

Всероссийский конкурс исследовательских и проектных работ
школьников «Высший пилотаж»

Анализ отзывов клиентов на банковские каналы обслуживания и продукты/услуги

Проектная работа

Направление «Computer Science»

Авторы: Рукавица Артём Кириллович,

Федченко Анастасия Кирилловна,

учащиеся 11ого класса,

ГБОУ «Школа №1533 «ЛИТ»,

Москва

2022 г.

1. Содержание

1. Содержание	2
2. Аннотация	3
3. Введение	4
3.1 Анализ предметной области	4
3.2 Актуальность	5
3.3 Постановка задачи	6
3.4 Описание решаемых задач	7
4. Анализ существующих решений	8
5. Дорожная карта	8
6. Методы исследования	9
6.1 Критерий согласия Пирсона (χ^2 , хи-квадрат)	9
6.2 Метод опорных векторов	10
6.3 Градиентный бустинг	10
7. Описание разработанного решения/паспорт изделия	11
7.1 Страница «Средняя оценка по всем банкам»	11
7.2 Средние оценки категорий по банкам и регионам	12
7.3 Страница «Важные слова для различных категорий»	15
7.4 Страница «Интерактивная карта банковских отделений страны»	16
8. Выводы	18
9. Перспективы дальнейшей разработки	19
10. Список источников	20
11. Приложения	22
А. Sequence Diagram (диаграмма последовательностей) для сбора отзывов с сайта https://yandex.ru/maps	22
Б. Sequence Diagram (диаграмма последовательностей) для сбора отделений с сайта https://1000bankov.ru	23
В. Class Diagram (диаграмма классов)	24
Г. Activity diagram (диаграмма деятельности) для сбора отзывов	25
Д. Activity diagram (диаграмма деятельности) для сбора отделений	26
Е. Component diagram (диаграмма компонентов)	27

2. Аннотация

Цифровые технологии открыли новые возможности и установили абсолютно новые правила игры для компаний и пользователей, в результате чего конкуренция сместилась от создания лучшего продукта или услуги к созданию лучшего клиентского опыта.

Актуальность нашего проекта заключается в отсутствии в открытом доступе программ или сайтов, которые занимаются сравнением клиентского опыта в банковской сфере РФ. Целью нашей работы является создание web-приложения, позволяющего пользователю сравнить по нескольким показателям любой банк с рынком в среднем и с лучшими игроками рынка, определить области улучшения и дальнейшего развития каналов обслуживания, продуктов и услуг по определенным критериям: удобство офиса, сеть банкоматов, уровень сервиса, персонал, продукты и услуги, дистанционные каналы обслуживания.

Для создания веб-приложения мы воспользовались языком программирования Python, фреймворком для визуализации приложений с использованием машинного обучения streamlit, фреймворком scikit-learn для предиктивного анализа данных, библиотекой plotly для создания интерактивных диаграмм и модулем folium для создания карты банковских отделений.

Результатом нашей работы является сайт, предоставляющий сервис визуализации анализа, проведённого на собранных данных. На нашем сайте пользователю доступно три типа диаграмм и интерактивная карта: горизонтальная гистограмма (для отображения средних оценок и важных слов), диаграмма в формате “паутина” (для удобного отображения средних оценок по категориям), диаграмма в формате “торнадо” (для отображения всех распределений положительных и отрицательных отзывов по категориям и в среднем), интерактивная карта банковских отделений с подробной аналитикой для каждого отделения из базы данных.

Нами создан продукт, принятый в котором подход применим к оцениванию и сравнению различных видов обслуживания, например, приложения банков и доставки продуктов, сети продовольственных супермаркетов и пр. Для построения аналитики нужны лишь отзывы, а также их категоризация.

3. Введение

3.1 Анализ предметной области

Цифровые технологии открыли новые возможности и установили абсолютно новые правила игры для компаний и пользователей, в результате чего конкуренция сместилась от создания лучшего продукта/услуги к созданию лучшего клиентского опыта. Качество взаимодействия бизнеса с людьми вычисляется с помощью индекса потребительской лояльности (NPS). *NPS* — индекс определения приверженности потребителей товару или компании (индекс готовности рекомендовать), используется для оценки готовности к повторным покупкам. На представленной диаграмме красными маркерами отмечены банки-лидеры по NPS; серыми маркерами отмечены отстающие по этому показателю. На графике мы можем увидеть, насколько изменился чистый процентный доход в зависимости от страны и NPS за три года. Обратим внимание на то, что у лидеров по NPS чистый процентный доход увеличился; у отстающих банков же он либо остался на прежнем уровне, либо уменьшился. Исходя из вышесказанного, клиентский опыт очень важен, и поэтому наш проект нацелен на его анализ в разных банках.



Изображение 1. Сравнение стран по уровню NPS в банковской сфере

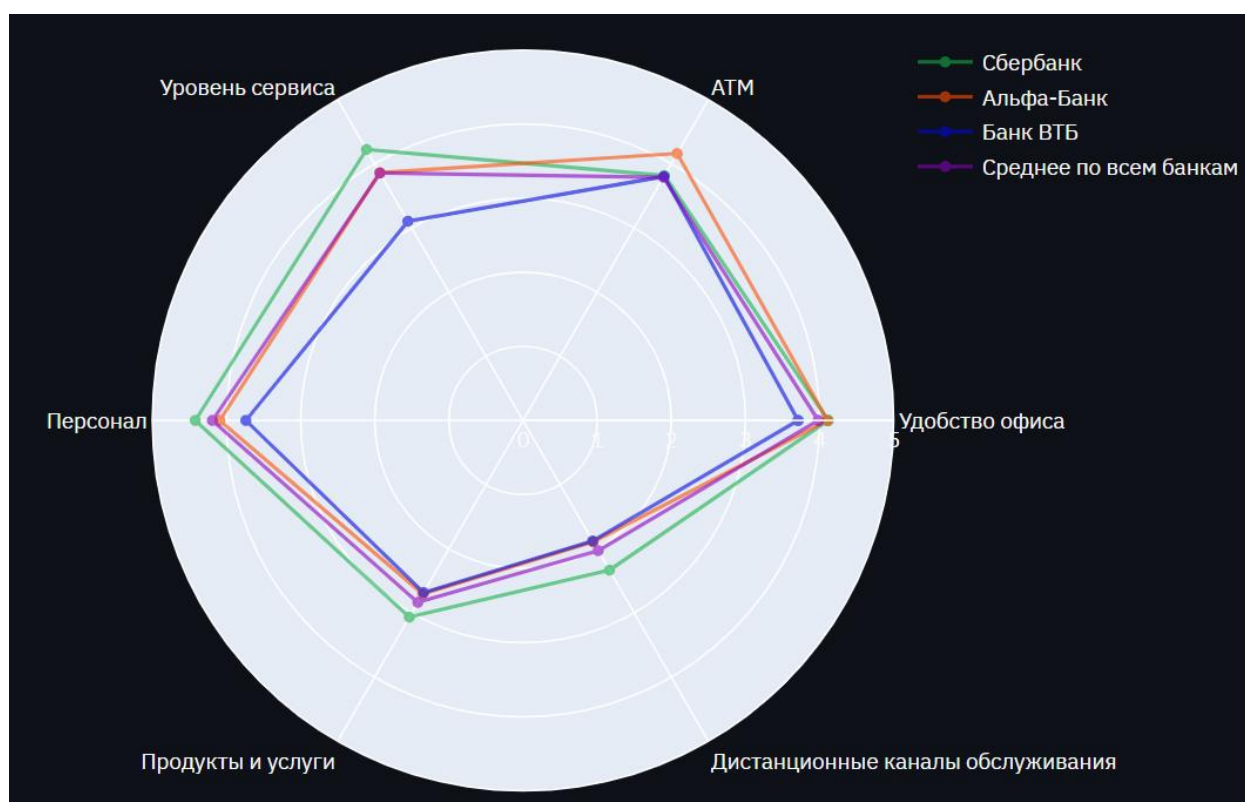
Источник: исследование Bain & Company “In search of customers who love their bank”

3.2 Актуальность

Лучший клиентский опыт включает в себя простые, умные и беспроblemные точки контакта пользователя с компанией, персонафикацию, простоту и понятность использования, позитивные эмоции от процесса использования продукта/услуги (не просто функциональные свойства) в условиях постоянно изменяющихся предпочтений пользователей и конкурентной среды.

Банки – традиционная и консервативная отрасль (в которой, конечно, есть и свои технологические лидеры), находящаяся в начале пути от предоставления традиционных услуг и создания экономически эффективных процессов к человеко-центричному дизайну.

Итак, *причиной актуальности* нашего продукта является отсутствие в открытом доступе программ/сайтов, занимающихся сравнением клиентского опыта в различных банках. Наша программа упрощает пользователю процесс сравнения нескольких банков по разным критериям. Вместо того чтобы искать информацию про каждый банк (т.е. открывать официальные сайты, форумы, отзывы на конкретные услуги), тратя большое количество времени, можно воспользоваться нашим сайтом и получить всю нужную информацию быстро и удобно, получив результат анализа в виде графика в формате “паутина” (пример на фотографии ниже).



Изображение 2. Пример графика в формате «паутина»

3.3 Постановка задачи

Целью нашей работы является создание web-приложения, позволяющего пользователю сравнить любой банк с рынком в среднем и с лучшими игроками рынка, определить области улучшения и дальнейшего развития каналов обслуживания, продуктов и услуг по определенным критериям: удобство офиса, банкоматы, касса, уровень сервиса, персонал, продукты и услуги, дистанционное обслуживание. В качестве входных данных мы используем отзывы с картографического сервиса Яндекс.Карты (www.yandex.ru/maps) на банковские отделения. На основе полученных данных создаем модель оценки качества банковских каналов, продуктов и услуг с визуализацией результата.

Вместе с техническим заданием мы получили от научного руководителя список банков, отзывы на отделения которых нам необходимо проанализировать. Ниже приведена таблица, содержащая список с названиями 24 банков:

Таблица 1. Список банков для анализа

1	Сбербанк
2	Банк ВТБ
3	Альфа-Банк
4	Газпромбанк
5	Россельхозбанк
6	Почта Банк
7	Банк Открытие
8	Росбанк
9	Совкомбанк
10	Райффайзенбанк
11	Промсвязьбанк
12	Банк Хоум Кредит
13	Банк ДОМ.РФ
14	Уралсиб
15	ЮниКредит
16	Сетелем Банк
17	Русфинанс Банк
18	Ренессанс Кредит
19	Московский кредитный банк
20	Банк Санкт-Петербург
21	СМП Банк

22	Московский индустриальный банк
23	Уральский банк реконструкции и развития
24	Ситибанк

Для построения аналитики необходимо разделить отзывы пользователей на темы. Данная категоризация была предложена научным руководителем.

1. **Удобство офиса** (удобство расположения, транспортная доступность, наличие парковки, время работы, простота навигации внутри офиса, электронная очередь/возможность записаться в офис, комфортная зона ожидания, состояние и привлекательность интерьера офиса, чистота, наличие кулера с водой/туалета для клиентов, детского уголка, Wi-Fi и т.п.);

2. **Банкоматы** (их наличие, круглосуточная доступность, достаточность количества устройств, функциональность - выдача/прием наличных/платежи/переводы и платежи и т.п., работоспособность - в банкомате есть наличные/внесение наличных работает/у банкомата есть связь/он не висит и т.д.);

3. **Уровень сервиса** (время ожидания, скорость обслуживания, простота документов, доступность безбумажной/электронной подписи клиента, доступность нужной клиенту услуги, уполномоченность персонала решать вопросы по претензиям на месте без перенаправления в головной офис);

4. **Персонал** (вежливость, опрятность, клиентоориентированность/заинтересованность в решении вопроса клиента, компетентность, решение вопроса клиента без перенаправления к другому специалисту и т.п.);

5. **Продукты и услуги** (полнота и доступность, стоимость, прозрачность условий, удобство использования и др. характеристики стандартных продуктов и услуг банка);

6. **Дистанционные каналы обслуживания** (удобство интернет-банка (ИБ) и мобильного банка (МБ), простота подключения, доступность операций в ИБ/МБ без необходимости визита в офис, простота совершения операций в ИБ/МБ, ограничения по операциям в ИБ/МБ и пр.);

7. **Прочее**. В эту категорию следует относить отзывы, которые не принадлежат ни к одной категории выше.

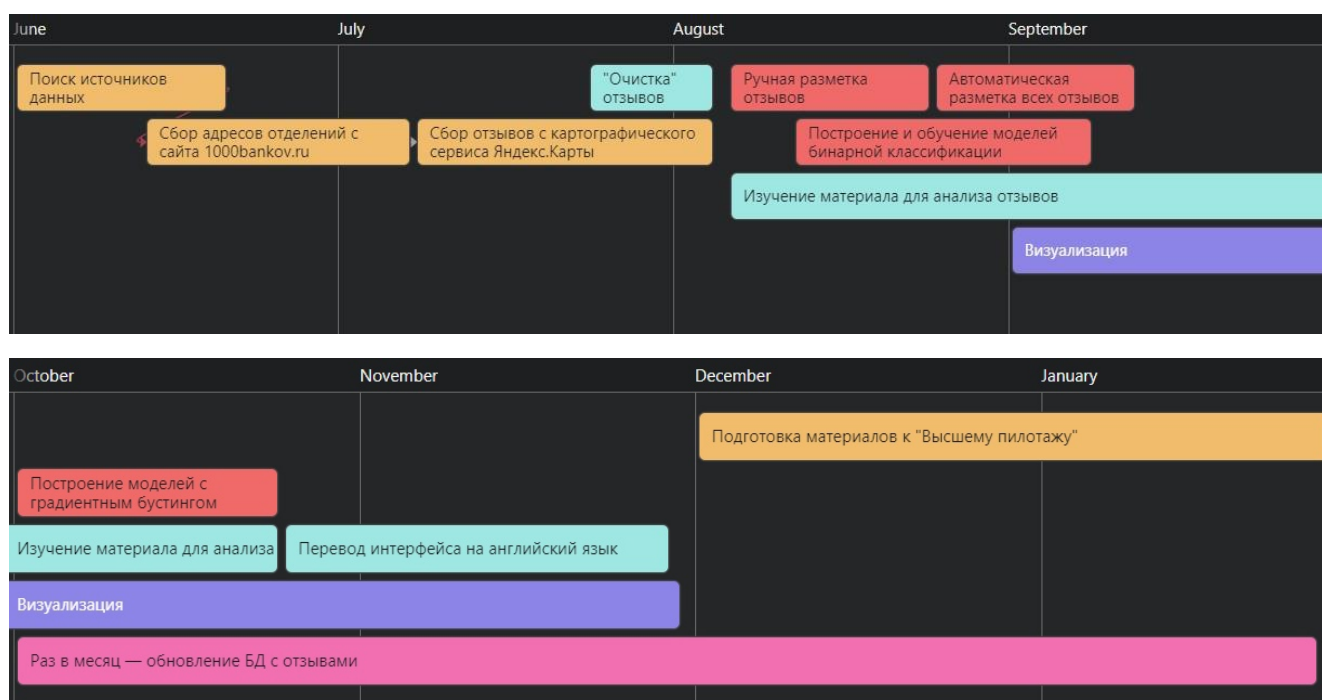
3.4 Описание решаемых задач

1. упрощение процесса сравнения клиентского опыта в разных банках;
2. выделение наиболее важных критериев оценки по мнению клиентов;
3. определение областей улучшения и дальнейшего развития каналов обслуживания, продуктов и услуг.

4. Анализ существующих решений

На данный момент существует множество программ, нацеленных на сравнение любого банка с рынком в среднем и с лучшими игроками рынка, определение области улучшения и дальнейшего развития каналов обслуживания, продуктов и услуг, однако все они являются частями дорогих систем и недоступны для использования простым людям. Именно поэтому мы не знаем, как выглядят и какими качествами обладают аналоги.

5. Дорожная карта



Изображения 3-4. Дорожная карта проекта

Кратко опишем, как решается наша задача.

1. Сбор адресов отделений банков с сайта <https://1000bankov.ru>;
2. Сбор отзывов с сайта <https://yandex.ru/maps>;
3. Подготовка отзывов к анализу: удаление стоп-слов, знаков препинания, приведение слов в начальную форму;
4. Ручной лейблинг отзывов на случайной выборке, состоящей из 1600 отзывов, на основе категоризации из «Постановки задачи»;
5. Обучение моделей бинарной классификации на сервисе Google Colab на основе пролейбленных отзывов из пункта 4 при помощи метода опорных векторов (см. 6.2);
6. Категоризация всех отзывов при помощи обученных моделей;

7. Визуализация: создание горизонтальных гистограмм для отображения средних оценок и важных слов (на основе градиентного бустинга, см. 6.3), диаграммы в форматах «паутина» и «торнадо» (с применением критерия согласия Пирсона, см 6.1), интерактивной карты банковских отделений.

6. Методы исследования

6.1 Критерий согласия Пирсона (χ^2 , хи-квадрат)

Критерий согласия Пирсона проверяет значимость расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот. Он используется для проверки нулевой гипотезы о подчинении наблюдаемой случайной величины определенному теоретическому закону распределения. Нулевая гипотеза заключается в том, что частоты согласованы, то есть фактические данные не противоречат ожидаемым.

Критерий согласия Пирсона считается по данной формуле:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e},$$

Формула 1. Критерий согласия Пирсона

где f_o и f_e — наблюдаемые и ожидаемые частоты соответственно. Суммирование производится по всем ячейкам таблицы.

В нашей задаче мы используем критерий согласия Пирсона при проверке расхождения между частотами в двух ситуациях:

1) Сравнение долей положительных и отрицательных отзывов в одном банке каждой категории и итоговых долей всех категорий:

Нулевая гипотеза такова: Распределение долей положительных и отрицательных отзывов в категории имеет то же распределение, что по всем отзывам банка в целом.

Пусть распределение долей положительных и отрицательных отзывов по всем отзывам банка — это ожидаемое распределение, а распределение долей положительных и отрицательных отзывов по отзывам конкретной категории — фактическое распределение.

Если вероятность p , которая посчитана при помощи функции `scipy.stats.chi2_contingency`, больше 0.05, то мы не показываем выбранную категорию, так как она с высокой вероятностью похожа на распределение долей по всем отзывам, то есть нулевая гипотеза оказалась правдивой.

Если вероятность $p \leq 0.05$, то распределения различаются, значит, доли выбранной категории стоит показать, то есть мы опровергли нулевую гипотезу.

2) Сравнение долей положительных и отрицательных отзывов каждой категории между банками.

Нулевая гипотеза такова: распределение долей положительных и отрицательных отзывов в пределах одной категории между всеми банками совпадают.

Мы попарно проверяем распределения между банками. Фактическим и ожидаемым распределениями мы ставим соответственно положительные и отрицательные доли отзывов на каждую пару банков. Если найдется хотя бы одно, вероятность p которого ≤ 0.05 , то мы показываем весь ряд, так как распределения с высокой вероятностью различны, то есть нулевая гипотеза опровергнута.

6.2 Метод опорных векторов

Метод опорных векторов — это алгоритм, позволяющий определить гиперплоскость (в частном случае — линию), которая распределяет данные на два класса. Разделяющей гиперплоскостью будет гиперплоскость, создающая наибольшее расстояние до двух параллельных гиперплоскостей.

Мы воспользовались классификатором на основе метода опорных векторов для создания двух типов моделей:

- определение отзыва к определенной категории;
- определение принадлежности слова к позитивным или негативным отзывам.

Для создания моделей первого типа мы "пролейблили" 1600 отзывов по категориям.

Входные данные: "чистый" отзыв и 0 либо 1 (0 — отзыв не принадлежит категории, 1 — отзыв принадлежит категории).

Каждая из шести моделей определяет принадлежность или непринадлежность отзыва к конкретной категории.

Для создания моделей второго типа мы в качестве входных данных использовали отзывы по каждой из категорий и 0 либо 1 (0 — оценка отзыва 1 или 2 (т.е. он негативный), 1 — оценка отзыва 4 или 5 (т.е. он положительный)).

Каждая из шести моделей может определить, в каких отзывах встречается слово: в положительных или отрицательных.

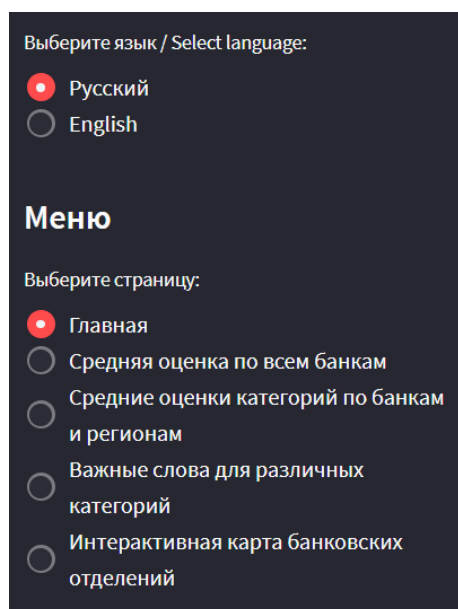
6.3 Градиентный бустинг

Градиентный бустинг — это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Градиентный бустинг позволяет обучать алгоритмы последовательно (т.е. каждый следующий учится на ошибках предыдущего).

Мы используем градиентный бустинг для определения наиболее важных слов в каждой из категорий.

7. Описание разработанного решения/паспорт изделия

При переходе на сайт пользователю предлагается выбрать язык (русский/английский) и одну из страниц: средняя оценка по всем банкам, сыредние оценки категорий по банкам и регионам, важные слова для различных категорий, интерактивная карта банковских отделений.



Выберите язык / Select language:

☒ Русский

☐ English

Меню

Выберите страницу:

☒ Главная

☐ Средняя оценка по всем банкам

☐ Средние оценки категорий по банкам и регионам

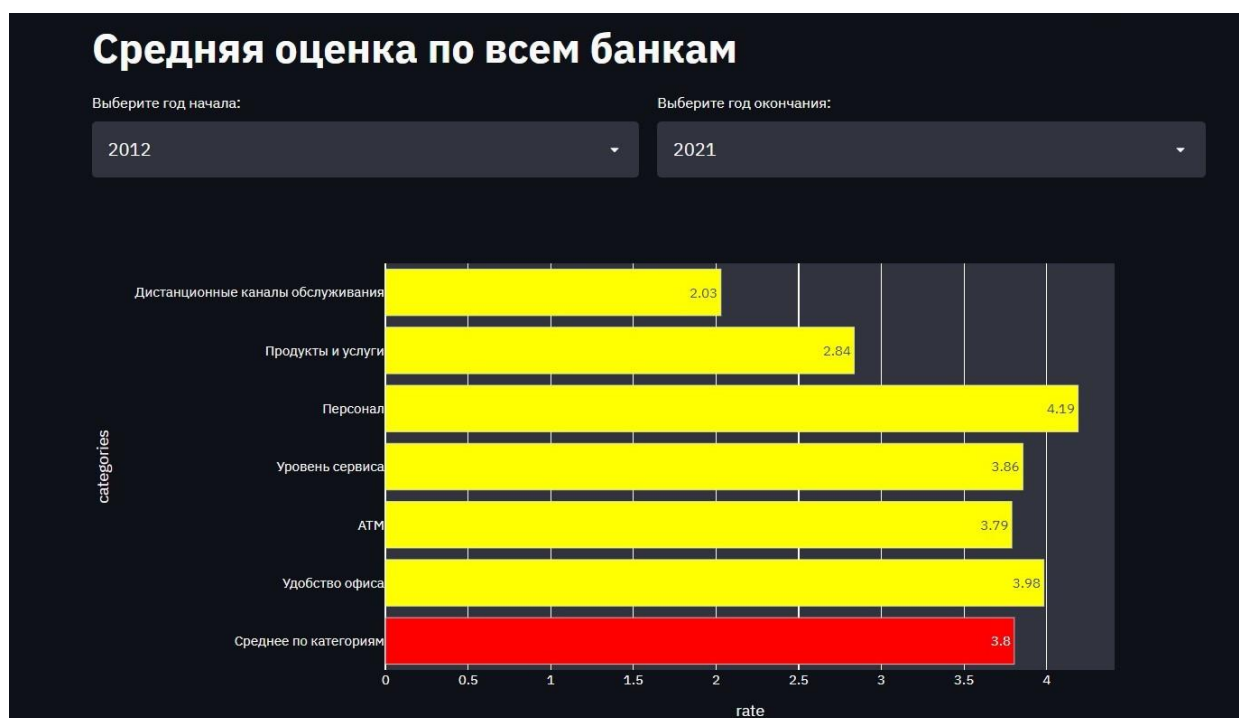
☐ Важные слова для различных категорий

☐ Интерактивная карта банковских отделений

Изображение 5. Меню на сайте

7.1 Страница «Средняя оценка по всем банкам»

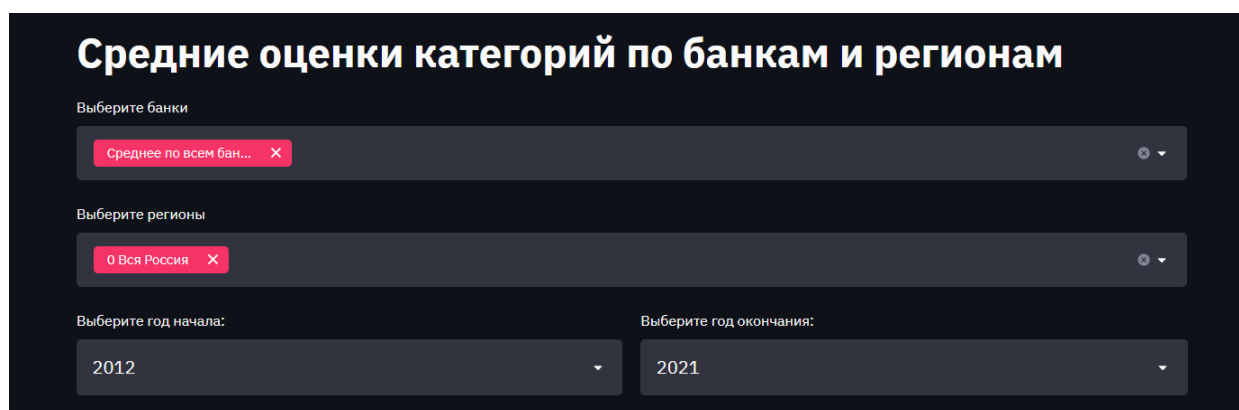
При переходе на первую страницу пользователь может увидеть диаграмму средней оценки по всем банкам. Имеется возможность выбора периода времени, за который будут выведены данные. Это диаграмма отлично подходит для анализа изменения клиентского опыта во всех банках в целом за определённый период времени.



Изображение 6. Горизонтальная гистограмма на странице «Средняя оценка по всем банкам»

7.2 Средние оценки категорий по банкам и регионам

На второй странице пользователю доступна диаграмма средних оценок категорий по банкам и регионам, представленная в формате “паутинки”. Такой формат позволяет легко сравнить между собой несколько банков. Для данной диаграмме также возможен выбор периода времени, банков (не больше 4) и регионов, которые будут соответственно распространяться и на другие диаграммы этой страницы.

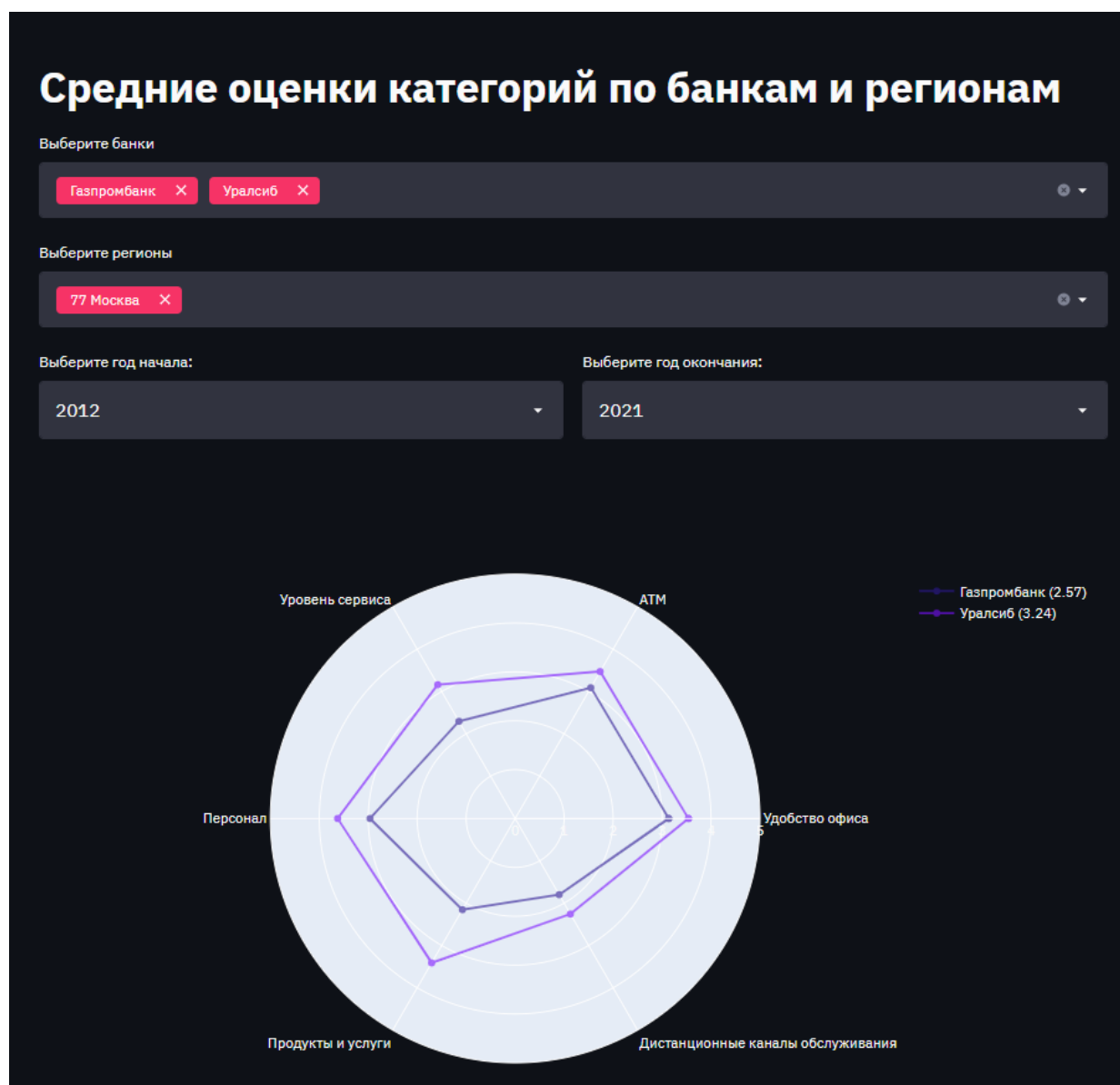


The screenshot shows a web form titled "Средние оценки категорий по банкам и регионам" (Average category ratings by banks and regions). The form is set against a dark background with light-colored text and input fields. It contains four main sections for parameter selection:

- Выберите банки** (Select banks): A dropdown menu showing "Среднее по всем банкам..." (Average for all banks...) with a red 'X' icon to clear the selection and a small gear icon to the right.
- Выберите регионы** (Select regions): A dropdown menu showing "0 Вся Россия" (0 All Russia) with a red 'X' icon to clear the selection and a small gear icon to the right.
- Выберите год начала:** (Select start year): A dropdown menu showing "2012".
- Выберите год окончания:** (Select end year): A dropdown menu showing "2021".

Изображение 7. Пример заданных параметров для диаграммы формата «паутина»

Выше представлены начальные значения параметров.



Изображение 8. Пример диаграммы в формате «паутина»

На этом изображении Вы можете увидеть пример выбора двух банков и региона.

Ниже находится диаграмма сравнения процентного соотношения отзывов с положительной и отрицательными оценками для каждого выбранного банка по категориям в формате торнадо. Эта диаграмма представлена в трёх версиях. Первая показывает все распределения оценок, а остальные две показывают лишь те, которые существенно отличаются от итогового среднего в банке либо отличают выбранные банки по категориям между собой.

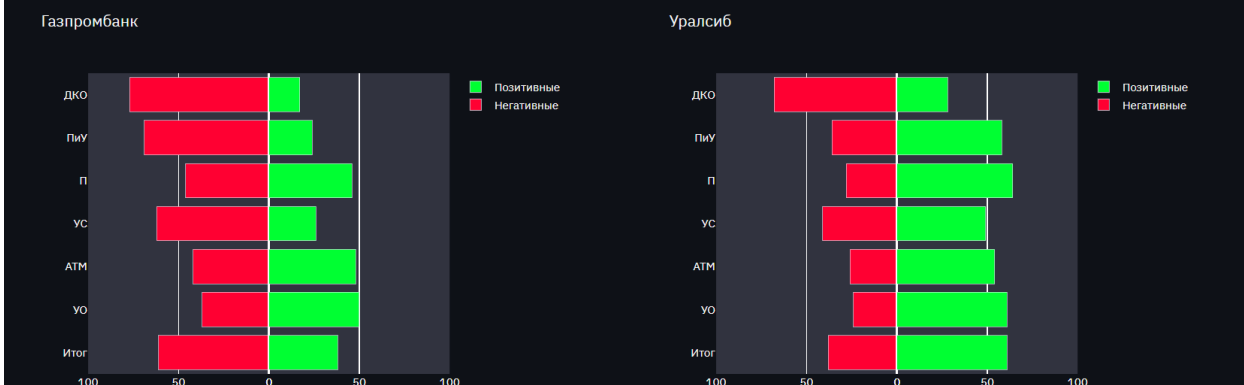
Диаграмма сравнения процентного соотношения отзывов с положительной и отрицательной оценками для каждого выбранного банка по категориям

Положительные — оценка 4 или 5

Отрицательные — оценка 1 или 2

Выберите тип диаграммы "торнадо"

- ☒ Показать все категории
- ☐ Показать только те категории, которые существенно отличаются от итогового распределения внутри банка
- ☐ Показать только те, которые сильно отличают выбранные банки



Изображение 9. Диаграмма в формате «торнадо» (вариант 1)

Пользователь может самостоятельно выбрать, какую именно версию смотреть.

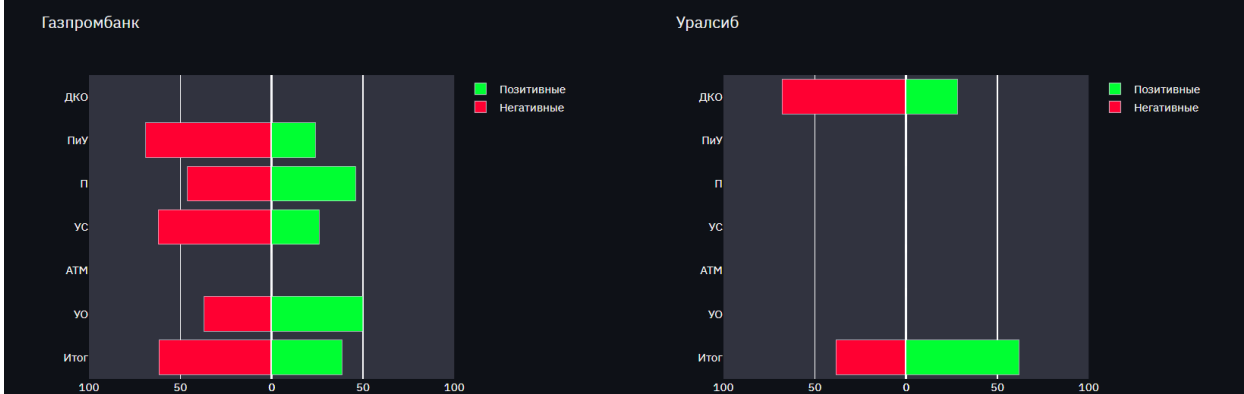
Диаграмма сравнения процентного соотношения отзывов с положительной и отрицательной оценками для каждого выбранного банка по категориям

Положительные — оценка 4 или 5

Отрицательные — оценка 1 или 2

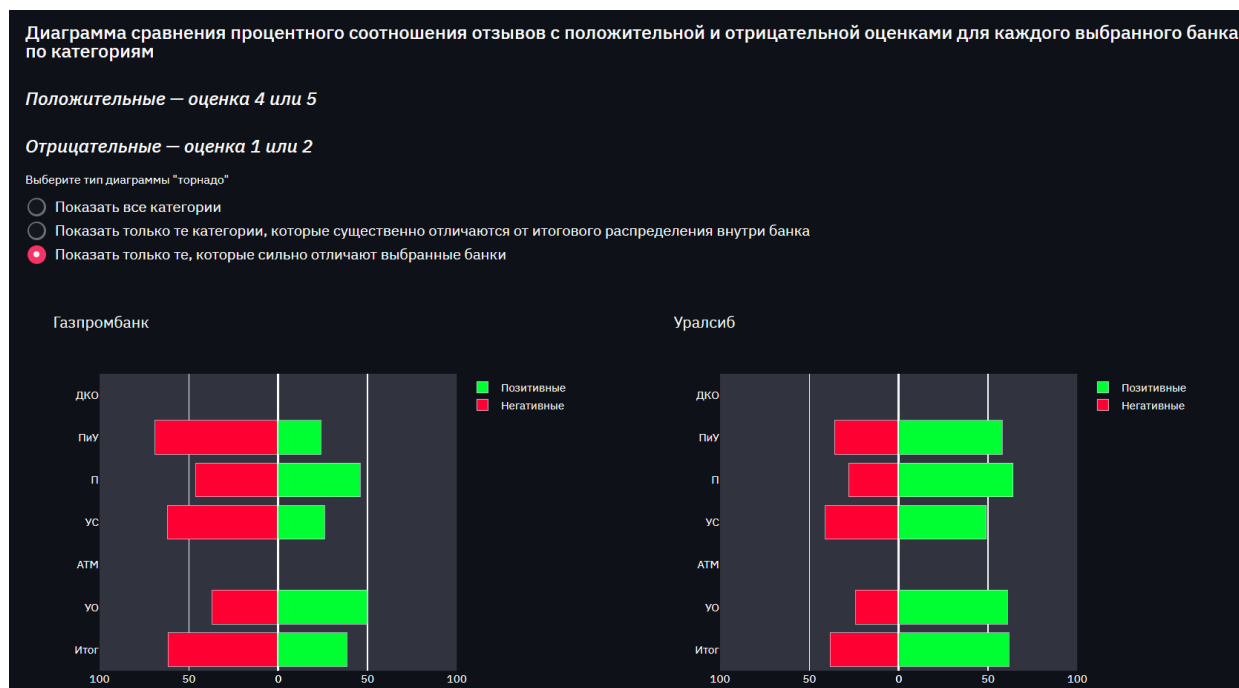
Выберите тип диаграммы "торнадо"

- ☐ Показать все категории
- ☒ Показать только те категории, которые существенно отличаются от итогового распределения внутри банка
- ☐ Показать только те, которые сильно отличают выбранные банки



Изображение 10. Диаграмма в формате «торнадо» (вариант 2)

При просмотре этой диаграммы, можем сделать вывод, что в Газпромбанке схожи с средним только две категории, а у Уралсиба, наоборот, отличается от итога только одна.



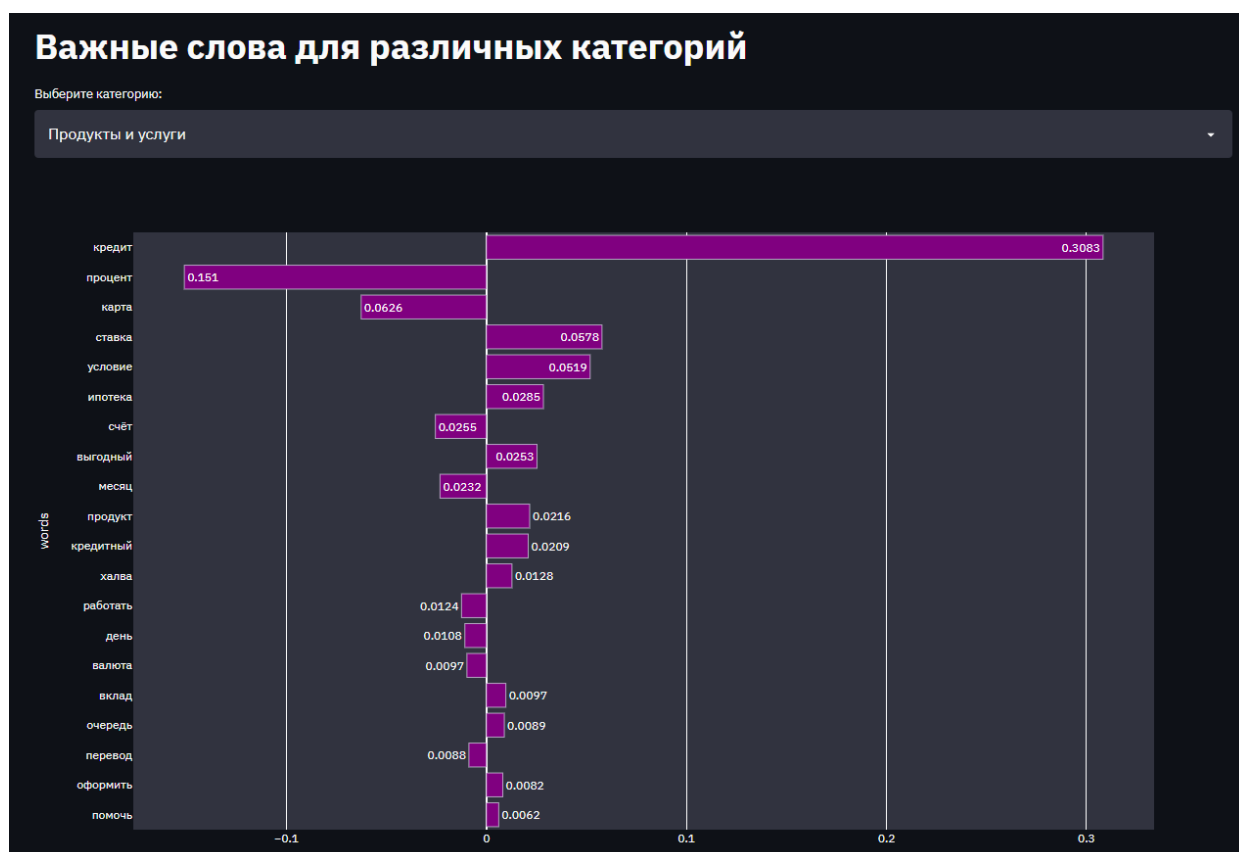
Изображение 11. Диаграмма в формате «торнадо» (вариант 3)

И здесь мы можем увидеть, соотношение отзывов в выбранных нами банках совпадает только в двух из шести категориях.

7.3 Страница «Важные слова для различных категорий»

Когда мы создавали модели для категоризации отзывов, выделялись отдельные слова, которые являются для определения отношения отзыва к той или иной категории. Мы решили составить по ним диаграмму с учётом их важности, а ещё мы учли, как те или иные слова влияют на оценку, оставленную пользователем. В результате чего у нас получилась такая диаграмма. В выпадающем списке можно выбрать категорию, для которой пользователь хочет посмотреть диаграмму.

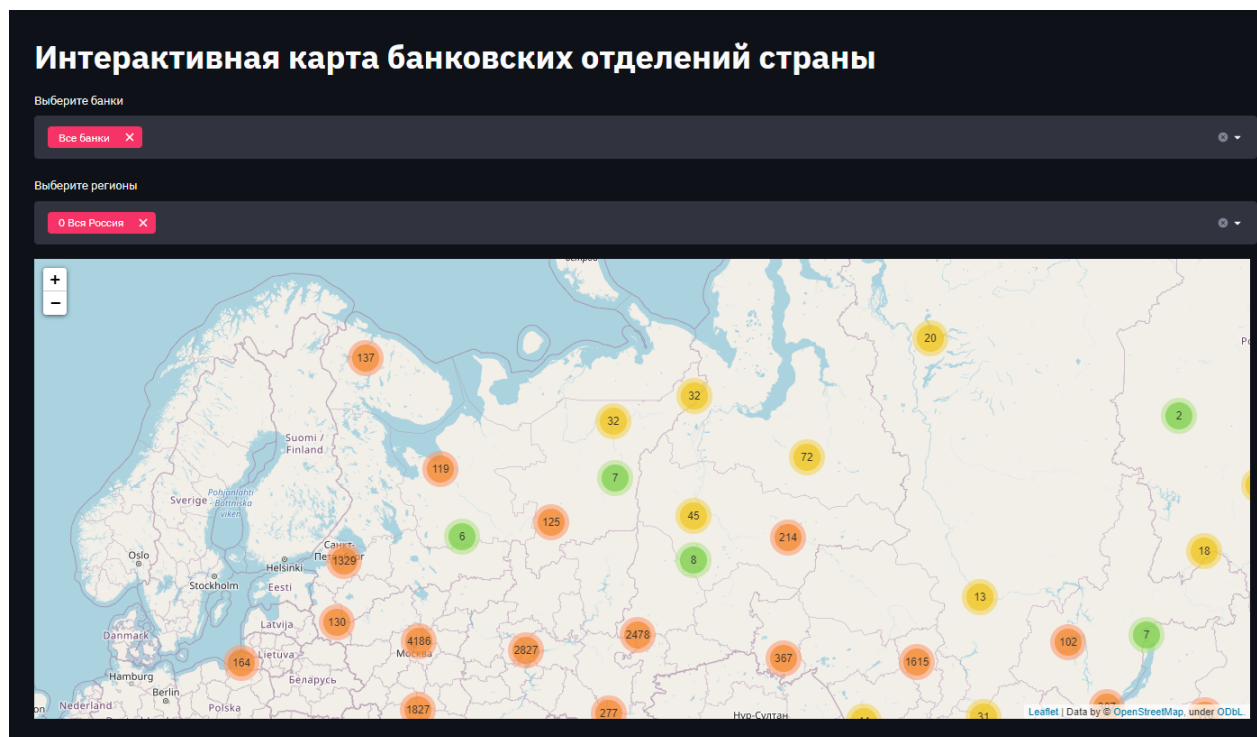
Нами представлены только 20 самых важных слов, они расположены построчно по убыванию важности. Каждой лексической единице соответствует коэффициент важности.



Изображение 12. Важные слова для категории «Продукты и услуги»

7.4 Страница «Интерактивная карта банковских отделений страны»

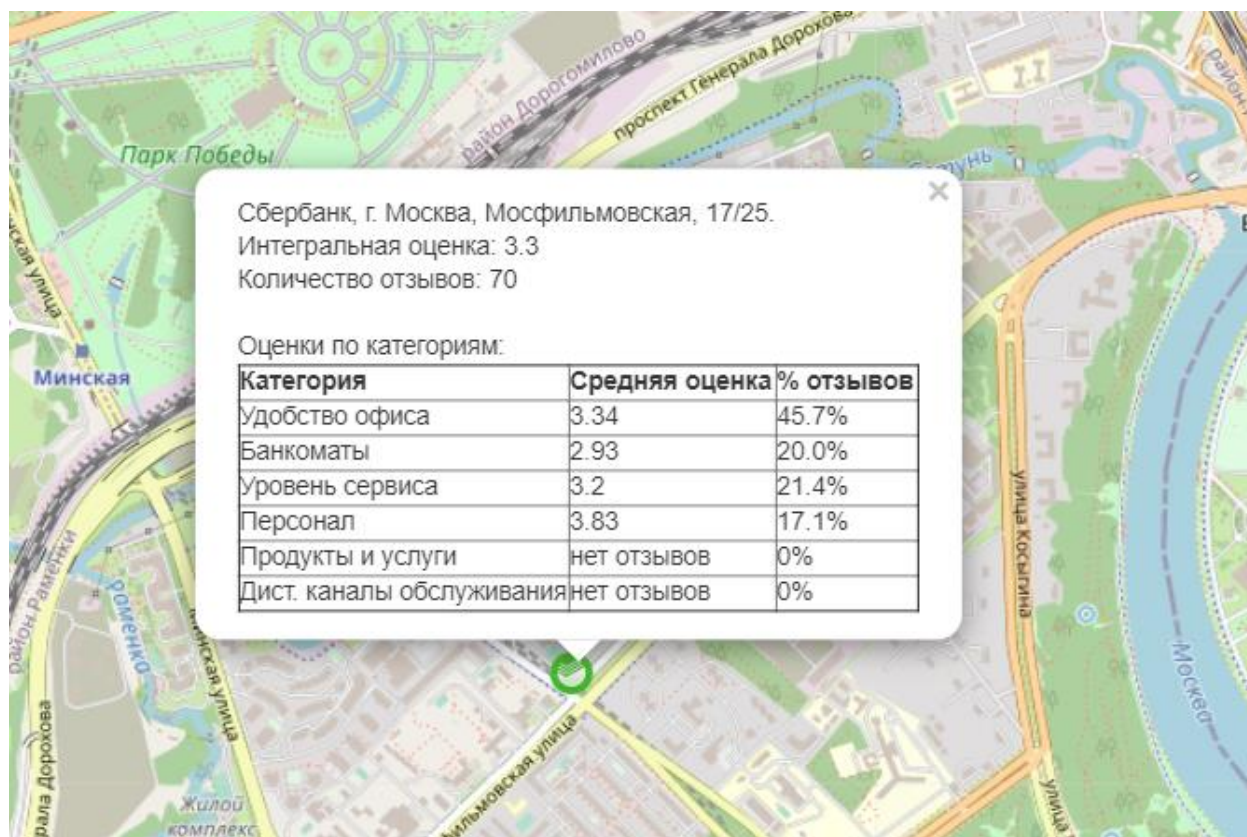
На четвёртой странице пользователю доступен выбор банков и регионов для отображения банковских отделений на карте.



Изображение 13. Интерактивная карта банковских отделений страны

При выборе отделения пользователь может увидеть его подробную характеристику.

Отображается название банка, адрес, средняя оценка, количество отзывов, средние оценки и % отзывов (от общего количества) по категориям.



8. Выводы

За последние три года уровень предоставления банковских услуг в отделениях увеличился (средний показатель вырос на 0.5 балла из 5).

Таблица 2. Сравнение средних оценок и долей категорий за 2017-2018гг. и 2020-2021гг.

	Средняя оценка за 2017-2018	Доля категории во всех отзывах (2017-2018)	Средняя оценка за 2020-2021	Доля категории во всех отзывах (2020-2021)	Изменение средней оценки	Изменение доли категории
Удобство офиса	3.59	16.8%	4.02	17.6%	+0.43	+0.8%
Банкоматы	3.26	15.1%	3.91	9.8%	+0.65	-5.3%
Уровень сервиса	3.3	35.6%	3.87	30%	+0.57	-5.6%
Персонал	3.63	24%	4.22	28.4%	+0.59	+4.4%
Продукты и услуги	2.38	6.9%	2.77	6.8%	+0.39	-0.1%
Дист. каналы обслуживания	1.63	2.2%	2.03	1.5%	+0.4	-0.7%

Рост средней оценки вызван огромным ростом средних оценок в каждой из категорий. За три года предпочтения клиентов банковских отделений изменились: пользователи стали чаще писать про персонал, но реже про банкоматы и уровень сервиса.

Изменения средних оценок вызваны резким скачком улучшения предоставления услуг у главного банка России – Сбербанк (занимает ~50% банковского рынка РФ). У остальных банков из топ-10 списка, который приведен в «Постановке задачи», особых изменений в оценках нет. Сбербанк смело можно назвать драйвером изменений на всем рынке за три года.

Можно сказать, что Сбербанк и Тинькофф, который нацелен на дистанционное обслуживание, задали новый тренд в дистанционных каналах обслуживания (средняя оценка за 2017-2018 у Сбербанка 1.79, 2020-2021 – 2.86). Несмотря на уменьшение доли данной категории во всех отзывах люди стали чаще оценивать приложения (скорее всего, отзывы на приложения клиенты оставляют в магазинах Apple Store и Play Market) и колл-центры банков; у многих банков за этот период дистанционные каналы появились. Клиенты банков оставляют отзывы на данную категорию лишь в случае, когда они посещают отделение банка по причине плохой работы приложения или колл-центра, именно поэтому эта категория имеет самый низкий средний балл (т.е. люди жалуются).

Клиенты банков стали гораздо чаще стали оставлять отзывы на картографических сервисах (например, Яндекс.Карты, отзывы с которого мы обзоредали). За 2017-2018 годы пользователи написали около 25 тысяч отзывов, а за 2020-2021 – уже 380 тысяч. Как мы можем заметить, количество отзывов за последний период увеличилось более чем в 15 раз.

9. Перспективы дальнейшей разработки

1. Анализ отзывов пользователей на мобильные приложения банков из магазина приложений Google Play;
2. Анализ развернутых отзывов клиентов об услугах банков в «Народном рейтинге» на сайте banki.ru.

Эта задача является более сложной, так как на «Народном рейтинге» пользователи в основном пишут отзывы не о конкретных плюсах (компетентный персонал, отсутствие очередей и пр.) или минусах отделения (неудобные часы работы, неработающие банкоматы и пр.), а о банковском обслуживании в целом.

10. СПИСОК ИСТОЧНИКОВ

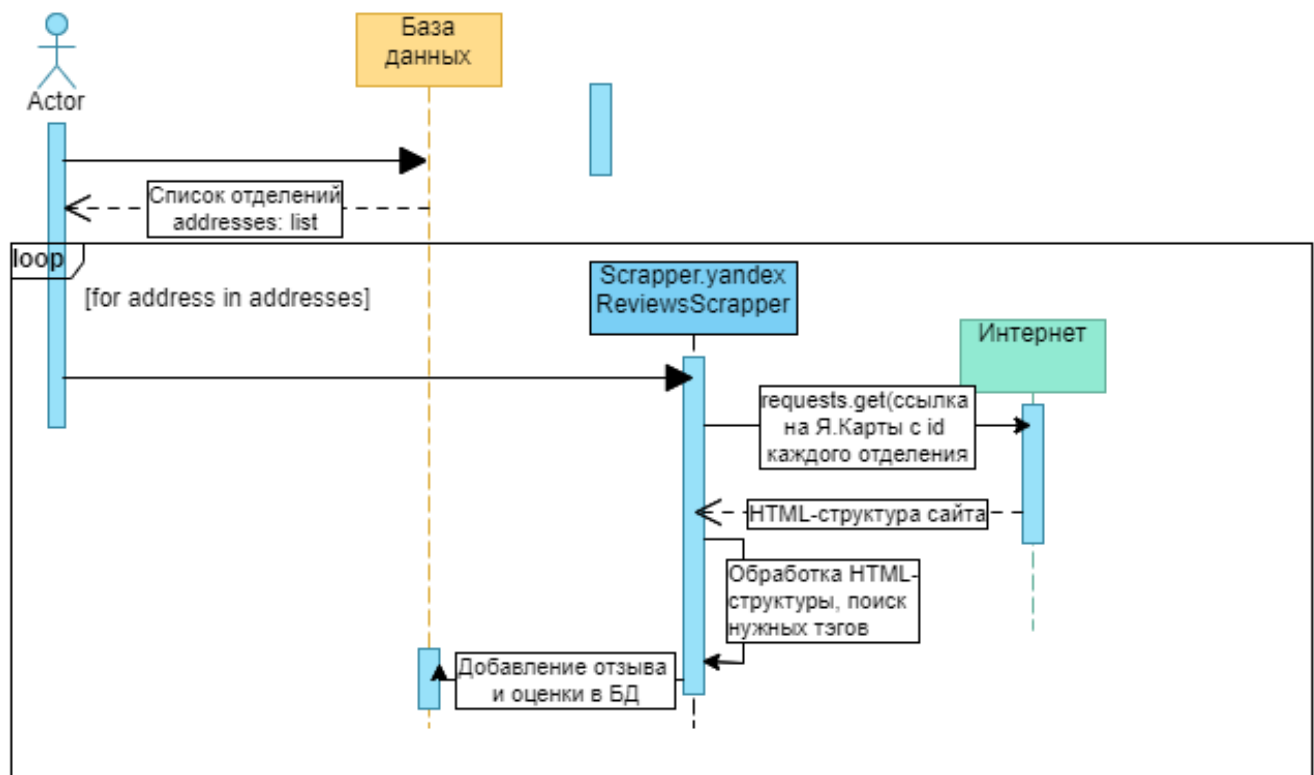
1. ООО «Яндекс»: О сервисе. API поиск по организациям [Электронный ресурс]: 2020 — Режим доступа: <https://yandex.ru/dev/maps/geosearch/doc/concepts/about.html?from=geosearch> (Дата обращения 01.07.2021)
2. Voximplant: Основы Natural Language Processing для текста / Блог компании Voximplant / Хабр [Электронный ресурс]: 15.04.2019 — Режим доступа: <https://habr.com/ru/company/Voximplant/blog/446738/> (Дата обращения 21.08.2021)
3. Mikhail Korobov Revision: Документация — Морфологический анализатор pymorphy2 [Электронный ресурс]: 2013-2020 — Режим доступа: <https://pymorphy2.readthedocs.io/en/latest/user/index.html> (Дата обращения 21.07.2021)
4. Список регионов (субъектов, областей) России 2021 РФ с кодами согласно данным ФНС по алфавиту [Электронный ресурс]: 2020 — Режим доступа: <https://www.sites.google.com/site/ruregdatav1/spisok-regionov-rossii-s-kodamy> (Дата обращения 20.01.2021)
5. UdeMy: Бесплатное учебное руководство по теме “Обработка и анализ данных” [Электронный ресурс]: 2021 — Режим доступа: <https://www.udemy.com/course/knime-bootcamp/> (Дата обращения 05.09.2021)
6. Working With Text Data — scikit-learn 0.24.2 documentation [Электронный ресурс]: 2020 — Режим доступа: https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html (Дата обращения 10.09.2021)
7. 3.1 Cross-validation: evaluating estimator performance — scikit-learn 0.24.2 documentation [Электронный ресурс]: 2020 — Режим доступа: https://scikit-learn.org/stable/modules/cross_validation.html (Дата обращения 10.09.2021)
8. Receiver Operating Characteristic (ROC) with cross validation — scikit-learn 0.24.2 documentation [Электронный ресурс]: 2020 — Режим доступа: https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc_crossval.html (Дата обращения 11.09.2021)
9. Краткий обзор алгоритма машинного обучения Метод Опорных Векторов (SVM) [Электронный ресурс]: 01.11. 2018 — Режим доступа: <https://habr.com/ru/post/428503/> (Дата обращения 10.09.2021)
10. Welcome to Streamlit — Streamlit 0.82.0 documentation [Электронный ресурс]: 2021 — Режим доступа: <https://docs.streamlit.io/en/stable/> (Дата обращения 30.09.2021)

11. Plotly Python Graphic Library | Python | Plotly [Электронный ресурс]: 2021 — Режим доступа: <https://plotly.com/python/> (Дата обращения 01.10.2021)
12. Классические методы статистики: критерий хи-квадрат [Электронный ресурс]: 05.08.2012 — Режим доступа: <https://r-analytics.blogspot.com/2012/08/blog-post.html?m=1> (Дата обращения 15.09.2021)
13. Running Chi-Square Tests with Die Roll Data in Python | By Jake Huneycutt | Towards Data Science [Электронный ресурс]: 07.05.2018 — Режим доступа: <https://towardsdatascience.com/running-chi-square-tests-in-python-with-die-roll-data-b9903817c51b> (Дата обращения 15.09.2021)
14. Критерий согласия Пирсона χ^2 (Хи-квадрат) | statanaliz.info [Электронный ресурс]: 07.10.2019 — Режим доступа: <https://statanaliz.info/statistica/proverka-gipotez/kriterij-soglasiya-pirsona-khi-kvadrat/> (Дата обращения 15.09.2021)
15. Random Forest Feature Importance Computed in 3 Ways with Python | MLJAR Automated Machine Learning [Электронный ресурс]: 2021 — Режим доступа: <https://mljar.com/blog/feature-importance-in-random-forest/> (Дата обращения 15.10.2021)
16. Градиентный бустинг — просто о сложном [Электронный ресурс]: 27.11.2018 — Режим доступа: <https://neurohive.io/ru/osnovy-data-science/gradientyj-busting/> (Дата обращения 20.10.2021)
17. Статистический анализ гипотез в Python с корреляцией Anova, Хи-квадрат и Пирсона - pythobyte.com [Электронный ресурс]: 2021 — Режим доступа: <https://pythobyte.com/statistical-hypothesis-analysis-in-python-with-anovas-chi-square-and-pearson-correlation-be15ad06/> (Дата обращения 15.09.2021)
18. Folium — Folium 0.12.1 documentation [Электронный ресурс]: 2021 — Режим доступа: <https://python-visualization.github.io/folium/> (Дата обращения 01.10.2021)

11. Приложения

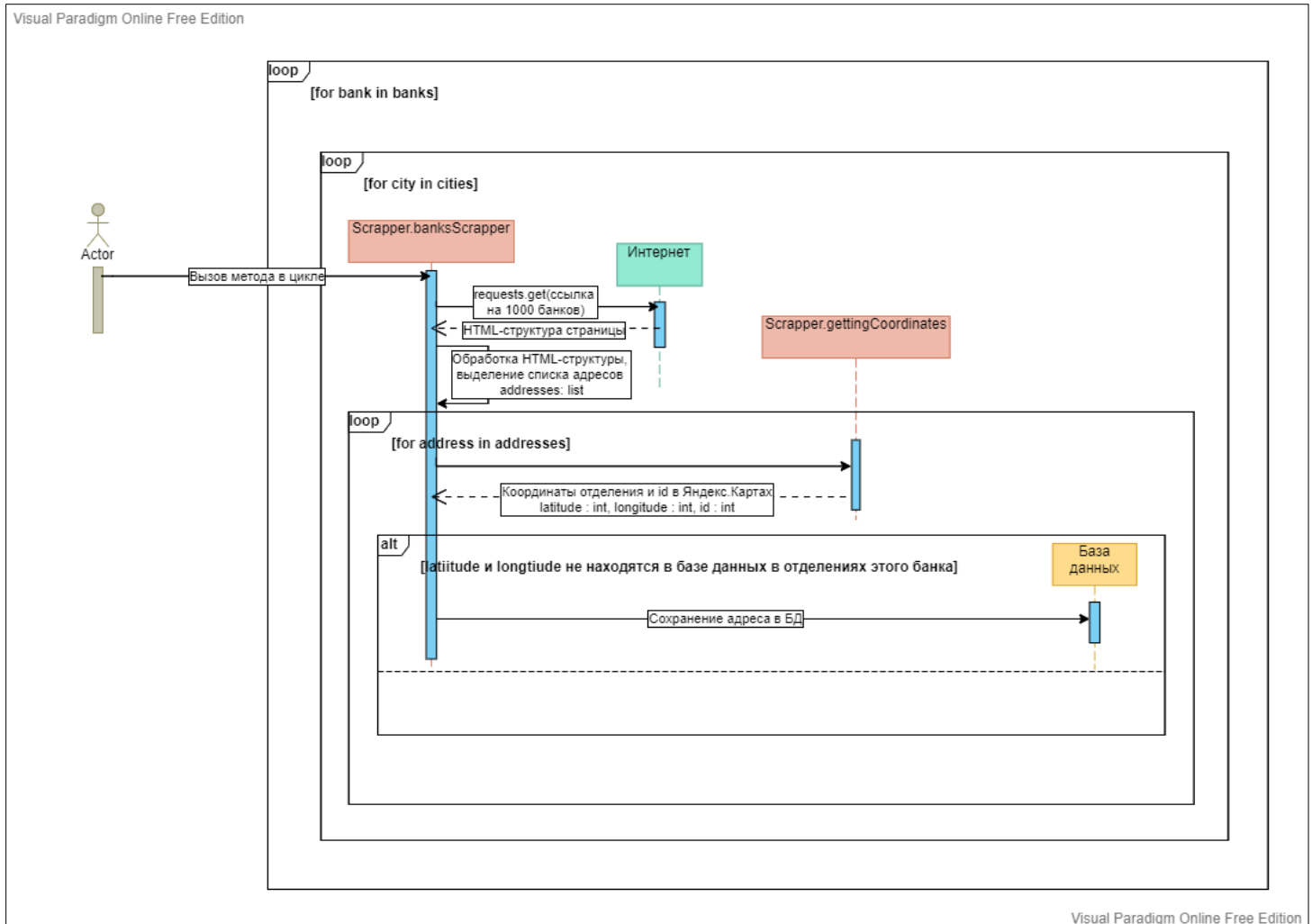
A. Sequence Diagram (диаграмма последовательностей) для сбора отзывов с сайта <https://yandex.ru/maps>

Visual Paradigm Online Free Edition



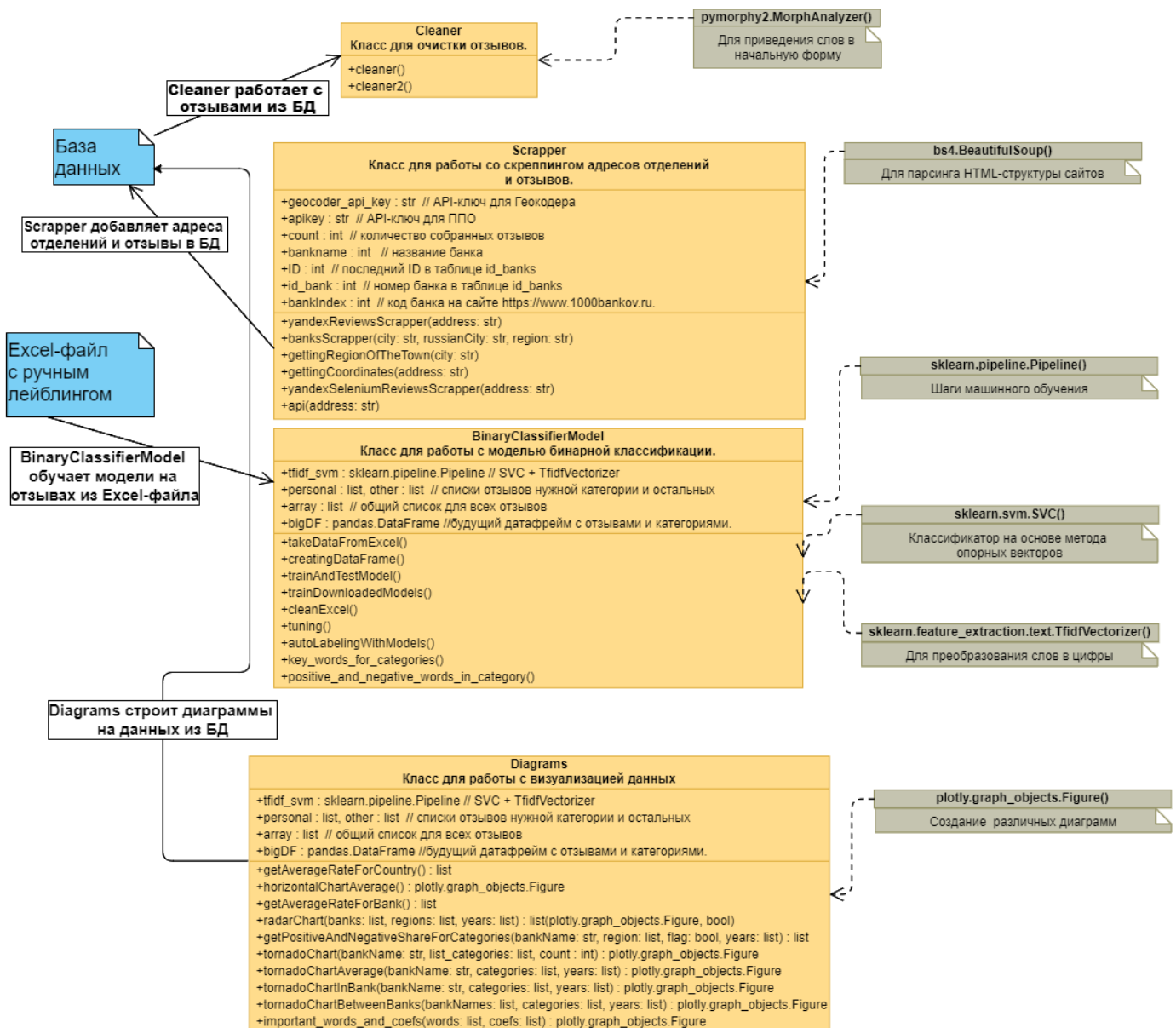
Visual Paradigm Online Free Edition

Б. Sequence Diagram (диаграмма последовательностей) для сбора отделений с сайта <https://1000bankov.ru>



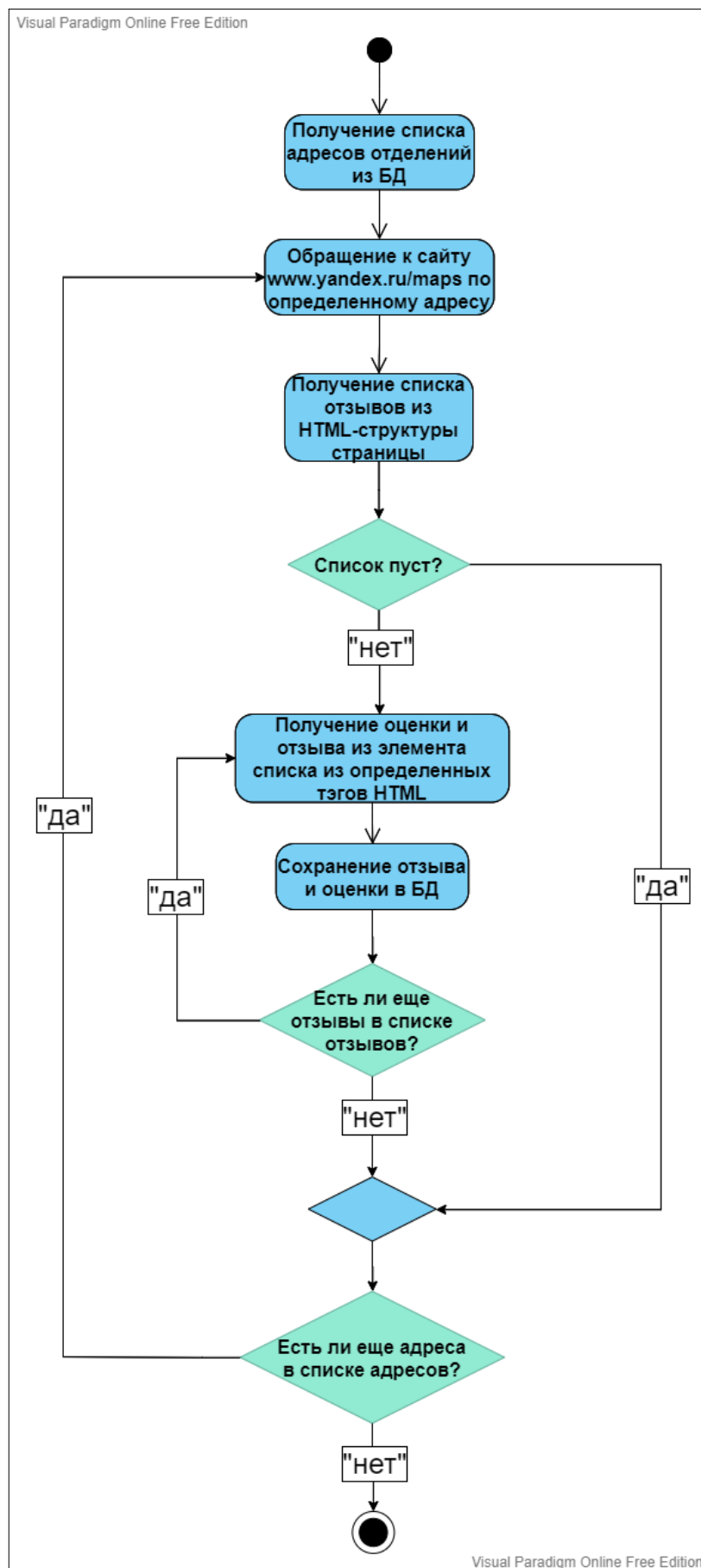
B. Class Diagram (диаграмма классов)

Visual Paradigm Online Free Edition

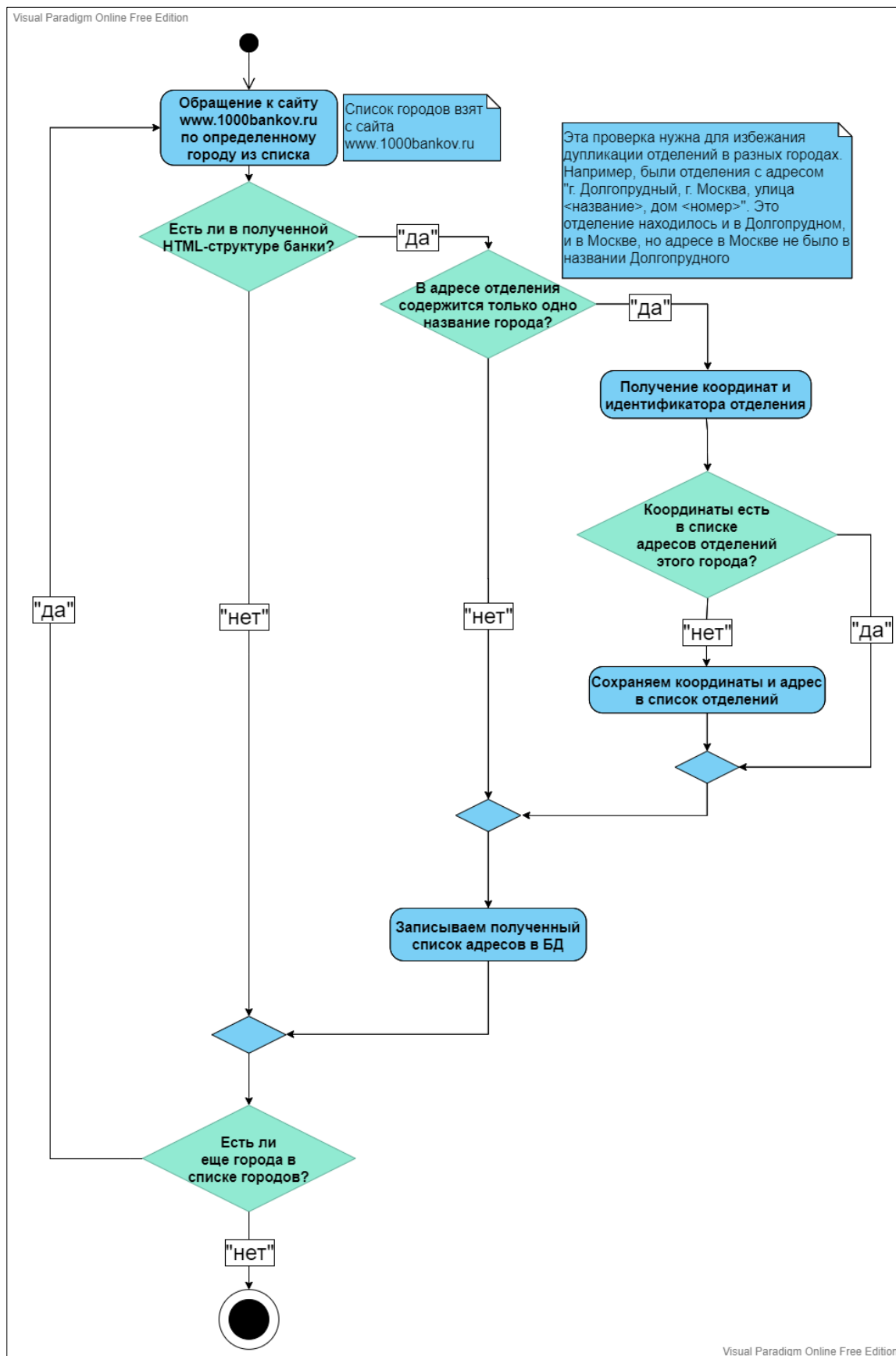


Visual Paradigm Online Free Edition

Г. Activity diagram (диаграмма деятельности) для сбора отзывов



Д. Activity diagram (диаграмма деятельности) для сбора отделений



E. Component diagram (диаграмма компонентов)

