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)

Инженерия данных на транспорте

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: 02.09.2026

1. Общие сведения о дисциплине (модуле).

Целями освоения учебной дисциплины являются формирование у обучающихся системных знаний и практических навыков в области инженерии данных для решения задач транспортной отрасли. В рамках курса рассматриваются жизненный цикл данных в транспортных системах, методы сбора, хранения и предобработки структурированных и неструктурированных данных, а также основы построения конвейеров данных (ETL/ELT) для последующего применения моделей машинного обучения и эконометрического анализа. Особое внимание уделяется работе с транспортными логистическими потоками, телематическими данными, базами данных ГЛОНАСС/GPS, а также автоматизации процессов подготовки признаков для прогнозных моделей.

Задачами освоения дисциплины являются:

- Формирование теоретических знаний об источниках данных на транспорте (GPS-треки, логистические системы, АСУ, метеоданные) и их специфике.

- Освоение методов первичного анализа данных (EDA) для выявления структуры, пропусков, выбросов и аномалий в транспортных данных.

- Развитие практических навыков преобразования и нормализации данных для обеспечения их совместимости с различными моделями машинного обучения.

- Формирование компетенций по автоматизации процессов очистки и подготовки признаков с использованием инструментов Python.

- Освоение критериев выбора модели в зависимости от типа решаемой задачи (классификация, регрессия, кластеризация) и характера транспортных данных.

- Приобретение навыков оценки качества моделей и диагностики переобучения/недообучения применительно к задачам прогнозирования на транспорте.

Дисциплина предназначена для формирования компетенций, необходимых для эффективной работы с данными на этапе их подготовки и первичного моделирования в транспортных организациях.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ИД-2 - Способен планировать и организовывать аналитические работы с использованием технологий больших данных;

ИД-3 - Способен разрабатывает, поддерживать и развивать ETL/ELT-процессы;

ОПК-3 - Способен использовать современные информационные технологии и программно-аппаратные средства, в том числе отечественного производства, при решении задач профессиональной деятельности.

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

Знать:

- Основные источники и типы данных в транспортной отрасли (телематика, GPS/ГЛОНАСС, логистические системы, метеоданные), их структуру, форматы и проблематику качества (пропуски, шум, выбросы).
- Критерии выбора модели машинного обучения (регрессия, классификация, кластеризация) в зависимости от типа транспортной задачи и характера данных, а также основные метрики оценки качества моделей.
- Алгоритмы автоматизации работы с признаками, включая методы создания новых признаков на основе существующих, выделение географических и временных характеристик, формирование агрегированных статистик.

Уметь:

- Проводить первичный анализ транспортных данных (EDA): выявлять структуру, пропуски, выбросы и аномалии; визуализировать распределения и корреляции.
- Выполнять необходимые преобразования данных для совместимости с выбранной моделью: масштабирование, кодирование категориальных признаков, работа с пропусками, обработка временных рядов.
- Выбирать и обосновывать применение подходящей модели машинного обучения для транспортной задачи, оценивать ее качество с использованием релевантных метрик и диагностировать переобучение/недообучение.

Владеть:

- Навыками работы с библиотеками Python (Pandas, NumPy) для выполнения преобразований данных: фильтрация, агрегация, слияние, очистка и подготовка признаков.
- Навыками визуализации данных (Matplotlib, Seaborn) и построения пайплайнов предобработки для решения задач прогнозирования на транспорте.

- Навыками применения моделей классического машинного обучения (линейная регрессия, логистическая регрессия, деревья решений) для решения транспортных задач с использованием библиотеки Scikit-learn.

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):	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 24 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	Тема 1. Введение в инженерию данных на транспорте. Рассматриваемые вопросы: - Понятие инженерии данных и ее место в Data Science. - Специфика транспортных данных: GPS/ГЛОНАСС, телематика, логистические системы. - Источники данных в транспортной отрасли. - Жизненный цикл данных: сбор, хранение, обработка, использование.
2	Тема 2. Источники данных и форматы хранения. Рассматриваемые вопросы: - Реляционные базы данных (SQL) и NoSQL-решения на транспорте. - Работа с файловыми форматами: CSV, JSON, Excel. - Особенности потоковых данных на транспорте. - Основы работы с API транспортных сервисов (обзорно).
3	Тема 3. Предварительный анализ данных (EDA). Рассматриваемые вопросы: - Методы описательной статистики для транспортных данных. - Визуализация распределений признаков. - Выявление пропусков и дубликатов в данных. - Обнаружение аномалий и выбросов (базовые методы). - Анализ корреляций между признаками.
4	Тема 4. Предобработка данных. Рассматриваемые вопросы: - Методы работы с пропущенными значениями (удаление, заполнение). - Кодирование категориальных переменных. - Масштабирование и нормализация признаков (StandardScaler, MinMaxScaler). - Обработка временных рядов в транспортных задачах (базовые подходы).
5	Тема 5. Проектирование признаков для транспортных задач. Рассматриваемые вопросы: - Понятие проектирования признаков. - Создание новых признаков на основе существующих. - Извлечение признаков из даты и времени (час, день недели, месяц). - Выделение географических признаков - обзорно. - Формирование агрегированных статистик (средние, суммы, скользящие окна).
6	Тема 6. Критерии выбора моделей машинного обучения. Рассматриваемые вопросы: - Классификация задач машинного обучения на транспорте. - Регрессионные модели для прогнозирования (линейная регрессия). - Классификационные модели для оценки рисков (логистическая регрессия). - Кластеризация для сегментации (K-Means) - обзорно. - Выбор модели в зависимости от типа задачи и характера данных.
7	Тема 7. Оценка качества моделей и диагностика переобучения. Рассматриваемые вопросы: - Метрики оценки регрессионных моделей (MAE, RMSE, R?). - Метрики оценки классификационных моделей (Accuracy, Precision, Recall, F1). - Понятие переобучения и недообучения. - Разделение данных на обучающую и тестовую выборки. - Базовые методы борьбы с переобучением (регуляризация).
8	Тема 8. Введение в эконометрическое моделирование. Рассматриваемые вопросы: - Понятие эконометрики и ее задачи. - Классическая линейная регрессионная модель.

№ п/п	Тематика лекционных занятий / краткое содержание
	- Оценка параметров модели (метод наименьших квадратов). - Интерпретация коэффициентов регрессии. - Применение эконометрических моделей на транспорте.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Введение в Python для инженерии данных. Рассматриваемые вопросы: - Установка и настройка среды разработки. - Базовые конструкции языка Python. - Знакомство с библиотеками Pandas и NumPy. - Создание и работа с DataFrame. - Импорт и экспорт данных из различных форматов (CSV, Excel).
2	Тема 2. Основы SQL для работы с транспортными данными. Рассматриваемые вопросы: - Подключение к учебной базе данных (SQLite). - Написание базовых запросов: SELECT, WHERE, ORDER BY. - Фильтрация и сортировка данных. - Агрегация данных с GROUP BY. - Объединение таблиц с JOIN.
3	Тема 3. Работа с API транспортных сервисов. Рассматриваемые вопросы: - Знакомство с REST API (базовые принципы). - Получение данных с помощью библиотеки requests. - Обработка JSON-ответов. - Загрузка данных о погоде или дорожной обстановке (упрощенный кейс).
4	Тема 4. Загрузка и первичное знакомство с транспортными данными. Рассматриваемые вопросы: - Загрузка реального набора транспортных данных (датасет). - Первичный просмотр данных (head, info, describe). - Определение типов данных и структуры таблиц. - Формулировка гипотез о данных.
5	Тема 5. Первичный анализ транспортных данных (EDA). Рассматриваемые вопросы: - Описательная статистика признаков. - Визуализация распределений с помощью Matplotlib и Seaborn. - Выявление пропусков и дубликатов. - Анализ корреляционной матрицы.
6	Тема 6. Обработка пропусков и выбросов. Рассматриваемые вопросы: - Анализ структуры пропусков в данных. - Удаление записей с пропусками. - Заполнение пропусков средними, медианой и модой. - Методы обнаружения и обработки выбросов.
7	Тема 7. Кодирование категориальных признаков. Рассматриваемые вопросы:

№ п/п	Тематика практических занятий/краткое содержание
	- Выявление категориальных переменных в транспортных данных. - Применение Label Encoding. - Применение One-Hot Encoding. - Сравнение различных подходов к кодированию.
8	Тема 8. Масштабирование и нормализация данных. Рассматриваемые вопросы: - Стандартизация признаков (StandardScaler). - Нормализация признаков (MinMaxScaler). - Применение масштабирования к данным. - Влияние масштабирования на качество моделей (демонстрация).
9	Тема 9. Проектирование признаков на практике. Рассматриваемые вопросы: - Создание новых признаков на основе даты и времени. - Выделение географических признаков - упрощенно. - Расчет расстояний и времени в пути. - Формирование агрегированных признаков (средние, суммы).
10	Тема 10. Визуализация транспортных данных. Рассматриваемые вопросы: - Построение графиков временных рядов. - Визуализация географических данных на картах. - Создание интерактивных визуализаций. - Представление результатов анализа.
11	Тема 11. Построение и оценка регрессионных моделей. Рассматриваемые вопросы: - Постановка задачи регрессии (прогнозирование времени доставки). - Реализация линейной регрессии с использованием Scikit-learn. - Оценка качества модели (MAE, RMSE, R?). - Интерпретация коэффициентов регрессии.
12	Тема 12. Построение и оценка классификационных моделей. Рассматриваемые вопросы: - Постановка задачи классификации (прогнозирование задержек). - Реализация логистической регрессии с использованием Scikit-learn. - Оценка качества модели (Accuracy, Precision, Recall, F1). - Построение матрицы ошибок.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение литературы и документации по библиотекам Pandas, NumPy, Scikit-learn
2	Работа с официальной документацией и материалами по SQL (базовый уровень).
3	Изучение источников открытых данных по транспортной тематике (подготовка обзора 2-3 источников).
4	Выполнение домашних заданий по предобработке данных и EDA.
5	Подготовка к промежуточной аттестации.
6	Подготовка к промежуточной аттестации.
7	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Системы баз данных: организация, инженерия, ведение : учебное пособие / О. В. Тараканов, Ю. А. Паршенкова, М. Ю. Конышев [и др.]. — Москва : РТУ МИРЭА, 2023. — 373 с. — ISBN 978-5-7339-1767-2.	https://e.lanbook.com/book/368672

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

MS Office (Excel, Word)

Notepad++

Браузер Chrome

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

Инструменты визуального моделирования

Офисный пакет приложений

Система контроля версий Git

Браузер

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:

Зачет 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