

Отчет по лабораторной работе №0 по курсу «Искусственный Интеллект»

Студент группы 8О-306 МАИ *Недосеков Иван*, №18 по списку

Контакты: `ivan-nedd@mail.ru`

Работа выполнена: 25 мая 2022 г.

Преподаватель: Самиров А.

Отчет сдан:

Итоговая оценка:

Подпись преподавателя:

1. Тема работы

Обработка данных

2. Цель работы

В данной лабораторной работе, вы выступаете в роли предприимчивого начинающего стартапера в области машинного обучения. Вы заинтересовались этим направлением и хотите предложить миру что-то новое и при этом неплохо заработать. От вас требуется определить задачу которую вы хотите решить и найти под нее соответствующие данные. Так как вы не очень богаты, вам предстоит руками проанализировать данные, визуализировать зависимости, построить новые признаки и сказать хватит ли вам этих данных, и если не хватит найти еще. Вы готовитесь представить отчет ваши партнерам и спонсорам, от которых зависит дальнейшая ваша судьба. Поэтому тщательно работайте:) И главное, день промедления и вас опередит ваш конкурент, да и спланированная работа отразится на репутации

3. Ход решения

Все началось с нахождения датасета на kaggle.

Идея выяснить курит ли человек по его медицинским показателям показалась мне интересной.

После скачивания и установки необходимых библиотек, приступил к обработке данных. Привел категориальные признаки к числовому значению, убрал столбцы не дающие информацию об цели изучения. Приступил к анализу данных, оказалось что в данных много шумов и есть сильно скоррелированные данные, все это предстояло вычистить из датасета. В итоге получился не очень сбалансированный в плане целевых признаков датасет.

4. Выводы

В данной лабораторной работе я научился обрабатывать и анализировать данные датасета. Библиотека pandas сильно в этом помогает, правда большинство конструкций не очевидно с ходу.