Министерство науки и высшего образования Российской Федерации  
Пензенский государственный университет  
Кафедра вычислительная техника

**ОТЧЕТ**

по лабораторной работе №4

по дисциплине «Программные средства кибертехнических систем»

на тему «Изучение основных принципов построения обучаемых алгоритмов»

Выполнили:

студенты группы 22ВВП1

Хоссейни Нежад С.А.С.М.

Амиров И.Р.

Сергунов М. Р.

Приняли:

Зинкин С.А.

Карамышева Н.С.

Пенза 2025

**Цель работы**

Изучить основные принципы построения обучаемых алгоритмов

**Задание**

1. Получить описание заданной предметной области и набор соответствующих ей данных (варианты заданий приведены в конце пособия).

2. Провести исследование в соответствии с планом работы в рамках первой фазы исследования.

3. Подготовить отчёт по результатам первой фазы работ.

Отчёт должен содержать информацию в соответствии со следующей структурой:

1. Бизнес цель.

2. Ресурсы и риски.

3. Цель анализа данных.

4. План проекта.

Вариант 7

Задача определения принадлежности пикселя на фотографии к коже человека или другим объектам (4-й столбец набора данных {1 – кожа, 2 – не кожа}) по его цвету в кодировке RGB:

Blue – первый столбец {0...255}

Green – второй столбец {0...255}

Red – третий столбец {0...255}

Ссылка на репозиторий: https://archive.ics.uci.edu/ml/datasets/Skin+Segmentation

**Ход работы**

**1. Бизнес-цель**

**Цель проекта:** Разработать систему, способную автоматически определять, принадлежит ли пиксель изображения к коже человека (метка «1») или не к коже (метка «2») на основе его значений в цветовой модели RGB.

**Обоснование:**

* **Повышение эффективности обработки изображений.** Автоматическая сегментация кожи может применяться в системах видеонаблюдения, биометрической идентификации, медицинской диагностике, косметических приложениях и рекламных технологиях.
* **Улучшение качества анализа.** Чёткое выделение областей кожи позволяет ускорить и улучшить дальнейшую обработку изображений, например, при распознавании лиц или анализе экспрессии эмоций.
* **Коммерческая ценность.** Продукт, основанный на надёжном алгоритме сегментации, может стать конкурентным решением для компаний, работающих в сфере обработки изображений и аналитики данных.

## **2. Ресурсы и риски**

### **Ресурсы**

* **Набор данных:** Используется общедоступный датасет Skin Segmentation, содержащий пиксели с указанием значений Blue, Green, Red и меткой принадлежности к коже/не к коже.
* **Инструментарий:** Программное обеспечение и библиотеки для анализа данных и машинного обучения (например, Python, библиотеки NumPy, Pandas, scikit-learn, Matplotlib).
* **Аппаратные средства:** Компьютеры с достаточной вычислительной мощностью для предварительной обработки данных, обучения моделей и визуализации результатов.

### **Риски**

* **Качество данных:** Возможное наличие шумов, пропущенных или ошибочных значений, что может повлиять на точность модели. Набор данных может не охватывать достаточное разнообразие оттенков кожи и условий освещенности.
* **Недостаток данных:** При ограниченном объёме данных может возникнуть риск переобучения модели, что негативно скажется на её обобщающей способности.
* **Ограниченные ресурсы:** В случае нехватки вычислительных ресурсов возможно замедление разработки.

## **3. Цель анализа данных**

**Техническая цель:** Разработать алгоритм машинного обучения, который по входным значениям каналов Blue, Green, Red определяет принадлежность пикселя к коже человека (метка «1») или к другим объектам (метка «2»).

**Метрики качества:**

* Точность классификации (accuracy)
* Полнота (recall) и точность (precision)
* ROC-кривая и AUC для оценки разделительной способности модели

**Критерий успеха:** Пороговое значение метрики (точность не ниже 90 %) для демонстрации практической применимости алгоритма.

## **4. План проекта**

Для реализации проекта предлагается следующий план работ, основанный на методологии CRISP-DM:

1. **Понимание бизнес-целей и процессов (Business Understanding):**
   * Формулировка задачи с точки зрения бизнеса и технических требований.
   * Определение ключевых заинтересованных сторон, таких как заказчик, эксперты по данным и конечные пользователи.
   * Согласование ожидаемых результатов и критериев успеха.
2. **Изучение данных (Data Understanding):**
   * Сбор исходного набора данных из открытых источников (UCI Repository).
   * Проведение первичного анализа: изучение структуры данных, распределения значений RGB, проверка наличия пропусков и аномалий.
   * Выявление первичных гипотез относительно разделения классов.
3. **Подготовка данных (Data Preparation):**
   * Очистка данных от ошибок и пропусков.
   * Преобразование типов данных, нормализация признаков при необходимости.
   * Формирование обучающей, валидационной и тестовой выборок.
4. **Моделирование (Modeling):**
   * Выбор подходящих алгоритмов классификации (например, логистическая регрессия, деревья решений, SVM).
   * Проведение экспериментов с различными моделями и подбор гиперпараметров.
   * Оценка качества моделей с использованием выбранных метрик.
5. **Оценка (Evaluation):**
   * Сравнение полученных результатов с поставленными бизнес-целями.
   * Анализ рисков внедрения модели, проверка устойчивости результатов.
   * Подготовка отчёта с рекомендациями по дальнейшему развитию.
6. **Внедрение (Deployment):**
   * Разработка плана развертывания модели в рабочей среде.
   * Организация мониторинга качества работы модели и разработка стратегии её дообучения при необходимости.
   * Подготовка финальной документации и обучающих материалов для конечных пользователей.

**Вывод**

В результате выполнения лабораторной работы изучили основные принципы построения обучаемых алгоритмов