

Федеральное государственное образовательное бюджетное учреждение
высшего образования

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Кафедра искусственного интеллекта
Факультет информационных технологий и анализа больших данных**

Документ подписан усиленной неквалифицированной электронной подписью
Организация: Финансовый университет при Правительстве РФ
Утверждено: Проректор по учебной и методической работе Е.А. Каменева
Сертификат: +75q1DEWL/2hv2D7ogm0lIGxHLNeyLIA
Дата: 22.09.2025 г.

М.В. Коротеев, Н.А. Андриянов

Машинное обучение

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:

01.03.02 - Прикладная математика и информатика,

Образовательная программа «Прикладное машинное обучение»

Профиль:

«Прикладное машинное обучение»

Рекомендовано

*Факультет информационных технологий и анализа больших данных
(протокол № 03 от 16.12.2025 г.)*

Одобрено

*Кафедра искусственного интеллекта
(протокол № 4 от 05.12.2025 г.)*

© Москва 2026

СОДЕРЖАНИЕ

1.	Наименование дисциплины	3
2.	Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3.	Место дисциплины в структуре образовательной программы	5
4.	Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	5
5.	Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	6
5.1.	Содержание дисциплины	6
5.2.	Учебно-тематический план	8
5.3.	Содержание семинаров	9
6.	Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	11
6.1.	Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	11
6.2.	Перечень вопросов, заданий, тем для подготовки к текущему контролю	13
7.	Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	17
8.	Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	26
9.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	26
10.	Методические указания для обучающихся по освоению дисциплины	28
11.	Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	30
12.	Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	31

1. Наименование дисциплины

«Машинное обучение».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПKN-2	Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля	демонстрирует знание базовых математических моделей, применяемых в различных предметных областях	знать: математические основы ключевых классов моделей машинного обучения (линейные, вероятностные, деревья решений, ансамбли, базовые нейросети); принципы построения и оптимизации математических моделей; условия корректного применения моделей, типовые ограничения, допущения, области устойчивости и чувствительности уметь: выбирать подходящий класс модели на основе анализа структуры задачи и свойств данных; оценивать применимость моделей с точки зрения интерпретируемости, вычислительной сложности и риска переобучения; аргументировать выбор модели на теоретическом и прикладном уровне
		Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи	знать: инструментальные средства и библиотеки для реализации моделей машинного обучения (NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow); методы предобработки данных: нормализация, выбор признаков, трансформации, очистка, заполнение пропусков; основные подходы к валидации моделей, оценке качества и подбору гиперпараметров уметь: реализовывать и модифицировать алгоритмы машинного обучения в виде программного модуля; выполнять

			полный цикл разработки модели: загрузка данных, подготовка, обучение, валидация, оптимизация; оформлять результаты работы в виде репрезентативной программной системы (скрипта, модуля, сервиса) с использованием распространённых библиотек
		Владеет методологией математического моделирования для решения профессиональных задач	знать: современные методы моделирования в выбранной предметной области (регрессия, классификация, кластеризация, прогнозирование, интерпретируемые модели); принципы построения комплексных моделей, сочетания нескольких подходов (ансамблирование, каскадирование, пайплайны обработки); методы анализа результатов моделирования, интерпретации параметров и качественных метрик уметь: проводить комплексное моделирование предметной области: формулировать постановку задачи, выбирать метод, разрабатывать модель и оценивать её качество; интерпретировать и эффективно представлять результаты моделирования в виде отчётов, метрик, графиков, сравнения моделей; делать обоснованные выводы и рекомендации на основе математических результатов, включая ограничения и возможные риски

3. Место дисциплины в структуре образовательной программы

Дисциплина «Машинное обучение» относится к «Общефакультетскому (предпрофильному) циклу».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 3 (в часах)	Семестр 4 (в часах)
Общая трудоёмкость дисциплины	8/288	144	144
Контактная работа- Аудиторные занятия	100	50	50
Лекции	32	16	16
Семинары, практические занятия	68	34	34
Самостоятельная работа	188	94	94
В том числе курсовая работа/курсовой проект	24	-	-
Вид текущего контроля	Контрольная работа	Контрольная работа	
Вид промежуточной аттестации	Зачет, Экзамен	Зачет	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Введение в машинное обучение и роль ML в системах искусственного интеллекта

Основные понятия и цели машинного обучения. Связь с данными, статистикой и ИИ. Типы задач: обучение с учителем, без учителя и с подкреплением. Частичное и гибридное обучение. Структура и источники данных. Основные этапы ML-проекта: постановка задачи, подготовка данных, обучение, оценка и интерпретация. Обзор библиотек Scikit-learn, PyTorch, TensorFlow. Применения ML в разных отраслях.

Тема 2. Регрессия и линейные модели

Постановка задачи регрессии. Линейная и полиномиальная регрессия. Функции потерь, градиентный спуск и его варианты. Регуляризация: L1, L2, ElasticNet. Интерпретация коэффициентов. Основные метрики регрессии: R -квадрат, MAE, MSE, RMSE, MAPE. Практика реализации регрессионных моделей в Scikit-learn. Прогнозирование трендов, применение к временным данным.

Тема 3. Классификация и вероятностные модели

Понятие классификации. Логистическая регрессия, перцептрон, многоклассовые схемы («один против всех»). Наивные байесовские модели. Вероятностная интерпретация решений, softmax и sigmoid функции. Метрики классификации: accuracy, precision, recall, F1-мера, ROC-AUC, PR-кривая. Работа с несбалансированными классами. Кастомизация классификаторов и интерпретация вероятностей. Применение классификации в прикладных задачах.

Тема 4. Алгоритмы обучения с учителем: классические и ансамблевые модели

SVM с разными ядрами. Метод ближайших соседей. Деревья решений. Ансамблирование: беггинг, бустинг, стекинг. Случайный лес, XGBoost, CatBoost. Особенности применения разных классов моделей, достоинства и ограничения. Настройка гиперпараметров, сравнение моделей. Частичное обучение и расширяемость моделей.

Тема 5. Предварительный анализ, очистка данных и построение признаков

Сбор и оценка качества данных. Структура реляционных и табличных данных. Анализ распределений, визуализация, выявление выбросов, аномалий, пропусков. Несбалансированность классов. Очистка данных: удаление, заполнение, импутация. Обработка категориальных признаков, нормализация и стандартизация. Создание производных признаков. Оценка важности признаков. Воспроизводимость преобразований данных.

Тема 6. Обучение без учителя: кластеризация и понижение размерности

Кластеризация: K-средних, иерархическая, агломеративная, DBSCAN. Выбор числа кластеров и метрики качества: Silhouette, Davies–Bouldin. Интерпретация кластеров и применение в аналитике. Понижение размерности: PCA, t-SNE, UMAP. Использование методик для визуализации, ускорения и интерпретации моделей. Рабочие кейсы и ограничения.

Тема 7. Обнаружение аномалий и выявление скрытых закономерностей

Типы аномалий. Методы обнаружения: статистические подходы, LOF, Isolation Forest, One-Class SVM. Проверка устойчивости моделей, synthetic/adversarial data. Ассоциативные правила: Apriori, FP-Growth. Метрики lift, support, confidence. Применение в бизнес-аналитике, ретейле и ИИ-системах. Интерпретация закономерностей и их использование для принятия решений.

Тема 8. Практическое применение ML-моделей и продвинутые техники

Стохастический и пакетный градиентный спуск. Ансамблирование и оптимизация моделей. Конвейеризация ML-процессов (Pipelines). Обучение на малых данных: трансферное, few-shot. Обучение на больших данных: распределённые вычисления. Оценка устойчивости моделей, детекция дрейфа. Интерпретация результатов (Explainability). AutoML. Представление результатов заказчику. Работа с разными типами данных: таблицы, тексты, изображения.

5.2. Учебно-тематический план

№ п/п	Наименование тем(разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа
			Общая, в т.ч.:	Лекции	Семинары, практическ ие занятия	
1	Введение в машинное обучение и роль ML в системах искусственного интеллекта	20	6	2	4	14
2	Регрессия и линейные модели	26	8	2	6	18
3	Классификация и вероятностные модели	26	8	2	6	18
4	Алгоритмы обучения с учителем: классические и ансамблевые модели	44	16	4	12	28
5	Предварительный анализ, очистка данных и построение признаков	44	18	8	10	26
6	Обучение без учителя: кластеризация и понижение размерности	44	18	6	12	26
7	Обнаружение аномалий и выявление скрытых закономерностей	36	10	4	6	26
8	Практическое применение ML-моделей и продвинутые техники	48	16	4	12	32
	Итого	288	100	32	68	188

5.3. Содержание семинаров

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Введение в машинное обучение и роль ML в системах искусственного интеллекта	Входной контроль. Знакомство с технологическим стеком ML на Python: numpy, pandas, matplotlib, scikit-learn. Разбор примерного ML-проекта: сбор данных, подготовка, обучение простой модели. Практика базового EDA: просмотр данных, вычисление простых статистик, построение графиков.	Решение и обсуждение задач
Регрессия и линейные модели	Реализация линейной регрессии вручную: формирование функции потерь, градиентный спуск. Создание полиномиальной регрессии. Построение моделей в scikit-learn и сравнение с ручной реализацией. Анализ результатов на реальном датасете: интерпретация коэффициентов, оценка качества.	Решение и обсуждение задач
Классификация и вероятностные модели	Построение логистической регрессии своими руками и с помощью библиотек. Отработка подходов к многоклассовой классификации. Практика работы с вероятностными выводами и порогами классификации.	Решение и обсуждение задач
Алгоритмы обучения с учителем: классические и ансамблевые модели	Построение нескольких моделей (SVM, деревья решений, kNN) на одном датасете и сравнение результатов. Создание ансамблей: случайный лес, градиентный бустинг. Построение сложного регрессора/классификатора и анализ гиперпараметров.	Решение и обсуждение задач, соревнование
Предварительный анализ, очистка данных и построение признаков	Разбор качества данных: пропуски, выбросы, несбалансированность. Очистка данных: кодирование категориальных признаков, нормализация, обработка пропусков. Построение новых признаков и оценка их влияния на качество модели. Визуальное сравнение моделей с разными наборами признаков.	Решение и обсуждение задач
Обучение без учителя: кластеризация и понижение размерности	Решение задач кластеризации методами K-means, иерархической и агломеративной кластеризации. Выбор числа кластеров, интерпретация результатов. Понижение размерности с помощью PCA, t-SNE, UMAP. Применение понижения размерности для визуализации многомерных данных.	Решение и обсуждение задач
Обнаружение аномалий и выявление	Обнаружение аномалий: LOF, Isolation Forest, One-Class SVM на искусственных и реальных данных.	Решение и обсуждение задач

скрытых закономерностей	Поиск ассоциативных правил: FP-Growth, Apriori и интерпретация результатов. Использование обнаружения аномалий для диагностики данных перед обучением моделей.	
Практическое применение ML-моделей и продвинутые техники	Создание ансамблей моделей и оптимизация гиперпараметров. Построение ML-конвейера (pipeline) со всеми этапами подготовки данных. Классификация текстов: сравнение методов векторизации (Bag-of-Words, TF-IDF). Классификация изображений: представление данных, простые модели. Визуализация и интерпретация результатов в контексте прикладной задачи.	Решение и обсуждение задач, проект

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Введение в машинное обучение и роль ML в системах искусственного интеллекта	Изучение истории машинного обучения как дисциплины, связи с другими областями знания. Изучение структуры ML-проекта, роли данных, этапов подготовки и проверки. Освоение базовых библиотек Python для анализа данных (numpy, pandas, matplotlib, scikit-learn) по документации. Разбор примеров типовых задач: классификация, регрессия, кластеризация. Изучение связей ML со статистикой и искусственным интеллектом.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Регрессия и линейные модели	Знакомство с нормированием данных, адаптивной скоростью обучения в ручной реализации метода градиентного спуска. Векторизация вычислений в модели множественной регрессии. Использование множественной регрессии для моделирования полиномиальных признаков и нелинейных зависимостей (корень, экспонента, синус, логарифм, обратная пропорциональность). Изучение расширенных методов регуляризации (L1, L2, ElasticNet) и их влияния на модель. Практика масштабирования данных и анализа нелинейных трансформаций признаков. Реализация регрессии для временных рядов и оценка устойчивости результатов. Сравнение моделей по метрикам, анализ ошибок и интерпретация коэффициентов.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Классификация и вероятностные модели	Реализация многоклассовых схем классификации («один против всех», softmax). Сравнение различных вероятностных моделей: логистическая регрессия, наивный байес. Анализ порогов классификации и построение ROC/PR-кривых. Оценка классификаторов на несбалансированных выборках.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Алгоритмы обучения с учителем: классические и ансамблевые модели	Сравнение SVM, деревьев решений, kNN и ансамблей на одном датасете. Изучение влияния гиперпараметров на качество моделей. Реализация ансамблей (Random Forest, Gradient Boosting) и их	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение

	анализ. Сравнение производительности моделей на больших данных. Визуализация результатов классификации. Методы кросс-валидации.	домашних заданий к каждому занятию.
Предварительный анализ, очистка данных и построение признаков	Глубокий EDA: распределения, корреляции, выбросы, пропуски, визуализация. Сравнение стратегий нормализации и их влияния на обучение. Создание производных признаков и проверка их значимости. Алгоритмический отбор признаков, оценка важности признаков на основе моделей.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Обучение без учителя: кластеризация и понижение размерности	Решение задачи кластеризации на реальных данных, оценка числа кластеров. Применение PCA, t-SNE, UMAP для анализа структуры данных. Понижение размерности как шаг инжиниринга признаков и подготовка к моделированию. Интерпретация кластеров и визуализация результатов.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Обнаружение аномалий и выявление скрытых закономерностей	Применение методов LOF, Isolation Forest, One-Class SVM к реальным данным. Поиск ассоциативных правил (Apriori, FP-Growth) и интерпретация результатов. Выявление скрытых закономерностей в данных и проверка гипотез. Использование методов обнаружения аномалий в качестве этапа предварительного анализа.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Практическое применение ML-моделей и продвинутые техники	Применение XGBoost, CatBoost и сравнение с базовыми моделями. Оптимизация гиперпараметров и построение ML-конвейера (pipeline). Знакомство с обработкой текстов (TF-IDF, bag-of-words) и классификацией текстов. Векторизация изображений и построение простой модели классификации изображений. Интерпретация результатов и подготовка визуализаций.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные задания контрольной работы

Для решения используется Jupyter Notebook

1. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Описать основные количественные характеристики датасета. Визуализировать распределение каждого признака, входящего в датасет. Визуализировать совместное распределение каждого признака и целевой переменной. Сделать вывод о линейной разделимости классов.
2. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Построить модель парной классификации каждого признака и целевой переменной. Оценить метрики эффективности четырех моделей классификации. Сделать вывод о значимости каждого признака для классификации.
3. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод деревьев решений. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.
4. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод логистической регрессии. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.
5. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод опорных векторов. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.
6. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод многослойного перцептрона. Построить отчет о классификации, оценить основные метрики

классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.

7. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Построить модель классификации объектов по всем четырем признакам, используя любой метод классификации. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности классификации.

8. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Построить модель классификации объектов по всем четырем признакам, используя любой метод. Оценить эффективность модели при помощи перекрестной проверки (кросс-валидации). Сделать вывод об эффективности классификации.

9. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Понизить размерность датасета до двух измерений. Визуализировать получившийся набор данных учитывая значение целевой переменной. Сделать вывод о линейной разделимости классов.

10. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Построить ансамблевую модель (случайный лес, градиентный бустинг). Сравнить качество классификации с базовой моделью. Сделать вывод о влиянии ансамблей.

11. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Построить модель классификации и вычислить матрицу ошибок. Провести её интерпретацию. Сделать вывод об источниках ошибок классификатора.

12. Загрузить встроенный в библиотеку sklearn датасет “Ирисы”. Построить кривую ROC и PR-кривую для одного из классов. Сделать вывод о качестве классификации на разных порогах.

13. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Описать основные количественные характеристики датасета. Охарактеризовать шкалу измерения каждого признака (вид шкалы, минимальное, максимальное значение, количество значений). Визуализировать совместное распределение каждого признака и целевой переменной. Сделать вывод о значимости каждого признака для регрессии.

14. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Построить модель множественной регрессии. Оценить метрики эффективности регрессии. Сделать вывод об эффективности получившейся модели.

15. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Построить модель регрессии по всем признакам, используя любой метод регрессии. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности регрессии.
16. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Построить модель регрессии по всем признакам, используя любой метод. Оценить эффективность модели при помощи перекрестной проверки (кросс-валидации). Сделать вывод об эффективности регрессии.
17. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Понизить размерность датасета до двух измерений. Визуализировать получившийся набор данных учитывая значение целевой переменной. Сделать вывод.
18. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Построить регуляризованную регрессию (Lasso, Ridge, ElasticNet). Сравнить метрики с обычной регрессией. Сделать вывод о влиянии регуляризации.
19. Загрузить встроенный в библиотеку sklearn датасет “Бостон”. Выполнить отбор признаков (feature importance, Lasso, пермутационная важность). Оценить влияние отбора на качество модели.
20. Загрузить встроенный в библиотеку sklearn датасет “Цифры” (MNIST). Построить модель классификации объектов, используя метод многослойного перцептрона. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.
21. Загрузить встроенный в библиотеку sklearn датасет “Цифры” (MNIST). Построить модель классификации объектов, используя любой метод классификации. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности классификации.
22. Загрузить встроенный в библиотеку sklearn датасет “Цифры” (MNIST). Построить модель классификации объектов, используя любой метод. Оценить эффективность модели при помощи перекрестной проверки (кросс-валидации). Сделать вывод об эффективности классификации.
23. Загрузить встроенный в библиотеку sklearn датасет “Цифры” (MNIST). Понизить размерность датасета до двух измерений. Визуализировать

получившийся набор данных учитывая значение целевой переменной. Сделать вывод о линейной разделимости классов.

24. Загрузить встроенный в библиотеку sklearn датасет “Цифры” (MNIST). Построить SVM-классификатор с различными ядрами (линейным, rbf). Сравнить метрики. Сделать вывод о влиянии выбора ядра.

25. Загрузить встроенный в библиотеку sklearn датасет “Цифры” (MNIST). Построить случайный лес и сравнить качество классификации с MLP-классификатором. Сделать вывод о преимуществе/недостатке ансамблей.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры искусственного интеллекта Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. *Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.*

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

ПKN-2 Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля

1) демонстрирует знание базовых математических моделей, применяемых в различных предметных областях

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: математические основы ключевых классов моделей машинного обучения (линейные, вероятностные, деревья решений, ансамбли, базовые нейросети); принципы построения и оптимизации математических моделей; условия корректного применения моделей, типовые ограничения, допущения, области устойчивости и чувствительности

Уметь: выбирать подходящий класс модели на основе анализа структуры задачи и свойств данных; оценивать применимость моделей с точки зрения интерпретируемости, вычислительной сложности и риска переобучения; аргументировать выбор модели на теоретическом и прикладном уровне

Типовые контрольные задания

1. По текстовому описанию трёх прикладных задач (регрессия, бинарная и многоклассовая классификация) выбрать подходящие классы моделей (линейные, вероятностные, деревья решений, ансамбли, базовые нейросети), записать их математическую постановку (модель, функция потерь) и указать ключевые ограничения и допущения применения для каждой задачи

2. Для заданного датасета проанализировать структуру задачи и данных, обосновать выбор подходящего класса модели среди нескольких альтернатив, обучить модели и, сравнив интерпретируемость, вычислительную сложность и риск переобучения, аргументировать выбор итоговой модели

2) Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: инструментальные средства и библиотеки для реализации моделей машинного обучения (NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow); методы предобработки данных: нормализация, выбор признаков, трансформации, очистка, заполнение пропусков; основные подходы к валидации моделей, оценке качества и подбору гиперпараметров

Уметь: реализовывать и модифицировать алгоритмы машинного обучения в виде программного модуля; выполнять полный цикл разработки модели: загрузка данных, подготовка, обучение, валидация, оптимизация; оформлять результаты работы в виде репрезентативной программной системы (скрипта, модуля, сервиса) с использованием распространённых библиотек

Типовые контрольные задания

1. По описанию исходных данных и постановки задачи предложить необходимый набор этапов предобработки, выбрать схему валидации и обосновать использование инструментальных средств
2. Загрузить встроенный в библиотеку sklearn датасет, реализовать полный цикл разработки модели: загрузка и подготовка данных, выбор и обучение алгоритма, оценка качества по кросс-валидации, оптимизация гиперпараметров методом поиска по сетке, оформить решение в виде программного модуля и сделать вывод об эффективности полученной модели

3) Владеет методологией математического моделирования для решения профессиональных задач

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: современные методы моделирования в выбранной предметной области (регрессия, классификация, кластеризация, прогнозирование, интерпретируемые модели); принципы построения комплексных моделей, сочетания нескольких подходов (ансамблирование, каскадирование, пайплайны обработки); методы анализа результатов моделирования, интерпретации параметров и качественных метрик

Уметь: проводить комплексное моделирование предметной области: формулировать постановку задачи, выбирать метод, разрабатывать модель и оценивать её качество; интерпретировать и эффективно представлять результаты моделирования в виде отчётов, метрик, графиков, сравнения моделей; делать обоснованные выводы и рекомендации на основе математических результатов, включая ограничения и возможные риски

Типовые контрольные задания

1. Для выбранной предметной области сформулировать постановку задачи, предложить схему комплексного моделирования, определить набор метрик качества и описать возможные ограничения и риски использования такой модели
2. На основе открытого датасета спроектировать и реализовать комплексное решение: собрать пайплайн предобработки и моделирования, обучить не менее двух различных моделей, сравнить их по нескольким метрикам, визуализировать результаты, подготовить краткий отчет с интерпретацией параметров и метрик и сформулировать обоснованные выводы и рекомендации по применению модели

Примеры практико-ориентированных заданий

Практическая часть представляет собой задачу по анализу конкретного датасета. Для каждого задания будет необходимо:

1. Загрузить датасет в Python.
2. Описать набор данных и решаемую задачу.
3. Выделить целевую переменную и факторные переменные.
4. Удалить ненужные данные, проанализировать отсутствующие значения.
5. Прокомментировать количественные параметры датасета.
6. Разбить выборку на обучающую и тестовую.
7. Работа по вариантам.

Вариант 1. Очистка данных и обучение моделей.

Данный вариант предполагает фокусировку на обучении нескольких видов моделей обучения с учителем. В зависимости от набора данных, может предполагаться задача классификации и регрессии. Необходимо после минимальной подготовки датасета к обучению обучить несколько моделей и сравнить их эффективность.

Вариант 2. Описательный анализ и визуализация данных.

Данный вариант предполагает фокусировку на исследовании данных и визуализации. При решении этого варианта следует провести как можно более подробный описательный анализ данных с использованием максимального спектра средств визуализации. При этом следует делать значимые выводы об обнаруженных в данных закономерностях.

Вариант 3. Построение модели и оптимизация гиперпараметров.

Данный вариант предполагает фокусировку на процессе улучшения эффективности модели обучения с учителем. Студенту следует подготовить датасет к обучению, обучить одну из моделей с учителем со значениями гиперпараметров по умолчанию, получить значение эффективности. После этого вручную или автоматически подобрать значения гиперпараметров таким образом, чтобы получить максимальный прирост эффективности.

Вариант 4. Выбор признаков.

Данный вариант предполагает фокусировку на улучшении модели путем ввода новых признаков в модель. Следует подготовить модель к обучению, обучить модель и зафиксировать начальный уровень эффективности. Затем следует исследовать влияние исключения существующих и введения новых признаков в модель на эффективность. Как вариант можно рассматривать введение полиномиальных признаков. Следует стремиться к максимальному увеличению эффективности модели.

Вариант 5. Использование методов обучения без учителя для улучшения качества моделей с учителем.

Данный вариант предполагает использование методов обучения без учителя (кластеризация, понижение размерности, обнаружение аномалий) в качестве этапа подготовки данных перед обучением модели с учителем. Студенту следует обучить базовую модель и зафиксировать исходную эффективность. После этого применить один или несколько методов обучения без учителя (PCA, UMAP, кластеризацию, обнаружение аномалий) для преобразования данных, выполнить повторное обучение и сравнить результаты. По итогам работы необходимо сделать вывод об изменении скорости, устойчивости и эффективности обучения с учителем.

Вариант 6. Построение полного конвейера машинного обучения (Pipeline)

Данный вариант предполагает фокусировку на создании сквозного конвейера обработки данных и обучения модели. Студенту следует подготовить датасет к обучению, выбрать необходимые этапы предобработки (масштабирование, кодирование, отбор признаков) и объединить их с моделью в единый Pipeline. После этого необходимо обучить модель, провести оценку эффективности с использованием кросс-валидации и сравнить работу конвейера с ручной последовательной реализацией этапов. По итогам работы следует сделать вывод о преимуществах использования конвейера в задачах машинного обучения.

Вариант 7. Исследование интерпретируемости моделей машинного обучения

Данный вариант предполагает фокусировку на изучении интерпретируемости моделей и объяснении их решений. Студенту следует выбрать подходящую модель (линейную, дерево решений или ансамблевую), обучить её на подготовленных данных и выполнить анализ важности признаков. После этого необходимо визуализировать вклад признаков в предсказание (например, с помощью графиков зависимости или значимости признаков) и интерпретировать полученные результаты. По итогам работы следует сделать вывод о прозрачности модели и возможности доверия к её прогнозам.

Вариант 8. Сравнение ансамблей и базовых моделей машинного обучения

Данный вариант предполагает фокусировку на сравнении эффективности ансамблевых моделей с базовыми алгоритмами. Студенту следует обучить простую модель (например, логистическую регрессию или дерево решений) и зафиксировать базовую эффективность. Затем необходимо обучить одну или несколько ансамблевых моделей (Random Forest, Gradient Boosting, XGBoost, CatBoost), сравнить результаты по точности, устойчивости и времени обучения. По итогам работы следует сделать вывод о преимуществах и недостатках ансамблей по сравнению с базовыми моделями.

Примерные вопросы для подготовки к зачету, экзамену

Примерные вопросы для подготовки к зачету

1. Понятие машинного обучения. Отличие машинного обучения от других областей программирования.
2. Классификация задач машинного обучения. Примеры задач из различных классов.
3. Основные понятия машинного обучения: набора данных, объекты, признаки, атрибуты, модели, параметры.
4. Структура и представление данных для машинного обучения.
5. Инструментальные средства машинного обучения.
6. Задача регрессии: постановка, математическая формализация.
7. Метод градиентного спуска для парной линейной регрессии.
8. Понятие функции ошибки: требования, использование, примеры.
9. Множественная и нелинейная регрессии.

10. Нормализация признаков в задачах регрессии.
11. Задача классификации: постановка, математическая формализация.
12. Метод градиентного спуска для задач классификации.
13. Логистическая регрессия в задачах классификации.
14. Множественная и многоклассовая классификация. Алгоритм “один против всех”.
15. Метод опорных векторов в задачах классификации.
16. Понятие ядра и виды ядер в методе опорных векторов.
17. Метод решающих деревьев в задачах классификации.
18. Метод k ближайших соседей в задачах классификации.
19. Однослойный перцептрон в задачах классификации.
20. Проблемы недообучения и переобучения. Понятие обобщающей способности модели.
21. Метрики эффективности моделей классификации: accuracy, precision, recall, F-метрика.
22. Сравнение основных методов обучения с учителем: преимущества, ограничения и области применимости.

Примерные вопросы для подготовки к экзамену

1. Метрики эффективности и функции ошибки: назначение, примеры, различия.
2. Понятие набора данных (датасета) в машинном обучении. Требования, представление. Признаки и объекты.
3. Шкалы измерения признаков. Виды шкал, их характеристика.
4. Понятие чистых данных. Определение, очистка данных.
5. Основные этапы проекта по машинному обучению.
6. Предварительный анализ данных: задачи, методы, цели.
7. Проблема отсутствующих данных: причины, исследование, пути решения.
8. Проблема несбалансированных классов: исследование, пути решения.

9. Понятие параметров и гиперпараметров модели. Обучение параметров и гиперпараметров. Поиск по сетке.
10. Понятие недо- и переобучения. Определение, пути решения.
11. Диагностика модели машинного обучения. Методы, цели.
12. Проблема выбора модели машинного обучения. Сравнение моделей.
13. Измерение эффективности работы моделей машинного обучения. Метрики эффективности.
14. Метрики эффективности моделей классификации. Виды, характеристика, выбор.
15. Метрики эффективности моделей регрессии. Виды, характеристика, выбор.
16. Перекрестная проверка (кросс-валидация). Назначение, схема работы.
17. Методы векторизации текстов для задач машинного обучения.
18. Представление графической информации в моделях машинного обучения.
19. Задачи без учителя. Кластеризация. Метод k средних.
20. Задачи без учителя. Обнаружение аномалий.
21. Задачи без учителя. Понижение размерности. Метод PCA.
22. Конвейеры в библиотеке sklearn. Назначение, использование.
23. Использование методов визуализации данных для предварительного анализа.
24. Исследование коррелированности признаков: методы, цели, выводы.
25. Решкалирование данных. Виды, назначение, применение. Нормализация и стандартизация данных.
26. Преобразование категориальных признаков в числовые.
27. Ансамблевые модели машинного обучения. Виды ансамблирования.
28. Конвейеризация моделей машинного обучения.
29. Методы визуализации данных для машинного обучения.
30. Задача выбора модели. Оценка эффективности, валидационный набор.

31. Интерпретируемость моделей машинного обучения: методы оценки важности признаков и способы визуализации.

32.т Современные методы понижения размерности: t-SNE, UMAP. Назначение, особенности применения.

33.т Автоматизация машинного обучения (AutoML). Основные идеи, преимущества и ограничения.

34.т Использование методов обучения без учителя для улучшения качества моделей с учителем: цели, примеры, ограничения.

Пример экзаменационного билета

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

1. Представление графической информации в моделях машинного обучения **(20 баллов)**

2. Построение модели и оптимизация гиперпараметров.

Датасет: <https://www.kaggle.com/nandvard/microsoft-data-science-capstone> **(40 баллов)**

Примерные темы курсовой работы /курсового проекта

1) Deskриптивный анализ численного набора данных с использованием технологий визуализации.

2) Deskриптивный анализ категориального набора данных с использованием технологий визуализации.

3) Deskриптивный анализ смешанного набора данных с использованием технологий визуализации.

4) Сравнение методов регрессии на реальных наборах данных.

5) Сравнение методов классификации на реальных наборах данных.

6) Предварительный анализ данных и построение признаков в задачах обработки финансовой и экономической информации.

7) Предварительный анализ данных и построение признаков в задачах распознавания голоса.

- 8) Предварительный анализ данных и построение признаков в задачах распознавания текста.
- 9) Предварительный анализ данных и построение признаков в задачах распознавания объектов на фотографии.
- 10) Предварительный анализ данных и построение признаков в задачах сжатия информации.
- 11) Предварительный анализ данных и построение признаков в задачах повышения качества изображений.
- 12) Предварительный анализ данных и построение признаков в задачах идентификации личности по голосу.
- 13) Предварительный анализ данных и построение признаков в задачах идентификации личности по изображению.
- 14) Предварительный анализ данных и оценка важности признаков с помощью статистических и модельных методов.
- 15) Построение и визуализация многоуровневых признаков (feature engineering pipelines) в задачах регрессии и классификации.
- 16) Сравнение методов понижения размерности для числовых и смешанных данных (PCA, t-SNE, UMAP).
- 17) Исследование методов обнаружения аномалий на реальных данных (Isolation Forest, LOF, One-Class SVM).
- 18) Оценка эффективности ансамблей моделей (Random Forest, Gradient Boosting, CatBoost, XGBoost) на прикладных задачах.
- 19) Построение ML-конвейера (pipeline) от данных до финальной модели с использованием scikit-learn.
- 20) Сравнение методов векторизации текстов (Bag-of-Words, TF-IDF, Word Embeddings) в задачах классификации.
- 21) Сравнение методов предварительной обработки и классификации изображений (простые фильтры, PCA-кодирование, MLP-классификация).

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Нормативные правовые акты:

Кодекс этики в сфере ИИ <https://ethics.a-ai.ru/>

Основная литература:

1. Колдаев, Виктор Дмитриевич Структуры и алгоритмы обработки данных : Учебное пособие / Московский институт электронной техники 1 Москва : Издательский Центр РИОР, 2021 296 с. (Высшее образование: Бакалавриат) ВО - Бакалавриат <https://znanium.com/catalog/document?id=398591> ISBN 978-5-369-01264-2 ISBN 978-5-16-101275-8 (электр. издание) ISBN 978-5-16-009012-2 (ISBN соиздателя) . [БИК ID: RU\infra-m\znanium\bibl\1230215]
2. Коротеев, М.В. Основы машинного обучения на Python : Учебник / М.В. Коротеев Электрон. дан. Москва : КноРус, 2025 431 с. Режим доступа: book.ru Internet access <https://book.ru/book/957785> ISBN 978-5-406-14728-3. [БИК ID: RU\bookru\bibl\957785]

Дополнительная литература:

1. Черенкова, И. А. Основы алгоритмизации и программирования на языке высокого уровня [Электронный ресурс] : учебное пособие / Черенкова И. А., Кишкинова О. А., Миндлин Ю. Б. Москва : МГАВМиБ им. К.И. Скрябина, 2022 126 с. Книга из коллекции МГАВМиБ им. К.И. Скрябина - Информатика <https://e.lanbook.com/book/331409> ISBN 978-5-4443-0235-4. [БИК ID: RU-LAN-BOOK-331409]
2. Златопольский, Д.М. Программирование: типовые задачи алгоритмы методы — 5-е изд. : Не указано / Д.М. Златопольский Электрон. дан. Москва : Лаборатория знаний, 2025 225 с. Режим доступа: book.ru Internet access <https://book.ru/book/957129> ISBN 978-5-93208-832-6. [БИК ID: RU\bookru\bibl\957129]

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>

3. Электронно-библиотечная система "Университетская библиотека ОНЛАЙН"
<http://biblioclub.ru/>
4. <http://znanium.com> - ЭБС издательства «ИНФРА-М»
5. •Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
6. •Электронно-библиотечная система издательства Проспект
<http://ebs.prospekt.org/books>
7. •Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>
8. •Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
9. Электронная библиотека <http://grebennikon.ru>
10. •Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
11. •Национальная электронная библиотека <http://нэб.рф/>
12. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
13. Pylru 1.0.9 <https://pypi.python.org/pypi/pylru>
14. Python Documentation <http://python.org/doc/>
15. Python Standard Library <https://docs.python.org/2/library/>
16. <http://pandas.pydata.org/>

10. Методические указания для обучающихся по освоению дисциплины

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного списка. Кроме этого, необходимо активно работать с Интернет-источниками и пособиями других авторов, помогающими усвоить материал отдельных разделов программы.

Необходимо конспектировать лекции, пометая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на консультации.

При подготовке к семинарским занятиям необходимо изучить вопросы, вынесенные на самостоятельное изучение, так как семинарские занятия предполагают их обсуждение и дискуссию по теме; кроме того, задания для самостоятельной работы необходимы для того, чтобы успешно выполнить самостоятельные задания на семинарах.

Индивидуальные задания для работы на компьютере, файлы с выполненными заданиями необходимо хранить в личной сетевой папке в компьютерной сети вуза.

Для выполнения кейса задачи, рекомендуется использовать язык программирования Python и специализированные библиотеки для анализа данных и машинного обучения.

Студент должен прислать рабочий notebook с решением кейса задачи, комментариями кода и аналитическими выводами до конца экзамена на электронную почту преподавателя.

Критерии оценки практической экзаменационной работы:

1. Структурированность отчета. В работе должна прослеживаться четкая структура - подготовительный этап, анализ данных, построение простых моделей, сравнение и анализ моделей, выводы, построение моделей с учетом выводов, итоговый результат.
2. Наличие выводов. Работа должна содержать текстовые замечания, поясняющие каждый шаг работы студента: что делается, зачем и какую информацию это нам дает. Оценивается полнота и адекватность выводов.
3. Визуализация. Работа должна демонстрировать навыки студента визуализировать информацию. Особенно на этапах описательного анализа и анализа обучаемости модели. Оценивается разнообразие, наглядность и информативность визуализации.
4. Использование метрик эффективности. Оценивается разнообразие и адекватность задаче примененных метрик эффективности (включая время обучения) а также полнота сравнения и правильность выводов из сравнения моделей по разным метрикам.

5. Валидность результатов. Студент должен продемонстрировать умение оценивать достоверность измерения метрик моделей и повышать ее с использованием перекрестной проверки (кросс-валидации). Использование k-fold cross validation является предпочтительным методом измерения эффективности модели. Если происходит выбор модели, то ее итоговая эффективность должна измеряться на чистом наборе данных.

Список датасетов, использующихся на экзамене:

- 1.<https://www.kaggle.com/uciml/mushroom-classification>
- 2.<https://www.kaggle.com/lodetomasi1995/income-classification>
- 3.<https://www.kaggle.com/uciml/glass>
- 4.<https://www.kaggle.com/uciml/german-credit>
- 5.<https://www.kaggle.com/zaurbegiev/my-dataset>
- 6.<https://www.kaggle.com/kaushiksuresh147/customer-segmentation>
- 7.<https://www.kaggle.com/deepu1109/star-dataset>
- 8.<https://www.kaggle.com/vinesmsuic/star-categorization-giants-and-dwarfs>
- 9.<https://www.kaggle.com/shebrahimi/financial-distress>
- 10.<https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction>
- 11.<https://www.kaggle.com/amir75/caesarean-section-classification>
- 12.<https://www.kaggle.com/sachinsharma1123/performance-prediction>
- 13.<https://www.kaggle.com/ninzaami/loan-predication>
- 14.<https://www.kaggle.com/caparrini/beatsdataset>
- 15.<https://www.kaggle.com/zhiruo19/covid19-symptoms-classification>
- 16.<https://www.kaggle.com/kunalvsingh93/banking-modelclassification>
- 17.<https://www.kaggle.com/mansoordaku/ckdisease>
- 18.<https://www.kaggle.com/mnassrib/telecom-churn-datasets>
- 19.<https://www.kaggle.com/akshayksingh/kidney-disease-dataset>
- 20.<https://www.kaggle.com/henriqueyamahata/bank-marketing>

21. <https://www.kaggle.com/maajdl/yeh-concret-data>
22. <https://www.kaggle.com/hellbuoy/car-price-prediction>
23. <https://www.kaggle.com/rhuebner/human-resources-data-set>
24. <https://www.kaggle.com/loveall/appliances-energy-prediction>
25. <https://www.kaggle.com/elikplim/forest-fires-data-set>
26. <https://www.kaggle.com/nandvard/microsoft-data-science-capstone>
27. <https://www.kaggle.com/shebrahimi/financial-distress>
28. <https://www.kaggle.com/aungpyaeap/beauty>
29. <https://www.kaggle.com/vbmokin/ammonium-prediction-in-river-water>
30. <https://www.kaggle.com/veer06b/marrket-mix-dataset>

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

Комплект лицензионного программного обеспечения:

1. Python 3.8
2. Jupyter Notebook (Windows 10)
3. Anaconda
4. Kaspersky

Современные профессиональные базы данных и информационные справочные системы:

1. Язык программирования Python 3. <https://pythonworld.ru/>

Сертифицированные программные и аппаратные средства защиты информации:

1. не предусмотрены

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

1. Учебная аудитория для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), набор демонстрационного оборудования (проектор, экран)

2. Компьютерный класс для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенный оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), персональные компьютеры, набор демонстрационного оборудования (проектор, экран)