

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

Направление подготовки/специальность: 09.04.03 Прикладная информатика

Наименование образовательной программы: Прикладная информатика в энергетике

Уровень образования: высшее образование - магистратура

Форма обучения: Очная

Рабочая программа дисциплины
АНАЛИЗ ДАННЫХ

Блок:	Блок 1 «Дисциплины (модули)»
Часть образовательной программы:	Часть, формируемая участниками образовательных отношений
№ дисциплины по учебному плану:	Б1.Ч.05
Трудоемкость в зачетных единицах:	2 семестр - 4;
Часов (всего) по учебному плану:	144 часа
Лекции	2 семестр - 16 часов;
Практические занятия	не предусмотрено учебным планом
Лабораторные работы	2 семестр - 32 часа;
Консультации	2 семестр - 2 часа;
Самостоятельная работа	2 семестр - 93,5 часа;
в том числе на КП/КР	не предусмотрено учебным планом
Иная контактная работа	проводится в рамках часов аудиторных занятий
включая:	
Индивидуальный проект	
Дискуссия	
Промежуточная аттестация:	
Экзамен	2 семестр - 0,5 часа;

Москва 2024

ПРОГРАММУ СОСТАВИЛ:

Преподаватель

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Гашо Е.Г.
	Идентификатор	R913da1fa-GashoYG-eb0efe14

Е.Г. Гашо

СОГЛАСОВАНО:

Руководитель
образовательной
программы

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Щербатов И.А.
	Идентификатор	R6b2590a8-ShcherbatovIA-d91ec17

И.А. Щербатов

Заведующий выпускающей
кафедрой

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Яворовский Ю.В.
	Идентификатор	R7e35b260-YavorovskyYV-dabb149

Ю.В.
Яворовский

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины: познакомить студентов с современными инструментами, процессами и технологиями аналитики данных. В рамках курса необходимые навыки будут закреплены в low-code платформе Loginom. Также важным является знакомство студентов с приемами проектирования сценариев анализа данных для решения прикладных задач предприятий энергетики.

Задачи дисциплины

- Освоение принципов, стандартов технологии анализа данных;
- Освоение инструментов преобразования, визуализации, очистки данных и стандартизации нормативно-справочной информации;
- Освоение инструментов загрузки и связывания данных из разных источников (файлов, баз-данных и веб-сервисов), работа с внешними библиотеками;
- Освоение понятий Data Mining, Big Data, Knowledge Discovery.

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-1 Способен участвовать в проектировании информационных систем для предприятий энергетики	ИД-4 _{ПК-1} Осуществляет решение задач бизнес-анализа для проектирования информационных систем	знать: - Основные принципы, технологии и стандарты бизнес-анализа. уметь: - Проектировать сценарии анализа данных сверху вниз для работы с любыми данными, соответствующей заложенной структуре; - Разрабатывать сценарии обработки данных в аналитической платформе; - Использовать на практике современные low-code аналитические платформы, востребованные в российских проектных компаниях.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ВО

Дисциплина относится к основной профессиональной образовательной программе Прикладная информатика в энергетике (далее – ОПОП), направления подготовки 09.04.03 Прикладная информатика, уровень образования: высшее образование - магистратура.

Требования к входным знаниям и умениям:

- знать Системы управления базами данных
- знать Методы математического моделирования
- знать Методы теории принятия решений
- уметь Ставить задачи поддержки принятия решений для предприятий энергетики
- уметь Формализовывать решения прикладных задач на основе математических методов

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 4 зачетных единицы, 144 часа.

№ 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 в аналитику данных	16	2	2	4	-	-	-	-	-	-	10	-	<u>Самостоятельное изучение теоретического материала:</u> Самостоятельное изучение теоретического материала по разделу "Введение в аналитику данных" <u>Изучение материалов литературных источников:</u> [2], 4-15	
1.1	Анализ данных как процесс	8		1	2	-	-	-	-	-	-	5	-		
1.2	Структурированные данные	8		1	2	-	-	-	-	-	-	5	-		
2	Технологии и инструменты аналитики данных	16		2	4	-	-	-	-	-	-	-	10	-	<u>Самостоятельное изучение теоретического материала:</u> Самостоятельное изучение теоретического материала по разделу "Технологии и инструменты аналитики данных" <u>Изучение материалов литературных источников:</u> [2], 21-39
2.1	Технологии и прикладные задачи аналитики данных	8		1	2	-	-	-	-	-	-	5	-		
2.2	Инструменты аналитики данных. Большие данные. Наука о данных	8		1	2	-	-	-	-	-	-	5	-		
3	Методы подготовки данных к анализу	22		4	8	-	-	-	-	-	-	-	10	-	<u>Самостоятельное изучение теоретического материала:</u> Самостоятельное изучение теоретического материала по разделу "Методы подготовки данных к анализу" <u>Изучение материалов литературных источников:</u> [2], 42-83
3.1	Процесс ETL	11		2	4	-	-	-	-	-	-	5	-		
3.2	Трансформация данных	11		2	4	-	-	-	-	-	-	5	-		
4	Визуализация данных	16		2	4	-	-	-	-	-	-	10	-		
4.1	Цели, задачи, этапы и	16	2	4	-	-	-	-	-	-	10	-	<u>Самостоятельное изучение теоретического материала:</u>		

	методы визуализации												Самостоятельное изучение теоретического материала по разделу "Визуализация данных" <u>Изучение материалов литературных источников:</u> [1], 6-28
5	Стандартизация и очистка нормативно-справочной информации	16	2	4	-	-	-	-	-	-	10	-	<u>Самостоятельное изучение теоретического материала:</u> Стандартизация и очистка нормативно-справочной информации <u>Изучение материалов литературных источников:</u> [1], 32-91
5.1	Понятия очистки и стандартизации данных	16	2	4	-	-	-	-	-	-	10	-	<u>Самостоятельное изучение теоретического материала:</u> Модели Data Mining <u>Изучение материалов литературных источников:</u> [3], 4-42
6	Модели Data Mining	22	4	8	-	-	-	-	-	-	10	-	
6.1	Описательные модели Data Mining: кластеризация, ассоциативные правила	11	2	4	-	-	-	-	-	-	5	-	
6.2	Предсказательные модели Data Mining: классификация, регрессия	11	2	4	-	-	-	-	-	-	5	-	
	Экзамен	36.0	-	-	-	-	2	-	-	0.5	-	33.5	
	Всего за семестр	144.0	16	32	-	-	2	-	-	0.5	60	33.5	
	Итого за семестр	144.0	16	32	-	2	-	-	0.5	0.5	93.5		

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 в аналитику данных

1.1. Анализ данных как процесс

Термин «анализ данных». Концепция анализа данных Дж. Тьюки. Этапы анализа данных по Тьюки. Процесс анализа данных по Тьюки. Составляющие аналитики данных. Принципы анализа данных. Этапы моделирования. Методология CRISP-DM. Преимущества методологии..

1.2. Структурированные данные

Формы представления данных. Типы данных. Номинальные и ординальные переменные. Числовые, дискретные и непрерывные данные. Представление наборов данных. Подготовка данных к анализу. Особенности бизнес-данных. Принципы формализации данных. Методы сбора данных. Информативность данных. Требования к данным..

2. Технологии и инструменты аналитики данных

2.1. Технологии и прикладные задачи аналитики данных

Предпосылки изменения подходов к анализу данных. Особенности задач нового типа. Трактовки термина «бизнес-аналитика». Knowledge Discovery— обнаружение знаний. Data Mining. Классы моделей Data Mining. Машинное обучение..

2.2. Инструменты аналитики данных. Большие данные. Наука о данных

Причины распространения KDD и Data Mining. Современные программные инструменты аналитики данных. Инструменты традиционного BI (Business Intelligence). Настольные пакеты и библиотеки. Облачные сервисы. Аналитические платформы..

3. Методы подготовки данных к анализу

3.1. Процесс ETL

ETL-процесс. Преобразование данных как часть ETL-операций..

3.2. Трансформация данных

Основные методы трансформации данных: преобразование временных данных, квантование, сортировка, слияние, группировка и др. Группировка данных. Функции агрегации. Преимущества группировки. Операции с датой и временем. Общие сведения об обогащении данных. Слияние (Внутреннее соединение. Левое и правое соединение. Полное соединение. Разность). Соединение. Дополнение данных. Объединение. Цели квантования. Результат квантования. Выбор числа интервалов. Методы квантования. Пример квантования. Преобразование упорядоченных данных. Скользящее окно. Структура данных. Представление данных. Некорректная структура. Понятие транспонирования. Транспонирование на этапе ETL. Обратное транспонирование..

4. Визуализация данных

4.1. Цели, задачи, этапы и методы визуализации

Визуализация и бизнес-аналитика. Цели и задачи визуализации. Этапы визуализации. Методы визуализации. Визуализатор "Статистика" в Logiplot. Визуализаторы общего назначения. Таблицы. Графики. Диаграммы. Круговые диаграммы. Гистограммы. Визуализатор "Диаграмма" в Logiplot. Сложные визуализаторы общего назначения. OLAP-

анализ. Географические карты. Тепловые карты. Плоское дерево. Диаграмма связей. Облако тегов. Пузырьковая диаграмма. Диаграмма рассеяния. Визуализатор "OLAP-куб" в Loginom..

5. Стандартизация и очистка нормативно-справочной информации

5.1. Понятия очистки и стандартизации данных

Объемы данных. Качество данных. Причины загрязнения. Виды ошибок в данных. Понятия очистки и стандартизации данных. Причины необходимости очистки. Показатели качества данных. Основные методы очистки. Использование справочников и таблиц замены. Анализ строк. Регулярные выражения. Частотный анализ. Контрольные числа. Регулярные выражения, история возникновения. Диалекты регулярных выражений. Возможности регулярных выражений. Инструменты для проверки. Метасимволы. Квантификаторы. Регулярные выражения в Loginom..

6. Модели Data Mining

6.1. Описательные модели Data Mining: кластеризация, ассоциативные правила

Цели кластеризации. Алгоритм кластеризации k-means. Сети и карты Кохонена. Проблемы алгоритмов кластеризации. Введение в ассоциативные правила. Значимость ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.

6.2. Предсказательные модели Data Mining: классификация, регрессия

Применение классификации и регрессии. Классификация данных. Обзор методов классификации и регрессии. Статистические методы (Линейная и логистическая регрессия). Методы, основанные на обучении (Деревья решений, Искусственные нейронные сети, Метод k ближайших соседей)..

3.3. Темы практических занятий

не предусмотрено

3.4. Темы лабораторных работ

1. Основы работы в Loginom;
2. Работа с переменными в Loginom;
3. Работа с основными компонентами управления Loginom;
4. Трансформация данных: группировка, преобразование даты и обогащение данных;
5. Трансформация данных: квантование, скользящее окно и транспонирование данных;
6. Визуализаторы общего назначения, простые: таблицы, графики, диаграммы, круговые диаграммы, гистограммы.;
7. Визуализаторы общего назначения, сложные: OLAP-анализ, географические карты, тепловые карты, плоское дерево, диаграмма связей, облако тегов, пузырьковая диаграмма, диаграмма рассеяния.;
8. Методы очистки данных;
9. Основы регулярных выражений. Объединение методов;
10. Описательные модели Data Mining: кластеризация, ассоциативные правила;
11. Предсказательные модели Data Mining: классификация, регрессия.

3.5 Консультации

Групповые консультации по разделам дисциплины (ГК)

1. Консультация по разделу "Введение в аналитику данных"
2. Консультация по разделу "Технологии и инструменты аналитики данных"
3. Консультация по разделу "Методы подготовки данных к анализу"
4. Консультация по разделу "Визуализация данных"
5. Консультация по разделу "Стандартизация и очистка нормативно-справочной информации"
6. Консультация по разделу "Модели Data Mining"

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	4	5	6	
Знать:								
Основные принципы, технологии и стандарты бизнес-анализа	ИД-4ПК-1	+	+	+		+	+	Дискуссия/Устный опрос № 1
Уметь:								
Использовать на практике современные low-code аналитические платформы, востребованные в российских проектных компаниях	ИД-4ПК-1				+		+	Индивидуальный проект/Контрольная работа № 3
Разрабатывать сценарии обработки данных в аналитической платформе	ИД-4ПК-1				+		+	Индивидуальный проект/Контрольная работа № 2
Проектировать сценарии анализа данных сверху вниз для работы с любыми данными, соответствующей заложенной структуре	ИД-4ПК-1	+	+	+		+	+	Индивидуальный проект/Контрольная работа № 1

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

4.1. Текущий контроль успеваемости

2 семестр

Форма реализации: Компьютерное задание

1. Контрольная работа № 1 (Индивидуальный проект)
2. Контрольная работа № 2 (Индивидуальный проект)
3. Контрольная работа № 3 (Индивидуальный проект)

Форма реализации: Проверка задания

1. Устный опрос № 1 (Дискуссия)

Балльно-рейтинговая структура дисциплины является приложением А.

4.2 Промежуточная аттестация по дисциплине

Экзамен (Семестр №2)

В диплом выставляется оценка за 2 семестр.

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

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Печатные и электронные издания:

1. Кузьмин В. И., Гадзаов А. Ф.- "Методы анализа данных", (2-е изд., перераб. и доп.), Издательство: "РТУ МИРЭА", Москва, 2020 - (155 с.)
<https://e.lanbook.com/book/171433>;
2. Прокопенко Н. Ю.- "Аналитические информационные системы поддержки принятия решений", Издательство: "ННГАСУ", Нижний Новгород, 2020 - (142 с.)
<https://e.lanbook.com/book/164866>;
3. Митина О. А., Юрченков И. А.- "Технологии организации, обработки и хранения статистических данных", Издательство: "РТУ МИРЭА", Москва, 2019 - (163 с.)
<https://e.lanbook.com/book/171511>.

5.2 Лицензионное и свободно распространяемое программное обеспечение:

1. СДО "Прометей";
2. Office / Российский пакет офисных программ;
3. Windows / Операционная система семейства Linux.

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. Национальная электронная библиотека - <https://rusneb.ru/>

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Тип помещения	Номер аудитории, наименование	Оснащение
Учебные аудитории для проведения лекционных занятий и текущего контроля	Г-404, Учебная аудитория	парта со скамьей, стол преподавателя, стул, доска меловая, компьютерная сеть с выходом в Интернет
Учебные аудитории для проведения лабораторных занятий	Ж-207, Компьютерный класс ИВЦ	стол, стул, вешалка для одежды, доска меловая, компьютерная сеть с выходом в Интернет, компьютер персональный, кондиционер
Учебные аудитории для проведения промежуточной аттестации	Г-407, Учебная аудитория	парта, стол преподавателя, стул, доска меловая
Помещения для самостоятельной работы	НТБ-303, Компьютерный читальный зал	стол компьютерный, стул, стол письменный, вешалка для одежды, компьютерная сеть с выходом в Интернет, компьютер персональный, принтер, кондиционер
Помещения для консультирования	В-204, Кабинет сотрудников каф. "ПТС"	стеллаж, стол преподавателя, стол для оргтехники, стул, шкаф для документов, шкаф для одежды, компьютер персональный, принтер, холодильник
Помещения для хранения оборудования и учебного инвентаря	В-206, Кабинет сотрудников каф. "ПТС"	стул, шкаф для документов, стол письменный, кондиционер, дипломные и курсовые работы студентов

БАЛЛЬНО-РЕЙТИНГОВАЯ СТРУКТУРА ДИСЦИПЛИНЫ

Анализ данных

(название дисциплины)

2 семестр

Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:

КМ-1 Контрольная работа № 1 (Индивидуальный проект)

КМ-2 Устный опрос № 1 (Дискуссия)

КМ-3 Контрольная работа № 2 (Индивидуальный проект)

КМ-4 Контрольная работа № 3 (Индивидуальный проект)

Вид промежуточной аттестации – Экзамен.

Номер раздела	Раздел дисциплины	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4
		Неделя КМ:	4	8	12	16
1	Введение в аналитику данных					
1.1	Анализ данных как процесс		+	+		
1.2	Структурированные данные		+	+		
2	Технологии и инструменты аналитики данных					
2.1	Технологии и прикладные задачи аналитики данных		+	+		
2.2	Инструменты аналитики данных. Большие данные. Наука о данных		+	+		
3	Методы подготовки данных к анализу					
3.1	Процесс ETL		+	+		
3.2	Трансформация данных		+	+		
4	Визуализация данных					
4.1	Цели, задачи, этапы и методы визуализации				+	+
5	Стандартизация и очистка нормативно-справочной информации					
5.1	Понятия очистки и стандартизации данных		+	+		
6	Модели Data Mining					
6.1	Описательные модели Data Mining: кластеризация, ассоциативные правила		+	+	+	+

6.2	Предсказательные модели Data Mining: классификация, регрессия	+	+	+	+
Вес КМ, %:		25	25	25	25