**Наименование работы**:

Задача кластеризации и её метрики качества

**Цель работы:**

Ознакомиться с задачей кластеризации данных, методами её решения и оценки на практических примерах

**Задачи**:

1. В соответствие с вариантом выбрать набор данных для задачи кластеризации;
2. Обучить несколько различных моделей:
3. метод К-средних;
4. аггломеративная кластеризация;
5. Оценить каждую из обученных моделей по следующим метрикам:
6. Silhouette coefficient;
7. Davies Bouldin Index;
8. Интерпретировать результаты метрик качества, визуализировать кластеры.

Для реализации рекомендуется использовать язык программирования Python 3.x и библиотеки Pandas, Numpy, Sklearn, Matplotlib

**Оформление результатов**:

Результаты лабораторной работы оформляются в виде отчета в формате PDF.

**Структура отчета:**

* 1. Титульный лист;
  2. Основная часть;
  3. Заключение.

**В основной части** приводитсяописание выполнения каждой из поставленных задач в виде текста и скриншотов программного кода

**В заключении** приводятся практически значимыевыводы по проделанной работе

Таблица 1 - наборы данных

|  |  |  |
| --- | --- | --- |
| **№** | **Датасет** | **Описание** |
| 1 | Walmart Recruiting: Trip Type Classification  h [ttps://www.kaggle.com/c/walmart-recruiting-](https://www.kaggle.com/c/walmart-recruiting-trip-type-classification/data)t [rip-type-classification/data](https://www.kaggle.com/c/walmart-recruiting-trip-type-classification/data) | Данные о посещениях магазина Walmart покупателями.  Целевой признак: тип посещения магазина. |
| 2 | Home Credit Default Risk  h [ttps://www.kaggle.com/c/home-credit-defaul](https://www.kaggle.com/c/home-credit-default-risk/data)t [-risk/data](https://www.kaggle.com/c/home-credit-default-risk/data) | Данные о заемщиках банка.  Целевой признак: способен ли заёмщик выплатить кредит |