

# КОНКУРСНОЕ ЗАДАНИЕ

***ДЛЯ МУНИЦИПАЛЬНОГО ЧЕМПИОНАТА***

***ЧЕМПИОНАТНОГО ЦИКЛА 2024 Г.***

# компетенции

**«МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ»**

**для возрастной категории 14-16 года**

*Конкурсное задание включает в себя следующие разделы:*

1. [Форма участия в конкурсе: 2](#_bookmark0)
2. [Общее время на выполнение задания: 2](#_bookmark1)
3. [Задание для конкурса 2](#_bookmark2)
4. [Модули задания и необходимое время 2](#_bookmark3)
5. [Критерии оценки. 5](#_bookmark4)
6. **Форма участия в конкурсе**: Индивидуальный конкурс

# Общее время на выполнение задания: 3 ч.

# Задание для конкурса

Общее описание:

В рамках конкурсного задания на основе большого набора данных, конкурсантам необходимо будет выполнить предобработку больших данных, представленных в «ненормализованном» виде, далее необходимо будет построить модель данных и провести машинное обучение на основе больших данных, произвести проверку модели данных на тестовой нормализованной выборке, показать (визуализировать) зависимости, разработать бота, который по указанным параметрам вычисляет результат по запросу пользователя, в качестве завершающего этапа конкурсного задания конкурсанту необходимо разработать документацию по выполненному решению.

Конкурсное задание является секретным и доступно для ознакомления экспертам в день проведения, участникам – в день проведения за 15 минут до начала модуля.

# Модули задания и необходимое время

Модули конкурсного задания выполняются последовательно. Время выполнения каждого модуля представлено в таблице 1.

Таблица 1. Время выполнения заданий каждого модуля

|  |  |  |  |
| --- | --- | --- | --- |
| **Наименование модуля** | | **Соревновательный день (С1)** | **Время на задание** |
| **A** | Парсинг и предобработка  данных | С1 | 1,5 часа |
| **B** | Разведочный анализ данных | С1 | 1,5 часа |

*Модуль A: Парсинг и предобработка данных.*

Данные представлены в одном из форматов: txt, pdf, doc, docx, xls, xslx, xml, csv. Необходимо преобразовать исходный набор данных к виду, пригодному для загрузки.

Примерные этапы задания в модуле A:

* 1. Предобработка данных и выделение значимых атрибутов;
  2. Разбиение сложных атрибутов;
  3. Очистка данных от служебной информации.

*Модуль B: Разведочный анализ данных.*

Необходимо провести предобработку данных, например, разбить сложные атрибуты, очистить от технической информации, удалить спецсимволы. В обработанном наборе данных выделяются значимые атрибуты для дальнейшего анализа и построения модели машинного обучения.

Примерные этапы задания в модуле B:

1. Дополнение недостающими данными;
2. Формирование словарей (мешков слов);
3. Предварительный анализ данных.

# Критерии оценки.

В данном разделе определены критерии оценки и количество начисляемых баллов (субъективные и объективные) таблица 2. Общее количество баллов задания/модуля по всем критериям оценки составляет 30.

Таблица 2. Критерии оценки и количество начисляемых баллов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Критерий** | | **Баллы** | | |
| **Судейские аспекты** | **Объективная оценка** | **Общая оценка** |
| **A** | Парсинг и предобработка данных | 0 | 10 | 10 |
| **B** | Разведочный анализ данных | 0 | 20 | 20 |
| **Итого** | | 0 | 30 | 30 |