
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол №2

**Рабочая программа дисциплины (модуля)
«Машинное обучение»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Прикладной искусственный интеллект

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2026

**Москва
2026**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения	5
3. Тематический план	10
4. Содержание дисциплины (модуля)	10
5. Учебно-методическое обеспечение	11
6. Материально-техническое обеспечение	11
7. Методические и оценочные материалы	13

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Машинное обучение» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Машинное обучение» позволяет студентам научиться разрабатывать модели, способные анализировать большие объемы данных и делать предсказания, что находит применение в различных отраслях. Кроме того, знание методов машинного обучения способствует автоматизации процессов и улучшению принятия решений.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект» и входит в обязательную часть Блока 1.

Дисциплина (модуль) изучается на 2 курсе в 3 семестрах, доступна к изучению после успешного освоения дисциплин (модулей): «Введение в статистику», «Искусственный интеллект».

Цель изучения дисциплины (модуля): заключается в формирование у студентов теоретических знаний и практических навыков по основам машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

Задачи изучения дисциплины (модуля):

- освоение основ теории и алгоритмов машинного обучения;
- развитие умений работы с данными и построения моделей;
- формирование навыков программирования и критического мышления.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- ключевые понятия и термины (например, обучающая и тестовая выборки, переобучение, регуляризация);
- различные типы машинного обучения (обучение с учителем, обучение без учителя, обучение с подкреплением);
- основные алгоритмы машинного обучения (линейные модели, деревья решений, ансамбли моделей, нейронные сети и др.);
- принципы работы алгоритмов и их применимости к различным типам данных;
- методы оценки производительности моделей (точность, полнота, F-мера, ROC-AUC и др.);
- принципы кросс-валидации и настроек гиперпараметров;
- популярные библиотеки для машинного обучения (например, Scikit-Learn, TensorFlow, Keras, PyTorch);

— основы работы с языком программирования Python и его библиотеками для работы с данными (NumPy, Pandas, Matplotlib).

уметь:

- собирать, очищать и преобразовывать данные для машинного обучения;
- визуализировать данные для выявления закономерностей и аномалий;

- выбирать и применять подходящие алгоритмы машинного обучения для решения конкретных задач;
- настраивать гиперпараметры моделей и проводить их оценку;
- интерпретировать результаты работы моделей и делать выводы на основе полученных данных;
- выявлять и устранять проблемы, такие как переобучение и недообучение.

владеть навыками:

- использования библиотек для анализа данных и машинного обучения в языке программирования Python;
- критического анализа результатов и обоснования выбора алгоритмов и методов, используемых в проекте;
- самостоятельного и систематического обновления профессиональных знаний в области машинного обучения на основе анализа актуальных исследований и технологических тенденций.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Код и содержание компетенции	Индикатор компетенции и перечень планируемых результатов обучения по дисциплине (модулю)
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1. Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач.
	УК-2.2. Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения.
	УК-2.3. Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями.
УК-3. Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде (соответствует УНК-02)	УК-3.1. (УНК-02.01.) Принимает участие в групповом взаимодействии в ходе профессиональной деятельности
	УК-3.2. (УНК-02.02.) Проявляет лидерство и осуществляет наставничество
УК-4. Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах) (соответствует УНК-01)	УК-4.1. (УНК-01.01.) Осуществляет коммуникации в профессиональной деятельности УНК-01.01.01. У-3 Способен формулировать и понимать технологические и бизнес-требования
	УК-4.2. (УНК-01.02.) Демонстрирует владение профессиональной культурой УНК-01.02.01. У-3. Умеет грамотно оформлять документацию и вести коммуникацию в соответствии с профессиональными стандартами
УК-6. Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни (соответствует УНК-04)	УК-6.1. (УНК-04.01.) Планирует и осуществляет самообучение УНК-04.01.01. У-1. Умеет самостоятельно определять области для профессионального развития и формулировать цели обучения УНК-04.01.01. У-2. Способен разрабатывать план самообучения, выбирать подходящие источники и методы обучения УНК-04.01.01. У-3. Умеет организовывать свое время для регулярного повышения квалификации УНК-04.01.01. У-4. Способен оценивать эффективность проведенного обучения и корректировать план при необходимости УНК-04.01.01. 3-1. Знает современные методы и ресурсы для самостоятельного обучения УНК-04.01.01. 3-2. Понимает принципы постановки целей и планирования личностного развития УНК-04.01.01. 3-3. Знает основы саморегуляции и мотивации для поддержания постоянного профессионального роста

ОПК-4. Способен находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем	ОПК-4.1. Знает основные классы математических алгоритмов, методы их анализа, критерии выбора и оценки вычислительной сложности, а также принципы программной реализации.
	ОПК-4.2. Умеет находить и анализировать математические алгоритмы, выбирать подходящий алгоритм для поставленной задачи и реализовывать его с использованием современных вычислительных систем.
	ОПК-4.3. Имеет практический опыт программной реализации, тестирования и оценки корректности и эффективности математических алгоритмов.
ОПК-6. Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения	ОПК-6.1. Знает принципы алгоритмизации, основные структуры данных, парадигмы программирования, методы разработки, тестирования и отладки компьютерных программ.
	ОПК-6.2. Умеет разрабатывать, документировать, тестировать и отлаживать алгоритмы и компьютерные программы, пригодные для практического применения, с использованием современных языков и сред разработки.
	ОПК-6.3. Имеет практический опыт разработки и сопровождения компьютерных программ и программных комплексов для решения прикладных задач.
МО-1. Способен проектировать, разрабатывает и улучшать алгоритмы и модели машинного обучения	МО-1.1. Применяет классические методы и модели машинного обучения (МО) (базовый) МО-1.1 3-1. Знает алгоритмы обучения с учителем для регрессии и классификации. Линейные методы, ансамблевые методы, методы на опорных векторах. МО-1.1 3-1. Знает алгоритмы обучения без учителя. Методы кластеризации и поиска аномалий. МО-1.1 3-1. Знает основные ограничения классического МО и области применения. МО-1.1 У-1. Умеет применять стандартные библиотеки для обучения с учителем и без учителя. МО-1.1 У-2. Умеет использовать инструменты очистки данных и предварительной подготовки данных для статистического обучения (масштабирование, корреляционный анализ, заполнение пропусков в данных). МО-1.1 У-3. Умеет самостоятельно строить пайплайн очистки данных, обучения, подбора гиперпараметров и валидации моделей МО.
	МО-1.2. Применяет глубокие нейронные сети (средний) МО-1.2. 3-1. Знает теоретические основы нейросетевых алгоритмов. МО-1.2. 3-2. Знает методы обучения нейронных сетей с помощью градиентного спуска и обратного распространения ошибки. МО-1.2. 3-3. Знает популярные архитектуры глубоких нейронных сетей и их применение: сверточные, рекуррентные, сети с памятью. МО-1.2. 3-4. Знает механизм внимания в нейронных сетях и архитектуру трансформера. МО-1.2. У-1. Умеет применять стандартные библиотеки, реализующие основные архитектуры глубокого обучения. МО-1.2. У-2. Умеет разрабатывать нейронные сети с вниманием. МО-1.2. У-3. Умеет использовать предварительно обученные нейронные сети и перенос обучения.

	МО-1.3. Применяет автоматическое машинное обучение (средний) МО-1.3. 3-1. Знает алгоритмы автоматизации работы с признаками. МО-1.3. 3-1. Знает алгоритмы автоматической настройки гиперпараметров. МО-1.3. 3-1. Знает алгоритмы автоматического поиска структуры модели (нейронных сетей и других графовых моделей). МО-1.3. У-1. Умеет обосновывать способы и варианты применения алгоритмов автоматического машинного обучения в задачах ИИ, включая их преобразование и адаптацию к специфике задачи МО-1.3. У-1. Умеет применять методы повышения устойчивости, надежности, безопасности алгоритмов автоматического машинного обучения для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ
ППК-ДА2. Способен подготавливать и обрабатывать данные для аналитических исследований	ППК-ДА2.1. – Очищает, нормализует и преобразует данные для анализа (базовый) ППК-ДА2.1. 3-1. Знает методы предобработки структурированных и неструктурированных данных ППК-ДА2.1. 3-2. Знает алгоритмы нормализации и стандартизации данных ППК-ДА2.1. 3-3. Знает техники обработки пропущенных значений и выбросов ППК-ДА2.1. У-1. Умеет применять методы очистки данных в зависимости от их типа ППК-ДА2.1. У-2. Умеет преобразовывать данные в требуемые форматы
	ППК-ДА2.2. – Обеспечивает соответствие данных требованиям заказчика (базовый) ППК-ДА2.2. 3-1. Знает критерии качества данных для различных аналитических задач ППК-ДА2.2. 3-2. Знает отраслевые стандарты представления данных ППК-ДА2.2. 3-3. Знает методы верификации и валидации данных ППК-ДА2.2. У-1. Умеет проверять данные на соответствие спецификациям ППК-ДА2.2. У-2. Умеет разрабатывать чек-листы контроля качества данных
	ППК-ДА2.3. – Формирует датасеты, пригодные для моделирования (средний) ППК-ДА2.2. 3-1. Знает принципы формирования обучающих и тестовых выборок ППК-ДА2.2. 3-2. Знает методы балансировки данных ППК-ДА2.2. 3-3. Знает техники аугментации данных ППК-ДА2.2. У-1. Умеет создавать репрезентативные выборки данных ППК-ДА2.2. У-2. Умеет применять методы семплирования

ППК-ДА3. Способен проводить аналитические исследования с применением технологий больших данных	ППК-ДА3.1. – Применяет методы машинного обучения и статистического анализа (базовый) ППК-ДА3.1. 3-1. Знает основные алгоритмы машинного обучения ППК-ДА3.1. 3-2. Знает методы статистического анализа данных ППК-ДА3.1. 3-3. Знает критерии выбора алгоритмов для различных задач ППК-ДА3.1. У-1. Умеет реализовывать алгоритмы машинного обучения ППК-ДА3.1. У-2. Умеет интерпретировать результаты статистического анализа
	ППК-ДА3.2. – Обеспечивает соответствие результатов анализа бизнес-задачам заказчика (средний) ППК-ДА3.2. 3-1. Знает методы перевода бизнес-требований в аналитические задачи ППК-ДА3.2. 3-2. Знает ключевые бизнес-метрики в предметной области ППК-ДА3.2. 3-3. Знает принципы интерпретации результатов для бизнес-пользователей ППК-ДА3.2. У-1. Умеет адаптировать аналитические модели под бизнес-потребности ППК-ДА3.2. У-2. Умеет оценивать экономический эффект от аналитических решений
	ППК-ДА3.3. – Подготавливает отчеты и визуализации для презентации результатов (средний) ППК-ДА3.3 3-1. Знает принципы эффективной визуализации данных ППК-ДА3.3 3-2. Знает инструменты создания аналитических отчетов ППК-ДА3.3 3-3. Знает методы сторителлинга на основе данных ППК-ДА3.3 У-1. Умеет выбирать оптимальные типы визуализации ППК-ДА3.3 У-2. Умеет создавать интерактивные дашборды
МО-2. Способен проектировать и внедрять системы автоматизации жизненного цикла машинного обучения (MLOps)	МО-2.1. Проектирует и настраивает CI/CD-пайплайны для машинного обучения (базовый) МО-2.1. 3-1. Знает принципы CI/CD применительно к моделям машинного обучения (тестирование, сборка, развертывание). МО-2.1. 3-2. Знает методы интеграции инструментов MLOps в существующую инфраструктуру. МО-2.1. У-1. Умеет настраивать CI/CD-пайплайны для МО с использованием инструментов.
	МО-2.2. Организует версионирование моделей, данных и экспериментов (средний) МО-2.2. 3-1. Понимает системы версионирования моделей и данных. МО-2.2. У-1. Умеет работать с системами версионирования. МО-2.2. У-2. Умеет использовать инструменты MLOps для воспроизводимости и масштабируемости экспериментов.

	МО-2.3. Внедряет системы мониторинга моделей в продуктивную эксплуатацию (средний) МО-2.3. З-1. Знает метрики мониторинга моделей МО. МО-2.3. У-1. Умеет развертывать системы мониторинга моделей МО. МО-2.3. У-2. Умеет настраивать сообщения от мониторинга (altering).
	МО-2.4. Автоматизирует процессы переобучения и развертывания моделей (продвинутый) МО-2.4. З-1. Знает стратегии автоматизации переобучения МО-2.4. У-1. Умеет автоматизировать переобучение моделей и А/В-тестирование развертываний
УНК-03 – Способен применять методы стратегического планирования в профессиональной деятельности	УНК-03.01. Демонстрирует системное мышление при решении задач профессиональной деятельности (базовый) УНК-03.01.01. У-1. Умеет анализировать задачи, учитывая взаимосвязи между их компонентами.
	УНК-03.02. Осуществляет анализ и планирование деятельности (базовый) УНК-03.02.01. У-2. Способен разрабатывать планы действий с учетом поставленных целей и имеющихся ресурсов.

3. Тематический план

№п/ п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятел ьная работа	
		Лекции	Практически е занятия			
1	Основные термины машинного обучения и смежных областей. Оценка качества моделей	8	8		8	Домашнее задание
2	Линейные модели	12	12		12	Домашнее задание
3	Деревья решений и модели, основанные на них	20	24		18	Домашнее задание Коллоквиум Проект
4	Обучение без учителя, работа с признаками	20	26		18	Домашнее задание Коллоквиум
	Экзамен			4		
Итого:		60	70	4	56	
Объем дисциплины (модуля) (в ак. ч.)		190				
Объем дисциплины (модуля) (в зач. ед.)		5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Основные термины машинного обучения и смежных областей. Оценка качества моделей	Метрики и функционалы потерь в задачах регрессии и классификации. Контроль качества и выбор модели. Метрические алгоритмы
2	Линейные модели	Линейная регрессия. Логистическая регрессия, метод опорных векторов
3	Деревья решений и модели, основанные на них	Решающие деревья. Сложность алгоритмов. Ансамбли алгоритмов Случайные леса. Градиентный бустинг. Аплифт-моделирование
4	Обучение без учителя, работа с признаками	Обучение без учителя. Кластеризация и методы понижения размерности. Обнаружение аномалий. Генерация и отбор признаков. Интерпретация моделей и диагностика сдвига данных: Частые вопросы на собеседованиях по темам машинного обучения

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. — 2-е изд. — Москва : ДМК Пресс, 2023. — 401 с. — ISBN 978-5-89818-300-4. — URL: <https://znanium.ru/catalog/product/2085038>.
2. Мэрфи, К. П. Вероятностное машинное обучение: введение : практическое руководство / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. — Москва : ДМК Пресс, 2023. — 990 с. — ISBN 978-5-93700-119-1. — URL: <https://znanium.ru/catalog/product/2109489>.
3. Шолле, Ф. Глубокое обучение на Python : практическое руководство / Ф. Шолле. - Санкт-Петербург : Питер, 2023. - 576 с. - (Библиотека программиста). - ISBN 978-5-4461-1909-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123389>.

Дополнительная литература:

1. Дьяконов, А.Г. Машинное обучение и анализ данных / А.Г. Дьяконов. — URL: https://github.com/Dyakonov/MLDM_BOOK/blob/main/README.md.
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>.
3. Введение в статистическое обучение с примерами на языке R / Джеймс Г. [и др.]. — 2-е изд., испр. — Москва : ДМК Пресс, 2018. — 455 с. : ил. — Библиогр. : с. 450—455.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		

Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Машинное обучение» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары и домашние задания, проект, коллоквиумы, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Проект – исследовательская работа по курсу и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Коллоквиум – устные ответы на вопросы, список которых известен студенту заранее.

В процессе подготовки к коллоквиуму необходимо проанализировать учебные материалы, ознакомившись с лекциями, учебниками и дополнительными источниками, акцентируя внимание на ключевых темах. Рекомендуется создать структурированные конспекты, выделяя основные идеи, термины и формулы.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Машинное обучение»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *экзамена*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	
8	Отлично	
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы,
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине, но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Машинное обучение» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	30%	10	Набор задач по темам недели
Коллоквиум	30%	3	Устные ответы на вопросы, список которых известен студенту заранее
Проект	20%	1	Исследовательская работа по дисциплине (модулю) и презентация результатов
Экзамен	20%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Машинное обучение»: « $0,3 \times \text{среднее за домашние задания} + 0,3 \times \text{среднее за коллоквиумы} + 0,2 \times \text{Проект} + 0,2 \times \text{Экзамен}$ ».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные вопросы к коллоквиуму

Коллоквиум 1: Линейные модели

Вопросы:

1. Что такое линейная регрессия и в каких задачах она применяется?
2. Как интерпретировать коэффициенты в линейной модели?
3. Что такое функция потерь в линейной регрессии? Какая функция используется чаще всего?
4. В чём разница между линейной и логистической регрессией?

5. Как логистическая регрессия преобразует линейную комбинацию признаков в вероятность?
6. Что такое S-образная (сигмоидная) функция и зачем она нужна?
7. Объясните, как работает метод опорных векторов (SVM) в задаче классификации.
8. Что такое зазор (margin) и почему SVM старается его максимизировать?
9. Какое влияние оказывает параметр C в SVM? Что происходит при слишком малом или слишком большом значении?
10. Зачем нужна регуляризация в линейных моделях? Чем отличаются L1 и L2 регуляризация?

Критерии оценивания:

- Точность и корректность определений и объяснений.
- Понимание математической и логической основы моделей.
- Умение объяснить, как и зачем применяются модели на практике.
- Способность сравнивать подходы и аргументировать выбор метода.
- Чёткость, логичность и структурированность ответа.

Коллоквиум 2: Деревья решений и ансамбли

Вопросы:

1. Как строится решающее дерево? Какие критерии используются для разбиения узлов?
2. Что такое прирост информации (information gain) и неоднородность Джини?
3. Почему решающие деревья склонны к переобучению? Какие способы борьбы существуют?
4. Как работает алгоритм случайного леса (Random Forest)?
5. Почему в случайном лесе используются подвыборки признаков при каждом разбиении?
6. Как устроен градиентный бустинг (например, XGBoost)? Чем он отличается от случайного леса?
7. Что такое остатки (residuals) и как они используются в бустинге?
8. Как ансамбли моделей повышают качество предсказаний по сравнению с отдельными деревьями?
9. Что такое аплифт-моделирование? В каких задачах оно применяется?
10. Приведите пример бизнес-задачи, где аплифт-модель полезнее обычной модели отклика.

Критерии оценивания:

- Понимание принципов построения деревьев и ансамблей.
- Умение объяснить разницу между Random Forest и Gradient Boosting.
- Способность описать, как избежать переобучения и оценить важность признаков.
- Понимание специфических подходов, таких как аплифт-моделирование.
- Ясность, логичность и правильное использование терминологии.

Коллоквиум 3: Обучение без учителя и работа с признаками

Вопросы:

1. Что такое обучение без учителя? Приведите примеры задач.
2. Как работает алгоритм k-means? Как он обновляет центры кластеров?
3. Какие проблемы есть у k-means? В каких случаях он работает плохо?
4. Как определить оптимальное число кластеров (например, метод локтя или силуэт)?
5. Что такое метод главных компонент (PCA)? Как он помогает при работе с данными?

6. Как интерпретировать главные компоненты и что такое доля объяснённой дисперсии?
7. Какие методы обнаружения аномалий вы знаете? Опишите один из них.
8. В чём идея изоляционного леса (Isolation Forest)? Почему он эффективен для аномалий?
9. Какие способы генерации и отбора признаков существуют?
10. Что такое сдвиг данных (data drift)? Почему его важно отслеживать в реальных системах?

Критерии оценивания:

- Понимание алгоритмов кластеризации, понижения размерности и обнаружения аномалий.
- Умение объяснить, зачем нужны методы работы с признаками.
- Осознание важности мониторинга моделей в продакшене (например, data drift).
- Способность привести практические примеры и объяснить логику работы методов.
- Чёткость, последовательность и грамотное использование терминов.

Примерное задание для проекта

Задание для проекта, кейс от АО «Лаборатория Касперского»

Студентам предлагается разработать ML-модуль для выявления подозрительных командных строк в SIEM-системе или ML-модуль первичной оценки событий безопасности на выбор:

1. Описание задания:

Представьте, что вы вошли в команду, разрабатывающую интеллектуальные функции для системы мониторинга информационной безопасности.

SIEM-система — это программная платформа, которая централизованно собирает события безопасности от компьютеров, серверов, сетевого оборудования и приложений. SIEM приводит события из разных источников к единому формату, хранит и сопоставляет их, выявляет потенциально опасные ситуации и формирует предупреждения для аналитиков информационной безопасности.

За короткий промежуток времени в SIEM могут поступать тысячи событий. Часть из них содержит сведения о запущенных процессах и командных строках. Командная строка показывает, какая программа была запущена и с какими параметрами. Некоторые команды могут указывать на попытку получить несанкционированный доступ, изменить настройки системы, запустить нежелательный процесс или скрыть подозрительную активность.

Проверить все события вручную невозможно. Поэтому ваша задача — разработать учебный ML-модуль, который будет автоматически оценивать командные строки и выделять записи, требующие внимания аналитика.

2. Этапы работы:

- *Подготовка данных* — загрузить набор обезличенных или синтетических командных строк, проверить данные и подготовить их для анализа.
- *Базовое правило* — разработать простой алгоритм, который выявляет подозрительные записи по заранее заданным признакам.
- *ML-модель* — преобразовать текст командных строк в признаки и обучить модель классифицировать их как обычные или подозрительные.
- *Оценка качества* — рассчитать метрики, изучить ошибочные прогнозы и сравнить ML-модель с решением на основе правил.
- *Прототип модуля* — реализовать программу, которая принимает командную строку, присваивает ей оценку риска и показывает, требуется ли проверка аналитика.

3. Используемые ресурсы:

- обезличенный или синтетический набор из 1 000–5 000 командных строк;
- учебная метка для каждой записи: «обычная» или «подозрительная»;
- описание полей набора данных;
- отдельная тестовая выборка, содержащая ранее не встречавшиеся варианты командных строк;
- шаблон Jupyter Notebook;
- Python 3;
- библиотеки Pandas, NumPy, scikit-learn и Matplotlib;
- компьютер не менее чем с 4 ГБ оперативной памяти;
- краткая справка о SIEM-системах и анализе событий безопасности.

4. Итоговая проверка перед сдачей:

- корректно загрузили и исследовали данные;
- проверили данные на пропуски, дубли и возможные ошибки разметки;
- разделили данные на обучающую и тестовую выборки;
- исключили попадание тестовых примеров в обучение;
- разработали понятное базовое правило;
- обучили хотя бы одну простую модель классификации;
- проверили модель на ранее не встречавшихся примерах;
- сравнили ML-модель с решением на основе правил;
- показали ситуации, в которых модель обнаруживает события, не найденные правилом;
- проанализировали ложные срабатывания и пропущенные события;
- подготовили воспроизводимый код и инструкцию по запуску.

5. По итогам выполнения проекта студенты должны представить:

- Jupyter Notebook с загрузкой и анализом данных;
- Результаты проверки данных на пропуски, дубли и ошибки;
- Описание признаков командных строк;
- Базовый классификатор на основе заданных вручную правил;
- ML-модель классификации;
- Сравнение результатов правила и ML-модели;
- Значения precision, recall и F1-score;
- Матрицу ошибок;
- Анализ не менее пяти ошибочных прогнозов;
- Примеры ранее не встречавшихся командных строк и результаты их оценки;
- Краткое объяснение ограничений решения;
- README с инструкцией по запуску;
- Презентацию на 5–7 слайдов.

Примерные домашние задания

Домашнее задание по теме «Основные термины машинного обучения и смежных областей. Оценка качества моделей»

Задание 1: Метрики и функционалы потерь

Найдите примеры функционалов потерь, используемых в задачах регрессии и классификации. Опишите их особенности и применения. Кроме того, объясните понятие средней гипотезы и ее связь с функционалами потерь.

Задание 2: Контроль качества и выбор модели

Опишите методы контроля качества моделей машинного обучения, такие как валидация и кросс-валидация. Explain понятие ошибки внутри и вне выборки и ошибки

обобщения. Кроме того, объясните неравенство Хёфдинга и его применение в контроле качества моделей.

Задание 3: Метрические алгоритмы

Найдите примеры метрических алгоритмов, используемых в задачах классификации и регрессии. Опишите их особенности и применения. Кроме того, объясните понятие размерности Вапника-Червоненкиса и ее связь с метрическими алгоритмами.

Домашнее задание по теме «Линейные модели»

Задание 1: Линейная регрессия

Дан набор данных, содержащий информацию о цене на жилье и площади квартир в городе. Используя линейную регрессию, построить модель, которая позволяет предсказать цену на жилье на основе площади квартиры.

- Найдите коэффициенты линейной регрессии (β_0 , β_1) используя метод наименьших квадратов.
- Оцените качество модели с помощью коэффициента детерминации (R^2).
- Дайте интерпретацию результатов и объясните, как можно использовать эту модель для предсказания цены на жилье.

Задание 2: Логистическая регрессия и SVM

Дан набор данных, содержащий информацию о клиентах банка и их кредитной истории. Используя логистическую регрессию и SVM, построить модели, которые позволяют предсказать вероятность того, что клиент вернет кредит.

- Найдите коэффициенты логистической регрессии (β_0 , β_1) используя метод максимального правдоподобия.
- Построить модель SVM с линейным ядром и оценить ее качество с помощью коэффициента точности.
- Дайте интерпретацию результатов и объясните, как можно использовать эти модели для предсказания вероятности возврата кредита.

Задание 3: Линейные модели

Дан набор данных, содержащий информацию о влиянии различных факторов на результаты студентов в университете. Используя линейные модели, построить модель, которая позволяет предсказать результаты студентов на основе следующих факторов: возраст, пол, количество посещенных занятий и средний балл по предметам.

- Найдите коэффициенты линейной регрессии (β_0 , β_1 , ..., β_n) используя метод наименьших квадратов.
- Оцените качество модели с помощью коэффициента детерминации (R^2).
- Дайте интерпретацию результатов и объясните, как можно использовать эту модель для предсказания результатов студентов.

Домашнее задание по теме: «Деревья решений и модели, основанные на них»

Задача 1. Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10).

Задача 2. Примените метод SVM (например, из библиотеки `sklearn`) для датасета `blobs2`. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF).

Задача 3. Найдите максимум функции с помощью алгоритма кросс-энтропийного поиска, изображая распределение на каждом шаге.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция	Индикатор
1.	В разделе «Основные термины» укажите метод определения круга задач ML.	Анализ требований	УК-2	УК-2.2-У-1
2.	При оценке качества укажите метрику для задачи регрессии.	MSE / MAE / RMSE	МО-1	МО-1.1-3-1
3.	В разделе «Основные термины» укажите метрику для бинарной классификации.	Accuracy / F1 / ROC-AUC	МО-1	МО-1.1-3-1
4.	При контроле качества укажите метод проверки модели на новых данных.	Кросс-валидация	МО-1	МО-1.1-У-3
5.	В разделе «Основные термины» укажите алгоритм на основе метрик.	k-NN / Метрические алгоритмы	МО-1	МО-1.1-3-1
6.	При анализе данных укажите метод интерпретации результатов статистики.	Статистический тест	ППК-ДА3	ППК- ДА3.1-У-2
7.	В разделе «Основные термины» укажите понятие компромисса смещения и разброса.	Bias-Variance Tradeoff	МО-1	МО-1.1-3-1
8.	При коммуникации укажите способ объяснения метрик бизнесу.	Визуализация / Отчет	УК-4	УК-4.1-У-3
9.	В разделе «Основные термины» укажите метод планирования оценки модели.	План валидации	УНК-3	УНК-3.1-У- 1
10.	При подготовке данных укажите метод обработки пропусков.	Imputation / Drop	ППК-ДА2	ППК- ДА2.1-У-1
11.	В разделе «Основные термины» укажите критерий соответствия данных бизнесу.	Бизнес-метрики	ППК-ДА3	ППК- ДА3.2-3-1
12.	При самообучении укажите источник изучения новых метрик.	Документация / Статьи	УК-6	УК-6.1-У-2
13.	В разделе «Линейные модели» укажите алгоритм для предсказания непрерывной величины.	Линейная регрессия	МО-1	МО-1.1-3-1
14.	При реализации укажите библиотеку для линейных моделей в Python.	scikit-learn	МО-1	МО-1.1-У-1
15.	В разделе «Линейные модели» укажите выход логистической регрессии.	Вероятность / 0 или 1	МО-1	МО-1.1-3-1
16.	При выборе алгоритма укажите метод разделения классов гиперплоскостью.	SVM / Метод опорных векторов	МО-1	МО-1.1-3-1
17.	В разделе «Линейные модели» укажите необходимость масштабирования признаков.	StandardScaler / Normalization	ППК-ДА2	ППК- ДА2.1-3-2

18.	При регуляризации укажите метод борьбы с переобучением в линейных моделях.	L1 / L2 регуляризация	МО-1	МО-1.1-3-1
19.	В разделе «Линейные модели» укажите принцип реализации алгоритма.	Градиентный спуск	ОПК-6	ОПК-6.1-3-1
20.	При подготовке данных укажите метод масштабирования для регрессии.	Z-score / MinMax	ППК-ДА2	ППК-ДА2.1-У-2
21.	В разделе «Линейные модели» укажите метод проверки значимости коэффициентов.	P-value / T-тест	ППК-ДА3	ППК-ДА3.1-3-2
22.	При планировании укажите способ изучения новых линейных методов.	Учебный план	УК-6	УК-6.1-У-1
23.	В разделе «Линейные модели» укажите метод подбора гиперпараметров.	GridSearch / CrossVal	МО-1	МО-1.3-3-1
24.	При ресурсах укажите фактор учета при обучении линейной модели.	Память / CPU	УНК-3	УНК-3.2-У-2
25.	В разделе «Деревья решений» укажите критерий разделения узла дерева.	Gini / Entropy / MSE	МО-1	МО-1.1-3-1
26.	При ансамблировании укажите метод построения случайного леса.	Bagging / Бутстрэппинг	МО-1	МО-1.1-3-1
27.	В разделе «Деревья решений» укажите алгоритм последовательного обучения моделей.	Градиентный бустинг	МО-1	МО-1.1-3-1
28.	При реализации укажите библиотеку для бустинга.	XGBoost / CatBoost / LightGBM	МО-1	МО-1.1-У-1
29.	В разделе «Деревья решений» укажите метод оценки важности признаков.	Feature Importance	ППК-ДА2	ППК-ДА2.3-3-1
30.	При моделировании укажите цель аплифт-моделирования.	Оценка воздействия вмешательства	МО-1	МО-1.1-3-1
31.	В разделе «Деревья решений» укажите метод оценки сложности алгоритма.	$O(N \log N)$ / Глубина дерева	ОПК-4	ОПК-4.1-3-1
32.	При подготовке данных укажите метод кодирования категориальных признаков для деревьев.	Label Encoding / One-Hot	ППК-ДА2	ППК-ДА2.1-У-1
33.	В разделе «Деревья решений» укажите систему версионирования моделей.	MLflow / DVC	МО-2	МО-2.2-У-1
34.	при взаимодействии укажите роль в команде при ревью кода модели.	Рецензент / Разработчик	УК-3	УК-3.1-У-1
35.	В разделе «Деревья решений» укажите обоснование применения AutoML.	Экономия времени / Поиск лучшего	МО-1	МО-1.3-У-1

36.	При визуализации укажите способ отображения решений дерева.	Дерево решений / Граф	ППК-ДА3	ППК-ДА3.3-У-1
37.	В разделе «Деревья решений» укажите план развертывания ансамбля.	План деплоя	УНК-3	УНК-3.2-У-2
38.	В разделе «Обучение без учителя» укажите алгоритм кластеризации.	K-Means / K-средних	МО-1	МО-1.1-3-1
39.	При понижении размерности укажите метод линейной проекции.	PCA / Метод главных компонент	МО-1	МО-1.1-3-1
40.	В разделе «Обучение без учителя» укажите метод поиска аномалий.	Isolation Forest / LOF	МО-1	МО-1.1-3-1
41.	При генерации признаков укажите метод создания новых фичей.	Feature Engineering	ППК-ДА2	ППК-ДА2.1-3-1
42.	В разделе «Обучение без учителя» укажите формат преобразования данных.	DataFrame / Array	ППК-ДА2	ППК-ДА2.1-У-2
43.	При интерпретации укажите инструмент объяснения предсказаний.	SHAP / LIME	МО-1	МО-1.1-3-1
44.	В разделе «Обучение без учителя» укажите метрику обнаружения сдвига данных.	PSI / Distance metric	МО-2	МО-2.3-3-1
45.	При CI/CD укажите этап пайплайна для ML.	Training / Deployment	МО-2	МО-2.1-3-1
46.	В разделе «Обучение без учителя» укажите способ развертывания мониторинга.	Dashboard / Alerting	МО-2	МО-2.3-У-1
47.	При автоматизации укажите метод переобучения модели.	Automated Retraining	МО-2	МО-2.4-У-1
48.	В разделе «Обучение без учителя» укажите стандарт документации экспериментов.	Experiment Tracking	УК-4	УК-4.2-У-3
49.	При отладке укажите метод поиска ошибок в пайплайне данных.	Logging / Debugging	ОПК-6	ОПК-6.2-У-2
50.	В разделе «Основные термины» укажите норму защиты персональных данных.	152-ФЗ / GDPR	УК-2	УК-2.1-3-1