

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

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