

**Вашакидзе Нателла Семеновна, Филиппова Галина Викторовна,
Рауш Наталья Леонидовна**, Старшие преподаватели,
Сахалинский государственный университет, Южно-Сахалинск

Осипов Геннадий Сергеевич, д.т.н., профессор,
Сахалинский государственный университет, Южно-Сахалинск

ИССЛЕДОВАНИЕ КЛАССИЧЕСКОЙ ЗАДАЧИ КЛАССИФИКАЦИИ ИРИСОВ ФИШЕРА СОВРЕМЕННЫМИ ИНСТРУМЕНТАЛЬНЫМИ СРЕДСТВАМИ

Аннотация: Проведено исследование классической задачи классификации ирисов Фишера с помощью современных инструментальных средств интеллектуальной обработки данных и машинного обучения.

Ключевые слова: решение задач классификации и кластеризации.

Введение.

Целью работы является исследование классической задачи классификации ирисов [1] с помощью современной аналитической low-code платформы интеллектуального анализа данных Loginom [2] (внесена в реестр [российских программ](#)) и системы символьной математики Wolfram Mathematica [3], ориентированной на решение задач машинного обучения и искусственного интеллекта.

В исследовании используется обучающая выборка [4], фрагмент которой представлен в таблице 1.

Таблица 1

Фрагмент исходных данных

#	Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Класс
1	5,1	3,5	1,4	0,2	Iris-setosa
2	4,9	3	1,4	0,2	Iris-setosa
3	4,7	3,2	1,3	0,2	Iris-setosa
4	4,6	3,1	1,5	0,2	Iris-setosa
5	5	3,6	1,4	0,2	Iris-setosa

Исследование задачи в среде Loginom

Ключевым элементом в аналитической платформе Loginom является сценарий, который для данной задачи представлен на рисунке 1.

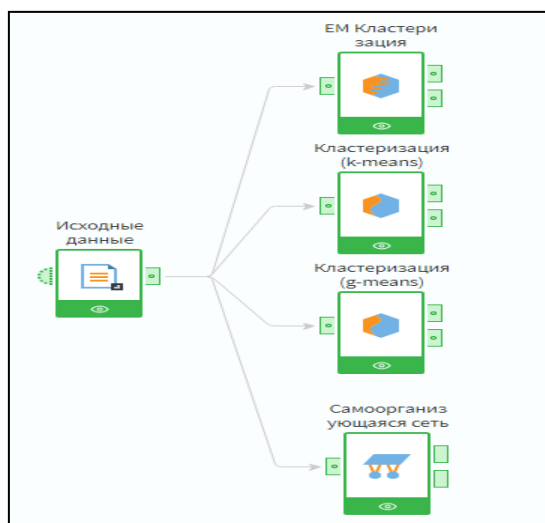


Рис. 1 Диаграмма сценария



Далее представлены результаты тестовой кластеризации ирисов с помощью искусственной самоорганизующейся нейронной сети. Так на рисунке 2 представлены функции распределения объектов по кластерам.

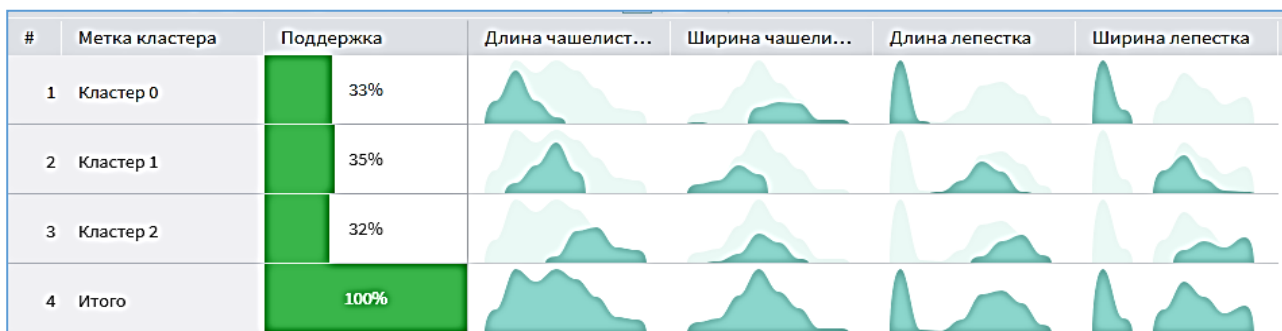


Рис. 2 Функции распределения

На рисунке 3 приведена информация о центрах кластеров.

#	Метка кластера	Поддержка	Длина чашелист...	Ширина чашели...	Длина лепестка	Ширина лепестка
1	Кластер 0	33%	5,02	3,43	1,47	0,247
2	Кластер 1	35%	5,79	2,66	4,36	1,42
3	Кластер 2	32%	6,74	3,09	5,47	1,98
4	Итого	100%				

Рис. 3 Центры кластеров

Величины стандартных отклонений параметров ирисов по каждому из кластеров представлены на рисунке 4.

#	Метка кластера	Поддержка	Длина чашелист...	Ширина чашели...	Длина лепестка	Ширина лепестка
1	Кластер 0	33%	0,35	0,38	0,17	0,107
2	Кластер 1	35%	0,39	0,25	0,56	0,31
3	Кластер 2	32%	0,49	0,26	0,64	0,34
4	Итого	100%	0,83	0,43	1,76	0,76

Рис. 4 Стандартные отклонения по кластерам

На рисунке 5 приведена обобщенная диаграмма кластеров с параметрами ирисов в них.

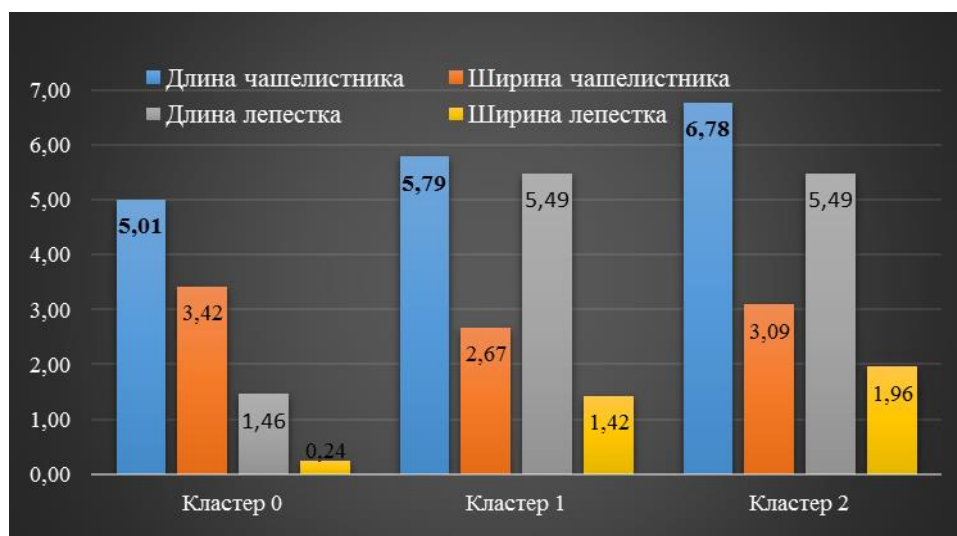


Рис. 5 Диаграммы кластеров по параметрам ирисов



Двойственная диаграмма распределения параметров ирисов по кластерам представлена на рисунке 6.

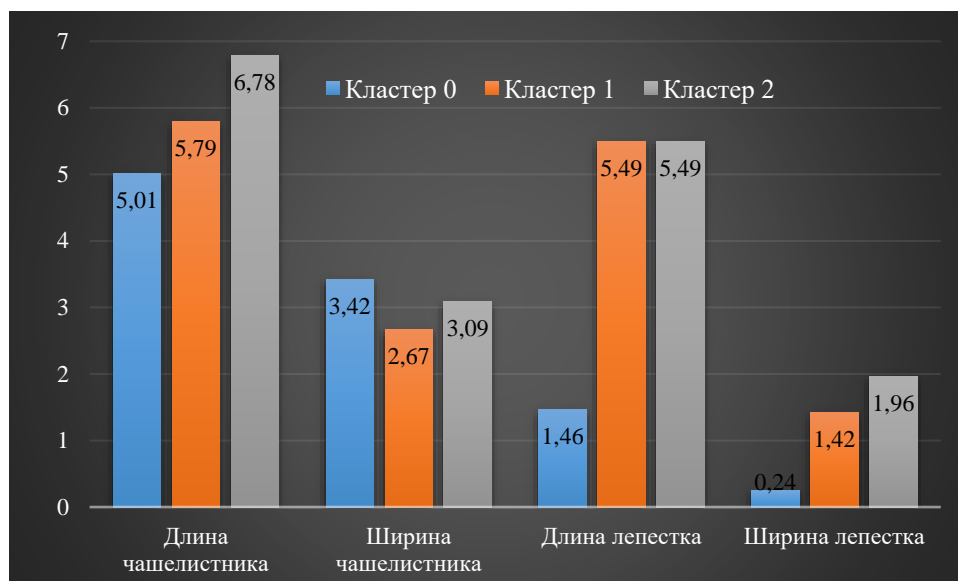


Рис. 6 Распределения параметров ирисов по кластерам

Исследование задачи в среде Wolfram Mathematica

В системе символьной математики (компьютерной алгебры) Wolfram Mathematica решались задачи синтеза классификатора одними из методов машинного обучения и разбиение на кластеры с обеспечением возможности последующей классификации новых объектов.

На рисунке 7 приведены основные операторы считывания исходных данных (обучающей выборки) из текстового файла.

```

Ввод исходных данных

path = StringJoin[NotebookDirectory[], "Ирисы.txt"];
      [соединить с... |директория файла блокнота]

data = Import[path, "Data"];
      [импорт]
x = data[[All, 1 ;; 4]];
      [всё]
y = Flatten[data[[All, 5 ;; 5]]];
      [уплостить] [всё]
    
```

Рис. 7 Операторы ввода исходных данных

Для обучения классификатора использовалась многослойная нейронная сеть. на рисунке 8 представлен соответствующий оператор.

```

Обучение нейросетевого классификатора

Классификатор = Classify[x -> y, Method -> "NeuralNetwork"] ;
      [классифицировать] [метод]

ClassifierFunction [
  [ ]
  Input type: Mixed (number: 4)
  Classes: setosa, versicolor, virginica
  Method: NeuralNetwork
  Number of training examples: 150
]
    
```

Рис. 8 Обучение классификатора



Рисунок 9 демонстрирует этап тестирования нейросетевого классификатора по обобщенным критериям качества обучения.

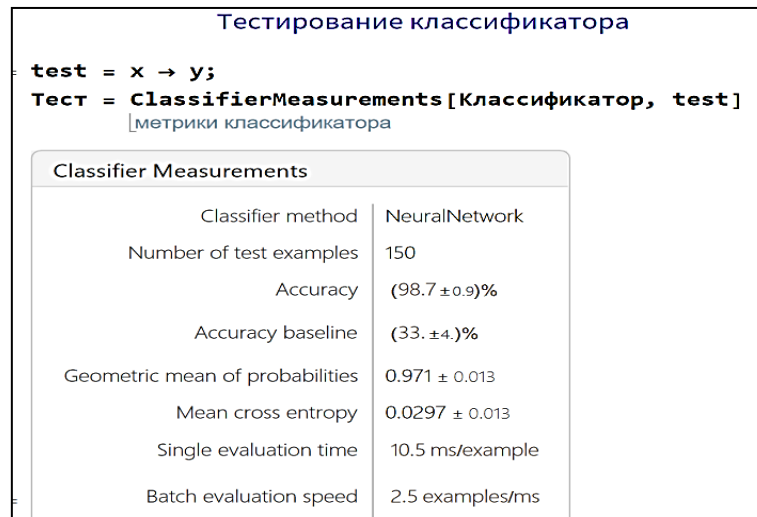


Рис. 9 Оценка качества обучения классификатора

На рисунке 10 представлена матрица ошибок обучения на используемой обучающей выборке.

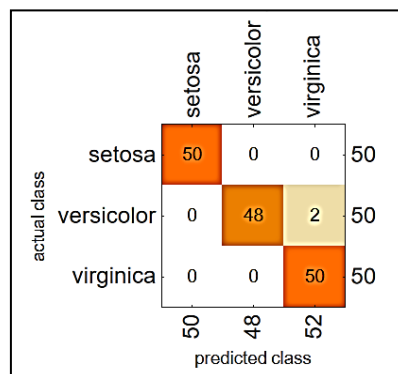


Рис. 10 Матрица ошибок обученного классификатора

Отметим, что «ошибочная» классификация двух ирисов не является следствием недостатка используемого метода машинного обучения. Практически все классические методы классификации и кластеризации выдают такое же распределение объектов на классы и кластеры, соответственно. Вероятно, причина этого восходит к качеству первоначальной стандартной классификации ирисов, которая и используется в качестве обучающей выборки.

На рисунке 11 даны операторы простого разбиения на 3 кластера и классификатор отнесения к одному из кластеров.

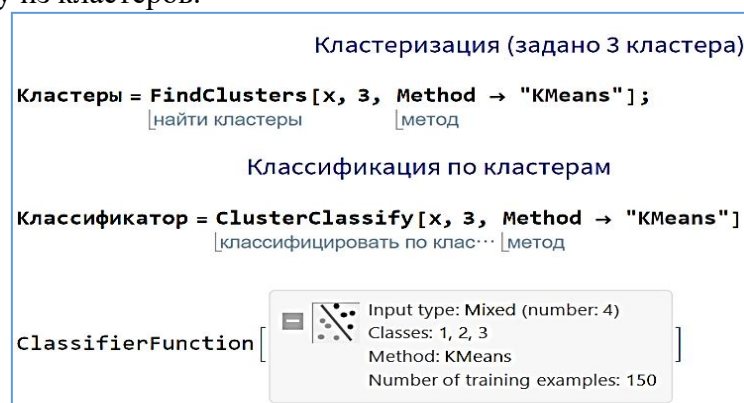


Рис. 11 Кластеризация и классификатор по кластерам



Очевидно, итогом построения классификатора является обеспечение возможности отнесения нового объекта к одному из известных классов. Соответствующий оператор представлен на рисунке 12.

```
Классификация нового объекта  
Классификатор[{5.1, 3.5, 1.4, 0.2}, "Probabilities"]  
<| 1 → 4.80024 × 10-15, 2 → 1., 3 → 1.19614 × 10-9 |>
```

Рис. 12 Классификация нового объекта

В данном случае новый объект отнесен ко 2 классу.

Заключение

В работе представлены основные этапы, схемы и операторы решения классической задачи классификации ирисов Фишера на базе современных инструментальных сред – аналитической low-code платформы интеллектуального анализа данных Loginom и системы символьной математики Wolfram Mathematica, ориентированной на решение задач машинного обучения систем искусственного интеллекта. В качестве моделей обучения использовалась самоорганизующаяся и многослойная нейронные сети.

Список литературы:

1. The iris data set: In search of the source of virginica. URL: <https://www.semanticscholar.org/paper/The-iris-data-set%3A-In-search-of-the-source-of-Unwin-Kleinman/4599862ea877863669aba8e63a3c707a787d5d7e> (дата обращения 23.02.2024).
2. Low-Code Платформа для реализации аналитических процессов. URL: <https://loginom.ru/> (дата обращения 23.02.2024).
3. Русскоязычная версия Wolfram. URL: <https://www.wolfram.com/russian/?source=frontpage-stripe> (дата обращения 23.02.2024)
4. UCI Machine Learning Repository. URL: <https://archive.ics.uci.edu/dataset/53/iris> (дата обращения 23.02.2024)

