МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ)

Рабочая программа дисциплины (модуля),
как компонент образовательной программы
базового высшего образования
по направлению подготовки
09.03.01 Информатика и вычислительная техника,
утвержденной первым проректором ФГАОУ ВО
РУТ (МИИТ) Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Основы анализа данных

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): Цифровая инженерия транспортных процессов

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 ФГАОУ ВО РУТ (МИИТ)
ID подписи: 937226
Подписал: руководитель образовательной программы
Проневич Ольга Борисовна
Дата: 07.09.2026

1. Общие сведения о дисциплине (модуле).

Целями освоения учебной дисциплины являются формирование у обучающихся базовых теоретических знаний и практических навыков в области анализа данных и искусственного интеллекта для решения прикладных бизнес-задач. В рамках курса рассматриваются основные этапы жизненного цикла аналитического исследования: от постановки задачи и сбора данных до построения моделей машинного обучения и интерпретации результатов. Особое внимание уделяется применению методов статистического анализа, визуализации данных и построению прогнозных моделей с использованием современных библиотек языка Python.

Задачами освоения дисциплины являются:

Формирование теоретических знаний об основных алгоритмах анализа данных, методах машинного обучения и искусственного интеллекта, их классификации и областях применения.

Освоение инструментов и библиотек языка Python для обработки, преобразования и визуализации данных (Pandas, NumPy, Matplotlib, Seaborn).

Формирование практических навыков применения методов регрессии, классификации, кластеризации и снижения размерности для решения прикладных задач.

Приобретение навыков постановки аналитических задач, выбора адекватных методов анализа и интерпретации полученных результатов.

Развитие компетенций по оценке качества моделей машинного обучения, настройке гиперпараметров и предотвращению переобучения.

Формирование навыков работы с данными различных типов (числовые, категориальные, текстовые, временные ряды) и их подготовки к анализу.

Дисциплина предназначена для формирования компетенций, необходимых для проведения базовых аналитических исследований и внедрения методов искусственного интеллекта в практическую деятельность транспортных организаций.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-4 - Способен решать задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и технологий искусственного интеллекта, а также с учетом основных требований информационной безопасности;

ППК-ДАЗ - Способен проводить аналитические исследования с применением технологий больших данных.

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

Уметь:

Применять библиотеки Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn) для обработки и визуализации данных.

Выполнять разведочный анализ данных (EDA): очистка, преобразование, анализ распределений и взаимосвязей.

Строить и оценивать модели машинного обучения для решения задач регрессии, классификации и кластеризации.

Интерпретировать результаты анализа и формулировать выводы и рекомендации для бизнес-задач.

Знать:

Основные этапы жизненного цикла аналитического исследования и принципы постановки аналитических задач.

Основные алгоритмы машинного обучения: регрессия, классификация, кластеризация, снижение размерности, деревья решений и ансамблевые методы.

Методы оценки качества моделей, метрики и способы предотвращения переобучения.

Инструменты и библиотеки Python для обработки, визуализации и анализа данных.

Владеть:

Навыками обработки и подготовки данных различных типов к аналитическому исследованию.

Методами применения базовых алгоритмов машинного обучения для решения прикладных задач.

Навыками визуализации данных и представления результатов анализа в форме, доступной для принятия управленческих решений.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 2 з.е. (72 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):	32	32
В том числе:		
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 40 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.

Не предусмотрено учебным планом

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Введение в предметную область аналитики больших данных: хайп или нераскрытый потенциал Рассматриваются вопросы: Данные, информация, знания – в чем отличия. Что есть большие данные. Что есть аналитика. Что за зверь такой: data-driven организация? Ожидания рынка vs результаты. Несколько основных ловушек и извлечённых из них опыт.

№ п/п	Тематика практических занятий/краткое содержание
	Пересмотр приоритетов за последние 10 лет. Новые игроки на рынке и чем они характеризуются.
2	Тема 2. Стратегия управления данными: искусство видеть лес за деревьями Рассматриваются вопросы: Стратегия управления данными, как составная часть культуры современной организации. Почему это важный вопрос? Что есть стратегия управления данными, кто за неё отвечает в организации. Кто такой владелец данных, его сфера ответственности и полномочия. Модель данных, что это такое и зачем она нужна. Основные заинтересованные стороны. Концепция Self-Service BI. Техническая инфраструктура. Революция open-source и доступность технологий. Как измерить эффективность стратегии управления данными.
3	Тема 3. Хранилища данных: первый шаг к аналитике и зачем все так усложнять Рассматриваются вопросы: Разберемся в предмете: что есть хранилище данных, зачем оно нужно. Что за зверь такой – ETL. Определение источников и загрузка данных. Какие виды преобразования и объединения данных существуют, в чём их принципиальные отличия и на что следует обращать внимание. Что такое витрины данных.
4	Тема 4. Постановка задачи: как не ошибиться с выбором цели и инструментов Рассматриваются вопросы: С чего начать, с какой стороны подойти к задаче анализа данных. Основные термины и понятия. Как правильно ставить задачу и формулировать цель анализа и почему это так важно. Как оценить достаточно ли вам данных для анализа, где и как данные следует добывать, на что обращать внимание. Понятие качества данных. Методики и инструменты обеспечения и контроля качества данных. Выработка гипотез и выбор методов.
5	Тема 5. Подготовка данных: искусство есть слона по частям Рассматриваются вопросы: Какие данные бывают, какие у них свойства, почему важно разобраться в их природе (рассмотрим на примере ошибок, допускаемых аналитиками при работе вслепую). Что такое метаданные или бизнес-глоссарий и почему обязательно сопровождать аналитическую задачу созданием соответствующего репозитория метаданных. Какие есть способы подготовки данных, на что следует обращать внимание, как не ошибиться.
6	Тема 6. Введение в статистический анализ: разбираемся в основах Рассматриваются вопросы: Основные понятия: генеральная совокупность и выборка. Какие основные статистические показатели используются в аналитике и почему (рассмотрим на известных исторических примерах). Что такое сводная таблица и как она используется в работе (рассмотрим на примере задачи финансового планирования).
7	Тема 7. Обзор описательных статистик: как заставить данные говорить (часть 1) Рассматриваются вопросы: Какие существуют описательные статистики и для чего используются. Почему ими надо аккуратно пользоваться. Почему иногда полезно выполнять нормализацию данных.

№ п/п	Тематика практических занятий/краткое содержание
	Какие инструменты нам в этом помогут, как правильно представить результатная статистического анализа.
8	Тема 8. Основы визуализации и story telling: как заставить данные говорить (часть 2) Рассматриваются вопросы: Декодирование результатов аналитической работы. Почему это важно (аналитика не игрушка, она должна быть подчинена реальным бизнес-целям и показателям). Как показывать положительные и отрицательные значения, как выводить данные по категориальным аналитикам, как показывать динамику и прогноз, как отображать выбросы и пропуски, как расставить акценты и выделить самое важное (рассмотрим на реальных кейсах нескольких компаний).
9	Тема 9. Аналитика: сравнение групп данных Рассматриваются вопросы: Для чего и где применяется. Понятие значимости различий при сравнении групп (выборок) данных. Зависимые и независимые выборки. Таблицы перекрестной классификации. Сравнение независимых выборок. Сравнение зависимых выборок. Анализ последовательностей.
10	Тема 10. Аналитика: анализ взаимосвязей, поиск скрытого от человеческого глаза Рассматриваются вопросы: Зачем нужно анализировать связи в данных (примеры задач). Зависимые и независимые переменные. Статистические взаимосвязи и их значимость для анализа. Корреляции. Регрессия и её виды. Факторный анализ. Анализ надежности согласованности. Понятие проклятия размерности.
11	Тема 11. Аналитика: классификация и что это даёт Рассматриваются вопросы: Зачем нужна классификация (примеры задач). Виды классификации. Обзор используемых методов: деревья решений, случайный лес, наивный байесовский классификатор, линейная и логистическая регрессия. Кластеризация (и чем принципиально отличается от классификации). О чем важно помнить при выполнении кластеризации.
12	Тема 12. Аналитика: прогнозирование, хотим управлять будущим Рассматриваются вопросы: Что есть прогнозирование, где и как используется. Временные ряды. Что такое сезонность и как её найти в данных. Методы прогнозирования и их сравнение: линейная регрессия, экспоненциальное сглаживание, фильтр Кальмана, модель сквозящего среднего, авторегрессионная модель, ARIMA.
13	Тема 13. Машинное обучение: где и как применяется Рассматриваются вопросы: Зачем нужно машинное обучение, что скрывается под этим названием. Методы машинного обучения для изучения и подготовки данных. Примеры задач и методов их решения.

№ п/п	Тематика практических занятий/краткое содержание
	Обучение с подкреплением. A/B тестирование.
14	Тема 14. Машинное обучение и Искусственный интеллект: а есть ли разница Рассматриваются вопросы: Почему современные компании постоянно используют эти термины как будто они значат одно и то же? Есть ли разница на самом деле? Задачи применения искусственного интеллекта, как более широкий класс задач. Понятие нейронных сетей. Проблема интерпретируемости. Примеры применения решений, основанных на искусственном интеллекте, которые окружают нас уже сегодня. Будущее ИИ и ML.
15	Тема 15. Регулирование данных: в любой сфере должны быть свои законы Рассматриваются вопросы: Актуальные вопросы регулирования данных, стоящие на повестке дня в мире и России, в частности. Какие существуют нормативы на сегодня, как контролируется факт их выполнения. Этические проблемы работы с данными и применения машинного обучения / Искусственного интеллекта (рассмотрим на примере нескольких реальных кейсов – как используется и с какими последствиями приходится иметь дело). Формирование этической культуры работы с данными в организации, на что обратить внимание.
16	Тема 16. Введение в ИИ Агенты Рассматриваются вопросы: Современное состояние ИИ и практические инструменты применения ИИ

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение литературы и документацией по пройденным темам
2	Подготовка к промежуточной аттестации
3	Подготовка к текущему контролю
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	Балабаева, Н. П. Математические основы анализа данных : учебное пособие / Н. П. Балабаева, О. В. Осипов, Е. А. Энбом. — Самара : ПГУТИ, 2025. — 272 с.	https://e.lanbook.com/book/517376
2	Основы статистического анализа данных : учебное пособие / Т. Г. Авачева, М. Н. Дмитриева, Н. В.	https://e.lanbook.com/book/513194

	Дорошина, О. А. Милованова. — 2-е изд., испр. и доп. — Рязань : РязГМУ, 2025. — 167 с.	

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

https://habr.com/ru - база знаний в виде статей, обзоров

http://library.miit.ru/ - электронно-библиотечная система научно-технической библиотеки МИИТ

https://superset.apache.org/user-docs/intro/#documentation – документация Apache Superset

https://www.postgresql.org/ - официальная страница PostgreSQL

https://yandex.cloud/ru/docs/datalens/ - официальная документация Yandex Datalens

https://docs.docker.com/ - официальная документация Docker

https://www.metabase.com/docs/latest/ - официальная документация Metabase

https://tableau.pro/ - официальный сайт Tableau с лучшими практиками проектирования дашбордов

https://revealthedata.com/blog/ - Блог о визуализации данных и BI-системах

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

MS Office (Excel, Word, PowerPoint)
Apache Superset
Yandex Datalens
Metabase
PostgreSQL
Docker
Apache Airflow
Браузер Chrome / Яндекс браузер
Draw.io

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

Компьютер преподавателя с доступом к сети Интернет

Компьютеры студентов с установленным Python-окружением и табличным процессором

Доступ к учебной базе событийных данных и экспериментальным выборкам

Доступ к PostgreSQL и BI-системе для визуализации результатов

Доступ к облачной виртуальной машине или Docker-контейнерам для воспроизводимого анализа

Проектор и экран для проектора

Маркерная доска

9. Форма промежуточной аттестации:

Экзамен в 1 семестре.

10. Оценочные материалы.

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

Авторы:

руководитель образовательной программы О.Б. Проневич

Согласовано:

Директор Д.В. Паринов

Руководитель образовательной программы О.Б. Проневич

Председатель учебно-методической комиссии Д.В. Паринов