Конкурсное задание

AtomSkills 2019

**Машинное обучение и большие данные**

**С2**

**Модуль 3 Подбор переменных и алгоритмов для модели**

## СОДЕРЖАНИЕ

Модуль 3 данного Конкурсного задания состоит из следующей документации / файлов:

1. C2\_3\_AS2019\_BDML.pdf (Инструкция к модулю 3)
2. Видеоролик (Знакомство с системой ПУЗК)
3. SOP1.txt, SOP2.txt, SOP3.txt (Исходные данные)
4. target1.csv, target2.csv, target3.csv (Целевая переменная)
5. Описание ПУЗК\_С2\_M3.doc (Описание технологии полуавтоматизированного УЗК)

## ВВЕДЕНИЕ

## В этом модуле Вам предстоит выбрать переменные из исходного набора данных, при необходимости создать новые переменные и приступить к подбору алгоритма для определения целевой переменной.

## ИНСТРУКЦИЯ участнику

К концу этого модуля, у Вас должны быть достигнуты следующие результаты:

### ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

|  |
| --- |
| 1. ВЫБОР ПЕРЕМЕННЫХ ИЗ ИСХОДНОГО НАБОРА ДАННЫХ: ОБОСНОВАНИЕ (ЛОГИКА, АНАЛИЗ ДАННЫХ) |
| 1. Представлена (в текстовом или графическом виде) логика выбора переменных для построения модели, отражено в отчете.  2. Переменные обоснованы исходя из предметной области, природы данных, отражено в отчете. |

|  |
| --- |
| 2. ВЫБОР И ПОСТРОЕНИЕ НОВЫХ ПЕРЕМЕННЫХ; ТЕСТИРОВАНИЕ ЗНАЧИМОСТИ ВЫБРАННЫХ ПЕРЕМЕННЫХ |
| 1. Описаны необходимые шаги по обработке данных для получения переменных, отражено в отчете  2. Написана функция для получения переменных, отражено в отчете.  3. Представлена и оценена значимость выбранных переменных, отражено в отчете. |

|  |
| --- |
| 3. АРГУМЕНТИРОВАННЫЙ ВЫБОР АЛГОРИТМА ДЛЯ КЛАСТЕРИЗАЦИИ |
| 1. Алгоритм выбран на основе сравнения эффективности не менее двух вариантов на существующих данных, подтверждено графически и отражено в отчете.  2. Описана в текстовом виде логика выбора алгоритма, отражено в отчете. |

|  |
| --- |
| 4. РЕЗУЛЬТАТ КЛАСТЕРНОГО АНАЛИЗА |
| 1. Представлен файл формата .csv  2. Файл содержит вектор из 0 и 1 (нет\есть дефект) |

|  |
| --- |
| 5. 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. Отчет должен быть предоставлен в папке C[X]\_M3, где [Х] – номер участника, который должен быть размещен 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]\_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. |