**СЕССИЯ 4**

**Исходные файлы:**

1. Data.zip (Набор данных)
2. Сессия 4.docx (Инструкция к 4 сессии)

**Результаты работы:**

1. Data.zip (Обработанные данные)
2. C[X]\_M4 (Папка с отчетом о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

**ЗАДАНИЕ**

|  |
| --- |
| 4.1 РАЗБИЕНИЕ ДАННЫХ НА ОБУЧАЮЩУЮ И ТЕСТИРУЮЩУЮ ВЫБОРКИ |
| Определите принцип разделения данных на обучающую и тестирующую выборки. Даже если вы выбираете случайное разделение данных, необходимо обосновать выбор данного принципа |

|  |
| --- |
| 4.2 ВИЗУАЛЬНОЕ ОБОСНОВАНИЕ ВЫБОРА ПЕРЕМЕННЫХ/ РАБОТОСПОСОБНОСТИ МОДЕЛИ |
| Представьте графики влияния переменных на целевую переменную.  Представьте график матрицы ошибок на обучающей выборке, отражено в отчете. |

|  |
| --- |
| 4.3 АРГУМЕНТИРОВАННЫЙ ВЫБОР МОДЕЛИ КЛАССИФИКАЦИИ |
| Выберите алгоритм классификации на основе сравнения эффективности не менее трех вариантов на существующих данных, отражено в отчете.  Представьте матрицу ошибок или ROC-кривую, определите точность моделей.  Опишите в текстовом виде логику выбора модели. |

|  |
| --- |
| 4.4 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]\_M4, 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]\_M4.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]\_M4.HTML( или PDF), где [Х] – номер участника, экспортированная тетрадь (или аналог) в статический формат |