Министерство образования Республики Беларусь

Учреждение Образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Кафедра ИИТ

Лабораторная работа № 3

По дисциплине «Статистические основы индуктивного вывода»

По теме «Кластерный анализ методом k-средних»

Выполнил: Ахроров М. Д.

Принял: Федосенко В. А.

Минск 2022

# Цель работы

Научиться импортировать данные, проводить предобработку данных, применять к данным метод кластерного анализа (k-средних) и интерпретировать полученные на основе выделения кластеров знания с точки зрения бизнес-контекста.

Формируемые знания, умения и навыки: изучить, как реализуtтся метод k-средних на языке Python, получить навыки выделения бизнес-контекста из результатов кластерного анализа

# Необходимо:

1. С сайта https://www.kaggle.com/ импортировать один из наборов данных.

2. Используя возможности языка Python, провести кластеризацию данных методом k-средних, выдвинуть гипотезы о бизнес-контексте результатов анализа.

# Бизнес-контекст

Опишите выбранный набор данных: контекст данных, какие переменные есть в наборе, какое количество кластеров было выбрано для данных, в чем отличия полученных кластеров друг от друга, как можно использовать в бизнесе полученное разбиение на кластеры.

# Ход работы

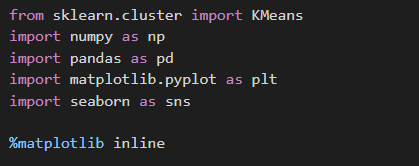


Рисунок 1. Подключение необходимых библиотек

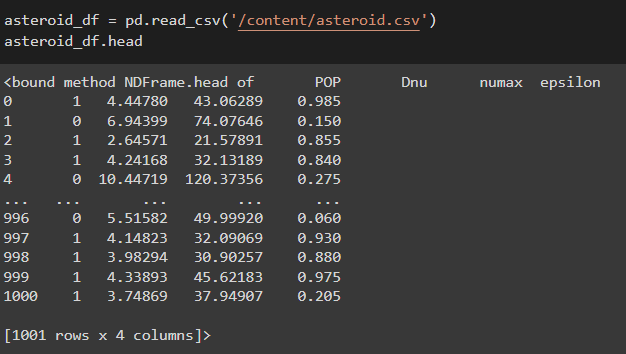


Рисунок 2. Импорт датасета



Рисунок 3. Описывание модель



Рисунок 4. Проведение моделирования

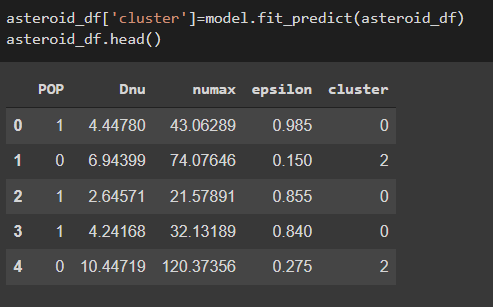


Рисунок 5. Добавление в исходный набор новый столбец с номером кластера наблюдения

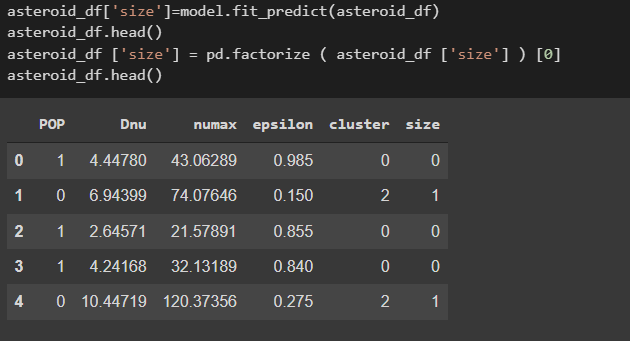


Рисунок 6. Добавление в исходный набор новый столбец, преобразовав переменную в фактор



Рисунок 7. Строение графика

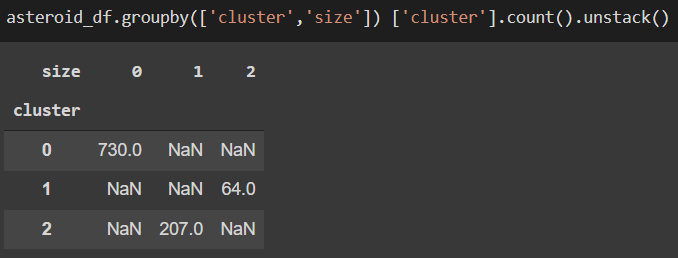


Рисунок 8. Составление сводной таблицы по данным класса и кластера

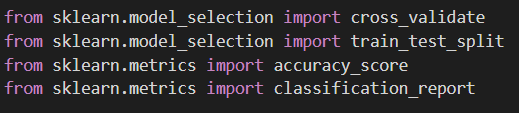


Рисунок 9. Подключение методов

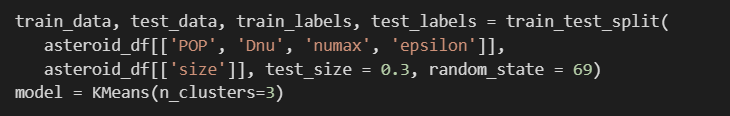


Рисунок 10. Разбиение исходного множества данных на тренировочную и тестовую выборки (в соотношении 7/3) и соответствующие им верные значения классов

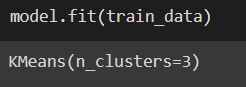


Рисунок 11. Проведение кластеризации на тренировочной выборке



Рисунок 12. Определение кластеров по построенной модели на тестовых данных

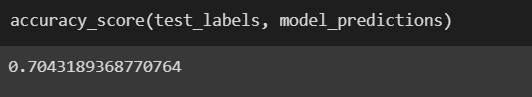


Рисунок 13. Сравнение с помощью метрики accuracy предсказанные значения с исходной классификацией тестовой выборки

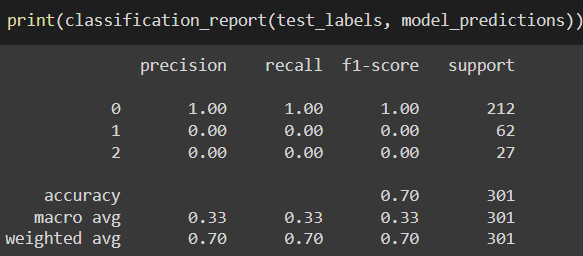


Рисунок 14. Отчет по качеству

# Выводы

В ходе данной работы были получены навыки импортирования данных, провождение предобработки данных, применение к данным метод кластерного анализа (k-средних) и интерпретирование полученных на основе выделения кластеров знания с точки зрения бизнес-контекста. Формировал знания, умения и навыки: изучил, как реализуtтся метод k-средних на языке Python, получил навыки выделения бизнес-контекста из результатов кластерного анализа.