**Создать модель, собирающую тот набор фильмов, которые пользователь скорее всего посмотрит.**

**Шаг 1. Business Understanding**

Понять, что конкретно мы хотим получить. Как мы видим результат нашей деятельности. Определить критерии успеха проекта. Насколько точной должна быть наша модель, определяющая список фильмов.

Проанализировать наши ресурсы. Сможем ли мы при имеющейся команде сотрудников достигнуть результата работы, отвечающего всем требованиям и ограничениям, либо нужно привлекать дополнительные мощности, трудовые ресурсы (аутсорсинг, краудсорсинг)

Провести анализ рисков. Удастся ли нам добиться той точности модели, которая нам нужна. Достаточно ли данных в источниках, на основании которых мы будем строить нашу модель.

Провести анализ затрат/выгод. Каковы будут затраты на реализацию данного проекта на наших мощностях, либо при привлечении аутсорса, краудсорса. Какова прогнозируемая прибыль.

После анализа вышеперечисленного, если нас все устраивает, то прикинуть примерный план действий.

**Шаг 2. Data Understanding**

Ввести в курс дела команду. Обсудить все ли имеется у команды для выполнения задачи и какие могут быть непредвиденные обстоятельства.

Обсудить план действий и выбрать инструменты. Какие методы помогут достигнуть максимального качества модели при минимальных затратах. Как будет представлен результат. Необходима ли интеграция нашей модели с какими-либо нашими сайтами/платформами.

Если все хорошо, то начинаем работу.

Собираем данные со всех необходимых источников данных. Описываем их качество и количество. Анализируем данные на ошибки, пропуски, недостатки метаданных.

**Шаг 3. Data Preparation**

Определяем основные атрибуты, гипотезы, множества и подмножества. Определяем стак данных, необходимых для качественного расчета. После выбора данных разделяем их на основную и тестовую выборки.

Избавляемся от ошибок, убираем дубликаты, по возможности заполняем пропуски, расширяем данные новыми признаками. Проверяем корреляцию данных с итоговым показателем (Например, как сильно взаимосвязаны история по пользователям и истории их просмотров с прогнозируемым списком фильмов). Определяем приоритетность данных для того, чтобы понять, что конкретно нам использовать.

Сохраняем данные для последующего построения модели.

**Шаг 4. Modeling**

Выбираем метод моделирования (или несколько методов), строим модель, проводим расчеты. Определяем точность модели (если их несколько, то выбираем самую точную, либо при одинаковой точности выбираем ту, которая требует меньше энергозатрат). Проводим эксперимент как работает модель для наших условий. Насколько точна вероятность просмотра конкретного фильма из списка пользователем, хватает ли заданных источников данных, есть ли ошибки в модели. Описываем результаты работы для дальнейшего анализа.

**Шаг 5. Evaluation**

Проанализировать итоги результатов моделирования. Достигла ли модель требуемых результатов. Какие результаты показывает нам модель и не появились ли в процессе какие-либо новые идеи (привлечение дополнительных данных, смена модели, методы снижения затрат).

Нет ли каких-либо трудностей и дополнительных вопросов. Есть ли какие-нибудь недочеты, стоит ли их исправлять, и если стоит, то как именно это сделать.

Сможем ли мы предоставить результат своей работы, и сможем ли мы его грамотно принять и в дальнейшем беспроблемно с ним работать.

Если все хорошо, то спланировать алгоритм внедрения.

**Шаг 6. Deployment**

Если у нас было несколько моделей, то составляем план внедрения для каждой из них. Определяем для какой модели осуществление внедрения пройдет самым эффективным способом.

Предполагаем возможные риски и проблемы во время внедрения и заранее планируем свои действия при возникновении этих проблем. (Возможны проблемы с ПО, несовместимость версий, недостаток ресурсов для обслуживания)

Определяем дальнейшее взаимодействие с моделью. Будет ли ей требоваться поддержка с нашей стороны в дальнейшем. Будет ли требоваться актуализация модели и что для этого придется сделать. Как часто будет требоваться актуализация (Если часто, то может быть стоит переделать нашу модель так, чтобы ее было проще адаптировать к изменяющимся условиям). Будет ли работать модель при обновлении версий ПО.

Провести итоговый обзор проекта.

**Роли в команде.**

**Шаг 4. Modeling**

В первую очередь нам необходим DS-специалист, так как именно он будет строить модели. То есть самая важная задача у него. К нему на помощь можно приставить Аналитика (Так как он обладает способностью проверить значимость результата и вытянуть смысл из полученных метрик, чтобы понять, как их интерпретировать), а также Data Engeneera (Так как он понимает потребности DS-специалиста и умеет работать с большими данными). Также можно привлечь разработчика, так как он умеет провести интеграции, настроить получение требуемых данных, встроить модель и ее протестировать. И, конечно, администрирование ляжет на плечи менеджера продукта / CDO так как эти люди следят за продвижением всего продукта по роудмапу и продвижением аналитики по всей компании.

**Шаг 5. Evaluation**

На данном этапе самым ценным кадром будет аналитик, так как он сумеет проанализировать итоги результатов моделирования, достигла ли модель требуемых результатов. Ему в помощь придут DS-специалист, так как он умеет извлекать важную информацию и инсайты. Также нам потребуется разработчик, который сможет нам сказать как встроить эту модель и с ней в дальнейшем работать и еще архитектор баз данных и специалист-разработчик баз данных для того, чтобы они нас проконсультировали в вопросах оптимального хранения данных нашей модели, легкого доступа к ним, а также выбрать технологии и оптимизацию (Но больше они, конечно потребуются на 6 этапе). Ну и куда же без администрирования. Менеджер продукта / CDO следят за продвижением всего продукта по роудмапу и продвижением аналитики по всей компании.