

Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Национальный исследовательский университет «Высшая школа экономики»

Факультет Санкт-Петербургская Школа экономики и менеджмента

Образовательная программа «Экономика»

направление подготовки: 38.03.01 Экономика

(уровень: бакалавриат)

**О Т Ч Е Т**  
**по производственной практике**

Выполнил студент 4 курса 182 группы  
Соснин Юрий Алексеевич  
(ФИО)

---

(подпись)

Проверил:

---

(должность, ФИО руководителя от факультета)

---

(оценка)

---

(подпись)

---

(дата)

Санкт-Петербург  
2022

## **СОДЕРЖАНИЕ**

1. Введение
2. Содержательная часть.
  - 2.1. Краткая характеристика организации
  - 2.2. Описание профессиональных задач, решенных студентом на практике
  - 2.3. Исполненное индивидуальное задание.
3. Заключение
4. Список литературы
5. Приложения

## **ВВЕДЕНИЕ**

В период с 10 января 2022 года по 20 марта 2022 года мной была пройдена производственная практика в компании «Самокат» (ООО «Умное пространство»). Практика проходила в отделе машинного обучения, в команде, занимающейся рекомендательными системами, в онлайн-формате.

Цели производственной практики, согласно программе практики - развить навыки полученных компетенций за период теоретического обучения на основании непосредственного знакомства с профессиональной деятельностью на базе профильной организации.<sup>1</sup>

Руководителями практики от НИУ ВШЭ были поставлены следующие задачи общего характера:

1. Ознакомиться с программой практики и пройти инструктаж по технике безопасности.
2. Ознакомиться с подразделением, в котором проходит практика: изучение действующих в подразделении нормативно-правовых актов по его функциональному назначению, структуре организации, основным показателям подразделения и его результатам.
3. Провести анализ внешней среды компании.
4. Провести анализ деятельности организации, включая анализ экономических показателей в динамике.
5. Провести анализ конкурентоспособности организации и разработать предложения по ее повышению на основе результатов анализа.
6. Провести анализ бизнес-процесса, в который был включен студент в период прохождения практики.
7. Собрать количественные и/или качественные данные о деятельности организации. Сформулировать и обосновать предложения по совершенствованию деятельности организации, направлению ее развития.

Руководитель практики от организации поставил задачи индивидуального характера:

1. Изучить литературу по теме рекомендательных систем
2. Имплементировать алгоритм анализа данных из одной из статей
3. Проверить работу алгоритма на общедоступных данных

---

<sup>1</sup> Программа практики Образовательная программа «Экономика» URL:  
<https://spb.hse.ru/mirror/pubs/share/544988522.pdf>

4. Применить алгоритм к данным компании, рассчитать метрики качества, сравнить с использующимися методами
5. Сделать вывод о целесообразности использования нового метода

## СОДЕРЖАТЕЛЬНАЯ ЧАСТЬ

### КРАТКАЯ ХАРАКТЕРИСТИКА ОРГАНИЗАЦИИ

«Самокат» — российский сервис экспресс-доставки продуктов питания и товаров для дома, совместное предприятие «Сбера» и VK, с штаб-квартирой в Санкт-Петербурге.<sup>2</sup>

Организационно «Самокат» состоит из двух юридических лиц: ООО «Умный ритейл» и ООО «Умное пространство». Первое подразделение занимается инфраструктурой (склады, курьеры, и т.п.), второе - IT-подразделение, занимающегося разработкой приложения, аналитикой, а также задачами анализа данных. В частности, отдел машинного обучения занимается прогнозированием спроса и его сезонности, регулированием минимального чека, а также анализом поведения потребителей, для совершенствования рекомендательных систем внутри приложения. В этой части команды и была пройдена практика.

Внутри отделов подразделений преобладает плоская (горизонтальная) организационная структура, присутствует минимальное число уровней организационной иерархии. Отдельные сотрудники имеют сферы профессиональных интересов и определенные задачи, которыми они занимаются ежедневно, однако не закреплены за ними и периодически могут участвовать в других проектах внутри «ML-команды».

### ОПИСАНИЕ ПРОФЕССИОНАЛЬНЫХ ЗАДАЧ, РЕШЕННЫХ СТУДЕНТОМ НА ПРАКТИКЕ

В этом разделе будут описаны результаты выполнения заданий общего характера, результаты и процесс выполнения индивидуального задания в следующем разделе.

Во-первых, стоит составить характеристику **внешней среды**, в которой функционирует компания. Продажи на рынке eGrocery, то есть рынке онлайн-продажи товаров повседневного спроса, в том числе продуктов питания, в 2021 года составили 329 млрд рублей, показав более чем двукратный рост по сравнению с 2020 годом, когда они составили 155 млрд рублей. В декабре 2021 года эксперты предсказывали и дальнейший рост, давая оценку в 500-600 миллиардов рублей к 2023 году.<sup>3</sup>

---

<sup>2</sup> Официальный сайт «Самоката» / URL: samokat.ru

<sup>3</sup> Выпуск eGrocery бюллетеня Январь 2022 / URL: datainsight.ru/eGrocery\_January\_2022

Основными драйверами роста были названы расширяющаяся география – доставка начинает работать во все большем числе городов России – а также экстенсивная маркетинговая активность. Тем не менее, причиной настоящего расцвета онлайн-ритейла послужила пандемия коронавируса, резко повысившая спрос людей на «бесконтактный» способ приобретения товаров первой необходимости. Именно в 2020 году рынок показал наибольший рост.

Согласно отчету tadviser<sup>4</sup>, на рынке eGrocery присутствует 6 крупных компаний, занимающих около 80% этого рынка.

- холдинг O2O (совместное предприятие «Сбера» и VK, включает сервисы Delivery Club, «Самокат»)
- «Сбермаркет»
- «Вкусвилл»
- X5 Group (сервисы Vprok.ru, доставка «Пятерочки» и «Перекрестка»)
- «Яндекс» («Еда», «Лавка»)
- «Севергрупп» («Лента», «Утконос»)

«Самокат» – относительно молодая компания, основанная в 2018 году. Изначально проект позиционировал себя как инновация на рынке доставки, заявляя беспрецедентно быстрое время доставки, от 15 до 30 минут. Основное **конкурентное преимущество** компании, позволяющее реализовать такую скорость – это система дарксторов. Даркстор (dark store) – это такой формат склада, на котором товары хранятся в собранном виде, готовые для доставки курьерами. Он выглядит как обычный магазин с штучной выкладкой товаров, но недоступен непосредственно для покупателей, будучи предназначенным только для курьеров.

Согласно отчету CM Ventures, «Самокат» в 2021 являлся одной из четырех крупнейших компаний по быстрой доставке в мире, наряду с Dmart, «Яндекс.Лавкой», и Getir.<sup>5</sup>

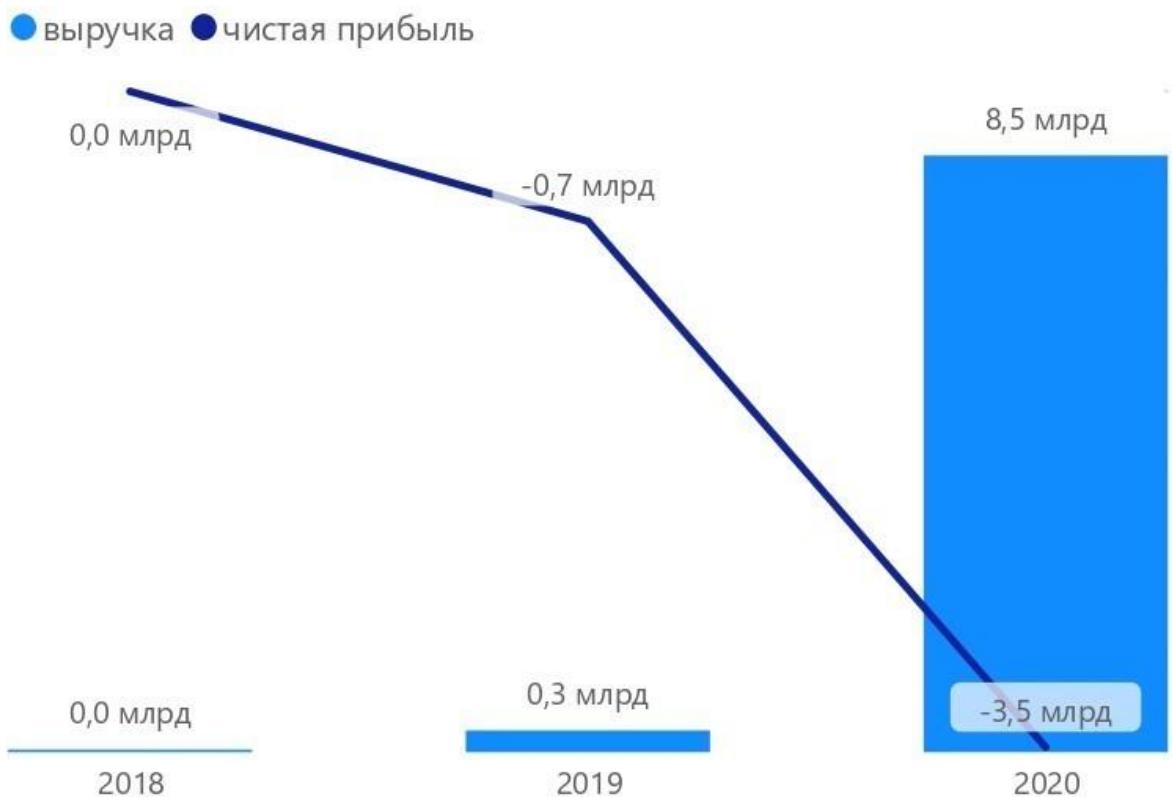
Что касается **экономических показателей**, можно провести оценку основных финансовых показателей за 2018-2021 годы. Начав с 0,3 млрд рублей в 2019, выручка в 2020 составила 8,5 миллиарда рублей, а за 9 месяцев 2021 – уже превысила 23 млрд. «Самокат» работает в 46 городах России, а также, на момент начала 2022, в США; функционируют около 400 дарксторов. Прибыль при этом остается отрицательной, и финансируется за счет добавочного капитала.

---

<sup>4</sup> [tadviser.ru/index.php/Статья:Интернет-продажи\\_продуктов\\_питания\\_в\\_России](http://tadviser.ru/index.php/Статья:Интернет-продажи_продуктов_питания_в_России)

<sup>5</sup> [e-pepper.ru/news/investory-cm-ventures-rossiyskiy-rynok-bystroy-dostavki-budet-rasti-eshchye-10-let.html](http://e-pepper.ru/news/investory-cm-ventures-rossiyskiy-rynok-bystroy-dostavki-budet-rasti-eshchye-10-let.html)

## Выручка и чистая прибыль Самокат (ООО «Умный ритейл») 2018-2020гг. руб.



Источник: vc.ru<sup>6</sup>

Подробнее стоит рассказать о **бизнес-процессе**, в который я был непосредственно включен. Рекомендательная система – это программа, предсказывающая на основе данных о пользователе его будущие предпочтения. В контексте онлайн-ритейла – это раздел в приложении, показывающий товары, похожие на уже внесенные пользователем в корзину или на прошлую историю покупок этого пользователя. Постановка задачи рекомендации в контексте машинного обучения – как, имея матрицу пользователей и предметов, заполненную в очень низкой степени (как правило пользователи взаимодействуют лишь с очень небольшой долей предметов), наилучшим образом предсказать спрос пользователя на предметы, с которыми он еще не взаимодействовал.

Процесс разработки такой системы будет описан позже, в разделе «Исполненное индивидуальное задание», однако стоит сказать о способах оценки качества полученного результата, так как это непосредственно связано с соответствующим бизнес-процессом.

<sup>6</sup> <https://vc.ru/finance/247397-sezon-otchetov-chast-2-itogi-deyatelnosti-krupneyshih-sluzhb-dostavki-edy-rossii-za-2020-g>

Существует два фундаментальных способа оценки рекомендательной системы: офлайн и онлайн оценка. Оффлайн оценка проводится непосредственно во время разработки, это определенная метрика, которая показывает качество предсказания на существующем наборе данных. Примерами таких метрик для задачи бинарного предсказания (потенциальной покупки предмета, в отличие, например, от оценки фильма) могут быть Hit Rate и nDCG (normalized discounted cumulative gain). Они показывают, насколько много предметов, предложенных системой, на самом деле были куплены пользователями.

Онлайн оценка – это следующий этап оценки рекомендательной системы, она проводится аналитиками с помощью А/В тестов. Для этого реальные пользователи приложения случайно делятся на две группы: группу воздействия и группу контроля. Для группы контроля новая версия системы (подлежащая оценке) вводится, для группы контроля – нет. После определенного периода проводится оценка с использованием бизнес-метрик, таких как Click-through rate, Conversion, Sales & Revenue. Уже на основе этого этапа тестирования принимается решение о вводе системы в эксплуатацию.

Я во время своей практики занимался только разработкой и офлайн-стадией оценки, в этапе тестов принять участие, к сожалению, не представилось возможности ввиду его длительности.

### ИСПОЛНЕННОЕ ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ

Моим индивидуальным заданием от организации была проверка гипотезы о том, что применение гиперболических методов в процессе рекомендации будет существенно улучшать метрики онлайн-этапа на данных компаний.

Первой задачей было изучение научной статьи «HyperML: A Boosting Metric Learning Approach in Hyperbolic Space for Recommender Systems» (Tran, L. V., Tay, Y., Zhang, S., Cong, G., & Li, X., 2019).

Задача Metric Learning заключается в том, чтобы на основе данных создать модель, помещающую предметы и пользователей в определенное многомерное пространство, в котором предметы, понравившиеся пользователям, находятся ближе к ним в смысле расстояния, а не понравившиеся – дальше. Например, проекции таких пространств можно увидеть в Приложении. Таким образом, похожие предметы «притягиваются» друг к другу, как и похожие пользователи. Чтобы порекомендовать покупателю товары, нужно взять наиболее близко расположенные к нему предметы, у них будет выше шанс ему понравиться.

Инновация статьи заключается в том, что авторы используют не обычные Евклидовы расстояния для определения сходства объектов, а гиперболические расстояния модели Пуанкарé. По результатам статьи, авторы приходят к выводу о том, что такие расстояния лучше отражают структуру пространства предметов и пользователей, чем обычные, и производят лучшие результаты. Моей задачей являлось проверка этой гипотезы на данной компании «Самокат».

Для этого мной на языке Python была имплементирована модель Collaborative Metric Learning (CML), а также ее гиперболическая версия из статьи. Отдельные компоненты были взяты из существующих библиотек с открытым исходным кодом.

После создания моделей была проведена их офлайн оценка, сравнение друг с другом и стандартным бейзлайном ALS, работающим по другому принципу. Значения метрик приведены в таблице.

| Модель         | Hit Rate at 10 | Hit Rate at 10 (global) | nDCG at 10    |
|----------------|----------------|-------------------------|---------------|
| ALS            | 0.524          | 0.0485                  | 0.2875        |
| CML            | <b>0.56</b>    | 0.051                   | <b>0.3169</b> |
| Hyperbolic CML | 0.5435         | <b>0.0653</b>           | 0.3094        |

Из таблицы видно, что по 2 из 3 метрик, результаты гиперболического варианта оказываются хуже, чем у стандартного CML. Однако результаты, тем не менее, лучше бейзлайна. Получившиеся значения в целом соответствуют другой статье, «Where are we in embedding spaces? A Comprehensive Analysis on Network Embedding Approaches for Recommender Systems» (Zhang, S., Chen, H., Ming, X., Cui, L., Yin, H., & Xu, G., 2021), авторы которой экспериментировали с гиперболическими моделями. По их выводам, гиперболические модели далеко не всегда применимы, и стандартные методы в большинстве случаев оказываются более предпочтительными.

Таким образом, гипотеза о том, что новая модель сможет улучшить рекомендации оказалось отвергнута. Было принято решение пока не переходить к онлайн стадии тестирования с помощью А/В теста, до возможного совершенствования работы модели.

## ЗАКЛЮЧЕНИЕ

За время прохождения практики я применил навыки, полученные на образовательной программе «Экономика», непосредственно ознакомился с деятельностью профильной

организации. Мной были применены мои знания в области машинного обучения для реализации прикладной задачи. К сожалению, достичь положительного результата не удалось, однако проведенное исследование тем не менее имеет смысл, исключая одно из потенциальных направлений развития.

Кроме того, был проведен анализ компании, ее внешней среды, конкурентоспособности, и одного бизнес процесса, в который я был включен.

## Список литературы

- Tran, L. V., Tay, Y., Zhang, S., Cong, G., & Li, X. (2019). HyperML: A Boosting Metric Learning Approach in Hyperbolic Space for Recommender Systems. ArXiv:1809.01703 [Cs, Stat]. <http://arxiv.org/abs/1809.01703>
- Zhang, S., Chen, H., Ming, X., Cui, L., Yin, H., & Xu, G. (2021). Where are we in embedding spaces? A Comprehensive Analysis on Network Embedding Approaches for Recommender Systems. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2223–2231. <https://doi.org/10.1145/3447548.3467421>
- Программа практики Образовательная программа «Экономика» / URL:  
<https://spb.hse.ru/mirror/pubs/share/544988522.pdf>
- Официальный сайт «Самоката» / URL: samokat.ru
- Выпуск eGrocery бюллетеня Январь 2022 // Data Insight / URL:  
[https://www.datainsight.ru/eGrocery\\_January\\_2022](https://www.datainsight.ru/eGrocery_January_2022)
- Бум спроса на инновации для розничной торговли: как меняется объем инвестиций в RetailTech в России и в мире // New Retail / URL: [https://new-retail.ru/business/bum\\_sprosa\\_na\\_innovatsii\\_dlya\\_roznichnoy\\_torgovli\\_kak\\_menyaeetsya\\_obem\\_investitsiy\\_v\\_retailtech\\_v\\_ros2682/](https://new-retail.ru/business/bum_sprosa_na_innovatsii_dlya_roznichnoy_torgovli_kak_menyaeetsya_obem_investitsiy_v_retailtech_v_ros2682/)
- Интернет-продажи продуктов питания в России // TAdviser / URL:  
[https://www.tadviser.ru/index.php/Статья:Интернет-продажи\\_продуктов\\_питания\\_в\\_России](https://www.tadviser.ru/index.php/Статья:Интернет-продажи_продуктов_питания_в_России)
- Сезон отчетов. Часть 2. Итоги деятельности крупнейших служб доставки еды России за 2020 г // vc.ru / URL: <https://vc.ru/finance/247397-sezon-otchetov-chast-2-itogi-deyatelnosti-krupneyshih-sluzhb-dostavki-edy-rossii-za-2020-g>
- Дарксторы привлекают инвестиции. Аналитики изучили мировой рынок экспресс-доставки продуктов на примере крупнейших игроков в сегменте // Inc. / URL:  
<https://incrussia.ru/understand/dark-store-research/>

## Приложения

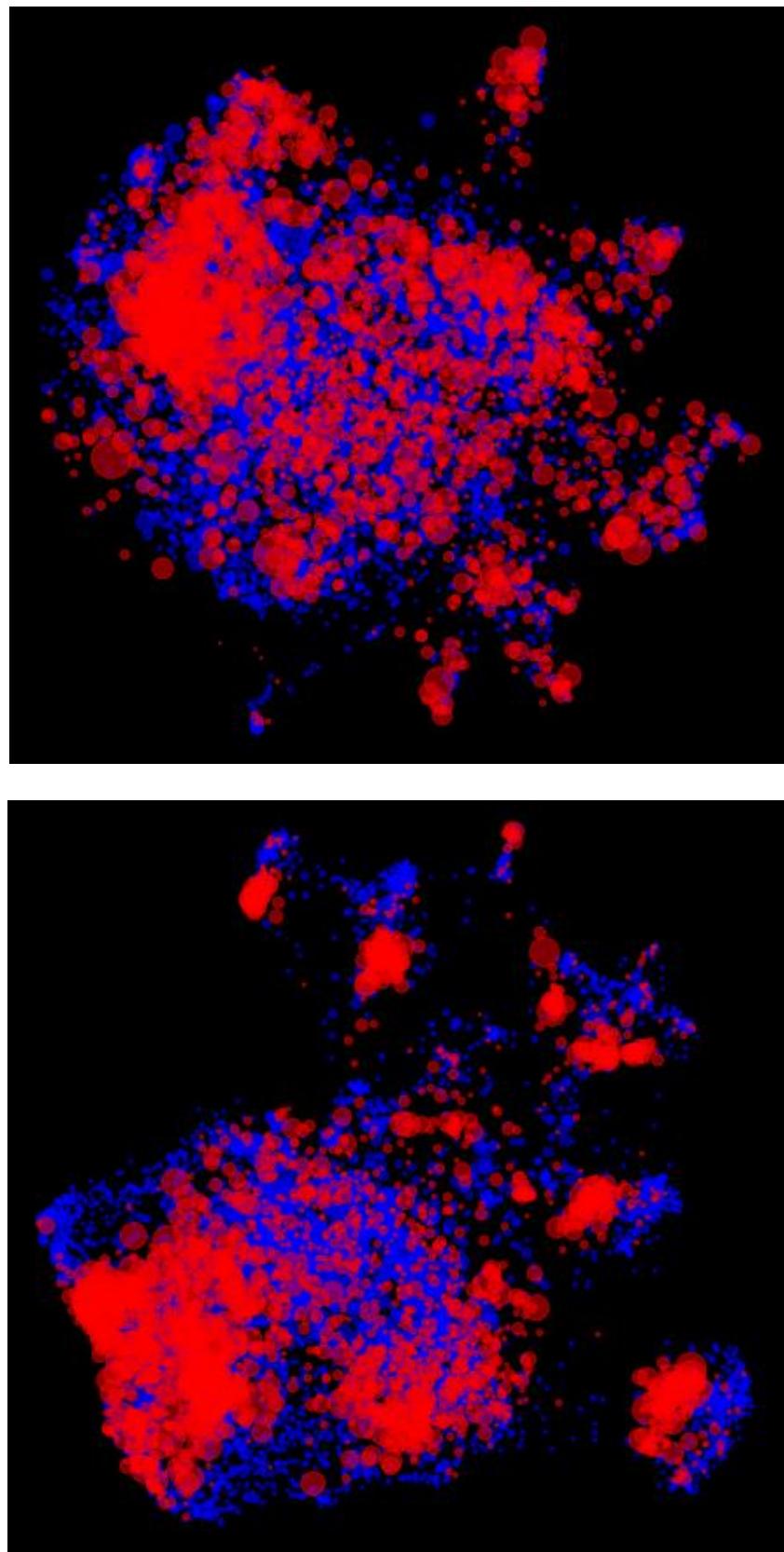


Иллюстрация получившихся эмбеддингов. Сверху – Hyperbolic CML, снизу – CML.

Синим цветом выделены предметы, красным - пользователи.

**Подтверждение выполнения**

|   |   |   |
|---|---|---|
| Сосин Ю.А.  | Директор по персоналу и<br>организационным<br>изменениям, Глотова Ю.Н.                  | Доцент Департамента<br>экономики, Артемова Д.И. |
|  | <br>МП |   |
| « 20 » марта 2022 г.  | 2022 г.   | « _____ » 2022 г.                               |