



Выявление поляризации и нейтральности текстов в новостном потоке

Polarization and neutrality detection of texts in the news flow

Авдеев Роман Артемович

Научный руководитель: д.ф.-м.н. Воронцов Константин Вячеславович

Московский государственный университет имени М. В. Ломоносова
Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования

Москва, 2023

Цель исследования

В данной работе предлагается способ определения поляризации/нейтральности текстов в новостном потоке.

Решается задача разделения множества новостных сообщений на кластеры-мнения, выделения отдельных кластеров нейтральных и нерелевантных документов.

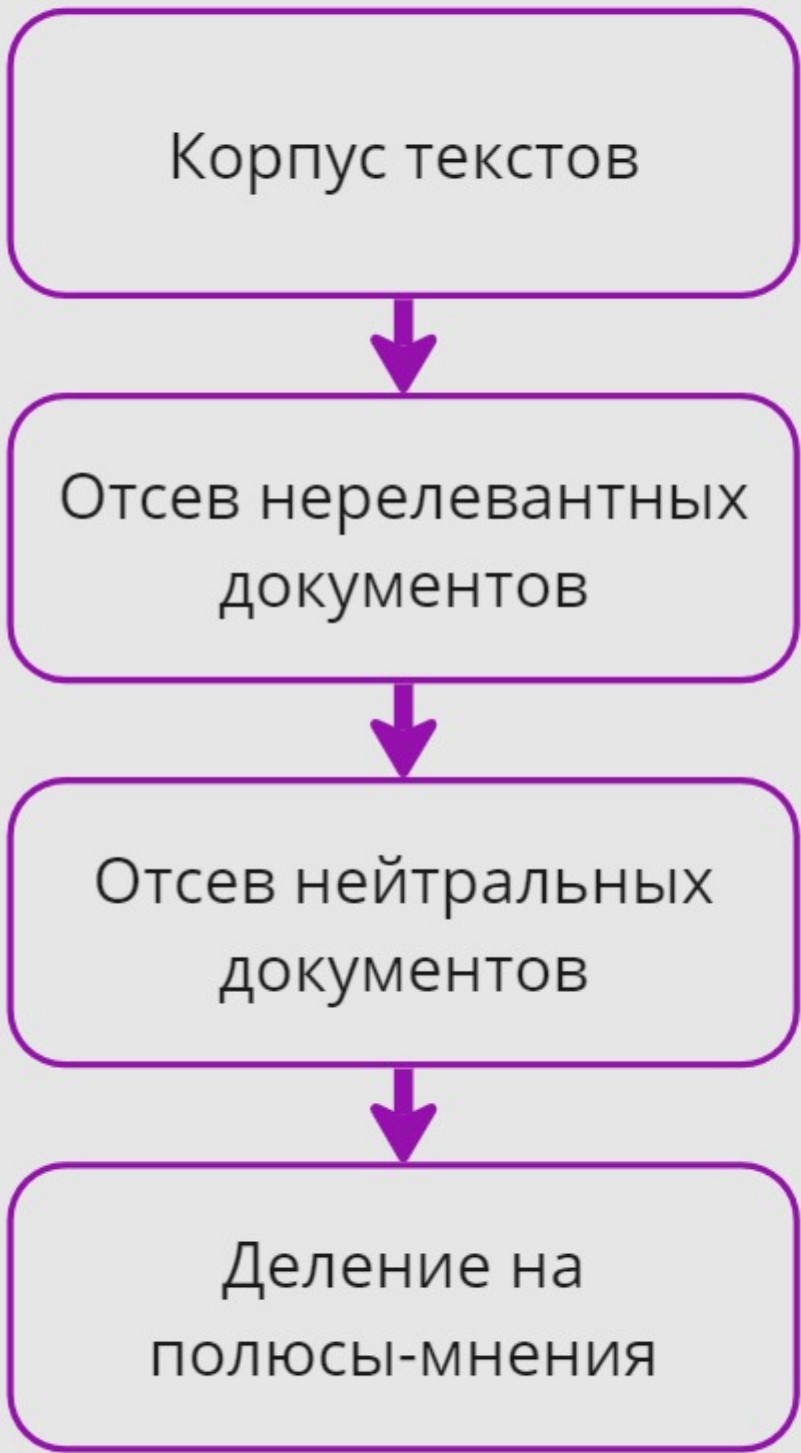
Данные

Используемый датасет представляет собой набор из 30-ти корпусов новостных сообщений из рубрик «Политика» и «Происшествия».

В каждом корпусе от 8-ми до 33-х документов.
Всего 452 документа.



Модель и текущие результаты



Метрики:

- Критерий M1: точность и полнота кластеризации мнений
- Критерий M2: точность и полнота отсева нейтральных документов
- Критерий M3: точность и полнота отсева нерелевантных документов
- Критерий M4: точность определения числа мнений

При независимом применении каждого из этапов:

	M1	M2	M3	M4
Алгоритм	0.65	0.65	0.856 / 0.907	0.695
Разметка	0.64	0.62	0.93	0.69

При последовательном применении каждого из этапов:

	M1	M2	M3	M4
Алгоритм	0.67/0.68	0.63/0.65	0.85 / 0.97	0.73/0.78
Разметка	0.64	0.62	0.93	0.69