МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ)

Рабочая программа дисциплины (модуля),
как компонент образовательной программы
высшего образования - программа магистратуры
по направлению подготовки
27.04.02 Управление качеством,
утвержденной первым проректором ФГАОУ ВО
РУТ (МИИТ) Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Методы анализа больших данных

Направление подготовки: 27.04.02 Управление качеством

Направленность (профиль): Управление качеством в производственно-технологических системах

Форма обучения: Очно-заочная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная ФГАОУ ВО РУТ (МИИТ)
ID подписи: 581797
Подписал: заведующий кафедрой Гуськова Марина Федоровна
Дата: 01.09.2026

1. Общие сведения о дисциплине (модуле).

Целями освоения учебной дисциплины являются изучение проблем и технологий работы с большими данными (Big Data), формирование представления о возможностях и ограничениях основных методов обработки и анализа больших данных. К предметным областям, на которых основывается данная учебная дисциплина, относятся математика, статистика, информатика, машинное обучение, нейронные сети, кластерный и интеллектуальный анализ, базы данных и визуализация. Учебный курс носит междисциплинарный характер, а методы анализа могут сильно отличаться друг от друга и зависеть от конкретной предметной области.

Задачи учебного курса: изучение теоретических основ хранения и анализа больших данных, включая базовые элементы статистического программирования и интеллектуального анализа больших наборов данных, проведение расчетов с применением технологий анализа больших данных и решение широкого спектра прикладных задач обработки больших данных.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

УК-4 - Способен применять современные коммуникативные технологии, в том числе на иностранном (ых) языке (ах), для академического и профессионального взаимодействия.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

основные условия равновесия тел и их систем на основе математического аппарата, используемого для этих целей.

Уметь:

выполнять математические операции и действия на основе уравнений равновесия тел и их систем

Владеть:

основами теории статического равновесия на основе законов статики

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 9 з.е. (324 академических часа(ов).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №1
Контактная работа при проведении учебных занятий (всего):	40	40
В том числе:		
Занятия лекционного типа	8	8
Занятия семинарского типа	32	32

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 284 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	Раздел 1. Введение в науку о больших данных. История и причины появления термина «большие данные» (Big Data). Характеристики и источники Big Data. Четыре основных типа больших данных. Аналитика больших данных. Задачи, решаемые Big Data. Понятие жизненного цикла аналитики больших данных.
2	Раздел 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных. Классификация задач анализа больших данных. Основные характеристики и типы больших данных. Алгоритмы обработки первичной обработки больших данных. Современные подходы к обработке и хранению больших данных. Процесс анализа больших данных. Общая схема анализа. Извлечение и

№ п/п	Тематика лекционных занятий / краткое содержание
	визуализация больших данных. Этапы моделирования. Процесс построения моделей. Формы представления, типы и виды больших данных. Представления наборов больших данных. Системы управления базами данных. Использование языка SQL (DDL) для создания структур больших данных. Запросы на языке SQL (DML). NoSQL хранилища больших данных. Хранилища больших данных ключ-значение. Документо-ориентированные хранилища больших данных. Графовые хранилища больших данных. Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.
3	Раздел 3. Интеллектуальный анализ данных. Введение в Data Mining. Большие данные и метаданные. Методы и стадии Data Mining. Задачи Data Mining. Извлечение больших данных, информации и знаний. Сферы применения Data Mining. Технологии KDD и Data Mining. Подготовка больших данных к анализу. Методика извлечения знаний. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. Программное обеспечение в области анализа больших данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. Машинное обучение и разработка больших данных. Основные аналитические методы обработки больших данных. Машинное обучение и майнинг больших данных. Распознавание образов. Нейронные сети как реализация алгоритмов машинного обучения. Облачные платформы интеллектуального анализа больших данных. Разработка методов, алгоритмов и программных средств интеллектуального анализа больших данных: Hadoop, Spark и IoT. Краудсорсинг — категоризация и обогащение больших данных силами широкого, неопределённого круга лиц, привлечённых на основании публичной оферты, без вступления в трудовые отношения. Смешение и интеграция больших данных — набор техник, позволяющих интегрировать разнородные большие данные из разнообразных источников для возможности глубинного анализа. Цифровая обработка сигналов и обработка естественного языка (включая тональный анализ). Сетевой анализ и оптимизация больших данных. Генетические алгоритмы. Имитационное моделирование. Пространственный анализ — класс методов, использующих топологическую, геометрическую и географическую информацию в больших данных.
4	Раздел 4. Статистический анализ, классификация и прогнозирование. Статистический подход к анализу больших данных. Точечное и интервальное оценивание. Обработка больших данных для контроля качества процессов и технологии six sigma. Анализ последовательностей. Отбор признаков и снижение размерности. Работа с выбросами и пропущенными значениями. Метрический и линейный классификаторы. ROC–кривая. A/B-тестирование. Анализ временных рядов. Классификация и кластеризация. Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации. Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов. Алгоритмизация больших данных. Метод главных компонент для сжатия больших данных. Иерархические методы кластерного анализа больших данных. Неиерархические методы кластерного анализа больших данных. Поиск признаковых зависимостей и частых множеств. Мультимодальная кластеризация. Предиктивные алгоритмы. Прогнозирование поведения больших данных.
5	Раздел 5. Визуализация, интерпретация, проектная организация больших данных. Способы визуального представления больших данных. Типы, задачи и виды визуализации. Графики, диаграммы, инфографика. Интерактивный сторителлинг, дашборды. Язык Python и его возможности. Дедупликация больших данных. Методы визуализации. Языки визуального моделирования. Когнитивная информатика и экономика знаний. Требования к специалистам в

№ п/п	Тематика лекционных занятий / краткое содержание
	области интеллектуальной обработки больших данных для бизнеса. Применение технологий больших данных для задач управления качеством.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	Раздел 1. Введение в науку о больших данных. История и причины появления термина «большие данные» (Big Data). Характеристики и источники Big Data. Четыре основных типа больших данных. Аналитика больших данных. Задачи, решаемые Big Data. Понятие жизненного цикла аналитики больших данных.
2	Раздел 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных. Классификация задач анализа больших данных. Основные характеристики и типы больших данных. Алгоритмы обработки первичной обработки больших данных. Современные подходы к обработке и хранению больших данных. Процесс анализа больших данных. Общая схема анализа. Извлечение и визуализация больших данных. Этапы моделирования. Процесс построения моделей. Формы представления, типы и виды больших данных. Представления наборов больших данных. Системы управления базами данных. Использование языка SQL (DDL) для создания структур больших данных. Запросы на языке SQL (DML). NoSQL хранилища больших данных. Хранилища больших данных ключ-значение. Документо-ориентированные хранилища больших данных. Графовые хранилища больших данных. Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.
3	Раздел 3. Интеллектуальный анализ данных. Введение в Data Mining. Большие данные и метаданные. Методы и стадии Data Mining. Задачи Data Mining. Извлечение больших данных, информации и знаний. Сферы применения Data Mining. Технологии KDD и Data Mining. Подготовка больших данных к анализу. Методика извлечения знаний. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. Программное обеспечение в области анализа больших данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. Машинное обучение и разработка больших данных. Основные аналитические методы обработки больших данных. Машинное обучение и майнинг больших данных. Распознавание образов. Нейронные сети как реализация алгоритмов машинного обучения. Облачные платформы интеллектуального анализа больших данных. Разработка методов, алгоритмов и программных средств интеллектуального анализа больших данных: Hadoop, Spark и IoT. Краудсорсинг — категоризация и обогащение больших данных силами широкого, неопределённого круга лиц, привлечённых на основании публичной оферты, без вступления в трудовые отношения. Смешение и интеграция больших данных — набор техник, позволяющих интегрировать разнородные большие данные из разнообразных источников для возможности глубинного анализа. Цифровая обработка сигналов и обработка естественного языка (включая тональный анализ). Сетевой анализ и оптимизация больших данных. Генетические алгоритмы. Имитационное моделирование. Пространственный анализ — класс методов, использующих топологическую, геометрическую и географическую информацию в больших данных.

№ п/п	Наименование лабораторных работ / краткое содержание
4	Раздел 4. Статистический анализ, классификация и прогнозирование. Статистический подход к анализу больших данных. Точечное и интервальное оценивание. Обработка больших данных для контроля качества процессов и технологии six sigma. Анализ последовательностей. Отбор признаков и снижение размерности. Работа с выбросами и пропущенными значениями. Метрический и линейный классификаторы. ROC–кривая. A/B-тестирование. Анализ временных рядов. Классификация и кластеризация. Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации. Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов. Алгоритмизация больших данных. Метод главных компонент для сжатия больших данных. Иерархические методы кластерного анализа больших данных. Неиерархические методы кластерного анализа больших данных. Поиск признаковых зависимостей и частых множеств. Мультимодальная кластеризация. Предиктивные алгоритмы. Прогнозирование поведения больших данных.
5	Раздел 5. Визуализация, интерпретация, проектная организация больших данных. Способы визуального представления больших данных. Типы, задачи и виды визуализации. Графики, диаграммы, инфографика. Интерактивный сторителлинг, дашборды. Язык Python и его возможности. Дедупликация больших данных. Методы визуализации. Языки визуального моделирования. Когнитивная информатика и экономика знаний. Требования к специалистам в области интеллектуальной обработки больших данных для бизнеса. Применение технологий больших данных для задач управления качеством.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Подготовка к лабораторным работам.
2	Работа с лекционным материалом.
3	Самостоятельное изучение учебной литературы.
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Введение в большие данные и анализ информации [Электронный ресурс] : учебное пособие / В. В. Конкина, А. Б. Борисенко, И. Л. Коробова. – Тамбов : Издательский центр ФГБОУ ВО «ТГТУ», 2024. – 1 электрон. опт. диск (CD-ROM). – Системные требования :	https://tstu.ru/book/elib1/pdf/2024/Konkina.pdf

	ПК не ниже Pentium IV ; CD-ROM-дисковод ; 1,32 Мб ; RAM ; Windows 95/98/XP ; мышь. – Загл. с экрана. ISBN 978-5-8265-2749-8	
2	Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. – СПб: Университет ИТМО, 2018. – 52 с.	https://books.ifmo.ru/file/pdf/2326.pdf

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

http://library.miit.ru/ - электронно-библиотечная система Научно-технической библиотеки МИИТ. http://elibrary.ru/ - научно-электронная библиотека.Поисковые системы: Yandex, Google, Mail. Облачные хранилища информации: Яндекс диск https://disk.yandex.ru, облако mail.ru, dropbox.com или другие.

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

Информационно-справочные и поисковые системы:

Internet Explorer, Yandex, Rambler, Mail, Opera,

http://www.efqm.org – интернет-портал Европейского фонда по менеджменту качества (EFQM).

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Для проведения практических занятий требуется аудитория, оснащенная техническими средствами. Для проведения лекционных занятий требуется аудитория, оснащенная мультимедиа аппаратурой и ПК с подключением к сети INTERNET.

9. Форма промежуточной аттестации:

Зачет в 1 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

доцент, доцент, к.н. кафедры
«Менеджмент качества» И.С. Кравчук

Согласовано:

Заведующий кафедрой СКиУК М.Ф. Гуськова

Председатель учебно-методической
комиссии М.Ф. Гуськова