

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



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.

	A	B	C
1	продукт, сладость, хруст, класс		
2	яблоко,7,7,0		
3	салат,2,5,1		
4	бекон,1,2,2		
5	банан,9,1,0		
6	орехи,1,5,2		
7	рыба,1,1,2		
8	сыр,1,1,2		
9	виноград,8,1,0		
10	морковь,2,8,1		
11	апельсин,6,1,0		
12	ананас,7,2,0		
13	гранат,6,3,0		
14	яйца,1,2,2		
15	капуста,2,9,1		
16	красная икра,1,3,2		
17	редис,3,10,1		
18	кабочок,2,6,1		
19	персик,7,4,0		
20	творог,2,2,2		
21	тыква,4,6,1		
22			

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 K=5, результатом программы было следующее:

```

D:\Documents\МММ\lab4\venv\Scripts\python.exe D:/Documents/МММ/lab4/main.py
Точность метрического классификатора, когда тестовая выборка составляет 40.0 % данных при k= 5 : 0.5

```

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

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

Точность метрического классификатора, когда тестовая выборка составляет 20.0 % данных при k= 3 : 1.0

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

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

Точность классификатора sklearn, когда тестовая выборка составляет 20.0 % данных при k= 3 : 1.0

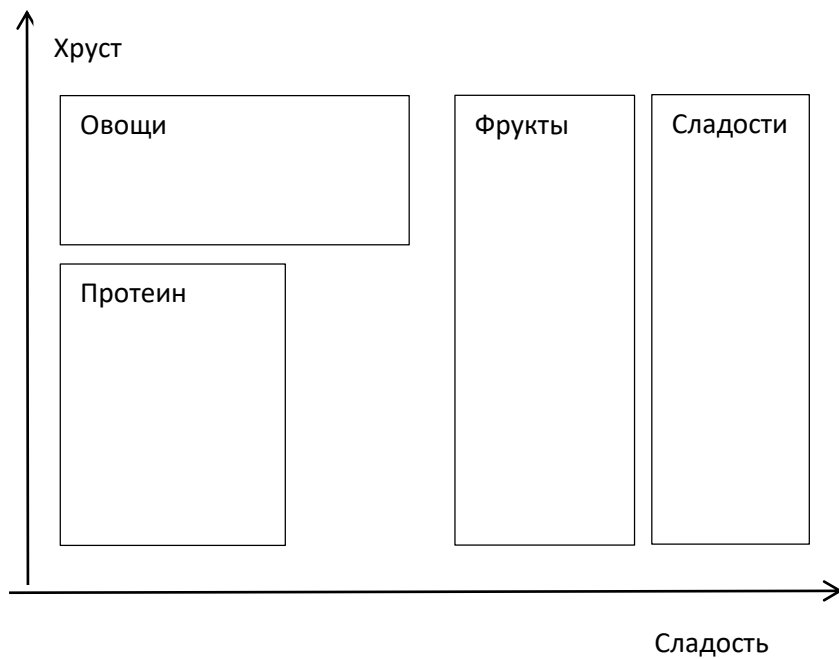
Рис.4. Результат классификатора sklearn при тестовой выборке в 20% и K=3

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.

	п	д	с
1	продукт,сладость,хруст,класс		
2	яблоко,7,7,0		
3	салат,2,5,1		
4	бекон,1,2,2		
5	пахлава,10,7,3		
6	банан,9,1,0		
7	орехи,1,5,2		
8	помадная конфета,10,1,3		
9	рыба,1,1,2		
10	сыр,1,1,2		
11	виноград,8,1,0		
12	чурчела,9,3,3		
13	капуста,2,9,1		
14	морковь,2,8,1		
15	козинаки,9,10,3		
16	апельсин,6,1,0		
17	вафли,8,9,3		
18	гранат,6,3,0		
19	яйца,1,2,2		
20	красная икра,1,3,2		
21	редис,3,10,1		
22	кабочок,2,6,1		
23	чакчак,10,8,3		
24	персик,7,4,0		
25	творог,2,2,2		
26	тыква,4,6,1		
27			

Рис.5. Файл dataset2.csv

Данный класс располагается по отношению к другим следующим образом:



5. Далее результаты работы классификаторов получилось визуализировать.

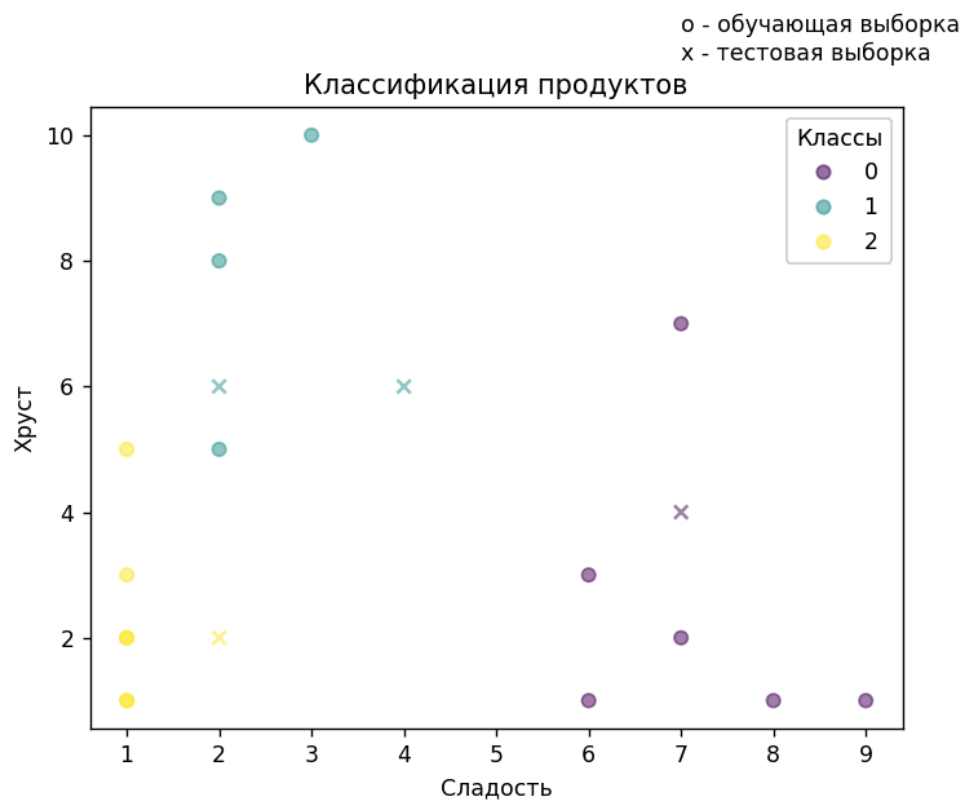


Рис.6. Визуализация работы метрического классификатора на наборе с 3 классами при тестовой выборке в 20% и  $K=3$

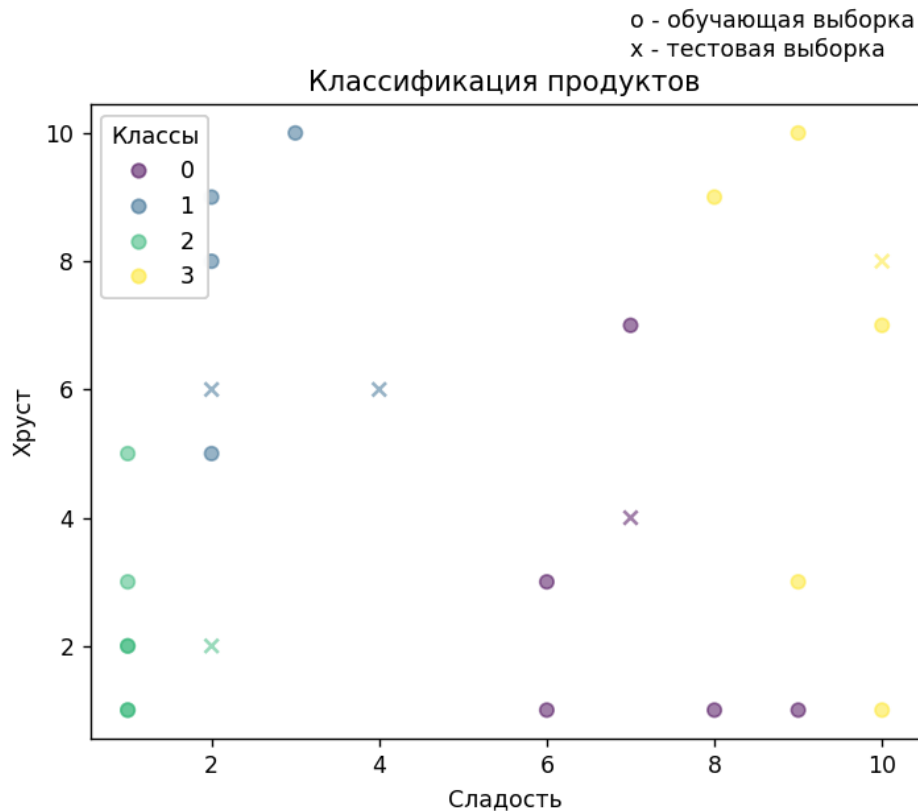


Рис.7. Визуализация работы классификатора sklearn на наборе с 4 классами при тестовой выборке в 20% и K=3

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.