/blog/mobile-application-architecture/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

29 Client-server architecture – system design [Электронный ресурс] // GeeksforGeeks. — 2024. — URL: https://www.geeksforgeeks.org/

client-server-architecture-system-design/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

30 Best practices - oauth for mobile apps [Электронный ресурс] // Curity. — 2025. — URL: https://curity.io/resources/learn/oauth-for-mobile-apps-best-practices/ (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.

31 Zinkus, M. Data security on mobile devices: Current state of the art, open problems, and proposed solutions / M. Zinkus, T. M. Jois, M. Green // arXiv preprint arXiv:2105.12613. — 2021.

32 Fisman, R. Speed Dating Experiment Dataset [Электронный ресурс] / R. Fisman, S. S. Iyengar. — Kaggle. — URL: https://www.kaggle.com/datasets/annavictoria/speed-dating-experiment (Дата обращения 30.04.2025). Загл. с экр. Яз. англ.