

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)

Большие данные на транспорте

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте (Российско-Китайская программа)

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: 35225

Подписал: заместитель директора Игольников Борис
Владимирович

Дата: 16.10.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины является формирование у обучающихся системных знаний и практических навыков в области анализа больших данных на транспорте, включая понимание методов перевода бизнес-требований в аналитические задачи, ключевых бизнес-метрик транспортной отрасли, критериев выбора алгоритмов для различных задач, а также принципов причинно-следственного вывода и визуализации результатов для поддержки принятия решений.

Задачами освоения дисциплины являются:

Формирование теоретических знаний о методах перевода бизнес-требований в аналитические задачи в транспортной отрасли.

Освоение практических навыков применения стандартных библиотек глубокого обучения для решения транспортных задач.

Приобретение умений выбирать классические модели машинного обучения (регрессия, классификация) для различных аналитических задач на транспорте, а также применять методы причинно-следственного вывода для интерпретации результатов.

Освоение методов визуализации результатов анализа и создания интерактивных дашбордов.

Приобретение навыков работы с ключевыми бизнес-метриками транспортной отрасли.

Дисциплина предназначена для формирования компетенций, необходимых для решения профессиональных задач, связанных с проведением аналитических исследований с применением технологий больших данных, проектированием алгоритмов машинного обучения и применением методов статистического анализа на транспорте.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ПК-1 - Способен проектировать, разрабатывает и улучшать алгоритмы и модели машинного обучения;

ПК-6 - Способен проводить аналитические исследования с применением технологий больших данных.

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

Знать:

Методы перевода бизнес-требований в аналитические задачи и ключевые бизнес-метрики в предметной области транспортной отрасли.

Критерии выбора алгоритмов для различных задач анализа больших данных и принципы причинно-следственного вывода.

Методы визуализации результатов и создания интерактивных дашбордов для представления аналитических данных.

Принципы причинно-следственного вывода и методы их применения для интерпретации результатов анализа на транспорте.

Уметь:

Применять стандартные библиотеки, реализующие основные архитектуры глубокого обучения, для решения транспортных задач.

Выбирать оптимальные типы визуализации и создавать интерактивные дашборды для представления результатов.

Визуализировать результаты экспериментов и применять принципы причинно-следственного вывода для интерпретации данных.

Применять принципы причинно-следственного вывода для интерпретации данных и формулирования выводов.

Владеть:

Навыками перевода бизнес-требований в аналитические задачи и выбора алгоритмов для их решения.

Навыками создания аналитических отчетов и интерактивных дашбордов для транспортных данных.

Навыками применения принципов причинно-следственного вывода и методов статистического анализа для интерпретации результатов.

Навыками выбора классических моделей машинного обучения (регрессия, классификация) для транспортных задач.

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. Установка и настройка среды для работы с большими данными. Рассматриваемые вопросы: - Установка и настройка PySpark (локальный режим, Databricks Community Edition). - Установка Jupyter Notebook и необходимых библиотек. - Загрузка транспортных датасетов (GTFS, данные GPS-треков). - Первичное знакомство со Spark DataFrame и Spark SQL.
2	Тема 2. Базовые операции с большими данными в PySpark. Рассматриваемые вопросы: - Чтение данных из различных источников (CSV, Parquet, JSON). - Базовые трансформации: выбор колонок, фильтрация, сортировка. - Агрегации и группировки (groupBy, agg). - Создание и использование временных представлений (Spark SQL).
3	Тема 3. Очистка и предобработка больших транспортных данных в PySpark. Рассматриваемые вопросы: - Обработка пропущенных значений в больших данных (fillna, dropna).

№ п/п	Тематика практических занятий/краткое содержание
	- Выявление и удаление дублей (dropDuplicates). - Работа с выбросами в масштабах больших данных. - Преобразование типов и форматов данных. - Оптимизация производительности при работе с большими данными.
4	Тема 4. Описательная статистика и визуализация транспортных данных. Рассматриваемые вопросы: - Расчет статистических характеристик в PySpark (describe, approxQuantile). - Визуализация распределений с помощью Matplotlib/Seaborn на основе агрегированных данных. - Построение корреляционных матриц. - Интерпретация статистических показателей для бизнеса.
5	Тема 5. Трансформация бизнес-требований в аналитические задачи (практикум). Рассматриваемые вопросы: - Анализ бизнес-кейса из транспортной отрасли (например, прогнозирование пассажиропотока). - Выявление потребностей заказчика и формулировка бизнес-задачи. - Перевод бизнес-требований в аналитические задачи. - Определение KPI и метрик успеха проекта. - Подготовка краткого технического задания.
6	Тема 6. Критерии выбора алгоритмов для транспортных задач (практикум). Рассматриваемые вопросы: - Анализ характеристик данных для транспортной задачи. - Сравнение алгоритмов для задачи прогнозирования/классификации. - Обоснование выбора алгоритма с учетом ограничений (время обучения, интерпретируемость). - Подготовка аналитической записки с обоснованием выбора.
7	Тема 7. Проверка статистических гипотез на практике. Рассматриваемые вопросы: - Формулировка гипотез для транспортного кейса (например, влияние нового маршрута на пассажиропоток). - Применение t-теста и критерия ?? в Python (SciPy). - Интерпретация p-value и уровня значимости. - Формулировка выводов для бизнес-решений.
8	Тема 8. Планирование и проведение А/В-тестирования. Рассматриваемые вопросы: - Постановка задачи для А/В-теста на транспорте (например, изменение расписания). - Расчет размера выборки для достижения заданной мощности. - Проведение А/В-теста с использованием инструментов. - Сбор и первичный анализ данных.
9	Тема 9. Анализ результатов А/В-тестирования. Рассматриваемые вопросы: - Статистическая обработка результатов А/В-теста. - Визуализация результатов эксперимента. - Интерпретация p-value и доверительных интервалов. - Принятие решения на основе результатов теста. - Формулировка рекомендаций для бизнеса.
10	Тема 10. Применение причинно-следственного вывода (Causal Inference). Рассматриваемые вопросы: - Анализ транспортного кейса на предмет причинно-следственных связей. - Применение методов Causal Inference (например, Difference-in-Differences). - Интерпретация результатов и формулировка выводов. - Подготовка отчета с рекомендациями.

№ п/п	Тематика практических занятий/краткое содержание
11	Тема 11. Прогнозирование пассажиропотока с использованием LSTM. Рассматриваемые вопросы: - Подготовка временных рядов для модели. - Построение и обучение LSTM-модели (Keras/TensorFlow). - Оценка качества прогноза (MAE, RMSE, MAPE). - Визуализация прогнозов и интерпретация результатов.
12	Тема 12. Анализ и оптимизация маршрутов с использованием больших данных. Рассматриваемые вопросы: - Анализ данных о движении транспорта (GPS-треки). - Выявление загруженных участков и временных пиков. - Применение методов оптимизации маршрутов. - Визуализация результатов оптимизации на карте.
13	Тема 13. Валидация и оценка качества моделей на больших данных. Рассматриваемые вопросы: - Применение временной кросс-валидации для прогнозных моделей. - Расчет метрик качества для различных моделей. - Анализ остатков и выявление систематических ошибок. - Формулировка рекомендаций по улучшению модели.
14	Тема 14. Комплексный проект: постановка задачи и анализ данных. Рассматриваемые вопросы: - Формулировка бизнес-задачи на основе реального транспортного кейса. - Перевод бизнес-требований в аналитические задачи. - Сбор и предобработка больших данных в PySpark. - Разведочный анализ данных и формулировка гипотез. - Выбор алгоритмов и обоснование выбора.
15	Тема 15. Комплексный проект: реализация, интерпретация и подготовка отчета. Рассматриваемые вопросы: - Реализация ETL-пайплайна в PySpark. - Применение статистических методов и ML-моделей. - Применение методов причинно-следственного вывода. - Интерпретация результатов и формулировка выводов. - Подготовка итогового отчета и презентации.
16	Тема 16. Комплексный проект: защита. Рассматриваемые вопросы: - Доработка проекта по результатам проверки. - Подготовка финальной версии отчета. - Публичная защита проекта. - Ответы на вопросы и обсуждение результатов.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение документации по инструментам работы с большими данными.
2	Подготовка к практическим занятиям и выполнение домашних заданий.
3	Работа с реальными транспортными датасетами.
4	Изучение кейсов применения больших данных на транспорте.
5	Изучение принципов причинно-следственного вывода.

№ п/п	Вид самостоятельной работы
6	Выполнение комплексного проекта.
7	Подготовка к промежуточной аттестации и текущему контролю.
8	Выполнение курсовой работы.
9	Подготовка к промежуточной аттестации.
10	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых работ

1. Перевод бизнес-требований транспортной компании в аналитические задачи: методология и практика.
2. Ключевые бизнес-метрики транспортной отрасли: анализ, расчёт и интерпретация.
3. Разведочный анализ больших транспортных данных с использованием PySpark.
4. Очистка и предобработка больших транспортных данных: методы и оптимизация.
5. Проверка статистических гипотез в транспортных исследованиях.
6. Планирование и проведение А/В-тестирования на транспорте.
7. Анализ результатов А/В-тестирования транспортных решений.
8. Причинно-следственный вывод в анализе транспортных данных.
9. Прогнозирование пассажиропотока с использованием LSTM-моделей.
10. Сравнительный анализ классических моделей машинного обучения для прогнозирования транспортных показателей.
11. Прогнозирование задержек транспорта на основе больших данных.
12. Валидация и оценка качества прогнозных моделей на больших транспортных данных.
13. Анализ и оптимизация маршрутов транспортной сети с использованием больших данных.
14. Выявление временных пиков и загруженных участков транспортной сети на основе больших данных.
15. Оптимизация расписания общественного транспорта на основе анализа больших данных.
16. Применение методов кластеризации для анализа транспортных потоков.
17. Разработка интерактивных дашбордов для анализа транспортных данных.

18. Визуализация результатов анализа больших транспортных данных: методы и инструменты.

19. Визуализация пространственно-временных транспортных данных.

20. Разработка ETL-пайплайна для обработки больших транспортных данных в PySpark.

21. Анализ данных GTFS для оценки качества транспортного обслуживания.

22. Анализ данных GPS-треков для мониторинга транспортных средств.

23. Применение больших данных для анализа пассажиропотока городского транспорта.

24. Оценка влияния транспортных решений на основе причинно-следственного вывода.

25. Комплексный проект: анализ больших транспортных данных с разработкой ML-модели и дашборда.

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

№ п/п	Библиографическое описание	Место доступа
1	Горшенин, А. К. Методы машинного обучения. Классические алгоритмы : учебное пособие / А. К. Горшенин. — Москва : РТУ МИРЭА, 2026. — 121 с. — ISBN 978-5-7339-2955-2.	https://e.lanbook.com/book/523312
2	Асалханова, Т. Н. Организация производственных процессов в путевом хозяйстве с использованием информационных систем управления железнодорожным транспортом : учебно-методическое пособие / Т. Н. Асалханова, И. Г. Карпов. — Иркутск : ИрГУПС, 2022. — 100 с.	https://e.lanbook.com/book/284525
3	Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов. — Москва : МИСИ – МГСУ, 2020. — 46 с. — ISBN 978-5-7264-2193-3.	https://e.lanbook.com/book/145102

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).

Python 3.x
Jupyter Notebook / Jupyter Lab
Система контроля версий Git
Офисный пакет приложений
Браузер
PySpark
Apache Spark
TensorFlow / Keras / PyTorch
Databricks Community Edition

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

Компьютер преподавателя
Компьютеры студентов
Проектор
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:

заместитель директора

Б.В. Игольников

Согласовано:

Директор

Д.В. Паринов

Заместитель директора

Б.В. Игольников

Председатель учебно-методической
комиссии

Д.В. Паринов