



code: Букля

Алгоритм для поиска новых
тенденций и проблем бухгалтера

ФИЦ: ХАКАТОН

2024

Техническая информация

Для реализации кода использовалась следующие библиотеки:

● **SQLITE3**



● **PANDAS**



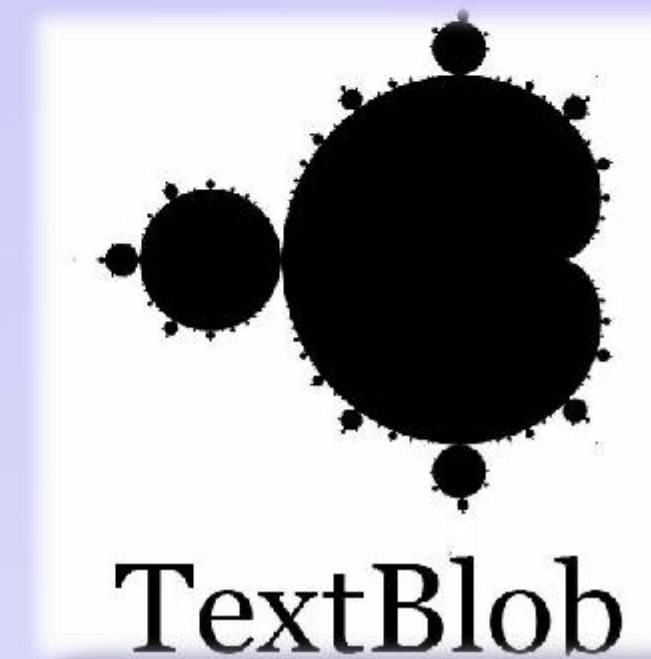
● **SKLEARN**



● **SPACY**



● **TextBlob**



Описание продукта

Определение трендов в области бухгалтерского учёта и выявление проблем бухгалтеров.

1. Анализ периодов : Пользователь выбирает «сегодня», «неделя» или «месяц». Код использует метки для фильтрации данных
1. Обработка текста : Используемая библиотека spaCy для лемматизации и очистки текста
1. Анализ тенденций : самые популярные слова/темы, выведенные в консоль
1. Гибкость : Код легко адаптируется для других форматов вывода (например, веб-интерфейса)

Ценность

- Усиление финансовой прозрачности ESG – отчетность
- Быстрая адаптация к перемена в налоговой политики
- Активная цифровизация и автоматизация

Реализация

- Определение каналов для сбора
- Создание инструментов для сбора информации
- Анализ и обработка данных
- Создание алгоритма для поиска облака ключевых слов и выделение популярных трендов

Проблема

- Законодательные изменения
- Проблемы с автоматизацией
- Перегрузка
- Ответственность за ошибки
- Дефицит кадров и высокая оплата труда
- Вопросы коммуникации с налоговыми органами

Функционал и фичи

Подключение к базе данных

Мы подключаемся к базе данных SQL, которая содержит сообщения с форумов

Выбор периода и категории

Пользователь выбирает период: «1 час», «1 день», «неделя», «месяц», «6 месяцев», «год».
Можно также отфильтровать по категориям (например, налоги, отчетность)

Обработка и анализ данных

Мы используем компонент Spacy для лемматизации текста
Применения анализа частоты слов (для определения трендов)
Используем TextBlob для анализа тональности
(положительная/отрицательная/нейтральная)

Генерация облака слов

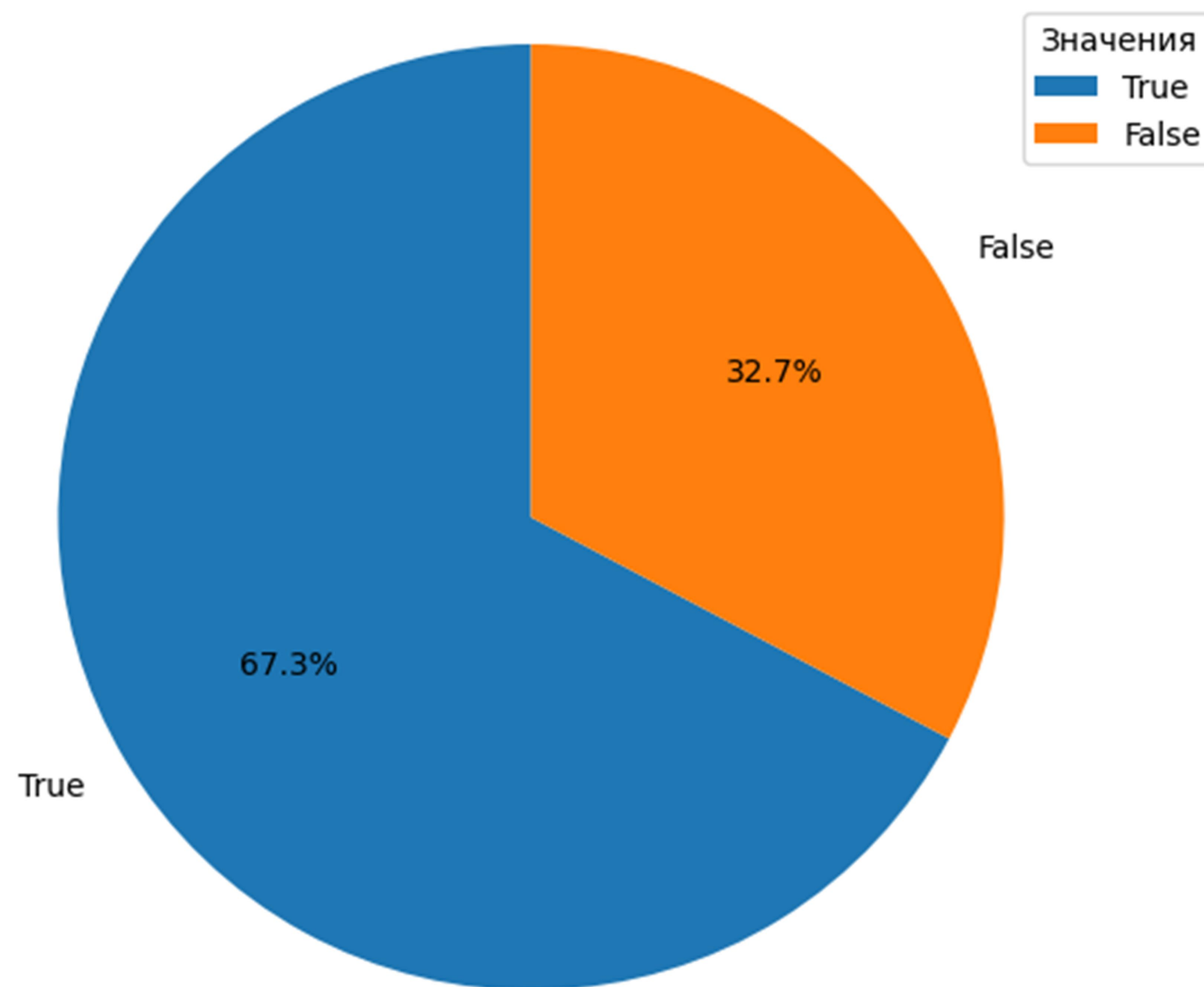
Мы визуализируем тенденции с помощью облаков слов.
что позволяет видеть самые популярные темы

Результаты вывода

Показываем топ-10 популярных трендов
Выводим информацию о распределении тональности
Отображаем облако слов

Налоговые отчеты

Количество проблем, связанных с отправкой отчетности в налоговые органы



Огромное количество комментариев на форуме относится к отправке отчетности в налоговые органы. По нынешним данным об этом 68% текста.

Блокировка по 115 -ФЗ



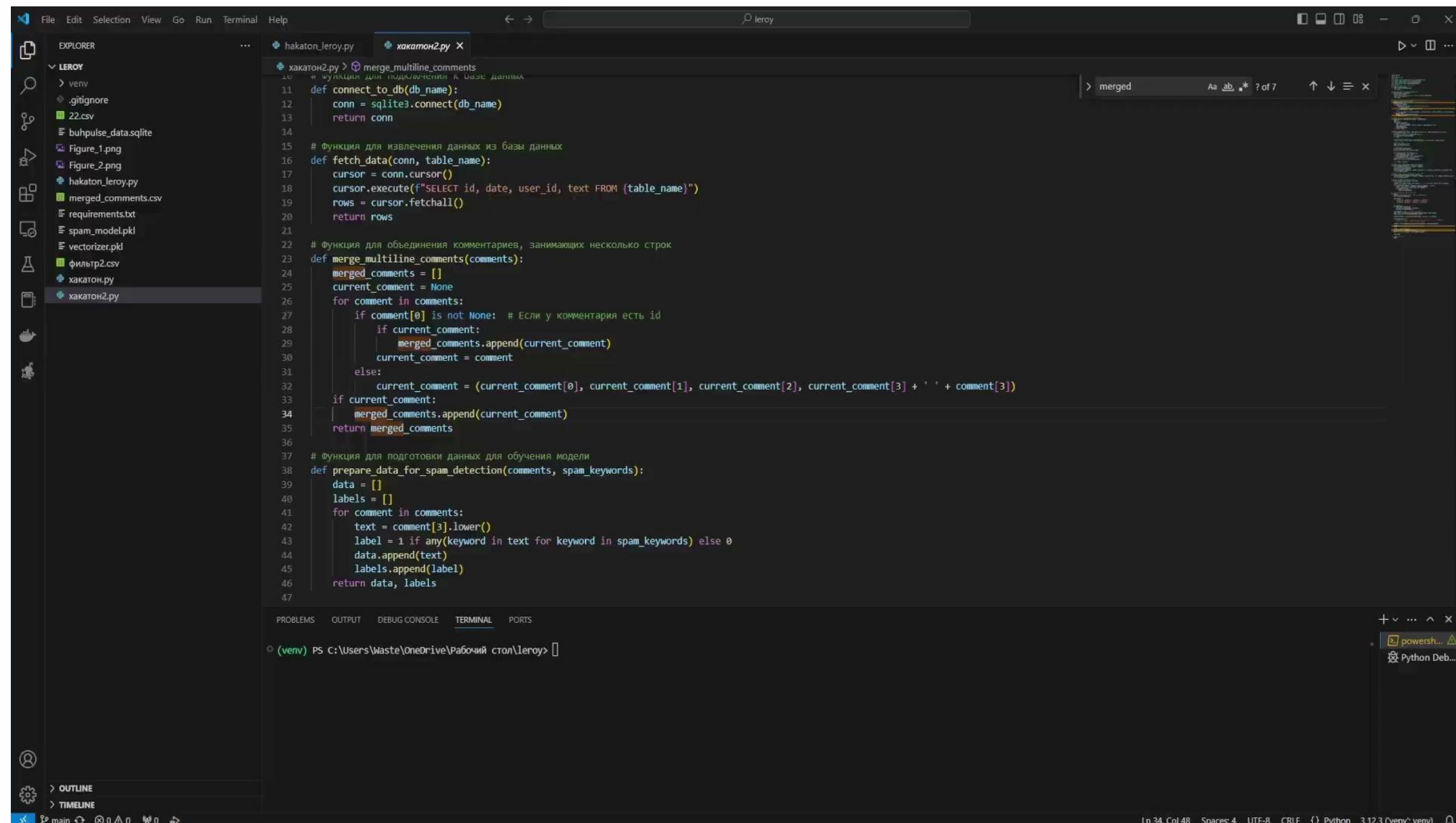
6% сообщения
относится к
блокировке
счетов по 115-ФЗ

Автоматизация бухгалтерии

Обсуждение проблем
автоматизации 5%



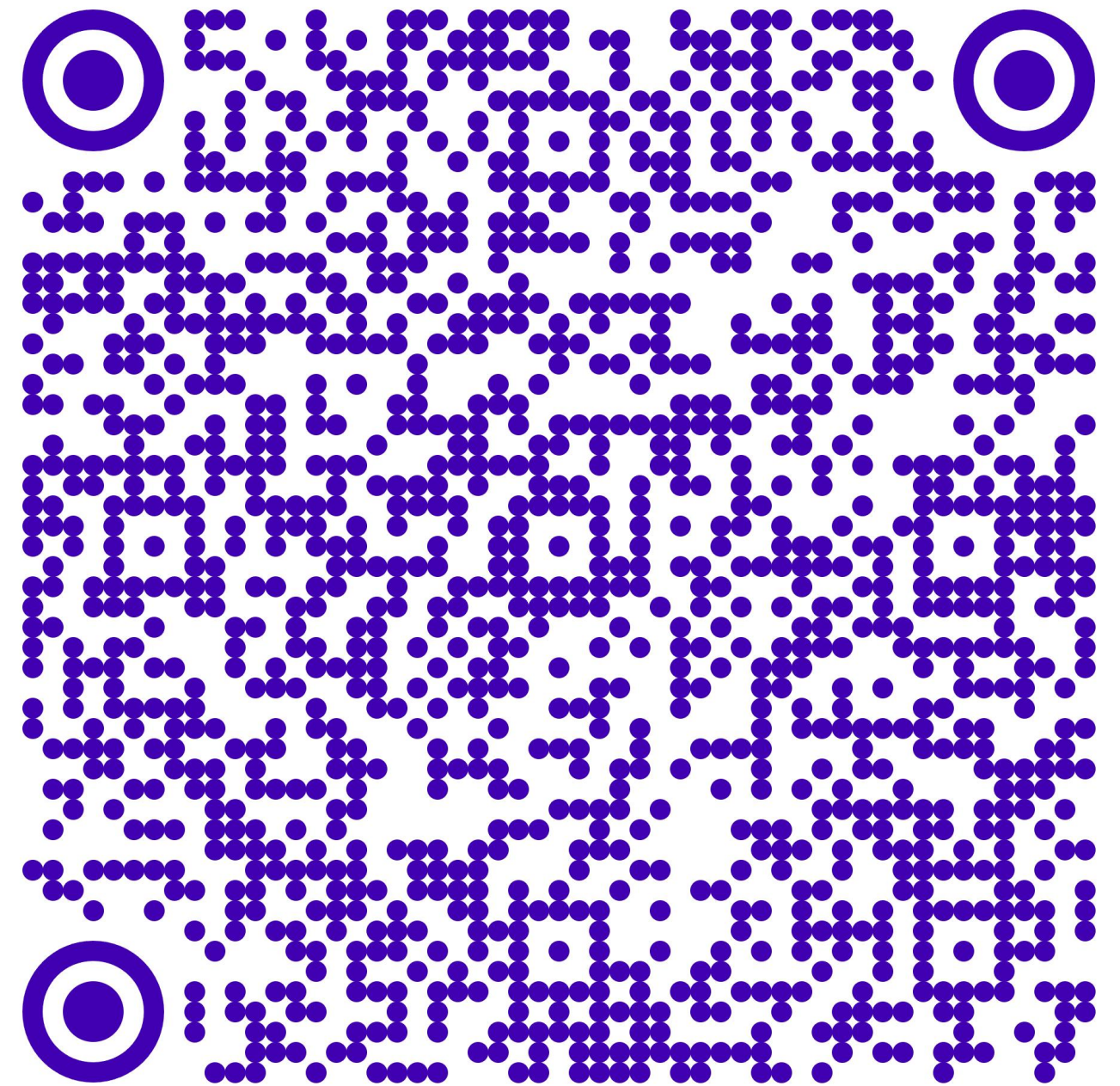
Демонстрация проекта



```
hakaton2.py > merge_multiline_comments
10 # функция для подключения к базе данных
11 def connect_to_db(db_name):
12     conn = sqlite3.connect(db_name)
13     return conn
14
15 # функция для извлечения данных из базы данных
16 def fetch_data(conn, table_name):
17     cursor = conn.cursor()
18     cursor.execute(f"SELECT id, date, user_id, text FROM {table_name}")
19     rows = cursor.fetchall()
20     return rows
21
22 # функция для объединения комментариев, занимающих несколько строк
23 def merge_multiline_comments(comments):
24     merged_comments = []
25     current_comment = None
26     for comment in comments:
27         if comment[0] is not None: # Если у комментария есть id
28             if current_comment:
29                 merged_comments.append(current_comment)
30                 current_comment = comment
31             else:
32                 current_comment = (comment[0], comment[1], comment[2], comment[3] + ' ' + comment[3])
33         if current_comment:
34             merged_comments.append(current_comment)
35     return merged_comments
36
37 # функция для подготовки данных для обучения модели
38 def prepare_data_for_spam_detection(comments, spam_keywords):
39     data = []
40     labels = []
41     for comment in comments:
42         text = comment[3].lower()
43         label = 1 if any(keyword in text for keyword in spam_keywords) else 0
44         data.append(text)
45         labels.append(label)
46     return data, labels
47
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

(venv) PS C:\Users\Waste\OneDrive\Рабочий стол\leroy>



[Ссылка на видеоматериалы](#)

Roadmap проекта

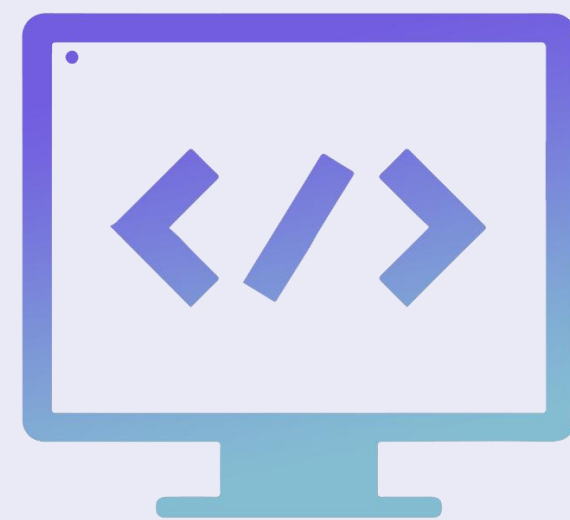
01 этап

Стратегическое
планирование. Ресерч
бух. рынка



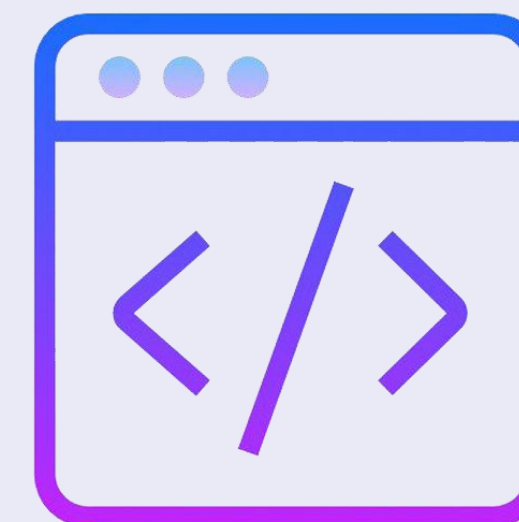
02 этап

Проработка и
тестирование MVP



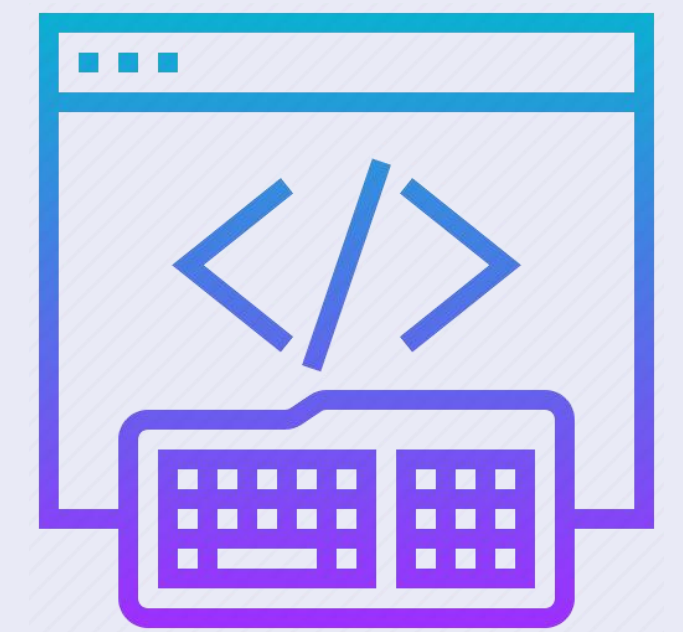
03 этап

Запуск кода и устранение
недочетов. Оптимизация



04 этап

**Релиз. Адаптация
кода под девайсы**



Конкурентные преимущества

Актуальность и скорость анализа

- Данные в реальном времени: инструмент позволяет анализировать информацию с форумов, чатов и других источников практически моментально, благодаря поддержке реального времени.
- Обновление контента: возможность оперативно генерировать контент-планы и рекомендации по актуальным темам дает конкурентное преимущество перед сервисами, которые работают с устаревшей информацией.

Повышение кач-ва контента

- Предложения контент-планов: возможность автоматически предлагать идеи для контент-маркетинга на основе популярных трендов и запросов дает маркетологам БухПлюс явное преимущество в создании релевантного материала.
- Примеры сообщений пользователей: использование реальных данных из сообщений на форумах увеличивает доверие и создает ощущение глубокого понимания проблем клиента.

Индивидуальность услуг

- Фильтры по категориям: возможность фильтровать данные по интересующим категориям (налоги, автоматизация, отчетность и т.д.) делает инструмент гибким и полезным для разных сегментов аудитории.
- Адаптация под запросы пользователя: пользователь может выбрать период анализа и получить подборку материалов или трендов, релевантных для его задач, что повышает уровень вовлеченности.

Усиление аналитической экспертизы

- Тональный анализ: возможность выявлять эмоциональную окраску упоминаний о компании БухПлюс или других темах помогает улучшать клиентский сервис.
- Идентификация проблем: Программа не только показывает тренды, но и помогает определить текущие "боли" клиентов, предоставляя решения.

Инновационность и высокотехнологичность

- Использование NLP: анализ данных с помощью искусственного интеллекта (например, лемматизация текста, тональный анализ) делает инструмент передовым решением на рынке.
- Визуализация данных: Облако слов, графики популярности тем и динамика трендов позволяют наглядно и просто воспринимать информацию, что повышает ее ценность.

Конкуренция через ценность

- Прямое взаимодействие с клиентами: благодаря анализу отзывов о компании, инструмент позволяет быстро реагировать на негативные упоминания, улучшая репутацию.
- Прозрачность и подкрепление ссылками: автоматическое добавление гиперссылок к источникам демонстрирует открытость компании и подчеркивает ее профессионализм.

Выводы и рекомендации

Рекомендовано использование модели BERT
К сожалению, в рамках текущего Хакатона, данную модель обучить нельзя ввиду сжатых сроков (требуется ~ 3-4 дней)

В дальнейшем следует расширить модель функционалом для анализа отзывов о предоставленных БухПлюс услугах

Интеграция с WebSockets или другой системой для мгновенного обновления данных

Необходим детальный анализ сайтов и форумов для парсинга бухгалтерской информации

Рекомендованные сайты

- [Audit-it.ru](https://audit-it.ru)
 - [Glavbuh.ru](https://glavbuh.ru)
 - forum.klerk.ru
 - buh.ru/forum/
-

**Данная модель позволяет
эффективно находить решения для
актуальных проблемы бухгалтерии**



Панина Валерия

Маркетолог & Промт-инженер



Сиротенко Никита

Аналитик & Код-инженер



Ковалева Екатерина

ML инженер & Бух. консультант