1. Выбрать данные разметить лейблы (н-р: текстовые сообщения, метка спам- не спам) Загрузить в модель (модели, которые можно потестировать в рамках проекта по классификации: - KNN - Decision Tree – Boosting) . Подставить модели в классификатор и сравнить предсказания. С помощью метрики показать как хорошо модели предсказывают и сделать визуализацию данных.

Блокнот по классификации

<https://github.com/nstsj/ML_for_NLP/blob/34d4b4c27c3ee6d24624a9b41f2caca6f383f4ad/2_classification/clf.ipynb>

1. Сделать слайды-презентацию
2. Постановка общей задачи.
3. Кому такое могло бы быть нужно.
4. К какой задачи NLP можно свести.
5. Какие методы обычно используются, какие аналогичные системы существуют.
6. Какой метод выбрали Вы, почему?
7. Какая общая схема решения задачи.
8. Как собирались и обрабатывались данные.
9. Как размечались данные или какая разметка данных использовалась.
10. Какие дополнительные ресурсы.
11. Эксперименты, методы в основе построения системы(можно просто перечислить), тьюнинг и т.п.
12. Оценка качества.
13. Разбор ошибок и обсуждение результатов (сильных и слабых сторон тех технологий, которые вы использовали)