Федеральное государственное бюджетное образовательное учреждение высшего образования «Алтайский государственный университет»

УТВЕРЖДАЮ

Директор Института математики

и информационных технологий

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Е.В. Журавлев

«29» октября 2021 г.

**РАБОЧАЯ ПРОГРАММА МОДУЛЯ**

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

**Барнаул, 2021**

|  |  |
| --- | --- |
| **Перечень сведений о рабочей программе модуля** | **Учетные данные** |
| **Образовательная программа**  Инженерия искусственного интеллекта | **Код ОП**  09.04.01 |
| **Направление подготовки**  Информатика и вычислительная техника | **Код направления и уровня подготовки**  09.04.01 |

**Рекомендовано учебно-методическим советом Института математики и информационных технологий**

Протокол №3 от 28.09.2021 г.

1. **ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Инжиниринг данных** 
   1. **Аннотация содержания модуля**

Модуль «Инжиниринг данных» состоит из одноименной дисциплины, посвященной подготовке данных для моделей машинного обучения. Рассматриваются особенности работы с данными в различных форматах на языке Python. Уделяется внимание инструментам и технологиям загрузки данных из интернета и социальных сетей. Подробно изучаются методы очистки данных и соответствующие библиотеки на Python.

* 1. **Структура и объем модуля**

Таблица 1

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Перечень дисциплин модуля в последовательности их освоения** | **Объем дисциплин модуля и всего модуля в зачетных единицах и часах** |
| 1. | Инжиниринг данных | 3 з.е. / 108 ч. |
| **ИТОГО по модулю:** | | **3 з.е. / 108 ч.** |

* 1. **Последовательность освоения модуля в образовательной программе**

|  |  |
| --- | --- |
| **Пререквизиты модуля** | *отсутствуют* |
| **Постреквизиты и корреквизиты модуля** | *отсутствуют* |

* 1. **Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю**

Таблица 2.1

|  |  |  |
| --- | --- | --- |
| **Перечень дисциплин модуля** | **Код и наименование компетенции** | **Планируемые индикаторы достижения компетенций** |
| 1 | 2 | 3 |
| Инжиниринг данных | 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. |

Таблица 2.2

|  |  |  |  |
| --- | --- | --- | --- |
| **Перечень дисциплин модуля** | **Код и наименование компетенции** | **Индикаторы достижения компетенции** | **Планируемые результаты обучения** |
| 1 | 2 | 3 | 4 |
| Инжиниринг данных | 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 |
| 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. **Форма обучения**

Обучение по дисциплине модуля может осуществляться в очной форме.

**РАЗДЕЛ 2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН МОДУЛЯ**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

ИНЖИНИРИНГ ДАННЫХ

Рабочая программа дисциплины составлена авторами:

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

**Рекомендовано учебно-методическим советом Института математики и информационных технологий**

Протокол №3 от 28.09.2021 г.

**2. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ** Инжиниринг данных

**2.1. Технологии обучения, используемые при изучении дисциплины модуля**

* Традиционная (репродуктивная) технология;
* Смешанная модель обучения с использованием онлайн-курса;
* Исключительно электронного обучения с использованием онлайн-курса;

**2.2. Содержание дисциплины**

Таблица 1.3

|  |  |  |
| --- | --- | --- |
| **Код**  **раздела, темы** | **Раздел, тема**  **дисциплины** | **Содержание** |
| 1 | Работа с данными в Python. | Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python. Работа с изображениями, видео и звуковыми файлами. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний. |
| 2 | Подготовка данных для систем машинного обучения. | Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных. Общедоступные платформы для хранения данных. Подход Data-Centric AI. |
| 3 | Параллельная и распределенная обработка данных. | Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop, HDFS. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame. |

2.3. Программа дисциплины реализуется на государственном языке Российской Федерации.

**2.4. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ Инжиниринг данных**

**Электронные ресурсы (издания)**

1. Онлайн-курс “Прикладное программирование на языке Python”. URL: https://openedu.ru/course/urfu/PYAP/ (дата обращения: 05.10.2021).
2. Real Python Tutorials. URL: https://realpython.com/ (дата обращения: 05.10.2021).
3. Jake VanderPlas. Python Data Science Handbook. URL: <https://jakevdp.github.io/PythonDataScienceHandbook/> (дата обращения: 05.10.2021).
4. Jeroen Janssens. Data Science at the Command Line. URL: <https://www.datascienceatthecommandline.com/> (дата обращения: 05.10.2021).
5. Andrew Ng. MLOps: From Model-centric to Data-centric AI. URL: <https://www.deeplearning.ai/wp-content/uploads/2021/06/MLOps-From-Model-centric-to-Data-centric-AI.pdf> (дата обращения: 05.10.2021).
6. Spark SQL, DataFrames and Datasets Guide. URL: <https://spark.apache.org/docs/latest/sql-programming-guide.html> (дата обращения: 05.10.2021).

**Профессиональные базы данных, информационно-справочные системы**

1. Applied Science & Technology Source. EBSCO publishing <http://search.ebscohost.com>
2. Wiley Online Library http://onlinelibrary.wiley.com/
3. Гугл Академия https://scholar.google.ru/

**Материалы для лиц с ОВЗ**

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

**Базы данных, информационно-справочные и поисковые системы**

1. Academic Search Ultimate EBSCO publishing – <http://search.ebscohost.com>
2. eBook Collections Springer Nature – <https://link.springer.com/>
3. Гугл Академия – <https://scholar.google.ru/>
4. СПС КонсультантПлюс (инсталлированный ресурс АлтГУ или <http://www.consultant.ru/>)
5. Электронная база данных «Scopus» (<http://www.scopus.com>)
6. Электронная библиотечная система Алтайского государственного университета (http://elibrary.asu.ru/)
7. Электронный научный архив УрФУ <https://elar.urfu.ru/>
8. Зональная научная библиотека (УрФУ) - <http://lib.urfu.ru/>
9. Портал информационно-образовательных ресурсов УрФУ <https://study.urfu.ru/>
10. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
11. Университетская библиотека ONLINE – <https://biblioclub.ru/>
12. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks) <http://www.bibliocomplectator.ru/available>
13. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
14. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>

**2.5. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

Инжиниринг данных

**Сведения об оснащенности дисциплины специализированным и лабораторным оборудованием и программным обеспечением**

Таблица 3.1

|  |  |  |  |
| --- | --- | --- | --- |
| **№  п/п** | **Виды занятий** | **Оснащенность специальных помещений и помещений для самостоятельной работы** | **Перечень программного обеспечения** |
| 1 | Лекции; Практические занятия. | Компьютерный класс.  Мультимедийный проектор с экраном.  Сетевое оборудование.  Локальная сеть с выходом в глобальную сеть Internet. | В курсе используется бесплатно распространяемое программное обеспечение:   1. Python – https://www.python.org/ 2. Веб - среда разработки для языка программирования Python: google colab - https://colab.research.google.com/ 3. Anaconda solution - https://www.anaconda.com/ 4. Система распределенной обработки данных Apache Spark – https://spark.apache.org/ |