

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт искусственного интеллекта, робототехники и системной инженерии



УТВЕРЖДАЮ
Проректор по образовательной деятельности КФУ

Турилова Е.А.

20 23 г.



Программа дисциплины
Основы BigData и DataMining

Направление подготовки: 15.03.06 - Мехатроника и робототехника

Профиль подготовки: Робототехника и искусственный интеллект

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): ведущий инженер-программист Державин Д.В. (НИЦ "ЦП "Специальная робототехника и ИИ"", Институт вычислительной математики и информационных технологий), DVDerzhavin@kpfu.ru.

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-1	Способен разрабатывать программное обеспечение, необходимое для обработки информации и управления в мехатронных и робототехнических системах, а также для их проектирования

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- основные принципы и методы разработки программного обеспечения, которое работает с большими данными;
- основные концепции Big Data и Data Mining;
- особенности разработки распределенных программ.

Должен уметь:

- разрабатывать программное обеспечение, с использованием алгоритмов машинного обучения;
- следовать основным шаблонам проектирования систем, анализирующих большие данные.

Должен владеть:

- навыками построения и моделирования систем работы с большими данными;
- навыками анализа и обработки больших объемов данных, ошибок и отладки.

Должен демонстрировать способность и готовность:

Применять полученные знания и навыки в практической деятельности

Дисциплина связана со следующими дисциплинами: “цифровая обработка сигналов”, “основы программирования C/C++”, “основы Python”, “курсовая работа по профилю подготовки”, “компьютерные игры и стрессоустойчивое проектирование”, “основы машинного обучения”, “нейросети, генеративные платформы и глубокое обучение”, “основы машинного зрения и обработки сенсорных данных”, “технологическая (проектно-технологическая) практика”, “выполнение, подготовка к процедуре защиты и защита выпускной квалификационной работы”.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел " Б1.В.ДВ.03.01 Дисциплины (модули)" блока ИИ-сопряженных и обеспечивающих дисциплин 15.03.06 "Мехатроника и робототехника" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2 курсе в 3 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 8 зачетные единицы на 288 часов.

Контактная работа - 108 часа, в том числе лекции - 54 часов, практические занятия - 54 часов, лабораторные работы - 0 часов, контроль самостоятельной работы - 0 часов.

Самостоятельная работа - 126 часа.

Контроль экзамен - 54 часа.

Форма промежуточного контроля дисциплины: экзамен в 3 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины /	ем		Виды и часы контактной работы, их трудоемкость (в часах)	л	ял	ел
		с	с				

модуля			Лекции, всего	в т.ч. лекции в эл.форме	Практические занятия, всего	в т.ч. практические в эл.форме	Лабораторные работы, всего	в т.ч. лабораторные в эл.форме	
1.	Тема 1. Введение в Big Data и Data Mining	3	2	0	2	0	0	0	6
2.	Тема 2. Распределенные системы	3	2	0	2	0	0	0	8
3.	Тема 3. Технологии хранения	3	2	0	2	0	0	0	8
4.	Тема 4. Основы параллельной обработки данных	3	2	0	2	0	0	0	8
5.	Тема 5. MapReduce. Spark	3	2	0	2	0	0	0	8
6.	Тема 6. Hadoop и Hive	3	4	0	4	0	0	0	8
7.	Тема 7. Анализ данных	3	4		4				8
8.	Тема 8. Data Mining	3	4		4				8
9.	Тема 9. Программное обеспечение в области анализа данных	3	4		4				8
10.	Тема 10. Ассоциативные правила	3	4		4				8
11.	Тема 11. Кластеризация	3	4		4				8
12.	Тема 12. Классификация и регрессия	3	5		5				10
13.	Тема 13. Решающие деревья	3	5		5				10
14.	Тема 14. Нейронные сети	3	5		5				10
15.	Тема 15. Архитектуры нейронных сетей	3	5		5				10
	Итого	3	54	0	54	0	0	0	126

4.2 Содержание дисциплины (модуля)

Тема 1. Введение в Big Data и Data Mining

Определение дисциплины, ее цели и задачи. Важность изучения в современном мире. Понятие больших данных. Проблемы, связанные с большими данными.

Тема 2. Распределенные системы

Понятие распределенных систем. Обзор различных типов распределенных систем, мотивация создания распределенных систем. Задачи распределенных систем: доступ к ресурсам, прозрачность, открытость, масштабируемость. Способы организации распределенных систем: аппаратные и программные решения. Связь в распределенных системах.

Тема 3. Технологии хранения

Распределенные хранилища. NoSql хранилища. OLAP

Тема 4. Основы параллельной обработки данных

Основные характеристики параллельной программы: ускорение, эффективность, предел масштабирования. 1-й закон Амдала. 2-й закон Амдала. Закон Густафсона-Барсиса. Параллелизм задач. Параллелизм данных. Этапы разработки параллельных алгоритмов.

Тема 5. Алгоритм MapReduce. Spark

Алгоритм MapReduce. Стадия Map. Стадия Shuffle. Стадия Reduce. Spark. Отличия от MapReduce.

Тема 6. Hadoop и Hive

Файловая система GFS. Hadoop. HDFS. Hive.

Тема 7. Анализ данных

Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных

Тема 8. Data Mining

KDD. Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность применения технологий Data Mining для обработки больших объемов информации.

Тема 9. Программное обеспечение в области анализа данных

Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования.

Тема 10. Ассоциативные правила

Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка ассоциативного правила. Достоверность ассоциативного правила. Поиск ассоциативных правил. Алгоритм Apriori. Иерархические ассоциативные правила.

Тема 11. Кластеризация

Понятие кластеризации. Применение кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Метрики расстояний. Алгоритм k средних.

Тема 12. Классификация и регрессия

Задача классификации. Задача регрессии. Обзор методов классификации и регрессии. Методы, основанные на обучении, разнообразие подходов. Применение в Data mining.

Тема 13. Решающие деревья

Решающее дерево. Стратегия «разделяй и властвуй». Критерии ветвления. Критерий Джини. Деревья регрессии. ID3, C4.5, CART. Решающие деревья и линейные модели.

Тема 14. Нейронные сети

Нейронные сети. Глубинное обучение. Линейный слой. Функция активации. Вычислительный граф. Алгоритм обратного распространения ошибки. Различные функции активации. Инициализация весов. Регуляризация нейронных сетей.

Тема 15. Архитектуры нейронных сетей

Сверточные нейронные сети. Рекуррентные нейронные сети. Генеративно-состязательные нейронные сети. Нейронные сети типа трансформер. Фреймворки и библиотеки для работы с нейросетями.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;

- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

База данных научной электронной библиотеки - <https://elibrary.ru/>

Электронно-библиотечная система Znanium - <https://znanium.com/>

Профессиональный интернет-ресурс по машинному обучению - <http://www.machinelearning.ru/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	Лекционные занятия проходят в интерактивной форме, предполагающей вовлечение обучающихся в обсуждение всех предложенных тем. Применяются такие формы лекционных занятий как лекция-презентация, лекция-дискуссия, проблемная лекция, видео-лекция. Студенты активно участвуют в конструировании знаний во время круглых столов, дискуссионных площадок.
практические занятия	Практические занятия, семинары являются одной из основных форм образовательного процесса, ориентированной на усвоение студентами теоретического материала и выработку практических компетенций. Основной целью практических занятий является комплексный контроль усвоения пройденного материала, хода выполнения студентами самостоятельной работы и рассмотрение наиболее сложных и спорных вопросов в рамках темы занятия. Подготовка к семинарам предполагает самостоятельную работу студентов по изучению материала по конкретной теме.
самостоятельная работа	Самостоятельная работа преследует цель закрепить, углубить и расширить знания, полученные студентами в ходе аудиторных занятий, а также сформировать навыки работы с научной, учебной и учебно-методической литературой, развивать творческое, продуктивное мышление обучаемых, их креативные качества, формирование общекультурных и профессиональных компетенций.
экзамен	Экзамен проводится в письменной форме. В билет включаются тестовые вопросы и задачи. Студенту дается 90 минут для выполнения своего варианта экзаменационного задания.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;

- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;

- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;

- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;

- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;

- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;

- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:

- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;

- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;

- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 15.03.06 "Мехатроника и робототехника" и профилю подготовки "Робототехника и искусственный интеллект".

Приложение №1
к рабочей программе дисциплины (модуля)
«Основы BigData и DataMining»

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«Казанский (Приволжский) федеральный университет»

Институт искусственного интеллекта, робототехники и системной инженерии

Фонд оценочных средств по дисциплине

Б1.В.ДВ.03.01 Основы BigData и DataMining

Направление подготовки: 15.03.06 «Мехатроника и робототехника»

Профиль: Робототехника и искусственный интеллект

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

СОДЕРЖАНИЕ

1. СООТВЕТСТВИЕ КОМПЕТЕНЦИЙ ПЛАНИРУЕМЫМ РЕЗУЛЬТАТАМ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)
2. КРИТЕРИИ ОЦЕНИВАНИЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ
3. РАСПРЕДЕЛЕНИЕ ОЦЕНОК ЗА ФОРМЫ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНУЮ АТТЕСТАЦИЮ
4. ОЦЕНОЧНЫЕ СРЕДСТВА, ПОРЯДОК ИХ ПРИМЕНЕНИЯ И КРИТЕРИИ ОЦЕНИВАНИЯ
 - 4.1. ОЦЕНОЧНЫЕ СРЕДСТВА ТЕКУЩЕГО КОНТРОЛЯ
 - 4.1.1. Устный опрос по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.
 - 4.1.1.1. Порядок проведения и процедура оценивания
 - 4.1.1.2. Критерии оценивания
 - 4.1.1.3. Содержание оценочного средства
 - 4.1.2. Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.
 - 4.1.2.1. Порядок проведения и процедура оценивания
 - 4.1.2.2. Критерии оценивания
 - 4.1.2.3. Содержание оценочного средства
 - 4.2. ОЦЕНОЧНЫЕ СРЕДСТВА ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
 - 4.2.1. Вопросы по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.
 - 4.2.1.1. Порядок проведения и процедура оценивания
 - 4.2.1.2. Критерии оценивания
 - 4.2.1.3. Содержание оценочного средства
 - 4.2.2. Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.
 - 4.2.2.1. Порядок проведения и процедура оценивания
 - 4.2.2.2. Критерии оценивания
 - 4.2.2.3. Содержание оценочного средства

1. Соответствие компетенций планируемым результатам обучения по дисциплине
(модулю)

Код и наименование	Индикаторы достижений компетенций	Оценочные средства текущего контроля и промежуточной аттестации
<p>ПК-1 Способен разрабатывать программное обеспечение, необходимое для обработки информации и управления в мехатронных и робототехнических системах, а также для их проектирования</p>	<p>ПК-1. И-1: знает основные принципы и методы разработки программного обеспечения в области анализа больших данных для управляющих и информационных систем в мехатронике и робототехнике</p> <p>ПК-1. И-2: умеет разрабатывать программное обеспечение, с применением алгоритмов анализа больших данных, для управления мехатронными и робототехническими системами</p> <p>ПК-1. И-3: владеет навыками проектирования и моделирования мехатронных систем, включая специализированные программные пакеты с алгоритмами анализа больших данных</p>	<p>Текущий контроль: Устный опрос по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.</p> <p>Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.</p> <p>Промежуточная аттестация: Вопросы по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”,</p>

“Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.

Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.

2. Критерии оценивания сформированности компетенций

Компетенция	Высокий уровень (отлично) (86-100 баллов)	Средний уровень (хорошо) (71-85 баллов)	Низкий уровень (удовлетворительно) (56-70 баллов)	Ниже порогового уровня (неудовлетво рительно) (0-55 баллов)
ПК-1 И-1	Знает принципы организации и управления проектами в области анализа больших данных, фундаментальные принципы программирования и разработки программного обеспечения, включая алгоритмы, структуры данных, управление памятью и основы языков программирования.	Знает фундаментальные принципы программирования и разработки программного обеспечения в области анализа больших данных, включая алгоритмы, структуры данных, управление памятью и основы языков программирования.	Знает фундаментальные принципы программирования и разработки программного обеспечения в области анализа больших данных, структуры данных и основы языков программирования.	Знает на крайне низком уровне принципы организации и управления проектами в области анализа больших данных, фундаментальные принципы программирования и разработки программного обеспечения, включая алгоритмы, структуры данных, управление памятью и основы языков программирования

ПК-1 И-2	Умеет работать с инструментарием для проектирования, разработки и тестирования программного обеспечения в области анализа больших данных, интерпретировать результаты и принимать решения на основе полученных данных.	Умеет работать с инструментарием для проектирования, разработки и тестирования программного обеспечения в области анализа больших данных.	Умеет работать с инструментарием для разработки программного обеспечения в области анализа больших данных.	Умеет на крайне низком уровне работать с инструментарием для проектирования, разработки и тестирования программного обеспечения в области анализа больших данных, интерпретировать результаты и принимать решения на основе полученных данных.
----------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

ПК-1 И-3	Владеет навыками анализа и обработки больших объемов данных, интерпретации результатов и принятия решений на основе полученных данных, языками программирования и методами разработки программного обеспечения, управления проектами.	Владеет навыками анализа и обработки больших объемов данных, интерпретации результатов и принятия решений на основе полученных данных, языками программирования и методами разработки программного обеспечения.	На базовом уровне владеет навыками анализа и обработки больших объемов данных, интерпретации результатов и принятия решений на основе полученных данных.	Не обладает или владеет на крайне низком уровне навыками анализа и обработки больших объемов данных, интерпретации результатов и принятия решений на основе полученных данных, языками программирования и методами разработки программного обеспечения, управления проектами.
----------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

3. Распределение оценок за формы текущего контроля и промежуточную аттестацию

3 семестр:

Текущий контроль:

1. Устный опрос по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей” - 20 баллов.

2. Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”. - 30 баллов

Итого 50 баллов

Промежуточная аттестация – экзамен

Экзамен проходит в письменной форме. Студенту предоставляется 90 минут на письменный ответ по билету. Каждый билет содержит два вопроса и одну задачу посвященные темам дисциплины, предусмотренные Учебной программой. Ответ на каждый вопрос оценивается в 15 баллов, ответ на задачу оценивается в 20 баллов.

Итоговая оценка за экзамен определяется путем суммирования баллов за все правильно отвеченные вопросы билета.

Итого 50 баллов

Общее количество баллов по дисциплине за текущий контроль и промежуточную аттестацию: $50+50=100$ баллов.

Соответствие баллов и оценок:

86-100 – отлично

71-85 – хорошо

56-70 – удовлетворительно

0-55 – неудовлетворительно

4. Оценочные средства, порядок их применения и критерии оценивания

4.1. Оценочные средства текущего контроля

4.1.1. Устный опрос по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей””.

4.1.1.1. Порядок проведения и процедура оценивания

В рамках данного курса студенты, помимо изучения теоретического материала и разбора практических примеров должны показать степень усвоения рассмотренного вопроса занятий путем устного опроса. Теоретические материалы и практические примеры студенты совместно с преподавателем изучают на лекционных и практических занятиях соответственно.

4.1.1.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если обучающийся: - Правильно ответил на все вопросы и обосновал свой ответ.

Баллы в интервале 71-85% от максимальных ставятся, если обучающийся:

- Правильно ответил на все вопросы, но при этом не обосновал свой ответ;
- Обосновал свой ответ, но не раскрыл его полностью.

Баллы в интервале 56-70% от максимальных ставятся, если обучающийся:

- Ответил не на все вопросы;
- Ответил на все вопросы, но меньшая часть ответов являются ошибочными.

Баллы в интервале 0-55% от максимальных ставятся, если обучающийся:

- Не ответил на большую часть вопросов;

4.1.1.3. Содержание оценочного средства

1. Какие проблемы, связанные с большими данными, вы знаете?
2. Что такое распределенная система? Какие основные задачи распределенной системы?
3. Какие существуют аппаратные способы организации распределенных систем?

4. Какие существуют программные способы организации распределенных систем?
5. Какие существуют этапы разработки параллельных алгоритмов?
6. 1-ый закон Амдала. 2-й закон Амдала.
7. Чем отличаются параллелизм задач и параллелизм данных? Приведите примеры.
8. Какие стадии алгоритма MapReduce вы знаете?
9. Какие задачи решает Spark?
10. Какие задачи решает Hadoop?
11. Какие задачи решает Hive?
12. Что такое ассоциативные правила?
13. Что такое кластеризация?
14. Дайте определение классификации. Для чего используется классификация в задачах анализа данных?
15. Что такое OLAP?
16. Дайте определение регрессии. Для чего используется регрессия в задачах анализа данных?
17. Для чего используется кластеризация в задачах анализа данных?
18. Какие метрики расстояний вы знаете?
19. Какие алгоритмы построения решающих деревьев вы знаете?
20. Что такое свертка?
21. Что такое функция активации?
22. В чем заключается основная идея алгоритма Apriori?
23. Как осуществляется подготовка данных к анализу?
24. Что такое HDFS? Какие проблемы HDFS решает?
25. Как работает метод обратного распределения ошибки?

4.1.2. Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.

4.1.2.1. Порядок проведения и процедура оценивания

Задачи являются одной из форм текущего контроля. Задачи включают в себя задания, которые охватывают все темы курса, поэтому соответствуют ПК-1, знания.

Каждый из вариантов включает в себя 2 задачи, каждый из которых оценивается в 15 баллов. В случае неверно выполненного задания результат ответа признается равным 0.

Общий итог тестирования рассчитывается путем суммирования баллов за правильные ответы. Задачи даются в конце семестра после того, как обучающиеся освоили все темы курса.

4.1.2.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если обучающийся:

- студент полностью решил обе задачи;
- студент полностью решил одну задачу и с небольшими ошибками вторую.

Баллы в интервале 71-85% от максимальных ставятся, если обучающийся:

- студент с небольшими ошибками решил обе задачи;
- студент полностью решил одну задачу и со значимыми ошибками вторую.

Баллы в интервале 56-70% от максимальных ставятся, если обучающийся:

- студент полностью решил только одну задачу;
- студент со значимыми ошибками решил одну задачу и с небольшими ошибками решил вторую.

Баллы в интервале 0-55% от максимальных ставятся, если обучающийся:

- студент со значимыми ошибками решил обе задачи;
- студент не решил одну задачу и со значимыми ошибками решил вторую.

4.1.2.3. Содержание оценочного средства

Пример вариантов тестирования:

ВАРИАНТ 1.

1. Используя файл с данными, полученный от преподавателя, построить модель линейной регрессии. Данные не содержат выбросов. Визуализировать результат.
2. Используя файл с данными, полученный от преподавателя, разделите выборку на обучающую и тестовую в соотношении 70/30 и обучите модель решающего дерева. Выведите долю верных прогнозов.

ВАРИАНТ 2.

1. Используя файл с данными, полученный от преподавателя, построить модель метода k ближайших соседей. Визуализируйте результат.
2. Используя файл с данными, полученный от преподавателя, разделите выборку на обучающую и тестовую в соотношении 80/20 и обучите модель линейного классификатора. Выведите долю верных прогнозов.

База задач

1. Используя файл с данными, полученный от преподавателя, построить модель линейной регрессии. Данные не содержат выбросов. Визуализировать результат.
2. Используя файл с данными, полученный от преподавателя, разделите выборку на обучающую и тестовую в соотношении 70/30 и обучите модель решающего дерева. Выведите долю верных прогнозов.
3. Используя файл с данными, полученный от преподавателя, построить модель метода k ближайших соседей. Визуализируйте результат.
4. Используя файл с данными, полученный от преподавателя, разделите выборку на обучающую и тестовую в соотношении 80/20 и обучите модель линейного классификатора. Выведите долю верных прогнозов.
5. Постройте модель дерева решений, используя файл с данными, полученный от преподавателя. Выведите процент правильных ответов на обучающей и тестовой выборках.

4.2. Оценочные средства промежуточной аттестации

Экзамен проходит в письменной форме. Студенту предоставляется 90 минут на письменный ответ по билету. Каждый билет содержит 2 теоретических вопроса и задание, охватывающие все темы дисциплины, предусмотренные Учебной программой.

Билет состоит из двух частей: теоретической (тестовой) и практической (задачи). В билет входят:

- Вопросы;
- Задачи;

Первая часть включает в себя 2 теоретических вопроса. Каждый тестовый вопрос оценивается в 10 баллов.

Далее идет одна задача, которая показывает умение обучающегося анализировать информацию, работать с ней, разрабатывать на основе нее программное обеспечение. При оценке задачи учитывается полнота ответа, его логичность, правильность решения. Решение задачи оценивается в 30 баллов.

Итоговая оценка за экзамен определяется путем суммирования баллов за все правильно выполненные задания билета.

Результат экзамена оценивается так:

86-100 – отлично

71-85 – хорошо

56-70 – удовлетворительно

0-55 – неудовлетворительно

4.2.1. Вопросы по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.

4.2.1.1. Порядок проведения и процедура оценивания

Вопросы являются одной из форм промежуточной аттестации. Вопросы включают себя теоретическое изложение указанной темы, что соответствует ПК-1.

Каждый из вариантов включает в себя 2 вопроса, каждый из которых оценивается в 10 баллов. В случае неверно выполненного задания результат ответа признается равным 0.

Общий итог рассчитывается путем суммирования баллов за правильные ответы. Вопросы даются в конце семестра в рамках экзаменационного билета.

4.2.1.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если обучающийся:

- студент полностью ответил на оба вопроса;
- студент полностью ответил на один вопроса и с небольшими ошибками на второй.

Баллы в интервале 71-85% от максимальных ставятся, если обучающийся:

- студент с небольшими ошибками ответил на оба вопроса;
- студент полностью решил ответил на один вопроса и со значимыми ошибками на второй.

Баллы в интервале 56-70% от максимальных ставятся, если обучающийся:

- студент полностью ответил только на один вопрос;

- студент со значимыми ошибками ответил на один вопрос и с небольшими ошибками ответил на второй.

Баллы в интервале 0-55% от максимальных ставятся, если обучающийся:

- студент не ответил ни на один вопрос;
- студент со значимыми ошибками ответил на оба вопроса;
- студент не ответил на один вопроса и со значимыми ошибками ответил на второй.

4.2.1.3. Содержание оценочного средства

Пример вариантов тестирования:

ВАРИАНТ 1.

1. Большие данные. Проблемы, связанные с большими данными.
2. Решающее дерево. Стратегия «разделяй и властвуй». Критерии ветвления.

ВАРИАНТ 2.

1. Распределенные системы. Задачи распределенных систем.
2. Нейронные сети. Функция активации. Алгоритм обратного распространения ошибки.

Список вопросов

1. Большие данные. Проблемы, связанные с большими данными.
2. Распределенные системы. Задачи распределенных систем.
3. Решающее дерево. Стратегия «разделяй и властвуй». Критерии ветвления.
4. Нейронные сети. Функция активации. Алгоритм обратного распространения ошибки.
5. Способы организации распределенных систем.
6. Связь в распределенных системах.
7. NoSql хранилища. OLAP.
8. Основные характеристики параллельной программы. 1-й закон Амдала.
9. 2-й закон Амдала. Закон Густафсона-Барсиса.
10. Параллелизм задач. Параллелизм данных. Этапы разработки параллельных алгоритмов.
11. MapReduce. Стадии алгоритма MapReduce.
12. Spark. Отличия от MapReduce.
13. Hadoop. HDFS.
14. Общая схема анализа. Формы представления данных.
15. Подготовка данных к анализу. Методика извлечения знаний.
16. Ассоциативные правила. Поддержка и достоверность ассоциативного правила.
17. Алгоритм Apriori. Иерархические ассоциативные правила.
18. Понятие кластеризации. Методы кластеризации.
19. Метрики расстояний. Алгоритм k средних.
20. Задача классификации. Задача регрессии. Применение в Data mining.

4.2.2. Задачи по темам: “Введение в Big Data и Data Mining”, “Распределенные системы”, “Технологии хранения”, “Основы параллельной обработки данных”, “MapReduce. Spark”, “Hadoop и Hive”, “Анализ данных, Data Mining”, “Программное обеспечение в области анализа данных”, “Ассоциативные правила”, “Кластеризация”, “Классификация и регрессия”, “Решающие деревья”, “Нейронные сети”, “Архитектуры нейронных сетей”.

4.1.2.1. Порядок проведения и процедура оценивания

Задачи являются одной из форм промежуточной аттестации. Задачи включают себя письменное изложение решения задания затрагивающие темы дисциплины, что соответствует ПК-1.

Каждый из вариантов включает в себя одну задачу, которая оценивается в 30. В случае неверно выполненного задания результат ответа признается равным 0.

Общий итог рассчитывается путем суммирования баллов за правильные ответы. Задание дается в конце семестра в рамках экзаменационного билета.

4.2.2.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если обучающийся:

- студент полностью ответил на задачу;
- студент ответил на задачу с не критичными ошибками;

Баллы в интервале 71-85% от максимальных ставятся, если обучающийся:

- студент ответил на задачу с небольшими ошибками;

Баллы в интервале 56-70% от максимальных ставятся, если обучающийся:

- студент ответил на задачу со значимыми ошибками;

Баллы в интервале 0-55% от максимальных ставятся, если обучающийся:

- студент не ответил на задачу;
- студент ответил на задачу с серьёзными ошибками;

4.2.2.3. Содержание оценочного средства

Пример варианта задачи:

ВАРИАНТ 1.

1. Используя файл с данными, полученный от преподавателя, разделите выборку на обучающую и тестовую в соотношении 80/20. Если выборка содержит пропущенные значения, их необходимо удалить. Обучите 3 классификатора kNN, укажите число соседей равным 3, 10, 25 соответственно. Выведите в консоль долю правильных ответов на обучающей и тестовой выборках.

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 15.03.06 – Мехатроника и робототехника

Профиль подготовки: Робототехника и искусственный интеллект

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Основная литература:

1. Протодяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / А. В. Протодяконов, П. А. Пылов, В. Е. Садовников. - Москва ; Вологда : Инфра-Инженерия, 2022. - 392 с. - ISBN 978-5-9729-1006-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1902689> (дата обращения: 21.09.2023). – Режим доступа: по подписке.

2. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. - 2-е изд. - Москва.:ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2085038> (дата обращения: 14.09.2023). – Режим доступа: по подписке.

3. Замятин, А. В. Интеллектуальный анализ данных : учебное пособие / А. В. Замятин. - Томск : Издательский Дом Томского государственного университета, 2020. - 196 с. - ISBN 978-5-94621-898-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1864765> (дата обращения: 21.09.2023). – Режим доступа: по подписке.

Дополнительная литература:

1. Коэльо, Л. Построение систем машинного обучения на языке Python : практическое руководство / Л. Коэльо, В. Ричарт ; пер. с англ. А. А. Слинкина. - 3-е изд. - Москва : ДМК Пресс, 2023. - 304 с. - ISBN 978-5-89818-331-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2102618> (дата обращения: 14.09.2023). – Режим доступа: по подписке.

2. Рамальо, Л. Python. К вершинам мастерства / Лучано Рамальо ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2016. - 768 с. - ISBN 978-5-97060-384-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028052> (дата обращения: 14.09.2023). – Режим доступа: по подписке.

3. Зарова, Е. В. Методы Data mining в обработке и анализе статистических данных (решения в R) : монография / Е.В. Зарова. — Москва : ИНФРА-М, 2021. — 232 с. : ил. - ISBN 978-5-16-016814-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1240276> (дата обращения: 21.09.2023). – Режим доступа: по подписке.

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 15.03.06 – Мехатроника и робототехника

Профиль подготовки: Робототехника и искусственный интеллект

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Microsoft Windows 10 Профессиональная

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

PyCharm Community Edition

Kaspersky Endpoint Security для Windows