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)

Прикладной анализ 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: 08.09.2026

1. Общие сведения о дисциплине (модуле).

Цель дисциплины:

Формирование компетенций по выявлению скрытых закономерностей, аномалий и трендов в транспортных данных, а также по построению, обучению и интерпретации моделей классического машинного обучения и эконометрики для преобразования результатов интеллектуального анализа в обоснованные бизнес-рекомендации и управленческие решения в транспортной отрасли.

Задачи дисциплины: изучение методов статистического анализа и визуализации для поиска аномалий, трендов и паттернов в пассажиропотоках, телеметрии подвижного состава и геопространственных данных транспортных систем; освоение методологии построения и диагностики эконометрических моделей (OLS, Logit/Probit, модели с фиксированными эффектами) для оценки эластичности спроса, вероятности событий и эффективности транспортной политики; приобретение практических навыков применения библиотек Python (Scikit-learn, Statsmodels) для обучения моделей классического машинного обучения (логистическая регрессия, деревья решений, SVM, ансамблевые методы) на реальных отраслевых данных; развитие умений решать проблему несбалансированных классов и выбирать адекватные метрики качества (Precision, Recall, F1, MAE, RMSE) применительно к транспортным задачам — прогноз отказов оборудования, опозданий рейсов, классификация опасного вождения; овладение техниками интерпретации коэффициентов моделей, оценки важности признаков и предельных эффектов для формирования доказательной базы при принятии операционных и стратегических решений; формирование способности трансформировать формальные результаты моделирования (скор-карты, прогнозы временных рядов, оценки вероятностей) в конкретные бизнес-инсайты: регламенты диспетчерского реагирования, обоснование субсидирования маршрутов, планы предиктивного технического обслуживания, схемы размещения инфраструктуры контроля безопасности дорожного движения; выработка навыков презентации и защиты аналитических выводов перед условным «бизнес-заказчиком» с расчетом экономического эффекта от внедрения разработанных моделей.

2. Планируемые результаты обучения по дисциплине (модулю).

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

МП-1 - Способен сопровождать развитие существующего ИТ-продукта;

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.

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

Знать:

Методы поиска аномалий, трендов и паттернов в данных: статистические подходы (правило трёх сигм, IQR), визуальные методы, методы пространственного анализа.

Методы обучения моделей классического машинного обучения: линейная и логистическая регрессия, деревья решений, метод опорных векторов (SVM), ансамблевые методы (Random Forest, Gradient Boosting).

Методологию построения и диагностики эконометрических моделей: OLS, Logit/Probit, модели с фиксированными и случайными эффектами, тесты на гетероскедастичность и мультиколлинеарность.

Методы работы с временными рядами: декомпозиция, экспоненциальное сглаживание (ETS, Holt-Winters), оценка качества прогнозов (MAE, MAPE).

Метрики качества моделей классификации и регрессии: Accuracy, Precision, Recall, F1, ROC-AUC, MAE, RMSE, R?, скорректированный R?.

Методы борьбы с несбалансированными классами: SMOTE, class_weight, выбор адекватных метрик.

Уметь:

Применять библиотеки Python (Scikit-learn, Statsmodels, Pandas, NumPy) для обучения моделей машинного обучения и эконометрики на реальных данных.

Выполнять поиск аномалий, трендов и паттернов в транспортных данных с использованием статистических методов и визуализации.

Строить и диагностировать эконометрические модели (OLS, Logit/Probit, модели с фиксированными эффектами) для оценки влияния факторов на транспортные показатели.

Обучать модели классического машинного обучения для решения задач прогнозирования, классификации и кластеризации.

Выбирать адекватные метрики качества в зависимости от типа задачи и бизнес-контекста.

Интерпретировать результаты моделей и трансформировать их в бизнес-рекомендации.

Владеть:

Навыками применения методов статистического анализа и машинного обучения для решения прикладных транспортных задач.

Навыками работы с библиотеками Python: Scikit-learn, Statsmodels, Pandas, NumPy, Matplotlib, Seaborn, Plotly, GeoPandas.

Навыками интерпретации коэффициентов моделей, оценки важности признаков и предельных эффектов для формирования доказательной базы управленческих решений.

Навыками трансформации формальных результатов моделирования в конкретные бизнес-инсайты: регламенты, планы обслуживания, рекомендации по оптимизации.

Навыками презентации и защиты аналитических выводов перед заказчиком с расчетом экономического эффекта от внедрения разработанных моделей.

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 №3
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. Среда разработки и Data Engineering для транспортных данных (Часть 1: Подключение к источникам) Рассматриваются вопросы: Настройка рабочего окружения (Jupyter / VS Code) и системы контроля версий (Git). Подключение к реляционным базам данных АСУ транспортного предприятия (расписания, данные валидаторов). Загрузка данных из открытых API городских транспортных систем и CSV-выгрузок IoT-устройств. Первичный осмотр данных: типы полей, пропуски в геотреках транспорта.
2	Тема 2. Среда разработки и Data Engineering для транспортных данных (Часть 2: Обработка и очистка) Рассматриваются вопросы: Обработка аномальных значений в телеметрии (скорость > 200 км/ч для городского автобуса). Нормализация временных меток из разных часовых поясов (GPS-трекеры) и приведение к единой сетке расписания. Объединение таблиц (SQL / Pandas merge): стыковка данных об оплате проезда с данными о рейсах. Создание витрины данных «Пассажиропоток на маршруте в час пик».
3	Тема 3. Поиск аномалий, трендов и паттернов в данных (Часть 1: Статистический подход и визуализация) Рассматриваются вопросы: Методы поиска аномалий в пассажиропотоке: правило трёх сигм и метод межквартильного размаха (IQR) для выявления резких падений спроса. Визуализация временных рядов (Matplotlib, Seaborn, Plotly) для обнаружения «провалов» в движении поездов метро в дневное время. Выявление паттернов недельной сезонности: сравнение буднего дня и выходного для загруженности остановочных пунктов.

№ п/п	Тематика практических занятий/краткое содержание
	Формирование гипотез бизнес-инсайтов (АД-3): является ли «аномалия» сбоем датчика или реальным событием (ДТП, митинг).
4	Тема 4. Поиск аномалий, трендов и паттернов в данных (Часть 2: Геопространственные и событийные данные) Рассматриваются вопросы: Анализ плотности трафика с помощью библиотек пространственного анализа (GeoPandas, Folium). Поиск систематических отклонений транспорта от маршрута (паттерн «срезание углов») на основе сопоставления плановых и фактических треков. Кросс-корреляция данных дорожного трафика и погодных условий: как осадки влияют на скорость потока. Преобразование результатов анализа в рекомендации по корректировке расписания (АД-3).
5	Тема 5. Основы эконометрического моделирования спроса на транспорт Рассматриваются вопросы: Спецификация линейной регрессионной модели спроса (зависимость пассажиропотока от времени суток и интервала движения). Обучение модели через МНК (OLS) с использованием библиотеки Statsmodels. Интерпретация коэффициентов: эластичность спроса по интервалу движения транспорта. Проверка значимости предикторов (p-value) и качества модели (R-squared, adjusted R-squared). Интерпретация результатов для бизнеса: какой выигрыш даст сокращение интервала на 1 минуту (АД-3).
6	Тема 6. Диагностика и устранение проблем регрессионных моделей на транспортных данных Рассматриваются вопросы: Анализ остатков на гетероскедастичность (тест Бройша-Пагана) и мультиколлинеарность (VIF) факторов (цена билета, доходность рейса). Работа с фиктивными переменными (dummy variables) для учета типа подвижного состава (автобус/троллейбус/трамвай). Применение робастных стандартных ошибок (HC) для коррекции выводов. Вывод бизнес-инсайта: является ли тип транспорта значимым фактором выручки (АД-3).
7	Тема 7. Временные ряды в транспорте: прогнозирование и тренды Рассматриваются вопросы: Декомпозиция временного ряда пассажиропотока аэропорта: выделение тренда, сезонности и остатков. Модели экспоненциального сглаживания (ETS, Holt-Winters) для краткосрочного прогноза загруженности парковок. Оценка качества прогноза (MAE, MAPE) на отложенной выборке. Формулировка рекомендаций для службы эксплуатации: необходимое количество персонала в пиковые часы (АД-3).
8	Тема 8. Логистическая регрессия и оценка вероятности событий на транспорте (Часть 1) Рассматриваются вопросы: Постановка задачи бинарной классификации: прогнозирование вероятности опоздания рейса (label: 1/0). Выбор признаков: загрузка дорог, время суток, стаж водителя. Обучение модели логистической регрессии (библиотека Scikit-learn). Метрики качества: Accuracy, Precision, Recall, F1-score применительно к цене ошибки (ложный выезд резервного автобуса vs срыв расписания).
9	Тема 9. Логистическая регрессия и оценка вероятности событий на транспорте (Часть 2) Рассматриваются вопросы:

№ п/п	Тематика практических занятий/краткое содержание
	Визуализация ROC-AUC кривой и выбор порога классификации для минимизации финансовых потерь компании. Интерпретация коэффициентов логистической регрессии (log odds) для операторов диспетчерской службы. Создание скоринговой карты «риска опоздания» на основе модели. Преобразование математического вывода в управленческий регламент: при каком score необходимо вмешательство диспетчера (АД-3).
10	Тема 10. Деревья решений и ансамблирование в задачах оценки ресурса деталей транспорта Рассматриваются вопросы: Применение Decision Trees и Random Forest для классификации состояний узлов вагона (требуется ремонт / исправен) на основе данных телеметрии (вибрация, температура). Сравнение важности признаков (Feature Importance), интерпретируемость вывода модели для инженерного состава. Преобразование «машинного» признака в бизнес-рекомендацию: какие узлы чаще выходят из строя при экстремальных температурах. Библиотека Scikit-learn: обучение и подбор гиперпараметров (max_depth).
11	Тема 11. Метод опорных векторов (SVM) и кластеризация ДТП Рассматриваются вопросы: Классификация тяжести ДТП (легкое / тяжелое) методом SVM с различными ядрами. Метод главных компонент (PCA) для визуализации многомерных факторов аварийности. Метод k-средних (K-Means) для выявления «очагов аварийности» в городе по координатам. Вывод рекомендаций для дорожной инспекции: где установить камеры контроля скорости в первую очередь (АД-3).
12	Тема 12. Базовые модели эконометрики дискретного выбора на транспорте Рассматриваются вопросы: Пробит- и логит-модели множественного выбора: как пассажир выбирает вид транспорта (автобус, каршеринг, такси) в зависимости от цены и времени в пути. Расчет предельных эффектов (marginal effects) для оценки чувствительности к изменению тарифа. Оценка моделей в Statsmodels. Бизнес-инсайт: насколько нужно субсидировать проезд, чтобы пересадить 10% автомобилистов на общественный транспорт (АД-3).
13	Тема 13. Панельные данные и модели с фиксированными эффектами (FE) Рассматриваются вопросы: Сборка панельных данных: показатели транспортных компаний города за 5 лет. Построение модели с фиксированными эффектами для оценки влияния обновления подвижного состава на операционные расходы. Сравнение моделей: сквозная (Pooled), FE и случайные эффекты (RE) через тест Хаусмана. Интерпретация результатов как доказательной базы для получения субсидий (АД-3).
14	Тема 14. Метрики классификации и работа с несбалансированными данными (Отказы оборудования) Рассматриваются вопросы: Проблема несбалансированных классов на транспорте: поломка двигателя (1%) против штатной работы (99%). Методы борьбы с дисбалансом: SMOTE, class_weight в моделях машинного обучения. Выбор метрики качества: почему Accuracy бесполезна, а ключевое значение имеет Recall (не пропустить поломку). Расчёт затрат на ложноположительное и ложноотрицательное срабатывание системы предсказания ремонтов (АД-3).

№ п/п	Тематика практических занятий/краткое содержание
15	Тема 15. Комплексный практикум: предсказание времени прибытия транспорта (ETA) Рассматриваются вопросы: Сборка пайплайна обработки данных (Pipelines в Sklearn). Обучение моделей регрессии для предсказания времени в пути (Linear Regression, Decision Tree Regressor, Gradient Boosting). Сравнение моделей по MAE и RMSE с использованием кросс-валидации. Умение применять библиотеки: инференс лучшей модели. Бизнес-интерпретация: приложение для пассажиров с прогнозом времени подачи автобуса (АД-3).
16	Тема 16. От ML-модели к управленческому решению: итоговый кейс по компетенциям Рассматриваются вопросы: Разработка прототипа скоринговой модели для транспортного оператора (задача на выбор: прогноз увольнения водителей, классификация опасного вождения, эконометрический анализ топливной эффективности). Упаковка модели в формат, пригодный для демонстрации заказчику (дашборд на Streamlit / Plotly Dash). Защита решения: рассказ о бизнес-инсайтах для виртуального «Совета директоров». Оценка экономического эффекта от внедрения рекомендаций на основе построенной модели (АД-3).

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение литературы по машинному обучению.
2	Работа с документацией библиотек Scikit-learn, XGBoost, LightGBM.
3	Решение задач с Kaggle (классификация, регрессия).
4	Подготовка к текущему контролю.
5	Выполнение курсовой работы.
6	Подготовка к промежуточной аттестации.
7	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых работ

Прогнозирование пассажиропотока на маршруте общественного транспорта с использованием моделей временных рядов.

Классификация опасного вождения на основе данных телеметрии подвижного состава.

Прогнозирование вероятности опоздания рейса с использованием методов машинного обучения.

Выявление «очагов аварийности» на дорожной сети с использованием кластеризации геопространственных данных.

Оценка эластичности спроса на общественный транспорт с помощью эконометрических моделей.

Прогнозирование отказов оборудования подвижного состава с использованием ансамблевых методов машинного обучения.

Анализ эффективности субсидирования маршрутов с использованием панельных данных и моделей с фиксированными эффектами.

Построение модели предсказания времени прибытия (ETA) с оценкой качества прогнозирования.

Сравнительный анализ моделей классификации для прогнозирования увольнения водителей транспортной компании.

Разработка системы мониторинга загруженности парковок на основе прогнозных моделей временных рядов.

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

№ п/п	Библиографическое описание	Место доступа
1	Лабковская, Р. Я. Анализ больших данных : учебное пособие / Р. Я. Лабковская, П. В. Косов. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2025. — 152 с. — ISBN 978-5-89160-366-0.	https://e.lanbook.com/book/508654
2	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

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:

Зачет в 3 семестре.

Курсовая работа в 3 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