



Национальный  
исследовательский

**Томский  
государственный  
университет**



# Конкурсное задание

## Компетенция

### «Машинное обучение и большие данные»

Конкурсное задание включает в себя следующие разделы:

1. Введение
2. Формы участия в конкурсе
3. Задание для конкурса. Модули задания и необходимое время
4. Критерии оценки

Количество часов на выполнение задания: 18 ч.

## **1. ВВЕДЕНИЕ**

1.1. Название и описание профессиональной компетенции.

1.1.1 Название профессиональной компетенции:

«Машинное обучение и большие данные».

1.1.2. Описание профессиональной компетенции.

Big data, или "большие данные" – сегодня повсюду. Сотни миллионов датчиков в смартфонах, на транспорте и со спутников днем и ночью собирают бесчисленные потоки данных. Это огромные массивы информации, из которых можно узнать много полезного.

И это уже возможно - благодаря новейшей профессии специалиста по большим данным. Ее обладатель должен быть отличным программистом и управленцем, разбираться в статистике.

Специалисты в области "больших данных" умеют задавать правильные вопросы, создавать алгоритмы для обработки информации и получать нужные выводы – это люди, которые знают больше других.

Количество источников "больших данных" постоянно растет, и поэтому специалисты в этой области становятся все более востребованными.

1.2. Область применения

1.2.1. Каждый Эксперт и Участник обязан ознакомиться с данным Конкурсным заданием.

1.3. Сопроводительная документация

1.3.1. Поскольку данное Конкурсное задание содержит лишь информацию, относящуюся к соответствующей профессиональной компетенции, его необходимо использовать совместно со следующими документами:

- «WorldSkills Russia», Техническое описание. Организация эффективного производства;
- «WorldSkills Russia», Правила проведения чемпионата
- Принимающая сторона – Правила техники безопасности и санитарные нормы.

## 2. ФОРМЫ УЧАСТИЯ В КОНКУРСЕ

Командный конкурс.

## 3. ЗАДАНИЕ ДЛЯ КОНКУРСА МОДУЛИ КОНКУРСНОГО ЗАДАНИЯ

Общее описание:

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

**Конкурсное задание является секретным и доступно для ознакомления экспертам в день С-2, участникам – в день С-1.**

Модули и время сведены в таблице 1.

Таблица 1.

| № п/п | Наименование модуля                               | Время на задание |
|-------|---|------------------|
| 1     | Модуль 1: Подготовка репозитория хранения данных. | 3 часа           |
| 2     | Модуль 2: Предобработка больших данных.           | 3 часа           |
| 3     | Модуль 3: Визуализация зависимостей               | 3 часа           |
| 4     | Модуль 4: Разработка математического аппарата     | 3 часа           |
| 5     | Модуль 5: Разработка прикладного решения          | 3 часа           |
| 6     | Модуль 6: Документирование                        | 3 часа           |
|       | Итого:  | 18 часов         |

#### 4. 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 100.

Таблица 2.

| Критерий     |   | Баллы     |             |             |
|--------------|---|-----------|-------------|-------------|
|              |   | Judge     | Object      | Всего       |
| <b>A</b>     | <b>Подготовка репозитория хранения данных</b> |           | <b>9</b>    | <b>9</b>    |
| <b>B</b>     | <b>Предобработка больших данных</b>           |           | <b>19,5</b> | <b>19,5</b> |
| <b>C</b>     | <b>Визуализация зависимостей</b>              | <b>4</b>  | <b>15</b>   | <b>19</b>   |
| <b>D</b>     | <b>Разработка математического аппарата</b>    | <b>4</b>  | <b>14</b>   | <b>18</b>   |
| <b>E</b>     | <b>Разработка прикладного решения</b>         | <b>4</b>  | <b>14,5</b> | <b>18,5</b> |
| <b>F</b>     | <b>Документирование</b>                       | <b>8</b>  | <b>8</b>    | <b>16</b>   |
| <b>Всего</b> |   | <b>20</b> | <b>80</b>   | <b>100</b>  |