

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 - программы магистратуры
по направлению подготовки
09.04.01 Информатика и вычислительная техника,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Анализ данных и машинное обучение

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Искусственный интеллект и предиктивная аналитика в транспортных системах

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: 5665
Подписал: заведующий кафедрой Нутович Вероника
Евгеньевна
Дата: 01.09.2024

1. Общие сведения о дисциплине (модуле).

Цель дисциплины «Анализ данных и машинное обучение» заключается в освоении профессиональных компетенций в области разработки и применения искусственного интеллекта и методов машинного обучения, пригодных для практического применения и анализа данных.

Задачей дисциплины является формирование у обучающихся навыков проектирования моделей искусственного интеллекта, подготовки данных для обучения и отладки средствами современных библиотек языка программирования Python

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-2 - Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач;

ОПК-3 - Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями;

ОПК-4 - Способен применять на практике новые научные принципы и методы исследований;

ПК-2 - Способен осуществить сбор, очистку, подготовку и разметку данных используя методологию ETL для дальнейшего обучения моделей искусственного интеллекта;

ПК-3 - Способен спроектировать, разработать, обучить, оценить и развернуть модели искусственного интеллекта в соответствии с методологией MLOps.

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

Знать:

- основные задачи, решаемые методами машинного обучения и анализа данных;
- принципы построения моделей искусственного интеллекта;
- метода сбора, подготовки и преобразования данных для последующего обучения модели искусственного интеллекта;

- этапы полного цикла работы с данными, в соответствии с ETL-концепцией;
- подходы к обучению моделей искусственного интеллекта;
- возможности инструментов для разработки технологий искусственного интеллекта.

Уметь:

- оценивать качество данных, проводить анализ и отбор признаков;
- проектировать, обучать, развертывать модели искусственного интеллекта;
- тестировать и оценивать работоспособность модели искусственного интеллекта;
- применять инструменты разработки моделей;
- применять инструменты для визуализации работы моделей;
- интегрировать программные модули и библиотеки.

Владеть:

- навыками построения модели искусственного интеллекта для прикладных задач и задач профессиональной деятельности;
- навыком оптимизации модели путем подбора параметров;
- навыком использования специализированных библиотек языка Python для анализа данных и машинного обучения;

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 10 з.е. (360 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	№2	№3
Контактная работа при проведении учебных занятий (всего):	96	48	48	
В том числе:				
Занятия лекционного типа	32	16	16	
Занятия семинарского типа	64	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 264 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	<p>Введение в искусственный интеллект.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- история искусственного интеллекта;- тест Тьюринга;- слабый и сильный ИИ;- классификация методов искусственного интеллекта;- области применения ИИ.
2	<p>Экспертные системы.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- структура экспертной системы;- классификация экспертных систем;- неопределенность знаний и коэффициент доверия в экспертных системах;- этапы разработки экспертных систем.
3	<p>Эволюционное моделирование.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- генетические и эволюционные алгоритмы;- методы селекции в генетических алгоритмах;- методы скрещивания в генетических алгоритмах.
4	<p>Введение в машинное обучение.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- основы машинного обучения;- классификация типов систем машинного обучения;- методы обучения с учителем и без учителя;- основные проблемы машинного обучения;- проблема переобучения и недообучения;- примеры задач машинного обучения.

№ п/п	Тематика лекционных занятий / краткое содержание
5	<p>Методы предобработки данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - оценка качества данных; - анализ и отбор признаков; - виды признаков; - нормировка признаков.
6	<p>Методы снижения размерности данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - метод главных компонент; - учет и заполнение пропусков; - поиск аномалий; - татистический анализ данных; - отбеливание данных.
7	<p>Метод опорных векторов SVM.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - линейная и нелинейная классификация; - классификация с мягким зазором; - выявление выбросов.
8	<p>Линейная регрессия;</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - метод наименьших квадратов; - принцип максимизации максимального правдоподобия; - разложение ошибки на смещение и разброс.
9	<p>Логистическая регрессия.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - градиентный спуск; - метрики, используемые в задачах регрессии: среднеквадратичная ошибка, средняя абсолютная ошибка, коэффициент детерминации.
10	<p>Классификация.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - задачи и функция классификации; - линейная разделимость признаков; - геометрическая интерпретация отступа; - метрики, используемые в задачах классификации: верность, матрица ошибок, точность, полнота, F-мера, ROC-кривая, PR-кривая, площадь ROC-кривой.
11	<p>Кластеризация.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - задача кластеризации данных; - обзор алгоритмов кластеризации; - Алгоритм k-средних; - иерархическая кластеризация; - извлечение признаков из алгоритмов кластеризации.
12	<p>Деревья и градиентный бустинг.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - деревья решений; - принципы построения дерева решений; - прирост информации и энтропия; - деревья для задач регрессии и классификации; - градиентный бустинг; - бустинг деревьев;

№ п/п	Тематика лекционных занятий / краткое содержание
	<ul style="list-style-type: none"> - алгоритмы бустинга; - функции потерь.
13	Анализ временных рядов. Рассматриваемые вопросы: <ul style="list-style-type: none"> - анализ данных временных рядов; - извлечение признаков из временного ряда.
14	Ансамбли моделей. Рассматриваемые вопросы: <ul style="list-style-type: none"> - ансамблирование моделей; - случайный лес; - бэггинг моделей; - ошибка при использовании бэггинга; - стэкинг моделей; - блэндинг моделей.
15	Введение в глубокое обучение. Рассматриваемые вопросы: <ul style="list-style-type: none"> - сравнение глубокого обучения и классических подходов к машинному обучению; - повторное использование заранее обученных слоев; - глубокое обучение с подкреплением.
16	Оценка моделей. Рассматриваемые вопросы: <ul style="list-style-type: none"> - методики оценки моделей; - инструменты мониторинга работы моделей; - инструменты визуализации работы модели.
17	Развертывание моделей. Рассматриваемые вопросы: <ul style="list-style-type: none"> - форматы моделей; - хранилища моделей; - подход MLOps.
18	Процесс управления полным циклом работы с данными Рассматриваемые вопросы: <ul style="list-style-type: none"> - концепция ETL (Extract, Transform, Load).

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	Визуализация данных в Matplotlib. В результате выполнения лабораторных работ студент получает навыки отображения данных, построения линейных и трехмерных графиков, точечных диаграмм средствами библиотеки Matplotlib
2	Визуализация данных в Plotly. В результате выполнения лабораторных работ студент получает навыки отображения интерактивных данных, динамических графиков средствами библиотеки Plotly
3	Визуализация данных в Seaborn. В результате выполнения лабораторных работ студент получает навыки отображения данных и визуализации парных взаимосвязей средствами библиотеки Seaborn.

№ п/п	Наименование лабораторных работ / краткое содержание
4	Анализ и оценка данных в Pandas. В результате выполнения лабораторных работ студент получает навыки первичного анализа, обработки и оценки данных, манипулирования и сортировки в Pandas
5	Анализ и оценка данных в NumPy. В результате выполнения лабораторных работ студент получает навыки первичного анализа, обработки и оценки данных, работы с векторами и матрицами в NumPy.
6	Анализ и оценка данных в SciPy. В результате выполнения лабораторных работ студент получает навыки первичного анализа, обработки и оценки данных, поиска минимумов и максимумов функций в SciPy.
7	Задачи регрессии в XGBoost В результате выполнения лабораторных работ студент получает навыки использования градиентного бустинга в задачах регрессии средствами библиотеки XGBoost
8	Задачи классификации в XGBoost В результате выполнения лабораторных работ студент получает навыки использования градиентного бустинга в задачах классификации средствами библиотеки XGBoost
9	Задачи регрессии в LightGBM В результате выполнения лабораторных работ студент получает навыки использования градиентного бустинга в задачах регрессии средствами библиотеки LightGBM
10	Задачи классификации в LightGBM В результате выполнения лабораторных работ студент получает навыки использования градиентного бустинга в задачах классификации средствами библиотеки LightGBM
11	Задачи регрессии в CatBoost В результате выполнения лабораторных работ студент получает навыки использования градиентного бустинга в задачах регрессии средствами библиотеки CatBoost
12	Задачи классификации в CatBoost В результате выполнения лабораторных работ студент получает навыки использования градиентного бустинга в задачах классификации средствами библиотеки CatBoost
13	Задачи классификации с библиотекой Scikit Learn В результате выполнения лабораторных работ студент получает навыки первичного анализа, обработки и оценки данных, последующего обучения модели для задач классификации средствами библиотеки Scikit Learn
14	Задачи кластеризации с библиотекой Scikit Learn В результате выполнения лабораторных работ студент получает навыки первичного анализа, обработки и оценки данных, последующего обучения модели для задач кластеризации средствами библиотеки Scikit Learn
15	Задачи регрессии с библиотекой Scikit Learn В результате выполнения лабораторных работ студент получает навыки первичного анализа, обработки и оценки данных, последующего обучения модели для задач регрессии средствами библиотеки Scikit Learn
16	Глубокое обучение с фреймворком PyTorch В результате выполнения лабораторных работ студент получает навыки использования и отладки методов глубокого обучения искусственной нейронной сети средствами фреймворка PyTorch
17	Глубокое обучение с фреймворком TensorFlow В результате выполнения лабораторных работ студент получает навыки использования и отладки методов глубокого обучения искусственной нейронной сети средствами фреймворка TensorFlow
18	Глубокое обучение с фреймворком Keras В результате выполнения лабораторных работ студент получает навыки использования и отладки методов глубокого обучения искусственной нейронной сети средствами фреймворка Keras

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Установка и настройка среды разработки. В ходе практических занятий студенты освежают навыки использования среды разработки PyCharm для написания, отладки и запуска программ на языке Python
2	Основы программирования на языке Python. В ходе практических занятий студенты повторяют основы программирования на языке Python, синтаксис языка программирования, встроенные функции, типы и структуры данных.
3	Установка и использование библиотеки Matplotlib В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки Matplotlib
4	Установка и использование библиотеки Plotly В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки Plotly
5	Установка и использование библиотеки Seaborn В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки Seaborn
6	Установка и использование библиотеки Pandas В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки Pandas
7	Установка и использование библиотеки NumPy В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки NumPy
8	Установка и использование библиотеки SciPy В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки SciPy
9	Установка и использование библиотеки XGBoost В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки XGBoost
10	Установка и использование библиотеки LightGBM В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки LightGBM
11	Установка и использование библиотеки CatBoost В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки CatBoost
12	Установка и использование библиотеки Scikit Learn В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения. Изучают функционал библиотеки Scikit Learn
13	Установка Jupyter Notebook В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков

№ п/п	Тематика практических занятий/краткое содержание
	открытового и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения
14	Установка фреймворка PyTorch В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения
15	Установка фреймворка TensorFlow В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения
16	Установка фреймворка Keras В ходе практических занятий студенты знакомятся с установкой внешних пакетов поставщиков открытого и свободно распространяемого исходного кода на локальную машину и настройку развернутого приложения

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение рекомендованной литературы.
2	Подготовка к лабораторным и практическим занятиям.
3	Подготовка к промежуточной аттестации.
4	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Келлехер, Д. Наука о данных: базовый курс / Джон Келлехер, Брендан Тирни ; пер. с англ.. - Москва : Альпина Паблишер, 2020. - 222 с. - ISBN 978-5-9614-3170-4. - Текст : электронный	https://znanium.ru/catalog/product/1221800 (дата обращения: 09.04.2025)
2	Андропова, О. Ю. Искусственный интеллект и язык программирования Python : учебное пособие / О. Ю. Андропова, И. И. Васильева, Н. А. Гнездилова. — Елец : ЕГУ им. И.А. Бунина, 2024. — 106 с. — ISBN 978-5-00151-413-8. — Текст : электронный	https://e.lanbook.com/book/462479 (дата обращения: 09.04.2025)
3	Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. —	https://e.lanbook.com/book/131721 (дата обращения: 09.04.2025)

	Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный	
4	Ма, К. Трехмерное глубокое обучение на Python : руководство / К. Ма, В. Хегде, Л. Йольян ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2023. — 226 с. — ISBN 978-5-93700-202-0. — Текст : электронный	https://e.lanbook.com/book/455303 (дата обращения: 09.04.2025)
5	Митяков, Е. С. Искусственный интеллект и машинное обучение : учебное пособие для вузов / Е. С. Митяков, А. Г. Шмелева, А. И. Ладынин. — Санкт-Петербург : Лань, 2025. — 252 с. — ISBN 978-5-507-51465-6. — Текст : электронный	https://e.lanbook.com/book/450827 (дата обращения: 09.04.2025)
6	Гафаров, Ф. М. Нейронные сети в PyTorch : учебное пособие / Ф. М. Гафаров, А. Ф. Гилемзянов. - Казань : Казанский федеральный университет, 2024. - 106 с. - Текст : электронный	https://znanium.ru/catalog/product/2173433 (дата обращения: 09.04.2025)
7	Антонио, Д. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / Д. Антонио, П. Суджит ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2018. — 294 с. — ISBN 978-5-97060-573-8. — Текст : электронный	https://e.lanbook.com/book/111438 (дата обращения: 09.04.2025)
8	Потопахин, В. В. Романтика искусственного интеллекта / В. В. Потопахин. — Москва : ДМК Пресс, 2017. — 170 с. — ISBN 978-5-97060-476-2. — Текст : электронный	https://e.lanbook.com/book/93578 (дата обращения: 09.04.2025)

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

- Научно-техническая библиотека РУТ (МИИТ) (<http://library.miit.ru>);
- Информационный портал Научная электронная библиотека eLIBRARY.RU (www.elibrary.ru);
- Образовательная платформа «Юрайт» (<https://urait.ru/>);
- Электронно-библиотечная система издательства «Лань» (<http://e.lanbook.com/>);

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

Прикладное программное обеспечение

Браузер Microsoft Internet Explorer или его аналоги

Пакет офисных программ Microsoft Office или его аналоги

Среда разработки PyCharm Community Edition

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.

Для практических занятий – наличие персональных компьютеров вычислительного класса.

9. Форма промежуточной аттестации:

Зачет во 2 семестре.

Экзамен в 3 семестре.

10. Оценочные материалы.

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

Авторы:

заведующий кафедрой, доцент, к.н.
кафедры «Цифровые технологии
управления транспортными
процессами»

В.Е. Нутович

старший преподаватель кафедры
«Цифровые технологии управления
транспортными процессами»

Е.А. Заманов

Согласовано:

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической
комиссии

Н.А. Андриянова