

# **1 Проект 23: Кластеризация текстов документов**

## **1.1 Категория**

Кластеризация

## **1.2 Описание**

Группировка текстовых документов по темам.

## **1.3 План реализации**

1. **Сбор данных:** Датасет текстов (например, 20 Newsgroups).
2. **Предобработка:** Токенизация, удаление стоп-слов.
3. **Модель:** LDA или K-Means.
4. **Оценка:** Silhouette score.