

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
11.04.02 Инфокоммуникационные технологии и
системы связи,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Обработка и анализ данных в инфокоммуникационных системах

Направление подготовки: 11.04.02 Инфокоммуникационные
технологии и системы связи

Направленность (профиль): Инфокоммуникационные и нейросетевые
технологии передачи и анализа больших
данных

Форма обучения: Очная

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

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 167783
Подписал: руководитель образовательной программы
Киселёва Анастасия Сергеевна
Дата: 07.08.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины является формирование у обучающихся компетенций в соответствии с требованиями образовательного стандарта и формирование у студентов глубоких знаний и практических навыков в области обработки, анализа и интерпретации данных, полученных из различных инфокоммуникационных систем.

Задачами дисциплины являются:

Освоение методов и инструментов для сбора, хранения и обработки данных, включая работу с базами данных и облачными решениями.

Разработка навыков анализа больших данных с использованием статистических методов и алгоритмов машинного обучения для извлечения полезной информации.

Изучение принципов визуализации данных и представления результатов анализа, что позволяет делать обоснованные выводы и принимать решения на основе полученных данных.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ПК-3 - Способен совершенствовать и разрабатывать новые методы, модели, алгоритмы, технологии и инструментальные средства работы с большими данными.

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

Знать:

- основные архитектуры систем обработки больших данных (Hadoop, Spark, Lambda, Kappa); модели данных (реляционная, иерархическая, объектная); методы ETL (извлечение, преобразование, загрузка); принципы распределённых вычислений и CAP-теорему; виды хранилищ данных (DWH, Data Lake, Data Mart); OLAP-технологии и инструменты визуализации.

Уметь:

- выбирать архитектуру обработки данных в зависимости от задачи; применять методы очистки, трансформации и интеграции данных; оценивать производительность систем и использовать инструменты мониторинга; проектировать конвейеры данных с использованием современных фреймворков.

Владеть:

- методами проектирования хранилищ данных; навыками работы с инструментами ETL и распределённой обработки (Spark, MapReduce); приёмами визуализации и построения отчётов; подходами к масштабированию и обеспечению отказоустойчивости.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 6 з.е. (216 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):	48	48
В том числе:		
Занятия лекционного типа	16	16
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 168 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	Введение в обработку данных. Рассматриваемые вопросы: Основные понятия и определения Цикл обработки данных Роль данных в инфокоммуникационных системах.
2	Методы сбора данных. Рассматриваемые вопросы: Традиционные и современные методы сбора Использование API и веб-скрейпинг Этические аспекты сбора данных.
3	Хранение данных. Рассматриваемые вопросы: Реляционные и нереляционные базы данных Облачные технологии для хранения данных Архитектура хранения данных.
4	Обработка и очистка данных Рассматриваемые вопросы: Принципы очистки данных Обработка пропусков и аномалий Форматы и стандарты данных.
5	Анализ данных. Рассматриваемые вопросы: Описательная и inferential статистика Алгоритмы машинного обучения Анализ временных рядов.
6	Визуализация данных Принципы визуализации и восприятия информации Использование инструментов визуализации (Tableau, Power BI) Создание интерактивных дашбордов.
7	Интерпретация результатов анализа Рассматриваемые вопросы: Методология интерпретации данных Принятие решений на основе анализа Коммуникация результатов заинтересованным сторонам.
8	Будущее обработки данных Рассматриваемые вопросы: Тренды в области больших данных и ИИ Этика и конфиденциальность данных Перспективы развития инфокоммуникационных технологий.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Введение в инструменты обработки данных. Рассматриваемые вопросы: Установка и настройка Python и R Обзор библиотек для анализа данных (Pandas, NumPy, dplyr)ой.

№ п/п	Тематика практических занятий/краткое содержание
2	Сбор данных с использованием API. Рассматриваемые вопросы: Работа с RESTful API Извлечение данных из открытых источников (например, Twitter, OpenWeather).
3	Веб-скрейпинг. Рассматриваемые вопросы: Основы веб-скрейпинга с BeautifulSoup и Scrapy Практика извлечения данных с веб-страниц.
4	Работа с реляционными базами данных. Рассматриваемые вопросы: Основы SQL: создание, чтение, обновление и удаление данных Выполнение сложных запросов и объединение таблиц
5	Работа с нереляционными базами данных. Рассматриваемые вопросы: Введение в MongoDB: создание и управление коллекциями Запросы и агрегации в MongoDB.
6	Очистка и подготовка данных Рассматриваемые вопросы: Устранение пропусков и аномалий в данных Преобразование данных: нормализация и стандартизация.
7	Анализ данных с использованием статистических методов Рассматриваемые вопросы: Применение описательной статистики на реальных данных Проведение гипотез и тестов значимости.
8	Основы машинного обучения Рассматриваемые вопросы: Обучение моделей на примерах (регрессия, классификация) Оценка качества моделей (метрики, кросс-валидация)
9	Анализ временных рядов. Рассматриваемые вопросы: Построение и анализ временных рядов Прогнозирование на основе исторических данных.
10	Визуализация данных с использованием Matplotlib и Seaborn. Рассматриваемые вопросы: Создание графиков и диаграмм Настройка визуализаций для лучшего восприятия информации
11	Создание интерактивных дашбордов. Рассматриваемые вопросы: Использование Dash или Streamlit для создания веб-приложений Интеграция визуализаций и данных в дашборд.
12	Работа с большими данными. Рассматриваемые вопросы: Введение в Apache Spark и его экосистему Обработка больших объемов данных с использованием PySpark
13	Этика и конфиденциальность данных. Рассматриваемые вопросы: Практические аспекты защиты данных Анонимизация и псевдонимизация данных.

№ п/п	Тематика практических занятий/краткое содержание
14	Применение машинного обучения к реальным задачам Рассматриваемые вопросы: Проект по решению практической задачи с использованием алгоритмов ML Работа в группах над проектами
15	Интеграция данных из разных источников Рассматриваемые вопросы: Объединение данных из различных баз и форматов Применение ETL-процессов (Extract, Transform, Load).
16	Презентация и защита проектов Рассматриваемые вопросы: Подготовка презентации результатов анализа Защита проектов перед аудиторией и получение обратной связи.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Самостоятельное изучение и конспектирование отдельных тем учебной литературы, связанных с разделами дисциплины
2	Работа с лекционным материалом
3	Подготовка к практическим занятиям
4	Выполнение курсовой работы.
5	Подготовка к промежуточной аттестации.
6	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых работ

Курсовая работа на тему: "Применение алгоритмов машинного обучения для анализа текстовых данных". Исходные данные выбираются согласно вариантам:

Вариант 1:

Анализ тональности отзывов о продуктах

Разработать модель машинного обучения для классификации отзывов (положительные, отрицательные, нейтральные) с использованием методов обработки естественного языка (NLP).

Вариант 2:

Классификация новостных статей по категориям

Создать модель, которая будет автоматически классифицировать новостные статьи по темам (например, политика, спорт, экономика) с использованием алгоритмов классификации.

Вариант 3:

Извлечение ключевых слов из текстов

Реализовать алгоритм для автоматического извлечения ключевых слов и фраз из документов, используя методы TF-IDF и LDA (Latent Dirichlet Allocation).

Вариант 4:

Создание чат-бота на основе NLP

Разработать простого чат-бота, который будет отвечать на вопросы пользователей, используя алгоритмы обработки естественного языка и машинного обучения.

Вариант 5:

Анализ и визуализация данных о постах Вконтакте

Провести анализ постов Вконтакте на определенную тему, используя машинное обучение для выявления трендов и паттернов, а также визуализировать результаты.

Вариант 6:

Кластеризация текстов для выявления тем

Использовать алгоритмы кластеризации (например, K-means) для группировки текстов по схожести и выявления тем, обсуждаемых в данных.

Вариант 7:

Обнаружение фейковых новостей

Разработать модель, которая будет классифицировать статьи как подлинные или фейковые, используя различные алгоритмы машинного обучения и наборы данных о новостях.

Вариант 8:

Анализ эмоциональной окраски текстов

Исследовать эмоциональную окраску текстов (например, литературных произведений или статей) с помощью алгоритмов машинного обучения и визуализировать результаты.

Вариант 9:

Обработка и анализ отзывов о фильмах

Создать модель для анализа отзывов о фильмах, выявления популярных тем и трендов, а также прогнозирования оценок фильмов на основе текстовых данных.

Вариант 10:

Сравнительный анализ алгоритмов для задач NLP

Провести сравнительный анализ различных алгоритмов машинного обучения (например, SVM, Random Forest, нейронные сети) для решения задач классификации текстов и выявить наиболее эффективные подходы.

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

№ п/п	Библиографическое описание	Место доступа
1	Вольфсон, М. Б. Анализ данных : учебно-методическое пособие / М. Б. Вольфсон. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2023. — 69 с.	https://e.lanbook.com/book/381533
2	Ланских, Ю. В. Интеллектуальный анализ данных : учебное пособие / Ю. В. Ланских, В. Г. Ланских. — Киров : ВятГУ, 2023. — 240 с.	https://e.lanbook.com/book/408569
3	Золкин, А. Л. Математическое моделирование и анализ данных : учебное пособие для вузов / А. Л. Золкин, М. В. Сартаков. — Санкт-Петербург : Лань, 2025. — 128 с. — ISBN 978-5-507-51616-2.	https://e.lanbook.com/book/455660

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

Информационный портал Научная электронная библиотека eLIBRARY.RU (www.elibrary.ru);

Единая коллекция цифровых образовательных ресурсов (<http://window.edu.ru>);

Научно-техническая библиотека РУТ (МИИТ) (<http://library.miit.ru>);

Поисковые системы «Яндекс» для доступа к тематическим информационным ресурсам;

Электронно-библиотечная система издательства «Лань» — <http://e.lanbook.com/>;

Электронно-библиотечная система ibooks.ru — <http://ibooks.ru/>;

Электронно-библиотечная система «УМЦ» — <http://www.umczdt.ru/>;

Электронно-библиотечная система «BOOK.ru» — <http://www.book.ru/>;

Электронно-библиотечная система «ZNANIUM.COM» — <http://www.znanium.com/>

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

1. Операционная система windows microsoft office 2003 и выше;
2. Браузер Internet Explorer 8.0 и выше с установленным Adobe Flash player версии 10.3 и выше;
3. Adobe acrobat.

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

Учебные аудитории для проведения учебных занятий, оснащенные компьютерной техникой и наборами демонстрационного оборудования.

9. Форма промежуточной аттестации:

Зачет в 1 семестре.

Курсовая работа в 1 семестре.

10. Оценочные материалы.

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

Авторы:

доцент, к.н. кафедры «Системы
управления транспортной
инфраструктурой»

И.М. Губенко

Согласовано:

Директор

Д.В. Паринов

Руководитель образовательной
программы

А.С. Киселёва

Председатель учебно-методической
комиссии

Д.В. Паринов