СЕССИЯ 2

СОДЕРЖАНИЕ

Сессия 2 данного Конкурсного задания состоит из следующей документации / файлов:

- Инструкция ко второй сессии
- Файл, содержащий целевые переменные
- Архив, содержащий информацию с датчиков по каждому пациенту
- Файл с описанием признаков в датасете
- Файл, содержащий предложенный разработчиками датасет

ВВЕДЕНИЕ

На этой сессии, вы будете продолжать исследования, опираясь на то, что вы уже разработали. **Если вы не выполнили задание в предыдущей сессии, не выполняйте его сейчас, воспользуйтесь предложенным разработчиками датасетом.** На этой сессии Вам необходимо изучить данные и выполнить их предобработку.

ИНСТРУКЦИЯ УЧАСТНИКУ

Убедитесь, что вы сохранили отчет о проделанной работе. К концу этой сессии у вас должны быть достигнуты следующие практические результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

2.1. Устранение дубликатов, пустых записей
Из исходных данных необходимо убрать пустые и дублирующие записи.
2.2. Обработка пропущенных значений, выбросов
Необходимо проанализировать количество пропущенных значений и выполнить их обработку, проанализировать значения, которыми заполнены признаки и обработать их при необходимости (выбросы, категориальные признаки и т.п.). Визуализируйте результат.
2.3. Обработка аномалий (аномальные объекты)
Необходимо проверить данные на наличие аномальных объектов, провести анализ и обработку аномалий.
2.4. Получение описательных статистик и графиков распределения всех признаков из итогового набора полей
Необходимо получить различные описательные статистики всех признаков из итогового набора полей. Необходимо построить графики распределения всех признаков из итогового набора полей. Дайте заключение по характеру распределения, сделайте предположение о характере зависимостей между признаками.
2.5. Подготовка отчета
Подготовьте отчет, содержащий результаты, полученные в пп. 2.1-2.4. Загрузите для проверки отчет, а также исходный программный код.