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

КОНКУРСНОЕ ЗАДАНИЕ КОМПЕТЕНЦИИ

«Машинное обучение и большие данные»

Юниоры

Итоговый (межрегиональный) этап Чемпионата по профессиональному мастерству «Профессионалы» в 2025 г

Республика Мордовия

2025 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 3](#_Toc195187549)

[1.1. ОБЩИЕ СВЕДЕНИЯ О ТРЕБОВАНИЯХ КОМПЕТЕНЦИИ 3](#_Toc195187550)

[1.2. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ ЗАДАЧ СПЕЦИАЛИСТА ПО КОМПЕТЕНЦИИ «МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ» 3](#_Toc195187551)

[1.3. ТРЕБОВАНИЯ К СХЕМЕ ОЦЕНКИ 9](#_Toc195187552)

[1.4. СПЕЦИФИКАЦИЯ ОЦЕНКИ КОМПЕТЕНЦИИ 9](#_Toc195187553)

[1.5. КОНКУРСНОЕ ЗАДАНИЕ 10](#_Toc195187554)

[1.5.1. Разработка/выбор конкурсного задания 10](#_Toc195187555)

[1.5.2. Структура модулей конкурсного задания 10](#_Toc195187556)

[2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ 15](#_Toc195187558)

[2.1 Личный инструмент конкурсанта 15](#_Toc195187559)

[3. Приложения 15](#_Toc195187560)

**ИСПОЛЬЗУЕМЫЕ СОКРАЩЕНИЯ**

1. ФГОС – Федеральный государственный образовательный стандарт
2. ПС – Профессиональный стандарт
3. КЗ – Конкурсное задание
4. ИЛ – Инфраструктурный лист
5. ML – Машинное обучение
6. DL – Глубокое обучение
7. AI – Искусственный интеллект
8. NLP – Обработка естественного языка
9. SMOTE – Техника синтетического увеличения малочисленного класса
10. ROC – Характеристика работы приемника
11. AUC – Площадь под кривой
12. MSE – Среднеквадратическая ошибка
13. MAE –Средняя абсолютная ошибка
14. MAPE – Средняя абсолютная процентная ошибка
15. API – Программный интерфейс
16. EDA –Исследовательский анализ данных
17.

1. ОСНОВНЫЕ ТРЕБОВАНИЯ КОМПЕТЕНЦИИ

1.1. ОБЩИЕ СВЕДЕНИЯ О ТРЕБОВАНИЯХ КОМПЕТЕНЦИИ

Требования компетенции (ТК) «Машинное обучение 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 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 «МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ»

*Перечень видов профессиональной деятельности, умений и знаний, и профессиональных трудовых функций специалиста (из ФГОС/ПС/ЕТКС.) и базируется на требованиях современного рынка труда к данному специалисту*

*Таблица №1*

**Перечень профессиональных задач специалиста**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Раздел** | **Важность в %** |
| 1 | **Планирование и организация аналитических работ с использованием технологий больших данных** | 22% |
| Специалист должен знать и понимать:* Предметная область анализа больших данных в соответствии с требованиями заказчика;
* Основы планирования аналитических работ;
* Стандарты проведения анализа данных;
* Методы и инструментальные средства управления аналитическими проектами по исследованию больших данных;
* Содержание и последовательность выполнения этапов аналитического проекта по исследованию больших данных;
* Типы анализа больших данных, виды аналитики;
* Теоретические и прикладные основы анализа больших данных;
* Современные методы и инструментальные средства анализа больших данных;
* Теория вероятностей и математическая статистика;
* Методы интерпретации и визуализации анализа больших данных
 |  |
| Специалист должен уметь:* Представлять содержание и результаты работ по анализу больших данных;
* Планировать аналитические работы с использованием технологий больших данных;
* Проводить аналитические работы с использованием технологий больших данных;
* Проводить анализ больших данных;
* Осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных
 |  |
| 2 | **Подготовка данных для проведения аналитических работ по исследованию больших данных** | 27% |
| Специалист должен знать и понимать:* Предметная область анализа;
* Теоретические и прикладные основы анализа больших данных;
* Современные методы и инструментальные средства анализа больших данных;
* Типы больших данных: метаданные, полуструктурированные, структурированные, неструктурированные;
* Виды источников данных: созданные человеком, созданные машинами;
* Источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования;
* Методы извлечения информации и знаний из гетерогенных, мультиструктурированных, неструктурированных источников, в том числе при потоковой обработке;
* Российские и международные стандарты информационной безопасности;
* Режимы получения и обработки данных, поддержка режима реального времени;
* Технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памяти
 |  |
| Специалист должен уметь:* Осуществлять взаимодействие с внутренними и внешними поставщиками данных из гетерогенных источников;
* Разрабатывать и оценивать модели больших данных;
* Использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени;
* Производить очистку данных для проведения аналитических работ;
* Проводить интеграцию и преобразование больших объемов данных;
* Оценивать соответствие наборов данных задачам анализа больших данных
 |  |
| 3 | **Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика** | 23% |
| Специалист должен знать и понимать:* Предметная область анализа;
* Теория принятия решений;
* Математическое моделирование;
* Теоретические и прикладные основы анализа больших данных;
* Технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, планирование экспериментов, карты контроля качества;
* Статистический анализ: метод многовариантного тестирования, корреляционный анализ, регрессионный анализ;
* Статистические методы: параметрические, непараметрические, управляемые, неуправляемые, полууправляемые, кластеризация;
* Семантический анализ: обработка естественного языка, сентиментный анализ, анализ текста;
* Алгоритмы машинного обучения: обучение с учителем, обучение без учителя, полууправляемое обучение, обучение с подкреплением;
* Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация;
* Методы и модели классификации: логистическая регрессия, деревья решений, предредукция, постредукция, модели, основанные на правилах, вероятностные классификаторы, усиление энтропии информации;
* Фильтрация шумовых выбросов, виды шумовых выбросов: глобальный, контекстуальный, коллективный;
* Анализ изображений, анализ сетей, анализ пространственных данных, анализ временных рядов;
* Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма;
* Распределенный анализ данных;
* Методы разработки отчетной аналитической документации
 |  |
|  | Специалист должен уметь:* Разрабатывать и оценивать модели больших данных;
* Программировать на языках высокого уровня, ориентированных на работу с большими данными: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных;
* Адаптировать и развертывать модели в предметной среде;
* Решать задачи классификации, кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данных;
* Решать проблемы переобучения и недообучения алгоритма;
* Формировать предложения по использованию результатов анализа;
* Оформлять результаты аналитического исследования для представления заказчику;
* Осуществлять поиск информации о новых и перспективных методах анализа больших данных, выполнять сравнительный анализ методов
 |  |
| 4 | **Разработка продуктов на основе встроенной аналитики больших данных** | 17% |
| Специалист должен знать и понимать:* Локальные и глобальные потребности в создании новых и модернизации существующих продуктов на основе встроенной аналитики больших данных;
* Существующие и перспективные методы и программный инструментарий технологий больших данных;
* Существующий опыт разработки и использования продуктов и услуг на основе технологий больших данных;
* Современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников;
* Источники больших данных, интенсивность генерации данных источниками;
* Технические средства и среды сбора, хранения и обработки больших данных;
* Принципы и методы управления защитой и обеспечением конфиденциальности больших данных;
* Современные и перспективные средства визуализации и интерпретации больших данных;
* Машинное обучение;
* Математическое моделирование;
* Теория принятия решений;
* Высокопроизводительные и распределенные вычисления
 |  |
|  | Специалист должен уметь:* Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных;
* Проводить маркетинговые исследования в области информационных продуктов и услуг;
* Разрабатывать конкурсную, проектную и рабочую документацию на разработку новых продуктов;
* Проводить технико-экономическое обоснование разработки новых продуктов;
* Оценивать экономические параметры технологий больших данных;
* Осуществлять разработку программно-аппаратных компонентов и систем;
* Осуществлять математическое и информационное моделирование;
* Проводить аналитические работы на основе технологий больших данных;
* Разрабатывать научно-техническую документацию;
* Проводить презентации, подготавливать публикации по итогам проектных работ
 |  |
| 5 | **Разработка сервисов на основе аналитики больших данных** | 11% |
| Специалист должен знать и понимать:* Существующий опыт разработки и использования продуктов и услуг на основе технологий больших данных;
* Особенности социально-экономической сферы использования сервисов на основе больших данных;
* Правовые и этические аспекты предоставления услуг на основе аналитики больших данных;
* Современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников;
* Источники больших данных, интенсивность генерации данных источниками;
* Технические средства и среды сбора, хранения и обработки больших данных;
* Современные и перспективные средства визуализации и интерпретации больших данных;
* Высокопроизводительные и распределенные вычисления;
* Методы маркетинговых исследований;
* Методы сравнительного анализа;
* Основы инновационной деятельности и управления инновациями;
* Основы управления информационно-технологическими проектами
 |  |
|  | Специалист должен уметь:* Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных;
* Проводить маркетинговые исследования в области аналитических услуг на основе технологий больших данных;
* Разрабатывать конкурсную, проектную и рабочую документацию на разработку аналитических услуг на основе технологий больших данных;
* Выполнять технико-экономическое обоснование разработки и использования услуг на основе технологий больших данных;
* Оценивать экономические параметры технологий больших данных;
* Разрабатывать научно-техническую документацию по проектам в области больших данных;
* Проводить презентации, подготавливать публикации по итогам проектных работ в области больших данных
 |  |

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** | 22 |  |  |  |  | 22 |
| **2** |  | 27 |  |  |  | 27 |
| **3** |  |  | 23 |  |  | 23 |
| **4** |  |  |  | 17 |  | 17 |
| **5** |  |  |  |  | 11 | 11 |
| **Итого баллов за критерий/модуль** | 22 | 27 | 23 | 17 | 11 | **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** | **Парсинг и предобработка данных** | Экспертная оценка практических результатов, согласно измеримым и судейским аспектам критериев оценки |
| **Б** | **Разведочный анализ данных** | Экспертная оценка практических результатов, согласно измеримым и судейским аспектам критериев оценки |
| **В** | **Построение, обучение и оптимизация модели** | Автоматизированная оценка |
| **Г** | **Разработка программного продукта** | Экспертная оценка практических результатов, согласно измеримым и судейским аспектам критериев оценки |
| **Д** | **Разработка средств интеграции и поддержки готового решения** | Экспертная оценка практических результатов, согласно измеримым и судейским аспектам критериев оценки |

1.5. КОНКУРСНОЕ ЗАДАНИЕ

Общая продолжительность Конкурсного задания[[1]](#footnote-1): 10 ч.

Количество конкурсных дней: 3 дня

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 модулей, включает обязательную к выполнению часть (инвариант) – 3 модуля, и вариативную часть – 2 модуля. Общее количество баллов конкурсного задания составляет 100.

1.5.2. Структура модулей конкурсного задания

ОСНОВНАЯ ЗАДАЧА — построение интеллектуальной системы управления активами компании на основе комплексного анализа**.**

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

Время на выполнение модуля – 2 часа

**Описание модуля A:**

В рамках данного модуля необходимо загрузить данные из различных источников. Изначально данные могут храниться в реляционных и нереляционных базах данных, в файлах или других источниках данных, таких как файлы JSON, XML и др. Провести необходимые преобразования данных, которые требуется выполнить перед загрузкой их в единую централизованную базу данных. Это может включать очистку данных, преобразование форматов, объединение данных из разных источников и т. д. Создать целевую базу данных для аналитики и загрузить в неё подготовленные данные. Обеспечить целостность и актуальность информации в базе данных для последующего выполнения аналитических запросов и отчётов.

**Модуль Б. Разведочный анализ данных (инвариант)**

Время на выполнение модуля – 2 часа

**Описание модуля Б:**

Для своевременного принятия решения на основе данных, минимизации ошибок и быстрого достижения целей необходимо построить аналитический дашборд. Интерактивная информационная панель должна наглядно представлять, визуализировать и анализировать данные. Для этого необходимо подключить к системе источник данных, т.е. подключиться к базе данных, настроить регламент обновления и уровни доступа. Аналитическая система должна позволить пользователю:

* автоматизировать работу с данными;
* анализировать все в реальном времени;
* видеть все главные метрики на одном экране;
* выполнять визуализацию с учетом фильтрации данных;
* легко считывать информацию и делать выводы.

Разработанная информационная панель (дашборд) должна позволить пользователю проводить анализ:

* текущего финансового состояния компании;
* прогнозных показателей и рисков;
* эффективности управления оборотным капиталом;
* паттернов и причинно-следственных связей;
* платежного поведения.

**Модуль В. Построение, обучение и оптимизация модели (инвариант)**

Время на выполнение модуля – 2 часа

**Описание модуля В:**

Задание модуля заключается в разработке и оптимизации моделей машинного обучения для предиктивного анализа финансовых потоков компании. Целью является построение точных и интерпретируемых моделей, способных прогнозировать ключевые показатели на основе исторических финансовых и операционных данных. В рамках модуля необходимо реализовать полный цикл обработки данных: от их преобразования до выбора признаков, обучения моделей, их валидации и тестирования. Предусматривается построение нескольких моделей с использованием различных алгоритмов машинного обучения, применимые к характеру данных. На основе заданных метрик качества необходимо провести сравнительный анализ полученных моделей и выбрать оптимальную по совокупности критериев, включая точность прогноза, устойчивость к шуму, интерпретируемость и способность к обобщению. Итоговая модель должна быть подготовлена к интеграции в бизнес-процессы компании с возможностью масштабирования и регулярного обновления на основе новых данных для поддержки принятия управленческих решений и повышения эффективности управления активами.

**Модуль Г. Разработка программного продукта (вариатив)**

Время на выполнение модуля – 2 часа

**Описание модуля Г:**

На данном этапе конкурсантам необходимо выполнить автоматизацию процессов сбора, обработки и анализа данных, что позволит более эффективно проводить аналитику данных и принимать оперативные решения на основе полученной информации. Для решения поставленной задачи конкурсантам необходимо составить план и обеспечить автоматизацию процессов извлечения, трансформации и загрузки данных из различных источников компании в хранилище данных. Обеспечить автоматизацию запуска задач анализа данных, таких как вычисление ключевых метрик и создания отчетов или визуализации для мониторинга оборотного капитала компании. Обеспечить мониторинг и управление рабочими процессами для быстрого реагирования на изменения и проблемы в процессе анализа данных. Требуется выполнить повторное обучение моделей на основе обнаруженного дрейфа и обновленных данных.

Также в рамках данного модуля также требуется разработать программное решение, реализующее предиктивный анализ на основе моделей машинного обучения. Модуль должен обеспечивать загрузку, предварительную обработку финансовых и операционных данных и генерацию прогноза ключевых показателей.

**Модуль Д. Разработка средств интеграции и поддержки готового решения (вариатив)**

Время на выполнение модуля – 2 часа

**Описание модуля Д:**

Для разработанного приложения ПП составьте пользовательскую документацию, представляющую собой руководство пользователя. Руководство должно описывать каждую функцию программы, а также шаги, которые нужно выполнить для использования этой функции. Пользовательская документация должна предоставлять инструкции о том, что делать в случае возникновения проблем. Очень важно, чтобы документация не вводила в заблуждение и была актуальной. Руководство должно иметь чёткую структуру. Необходимо создать презентацию, охватывающую все результаты

выполнения задания.

Кроме того, на данном этапе конкурсантам необходимо презентовать полученные решения заказчику или другим заинтересованным сторонам, чтобы помочь им принимать обоснованные решения на основе данных. Презентация решения должна представлять собой наглядный и информативный документ, который демонстрирует основные характеристики, функциональные возможности и преимущества разработанного дашборда и программного продукта. Вот основные разделы, которые могут быть включены в такую презентацию: введение, описание данных, функциональные возможности, преимущества прикладного решения, примеры использования, развертывание и поддержка, сводка и заключение.

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

2.1 Личный инструмент конкурсанта

1. Конкурсанты могут использовать защиту для ушей.
2. Конкурсанты могут принести с собой свои клавиатуры, мышки и коврики для мышек. Все принесенные клавиатуры, мышки и коврики должны быть предварительно сданы на проверку Техническому эксперту. Запрещено использование Клавиатур и мышек с подключением по беспроводным каналам. Устройства ввода не должны быть программируемыми.

### 2.2.Материалы, оборудование и инструменты, запрещенные на площадке

1. Конкурсанты не должны приносить:
* Дополнительные программы.
* Мобильные телефоны.
* Портативные электронные устройства (планшеты, и т.п.).
* Устройства для хранения информации (флэш-накопители, диски, и т.п.).
1. Оборудование не должно иметь доступ к внутренним устройствам для хранения информации. Организаторы соревнования проверят, что доступ был заблокирован.

3. Приложения

Приложение 1. Инструкция по заполнению матрицы конкурсного задания

Приложение 2. Матрица конкурсного задания

Приложение 3. Инструкция по охране труда

1. *Указывается суммарное время на выполнение всех модулей КЗ одним конкурсантом.* [↑](#footnote-ref-1)
2. *Указываются особенности компетенции, которые относятся ко всем возрастным категориям и чемпионатным линейкам без исключения.* [↑](#footnote-ref-2)