

1 Проект 8: Кластеризация новостей

1.1 Категория

Кластеризация

1.2 Описание

Группировка новостных статей по темам с использованием алгоритмов кластеризации текста.

1.3 План реализации

- Сбор данных:** Датасет новостей (например, Kaggle News).
- Предобработка:** Токенизация, удаление стоп-слов.
- Векторизация:** TF-IDF или BERT embeddings.
- Модель:** K-Means или LDA.
- Оценка:** Silhouette score, тематическая интерпретация.