Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**ЛАБОРАТОРНАЯ РАБОТА №6**

**Дисциплина: Обработка больших данных**

Работу выполнила: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_В.В.Гончаренко

Направление подготовки: 02.03.02 Фундаментальная информатика и информационные технологии

Преподаватель: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Т.А.Приходько

**Цель работы:** закрепить знания об алгоритмах классификации и кластеризации данных, ознакомиться с некоторыми функциями языка R, осуществляющими этот вид анализа, принципами их работы. Научиться визуализировать результаты работы функций кластерного анализа и классификаторов, интерпретировать полученные результаты. Научиться выполнять классификацию на основе формулы Байеса и деревьев решений

**Ход работы:**

**Часть 1.** **Задачи классификации и кластеризации**

Из исходного csv файла данные были импортированы в RStduio.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1 – Исходная таблица с данными

В данной задаче переменные различны, поэтому необходимо выполнить нормализацию данных. Для стандартизации исключим колонки n\_ и type. После нормализации вернем отброшенные столбцы.

Изображение выглядит как текст, окно

Автоматически созданное описание

Рисунок 3 – Стандартизированная таблица с данными

По полученным данным построим дендрограмму.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 4 – Дендограмма, построенная по данным

Целесообразнее выделить 4 кластеров, выделим их красным цветом.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 5 – Дендограмма с кластерами.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 6 – Диаграмма «Каменная осыпь».

График имеет резкие пики и спады, это указывает на то, что происходят более резкие изменения в размерах «камней» или в количестве материала.

Далее представим списки городов, разбитых по кластерам.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 7 – Типы семей, разбитые по кластерам

Далее представим распределения характеристик по всем кластерам.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 8 – Распределение характеристик в каждом кластере.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 5 – Распределение характеристики Хлеб.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 6 – Распределение характеристики Овощи.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 7 – Распределение характеристики Фрукты.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 8 – Распределение характеристики Мясо.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 9 – Распределение характеристики Птица.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 10 – Распределение характеристики Молоко.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 10 – Распределение характеристики Вино.

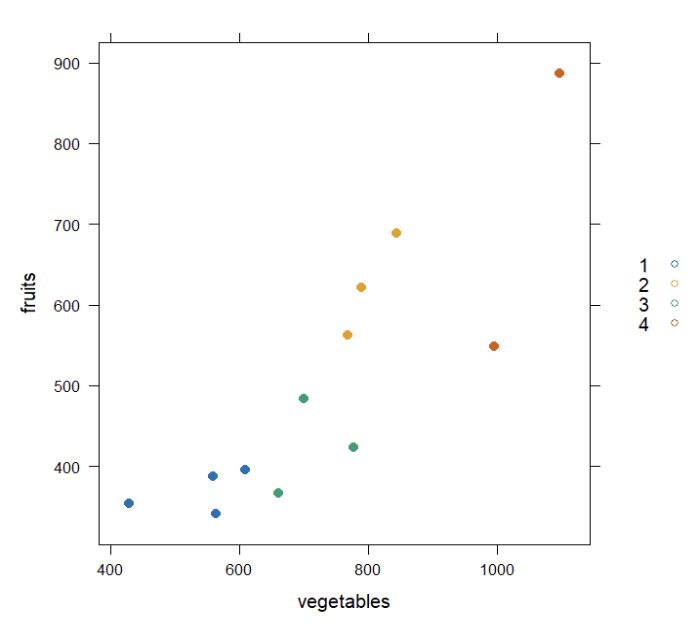


Рисунок 10 – Диаграмма рассеивания для фруктов и овощей.

После были построены «ящики с усами», показанные на рисунке 11.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 11 – Диаграмма ящик с усами.

На нем мы можем видеть, как распределяются индексы потребления хлеба по кластерам.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 12 – График, показывающий распределение мяса и овощей относительно вина.

Можно сделать вывод, мясо с вином пьют чаще чем с овощами.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 13 – Трехмерный график

Данный график классифицирует типы семей по хлеба овощей и мяса

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 14 – Зависимость овощей и фруктов в кластерах.

На данном шаге была произведена кластеризация k-means. Данная кластеризация позволяет отследить взаимосвязь двух непрерывных переменных. В данном случае – овощей и фруктов в кластерах.