

Титульный лист материалов по дисциплине
(заполняется по каждому виду учебного материала)

ДИСЦИПЛИНА **Модели и методы предиктивной аналитики**
(полное наименование дисциплины без сокращений)

ИНСТИТУТ **информационных технологий**

КАФЕДРА **Прикладной математики**
(полное наименование кафедры)

ВИД УЧЕБНОГО
МАТЕРИАЛА **Практические занятия**
(в соответствии с пп. I-III)

ПРЕПОДАВАТЕЛЬ **Парамонов Александр Александрович**
(фамилия, имя, отчество)

СЕМЕСТР **7, 2025-2026**
(указать семестр обучения, учебный год)

Задание 3. Обработка текста

Цель:

Обработать текстовые данные и проанализировать их.

Этапы выполнения задания:

- 1. Провести предобработку данных:**
 - Провести токенизацию (разбиение на слова/предложения, в зависимости от объёма текста).
 - Удалить стоп-слова (предлоги, союзы) и пунктуации.
 - Удалить лишние/мусорные термины (ссылки, эмодзи и т.д.).
 - Провести стемминг и лемматизацию.
- 2. Провести частотный анализ** – выявить топ 20 часто встречаемых слов.
- 3. Определить эмоциональную окраску текстов** (позитивный/негативный/нейтральный) с помощью готовых библиотек или обученной модели.
- 4. Выявите ключевые сущности** – имена людей, компании, даты, географические названия и т.д..
- 5. Определите основные темы в анализируемом тексте и классифицируйте его** на основе проведённого анализа (например, спорт, ИТ, и т.д.)

Результат работы:

После выполнения задания студент должен иметь результаты проведенного анализа, включающий реализацию выполнения поставленных задач. Отчёт о проделанной работе.