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.1.13 | Спортивный анализ 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 Спортивный анализ данных** 
   1. **Аннотация содержания модуля**

Модуль «Спортивный анализ данных» состоит из одноименной дисциплины.

Цель освоения дисциплины «Спортивный анализ данных» ознакомить студентов с современной платформой для проведения соревнований "Kaggle .com". Студенты узнают, как использовать различные алгоритмы и методы анализа данных для решения конкретных прикладных задач. На практике рассматриваются все типы задач анализа данных: анализ табличных данных, анализ временных рядов, обработка естественного языка, обработка изображений. В курсе рассматриваются методы анализа базовых обученных моделей, выбора и проверки новых возможностей, оптимальные методы поиска лучшего алгоритма решения задачи.

* 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. | Спортивный анализ данных | 6 з.е. / 216 ч. |
| **ИТОГО по модулю:** | | **6 з.е. / 216 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 |
| Спортивный анализ 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 |
| Спортивный анализ данных | ПК-6. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях | ПК-6.1. Осуществляет руководство проектом по построению комплексных систем на основе аналитики больших данных в различных отраслях | ПК-6.1. З-1. Знает методологию и принципы руководства проектом по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных  ПК-6.1. З-2. Знает специфику сфер и отраслей, для которых реализуется проект по аналитике больших данных  ПК-6.1. У-1. Умеет решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных  ПК-6.1. У-2. Умеет сосредотачивать внимание на целях, достижение которых обеспечивает большую отдачу и сильное воздействие  ПК-6.1. У-3. Умеет формировать матрицу приоритетов, включая критерии отбора проектов для реализации |

* 1. **Форма обучения**

Обучение по дисциплине модуля может осуществляться в очной форме.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

СПОРТИВНЫЙ АНАЛИЗ ДАННЫХ

Рабочая программа дисциплины составлена авторами:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ п/п** | **Фамилия Имя Отчество** | **Ученая степень, ученое звание** | **Должность** | **Подразделение** |
| 1 | Пономарев И.В. | К.ф.-м.н., доцент | Доцент каф. математического анализа | Кафедра математического анализа |

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

Протокол №3 от 28.09.2021 г.

**2. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ Спортивный анализ данных**

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

* Традиционная (репродуктивная) технология *(ориентирована на передачу знаний и умений, обеспечивающая усвоение обучающимися содержания обучения, проверку и оценку его качества на репродуктивном уровне);*
* Смешанная модель обучения с использованием онлайн-курса УрФУ;

**2.2. Содержание дисциплины**

Таблица 1.3

|  |  |  |
| --- | --- | --- |
| **Код**  **раздела, темы** | **Раздел, тема**  **дисциплины\*** | **Содержание** |
| 1 | Обзор инструментов применяемых для анализа данных и машинного обучения | Знакомство с инструментами: jupyter notebook, google colaboratory, kaggle kernel. Базовый функционал и приемы работы. Основные понятия Data Science и Machine Learning |
| 2 | Обзор библиотеки Numpy | Применение библиотеки numpy для операция над тензорами и решения задач линейной алгебры и генерации тензоров различной размерности по заданным правилам |
| 3 | Работа с библиотекой Pandas (Matplotlib, seaborn) | Библиотека для работы с табличными данными pandas: чтение и запись файлов, методы просмотра данных, индексация, срезы, фильтрация, агрегация и сортировка данных, визуализация данных, разведывательный анализ данных, базовые статистики. |
| 4 | Модели машинного обучения для решения задач классификации | Постановка задачи классификации. Изучение алгоритмов классификации в машинном обучении: логистическая регрессия, машины опорных векторов, наивный Байес, К ближайших соседей, деревья решений, ансамблевые модели |
| 5 | Модели машинного обучения для решения задач регрессии | Постановка задачи регрессии. Изучение алгоритмов регрессии в машинном обучении: линейная регрессия, машины опорных векторов, К ближайших соседей, деревья решений, ансамблевые модели. Регуляризация моделей машинного обучения |
| 6 | Метрики качества при решении задач классификации и регрессии | Изучение метрик качества моделей машинного обучения для классификации и регрессии. Рассмотрение сильных и слабых сторон различных метрик для различных задач. Применение изученных метрик |
| 7 | Предварительная обработка данных при решении задач анализа данных | Методы предварительной обработки: заполнение пропусков в данных, обработка выбросов в данных. |
| 8 | Преобразование и создание новых признаков для решения задач анализа данных | Изменение непрерывных типов данных: логарифмирование, построение полиномов, дискретизация данных, стандартизация и нормализация данных. Различные методы преобразование категориальных данных. |
| 9 | Методы выбора лучших признаков для решения задач классификации и регрессии | Методы отбора лучших для решения поставленной задачи признаков в задачах классификации и регрессии: на основе статистических подходов и методов машинного обучения |
| 10 | Методы выбора лучших моделей и алгоритмов для решения поставленных задач анализа данных | Приемы отбора лучших моделей для решения поставленных задач. Различные методики разбиения выборок для более стабильной и правдоподобной оценки |
| 11 | Методы подбора гипер параметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных | Инструменты построения конвейеров обработки данных в машинном обучении. Методы и инструменты автоматического подбора гипер параметров моделей машинного обучения, а также подбора самих моделей и методов предварительной обработки |
| 12 | Работа с временными рядами | Понятие временного ряда и его составляющие элементы. Методы прогнозирования на основе эконометрических подходов. Методы прогнозирования на основе регрессионного подхода и моделей машинного обучения. Генерация новых временных признаков для временного ряда. |
| 13 | Задачи обучения без учителя: понижение размерности | Обзор задач обучения без учителя и их применимость на практике. Методы понижения размерности и их практические приложения |
| 14 | Задачи обучения без учителя: кластеризация | Обзор методов кластеризации и их сравнительные характеристики. Методы оценки оптимального числа кластеров. Кластеризация для сегментации и анализа. Кластеризация, как метод сжатия информации и снижения размерности. |
| 15 | Задачи обучения без учителя: поиск аномалий | Постановка задачи поиска аномалий и выявления новизны. Методы поиска аномалий на основе подходов машинного обучения. |
| 16 | Обработка естественного языка: предварительная обработка | Знакомство с задачами обработки естественного языка.: Методы первичной предварительной обработки текстов: очистка, токенизация, лемматизация, стемминг. |
| 17 | Обработка естественного языка: векторные модели | Преобразование текста в векторное пространство. Статистические методы построения векторных моделей: мешок слов, OHE, tf-idf. Тематическое моделирование: латентное размещение Дирихле (LDA), латентный семантический анализ (LSA). Нейросетевые модели: word2vec, fastText, GloVe. |
| 18 | Обработка естественного языка: типы решаемых задач | Рассмотрение спектра задач обработки естественного языка: моделирование языка, классификация, поиск именованных сущностей, суммаризация, генерация. |
| 19 | Введение в нейронные сети | Базовые понятия нейронных сетей: нейрон, обучение нейрона, нейронная сеть, обучение нейронной сети. Функции потерь и метрики качества. |
| 20 | Нейронные сети для решения задач регрессии | Применение нейронных сетей для решения задач регрессии. Специфические для регрессии функции активации нейронов и метрики качества. Методы предварительной обработки данных. |
| 21 | Нейронные сети для решения задач классификации | Применение нейронных сетей для решения задач классификации. Специфические для классии функции активации нейронов и метрики качества. Методы предварительной обработки данных. |
| 22 | Нейронные сети для решения задач обработки изображений | Знакомство с задачами обработки изображений. Изучение сверточных нейронных сетей. Понятие ядра, свертки и подвыборки. |
| 23 | Нейронные сети для решения задач обработки изображений: современные архитектуры | Изучение современных архитектур нейронных сетей для решения задач обработки изображений: VGG-16, VGG-19, ResNet, Inception, Xception, DenseNet, MobileNet, EfficientNet. |
| 24 | Нейронные сети для решения задач обработки естественного языка | Архитектуры нейронных сетей для решения задач обработки естественного языка: одномерные сверточные сети, рекуррентные нейронные сети. |
| 25 | Нейронные сети для решения задач обработки естественного языка: современные архитектуры | Обзор современных модификаций нейронных сетей для решения задач обработки естественного языка: трансформеры, механизм внимания, BERT |
| 26 | Нейронные сети для решения задач предсказания временных рядов | Применение нейронных сетей для решения задачи прогнозирования временных рядов: одномерные сверточные нейронные сети, рекуррентные нейронные сети. WaveNet для прогнозирование временного ряда |

2.3. Программа дисциплины реализуется на государственном языке Российской Федерации.

**2.4. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ Спортивный анализ данных**

**Электронные ресурсы (издания)**

1. Хливненко, Л. В. Практика нейросетевого моделирования : учебное пособие для вузов / Л. В. Хливненко, Ф. А. Пятакович. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 200 с. — ISBN 978-5-8114-8264-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/173811 (дата обращения: 05.10.2021).
2. Клетте, Р. Компьютерное зрение. Теория и алгоритмы : учебник / Р. Клетте ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 506 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131691 (дата обращения: 08.10.2021).
3. Тарасов, И. Е. Статистический анализ данных в информационных системах : учебно-методическое пособие / И. Е. Тарасов. — Москва : РТУ МИРЭА, 2020. — 96 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/163854 (дата обращения: 06.10.2021).
4. Ростовцев, В. С. Искусственные нейронные сети : учебник / В. С. Ростовцев. — Санкт-Петербург : Лань, 2019. — 216 с. — ISBN 978-5-8114-3768-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/122180 (дата обращения: 08.10.2021). — Режим доступа: для авториз. пользователей.

**Печатные издания**

1. Жерон, Орельен, Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем, 2-е изд. Пер. с англ. - СпБ.: ООО "Диалектика': 2020. - 1040 с.: ил. - Парал. тит. англ.
2. Копец Дэвид, Классические задачи Computer Science на языке Python. - СПб.: Питер, 2020. - 256 с.: ил. - (Серия «Библиотека программиста»).
3. Элбон Крис, Машинное обучение с использованием Python. Сборник рецептов: Пер. с англ. — СПб.: БХВ-Петербург, 2019. — 384 с.: ил.
4. Вейдман Сет, Глубокое обучение: легкая разработка проектов на Python. — СПб.: Питер, 2021. — 272 с.: ил. — (Серия «Бестселлеры O’Reilly»).
5. Микелуччи У., Прикладное глубокое обучение. Подход к пониманию глубоких нейронных сетей на основе метода кейсов: Пер. с англ. - СПб.: БХВ-Петербург, 2020. - 368 с.: ил.
6. Шолле Франсуа, Глубокое обучение на Python. — СПб.: Питер, 2018. — 400 с.: ил. — (Серия «Библиотека программиста»).
7. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с анг. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил.
8. Хобсон Лейн, Ханнес Хапке, Коул Ховард, Обработка естественного языка в действии. — СПб.: Питер, 2020. — 576 с.: ил. — (Серия «Для профессионалов»)
9. Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони, Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. — СПб.: Питер, 2019. — 368 с.: ил. — (Серия «Бестселлеры O’Reilly»).

**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**

Спортивный анализ 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. | MS PowerPoint, MS Word, MS Excel. MS Teams    Используется бесплатно-распространяемое программное обеспечение:   1. Python – <https://www.python.org/> |