|  |  |
| --- | --- |
|  |  |

КОНКУРСНОЕ ЗАДАНИЕ КОМПЕТЕНЦИИ

«НЕЙРОСЕТИ И БОЛЬШИЕ ДАННЫЕ»

Итогового (межрегионального) этапа чемпионата по профессиональному мастерству «Профессионалы» в 2024г

Конкурсное задание разработано экспертным сообществом и утверждено Менеджером компетенции, в котором установлены нижеследующие правила и необходимые требования владения профессиональными навыками для участия в соревнованиях по профессиональному мастерству.

**Конкурсное задание включает в себя следующие разделы:**

[1. ОСНОВНЫЕ ТРЕБОВАНИЯ КОМПЕТЕНЦИИ 4](#_Toc142037183)

[1.1. Общие сведения о требованиях компетенции 4](#_Toc142037184)

[1.2. Перечень профессиональных задач специалиста по компетенции «Нейросети и большие данные» 4](#_Toc142037185)

[1.3. Требования к схеме оценки 6](#_Toc142037186)

[1.4. Спецификация оценки компетенции 6](#_Toc142037187)

[1.5. Конкурсное задание 7](#_Toc142037188)

[1.5.1. Разработка/выбор конкурсного задания 7](#_Toc142037189)

[1.5.2. Структура модулей конкурсного задания (инвариант/вариатив) 7](#_Toc142037190)

[2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ 12](#_Toc142037191)

[2.1. Личный инструмент конкурсанта 13](#_Toc142037192)

[2.2.Материалы, оборудование и инструменты, запрещенные на площадке 13](#_Toc142037193)

[3. ПРИЛОЖЕНИЯ 14](#_Toc142037194)

**ИСПОЛЬЗУЕМЫЕ СОКРАЩЕНИЯ**

*1. API (Application Programming Interface или интерфейс программирования приложений) — это совокупность инструментов и функций в виде интерфейса для создания новых приложений, благодаря которому одна программа будет взаимодействовать с другой.*

1. ОСНОВНЫЕ ТРЕБОВАНИЯ КОМПЕТЕНЦИИ

1.1. ОБЩИЕ СВЕДЕНИЯ О ТРЕБОВАНИЯХ КОМПЕТЕНЦИИ

Требования компетенции «Нейросети и большие данные» определяют знания, умения, навыки и трудовые функции, которые лежат в основе наиболее актуальных требований работодателей отрасли.

Целью соревнований по компетенции является демонстрация лучших практик и высокого уровня выполнения работы по соответствующей рабочей специальности или профессии.

Требования компетенции являются руководством для подготовки конкурентоспособных, высококвалифицированных специалистов / рабочих и участия их в конкурсах профессионального мастерства.

В соревнованиях по компетенции проверка знаний, умений, навыков и трудовых функций осуществляется посредством оценки выполнения практической работы.

Требования компетенции разделены на четкие разделы с номерами и заголовками, каждому разделу назначен процент относительной важности, сумма которых составляет 100.

1.2. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ ЗАДАЧ СПЕЦИАЛИСТА ПО КОМПЕТЕНЦИИ «НЕЙРОСЕТИ И БОЛЬШИЕ ДАННЫЕ»

**Таблица №1 - Перечень профессиональных задач специалиста**

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 1 | **Разработка модулей программного обеспечения для компьютерных систем** | 42,5 |
| Специалист должен знать и понимать:  - основные этапы разработки программного обеспечения;  - основные принципы технологии структурного и объектно-ориентированного программирования;  - способы оптимизации и приемы рефакторинга;  - основные принципы отладки и тестирования программных продуктов. |
| Специалист должен уметь:  - осуществлять разработку кода программного модуля на языках низкого и высокого уровней;  - создавать программу по разработанному алгоритму как отдельный модуль; |

**Продолжение таблицы №1** *-* **Перечень профессиональных задач специалиста**

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 1 | - выполнять отладку и тестирование программы на уровне модуля;  - осуществлять разработку кода программного модуля на современных языках программирования;  - уметь выполнять оптимизацию и рефакторинг программного кода;  - оформлять документацию на программные средства. | 42,5 |
| 2 | **Разработка и отладка программного кода** | 24 |
| Специалист должен знать и понимать:  - методы и приемы формализации поставленных задач;  - методы и приемы алгоритмизации поставленных задач;  - синтаксис выбранного языка программирования, особенности программирования на этом языке, стандартные библиотеки языка программирования;  - методологии разработки компьютерного программного обеспечения  - методы повышения читаемости программного кода;  - основные стандарты оформления технической документации на компьютерное программное обеспечение;  - методы и приемы отладки программного кода. |
| Специалист должен уметь:  - применять алгоритмы решения типовых задач в области разработки;  - использовать методы и приемы формализации поставленных задач;  - использовать методы и приемы алгоритмизации поставленных задач;  - применять выбранные языки программирования для написания программного кода;  - использовать выбранную среду программирования;  - использовать возможности имеющейся технической и/или программной архитектуры для написания программного кода;  - применять нормативно-технические документы, определяющие требования к оформлению программного кода;  - выявлять ошибки в программном коде;  - применять методы и приемы отладки программного кода;  - интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов; |
| 3 | **Анализ данных и построение моделей машинного обучения** | 33,5 |
| Специалист должен знать и понимать:  - приемы анализа данных;  - принципы работы с графическими данными;  - принципы работы с текстовыми данными;  - принципы работы с аудио данными;  - принципы работы с видео данными;  - различные методы и алгоритмы машинного обучения;  - критерии качества моделей машинного обучения;  - последовательность разработки моделей машинного обучения;  - какие методы машинного обучения применять в зависимости от исходных данных;  - как работать с различными выборками данных;  - как использовать различные программные средства для разработки и улучшения моделей. |

**Окончание таблицы №1 - Перечень профессиональных задач специалиста**

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 3 | Специалист должен уметь:  - структурировать данные;  - проводить нормализацию и подготовку данных;  - выделять признаки, свойства и характеристики объектов в данных;  - осуществлять операции с большими данными;  - проводить визуальный анализ данных;  - применять классические алгоритмы машинного обучения:   * обучение без учителя (уменьшение размерности, поиск правил, кластеризация); * обучение с учителем (регрессия, классификация);   - применять методы глубокого обучения и Нейросети:   * Перцептроны; * Сверточные нейросети; * Реккурентные сети. | 33,5 |

1.3. ТРЕБОВАНИЯ К СХЕМЕ ОЦЕНКИ

Сумма баллов, присуждаемых по каждому аспекту, должна попадать в диапазон баллов, определенных для каждого раздела компетенции, обозначенных в требованиях и указанных в таблице №2.

**Таблица №2** *-* **Матрица пересчета требований компетенции в критерии оценки**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Критерий/Модуль** | | | | | | **Итого баллов за раздел ТРЕБОВАНИЙ КОМПЕТЕНЦИИ** |
| **Разделы ТРЕБОВАНИЙ КОМПЕТЕНЦИИ** |  | **A** | **Б** | **В** | **Г** |  |
| **1** | 14 | 17,5 | 11 | 0 | 42,5 |
| **2** | 14 | 1 | 1 | 8 | 24 |
| **3** | 0 | 33,5 | 0 | 0 | 33,5 |
| **Итого баллов за критерий/модуль** | | 28 | 52 | 12 | 8 | **100** |

1.4. СПЕЦИФИКАЦИЯ ОЦЕНКИ КОМПЕТЕНЦИИ

Оценка Конкурсного задания будет основываться на критериях, указанных в таблице №3:

**Таблица №3 -****Оценка конкурсного задания**

|  |  |  |
| --- | --- | --- |
| **Критерий** | | **Методика проверки навыков в критерии** |
| **А** | **Анализ и предобработка данных** | Экспертам необходимо проверить проведенный анализ и подготовку данных участника, после чего происходит объективная оценка по критериям.  Проверка осуществляется на рабочих местах конкурсантов. |
| **Б** | **Разработка модели машинного обучения** | Экспертам необходимо проанализировать исходный код нейросети, обучающий набор данных и документацию, выполненные конкурсантам и произвести объективные и субъективные оценки по критериям.  Проверка осуществляется на рабочих местах конкурсантов, также отдельные аспекты проверяются на компьютере группы оценки. |
| **В** | **Тестирование разработанной модели** | Экспертам необходимо провести тестирование работ участников на рабочих местах, путем воспроизведения подготовленных конкурсантами тестов, оценивание производится по объективным критериям. |
| **Г** | **Презентация решения** | Эксперты выслушивают презентации, оценивают содержание и выступление конкурсантов с помощью объективных и субъективных критериев. |

1.5. КОНКУРСНОЕ ЗАДАНИЕ

Общая продолжительность Конкурсного задания: 16 ч.

Количество конкурсных дней: 3 дней

Вне зависимости от количества модулей, КЗ должно включать оценку по каждому из разделов требований компетенции.

Оценка знаний участника должна проводиться через практическое выполнение Конкурсного задания. В дополнение могут учитываться требования работодателей для проверки теоретических знаний / оценки квалификации.

1.5.1. Разработка/выбор конкурсного задания

Конкурсное задание состоит из 4 модулей, включает обязательную к выполнению часть (инвариант) – 3 модуля: анализ и предобработка данных, разработка модели машинного обучения, тестирование разработанной модели; и вариативная часть – 1 модуль: презентация решения. Общее количество баллов конкурсного задания составляет 100.

Обязательная к выполнению часть (инвариант) выполняется всеми регионами без исключения на всех уровнях чемпионатов.

Количество модулей из вариативной части, выбирается регионом самостоятельно в зависимости от материальных возможностей площадки соревнований и потребностей работодателей региона в соответствующих специалистах. В случае если ни один из модулей вариативной части не подходит под запрос работодателя конкретного региона, то вариативный модуль формируется регионом самостоятельно под запрос работодателя. При этом время на выполнение модуля и количество баллов в критериях оценки по аспектам не меняются.

1.5.2. Структура модулей конкурсного задания (инвариант/вариатив)

**Введение**

В настоящий момент множество полезной информации достаточно легко найти в сети Интернет. Различные статьи, заметки, научные работы – все это находится в свободном доступе, публикуется на всевозможных ресурсах, журналах и форумах. Однако, у такого изобилия существуют и недостатки – никакого количества времени не хватит, чтобы полностью изучить всю информацию, по определенной тематике, размещенную в сети Интернет. Однако, благодаря аннотациям в начале статей можно фильтровать наиболее важные и полезные материалы. К несчастью, не во всех произведениях встречаются краткие выжимки в начале. К счастью, данная проблема решается с помощью нейронных сетей.

Помимо отсутствия аннотаций, также довольно велико количество схожих между собой по содержанию статей, что в свою очередь увеличивает время получения полезной информации из различных источников.

Таким образом, в рамках данного чемпионата, участникам необходимо разработать нейронные сети, выполняющие функцию составления аннотации загружаемой в нее научной статьи и функцию сравнения схожести научных статей между собой на основании показателей нейронной сети. Помимо разработки нейронной сети, необходимо разработать простое графическое приложение, которое позволит удобно взаимодействовать с нейронными сетями и визуализировать результаты работы второй сети.

В процессе работы участники должны вести сопроводительную документацию, а также отчеты о проделанной работе в рамках модуля.

После разработки нейронной сети, участникам протестировать точность и работоспособность их программного продукта.

На финальном этапе конкурсанты должны подготовить презентацию, включающую основные этапы их работы, результаты тестирования, а также документацию на разработанный ими программный продукт.

**Модуль А. Анализ и предобработка данных (инвариант)**

**Время на выполнения модуля:**6 часов

**Краткое описание задания:** в рамках данного модуля участникам необходимо ознакомиться с представленным набором данных, проанализировать его структуру, содержимое и подготовить данные для обучения нейросети.

**Задание:**

Конкурсантам предоставляется набор данных, содержащий научные статьи, собранные в папки с файлами text, annotation и tаgs. Предоставленный набор данных необходимо отфильтровать по определенным параметрам:

* заспамленность;
* кол-во символов аннотации со спецсимволами;
* кол-во символов аннотации без спецсимволов;
* кол-во символов текста со спецсимволами;
* кол-во символов текста без спецсимволов.

Значения этих параметров необходимо определить самостоятельно проанализировав разметку к файлу. Для получения значения параметров и фильтрации статей по ним допустимо использовать любые применимые алгоритмы.

По результатам выполнения модуля конкурсанты должны предоставить отфильтрованный по параметрам набор данных, отчет о работе, включающий описание проделанной работы и описание итогового набора данных, также, при наличии, необходимо предоставить программный алгоритм, с помощью которого проводилась очистка и включить его описание в отчет о работе.

**Модуль Б. Разработка модели машинного обучения (инвариант)**

**Время на выполнения модуля:**7 часов

**Краткое описание задания:** в рамках данного модуля необходимо разработать модель, которая будет решать поставленную задачу, а также интерфейсы взаимодействия с ней.

**Задание:**

Конкурсантам необходимо подготовить нейросетевую модель, которая будет составлять аннотацию загружаемой в нее научной статьи.

Также конкурсантам необходимо реализовать отдельную нейронную сеть, которая будет осуществлять сравнение схожести научных статей между собой, и на основании показателей нейронной сети, формировать графическое представление схожести с остальными.

Также, конкурсантам необходимо разработать API для взаимодействия с разработанными нейросетевыми моделями, затем разработать графическое приложение для работы пользователя и демонстрации графического представления схожести статей.

По результатам работы конкурсанты должны предоставить отчет о проделанной работе, включающий в себя обоснование выбранных архитектур нейронных сетей, исходный код нейронных сетей и итоговые обученные модели, приложение и его исходный код, руководство пользователя приложения, включающее инструкции по развертке, API и его документацию.

**Модуль В. Тестирование разработанной модели (инвариант)**

**Время на выполнения модуля:**2 часа

**Краткое описание задания:** в рамках данного модуля необходимо покрыть разработанный программный продукт тестированием

**Задание:**

Конкурсантам необходимо реализовать последовательный и автоматизированный способ тестирования приложений, наполнить его тест-кейсами, которые будут проверять основной функционал программы и точность работы нейронной сети. В рамках создания системы автоматизированного тестирования участникам необходимо настроить тестовый пайплайн.

По результатам выполнения модуля, конкурсанты должны предоставить набор автоматизированных тестов, тест-кейсы с результатами модульного тестирования и отчет о проведенном тестировании.

**Модуль Г. Презентация решения (вариатив)**

**Время на выполнения модуля:**1 час

**Краткое описание задания:** в рамках данного модуля участникам необходимо подготовить презентацию своего решения, в которой необходимо отразить результаты тестирования, обоснование выбора алгоритмов, а также продемонстрировать работу решения.

**Задание:**

В рамках данного модуля конкурсантам необходимо подготовить презентацию реализованного проекта. Необходимо отразить следующие пункты: информация о работе с данными, которые использовались для обучения нейронных сетей; какие алгоритмы и какой тип нейронных сетей был выбран при разработке; документацию к разработанному API; результаты тестирования нейронных сетей. Далее презентацию необходимо защитить перед экспертами.

2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИ

В ходе проведения чемпионата есть вероятность возникновения спорных ситуаций, дающих преимущества некоторым участникам над другими. В таблице 4 представлены наиболее частые проблемы, а также алгоритм действия, при выявлении подобных ситуаций:

**Таблица №4 -****Решение спорных ситуаций**

|  |  |
| --- | --- |
| **Вопрос** | **Специальные правила** |
| Сохранение решения конкурсантами | Решение должно быть сохранено по указанному в Конкурсном задании пути, в случае нарушения инструкций решение не проверяется.  В случае технической невозможности выполнить требование выше – необходимо обратиться к ТАП и Главному эксперту, после чего получить указания по обновленному пути сохранения |
| Временной регламент | В случае выполнения участником модуля Конкурсного задания в непредназначенное для этого время, в том числе во время выполнения другого модуля, такое решение не будет оценено. |
| Программное обеспечение | Конкурсанты могут создавать программные продукты, оформлять инструкции или делать заметки, находясь на рабочей площадке, однако их никогда нельзя забирать с рабочей площадки. За использование материалов, файлов, подготовленных вне конкурсного времени и за пределами конкурсной площадки, в том числе шпаргалок, материалов, полученных в сети Интернет (если иное не указано в Конкурсном задании), субкритерии, к котором относится выполненная заранее заготовка, не оцениваются. |
| Отказ оборудования | Если имеется явное доказательство того, что конкурсанты сами причинили ущерб оборудованию, им не будет предоставляться замена и дополнительное время. |
| Поведение конкурсантов | Общение участников во время выполнения Конкурсного задания запрещено, в случае неоднократного нарушения запрета участники дисквалифицируются на оставшееся время конкурсного дня. |
| Использование интернета | Конкурсанты имеют право использовать интернет на своих рабочих местах с целью дозагрузки понадобившихся библиотек и фреймворков, дата выпуска и верификации которых произведена не позднее чем за 1 месяц до чемпионата Конкурсантам запрещается использовать интернет для действий, описанных в пункте «Программное обеспечение», для использования социальных сетей и иных сервисов с авторизацией, на которых могут быть размещены заранее подготовленные материалы для решения конкурсного задания |

На площадке реализованы дополнительные правила:

1) на площадке необходимо обеспечена видеофиксация мониторов конкурсантов (видеозахват рабочих столов на обоих мониторах), фиксация действий пользователя за клавиатурой (кейлоги) и контроль сетевого трафика (проксирование) с выгрузкой видеозаписей и данных на выделенный сервер;

2) на площадке производится выгрузка работ конкурсантов на флеш-накопитель техническим экспертом;

3) любая фото-, видеосъемка СМИ допускается только после согласования с Главным экспертом и Индустриальным экспертом.

2.1. Личный инструмент конкурсанта

Список оборудования неопределенный:

Конкурсант может привезти индивидуальное периферийное оборудование по списку: мышь, клавиатура, проводная гарнитура.

Ко всей периферии применяется следующее требование: оборудование не должно иметь возможности беспроводного подключения, а также заранее программируемых команд (макросов).

2.2.Материалы, оборудование и инструменты, запрещенные на площадке

Для участников соревнований: телефоны и иные средства связи, ноутбуки, смарт-часы, Bluetooth-гарнитуры, средства фото-, аудио-, видеозаписи, средства электронного переноса информации (USB-накопители).

Для экспертного сообщества: запрещено использование телефонов или иных средств связи, ноутбуков, Bluetooth-гарнитур, средств фото-, аудио-, видеозаписи во время формирования итоговой версии конкурсного задания и критериев оценивания (во время внесения 30% изменений) и во время оценивания работ конкурсантов. В остальное время ограничений нет.

3. Приложения

Приложение №1 Инструкция по заполнению матрицы конкурсного задания

Приложение №2 Матрица конкурсного задания

Приложение №3 Инструкция по охране труда и технике безопасности по компетенции «Нейросети и большие данные»

Приложение №4 Инструкция по проведению оценки работ конкурсантов по компетенции «Нейросети и большие данные»

Приложение №5 Свод отраслевых (индустриальных) стандартов по компетенции «Нейросети и большие данные»