

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 22.05.2026 14:55:10

Уникальный программный ключ:

ca953a01204891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования

«Российский университет дружбы народов имени Патриса Лумумбы»

Факультет искусственного интеллекта

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

(наименование дисциплины/модуля)

Рекомендована МССН для направлений подготовки:

**02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ;**

09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗРАБОТКА И ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Методы машинного обучения» входит в программу бакалавриата «Искусственный интеллект: разработка и обучение интеллектуальных систем» по направлениям подготовки 02.03.02 Фундаментальная информатика и информационные технологии и 09.03.03 Прикладная информатика, и изучается в 5, 6, 7, 8 семестрах 3, 4 курсов. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 12 разделов и 144 тем и направлена на изучение теоретических основ и практических методов машинного обучения: математических моделей обучения по данным (PAC-learning, VC-размерность, bias-variance tradeoff), линейных моделей (регрессия, классификация), методов оптимизации (градиентный спуск и его вариации), ансамблевых методов (бэггинг, бустинг, стекинг), методов обучения без учителя (кластеризация, снижение размерности, обнаружение аномалий), вероятностных моделей (наивный Байес, EM-алгоритм, гауссовские смеси), методов работы с временными рядами и последовательными данными, основ обучения с подкреплением, а также промышленных практик разработки и эксплуатации ML-систем (MLOps, мониторинг, интерпретируемость, справедливость), применяемых на всех стадиях жизненного цикла ИИ-решений.

Целью освоения дисциплины является формирование у студентов системных знаний и практических навыков в области машинного обучения на уровне, достаточном для самостоятельного проектирования, реализации и эксплуатации ML-решений: от анализа осуществимости требований и подготовки данных до обучения моделей, оценки качества, развёртывания и мониторинга в промышленной среде, включая способность обосновывать выбор метода, оценивать компромиссы между метриками, обеспечивать воспроизводимость и справедливость, а также документировать весь процесс в соответствии с профессиональными стандартами.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Методы машинного обучения» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

| Шифр | Компетенция | Индикаторы достижения компетенции (в рамках данной дисциплины) |
|-------|--|--|
| УК-12 | Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных | УК-12.2 Способен проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных; |
| ОПК-1 | Способен применять фундаментальные знания, | ОПК-1.2 Умеет строить математические модели процессов и явлений, применять методы численного анализа и |

| Шифр | Компетенция | Индикаторы достижения компетенции (в рамках данной дисциплины) |
|-------------|---|--|
| | полученные в области математических и естественных наук, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности | оптимизации для решения задач машинного обучения и обработки данных; ОПК-1.3 Владеет навыками проведения вычислительных экспериментов, анализа их результатов и обоснования выбора математического аппарата для решения конкретных профессиональных задач в области ИИ; |
| ОПК-2 | Способен понимать принципы работы современных информационных технологий и применять компьютерные методы, современное программное обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности | ОПК-2.2 Умеет применять современное программное обеспечение (в том числе отечественного происхождения), фреймворки машинного обучения и инструменты обработки данных для решения задач в области ИИ; |
| ОПК-3 | Способен разрабатывать алгоритмические и программные решения в области системного и прикладного программирования, математических и информационных моделей, баз данных, средств тестирования, пригодные для практического применения | ОПК-3.3 Владеет навыками создания, тестирования и отладки алгоритмических и программных решений для систем ИИ, включая разработку пайплайнов обработки данных и обучения моделей; |
| ОПК-6 | Способен анализировать и разрабатывать организационно-технические процессы с применением методов системного анализа, математического моделирования и технологий искусственного интеллекта | ОПК-6.2 Умеет анализировать предметную область с позиции системного подхода, определять требования к ИИ-системе, формализовывать бизнес-задачи в задачи машинного обучения; |
| ОПК-7 | Способен решать задачи профессиональной деятельности на основе информационной культуры, применяя методы сбора, обработки, анализа и интерпретации данных с использованием информационно-коммуникационных технологий | ОПК-7.3 Владеет навыками подготовки данных для обучения моделей машинного обучения, оценки качества и репрезентативности обучающих выборок, документирования процессов работы с данными; |
| ПК-1 | Способен анализировать требования к программному обеспечению систем ИИ, разрабатывать технические спецификации и техническое задание на систему | ПК-1.1 Анализирует возможности реализации функциональных и нефункциональных требований к ПО систем ИИ, выявляет противоречия и ограничения; ПК-1.2 Разрабатывает технические спецификации на программные компоненты систем ИИ и описывает их взаимодействие; |
| ПК-3 | Способен разрабатывать и реализовывать стратегии тестирования и контроля качества программного обеспечения систем ИИ | ПК-3.1 Верифицирует требования к ПО систем ИИ, определяет требования к тестам и критерии приёмки; ПК-3.3 Оценивает результаты тестирования, реализует процесс контроля качества ПО систем ИИ; |
| ВД-1 | Способен осуществлять поиск, | ВД-1.2 Обосновывает способы и варианты применения |

| Шифр | Компетенция | Индикаторы достижения компетенции (в рамках данной дисциплины) |
|-------------|--|---|
| | сбор, очистку и предварительный анализ данных | методов предварительного анализа данных в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи; BD-1.3 Применяет методы анализа данных для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ; BD-1.4 Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных; BD-1.5 Отбирает признаки данных, значимые для исследования; |
| BD-2 | Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных | BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения; |
| FC-1 | Способен проводить передовые исследования в области архитектур, алгоритмов МО, оптимизации и математики | FC-1.1 Разрабатывает фундаментальные основы и новые алгоритмы машинного обучения; |
| FC-5 | Способен проводить передовые исследования в области безопасности, доверия и объяснимости | FC-5.2 Обеспечивает объяснения причин принятия тех или иных решений в результатах работы искусственного интеллекта; |
| LC-1 | Способен проводить анализ бизнес-проблем с оценкой перспективности применения ИИ для их решения, осуществлять постановку задачи, формулировать требования к системе ИИ | LC-1.2 Выбирает оптимальные технологии под конкретные требования проекта внедрения ИИ; |
| LC-2 | Способен проводить эксперименты на данных, формулировать гипотезы исследования, строить (обучать, дообучать) модели ИИ с оценкой их качества и анализом ошибок, обеспечивать воспроизводимость и масштабируемость исследований на данных | LC-2.1 Проводит эксперименты с моделями ИИ, оценивает их качество (точность, производительность); LC-2.2 Проводит эксперименты на данных и визуализирует результаты с применением технологий анализа данных (статистического анализа), методов и алгоритмов МО; |
| MF-1 | Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ | MF-1.1 Обосновывает способы и варианты применения методов и моделей в задачах искусственного интеллекта, включая их модификацию и адаптацию к специфике задачи; MF-1.3 Применяет аппарат теории вероятностей для исследования методов и моделей машинного обучения; |
| MF-3 | Способен применять современные методы оптимизации для обучения моделей машинного обучения, настройки гиперпараметров и решения задач искусственного интеллекта | MF-3.2 Применяет методы оптимизации для настройки гиперпараметров моделей машинного обучения, включая использование методов поиска (поиск по сетке, случайный поиск) и байесовской оптимизации; |
| MF-4 | Способен применять статистические методы для анализа данных, валидации | MF-4.1 Применяет статистические методы анализа и машинного обучения для решения задач анализа данных и проведения экспериментов на данных; |

| Шифр | Компетенция | Индикаторы достижения компетенции (в рамках данной дисциплины) |
|------|---|--|
| | моделей машинного обучения и проведения экспериментов в области ИИ | MF-4.2 Способен применять статистические методы для построения предсказательных моделей, включая методы для анализа и прогнозирования временных рядов, а также моделирования нестационарных случайных процессов; MF-4.3 Способен применять статистические методы для оценки качества моделей ИИ, включая метрики и критерии для регрессии, классификации и кластеризации, а также для проведения статистических тестов для сравнения моделей; |
| ML-2 | Способен применять фундаментальные принципы и методы машинного обучения, включая подготовку данных, оценку качества моделей и работу с признаками | ML-2.1 Различает основные типы задач МО и применяет на практике принципы их решения; ML-2.2 Применяет методы предварительной обработки данных и работы с признаками; ML-2.3 Решает проблемы несбалансированных данных и оценивает качество моделей; |
| ML-3 | Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения | ML-3.1 Обосновывает способы и варианты применения классических методов и моделей МО в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи; ML-3.2 Эффективно применяет классические методы и модели МО для обеспечения достижимости функциональных характеристик систем ИИ; ML-3.3 Оценивает результативность применения классических методов и моделей МО в задачах ИИ на основе сопоставления с аналогами; |
| ML-4 | Способен применять методы обучения без учителя для анализа данных и выявления скрытых закономерностей | ML-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач; ML-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил; ML-4.3 Оценивает качество результатов обучения без учителя; |
| ML-6 | Способен применять алгоритмы обучения с подкреплением | ML-6.1 Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи; |
| PL-1 | Способен применять язык программирования Python для решения задач в области ИИ | PL-1.2 Осуществляет выбор инструментов разработки на Python, приемлемых для создания прикладной системы обработки научных данных, машинного обучения и визуализации с заданными требованиями; |
| SS-1 | Способен учитывать философские, когнитивные и социальные основания концепций ИИ в профессиональной деятельности | SS-1.1 Учитывает в разработке и эксплуатации систем ИИ философские основания концепций интеллекта, языка, знания, агентности; SS-1.2 Применяет методики работы с этическими и социальными рисками, возникающими на разных стадиях жизненного цикла ИИ; |
| SS-3 | Способен к критическому анализу, метарефлексии и переносу знаний при работе с системами ИИ | SS-3.1 Учитывает в работе когнитивные искажения человека и примеры их проявления при работе с данными и ИИ, выявляет предвзятости систем ИИ, аргументированно оценивает надежность данных и выдачи ИИ, применяет базовые принципы критического мышления (оценка источников, проверка аргументов, отличие факта от интерпретации); SS-3.2 Определяет релевантность применения ИИ для решения конкретных задач, анализирует поведение ИИ в техническом, социальном и правовом контекстах, переносит идеи и методы за пределы исходной предметной области; |

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Методы машинного обучения» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Методы машинного обучения».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

| Шифр | Наименование компетенции | Предшествующие дисциплины/модули, практики* | Последующие дисциплины/модули, практики* |
|-------|--|---|--|
| УК-12 | Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных | Программирование на языке Python; Введение в искусственный интеллект; Методы разработки решений на основе искусственного интеллекта (Git, Docker); Введение в базы данных; Статистические методы и первичный анализ данных; Технологическая (проектно-технологическая) практика (учебная); Эксплуатационная практика (учебная); | |
| ОПК-6 | Способен анализировать и разрабатывать организационно-технические процессы с применением методов системного анализа, математического моделирования и технологий искусственного интеллекта | Введение в искусственный интеллект; Искусственный интеллект и когнитивная психология; Этика и безопасность использования искусственного интеллекта; | |
| ОПК-1 | Способен применять фундаментальные знания, полученные в области математических и естественных наук, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности | Линейная алгебра; Дискретная математика; Математический анализ; Теория вероятностей и математическая статистика; Дифференциальные уравнения; Численная линейная алгебра; Статистические методы и первичный анализ данных; | |
| ОПК-2 | Способен понимать принципы работы современных информационных технологий и применять компьютерные методы, современное программное | История и теория программирования; Введение в искусственный интеллект; Программирование на языке Python; | |

| Шифр | Наименование компетенции | Предшествующие дисциплины/модули, практики* | Последующие дисциплины/модули, практики* |
|-------|---|--|--|
| | обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности | | |
| ОПК-3 | Способен разрабатывать алгоритмические и программные решения в области системного и прикладного программирования, математических и информационных моделей, баз данных, средств тестирования, пригодные для практического применения | Дискретная математика; История и теория программирования; Программирование на языке C++; Алгоритмы и структуры данных; Введение в базы данных; Программирование на языке Python; Методы разработки решений на основе искусственного интеллекта (Git, Docker); Технологическая (проектно-технологическая) практика (учебная); | |
| ОПК-7 | Способен решать задачи профессиональной деятельности на основе информационной культуры, применяя методы сбора, обработки, анализа и интерпретации данных с использованием информационно-коммуникационных технологий | Технологическая (проектно-технологическая) практика (учебная); Теория вероятностей и математическая статистика; Статистические методы и первичный анализ данных; Введение в базы данных; | |
| ПК-1 | Способен анализировать требования к программному обеспечению систем ИИ, разрабатывать технические спецификации и техническое задание на систему | Эксплуатационная практика (учебная); Технологическая (проектно-технологическая) практика (учебная); Правоведение; Введение в искусственный интеллект; Искусственный интеллект и когнитивная психология; Этика и безопасность использования искусственного интеллекта; История и теория программирования; Программирование на языке C++; Методы разработки решений на основе искусственного интеллекта (Git, Docker); Введение в базы данных; | |
| ПК-3 | Способен разрабатывать и реализовывать стратегии тестирования и контроля качества программного обеспечения систем ИИ | Технологическая (проектно-технологическая) практика (учебная); Эксплуатационная практика (учебная); Теория вероятностей и математическая статистика; Этика и безопасность использования искусственного | |

| Шифр | Наименование компетенции | Предшествующие дисциплины/модули, практики* | Последующие дисциплины/модули, практики* |
|------|--|---|--|
| | | интеллекта; Статистические методы и первичный анализ данных; Программирование на языке Python; Методы разработки решений на основе искусственного интеллекта (Git, Docker); | |
| SS-1 | Способен учитывать философские, когнитивные и социальные основания концепций ИИ в профессиональной деятельности | История и теория программирования; Введение в искусственный интеллект; Искусственный интеллект и когнитивная психология; Лингвистические основы анализа естественного языка; Этика и безопасность использования искусственного интеллекта; | |
| SS-3 | Способен к критическому анализу, метарефлексии и переносу знаний при работе с системами ИИ | Эксплуатационная практика (учебная); Теория вероятностей и математическая статистика; Искусственный интеллект и когнитивная психология; Этика и безопасность использования искусственного интеллекта; Статистические методы и первичный анализ данных; Правоведение; Введение в искусственный интеллект; Введение в компьютерное зрение; | |
| MF-1 | Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ | Линейная алгебра; Математический анализ; Теория вероятностей и математическая статистика; Статистические методы и первичный анализ данных; Эксплуатационная практика (учебная); | |
| MF-3 | Способен применять современные методы оптимизации для обучения моделей машинного обучения, настройки гиперпараметров и решения задач искусственного интеллекта | Эксплуатационная практика (учебная); Численная линейная алгебра; | |
| MF-4 | Способен применять статистические методы для анализа данных, валидации моделей машинного обучения и проведения экспериментов в области ИИ | Теория вероятностей и математическая статистика; Статистические методы и первичный анализ данных; Дифференциальные уравнения; | |
| BD-1 | Способен осуществлять | Эксплуатационная практика | |

| Шифр | Наименование компетенции | Предшествующие дисциплины/модули, практики* | Последующие дисциплины/модули, практики* |
|------|---|---|--|
| | поиск, сбор, очистку и предварительный анализ данных | (учебная); Статистические методы и первичный анализ данных; | |
| BD-2 | Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных | Статистические методы и первичный анализ данных; Введение в базы данных; Методы разработки решений на основе искусственного интеллекта (Git, Docker); Технологическая (проектно-технологическая) практика (учебная); | |
| ML-2 | Способен применять фундаментальные принципы и методы машинного обучения, включая подготовку данных, оценку качества моделей и работу с признаками | Введение в искусственный интеллект; Статистические методы и первичный анализ данных; | |
| ML-3 | Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения | Эксплуатационная практика (учебная); | |
| ML-4 | Способен применять методы обучения без учителя для анализа данных и выявления скрытых закономерностей | | |
| ML-6 | Способен применять алгоритмы обучения с подкреплением | | |
| PL-1 | Способен применять язык программирования Python для решения задач в области ИИ | Технологическая (проектно-технологическая) практика (учебная); Программирование на языке Python; Алгоритмы и структуры данных; Статистические методы и первичный анализ данных; | |
| LC-1 | Способен проводить анализ бизнес-проблем с оценкой перспективности применения ИИ для их решения, осуществлять постановку задачи, формулировать требования к системе ИИ | Введение в искусственный интеллект; Технологическая (проектно-технологическая) практика (учебная); | |
| LC-2 | Способен проводить эксперименты на данных, формулировать гипотезы исследования, строить (обучать, дообучать) модели ИИ с оценкой их качества и анализом ошибок, обеспечивать | Эксплуатационная практика (учебная); Статистические методы и первичный анализ данных; | |

| Шифр | Наименование компетенции | Предшествующие дисциплины/модули, практики* | Последующие дисциплины/модули, практики* |
|-------------|---|---|---|
| | воспроизводимость и масштабируемость исследований на данных | | |
| FC-1 | Способен проводить передовые исследования в области архитектур, алгоритмов МО, оптимизации и математики | Эксплуатационная практика (учебная); Линейная алгебра; Математический анализ; Теория вероятностей и математическая статистика; Введение в искусственный интеллект; Численная линейная алгебра; | |
| FC-5 | Способен проводить передовые исследования в области безопасности, доверия и объяснимости | Эксплуатационная практика (учебная); Правоведение; Этика и безопасность использования искусственного интеллекта; | |

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Методы машинного обучения» составляет «15» зачетных единиц.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

| Вид учебной работы | ВСЕГО, ак.ч. | | Семестр(-ы) | | | |
|--|----------------|------------|-------------|------------|------------|------------|
| | | | 5 | 6 | 7 | 8 |
| <i>Контактная работа, ак.ч.</i> | 291 | | 85 | 85 | 65 | 56 |
| Лекции (ЛК) | 61 | | 17 | 17 | 13 | 14 |
| Лабораторные работы (ЛР) | 0 | | 0 | 0 | 0 | 0 |
| Практические/семинарские занятия (СЗ) | 230 | | 68 | 68 | 52 | 42 |
| <i>Самостоятельная работа обучающихся, ак.ч.</i> | 168 | | 32 | 68 | 16 | 52 |
| <i>Контроль (экзамен/зачет с оценкой), ак.ч.</i> | 81 | | 27 | 27 | 27 | 0 |
| Общая трудоемкость дисциплины | ак.ч. | 540 | 144 | 180 | 108 | 108 |
| | зач.ед. | 15 | 4 | 5 | 3 | 3 |

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--|-------------------|--|---|----------------------|-------------------------|
| Раздел 1 | Введение в машинное обучение и линейные модели | 1.1 | Введение в машинное обучение. Типы задач и постановка | Определение ML. Типы задач: обучение с учителем (регрессия, классификация), без учителя (кластеризация, снижение размерности), с подкреплением. Формализация: пространство признаков, целевая переменная, гипотеза, функция потерь. Этапы ML-проекта: постановка задачи → данные → модель → оценка → развёртывание. Анализ осуществимости: когда ML подходит, а когда нет | ЛК | MF-1.1, ML-2.1, ПК-1.1 |
| | | 1.2 | Теоретические основы обучения: bias-variance, обобщающая способность | Bias-variance tradeoff: декомпозиция ошибки. Переобучение (overfitting) и недообучение (underfitting). Тренировочная и тестовая ошибки. Обобщающая способность. Регуляризация как способ борьбы с переобучением. PAC-learning и VC-размерность (обзор). Проклятие размерности. Связь с объёмом данных и сложностью модели | ЛК | MF-1.1, MF-1.3, FC-1.1 |
| | | 1.3 | Линейная регрессия и градиентный спуск | Линейная регрессия: модель, функция потерь MSE, нормальные уравнения. Градиентный спуск: идея, learning rate, сходимость. Стохастический GD и мини-батч GD. Регуляризация: Ridge (L2), Lasso (L1), Elastic Net. Полиномиальная регрессия. MLE-интерпретация линейной регрессии | ЛК | MF-1.3, MF-3.2, ОПК-1.2 |
| | | 1.4 | Практикум: EDA и подготовка данных для ML | Загрузка реального датасета (Boston Housing / California Housing). EDA: описательная статистика, визуализация распределений, корреляция. Обработка пропусков и выбросов. Масштабирование: StandardScaler, MinMaxScaler. Train/test split. Фиксация random seed для воспроизводимости | СЗ | BD-1.2, BD-1.3, LC-2.2 |
| | | 1.5 | Практикум: линейная регрессия — реализация с нуля | Реализация линейной регрессии на NumPy: нормальные уравнения, градиентный спуск. Визуализация: ландшафт функции потерь, траектория GD. Влияние learning rate: слишком большой, слишком маленький. Сравнение с sklearn.linear_model.LinearRegression | СЗ | MF-3.2, ОПК-1.3 |
| | | 1.6 | Практикум: регуляризация и полиномиальные признаки | Ridge и Lasso на sklearn. Влияние параметра α : кривые обучения. Lasso для отбора признаков: анализ нулевых | СЗ | MF-1.3, BD-1.4, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|--|---|----------------------|-------------------------|
| | | | коэффициентов. Полиномиальные признаки (PolynomialFeatures): от недообучения к переобучению. Визуализация bias-variance tradeoff на примере полиномиальной регрессии | | MF-4.1 |
| | | 1.7 Практикум: метрики регрессии и валидация | Метрики: MSE, RMSE, MAE, R^2 , MAPE. Кросс-валидация: k-fold, Leave-One-Out. Кривые обучения (learning curves): диагностика переобучения/недообучения. Выбор гиперпараметров: GridSearchCV, RandomizedSearchCV. Практика: выбор лучшей модели для датасета с обоснованием | СЗ | ML-2.2, ПК-3.1, ОПК-1.3 |
| | | 1.8 Практикум: логистическая регрессия — теория и реализация | Задача бинарной классификации. Логистическая функция (сигмоида). Функция потерь: бинарная кросс-энтропия. Вывод из MLE. Градиентный спуск для логистической регрессии. Реализация на NumPy. Разделяющая гиперплоскость. Многоклассовая: One-vs-Rest, Softmax | СЗ | MF-1.3, MF-3.2, ML-2.1 |
| | | 1.9 Практикум: метрики классификации | Матрица ошибок (confusion matrix). Accuracy, Precision, Recall, F1-score. Macro/Micro/Weighted averaging. ROC-кривая и AUC. Precision-Recall кривая. Порог принятия решения: выбор порога для бизнес-задачи. Практика: оценка модели на несбалансированном датасете | СЗ | ML-2.2, SS-3.1, ПК-3.1 |
| | | 1.10 Практикум: работа с несбалансированными классами | Проблема дисбаланса: влияние на метрики. Стратегии: oversampling (SMOTE), undersampling, class weights. Стратифицированная кросс-валидация. Выбор метрики: F1 vs. AUC-PR. Практика: сравнение стратегий на датасете с дисбалансом 1:50 (fraud detection) | СЗ | BD-1.5, ML-2.2, SS-3.1 |
| | | 1.11 Практикум: SVM — метод опорных векторов | Линейный SVM: максимальный зазор, опорные векторы. Мягкий зазор (soft margin): параметр C. Ядровый трюк: полиномиальное, RBF. Параметр γ в RBF. Практика: визуализация разделяющей границы для разных ядер. Подбор C и γ . Сравнение SVM с логистической регрессией | СЗ | ML-2.1, MF-1.3 |
| | | 1.12 Практикум: метод k ближайших соседей (k-NN) | k-NN: принцип, метрики расстояния (евклидова, манхэттенская, Минковского). Выбор k: bias-variance. Взвешенный k-NN. Проклятие размерности для k-NN. KD-tree и Ball-tree для ускорения. Практика: классификация и регрессия с k-NN, визуализация границ решений | СЗ | ML-2.1, MF-4.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--------------------------------------|-------------------|---|---|----------------------|-------------------------|
| | | 1.13 | Практикум: pipelines в scikit-learn | sklearn.pipeline.Pipeline: цепочка трансформаций и модели. ColumnTransformer для разных типов признаков. Интеграция предобработки в кросс-валидацию (предотвращение утечки данных). Сохранение пайплайна (joblib). Практика: создание полного пайплайна для реального датасета | СЗ | ML-3.1, ОПК-3.3, ПК-1.2 |
| | | 1.14 | Практикум: воспроизводимость и документирование экспериментов | Фиксация random seed. Версионирование данных и кода. Логирование экспериментов: MLflow Tracking (обзор). Структура отчёта по эксперименту: задача, данные, метод, метрики, выводы. Model Card (шаблон). Практика: проведение эксперимента с логированием в MLflow | СЗ | LC-2.2, LC-1.2, FC-5.2 |
| | | 1.15 | Практикум: мини-проект — сравнение линейных моделей | Сквозная задача: реальный датасет → EDA → предобработка → обучение нескольких моделей (LinearRegression, Ridge, Lasso, LogisticRegression, SVM, k-NN) → кросс-валидация → сравнение → выбор → документирование. Обоснование выбора модели | СЗ | ML-2.2, ПК-3.3, LC-2.1 |
| Раздел 2 | Деревья решений и ансамблевые методы | 2.1 | Деревья решений: принцип и алгоритмы | Дерево решений: структура (узлы, рёбра, листья). Критерии разбиения: Gini impurity, информационный выигрыш (энтропия), MSE для регрессии. Алгоритмы: CART, ID3, C4.5. Регуляризация деревьев: max_depth, min_samples_split, min_samples_leaf, max_features. Pruning (обзор). Интерпретируемость деревьев: feature importance | ЛК | ML-2.1, MF-1.3, ОПК-6.2 |
| | | 2.2 | Ансамблевые методы: бэггинг и случайный лес | Идея ансамблирования: «мудрость толпы». Бэггинг (Bootstrap Aggregating): бутстрэп-выборки, агрегация предсказаний. Случайный лес (Random Forest): бэггинг + случайный подвыбор признаков. Out-of-Bag оценка. Feature importance: impurity-based и permutation importance. Параметры: n_estimators, max_features, max_depth | ЛК | ML-2.1, ML-2.3, MF-4.1 |
| | | 2.3 | Ансамблевые методы: бустинг | Бустинг: идея последовательного обучения «слабых» моделей. AdaBoost: перевзвешивание примеров. Градиентный бустинг (GBM): аппроксимация антиградиента. XGBoost: регуляризация, гистограммный метод, параллелизация. LightGBM: leaf-wise growth, эффективность. CatBoost: обработка категорий, ordered boosting. Сравнение библиотек | ЛК | ML-2.1, ML-2.3, ОПК-2.2 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы | |
|---------------|---------------------------------|-------------------|--|--|------------------------|------------------------|
| | | 2.4 | Практикум: деревья решений — обучение и визуализация | Обучение DecisionTreeClassifier/Regressor. Визуализация дерева (plot_tree, export_graphviz). Влияние глубины на переобучение. Feature importance. Практика: обучение дерева для задачи классификации, анализ правил принятия решений | СЗ | ML-2.1, ОПК-6.2 |
| | | 2.5 | Практикум: Random Forest — обучение и настройка | Обучение RandomForestClassifier. Влияние n_estimators: кривая OOB-error. Влияние max_features. Feature importance: impurity vs. permutation. Практика: сравнение дерева и леса по метрикам и устойчивости. Partial dependence plots (PDP) | СЗ | ML-2.1, ML-2.3 |
| | | 2.6 | Практикум: XGBoost — обучение и тонкая настройка | Установка XGBoost. Основные параметры: n_estimators, max_depth, learning_rate, subsample, colsample_bytree, reg_alpha, reg_lambda. Early stopping: eval_set, early_stopping_rounds. Практика: подбор гиперпараметров (Optuna / RandomizedSearchCV). Кривые обучения | СЗ | ML-2.3, MF-3.2 |
| | | 2.7 | Практикум: LightGBM и CatBoost | LightGBM: leaf-wise vs. depth-wise, num_leaves, обработка категориальных признаков. CatBoost: встроенная обработка категорий, ordered boosting, Pool. Практика: сравнение XGBoost, LightGBM, CatBoost на одном датасете по метрикам и времени обучения | СЗ | ML-2.3, ОПК-2.2 |
| | | 2.8 | Практикум: стекинг и блендинг | Стекинг (Stacking): мета-модель поверх базовых. Реализация: StackingClassifier в sklearn. Блендинг: holdout set для метапризнаков. Многоуровневый стекинг. Практика: построение стекинга из RF + XGBoost + LightGBM, сравнение с отдельными моделями | СЗ | ML-2.3, ML-4.1 |
| | | 2.9 | Практикум: подбор гиперпараметров — Optuna | Подходы: GridSearch, RandomSearch, байесовская оптимизация. Optuna: study, trial, suggest_float/int/categorical. Pruning неперспективных испытаний. Визуализация: importance plot, optimization history. Практика: оптимизация гиперпараметров GBM на реальном датасете | СЗ | ML-3.2, MF-3.2 |
| | | 2.10 | Практикум: feature engineering для табличных данных | Создание признаков: взаимодействия, агрегаты, полиномиальные. Target encoding для категорий. Работа с датами: выделение компонентов, циклические признаки. Feature selection: mutual information, permutation importance, Boruta. Практика: feature engineering для соревновательного датасета | СЗ | BD-1.4, ML-3.1, MF-4.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--|-------------------|---|--|----------------------|-------------------------|
| | | 2.11 | Практикум: интерпретируемость ансамблей — SHAP | SHAP (SHapley Additive exPlanations): теоретическая основа (значения Шепли). TreeSHAP для деревьев. SHAP summary plot, dependence plot, waterfall plot, force plot. Глобальная и локальная интерпретируемость. Практика: анализ предсказаний модели GBM с помощью SHAP | СЗ | ОПК-6.2, SS-3.2, ML-6.1 |
| | | 2.12 | Практикум: анализ ошибок модели | Error analysis: на каких примерах модель ошибается? Построение confusion matrix по подгруппам. Визуализация ошибочных примеров. Residual analysis для регрессии. Связь ошибок с качеством данных. Формулирование рекомендаций по улучшению | СЗ | ПК-3.3, ML-2.2, SS-3.1 |
| | | 2.13 | Практикум: формулирование критериев приёмки ML-модели | Трансляция бизнес-требований в метрики ML. Пороги метрик. Тестирование на edge cases. Регрессионные тесты модели. Составление плана тестирования: unit-тесты пайплайна, тесты данных, тесты метрик. Документирование: техническая спецификация ML-компонента | СЗ | ПК-3.1, ПК-1.2, LC-2.1 |
| | | 2.14 | Практикум: Kaggle-соревнование по табличным данным | Участие в мини-соревновании: реальный датасет, фиксированная метрика, лидерборд. Полный цикл: EDA → feature engineering → модели → ансамбль → подбор гиперпараметров → submission. Ретроспектива: что сработало, что нет. Обсуждение: соревнования vs. промышленная разработка | СЗ | ML-4.2, ML-4.3, ОПК-7.3 |
| Раздел 3 | Обучение без учителя: кластеризация и снижение размерности | 3.1 | Кластеризация: методы и метрики | Задача кластеризации. K-Means: алгоритм, инициализация (k-means++), выбор k (метод локтя, силуэтный анализ). Иерархическая кластеризация: агломеративная, дендрограмма, linkage. DBSCAN: плотностная кластеризация, параметры eps и min_samples. Метрики: Silhouette, Calinski-Harabasz, Davies-Bouldin; внешние: ARI, NMI | ЛК | ML-2.1, MF-4.1 |
| | | 3.2 | Снижение размерности: PCA, t-SNE, UMAP | PCA: ковариационная матрица, собственные значения, доля объяснённой дисперсии, выбор числа компонент. Связь с SVD. t-SNE: принцип, перплексия, ограничения (нетранзитивность расстояний, непригодность для новых точек). UMAP: преимущества над t-SNE (скорость, глобальная структура). Применение: визуализация, | ЛК | MF-4.1, PL-1.2, ML-2.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|--|---|----------------------|------------------------|
| | | | предобработка, обнаружение аномалий | | |
| | | 3.3 Практикум: K-Means — реализация и визуализация | Реализация K-Means на NumPy. sklearn KMeans. Метод локтя (inertia plot). Силуэтный анализ (silhouette_samples). Визуализация кластеров в 2D. Mini-batch K-Means для больших данных. Практика: сегментация клиентов по покупательскому поведению | СЗ | ML-2.1, MF-4.1 |
| | | 3.4 Практикум: иерархическая кластеризация | Агломеративная кластеризация: single, complete, average, Ward linkage. Построение и анализ дендрограммы (scipy.cluster.hierarchy). Выбор числа кластеров по дендрограмме. Сравнение с K-Means. Практика: кластеризация текстов по TF-IDF с визуализацией дендрограммы | СЗ | ML-2.1, MF-4.1 |
| | | 3.5 Практикум: DBSCAN и плотностная кластеризация | DBSCAN: core points, border points, noise. Выбор eps (k-distance plot) и min_samples. Преимущества: произвольная форма кластеров, обнаружение выбросов. HDBSCAN (обзор). Практика: кластеризация геоданных (GPS-точки), обнаружение аномалий | СЗ | ML-2.1, BD-1.5 |
| | | 3.6 Практикум: PCA — анализ и визуализация | PCA в sklearn: fit_transform, explained_variance_ratio_. Визуализация в пространстве первых двух компонент. Biplot: проекция признаков. Выбор числа компонент: scree plot, порог дисперсии. PCA как предобработка для классификации. Практика: PCA для датасета с большим числом признаков | СЗ | MF-4.1, PL-1.2 |
| | | 3.7 Практикум: t-SNE и UMAP для визуализации | t-SNE: влияние perplexity, n_iter. Визуализация MNIST, Fashion-MNIST. UMAP: параметры n_neighbors, min_dist. Сравнение: PCA vs. t-SNE vs. UMAP по сохранению структуры. Практика: визуализация эмбедингов и кластеров высокоразмерных данных | СЗ | PL-1.2, ML-2.1 |
| | | 3.8 Практикум: обнаружение аномалий | Подходы: статистические (z-score, IQR), модельные (Isolation Forest, Local Outlier Factor, One-Class SVM). Isolation Forest: принцип, параметры. LOF: плотностный подход. Метрики: precision/recall при наличии меток, визуальная оценка. Практика: обнаружение мошенничества в транзакциях | СЗ | ML-2.1, BD-1.5, MF-4.1 |
| | | 3.9 Практикум: гауссовские смеси и EM- | GMM: модель, параметры (средние, ковариации, веса). EM- | СЗ | MF-1.3, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--|-------------------|--|--|----------------------|------------------------|
| | | | алгоритм | алгоритм: E-шаг, M-шаг, сходимость. Выбор числа компонент: BIC, AIC. Мягкая кластеризация: вероятности принадлежности. Сравнение с K-Means. Практика: кластеризация с GMM, визуализация вероятностных границ | | ML-2.1 |
| | | 3.10 | Практикум: кластеризация текстов | TF-IDF для векторизации текстов. Кластеризация: K-Means на TF-IDF. Визуализация: PCA/UMAP проекция кластеров. Определение тем кластеров по ключевым словам. Сравнение с LDA (Latent Dirichlet Allocation, обзор). Практика: тематическая кластеризация новостей | СЗ | ML-2.1, MF-4.2 |
| | | 3.11 | Практикум: оценка и сравнение методов кластеризации | Сравнение K-Means, иерархического, DBSCAN, GMM на нескольких датасетах с разной структурой (шары, полумесяцы, кольца, зашумлённые). Метрики: Silhouette, ARI (при наличии меток). Визуализация результатов. Критерии выбора метода в зависимости от задачи | СЗ | ML-2.2, SS-3.1 |
| | | 3.12 | Практикум: применение снижения размерности в ML-пайплайне | PCA как этап предобработки: PCA + логистическая регрессия. Выбор числа компонент через кросс-валидацию. Ускорение обучения при высокой размерности. Практика: Pipeline(PCA → Classifier), сравнение с обучением без PCA | СЗ | ML-3.1, MF-4.1 |
| | | 3.13 | Практикум: итоговый проект семестра 5 — полный ML-пайплайн | Сквозная задача: реальный датасет → EDA → предобработка (Pipeline) → обучение нескольких моделей (линейные, деревья, ансамбли, кластеризация) → оценка → интерпретация (SHAP) → документирование (Model Card, отчёт). Презентация с обоснованием решений | СЗ | ML-4.1, ПК-3.3, LC-2.1 |
| Раздел 4 | Продвинутые методы обучения с учителем | 4.1 | Наивный байесовский классификатор и вероятностные модели | Наивный Байес: предположение условной независимости. Гауссов, мультиномиальный, Бернулли NB. Оценка параметров MLE, сглаживание Лапласа. Применения: спам-фильтрация, классификация текстов. Связь с генеративными и дискриминативными моделями | ЛК | MF-1.3, ML-2.1, FC-1.1 |
| | | 4.2 | Многоклассовая и многометочная классификация | Стратегии многоклассовой классификации: One-vs-Rest, One-vs-One, Softmax. Многометочная (multi-label): Binary Relevance, Classifier Chains. Иерархическая классификация (обзор). Метрики: multi-class accuracy, macro/micro F1, Hamming loss, subset accuracy. Калибровка вероятностей: Platt scaling, isotonic regression | ЛК | ML-2.1, ML-2.2, SS-3.1 |
| | | 4.3 | Методы отбора и генерации признаков | Фильтрующие методы: корреляция, mutual information, chi- | ЛК | BD-1.4, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|--|---|----------------------|------------------------|
| | | | squared, ANOVA F-test. Обёрточные: forward selection, backward elimination, recursive feature elimination (RFE). Встроенные: L1, feature importance. Генерация: PolynomialFeatures, target encoding, aggregation features. Feature store: концепция хранения признаков. AutoML feature generation (обзор: Featuretools) | | ML-3.1, BD-2.1 |
| | | 4.4 Практикум: наивный Байес для текстовой классификации | Загрузка текстового датасета (20 Newsgroups). Векторизация: CountVectorizer, TfidfVectorizer. Обучение MultinomialNB. Оценка: accuracy, F1, confusion matrix. Визуализация наиболее информативных слов. Сравнение с логистической регрессией | СЗ | ML-2.1, MF-1.3 |
| | | 4.5 Практикум: калибровка вероятностей | Проверка калибровки: calibration_curve. Reliability diagram. Методы: CalibratedClassifierCV (Platt, isotonic). Brier score. Практика: калибровка GBM, сравнение до/после. Обсуждение: когда нужна калибровка (медицина, финансы, скоринг) | СЗ | ML-2.2, SS-3.2 |
| | | 4.6 Практикум: multi-label классификация | Преобразование задачи: MultiOutputClassifier, ClassifierChain. Метрики: Hamming loss, subset accuracy, macro F1. Практика: multi-label классификация на датасете тегов (фильмы, статьи). Сравнение стратегий | СЗ | ML-2.1, ML-2.2 |
| | | 4.7 Практикум: recursive feature elimination и отбор признаков | RFE (sklearn): пошаговый отбор. RFECV: RFE с кросс-валидацией. Сравнение с mutual_info, permutation importance. SelectKBest, SelectFromModel. Практика: отбор признаков для датасета с > 100 признаками, оценка влияния на качество модели | СЗ | BD-1.4, ML-3.1 |
| | | 4.8 Практикум: автоматическая генерация признаков (Featuretools) | Featuretools: entity sets, relationships, deep feature synthesis. Примитивы: aggregation (mean, count, max), transform (month, year, haversine). Практика: автоматическая генерация признаков для реляционного датасета. Отбор полезных из сгенерированных. Сравнение с ручным feature engineering | СЗ | BD-1.4, ML-3.1, BD-2.1 |
| | | 4.9 Практикум: метрики для различных бизнес-задач | Выбор метрики для конкретной задачи: кредитный скоринг (precision@threshold), медицинская диагностика (recall), рекомендации (NDCG@K, MAP@K). Пользовательские функции потерь. Cost-sensitive learning. Практика: анализ | СЗ | ML-2.2, ПК-1.1, SS-3.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|-------------------|--|----------------------|------------------------|
| | | | трёх задач, обоснование выбора метрики, настройка порога | | |
| | | 4.10 | Практикум: обработка категориальных признаков высокой кардинальности | СЗ | BD-1.3, ML-3.1 |
| | | 4.11 | Практикум: работа с пропусками — продвинутые методы | СЗ | BD-1.3, ML-3.1 |
| | | 4.12 | Практикум: утечка данных (data leakage) — обнаружение и предотвращение | СЗ | ML-4.3, BD-1.2, SS-3.1 |
| | | 4.13 | Практикум: cross-validation — продвинутые стратегии | СЗ | ML-4.3, ОПК-1.3 |
| | | 4.14 | Практикум: спецификация ML-компонента | СЗ | ПК-1.2, LC-2.1, ML-3.1 |
| | | 4.15 | Практикум: мини-проект — ML-решение для бизнес-задачи | СЗ | ML-4.1, ПК-1.1, ПК-3.3 |
| Раздел 5 | Обучение без учителя: | 5.1 | Матричные разложения и тематические | ЛК | MF-4.2, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--|-------------------|--|--|----------------------|------------------------|
| | продвинутое методы и тематические модели | | модели | Применение: тематическое моделирование, рекомендации. LDA (Latent Dirichlet Allocation): генеративная модель тем, гиббсовский сэмплинг (обзор). Сравнение NMF и LDA. Embedding-based методы: BERTopic (обзор). Когерентность тем (coherence score) | | ML-2.1 |
| | | 5.2 | Рекомендательные системы | Коллаборативная фильтрация: user-based, item-based. Матричная факторизация: SVD, ALS. Контентная фильтрация: TF-IDF + cosine similarity. Гибридные подходы. Implicit feedback: BPR (обзор). Метрики: Precision@K, Recall@K, NDCG@K, MAP@K. Проблема холодного старта | ЛК | ML-2.1, MF-4.2, BD-2.1 |
| | | 5.3 | Semi-supervised и self-supervised learning | Semi-supervised: label propagation, pseudo-labeling, co-training. Мотивация: дорогая разметка, малая доля размеченных данных. Self-supervised: pretext tasks, contrastive learning (обзор). Связь с transfer learning и foundation models | ЛК | ML-2.1, FC-1.1, SS-1.1 |
| | | 5.4 | Практикум: NMF для тематического моделирования | NMF на TF-IDF матрице документов. Выбор числа тем. Интерпретация: топ-слова каждой темы. Визуализация тем. Сравнение с LDA (sklearn LatentDirichletAllocation). Coherence score. Практика: тематический анализ корпуса новостей | СЗ | MF-4.2, ML-2.1 |
| | | 5.5 | Практикум: LDA и BERTopic | LDA в gensim: обучение, coherence, визуализация (pyLDAvis). BERTopic: embedding + HDBSCAN + c-TF-IDF. Сравнение: NMF vs. LDA vs. BERTopic. Практика: тематический анализ коллекции отзывов, сравнение интерпретируемости | СЗ | MF-4.2, FC-1.1 |
| | | 5.6 | Практикум: рекомендательная система — коллаборативная фильтрация | Библиотека Surprise: SVD, SVD++, NMF. Оценка: RMSE, MAE. Кросс-валидация. Предсказание рейтинга. Top-N рекомендации. Практика: рекомендательная система на датасете MovieLens | СЗ | ML-2.1, MF-4.2, BD-2.1 |
| | | 5.7 | Практикум: контентная и гибридная рекомендации | Контентная: TF-IDF / эмбединги описаний → cosine similarity. Гибридная: взвешенная комбинация. Практика: гибридная рекомендательная система (collaborative + content). Обсуждение: cold start, scalability | СЗ | ML-2.1, MF-4.2 |
| | | 5.8 | Практикум: semi-supervised learning | Label Propagation и Label Spreading (sklearn). Pseudo- | СЗ | ML-2.1, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|-------------------|--|--|----------------------|------------------------|
| | | | | labeling: обучение на размеченных → предсказание для неразмеченных → добавление уверенных → переобучение. Практика: классификация с 10% размеченных данных, сравнение supervised vs. semi-supervised | | FC-1.1 |
| | | 5.9 | Практикум: manifold learning — Isomap, LLE, Spectral Embedding | Isomap: геодезические расстояния. LLE (Locally Linear Embedding): линейная реконструкция соседей. Spectral Embedding: лапласиан графа. Сравнение с PCA, t-SNE, UMAP на нелинейных датасетах (Swiss Roll). Практика: визуализация высокоразмерных данных | СЗ | MF-4.1, ML-2.1 |
| | | 5.10 | Практикум: обнаружение аномалий — продвинутые методы | Autoencoders для обнаружения аномалий (обзор): ошибка реконструкции как anomaly score. Isolation Forest на больших данных. Ensemble подходы. Практика: обнаружение аномалий в сетевом трафике / IoT-данных | СЗ | ML-2.1, BD-1.5 |
| | | 5.11 | Практикум: метрики для задач без учителя | Проблема оценки без меток. Внутренние метрики: Silhouette, BIC/AIC, coherence. Внешние (при наличии меток): ARI, NMI, Fowlkes-Mallows. Стабильность кластеризации. Практика: комплексная оценка результатов кластеризации несколькими метриками | СЗ | ML-2.2, ОПК-1.3 |
| | | 5.12 | Практикум: работа с текстовыми данными для ML | Предобработка: токенизация, стемминг, лемматизация, стоп-слова. Векторизация: Bag-of-Words, TF-IDF. Классификация текстов: NB, SVM, LR на TF-IDF. Embedding-based: sentence-transformers (обзор). Практика: классификация текстов с полным пайплайном | СЗ | MF-4.2, BD-1.3 |
| | | 5.13 | Практикум: работа с изображениями для ML (без DL) | Признаки изображений: гистограммы цвета, HOG, LBP. Классификация: SVM/RF на ручных признаках. Transfer learning embeddings: извлечение признаков из предобученной CNN (ResNet feature extractor) + классификатор sklearn. Практика: классификация изображений с помощью ML на CNN-эмбедингах | СЗ | MF-4.2, ML-2.1 |
| | | 5.14 | Практикум: итоговый проект семестра 6 — кейс unsupervised + supervised | Сквозная задача: кластеризация данных → разведочный анализ кластеров → использование кластеров как признаков → обучение классификатора → рекомендации. Полная документация: отчёт, Model Card, спецификация. Презентация | СЗ | ML-4.1, LC-2.1, ПК-3.3 |
| Раздел 6 | Временные ряды, | 6.1 | Временные ряды: основы анализа и | Определение временного ряда. Компоненты: тренд, | ЛК | MF-4.2, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--|-------------------|--|--|----------------------|------------------------|
| | последовательности и специальные типы данных | | классические методы | сезонность, шум. Декомпозиция: аддитивная, мультипликативная (STL). Стационарность: тест Дики-Фуллера (ADF). Автокорреляция: ACF, PACF. Классические модели: AR, MA, ARMA, ARIMA, SARIMA. Выбор порядка: AIC, BIC | | ML-2.1 |
| | | 6.2 | Прогнозирование временных рядов методами ML | ML для временных рядов: создание лаговых признаков, скользящих статистик, календарных. Проблема утечки из будущего: temporal split, TimeSeriesSplit. Модели: LightGBM на лаговых признаках, Prophet (Facebook), модели из sktime. Мультишаговое прогнозирование: direct, recursive, multi-output | ЛК | MF-4.2, ML-2.1, BD-1.4 |
| | | 6.3 | Практикум: анализ и декомпозиция временного ряда | Загрузка временного ряда (pandas DatetimeIndex). Визуализация: линейный график, подграфики по годам/месяцам. Декомпозиция (seasonal_decompose, STL). Тест стационарности (ADF). ACF и PACF (statsmodels). Практика: анализ реального временного ряда (продажи, трафик, погода) | СЗ | MF-4.2, PL-1.2 |
| | | 6.4 | Практикум: ARIMA и SARIMA | Выбор порядка (p, d, q) по ACF/PACF. auto_arima (pmdarima). Обучение ARIMA/SARIMA. Диагностика остатков (plot_diagnostics). Прогнозирование. Метрики: MAE, RMSE, MAPE. Практика: прогнозирование продаж на 12 месяцев | СЗ | MF-4.2, ML-2.2 |
| | | 6.5 | Практикум: Prophet | Prophet (Meta): установка, обучение. Компоненты: тренд (changepoints), сезонность (yearly, weekly, daily), праздники. Настройка: сезонность Фурье, changepoint_prior_scale. Практика: прогнозирование с Prophet, сравнение с ARIMA | СЗ | MF-4.2, ML-2.1 |
| | | 6.6 | Практикум: ML на лаговых признаках для временных рядов | Создание признаков: лаги (lag_1, lag_7, lag_30), скользящие статистики (rolling mean, std), календарные (day_of_week, month, is_holiday). TimeSeriesSplit для валидации. LightGBM на лаговых признаках. Сравнение с ARIMA и Prophet. Практика: прогнозирование спроса | СЗ | MF-4.2, BD-1.4 |
| | | 6.7 | Практикум: мультишаговое прогнозирование | Direct strategy: отдельная модель на каждый горизонт. Recursive: предсказание → подстановка как лаг → следующее предсказание. MultiOutputRegressor. Практика: прогноз на 7 дней вперед, сравнение стратегий, анализ | СЗ | MF-4.2, ML-2.2 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---|-------------------|--|---|----------------------|-------------------------|
| | | | | нарастания ошибки | | |
| | | 6.8 | Практикум: классификация временных рядов | Представления: raw, статистические признаки, shapelets (обзор). Библиотека tslearn / sktime: KNN на DTW, TimeSeriesForest. Практика: классификация активности человека (HAR) по данным акселерометра | СЗ | MF-4.2, ML-2.1 |
| | | 6.9 | Практикум: обнаружение аномалий во временных рядах | Подходы: статистические (z-score на скользящем окне), ML (Isolation Forest на лаговых признаках), модельные (прогноз + порог на ошибку). Changepoint detection: PELT, Bayesian (ruptures). Практика: обнаружение аномалий в серверных метриках (CPU, latency) | СЗ | MF-4.2, BD-1.5 |
| | | 6.10 | Практикум: работа с геоданными | Признаки: расстояния (haversine), геохеш, пространственные агрегаты. Визуализация: folium, plotly mapbox. Кластеризация: DBSCAN/HDBSCAN на GPS. Практика: анализ пространственных данных (такси, доставка), построение признаков для ML-модели | СЗ | MF-4.2, BD-1.4 |
| | | 6.11 | Практикум: работа с табличными данными с разнородными типами | Комбинирование числовых, категориальных, текстовых, временных, пространственных признаков в одном Pipeline. ColumnTransformer для отдельной обработки. Практика: построение модели для датасета с разнородными типами данных (объявления, транзакции) | СЗ | ML-3.1, BD-1.3 |
| | | 6.12 | Практикум: A/B-тестирование и причинно-следственный вывод (обзор) | A/B-тест: дизайн, расчёт размера выборки, анализ (t-тест, bootstrap). Uplift modeling: идея, CATE (обзор). Причинно-следственный вывод: DoWhy (обзор). Связь ML и каузального вывода. Практика: проведение A/B-теста для оценки ML-модели | СЗ | MF-4.1, SS-3.1, ОПК-1.3 |
| | | 6.13 | Практикум: итоговый проект семестра 6 — прогнозирование на реальных данных | Сквозная задача: временной ряд из открытых источников → анализ → feature engineering → обучение нескольких моделей (ARIMA, Prophet, LightGBM) → мультишаговый прогноз → оценка → документирование. Презентация | СЗ | ML-4.1, MF-4.2, ПК-3.3 |
| Раздел 7 | Вероятностные модели и байесовский подход | 7.1 | Байесовские методы в ML | Байесовский вывод: априорное, правдоподобие, апостериорное. MAP vs. MLE. Сопряжённые распределения. Связь регуляризации с байесовским априором. Байесовская линейная регрессия. Гауссовские процессы: ядро, апостериорное распределение, неопределённость предсказания (обзор). Байесовская | ЛК | FC-1.1, MF-1.3, ML-2.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|-------------------|--|----------------------|-------------------------|
| | | | оптимизация гиперпараметров | | |
| | | 7.2 | Графические модели и вероятностные программы | ЛК | FC-1.1, MF-1.3, ОПК-6.2 |
| | | 7.3 | Практикум: байесовская линейная регрессия | СЗ | FC-1.1, MF-1.3 |
| | | 7.4 | Практикум: гауссовские процессы для регрессии | СЗ | FC-1.1, ML-2.1 |
| | | 7.5 | Практикум: байесовская оптимизация гиперпараметров | СЗ | ML-3.2, FC-1.1 |
| | | 7.6 | Практикум: РуМС — вероятностное программирование | СЗ | FC-1.1, MF-1.3 |
| | | 7.7 | Практикум: байесовские сети | СЗ | FC-1.1, ОПК-6.2 |
| | | 7.8 | Практикум: EM-алгоритм — глубокое погружение | СЗ | MF-1.3, ML-2.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---|-------------------|---|---|----------------------|-------------------------|
| | | | | монотонность ELBO. Практика: реализация EM для GMM с нуля, визуализация итераций | | |
| | | 7.9 | Практикум: моделирование неопределённости и доверительные интервалы | Источники неопределённости: алеаторическая (шум данных), эпистемическая (незнание модели). Методы оценки: bootstrap prediction intervals, conformal prediction (обзор), MC Dropout (обзор). Практика: построение доверительных интервалов для предсказаний GBM через bootstrap | СЗ | FC-1.1, ML-2.2 |
| | | 7.10 | Практикум: conformal predictionж | Conformal prediction: идея гарантированного покрытия. Split conformal: calibration set → nonconformity scores → prediction sets. Для классификации: prediction sets. Для регрессии: prediction intervals. Библиотека MAPIE. Практика: построение prediction intervals с гарантированным покрытием | СЗ | FC-1.1, ML-2.2, ОПК-1.3 |
| | | 7.11 | Практикум: мини-проект — модель с оценкой неопределённости | Сквозная задача: обучение модели → оценка неопределённости (bootstrap / conformal / GP) → принятие решений с учётом неопределённости (отказ от предсказания при высокой неопределённости). Документирование: Model Card с описанием неопределённости | СЗ | FC-1.1, ПК-3.3, ML-4.1 |
| Раздел 8 | Обучение с подкреплением и специальные задачи | 8.1 | Введение в обучение с подкреплением | Формализация: агент, среда, состояние, действие, награда, политика, функция ценности. MDP (Markov Decision Process). Уравнение Беллмана. Типы методов: model-based vs. model-free, value-based vs. policy-based. Exploration vs. exploitation. Многорукий бандит. Epsilon-greedy стратегия | ЛК | ML-2.1, MF-1.1, SS-1.1 |
| | | 8.2 | Методы обучения с подкреплением | Q-learning: алгоритм, таблица Q-значений, сходимость. SARSA: on-policy альтернатива. Deep Q-Network (DQN, обзор): нейросеть вместо таблицы, experience replay, target network. Policy Gradient: REINFORCE. Actor-Critic (обзор). Применения: игры, робототехника, рекомендации, оптимизация | ЛК | ML-2.1, MF-3.2 |
| | | 8.3 | Практикум: многорукий бандит | Реализация: epsilon-greedy, UCB, Thompson Sampling. Среда: набор автоматов с разными вероятностями. Визуализация: cumulative regret. Сравнение стратегий. Обсуждение: связь с A/B-тестированием, рекомендациями | СЗ | ML-2.1, MF-3.2 |
| | | 8.4 | Практикум: Q-learning в среде | Установка Gymnasium (OpenAI Gym). Среда: FrozenLake, | СЗ | ML-2.1, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|---|---|----------------------|------------------------|
| | | Gymnasium | Taxi. Q-learning: реализация, обновление таблицы, визуализация Q-таблицы. Параметры: learning rate, discount factor, epsilon decay. Визуализация обучения: награда по эпизодам | | MF-3.2 |
| | | 8.5 Практикум: SARSA и сравнение с Q-learning | SARSA: on-policy алгоритм, отличие от Q-learning. Реализация. Сравнение Q-learning и SARSA на среде CliffWalking: разные стратегии (cautious vs. optimal path). Визуализация | СЗ | ML-2.1 |
| | | 8.6 Практикум: Deep Q-Network (DQN) — обзорная реализация | DQN: нейросеть для аппроксимации Q-функции. Experience replay buffer. Target network. Реализация упрощенного DQN на PyTorch для CartPole. Обучение и визуализация. Обсуждение: масштабирование на сложные среды | СЗ | ML-2.1, MF-3.2 |
| | | 8.7 Практикум: AutoML — автоматизация подбора модели | Концепция AutoML: автоматический подбор предобработки, модели и гиперпараметров. Инструменты: auto-sklearn, FLAML, H2O AutoML (обзор). Практика: запуск AutoML на реальном датасете, анализ результатов, сравнение с ручным подходом. Обсуждение: когда AutoML полезен, а когда нет | СЗ | ML-3.2, УК-12.2 |
| | | 8.8 Практикум: Active Learning | Идея: модель запрашивает разметку наиболее информативных примеров. Стратегии: uncertainty sampling, query-by-committee, expected model change. Библиотека modAL. Практика: сравнение random sampling vs. active learning на задаче классификации. Экономия разметки | СЗ | ML-2.1, FC-1.1 |
| | | 8.9 Практикум: transfer learning для табличных данных | Предобученные эмбединги как признаки. TabNet (обзор): self-supervised pretraining на неразмеченных данных. Fine-tuning. Практика: использование предобученных эмбедингов категориальных признаков для улучшения модели на маленьком датасете | СЗ | ML-2.1, FC-1.1 |
| | | 8.10 Практикум: мультимодальные данные | Объединение разнородных данных: табличные + текстовые + изображения. Стратегии: отдельные модели → конкатенация предсказаний, early fusion (конкатенация эмбедингов), late fusion (ансамбль). Практика: мультимодальная классификация (описание + изображение) | СЗ | MF-4.3, ML-2.1 |
| | | 8.11 Практикум: итоговый проект семестра 7 | Сквозная задача: выбор задачи из предложенных (RL-среда, рекомендации, временные ряды + ML, модель с | СЗ | ML-4.1, ПК-3.3, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---|-------------------|---|---|----------------------|-------------------------|
| | | | | неопределённостью). Полный цикл: постановка → данные → модель → оценка → документирование. Презентация с акцентом на обоснование подхода | | LC-2.1 |
| Раздел 9 | Интерпретируемость, справедливость и ответственный ML | 9.1 | Интерпретируемость и объяснимость моделей ML | Интерпретируемые модели: линейные, деревья, правила. Post-hoc объяснения: feature importance (permutation, SHAP), PDP, ICE. Локальные: LIME, SHAP waterfall. Глобальные: SHAP summary. Attention и gradient-based (обзор для нейросетей). Уровни объяснимости: для разработчика, пользователя, регулятора | ЛК | ML-6.1, ОПК-6.2, SS-3.2 |
| | | 9.2 | Справедливость и ответственная разработка ML | Формальные определения справедливости: demographic parity, equalized odds, predictive parity. Невозможность одновременного выполнения. Источники предвзятости: данные, модель, оценка. Методы митигации: preprocessing (resampling, reweighting), in-processing (adversarial debiasing), post-processing (threshold adjustment). Fairlearn, AIF360. Model Card как инструмент ответственности | ЛК | SS-1.2, SS-3.2, FC-5.2 |
| | | 9.3 | Практикум: SHAP — глубокое погружение | SHAP: TreeSHAP, KernelSHAP. Summary plot, dependence plot, waterfall, force plot. Глобальная и локальная интерпретация. Выявление bias через SHAP: различия объяснений по подгруппам. Практика: полный анализ интерпретируемости GBM-модели с SHAP | СЗ | ML-6.1, ОПК-6.2 |
| | | 9.4 | Практикум: LIME и counterfactual explanations | LIME: локальное объяснение через интерпретируемую модель-заменитель. Counterfactual explanations: «что нужно изменить, чтобы получить другое предсказание?». DiCE (Microsoft). Практика: LIME и counterfactual для модели кредитного скоринга | СЗ | ML-6.1, ОПК-6.2 |
| | | 9.5 | Практикум: измерение справедливости модели | Fairlearn: вычисление метрик справедливости по подгруппам (MetricFrame). Demographic parity difference, equalized odds difference. Визуализация disparities. Практика: анализ справедливости модели скоринга по полу/расе/возрасту | СЗ | SS-3.2, FC-5.2 |
| | | 9.6 | Практикум: митигация предвзятости | Методы: resampling (oversampling minority), reweighting (sample_weight), threshold adjustment (Fairlearn ThresholdOptimizer). Adversarial debiasing (обзор). Практика: применение митигации, измерение trade-off accuracy vs. | СЗ | SS-3.2, FC-5.2 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---|-------------------|--|--|----------------------|-------------------------|
| | | | | fairness | | |
| | | 9.7 | Практикум: Model Card и Datasheet | Составление Model Card (Mitchell et al.): назначение, метрики, ограничения, этические аспекты, оценка по подгруппам. Datasheet for Datasets (Geburu et al.): источник, состав, предвзятости. Практика: заполнение Model Card и Datasheet для учебного проекта | СЗ | FC-5.2, LC-2.2 |
| | | 9.8 | Практикум: юнит-тесты и регрессионные тесты для ML | Тестирование ML-кода: тесты предобработки (pytest), тесты модели (determinism, invariance, directional expectations). Регрессионные тесты: метрика не хуже порога. Great Expectations для тестирования данных (обзор). Практика: написание тестов для ML-пайплайна | СЗ | ОПК-3.3, ПК-3.1 |
| | | 9.9 | Практикум: приёмочное тестирование ML-модели | Чек-лист приёмки: метрики \geq порога, справедливость, интерпретируемость, латентность, размер модели. Тестирование на edge cases и adversarial examples. Документирование результатов тестирования. Практика: проведение приёмочного тестирования для готовой модели | СЗ | ПК-3.1, ПК-3.3 |
| | | 9.10 | Практикум: итоговый проект — ответственный ML-пайплайн | Сквозная задача: модель для чувствительной области (кредиты, здоровье, HR) → обучение → SHAP-анализ → оценка справедливости → митигация → тесты → Model Card → Datasheet. Презентация с акцентом на ответственность | СЗ | ML-4.1, FC-5.2, SS-3.2 |
| Раздел 10 | MLOps: развёртывание и мониторинг моделей | 10.1 | Жизненный цикл ML-модели в продакшене | Стадии: разработка → экспериментирование → развёртывание → мониторинг → переобучение → вывод из эксплуатации. Experiment tracking: MLflow. Model registry: версионирование моделей. CI/CD для ML. Feature store (обзор). Ролевая модель: Data Scientist, ML Engineer, MLOps Engineer | ЛК | LC-1.2, ML-4.1, SS-1.2 |
| | | 10.2 | Развёртывание моделей: паттерны и инструменты | Паттерны: batch prediction (scoring job), online prediction (REST API), streaming prediction. Инструменты: FastAPI + Docker, MLflow Models, BentoML (обзор), Triton (обзор). A/B-тестирование и canary deployment. Shadow mode. Blue-green deployment | ЛК | ML-4.2, PL-1.2, ОПК-2.2 |
| | | 10.3 | Мониторинг ML-моделей в продакшене | Дрейф данных (data drift): covariate shift, prior probability shift. Дрейф модели (concept drift). Методы обнаружения: PSI, KL-divergence, KS-test, ADWIN. Мониторинг метрик: | СЗ | ML-4.3, LC-2.2 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|-------------------|--|--|----------------------|------------------------|
| | | | | accuracy degradation. Инструменты: Evidently AI, NannyML (обзор). Стратегии переобучения: scheduled, trigger-based | | |
| | | 10.4 | Практикум: MLflow — трекинг экспериментов | MLflow Tracking: logging parameters, metrics, artifacts. MLflow UI. Автоматическое логирование (autolog). Сравнение экспериментов. Практика: обучение нескольких моделей с логированием в MLflow, выбор лучшей | СЗ | LC-1.2, ML-4.1 |
| | | 10.5 | Практикум: MLflow — Model Registry и версионирование | MLflow Model Registry: регистрация модели, версии, стадии (Staging, Production, Archived). Переход между стадиями. Загрузка модели из реестра. Практика: полный workflow от логирования до регистрации модели в реестре | СЗ | LC-1.2, ML-4.2 |
| | | 10.6 | Практикум: развёртывание модели как REST API | FastAPI сервис предсказаний: загрузка модели, endpoint /predict, валидация входов (Pydantic), обработка ошибок. Контейнеризация (Dockerfile). Docker Compose с Redis для кэширования. Health check. Практика: развёртывание sklearn/LightGBM модели | СЗ | ML-4.2, ОПК-3.3 |
| | | 10.7 | Практикум: batch prediction pipeline | Сценарий: ежедневное скоринг всех клиентов. Pipeline: загрузка данных из БД → загрузка модели из MLflow → предсказание → сохранение результатов. Планирование: cron / Airflow (обзор). Практика: реализация batch prediction pipeline с логированием | СЗ | ML-4.2, LC-2.2 |
| | | 10.8 | Практикум: мониторинг дрейфа данных (Evidently AI) | Evidently: Data Drift Report, Data Quality Report, Model Performance Report. Интеграция в CI/CD: автоматическая проверка дрейфа. Практика: обнаружение дрейфа на синтетическом примере, настройка порогов, генерация отчётов | СЗ | ML-4.3, LC-2.2 |
| | | 10.9 | Практикум: стратегии переобучения модели | Scheduled retraining: переобучение по расписанию. Triggered retraining: по сигналу дрейфа. Инкрементальное обучение (partial_fit, warm_start). Online learning (обзор). Практика: автоматизация переобучения при обнаружении дрейфа | СЗ | ML-4.3, LC-2.2 |
| | | 10.10 | Практикум: мини-проект — полный MLOps-пайплайн | Сквозная задача: данные → обучение с MLflow → развёртывание (FastAPI + Docker) → мониторинг (Evidently) → переобучение по сигналу. Документирование: архитектурная диаграмма, инструкция по развёртыванию. Презентация | СЗ | ML-4.1, ML-4.2, ML-4.3 |
| Раздел | Продвинутые темы и | 11.1 | Масштабирование ML: большие данные | ML на больших данных: out-of-core learning (partial_fit), | ЛК | ОПК-2.2, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|-------------------|---|--|----------------------|------------------------|
| 11 | масштабирование | | и распределённое обучение | Dask-ML, Spark MLlib (повторение). Распределённое обучение: data parallelism (обзор). Ускорение инференса: ONNX, TensorRT (обзор). Квантизация моделей (обзор). Edge deployment (обзор) | | ML-3.3, PL-1.3 |
| | | 11.2 | Продвинутое ансамблевые методы и мета-обучение | Двухуровневый стекинг. Blending на holdout. Ансамбль разнородных моделей (табличные + текстовые + нейросетевые). Meta-learning: learning to learn (обзор). Few-shot learning (обзор). Neural Architecture Search (обзор). Связь с AutoML | ЛК | ML-2.3, FC-1.1, ML-3.2 |
| | | 11.3 | Практикум: out-of-core learning | sklearn partial_fit: SGDClassifier, MiniBatchKMeans. Чтение данных чанками (pd.read_csv chunks). Dask-ML: distributed training. Практика: обучение модели на датасете, не помещающемся в память | СЗ | ML-3.3, ОПК-2.2 |
| | | 11.4 | Практикум: ONNX — экспорт и ускорение инференса | Экспорт sklearn/LightGBM модели в ONNX (skl2onnx, onnxmltools). ONNX Runtime: загрузка и инференс. Замер ускорения: ONNX vs. native predict. Практика: конвертация пайплайна в ONNX, интеграция с FastAPI-сервисом | СЗ | ML-3.3, PL-1.3 |
| | | 11.5 | Практикум: двухуровневый стекинг на реальных данных | Уровень 1: разнородные модели (LR, RF, XGBoost, CatBoost, SVM). Out-of-fold predictions. Уровень 2: мета-модель (LR / Ridge). Оценка через nested CV. Практика: сравнение одноуровневого и двухуровневого стекинга | СЗ | ML-2.3, ML-4.1 |
| | | 11.6 | Практикум: воспроизводимость ML-проекта — чек-лист | Полный чек-лист: фиксация seed, версионирование данных (DVC), версионирование кода (Git), фиксация зависимостей (requirements.txt / poetry.lock), Docker, MLflow. Практика: проверка воспроизводимости учебного проекта по чек-листу | СЗ | LC-2.2, LC-1.2 |
| | | 11.7 | Практикум: code review для ML-кода | Специфика code review в ML: проверка утечки данных, корректность валидации, reproducibility, тесты, документация. Чек-лист ML code review. Практика: взаимное ревью ML-проектов по чек-листу | СЗ | ОПК-3.3, ML-4.3 |
| | | 11.8 | Практикум: подготовка ML-решения к передаче заказчику | Формирование delivery package: обученная модель, API, документация (Model Card, README, API spec), тесты, инструкция по развёртыванию. Приёмочное тестирование. Обучение пользователей (обзор). Практика: подготовка полного пакета для передачи | СЗ | ML-4.1, ПК-3.3, LC-2.1 |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|--|-------------------|---|---|----------------------|------------------------|
| | | 11.9 | Практикум: ретроспектива ML-проекта | Ретроспектива: что пошло хорошо, что можно улучшить, какие уроки извлечены. Технический долг в ML. Метрики процесса: время от данных до модели, число итераций. Практика: ретроспектива по учебным проектам семестров 5–8 | СЗ | ML-4.3, SS-1.2 |
| Раздел 12 | Итоговый проект: промышленное ML-решение | 12.1 | Постановка итогового проекта и выбор задачи | Требования к итоговому проекту: реальная задача, полный цикл (данные → модель → развёртывание → мониторинг), документация, презентация. Выбор задачи из предложенных или своей. Формулирование бизнес-постановки и трансляция в ML. Определение метрик успеха | ЛК | ПК-1.1, ML-4.1 |
| | | 12.2 | Планирование проекта и обследование данных | Декомпозиция проекта: этапы, сроки, deliverables. Обследование данных: источники, объём, качество, доступность. Анализ осуществимости. Формулирование ТЗ: функциональные и нефункциональные требования. Критерии приёмки | ЛК | ПК-1.1, ПК-1.2, ПК-3.1 |
| | | 12.3 | Практикум: EDA и baseline для итогового проекта | Загрузка и профилирование данных. Полный EDA. Обучение baseline модели (простейшая: mean, majority class, линейная). Фиксация baseline метрик. Обсуждение: превышение baseline как минимальный критерий успеха | СЗ | BD-1.2, ML-2.2 |
| | | 12.4 | Практикум: feature engineering и предобработка | Разработка пайплайна предобработки (Pipeline + ColumnTransformer). Feature engineering для конкретной задачи. Тестирование пайплайна. Документирование: описание признаков, трансформаций, обоснование решений | СЗ | BD-1.4, ML-3.1 |
| | | 12.5 | Практикум: обучение и подбор моделей | Обучение нескольких моделей. Подбор гиперпараметров (Optuna). Стекинг/ансамблирование. Кросс-валидация. Анализ ошибок. SHAP-интерпретация. Оценка справедливости | СЗ | ML-2.3, ML-6.1 |
| | | 12.6 | Практикум: развёртывание и тестирование | Развёртывание модели (FastAPI + Docker). Написание тестов: unit, integration, приёмочные. Мониторинг: настройка проверки дрейфа. MLflow: логирование финальной модели | СЗ | ML-4.2, ОПК-3.3 |
| | | 12.7 | Практикум: документирование итогового проекта | Model Card. Datasheet. Техническая спецификация. README с инструкцией по развёртыванию. Архитектурная диаграмма. План тестирования | СЗ | LC-2.1, FC-5.2, ПК-1.2 |
| | | 12.8 | Практикум: подготовка презентации и | Подготовка презентации: бизнес-контекст → данные → | СЗ | ML-4.1, |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы * | Формируемые индикаторы |
|---------------|---------------------------------|-------------------|-------------------------------------|--|----------------------|------------------------------|
| | | | peer review | метод → результаты → ограничения → дальнейшие шаги. Peer review проектов: обратная связь по чек-листу | | SS-3.2 |
| | | 12.9 | Практикум: защита итогового проекта | Защита: презентация (15 мин) + ответы на вопросы (10 мин). Критерии: качество модели, обоснованность решений, полнота пайплайна, документация, интерпретируемость, справедливость, воспроизводимость. Подведение итогов курса | СЗ | ML-4.1, ПК-3.3, LC-2.1 |

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

| Тип аудитории | Оснащение аудитории | Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости) |
|----------------------------|---|--|
| Лекционная | Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций. | |
| Семинарская | Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций. | Персональные компьютеры, необходимое ПО |
| Для самостоятельной работы | Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС. | Персональные компьютеры, необходимое ПО |

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Основы машинного обучения и нейронные сети. учебное пособие [Электронный ресурс] / Бобрикова Екатерина Васильевна [и др.]. - М.: РУДН, 2024. 124 с. ISBN 978-5-209-12322-4 URL: https://mega.rudn.ru/MegaPro/UserEntry?Action=Link_FindDoc&id=517203&idb=0

2. Татарникова, Татьяна Михайловна. Методы машинного обучения: учебное пособие. - Санкт-Петербург: ГУАП, 2023. - 99 с.: ил., табл.; 21 см.; ISBN 978-5-8088-1885-9

Дополнительная литература:

1. Фомин, Владимир Владимирович. Методы машинного обучения: лабораторный практикум. - Санкт-Петербург: ГУАП, 2024 (Санкт-Петербург). - 37 с.: ил.

2. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662>

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН
<https://mega.rudn.ru/MegaPro/Web>
- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
- ЭБС «Юрайт» <http://www.biblio-online.ru>
- ЭБС «Консультант студента» www.studentlibrary.ru
- ЭБС «Знаниум» <https://znanium.ru/>

2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>
- Springer Nature Link <https://link.springer.com/>
- Wiley Journal Database <https://onlinelibrary.wiley.com/>
- Научометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Методы машинного обучения».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**