

СОДЕРЖАНИЕ

Введение.....	5
Теоретический материал к лабораторной работе №. Бинарная классификация фактографических данных.....	6
1.1. Постановка задачи классификации данных на примере текстовых документов.....	6
1.1.1. Способы формирования обучающих и тестовых выборок	7
1.1.2. Показатели качества классификации.....	8
1.1.3. Базовые методы классификации.....	11
Теоретический материал к лабораторной работе №2. Предварительная обработка текстовых данных.....	15
Теоретический материал к лабораторной работе №3. Классификация текстовых данных.....	16
3.1. Методы многоклассовой классификации.....	16
3.1.1. Регуляризация модели.....	21
3.1.2. Микро- и макроусреднение показателей качества.....	22
Теоретический материал к лабораторной работе №4. Кластеризация данных.....	22
4.1. Задача кластеризации данных.....	22
4.1.1. Иерархические методы кластеризации.....	24
4.1.2. Метод k -средних.....	26
Лабораторные работы.....	27
Лабораторная работа №1 Бинарная классификация фактографических данных.....	27
Методические указания к лабораторной работе №1.....	29
Генерация выборки.....	31
Обучение модели и классификация.....	33
Лабораторная работа №2. Предварительная обработка текстовых данных	35
Методические указания к лабораторной работе №2.....	37
Загрузка выборки.....	38
Векторизация.....	39
TF- и TF-IDF взвешивание.....	41
Классификация.....	41
Pipeline.....	41