КОНКУРСНОЕ ЗАДАНИЕ КОМПЕТЕНЦИИ

«Машинное обучение и большие данные»

2023 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](#_Toc124422965)

[1.1. ОБЩИЕ СВЕДЕНИЯ О ТРЕБОВАНИЯХ КОМПЕТЕНЦИИ 2](#_Toc124422966)

[1.2. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ ЗАДАЧ СПЕЦИАЛИСТА ПО КОМПЕТЕНЦИИ «Машинное обучение и большие данные» 2](#_Toc124422967)

[1.3. ТРЕБОВАНИЯ К СХЕМЕ ОЦЕНКИ 4](#_Toc124422968)

[1.4. СПЕЦИФИКАЦИЯ ОЦЕНКИ КОМПЕТЕНЦИИ 4](#_Toc124422969)

[1.5.2. Структура модулей конкурсного задания (инвариант/вариатив) 7](#_Toc124422970)

[2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ 8](#_Toc124422971)

[2.1. Личный инструмент конкурсанта 8](#_Toc124422972)

[3. Приложения 8](#_Toc124422973)

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»

**Таблица 1. Перечень профессиональных задач специалиста**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Раздел** | **Важность в %** |
| 1 | **Планирование и организация аналитических работ с использованием технологий больших данных** | 19% |
| - Специалист должен знать и понимать:Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данныхВозможности использования свободно распространяемого программного обеспечения для анализа больших данныхПредметная область анализа больших данных в соответствии с требованиями заказчикаОсновы планирования аналитических работСтандарты проведения анализа данныхМетоды и инструментальные средства управления аналитическими проектами по исследованию больших данныхСодержание и последовательность выполнения этапов аналитического проекта по исследованию больших данныхСодержание этапов жизненного цикла больших данныхТипы анализа больших данных, виды аналитикиТеоретические и прикладные основы анализа больших данныхСовременные методы и инструментальные средства анализа больших данныхТеория вероятностей и математическая статистикаИсточники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследованияМетоды интерпретации и визуализации анализа больших данныхТехнологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологииТехнологии подготовки и проведения презентацийПравила деловой переписки |  |
| - Специалист должен уметь:Проводить переговоры при определении содержания аналитических работ с использованием технологий больших данныхПредставлять содержание и результаты работ по анализу больших данныхВести протоколы мероприятий по анализу больших данныхПланировать аналитические работы с использованием технологий больших данныхПроводить аналитические работы с использованием технологий больших данныхПроводить анализ больших данныхОсуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных |  |
| 2 | **Подготовка данных для проведения аналитических работ по исследованию больших данных** | 26% |
| - Специалист должен знать и понимать:Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данныхПредметная область анализаТеоретические и прикладные основы анализа больших данныхСовременные методы и инструментальные средства анализа больших данныхСовременный опыт использования анализа больших данныхТипы больших данных: метаданные, полуструктурированные, структурированные, неструктурированныеВиды источников данных: созданные человеком, созданные машинамиИсточники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследованияМетоды извлечения информации и знаний из гетерогенных, мультиструктурированных, неструктурированных источников, в том числе при потоковой обработкеРоссийские и международные стандарты информационной безопасностиСовременная технологическая инфраструктура высокопроизводительных и распределенных вычисленийРежимы получения и обработки данных, поддержка режима реального времениТехнологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памятиОблачные технологии, облачные сервисыМетоды оценки временных и стоимостных характеристик технологий больших данныхТехнологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологииПравила деловой переписки |  |
| - Специалист должен уметь:Определять требования к поставщикам данных из гетерогенных источниковОсуществлять взаимодействие с внутренними и внешними поставщиками данных из гетерогенных источниковРазрабатывать и оценивать модели больших данныхИспользовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времениПроизводить очистку данных для проведения аналитических работПроводить интеграцию и преобразование больших объемов данныхОценивать соответствие наборов данных задачам анализа больших данныхОценивать стоимость данных для проведения аналитических работ |  |
| 3 | **Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика** | 19% |
| - Специалист должен знать и понимать:Содержание и последовательность выполнения этапов аналитического проектаОсновы управления аналитическими работамиОсновы управления малыми аналитическими группамиВозможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данныхПредметная область анализаТеория принятия решенийМатематическое моделированиеТеоретические и прикладные основы анализа больших данныхСовременный опыт использования анализа больших данныхТехнологии 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 качестваНейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерностиСтатистические моделиСтатистический анализ: метод многовариантного тестирования, корреляционный анализ, регрессионный анализСтатистические методы: параметрические, непараметрические, управляемые, неуправляемые, полууправляемые, кластеризацияСемантический анализ: обработка естественного языка, сентиментный анализ, анализ текстаАлгоритмы машинного обучения: обучение с учителем, обучение без учителя, полууправляемое обучение, обучение с подкреплениемМашинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрацияМетоды и модели классификации: логистическая регрессия, деревья решений, предредукция, постредукция, модели, основанные на правилах, вероятностные классификаторы, усиление энтропии информацииФильтрация шумовых выбросов, виды шумовых выбросов: глобальный, контекстуальный, коллективныйАнализ изображений, анализ сетей, анализ пространственных данных, анализ временных рядовМетоды идентификации шаблоновМетоды оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритмаРаспределенный анализ данныхАнализ данных в реальном времениПравила деловой перепискиМетоды разработки отчетной аналитической документации |  |
|  | - Специалист должен уметь:Планировать аналитические работы с использованием технологий больших данныхПроводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группамиИспользовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работПроводить сравнительный анализ методов и инструментальных средств анализа больших данныхРазрабатывать и оценивать модели больших данныхПрограммировать 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 данныхАдаптировать и развертывать модели в предметной средеРешать задачи классификации, кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данныхРешать проблемы переобучения и недообучения алгоритмаФормировать предложения по использованию результатов анализаОформлять результаты аналитического исследования для представления заказчикуРазъяснять заказчику результаты аналитической работыОсуществлять поиск информации о новых и перспективных методах анализа больших данных, выполнять сравнительный анализ методов |  |
| 4 | **Разработка продуктов на основе встроенной аналитики больших данных** | 18% |
| - Специалист должен знать и понимать:Состояние и перспективы развития в Российской Федерации и в мире информационных технологий нового поколения, предназначенных для экономически эффективного извлечения полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, а также продуктов и услуг на их основеЛокальные и глобальные потребности в создании новых и модернизации существующих продуктов на основе встроенной аналитики больших данныхСуществующие и перспективные методы и программный инструментарий технологий больших данныхСуществующий опыт разработки и использования продуктов и услуг на основе технологий больших данныхСовременные и перспективные методы сбора, хранения и передачи данных из гетерогенных источниковИсточники больших данных, интенсивность генерации данных источникамиТехнические средства и среды сбора, хранения и обработки больших данныхПринципы и методы управления защитой и обеспечением конфиденциальности больших данныхОсновы правового регулирования оборота и использования больших данныхСовременные и перспективные средства визуализации и интерпретации больших данныхСистемная инженерияМашинное обучениеМатематическое моделированиеТеория принятия решенийВысокопроизводительные и распределенные вычисленияМетоды маркетинговых исследованийМетоды сравнительного анализаОсновы инновационной деятельности и управления инновациями в сфере информационных технологийОсновы управления информационно-технологическими проектамиОсновы управления проектными коллективамиОсновы управления взаимоотношениями с партнерамиПоказатели эффективности технологий больших данныхОсновы охраны авторских прав и интеллектуальной собственности в сфере информационных технологийПравила деловой переписки |  |
|  | - Специалист должен уметь:Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данныхПроводить маркетинговые исследования в области информационных продуктов и услугРазрабатывать конкурсную, проектную и рабочую документацию на разработку новых продуктовПроводить технико-экономическое обоснование разработки новых продуктовОценивать экономические параметры технологий больших данныхОсуществлять разработку программно-аппаратных компонентов и системОсуществлять математическое и информационное моделированиеПроводить аналитические работы на основе технологий больших данныхФормировать коллектив исполнителейУправлять взаимоотношениями с исполнителями и соисполнителями проектных работУправлять коллективом исполнителейУправлять исполнением проектных работРазрабатывать научно-техническую документациюПроводить согласование с заказчиком содержания и отчетной документации проектных работПроводить презентации, подготавливать публикации по итогам проектных работ |  |
| 5 | **Разработка сервисов на основе аналитики больших данных** | 18% |
| - Специалист должен знать и понимать:Состояние и перспективы развития в Российской Федерации и в мире информационных технологий нового поколения, предназначенных для экономически эффективного извлечения полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, а также продуктов и услуг на их основеЛокальные и глобальные потребности в создании новых и модернизации существующих сервисов на основе встроенной аналитики больших данныхСуществующие и перспективные методы и программный инструментарий технологий больших данныхПринципы и методы управления защитой и обеспечением конфиденциальности больших данныхСуществующий опыт разработки и использования продуктов и услуг на основе технологий больших данныхОсобенности социально-экономической сферы использования сервисов на основе больших данныхОсновы науки о сервисахСервисы и системы обслуживанияСервис-ориентированная архитектураТехнология веб-сервисов и мобильных сервисовМоделирование процессов предоставления услугУправление качеством обслуживанияЭффективность сервисовСервисные инновацииПравовые и этические аспекты предоставления услуг на основе аналитики больших данныхСовременные и перспективные методы сбора, хранения и передачи данных из гетерогенных источниковИсточники больших данных, интенсивность генерации данных источникамиТехнические средства и среды сбора, хранения и обработки больших данныхСовременные и перспективные средства визуализации и интерпретации больших данныхВысокопроизводительные и распределенные вычисленияМетоды маркетинговых исследованийМетоды сравнительного анализаОсновы инновационной деятельности и управления инновациямиОсновы управления информационно-технологическими проектамиОсновы управления проектными коллективамиОсновы управления взаимоотношениями с партнерамиПоказатели эффективности технологий больших данныхОсновы охраны авторских прав и интеллектуальной собственности в сфере информационных технологийПравила деловой переписки |  |
|  | - Специалист должен уметь:Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данныхПроводить маркетинговые исследования в области аналитических услуг на основе технологий больших данныхРазрабатывать конкурсную, проектную и рабочую документацию на разработку аналитических услуг на основе технологий больших данныхВыполнять технико-экономическое обоснование разработки и использования услуг на основе технологий больших данныхОценивать экономические параметры технологий больших данныхРазрабатывать программно-аппаратные компоненты и системы на основе технологий больших данныхПроводить математическое и информационное моделированиеВыполнять аналитические работы на основе технологий больших данныхФормировать коллектив исполнителей проектов в области больших данныхУправлять взаимоотношениями с исполнителями и соисполнителями проектных работ в области больших данныхУправлять коллективом исполнителей проектов в области больших данныхУправлять исполнением проектных работ в области больших данныхРазрабатывать научно-техническую документацию по проектам в области больших данныхСогласовывать с заказчиком содержание проектных работ в области больших данных и отчетную документацию по нимПроводить 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** | 19 |  |  |  |  | 19 |
| **2** |  | 26 |  |  |  | 26 |
| **3** |  |  | 19 |  |  | 19 |
| **4** |  |  |  | 18 |  | 18 |
| **5** |  |  |  |  | 18 | 18 |
| **Итого баллов за критерий/модуль** | 19 | 26 | 19 | 18 | 18 | **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**

Возрастной ценз: 16–22 года.

Общая продолжительность Конкурсного задания[[1]](#footnote-1): 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**

 **(https://disk.yandex.ru/i/CxkVbLaP\_WwFNQ)**

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.

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 вариативные модули формируется регионом самостоятельно под запрос работодателя. При этом, время на выполнение модулей и количество баллов в критериях оценки по аспектам не меняются.

**Таблица 4. Матрица конкурсного задания**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Обобщенная трудовая функция** | **Трудовая функция** | **Нормативный документ/ЗУН** | **Модуль** | **Константа/****вариатив** | **ИЛ** | **КО** |
| Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры | Планирование и организация аналитических работ с использованием технологий больших данных | ПС: 06.042; ФГОС СПО 09.02.07 Информационные системы и программирование | Модуль 1 – Парсинг и предобработка данных | Константа  | [Раздел ИЛ 1](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)[Раздел ИЛ 2](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Рабочая_площадка_М2)[Раздел ИЛ 3](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Модуль3)[Раздел ИЛ 4](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль5)[Раздел ИЛ 5](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль6) | 19 |
| Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры | Подготовка данных для проведения аналитических работ по исследованию больших данных | ПС: 06.042; ФГОС СПО 09.02.07 Информационные системы и программирование | Модуль 2 - Разведочный анализ данных | Константа | [Раздел ИЛ 1](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)[Раздел ИЛ 2](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Рабочая_площадка_М2)[Раздел ИЛ 3](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Модуль3)[Раздел ИЛ 4](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль5)[Раздел ИЛ 5](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль6) | [26](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#КО2!A1) |
| Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры | Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика | ПС: 06.042; ФГОС СПО 09.02.07 Информационные системы и программирование | Модуль 3 – Построение, обучение и оптимизация модели | Константа | [Раздел ИЛ 1](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)[Раздел ИЛ 2](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Рабочая_площадка_М2)[Раздел ИЛ 3](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Модуль3)[Раздел ИЛ 4](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль5)[Раздел ИЛ 5](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль6) | [19](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#'КО 3'!A1) |
| Управление разработкой продуктов, услуг и решений на основе больших данных | Разработка продуктов на основе встроенной аналитики больших данных | ПС: 06.042; ФГОС СПО 09.02.07 Информационные системы и программирование | Модуль 4 – Разработка программного продукта | Вариатив | [Раздел ИЛ 1](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)[Раздел ИЛ 2](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Рабочая_площадка_М2)[Раздел ИЛ 3](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Модуль3)[Раздел ИЛ 4](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль5)[Раздел ИЛ 5](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль6) | [18](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#КО5!A1) |
| Управление разработкой продуктов, услуг и решений на основе больших данных | Разработка сервисов на основе аналитики больших данных | ПС: 06.042; ФГОС СПО 09.02.07 Информационные системы и программирование | Модуль 5 – Разработка средств интеграции и поддержки готового решения | Вариатив | [Раздел ИЛ 1](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)[Раздел ИЛ 2](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Рабочая_площадка_М2)[Раздел ИЛ 3](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#Модуль3)[Раздел ИЛ 4](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль5)[Раздел ИЛ 5](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#модуль6) | [18](file:///C%3A%5CUsers%5Csmirnov.e%5CDownloads%5C6%20%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0.xlsx#КО6!A1) |

1.5.2. Структура модулей конкурсного задания (инвариант/вариатив)

**Модуль А. Парсинг и предобработка данных**

Время на выполнение модуля – 3 часа

**Задания:**

|  |
| --- |
| 1.1 Загрузка данных |
| Загрузить данные с ресурсов. Необходимо оставить не менее 40% от исходного количества записей. В наборах данных для последующего обучения должны быть включены записи по каждому месяцу, присутствующему в исходных наборах, в пропорциональном количестве. |

|  |
| --- |
| 1.2 Предобработка данных и выделение значимых атрибутов |
| Для каждого набора данных (по годам) предполагается применить методы кластерного анализа для создания групп сходных объектов. Такие кластеры в будущем помогут выявить динамику зависимости различных параметров для оптимизации ресурсов предприятия пассажирских перевозок.Необходимо определить, какие атрибуты имеют наибольшее влияние на определение таких групп объектов (кластеров), и оставить только их для последующего обучения. Также необходимо обосновать выбор дополнительных атрибутов и причину исключения каких-либо данных из исходного набора.  |

|  |
| --- |
| 1.3 Описание структуры набора данных |
| Для каждого атрибута подготовьте описание, содержащее текстовое представление (расшифровка, перевод, назначение). Выполнить проверку нормальности распределения значений в атрибутах. Для определения отклонений от нормы провести графическую и статистические оценки. Обосновать выбор методов оценки. На основе полученных результатов сделать вывод о типе распределения и мере скошенности. |

|  |
| --- |
| 1.4 Определение численных характеристик наборов данных |
| Определите количество записей в каждом наборе (по годам), удовлетворяющих следующим требованиям:1. Районы, из которых чаще всего производилась посадка пассажиров, где чаевых было оставлено не менее 15% от суммы оплаты за поездку (длина поездки не должна превышать 2 км);
2. Стоимость на километр поездки не превышала среднюю стоимость всех поездок на километр по такому же тарифу;
3. Среднее количество пассажиров на поездку, пользующихся услугами такси с самыми популярными тарифами.

Сравните полученные численные значения и выявите динамику изменений по годам. |

|  |
| --- |
| 1.5 Подготовка отчета |
| Подготовьте отчет о проделанной работе по итогам сессии в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip |

**Модуль Б. Разведочный анализ данных**

Время на выполнение модуля – 3 часа

**Задания:**

|  |
| --- |
| 2.1 Кластеризация набора данных |
| Выберите модель кластеризации данных. Задача кластеризации – сегментация поездок, определение схожих группы объектов в каждом наборе данных. В результате должно быть выявлено не менее 3 и не более 5 кластеров. При кластеризации могут получиться группы поездок, схожие по различным характеристикам. Необходимо выявить такие характеристики и дать подробное описание каждого кластера. |

|  |
| --- |
| 2.2 Вычисление показателей производительности кластеризации |
| Выберите метрику оценки качества кластеризации. Обоснуйте выбор методов иприемов. Выполните визуальный анализ кластерных структур и оценки качества кластеризации.Определите лучший алгоритм кластеризации на основе выбранной метрики. |

|  |
| --- |
| 2.3 Подготовка отчета |
| Подготовьте отчет о проделанной работе по итогам сессии в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip |

**Модуль В. Построение, обучение и оптимизация модели**

Время на выполнение модуля – 3 часа

**Задания:**

|  |
| --- |
| 3.1 Разведочный анализ данных |
| Проведите сравнительный анализ и сопоставьте выявленные кластеры в наборах данных по годам. Для каждого кластера выявите не менее двух характеристик, проявляющих устойчивые закономерности на протяжении трёх лет. Опишите эти закономерности. |

|  |
| --- |
| 3.2 Визуализация закономерностей |
| Определите динамику изменения выявленных характеристик и покажите её на графиках, используя методы и инструментальные средства визуализации данных. Визуализация должна наглядно отражать рост или снижение показателей для каждого кластера. |

|  |
| --- |
| 3.3 Прогнозирование динамики изменения характеристик |
| На основе выявленных характеристик сделайте прогноз изменения значений на последующий период. Прогноз должен отражать динамику минимум на два года вперёд. Визуализируйте результаты прогнозирования для каждой характеристики. |

|  |
| --- |
| 3.4 Подготовка отчета |
| Подготовьте отчет о проделанной работе по итогам сессии в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip |

**Модуль Г. Разработка программного продукта**

Время на выполнение модуля – 3 часа

**Задания:**

|  |
| --- |
| 4.1 Разработка API |
| Разработайте программный интерфейс для итоговой модели машинного обучения. APIдолжен позволить приложению пользователя получать доступ к регрессионной модели дляпрогнозирования динамики изменения характеристик перевозок по кластерам на последующий период времени. |

|  |
| --- |
| 4.2 Разработка приложения |
| Разработайте приложение с графическим интерфейсом, которое должно c помощьюразработанного API показывать значения характеристик в заданных кластерах на выбранный период времени.Приложение должно предоставлять справку по имеющимся командам и их параметрам. |

|  |
| --- |
| 4.3 Программная документация |
| Для разработанного приложения и API составьте программную документацию и руководство пользователя |

**Модуль Д. Разработка средств интеграции и поддержки готового решения**

Время на выполнение модуля – 3 часа

**Задания:**

|  |
| --- |
| 5.1 Разработка документации по системе |
| В ходе выполнения конкурсного задания вы создавали отчёты по результатам каждой из сессий. Для выполнения данного задания необходимо создать документацию, охватывающую всю вашу выполненную работу: обоснования выбранных решений и способов их реализации, скриншоты готового решения и всё то, что вы посчитаете нужным указать для ознакомления с вашей работой стороннему пользователю или специалисту. |

|  |
| --- |
| 5.2 Презентация результатов работы |
| Необходимо создать презентацию, охватывающую все результаты выполнения конкурсного задания. В ней должно быть указано ёмкое описание результатов работы с обоснованием выбора того или иного решения. Так же в презентации необходимо отразить скриншоты результатов своей работы. Презентация должна быть выполнена в формате «Печа-куча». |

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

2.1. Личный инструмент конкурсанта

### Участники могут использовать защиту для ушей.

### Участники 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. Участники 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 Критерии оценки

Приложение №4 Инструкция по охране труда и технике безопасности по компетенции «Машинное обучение и большие данные».

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)