

Команда 15, Big Bad Data

Было решено сконцентрироваться на выявлении основных проблем, по которым обращаются пользователи. После чего самые стандартные вопросы будут решаться путем подтягивания стандартных ответов, соответствующих тематике или ссылки на статью FAQ.

Для этого анализ проводился отдельно для клиентских сообщений и ответов технической поддержки. Для ускорения работы были построены тематические модели LDA из библиотеки gensim. Количество кластеров было взято в первом приближении. Далее построили отображение кластеров из данных клиентских сообщений в кластеры ответов тех. поддержки. Таким образом, имеем две кластеризации.

Пустим ответы из FAQ Tele2 в кластеризацию ответов, получим соответствие типовых ответов нашим кластерам. **Зачем это нужно и что имеем?** В первую очередь, поступает новый текстовый запрос, он кластеризуется по тематикам обращений, после подтягиваем наиболее подходящий кластер ответов поддержки (один/два, рекомендации, топ по вероятности). В случае высокой вероятности/достаточного количества слов, подтягиваем ответ из FAQ и закрываем инцидент, если клиент удовлетворен. Иначе можем подтягивать стандартные ответы тех. поддержки к данному отображению двух-трех топ кластеров. Это может работать автоматически или аналогично программе-суфлеру. Помимо этого, данная кластеризация может помочь выявлять спам-клиентов, некомпетентных работников поддержки. Это можно отследить по количеству повторных обращений пользователя на заданную тематику.

Таким образом, имеем макет суфлера/бота и повышения скорости обработки запросов.

Что не успели:

Bigartm, подбор оптимального количества тем, кластеризацию проблем по городам.