****

КОНКУРСНОЕ ЗАДАНИЕ КОМПЕТЕНЦИИ

«СПЕЦИАЛИСТ ПО АНАЛИЗУ ДАННЫХ (BI-АНАЛИТИК)»

Итоговый (Межрегиональный) этап Чемпионата

по профессиональному мастерству «Профессионалы»

*Калужская область, Калуга Технопарк*

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 2](#_Toc142037183)

[1.1. Общие сведения о требованиях компетенции 2](#_Toc142037184)

[1.2. Перечень профессиональных задач специалиста по компетенции «Специалист по анализу данных (BI-аналитик)» 2](#_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 13](#_Toc142037191)

[2.1. Личный инструмент конкурсанта 15](#_Toc142037192)

[2.2.Материалы, оборудование и инструменты, запрещенные на площадке 15](#_Toc142037193)

[3. ПРИЛОЖЕНИЯ 15](#_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 «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. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ ЗАДАЧ СПЕЦИАЛИСТА ПО КОМПЕТЕНЦИИ «СПЕЦИАЛИСТ ПО АНАЛИЗУ ДАННЫХ (BI-АНАЛИТИК)»

*Таблица №1*

**Перечень профессиональных задач специалиста**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Раздел** | **Важность в %** |
| **1** | **Разработка модулей программного обеспечения для компьютерных систем** | **10** |
|  | 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;  - 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. |  |
| **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 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** | **Управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации** | **25** |
|  | Специалист должен знать и понимать:  Теория вероятностей и математическая статистика  Современные методы и инструментальные средства анализа больших данных  Стандарты проведения анализа данных  Методы оценки временных и стоимостных характеристик технологий больших данных  Современная технологическая инфраструктура высокопроизводительных и распределенных вычислений  Методы интерпретации и визуализации больших данных |  |
|  | Специалист должен уметь:  Проводить презентации при консультировании заказчика, согласовании и утверждении требований к результатам аналитических работ с использованием технологий больших данных  Подготавливать документы, регламентирующие требования к результатам аналитического исследования с использованием технологий больших данных в соответствии с существующими регламентами организации |  |
| **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 заказчика  Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных  Современный опыт использования анализа больших данных  Теоретическая и прикладная информатика  Теоретические и прикладные основы анализа данных  Основы бизнес-интеллекта, типы систем бизнес-интеллекта  Теория принятия решений  Математическое моделирование  Типы анализа больших данных, виды аналитики |  |
|  | Специалист должен уметь:  Использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работ  Проводить сравнительный анализ методов и инструментальных средств анализа больших данных  Проводить анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования |  |
| **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 сетях  Базовая конфигурация системы защиты информации ИАС  Особенности применения программных и программно-аппаратных средств защиты информации в ИАС |  |
|  | Специалист должен уметь:  Конфигурировать параметры системы защиты информации ИАС в соответствии с ее эксплуатационной документацией  Обнаруживать 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 авторизации  Нормативные правовые акты Российской Федерации в области защиты информации  Организационные меры по защите информации |  |
|  | Специалист должен уметь:  Производить монтаж и диагностику компонентов ИАС  Использовать 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** | 6 | 0 | 0 | 0 | 4 | **10** |
| **2** | 2,5 | 5 | 9,5 | 0 | 3 | **20** |
| **3** | 8,5 | 4,5 | 1,5 | 7,5 | 3 | **25** |
| **4** | 4 | 12 | 13 | 6 | 0 | **35** |
| **5** | 0 | 2 | 0 | 3 | 0 | **5** |
| **6** | 0 | 3 | 0 | 2 | 0 | **5** |
| **Итого баллов за критерий/модуль** | | **21** | **26,5** | **24** | **18,5** | **10** | **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 данных  2. Проверка типов данных  3. Оценка объема данных  4. Оценка погрешности |
| **Б** | **Анализ количественных данных** | Критерии:  1. Точность расчетов описательных статистик и корреляций  2. Корректность работы дашборда (отсутствие ошибок при отображении информации) / корректность построения графиков и диаграмм |
| **В** | **Обработка и анализ текстовых данных** | Критерии:  1. Содержательная оценка точности распределения навыков по классам  2. Оценка доли неклассифицируемых навыков  3. Оценка показателей полноты и точности базы |
| **Г** | **Сетевой анализ данных** | Критерии:  1. Оценка правильности построения графа  2. Оценка удобства прочтения графа (выбор укладки)  3. Оценка описания метрик графа  4. Оценка предложенных рекомендаций |
| **Д** | **Презентация решения** | Эксперты выслушивают презентации, оценивают содержание и выступление конкурсантов с помощью объективных и субъективных критериев. |

1.5. КОНКУРСНОЕ ЗАДАНИЕ

Общая продолжительность Конкурсного задания: 15 ч.

Количество конкурсных дней: 3 дней

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

Оценка знаний конкурсанта 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-ти модулей, включает обязательную к выполнению часть (инвариант) – 4 модуля, и вариативную часть – 1 модуль. Предобработка количественных данных, анализ количественных данных, обработка и анализ текстовых данных, презентация решения, сетевой анализ данных.

Общее количество баллов конкурсного задания составляет 100.

1.5.2. Структура модулей конкурсного задания (инвариант/вариатив)

**Введение**

В рамках чемпионата конкурсанты решают задачу аналитики данных по добыче полезных. Конкурсанты определяют, какие параметры отражают отраслевую специфику, какие требования предъявляют к прогнозированию добычи и точности анализа данных по различным критериям. Эти данные могут быть использованы при разработке плана добычи нефтегазовых ресурсов, программ добычи месторождений.

**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). Необходимо произвести очистку от ошибок и поиск аномальных значений, выполнить работу с пропущенными значениями, преобразовать типы данных (при необходимости).

**Задание**

**Цель:** получить, очистить и проанализировать данные, связанные с добычей нефти и газа, с использованием открытой базы данных **Oil & Gas Production Data с Kaggle**. Устранить выбросы и прочие дефекты данных.

**Описание задачи:**

1. **Получение данных с открытого источника** "Oil & Gas Production Data"
2. **Обработка данных:**
   * Очистка данных от ошибок и аномалий.
   * Проверка и обработка пропущенных значений.
   * Преобразование типов данных при необходимости.

Результат модуля: файл в формате csv или xlsx с готовой к обработке базой данных. Подготовлен отчёт о работе. В ходе выполнения задания вы получите очищенный и готовый к анализу набор данных по добыче нефти и газа. Этот процесс включает загрузку данных из конкретного открытого источника, их очистку, обработку аномальных и пропущенных значений, а также приведение типов данных к необходимым формам.

**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).

**Цель:** провести анализ основных свойств данных, выявить общие закономерности, распределения и аномалии, а также сформулировать выводы с использованием инструментов визуализации, включая построение дашборда.

**Подробное описание выполнения задачи:**

1. **Анализ основных свойств данных:** расчёт основных статистических характеристик данных, построение графиков распределения для основных переменных, корреляционный анализ между различными параметрами.
2. **Поиск закономерностей и аномалий:** выявление закономерностей и аномалий, кластерный анализ для поиска групп сходных объектов.
3. **Математические методы анализа данных**
4. **Формулирование выводов и построение отчёта**
5. **Визуализация данных:** дашборд для визуализации ключевых метрик и результатов анализа.

Построенная информационно-аналитическая система должна предусматривать методы защиты данных.

Результат модуля: интерактивный дашборд с требуемыми параметрами и краткая аналитическая записка с основными показателями. Либо подробная аналитическая записка с основными показателями, графиками и диаграммами.

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

*Время на выполнения модуля: 4 часа*

**Краткое описание задания:**

1. Выполнить парсинг данных для сбора информации с указанных веб-ресурсов;
2. Формирование структуры набора данных;
3. Провести предварительную обработку данных;
4. Выполнить построение и отбор признаков;
5. Выполнить семантический анализ текста.
6. Провести визуальный анализ статистики текстовых признаков
7. Сделать выводы, в том числе с использованием инструментов визуализации.

**Задание:**

**Цель:** выполнить парсинг данных для сбора информации с указанных веб-ресурсов, провести предварительную обработку данных, построение и отбор признаков, семантический анализ текста и визуальный анализ статистики текстовых признаков. Сделать выводы, в том числе с использованием инструментов визуализации.

**Подробное описание выполнения задачи:**

1. **Парсинг данных:** сбор информации с веб-ресурсов, связанных с добычей нефти и газа.
2. **Формирование структуры набора данных:** определение необходимых полей и создайте структуры набора данных
3. **Предварительная обработка данных:** очистка данных от HTML-тегов, специальных символов и лишних пробелов и предварительная обработка.
4. **Построение и отбор признаков:** создайте числовые признаки на основе текстовых данных: количество слов, символов, заспамленность и др.
5. **Визуальный анализ статистики текстовых признаков:** постройте гистограммы и диаграммы распределения текстовых признаков.
6. **Формулирование выводов и визуализация:** сформулируйте выводы на основе проведенного анализа, подготовьте отчёт с результатами.

Построенная информационно-аналитическая система должна предусматривать методы защиты данных, описаны требования к программному обеспечению для работы с данными.

Результат модуля:

1. Таблица в формате xlsx с классом навыков, его содержанием и частотой встречаемости.

2. Аналитическая записка в формате docx с основными выводами

**Модуль Г. Сетевой анализ данных (Вариатив)**

*Время на выполнения модуля: 2 часа*

**Краткое описание задания:**

- Выявление связей и взаимодействия между пользователями, в том числе с использованием инструментов визуализации.

- Определение ключевых узлов и групп пользователей.

- Предложение рекомендаций по улучшению взаимодействия и привлечению новых пользователей на основе анализа данных.

**Задание:**

**Краткое описание задания:** выявление связей и взаимодействия между элементами с использованием базовых методов дискретной математики, в том числе реализация инструментов визуализации. Определение ключевых элементов графа. Предложение рекомендаций на основе анализа данных.

**Задание:**

У вас есть данные о нефтяном месторождении, включающие информацию о координатах скважин, их производительности, глубине, типе нефти и др. Ваша задача – реализовать инструмент для аналитика с функционалом построения сети взаимосвязей между скважинами на основе физических и технических характеристик скважин. Затем провести сетевой анализ для определения наиболее важных узлов сети, оптимальных путей для транспортировки нефти, выявления узких мест и оптимизации логистических процессов.

Результат модуля: создание модульного инструмента сетевого анализа с указанием основных метрик и их описанием или визуализация графа с использованием специализированного программного обеспечения и его описанием, а также с рекомендациями (в формате docx)

**Модуль Д. Презентация решения (Инвариант)**

*Время на выполнения модуля: 3 часа*

**Краткое описание задания:** необходимо создать презентацию, охватывающую все результаты выполнения задания. В ней должно быть указано ёмкое описание результатов работы с обоснованием выбора того или иного решения. Также в презентации необходимо отразить скриншоты результатов своей работы. Разрабатывать презентацию рекомендуется в Power Point или аналогичной среде. Подготовить устный доклад по результатам своей работы, включающие основные результаты по каждому модулю и выводы (не более 5 минут).

**Задание:**

**Краткое описание задания:** необходимо создать презентацию, охватывающую все этапы выполнения задания. В ней должно быть указано ёмкое описание результатов работы с обоснованием выбора того или иного решения. Также в презентации необходимо отразить скриншоты результатов своей работы. Разрабатывать презентацию рекомендуется в Power Point или аналогичной среде. Подготовить устный доклад по результатам своей работы, включающие основные результаты по каждому модулю и выводы (не более 5 минут).

**Задание:**

Необходимо представить полученные результаты: подготовить презентацию в формате pptx или pdf и устный доклад (не более 5 минут).

Основные элементы презентации:

1. Техническое задание (Название проекта, Параметры для выгрузки данных, метаданные выгрузки)
2. Описание готовой базы с основными метриками.
3. Аналитический отчет с основными результатами: описательные статистики, корреляционный анализ, анализ навыков.

2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИ

В ходе проведения чемпионата есть вероятность возникновения спорных ситуаций, дающих преимущества некоторым конкурсантам 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 конкурсантов во время выполнения Конкурсного задания запрещено, в случае неоднократного нарушения запрета конкурсанты 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 конкурсантов 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 №4 Инструкция по охране труда по компетенции «Специалист по анализу данных (BI-аналитик)».