

Чат бот: Персональный помощник Главы региона

B.D.Science

Декабрь 2023

Чатбот "Персональный помощник Главы региона" - инструмент созданный для обработки и фильтрации запросов населения относительно планов развития региона, с предоставлением нормативной документации. <https://github.com/Dimildizio/chat-bot-yakutia>.

1 Термины и пояснения

Следующие термины будут использованы в проекте:

- 1) **Итерация:** - это все работы, которые совершаются до старта очередного пилота
- 2) **БТ:** - бизнес-требования
- 3) **EDA:** Exploratory Data Analysis - исследовательский анализ данных
- 4) **‘Product Owner’, ‘Data Scientist’ :** - роли, которые заполняют соответствующие разделы

1.1 Команда

Команда B.D.Science

Дмитрий Диденко Data Scientist, Product Owner.

Игорь Шахматов Data Scientist.

Снежана Маркив Data Scientist.

Роман Кондрашев Product Owner.

Александра Попко Product Owner.

2 Цели и предпосылки

2.1 Зачем идем в разработку продукта?

Бизнес-цель ‘*Product Owner*’

- **Бизнес-цель:** внедрить современные языковые модели в процессобработку обращений граждан для более эффективного взаимодействия с гражданами и снижения нагрузки на сотрудников.

- Почему станет лучше, чем сейчас, от использования ML ‘*Product Owner*’ и ‘*Data Scientist*’
- Использование ML позволит уменьшить время, затрачиваемое на взаимодействие с гражданами. Сейчас ответственный сотрудник в соответствии с действующим законодательством обрабатывает и готовит ответ. Это требует значительного времени и работы с НПА.
- Что будем считать успехом итерации с точки зрения бизнеса ‘*Product Owner*’
 - Уменьшение времени затрачиваемого сотрудниками на подготовку ответа
 - Повышение удовлетворения граждан от взаимодействия с органом власти
 - Повышение грамотности граждан касаясь процессов происходящих в регионе и его развитии

2.2 Бизнес-требования и ограничения

Краткое описание БТ и ссылки на детальные документы с бизнес-требованиями ‘*Product Owner*’

- продукт должен обеспечивать взаимодействие граждан с чат-ботом для предоставления разъяснений касаясь развития регионов.
- продукт должен определять предмет обращения.
- продукт должен формировать ответ на основе имеющейся у него в доступе информации.
- продукт должен сохранять контекст беседы для уточнения информации.
- продукт должен исключать запросы, не имеющие отношения к работе администрации или содержащие ненормативную лексику.
- продукт должен использовать бота на базе социальной сети VK.com.
- продукт должен иметь базу данных PostgreSQL.
- продукт должен иметь реализацию на WebAPI.
- продукт должен использовать языковые модели доступные на бесплатной основе.
- система должна быстро обрабатывать запросы.
- продукт должен предоставлять форматированный текст в виде ответа.
- продукт должен предоставлять нормативно-правовые акты или ссылки на них в своих ответах.

Бизнес-ограничения ‘*Product Owner*’

- Временные ограничения на разработку продукта - до 17.12.2023
- *Что мы ожидаем от конкретной итерации* ‘*Product Owner*’.
- увеличение качества предоставляемого продуктом ответа
- Описание бизнес-процесса пилота, насколько это возможно
- как именно мы будем использовать модель в существующем бизнес-процессе? ‘*Product Owner*’
- Что считаем успешным пилотом? Критерии успеха и возможные пути развития проекта ‘*Product Owner*’
 - добавление новых направлений взаимодействия с гражданами - обработка файлов в pdf формате

- разработка автоматических систем по предоставлению ответов в других областях знаний. (1) веб-парсинг НПА источников из соответствующих областей знаний в текстовом формате и (2) и законодательных актов в pdf и jpg.

2.3 Что входит в скоуп проекта/итерации, что не входит

- На закрытие каких БТ подписываемся в данной итерации ‘Data Scientist’
 - Что не будет закрыто ‘Data Scientist’
 - Описание результата с точки зрения качества кода и воспроизводимости решения ‘Data Scientist’
 - Описание планируемого технического долга (что оставляем для дальнейшей продуктивизации) ‘Data Scientist’

2.4 Предпосылки решения

- Описание всех общих предпосылок решения, используемых в системе – с обоснованием от запроса бизнеса: какие блоки данных используем, горизонт прогноза, гранулярность модели, и др. ‘Data Scientist’

3 Методология

‘Data Scientist’

3.1 Постановка задачи

- Что делаем с технической точки зрения: рекомендательная система, поиск аномалий, прогноз, оптимизация, и др. ‘Data Scientist’
 - Диалоговая система на основе большой языковой модели

3.2 Блок-схема решения

- Блок-схема для бейзлайна и основного MVP с ключевыми этапами решения задачи: подготовка данных, построение прогнозных моделей, оптимизация, тестирование, закрытие технического долга, подготовка пилота, другое. ‘Data Scientist’

- [Пример возможной блок схемы]([LINK HERE](#))

Схема обязательно включает в себя архитектуру бейзлайна. Если бейзлайн и основной MVP отличаются несущественно, то это может быть одна блок-схема. Если значительно, то рисуем две: отдельно для бейзлайна, отдельно для основного MVP.

Если блок-схема шаблонна - т.е. её можно скопировать и применить к разным продуктам, то она некорректна. Блок-схема должна показывать схему решения для конкретной задачи, поставленной в части 2.

3.3 Этапы решения задачи

Примеры этапов ‘Data Scientist’:

1. Этап - Подготовка инфраструктуры разработки проекта
2. Этап - Подготовка данных
3. Этап - Подготовка языковых моделей
4. Этап - Интерпретация моделей (согл. с заказчиком)
5. Этап - Интеграция бизнес правил для расчета бизнес-метрик качества модели
6. Этап - Подготовка инференса модели по итерациям
7. Этап - Интеграция бизнес правил
8. Этап - Разработка оптимизатора (выбор оптимальной итерации)
9. Этап - Подготовка финального отчета для бизнеса

4 Подготовка пилота

4.1 Способ оценки пилота

- Краткое описание предполагаемого дизайна и способа оценки пилота ‘Product Owner’, ‘Data Scientist’

4.2 Что считаем успешным пилотом

- Рабочий чатбот что получает информацию из вк, обращается к БД, классифицирует, суммаризирует информацию и предоставляет ответ.

4.3 Подготовка пилота

- Наш продукт должен работать на одном сервере на локальной машине и предоставлять ответ в разумные рамки. Должен быть предоставлен доступ в интернет для получения запроса.

(Предусматриваем уточнение параметров пилота и установку ограничений по вычислительной сложности моделей.) ‘Data Scientist’

5 Внедрение

5.1 Архитектура решения

- Блок схема и пояснения: сервисы, назначения, методы API ‘Data Scientist’

5.2 Описание инфраструктуры и масштабируемости

- Какая инфраструктура выбрана и почему ‘Data Scientist’ - Плюсы и минусы выбора ‘Data Scientist’ - Почему финальный выбор лучше других альтернатив ‘Data Scientist’

5.3 Требования к работе системы

- SLA, пропускная способность и задержка ‘Data Scientist’

5.4 Безопасность системы

- Потенциальная уязвимость системы ‘Data Scientist’

5.5 Безопасность данных

- Нет ли нарушений GDPR и других законов ‘Data Scientist’

5.6 Издержки

- Расчетные издержки на работу системы в месяц ‘Data Scientist’

5.7 Integration points

- Описание взаимодействия между сервисами (методы API и др.) ‘Data Scientist’

5.8 Риски

- Риски галлюцинаций модели и предоставления некачественного ответа
 - Риски обработки персональных данных
 - Риск утечки чувствительной информации и информации для служебного пользования

References