

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.Ломоносова  
ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ  
КАФЕДРА МАТЕМАТИЧЕСКИХ МЕТОДОВ АНАЛИЗА ЭКОНОМИКИ



ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
**«УЛУЧШЕНИЕ ПОСЛЕДОВАТЕЛЬНЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ ДЛЯ  
ПЛАТФОРМ ЭЛЕКТРОННОЙ КОММЕРЦИИ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ  
МАШИННОГО ОБУЧЕНИЯ»**

Выполнил студент  
группы э401:  
Кабанов И. Д.  
Научный руководитель:  
Н.с. Елицур Д. А.

Москва – 2025

## Оглавление

<b>Введение.</b> .....	3
<b>Глава 1. Обзор моделей рекомендательных систем и их основных свойств в области электронной коммерции.</b> .....	8
1.1 Цели и задачи рекомендательных систем. ....	8
1.2 Полезные для бизнеса свойства рекомендательных систем. ....	11
1.3 Основные типы моделей рекомендательных систем. ....	14
1.4 Трансформерная модель последовательных рекомендаций. ....	17
<b>Глава 2. Построение трансформерной модели рекомендаций для её первичного анализа.</b> .....	21
2.1 Разведывательный анализ данных и их обработка. ....	21
2.2 Обучение последовательной рекомендательной модели. ....	23
2.3 Тестирование модели и первичный анализ её рекомендаций. ....	25
<b>Глава 3. Комплексный анализ последовательных рекомендаций построенной трансформерной модели.</b> .....	32
3.1 Визуальный анализ рекомендаций модели. ....	32
3.2 Зависимость рекомендаций модели от последовательной структуры пользовательской истории. ....	36
3.3 Тематическая согласованность рекомендаций с последовательностью истории пользователя. ....	38
<b>Заключение.</b> .....	41
<b>Библиография.</b> .....	43
<b>Приложения.</b> .....	47
Приложение 1. История и рекомендации пользователя 806550. ....	47
Приложение 2. История и рекомендации пользователя 733322. ....	48
Приложение 3. История и рекомендации пользователя 600927. ....	49

## **Введение.**

### **Актуальность исследования.**

По данным международной технологической компании BYYD<sup>1</sup>, в 2024 году россиянин проводил в интернете в среднем около 8 с половиной часов в день. То есть сейчас человек в России проводит около трети своей жизни на различных цифровых платформах, где он изучает нужную ему информацию, смотрит фильмы, слушает музыку, читает новости, общается с другими пользователями. Так как сейчас контента в Интернете колоссально много, он очень отличается по своему содержанию и качеству, пользователю самостоятельно довольно сложно найти подходящую для него информацию. Здесь человеку на помощь приходят алгоритмы, которые значимо упрощают его поиск, предлагая его вниманию подходящий контент, на основе его прошлого поведения. Такие алгоритмы и модели называют рекомендательные системы. Именно о них и пойдет речь в данной работе.

Область применения рекомендательных систем настолько обширна, что охватывает чуть ли не все коммерческие цифровые платформы: публикации в социальных сетях, рекламные посты в Интернете, персонализированные банковские услуги в приложениях, рекомендации на игровых и образовательных платформах, предложенные музыкальные треки и видео ролики на платформах, таких как Apple Music или YouTube. Этот список можно продолжать довольно долго. В данной работе я буду рассматривать применение рекомендательных систем в одной из самых популярных для них области – в электронной коммерции. К такой области относятся все цифровые платформы, где клиентам рознично продаются товары из одной категории (например, СберАптека) или сразу из нескольких категорий (Amazon, Wildberries, ЯндексМаркет). В этой сфере алгоритмы рекомендаций позволяют очень серьезно повышать доходы компаний, предлагая клиентам более релевантные для них товары. Приведу несколько примеров внедрения рекомендательных систем на сайты торговых площадок. По данным крупнейшей международной консалтинговой компании McKinsey<sup>2</sup>, внедрение алгоритмических рекомендаций увеличило доходы Amazon на 35%. По данным технологической компании Constructor<sup>3</sup>, их модели рекомендательных систем на основе искусственного интеллекта увеличили продажи

---

<sup>1</sup> <https://www.byyd.me/ru/blog/2024/02/digital-2024-datareportal/>

<sup>2</sup> <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>

<sup>3</sup> <https://constructor.com/solutions/recommendations>

на клиента американской компании Fisheries Supply на 15% и продажи за одно посещение российской мебельной компании Home24 на 5%. Основываясь на этих фактах, можно говорить о высокой актуальности моей темы. Также, помимо микроэкономических аспектов, можно указать и на макроэкономические последствия применения таких алгоритмов. Так как данные модели увеличивают продажи компаний, можно с уверенностью говорить о том, что они увеличивают спрос на товары производственных компаний. Из-за мультипликативного эффекта это стимулирует и общий спрос в экономике, что может привести к смягчению кризисов спроса, как это частично было в пандемию коронавируса.

Рассмотрев экономические выгоды от внедрения автоматизированных рекомендаций в бизнес онлайн-торговли, я бы хотел перейти к рассмотрению вопроса о важности учитывать последовательность покупок клиентов. Поскольку в каталоге многие товары являются комплементарными и часто покупаются вместе друг с другом, важно учитывать последовательную и временную связь между покупками для обеспечения наибольшей релевантности рекомендаций [Kang, McAuley, 2018]. Этот тезис я бы хотел пояснить на примере электроники. После покупки клиентом нового ноутбука Apple, можно руководствоваться логикой, что он любит технику Apple и он захочет купить смартфон iPhone из сопоставимой ценовой категории. Возможно, и правда, он любит электронику от Apple, ему нравится смартфон iPhone, но в данном случае он пришел именно за ноутбуком, потому что ноутбука у него нет, или его хочется сменить. Крайне маловероятно, что в этот же момент у него сломался телефон или тоже резко захотелось обновить его модель. Поскольку вся эта техника дорогостоящая, вероятно, у покупателя даже не будет денег на покупку смартфона дополнительно к ноутбуку. Более логично, в таких случаях рекомендовать комплементарные товары, такие как мышь для ноутбука или принтер, а к принтеру можно продать бумагу или картридж. Так как рекомендации часто появляются на экране после добавления очередного товара в корзину, то в таких случаях также важно рекомендовать товары, которые клиент может купить сразу после предмета, только что добавленного в корзину. Помимо краткосрочных закономерностей, нужно учитывать и долгосрочные паттерны покупок. Например, рекомендации онлайн-магазина для младенцев не должны вечно рекомендовать одному и тому же клиенту еду для новорожденных детей, а изменяться с возрастом ребенка.

До появления в 2017 году механизма трансформера [Vaswani et al., 2017] модели последовательных рекомендаций могли эффективно учитывать либо только самые последние товары в пользовательской истории, либо, наоборот, только долгосрочные

паттерны. После невероятного успеха этой нейронной сети в языковых задачах, были созданы модели аналогичной архитектуры для последовательных рекомендательных систем: Self-Attentive Sequential Recommendation (SASRec) [Kang, McAuley, 2018] и Bidirectional Encoder Representations for Transformer for Recommendation (Bert4Rec) [Sun et al, 2019]. Эти модели, благодаря основному свойству трансформеров – вниманию, могут запоминать как краткосрочные и долгосрочные тенденции покупок, увеличивая релевантность выданных рекомендаций.

Исходя из вышесказанных факторов, мотивация данного исследования заключается в следующем. Во-первых, в этой работе я собираюсь проанализировать эффективность передовой модели трансформера для рекомендательных систем на уникальных данных реальной компании PetCo, которые я самостоятельно предобработаю. Возможно, мои результаты будут отличаться от авторов исходных статей про SASRec и аналогичных моделей. В частности, авторы статьи [Ferrari et. al, 2019] пришли к выводу о том, что на менее популярных датасетах простые модели рекомендаций иногда обходят по метрикам сложные модели нейросетей, презентованные на знаменитых конференциях. Вполне вероятно, что такое ждет и SASRec на моих данных. Во-вторых, важно оценить соответствие рекомендаций модели потребностям бизнеса. Было доказано, что высокое значение стандартных метрик (Normalized discounted cumulative gain (NDCG)@k, Mean Average Precision (MAP)@k, HitRate@k] абсолютно не обеспечивает «хорошие» свойства модели [Abdollahpouri H. et al, 2019]. В-третьих, необходимо содержательно проанализировать выданную моделью рекомендации (понять логику модели) для ее применения в бизнесе. В исследовательской среде машинного обучения пришли к единогласному мнению [Wagstaff, 2012; Rudin and Wagstaff, 2014; Lo Piano, 2020; Ashoori and Weisz, 2019; Thiebes et al., 2020; Spiegelhalter, 2020; Brundage et al., 2020], что интерпретируемость является ключевым фактором доверия к моделям искусственного интеллекта. По этой причине возможность донести логику модели до менеджеров является одним из ключевых факторов её применимости в бизнесе. Довольно естественно ожидать, что бизнесмен не будет рисковать своими деньгами для внедрения модели, которую он не понимает и которой, следовательно, он не доверяет.

Новизна моего исследования заключается в двух основных вещах. Во-первых, в подавляющем большинстве статей про рекомендательные модели трансформера, используются одни и те же наиболее популярные датасеты: Amazon, MovieLens, Netflix (см. Рисунок 1)

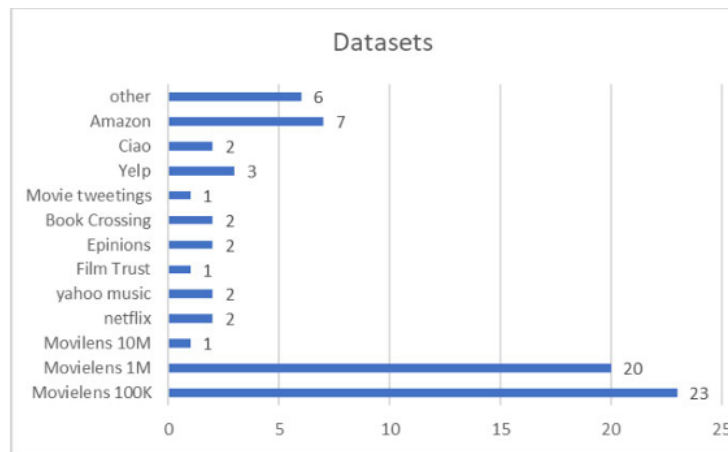


Рисунок 1. Популярные данные для обучения моделей рекомендательных систем. Источник: [SAIFUDIN, WIDIYANINGTYAS, 2024]

Мое исследование будет основано на таком датасете, на котором еще никто в научной среде глубоко не анализировал поведение модели SASRec – на каталоге онлайн-магазина товаров для животных компании PetCo. Возможно, этот фактор приведет к неожиданным результатам, как я об этом писал ранее. Во-вторых, я проведу более глубокий и подробный содержательный анализ рекомендаций SASRec, чем встречается в научной среде. Мною будут использованы нестандартные метрики оценки качества рекомендаций, рассмотрение репрезентативных примеров рекомендаций для выдвижения исследовательских гипотез и изучение зависимости модели от последовательной структуры пользовательской истории.

## Цель и задачи.

*Цель работы* – Провести комплексный анализ качества и содержания рекомендаций модели SASRec для её внедрения в бизнес онлайн-ритейла товаров для животных

Исходя из данной цели можно сформулировать следующие понятия:

*Объект исследования* – модель SASRec для последовательной рекомендательной системы в онлайн зоомагазине.

*Предмет исследования* – качество и содержание рекомендаций модели SASRec в условиях специализированного ассортимента товаров для животных.

*Исследовательский вопрос* – какими потенциально полезными для бизнеса качествами обладает модель SASRec и как рекомендуемые ею товары зависят от пользовательской истории?

Для достижения поставленной цели были сформулированы следующие задачи:

1а. На основе анализа литературы выявить потенциально полезные свойства рекомендательных систем с точки зрения достижения целей бизнеса электронной

коммерции.

- 1b. На основе анализа литературы изучить основные типы моделей рекомендательных систем и сравнительные преимущества модели SASRec.
2. Провести обучение и тестирование модели SASRec для первичного анализа её свойств.
3. Сравнить рекомендации алгоритма SASRec с другими моделями с помощью ряда наглядных примеров. Проанализировать зависимость предлагаемых моделью рекомендаций от последовательной структуры пользовательской истории (веса внимания, тематическая схожесть).
4. На основе полученных результатов оценить целесообразность использования данной модели в сегменте онлайн зоо-ритейла. (Это часть будет описана в заключении.)

*Выдвигаемая гипотеза:* модель SASRec потенциально превосходит другие основные модели рекомендательных систем по важным свойствам (релевантность, персонализация, разнообразие, новизна) (Частично подтвердилась). Рекомендации модели SASRec учитывают не только последние товары в пользовательской истории, но и более ранние её элементы (Не подтвердилась). Рекомендации модели SASRec более похожи на последние товары в пользовательской истории. (Подтвердилась)

Данная работа построена следующим образом: в главе 1 представлен обзор на основе научной литературы потенциально полезных для бизнеса свойств рекомендательных систем и основных видов моделей рекомендаций; в главе 2 описаны процесс обработки датасета, обучения и тестирования модели SASRec, первичный анализ её свойств в сопоставлении с моделями-бейзлайнами; в главе 3 представлено визуальное сравнение моделей на ряде примеров, постановка и доказательства двух гипотез относительно содержания рекомендаций модели SASRec. Весь код доступен по [ссылке](#). Чрезмерно большие средства визуализации информации приведены в Приложении.

## **Глава 1. Обзор моделей рекомендательных систем и их основных свойств в области электронной коммерции.**

Рекомендательные системы являются одной из наиболее прикладных областей машинного обучения, в которой высокая прогнозная сила моделей может принести огромные доходы компании. Однако для достижения этой конечной цели бизнеса команде необходимо решить массу разнообразных задач касательно: выбора цели модели и соответствующей онлайн-метрики, выявления её желаемых свойств, выбор оффлайн метрик для оценивания этих свойств, выбор архитектуры модели, выбор способа её обучения и сбора данных для этого обучения, дальнейшее АВ-тестирование и принятие решение относительно дальнейшего её использования. Также стоит сказать, что применяемые модели не являются универсальными и эффект от их внедрения в разных областях бизнеса может быть сильно различен и оценивать его стоит по-разному [Jannach, Adomavicius, 2016]. Поэтому исследователю в этой области необходимо проанализировать особенно большое количество литературы на эту тему и адаптировать их подходы и результаты конкретно под свою задачу.

В данной главе я сначала провожу обзор целей и критериев оценки рекомендательных систем на основе литературы, подкрепляя их примерами из индустрии и личного потребительского опыта. Так, я определю для себя, как мне следует оценивать интересующую модель с остальными. Затем я рассматриваю аналогичные эмпирические работы, в которых подробно описываются алгоритмы машинного обучения для рекомендательных систем и результаты тестирования таких моделей. Таким образом, я смогу подробно изучить архитектуру интересующей меня модели SASRec и отобрать модели-бейзлайны для её сравнения.

### **1.1 Цели и задачи рекомендательных систем.**

В данной главе первый главный вопрос, на который нужно дать ответ, звучит следующим образом: «Какие свойства рекомендательных систем потенциально являются важными для внедрения в бизнес электронной коммерции?». Для ответа на этот вопрос логично сначала обсудить, для каких целей внедряются рекомендательные алгоритмы в бизнес онлайн-ритейла. Первостепенной целью коммерческой компании является извлечение прибыли. Даже обладая высоким авторитетом в обществе, заботясь о природе и поддерживая социально незащищенные слои общества, убыточная коммерческая компания рано или поздно будет вынуждена закрыться. Есть два основных способа увеличения



прибыли: увеличение доходов и снижение издержек. Основной статьей доходов торговых компаний, работают ли они в виде онлайн-платформ или в виде оффлайн торговых магазинов, являются доходы с продаж. Поэтому увеличение продаж торгового бизнеса может очень значительно увеличить его прибыль и улучшить финансовое положение этой компании. Именно на увеличение продаж в основном направлены рекомендательные системы в отрасли электронной коммерции [Jannach, Adomavicius, 2016]. Однако этот эффект может быть не прямым, а косвенным - через посреднические показатели, которые напрямую связаны с уровнем продаж. В индустрии электронной коммерции встречаются следующие цели внедрения рекомендательных систем:

- Непосредственное увеличение продаж через показ релевантных товаров и стимулирование дополнительных покупок.

Пример: Сеть продуктовых магазинов «ВкусВилл»<sup>4</sup> позволяет оформить доставку продуктов через их онлайн-приложение. После добавления очередного продукта в корзину (жареной картошки) клиенту предлагается купить товар, сопутствующий добавленному (кетчуп). Если изначально пользователь не думал о покупке кетчупа, то рекомендации приложения напрямую увеличили продажи «ВкусВилла».

- Непосредственное увеличение продаж через снижения временных издержек поиска

Пример: Немецкий онлайн-магазин одежды «Zalando» предлагает клиентам одежду с учетом их любимого стиля и ценового сегмента.<sup>5</sup> Если клиентке не подходят по её критериям платья популярных брендов, а её нужно более редкое и персонализированное платье, например, с очень большим вырезом на спине, то она может очень долго пролистывать товары в каталоге, отфильтровывая каждый просмотренный предмет. Если такой процесс на самом деле будет продолжаться долго, то девушка может прекратить поиск и покинуть сайт, так и не сделав покупку. Персонализированная рекомендательная система на основе прошлых покупок клиентки должна предложить ей короткий список платьев, среди которых есть то, что соответствует её критериям. Таким образом, девушка все-таки найдет нужный ей товар и совершит покупку.

- Улучшение пользовательского опыта => Увеличение продаж

Пример: Высоко персонализированная рекомендательная система Netflix способна также предлагать к просмотру нишевые фильмы из его любимой категории, о которых он раньше

---

<sup>4</sup> <https://vkusvill.ru/>

<sup>5</sup> [https://www.drapersonline.com/insight/zalando-on-using-ai-to-create-personalised-experiences-for-consumers?utm\\_source=chatgpt.com](https://www.drapersonline.com/insight/zalando-on-using-ai-to-create-personalised-experiences-for-consumers?utm_source=chatgpt.com)

никогда не слышал [Gomez-Uribe, Hunt, 2015]. Так, пользователь может быть приятно удивлен, останется доволен сервисом и продолжит покупать фильмы только здесь, чтобы также получать неожиданно приятные рекомендации. Также, возможно, что клиент, удовлетворенный удобством сервиса, посоветует его своим знакомым, которые начнут покупать фильмы тоже в этом онлайн-магазине и тем самым тоже увеличат продажи бизнеса.

- Улучшение опыта продавцов => Увеличение продаж

Пример: Если система рекомендует пользователям не только товары популярных брендов, но и малоизвестных производителей, то такие мелкие продавцы чаще выставляют свой ассортимент на продажу на онлайн-платформе. Так, каталог пополняется более разнообразными по качеству и бренду товарами и благодаря сетевому клиенту привлекает большее количество клиентов на эту платформу, также увеличивая продажи сервиса. [Ye Z. et al, 2023]

- Монетизация через платные продвижения => Увеличение дохода не с продаж

Пример: Продавцы, выставляющие свои товары на WildBerries<sup>6</sup>, могут заплатить сервису за то, чтобы WildBerries чаще показывал эти товары пользователям на рекомендательных полках. При этом платформа может показывать определенный товар в рекомендациях только тем пользователям, для которых данный товар будет релевантен. Таким образом, сервис и получит плату за продвижение товара, и не снизит пользовательский опыт, оставляя релевантность рекомендаций на прежнем уровне.

Для того, чтобы поставить цель перед рекомендательной системой конкретного сервиса, нужно знать, на каком конкретном месте на онлайн-платформе она будет работать и какой онлайн-метрикой нужно будет ее измерять на АВ-тесте [Jannach, Jugovac, 2019]. На платформе онлайн-магазина PetCo, чьи данные я использую, рекомендательные алгоритмы находятся в разных местах: на главной странице, при поиске по ключевому слову или в категории, рядом с описанием конкретного товара или в специальном окне, всплывающем после добавления очередного товара в корзину. При этом в разных местах рекомендации выполняют разные функции, к ним соответственно нужно предъявлять разные требования и соответственно по-разному обучать. Помимо этого, у сайта могут быть свои нестандартные бизнес-цели, которым не соответствует конкретный рекомендательный алгоритм, могут быть инфраструктурные ограничения, о которых мне не известно, или

---

<sup>6</sup> <https://sellego.com/blog/stati/rekomendatelnye-polki-na-vb-chto-eto-i-kak-popast/>

внедряемая модель может использоваться в каскаде с другой. С этой точки зрения, я не могу самостоятельно обозначать конкретную цель внедрения рекомендательного алгоритма на сайте PetCo, равно, как я и не могу обозначать конкретную онлайн-метрику. Однако это не делает бессмысленным анализ потенциально внедряемой модели и важных свойств ее рекомендаций. На основе выделенных свойств и анализа содержания модельных рекомендаций держателю онлайн-сервиса будет проще составить своё мнение относительно данной модели, поставить для неё конкретную цель и выбрать онлайн-метрику, имея полную информацию о долгосрочных планах сервиса, других частях рекомендательной системы и его технических ограничениях.

## 1.2 Полезные для бизнеса свойства рекомендательных систем.

После внедрения модели в производство достаточно сложно измерить её эффект на конечный уровень доходов при помощи АВ-тестирования, поскольку доходы компании зависят от очень многих факторов (сезонности, инфляции, популярных трендов) и сложно отделить их влияние от воздействия новой рекомендательной модели. Поэтому в профессиональной области рекомендательных систем принято измерять эффект от внедрения моделей на более локальной онлайн-метрике, которая более чувствительна к изменению содержания рекомендаций и одновременно с этим напрямую влияет на уровень доходов [Jannach, Jugovac, 2019]. Так как полезные свойства модели первоочередно должны повлиять именно на онлайн-метрики, я считаю важным именно на их примерах обговорить особо важные свойства рекомендательных систем для сервисов электронной коммерции. В ближайших абзацах я буду представлять разные виды онлайн-метрик и указывать, какие полезные качества рекомендательных систем они должны измерять. Таким образом, на мой взгляд, обзор полезных свойств рекомендательных систем для бизнеса получится более конкретным и содержательным.

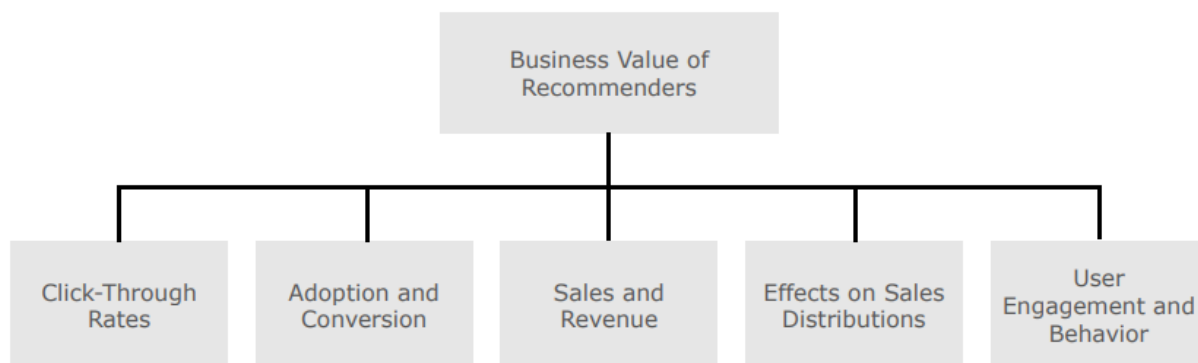


Диаграмма 1. Основные подходы к онлайн-оценке эффекта от внедрения рекомендательных систем. Источник: [Jannach, Jugovac, 2019]

Согласно методике статьи [Jannach, Jugovac, 2019], существует 5 основных подходов к оценке экономической выгоды от внедрения алгоритмов рекомендаций. (см. Диаграмма 1)

1) Первый из них называется кликабельностью (Click-through rate, CTR) и обозначает то, насколько часто пользователь кликает на рекомендацию. Этот метод более распространен на новостных и контент-платформах (Forbes News, YouTube), однако редко применяется и на торговой онлайн-платформе. Например, в статье [Katukuri et al., 2015] описано, что Ebay использовал именно кликабельность из-за её простоты для измерения релевантности рекомендации в окне «С этим товаром покупают также». Здесь основным свойством рекомендаций является релевантность, и определяется как факт клика на предложенный товар.

2) Так как покупатель далеко не всегда покупает товар после клика по нему, чаще используют другой измеритель релевантности – уровень принятия и конверсии, которые считают случаи совершения именно целевого действия. Например, в другом исследовании про Ebay [Chen, Canny, 2011] авторы считали клик релевантным только в том случае, если после него последовала покупка. В таком подходе релевантность рекомендаций также является основным свойством и определяется уже гораздо строже.

3) Когда измеряется непосредственно рост продаж от введения новой рекомендательной системы, он считается не только по продажам рекомендованных товаров. В работе [Dias et al., 2008] про рекомендательные алгоритмы в продуктовом интернет-магазине авторы пришли к выводу, что иногда рекомендации приносят гораздо больший доход вовсе не тогда, когда их покупают, а когда с помощью них покупатель узнает о новой для себя категории товаров. В таком случае важным свойством рекомендации является новизна и неожиданность в том смысле, что пользователь раньше не взаимодействовал с этим товаром, предложенный товар значительно отличается от его предыдущих покупок и одновременно с этим является релевантным с точки зрения кликабельности.

4) Также, рекомендательные системы могут изменять распределение покупаемых товаров, например, в сторону более дорогих и рентабельных для сервиса товаров или в сторону менее популярных товаров. Более широкое внимание в научной литературе получил второй случай. Когда пользователям предлагаются менее популярные товары, это говорит о том, что эти предложения стали более персонализированными. По результатам исследования Netflix [Gomez-Uribe, Hunt, 2015], одним из последствий персонализации является то, что более активно продаются нишевые малоизвестные товары. В частности, нишевые малоизвестные фильмы часто не обладают бюджетом для рекламы, поэтому персонализированные рекомендации стриминговой платформы являются критической механизмом, обеспечивающим просмотры этим нишевым фильмам. Этот механизм может привлекать и в онлайн-магазин большее число продавцов. Для оценки этого показателя тот же Netflix ввел онлайн-метрику «объем эффективного каталога», которые обозначает ту долю фильмов в каталоге, которые попадаетея пользователям в рекомендациях. Чем больше этот показатель, тем рекомендации более индивидуальные и покрывают большую долю каталога. Таким образом, важным свойством рекомендательной системы является персонализация, поскольку в данном случае она обеспечивает продажи более нишевых товаров из каталога.

5) Вторым последствием персонализации рекомендаций, согласно исследованию Netflix [Gomez-Uribe, Hunt, 2015], является большая вовлеченность пользователей и их удержание. Как это было описано во введении, благодаря персонализации Netflix смог снизить отток пользователей и экономить на этом ежегодно около \$1 миллиарда. Также, для поддержания вовлеченности необходимо разнообразие рекомендаций [Vargas, Castells, 2011]. Если все время пользователю будут предлагаться купить одни и те же товары, они ему наскучат, и он перестанет пользоваться сервисом. В частности, в электронной коммерции можно считать удержание пользователей – как долю пользователей, которые делают покупки хотя бы раз в месяц. Таким образом, важными свойствами для удержания пользователей являются персонализация и разнообразие выдачи рекомендаций.

Суммируя вышесказанное в двух последних частях, можно сделать вывод о том, что основными свойствами рекомендательных систем, которые могут оказаться важными и полезными для бизнеса электронной коммерции являются релевантность, персонализация, разнообразие и новизна рекомендаций. Конкретизация этих свойств и выбор соответствующих оффлайн-метрик для конкретного случая каталога товаров для животных будут представлены в третьей главе.

### 1.3 Основные типы моделей рекомендательных систем.

В данной части я рассмотрю основные виды моделей рекомендательных систем, разбив их на три группы в зависимости от того, как они учитывают время в пользовательской истории для ранжирования товаров.

1) Первая группа методов никак не учитывает время в пользовательской истории. Модели «не видят», какие товары были куплены только что, а какие были приобретены десятки действий назад.

- Эвристические методы

Эвристические методы [Sharma et al., 2023] как правило основаны на популярности покупаемых товаров и являются абсолютно не персонализированными. В частности, к ним относят рекомендации самых популярных товаров в каталоге или самых популярных внутри определенных категорий, посчитанных за какое-то время. Также, существуют системы, которые предлагают клиентам их же самые популярные товары. Однако этот метод крайне редко рассматривается в литературе, так как он абсолютно не способен предлагать товары, которые не фигурировали в пользовательской истории. Так как в основном используются датасеты, в которых пользователи редко покупают товары заново (игры, фильмы, бытовая техника), данный метод не представляет особого интереса.

- Коллаборативная фильтрация

Методы коллаборативной фильтрации используют исторические действия пользователей для моделирования близости между клиентами и товарами. Под словами «исторические действия» я имею виды элементы обратной связи (также фидбек), которые бывают явные (рейтинги) или не явные (клики, просмотры, покупки). Вторые фидбеки называют не


























				
				
				
				
				
				

Рисунок 2. Таблица обратной связи пользователей. Источник: Википедия

явными, потому что покупка товара вовсе не означает, что клиенту понравился этот товар. Возможно, он его больше никогда не купит. В данном случае пользователи похожи друг на

друга, если они оставляли похожий фидбек на одни и те же товары. И тогда модель рекомендует пользователю товары, которые понравились похожим на него клиентам. (User-based collaborative filtering, см. Рисунок 2). Аналогично, товары считаются близкими, если пользователи оставляют на них похожие фидбеки. И тогда модель рекомендует пользователю товар, который похож на только что понравившийся (Item-based Collaborative Filtering).

- Контентно-ориентированные рекомендации

Третий вид рекомендаций основан на представлении товаров в виде числовых векторов (эмбедингах) и рекомендации товаров, похожих по смыслу на те предметы, которые понравились пользователю [Pazzani, Billsus, 2007]. Так, векторное представление товаров строится при помощи категорий товара, его описания и изображения.

- Гибридные модели

Гибридные модели используют оба предыдущих варианта для усиления слабых сторон друг друга. Например, иногда используют обе модели независимо друг от друга, каждая модель предсказывает скор для каждой пары пользователь-товар, а итоговый скор для каждой пары берут как линейную комбинацию двух независимых оценок [Burke, 2002]. Второй вариант использует таблицу с историческими рейтингами (см. Рисунок 2) как дополнительные характеристики товаров дополнительно к основным, а потом применяется поиск похожего товара в новом пространстве большей размерности [Basu et. al., 1998]. Третий популярный вариант использует одну из моделей для отбора кандидатов, а другую для ранжирования этих кандидатов, которых уже гораздо меньше, чем наблюдений во всем каталоге. Так, например, устроена модель рекомендаций ресторанов EntreeC [Burke, 1999].

Отдельно здесь можно выделить нейросетевые модели. Благодаря их выдающейся способности представлять изображения и текст в виде векторных эмбедингов с сохранением их содержания, ряд гибридных рекомендательных алгоритмов используют именно нейронные сети для извлечения характеристик из контентных элементов товаров [Wang et. al., 2015; Wang et. al., 2017]

- Матричная факторизация

Методы матричной факторизации основаны на представлении пользователей и товаров в некотором латентном пространстве, получив для каждого пользователя  $u$  его векторное представление  $p_u$  и для каждого товара  $i$  его векторное представление  $q_i$  таким образом, чтобы их скалярное представление было приблизительно равно  $r_{ui}$  – рейтингу товара  $i$

пользователем  $u$  ( $r_{ui} \approx q_i^T p_u$ ). Прогноз этих рейтингов для всех пар (пользователь - товар) выставляется также относительно их скалярного представления ( $\widehat{r}_{ui} \approx q_i^T p_u$ ). Составление этих латентных векторов может быть произведен с помощью сингулярного разложения матрицы рейтингов (см. Рисунок 2) или оптимизацией соответствующей функции потерь. Двумя наиболее популярными, не нейросетевыми методами матричной факторизации являются Altering Least Squares (ALS) и Bayesian Personalized Ranking (BPR), которые отличаются именно функцией потерь.

Также и в области матричной факторизации создаются модели нейронных сетей, более точно представляющие предпочтения пользователей в виде векторов латентного пространства. В частности, модель NewMF [Xiangnan. et al, 2017] использует архитектуру автоэнкодеров для создания эмбеддингов пользователей и многослойный перцептрон для предсказания их рейтингов. Как я раньше упоминал во введении, в исследовании [Ferrari et. al., 2019] данная нейросетевая модель был в числе тех, кто показал более плохие результаты на ряде датасетов по сравнению с более примитивными моделями, в частности, по сравнению с коллаборативной фильтрацией (Item-based).

2) Вторая группа методов учитывает точное время, когда были произведены действия. В частности, модель TimeSVD [Koren, 2010] также создает эмбеддинги в латентном пространстве, однако в этом методе они зависят от времени  $p_u(t)$  и позволяют моделировать изменяющиеся во времени предпочтения пользователей. Такой метод позволяет фиксировать как долгосрочные изменения в предпочтениях (например, устаревание фильма в последние десятилетия), так и краткосрочные (например, большой интерес к кофе по утрам), что увеличивает точность рекомендаций на исторических данных.

3) Модели третьей группы «не видят» точного времени, когда пользователи совершают действия, но знают о их порядке по отношению друг к другу, поэтому они называются последовательными. Как уже было проиллюстрировано во введении, в рекомендациях для повышения релевантности предложенных товаров необходимо учитывать последовательную структуру пользовательской истории.

- Модели, основанные на марковских цепях

Модели марковских цепей моделируют матрицу перехода от одного товара к другому на основе историй покупок. Это предоставляет данной модели важный «контекст» для прогнозирования следующей покупки, позволяя ей превосходить по основным метрикам



точности непоследовательные модели [He et. al., 2017]. Однако такие алгоритмы обладают «короткой памятью», «запоминая» только последние несколько товаров.

- Рекуррентные нейронные сети

Модели рекомендаций, основанные на рекуррентных нейронных сетях, «учатся» запоминать всю историю пользователя, фиксируя её внутри одного вектора, который называется состоянием модели. Благодаря этому, модель получает одновременно более полное представление о пользовательской истории, однако и более размытое, так как внутри одного вектора (текущего состояния) модель уместит и важную, и второстепенную информацию о пользовательской истории. [Hidasi et. al., 2016].

В следующем пункте этой главы будет проведен обзор трансформерной модели рекомендаций SASRec, которая решает ключевые проблемы двух предыдущих последовательных моделей.

#### 1.4 Трансформерная модель последовательных рекомендаций.

Архитектура нейронной сети трансформер [Vaswani A. et al., 2017] 8 лет назад сделала большой прорыв в области обработки естественного языка. Основной частью её архитектуры стало «внимание» - свойство, позволяющее определять важность разных частей предложения и обогащать смысл слов их контекстом. Авторы статьи [Kang, McAuley, 2018] решили, что это свойство может быть очень полезным и в области последовательных рекомендаций. В предыдущей части этой главы было указано на то, что модели марковских цепей вынуждены уделять все «внимание» только на последние товары в пользовательской истории, в то время как модели рекуррентных нейронных сетей «запоминают» всю пользовательскую историю, не выделяя в ней важные и не важные части. Модель, предложенная в данной статье, использует свойство «самовнимание» трансформера, чтобы учить «запоминать» модель только важные части в истории покупок. Таким образом, модель может находить краткосрочные и долгосрочные паттерны в пользовательских покупках, добиваясь более точных результатов на основных оффлайн-метриках (HitRate, NDCG), по сравнению с моделями, не обладающими вниманием.

Несмотря на то, что моя работа не вносит технической новизны в модель, я бы хотел кратко описать архитектуру модели SASRec для лучшего понимания её свойств (см. Рисунок 3). Данная реализация модели была взята в библиотеке MTC rectools. (см. Рисунок 4) Нам дана последовательность пользовательской истории юзера  $u$  – последовательность товаров

$S^u = (S_1^u, S_2^u, S_3^u, \dots, S_{|S^u|}^u)$ . Модель принимает на вход все элементы истории, кроме последнего,  $(S_1^u, S_2^u, \dots, S_{|S^u|-1}^u)$ . А на выходе модель будет сверять свою выдачу с правильными ответами  $(S_2^u, S_3^u, \dots, S_{|S^u|}^u)$ , следующими сделанными покупками. Далее, полученные последовательности укорачиваются до последних  $n$  товаров или, наоборот, к началу добавляются паддинги («пустые» товары), чтобы все пользовательские истории  $S^u = (S_1^u, S_2^u, \dots, S_n^u)$  были одинаковой длины  $n$  (в нашем случае  $n=3$ ). Эмбединг каждого товара в истории является по координатной суммой нескольких эмбедингов: эмбединг позиции товара в истории, эмбединг  $id$  этого товара в каталоге и эмбединги категориальных переменных товара. ( $M$ - матрица эмбедингов  $id$  всех товаров в каталоге,  $M_{S_t^u}$  – эмбединг  $id$   $t$ -го товара в истории пользователя  $u$ ). Эти эмбединги будут оптимизироваться в процессе обучения модели. Далее, эти эмбединги передаются в слои трансформера, где они обогащаются эмбедингами других предшествующих товаров в этой истории. Например, как показано на картинке итоговые эмбединг картриджа содержит в себе информацию, что до него купили мышь, а еще раньше компьютер. Таким образом, мы получили итоговые эмбединги для каждой позиции в истории  $f^u = (f_1^u, f_2^u, f_3^u)$ . Расчетная вероятность  $\widehat{p}_{tq}$  того, что после  $t$ -го товара в истории пойдет товар  $q$  из каталога, равна

$$\widehat{p}_{tq} = \frac{\exp(\langle f_t^u, q \rangle)}{\sum_h \exp(\langle f_t^u, h \rangle)}, \quad (1)$$

где суммирование идет по  $h$  – каждому товару из каталога,

$f_t^u$  – эмбединг  $t$ -ой позиции в истории пользователя  $u$ .

Для каждого товара в истории функция потерь считается следующим образом.

$$l_t = -\log(\widehat{p}_{ts_{t+1}^u}) - \sum_{h \in A_{t+1}^u} \log(1 - \widehat{p}_{th}^u), \quad (2)$$

где  $A_{t+1}^u$  – некоторое подмножество товаров из каталога без  $s_{t+1}^u$ ,

$\widehat{p}_{tq}$  – вероятность того, что после  $t$ -ой позиции в истории пойдет товар  $q$ ,

$s_{t+1}^u$  – товар, который находится на позиции  $t+1$  в истории пользователя  $u$ .

Если модель предсказывает следующий товар идеально, то вероятность перехода к истинно следующему товару  $\widehat{p}_{ts_{t+1}^u} = 1$ , а ко всем остальным  $\widehat{p}_{th}^u = 0$ . Тогда  $l_t = 0$  и  $L = 0$ .

Множество  $A_{t+1}^u$  обозначает набор «негативных» товаров, с которыми пользователь  $u$  намеренно не провзаимодействовал в момент  $t+1$ . Как правильно, оно включает либо весь каталог без  $s_{t+1}^u$ , либо некоторое его подмножество (зависит от целей обучения). Понять,

какие товары из каталога пользователь не видел, а какие намеренно пропустил достаточно сложная задача. Общая функция потерь считается суммированием всех функций потерь  $l_t$  по каждому товару внутри пользовательской истории и по всем пользователям, не считая паддинги:

$$L = \sum_u \sum_{t=1}^n l_t$$

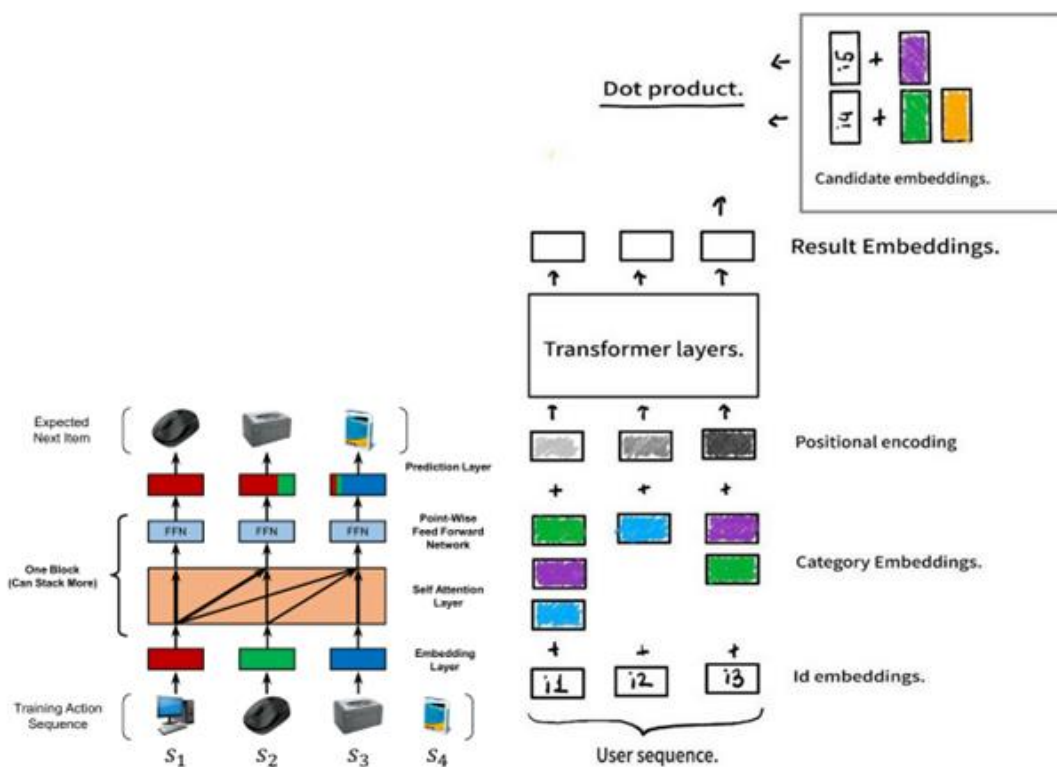


Рисунок 3. Рисунок. Устройство внимания SASRec.  
Источник: [Kang, McAuley, 2018]

Рисунок 4. Реализация модели SASRec от RecTools. Источник:  
github-репозиторий библиотеки RecTools

Таким образом, из обзора литературы можно сделать следующие выводы:

- 1) Конечная цель рекомендательных систем в электронной коммерции – это увеличение продаж, однако более локальная цель может быть иной, например, улучшение пользовательского опыта, и зависит от особенностей конкретного сервиса.
- 2) В силу отсутствия достаточной информации про бизнес онлайн-магазина «PetCo» я не могу ставить конкретную цель и онлайн-метрику для его рекомендательной системы, однако могу проанализировать потенциально важные свойства интересующей модели для её будущего внедрения на этой платформе.
- 3) Основными потенциально полезными свойствами рекомендательных систем являются релевантность, персонализация, разнообразие и новизна.

- 4) Последовательные модели, благодаря знанию об историческом «контексте» новой покупки, обычно делают более точные рекомендации на исторических данных, чем статические модели.
- 5) Трансформерная модель последовательных рекомендаций SASRec, благодаря поиску паттернов в пользовательских историях, обычно дает более точные рекомендации на исторических данных по сравнению с другими моделями, не обладающие свойством внимания.
- 6) Более продвинутые модели рекомендательных систем могут давать менее точные рекомендации на менее популярных наборах данных, поэтому их точность также необходимо проверять на данных конкретного бизнеса перед внедрением на платформу

## Глава 2. Построение трансформерной модели рекомендаций для её первичного анализа.

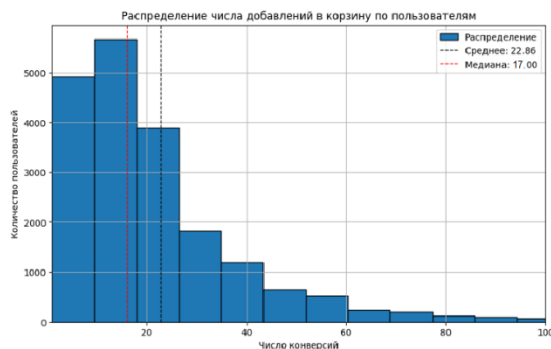
Данная глава посвящена построению трансформерной модели последовательных рекомендаций SASRec для более глубокого анализа её рекомендаций в следующей главе. Первоначально я провожу разведывательный анализ данных для их последующей обработки. Затем я осуществляю выбор конфигурации модели и её обучение. В завершении этой главы я проведу тестирование обученной модели с использованием других бейзлайн-моделей и оффлайн-метрик в соответствии с выделенными ранее потенциально важными свойствами рекомендательной модели. Тестирование будет сопровождено первичным анализом результатов, которые будут существенно дополнены в следующей главе.

### 2.1 Разведывательный анализ данных и их обработка.

Данные мне были предоставлены с сайта американского онлайн-магазина товаров для домашних животных PetCo. На данном сайте преимущественно крупные сторонние производители размещают свои товары для их розничной покупки. Данные представляют собой два датасета: первый содержит поведенческие действия покупателей за период с 06.06.2024 по 02.12.2024 (клики, поиск по сайту, просмотр товаров, добавления в корзину, покупки, просмотр предложенных рекомендаций); второй является каталогом и содержит описания всех товаров.

Конечное действие покупателя, которое представляет основной интерес для бизнеса, это покупка, однако покупки совершаются клиентами только в корзинах. Из-за покупок товаров группами (в среднем, по 4 товара), товары внутри корзины покупаются одновременно, поэтому достаточно сложно отследить последовательность, в которой бы клиент покупал товары отдельно. Возможный вариант решения этой проблемы заключается в восстановлении временной последовательности, в которой покупаемые товары добавляются в корзину. Поэтому в данном случае для технической простоты в качестве целевого действия были взяты добавления в корзину. Помимо простоты, добавление в корзину является хорошим приближенным показателем последующей покупки (в данном датасете около 60% добавлений в корзину завершается покупкой товара). (Далее иногда вместо слов «добавления в корзину» будет использовано слово конверсия, которое в данной работе будет означать тоже самое). За весь рассматриваемый период было совершено около 451 тысячи конверсий около 20 тысячами пользователей. Распределение числа конверсий по пользователям представлено на Гистограмма 1. Как на ней показано, в среднем люди

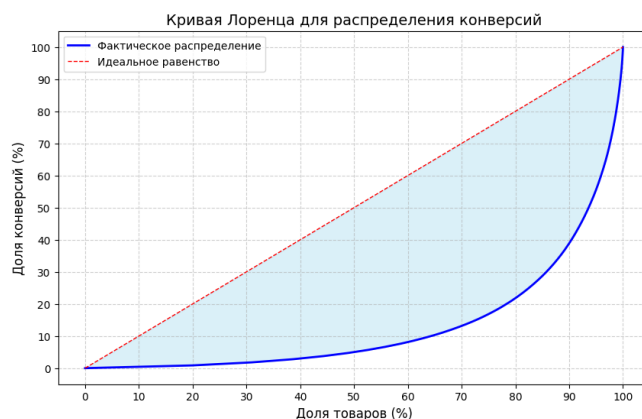
совершали 23 конверсии в своей истории, и половина всех клиентов сделали до 16



Гистограмма 1. Распределение числа конверсий по пользователям.  
Источник: расчета автора на основе [данных](#).

конверсий. В среднем, конверсии каждого пользователя приходятся на 15 товаров, что составляет  $5 \cdot 10^{-4}$  от всего каталога. В изначальной статье про SASRec [Kang, McAuley, 2018] похожие по плотности датасеты называют разреженными. Также стоит сказать и о том, что около половины всех конверсий (52%) приходится на повторные товары в истории. Вместе с этим, около 75% покупателей делают повторные конверсии (то есть хотя бы раз добавляют в корзину один и тот же товар 2 раза и больше). Таким образом, данный датасет является разреженным, и половина конверсий в нем приходится на повторяющиеся в истории товары.

Что касается каталога товаров, в нём содержится около 30 тысяч товаров, из которых пользователями добавляется в корзину около 18 тысяч, то есть 60%. В среднем, на один товар приходится 24 конверсии. Если обратить внимание на Гистограмма 2, можно заметить большую неравномерность в распределение конверсий по товарам (на 20% товаров приходится 80% конверсий). Это кажется довольно естественным, так как товары рекламируются также очень неравномерно. Так как используемая реализация модели использует только категориальные переменные, то название и изображение товара не могут учтены в модели. Из каталога для окончательного датасета я оставил только следующие



Гистограмма 2. Распределение конверсий по товарам.  
Источник: расчета автора на основе [данных](#).

категориальные переменные: категория, способ доставки, бренд и тип домашнего питомца. Пример такого описания одного из товаров каталога представлен на Рисунок 5.

<b>item_id</b>	100038
<b>Category</b>	[lovebird-cages, cockatiel-cages, cages-for-bi...
<b>Delivery</b>	[One Time Delivery]
<b>Brand</b>	[Prevue Pet Products]
<b>Pet</b>	[Bird]

Рисунок 5. Пример категориальных признаков товара из каталога. Источник: расчеты автора на основе [данных](#).

## 2.2 Обучение последовательной рекомендательной модели.

Все пользовательские истории были разделены на три выборки: последний элемент в каждой истории был отделен в тестовую выборку, предпоследний элемент был отправлен в валидационную выборку и все предшествующие элементы в историях составили обучающую выборку. На обучающей выборке модель обучалась, валидационная выборка использовалась мной для выбора лучшей конфигураций модели, а на тестовой выборке модель сравнивалась с моделями другого типа.

Многие гиперпараметры я выбрал такими, какие они использовались в исходной статье про SASRec и предлагались по умолчанию в библиотеке rectools, в частности, число блоков в self-attention слое (2), число голов в self-attention слое (4), доля dropout (0.2), шаг обучения (0.001). Однако по умолчанию в библиотеке rectools размерность латентного пространства эмбедингов больше, чем в изначальной статье SASRec, что связано с необходимостью хранить больше информации о товарах из-за включения в модель категориальных переменных. Поэтому мною было выбрано дефолтное значение этого гиперпараметра (256), используемое в rectools. Аналогично изначальной статье SASRec, мною была взята максимальная длина пользовательской истории примерно в 4 раза больше её средней длины, а именно 100. Все истории длиннее 100 элементов, как я описывал во введении, обрезались по последним 100 элементам. Модель могла включать либо оба вида эмбедингов (и вектора категориальных переменных, и вектора id товаров), либо только одни. Мною в модель были включены оба вида эмбедингов, так как модель сама способна извлекать нужную информацию из подаваемой на вход.

Единственная часть конфигурации модели, которые я менял, - это функция потерь. Помимо функции потерь softmax, которую я описывал в обзоре литературы, модель могла поддерживать ещё две - BCE и gBCE [Petrov, Macdonald, 2023]. В данном случае в softmax в качестве негативных товаров были рассмотрены все остальные товары в каталоге, с которыми пользователь в определенные моменты не взаимодействовал. Благодаря имеющимся вычислительным мощностям, в таком случае обучение заняло несколько минут. На валидационной выборке незначительно лучше себя показала модель с функцией потерь softmax, поэтому при тестировании использовалась именно её конфигурация. Стоит пояснить слова «лучше на валидационной выборке». При валидации модели я использовал метрику HitRate@5, она обозначает долю случаев, когда пользователь получил удачную рекомендацию в списке из 5 предложенных товаров. Например, модель порекомендовала пользователю 5 товаров: 001, 002, 003, 004, 005. Следующий купленный им товар был 002, значит, модель правильно предсказала следующее действие пользователя, и одна из 5 рекомендаций оказалась удачной. Так как исходя из функции потерь модель учится угадывать следующий купленный товар, то использовать для валидации именно эту метрику (долю успешных наборов рекомендаций) кажется мне наиболее разумным. Также поступили и авторы в оригинальной статье. На График 1 и График 2 представлен процесс обучения модели. В течение последних 10 эпох из 50 модель перестала улучшать на валидационной выборке свою функцию потерь и метрику точности, поэтому можно говорить об окончании обучения. Теперь модель готова к тестированию и сравнению с другими бейзлайн-моделями.

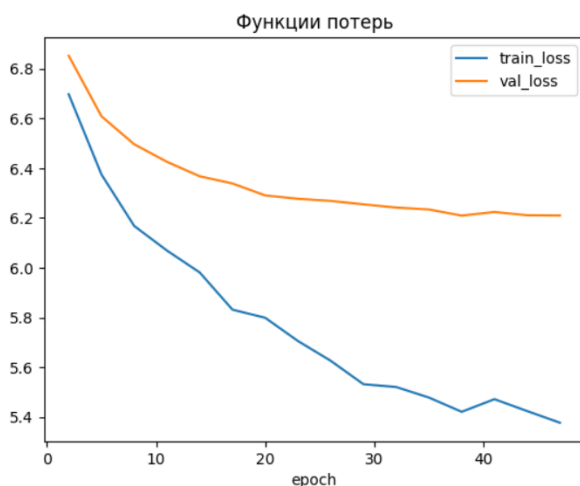


График 1. Изменение значений функций потерь во время обучения. Источник: расчеты автора на основе [данных](#)

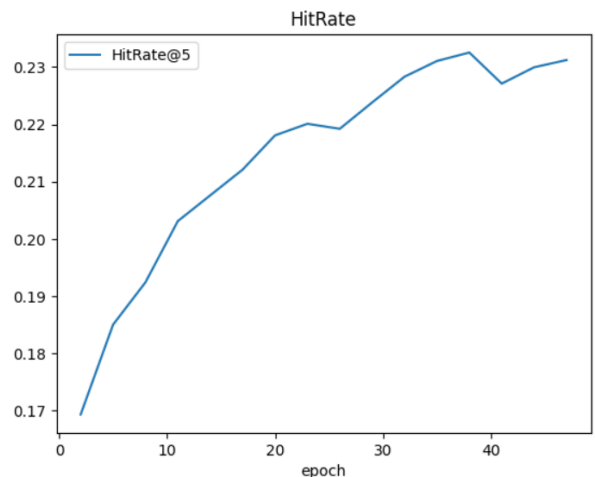


График 2. Изменение точности модели на валидационной выборке при обучении. Источник: расчеты автора на основе [данных](#)



## 2.3 Тестирование модели и первичный анализ её рекомендаций.

В качестве моделей-бейзлайнов были выбраны 3 эвристические модели: модель PopGlobal рекомендует всем пользователям одни и те же самые популярные товары во всем датасете; PopCat рекомендует всем пользователям одни и те же самые популярные товары по категориям (здесь под категорией имеются товары для одного вида домашних питомцев, то есть список рекомендаций состоит из самого популярного товара для собак, для кошек и так далее); PopUser рекомендует каждому пользователю список его самых популярных товаров.

Также, в качестве бейзланов была взята одна модель, основанная на коллаборативной фильтрации ItemKNN Colaborative Filtering (CF), и две модели использующие латентные векторные представления пользователей/товаров: Altering Least Squares (ALS) и Bayesian Personalized Ranking (BPR). В силу высоких временных затрат на обучение более сложных моделей они не были представлены для сравнения основной модели.

Дальнейшее повествование я бы хотел провести следующим образом. Так как выбор метрик для тестирования тесно связан с выбранными моделями и особенностями датасета, на мой взгляд, более разумно объединить описание выбранных метрик с анализом их конкретных значений, которые показывают модели. Для структурирования анализа метрики будут разбиты по группам относительно того, как свойства моделей они проверяют.

В обзоре литературы были выявлены следующие потенциально полезные свойства рекомендательной системы: релевантность, персонализация, разнообразие и новизна.

### а) Релевантность

Для оценки оффлайн-релевантности, то есть на исторических данных были выбраны наиболее общепринятые метрики HitRate@5 и NDCG@5 (Saifudin I, Widiyaningtyas, 2024). Так как в тестовой выборке для каждого пользователя только одно наблюдение, то HitRate тождественно равен метрике Recall и пропорционален метрике Precision. Метрика NDCG (Normalized Discounted Cumulative Gain), помимо появления удачной рекомендации в списке, также оценивает на каком месте в списке она находится и штрафует, если она находится не на первом месте. Для отдельного пользователя Discounted Cumulative Gain равен

$$DCG@5_{user} = \sum_{k=1}^5 \frac{rel(k)}{\log_2(k+1)},$$

где  $k$  – ранг товара в списке рекомендаций и  $rel(k)$  – индикатор удачной рекомендации.

Затем эта метрика нормируется (нормирующий множитель в данном случае равен 1) и усредняется по всеми пользователями. Если NDCG равен HitRate в данном случае, то это значит, что во всех списках рекомендаций удачные рекомендации встречаются только на первых местах в списках.

Таблица 1. Метрики точности рекомендаций. Источник: [расчета автора на основе данных](#).

Модель	HitRate@5	NDCG@5
a) PopGlobal	0,01	0,007
b) PopCat	0,008	0,005
c) PopUser	0,21	0,142
d) BPR	0,061	0,039
e) ALS	0,126	0,085
f) ItemKNN CF	0,164	0,113
g) SASRec	0,218	0,15
Улучшение с a-c	3,8%	5,6%
Улучшение с d-f	25,7%	32,7%

Как можно заметить на Таблица 1, модель SASRec по обоим метрикам значимо точнее предсказывает следующий добавленный в корзину товар по сравнению с другими моделями, то есть обладает наивысшей оффлайн-релевантностью. Примечательно, что HitRate модели PopUser всего на 0.8 п.п ниже, это подчеркивает тот факт, что в данном датасете клиенты часто добавляют в корзину товары повторно и их также релевантно рекомендовать. В этой таблице и далее цветные числа обозначают статистически значимое различие на уровне значимости 1%.

#### б) Персонализация

Персонализация обозначает, что рекомендованные товары, во-первых, отличаются у разных пользователей и, во-вторых, не являются глобально популярными. Как уже было рассмотрено в обзоре литературы, Netflix для первой использует «объем эффективного каталога» [Gomez-Uribe, Hunt, 2015], поэтому для проверки персонализации в первом смысле я использую аналогичную общепринятую метрику Coverage@5 – доля товаров из каталога, которые встречаются в рекомендациях. Исходя из результатов в Таблица 2, SASRec является второй по данной метрике, уступая по персонализации только PopUser. Это кажется естественным, что наиболее индивидуальным набором рекомендаций для пользователя являются его же любимые товары.

Метрика Average Recommendation Popularity (ARP) – считает среднюю историческую популярность рекомендованных товаров (Считает, какая доля от всех

исторических конверсий в датасете приходится в среднем на рекомендованный товар). Чем больше эта метрика, тем рекомендованный товар чаще встречался в историях пользователей и, соответственно, является более популярным и менее индивидуальным. Согласно Таблица 2, рекомендации SASRec являются наименее популярными в этой метрики, а значит, и наиболее персонализированными. Здесь также стоит отметить, что рекомендации модели PopUser являются вторыми по редкости. Это говорит о том, что наиболее популярные товары для отдельного пользователя являются глобально редкими. Этот важный факт стоит держать в уме для последующих размышлений.

Таблица 2. Оффлайн-метрики персонализации рекомендаций.  
Источник: расчета автора на основе [данных](#)

Модель	ARP@5	Novelty@5	Coverage@5
a) PopGlobal	2,3E-03	5,7	0,00
b) PopCat	1,7E-03	6,2	0,00
c) PopUser	4,1E-04	8,6	0,37
d) BPR	4,3E-04	7,9	0,06
e) ALS	7,1E-04	7,3	0,05
f) ItemKNN CF	6,3E-04	7,7	0,13
g) SASRec	3,6E-04	8,5	0,27
Улучшение с a-c	-12,0%	-1,6%	-26,1%
Улучшение с d-f	-15,3%	6,8%	106,2%

Более подробно стоит сказать про метрику Novelty@5, которая часто используется для подсчета новизны рекомендованных товаров.

$$Novelty_{user} = \frac{1}{|R_{user}|} \sum_{i \in R_{user}} -\log \frac{n_{users(i)}}{n_{users}},$$

где  $\frac{n_{users(i)}}{n_{users}}$  обозначает долю людей, которые раньше взаимодействовали с товаром  $i$ ,

$R_{user}$  обозначает набор товаров, рекомендованных для пользователя  $user$ .

Логика метрики Novelty: чем менее популярен рекомендованный товар (меньше людей с ним взаимодействовали), тем более вероятно, что он будет новым для пользователя, и Novelty тогда будет больше. Как мы заметили раньше, популярные товары каждого пользователя являются глобально редкими, и если их рекомендовать, то метрика Novelty у таких рекомендаций будет высокой (в Таблица 2 модель PopUser обладает наивысшей Novelty), но по сути эти товары будут очень старыми для пользователя. Поэтому я предлагаю рассматривать эту метрику с другой точки зрения, в качестве ещё одной метрики исторической непопулярности рекомендованных товаров, посчитанной только не по числу конверсий (как ARP), а по числу пользователей. Тогда: чем больше Novelty у модели, тем менее популярны рекомендованные ею товары и тем более они персонализированы. С этой

точки зрения, модель SASRec незначимо проигрывает в персонализации модели PopUser, но выигрывает у всех остальных моделей.

Так, можно сделать вывод о том, что модель SASRec является такой же очень высоко персонализированной, как и PopUser (по всем метрикам занимают первые два места), и лучше по этому свойству, чем все остальные модели.

### с) Разнообразие

Для подсчёта разнообразия товаров внутри списка рекомендаций я использовал одну из наиболее общепринятых метрик Intra-List Diversity (ILD) (среднее попарное косинусное расстояние между товарами в одном списке рекомендаций) [Du et al., 2021]:

$$ILD_{user} = \frac{2}{|R_{user}|(|R_{user}| - 1)} \sum_{\forall (1 \leq i \leq j \leq |R_{user}|)} (1 - \text{sim}(R_{user}[i], R_{user}[j]))$$

$\text{sim}(R_{user}[i], R_{user}[j])$  – косинусная близость между эмбедингами товаров с рангами  $i$  и  $j$ ,  $R_{user}$  обозначает набор товаров, рекомендованных для пользователя  $user$

В данном случае я использовал следующий алгоритм для расчета схожести между товарами:

- Эмбединги текстовых описаний товаров были получены при помощи языковой модели `nfhakim/topic-clustering-v1`<sup>7</sup>
- Эмбединги изображений товаров были получены при помощи модели `mobilenetv3_small_100`<sup>8</sup>
- Отдельно посчитана косинусная близость между текстовыми эмбедингами товаров
- Отдельно посчитана косинусная близость между эмбедингами картинок товаров
- Итоговая близость между парой товаров была получена как среднеарифметическое из близостей их текстов и близостей их картинок.

Согласно результатам в Таблица 3, SASRec предлагает почти самые однородные товары внутри одного списка рекомендаций (Static\_Diversity). По сравнению с SASRec, разнообразие ниже только у ALS. Однако более всего важно разнообразие рекомендаций не столько внутри одного списка, сколько на протяжении некоторого времени. Пользователю могут наскучить рекомендации, если ему постоянно будут попадаться похожие товары.

<sup>7</sup> <https://huggingface.co/nfhakim/topic-clustering-v1>

<sup>8</sup> [https://huggingface.co/timm/mobilenetv3\\_small\\_100.lamb\\_in1k](https://huggingface.co/timm/mobilenetv3_small_100.lamb_in1k)

Поэтому важно посмотреть на разнообразие рекомендаций во времени. Для оценки этого я воспользовался методом, описанном в статье [Lathia et al, 2010]. Я построил рекомендации для валидационного выборки, элементы в которой предшествовали элементам в тестовой выборке. Получил для каждого пользователя два листа рекомендаций по 5 элементов:  $L_1$  (рекомендации для валидационного наблюдения) и  $L_2$  (рекомендации для тестового наблюдения). Разнообразие для каждого пользователя будет считаться, как доля элементов нового листа  $L_2$ , которых нет в  $L_1$ :

$$Temporal\ diversity(L_1, L_2) = \frac{|L_2 \setminus L_1|}{|L_2|}$$

Таблица 3. Разнообразие рекомендаций моделей.  
Источник: расчета автора на основе [данных](#)

Модель	Static_Diversity@5	Temporal_Diversity@5
a) PopGlobal	0,41	0
b) PopCat	0,43	0
c) PopUser	0,36	0,07
d) BPR	0,32	0,82
e) ALS	0,27	0,91
f) ItemKNN CF	0,3	0,84
g) SASRec	0,29	0,58
Улучшение с а-с	<b>-34,30%</b>	<b>728,57%</b>
Улучшение с d-f	<b>-10,40%</b>	<b>-36,26%</b>

Потом метрика усредняется по всем пользователям. Эта метрика показывает, как обновляются товары в списке рекомендаций во времени. Чем этот показатель больше, тем более разнообразные во времени пользователь получает рекомендации. Как можно увидеть по Таблица 3, эвристические модели абсолютно не обновляются со временем в то время, как вторая группа моделей (d-f) обновляются очень быстро. SASRec здесь занимает среднее положение, постепенно обновляя список своих рекомендаций: в новом списке рекомендаций SASRec в среднем 3 из 5 товаров не новые.

Так, можно сделать вывод о том, что отдельные списки рекомендаций SASRec довольно однородные внутри и товары в них похожи, однако со временем списки постепенно обновляются, несколько поддерживая интерес клиентов. Однако модели второй группы потенциально могут намного лучше удерживать интерес пользователей.

d) Новизна

Как было сказано в обзоре литературы, важным свойством рекомендательной системы является новизна, чтобы пользователь мог открывать для себя новые категории товаров. В таком случае эти рекомендации должны быть еще и релевантными. Такое свойство рекомендательных систем называется «serendipity» и входит в список слов самых сложных для перевода слов<sup>9</sup>. Иногда его переводят на русский язык как «интуитивная прозорливость». Определение такого свойства не имеет однозначной интерпретации и нет одной общепринятой метрики, которая бы его измеряла [Kotkov, 2016]. Три важнейших свойства этого явления релевантность, новизна и неожиданность. Как я указывал немного ранее, непопулярный глобально товар не обязан быть новым для пользователя, особенно в данном датасете. Поэтому и популярные варианты метрики serendipity, основанные на этой идее, в данном случае не подойдут. Исходя из этого, я взял вариант расчета метрики serendipity, предложенный в статье [Ziarani, Ravanmehr, 2021].

Таблица 4. Неожиданность рекомендаций. Источник: [расчета автора на основе данных](#)

Модель	Serendipity@5
a) PopGlobal	0,44
b) PopCat	0,46
c) PopUser	0,34
d) BPR	0,35
e) ALS	0,33
f) ItemKNN CF	0,34
g) SASRec	0,32
Улучшение с a-c	<b>-30,50%</b>
Улучшение с d-f	<b>-9,30%</b>

В этой метрике для каждого пользователя считается среднее попарное косинусное расстояние между рекомендованным товаром и товаром в пользовательской истории, а потом усредняется по всем пользователям. Чем больше этот показатель, тем в среднем рекомендованный товар более не похож на историю пользователя, то есть является неожиданным. В этой метрике не фигурирует релевантность, однако более подходящей метрики для данного случая в научной литературе я не нашел. Как можно увидеть на Таблица 4, модель SASRec рекомендует пользователю наиболее ожидаемые для него товары. Следовательно, SASRec меньше всех остальных моделей помогает пользователю находить новые для него категории товаров.

<sup>9</sup> <http://www.todaytranslations.com/blog/most-untranslatable-word/>

Таким образом, можно сделать следующие первичные выводы касательно рекомендаций модели SASRec. Рекомендации модели SASRec по сравнению с другими рассмотренными моделями являются:

- Наиболее точными
- Наиболее персонализированными
- Достаточно однообразными и в одном списке, и во времени
- Наиболее ожидаемыми

### **Глава 3. Комплексный анализ последовательных рекомендаций построенной трансформерной модели.**

После первичного анализа свойств модели SASRec, проведенного в прошлой главе, в этой главе будет осуществлен более многосторонний анализ рекомендаций данной модели. Сначала на основе реализованной кластеризации пользовательских историй я продемонстрирую свойства данной рекомендательной системы на репрезентативной подборке примеров. Так, я смогу наглядно дополнить некоторые из ранее полученных результатов и выдвинуть новые гипотезы, которые будут проверяться в двух последующих пунктах данной главы. Подтверждение данных гипотез даст полезную информацию о том, как рекомендации модели SASRec зависят от последовательной структуры пользовательской истории.

#### **3.1 Визуальный анализ рекомендаций модели.**

В данном пункте я продемонстрирую действие интересующего рекомендательного алгоритма на подборке нескольких реальных примеров пользовательских историй. Однако для репрезентативности полученных выводов недостаточно просто взять случайные истории из датасета. Если такие истории окажутся очень схожи (например, будут содержать только товары для кошек), то нельзя экстраполировать выводы относительно них на другие группы историй (например, содержащие товары для других животных). Поэтому я кластеризую пользовательские истории по их содержанию и отбираю в подборку историй по одному примеру из нескольких самых крупных кластеров. Таким образом, я смогу экстраполировать выводы на другие примеры в самых крупных кластерах. Следовательно, такие выводы будут вполне применимы и на весь датасет.

Процесс кластеризации я начал с построения полноценных эмбедингов товаров. Эмбединги категориальных признаков товаров имеют размерность 1732, именно столько всего уникальных категорий встречается в каталоге. Товар  $i$  имеет единицы в своем категориальном эмбеддинге в тех столбцах, каким категориям он принадлежит, и нули во всех остальных столбцах. Далее, итоговый эмбеддинг каждого товара был получен конкатенацией эмбедингов его текстового описания, его категорий и его картинки. Так, я получил таблицу эмбедингов товаров размерности  $30\,000 \times 3628$  ( $30\,000$  – число товаров в каталоге и  $3628$  – размерность эмбединга каждого товара). Далее, эмбеддинг каждой пользовательской истории был получен по координатным усреднением эмбедингов товаров, которые в нее входят. Так, я получил таблицу эмбедингов историй  $20\,000 \times 3628$



(20 000 – число пользователей в датасете). Так как части эмбедингов пользователей состоят из разных по природе частей, я провел их стандартизацию и потом спроектировал их на двумерную плоскость с помощью алгоритма t-SNE.

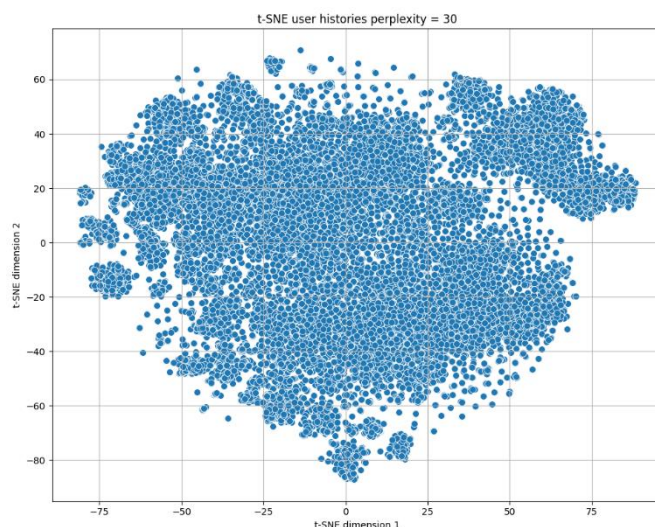


Рисунок 6. Спроектированные на плоскость эмбединги пользовательских историй. Источник: расчета автора на основе [данных](#)

На Рисунок 6 можно заметить, что на двумерной плоскости точки четко не разделены по кластерам. Это связано с тем, что пользовательские истории очень непрерывны по своему содержанию. То есть, помимо однородных историй, в датасете представлены и разнообразные истории, которые, например, содержат товары для разных видов питомцев. Так как кластеры не отделены четкими границами, использованный мною популярный алгоритм кластеризации DBSCAN при естественных гиперпараметрах объединял большую часть всех точек в один большой кластер. Чтобы разбить этот большой кластер на более

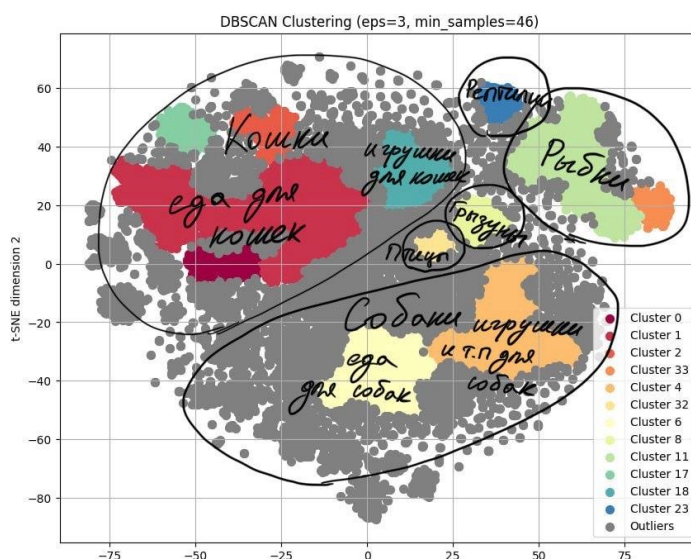


Figure 1. Карта отобранных кластеров пользовательских историй. Источник: расчета автора на основе [данных](#)

мелкие, я взял достаточно большим гиперпараметр `min_sample`, отвечающий за минимальное число соседей для образования кластера. В итоге, кластеризация при помощи алгоритма DBSCAN состоялась при гиперпараметрах `epsilon = 3` и `min_samples=46`. После отбора 10 самых крупных кластеров была получена следующая карта пользовательских историй (см. Figure 1). Подписями обозначены, какие товары в основном наполняют соответствующую историю. Для полноты картины к 10 самым крупным кластерам были также добавлены кластеры историй с товарами для птиц и грызунов, чтобы в примерах присутствовали товары для всех видов животных из каталога. Далее, из каждого кластера было взято по одной случайной истории и так получилась финальная репрезентативная выборка из 12 историй.

В целях избежания избытка рисунков и расфокусировки внимания между примерами, из 12 примеров я отобрал три примера, анализ которых я проведу более детально (Эти три примера расположены в Приложении). Данные 3 примера содержат все свойства моделей, которые я бы хотел наглядно показать. Выводы, полученные на основе этих трех примерах, применимы и к остальным 9 пользовательским историям, следовательно, и на весь датасет. Читающий также может в этом убедиться при обращении к [ноутбуку](#), где эти оставшиеся 9 историй расположены.

Также для избежания расфокусировки внимания я провел сравнение модели SASRec только с двумя из ранее протестированных моделей. В качестве двух таких моделей были взяты модели PopUser и BPR, так как они больше всех (согласно посчитанным метрикам) должны контрастировать с моделью SASRec по разнообразию и неожиданности/новизне.

Товары в пользовательской истории будут расположены в следующем порядке. Слева-сверху находятся самые старые товары (самые первые), которые были добавлены в корзину раньше всех. Время (и последовательность) идет в списке слева направо, сверху вниз. То есть самый правый товар в самой нижней строке был добавлен в корзину последним.

Пример 1:

Как можно заметить на (Приложение 1), в начале истории пользователя 806550 встречаются только аквариумы в то время, как в конце встречается только еда для собак. Рекомендации SASRec, в отличие от остальных моделей, представлены только едой для собак, которая похожа на конец пользовательской истории. Также стоит заметить, что модель BPR рекомендует корм для рыб (первый товар) и фильтр для воды (третий товар)

несмотря на то, что их не было в пользовательской истории. Эти товары можно назвать релевантными и неожиданными для пользователя.

#### Пример 2:

История пользователя 733322 (Приложение 2) очень разнообразная и содержит товары для рыбок (корм, фильтр для аквариума, декор для аквариума), для собак (игрушки, корм) и только в конце присутствует корм для птиц. Рекомендации SASRec содержат только корм для птиц, в отличие от других разнообразных рекомендаций. Здесь также BPR рекомендует товары релевантные, но ранее не встречающиеся в истории – жидкости для аквариума (товары 4 и 5).

#### Пример 3:

Как можно увидеть на Приложение 3, в истории пользователя 600927 встречаются разные категории товаров для собак (игрушки, корм, клетки, жевательные принадлежности). Рекомендации SASRec снова рекомендуют одинаковые жевательные принадлежности, все похожие на последние товары в истории. Здесь в рекомендациях BPR также фигурирует один необычный для истории товар, а именно собачий ошейник (второй). Он неожиданен только немного, так как в данной истории один раз встретился собачий поводок.

Как я сказал перед анализом приведенных примеров, из 12 полученных примеров я выбрал именно эти три, так как они наиболее сжато демонстрируют основные закономерности в рекомендациях моделей. Остальные 9 примеров их не противоречат, в частности, во всех случаях SASRec рекомендует товары более однородные и похожие, как минимум, на один из трех последних товаров в истории. Также, в двух из 9 оставшихся случаев BPR рекомендует товары, ранее не встречавшиеся в пользовательской истории (Морскую соль - пользователю 6626860 и лекарство для слуха для собак – пользователю 7089124). Поэтому проведенный визуальный анализ можно считать репрезентативным и хорошим подтверждением некоторых результатов, полученных в прошлой главе. А именно:

- Рекомендации модели SASRec являются логичными и релевантными с точки зрения соответствия покупательскому профилю пользователя.
- Рекомендации модели SASRec являются очень однообразными внутри одного списка (то есть рекомендации взяты из одной или двух подкатегорий товаров)

- Рекомендации модели SASRec взяты из подкатегорий, ранее встречающихся в конверсиях пользователя, поэтому являются ожидаемыми и не новыми для него (по сравнению с рекомендациями модели BPR).
- Вероятно, рекомендации модели SASRec более похожи на последние товары в пользовательской истории. Эта гипотеза будет проверена в последнем пункте этой главы.

### 3.2 Зависимость рекомендаций модели от последовательной структуры пользовательской истории.

Как было указано в обзоре литературы, основным выделяющимся свойством архитектуры модели трансформер и, в частности, модели SASRec является «внимание». В оригинальной статье, посвященной SASRec [Kang, McAuley, 2018], была построена тепловая карта весов. Благодаря ей, авторы сделали вывод о том, что SASRec уделяет большие веса своего внимания последним элементам в пользовательской истории и совсем не распространяет свое внимание на начальные элементы в истории (то есть те, с которыми были взаимодействие раньше всех). Причем, часть истории, на которую отводятся большие веса модели, напрямую зависит от средней длины пользовательской истории в датасете. Чем более длинные истории в среднем представлены в датасете, тем более старые элементы получают высокий вес внимания. Аналогично, этой статье я также построил тепловую карту средних весов модели для последних 40 элементов истории (см. Figure 2). (При усреднении весов веса паддингов не учитывались.) Вертикальная ось обозначает время: сколько конверсий сделал пользователь к настоящему моменту. Горизонтальная ось обозначает позицию товара в истории конверсий. Пусть  $k$  – координата по вертикальной оси, а  $m$  – координата по горизонтальной оси. Тогда значение элемента в клетке с координатами  $[k, m]$  обозначает, с каким весом в истории учитывается товар на позиции  $m$ , если всего на данный момент было добавлено в корзину только  $k$  товаров из всей истории. Например, возьмем историю с длиной 40. На строчке 32 обозначены веса товаров в пользовательской истории в тот момент, когда было сделано только 32 конверсии из 40. В таком случае товары после 32-го получают нулевые веса, так как они еще не были добавлены в корзину. Товары с 25 по 32 получают наибольшие веса, так как они самые последние в истории. А товары с 1 по 18 получают околонулевые веса, потому что модель считает эти товары не актуальными и не важными для создания текущего покупательского образа клиента, а именно для предсказания 33-го товара в истории. Если бы модель учитывала только последние товары в истории, то тепловая карта была бы диагональная. В данном случае модель придает не

околонулевые веса примерно 10-15 последним товарам в истории (которые находятся слева от главной диагонали на расстоянии 10-15 клеток). Из этого можно сделать предположение о том, что модель учитывает в истории только последние 10-15 товаров и совсем не

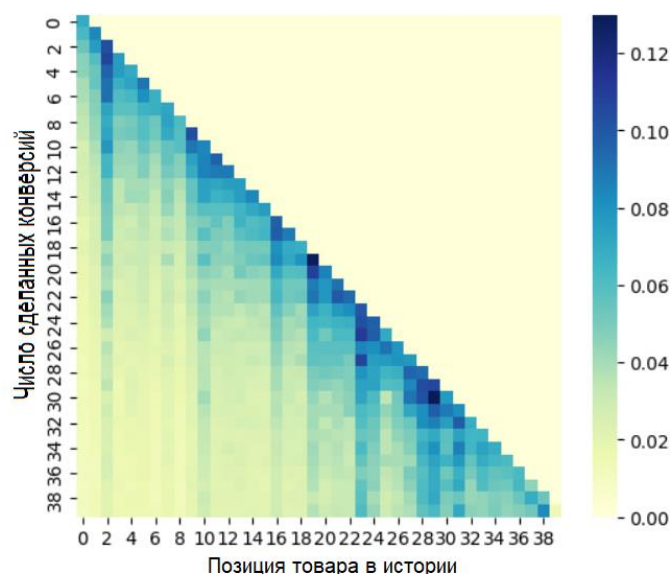


Figure 2. Тепловая карта весов внимания SASRec. Источник: [расчета автора на основе данных](#)

учитывает более ранние товары. Поэтому я ставлю следующую гипотезу: «На данном датасете более ранние товары в истории (которые были добавлены в корзину до последних 15 элементов) не влияют на рекомендации уже обученной модели SASRec». Для проверки этой гипотезы я сначала составил подвыборку только тех пользовательских историй, длина которых больше 25, и оставил в них только последние 15 элементов, убрав все предыдущие товары. Если в истории больше 25 элементов и оставляются последние 15, это значит, что вырезается больше трети всех осуществленных конверсий и это более вероятно повлияет на выдаваемые рекомендации. Так, получилась подвыборка пользователей с укороченными историями, содержащая около 30% от всех пользователей. Далее, не переобучая модель, для данной подвыборки пользователей с помощью модели SASRec я построил рекомендации для укороченных и изначальных пользовательских историй. Оба набора рекомендаций я протестировал на тестовой выборке и рассчитал те же самые описательные метрики. Как можно заметить на (Таблица 5), удаление старых рекомендаций значимо не повлияло ни на точность рекомендаций, ни на остальные метрики.

Таблица 5. Результаты тестирования модели SASRec на историях разной длины. Источник: расчета автора на основе [данных](#).

Товары в истории	HitRate@5	NDCG@5	ARP@5	Novelty@5	Coverage@5	Static_Diversity@5	Temporal_Diversity@5	Serendipity@5
Оригинальные	0,078	0,053	3,98E-04	8,38	0,173	0,265	0,568	0,344
Последние 15	0,076	0,052	4,12E-04	8,4	0,17	0,278	0,571	0,345
Разница	-2,56%	-1,89%	3,52%	0,24%	-1,73%	4,91%	0,53%	0,29%

Из этого можно сделать вывод о том, что на данном датасете SASRec учитывает только последние 15 товаров в пользовательской истории, а предыдущие конверсии значимо не влияют на основные свойства выдаваемых рекомендаций.

### 3.3 Тематическая согласованность рекомендаций с последовательностью истории пользователя.

В первом пункте этой главы на основе визуального анализа ряда репрезентативных примеров была выдвинута следующая гипотеза: «В данном датасете товары, предлагаемые моделью SASRec, тематически более похожи на последние товары в истории пользователя». Под тематической схожестью товаров я имею в виду их принадлежность к одним содержательным подкатегориям товаров, например, «сухой корм для собак» или «фильтры для воды». В прошлом пункте тепловая карта весов показала, что последние (наиболее новые) товары в пользовательской истории более важны для составления рекомендаций, чем более старые конверсии. Большой вес товара при составлении рекомендаций не обязательно должен вести к его тематической схожести с рекомендуемым товаром. Поэтому необходим более прямой анализ данного явления. Для этого я использовал следующую процедуру. Я составил подвыборку пользователей, чьи истории длиннее 20 элементов. Для них я посчитал среднее попарное расстояние (в косинусной метрике) между рекомендованными товарами и товарами в пользовательской истории аналогично тому, как была посчитана метрика serendipity в прошлой главе. Отличие данного процесса заключается в том, что попарные косинусные расстояния будут посчитаны не для всех товаров из истории вместе, а отдельно для каждой позиции. Например, согласно График 3, среднее косинусное расстояние рекомендованного товара от товара, находящегося в истории на 2 месте с конца, равно 0.31. На графике пунктирными линиями выделены критические значения в тесте на равенство средних с самым левым средним значением на графике. Так, на графике можно заметить, что SASRec рекомендует товары значимо более похожие на последние 6 товаров в истории. И более всего в среднем рекомендации похожи на самый последний товар в пользовательской истории. Чтобы можно было обобщить данный результат на все наблюдения в выборке, я построил аналогичный график для всех пользователей, которые сделали от 10 конверсий (75 % всех пользователей). Для большего

числа пользователей наблюдается тот же эффект (значимое сходство рекомендаций с последними 6-7 товарами в истории).

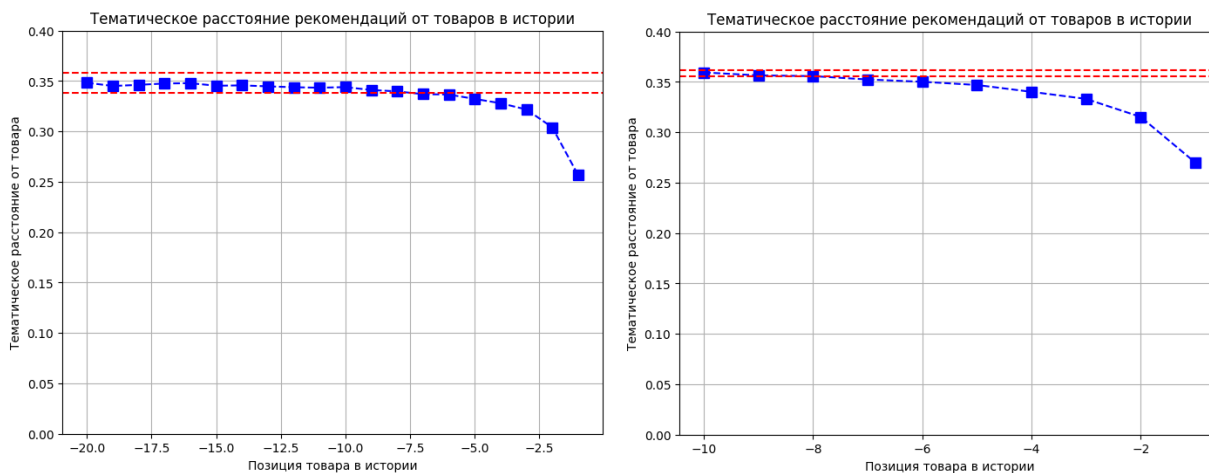


График 3. Тематическое расстояние рекомендаций от товаров в истории. Источник: расчета автора на основе [данных](#).

Также стоит сказать и об очевидном гетерогенном эффекте (см. График 4). Если в истории пользователя товары очень похожи друг на друга (левый график), то рекомендованный товар будет на одинаковом расстоянии от всех товаров в истории. Если же история разнообразна (правый график), то тематическое смещение рекомендаций к последним товарам будет заметно чуть больше, чем на первоначальном графике.

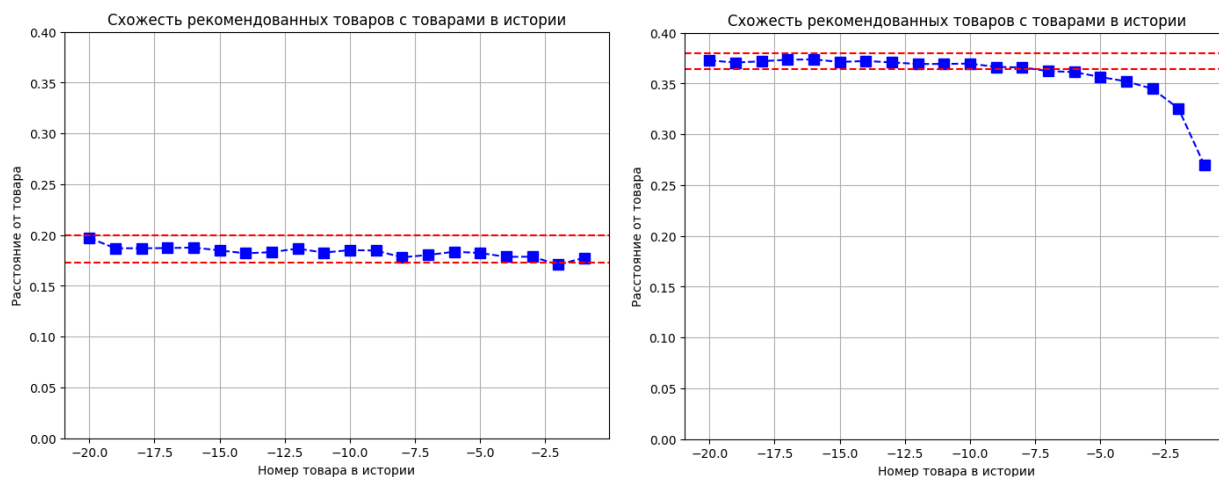


График 4. Расстояние рекомендаций от товаров в истории для однообразных и разнообразных историй. Источник: расчета автора на основе [данных](#).

Таким образом, поставленная гипотеза подтвердилась: в данном датасете товары, предлагаемые моделью SASRec, действительно тематически более похожи на последние 6

товаров в истории пользователя. И наиболее схож с рекомендациями в среднем именно тот товар, который был добавлен в корзину последним.



## Заключение.

После проведения комплексного анализа рекомендаций модели SASRec можно сказать о том, что она обладает рядом выдающихся качеств:

- 1) Данная модель генерирует высоко персонализированные рекомендации, которые принадлежат одной из подкатегорий соответствующей пользовательской истории.
- 2) SASRec точнее всех остальных алгоритмов предсказывает следующий добавленный в историю товар, что делает его потенциально очень релевантным с точки зрения добавлений в корзину.
- 3) Товары, рекомендованные трансформером, в основном похожи на самые последние добавленные в корзину товары. Это является хорошим свойством для генерации пользователям альтернативных товаров, которые они могут купить вместо только что добавленных продуктов. Вероятно, это не увеличит средний чек по сравнению с другими рассмотренными моделями, но увеличит лояльность клиентов и улучшит их опыт взаимодействия с платформой.

Однако есть и стороны, в которых SASRec серьёзно проигрывает более примитивным моделям:

- 1) Его рекомендации являются очень похожими друг на друга внутри одного листа и очевидными. Несмотря на постепенное обновление рекомендаций, пользователь, скорее всего, заскучает, если будет все время встречать ожидаемые для него товары.
- 2) Помимо потери интереса пользователей, SASRec не открывает для пользователей новые категории релевантных товаров. Впоследствии, это может привести к серьезному снижению среднего чека в отличие от внедрения, например, модели BPR, которая иногда создает для пользователя «вау» эффект.
- 3) На данном датасете SASRec не находит важные долгосрочные паттерны в историях. Это говорит о том, что либо он не считает их важными, либо их нет. В любом случае если бы целью внедрения SASRec на данную платформу было использование долгосрочных паттернов в пользовательской истории, то эта идея потерпела бы провал.

Таким образом, на мой взгляд, потенциально хорошей задачей для модели SASRec на данной платформе было бы предлагать альтернативные товары к только что добавленным в корзину. Получив приятные релевантные альтернативные варианты и улучшив опыт взаимодействия с платформой, пользователи могут начать пользоваться сайтом чаще или

порекомендовать его своим знакомым. Оба этих исхода потенциально приведут компанию PetCo к конечному желаемому результату – росту продаж.

## Библиография.

### A

- Abdollahpouri H. et al. The unfairness of popularity bias in recommendation //arXiv preprint arXiv:1907.13286. – 2019.

### B

- Burke, R.: 1999b 'Integrating Knowledge-Based and Collaborative-Filtering Recommender Systems'. In: Artificial Intelligence for Electronic Commerce: Papers from the AAAI Workshop (AAAI Technical Report WS-99-01), pp. 69-72.
- Burke R. Hybrid recommender systems: Survey and experiments //User modeling and user-adapted interaction. – 2002. – Т. 12. – С. 331-370.

### C

- Chen Y. and Canny J. F. Recommending ephemeral items at web scale. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, pages 1013–1022, 2011.

### D

- Dias M. B., Locher D., Li M., El-Deredy W., and Lisboa P. J. The value of personalised recommender systems to e-business: A case study. In Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys '08, pages 291–294, 2008.
- Du Y. et al. Is diversity optimization always suitable? Toward a better understanding of diversity within recommendation approaches //Information processing & management. – 2021. – Т. 58. – №. 6. – С. 102721.

### F

- Ferrari Dacrema M., Cremonesi P., Jannach D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches //Proceedings of the 13th ACM conference on recommender systems. – 2019. – С. 101-109.

### G

- Gomez-Uribe C. A. and Hunt N. The Netflix recommender system: Algorithms, business value, and innovation. Transactions on Management Information Systems, 6(4):13:1–13:19, 2015.

## H

- He X., Liao L., Zhang H., Nie L., Hu X., and Chua T.-S. Neural collaborative filtering. In Proceedings WWW '17. 173–182, 2017.
- He R., Kang W., and McAuley J. Translation-based recommendation. In RecSys, 2017.
- Hidasi B., Karatzoglou A., Baltrunas L., and Tikk D. Session-based recommendations with recurrent neural networks. In ICLR, 2016.

## J

- Jannach D., Adomavicius G. Recommendations with a purpose //Proceedings of the 10th ACM conference on recommender systems. – 2016. – C. 7-10.
- Jannach D., Jugovac M. Measuring the business value of recommender systems //ACM Transactions on Management Information Systems (TMIS). – 2019. – T. 10. – №. 4. – C. 1-23.

## K

- Kang W. C., McAuley J. Self-attentive sequential recommendation //2018 IEEE international conference on data mining (ICDM). – IEEE, 2018. – C. 197-206.
- Katukuri J., Konik T., Mukherjee R., and Kolay S. Post-purchase recommendations in large-scale online marketplaces. In Proceedings of the 2015 IEEE International Conference on Big Data, Big Data '15, pages 1299–1305, 2015.
- Koren Y. Collaborative filtering with temporal dynamics. Communications of the ACM, 2010.
- Kotkov D., Wang S., Veijalainen J. A survey of serendipity in recommender systems //Knowledge-Based Systems. – 2016. – T. 111. – C. 180-192.

## L

- Lathia N. et al. Temporal diversity in recommender systems //Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. – 2010. – C. 210-217.
- Lee D. and Hosanagar K. How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment. Information Systems Research, 30(1):239–259, 2019.

## P

- Petrov A. V., Macdonald C. gsasrec: Reducing overconfidence in sequential recommendation trained with negative sampling //Proceedings of the 17th ACM Conference on Recommender Systems. – 2023. – C. 116-128.

## R

- Rio-Chanona R. M. et al. Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective //Oxford Review of Economic Policy. – 2020. – T. 36. – №. Supplement\_1. – C. S94-S137.

## S

- Saifudin I., Widiyaningtyas T. Systematic literature review on recommender system: Approach, problem, evaluation techniques, datasets //IEEE Access. – 2024. – T. 12. – C. 19827-19847.
- Sharma R., Gopalani D., Meena Y. An anatomization of research paper recommender system: Overview, approaches and challenges //Engineering Applications of Artificial Intelligence. – 2023. – T. 118. – C. 105641.
- Sun F. et al. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer //Proceedings of the 28th ACM international conference on information and knowledge management. – 2019. – C. 1441-1450.

## V

- Vargas S., Castells P. Rank and relevance in novelty and diversity metrics for recommender systems //Proceedings of the fifth ACM conference on Recommender systems. – 2011. – C. 109-116.
- Vaswani A. et al. Attention is all you need //Advances in neural information processing systems. – 2017. – T. 30.

## W

- Wang H., Wang N., and Yeung D. Collaborative deep learning for recommender systems. In SIGKDD, 2015.
- Wang S., Wang Y., Tang J., Shu K., Ranganath S., and Liu H. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In WWW, 2017.

## **Y**

- Ye Z. et al. Seller-side Outcome Fairness in Online Marketplaces //arXiv preprint arXiv:2312.03253. – 2023.

## **Z**

























- Zhang S. et al. Deep learning based recommender system: A survey and new perspectives //ACM computing surveys (CSUR). – 2019. – T. 52. – №. 1. – C. 1-38.
- Ziarani R. J., Ravanmehr R. Serendipity in recommender systems: a systematic literature review //Journal of Computer Science and Technology. – 2021. – T. 36. – C. 375-396.

# Приложения.

## Приложение 1. История и рекомендации пользователя 806550.

Источник: расчета автора на основе данных

user 806550 conversions from cluster 6

									
ID: 5027894 Imagitarium Sens Hexagonal Aquarium Kit, 5.7 GAL	ID: 5015813 Imagitarium Freshwater Aquarium	ID: 5001245 Aqueon Minibow LED Desktop Fish Aquarium Kit in Black	ID: 5001245 Aqueon Minibow LED Desktop Fish Aquarium Kit in Black	ID: 5068878 Aqueon Frametess Cube Aquarium	ID: 5024275 Seachem Stability Water Conditioner	ID: 5090802 Barkworthies Odor Free Bully Sticks Small Dog Chews	ID: 5090802 Barkworthies Odor Free Bully Sticks Small Dog Chews	ID: 5090802 Barkworthies Odor Free Bully Sticks Small Dog Chews	ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew
									
ID: 5085587 Hill's Science Diet Adult Sensitive Stomach & Skin Small Bites Chicken Recipe Dry Dog Food	ID: 5051123 Aqueon Standard Glass Aquarium Tank 5.3 Gallon	ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew	ID: 5085587 Hill's Science Diet Adult Sensitive Stomach & Skin Small Bites Chicken Recipe Dry Dog Food	ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew	ID: 5172044 Imagitarium Aquatic Lava Rock Kit	ID: 5146197 Better Belly Triple Flavor Kabob Chews for Dogs	ID: 5085587 Hill's Science Diet Adult Sensitive Stomach & Skin Small Bites Chicken Recipe Dry Dog Food	ID: 5055398 Hill's Science Diet Adult Sensitive Stomach & Skin Small Bites Chicken Recipe Dry Dog Food	ID: 5085587 Hill's Science Diet Adult Sensitive Stomach & Skin Small Bites Chicken Recipe Dry Dog Food
									
ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew	ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew	ID: 5043283 Purina Pro Plan Sensitive Skin and Stomach Salmon & Rice Formula Dry Dog Food	ID: 5029174 Purina Pro Plan Sensitive Skin and Stomach Salmon & Rice Formula Small Breed Dry Dog Food						

## SASRec Recommendations

				
ID: 5029174 Purina Pro Plan Sensitive Skin & Stomach, Salmon & Rice Formula Small Breed Dry Dog Food	ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew	ID: 107860 Purina Pro Plan High Protein with Probiotics Shredded Blend Chicken and Rice Formula Dry Dog Food	ID: 5026321 Good Lovin' Braided Bully Stick Dog Chew	ID: 5043283 Purina Pro Plan Sensitive Skin and Stomach Salmon & Rice Formula Dry Dog Food

## PopUserModel Recommendations

				
ID: 5051420 Good Lovin' Traditional Beef Bully Bites Dog Chew	ID: 5085587 Hill's Science Diet Adult Sensitive Stomach & Skin Small Bites Chicken Recipe Dry Dog Food	ID: 5090802 Barkworthies Odor Free Bully Sticks Small Dog Chews	ID: 5001245 Aqueon Minibow LED Desktop Fish Aquarium Kit in Black	ID: 5015813 Imagitarium Freshwater Aquarium

## BPR Recommendations

				
ID: 3985 Hikari Tropical Sinking Wafers for Catfish, Loaches & Bottom Feeders	ID: 5057956 Aqueon LED 10 Gallon Aquarium Kit	ID: 511818 Sicce Shark ADV 400 Internal Filter	ID: 115417 Zoo Med Natural Full Cork Round Reptile Hideaway	ID: 2896 API Stress Coat Aquarium Water Conditioner

## Приложение 2. История и рекомендации пользователя 733322.

Источник: расчета автора на основе [данных](#).

user 733322 conversions from cluster 11

ID: 5051712	ID: 5018065	ID: 5106965	ID: 5024651	ID: 5024653	ID: 5068958	ID: 5024654	ID: 5024654	ID: 5024654	ID: 5024648
ID: 5024653	ID: 5018067	ID: 5024651	ID: 5197746	ID: 5051724	ID: 5051712	ID: 5051706	ID: 5024653	ID: 5024651	ID: 115852
ID: 5020948	ID: 3964	ID: 116578	ID: 5169754	ID: 5234298	ID: 3985	ID: 167828	ID: 107828	ID: 125896	ID: 107828

### BPR Recommendations

ID: 5196296	ID: 5024605	ID: 115417	ID: 2896	ID: 5184226

### PopUserModel Recommendations

ID: 107828	ID: 5024651	ID: 5024653	ID: 5024654	ID: 5051712

### SASRec Recommendations


































ID: 107828	ID: 125896	ID: 107827	ID: 126599	ID: 6084







## Приложение 3. История и рекомендации пользователя 600927.

Источник: расчета автора на основе [данных](#)

 user 600927 conversions from cluster 4

### BPR Recommendations

				
---	---	---	---	---

### PopUserModel Recommendations

				
---	---	---	---	---

### SASRec Recommendations

				
---	---	---	---	---