**Понимание задачи (1 неделя)**

На этом этапе задача заключается в четком определении алгоритмов, которые будут использоваться для построения системы рекомендаций. Мы начали с анализа существующих подходов к рекомендациям (коллаборативная фильтрация, контентная фильтрация и гибридные модели). Для этого применяли литературный обзор и изучали текущие исследования. Результатом данного этапа стало формулирование технического задания, в котором были обозначены требования к системе, ожидаемые результаты и ключевые метрики для оценки успешности работы алгоритмов. Данный процесс обеспечил четкое направление для дальнейших этапов проекта.

**Отчет о работе за 2 недели: Сбор и предобработка данных**

**Неделя 1: Создание среды разработки и начало сбора данных**

В первую неделю второго этапа проекта мы сосредоточились на создании среды для разработки и тестирования модели. Были выбраны основные технологии и инструменты, такие как Jupyter Notebook, который обеспечил интерактивную среду для анализа данных и разработки моделей, а также Anaconda для управления библиотеками и зависимостями.

Также начат процесс сбора данных. Изучены доступные внутренние источники данных, включая базы данных пользователей и просмотров. Подключение к внешним источникам (API) было организовано для получения дополнительной информации о фильмах, таких как жанры, рейтинги и другие метаданные. Первые запросы к API были выполнены, и начата интеграция полученных данных с внутренними источниками.

**Неделя 2: Предобработка данных**

На второй неделе мы продолжили работу над сбором данных и уделили значительное внимание их предобработке. Используя библиотеку Pandas, проведена очистка данных: удалены дубликаты, обработаны пропуски в данных путем заполнения их средними значениями или медианами, а также нормализованы значения в числовых колонках для лучшего представления данных.

Была выполнена проверка качества данных в целях выявления и устранения ошибок. В результате работы собраны и очищены данные, что обеспечивало качественную основу для дальнейшего анализа и построения модели. Завершение этого этапа дало нам уверенность в том, что данные готовы к следующему шагу в разработке системы рекомендаций.

**Отчет о работе за 2 недели: Исследовательский анализ данных**

**Неделя 1: Визуализация данных**

На первом этапе исследовательского анализа данных мы сосредоточились на визуализации собранной информации. С помощью библиотек Matplotlib и Seaborn были созданы разнообразные графики, включая гистограммы и диаграммы рассеяния, для выявления паттернов и тенденций в предпочтениях пользователей.

Анализ визуализаций позволил нам выявить зависимости между различными переменными, такими как жанры фильмов, количество просмотров и рейтинги. Наблюдения за графиками помогли сформулировать гипотезы о том, как различные факторы могут влиять на пользовательские предпочтения, что стало основой для более глубокого анализа.

**Неделя 2: Сегментация пользователей**

Во вторую неделю работы мы углубились в анализ поведения пользователей через сегментацию. Была проведена кластеризация пользователей на основе их предпочтений и активности, что дало возможность лучше понять различные группы пользователей и их характеристики. Используя полученные данные о взаимодействиях пользователей с фильмами, мы выделили ключевые сегменты, такие как «любители жанра», «активные зрители» и «редкие пользователи».

Кроме того, были сделаны выводы о тенденциях, замеченных в разных сегментах, что помогло сформировать более детальное представление о предпочтениях пользователей. Эти шаги позволили глубже понять структуру данных и подготовить необходимые выводы для следующего этапа разработки модели.

**Отчет о работе за 3 недели: Построение модели**

**Неделя 1: Разработка архитектуры модели**

В первой неделе работы мы сосредоточились на разработке архитектуры модели рекомендаций. Исходя из полученного на предыдущем этапе анализа данных, мы определили, что комбинированный подход, использующий как коллаборативную фильтрацию, так и контентные фильтры, будет наиболее эффективным. В качестве основных методов коллаборативной фильтрации были выбраны алгоритмы на основе K-ближайших соседей (KNN), а для контентного фильтра рекомендовали использовать характеристику фильмов, такие как жанр, режиссер и актерский состав.

**Неделя 2: Имплементация и обучение моделей**

На второй неделе мы приступили к реализации выбранных алгоритмов. С использованием библиотеки Surprise для коллаборативной фильтрации мы настроили и обучили модель на собранных данных. В процессе имплементации мы провели предварительный анализ производительности каждого из алгоритмов, чтобы понять, как они справлялись с нашими данными. Это позволило нам оценить, какие методы показывают лучшие результаты и какие аспекты модели требуют дополнительной настройки.

**Неделя 3: Настройка гиперпараметров и оценка производительности**

Третью неделю мы посвятили настройке гиперпараметров выбранных моделей. Для этого мы использовали метод кросс-валидации, который помог выявить оптимальные параметры для каждой модели. Процесс включал в себя тестирование различных комбинаций гиперпараметров и оценку их влияния на точность рекомендаций. В результате удалось значительно улучшить производительность моделей, что было подтверждено с помощью метрик, таких как MAE (средняя абсолютная ошибка) и RMSE (корень из средней квадратичной ошибки). Завершив этот этап, мы получили модели, готовые к дальнейшему тестированию и интеграции.

**Отчет о работе за 2 недели: Оценка модели и интерпретация результатов**

**Неделя 1: Применение и анализ метрик**

В первую неделю мы сосредоточились на применении различных метрик для оценки производительности построенной модели рекомендаций. Основными метриками, которые мы использовали, были RMSE (корень из среднеквадратичной ошибки) и Precision@K, которые позволили нам измерить точность рекомендаций в различных аспектах. На основе тестовой выборки мы провели детальный анализ, который позволил выявить как сильные, так и слабые стороны модели.

Мы обнаружили, что модель демонстрирует высокую точность в рекомендациях для пользователей с обширной историей оценок, однако у нее были проблемы с рекомендациями для пользователей с ограниченными или новым набором данных. Это показало, что коллаборативная фильтрация зависит от наличия достаточного количества информации о пользователях и элементах.

**Неделя 2: Интерпретация результатов и оптимизация**

На второй неделе мы провели интерпретацию результатов тестирования, сосредоточившись на факторах, влияющих на качество рекомендаций. Мы выделили ключевые аспекты, такие как количество взаимодействий пользователя с элементами, а также разнообразие контентных характеристик фильмов. Это помогло нам понять, какие конкретные факторы следует улучшить для повышения общей производительности модели.

Основываясь на полученных данных, мы провели оптимизацию модели. Это включало пересмотр исходных алгоритмов, их комбинаций и гиперпараметров, чтобы улучшить точность рекомендаций для пользователей с ограниченной информацией. Мы эксперименировав с различными подходами, включая смешение коллаборативной и контентной фильтрации, что дало положительные результаты. На этом этапе мы приготовили модель к финальному тестированию и внедрению, что позволило бы значительно повысить ее продуктивность и адаптивность к различным сценариям.