



Система генерации отзывов на основе GPT-2

Учебная задача

URFUML2024_PP_TEAM-5

URFUML2024_TEAM-5



Кравцов
Александр

[GitHub](#)



Чашников
Сергей

[GitHub](#)



Ефимович
Евгений

[GitHub](#)



Ильин
Виктор

[GitHub](#)



Салов
Антон

[GitHub](#)



Крючков
Владислав

[GitHub](#)

ВВЕДЕНИЕ

- Современные технологии активно меняют способы взаимодействия с информацией, создавая новые возможности для автоматизации задач. В условиях растущего объема данных особенно актуальными становятся системы, способные упрощать и ускорять обработку информации. Одной из таких задач является генерация текстовых отзывов, которая может быть полезной для бизнеса и пользователей, предоставляя релевантные и персонализированные описания на основе различных параметров.
- Разработка нейронной сети, способной генерировать текстовые отзывы на основе таких входных данных, как категория места, средний рейтинг и ключевые слова, позволяет автоматизировать этот процесс. Это решение не только экономит время, но и повышает качество предоставляемой информации, делая её более доступной и понятной.

ПЛАН РЕАЛИЗАЦИИ ПРОЕКТА

Этап	Ответственный	Цель	Результат
1. Анализ и исследование данных	Kravtsov A.V.	Изучить данные, выявить проблемы, подготовить базовые статистики.	Визуализация структуры данных, список проблем, скрипт выгрузки.
2. Очистка данных	Chashnikov S.Yu.	Устранить проблемы и подготовить данные для обучения.	Чистые, нормализованные данные, скрипт чистки.
3. Генерация данных и эмбедингов	Efimovich E.A.	Создать векторное представление данных.	Финальный датасет, векторные представления текстов.
4. Создание и обучение моделей	Salov A.S.	Построить и обучить модели генерации текстов.	Обученные модели с первичной оценкой качества.
5. Улучшение моделей и интерфейс	Kryuchkov V.V.	Оптимизировать модели и создать интерфейс.	Оптимизированная модель с готовым приложением.
6. Презентация и демонстрация	Ilyin V.B.	Представить результаты работы.	Полноценная презентация проекта с демонстрацией.

Более детально можно
ознакомится по [ссылке](#)

ЗАДАЧИ ПРОЕКТНОЙ РАБОТЫ

- **1. Изучение существующих моделей нейронных сетей**
 - - Исследовать предобученные модели, предназначенные для написания краткого конспекта вводимого текста, с акцентом на модели семейства GPT-2.
 - - Оценить их применимость для автоматического создания резюме текста.
- **2. Разработка Python-скрипта**
 - - Написать скрипт на Python, реализующий преобразование вводимого текста в краткий конспект.
 - - Использовать библиотеку машинного обучения transformers, чтобы интегрировать предобученную модель GPT-2.
- **3. Создание Web-приложения**
 - - Разработать Web-приложение с пользовательским интерфейсом для работы со скриптом.
 - - Обеспечить удобный ввод текста и отображение сгенерированного краткого конспекта.
 - - Использовать фреймворки, такие как FastAPI или Streamlit, для реализации интерфейса.
- **4. Проверка и тестирование**
 - - Провести тестирование приложения на разных текстах для оценки качества сгенерированных конспектов.
 - - Проверить работоспособность интерфейса и корректность обработки вводимых данных.
- **5. Подготовка документации**
 - - Создать подробную документацию по проекту, включая описание используемой модели, архитектуры приложения и инструкций по запуску.
 - - Включить раздел с рекомендациями для дальнейшего улучшения.

Результат:

Готовое приложение, позволяющее преобразовывать текст в краткий конспект, с проверенной функциональностью и сопровождающей документацией.

АНАЛИЗ ПРОБЛЕМЫ

- Генерация текстовых отзывов для различных мест — сложная и трудоемкая задача, особенно когда требуется учитывать множество параметров, таких как категория места, средний рейтинг и ключевые слова.
- Рост числа платформ для отзывов и ожиданий пользователей к качеству текста создаёт следующие проблемы:
 - 1. Трудоемкость ручной генерации:
 - Написание отзывов вручную занимает много времени и требует творческого подхода.
 - 2. Непоследовательность и субъективность:
 - Люди часто используют разные стили и подходы, что затрудняет создание отзывов с единообразной структурой.
 - 3. Необходимость персонализации:
 - Современные пользователи ожидают, что отзывы будут соответствовать их ожиданиям и контексту, что трудно достичь без использования технологий.
 - Использование нейронных сетей, таких как GPT-2, позволяет автоматизировать процесс написания отзывов, создавая последовательные и персонализированные тексты с учётом заданных параметров. Это значительно ускоряет процесс и улучшает его результат.

ОПИСАНИЕ РЕШЕНИЯ

- Для решения обозначенной проблемы мы предлагаем Web-приложение «Генератор отзывов», основанное на модели нейронной сети GPT-2. Основной функцией приложения является автоматическая генерация текстовых отзывов для различных мест на основе входных параметров, таких как категория места, средний рейтинг и ключевые слова.
- Ключевой особенностью приложения является его архитектура в виде Web-приложения. Такой подход избавляет от необходимости разрабатывать отдельные версии для различных платформ и упрощает доступ для пользователей. Приложение не требует установки и позволяет разработчику оперативно обновлять функционал, добавлять новые возможности и исправлять ошибки.
- Инновационность решения заключается в применении трансформерной модели GPT-2 для генерации персонализированных и релевантных текстов, что значительно ускоряет процесс создания отзывов и повышает их качество

ИСПОЛЬЗУЕМЫЕ ТЕХНОЛОГИИ И ИНСТРУМЕНТЫ

1. Язык программирования:

Python – основной язык для реализации обработки данных и логики приложения.

2. Предобученная модель:

GPT-2 – нейронная сеть на основе трансформеров для генерации текстовых отзывов.

3. Система контроля версий:

GitHub – для хранения кода, управления версиями и совместной работы над проектом.

4. Библиотека для создания Web-приложений:

Streamlit – для разработки удобного пользовательского интерфейса.

ПРАКТИЧЕСКАЯ ЦЕННОСТЬ И ПРИМЕНИМОСТЬ

- Приложение «Генератор отзывов» работает без необходимости установки на устройство пользователя, обеспечивая доступность и удобство использования.
- Сгенерированные отзывы могут быть применены в различных сферах:
 - **Для бизнеса:**
 - автоматическое создание отзывов для онлайн-платформ, что упрощает заполнение карточек мест и привлечение клиентов.
 - **Для маркетинга:**
 - генерация отзывов для анализа предпочтений пользователей и тестирования гипотез.
 - **Для пользователей:**
 - получение персонализированных текстов, которые отражают основные характеристики мест.
 - Это решение помогает сократить время на ручное написание отзывов, улучшить их качество и повысить продуктивность работы с текстами, предоставляя универсальный инструмент для разных задач.

