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

КОНКУРСНОЕ ЗАДАНИЕ КОМПЕТЕНЦИИ

«Машинное обучение и большие данные»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ этап Чемпионата по профессиональному мастерству «Профессионалы»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

регион проведения

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 4](#_Toc142037183)

[1.1. Общие сведения о требованиях компетенции 4](#_Toc142037184)

[1.2. Перечень профессиональных задач специалиста по компетенции «\_\_\_\_\_\_\_\_\_» 4](#_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) 8](#_Toc142037190)

[2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ 9](#_Toc142037191)

[2.1. Личный инструмент конкурсанта 9](#_Toc142037192)

[2.2.Материалы, оборудование и инструменты, запрещенные на площадке 9](#_Toc142037193)

[3. ПРИЛОЖЕНИЯ 9](#_Toc142037194)

**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

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** | 19% |
| 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** | 26% |
| 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** | 19% |
| 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:   * 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** | 20 |  |  |  |  | 20 |
| **2** |  | 22 |  |  |  | 23 |
| **3** |  |  | 28 |  |  | 28 |
| **4** |  |  |  | 19 |  | 18 |
| **5** |  |  |  |  | 11 | 11 |
| **Итого баллов за критерий/модуль** | | 20 | 23 | 28 | 18 | 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 дней

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

Оценка знаний участника должна проводиться через практическое выполнение Конкурсного задания. В дополнение могут учитываться требования работодателей для проверки теоретических знаний / оценки квалификации.

1.5.1. Разработка/выбор конкурсного задания

Конкурсное задание состоит из 5 модулей, включает обязательную к выполнению часть (инвариант) – 3 модуля, и вариативную часть – 2 модуля. Общее количество баллов конкурсного задания составляет 100.

1.5.2. Структура модулей конкурсного задания

Туризм – комплексная индустрия, которая охватывает весь мир и множество участников цепочки создания ценности. Каждому из них приходится принимать очень много решений, которые непосредственно влияют на качество услуг, рыночное взаимодействие игроков и совместимость сервисов друг с другом. До недавнего времени большинство решений принимались человеком, но мир стремительно меняется, и всем нужна более высокая гибкость. Необходимо принимать решения очень быстро и уметь менять их на ходу. Управлять этим процессом непросто.

Участникам необходимо применить технологии AI/ML для создания решения, которое позволит определять приоритеты развития туристской инфраструктуры, конкурировать за инвестиции и оценивать вклад туризма в экономику региона. Разработанное решение должно предоставить возможность в получении детальной картины о туристическом потоке и его ключевых показателях и оказывать помощь в принятии решений при планировании путешествия.

Таким образом, на данном чемпионате ВЫ — разработчик инструментов аналитического центра работы над проектом, цель которого заключается в визуализации показателей внутреннего туризма с помощью анализа данных.

ОСНОВНАЯ ЗАДАЧА — создать эффективную, визуально привлекательную и технически продвинутую аналитическую панель для специалистов туристической отрасли, а также разработать систему рекомендаций отелей для путешественников.

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

Время на выполнение модуля – 2 часа

**Описание модуля A:**

**Исходные файлы:**

1. Data (Папка, содержащий данные о туризме)
2. Description.docx ( Описание исходных файлов)
3. Модуль А.pdf (Инструкция к модулю А)
4. Папка Исходные файлы содержит исходные файлы (использовать в случае неудачи восстановления данных из дампов)

**Результаты работы:**

1. Дамп централизованной БД
2. Report\_А.html, Report\_А.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

В рамках данного модуля необходимо загрузить данные из различных источников. Изначально данные хранятся в реляционной и нереляционной базах данных (MongoDB, PostgreSQL) и в файле в формате tskv. Участникам требуется восстановить БД из дампов, загрузить все необходимые данные из восстановленных баз данных и файла для проведения предварительной обработки.

Проведите 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, обработка выбросов и т. д.

Создайте целевую базу данных для аналитики и загрузить в неё подготовленные данные.

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Папка Data должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля Б:**

**Исходные файлы:**

Модуль Б.pdf (Инструкция к модулю Б)

**Результаты работы:**

1. Интерактивная панель (дашборд)
2. Report\_Б.html или Report\_Б.docx(Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

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;
* янализировать 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 ключевых показателей для развития туристической инфраструктуры и бизнеса.

Разработанный дашборд должен позволить специалистам из сферы туризма получить такую информацию как:

* Топ 10 регионов по количеству отзывов с возможностью выбора способа сортировки по тональности;
* Доля туристического потока иностранных граждан в РФ;
* Типы отдыха в каждом ФО;
* План и факт притока туристов с возможностью фильтрации по регионам;
* Демонстрация основные показателей туристической инфраструктуры регионов на карте

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены все sql-запросы для визуального анализа данных и результаты их выполнения. Результаты работы должны состоять из отчетов в формате .html или .docx. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля В:**

**Исходные файлы:**

1. Модуль В.pdf (Инструкция к модулюВ)
2. geo-reviews-dataset-2023.csv — файл с отзовами туристов об отелях

**Результаты работы:**

1. Математическая модель
2. Report\_В.html или Report\_В.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

На данном этапе участникам необходимо разработать рекомендательную систему для бронирования отелей на основе отзывов гостей. Для решения задачи необходимо очистить и предобработать текстовые данные. Выполнить обработку отзывов (например, исправить орфографические ошибки, удалить стоп-слова) и дополнить набор дополнительными атрибутами для анализа (например, количество положительных и отрицательных слов в отзывах). Требуется провести анализ данных, чтобы определить наиболее важные признаки для оценки отелей. Необходимо выбрать алгоритм, который является наиболее эффективным в решении поставленной задачи. Для предложения пользователям наиболее подходящих отелей на основе их предпочтений требуется построить модель машинного обучения. Оценить качество модели с учетом выбранной метрики. Выполнить тестирование оптимизацию модели.

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Папка Data должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля Г:**

**Исходные файлы:**

1. Модуль Г.pdf (Инструкция к модулю Г)
2. Data (Результаты предыдущего модуля)

**Результаты работы:**

1. Data (Результирующие файлы)
2. Report\_Г.html или Report\_Г.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

На данном этапе участникам 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 участникам необходимо составить план и обеспечить автоматизацию процессов извлечения, трансформации и загрузки данных из различных источников компании в хранилище данных. Обеспечить автоматизацию загрузки актуальных данных из баз данных. В хранилище должны загружаться предобработанные данные, добавленные в базы данных за последнюю неделю. Действия ETL-пцесса должны подчиняется основным принципам построения DAG: отсутствие циклов, каждый узел имеет направление, возможность параллельных шагов.

В рамках данного модуля также требуется разработать программное решение для предложения потребителю наиболее подходящих ему отелей. Программа должна предложить пользователям наиболее подходящие варианты для бронирования отеля на основе их предпочтений.

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Папка Data должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля Д:**

**Исходные файлы:**

1. Модуль Д.pdf (Инструкция к модулю Д)
2. Data (Результаты предыдущего модуля)

**Результаты работы:**

1. Data (Результирующие файлы)
2. Project.pdf (Презентация)
3. Readme.txt (Дополнительные комментарии)
4. Report\_ Д.docx (Программная документация)

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 участникам необходимо презентовать полученные решения заказчику или другим заинтересованным сторонам, чтобы помочь им принимать обоснованные решения на основе данных. 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. Разрабатывать презентацию рекомендуется в Power Point или аналогичной среде. Опишите перспективы улучшения Вашего решения. Подготовить устный доклад по результатам своей работы, включающие основные результаты по каждому модулю и выводы (не более 5 минут). Устные представления докладов – за 40 мин до окончания сессии.

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

2.1 Личный инструмент конкурсанта

1. Участники могут использовать защиту для ушей.
2. 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. Инструкция по охране труда

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** | 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** | 20 |  |  |  |  | 20 |
| **2** |  | 22 |  |  |  | 23 |
| **3** |  |  | 28 |  |  | 28 |
| **4** |  |  |  | 19 |  | 18 |
| **5** |  |  |  |  | 11 | 11 |
| **Итого баллов за критерий/модуль** | | 20 | 23 | 28 | 18 | 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[[3]](#footnote-3): 10 ч.

Количество конкурсных дней: 3 дней

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

Оценка знаний участника должна проводиться через практическое выполнение Конкурсного задания. В дополнение могут учитываться требования работодателей для проверки теоретических знаний / оценки квалификации.

1.5.1. Разработка/выбор конкурсного задания

Конкурсное задание состоит из 5 модулей, включает обязательную к выполнению часть (инвариант) – 3 модуля, и вариативную часть – 2 модуля. Общее количество баллов конкурсного задания составляет 100.

1.5.2. Структура модулей конкурсного задания

Туризм – комплексная индустрия, которая охватывает весь мир и множество участников цепочки создания ценности. Каждому из них приходится принимать очень много решений, которые непосредственно влияют на качество услуг, рыночное взаимодействие игроков и совместимость сервисов друг с другом. До недавнего времени большинство решений принимались человеком, но мир стремительно меняется, и всем нужна более высокая гибкость. Необходимо принимать решения очень быстро и уметь менять их на ходу. Управлять этим процессом непросто.

Участникам необходимо применить технологии AI/ML для создания решения, которое позволит определять приоритеты развития туристской инфраструктуры, конкурировать за инвестиции и оценивать вклад туризма в экономику региона. Разработанное решение должно предоставить возможность в получении детальной картины о туристическом потоке и его ключевых показателях и оказывать помощь в принятии решений при планировании путешествия.

Таким образом, на данном чемпионате ВЫ — разработчик инструментов аналитического центра работы над проектом, цель которого заключается в визуализации показателей внутреннего туризма с помощью анализа данных.

ОСНОВНАЯ ЗАДАЧА — создать эффективную, визуально привлекательную и технически продвинутую аналитическую панель для специалистов туристической отрасли, а также разработать систему рекомендаций отелей для путешественников.

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

Время на выполнение модуля – 2 часа

**Описание модуля A:**

**Исходные файлы:**

1. Data (Папка, содержащий данные о туризме)
2. Description.docx ( Описание исходных файлов)
3. Модуль А.pdf (Инструкция к модулю А)
4. Папка Исходные файлы содержит исходные файлы (использовать в случае неудачи восстановления данных из дампов)

**Результаты работы:**

1. Дамп централизованной БД
2. Report\_А.html, Report\_А.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

В рамках данного модуля необходимо загрузить данные из различных источников. Изначально данные хранятся в реляционной и нереляционной базах данных (MongoDB, PostgreSQL) и в файле в формате tskv. Участникам требуется восстановить БД из дампов, загрузить все необходимые данные из восстановленных баз данных и файла для проведения предварительной обработки.

Проведите 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, обработка выбросов и т. д.

Создайте целевую базу данных для аналитики и загрузить в неё подготовленные данные.

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Папка Data должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля Б:**

**Исходные файлы:**

Модуль Б.pdf (Инструкция к модулю Б)

**Результаты работы:**

1. Интерактивная панель (дашборд)
2. Report\_Б.html или Report\_Б.docx(Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

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;
* янализировать 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 ключевых показателей для развития туристической инфраструктуры и бизнеса.

Разработанный дашборд должен позволить специалистам из сферы туризма получить такую информацию как:

* Топ 10 регионов по количеству отзывов с возможностью выбора способа сортировки по тональности;
* Доля туристического потока иностранных граждан в РФ;
* Типы отдыха в каждом ФО;
* План и факт притока туристов с возможностью фильтрации по регионам;
* Демонстрация основные показателей туристической инфраструктуры регионов на карте

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены все sql-запросы для визуального анализа данных и результаты их выполнения. Результаты работы должны состоять из отчетов в формате .html или .docx. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля В:**

**Исходные файлы:**

1. Модуль В.pdf (Инструкция к модулюВ)
2. geo-reviews-dataset-2023.csv — файл с отзовами туристов об отелях

**Результаты работы:**

1. Математическая модель
2. Report\_В.html или Report\_В.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

На данном этапе участникам необходимо разработать рекомендательную систему для бронирования отелей на основе отзывов гостей. Для решения задачи необходимо очистить и предобработать текстовые данные. Выполнить обработку отзывов (например, исправить орфографические ошибки, удалить стоп-слова) и дополнить набор дополнительными атрибутами для анализа (например, количество положительных и отрицательных слов в отзывах). Требуется провести анализ данных, чтобы определить наиболее важные признаки для оценки отелей. Необходимо выбрать алгоритм, который является наиболее эффективным в решении поставленной задачи. Для предложения пользователям наиболее подходящих отелей на основе их предпочтений требуется построить модель машинного обучения. Оценить качество модели с учетом выбранной метрики. Выполнить тестирование оптимизацию модели.

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Папка Data должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля Г:**

**Исходные файлы:**

1. Модуль Г.pdf (Инструкция к модулю Г)
2. Data (Результаты предыдущего модуля)

**Результаты работы:**

1. Data (Результирующие файлы)
2. Report\_Г.html или Report\_Г.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

На данном этапе участникам 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 участникам необходимо составить план и обеспечить автоматизацию процессов извлечения, трансформации и загрузки данных из различных источников компании в хранилище данных. Обеспечить автоматизацию загрузки актуальных данных из баз данных. В хранилище должны загружаться предобработанные данные, добавленные в базы данных за последнюю неделю. Действия ETL-пцесса должны подчиняется основным принципам построения DAG: отсутствие циклов, каждый узел имеет направление, возможность параллельных шагов.

В рамках данного модуля также требуется разработать программное решение для предложения потребителю наиболее подходящих ему отелей. Программа должна предложить пользователям наиболее подходящие варианты для бронирования отеля на основе их предпочтений.

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут

представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Папка Data должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов.

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

Время на выполнение модуля – 2 часа

**Описание модуля Д:**

**Исходные файлы:**

1. Модуль Д.pdf (Инструкция к модулю Д)
2. Data (Результаты предыдущего модуля)

**Результаты работы:**

1. Data (Результирующие файлы)
2. Project.pdf (Презентация)
3. Readme.txt (Дополнительные комментарии)
4. Report\_ Д.docx (Программная документация)

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 участникам необходимо презентовать полученные решения заказчику или другим заинтересованным сторонам, чтобы помочь им принимать обоснованные решения на основе данных. 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. Разрабатывать презентацию рекомендуется в Power Point или аналогичной среде. Опишите перспективы улучшения Вашего решения. Подготовить устный доклад по результатам своей работы, включающие основные результаты по каждому модулю и выводы (не более 5 минут). Устные представления докладов – за 40 мин до окончания сессии.

2. СПЕЦИАЛЬНЫЕ ПРАВИЛА КОМПЕТЕНЦИИ*[[4]](#footnote-4)*

2.1 Личный инструмент конкурсанта

1. Участники могут использовать защиту для ушей.
2. 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. 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)
3. *Указывается суммарное время на выполнение всех модулей КЗ одним конкурсантом.* [↑](#footnote-ref-3)
4. *Указываются особенности компетенции, которые относятся ко всем возрастным категориям и чемпионатным линейкам без исключения.* [↑](#footnote-ref-4)