

Решение кейса <<Барьеры безопасности>>

Команда Stardust Crusaders

Препроцессинг

- Очистка данных (символы, цифры)
- Сепарация данных
- Составление словаря уникальных слов с подсчетом частоты
- Создание искусственных сообщений для тестирования моделей

Различные подходы к построению модели

1. Ручная кластеризация
2. Word2Vec + классические алгоритмы кластеризации:
 - i. Реализация через платформу KNIME
 - ii. Реализация через python-nltk

Ручная кластеризация на основе ключевых слов

Word2Vec + Kmeans (KNIME)

Word2Vec + Kmeans (Python)

Гипотезы

Для выдвижения гипотез использовались открытые данные:

Выдвинутые гипотезы

Сравнение моделей и выводы: