**СЕССИЯ 3**

**Исходные файлы:**

1. Data.zip (Набор данных)
2. Сессия 3.docx (Инструкция к 3 сессии)

**Результаты работы:**

1. Data.zip (Обработанные данные)
2. C[X]\_M3 (Папка с отчетом о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

**ЗАДАНИЕ**

|  |
| --- |
| 3.1 ОТБОР ПРИЗНАКОВ |
| Определите, какие признаки имеют наибольшее влияние на переменную readmitted, означающую будет ли пациент повторно принят в какую-либо больницу, и оставьте только их для обучения. Могут остаться и все исходные признаки. Необходимо обосновать выбор признаков, оставленных для обучения. |

|  |
| --- |
| 3.2 ВЫБОР И ПОСТРОЕНИЕ НОВЫХ ПЕРЕМЕННЫХ; ТЕСТИРОВАНИЕ ЗНАЧИМОСТИ ВЫБРАННЫХ ПЕРЕМЕННЫХ |
| Опишите необходимые шаги по обработке данных для получения новых переменных.  Напишите функцию для получения переменных.  Представьте и оцените значимость выбранных переменных. |

|  |
| --- |
| 3.3 АРГУМЕНТИРОВАННЫЙ ВЫБОР АЛГОРИТМА КЛАСТЕРИЗАЦИИ |
| Выберите алгоритм кластеризации на основе сравнения эффективности не менее трех вариантов на существующих данных с использованием метрики «Silhouette».  Опишите в текстовом виде логику выбора алгоритма. |

|  |
| --- |
| 3.4 ПОДГОТОВКА ОТЧЕТА |
| Предоставьте отчет о проделанной работе. Внимание: оценка модуля будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода и результат работы программы.  Отчет должен быть предоставлен в папке C[X]\_M3, где [Х] – номер участника, который должен быть размещен на рабочем столе ноутбука участника.  Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]\_M3.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]\_M3.HTML( или PDF), где [Х] – номер участника, экспортированная тетрадь (или аналог) в статический формат |