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

КОНКУРСНОЕ ЗАДАНИЕ КОМПЕТЕНЦИИ

«СПЕЦИАЛИСТ ПО АНАЛИЗУ ДАННЫХ (BI-АНАЛИТИК)»

**Финал чемпионата высоких технологий 2024**

2024 г.

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](#_Toc172549351)

[1.1. ОБЩИЕ СВЕДЕНИЯ О ТРЕБОВАНИЯХ КОМПЕТЕНЦИИ 4](#_Toc172549352)

[1.2. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ ЗАДАЧ СПЕЦИАЛИСТА ПО КОМПЕТЕНЦИИ «СПЕЦИАЛИСТ ПО АНАЛИЗУ ДАННЫХ (BI-АНАЛИТИК)» 4](#_Toc172549353)

[1.3. ТРЕБОВАНИЯ К СХЕМЕ ОЦЕНКИ 12](#_Toc172549354)

[1.4. СПЕЦИФИКАЦИЯ ОЦЕНКИ КОМПЕТЕНЦИИ 13](#_Toc172549355)

[1.5. КОНКУРСНОЕ ЗАДАНИЕ 15](#_Toc172549356)

[1.5.1. Разработка/выбор конкурсного задания 15](#_Toc172549357)

[1.5.2. Структура модулей конкурсного задания 15](#_Toc172549358)

[2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ 25](#_Toc172549359)

[2.1. Личный инструмент конкурсанта 26](#_Toc172549360)

[2.2.Материалы, оборудование и инструменты, запрещенные на площадке 27](#_Toc172549361)

[2.3 Индустриальные стандарты 27](#_Toc172549362)

[3. Приложения 27](#_Toc172549363)

**ИСПОЛЬЗУЕМЫЕ СОКРАЩЕНИЯ**

1. ФГОС – Федеральный государственный образовательный стандарт
2. ПС – Профессиональный стандарт
3. КЗ – Конкурсное задание
4. ИЛ – Инфраструктурный лист
5. ТК – Требования компетенции
6. ПО – Программное обеспечение
7. ООО – Объектно-ориентированное программирование
8. ПП – Программный продукт
9. ПМ – Программный модуль
10. ЯП – Языки программирования
11. ИАС – Информационно-аналитическая система

1. ОСНОВНЫЕ ТРЕБОВАНИЯ КОМПЕТЕНЦИИ

1.1. ОБЩИЕ СВЕДЕНИЯ О ТРЕБОВАНИЯХ КОМПЕТЕНЦИИ

Требования компетенции (ТК) «Специалист по анализу данных (BI-аналитик)» 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 «СПЕЦИАЛИСТ ПО АНАЛИЗУ ДАННЫХ (BI-АНАЛИТИК)»

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

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 1 | **Разработка модулей ПО для компьютерных систем** | 10 |
| Специалист должен знать и понимать:- основные этапы разработки ПО;- основные принципы технологии структурного и ООП;- способы оптимизации и приемы рефакторинга;- основные принципы отладки и тестирования ПП. |
| Специалист должен уметь:- осуществлять разработку кода ПМ на языках низкого и высокого уровней;- создавать программу по разработанному алгоритму как отдельный модуль; - выполнять отладку и тестирование программы на уровне модуля;- осуществлять разработку кода ПМ на современных языках программирования;- уметь выполнять оптимизацию и рефакторинг программного кода;- оформлять документацию на программные средства. |

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 2 | **Разработка и отладка программного кода** | 20 |
| Специалист должен знать и понимать:- методы и приемы формализации поставленных задач;- методы и приемы алгоритмизации поставленных задач;- синтаксис выбранного ЯП, особенности программирования на этом языке, стандартные библиотеки языка программирования;- методологии разработки компьютерного ПО- методы повышения читаемости программного кода;- основные стандарты оформления технической документации на компьютерное ПО;- 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 ЯП 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; |

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 3 | **Управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации** | 25 |
| Специалист должен знать и понимать:Теория вероятностей и математическая статистикаСовременные методы и инструментальные средства анализа больших данныхМетоды трансляции больших данныхСтандарты проведения анализа данныхМетоды оценки временных и стоимостных характеристик технологий больших данныхСовременная технологическая инфраструктура высокопроизводительных и распределенных вычисленийМетоды интерпретации и визуализации больших данныхСпециалист должен уметь:Проводить презентации при консультировании заказчика, согласовании и утверждении требований к результатам аналитических работ с использованием технологий больших данныхПодготавливать документы, регламентирующие требования к результатам аналитического исследования с использованием технологий больших данных в соответствии с существующими регламентами организации |

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 4 | **Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры** | 35 |
| Специалист должен знать и понимать:Предметная область анализа больших данных в соответствии с требованиями заказчикаВозможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данныхСовременный опыт использования анализа больших данныхТеоретическая и прикладная информатикаТеоретические и прикладные основы анализа данныхОсновы бизнес-интеллекта, типы систем бизнес-интеллектаТеория принятия решенийМатематическое моделированиеТипы анализа больших данных, виды аналитики |
| Специалист должен уметь:Использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работПроводить сравнительный анализ методов и инструментальных средств анализа больших данныхПроводить анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования |

|  |  |  |
| --- | --- | --- |
| **№** | **Раздел** | **Важность в %** |
| 5 | **Решение задач автоматизации информационно-аналитической деятельности с использованием информационно-аналитических систем в защищенном исполнении** | 5 |
| Специалист должен знать и понимать:Типовые средства и методы защиты информации в локальных и глобальных вычислительных сетяхБазовая конфигурация системы защиты информации ИАСОсобенности применения программных и программно-аппаратных средств защиты информации в ИАС |
| Специалист должен уметь:Конфигурировать параметры системы защиты информации ИАС в соответствии с ее эксплуатационной документациейОбнаруживать и устранять неисправности системы защиты информации ИАС согласно эксплуатационной документации |
| 6 | **Проектирование информационно-аналитических систем в защищенном исполнении** | 5 |
| Специалист должен знать и понимать:Типовые средства, методы и протоколы идентификации, аутентификации и авторизацииНормативные правовые акты Российской Федерации в области защиты информацииОрганизационные меры по защите информации |
| Специалист должен уметь:Производить монтаж и диагностику компонентов ИАСИспользовать типовые криптографические средства защиты информации, в том числе средства электронной подписи |

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** | **Д** |  |
| **1** | **2** | **2** | **0** | **4** | **2** | **10** |
| **2** | **2** | **6** | **6** | **3** | **3** | **20** |
| **3** | **6,5** | **3** | **3,5** | **5** | **7** | **25** |
| **4** | **8,5** | **11** | **8,5** | **0** | **7** | **35** |
| **5** | **0** | **0** | **0** | **5** | **0** | **5** |
| **6** | **0** | **0** | **0** | **5** | **0** | **5** |
| **Итого баллов за критерий/модуль** | **19** | **22** | **18** | **22** | **19** | **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** | **Предобработка количественных данных** |  Критерии:1. Выбор источников данных2. Проверка типов данных3. Оценка объема данных4. Оценка погрешности |
| **Б** | **Анализ количественных данных** |  Критерии:1. Точность расчетов описательных статистик и корреляций2. Корректность построения графиков и диаграмм3. Корректность работы дашборда (отсутствие ошибок при отображении информации) |
| **В** | **Обработка и анализ текстовых данных** |  Критерии:1. Содержательная оценка точности распределения навыков по классам2. Оценка доли неклассифицируемых навыков3. Оценка показателей полноты и точности базы |
| **Г** | **Жизненный цикл сбора данных и безопасность** | Критерии:1. Оценка анализа требований.2. Оценка разработки скриптов автоматического сбора данных.3. Оценка защиты данных и мониторинга. |
| **Д** | **Сетевой анализ данных**  |  Критерии:1. Оценка правильности построения графа2. Оценка удобства прочтения графа (выбор укладки)3. Оценка описания метрик графа4. Оценка предложенных рекомендаций |

1.5. КОНКУРСНОЕ ЗАДАНИЕ

Общая продолжительность Конкурсного задания: 15 ч.

Количество конкурсных дней: 3 дня

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 5 модулей, включает обязательную к выполнению часть (инвариант) – 4 модуля: предобработка количественных данных, анализ количественных данных, обработка и анализ текстовых данных, сетевой анализ; и вариативная часть – 1 модуль: жизненный цикл и безопасность. Общее количество баллов конкурсного задания составляет 100.

1.5.2. Структура модулей конкурсного задания

**Введение**

В рамках чемпионата конкурсанты решают задачу аналитики данных результатов ВСОШ для оценки общего уровня подготовленности школьников. Конкурсанты выполняют статистические исследования и определяют тенденцию развития олимпиадного движения в регионе. Эти данные могут быть использованы при разработке или обновлении образовательных программ.

**Модуль А. Предобработка количественных данных (инвариант)**

**Время на выполнения модуля:**2 часа

**Краткое описание задания:** написать запросы на выборку данных из базы данных либо работать с предоставленными данными (csv файл, содержащий следующие поля: id района, год, предмет, этап ВСОШ и результат). Необходимо произвести очистку от ошибок и поиск аномальных значений, выполнить работу с пропущенными значениями, преобразовать типы данных (при необходимости).

**Задание:** необходимо подготовить для обработки данные результатов ВСОШ по районам в Новгородской области.

1. **Сбор данных**: для сбора данных нужно использовать:
* предоставленный набор данных в формате csv, содержащий информацию о количестве участий разных субъектов области в различных этапах ВСОШ;
* произвести парсинг данных на указанном сервисе, содержащий аналогичные данные;
* объединение данных, полученных из разных источников.
1. **Предобработка данных**:
* очистка данных от ошибок и аномалий;
* обработка пропущенных значений;
* преобразование типов данных;
* оценка качества данных на полноту, точность, согласованность.
1. **Описание структуры набора данных:**
* определить и обосновать значимые атрибуты;
* обосновать выбор дополнительных атрибутов и причину исключения каких-либо данных из исходного набора;
* для каждого атрибута присутствует информация о количестве пустых значений;
* выполнить текстовое описание атрибутов.

**Результат модуля**:

1. Файл в формате csv или xlsx с готовой к обработке базой данных.
2. Отчёт, содержащий подробную информацию о выполненной работе.

**Модуль Б. Анализ количественных данных (инвариант)**

**Время на выполнения модуля:**4 часа

**Краткое описание задания:** необходимо провести анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, формулирование выводов, в том числе с использованием инструментов визуализации (построение дашборда).

**Задание:**

1. На этапе выполнения модуля А должны быть выбраны ключевые атрибуты.
2. **Анализ основных свойств данных:** расчёт основных статистических параметров выбранных характеристик с анализом результатов, построение графиков распределения для выбранных характеристик.
3. **Поиск зависимостей** одних показателей от других.
4. **Поиск закономерностей:** выявление закономерностей с анализом результатов.
5. **Определить тенденцию развития олимпиадного движения в регионе**. Проверить точность прогноза.
6. **Проверка нормальности распределения**: разделить данные на группы по районам, построить гистограммы результатов по каждому району.
7. **Выполнить математический анализ данных:** провести сравнительный анализ между результатами в двух различных районах с помощью F-теста или критерия Фишера/Стьюдента.
8. **Формулирование выводов и построение отчёта** написать аналитическую записку с основными показателями, графиками и диаграммами, и выводами.
9. **Визуализация данных:** построить интерактивный дашборд с требуемыми параметрами. Фильтрация данных по заданным параметрам. Визуализация должна наглядно отображать рост или снижение показателей.

**Результат модуля**:

1. Интерактивный дашборд с требуемыми параметрами.
2. Аналитическая записка с основными показателями, графиками и диаграммами, и выводами.

**Модуль В. Обработка и анализ текстовых данных (инвариант)**

**Время на выполнения модуля:**3 часа

**Краткое описание задания:**

Необходимо выполнить парсинг данных для сбора информации с указанных веб-ресурсов. Сформировать структуру набора данных. Провести предварительную обработку данных. Выполнить построение и отбор признаков. Выполнить семантический анализ текста. Провести визуальный анализ статистики текстовых признаков. Сделать выводы, в том числе с использованием инструментов визуализации.

**Задание:**

1. **Парсинг данных:** выполнить парсинг данных для сбора информации об участии школьников в олимпиаде по литературе: текст сочинения, оценка в баллах.
2. **Формирование структуры набора данных:** определение необходимых полей и создайте структуры набора данных.
3. **Предварительная обработка данных:** очистка данных и предварительная обработка (удаление стоп-слов, перевод в нижний регистр, лемматизация, стемминг, удаление аномалий и т.д.).
4. **Отбор признаков**: определение того, какие слова и лексические паттерны чаще всего применяются в текстах, получающих большее количество баллов и напротив определить ключевые моменты, которые характеризуют более слабые сочинения.
5. **Формулирование выводов**.

**Результат модуля**:

1. Таблица в формате xlsx или csv с необходимыми данными с веб ресурса.
2. Таблица с выявленными признаками и с частотой встречаемости частотой встречаемости.
3. Аналитическая записка в текстовом формате со всех проделанной работой.

**Модуль Г. Жизненный цикл сбора данных и безопасность (вариантив)**

**Время на выполнения модуля:** 3 часа

Так как информация о результатах участия субъектов регионов в ВСОШ является конфиденциальной, а данные появляются неравномерно, вам необходимо реализовать скрипт, который позволит сохранять и агрегировать на промежуточном сервере эти данные в защищенном виде.

**Краткое описание задания:** необходимо реализовать систему, обеспечивающую непрерывную выгрузку и агрегацию данных, создать прослойку для промежуточного хранения данных. На выгрузку вам даются данные и ключ, который также имеется на флеш-накопителе, выданном вам. Необходимо сделать систему для расшифрования данных с помощью этого ключа.

**Задание:**

1. **Анализ требований**: проведите подробный анализ требований, опишите функциональные требования к системе.
2. **Разработка скриптов сбора информации**: создайте скрипты для автоматического сбора и обработки данных из различных источников, что поможет оптимизировать процесс получения информации. Необходимо реализовать систему, обеспечивающую непрерывную выгрузку и агрегацию данных.
3. **Обеспечение защиты данных**: предусмотрены меры по обеспечению защиты данных от несанкционированного доступа. Реализованы методы защиты данных, представленных системах хранения данных. Предусмотрена возможность настройки доступа к информации в соответствии с ролями пользователей и уровнем их компетенций в анализе данных. Необходимо использовать данные и ключ, который также имеется на флеш-накопителе, выданном вам и разработать систему для расшифрования данных с помощью этого ключа.
4. **Реализация мониторинга**: добавьте функционал для мониторинга работы системы, чтобы оперативно реагировать на любые сбои или проблемы. Разработайте систему оповещений через электронные письма или мессенджеры для своевременного информирования администраторов.
5. **Разработать документацию по работе с системой**, включая инструкции по установке, настройке и использованию.
6. **Формирование рекомендаций**: определите ресурсы, необходимые для работы данной системы, чтобы сформировать рекомендации к горизонтальному и вертикальному масштабированию системы.

**Результат модуля**:

1. Программный модуль.
2. Документация на программный модуль.
3. Отчёт, содержащий подробную информацию о ходе выполнения работы.

**Модуль Д. Сетевой анализ данных (инвариант)**

**Время на выполнения модуля:**3 часа

**Краткое описание задания:** построение ориентированного графа в формате дерева для форматирования карты подготовки школьника к ВСОШ.

**Задание:**

Необходимо построить граф, демонстрирующий маршрутный лист подготовки школьника к заключительному этапу ВСОШ по информатике на основании предоставленного датасета. Граф должен иметь древовидную структуру и позволять определять очередность изучения тем.

Необходимо ответить на вопрос: на каком основании построенный маршрутный лист предлагает оптимальную образовательную траекторию.

**Результат модуля**:

1. Визуализированный граф (в формате jpeg или другом визуальном формате, позволяющем интерактивно взаимодействовать с графом).
2. Аналитическая записка в текстовом формате со всех проделанной работой и с выводами, с указанием основных метрик графа и их описанием.

2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ*[[1]](#footnote-1)*

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. |
| Временной регламент | 1. В случае выполнения конкурсантом модуля Конкурсного задания в непредназначенное для этого время, в том числе во время выполнения другого модуля, такое решение не будет оценено.2. За не прекращение выполнения Конкурсного задания по команде «СТОП» Главного эксперта или ответственных за хронометраж Экспертов в соответствии с временем выполнения Модуля или в иных ситуациях к Конкурсантам применяются штрафные санкции. |
| Программное обеспечение | 1. Конкурсанты могут создавать программные продукты, оформлять инструкции или делать заметки, находясь на рабочей площадке, однако их никогда нельзя забирать с рабочей площадки.2. 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), выполняемый модуль Конкурсного задания не оценивается.  |
| Отказ оборудования | Если имеется явное доказательство того, что конкурсанты сами причинили ущерб оборудованию, им не будет предоставляться замена и дополнительное время. |
| Поведение конкурсантов | 1. За использование ненормативной лексики устно во время выполнения Конкурсного задания или во время защиты своих работ, а также письменно в представленных к проверке файлах к Конкурсанту применяются штрафные санкции в виде снижения общей суммы баллов на 5 баллов. 2. Общение участников во время выполнения Конкурсного задания запрещено, в случае неоднократного нарушения запрета участники дисквалифицируются на оставшееся время конкурсного дня. |

Помимо описанных выше спорных ситуаций ниже приведен перечень указаний к организации работы на площадке проведения чемпионата:

1) при наличии технической возможности на площадке необходимо обеспечить видеофиксацию мониторов конкурсантов (видеозахват рабочих столов на обоих мониторах) с выгрузкой видеозаписей на выделенный сервер;

2) при наличии технической возможности на площадке для сохранения и выгрузки работ конкурсантов необходимо использовать систему контроля версий. При отсутствии технической возможности на площадке допускается выгрузка работ конкурсантов на флеш-накопитель техническим экспертом, с корректировкой соответствующих критериев в Критериях оценки;

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.

2.3 Индустриальные стандарты

Для соблюдения объективности и независимости оценки, в соответствии с положением «Об этике», каждый модуль должен иметь не менее 75% измеримых аспектов.

В случае разногласия экспертной группы в оценке измеримого аспекта, решение выносит руководитель группы оценки, задействованной в этом аспекте, и независимый эксперт.

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 «Специалист по анализу данных (BI-аналитик)»

Приложение №4 Инструкция по проведению оценки работ конкурсантов по компетенции «Специалист по анализу данных (BI аналитик)»

1. *Указываются особенности компетенции, которые относятся ко всем возрастным категориям и чемпионатным линейкам без исключения.* [↑](#footnote-ref-1)