

Анализ отзывов клиентов на банковские каналы обслуживания и продукты/услуги

Выполнили: Рукавица Артём Кириллович, Федченко Анастасия Кирилловна, 11 класс

ГБОУ Школа №1533 «ЛИТ», г. Москва

Научный руководитель: Чамров М.В., ПАО Банк «ФК Открытие», вице-президент, лидер трайба
необеспеченное кредитование

Инженерные науки, секция: IT

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.

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, уровень сервиса, персонал, продукты и услуги, дистанционные каналы обслуживания.

Причиной **актуальности** нашего продукта является отсутствие в открытом доступе программ/сайтов, занимающихся сравнением клиентского опыта в различных банках РФ. Мы предполагаем, что **целевой аудиторией** нашего сайта будут две группы людей. **Первая** — пользователи, которые смогут увидеть сильные и слабые стороны банков и выбрать подходящий для них. **Вторая** — это компании, которые хотят улучшить процесс предоставления услуг в условиях постоянно изменяющихся предпочтений пользователей конкурентной среды.

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.

Для создания нашего проекта 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.

Кратко опишем, как решается наша задача.

1. Сбор адресов отделений банков с сайта <https://1000bankov.ru> и отзывов с сайта <https://yandex.ru/maps>
2. Подготовка отзывов к анализу: удаление стоп-слов, знаков препинания, приведение слов в начальную форму;
3. Ручная разметка отзывов на случайной выборке, состоящей из 1600 отзывов, на основе категоризации;
4. Обучение моделей бинарной классификации на основе размеченных отзывов из пункта 4 при помощи метода опорных векторов;
5. Категоризация всех отзывов при помощи обученных моделей;

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

Результатом нашей работы является сайт, предоставляющий визуализацию проведённого анализа на собранных данных. На сайте пользователю доступно три типа диаграмм и интерактивная карта:

1. горизонтальная гистограмма (для отображения средних оценок и важных слов; см. рис. 2);
2. диаграмма в формате «паутина» (для удобного отображения средних оценок по категориям; см. рис. 1);
3. диаграмма в формате «торнадо» (для отображения всех распределений положительных и отрицательных отзывов по категориям и в среднем);
4. интерактивная карта банковских отделений РФ с подробной аналитикой для каждого отделения из базы данных (см. рис. 3).

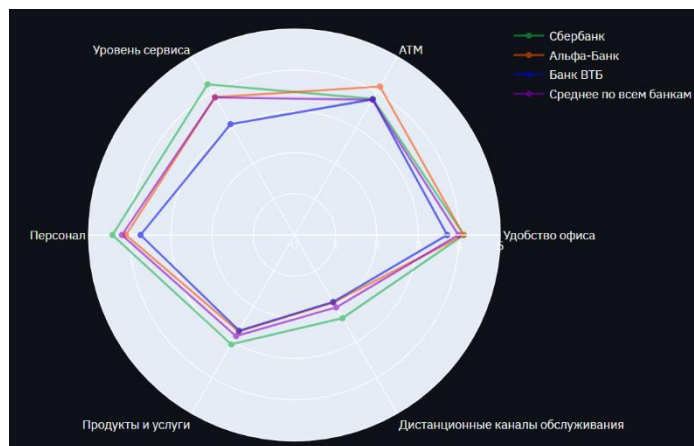


Рисунок 1. Пример диаграммы-«паутины»



Рисунок 2. Средние оценки по категориям по всем банкам

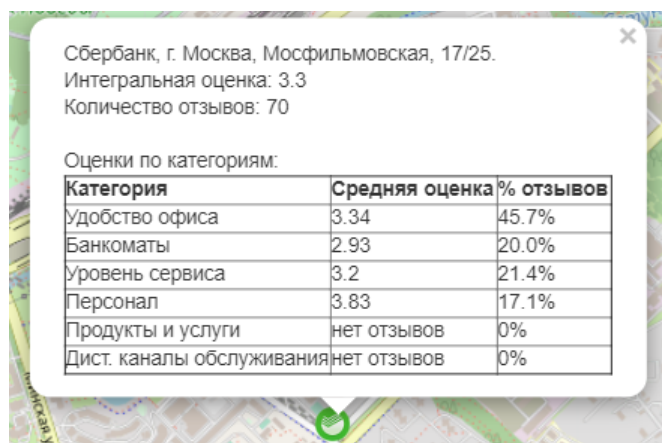


Рисунок 3. Подробная информация об одном из отделений банка на интерактивной карте

Нами создан продукт, подход которого 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 (т.е. про что пользователи пишут).

Литература:

1. 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 04.10.2021)
2. Elizabeth Mays. «Credit Scoring for Risk Managers. The Handbook for Lenders» 2004 г., с. 63-89
3. Краткий обзор алгоритма машинного обучения Метод Опорных Векторов (SVM) [Электронный ресурс]: 01.11. 2018 — Режим доступа: <https://habr.com/ru/post/428503/> (Дата обращения 05.09.2021)