



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.01.10 Машинное обучение и анализ больших данных

Направление подготовки: 01.03.02 Прикладная математика и информатика

Направленность (профиль): Системное программирование и компьютерные технологии

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		
Лекции	12		
Лабораторные занятия	12		
Практические (семинарские) занятия	12		
в т. ч. практическая подготовка	4		
Форма(ы) промежуточной аттестации	Экзамен — 0,3		
Контроль	9		
Иные формы работы			
Самостоятельная работа	98,7		

Всего часов: 144

Трудоемкость: 4 зачетных единиц.

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
УК-3	Знать: - стратегии сотрудничества для достижения поставленной цели; - особенности поведения разных групп людей, с которыми работает/взаимодействует;	Знает: - методы и стратегии сотрудничества при решении задач реализации машинного обучения и анализа данных.
	Уметь: - определять свою роль в команде; - устанавливать разные виды коммуникации (учебную, деловую, неформальную и др.); - оценивать последствия личных действий и планировать последовательность шагов для достижения заданного результата;	Умеет: - осуществлять личностные коммуникации для достижения задач разработки и внедрения технологий машинного обучения и анализа данных.
	Владеть: - навыками эффективного взаимодействия с другими членами команды, в т.ч. участия в обмене информацией, знаниями и опытом, в презента-	Владеет: - навыками эффективного командного взаимодействия для решения задач разработки и реализации систем машинного обучения и анализа данных.

	ции результатов работы команды.	
ПКС-1	<p>Знать:</p> <ul style="list-style-type: none"> – возможности существующей программно-технической архитектуры, современных и перспективных средств разработки программных продуктов, технических средств; – методы и приемы формализации задач, языки формализации функциональных спецификаций; – принципы построения и виды архитектуры программного обеспечения; – типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения; 	<p>Знает:</p> <ul style="list-style-type: none"> – 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.
	<p>Уметь:</p> <ul style="list-style-type: none"> – вырабатывать варианты реализации программного обеспечения; – применять методы и средства проектирования компьютерного программного обеспечения, структур данных, баз данных, программных интерфейсов; 	<p>Умеет:</p> <ul style="list-style-type: none"> – проводить оценку и обоснование рекомендуемых программных решений для построения систем на базе машинного обучения и анализа больших данных; – анализировать предметную область для построения проблемно-ориентированных моделей; – применять методы и средства построения баз знаний и нормализации данных для обучения.
	<p>Владеть:</p> <ul style="list-style-type: none"> – действиями по разработке и согласованию технических спецификаций на программные компоненты; – действиями по согласованию требований к программному обеспечению с заинтересованными сторонами, распределению заданий между программистами в соответствии с техническими спецификациями, осуществлению контроля выполнения заданий, формированию отчетности в соответствии с установленными регламентами. 	<p>Владеет:</p> <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.	77	7	7	7	56
1.	Тема 1. Основные понятия анализа данных. Обзор прикладных задач.	11	1	1	1	8
2.	Тема 2. Предварительный анализ данных. Описательная статистика.	11	1	1	1	8
3.	Тема 3. Генеральная и выборочная совокупности.	11	1	1	1	8
4.	Тема 4. Основные понятия корреляционного анализа.	11	1	1	1	8
5.	Тема 5. Основные понятия регрессионного анализа.	11	1	1	1	8
6.	Тема 6. Снижение размерности признакового пространства.	11	1	1	1	8
7.	Тема 7. Классификация многомерных наблюдений.	11	1	1	1	8
	Раздел 2. Машинное обучение. Применение ИИ в прикладных задачах.	55	5	5	5	40
8.	Тема 8. Задачи машинного обучения.	11	1	1	1	8
9.	Тема 9. Классификация, кластеризация, прогнозирование в машинном обучении.	11	1	1	1	8
10.	Тема 10. Реализация машинного обучения на современных ЭВМ.	11	1	1	1	8
11.	Тема 11. Распознавание образов. Сверточные нейронные сети.	11	1	1	1	8

12.	Тема 12. Анализ временных рядов и прогнозирование.	13,7	1	1	1	10,7
	<i>Контроль</i>	9				
	<i>Экзамен</i>	0,3				
	<i>Итого за 7 семестр</i>		12	12	12	98,7
	в т. ч. практическая подготовка	4				
	ИТОГО:	144	12	12	12	98,7

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.

А3. 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.

А4. 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.

А5. 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.

А6. 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.

А7. 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 моделей 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. История создания метода главных компонент.
2. Размерность пространства состояний. Поиск полным перебором.
3. Понятие эвристического алгоритма. Прикладные задачи, связанные с эвристическими алгоритмами.
1. Истоки и проблемы ИИ: историческая справка.
2. Обзор прикладных областей ИИ.
3. Логика высказываний: таблицы истинности для логических связок.
4. Логическое следствие и вывод в логике высказываний. Метод истинностных таблиц.
5. Метод резолюций в логике высказываний.
6. Исчисление высказываний. Синтаксис и семантика в логике высказываний.
7. Аксиоматический вывод и формальные теории. Формальный вывод.
8. Исчисление высказываний как разрешимая формальная теория.
9. Синтаксис и семантика в логике предикатов. Правильно построенные формулы (ППФ).
10. Представление ППФ в предваренной форме: алгоритм и пример его применения.
11. Пространство состояний задачи ИИ. Реализация поиска на графе состояний «в глубину» и «в ширину».
12. Эвристический поиск в пространстве состояний.
13. Декларативные и императивные языки программирования. LISP и Prolog, Clojure.
14. Метaprogramмирование и ИИ.

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

**Вопросы к экзамену
(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.
22. Задачи машинного обучения.
23. Классификация, кластеризация, прогнозирование в машинном обучении.
24. Реализация машинного обучения на современных ЭВМ.
25. Применение ИИ в прикладных задачах.
26. Распознавание образов. Сверточные нейронные сети.
27. Сети Хопфилда.
28. Автокодировщики и их применение.
29. Рекуррентные и рекурсивные нейронные сети.
30. Радиально-базисные нейронные сети.
31. ANFIS.
32. Свертка в обработке изображений.
33. Обработка изображений с применением машинного обучения.
34. Понятие обучающей выборки.
35. Обучение с учителем.
36. Обучение без учителя.
37. Обучение с подкреплением.
38. Анализ временных рядов и прогнозирование с применением машинного обучения.
39. Интеллектуальное управление в промышленности.
40. Определение термина «экспертная система»

- 41.Технология работы с экспертными системами
42.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: 01.09.2023).
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: 01.09.2023).
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: 01.09.2023).

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: 01.09.2023).

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, видеоуроки 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; государ-	Свободный доступ

		ственные образовательные стандарты; нормативные документы; каталог экскурсий и обучающих программ.	
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
- VurtualBox
- IPython 6.0 / JupyterLab

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.