





Инструмент для анализа клиентских отзывов

Проект в рамках всероссийской научнотехнологической программы по решению проектных задач в области искусственного интеллекта и смежных дисциплин «Сириус.ИИ»

АО «Тинькофф Б анк» Март, 2024

Проектная команда «SpectrUM»:

71 регион



Кончаков Павел 10 класс Капитан

71 регион



Шабаев Андрей 10 класс

71 регион



Григорьев Илья 10 класс







71 регион



Ведешкин Тимофей 10 класс



PRO AI SPECTRUM

Актуальность проекта:

1. Конкурентоспособность и адаптация:

Разработка инструмента для анализа клиентских отзывов позволит банку Тинькофф повысить свою конкурентоспособность за счет адаптации своих продуктов и услуг под потребности клиентов. Это особенно важно в контексте острой конкуренции на рынке банковских услуг.

2. Улучшение обслуживания и удовлетворенности клиентов:

Путем анализа отзывов пользователей банк сможет выявить слабые места в обслуживании и недостатки продуктов, что позволит ему принимать меры для их улучшения. Это в свою очередь повысит уровень удовлетворенности клиентов и способствует повышению лояльности.



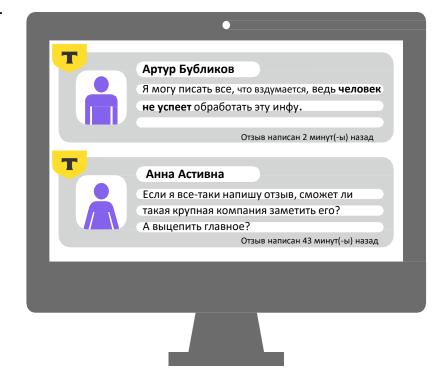
Проблематика проекта:

Возможное отсутствие оперативной реакции на изменения рыночной среды:

Без эффективного инструмента анализа клиентских отзывов банк может быть лишен возможности оперативно реагировать на изменения в требованиях рынка и предпочтениях клиентов, что затрудняет его конкурентоспособность и способность к инновациям. Отстутствие достаточного времени для выявление главной темы отзывов.

Недостаточная адаптация продуктов под потребности клиентов:

Отсутствие систематического анализа клиентских отзывов может приводить к тому, что банк не полностью учитывает потребности и ожидания своих клиентов при разработке и улучшении продуктов и услуг. Это может привести к разочарованию клиентов и потере лояльности.



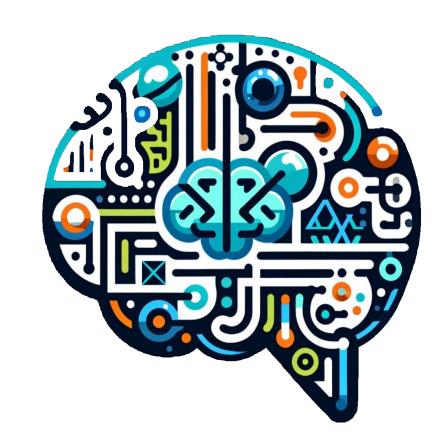
Идея проекта:



Создать инструмент, который сможет **обрабатывать** информационные ресурсы и **собирать** отзывы.

Отзывы будут проходить анализ нейронной сетью, помимо тега 'хороший отзыв' или 'плохой отзыв',

Мы хотим научить нейронную сеть **понимать отзыв** и **резюмировать** его, **выделяя** самое главное (что не хватает пользователю).





Главная цель проекта "Инструмент для анализа клиентских отзывов" заключается в создании эффективного и инновационного инструмента, который позволит банку Тинькофф систематически собирать, анализировать и использовать клиентскую обратную связь Анна Астивна



Азамат Про777

Ну ваще все кул, но вот кнопочка в вашем сервисе не нравится, мб исправите?

Задачи проекта:

- 1. Разработка алгоритма сбора данных.
- 2. Чистка данных от лишнего.
- 3. Извлечение признаков.
- 4. Анализ и интерпретация отзывовю



Артур Пирожок

Я переводила деньги своему продюссеру, мне не понравилось что во вкладке

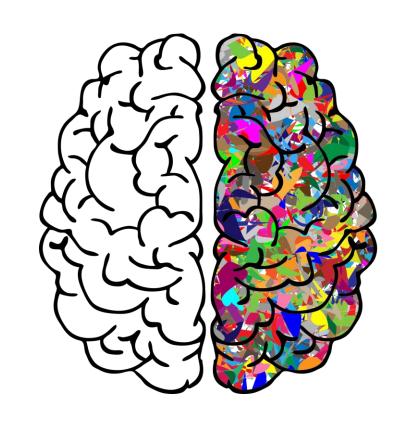
'перевод по договору' очень много инфы.

Мне не понравился новый дизайн аватарки Тинькофф, зачем Вам это?!



Анализ области:

- Где клиенты могут оставлять отзывы на продукты 'Тинькофф'?
 - 1. Социальные сети и сайты
 - 2. Анкеты и опросы
 - 3. Рейтинг на картах, в магазинах приложений (Apple Store, Google Play, RuStore и так далее)
- Как можно их собирать?
 - 1. Автоматизированные системы парсинга.
 - 2. За счёт размещения опросов на интернетресурсах



Анализ области:

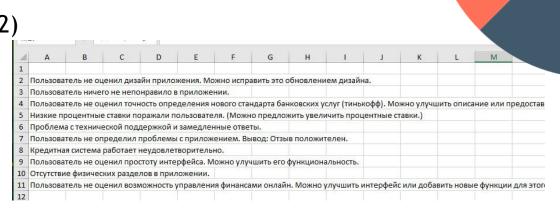


• Как можно их обрабатывать?

Анализ сообщений локальной нейронной сетью:

- (1) С целью выявления положительных и отрицательных отзывов.
- (2) С целью выявления основной проблемы отзыва.





^{*}скриншоты принадлежат команде Spectrum, сделаны во время этапов работы с проектом от АО «Тинькофф Банк»

Анализ области:



Сравнение решений разных компаний, сопоставление с нашей идеей:

Компания:	Решение:	
Ростелеком	Операторы отвечают на звонки вручную. Ввели речевую аналитику для улучшения клиентского опыта.	×
Steam Steam	Не хотят видеть и принимать критику на платформе. Имеют очень хитрые система обхода отрицательных отзывов.	X
Riot Games	Решают проблему отрицательных отзывов путём "хейта" других платформ (Steam, Epic games). Тем самым, переманивают игроков на свою платформу.	×
м МеДси	Анализируют отзывы клиентов, но делают это вручную. Не стремятся к созданию автоматизированных инструментов	×
Т Наше решение	С помощью систем парсинга собираем отзывы с интернет-ресурсов, используем ИИ для анализа семантики отызыва, берем главное и решаем проблему!	√

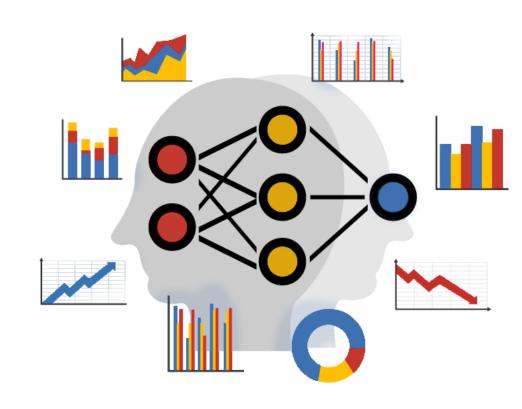
Первоначальный вариант решения:



- 1. Использование локальной, готовой, обученной на большой базе данных модели.
- 2. Вариативная система парсинга (сбор данных с разных интернет-ресурсов).
- 3. Написанный сценарий(код), умеющий отсылать запросы и интегрировать с ИИ-моделью.

Плюсы:

- Использование локальной нейронной сети позволит нам разворачивать ее прямо в коде Python, отсутствует необходимость использовать ВПН, клиенты браузеров для выгрузки ответов.
- Полностью автоматизированный инструмент, не требующий внедрения человека.



Какие модели лучше использовать?



Протестировав **GPT 3.5**, мы пришли к выводу, что использовать данную модель нецелесообразно (хоть и имели api_key для использования), так как:

Использование таких сервисов стоит денег, есть вероятность, что мы можем быть заблокированы за частые запросы. Немало важную деталь играет Использование ВПН для работы с сервисом.

Прибегнув к бесплатным моделям, которые мы можем использовать, выделили Mistral 7B Instruct v0.2, которую и использовали в дальнейшем. Zephyr beta 7b - нейросеть, которая очень быстро и круто работает с отзывами, пробовали на личном примере

Предполагаемый вариант для нашего проекта:

Сайга-Мистраль (третья русская нейросеть после YaGPT и GigaChat, публично доступная по API)

Интегрировать с ней хотим напрямую, в LMStudio не смогли найти бесплатную версию.

Слайд по пунктам №3 и №4

Наше решение на первый этап:

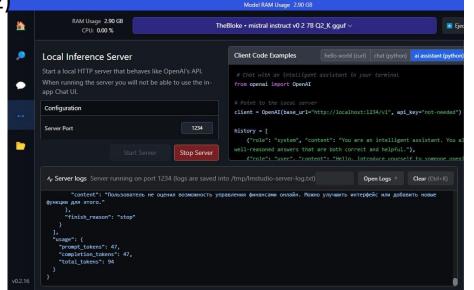


Идея, по нашему мнению, должна быть подкреплена практической частью. Мы готовы ее представить! Мы разработали инструмент, который работает на основе ИИ.

- 1 Наш код (1) посылает запросы с отзывами, которые получены путем парсинга.
- 2 Локальная нейронная сеть (2) обрабатывает их и **вычленяет главное**

(1) Код на языке **Python**

Планы на будущее: Определение наполненности отзывов. Чистка ненужного. Наш изначальный выбор: Mistral 7B Instruct v0.2 (2) Model RAM Usage 250 GB



Планы на будущее: Переход на сайгу-мистраль и взаимодействие с ней напрямую.

В рамках 1-ого этапа нам удалось:

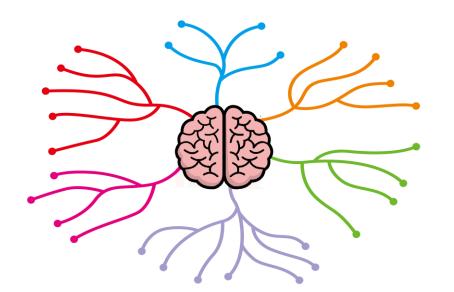


Обработанные данные формировать в таблицу, которая автоматически обновляется (3) на нашем сайте, который мы хотим разработать для удобной и синхронной работы сотрудников компании.

*Сотрудники Тинькофф смогут удобно работать с табличными данными и заниматься решением задач, исходя из проблем пользователя

Будущее проекта:

Работа над логикой сайта и дизайном



Добро пожаловать на Tinkoff-Sirius ИИ!

Зарегистрироваться Войти Таблица

Это содержимое вашей главной страницы.
© 2024 Tinkoff page

Excel Table

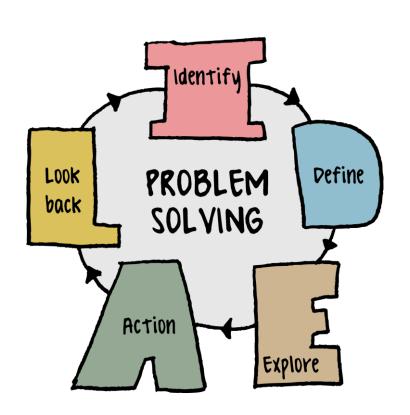
Unnamed: 0	
ользователь не оценил дизайн приложения. Можно исправить это обновлением дизайна.	
ользователь ничего не непонравило в приложении.	
Пользователь не оценил точность определения нового стандарта банковских услуг (тинькофф). Можно улучшить описание или предоставить более конкретные примерь	
Низкие процентные ставки поражали пользователя. (Можно предложить увеличить процентные ставки.)	
роблема с технической поддержкой и замедленные ответы.	
Пользователь не определил проблемы с приложением. Вывод: Отзыв положителен.	
редитная система работает неудовлетворительно.	
Пользователь не оценил простоту интерфейса. Можно улучшить его функциональность.	
тсутствие физических разделов в приложении.	
ользователь не оценил возможность управления финансами онлайн. Можно улучшить интерфейс или добавить новые функции для этого.	

Вернуться на главную

Видео-представление шагов работы:



В рамках первого этапа сделали наш ROADMAP, который представили в видео-формате.





сириус — Яндекс Диск (yandex.ru)

Работа со вторым этапом:

План реализация решения:

- -1 шаг(Поиск источников, с которых мы будем брать информацию):
- Выбрали такие источники, как banki.ru и sravni.ru.

Описание:

Мы хотели рассмотреть парсинг данных с соц.сетей, на основании статистики выявили, что там не так много полезных для нас отзывов, эту статистику можно найти в интернете.

Как мы парсили?

- Для первого источника (banki.ru) использовали bs4 и urllib
- Для второго источника (sravni.ru) использовали selenium и webdriver_manager, взаимодействовали через ChromeDriver

Подробнее о парсерах:

import requests from bs4 import BeautifulSoup import ison import unicodedata def clean html(html content): cleaned_content = ''.join(char for char in html_content if unicodedata.category(char)[0] != 'C') return cleaned_content def extract review body(url): try: response = requests.get(url) if response.status code == 200: encodings = ['utf-8', 'ISO-8859-1', 'windows-1252'] #ХЗ почему-то ругался на энкодинги, сразу ставим много for encoding in encodings: html_content = response.content.decode(encoding).replace('\x00', '') cleaned_content = clean_html(html_content) soup = BeautifulSoup(cleaned_content, 'html.parser') script_tags = soup.find_all('script', type='application/ld+json') for script tag in script tags: script_content = script_tag.contents[0].strip() script_data = json.loads(script_content) if script_data.get('@type', '') == 'Review': review_info = { 'review_name': script_data.get('name', ''), 'author_name': script_data.get('author', {}).get('name', ''), 'rating_value': script_data.get('reviewRating', {}).get('ratingValue', ''), 'review_body': script_data.get('author', {}).get('reviewBody', '') return review_info return "Review body not found." except UnicodeDecodeError as e: print(f"Failed to decode using {encoding}: {e}") print("Failed to decode the HTML content using any of the tried encodings.") print(f"Failed to fetch the webpage at {url}. Status code: {response.status_code}") return None except Exception as e: print(f"An error occurred while fetching the HTML content: {e}") return None url = "https://www.banki.ru/services/responses/bank/response/11397996/" review_body = extract_review_body(url) print("Review Body:") print(review_body)

Для первого источника (banki.ru) использовали bs4 и urllib Обычный парсер через get запросы по ссылке.

Была сложность с responce.

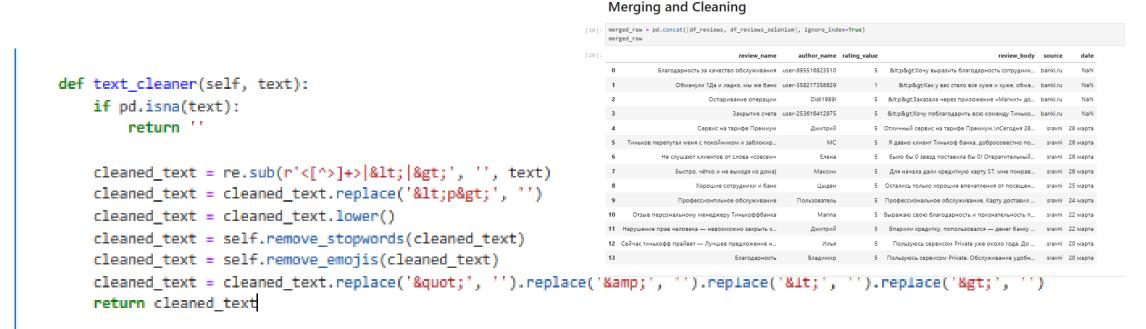
Извлекает данные:

Имя пользователя, дата написания, Рейтинг-оценку отзыва, сам отзыв.

Пример:

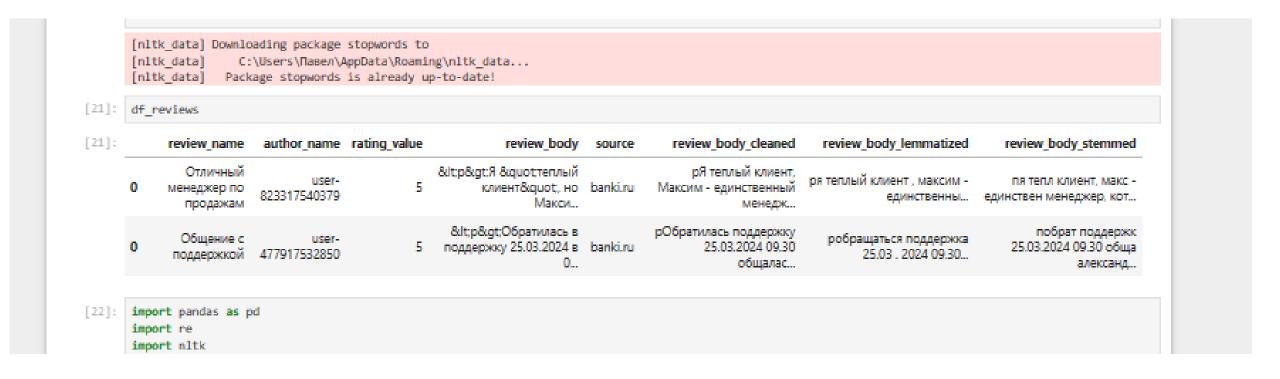
'name': 'Дмитрий', 'date': '22 марта', 'rating': 5, 'review_text': 'Впарили кредитку, попользовался — денег банку не принёс, да вот я гад такой.\'

- -2 шаг (Чистим от лишнего и структурируем собранные данные для последующего анализа)
 - Парсер banki.ru потребовал дополнительную чистку от HTML-символов для чистки.
 - Парсер sravni.ru устроен так, что после парсинга не требуется дополнительная чистка.



-3 шаг(Извлечение признаков)

1. Провели предварительную обработку текстов: удаление стоп-слов, лемматизацию / стемминг, удаление пунктуации.



-3 шаг(Извлечение признаков)

1. Преобразовали собранные тексты отзывов в векторное представление с использованием методов NLP, таких как TF-IDF.

Этот выбор обычно хорош для задач кластеризации текста, но может быть полезно также рассмотреть использование других методов, таких как Word2Vec или Doc2Vec, особенно если у будут очень большие объемы данных и мы захотим учитывать семантическое сходство между словами или предложениями.

```
def analyze_reviews(file_path):
    df = load_and_preprocess_data(file_path)
   num_reviews = len(df)
    vectorizer = TfidfVectorizer()
   X = vectorizer.fit_transform(df['processed_review'])
    kmeans = KMeans(n_clusters=10, n_init=10)
   df['cluster'] = kmeans.fit_predict(X)
    for i in range(10):
       cluster_text = " ".join(df[df['cluster'] == i]['processed_review'].values)
        generate_word_cloud(cluster_text, i)
    # Поиск и добавление по 5 наиболее схожих отзывов в каждый кластер
    for i in range(10):
        try:
            cluster_center_idx = cosine_similarity(np.asarray(X[df['cluster'] == i].toarray()), np.asarray(X[df['cluster'] == i].mean(axis=0)
            cluster_center_review = df[df['cluster'] == i].iloc[cluster_center_idx]['review_text']
            similar_reviews_indices = cosine_similarity(np.asarray(X[df['cluster'] == i].toarray()), np.asarray(X[df['cluster'] == i].mean(a
            similar_reviews = df[df['cluster'] == i].iloc[similar_reviews_indices][1:]['review_text'].values
            create_html_page(i, cluster_center_review, similar_reviews, num_reviews)
        except Exception as e:
            print(f"An error occurred while processing cluster {i}: {e}")
```

Использование KMeans для кластеризации векторов TF-IDF может дать хорошие результаты

-4 шаг (Анализ и интерпретация отзывов)

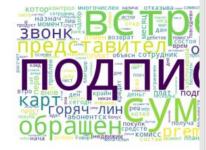
Использовали полученные признаки для выявления общих тем и тенденций в собранных отзывах. Сделали векторизацию на 10 кластеров. Анализа настроений нет - там надо применять что-то посерьезней ,хотя бы модель типа bert, появляются сложности с русским языком. Такой анализ в HTML-формате генерирует наша программа (только не 4, а 10 кластеров):

Анализ кластера 0

Средний отзыв кластера:

ответили. Тогда я спустя 10 двей обратилась на горячую линию Тинькофф! Сотрудница помогл бывка смс о пополении в сумме 1899 руб. Я благодарю бавк Тинькофф за возврат суммы за под

Похожие отзывы:



руб, покупка. Началась резбираться. А это подписка которая 1 руб продлизась. На сайте нигде 1 ответили. Тогда я спустя 10 двей обратилась на горячую лишно Тинькофф! Сотрудища помогл

мсплачиваюсь я за все переводом. Даже на рынке, когда продукты покупаю) Если переводнив

овмости квартиры в 9 000 000 руб. Ладво, мы согласились и на это. 5 Февраля 2024 был одоб полинсавия поговора. НО после звояка представителю выясиндось, что встреча не может состо окументы не прошли проверку (то есть, до 27.02 Сотрудники банка не проверяли мои докумен

Анализ кластера 1

Средний отзыв кластера:

Похожие отзывы:



(Персональный №6867601) с просьбой помочь разобраться. Какое же было у меня удивление, когда уж

персональному менеджеру 02.02.2024. Обращение выпало на выходной день. Дънтрий отнетил в понеда Потребовался интернет -жвайринг для продыж через социальные сети. Оставил заявку на сайте Тинькофф, т. К. У

для услугу. Я обратился в Поддержку, где сотруденки выслушали мою проблему и в целях сохранения предхожила бесплатный проход для членов семы в бизнес хал, что не было предусмотрено договором. Тыссе отпо-

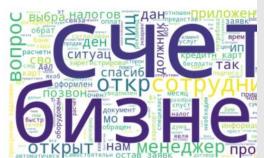
Гинькофф проволят профессиональные специалисты, такие как Лмитрий

Анализ кластера 2

Средний отзыв кластера:

Кочу поделиться опытом общения с персональным менеджером с подпиской Типькофф премнум. Я уж Оставил заянку на регистрацию ИП. Банк предоставил ако необходимую виформацию для того чтобы я мог само ныблюдать как банк улучшается с каждым вовым годом. С недавнего времени в оформал подписку тив бывка есть школа бинисы где бесплатию можно пройти курсы по открытию и развитию билиеса. Через день после менеджер, который в приоритетном порядке решает все возникающие проблемы. Так и случилось сего подписа, подписали необходимые документы, мне на телефон установили приложение Тиньсофф бизнес и вручил за покупку, а вторыя проблема когда 01.03.2024 Года в увидел, что мой период бесплатного обслуживан уведомление от налоговой о том что ип было зарегистрировано, вигоматически открылся счет для бизнеса. В прил ж полутур до достовный межен и простовные на пристовные н менеджером, объяснила как пользоваться приложением, дала свои контакты, проконсультировала по вопросам оп особенности услуги предоставляемые им бизнесу, всем кто думает открыть бизнес или уже открыл его рекоменду

Похожие отзывы:



овенедает, который в приоритетном порядке решает все возникающие проблемы. Так и случаюсь сего банка есть школа бизнеса где бесплатию можно пройти курсы по открытию и развитию бизнеса. Через день после ы покулку, а вторыя проблема когда 01.03.2024 Года в увидел, что мой первод бесплатного обслуживы водинсь, водинсали необходиямые документы, мне на телефом установили приложение Тинькофф бизнес и вручил менеджером, объяснила как пользоваться приложением, дала свои контакты, проконсультировала по вопросам оп

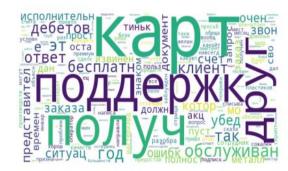
дкод ворождает высокую степень, доверня к Быку и стремление вользоваться их продуктами, котор условнем для оформления эквариита и ККМ). Спустя непродолжительное время, счет был открыт и с нами свика Тинькофф Банк является спонсором и банком КХЛ. И клиентам Тинькофф Премнум предлагает посец В месян. (Пакет). С проблемой естественно швето не разобрадся. Мы уже согласились с новыми условиями, чтобь друга. Об этой приятной акции я узнал от моего персонального менеджера [Дмитрия О. (6867601), Така передича оборудования (5 раз), таким образом сроки данного процесса растянулись. При всем этом, бинк перенапр банка были нарушены все заявленные условия и неоднократно. При попытке позвонить в банк и разъяснить ситуа

Анализ кластера 3

Средний отзыв кластера:

В ноябре 2023 года я посоветовал другу дебетовую карту Тинькофф. За это я получил бонус 1500Р, а друг наисегда бесплатное обслуживание. Мы оба довольны! Тех. Поддержка всегда помогает решать проблемы. Пользуюсь картой второй год, притекзий нет!

Похожие отзывы:



В ноябре 2023 года в посоветовал другу дебетовую карту Тинькофф. За это в получил бонус 1500Р, а друг навсегда бесплатное обслуживание. Мы оба довольны! Тех, Поддержка всегд

Я месяц пвонила в банк, более 5-ти раз точно! Уточияла, есть ли возможность получить дебетовую карту Тинькофф Блэк с платёжной системой Мастеркара. Меня убедили, что такие карть ещё есть в наличии и для её получения я должиа для начала получить карту с смстемой Мир. Несколько раз откладывала заказ карты, сомневалась. Но меня всё же убедили, что мастеркард с бесплатным обслуживанием мне помогут оформить, после получения карты Мир. Итог: карту в заказала, получила Мир. А теперь мне отказывают в выдачи карты Мастеркард, ссылаясь на то, что её можно получить только с подпиской премиум. Потрачено много премени на переговоры, но результат — ввод клиента и заблуждение, лишь бы выдать хоть какую-то карту. Считаю

По федеральным каналам крутили рекламу «акция! Приведи друга и получи 1500, а твой друг получит дебетовую карту с полностью бесплатным обслуживанием навсегда! " Меня попросил моя знакомая поучаствовать в этой акции. Я зарегестрировалась по её специальной персональной ссылке. Акция прошла, потому что за меня ей заплатили 1500 как и обещали. И пот совершению случайно я узнаю, что у меня списываются деньги! Звоины в Тивькофф, рассказываем ситуацию. И оператор отвечает "вам что-то не так объяснили. Карта плативя" Это как вообще? Пришлось её забложировать! А та моя знакомая, которая привела меня говорит "ин у меня, ин у моих знакомых никогда такого не было! Всё полностью у нас бесплитио! "Так причём извачально, когда мне представитель привёз эту карту (дебетовую естественно), то он зачем-то как подарок дал мне кредитично! И активировал её! Я отказывалась, но он очень настойчивый Пришлось снова звонить и Тинькофф и разбираться, блокировать кредитку. Так сотрудники Тинькофф долго передо мной изменялись, что вообще он мне дал ещё какую-то карту, потому ч

газдайтия: и пустые взаимения. Составил 2 запроса в поддержку с претекцией, мне «подарили» бесплатный выпуск карты, который и так предусмотрен, после еще 2 запросов сказали, что вичего компенсировать мне не будут По факту на пустом месте просто пытались меня убедить, что я не прав, а признав свою ошибку отделались пустыми извинениями Считаю качество сервиса в Тиньке очень упало, в буду безумно рад, если прочитав это ны пойдете в другой банка, откажетесь от каких-то продуктов и прочес, у всего есть цена, у такого отношения тоже должна быть, молю- не оформление Понщите дебетовые и премнальные карты в других бынках, после реструктуривщин им стало на клиентов плевать, по бизнес счету просмл не явонить,

Вывод:

Наше решение было успешно загружено на GitHub, вы можете опробовать его и посмотреть.

Будем ждать встречи 1 апреля в 13:00!