



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.01.13 Технологии машинного обучения и анализа больших данных

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): Компьютерное моделирование и анализ данных

Квалификация (степень): бакалавр

Форма обучения: очная

Институт: математики, естествознания и техники

Кафедра: математического моделирования, компьютерных технологий и информационной безопасности

	очная форма	очно-заочная форма	заочная форма
Курс	4		
Семестр	7, 8		
Лекции	18		
Лабораторные занятия	36		
Практические (семинарские) занятия	18		
в т. ч. практическая подготовка	8		
Форма(ы) промежуточной аттестации	Зачет		
Контроль	-		
Иные формы работы	-		
Самостоятельная работа	72		

Всего часов: 144

Трудоемкость: 4 зачетные единицы.

Разработчик(и) рабочей программы:

кандидат технических наук, доцент кафедры ММКТиИБ А.А. Петров

I. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ РАЗДЕЛ

Цель изучения дисциплины:

развитие у студента достаточно широкого взгляда на методы и технологии программирования систем машинного обучения и анализа больших данных;
подготовка студента к практической деятельности в области создания, внедрения и эксплуатации информационных систем на основе машинного обучения.

Задачи изучения дисциплины:

рассмотрение основных понятий искусственного интеллекта, искусственных нейронных сетей, машинного обучения и анализа больших данных;
рассмотрение технической постановки основных задач, решаемых с помощью современных систем машинного обучения;
ознакомление с основными моделями представления знаний и некоторыми интеллектуальными системами;
рассмотрение теоретических и практических вопросов создания информационных систем с применением ИИ.

Место дисциплины в структуре ОПОП: реализуется в рамках части, формируемой участниками образовательных отношений, блока Б1. Дисциплины (модули).

Планируемые результаты обучения по дисциплине:

Код компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения по дисциплине
ПКС-1	Знать: – возможности существующей программно-технической архитектуры; – методологию разработки программного обеспечения и технологию программирования; – методы и средства проектирования программного обеспечения; – типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения.	Знает: – возможности существующих программно-аппаратных средств для реализации информационных систем машинного обучения и обработки больших данных; – методологию разработки программного обеспечения для машинного обучения и обработки больших данных; – методы и средства построения баз знаний и нормализации данных для обучения.
	Уметь: – проводить оценку и обоснование рекомендуемых решений; – вырабатывать варианты реализации программного обеспечения;	Умеет: – проводить оценку и обоснование рекомендуемых программно-аппаратных решений для построения систем на базе машинного обучения и анализа больших данных; – анализировать предметную область для построения проблемно-ориентированных

	<ul style="list-style-type: none"> – применять методы и средства проектирования программного обеспечения, структур данных, программных интерфейсов. 	моделей; <ul style="list-style-type: none"> – применять методы и средства построения баз знаний и нормализации данных для обучения.
	Владеть: <ul style="list-style-type: none"> – анализом возможностей реализации требований к программному обеспечению; – навыками распределения заданий между программистами в соответствии с техническими спецификациями; – методами проектирования структур данных; – методами проектирования программных интерфейсов; – навыками осуществления обучения и наставничества. 	Владеет: <ul style="list-style-type: none"> – анализом возможностей реализации требований к программному обеспечению для построения систем на базе машинного обучения и анализа больших данных; – навыками разработки технических заданий и делегирования подзадач внутри команды разработчиков; – обширной теоретической базы для дальнейших возможностей самообразования.

II. СОДЕРЖАНИЕ И ОБЪЕМ ДИСЦИПЛИНЫ

с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу

Очная форма обучения

№ п/п	Наименование разделов и тем	Всего	Аудиторные занятия			Сам.раб.
			ЛК	ПЗ	ЛБ	
	Раздел 1. Анализ данных.	72	10	10	20	32
1.	Тема 1. Основные понятия анализа данных. Обзор прикладных задач.	8	2		2	4
2.	Тема 2. Предварительный анализ данных. Описательная статистика.	8	2		2	4
3.	Тема 3. Генеральная и выборочная совокупности.	12	2	2	4	4
4.	Тема 4. Основные понятия корреляционного анализа.	10	2	2	2	4
5.	Тема 5. Основные понятия регрессионного анализа.	12	2	2	4	4
6.	Тема 6. Снижение размерности признакового пространства.	12		2	4	6
7.	Тема 7. Классификация многомерных наблюдений.	10		2	2	6
	<i>Зачет</i>					
	<i>Итого за 7 семестр</i>	<i>72</i>	<i>10</i>	<i>10</i>	<i>20</i>	<i>32</i>
	Раздел 2. Машинное обучение.	34	4	4	8	18

8.	Тема 8. Задачи машинного обучения.	10	2		2	6
9.	Тема 9. Классификация, кластеризация, прогнозирование в машинном обучении.	12	2	2	2	6
10.	Тема 10. Реализация машинного обучения на современных ЭВМ.	12		2	4	6
	Раздел 3. Применение ИИ в прикладных задачах.	38	4	4	8	22
11.	Тема 11. Распознавание образов. Сверточные нейронные сети.	14	2	2	4	6
12.	Тема 12. Анализ временных рядов и прогнозирование.	10		2	2	6
13.	Тема 13. Интеллектуальное управление в промышленности.	14	2		2	10
	<i>Зачет</i>					
	<i>Итого за 8 семестр</i>	<i>72</i>	<i>8</i>	<i>8</i>	<i>16</i>	<i>40</i>
	в т. ч. практическая подготовка	8				
	ИТОГО:	144	18	18	36	72

Очно-заочная форма обучения не реализуется.

Заочная форма не реализуется.

III. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

Текущий контроль

Текущая аттестация проводится в форме контрольной работы, теста, реферата.

Типовой вариант контрольной работы

Задание 1. Установить и настроить систему для машинного обучения Scikit-Learn. Выполнить тестовые задания из документации (по выбору).

Задание 2. Выполнить кластеризацию методом k -средних на тестовом примере «ирисы Фишера».

Задание 3. Установить и настроить систему для машинного обучения TensorFlow внутри виртуальной машины VirtualBox.

Задание 4. Произвести обучение нейросети (перцептрон) для распознавания рукописных символов с использованием базы данных рукописных символов MNIST.

Типовой вариант тестовых заданий

A1. Аналитик это ...

- а) специалист в области анализа и моделирования**
- б) специалист в предметной области;
- в) человек, решающий определенные задачи;
- г) человек, который имеет опыт в программировании.

A2. Эксперт это ...

- а) специалист в области анализа и моделирование;
- б) специалист в предметной области;**
- в) человек, решать определенные задачи;
- г) человек, который имеет опыт в программировании.

A3. Задача классификации сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;**
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

A4. Задача регрессии сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;**
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

A5. Задача кластеризации заключается в ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.**

A6. Целью поиска ассоциативных правил является ...

- а) нахождения частых зависимостей между объектами или событиями;**
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

A7. До предполагаемых моделей относятся такие модели данных:

- а) модели классификации и последовательностей;**
- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
- в) классификации, кластеризации, исключений, итоговые и ассоциации;

- г) модели классификации, последовательностей и исключений.
- А8. В описательных моделях относятся следующие модели данных:
- а) модели классификации и последовательностей;
 - б) регрессивные, кластеризации, исключений, итоговые и ассоциации;**
 - в) классификации, кластеризации, исключений, итоговые и ассоциации;
 - г) модели классификации, последовательностей и исключений.
- А9. Модели классификации описывают ...
- а) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;**
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.
- А10. Модели последовательностей описывают ...
- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;**
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

Примерные темы рефератов

1. Применение методов машинного обучения для анализа больших данных в различных областях.
2. Обработка и анализ больших данных в реальном времени: методы и инструменты.
3. Алгоритмы кластеризации и их применение для анализа больших объемов данных.
4. Визуализация больших данных: методы и инструменты.
5. Проблемы безопасности при анализе больших данных и методы их решения.
6. Анализ больших данных в социальных сетях: выявление трендов и паттернов.
7. Применение методов глубокого обучения для анализа больших данных.
8. Большие данные в прогнозировании: методы и примеры использования.
9. Этика и правовые аспекты анализа больших данных.
10. Роль и значение анализа больших данных в современном мире: перспективы и вызовы.

Промежуточная аттестация обучающихся осуществляется в форме зачета с использованием следующих оценочных материалов: перечень вопросов к зачету.

Вопросы к зачету (7 семестр, очная форма обучения)

1. Основные понятия анализа данных.
2. Прикладные задачи, связанные с анализом данных.
3. Предварительный анализ данных.
4. Линейная регрессия.
5. Задача понижения размерности.
6. Принцип метода главных компонент.
7. Байесовский подход к распознаванию.
8. Байесовское решающее правило.
9. Байесовский наивный классификатор.
10. Задача кластеризации.
11. Метод k-средних.
12. Метод k-медиан.
13. Меры качества классификатора.
14. Основные принципы факторного анализа.
15. Описательная статистика.
16. Генеральная и выборочная совокупности.
17. Основные понятия корреляционного анализа.
18. Методы корреляционного анализа.
19. Основные понятия регрессионного анализа.
20. Снижение размерности признакового пространства.
21. Классификация многомерных наблюдений.

Вопросы к зачету с оценкой (8 семестр, очная форма обучения)

1. Задачи машинного обучения.
2. Классификация, кластеризация, прогнозирование в машинном обучении.
3. Реализация машинного обучения на современных ЭВМ.
4. Применение ИИ в прикладных задачах.
5. Распознавание образов. Сверточные нейронные сети.
6. Сети Хопфилда.
7. Автокодировщики и их применение.
8. Рекуррентные и рекурсивные нейронные сети.
9. Радиально-базисные нейронные сети.
10. ANFIS.
11. Свертка в обработке изображений.

12. Обработка изображений с применением машинного обучения.
13. Понятие обучающей выборки.
14. Обучение с учителем.
15. Обучение без учителя.
16. Обучение с подкреплением.
17. Анализ временных рядов и прогнозирование с применением машинного обучения.
18. Интеллектуальное управление в промышленности.
19. Определение термина «экспертная система»
20. Технология работы с экспертными системами
21. Структура ЭС

IV. ПЕРЕЧЕНЬ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

4.1. Основная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262> (дата обращения: 18.04.2024).
2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2021. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469022> (дата обращения: 18.04.2024).
3. Загорулько, Ю. А. Искусственный интеллект. Инженерия знаний : учебное пособие для вузов / Ю. А. Загорулько, Г. Б. Загорулько. — Москва : Издательство Юрайт, 2020. — 93 с. — (Высшее образование). — ISBN 978-5-534-07198-6. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/viewer/iskusstvennyy-intellekt-inzheneriya-znaniy-455500#page/1> (дата обращения: 18.04.2024).

4.2. Дополнительная литература

1. Бессмертный, И. А. Системы искусственного интеллекта : учебное пособие для вузов / И. А. Бессмертный. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 157 с. — (Высшее образование). — ISBN 978-5-534-07467-3. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/viewer/sistemy-iskusstvennogo-intellekta-451721#page/1> (дата обращения: 18.04.2024).

V. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

№ пп	Ссылка на информационный ресурс	Наименование разработки в электронной форме	Доступность
---------	------------------------------------	--	-------------

1.	https://infourok.ru/	Инфоурок: образовательный интернет-проект России. Включает: конспекты уроков, презентации, тесты, видеуроки и другие материалы по предметам школьной программы.	Свободный доступ
2.	http://edu.ru/	Российское образование: Федеральный портал. Включает ссылки на порталы и сайты образовательных учреждений; государственные образовательные стандарты; нормативные документы; каталог экскурсий и обучающих программ.	Свободный доступ
3.	www.intuit.ru/studies/courses	Информатика [Электронный ресурс] : открытые интернет-курсы «Интуит» //национальный открытый университет «Интуит»	Свободный доступ

VI. СОВРЕМЕННЫЕ ПРОФЕССИОНАЛЬНЫЕ БАЗЫ ДАННЫХ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ

№ пп	Ссылка на информационный ресурс	Наименование разработки в электронной форме	Доступность
1.	http://www.biblioclub.ru	Электронно-библиотечная система (ЭБС) Университетская библиотека онлайн	Регистрация через любой университетский компьютер. В дальнейшем предоставляется неограниченный индивидуальный доступ из любой точки, в которой имеется доступ к сети Интернет
2.	www.elibrary.ru	Российский информационный портал в области науки, технологии, медицины и образования	Свободный доступ
3.	https://urait.ru/	Образовательная платформа Юрайт — образовательный ресурс, электронная библиотека и интернет-магазин, где читают и покупают электронные и печатные учебники авторов — преподавателей ведущих университетов для всех уровней профессионального образования, а также пользуются видео- и аудиоматериалами, тестированием и сервисами для преподавателей, доступными 24 часа 7 дней в неделю.	Регистрация через любой университетский компьютер. В дальнейшем предоставляется неограниченный индивидуальный доступ из любой точки, в которой имеется доступ к сети Интернет

VII. ЛИЦЕНЗИОННОЕ И СВОБОДНО РАСПРОСТРАНЯЕМОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

При реализации учебной дисциплины применяется следующее лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Windows;
- LibreOffice;
- Google Chrome / Mozilla Firefox
- VrtualBox
- IPython 6.0 / JupyterLab
- SWI-Prolog

VIII. ОБОРУДОВАНИЕ И ТЕХНИЧЕСКИЕ СРЕДСТВА ОБУЧЕНИЯ, НЕОБХОДИМЫЕ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Учебные занятия проводятся в аудиториях, укомплектованных специализированной мебелью, в том числе стационарными или переносными техническими средствами обучения (проектор, экран, компьютер/ноутбук).

Лабораторные работы проводятся в специализированной лаборатории, оснащенной персональными компьютерами с возможностью работы системы виртуализации VirtualBox.

Самостоятельная работа проводится в кабинетах, оснащенных компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.