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 данных  Возможности использования свободно распространяемого программного обеспечения для анализа больших данных  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  Содержание этапов жизненного цикла 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  Источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования  Методы интерпретации и визуализации анализа больших данных  Технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии  Технологии подготовки и проведения презентаций  Правила деловой переписки |  |
| - Специалист должен уметь:  Проводить переговоры при определении содержания аналитических работ с использованием технологий больших данных  Представлять содержание и результаты работ по анализу больших данных  Вести протоколы мероприятий 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** | 26% |
| - Специалист должен знать и понимать:  Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных  Предметная область анализа  Теоретические и прикладные основы анализа больших данных  Современные методы и инструментальные средства анализа больших данных  Современный опыт использования 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  Современная технологическая инфраструктура высокопроизводительных и распределенных вычислений  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  Облачные технологии, облачные сервисы  Методы оценки временных и стоимостных характеристик технологий больших данных  Технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии  Правила деловой переписки |  |
| - Специалист должен уметь:  Определять требования к поставщикам данных из гетерогенных источников  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  Оценивать 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 проекта  Основы управления аналитическими работами  Основы управления малыми аналитическими группами  Возможности имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных  Предметная область анализа  Теория принятия решений  Математическое моделирование  Теоретические и прикладные основы анализа больших данных  Современный опыт использования 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  Нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности  Статистические модели  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 идентификации шаблонов  Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма  Распределенный анализ данных  Анализ данных в реальном времени  Правила деловой переписки  Методы разработки отчетной аналитической документации |  |
|  | - Специалист должен уметь:  Планировать аналитические работы с использованием технологий больших данных  Проводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группами  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  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  Разъяснять заказчику результаты аналитической работы  Осуществлять 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 основе  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  Основы правового регулирования оборота и использования больших данных  Современные и перспективные средства визуализации и интерпретации больших данных  Системная инженерия  Машинное обучение  Математическое моделирование  Теория принятия решений  Высокопроизводительные и распределенные вычисления  Методы маркетинговых исследований  Методы сравнительного анализа  Основы инновационной деятельности и управления инновациями в сфере информационных технологий  Основы управления информационно-технологическими проектами  Основы управления проектными коллективами  Основы управления взаимоотношениями с партнерами  Показатели эффективности технологий больших данных  Основы охраны авторских прав и интеллектуальной собственности в сфере информационных технологий  Правила деловой переписки |  |
|  | - 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  Формировать коллектив исполнителей  Управлять взаимоотношениями с исполнителями и соисполнителями проектных работ  Управлять коллективом исполнителей  Управлять исполнением проектных работ  Разрабатывать научно-техническую документацию  Проводить согласование с заказчиком содержания и отчетной документации проектных работ  Проводить 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 основе  Локальные и глобальные потребности в создании новых и модернизации существующих сервисов на основе встроенной аналитики больших данных  Существующие и перспективные методы и программный инструментарий технологий больших данных  Принципы и методы управления защитой и обеспечением конфиденциальности больших данных  Существующий опыт разработки и использования продуктов и услуг на основе технологий больших данных  Особенности социально-экономической сферы использования сервисов на основе больших данных  Основы науки о сервисах  Сервисы и системы обслуживания  Сервис-ориентированная архитектура  Технология веб-сервисов и мобильных сервисов  Моделирование процессов предоставления услуг  Управление качеством обслуживания  Эффективность сервисов  Сервисные инновации  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  Основы управления проектными коллективами  Основы управления взаимоотношениями с партнерами  Показатели эффективности технологий больших данных  Основы охраны авторских прав и интеллектуальной собственности в сфере информационных технологий  Правила деловой переписки |  |
|  | - 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 программно-аппаратные компоненты и системы на основе технологий больших данных  Проводить математическое и информационное моделирование  Выполнять аналитические работы на основе технологий больших данных  Формировать коллектив исполнителей проектов в области больших данных  Управлять взаимоотношениями с исполнителями и соисполнителями проектных работ в области больших данных  Управлять коллективом исполнителей проектов в области больших данных  Управлять исполнением проектных работ в области больших данных  Разрабатывать научно-техническую документацию по проектам в области больших данных  Согласовывать с заказчиком содержание проектных работ в области больших данных и отчетную документацию по ним  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** | 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:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)  [Раздел ИЛ 2](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Рабочая_площадка_М2)  [Раздел ИЛ 3](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Модуль3)  [Раздел ИЛ 4](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль5)  [Раздел ИЛ 5](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.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:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)  [Раздел ИЛ 2](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Рабочая_площадка_М2)  [Раздел ИЛ 3](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Модуль3)  [Раздел ИЛ 4](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль5)  [Раздел ИЛ 5](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль6) | [26](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.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:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)  [Раздел ИЛ 2](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Рабочая_площадка_М2)  [Раздел ИЛ 3](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Модуль3)  [Раздел ИЛ 4](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль5)  [Раздел ИЛ 5](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль6) | [19](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.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:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)  [Раздел ИЛ 2](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Рабочая_площадка_М2)  [Раздел ИЛ 3](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Модуль3)  [Раздел ИЛ 4](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль5)  [Раздел ИЛ 5](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль6) | [18](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.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:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#РАБОЧАЯ_ПЛОЩАДКА_КОНКУРСАНТОВ_М1)  [Раздел ИЛ 2](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Рабочая_площадка_М2)  [Раздел ИЛ 3](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#Модуль3)  [Раздел ИЛ 4](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль5)  [Раздел ИЛ 5](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.xlsx#модуль6) | [18](file:///C:\Users\smirnov.e\Downloads\6%20Матрица.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 и  приемов. 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. API  должен позволить приложению пользователя получать доступ к регрессионной модели для  прогнозирования динамики изменения характеристик перевозок по кластерам на последующий период времени. |

|  |
| --- |
| 4.2 Разработка приложения |
| Разработайте приложение с графическим интерфейсом, которое должно c помощью  разработанного 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)