федеральное государственное бюджетное образовательное учреждение высшего образования "Волгоградский государственный технический университет"

Факультет электроники и вычислительной техники

Кафедра «Системы автоматизированного проектирования и поискового конструирования»

**Лабораторная работа № 4: Классификация и кластеризация в R**

1. **Цель работы**

* Освоить основные методы классификации и кластеризации
* Приобрести основные навыки работы с классификацией и кластеризацией временных рядов в R

1. **Задачи**

* Углубить и закрепить знания по основным методам классификации и кластеризации.
* Научить делать классификацию и в R
* Совершенствовать навыки самостоятельной работы.

1. **План выполнения работы**

**3.1 Основные алгоритмы классификации и кластеризации**

**- Алгоритмы классификации:**

* Linear classifiers
  + Logistic regression
  + Naive Bayes classifier
  + Fisher’s linear discriminant
* Support vector machines
  + Least squares support vector machines
* Quadratic classifiers
* Kernel estimation
  + k-nearest neighbor
* Decision trees
  + Random forests
* Neural networks
* Learning vector quantization

**- Алгоритмы кластеризации:**

* **Linear clustering algorithm**
  + k-means clustering algorithm
  + Fuzzy c-means clustering algorithm
  + Hierarchical clustering algorithm
  + Gaussian(EM) clustering algorithm
  + Quality threshold clustering algorithm
* **Non-linear clustering algorithm**
  + MST based clustering algorithm
  + kernel k-means clustering algorithm
  + Density-based clustering algorithm

Для проверки алгоритмов классификации и кластеризации в R воспользуемся распространённым набором данных **IRIS**-классификация цветков ириса по небольшому набору параметров. **IRIS** состоит из данных о 150 экземплярах [ириса](https://ru.wikipedia.org/wiki/%D0%98%D1%80%D0%B8%D1%81_(%D1%80%D0%B0%D1%81%D1%82%D0%B5%D0%BD%D0%B8%D0%B5)), по 50 экземпляров из трёх видов — [Ирис щетинистый](https://ru.wikipedia.org/wiki/%D0%98%D1%80%D0%B8%D1%81_%D1%89%D0%B5%D1%82%D0%B8%D0%BD%D0%B8%D1%81%D1%82%D1%8B%D0%B9) (*Iris setosa*), [Ирис виргинский](https://ru.wikipedia.org/w/index.php?title=%D0%98%D1%80%D0%B8%D1%81_%D0%B2%D0%B8%D1%80%D0%B3%D0%B8%D0%BD%D1%81%D0%BA%D0%B8%D0%B9&action=edit&redlink=1) (*Iris virginica*) и [Ирис разноцветный](https://ru.wikipedia.org/w/index.php?title=%D0%98%D1%80%D0%B8%D1%81_%D1%80%D0%B0%D0%B7%D0%BD%D0%BE%D1%86%D0%B2%D0%B5%D1%82%D0%BD%D1%8B%D0%B9&action=edit&redlink=1) (*Iris versicolor*). Для каждого экземпляра измерялись четыре характеристики (в [сантиметрах](https://ru.wikipedia.org/wiki/%D0%A1%D0%B0%D0%BD%D1%82%D0%B8%D0%BC%D0%B5%D1%82%D1%80)):

1. Длина наружной доли [околоцветника](https://ru.wikipedia.org/wiki/%D0%9E%D0%BA%D0%BE%D0%BB%D0%BE%D1%86%D0%B2%D0%B5%D1%82%D0%BD%D0%B8%D0%BA) ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *sepal length*);
2. Ширина наружной доли околоцветника ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *sepal width*);
3. Длина внутренней доли околоцветника ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *petal length*);
4. Ширина внутренней доли околоцветника ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *petal width*).

**3.2. На основании набора данных IRIS требуется построить правило классификации, определяющее вид растения по данным измерений.**

**3.3 Кластеризация набора данных IRIS**