<https://habr.com/ru/company/lanit/blog/420499/>

<https://grouplens.org/datasets/movielens/> датасету для тестирования рекомендаций

**Обзор и постановка задачи**

Задача рекомендательной системы – проинформировать пользователя о товаре, который ему может быть наиболее интересен в данный момент времени. Клиент получает информацию, а сервис зарабатывает на предоставлении качественных услуг.

Рекомендательные системы – это про то, что предложить клиенту, чтобы сделать его счастливым.

**Алгоритмы**.  
  
Несмотря на множество существующих алгоритмов, все они сводятся к нескольким базовым подходам, которые будут описаны далее. К наиболее классическим относятся алгоритмы Summary-based (неперсональные), Content-based (модели основанные на описании товара), Collaborative Filtering (коллаборативная фильтрация), Matrix Factorization (методы основанные на матричном разложении) и некоторые другие.

**Неперсонализированные рекомендации**

Начнем с неперсонализированных рекомендаций, поскольку они самые простые в реализации. В них потенциальный интерес пользователя определяется просто средним рейтингом товара: «Всем нравится – значит понравится и вам».

#### Проблема холодного старта

Холодный старт – это типичная ситуация, когда ещё не накоплено достаточное количество данных для корректной работы рекомендательной системы (например, когда товар новый или просто его очень редко покупают). Если средний рейтинг посчитан по оценкам всего трёх пользователей (igor92, xyz\_111 и oleg\_s), такая оценка явно не будет достоверной, и пользователи это понимают. Часто в таких ситуациях рейтинги искусственно корректируют.

#### Актуальность рекомендаций

В некоторых случаях также важно учитывать «свежесть» рекомендации. Это особенно актуально для статей или постов на форумах.

Пример расчета рейтинга в журнале Hacker news:



где U = upvotes, D = downvotes, а P (Penalty) — дополнительная корректировка для имплементации иных бизнес-правил  
  
Расчет рейтинга в Reddit:



где U = число голосов «за», D = число голосов «против», T = время записи. Первое слагаемое оценивает «качество записи», а второе делает поправку на время.

**Content-based рекомендации**

Персональные рекомендации предполагают максимальное использование информации о самом пользователе, в первую очередь о его предыдущих покупках. Одним из первых появился подход content-based filtering.