**Методы машинного обучения.**

Весна 2022 / 2023

Авторы:

Назаров Д.А.

Борисов Р.

**Практическая работа 1.**

Цель:

1. Знакомство и получение навыков работы в среде Jupyter,
2. Знакомоство с базовыми операцими над наборами данных с помощью Numpy и Pandas

Порядок выполнения работы:

1. Загрузить любой набор данных с [https://www.kaggle.com](https://www.kaggle.com/datasets/uciml/student-alcohol-consumption), который содержал бы различные типы свойств: числовые, номинальные, категориальные.

В качестве примеров можно использовать:

* <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption>
* <https://www.kaggle.com/datasets/greeshmagirish/crime-against-women-20012014-india>
* <https://www.kaggle.com/datasets/fbi-us/california-crime?select=ca_offenses_by_city.csv>
* <https://www.kaggle.com/datasets/danofer/law-school-admissions-bar-passage>
* <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>
* <https://www.kaggle.com/datasets/mylesoneill/world-university-rankings>
* https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database

2. В среде Jupyter выполнить анализ выбранного набора данных:

1. Уметь создавать свои наборы данных Numpy и Pandas
2. Уметь показать и объяснить теническую структуру набора данных
3. Уметь показать и объяснить содержимое набора данных:
   * 1. Являются ли переменные номинальными или представяют собой категории.
     2. Имеются ли в данных какие-то особенности в виде опечаток, ошибок ввода (по возможности)
     3. Имеются ли пустые значения полей
     4. Статистические характеристики числовых полей (мин, макс, среднее,медиана).
4. Выполнить срезы набора данных как по именам полей, так и по их индексам.
5. Выполнить ряд выборок с фильтрацией по значниям полей (равенство, неравенство, частичный фильтр строк с помощью лямбда-функции)
6. Выполнить сортировку отфильтрованного набора данных
7. Выполнить группировку данных по некоторым признакам (на основе предварительного анализа содержимого набора данных )
8. Выполнить объединение таблиц с различнымы типами присоединения (если нет возможности выполнить эту операцию с загруженным набором, то допускается создать собственные наборы)

3. Результат выполнения работы должен быть отражен в тетради (notebook) Jupyter с текстовыми вставками пояснений и комментариев