

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__ г.

Программа дисциплины
Основы BigData и DataMining

Направление подготовки: 09.03.04 - Программная инженерия
Профиль подготовки: Разработчик искусственного интеллекта и когнитивных систем
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): заместитель директора центра Кокунин П.А. (Научно-исследовательский центр Центр превосходства Специальная робототехника и искусственный интеллект, Институт вычислительной математики и информационных технологий), PAKokunin@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:

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 алгоритмы и программы, пригодные для практического использования в области анализа больших данных;
- следовать основным шаблонам проектирования систем, анализирующих большие данные;
- представлять данные в требуемом формате с использованием информационных, компьютерных и сетевых технологий;
- применять основы информатики и программирования к проектированию, конструированию и тестированию программных продуктов в области больших данных.

Должен владеть:

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

Должен демонстрировать способность и готовность:

Применять полученные знания и навыки в практической деятельности

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.12 Дисциплины (модули)" основной профессиональной образовательной программы 09.03.04 "Программная инженерия (Разработчик искусственного интеллекта и когнитивных систем)" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2, 3 курсах в 4, 5 семестрах.

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 зачетных(ые) единиц(ы) на 288 часа(ов).

Контактная работа - 108 часа(ов), в том числе лекции - 60 часа(ов), практические занятия - 48 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 144 часа(ов).

Контроль (зачёт / экзамен) - 36 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 4 семестре; экзамен в 5 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 / модуля	Се- местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само- стоя- тель- ная ра- бота
			Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабора- торные работы, всего	Лабора- торные в эл. форме	
N	Разделы дисциплины / модуля	Се- местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само- стоя- тель- ная ра- бота
N	Разделы дисциплины / модуля	Се- местр	Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабора- торные работы, всего	Лабора- торные в эл. форме	Само- стоя- тель- ная ра- бота
1.	Тема 1. Введение в Big Data и Data Mining	4	4	0	4	0	0	0	8
2.	Тема 2. Распределенные системы	4	4	0	4	0	0	0	8
3.	Тема 3. Технологии хранения	4	4	0	4	0	0	0	8
4.	Тема 4. Основы параллельной обработки данных	4	4	0	4	0	0	0	8
5.	Тема 5. MapReduce. Spark	4	4	0	4	0	0	0	8
6.	Тема 6. Hadoop и Hive	4	4	0	4	0	0	0	8
7.	Тема 7. Анализ данных	4	4	0	4	0	0	0	8
8.	Тема 8. Data Mining	4	4	0	4	0	0	0	8
9.	Тема 9. Программное обеспечение в области анализа данных	4	4	0	4	0	0	0	8
10.	Тема 10. Ассоциативные правила	5	4	0	2	0	0	0	12
11.	Тема 11. Кластеризация	5	4	0	2	0	0	0	12
12.	Тема 12. Классификация и регрессия	5	4	0	2	0	0	0	12
13.	Тема 13. Решающие деревья	5	4	0	2	0	0	0	12
14.	Тема 14. Нейронные сети	5	4	0	2	0	0	0	12
4.2	Содержание дисциплины (модуля)								
	Тема 1. Введение в Big Data и Data Mining	5	4	0	2	0	0	0	12
	Тема 15. Архитектуры нейронных сетей								
	Определение дисциплины, ее цели и задачи. Важность изучения в современном мире. Понятие больших данных. Проблемы, связанные с большими данными.		60	0	48	0	0	0	144

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. 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-сопоставительные 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 №14-55-99бин/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

Журнал - http://www.creativeconomy.ru/mag_rp/

Журнал - http://www.basw-ngo.by/page.php?issue_id=2855

Правительство РФ - <http://government.ru/>

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)

Журнал - http://www.basw-ngo.by/page.php?issue_id=2855

Журнал - http://www.creativeconomy.ru/mag_rp/

Правительство РФ - <http://government.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.

Вид работ	Методические рекомендации
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.
зачет	Зачет проводится в письменной форме. В билет включаются вопросы. Студенту дается 90 минут для выполнения своего варианта задания.
экзамен	Экзамен проводится в письменной форме. В билет включаются 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.

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 09.03.04 "Программная инженерия" и профилю подготовки "Разработчик искусственного интеллекта и когнитивных систем".

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Разработчик искусственного интеллекта и когнитивных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Основная литература:

1. Протодьяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / А. В. Протодьяконов, 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: 29.01.2024). - Режим доступа: по подписке.
2. Цукалос, М. Golang для профи: работа с сетью, многопоточность, структуры данных и машинное обучение с Go : практическое руководство / М. Цукалос. - Санкт-Петербург : Питер, 2021. - 720 с. - (Серия 'Для профессионалов'). - ISBN 978-5-4461-1617-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1733707> (дата обращения: 29.01.2024)
3. Ярушкина, Н. Г. Интеллектуальный анализ временных рядов : учебное пособие / Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. - Москва : ФОРУМ : ИНФРА-М, 2022. - 160 с. - (Высшее образование). - ISBN 978-5-8199-0496-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1842559> (дата обращения: 29.01.2024). - Режим доступа: по подписке.

Дополнительная литература:

1. Коэльо, Луис Педро Построение систем машинного обучения на языке Python / Луис Педро Коэльо, Вилли Ричарт - Москва : ДМК Пресс, 2016. - 302 с. - ISBN 978-5-97060-330-7. - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970603307.html> (дата обращения: 29.01.2024). - Режим доступа : по подписке.
2. Шевченко, Л. Г. Программирование на PYTHON в среде IDLE : учебное пособие / Л. Г. Шевченко, Т. В. Дружинина. - Новосибирск : Изд-во НГТУ, 2020. - 195 с. - ISBN 978-5-7782-4215-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1866915> (дата обращения: 29.01.2024). - 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: 29.01.2024). - Режим доступа: по подписке.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.12 Основы BigData и DataMining

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

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Разработчик искусственного интеллекта и когнитивных систем

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 Windows XP (Volume License)

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

Kaspersky Endpoint Security для Windows