МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ

ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных систем и технологий

Кафедра «Измерительно-вычислительные комплексы»

Отчет по лабораторной работе №4

по дисциплине «Методы искусственного интеллекта»

Выполнила

ст. гр. ИСТбд-41 Карташова М.В.

Проверил:

Шишкин В.В.

Ульяновск, 2022

Задание:

1. Создать симулированный набор данных и записать его на диск в виде csv файла со следующими параметрами:

- продукт;

- сладость;

- хруст;

-класс.

|  |  |  |  |
| --- | --- | --- | --- |
| продукт | сладость | хруст | класс |
| Яблоко | 7 | 7 | Фрукт |
| салат | 2 | 5 | Овощ |
| бекон | 1 | 2 | Протеин |
| банан | 9 | 1 | Фрукт |
| орехи | 1 | 5 | Протеин |
| рыба | 1 | 1 | Протеин |
| сыр | 1 | 1 | Протеин |
| виноград | 8 | 1 | Фрукт |
| морковь | 2 | 8 | Овощ |
| апельсин | 6 | 1 | Фрукт |

Подготовить для классификации несколько примеров в соответствии с рисунком

Хруст

Фрукты

Овощи

Протеин

Сладость

2. Запрограммировать метрический классификатор по методу k-NN. Для проверки решить ту же задачу методом k-NN библиотеки sklearn.

3. Прочитать сгенерированный набор данных. Настроить классификатор. Провести эксперимент по классификации с контролем для подготовленных примеров.

4. По возможности результаты визуализировать.

5. Ввести в набор данных и примеры продукты еще одного класса (возможно изменив набор параметров) и повторить эксперимент.

6. По результатам эксперимента подготовить отчет по лабораторной работе.

Ход выполнения работы:

1. Был сгенерирован csv файл, в котором находятся данные из задания, плюс еще несколько примеров, соответствующие рисунку из задания. Здесь классу 0 соответствуют фрукты, 1 – овощи, 2 – протеин.

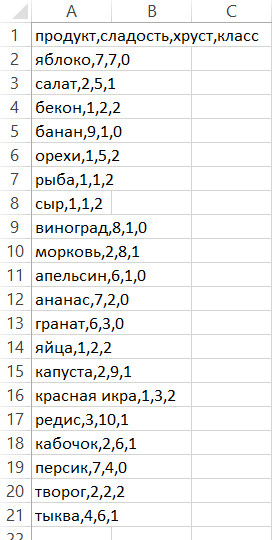


Рис.1. Файл dataset1.csv

2. Далее был запрограммирован метрический классификатор по методу k-NN. Его необходимо было настроить, для более точного прогнозирования класса. Так например, при тестовой выборке в 40% и К=5, результатом программы было следующее:



Рис.2. Результат классификатора при тестовой выборке в 40% и К=5

Методом подбора значений, получилось добиться результата работы классификатора, со 100% точностью:



Рис.3. Результат классификатора при тестовой выборке в 20% и К=3

3. Далее написана функция, решающая ту же задачу методом k-NN библиотеки sklearn.



Рис.4. Результат классификатора sklearn при тестовой выборке в 20% и К=3

4. Затем был сгенерирован второй аналогичный по структуре csv файл, но теперь в нем присутствуют больше записей, так как добавлен новый класс продуктов: 3 – сладости. Таким образом классу 0 соответствуют фрукты, 1 – овощи, 2 – протеин, 3 - сладости.

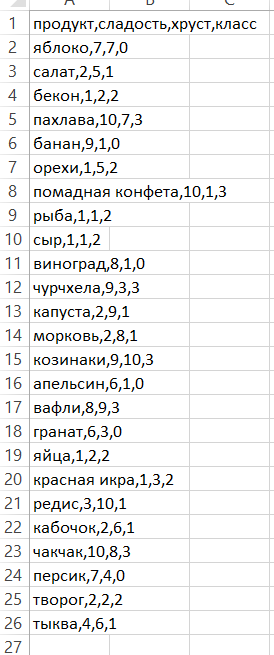


Рис.5. Файл dataset2.csv

Данный класс располагается по отношению к другим следующим образом:

Хруст

Сладости

Фрукты

Овощи

Протеин

Сладость

5. Далее результаты работы классификаторов получилось визуализировать.

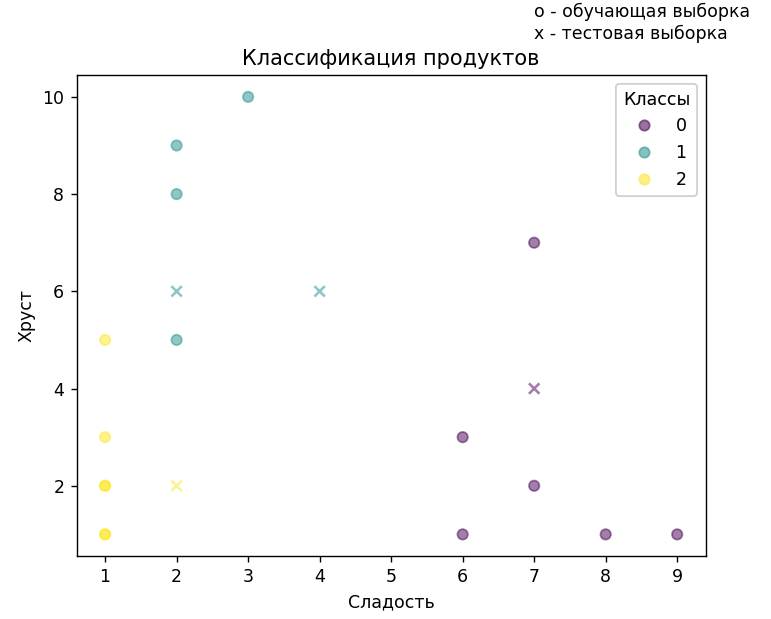


Рис.6. Визуализация работы метрического классификатора на наборе с 3 классами при тестовой выборке в 20% и К=3

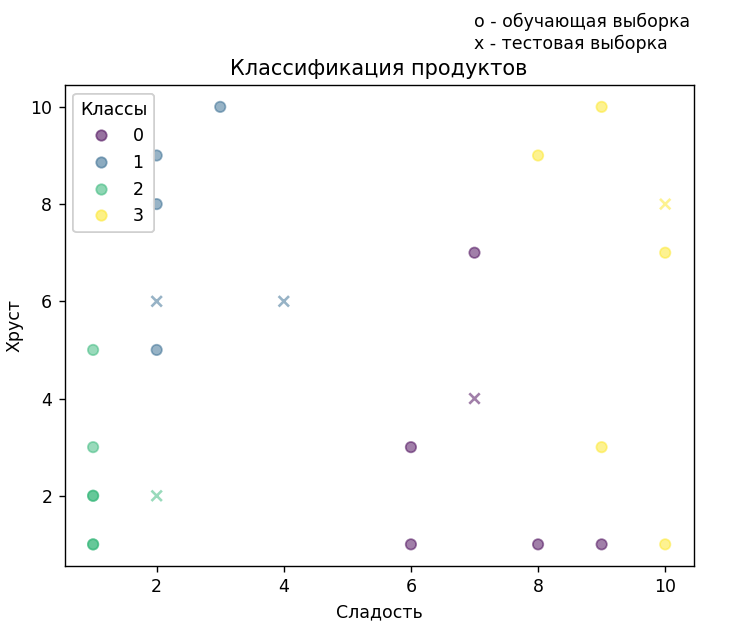


Рис.7. Визуализация работы классификатора sklearn на наборе с 4 классами при тестовой выборке в 20% и К=3

Данные рисунки полностью соответствуют графику распределения из задания (для 3 классов) и графику, спроектированному в л.р. для нового класса (для 4 классов).

Вывод: в данной лабораторной работе я реализовала алгоритм k ближайших соседей, а также научилась пользоваться методами библиотеки sklearn для тех же целей; затем визуализировала результаты. Метод knn – это алгоритм машинного обучения с учителем, который можно использовать для решения задач классификации. Он относит объекты к классу, которому принадлежит большинство из k его ближайших соседей в пространстве признаков.