

Автоматическая классификация обращений граждан

Липко Иван Юрьевич, г. Севастополь.

Автоматическая классификация обращений граждан *Рязанская область*

Источники данных:

- <https://git.asi.ru/tasks/classification-appeals> (только по рязанской области)

*Более подробное описание решения смотри в jupyter-тетрадке

Шаги решения/подходы

Разведка данных (исследование входных данных)

что из себя представляют данные, структура данных, распределение по целевым признакам, количественные и качественные характеристики, какие присутствуют зависимости

Чистка данных и подготовка данных

удаление шумов и неподходящих данных (длинные обращения, уменьшение количества категорий, тем, исполнителей)

создание признаков: Bag of Word, TF-IDF слов/н-грам

Поиск классификаторов

обучение, сравнение качества, отбор классификаторов
повторная работа с данными (чистка и проч.)

Общая схема функционирования

В исследовании

- Инициализация библиотек

- Загрузка, подготовка данных

- Обучение классификаторов

- Проверка на тестовых данных (предсказание)

В предполагаемом MVP

- Инициализация библиотек

- Загрузка предобученных классификаторов

- Предсказание на новых данных (основной режим)

- ДоОбучение классификаторов (при необходимости)

Использованные технологии

- Python
- Anaconda-Jupyter: Sklearn, Pandas, NumPy, NLTK, Seaborn, Matplotlib, lightgbm
- Среда для запуска jupyter-тетрадки и прототипа (кроссплатформенное решение)
- Сторонние сервисы не использовались

Итоги решения. Результативность

Умеем классифицировать:

наиболее популярные категории, *category* (качество 80-90%)

городская территория, дворовая территория, дороги, многоквартирные дома

наиболее популярные темы, *theme* (качество низкое)

надписи и рисунки на заборах и дворовых сооружениях, не работает уличное освещение, несанкционированные объявления надписи и рисунки на фасадах, повреждение или отсутствие дорожного покрытия

наиболее популярных исполнителей, *executor* (качество низкое)

министерство транспорта и автомобильных дорог рязанской области, префектура железнодорожного района, префектура октябрьского района, префектура советского района, управление благоустройства города

Поставленные **цели достигнуты не полностью.**

Итоги решения. Выводы

Поставленные цели достигнуты не полностью потому что:

- **недостаток данных** (для большинства классов данных не хватает)
- **отсутствует чёткое представление результата**, среды внедрения
- **следую принципу** *"лучше решить большую часть проблем чем не решить их вовсе"*
- **недостатка времени** (моё позднее включение в соревнование).

Результаты исследования можно использовать для дальнейшей работы над MVP.

Готовность программного решения для использования

Данное исследование **нельзя напрямую внедрить в действующие системы**, но возможно создать MVP.

Для дальнейших работ **необходимо:**

- **Видение от заказчика**; в каком виде требуется решение (интерфейс, формат вывода результатов, автоматизированное рабочее место оператора) (если его нет, то я могу предложить)
- **уточнение требований к разрабатываемой системе** (приоритеты, например, важнее правильно предсказать наиболее частые типы обращений или наиболее редкие и т.п.)
- **создать MVP**, используя код исследования
- **используются бесплатные библиотеки**. Авторские права на программы и данные остаются за их авторами.

Готовность программного решения для использования

Текущее решение-исследование **плохо масштабируемо**.

Присутствует код для приема данных и выдачи результатов ***внутри jupyter-тетрадки***.

Чтобы проверить текущие результаты исследования (**Jupyter-тетрадка**), необходимо выполнить все ячейки:

Главы [Инициализация](#)

Главы [Данные для обучения](#)

Главы [Алгоритмы категоризации](#)

Главы [Алгоритмы тематизации](#)

Главы [Алгоритмы выбора исполнителей](#)

Готовность команды участвовать в реализации

Я готов участвовать в реализации.

Мои ограничения (я один участник в команде):

- скорость работы
- отсутствие некоторых компетенций

Готов привлечь людей для работы над проектом (которые не являются участниками соревнования)

Состав Команды

Липко Иван Юрьевич, ivanlipko@yandex.ru

младший научный сотрудник, ФГАОУ ВО

"Севастопольский государственный университет",

педагог дополнительного образования, ГБОУ ЦДО

"Малая академия наук"



Портфолио:

1. <https://github.com/ivanlipko/>
 - https://github.com/ivanlipko/ml_boosters_contest_ny
 - https://github.com/ivanlipko/ds_laptop_analyze
2. <https://mlbootcamp.ru/profile/ivanlipko/>
3. Победитель регионального этапа Всероссийского конкурса «Цифровой прорыв» (участник команды разумеется)



datamasters.ru

Сделай мир вокруг себя
лучше!