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)

Введение в машинное обучение

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).

Целью освоения учебной дисциплины является формирование у обучающихся фундаментальных знаний и первичных практических навыков в области машинного обучения, включая понимание базовых алгоритмов обучения с учителем и без учителя, методов предобработки данных, оценки качества моделей и их применения для решения типовых прикладных задач.

Задачами освоения дисциплины являются:

1. Формирование теоретических знаний об основных алгоритмах машинного обучения, их классификации, областях применения и ограничениях.

2. Освоение практических навыков предобработки и анализа данных с использованием стандартных библиотек машинного обучения.

3. Приобретение умений применять алгоритмы обучения с учителем для задач регрессии и классификации.

4. Формирование компетенций по использованию алгоритмов обучения без учителя для задач кластеризации и поиска аномалий.

5. Освоение базовых методов оценки качества моделей и критериев качества данных для различных аналитических задач.

6. Приобретение навыков настройки гиперпараметров моделей с использованием стандартных подходов.

Дисциплина предназначена для формирования компетенций, необходимых для решения профессиональных задач, связанных с проектированием, разработкой и улучшением алгоритмов и моделей машинного обучения, а также подготовкой и обработкой данных для аналитических исследований.

2. Планируемые результаты обучения по дисциплине (модулю).

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

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

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

Знать:

1. Базовые алгоритмы обучения с учителем для регрессии и классификации, включая линейные методы, метод опорных векторов и

ансамблевые методы, а также основные ограничения классического машинного обучения и области их применения.

2. Базовые алгоритмы обучения без учителя (кластеризация, поиск аномалий), подходы к настройке гиперпараметров, а также основные принципы планирования экспериментов (рандомизация, контрольные группы).

3. Критерии качества данных для различных аналитических задач и базовые методы оценки качества моделей машинного обучения.

Уметь:

1. Применять стандартные библиотеки машинного обучения для реализации алгоритмов обучения с учителем и без учителя, включая предобработку данных и оценку качества моделей.

2. Выбирать и обосновывать выбор алгоритмов машинного обучения в зависимости от типа решаемой задачи и характеристик данных на базовом уровне.

3. Применять простейшие методы автоматического машинного обучения (AutoML) для базового подбора моделей и гиперпараметров.

Владеть:

1. Навыками предобработки и нормализации данных, включая работу с пропущенными значениями, кодирование категориальных признаков и масштабирование.

2. Навыками применения базовых алгоритмов машинного обучения для решения задач регрессии, классификации и кластеризации с использованием стандартных библиотек.

3. Навыками оценки качества моделей с использованием релевантных метрик, настройки гиперпараметров и интерпретации результатов с учетом критериев качества данных для различных аналитических 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 №1
Контактная работа при проведении учебных занятий (всего):	32	32
В том числе:		
Занятия лекционного типа	16	16
Занятия семинарского типа	16	16

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.

№ п/п	Тематика лекционных занятий / краткое содержание
1	Тема 1. Введение в машинное обучение. Рассматриваемые вопросы: - Определение машинного обучения, связь с Data Science и искусственным интеллектом. - Классификация задач машинного обучения: обучение с учителем, без учителя, с подкреплением. - Основная терминология: признаки, объекты, целевая переменная, выборка. - Обзор жизненного цикла проекта по машинному обучению. - Области применения машинного обучения и основные ограничения классических подходов.
2	Тема 2. Основы статистики и вероятности для машинного обучения. Рассматриваемые вопросы: - Основные понятия статистики и теории вероятностей. - Дискретные и непрерывные случайные величины. - Распределения вероятностей (нормальное, биномиальное, равномерное). - Проверка гипотез и доверительные интервалы. - Роль статистики в машинном обучении и интерпретации результатов.
3	Тема 3. Предобработка и подготовка данных. Рассматриваемые вопросы: - Методы очистки данных: обработка пропущенных значений (удаление, заполнение средним/медианой/модой). - Выявление и удаление выбросов с использованием межквартильного размаха и Z-оценки. - Нормализация (Min-Max) и стандартизация (Z-score) данных.

№ п/п	Тематика лекционных занятий / краткое содержание
	- Кодирование категориальных признаков (Label Encoding, One-Hot Encoding). - Критерии качества данных для различных аналитических задач.
4	Тема 4. Линейные модели для регрессии и классификации. Рассматриваемые вопросы: - Линейная регрессия: математическое описание, интерпретация коэффициентов, применение. - Логистическая регрессия для задач бинарной классификации. - Градиентный спуск и его вариации (стохастический, мини-пакетный). - Базовые метрики оценки качества линейных моделей (MSE, MAE, Accuracy). - Преимущества и ограничения линейных методов.
5	Тема 5. Метод опорных векторов (SVM). Рассматриваемые вопросы: - Постановка задачи классификации с использованием SVM. - Понятие разделяющей гиперплоскости и опорных векторов. - Линейно разделимые и неразделимые данные (мягкий зазор). - Ядровые функции (линейное, полиномиальное, RBF) и их роль в SVM. - Применение SVM для задач классификации.
6	Тема 6. Деревья решений и ансамблевые методы. Рассматриваемые вопросы: - Основы деревьев решений: алгоритм построения, критерии разделения (Gini, энтропия). - Преимущества (интерпретируемость) и недостатки (переобучение) деревьев решений. - Ансамблевые методы: бэггинг и случайный лес. - Бустинг: общая идея AdaBoost и Gradient Boosting. - Применение ансамблевых методов для повышения точности и устойчивости моделей.
7	Тема 7. Кластеризация и методы обучения без учителя. Рассматриваемые вопросы: - Постановка задач кластеризации. - Алгоритм K-means: принцип работы, выбор числа кластеров (метод локтя). - Иерархическая кластеризация и построение дендрограммы. - Методы поиска аномалий на основе расстояний. - Оценка качества кластеризации (силуэт, индекс Дэвиса-Болдина).
8	Тема 8. Оценка качества моделей, перекрестная проверка и настройка гиперпараметров. Рассматриваемые вопросы: - Методы оценки качества классификации: точность, полнота, F-мера, матрица ошибок. - Методы оценки качества регрессии: MSE, MAE, RMSE, R?. - Перекрестная проверка (k-fold кросс-валидация). - Методы борьбы с переобучением (регуляризация на концептуальном уровне, ограничение сложности моделей). - Алгоритмы настройки гиперпараметров: Grid Search, Random Search. - Интерпретация метрик для бизнес-задач.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Установка и настройка среды для машинного обучения. Рассматриваемые вопросы: - Установка и настройка сред разработки (Anaconda, VS Code).

№ п/п	Тематика практических занятий/краткое содержание
	- Установка библиотек для машинного обучения (numpy, pandas, matplotlib, seaborn, scikit-learn). - Создание и выполнение первого проекта на Python. - Проверка работоспособности инструментов.
2	Тема 2. Базовые операции с данными с использованием библиотек. Рассматриваемые вопросы: - Загрузка данных из CSV и Excel. - Базовые операции с данными (выборка, фильтрация, группировка). - Визуализация данных (гистограммы, boxplot, scatter plot). - Анализ структуры и статистических характеристик данных.
3	Тема 3. Обработка пропущенных значений и выбросов. Рассматриваемые вопросы: - Выявление пропущенных значений с помощью heatmap и info(). - Методы заполнения пропусков (среднее, медиана, мода, интерполяция). - Выявление и удаление выбросов с использованием IQR и Z-score. - Оценка влияния обработки на качество данных.
4	Тема 4. Нормализация и кодирование признаков. Рассматриваемые вопросы: - Применение нормализации (MinMaxScaler) и стандартизации (StandardScaler). - Кодирование категориальных признаков (One-Hot Encoding, Label Encoding). - Сравнение различных методов кодирования. - Оценка влияния предобработки на качество моделей.
5	Тема 5. Реализация линейной регрессии. Рассматриваемые вопросы: - Построение модели линейной регрессии с использованием LinearRegression. - Обучение модели и оценка качества (MSE, MAE, R?). - Интерпретация коэффициентов модели. - Визуализация результатов (график предсказаний vs фактических значений).
6	Тема 6. Реализация логистической регрессии. Рассматриваемые вопросы: - Построение модели логистической регрессии с использованием LogisticRegression. - Обучение модели и оценка качества (Accuracy, Precision, Recall, F1). - Интерпретация результатов классификации. - Построение матрицы ошибок и ROC-кривой.
7	Тема 7. Метод опорных векторов на практике. Рассматриваемые вопросы: - Построение модели SVM для классификации с линейным и RBF-ядром. - Настройка гиперпараметров (C, gamma) с использованием Grid Search. - Сравнение с логистической регрессией на одном датасете. - Оценка качества модели.
8	Тема 8. Деревья решений на практике. Рассматриваемые вопросы: - Построение дерева решений с использованием DecisionTreeClassifier. - Визуализация дерева решений (plot_tree). - Настройка глубины и других гиперпараметров (max_depth, min_samples_split). - Оценка качества модели и анализ переобучения.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Работа с лекционным материалом и дополнительной литературой.
2	Изучение документации по библиотекам машинного обучения.
3	Подготовка к практическим занятиям и выполнение домашних заданий.
4	Подготовка к промежуточной аттестации и текущему контролю.
5	Подготовка к промежуточной аттестации.
6	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Котельников, Е. В. Введение в машинное обучение и анализ данных : учебное пособие / Е. В. Котельников, А. В. Котельникова. — Киров : ВятГУ, 2023. — 68 с.	https://e.lanbook.com/book/390698
2	Агбозо, Э. Введение в аналитику больших данных и машинное обучение : учебное пособие / Э. Агбозо, Д. М. Балунгу. — Екатеринбург : УрФУ, 2024. — 84 с. — ISBN 978-5-7996-3898-6.	https://e.lanbook.com/book/522872

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:

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