

Наименование работы:

Задача кластеризации и её метрики качества

Цель работы:

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

Задачи:

- a) В соответствии с вариантом выбрать набор данных для задачи кластеризации;
- b) Обучить несколько различных моделей:
 - 1. метод К-средних;
 - 2. аггломеративная кластеризация;
- c) Оценить каждую из обученных моделей по следующим метрикам:
 - 1. Silhouette coefficient;
 - 2. Davies Bouldin Index;
- d) 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 приводится описание выполнения каждой из поставленных задач в виде текста и скриншотов программного кода

В заключении приводятся практически значимые выводы по проделанной работе

Таблица 1 - наборы данных

№	Датасет	Описание
1	Walmart Recruiting: Trip Type Classification https://www.kaggle.com/c/walmart-recruiting-trip-type-classification/data	Данные о посещениях магазина Walmart покупателями. Целевой признак: тип посещения магазина.
2	IEEE-CIS Fraud Detection https://www.kaggle.com/c/ieee-fraud-detection/data	Данные об онлайн-транзакциях. Целевой признак: является ли транзакция мошеннической
3	Home Credit Default Risk https://www.kaggle.com/c/home-credit-default-risk/data	Данные о заемщиках банка. Целевой признак: способен ли заёмщик выплатить кредит