

Кейс: анализ тональности



С чего все началось

- К вам пришёл владелец некоторого
- «IT Технологии», «Блокчейн», «Machine Learning», «Нейросети»
- Сделайте нам крутую систему анализа тональности текста

Первые шаги



A photograph of a young baby, seen from behind, taking its first steps on a dirt path. The baby is wearing a blue and white striped long-sleeved shirt and blue jeans. The background is filled with green foliage and trees.

Первые шаги

- Что вообще происходит?
- Подсказок больше не будет
- Спрашивайте!

A photograph of a young baby with light brown hair, seen from behind, walking away on a dirt path. The baby is wearing a blue and white striped long-sleeved shirt and dark blue jeans. The background is filled with green foliage and trees.

Важные аспекты

- Данных нет, специально собирать скорее всего никто не будет
- Проект отчасти исследовательский, конечное применение однозначно неясно, но есть варианты:
 - Онлайн-консультант
 - Проверка отзывов на сторонних сайтах и соцсетях
 - Выделение ключевых связок слов, определяющих отношение клиента



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

- Data
- Mining
- Action

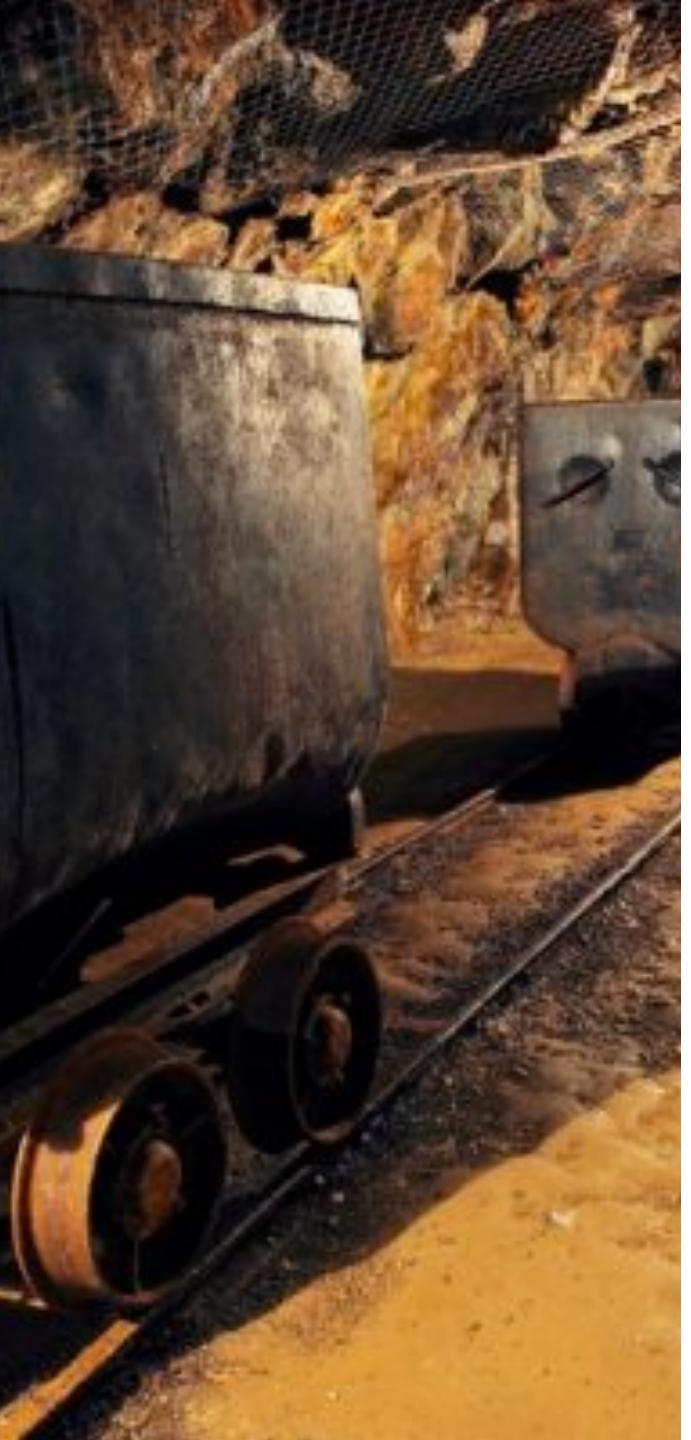


Data: Какие данные нам нужны?

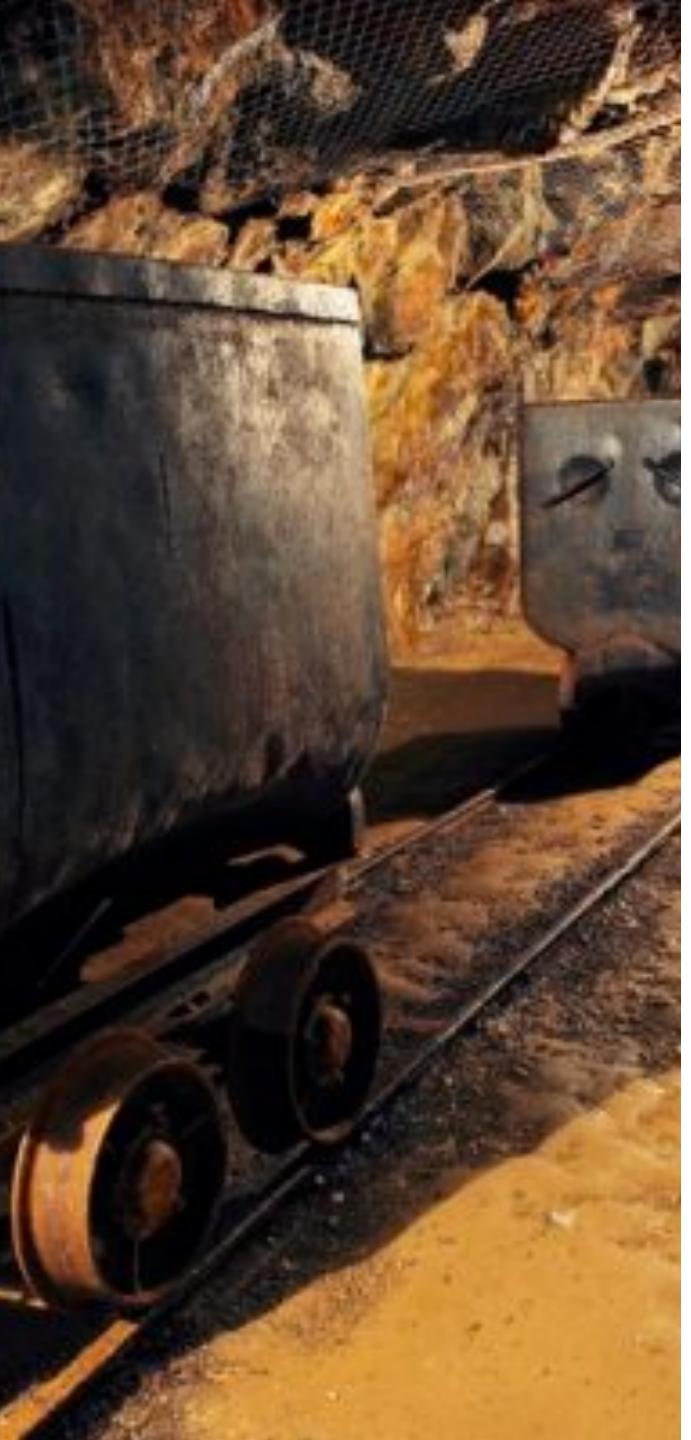


Data: Какие данные нам нужны?

- Тексты
- Фидбек
- Информация о пользователях



Mining: Где данные взять?



Mining: Где данные взять?

- Самим распарсить тематические сайты
- API соцсетей
- Тематические публичные датасеты
- Организовать разметку

Action: Что с этим всем делать



A photograph of Spider-Man (Tobey Maguire) in his classic red and blue suit, standing in front of a window with a brick building visible outside. He is looking slightly to the right.

Action: Что с этим всем делать

- Под разные применения нужны разные модели и метрики:
 - Онлайн консультант -- нужно вовремя переключаться на человека
 - Анализ соцсетей – нужно находить отзывы под заданную тематику
 - Поиск ключевых фраз – общий анализ данных



Action: А что дальше?



Action: А что дальше?

- Нужно проверить модели
- Какие эксперименты можно сделать в этих задачах?
- Где их нужно делать, а где нет?

Action: если полученный результат не нравится





Action: если полученный результат не нравится

- Как получить больше данных?
- Как можно более эффективно организовать разметку данных?