

Рабочая программа дисциплины (модуля) **Анализ данных**

Закреплена за кафедрой	Базовых дисциплин
Направление подготовки	27.03.04 Управление в технических системах
Профиль	Информационные технологии в управлении
Квалификация	Бакалавр
Форма обучения	очная
Общая трудоемкость	4 ЗЕТ
Часов по учебному плану	144
в том числе:	Формы контроля в семестрах:
аудиторные занятия	экзамен 6
самостоятельная работа	54
часов на контроль	61
	27

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	6 (3.2)		Итого	
Неделя	19			
Вид занятий	УП	РП	УП	РП
Лекции	18	18	18	18
Практические	36	36	36	36
Контроль самостоятельной работы	2	2	2	2
В том числе в форме практ.подготовки	22		22	
Итого ауд.	54	54	54	54
Контактная работа	56	56	56	56
Сам. работа	61	61	61	61
Часы на контроль	27	27	27	27
Итого	144	144	144	144

Программу составил(и):

к.т.н., Доцент, Уснунц-Кригер Т.Н.

Рабочая программа

Анализ данных

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - бакалавриат Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСиС» по направлению подготовки 27.03.04 Управление в технических системах (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

27.03.04 Управление в технических системах, УТС-25.plx Информационные технологии в управлении, утвержденного Ученым советом ВФ НИТУ "МИСиС" 26.12.2024, протокол № 4-24

Рабочая программа одобрена на заседании кафедры

Базовых дисциплин

Протокол от 14.05.2025 г., №9

Зав. кафедрой Уснунц-Кригер Т.Н. _____

1. ЦЕЛИ ОСВОЕНИЯ	
1.1	Целью освоения дисциплины являются овладение студентами моделями и методами интеллектуального анализа данных и машинного обучения в задачах поиска информации, обработки и анализа данных
2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	
Цикл (раздел) ОП:	Б1.В.ДВ.04
2.1	Требования к предварительной подготовке обучающегося:
2.1.1	Методы цифровой обработки
2.1.2	Протоколы сетей
2.1.3	Системы управления базами данных
2.1.4	Теория систем и системный анализ
2.1.5	Управление данными
2.1.6	Математика
2.1.7	Основы дискретной математики
2.1.8	Численные методы анализа
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:
2.2.1	Производственная практика
2.2.2	Моделирование систем управления
2.2.3	Выполнение, подготовка к процедуре защиты и защита выпускной квалификационной работы
2.2.4	Интернет-технологии
3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ	
УК-1: Способен осуществлять поиск, критический анализ и синтез информации, умение анализировать процессы и системы с использованием соответствующих аналитических, вычислительных и экспериментальных методов, применять системный подход для решения поставленных задач	
УК-1.2: Использует принципы системного подхода для решения поставленных задач	
Знать:	
УК-1.2-31 знает теорию таких дисциплин как математический анализ, дифференциальные уравнение, дискретная математика при изучении материалов курса	
ПК-1: Способен осуществлять обработку научно-технической информации и результатов исследований	
ПК-1.2: Владеет методами исследования с дальнейшей обработкой полученной информации, интерпретирует результаты и делает выводы	
Знать:	
ПК-1.2-31 модели и методы интеллектуального анализа данных и машинного обучения в задачах поиска информации, обработки и анализа данных	
УК-1: Способен осуществлять поиск, критический анализ и синтез информации, умение анализировать процессы и системы с использованием соответствующих аналитических, вычислительных и экспериментальных методов, применять системный подход для решения поставленных задач	
УК-1.2: Использует принципы системного подхода для решения поставленных задач	
Уметь:	
УК-1.2-У1 определять тип задач, строить модель и подбирать параметры методов	
ПК-1: Способен осуществлять обработку научно-технической информации и результатов исследований	
ПК-1.2: Владеет методами исследования с дальнейшей обработкой полученной информации, интерпретирует результаты и делает выводы	
Уметь:	
ПК-1.2-У1 использовать технологии и теорию искусственного интеллекта при разработке реальных автоматизированных систем обработки информации и управления	
УК-1: Способен осуществлять поиск, критический анализ и синтез информации, умение анализировать процессы и системы с использованием соответствующих аналитических, вычислительных и экспериментальных методов, применять системный подход для решения поставленных задач	
УК-1.2: Использует принципы системного подхода для решения поставленных задач	
Владеть:	

УК-1.2-В1 навыками применения математической модели и методов прикладной математики в анализе данных

ПК-1: Способен осуществлять обработку научно-технической информации и результатов исследований

ПК-1.2: Владеет методами исследования с дальнейшей обработкой полученной информации, интерпретирует результаты и делает выводы

Владеть:

ПК-1.2-В1 навыками анализа данных с помощью современных инструментальных средств

4. СТРУКТУРА И СОДЕРЖАНИЕ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература и эл. ресурсы	Примечание
	Раздел 1. Анализ данных					
1.1	Введение, основные понятия анализа данных /Лек/	6	2	УК-1.2 ПК-1.2	Э1 Э2	
1.2	Особенности методов анализа данных. Методы математической статистики. Типы статистических задач. /Лек/	6	2	УК-1.2 ПК-1.2	Э1 Э2	
1.3	Основные положения математической статистики /Пр/	6	6	УК-1.2 ПК-1.2	Э1 Э2	
1.4	Проработка лекционного материала. /Ср/	6	13	УК-1.2 ПК-1.2	Э1 Э2	
	Раздел 2. Многомерный анализ данных					
2.1	Корреляционный анализ. Регрессионный анализ. Дискриминантный анализ. Кластерный анализ. /Лек/	6	2	УК-1.2 ПК-1.2	Э1 Э2	
2.2	Линейная регрессия и классификация /Пр/	6	6	УК-1.2 ПК-1.2	Э1 Э2	
2.3	Метод наименьших квадратов: аналитическое решение и оптимизационный подход. /Пр/	6	6	УК-1.2 ПК-1.2	Э1 Э2	
2.4	Проработка лекционного материала. ДЗ Линейные методы классификации и регрессии /Ср/	6	12	УК-1.2 ПК-1.2	Э1 Э2	
	Раздел 3. Композиции алгоритмов					
3.1	Понятие алгоритма обработки данных. Представление алгоритмов. Анализ сложности алгоритмов. Блендинг алгоритмов. Понятие смещения и разброса. /Лек/	6	2	УК-1.2 ПК-1.2	Э1 Э2	
3.2	Понятие смещения и разброса (иллюстрация на примере линейных методов и решающих деревьев. Уменьшение разброса с помощью усреднения. /Пр/	6	6	УК-1.2 ПК-1.2	Э1 Э2	
3.3	Проработка лекционного материала. Метрики качества алгоритмов машинного обучения, кроссвалидация /Ср/	6	12	УК-1.2 ПК-1.2	Э1 Э2	
	Раздел 4. Особенности реальных данных					
4.1	Неполнота и противоречивость. Шумы и выбросы в данных. Методы поиска выбросов. Пропуски в данных, методы их восстановления /Лек/	6	2	УК-1.2 ПК-1.2	Э1 Э2	
4.2	Несбалансированные выборки: проблемы и методы борьбы. Задача отбора признаков, примеры подходов /Пр/	6	6	УК-1.2 ПК-1.2	Э1 Э2	
4.3	Проработка лекционного материала. /Ср/	6	12	УК-1.2 ПК-1.2	Э1 Э2	
	Раздел 5. Анализ частых множеств признаков и ассоциативных правил					
5.1	Поддержка и достоверность. Частые, замкнутые и максимальные частые множества. Алгоритм Априори. /Лек/	6	4	УК-1.2 ПК-1.2	Э1 Э2	

	Раздел 6. Кластеризация данных					
6.1	Простые эвристические подходы. Алгоритм K-Means. Проблема устойчивости результатов и важность грамотной инициализации, алгоритм K-Means++. Выбор числа кластеров. Оценка качества кластеризации. /Лек/	6	4	УК-1.2 ПК-1.2	Э1 Э2	
6.2	Кластерный анализ /Пр/	6	6	УК-1.2 ПК-1.2	Э1 Э2	
6.3	Проработка лекционного материала. ДЗ Кластеризация реальных данных /Ср/	6	12	УК-1.2 ПК-1.2	Э1 Э2	

5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

5.1. Вопросы для самостоятельной подготовки к экзамену (зачёту с оценкой)

Примерный перечень вопросов к экзамену:

1. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
2. Линейные пространства. Векторы и матрицы. Линейная независимость. Обратная матрица.
3. Производная и градиент функции. Градиентный спуск. Выпуклые функции.
4. Случайные величины. Дискретные и непрерывные распределения. Примеры.
5. Оценивание параметров распределений, метод максимального правдоподобия. Бутстрэппинг.
6. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
7. Метрики качества алгоритм регрессии и классификации.
8. Оценивание качества алгоритмов. Отложенная выборка, ее недостатки. Оценка полного скользящего контроля. Кросс-валидация. Leave-one-out.
9. Деревья решений. Методы построения деревьев. Их регуляризация.
10. Композиции алгоритмов. Разложение ошибки на смещение и разброс.
11. Случайный лес, его особенности.
12. Методы поиска выбросов в данных. Методы восстановления пропусков в данных. Работа с несбалансированными выборками.
13. Задача анализа потребительской корзины. Поддержка и достоверность. Частые, замкнутые и максимальные частые множества. Алгоритм Априори.
14. Задача кластеризации. Алгоритм K-Means. Оценки качества кластеризации.

5.2. Перечень работ, выполняемых по дисциплине (модулю, практике, НИР) - эссе, рефераты, практические и расчетно-графические работы, курсовые работы, проекты и др.

5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)

По дисциплине промежуточная аттестация предусмотрена в форме экзамена.

Экзаменационные билеты состоит из двух теоретических вопросов. Билеты хранятся на кафедре.

5.4. Методика оценки освоения дисциплины (модуля, практики. НИР)

ШКАЛА ОЦЕНИВАНИЯ ДОМАШНИХ ЗАДАНИЙ:

Оценка "зачтено" - задания выполнены полностью, расчеты выполнены верно, технически грамотно оформлены.

Оценка "не зачтено" - задания выполнены не в полном объеме, допущены ошибки в расчете и имеются недочеты в оформлении заданий.

ШКАЛА ОЦЕНИВАНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ:

а) «отлично» – студент показывает глубокие, исчерпывающие знания в объеме пройденной программы, уверенно действует по применению полученных знаний на практике, грамотно и логически стройно излагает материал при ответе, умеет формулировать выводы из изложенного теоретического материала, знает дополнительно рекомендованную литературу;

б) «хорошо» – студент показывает твердые и достаточно полные знания в объеме пройденной программы, допускает незначительные ошибки при освещении заданных вопросов, правильно действует по применению знаний на практике, четко излагает материал;

в) «удовлетворительно» – студент показывает знания в объеме пройденной программы, ответы излагает хотя и с ошибками, но уверенно исправляемыми после дополнительных и наводящих вопросов, правильно действует по применению знаний на практике;

г) «неудовлетворительно» – студент допускает грубые ошибки в ответе, не понимает сущности излагаемого вопроса, не умеет применять знания на практике, дает неполные ответы на дополнительные и наводящие вопросы.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

6.1. Рекомендуемая литература

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	Интеллектуальный анализ данных	https://www.elibrary.ru/download/elibrary_38522423_87347543.pdf
----	--------------------------------	---

Э2	Выпуклая оптимизация	https://stanford.edu/~boyd/cvxbook/
6.3 Перечень программного обеспечения		
П.1	MS Office,	
П.2	MS Teams	
6.4. Перечень информационных справочных систем и профессиональных баз данных		
И.1	Научная электронная библиотека eLIBRARY.ru – URL: https://elibrary.ru/	
И.2	Научная электронная библиотека МИСиС - URL: http://elibrary.misis.ru/login.php	
И.3	Электронная библиотечная система (ЭБС) – «Университетская библиотека онлайн» открытый круглосуточный доступ через интернет с регистрацией в библиотеке и вводом пароля.- URL: http://biblioclub.ru/	
7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ		
Ауд.	Назначение	Оснащение
11	Анализ данных	30 посадочных мест, лингафонное оборудование, 15 компьютеров для студентов. 1 компьютер для
6	Анализ данных	Компьютеры, доступ к интернету
8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ		
Весь курс разделен на самостоятельные взаимосвязанные части, т.е. имеет модульное построение. Развитие самостоятельности студентов достигается индивидуализацией вопросов для внутрисеместрового контроля знаний. Это обеспечивается методическими разработками, созданными в электронном формате, существенно повышающими эффективность самостоятельной работы студентов. Лекции проводятся с использованием мультимедийных технологий в специально оборудованных аудиториях, при этом лекционный материал демонстрируется с использованием графического редактора Power Point. На практических занятиях и при выполнении домашних занятий осваиваются как классические методы решения задач, так и с использованием пакетов прикладных программ. Такая возможность обеспечивается рациональным использованием времени при проведении лекций и практических занятий с широким привлечением мультимедийной техники, и современных пакетов прикладных программ. Дисциплина требует значительного объема самостоятельной работы. Отдельные учебные вопросы выносятся на самостоятельную проработку и контролируются посредством текущей аттестации. При этом организуются групповые и индивидуальные консультации.		