

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Декан РТФ

УТВЕРЖДАЮ /А.Н. Дедов/
(Ф.И.О. декана (директора института))

13.02.2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б.1.2.3 Введение в машинное обучение и интеллектуальные системы

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

09.03.02 Информационные системы и технологии

Квалификация выпускника

Бакалавр

(бакалавр/магистр/специалист)

Направленность

Интеллектуальные информационные системы и
технологии

Курс 2
Семестр 4

Распределение учебного времени

Трудоемкость по учебному плану	180 / 5	часов/зачетных единиц
Лекции	36	часов
Лабораторные работы	-	часов
Практические занятия	36	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	72	часов
Контактная работа по экзамену	-	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	108	часов
Самостоятельная работа по подготовке к экзамену	-	часов
Экзамен	-	семестр
Зачет	4	семестр
БРК, ДЗ	-	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 09.03.02 Информационные системы и технологии

Программу составили:

старший преподаватель	ИТРС	СОГЛАСОВАНО	Н.А. Конкин
(должность)	(кафедра)		(И.О. Фамилия)

РАССМОТРЕНА и ОДОБРЕНА на заседании кафедры, за которой закреплена дисциплина
Кафедра информационных технологий, радиотехники и связи

(наименование кафедры)		
13.01.2025	протокол №	2
(дата)		

Заведующий кафедрой	СОГЛАСОВАНО	Н.В. Рябова
		(И.О. Фамилия)

Рабочая программа СОГЛАСОВАНА с факультетом (институтом), выпускающей(ими) кафедрой(ами).

СООТВЕТСТВУЕТ действующей ОП.

Заведующий кафедрой	СОГЛАСОВАНО	Н.В. Рябова
		(И.О. Фамилия)

Председатель методической комиссии факультета (института), в который входит выпускающая кафедра

СОГЛАСОВАНО	А.Н. Дедов
	(И.О. Фамилия)

Эксперт(ы): Дудин Александр Николаевич , Исполнительный директор ООО «Трэвел Лайн Системс»

Рабочая программа проверена и зарегистрирована в УМЦ 17.02.2025 г.

Специалист учебно-методического центра СОГЛАСОВАНО /Т.А. Смирнова/

Раздел 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является достижение планируемых результатов обучения, соответствующих установленным в ОПОП индикаторам достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения
1. BD-1 Способен осуществлять поиск, сбор, очистку и предварительный анализ данных	BD-1.1. Применяет методы анализа данных для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ	знания: Знает основные подходы к обработке пропусков и выбросов в табличных данных, понимает их статистическую природу, причины возникновения и влияние на корректность анализа. Осознаёт различия между типами пропусков (MCAR, MAR, MNAR), видами выбросов и соответствующими методами их обработки. умения: Умеет применять статистические и алгоритмические методы заполнения пропусков (средним, медианой, регрессионными моделями, kNN-импутацией), а также корректно выявлять и устранять выбросы с использованием правил трёх сигм, межквартильного размаха, z-оценок и других методов. Способен адаптировать подход к особенностям конкретного набора данных. навыки: Владеет практическими навыками реализации процедур импутации и очистки данных в инструментах анализа данных, использует Python-библиотеки (Pandas, NumPy, Scikit-learn) для автоматизации процессов и обеспечивает корректность подготовки данных для дальнейшего анализа и моделирования.
	BD-1.2. Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных	знания: Знает основные методы понижения размерности и их математические принципы, включая PCA, t-SNE, UMAP и линейные/нелинейные преобразования признакового пространства. умения: Умеет применять инструменты анализа данных для проверки разведочных гипотез, исследовать взаимосвязи между признаками, визуализировать структуры данных и отбирать информативные признаки. навыки: Владеет практическими навыками использования современных библиотек Python (Pandas, Scikit-learn, umap-learn, OpenTSNE) для выполнения понижения размерности, подготовки данных к обучению моделей ИИ и повышения эффективности последующей обработки.

2. 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 машинного обучения и применяет на практике принципы их решения	знания: Знает ключевые типы задач машинного обучения — обучение с учителем, без учителя и с подкреплением, понимает их различия, области применения и характерные алгоритмические подходы. Осознаёт, какие виды данных и цели соответствуют каждому типу задач. умения: Умеет выбирать подходящий класс алгоритмов в зависимости от постановки задачи, проводить базовый анализ данных и подготавливать их для обучения модели. Применяет стандартные методы классификации, регрессии, кластеризации и снижения размерности с использованием готовых инструментов. навыки: Владеет практическими навыками работы с библиотекой Scikit-learn, уверенно использует её стандартные функции для решения типовых ML-задач, оценки качества моделей и проведения минимально необходимого эксперимента для подтверждения корректности выбранного подхода.
3. ML-3 Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения	ML-3.2. Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ	знания: Знает основные принципы задач ранжирования, их отличие от классификации и регрессии, а также особенности pointwise-подхода. Понимает, в каких практических ситуациях данные задачи применяются и какие метрики используются для оценки качества ранжирования. умения: Умеет выявлять прикладные задачи, формулировать их в терминах ранжирования, выбирать целевую функцию и подготавливать данные для модели pointwise-ранжирования. Способен корректно интерпретировать результаты и проводить первоначальную оценку качества построенного решения. навыки: Владеет практическими навыками применения простейших методов pointwise-ранжирования с использованием стандартных библиотек Python, умеет строить базовые модели ранжирования и интегрировать их в аналитический процесс для решения прикладных задач.

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к части, формируемой участниками образовательных отношений ОПОП.

Дисциплина является обязательной

Для продолжения формирования заявленных компетенций необходимы знания предшествующих практик: Учебная практика. Ознакомительная практика (BD-1), Учебная практика. Ознакомительная практика (ML-2), Учебная практика. Ознакомительная практика

(ML-3)

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих государственной итоговой аттестации в форме: Выполнение и защита выпускной квалификационной работы (BD-1), Выполнение и защита выпускной квалификационной работы (ML-2), Выполнение и защита выпускной квалификационной работы (ML-3)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Для формирования заявленных компетенций используются методологические технологии, реализующие деятельностный, личностно-ориентированный, практико-ориентированный подходы.

Основными стратегическими технологиями являются: лекционные занятия, практические занятия

На достижение конкретных целей обучения направлены применяемые тактические технологии: задания, информационные, классическая лекция, мини-проекты

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
№1 Введение в интеллектуальные системы и основы машинного	36	ML-2
Лекция. 1. Обзор интеллектуальных систем: понятие ИИ, эволюция от экспертных систем к ML. Различие типов задач ML (с учителем, без учителя, с подкреплением). 2. Основные принципы ML: этапы жизненного цикла модели (сбор данных, обучение, оценка). Введение в библиотеки (Scikit-Learn). 3. Типовые подходы к задачам ML: примеры классификации, регрессии и кластеризации с использованием готовых инструментов.	9	
Практическое занятие. - Решение простой задачи классификации с учителем с использованием Scikit-Learn (оценка модели метриками accuracy, precision). - Групповая работа: анализ реального кейса (например, предсказание цен) и выбор типа задачи ML.	9	
Задания для самостоятельной работы, в том числе выполнение: Задания для самостоятельной работы, в том числе выполнение: - Самостоятельный анализ 2-3 датасетов: определение типа задачи ML и подготовка отчета. - Чтение статей по истории ИИ (например, от Тьюринга к современным системам) и аннотация.	18	
№2 Подготовка и анализ данных для ML	36	BD-1
Лекция. 1. Поиск и сбор данных: источники (открытые датасеты, API), этические аспекты. Методы предварительного анализа (статистические характеристики). 2. Очистка данных: заполнение пропусков (импутация, медиана), удаление выбросов (Z-score, IQR) в табличных данных. 3. Работа с признаками: понижение размерности (PCA, t-SNE) с	9	

использованием библиотек; feature engineering.		
Практическое занятие. - Практика сбора данных из открытых источников (например, UCI Repository) и предварительный анализ в Pandas. - Очистка датасета: применение методов заполнения пропусков и удаления выбросов на реальных данных (например, Titanic dataset). - Понижение размерности: использование Scikit-Learn для PCA на многомерном датасете и визуализация результатов.	9	
Задания для самостоятельной работы, в том числе выполнение - Самостоятельный сбор и очистка датасета по выбранной теме (например, медицинские данные) с отчетом - Эксперименты с разными методами импутации и сравнение влияния на качество данных - Изучение библиотек (NumPy, Pandas) через онлайн-тutorиалы и мини-проект	18	
№3 Классические алгоритмы машинного обучения	45	ML-3
Лекция. 1. Алгоритмы обучения с учителем: линейная регрессия, логистическая регрессия – математические основы (градиентный спуск, loss-функции). 2. Деревья решений и ансамбли: Random Forest, Boosting – области применения, оценка качества (cross-validation, ROC-AUC). 3. Обучение без учителя: кластеризация (K-means), ассоциативные правила – математические принципы и примеры.	9	
Практическое занятие. - Реализация линейной регрессии в Scikit-Learn на датасете (предсказание), оценка модели. - Построение дерева решений и Random Forest: тюнинг гиперпараметров, анализ важности признаков. - Кластеризация данных: применение K-means и визуализация кластеров (matplotlib).	9	
Задания для самостоятельной работы, в том числе выполнение - Самостоятельная реализация простого алгоритма (например, KNN) без библиотек для понимания математики - Анализ ошибок моделей на разных датасетах и отчет по оценке качества - Чтение научных статей по классическим алгоритмам (например, по Random Forest)	27	
№4 Применение ML в интеллектуальных системах	63	ML-2, ML-3
Лекция. 1. Задачи ранжирования в ML: формулировка проблемы (pointwise подход), математические основы (логистическая регрессия для ранжирования). 2. Интеграция ML в интеллектуальные системы: примеры (рекомендательные системы, поисковые движки), оценка моделей в реальных сценариях. 3. Этика и перспективы: bias в данных, интерпретируемость моделей.	9	
Практическое занятие. - Формулировка задачи ранжирования (например, ранжирование поисковых результатов) и применение pointwise метода в Scikit-Learn. - Построение простой рекомендательной системы на основе	9	

ML-алгоритмов. - Оценка этических аспектов: анализ bias в очищенном датасете и корректировка.		
Задания для самостоятельной работы, в том числе выполнение Расчётно-Графическая Работа в рамках задач от АО "ММЗ" по теме: Разработка базового ИИ-инструмента для автоматизации контроля качества продукции предприятия АО «ММЗ» с использованием методов машинного обучения	45	
Иная контактная работа:	0	

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Подготовка к лабораторным занятиям включает ознакомление с планом занятия, работу с учебной и учебно-методической литературой, научными изданиями и электронными образовательными ресурсами, рекомендованными рабочей программой дисциплины. Содержание самостоятельной работы определяется рабочей программой дисциплины, оценочными и методическими материалами, заданиями и указаниями преподавателя. Самостоятельная работа может осуществляться в аудиторной и внеаудиторной формах. Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины (модуля), к электронным библиотечным системам, профессиональным базам данных и информационным справочным системам.

Контроль и аттестация

Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплины представлены в рабочей программе.

Формой промежуточной аттестации по дисциплине является зачет.

1) Необходимое программное обеспечение и инструментальные средства

Среды разработки

В качестве среды разработки для языка программирования Python рекомендуется использовать следующие решения:

1. **Visual Studio Code** (<https://code.visualstudio.com/>) — универсальный редактор с богатыми возможностями для Python-разработки
2. **PyCharm Community Edition** (<https://www.jetbrains.com/pycharm/>) — специализированная IDE для Python
3. **Jupyter Notebook/JupyterLab** (<https://jupyter.org/>) — интерактивная среда для исследовательской работы с данными

Облачные платформы для машинного обучения

Для организации удаленного коллективного решения задач машинного обучения рекомендуется использовать следующие платформы:

1. **Google Colaboratory** (<https://colab.research.google.com/>) — облачная среда с бесплатным доступом к GPU/TPU, интеграцией с Google Drive и поддержкой Jupyter Notebook
2. **Kaggle** (<https://www.kaggle.com/>) — платформа для соревнований по машинному обучению с большим разнообразием датасетов, примеров решений и возможностью работы в облачных ноутбуках с GPU

Kaggle Learn (<https://www.kaggle.com/learn>) — бесплатные интерактивные курсы по Python, машинному обучению и глубокому обучению

2) Содержание дисциплины включает изучение исследований и разработок, представленных в научных публикациях на конференциях уровня А*, статьях, опубликованных в журналах 1 уровня Белого списка научных журналов Минобрнауки России (УММ на основе публикаций в научных журналах 1 или 2 уровня Белого списка научных журналов Минобрнауки России, публикаций на конференциях уровня А*/А. Ниже приведен перечень конференций, рекомендуемых для подачи работ, соответствующих уровню подготовки студентов:

1. NeurIPS — Advances in Neural Information Processing Systems

Сайт: <https://neurips.cc/>

Краткое описание: ведущая конференция по машинному обучению; охватывает основы ML, оптимизацию, нейронные сети, прикладные задачи и исследования начального уровня.

2. ICML — International Conference on Machine Learning

Сайт: <https://icml.cc/>

Краткое описание: одна из ключевых конференций по машинному обучению; включает широкий спектр базовых методов, теории обучения и практических ML-приложений.

3. ICLR — International Conference on Learning Representations

Сайт: <https://iclr.cc/>

Краткое описание: конференция, посвящённая обучению представлений и глубокому обучению; подходит для знакомства студентов с современными архитектурами и методами обучения.

4. AAAI — AAAI Conference on Artificial Intelligence

Сайт: <https://aaai.org/conference/>

Краткое описание: крупная конференция по искусственному интеллекту; охватывает темы ML, NLP, CV и интеллектуальных систем на доступном уровне.

5. IJCAI — International Joint Conference on Artificial Intelligence

Сайт: <https://www.ijcai.org/>

Краткое описание: классическая конференция по ИИ; содержит материалы по машинному обучению, интеллектуальным агентам, планированию и reasoning.

6. KDD — ACM SIGKDD Conference on Knowledge Discovery and Data Mining

Сайт: <https://www.kdd.org/>

Краткое описание: крупнейшая конференция по data mining; содержит прикладные исследования, полезные студентам начального уровня ML.

7. EMNLP — Empirical Methods in Natural Language Processing

Сайт: <https://2024.emnlp.org/>

Краткое описание: конференция по обработке естественного языка; подходит для введения студентов в практические NLP-задачи.

8. CVPR — IEEE Conference on Computer Vision and Pattern Recognition

Сайт: <https://cvpr.thecvf.com/>

Краткое описание: ведущая конференция по компьютерному зрению; материалы демонстрируют практическое применение методов ML.

9. ICAPS — International Conference on Automated Planning and Scheduling

Сайт: <https://www.icaps-conference.org/>

Краткое описание: конференция по планированию и интеллектуальным системам принятия решений; релевантна при обучении базовым идеям агентного ИИ.

10. ACL — Annual Meeting of the Association for Computational Linguistics

Сайт: <https://aclweb.org/>

Краткое описание: ведущая NLP-конференция; включает исследования от простых моделей до современных LLM, полезно для начинающего уровня по теме интеллектуальных систем обработки текста.

3) Предусмотрено участие ИТ-практиков в различных форматах обучения (лекции, семинары с ИТ-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания и др.) и в качестве наставников.

В программе предусмотрено изучение кейсов (кейсы для изучения, практической и проектной работы) и реальных задач (проектов по разработке, внедрению, развитию цифровых технологий, ИТ-разработок и др.) из практического опыта Индустриальных партнеров в объеме не менее 1 кейса на 1 з.е.

РАСЧЁТНО-ГРАФИЧЕСКАЯ РАБОТА (РГР)

Тема: Разработка базового ИИ-инструмента для автоматизации контроля качества продукции предприятия АО «ММЗ» с использованием методов машинного обучения

1. Цель работы

Освоить базовые методы машинного обучения для решения прикладной задачи контроля качества. Научиться формулировать задачу, подготавливать данные, строить и обучать модели, оценивать качество решений и сравнивать алгоритмы.

2. Задачи работы

Студент должен:

1. Изучить особенности производственного процесса (условно или по данным, предоставленным партнёром).
2. Сформулировать задачу в терминах машинного обучения (классификация «дефект/норма» или ранжирование).

1. Подготовить набор данных:
 - очистка, обработка пропусков;
 - выделение признаков;
 - разделение на выборки.
2. Реализовать не менее трёх базовых моделей ML, например:
 - Logistic Regression;
 - Decision Tree / Random Forest.
3. Провести оценку моделей: accuracy, precision/recall, F1, ROC AUC.
4. Сравнить результаты и выбрать оптимальную модель.
5. Сделать выводы о применимости ML в промышленном контроле качества.

3. Объект исследования

Производственные данные предприятия АО «ММЗ», содержащие параметры изделий и бинарную метку качества (норма/дефект), либо синтетически созданный набор, отражающий реальные параметры:

- измерения (толщина, масса, сопротивление, температура);
- визуальные признаки (после простейшей предобработки, если данные изображений);
- категориальные параметры (материал, линия производства).

4. Требования к реализации

Данные:

Можно использовать:

- реальные данные партнёра (если предоставлены),
- открытые датасеты (например, производственные данные Kaggle),
- синтетические данные, созданные студентом (с описанием способа генерации).

Методы ML:

Реализовать минимум три алгоритма из Scikit-learn:

- LogisticRegression
- KNeighborsClassifier
- DecisionTreeClassifier / RandomForestClassifier
- (по желанию — SVM или Naive Bayes)

Этапы:

1. Предобработка данных:
 - удаление выбросов (IQR, Z-score),
 - заполнение пропусков (mean/median/knn),
 - масштабирование признаков.
2. Обучение моделей.
3. Оценка качества:
 - confusion matrix,
 - метрики качества, ROC-кривая,
4. Сравнение и выбор лучшего алгоритма.
5. Визуализация результатов.

4) Рекомендованные источники учебно-методической литературы из:

Дополнительные ссылки на источники информации (КРМ)

BD1:

- 1. ГОСТ Р 71476-2024 ИИ. Концепции и терминология искусственного интеллекта
- 2. Предварительная обработка данных в машинном обучении: инструкция, инструменты, полезные ресурсы для начинающих: <https://habr.com/ru/companies/skillfactory/articles/848858/>
- 3. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166> (дата обращения: 17.06.2020).

ML-2:

- 1. Системы искусственного интеллекта. Классификация алгоритмов и вычислительных методов ПНСТ953-2024
- 2. Системы искусственного интеллекта. Классификация систем искусственного интеллекта ГОСТ Р 59277-2020
- 3. Искусственный интеллект. Структура описания систем искусственного интеллекта использующих машинное обучение ПНСТ838-2023/ИСО/МЭК 23053:2022
- 4. *Машинное обучение*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/mashinnoye-obucheniye>
- 5. *Общие понятия МО* // Викиконспекты NEERC IFMO. Электронный ресурс. URL: https://neerc.ifmo.ru/wiki/index.php?title=Общие_понятия

ML-3:

1. *Ансамбли в машинном обучении*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/ansambli-v-mashinnom-obuchenii>. Яндекс Образование
2. *Градиентный бустинг*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/gradientnyj-busting> (дата обращения: 27.11.2025). Яндекс Образование
3. *XGBoost*. Электронный ресурс. NEERC IFMO Wiki. URL: <https://neerc.ifmo.ru/wiki/index.php?title=XGBoost>
4. *Временные ряды*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/vremennyye-ryady>. Яндекс Образование
5. *Аналитика временных рядов*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/analitika-vremennyh-ryadov>
6. *Модели вида ARIMA*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/modeli-vida-arima>

Альянса ИИ:

1. **Машинное обучение (Яндекс.Образование)** *Машинное обучение*. Электронный ресурс. Яндекс.Образование. URL: <https://education.yandex.ru/handbook/ml/article/mashinnoye-obucheniye>
2. **СТОХАСТИЧЕСКИЙ ГРАДИЕНТНЫЙ СПУСК (NEERC IFMO)** *Стохастический градиентный спуск*. Электронный ресурс. NEERC IFMO Wiki. URL: https://neerc.ifmo.ru/wiki/index.php?title=Стохастический_градиентный_спуск
3. **Хабр — Машинное обучение (категория)** *Машинное обучение — подборка статей*. Электронный ресурс. Хабр. URL: https://habr.com/ru/hubs/machine_learning/articles/
4. **Документация imbalanced-learn: SMOTE** *SMOTE — Synthetic Minority Over-sampling Technique*. Электронный ресурс. imbalanced-learn Documentation. URL: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html

1. **Хабр** — «Пошаговое введение в SMOTE» Коновалов А. *Как работает SMOTE: создание синтетических данных*. Электронный ресурс. Хабр. URL: <https://habr.com/ru/articles/545380/>
2. **sklearn: LinearRegression** *LinearRegression* — линейная регрессия. Электронный ресурс. Scikit-learn Documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
3. **Хабр** — **Линейная регрессия простым языком** Павлов С. *Линейная регрессия: подробное объяснение*. Электронный ресурс. Хабр. URL: <https://habr.com/ru/articles/514818/>
4. **sklearn: KNeighborsClassifier** *KNeighborsClassifier* — метод k ближайших соседей. Электронный ресурс. Scikit-learn Documentation. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
5. **Хабр** — **Метод ближайших соседей (подробно)** Назаров М. *KNN: детальный разбор алгоритма*. Электронный ресурс. Хабр. URL: <https://habr.com/ru/articles/801885/>
6. **sklearn: LogisticRegression** *LogisticRegression* — логистическая регрессия. Электронный ресурс. Scikit-learn Documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
7. **sklearn: Naive Bayes** *Naive Bayes* — наивный байесовский классификатор. Электронный ресурс. Scikit-learn Documentation. URL: https://scikit-learn.org/stable/modules/naive_bayes.html
8. **Хабр** — **Наивный Байес простыми словами** Шаров И. *Наивный байесовский классификатор: интуитивное объяснение*. Электронный ресурс. Хабр. URL: <https://habr.com/ru/articles/415963/>
9. **sklearn: Decision Trees** *Decision Trees* — деревья решений. Электронный ресурс. Scikit-learn Documentation. URL: <https://scikit-learn.org/stable/modules/tree.html>
10. **Хабр** — **Деревья решений** Прокопьев А. *Деревья решений: как работают и где применять*. Электронный ресурс. Хабр. URL: <https://habr.com/ru/articles/322034/>

5) Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины, к электронным библиотечным системам, профессиональным базам данных и информационным справочным системам. А также рекомендованные ОТКРЫТЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ (на платформе LMS Moodle (elearning.volgatech.net), интегрированного с федеральной платформой ГИС СЦОС online.edu.ru открытый образовательный портал (mooped.net) и ресурсы открытых образовательных платформ Яндекс-Практикум, Сберуниверситет.

Рекомендуемые образовательные ресурсы

Базовые курсы по Python

1. **"Поколение Python": курс для начинающих** (<https://stepik.org/course/58852/>)
 - Основы синтаксиса Python
 - Работа со встроенными типами данных
 - Функции и модули
1. **"Поколение Python": курс для продвинутых** (<https://stepik.org/course/68343/>)
 - ООП в Python
 - Работа с файлами
 - Обработка исключений

Курсы по машинному обучению и искусственному интеллекту

1. **"Введение в Data Science и машинное обучение"** (<https://stepik.org/course/4852/>)
 - Основы анализа данных
 - Классические алгоритмы ML

- Библиотеки NumPy, pandas, scikit-learn

1. "Нейронные сети и обработка текста" (<https://stepik.org/course/54098/>)

- Архитектуры нейронных сетей
- Natural Language Processing
- Практика на реальных задачах

1. "Введение в глубокое обучение" от HSE (<https://stepik.org/course/135003/>)

- Теория глубокого обучения
- PyTorch и TensorFlow
- Компьютерное зрение и NLP

Специализированные курсы

1. "MLOps: практики и инструменты" (<https://stepik.org/course/181476/>)

- Версионирование моделей и данных
- CI/CD для ML-проектов
- Мониторинг моделей в продакшене

Интерактивные платформы для практики

1. **LeetCode** (<https://leetcode.com/>) — алгоритмические задачи на Python
2. **HackerRank** (<https://www.hackerrank.com/domains/python>) — задачи по программированию и ML
3. **Kaggle Competitions** (<https://www.kaggle.com/competitions>) — соревнования по анализу данных

Раздел 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющих в библиотеке, или электронный адрес издания (ресурса) в сети Интернет
УЧЕБНЫЕ, УЧЕБНО-МЕТОДИЧЕСКИЕ И НАУЧНЫЕ ИЗДАНИЯ		
1.	Хахаев, И. А. Практикум по алгоритмизации и программированию на Python [Электронный ресурс] / Хахаев И. А. 2-е изд. Москва: ИНТУИТ, 2016. - 178 с.	https://e.lanbook.com/book/100377
2.	Сузи, Р. А. Язык программирования Python [Электронный ресурс] / Сузи Р. А. 2-е изд. Москва: ИНТУИТ, 2016. - 350 с. ISBN 5-9556-0058-2.	https://e.lanbook.com/book/100546
3.	Искусственный интеллект в интеллектуальных системах управления биотехнологическими робототехническими системами [Электронный ресурс] : Учебное пособие / Шафрай А. В. Москва: Ай Пи Ар Медиа, 2023. - 217 с. ISBN 978-5-4497-2071-9.	https://www.iprbookshop.ru/128381.html

4.	Макшанов, А. В. Большие данные. Big Data [Электронный ресурс] : учебник для вузов / Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н.; Журавлев А. Е., Тындыкарь Л. Н. 3-е изд., стер. Санкт-Петербург: Лань, 2023. - 188 с. ISBN 978-5-507-46866-9.	https://e.lanbook.com/book/322664
5.	Цехановский, В. В. Архитектор информационных систем: как проектировать большие системы [Электронный ресурс] / Цехановский В. В. Москва, Алматы: Ай Пи Ар Медиа, EDP Hub (Идипи Хаб), 2024. - 284 с. ISBN 978-5-4497-2315-4.	https://www.iprbookshop.ru/133454.html

6.2. Материально-техническая база и программное обеспечение

№№ п/п	Аудитории для проведения учебных занятий, самостоятельной работы и проведения государственной итоговой аттестации	Перечень основного оборудования	Программное обеспечение
1.	333б (II)	Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач
2.	333г (III)	Измерительный прибор "BerCut-E" (1), Комплекс ПАИК/77100/КПВ (1), Комплект дополнит.оборудования к ПАИК/7710/КПВ(автогенератор AnCom и автоответчи (1), Компьютер P4-3.0/2*256Mb/HDD 200Gb/128 6600GT/DVD-RW/KM/FDD/MBi945P/UPS (1), Ксерокс Canon FC-860 (1), Междисциплинарная лабораторная платформа в комплекте с аппаратно-программным кон (10), Монитор 19"Samsung 940N (LKSB) TFT (1), Принтер HP Laser Jet 1100 (1), Систем.блок Core2 DUOE6300/1024Mb*2/320Gb/DVD-RW/клав.мышь.ковр. (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач

3.	431 (III)	Доска маркерная 120x240 см. ауд.431 (2), Кабель VGA 30.5 М KRAMER (П-П) (1), Проектор мультимедийный Hitachi CP-X8150 (1), Экран настенный с электроприводом 400x300см (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач
4.	433 (III)	Акустический комплект с микшером CRATE PA 8FX (1), Микшерный пульт компактный малошумящий Behringer XENYX 1202FX (1), Подавитель шума ALTAIR NG-440 4-х канальный (1), Экран настенный с электроприводом 400*300 Braun (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач

Раздел 7. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ/ ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценивания индикаторов достижения компетенций направлены на:

- усвоение теоретического материала (объем знаний, глубина усвоения), предусмотренного рабочей программой;
- умение излагать материал (четкость, грамотность изложения материала, точность и полнота воспроизведения учебного материала);
- умение применять теоретические знания при решении практических заданий.

Шкала оценивания представлена ниже.

Уровень сформированности элементов компетенции	Критерии оценивания	Шкала оценивания
Пороговый уровень	Обучающийся имеет знания основного материала, проявляет умение логично его излагать, но может допускать неточности в изложении материала, недостаточно правильные формулировки, испытывает затруднения в выполнении практических заданий	Зачтено

7.1. Текущий контроль успеваемости

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины (модуля) и производится с применением технологии рейтингового контроля в соответствии с технологической картой дисциплины. Порядок составления технологической карты и алгоритм проведения процедуры оценивания видов деятельности обучающихся, направленных на освоение знаний, умений, навыков и/или опыта деятельности, по накопительной системе в баллах устанавливается положением о системе РИТМ в ФГБОУ ВО «ПГТУ»

7.2. Промежуточная аттестация обучающихся

Промежуточная аттестация обучающихся направлена на оценивание результатов обучения по дисциплине (модулю) и проводится с использованием фондов оценочных средств.

Примеры типовых контрольных заданий из базы фонда оценочных средств по образовательной программе.

Множественный выбор:

ML-2:

1. Укажите типы задач машинного обучения, в которых используется размеченный набор данных с известными целевыми значениями.

a) Классификация b) Регрессия c) Кластеризация d) Обучение с подкреплением

2. Определите методы заполнения пропущенных значений, которые используют статистические характеристики признака.

a) Импутация медианой признака b) Импутация средним значением признака c) Удаление всех строк с пропусками d) Заполнение нулями без анализа данных

3. Выберите метрики качества, которые используются для оценки моделей классификации в машинном обучении.

a) Accuracy b) Precision c) Mean Squared Error (MSE) d) R-squared (коэффициент детерминации)

4. Укажите ансамблевые методы машинного обучения, которые используют множество базовых моделей для улучшения качества предсказаний.

a) Random Forest b) Gradient Boosting c) Линейная регрессия d) Одиночное дерево решений

5. Определите методы понижения размерности, которые используются для визуализации и сжатия многомерных данных.

a) PCA (Principal Component Analysis) b) t-SNE (t-distributed Stochastic Neighbor Embedding) c) Нормализация признаков (Min-Max Scaling) d) Стандартизация (Z-score normalization)

BD-1:

6. Выберите источники данных, которые относятся к открытым ресурсам для задач машинного обучения.

a) UCI Machine Learning Repository b) Kaggle Datasets c) Внутренние корпоративные базы данных компаний d) Коммерческие API с платным доступом

Установление последовательности:

BD-1:

7. Расположите этапы жизненного цикла ML-модели в правильном порядке от начала до развертывания.

a) Сбор и подготовка данных (очистка, feature engineering) b) Обучение модели на тренировочных данных c) Оценка качества модели на тестовых данных d) Развертывание модели в продакшн

8. Установите последовательность шагов при очистке табличных данных от выбросов методом IQR.

a) Вычисление первого (Q1) и третьего (Q3) квартилей b) Вычисление межквартильного размаха $IQR = Q3 - Q1$ c) Определение границ выбросов ($Q1 - 1.5 \cdot IQR$, $Q3 + 1.5 \cdot IQR$) d) Удаление или обработка значений за пределами границ

ML-3:

9. Расположите этапы применения алгоритма K-means для кластеризации данных в порядке их выполнения.

a) Инициализация K центроидов кластеров b) Присвоение каждой точки ближайшему центроиду c) Пересчёт позиций центроидов как среднего точек кластера d) Проверка сходимости и повторение до стабилизации

10. Установите последовательность действий при построении и оценке модели Random Forest в Scikit-Learn.

a) Разделение данных на тренировочную и тестовую выборки b) Создание и обучение модели Random Forest c) Предсказание на тестовых данных и вычисление метрик d) Анализ важности признаков (feature_importances_)

11. Расположите шаги применения метода главных компонент (PCA) для понижения размерности в правильном порядке.

a) Стандартизация исходных признаков b) Вычисление ковариационной матрицы данных c) Нахождение собственных векторов и собственных значений d) Проецирование данных на главные компоненты

12. Установите последовательность этапов cross-validation (кросс-валидации) для оценки качества модели.

a) Разделение датасета на K равных фолдов b) Обучение модели на K-1 фолдах (тренировочная выборка) c) Оценка модели на оставшемся фолде (валидационная выборка) d) Усреднение метрик качества по всем итерациям

Установление соответствия:

ML-2, ML-3:

13. Установите соответствие между типами задач машинного обучения и их характеристиками.

Типы задач:

1. Обучение с учителем (Supervised Learning)
2. Обучение без учителя (Unsupervised Learning)
3. Обучение с подкреплением (Reinforcement Learning)

Характеристики: a) Использование размеченных данных с известными целевыми переменными b) Выявление скрытых структур в неразмеченных данных c) Обучение агента через систему вознаграждений и наказаний d) Построение генеративных моделей изображений e) Оптимизация гиперпараметров нейронных сетей

14. Установите соответствие между задачами машинного обучения и примерами их применения.

Задачи:

1. Классификация
2. Регрессия
3. Кластеризация

Примеры применения: a) Определение является ли email спамом или нет b) Предсказание стоимости недвижимости по характеристикам c) Сегментация клиентов интернет-магазина по поведению d) Перевод текста с одного языка на другой e) Генерация синтетических изображений лиц

15. Установите соответствие между метриками качества и типами задач, для которых они применяются.

Метрики:

1. Accuracy, Precision, Recall
2. Mean Squared Error (MSE), R-squared

3. Silhouette Score, Davies-Bouldin Index

Типы задач: а) Оценка качества моделей классификации б) Оценка качества моделей регрессии с) Оценка качества кластеризации d) Оценка скорости сходимости градиентного спуска е) Измерение информационного прироста в деревьях решений

16. Установите соответствие между методами обработки данных и их назначением.

Методы:

1. Стандартизация (Z-score normalization)
2. Нормализация (Min-Max Scaling)
3. Импутация

Назначение: а) Преобразование признаков к нулевому среднему и единичной дисперсии б) Масштабирование значений признака в диапазон [0, 1] с) Заполнение пропущенных значений в данных d) Удаление выбросов из датасета е) Создание новых признаков из существующих

17. Установите соответствие между алгоритмами машинного обучения и их ключевыми особенностями.

Алгоритмы:

1. Логистическая регрессия
2. Дерево решений
3. K-means

Особенности: а) Использование сигмоидной функции для бинарной классификации б) Рекурсивное разбиение пространства признаков правилами с) Итеративная группировка данных в K кластеров по близости d) Построение линейной комбинации признаков для регрессии е) Использование ядерных функций для нелинейного разделения

18. Установите соответствие между этапами предварительной обработки данных и их результатами.

Этапы:

1. Удаление дубликатов
2. Обработка выбросов методом Z-score
3. Feature Engineering

Результаты: а) Устранение полностью идентичных записей в датасете б) Выявление значений, отклоняющихся более чем на 3σ от среднего с) Создание новых признаков из комбинаций существующих d) Преобразование категориальных признаков в числовые е) Разделение данных на обучающую и тестовую выборки

19. Установите соответствие между параметрами алгоритма Random Forest и их влиянием на модель.

Параметры:

1. n_estimators (количество деревьев)
2. max_depth (максимальная глубина дерева)
3. min_samples_split (минимум примеров для разбиения узла)

Влияние: а) Определяет количество деревьев в ансамбле б) Ограничивает максимальную глубину каждого дерева с) Задаёт минимальное количество примеров для создания нового разбиения d) Контролирует скорость обучения модели е) Определяет количество признаков для построения каждого дерева

20. Установите соответствие между компонентами рекомендательных систем и методами машинного обучения.

Компоненты:

1. Контентная фильтрация (Content-based)
2. Коллаборативная фильтрация (Collaborative)
3. Гибридный подход

Методы: а) Рекомендации на основе характеристик объектов и профиля пользователя б) Анализ поведения множества пользователей для поиска схожих паттернов с) Комбинирование контентного и коллаборативного подходов d) Использование обучения с подкреплением для оптимизации кликов е) Применение генеративных моделей для создания новых рекомендаций

Перечень вопросов для проведения промежуточной аттестации

Блок 1. Основы машинного обучения и интеллектуальных систем (Модуль 1)

4. Определите машинное обучение. Назовите три основных типа задач машинного обучения и приведите пример каждой.
5. Что такое жизненный цикл модели машинного обучения? Опишите основные этапы.
6. Дайте определение искусственного интеллекта. В чём различие между слабым и сильным ИИ?
7. Перечислите основные библиотеки Python для машинного обучения. Какую роль играет Scikit-Learn?
8. Что такое обучение с подкреплением? Приведите пример реальной задачи, решаемой этим методом.
9. Сравните обучение с учителем и обучение без учителя по структуре данных и целям.
10. Что такое экспертные системы? Назовите их основные компоненты.
11. Опишите роль данных в машинном обучении. Что такое размеченные и неразмеченные данные?
12. Какие метрики используются для оценки качества моделей классификации? Приведите формулы.
13. Что такое переобучение и недообучение? Как их диагностировать и предотвращать?

Блок 2. Подготовка и анализ данных (Модуль 2)

14. Перечислите этапы подготовки данных для машинного обучения. Почему очистка данных важна?
15. Опишите методы заполнения пропусков в данных (импутация). В чём преимущества и недостатки каждого?
16. Что такое выбросы? Назовите статистические методы их обнаружения (Z-оценка, IQR).
17. Объясните суть понижения размерности. Для чего применяется метод главных компонент (PCA)?
18. Что такое признаковая инженерия? Приведите примеры создания новых признаков.
19. Как визуализировать многомерные данные после применения t-SNE или PCA?
20. Опишите процесс разведочного анализа данных (EDA). Какие библиотеки используются?
21. Что такое нормализация и стандартизация признаков? В каких случаях применяется каждый метод?
22. Как оценить корреляцию между признаками? Что делать при высокой мультиколлинеарности?
23. Назовите источники открытых датасетов. Как проверить качество собранных данных?

Блок 3. Классические алгоритмы и методы оценки (Модуль 3 + Модуль 4)

24. Опишите алгоритм линейной регрессии. Как интерпретировать коэффициенты модели?
25. Что такое логистическая регрессия? Как она применяется для бинарной классификации?

26. Объясните принцип работы метода k-ближайших соседей (KNN). Как выбрать оптимальное k?
27. Как устроен алгоритм дерева решений? Что такое энтропия и прирост информации?
28. Что такое случайный лес? В чём его преимущество перед одним деревом решений?
29. Опишите алгоритм K-средних. Как определить оптимальное число кластеров?
30. Что такое градиентный спуск? Назовите его разновидности и области применения.
31. Объясните кросс-валидацию. Почему она предпочтительнее простого разделения на train/test?
32. Сравните метрики ROC-AUC и F1-score. В каких задачах какая метрика предпочтительна?
33. Что такое регуляризация (L1, L2)? Как она помогает бороться с переобучением?

Блок 4. Применение ML: задачи ранжирования и интеллектуальные системы (Модуль 4)

34. Сформулируйте задачу ранжирования. В чём суть pointwise-подхода?
35. Как превратить задачу ранжирования в задачу классификации или регрессии?
36. Опишите применение машинного обучения в рекомендательных системах.
37. Что такое интеллектуальные информационные системы? Приведите примеры.
38. Как машинное обучение интегрируется в системы управления (например, автономные роботы)?
39. Назовите этические проблемы машинного обучения (bias, fairness). Как их минимизировать?
40. Что такое интерпретируемость моделей? Почему она важна в интеллектуальных системах?
41. Опишите роль OLAP и BI-систем в анализе данных. Чем они отличаются от ML?
42. Как используются онтологии в интеллектуальных системах? Приведите пример.
43. Сравните классические алгоритмы ML и нейронные сети по применимости и требованиям к данным