Конкурсное задание

AtomSkills 2019

**Машинное обучение и большие данные**

**С2**

**Модуль 4 Разработка модели.**

## СОДЕРЖАНИЕ

Модуль 4 данного Конкурсного задания состоит из следующей документации / файлов:

1. С2\_4\_AS2019\_BDML (Бумаж. вариант) (Инструкция к четвертому модулю)
2. SOP1.csv, SOP2.csv, SOP3.csv, (Исходные данные)
3. sop1\_target.csv, sop2\_target.csv, sop3\_target.csv (Исходные данные целевой переменной)
4. Папка “Results” с результатами контроля (Публичная контрольная выборка)
5. Папка “Targets” с результатами расшифровки (Публичная контрольная выборка)

## ВВЕДЕНИЕ

## В этом модуле Вам необходимо разработать и обучить модель, произвести разбиение данных на обучающую и проверочную выборку, а также, сделать предсказание целевой переменной на основе предоставленной публичной контрольной выборки.

## ИНСТРУКЦИЯ участнику

К концу этой сессии, у вас должны быть достигнуты следующие результаты:

### ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

|  |
| --- |
| 1 РАЗБИЕНИЕ ДАННЫХ НА ОБУЧАЮЩУЮ И ПРОВЕРОЧНУЮ ВЫБОРКУ |
| 1. Подготовлены файлы обучающей и проверочной выборки  2. Представлен алгоритм по формированию этих выборок  3. Отражено в отчете обоснование подбора выборок |

|  |
| --- |
| 2 ВИЗУАЛЬНОЕ ОБОСНОВАНИЕ ВЫБОРА ПЕРЕМЕННЫХ/ РАБОТОСПОСОБНОСТИ МОДЕЛИ |
| 1. Представлены графики влияния переменных на целевую переменную  2. Представлена матрица ошибок или ROC-кривая  3. Выводы отражены в отчете |

|  |
| --- |
| 3 ПРОВЕРКА ТОЧНОСТИ НА ПУБЛИЧНОЙ КОНТРОЛЬНОЙ ВЫБОРКЕ |
| 1. Предоставлен набор данных в установленном формате: идентификатор строки, предсказание целевой переменной  2. Предоставлен и документирован код создающий требуемый набор данных  3. В отчете должна быть представлена оценка точности модели на обучающей выборке  4. Указано количество несплошностей (дефектов) и их параметры (см. “Таблицу дефектов” из “Описания ПУЗК\_С1\_М1.docx”) |

|  |
| --- |
| 4 ОТЧЕТ |
| 1. Предоставлен отчет о проделанной работе. Внимание: оценка сессии 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.  2. Отчёт должен быть предоставлен в папке C[X]\_M4, где [Х] – номер команды, которая должна быть размещена на рабочем столе. 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 команды, экспортированная тетрадь (или аналог) в статический формат. |