**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ**

|  |  |
| --- | --- |
| **Код модуля** | **Модуль** |
| М.1.7 | Инжиниринг данных |

**Барнаул, 2021**

**1. СТРУКТУРА И ОБЪЕМ МОДУЛЯ** ИНЖИНИРИНГ ДАННЫХ

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Перечень дисциплин модуля в последовательности их освоения** | **Объем дисциплин модуля и всего модуля в зачетных единицах и часах** | **Форма итоговой промежуточной аттестации по дисциплинам модуля и в целом по модулю** |
| 1. | Инжиниринг данных | *3 з.е. /108 час.* | *Зачет* |
| **ИТОГО по модулю:** | | ***3 з.е. /108 час.*** |  |

1. **ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО МОДУЛЮ**

Не предусмотрено

**3. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ**

ИНЖИНИРИНГ ДАННЫХ

**Модуль** М 1.7 Инжиниринг данных

Оценочные материалы составлены автором(ами):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ п/п** | **Фамилия Имя Отчество** | **Ученая степень, ученое звание** | **Должность** | **Подразделение** |
| 1 | Хворова Л.А. | К. тех. н., доцент | Доцент каф. ТКПМ | Кафедра ТКПМ |

1. **ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ** ИНЖИНИРИНГ ДАННЫХ

Таблица 1.1

|  |  |  |
| --- | --- | --- |
| **Код и наименование компетенций, формируемые с участием дисциплины** | **Планируемые индикаторы достижения компетенций** | **Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине** |
| **1** | **2** | **3** |
| ОПК-3. Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями | ОПК-3.1. Знать: принципы, методы и средства анализа и структурирования профессиональной информации.  ОПК-3.2. Уметь: анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров.  ОПК-3.3. Владеть: методами подготовки научных докладов, публикаций и аналитических обзоров с обоснованными выводами и рекомендациями. | 1) Контрольная работа  2) Домашние работы  3) Выполнение практических работ  4) Зачет |

Таблица 1.2

|  |  |  |  |
| --- | --- | --- | --- |
| **Код и наименование компетенций, формируемые с участием дисциплины** | **Индикаторы достижения компетенции** | **Планируемые результаты обучения** | **Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине** |
| **1** | **2** | **3** | **4** |
| ПК-3. Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач | ПК-3.1. Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области  ПК-3.2. Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области  ПК-3.3. Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий | ПК-3.1. З-1. Знает классы методов и алгоритмов машинного обучения  ПК-3.1. У-1. Умеет ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения  ПК 3.2. З-1. Знает методы и критерии оценки качества моделей машинного обучения  ПК 3.2. У-1. Умеет определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области  ПК-3.3. З-1. Знает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий  ПК-3.3. У-1. Умеет разрабатывать унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий | 1) Контрольная работа  2) Домашние работы  3) Выполнение практических работ  4) Зачет |
| ПК-8. Способен разрабатывать и модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности в различных предметных областях | ПК-8.1. Разрабатывает программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях  ПК-8.2. Модернизирует программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях | ПК-8.1. З-1. Знает новые научные принципы и методы разработки программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач в различных предметных областях  ПК-8.1. У-1. Умеет разрабатывать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности для решения профессиональных задач в различных предметных областях  ПК-8.2. З-1. Знает особенности модернизации программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач в различных предметных областях  ПК-8.2. У-1. Умеет модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности для решения профессиональных задач в различных предметных областях | 1) Контрольная работа  2) Домашние работы  3) Выполнение практических работ  4) Зачет |

1. **ВИДЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ, ВКЛЮЧАЯ МЕРОПРИЯТИЯ ТЕКУЩЕЙ АТТЕСТАЦИИ**
   1. **Распределение объема времени по видам учебной работы**

Таблица 2

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Наименование дисциплины модуля** Инжиниринг данных | **Объем времени, отведенный на освоение дисциплины модуля** | | | | | | | | |
| **Аудиторные занятия, час.** | | | | **Промежуточная аттестация** (форма итогового контроля) | **Контактная работа** (час.) | **Самостоятельная работа студента, включая текущую аттестацию** (час.) | **Всего по дисциплине** | |
| **Занятия лекционного типа** | **Практические работы** | **Лабораторные работы** | **Всего** | **Час.** | **Зач. ед.** |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|  | Инжиниринг данных | 18 | 18 | 0 | 36 | Зачет | 41.65 | 66.35 | 108 | 3 |

* 1. **Виды СРС, количество и объем времени на контрольно-оценочные мероприятия СРС по дисциплине**

Контрольно-оценочные мероприятия СРС включают самостоятельное изучение материала, подготовку к аудиторным занятиям и мероприятиям текущего контроля, выполнение и оформление внеаудиторных мероприятий текущего контроля и подготовку к мероприятиям промежуточного контроля.

Таблица 3

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Вид самостоятельной работы студента по дисциплине модуля** | **Количество контрольно-оценочных мероприятий СРС** | **Объем контрольно-оценочных мероприятий СРС (час.)** |
| 1. | Подготовка к аудиторным занятиям и мероприятиям текущего контроля: лекционным, практическим занятиям. |  | 13,5 час. |
| 2 | Выполнение и оформление мероприятий текущего контроля: |  |  |
| 2.1 | Контрольная работа | 1 | 5 час. |
| 2.2 | Домашняя работа | 2 | 10 час. |
| 3. | Подготовка к зачету | зачет | 12 час. |
| 4. | Самостоятельное изучение материала |  | 25,85 час. |
| Итого на СРС по дисциплине: | | | 66,35 час. |

1. **ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)**

**Процедуры текущей и промежуточной аттестации по дисциплине**

|  |  |  |
| --- | --- | --- |
| **1.Лекции**: **коэффициент значимости совокупных результатов лекционных занятий – 0.5** | | |
| **Текущая аттестация на лекциях** | **Сроки – семестр,**  **учебная неделя** | **Максимальная оценка в баллах** |
| *Контрольная работа* | 1 сем. | *70* |
| *Самостоятельное изучение материала* | 1 сем. | *30* |
| **Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5** | | |
| **Промежуточная аттестация по лекциям –** *Зачет*  **Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5** | | |
| **2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5** | | |
| **Текущая аттестация на практических/семинарских занятиях** | **Сроки – семестр,**  **учебная неделя** | **Максимальная оценка в баллах** |
| *Выполнение и оформление практических работ* | 1 сем. | 50 |
| *Домашняя работа №1* | 1 сем. | 25 |
| *Домашняя работа №2* | 1 сем. | 25 |
| **Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1** | | |
| **Промежуточная аттестация по практическим/семинарским занятиям–** не предусмотрена  **Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – 0** | | |
| **3. Лабораторные занятия: Не предусмотрены**  **коэффициент значимости совокупных результатов лабораторных занятий – 0** | | |

1. **КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ** 
   1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

**Критерии оценивания учебных достижений обучающихся**

|  |  |
| --- | --- |
| **Результаты обучения** | **Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам** |
| Знания | Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью. |
| Умения | Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью. |
| Опыт /владение | Студент демонстрирует опыт в области изучения на уровне указанных индикаторов. |
| Личностные качества | Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов.  Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения.  Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения. |

* 1. Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

**Шкала оценивания достижения результатов обучения (индикаторов) по уровням**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Характеристика уровней достижения результатов обучения (индикаторов)** | | | | |
| **№ п/п** | **Содержание уровня выполнения критерия оценивания результатов обучения**  **(выполненное оценочное задание)** | **Шкала оценивания** | | |
| **Традиционная характеристика уровня** | | **Качественная характеристика уровня** |
| 1. | Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет | Отлично  (80-100 баллов) | Зачтено | Высокий (В) |
| 2. | Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения | Хорошо  (60-79 баллов) | Средний (С) |
| 3. | Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания | Удовлетворительно  (40-59 баллов) | Пороговый (П) |
| 4. | Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка | Неудовлетворительно  (менее 40 баллов) | Не зачтено | Недостаточный (Н) |
| 5. | Результат обучения не достигнут, задание не выполнено | Недостаточно свидетельств для оценивания | | Нет результата |

1. **СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ**

Задания по контрольно-оценочным мероприятиям в рамках текущей и промежуточной аттестации должны обеспечивать освоение и достижение результатов обучения (индикаторов) и предметного содержания дисциплины на соответствующем уровне.

**5.1. Описание контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля**

**5.1.1. Практические занятия**

|  |  |
| --- | --- |
| **Номер занятия** | **Примерный перечень тем практических занятий** |
| 1 | Библиотеки для работы с данными в Python: numpy, pandas. |
| 2 | Работа с текстовыми файлами разных форматов в Python: CSV, JSON, HTML. |
| 3 | Работа с базами данных в Python. |
| 4 | Работа с изображениями, видео и звуковыми файлами в Python. |
| 5 | Работа с файлами для хранения больших данных в Python. |
| 6 | Работа с графами знаний в Python. |
| 7 | Создание собственных наборов данных в Python. Очистка и подготовка данных. |
| 8 | Работа с данными в Apache Spark. |
| 9 | Использование SQL в Apache Spark. |

**5.1.2. Лабораторные занятия** *Не предусмотрено*

**5.1.3. Курсовая работа / Курсовой проект** *Не предусмотрено*

**5.1.4. Контрольная работа**

**Примерная тематика** контрольных работ***:***

1. Работа с данными в различных форматах.

**Примерные задания** в составе контрольных работ***:***

**Задание 1.** Формат данных CSV может быть использован как замена:

* **реляционным СУБД**
* нереляционным СУБД
* документо-ориентированным СУБД

**Задание 2.** Что обычно выступает разделителем столбцов в формате CSV:

* **Запятая**
* Точка с запятой
* Перенос строки

**Задание 3.** Какие элементы не используются в формате JSON в качестве структурных элементов:

* **Теги**
* Кавычки
* Двоеточие

**Задание 5.** Для какого языка программирования впервые использовался формат JSON:

* **JavaScript**
* Java
* Python

**Задание 6.** В каком виде хранятся данные в MongoDB:

* **BSON**
* JSON
* XML

**Задание 7.** Выберите наиболее подходящие характеристики MongoDB:

* **Ключ-значение, неструктурированность данных, нереляционные свойства**
* Ключ-значение, неструктурированность данных, реляционные свойства
* SQL, неструктурированность данных, реляционные свойства

**Задание 8.** Преобразуйте следующий код в формат JSON:

<companies>

<company>

<company-id>7707040070</company-id>

<name>Якорьбанк</name>

<shortname>Якорьбанк</shortname>

<name-other>Якорьбанк, платёжное устройство</name-other>

<address>Россия, Республика Татарстан, Зеленодольский район, село Нурлаты, улица Гагарина, 46</address>

<phone>

<type>phone</type>

<number>+7 (800) 999-99-90</number>

</phone>

<url>http://www.yakorbank.ru/</url>

<working-time>будни 8:30-18:00, сб 9:00-14:30</working-time>

<rubric-id>184106974</rubric-id>

<actualization-date>23.09.2019</actualization-date>

<coordinates>

<lat>55.616051</lat>

<lon>48.295532</lon>

</coordinates>

</company>

</companies>

**5.1.5. Домашняя работа**

**Примерная тематика** домашних работ***:***

*Домашняя работа №1:*

Создание собственного набора данных.

*Домашняя работа №2:*

Обработка данных в Apache Spark.

**Примерные задания** в составе домашних работ***:***

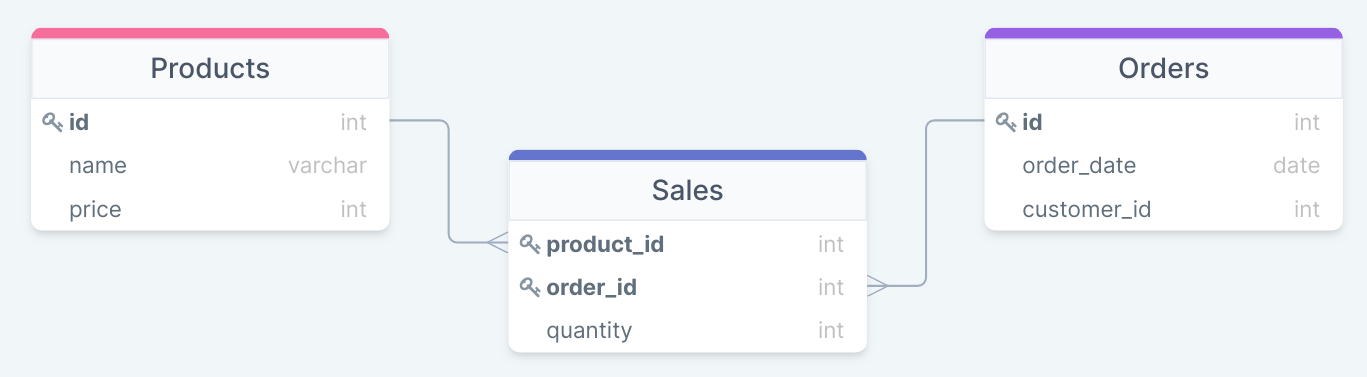
1. Подготовить собственный набор данных. Выберите задачу в одном из направлений создания системы искусственного интеллекту (компьютерное зрение, обработка естественного языка) и подготовьте для этой задачи набор данных для обучения с учителем. Соберите и очистите данные, проведите разметку. Готовый набор данных разместите на одной из общедоступных платформ для хранения данных по своему выбору. Подготовьте документацию к созданному набору данных.

Перечень задач, для которых рекомендуется подготовить набор данных:

* Классификация объектов на изображениях.
* Определение положения объектов на изображениях.
* Определение положения объектов в видео.
* Классификация текста на русском языке.
* Определение эмоциональной окраски текста на русском языке.

1. Создайте набор данных в Apache Spark и проведите его исследование с помощью Spark DataFrame API.

Схема данных выглядит следующим образом:



Онлайн-школа продает образовательные продукты: онлайн-курсы, книги, семинары и т.п. Описание и стоимость продуктов содержится в таблице **Products**. Когда клиент что-то покупает, создается заказ, который заносится в таблицу **Orders**. Заказ может содержать несколько продуктов, перечень продуктов в заказах содержится в таблице **Sales**.

Таблица **Products** - продукты онлайн-школы:

* **id** - идентификатор продукта
* **name** - название продукта
* **price** - стоимость продукта

Таблица **Oredrs** - заказы:

* **id** - идентификатор заказа
* **order\_date** - дата заказа
* **customer\_id** - идентификатор заказчика (таблица с заказчиками не создается для упрощения примера)

Таблица **Sales** - продажи:

* **product\_id** - идентификатор продукта, ссылка на таблицу Products, поле id
* **order\_date** - идентификатор заказа, ссылка на таблицу Orders, поле id
* **quantity** - количество продуктов в заказе

Ноутбук в облачной платформе Colab с заготовкой кода для домашней работы – <https://colab.research.google.com/drive/1MLilHIZ2CcBbCp_U7wKs_CAPmnx2F7Q6?usp=sharing>

**Задания для анализа:**

* Выведите список продуктов, которые не были проданы ни разу
* Определите, сколько продуктов любого типа было продано по дням.
* Определить, какая выручна от продуктов любого типа была получена по дням.

**5.1.6. Расчетная работа / Расчетно-графическая работа***. Не предусмотрено*

**5.1.7. Реферат / эссе / творческая работа** *Не предусмотрено*

**5.1.8. Проектная работа** *Не предусмотрено*

**5.1.9. Деловая (ролевая) игра /** **Дебаты / Дискуссия / Круглый стол** *Не предусмотрено*

**5.1.10. Кейс-анализ** *Не предусмотрено*

**5.2. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля**

**5.2.1. Зачет в форме независимого тестового контроля (НТК).** НТК по дисциплине модуля не проводится.

**5.2.2. Зачет в традиционной форме** (устные /письменные ответы на вопросы)

1. Библиотека pandas в Python.
2. Работа с данными в формате CSV в Python.
3. Работа с данными в формате JSON в Python.
4. Работа с данными в формате HTML в Python.
5. Работа с изображениями в Python.
6. Работа с видео в Python.
7. Работа с аудио в Python.
8. Работа с Parquet в Python.
9. Работа с графами знаний в Python.
10. Этапы и инструменты создания наборов данных для машинного обучения.
11. Загрузка данных с Web-сайтов.
12. Загрузка данных из социальных сетей.
13. Методы и инструменты подготовки данных.
14. Методы и инструменты очистки данных.
15. Разметка данных.
16. Общедоступные платформы для хранения данных.
17. Архитектура центров обработки данных.
18. Кластеры для параллельных и распределенных вычислений.
19. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop.
20. Распределенная файловая система HDFS.
21. Распределенная обработка данных в Apache Spark.
22. Работа с данными с использованием Apache Spark DataFrame.
23. Источники данных для Apache Spark DataFrame.
24. Обработка данных в Apache Spark DataFrame.
25. Использование SQL в Apache Spark DataFrame.