

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)

Оркестрация в машинном обучении

Направление подготовки: 09.03.01 Информатика и вычислительная техника

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

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: директор центра Проневич Ольга Борисовна

Дата: 29.09.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины является формирование у обучающихся практических навыков в области оркестрации пайплайнов машинного обучения, включая понимание процессов автоматизации последовательности задач по подготовке данных, обучению моделей, настройке гиперпараметров, балансировке данных и применению библиотек глубокого обучения для создания воспроизводимых и масштабируемых ML-пайплайнов.

Задачами освоения дисциплины являются:

Освоение практических навыков построения пайплайнов обработки данных и обучения моделей с использованием стандартных библиотек.

Формирование компетенций по автоматической настройке гиперпараметров моделей машинного обучения.

Приобретение умений применять методы балансировки данных в рамках оркестрируемых пайплайнов.

Освоение навыков применения библиотек глубокого обучения для решения прикладных задач.

Формирование навыков управления зависимостями и последовательностью выполнения задач в ML-пайплайнах.

Приобретение навыков оценки качества моделей и интерпретации результатов в рамках оркестрированных пайплайнов.

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):

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

ППК-ДА2 - Способен подготавливать и обрабатывать данные для аналитических исследований.

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

Знать:

Алгоритмы автоматической настройки гиперпараметров моделей машинного обучения, включая Grid Search, Random Search, Bayesian Optimization.

Методы балансировки данных для работы с несбалансированными выборками, включая техники ресемплинга и синтетической генерации данных.

Принципы построения пайплайнов машинного обучения и управления последовательностью выполнения задач.

Уметь:

Применять стандартные библиотеки, реализующие основные архитектуры глубокого обучения, для построения и обучения моделей.

Применять методы балансировки данных в рамках оркестрируемых пайплайнов.

Использовать инструменты автоматической настройки гиперпараметров для оптимизации моделей машинного обучения.

Владеть:

Навыками построения оркестрированных пайплайнов машинного обучения с использованием современных библиотек и инструментов.

Навыками применения методов балансировки данных и настройки гиперпараметров в составе пайплайнов.

Навыками оценки качества моделей и интерпретации результатов работы пайплайнов.

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 №2
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 оркестрации ML-пайплайнов. Рассматриваемые вопросы: - Установка и настройка библиотек для построения пайплайнов. - Настройка виртуального окружения. - Установка библиотек глубокого обучения. - Проверка работоспособности инструментов.
2	Тема 2. Знакомство с пайплайнами машинного обучения. Рассматриваемые вопросы: - Концепция пайплайнов в машинном обучении. - Компоненты пайплайна: предобработка, обучение, оценка. - Реализация простого пайплайна. - Визуализация и анализ результатов.
3	Тема 3. Построение пайплайна предобработки данных. Рассматриваемые вопросы: - Включение этапов предобработки в пайплайн. - Обработка пропущенных значений и выбросов. - Нормализация и кодирование признаков. - Интеграция предобработки с обучением.
4	Тема 4. Применение методов балансировки данных в пайплайне. Рассматриваемые вопросы:

№ п/п	Тематика практических занятий/краткое содержание
	<ul style="list-style-type: none"> - Включение методов балансировки в пайплайн. - Применение SMOTE и других методов ресемплинга. - Сравнение пайплайнов с балансировкой и без неё. - Оценка влияния балансировки на качество модели.
5	<p>Тема 5. Автоматическая настройка гиперпараметров в пайплайне.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Включение Grid Search в пайплайн. - Включение Random Search в пайплайн. - Сравнение эффективности методов. - Выбор оптимальных гиперпараметров.
6	<p>Тема 6. Байесовская оптимизация гиперпараметров.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Применение Bayesian Optimization. - Интеграция с пайплайном. - Сравнение с Grid Search и Random Search. - Анализ результатов оптимизации.
7	<p>Тема 7. Оркестрация пайплайнов с Scikit-learn Pipeline.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Создание сложных пайплайнов. - Использование FeatureUnion. - Управление зависимостями между этапами. - Оптимизация производительности пайплайна.
8	<p>Тема 8. Применение библиотек глубокого обучения в пайплайне.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Интеграция моделей глубокого обучения в пайплайн. - Использование Keras/TensorFlow в составе пайплайна. - Обучение нейронных сетей в пайплайне. - Оценка качества моделей.
9	<p>Тема 9. Оркестрация пайплайнов с PyTorch.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Интеграция PyTorch в пайплайн. - Построение и обучение модели PyTorch. - Включение в общий пайплайн. - Сравнение с Keras/TensorFlow.
10	<p>Тема 10. Оркестрация пайплайнов с использованием DVC.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Управление версиями данных и моделей. - Определение этапов пайплайна в DVC. - Воспроизводимость экспериментов. - Интеграция DVC с Git.
11	<p>Тема 11. Оркестрация пайплайнов с Apache Airflow.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Создание DAG для ML-пайплайна. - Определение задач и зависимостей. - Запуск и мониторинг выполнения. - Обработка ошибок в пайплайне.
12	<p>Тема 12. Оркестрация пайплайнов с MLflow.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Трекинг экспериментов в пайплайне. - Логирование параметров и метрик.

№ п/п	Тематика практических занятий/краткое содержание
	<ul style="list-style-type: none"> - Управление моделями в MLflow. - Интеграция с другими инструментами.
13	<p>Тема 13. Комплексный пайплайн: предобработка + обучение + оценка.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Сборка end-to-end пайплайна. - Включение всех этапов: предобработка, балансировка, настройка гиперпараметров, обучение, оценка. - Тестирование пайплайна. - Анализ результатов.
14	<p>Тема 14. Сравнительный анализ пайплайнов с различными моделями.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Построение пайплайнов для различных алгоритмов. - Сравнение качества моделей. - Анализ времени выполнения. - Выбор оптимального пайплайна.
15	<p>Тема 15. Оптимизация производительности пайплайнов.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Выявление узких мест в пайплайне. - Параллелизация этапов. - Использование кэширования. - Ускорение выполнения пайплайна.
16	<p>Тема 16. Комплексный проект.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Формулировка прикладной задачи на основе реального кейса. - Проектирование архитектуры пайплайна. - Реализация полного пайплайна с оркестрацией этапов. - Применение методов балансировки и настройки гиперпараметров. - Оценка качества и интерпретация результатов. - Защита проекта и презентация результатов.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение документации по библиотекам для построения пайплайнов.
2	Подготовка к практическим занятиям и выполнение домашних заданий.
3	Эксперименты с различными конфигурациями пайплайнов.
4	Работа с реальными наборами данных.
5	Изучение инструментов оркестрации.
6	Выполнение комплексного проекта.
7	Подготовка к промежуточной аттестации и текущему контролю.
8	Подготовка к промежуточной аттестации.
9	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	«Баланов, А. Н. DevOps: интеграция и автоматизация : учебное пособие для вузов / А. Н. Баланов. — 3-е изд., стер. — Санкт-Петербург : Лань, 2026. — 240 с. — ISBN 978-5-507-54640-4» (Баланов, А. Н. DevOps: интеграция и автоматизация : учебное пособие для вузов / А. Н. Баланов. — 3-е изд., стер. — Санкт-Петербург : Лань, 2026. — ISBN 978-5-507-54640-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/509963 (дата обращения: 28.09.2026). — Режим доступа: для авториз. пользователей. — С. 184.).	https://reader.lanbook.com/book/509963
2	«Колмогорова, С. С. Основы искусственного интеллекта : учебное пособие для студентов / С. С. Колмогорова. — Санкт-Петербург : СПбГЛТУ, 2022. — 108 с. — ISBN 978-5-9239-1308-8.» (Колмогорова, С. С. Основы искусственного интеллекта : учебное пособие для студентов / С. С. Колмогорова. — Санкт-Петербург : СПбГЛТУ, 2022. — ISBN 978-5-9239-1308-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/257804 (дата обращения: 28.09.2026). — Режим доступа: для авториз. пользователей. — С. 108.).	https://reader.lanbook.com/book/257804

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

Профессиональная база знаний. Документация библиотеки pandas.
<https://pandas.pydata.org/docs/>

Профессиональная база знаний. Документация библиотеки matplotlib.
<https://matplotlib.org/stable/users/index>

Профессиональная база знаний. Документация библиотеки seaborn.
<https://seaborn.pydata.org/tutorial.html>

Профессиональная база знаний. Документация библиотеки pandas_profiling. <https://pandas-profiling.github.io/pandas-profiling/docs/master/index.html#pandas-profiling>

Профессиональная база знаний. Документация scikit-learn https://scikit-learn.org/stable/user_guide.html

Профессиональная база знаний. Документация библиотеки lime <https://github.com/marcotcr/lime>

Электронно-библиотечная система Научно-технической библиотеки МИИТ - <http://library.miit.ru/>

Электронно-библиотечная система Лань - <https://e.lanbook.com/>

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

MS Office (Excel, Word)

Notepad++

Браузер Chrome

Anaconda

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

Компьютер преподавателя

Компьютеры студентов

Клавиатура

Мышка

Проектор

Экран для проектора

Маркерная доска

9. Форма промежуточной аттестации:

Зачет во 2 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

Директор ЦС IT - программ ВИШ

О.Б. Проневич

Председатель учебно-методической
комиссии

Д.В. Паринов