

Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»

Направление подготовки/специальность: 09.03.01 Информатика и вычислительная техника

Наименование образовательной программы: Технологии разработки интеллектуальных систем

Уровень образования: высшее образование - бакалавриат

Форма обучения: Заочная

Рабочая программа дисциплины
АНАЛИЗ ДАННЫХ

Блок:	Блок 1 «Дисциплины (модули)»
Часть образовательной программы:	Часть, формируемая участниками образовательных отношений
№ дисциплины по учебному плану:	Б1.Ч.01.05
Трудоемкость в зачетных единицах:	6 семестр - 3;
Часов (всего) по учебному плану:	108 часов
Лекции	6 семестр - 4 часа;
Практические занятия	6 семестр - 4 часа;
Лабораторные работы	не предусмотрено учебным планом
Консультации	6 семестр - 2 часа;
Самостоятельная работа	6 семестр - 96,8 часа;
в том числе на КП/КР	не предусмотрено учебным планом
Иная контактная работа	6 семестр - 0,9 часа;
включая: Тестирование Контрольная работа	
Промежуточная аттестация:	
Зачет с оценкой	6 семестр - 0,3 часа;

Москва 2024

ПРОГРАММУ СОСТАВИЛ:

Преподаватель

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Раскатова М.В.
	Идентификатор	R6bc62db2-RaskatovaMV-ead4381

М.В. Раскатова

СОГЛАСОВАНО:

Руководитель
образовательной программы

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Вишняков С.В.
	Идентификатор	R35b26072-VishniakovSV-02810d9

С.В. Вишняков

Заведующий выпускающей
кафедрой

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Вишняков С.В.
	Идентификатор	R35b26072-VishniakovSV-02810d9

С.В. Вишняков

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины: Изучение методов анализа данных и получении практических навыков в области обработки и анализа данных в бизнес аналитике.

Задачи дисциплины

- освоение существующих технологий подготовки данных к анализу;
- освоение основных методов анализа данных;
- приобретение навыков обработки и анализа данных в различных средах;
- изучение методов поиска в данных внутренних закономерностей, взаимосвязей, тенденций.

Формируемые у обучающегося **компетенции** и запланированные **результаты обучения** по дисциплине, соотнесенные с **индикаторами достижения компетенций**:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения
ПК-2 Способен осуществлять работу с информационными технологиями	ИД-1ПК-2 Применяет методы моделирования и осуществляет анализ результатов для моделирования работы вычислительных систем и сетей ЭВМ	знать: - основные понятия анализа данных; - принципы, методы и средства решения задач анализа данных: классификации и кластеризации. уметь: - искать и анализировать необходимую информацию для решения задач прогнозирования.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ВО

Дисциплина относится к основной профессиональной образовательной программе Технологии разработки интеллектуальных систем (далее – ОПОП), направления подготовки 09.03.01 Информатика и вычислительная техника, уровень образования: высшее образование - бакалавриат.

Базируется на уровне среднего общего образования.

Результаты обучения, полученные при освоении дисциплины, необходимы при выполнении выпускной квалификационной работы.

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

№ п/п	Разделы/темы дисциплины/формы промежуточной аттестации	Всего часов на раздел	Семестр	Распределение трудоемкости раздела (в часах) по видам учебной работы										Содержание самостоятельной работы/ методические указания	
				Контактная работа							СР				
				Лек	Лаб	Пр	Консультация		ИКР		ПА	Работа в семестре	Подготовка к аттестации /контроль		
КПР	ГК	ИККП	ТК												
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Введение. Основные понятия	23.80	6	1.0	-	-	-	0.50	-	0.30	-	22	-	<p><u>Подготовка к текущему контролю:</u> Работа ориентирована на изучение основных понятий, используемых при работе с данными</p> <p><u>Самостоятельное изучение теоретического материала:</u> Работа ориентирована на изучение литературных источников, конспектирование материалов</p> <p><u>Изучение материалов литературных источников:</u> [2], стр. 5-15</p>	
1.1	Введение. Основные понятия. Задачи анализа данных	11.90		0.5	-	-	-	0.25	-	0.15	-	11	-		
1.2	Данные. Методы, стадии, процесс Data Mining	11.90		0.5	-	-	-	0.25	-	0.15	-	11	-		
2	Классификация и кластеризация	25.80		1.0	-	1.0	-	0.50	-	0.30	-	23	-		<p><u>Подготовка к текущему контролю:</u> Повторение материала по разделу "Классификация и кластеризация"</p> <p><u>Изучение материалов литературных источников:</u> [1], стр. 60-75</p>
2.1	Классификация	12.40		0.5	-	0.5	-	0.25	-	0.15	-	11	-		
2.2	Кластеризация	13.40		0.5	-	0.5	-	0.25	-	0.15	-	12	-		
3	Задачи анализа данных	40.40		2.0	-	3	-	1.00	-	0.3	-	34.1	-		<p><u>Подготовка к текущему контролю:</u> Повторение материала по разделу "Задачи анализа данных"</p> <p><u>Самостоятельное изучение теоретического материала:</u> Изучение дополнительного материала по разделу "Задачи анализа данных"</p> <p><u>Изучение материалов литературных источников:</u> [1], стр. 55-65 [3], стр. 254-270</p>
3.1	Деревья решений	12.85		0.5	-	1	-	0.25	-	0.1	-	11	-		
3.2	Ассоциативные правила. Прогнозирование	13.2		0.5	-	1	-	0.5	-	0.1	-	11.1	-		
3.3	Визуальный анализ данных	14.35	1	-	1	-	0.25	-	0.1	-	12	-			

	Зачет с оценкой	18.0		-	-	-	-	-	-	-	0.3	-	17.7	
	Всего за семестр	108.00		4.0	-	4.0	-	2.00	-	0.90	0.3	79.1	17.7	
	Итого за семестр	108.00		4.0	-	4.0	2.00		0.90	0.3		96.8		

Примечание: Лек – лекции; Лаб – лабораторные работы; Пр – практические занятия; КПр – аудиторные консультации по курсовым проектам/работам; ИККП – индивидуальные консультации по курсовым проектам/работам; ГК- групповые консультации по разделам дисциплины; СР – самостоятельная работа студента; ИКР – иная контактная работа; ТК – текущий контроль; ПА – промежуточная аттестация

3.2 Краткое содержание разделов

1. Введение. Основные понятия

1.1. Введение. Основные понятия. Задачи анализа данных

Понятие Data Mining, возникновение, перспективы, проблемы. Отличия Data Mining от других методов анализа данных. Перспективы технологии Data Mining. Сферы применения технологии Data Mining. Задачи анализа данных Data Mining: классификация, кластеризация, прогнозирование, ассоциация, визуализация, анализ и обнаружение отклонений, оценивание, анализ связей, подведение итогов.

1.2. Данные. Методы, стадии, процесс Data Mining

Наборы данных и их атрибуты. Измерения, шкалы. Типы наборов данных. Форматы хранения. Классификация видов данных. Методы и стадии Data Mining. Процесс Data Mining. Начальные этапы. Анализ предметной области. Постановка задачи. Подготовка данных. Классификация методов Data Mining. Свойства методов Data Mining. Процесс Data Mining. Начальные этапы.

2. Классификация и кластеризация

2.1. Классификация

Классификация. Постановка задачи классификации и представление результатов. Виды классификации. Процесс классификации. Методы, применяемые для решения задач классификации. Точность классификации.

2.2. Кластеризация

Кластеризация. Постановка задачи. Сравнение задач классификации и кластеризации. Оценка качества кластеризации. Применение кластерного анализа. Процесс Data Mining. Начальные этапы. Метод деревьев решений. Преимущества. Процесс конструирования дерева решений. Алгоритмы, реализующие деревья решений. Области применения деревьев решений.

3. Задачи анализа данных

3.1. Деревья решений

Метод деревьев решений. Преимущества деревьев решений. Процесс конструирования дерева решений. Алгоритмы. Важные аспекты построения дерева решений. Области применения деревьев решений.

3.2. Ассоциативные правила. Прогнозирование

Ассоциативные правила. Формальная постановка задачи. Представление результатов. Границы поддержки и достоверности ассоциативного правила. Методы поиска ассоциативных правил. Алгоритм Apriori. Описание алгоритма, его разновидности. Свойство анти- монотонности.

3.3. Визуальный анализ данных

Визуализация данных. Постановка задачи. Методы и средства визуального представления информации, характеристики средств визуализации. Способы представления информации в одно-, двух-, трехмерном измерениях. Способы отображения информации в более чем трех измерениях. Качество визуализации. Основные тенденции в области визуализации.

3.3. Темы практических занятий

1. Деревья решений;
2. Кластерный анализ;
3. Прогнозирование. Визуальный анализ данных.

3.4. Темы лабораторных работ

не предусмотрено

3.5 Консультации

Групповые консультации по разделам дисциплины (ГК)

1. Рассмотрение основных понятий и терминов, применяемых при анализе данных.
Формулирование задач анализа данных
2. Обсуждение материалов по кейсам раздела "Классификация и кластеризация"
3. Обсуждение материалов по кейсам раздела "Задачи анализа данных"

3.6 Тематика курсовых проектов/курсовых работ

Курсовой проект/ работа не предусмотрены

3.7. Соответствие разделов дисциплины и формируемых в них компетенций

Запланированные результаты обучения по дисциплине (в соответствии с разделом 1)	Коды индикаторов	Номер раздела дисциплины (в соответствии с п.3.1)			Оценочное средство (тип и наименование)
		1	2	3	
Знать:					
принципы, методы и средства решения задач анализа данных: классификации и кластеризации	ИД-1ПК-2		+		Тестирование/Data Mining
основные понятия анализа данных	ИД-1ПК-2	+			Тестирование/Понятия анализа данных
Уметь:					
искать и анализировать необходимую информацию для решения задач прогнозирования	ИД-1ПК-2			+	Контрольная работа/Методы анализа данных

4. КОМПЕТЕНТНОСТНО-ОРИЕНТИРОВАННЫЕ ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ДИСЦИПЛИНЕ)

4.1. Текущий контроль успеваемости

6 семестр

Форма реализации: Компьютерное задание

1. Понятия анализа данных (Тестирование)
2. Data Mining (Тестирование)

Форма реализации: Письменная работа

1. Методы анализа данных (Контрольная работа)

Балльно-рейтинговая структура дисциплины является приложением А.

4.2 Промежуточная аттестация по дисциплине

Зачет с оценкой (Семестр №6)

Оценка определяется по совокупности результатов текущего контроля успеваемости в соответствии с Положением о балльно-рейтинговой системе для студентов НИУ «МЭИ»

В диплом выставляется оценка за 6 семестр.

Примечание: Оценочные материалы по дисциплине приведены в фонде оценочных материалов ОПОП.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Печатные и электронные издания:

1. Анализ данных : учебник для академического бакалавриата вузов по экономическим направлениям и специальностям / Нац. исслед. ун-т "Высшая школа экономики" ; ред. В. С. Мхитарян. – М. : Юрайт, 2017. – 490 с. – (Бакалавр. Академический курс). – ISBN 978-5-534-00616-2.;
2. Мидлтон, М. Р. Анализ статистических данных с использованием Microsoft Excel для Office XP : пер. с англ. / М. Р. Мидлтон. – М. : БИНОМ. Лаборатория знаний, 2013. – 296 с. – (Практикум). – ISBN 978-5-94774-216-9.;
3. Чубукова И. А.- "Data Mining", (2-е изд.), Издательство: "ИНТУИТ", Москва, 2016 - (470 с.)
<https://e.lanbook.com/book/100582>.

5.2 Лицензионное и свободно распространяемое программное обеспечение:

1. СДО "Прометей";
2. Office / Российский пакет офисных программ;
3. Windows / Операционная система семейства Linux;
4. Видеоконференции (Майнд, Сберджаз, ВК и др).

5.3 Интернет-ресурсы, включая профессиональные базы данных и информационно-справочные системы:

1. ЭБС Лань - <https://e.lanbook.com/>

2. ЭБС "Университетская библиотека онлайн" - http://biblioclub.ru/index.php?page=main_ub_red
3. Научная электронная библиотека - <https://elibrary.ru/>
4. Электронная библиотека МЭИ (ЭБ МЭИ) - <http://elib.mpei.ru/login.php>
5. Портал открытых данных Российской Федерации - <https://data.gov.ru>
6. База открытых данных Министерства труда и социальной защиты РФ - <https://rosmintrud.ru/opendata>
7. База открытых данных профессиональных стандартов Министерства труда и социальной защиты РФ - <http://profstandart.rosmintrud.ru/obshchiy-informatsionnyy-blok/natsionalnyy-reestr-professionalnykh-standartov/>
8. База открытых данных Министерства экономического развития РФ - <http://www.economy.gov.ru>
9. База открытых данных Росфинмониторинга - <http://www.fedsfm.ru/opendata>
10. Электронная открытая база данных "Polpred.com Обзор СМИ" - <https://www.polpred.com>

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Тип помещения	Номер аудитории, наименование	Оснащение
Учебные аудитории для проведения лекционных занятий и текущего контроля	Ж-417/6, Белая мультимедийная студия	стол компьютерный, доска интерактивная, компьютерная сеть с выходом в Интернет, мультимедийный проектор, компьютер персональный
	Ж-417/7, Световая черная студия	стул, компьютерная сеть с выходом в Интернет, микрофон, мультимедийный проектор, экран, оборудование специализированное, компьютер персональный
Учебные аудитории для проведения практических занятий, КР и КП	Ж-417/1, Компьютерный класс ИДДО	стол преподавателя, стол компьютерный, шкаф для документов, шкаф для одежды, стол письменный, компьютерная сеть с выходом в Интернет, доска маркерная передвижная, компьютер персональный, принтер, кондиционер, стенд информационный
Учебные аудитории для проведения промежуточной аттестации	Ж-417/1, Компьютерный класс ИДДО	стол преподавателя, стол компьютерный, шкаф для документов, шкаф для одежды, стол письменный, компьютерная сеть с выходом в Интернет, доска маркерная передвижная, компьютер персональный, принтер, кондиционер, стенд информационный
Помещения для самостоятельной работы	НТБ-303, Лекционная аудитория	стол компьютерный, стул, стол письменный, вешалка для одежды, компьютерная сеть с выходом в Интернет, компьютер персональный, принтер, кондиционер
Помещения для консультирования	Ж-200б, Конференц-зал ИДДО	стол, стул, компьютер персональный, кондиционер
Помещения для хранения оборудования и учебного инвентаря	Ж-417 /2а, Помещение для инвентаря	стеллаж для хранения инвентаря, экран, указка, архивные документы, дипломные и курсовые работы студентов, канцелярский принадлежности, спортивный инвентарь, хозяйственный инвентарь, запасные

		комплектующие для оборудования
--	--	--------------------------------

БАЛЛЬНО-РЕЙТИНГОВАЯ СТРУКТУРА ДИСЦИПЛИНЫ

Анализ данных

(название дисциплины)

6 семестр

Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:

КМ-1 Понятия анализа данных (Тестирование)

КМ-2 Data Mining (Тестирование)

КМ-3 Методы анализа данных (Контрольная работа)

Вид промежуточной аттестации – Зачет с оценкой.

Номер раздела	Раздел дисциплины	Индекс КМ:	КМ-1	КМ-2	КМ-3
		Неделя КМ:	6	11	16
1	Введение. Основные понятия				
1.1	Введение. Основные понятия. Задачи анализа данных		+		
1.2	Данные. Методы, стадии, процесс Data Mining		+		
2	Классификация и кластеризация				
2.1	Классификация			+	
2.2	Кластеризация			+	
3	Задачи анализа данных				
3.1	Деревья решений				+
3.2	Ассоциативные правила. Прогнозирование				+
3.3	Визуальный анализ данных				+
Вес КМ, %:			30	30	40