**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра САПР**

отчет

**по КУРСОВОЙ РАБОТЕ**

Тема: Анализ набора данных с характеристиками различных видов стекла

|  |  |  |
| --- | --- | --- |
| Студент гр. 1301 |  | Ахметзянов Д.А. |
| Руководитель |  | Боброва Ю.О. |

Санкт-Петербург

**2025ЗАДАНИЕ**

**на научно-исследовательскую работу**

|  |  |  |
| --- | --- | --- |
| Студент Ахметзянов Д.А. | | |
| Группа 1301 | | |
| Тема НИР: Анализ набора данных с характеристиками различных видов стекла | | |
| Задание на НИР:  На основе быбранного набора данных дать подробное описание вида данных, основные особенности. Написать классификатор для данного набора данных на основе логистической регрессии, случайного леса и нейронной сети. | | |
| Сроки выполнения НИР: 10.02.2025 - 24.03.2025 | | |
| Дата сдачи отчета: | | |
| Дата защиты отчета: | | |
|  | | |
| Студент |  | Ахметзянов Д.А. |
| Руководитель |  | Боброва Ю.О. |

**Аннотация**

Данная научно-исследовательская работа посвящена разработке и сравнительному анализу методов классификации для различных типов данных. Исследование включает анализ и подробное описание набора данных посвященного характеристикам различных типов стекла. В работе реализованы классификаторы на основе логистической регрессии, случайного леса и нейронной сети прямого распространения с обратным распространением ошибки.

**Summary**

This research work is devoted to the development and comparative analysis of classification methods for various types of data. The study includes analysis and detailed description of a dataset focused on the characteristics of different types of glass. The work implements classifiers based on logistic regression, random forest, and a feedforward neural network with backpropagation.

**содержание**

**введение**

Данная научно-исследовательская работа направлена на разработку и сравнительный анализ методов классификации применительно к набору данных о различных типах стекла. Основная цель исследования — определить наиболее эффективный метод классификации для данного набора данных путем сравнения трех различных подходов: логистической регрессии, случайного леса и нейронной сети прямого распространения.В рамках поставленной цели решаются следующие задачи:

1. Подробный анализ и описание набора данных о характеристиках различных типов стекла
2. Реализация и настройка трех различных классификаторов
3. Сравнительная оценка эффективности методов по критериям точности, устойчивости и вычислительной сложности
4. Выявление оптимального метода классификации для данного типа данных.
5. Описание набора данных

Используемый в данной работе набор данных взят с сайта <https://archive.ics.uci.edu/>, он предназначен для идентификации типов стекла и был создан для криминалистических исследований. Он содержит информацию о различных типах стекла, которые можно найти на месте преступления и использовать в качестве улик.

* Количество образцов: 214
* Количество атрибутов: 10 (включая ID) плюс целевой атрибут класса

**Атрибуты**:

* ID: Номер образца (от 1 до 214)
* RI: Коэффициент преломления
* Na: Содержание натрия (в весовых процентах соответствующего оксида)
* Mg: Содержание магния
* Al: Содержание алюминия
* Si: Содержание кремния
* K: Содержание калия
* Ca: Содержание кальция
* Ba: Содержание бария
* Fe: Содержание железа
* Type: Тип стекла (целевой атрибут)

**Классы стекла**:

* Оконное стекло, обработанное флоат-методом (building\_windows\_float\_processed)
* Оконное стекло, не обработанное флоат-методом (building\_windows\_non\_float\_processed)
* Автомобильное стекло, обработанное флоат-методом (vehicle\_windows\_float\_processed)
* Стекло для контейнеров (containers)
* Стекло для посуды (tableware)
* Стекло для фар (headlamps)

**Распределение классов**:

* Оконное стекло: 163 образца
  + Обработанное флоат-методом: 70 образцов
  + Не обработанное флоат-методом: 76 образцов
  + Автомобильное стекло: 17 образцов
  + Обработанное флоат-методом: 17 образцов
* Не оконное стекло: 51 образец
  + Контейнеры: 13 образцов
  + Посуда: 9 образцов
  + Фары: 29 образцов

Как видно данные распределены неравномерно что может привести к худшей точности классификации для тех классов, для которых набор данных меньше.

1. Анализ набора данных

Проанализируем каждый из параметров набора данных

1. Параметр: RI (Показатель преломления)
   1. Медианные значения

Медиана (общая): 1.5177

Cреднее значение: 1.5184

Медиана по классам:

* 1. Класс 1 (building\_windows\_float): 1.5178
  2. Класс 2 (building\_windows\_non\_float): 1.5171
  3. Класс 3 (vehicle\_windows\_float): 1.5177
  4. Класс 5 (containers): 1.5199
  5. Класс 6 (tableware): 1.5189
  6. Класс 7 (headlamps): 1.5165
  7. Средние значения и стандартные отклонения

Стандартное отклонение: 0.0030

По классам:

* 1. Класс 1 (building\_windows\_float): среднее = 1.5187, стд. откл. = 0.0023
  2. Класс 2 (building\_windows\_non\_float): среднее = 1.5186, стд. откл. = 0.0038
  3. Класс 3 (vehicle\_windows\_float): среднее = 1.5180, стд. откл. = 0.0019
  4. Класс 5 (containers): среднее = 1.5189, стд. откл. = 0.0033
  5. Класс 6 (tableware): среднее = 1.5175, стд. откл. = 0.0031
  6. Класс 7 (headlamps): среднее = 1.5171, стд. откл. = 0.0025

Как можно заметить для всех классов медианное значение близко к 1.52. А по тому что среднее значение близко к медиане и стандартное отклонение значительно меньше среднего можно понять, что изменения данного параметра можно считать погрешностью и при классификации не учитывать.

Рассмотрим распределение этого параметра по классам

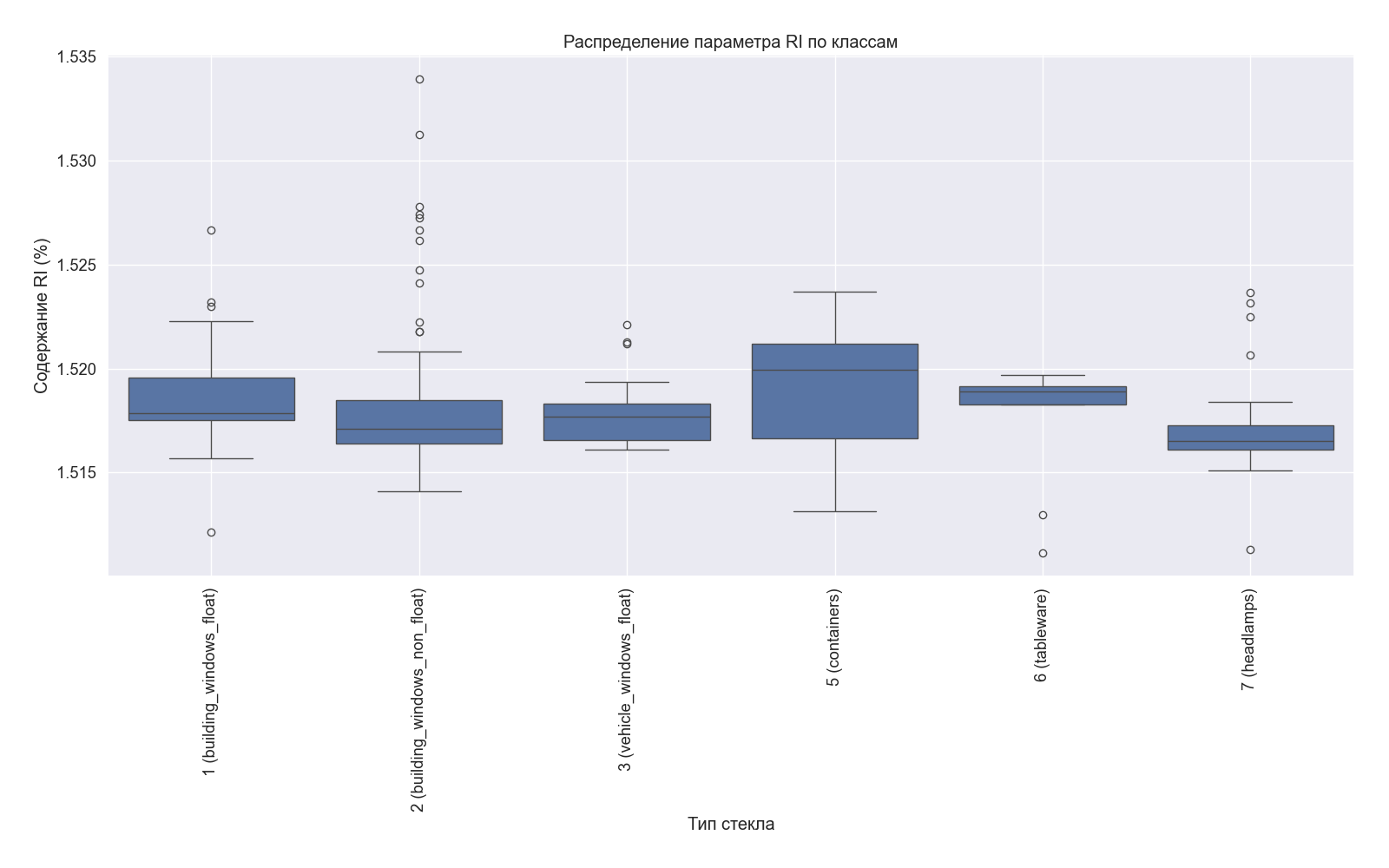


Рис. 1 Распределение параметра RI по классам

По рис. 1 видно, что сделанные ранее выводы верны и несмотря на выбросы данных, отклонение значений параметра минимально.

2. Параметр Na (Содержание натрия)

2.1. Медианные значения

Медиана (общая): 13.3000

Среднее значение: 13.4079

Медиана по классам:

1. Класс 1 (building\_windows\_float): 13.1950
2. Класс 2 (building\_windows\_non\_float): 13.1550
3. Класс 3 (vehicle\_windows\_float): 13.4200
4. Класс 5 (containers): 12.9700
5. Класс 6 (tableware): 14.4000
6. Класс 7 (headlamps): 14.3900
7. Средние значения и стандартные отклонения

Стандартное отклонение: 0.8166

По классам:

1. Класс 1 (building\_windows\_float): среднее = 13.2423, стд. откл. = 0.4993
2. Класс 2 (building\_windows\_non\_float): среднее = 13.1117, стд. откл. = 0.6642
3. Класс 3 (vehicle\_windows\_float): среднее = 13.4371, стд. откл. = 0.5069
4. Класс 5 (containers): среднее = 12.8277, стд. откл. = 0.7770
5. Класс 6 (tableware): среднее = 14.6467, стд. откл. = 1.0840
6. Класс 7 (headlamps): среднее = 14.4421, стд. откл. = 0.6864

Анализируя данные по содержанию натрия, можно заметить существенные различия между классами. Особенно выделяются классы 6 (tableware) и 7 (headlamps) с медианными значениями около 14.4, что значительно выше, чем у других классов. Стандартное отклонение составляет примерно 6% от среднего значения, что указывает на заметную вариативность этого параметра между классами.

Рассмотрим распределение этого параметра по классам

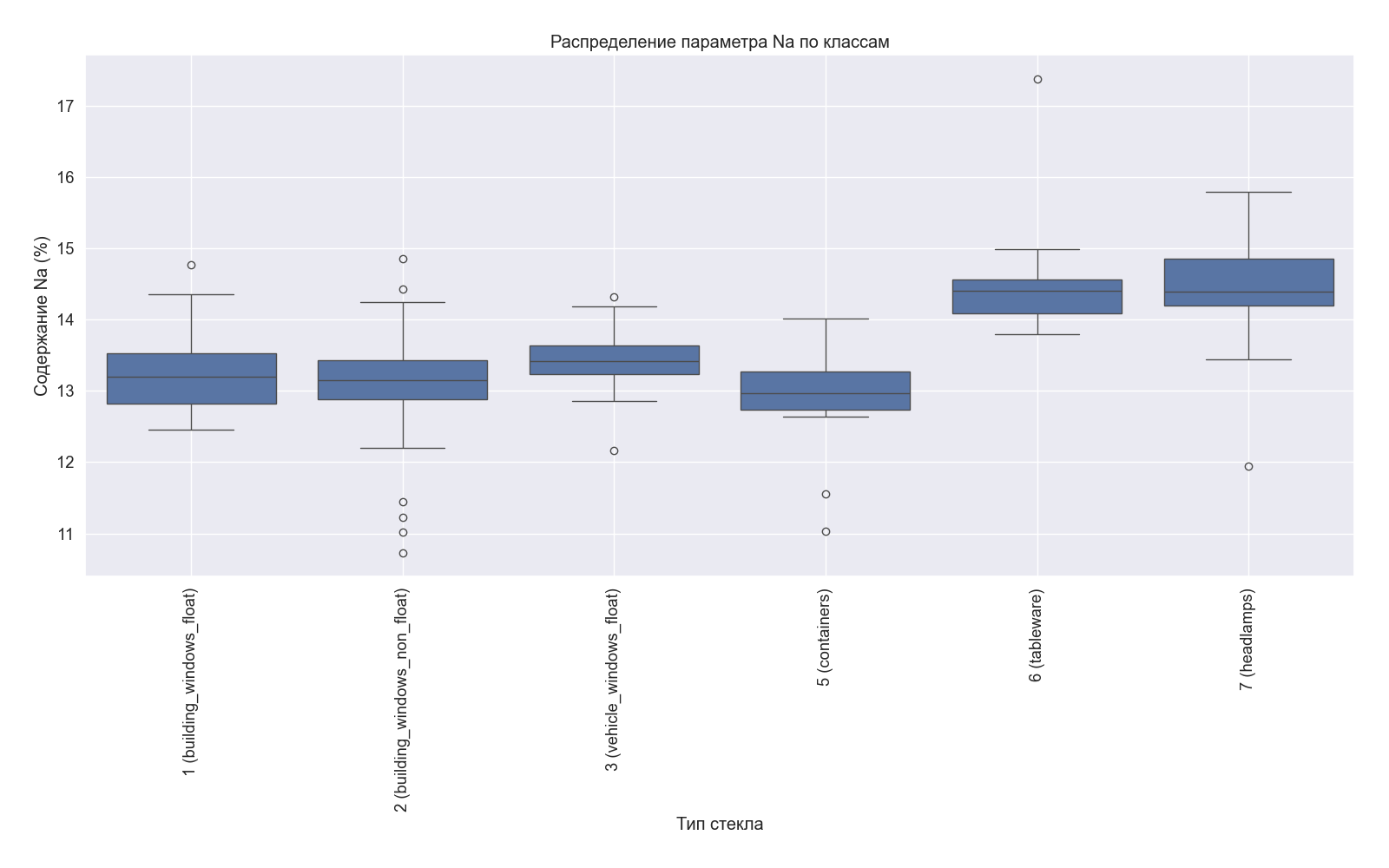


Рис. 2 Распределение параметра Na по классам

По рис. 2 видно, что содержание натрия может служить хорошим дискриминативным признаком, особенно для разделения классов 6 и 7 от остальных типов стекла.

1. **2. ВТОРОЙ раздел**

**2.1. Первый подраздел второго раздела**

**2.2. Второй подраздел второго раздела**

**3. третий раздел**

**3.1. Первый подраздел третьего раздела**

**3.2. Второй подраздел третьего раздела**

**заключение**

Кратко подвести итоги, проанализировать соответствие поставленной цели и полученного результата.

**список использованных источников**

***Ниже представлены примеры библиографического описания, В качестве названия источника в примерах приводится вариант, в котором применяется то или иное библиографическое описание.***

1. Иванов И. И. Книга одного-трех авторов. М.: Издательство, 2010. 000 с.

2. Книга четырех авторов / И. .И Иванов, П. П. Петров, С. С. Сидоров, В. В. Васильев. СПб.: Издательство, 2010. 000 с.

3. Книга пяти и более авторов / И. И. Иванов, П. П. Петров, С. С. Сидоров и др.. СПб.: Издательство, 2010. 000 с.

4. Описание книги под редакцией / под ред. И.И. Иванова СПб., Издательство, 2010. 000 с.

5. Иванов И.И. Описание учебного пособия и текста лекций: учеб. пособие. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2010. 000 с.

6. Описание методических указаний / сост.: И.И. Иванов, П.П. Петров. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2010. 000 с.

7. Иванов И.И. Описание статьи с одним-тремя авторами из журнала // Название журнала. 2010, вып. (№) 00. С. 000–000.

8. Описание статьи с четырьмя и более авторами из журнала / И. И. Иванов, П. П. Петров, С. С. Сидоров и др. // Название журнала. 2010, вып. (№) 00. С. 000–000.

9. Иванов И.И. Описание тезисов доклада с одним-тремя авторами / Название конференции: тез. докл. III международной науч.-техн. конф., СПб, 00–00 янв. 2000 г. / СПбГЭТУ «ЛЭТИ», СПБ, 2010, С. 000–000.

10. Описание тезисов доклада с четырьмя и более авторами / И. И. Иванов, П. П. Петров, С. С. Сидоров и др. // Название конференции: тез. докл. III международной науч.-техн. конф., СПб, 00–00 янв. 2000 г. / СПбГЭТУ «ЛЭТИ», СПБ, 2010, С. 000–000.

11. Описание электронного ресурса // Наименование сайта. URL: http://east-front.narod.ru/memo/latchford.htm (дата обращения: 00.00.2010).

12. ГОСТ 0.0–00. Описание стандартов. М.: Изд-во стандартов, 2010.

13. Пат. RU 00000000. Описание патентных документов / И. И. Иванов, П. П. Петров, С. С. Сидоров. Опубл. 00.00.2010. Бюл. № 00.

14. Иванов И.И. Описание авторефератов диссертаций: автореф. дисс. канд. техн. наук / СПбГЭТУ «ЛЭТИ», СПБ, 2010.

15. Описание федерального закона: Федер. закон [принят Гос. Думой 00.00.2010] // Собрание законодательств РФ. 2010. № 00. Ст. 00. С. 000–000.

16. Описание федерального постановления: постановление Правительства Рос. Федерации от 00.00.2010 № 00000 // Опубликовавшее издание. 2010. № 0. С. 000–000.

17. Описание указа: указ Президента РФ от 00.00.2010 № 00 // Опубликовавшее издание. 2010. № 0. С. 000–000.

**приложение А**

**название Приложения**