

# **Лабораторная работа 3. Рекомендации по отчету.**

## **Анализ набора данных "German Credit Data"**

### **1. Титульный лист**

Название работы: Анализ набора данных "German Credit Data"

Выполнил: [Фамилия Имя Отчество]

Группа: [номер группы]

Учебное заведение: [Название учебного заведения]

Факультет: [название факультета]

Лабораторная работа №: [номер]

Дата выполнения: [дата]

<https://edufpmi.bsu.by/course/view.php?id=625>

### **2. Введение**

В этом разделе кратко описываем:

- что планировали сделать?
- какие методы планировали применить?
- какие вопросы заинтересовали в наборе данных?

**Пример:**

Целью данной работы является проведение анализа набора данных "German Credit Data" для выявления ключевых факторов, влияющих на статус кредита, а также моделирования данных и их визуализации.

### **3. Загрузка и подготовка данных**

Источник данных:

- Описание файла, формат, расположение файла (локальный или через ссылку).

Структура данных:

- Количество строк и признаков, названия признаков.

Обработка пропущенных значений:

- Какие методы использовались?
- Заменены пропущенные значения, удалены или обработаны иным способом?

Кодирование категориальных переменных:

- Какие признаки были закодированы? Каким методом (например, one-hot, label encoding)?

**Пример:**

Данные были загружены из файла german.csv. В наборе присутствовали пропущенные значения, которые были заполнены медианными значениями для числовых признаков. Категориальные признаки были закодированы методом label encoding.

### **4. Анализ данных**

Описание числовых признаков:

- Использованные статистические показатели (среднее, медиана, минимум, максимум, стандартное отклонение).

Распределения признаков:

- Основные особенности распределений: асимметрии, выбросы, нормальность.

Анализ категориальных признаков:

- Частоты, наиболее часто встречающиеся категории.

Ключевые выводы:

- Какие признаки наиболее информативны? Есть ли взаимосвязи, важные для дальнейшего анализа?

## **5. Визуальный анализ**

Построенные графики:

- Гистограммы для числовых признаков
- Boxplot-ы для выявления выбросов
- Тепловая карта корреляций
- Барплоты для категориальных признаков

Интерпретация построенных графиков:

- Краткие комментарии к каждому графику и сделанные выводы.

## **6. Работа с базой данных SQLite**

Создание базы данных и таблицы:

- Описание структуры таблицы.

Вставка данных:

- Использованные команды или скрипты.

Выполненные SQL-запросы:

- Примеры выборок, группировок, агрегатов. Например:
  - Средний кредит по статусу
  - Количество клиентов по целям кредита
  - Распределение по возрастным группам

Результаты запросов:

- Краткое описание полученных данных или скриншоты.

## **7. Итоговые выводы**

Здесь подведите итоги анализа:

- Какие признаки оказались наиболее значимыми?
- Какие взаимосвязи обнаружены?
- Какие рекомендации можно дать на основе данных?
- Итоговые мысли по результатам работы.

## **8. Заключение**

Краткое резюме проделанной работы и возможные направления дальнейшего анализа или улучшений.