Модель для анализа открытых вопросов в исследованиях

*Описание кейса:*

Росатом проводит много опросов и исследований, чтобы изучать мнение сотрудников в разных разрезах. Опросы часто содержат много открытых вопросов, которые предполагают расширенные ответы респондентов.

Сейчас обработка идет вручную, силами эксперта-социолога/маркетолога, который на основе своих представлений и экспертизы осуществляет кластеризацию ответов. Это очень вдумчивая трудозатратная деятельность. У этого способа есть ряд рисков: высокая загрузка сотрудников, вовлечённых в исследование, вероятность ошибок в следствие человеческого фактора, например, эксперт перегружен или устал, необъективность и отсутствие единого критерия определения результатов, поскольку каждый эксперт анализирует, опираясь на свой опыт и экспертизу, а у каждого аналитика они могут быть разными, а также достаточно долгий срок обработки результатов.

*Диагностика проблемы:*

Идея – подключить к обработке ИИ. Это позволить исключить ошибки, субъективность позиции конкретного эксперта, увеличит скорость обработки информации, а также повысить универсальность и объективность результатов исследований.

В этом кейсе предлагается создать модель, анализирующую открытые вопросы респондентов и создающую на основе них выводы.

*Входные данные для решения кейса:*

Основной функционал:

* Модель должна уметь из массива ответов респондентов на открытый вопрос составлять облако слов с учетом их смысловой нагрузки и словоформ.
* Модель должна уметь из массива ответов респондентов на открытый вопрос составить одно (несколько) предложений.

*Предоставляемые данные:*

Обезличенный массив ответов респондентов на один открытый вопрос (2 тыс. ответов)

*Сценарий использования результата:*

1. По итогам опроса собранные результаты загружаются заказчиком в программу – вопрос и пул открытых ответов;
2. ПО генерирует облако тегов, тезисов или другие форматы по сценарию;
3. Предложение (предложения) должно содержать наиболее часто повторяющиеся в ответах респондентов слова, но объединённые смыслом. Не просто облако слов, а компиляцию из всех вариантов в виде законченного логически связанного предложения, несущего смысловую нагрузку.

*Технические требования:*

Работающий сайт, генерирующий полноценный документ + ссылка на репозиторий github или gitlab с исходным кодом проекта

* User-friendly интерфейс
* Хранение данных для генерирования документов в одном месте, которым можно будет обеспечить надлежащую защиту
* Семантический анализ
* Результаты в формате облака, тезисов и прочего по сценарию с возможностью выгрузки, в том числе, в форматах \*ppt, \*pdf и других, которые возможно открыть в разных ОС, в том числе, и в Astra Linux.

*Рекомендуемый стек:*

Python