МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ГЕОДЕЗИИ И КАРТОГРАФИИ
(МИИГАиК)

Факультет геоинформатики и информационной безопасности
Кафедра информатики и геоинформационных технологий

**ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ**

Сертификат: 5BFEE1D286E9C95A0C0D2C611E434BD5
Владелец: Степанченко Алексей Леонидович
Действителен: с 23.09.2026 по 17.12.2027

«УТВЕРЖДАЮ»
Проректор
А. Л. Степанченко
15 мая 2026 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
ОСНОВЫ КОМПЬЮТЕРНЫХ НАУК
«Б1.О.05.05 ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»

Уровень высшего образования: бакалавриат

Направление подготовки: 09.03.03 Прикладная информатика

Направленность (профиль) подготовки: Разработка программного обеспечения для обработки и анализа пространственных данных

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Год набора (приема на обучение): 2026

Срок получения образования: 4 года

Объем: в зачетных единицах: 4 з.е.
в академических часах: 144 ак.ч.

Москва, 2026

Разработчики:
Старший преподаватель кафедры Информатики и
геоинформационных технологий Стрельцов А. С.

Рабочая программа дисциплины (модуля) составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.03.03 Прикладная информатика, утвержденного приказом Минобрнауки от 19.09.2017 № 922

Согласование и утверждение

№	Подразделение	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	факультет Геоинформатики и информационной безопасности	Председатель ученого совета факультета	Матерухин А. В.	Рассмотрено	05.11.2025, № 12-25

Содержание

1. Цель и задачи освоения дисциплины (модуля)

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

3. Место дисциплины в структуре ОП

4. Объем дисциплины и виды учебной работы

5. Содержание дисциплины
5.1. Разделы, темы дисциплины и виды занятий
5.2. Содержание разделов, тем дисциплины и формы текущего контроля

6. Оценочные материалы текущего контроля

7. Оценочные материалы промежуточной аттестации

8. Порядок проведения промежуточной аттестации

9. Материально-техническое и учебно-методическое обеспечение дисциплины
9.1. Перечень основной и дополнительной учебной литературы
9.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся
9.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине
9.4. Оборудование и технические средства обучения

10. Методические указания по освоению дисциплины (модуля)

1. Цель и задачи освоения дисциплины (модуля)

Цель освоения дисциплины - формирование у обучающихся фундаментальных знаний в области машинного обучения и практических навыков разработки, обучения и развертывания моделей машинного обучения с использованием современных библиотек Python для решения задач анализа данных, прогнозирования и классификации в различных предметных областях.

Задачи изучения дисциплины:

- Освоение математических и программных основ анализа данных, включая работу с библиотеками NumPy для численных вычислений, Pandas для обработки и анализа структурированных данных, Matplotlib и Plotly для визуализации данных, а также изучение основных метрик оценки качества моделей в задачах регрессии и классификации.;
- Изучение и практическое применение классических алгоритмов машинного обучения с учителем, включая метод k-ближайших соседей (KNN), линейную и логистическую регрессию, решающие деревья, случайный лес, метод наивного Байеса, метод опорных векторов, а также освоение методов подбора оптимальных гиперпараметров и техник кросс-валидации.;
- Формирование навыков построения комплексных решений на основе ансамблевых методов машинного обучения, разработки систем классификации текстов и других типов данных, а также компетенций в области развертывания обученных ML-моделей в продуктивной среде с использованием современных инструментов и фреймворков..

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

Компетенции, индикаторы и результаты обучения

ОПК-1 Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности

ОПК-1.1 Разрабатывает и валидирует математические модели, основанные на естественнонаучных и общеинженерных принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

Знать:

ОПК-1.1/Зн1 Знание основных математических и физических законов и моделей, основанных на естественнонаучных и общеинженерных принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

Уметь:

ОПК-1.1/Ум1 Умение применять математические и физические законы и модели, основанные на естественнонаучных и общеинженерных принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

ОПК-1.1/Ум2 Решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования

Владеть:

ОПК-1.1/Нв1 Владение навыками применения основных математических и физических законов и моделей, основанных на естественнонаучных и общеинженерных принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

ОПК-1.2 Интегрирует естественнонаучные и общеинженерные знания с методами математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решений

Знать:

ОПК-1.2/Зн1 Знание методов математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решени

Уметь:

ОПК-1.2/Ум1 Умение применить методы математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решени

Владеть:

ОПК-1.2/Нв1 Владение навыками применения методов математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решени

ОПК-1.3 Планирует, проводит и интерпретирует теоретические и экспериментальные исследования, выбирая адекватные методы, собирая, обрабатывая и анализируя данные, и критически оценивает их достоверность

Знать:

ОПК-1.3/Зн1 Знание физическихзаконов и требований для планирования, проведения и интерпретации теоретических и экспериментальных исследований, выбора адекватных методов, обработки и анализа данных, и критической оценки их достоверности

Уметь:

ОПК-1.3/Ум1 Умение применить физические законы и требования для планирования, проведения и интерпретации теоретических и экспериментальных исследований, выбора адекватных методов, обработки и анализа данных, и критической оценки их достоверности

Владеть:

ОПК-1.3/Нв1 Владение навыками применения физических законов для планирования, проведения и интерпретации теоретических и экспериментальных исследований, выбора адекватных методов, обработки и анализа данных, и критической оценки их достоверности

ОПК-1.4 Использует современные цифровые инструменты для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

Знать:

ОПК-1.4/Зн1 Знание современных цифровых инструментов для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

Уметь:

ОПК-1.4/Ум1 Умение применения современных цифровых инструментов для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

Владеть:

ОПК-1.4/Нв1 Владение навыками применения современных цифровых инструментов для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

3. Место дисциплины в структуре ОП

Дисциплина (модуль) Б1.О.05.05 «Основы машинного обучения» относится к обязательной части образовательной программы и изучается в семестре(ах): 3.

В процессе изучения дисциплины студент готовится к решению типов задач профессиональной деятельности, предусмотренных ФГОС ВО и образовательной программой.

4. Объем дисциплины (модуля) и виды учебной работы

Период обучения	Общая трудоемкость (часы)	Общая трудоемкость (ЗЕТ)	Лекционные занятия (часы)	Практические занятия (часы)	Самостоятельная работа (часы)	Промежуточная аттестация (часы)
3 семестр	144	4	36	36	45	Экзамен (27)
Всего	144	4	36	36	45	27

5. Содержание дисциплины (модуля)

5.1. Разделы, темы дисциплины и виды занятий

(часы промежуточной аттестации не указываются)

Наименование раздела, темы	Всего	Лекционные занятия	Практические занятия	Самостоятельная работа	Планируемые результаты обучения, соотнесенные с результатами освоения программы
Раздел 1. Математические основы анализа данных	**6**	**2**	**2**	**2**	ОПК-1.3 ОПК-1.4
Тема 1.1. Библиотека NumPy для численных вычислений	3	1	1	1	
Тема 1.2. Статистические вычисления в NumPy	3	1	1	1	
Раздел 2. Визуализация и анализ данных	**12**	**4**	**4**	**4**	ОПК-1.3 ОПК-1.4
Тема 2.1. Визуализация данных с Matplotlib	3	1	1	1	
Тема 2.2. Интерактивная визуализация с Plotly	3	1	1	1	
Тема 2.3. Работа с табличными данными в Pandas	3	1	1	1	
Тема 2.4. Предобработка данных с Pandas	3	1	1	1	

Раздел 3. Основы теории машинного обучения	**9**	**3**	**3**	**3**	ОПК-1.3 ОПК-1.4
Тема 3.1. Введение в машинное обучение	3	1	1	1	
Тема 3.2. Метрики оценки качества моделей	3	1	1	1	
Тема 3.3. Валидация моделей и кросс-валидация	3	1	1	1	
Раздел 4. Методы обучения на основе расстояний	**6**	**2**	**2**	**2**	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 4.1. Метод k-ближайших соседей (KNN)	3	1	1	1	
Тема 4.2. Оптимизация гиперпараметров	3	1	1	1	
Раздел 5. Линейные модели	**12**	**4**	**4**	**4**	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 5.1. Линейная регрессия	6	2	2	2	
Тема 5.2. Логистическая регрессия	6	2	2	2	
Раздел 6. Древовидные модели	**16**	**5**	**5**	**6**	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 6.1. Решающие деревья для классификации	6	2	2	2	
Тема 6.2. Решающие деревья для регрессии	7	2	2	3	
Тема 6.3. Критерии останова и регуляризация деревьев	3	1	1	1	
Раздел 7. Ансамблевые методы	**21**	**6**	**6**	**9**	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 7.1. Случайный лес (Random Forest)	7	2	2	3	
Тема 7.2. Градиентный бустинг	7	2	2	3	
Тема 7.3. Стекинг и блендинг моделей	7	2	2	3	
Раздел 8. Вероятностные и дистанционные методы	**14**	**4**	**4**	**6**	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 8.1. Наивный байесовский классификатор	7	2	2	3	
Тема 8.2. Метод опорных векторов (SVM)	7	2	2	3	
Раздел 9. Практическое применение машинного обучения	**21**	**6**	**6**	**9**	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 9.1. Классификация текстовых данных	7	2	2	3	
Тема 9.2. Построение комплексных ML-решений	7	2	2	3	
Тема 9.3. Развертывание ML-моделей	7	2	2	3	
Итого	**117**	**36**	**36**	**45**	

5.2. Содержание разделов, тем дисциплин и формы текущего контроля

Раздел 1. Математические основы анализа данных

Тема 1.1. Библиотека NumPy для численных вычислений

Основы работы с многомерными массивами ndarray. Создание массивов различными способами. Индексация и срезы. Математические операции над массивами: поэлементные операции, матричные операции, векторизация вычислений. Функции агрегации. Broadcasting - правила транслирования. Линейная алгебра в NumPy.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 1.2. Статистические вычисления в NumPy

Генерация случайных чисел и распределений . Описательная статистика: среднее, медиана, дисперсия, стандартное отклонение. Корреляционный анализ. Работа с пропущенными значениями. Применение NumPy для предобработки данных в машинном обучении.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 2. Визуализация и анализ данных

Тема 2.1. Визуализация данных с Matplotlib

Основы построения графиков с библиотекой Matplotlib . Типы визуализаций: линейные графики, гистограммы, диаграммы рассеяния, box-plot. Настройка осей, легенд, заголовков. Создание подграфиков. Визуализация распределений и зависимостей в данных для машинного обучения.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 2.2. Интерактивная визуализация с Plotly

Создание интерактивных графиков с библиотекой Plotly . 3D-визуализации. Анимированные графики. Dashboard-визуализации. Визуализация результатов кластеризации и классификации. Построение интерактивных отчетов по результатам анализа.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 2.3. Работа с табличными данными в Pandas

Структуры данных Series и DataFrame . Загрузка данных из различных источников (CSV, Excel, SQL). Индексация и выборка данных. Группировка и агрегация. Объединение датафреймов. Работа с временными рядами. Pivot-таблицы.

Текущий контроль

Вид (форма) контроля

Просмотр выполненных работ

Тема 2.4. Предобработка данных с Pandas

Обработка пропущенных значений . Удаление дубликатов. Преобразование типов данных. Кодирование категориальных признаков. Нормализация и стандартизация данных. Feature engineering - создание новых признаков. Подготовка данных для машинного обучения.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 3. Основы теории машинного обучения

Тема 3.1. Введение в машинное обучение

Основные понятия и терминология машинного обучения. Типы задач: классификация, регрессия, кластеризация. Обучение с учителем и без учителя. Этапы разработки ML-модели. Проблемы переобучения и недообучения. Bias-variance tradeoff.

Тема 3.2. Метрики оценки качества моделей

Метрики для задач регрессии: MSE, RMSE, MAE, R^2 . Метрики для задач классификации: accuracy, precision, recall, F1-score. Матрица ошибок. ROC-кривая и AUC. Выбор оптимальной метрики в зависимости от бизнес-задачи. Интерпретация метрик качества.

Тема 3.3. Валидация моделей и кросс-валидация

Разделение данных на обучающую, валидационную и тестовую выборки. Методы кросс-валидации: k-fold, stratified k-fold, leave-one-out. Временная валидация для временных рядов. Оценка стабильности модели. Выбор оптимальной модели по результатам валидации.

Раздел 4. Методы обучения на основе расстояний

Тема 4.1. Метод k-ближайших соседей (KNN)

Алгоритм KNN для задач классификации и регрессии . Выбор метрики расстояния: евклидово, манхэттенское, расстояние Минковского. Влияние параметра k на качество модели. Взвешенный KNN. Масштабирование признаков. Применение KNN для рекомендательных систем.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 4.2. Оптимизация гиперпараметров

Методы подбора оптимальных гиперпараметров . Grid Search и Random Search. Байесовская оптимизация. Автоматизированный подбор гиперпараметров с помощью GridSearchCV и RandomizedSearchCV. Стратегии ускорения поиска. Визуализация результатов оптимизации.

Раздел 5. Линейные модели

Тема 5.1. Линейная регрессия

Метод наименьших квадратов . Множественная линейная регрессия. Полиномиальная регрессия. Регуляризация: Ridge, Lasso, ElasticNet. Градиентный спуск и его модификации. Диагностика модели: анализ остатков, проверка предположений.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 5.2. Логистическая регрессия

Бинарная и многоклассовая логистическая регрессия . Функция логистических потерь. Интерпретация коэффициентов. Регуляризация в логистической регрессии. Калибровка вероятностей. Применение для задач кредитного скоринга и медицинской диагностики.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 6. Древовидные модели

Тема 6.1. Решающие деревья для классификации

Алгоритм построения дерева решений . Критерии разбиения: Gini, энтропия. Обрезка дерева для борьбы с переобучением. Визуализация и интерпретация деревьев. Преимущества и недостатки деревьев решений.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 6.2. Решающие деревья для регрессии

Особенности деревьев регрессии . Критерии разбиения: MSE, MAE. Экстраполяционные свойства деревьев. Сравнение с линейными моделями. Применение для задач прогнозирования.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 6.3. Критерии останова и регуляризация деревьев

Параметры контроля сложности дерева . Минимальное количество объектов в листе. Максимальная глубина дерева. Минимальное улучшение критерия. Постпруning и препруning. Влияние параметров на качество модели.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 7. Ансамблевые методы

Тема 7.1. Случайный лес (Random Forest)

Принцип работы Random Forest . Бэггинг и случайные подпространства признаков. Out-of-bag оценка. Важность признаков. Параметры настройки: количество деревьев, глубина, количество признаков. Применение для задач с большим числом признаков.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 7.2. Градиентный бустинг

Основы бустинга . Алгоритм градиентного бустинга. XGBoost, LightGBM, CatBoost - сравнение реализаций. Настройка гиперпараметров бустинга. Early stopping. Обработка категориальных признаков.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 7.3. Стекинг и блендинг моделей

Методы комбинирования моделей . Усреднение предсказаний. Взвешенное голосование. Стекинг с мета-алгоритмом. Блендинг на валидационной выборке. Создание ансамблей для соревнований по машинному обучению.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 8. Вероятностные и дистанционные методы

Тема 8.1. Наивный байесовский классификатор

Теорема Байеса и ее применение . Гауссовский наивный Байес. Мультиномиальный и бернуллиевский варианты. Сглаживание Лапласа. Применение для классификации текстов и фильтрации спама.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 8.2. Метод опорных векторов (SVM)

Линейный SVM для разделимых и неразделимых выборок . Kernel trick и нелинейные SVM. Типы ядер: RBF, полиномиальное, сигмоидальное. Параметры C и gamma. SVM для регрессии (SVR). Многоклассовая классификация.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 9. Практическое применение машинного обучения

Тема 9.1. Классификация текстовых данных

Предобработка текста: токенизация, лемматизация, удаление стоп-слов. Векторизация текста: Bag of Words, TF-IDF, Word2Vec. Применение различных алгоритмов для классификации текстов. Анализ тональности. Классификация документов.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 9.2. Построение комплексных ML-решений

Создание end-to-end pipeline для машинного обучения . Автоматизация предобработки данных. Feature engineering и отбор признаков.

Тема 9.3. Развертывание ML-моделей

Сериализация обученных моделей с pickle и joblib. Создание REST API для моделей с Flask/FastAPI. Развертывание в облаке. Разработка графического интерфейса в Streamlit.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

6. Порядок проведения промежуточной аттестации

Порядок проведения промежуточной аттестации регламентируется локальными нормативными актами Университета:
- Положением о формах, периодичности, порядке текущего контроля успеваемости и промежуточной аттестации обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.6 «Организация и проведение промежуточной аттестации».
- Положением о порядке осуществления образовательного процесса, проведения текущего контроля успеваемости и промежуточной аттестации обучающихся с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.8 «Проведение промежуточной аттестации с применением ЭО и ДОТ».

Промежуточная аттестация: Экзамен, 3 семестр.
Экзамен по дисциплине проводится в формате письменного тестирования для итоговой оценки знаний.

Студентам будет предложен тест, состоящий из двадцати вопросов На выполнение всей работы отводится шестьдесят минут, в течение которых запрещается использование любых вспомогательных материалов и электронных устройств.

Итоговая оценка выставляется в зависимости от процента правильных ответов. Для получения оценки «отлично» необходимо верно ответить на 90-100% вопросов (18-20 правильных ответов). Оценка «хорошо» ставится за 70-89% правильных ответов (14-17), а

«удовлетворительно» — за 55-69% (11-13). Результат ниже 55% правильных ответов признается неудовлетворительным.

7. Оценочные материалы текущего контроля

Раздел 1. Математические основы анализа данных
Контролируемые ИДК: ОПК-1.3 ОПК-1.4

Тема 1.1. Библиотека NumPy для численных вычислений
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 1.2. Статистические вычисления в NumPy
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 2. Визуализация и анализ данных
Контролируемые ИДК: ОПК-1.3 ОПК-1.4

Тема 2.1. Визуализация данных с Matplotlib
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 2.2. Интерактивная визуализация с Plotly
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 2.3. Работа с табличными данными в Pandas
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 2.4. Предобработка данных с Pandas
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 4. Методы обучения на основе расстояний
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 4.1. Метод k-ближайших соседей (KNN)
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 5. Линейные модели
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 5.1. Линейная регрессия
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 5.2. Логистическая регрессия
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 6. Древовидные модели
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 6.1. Решающие деревья для классификации
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 6.2. Решающие деревья для регрессии

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 6.3. Критерии останова и регуляризация деревьев
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 7. Ансамблевые методы
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 7.1. Случайный лес (Random Forest)
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 7.2. Градиентный бустинг
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 7.3. Стекинг и блендинг моделей
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 8. Вероятностные и дистанционные методы
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 8.1. Наивный байесовский классификатор
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 8.2. Метод опорных векторов (SVM)
Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 9. Практическое применение машинного обучения
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 9.1. Классификация текстовых данных
Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 9.3. Развертывание ML-моделей
Форма контроля/оценочное средство: Просмотр выполненных работ

8. Оценочные материалы промежуточной аттестации

3 семестр, Экзамен
Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Вопросы/Задания:

1. Как в NumPy называется набор правил для выполнения математических операций над массивами разной формы (например, при сложении массива и скаляра)?

2. Как называется процесс приведения слов к их базовой (словарной) форме при предобработке текста?

3. Какая метрика качества классификации вычисляется как отношение верно предсказанных положительных объектов ко всем объектам, предсказанным моделью как положительные?

4. Как называется ситуация, когда модель показывает хорошие результаты на обучающей выборке, но плохие результаты на тестовой?

5. Дополните предложение терминами:
Метод регуляризации линейной регрессии _____ способен полностью обнулять коэффициенты наименее значимых признаков за счёт штрафа по норме L1, в отличие от _____, который лишь уменьшает величину коэффициентов, но не обнуляет их (штраф по норме L2).

6. Дайте определение переобучения (overfitting) и недообучения (underfitting) модели машинного обучения, поясните, как каждая из этих проблем проявляется в поведении ошибки на обучающей и тестовой выборках.

7. Опишите принцип работы алгоритма k-ближайших соседей (KNN) и объясните, как один и тот же алгоритм адаптируется для решения задачи регрессии и задачи классификации – что при этом меняется, а что остаётся общим.

8. Раскройте содержание метода наименьших квадратов при обучении линейной регрессии: дайте определение, поясните, что именно минимизируется, и опишите, как в эту схему встраивается регуляризация (Ridge, Lasso, ElasticNet).

9. Опишите принцип работы ансамблевого метода "случайный лес" (Random Forest): объясните роль бэггинга, случайного подпространства признаков и out-of-bag оценки.

10. Раскройте содержание процесса подготовки текстовых данных для классификации: опишите этапы предобработки текста (токенизация, удаление стоп-слов, лемматизация) и по крайней мере два способа векторизации текста (Bag of Words, TF-IDF, Word2Vec), укажите их различие.

11. Расположите этапы построения и оценки модели машинного обучения с использованием библиотек scikit-learn в правильной последовательности.
1. Обучение модели на обучающей выборке методом fit().
2. Разделение данных на обучающую и тестовую выборки.
3. Предобработка данных (масштабирование признаков, кодирование категориальных переменных).
4. Получение предсказаний на тестовой выборке методом predict().
5. Вычисление метрик качества модели по предсказаниям и истинным значениям тестовой выборки.

12. Расположите этапы работы наивного байесовского классификатора при классификации нового объекта в правильной последовательности.
1. Отнесение объекта к классу с максимальной апостериорной вероятностью.
2. Вычисление апостериорной вероятности каждого класса по теореме Байеса на основе значений признаков объекта.
3. Оценка по обучающей выборке априорных вероятностей классов и условных вероятностей признаков при каждом классе.

13. Расположите этапы развёртывания обученной ML-модели в виде REST API в правильной последовательности.
1. Создание конечной точки (endpoint) веб-приложения, принимающей входные данные и возвращающей предсказание модели.
2. Обучение и сохранение (сериализация) модели в файл, например средствами pickle или joblib.
3. Загрузка сохранённой модели при запуске веб-приложения (например, на Flask или FastAPI).

14. Расположите этапы построения градиентного бустинга в правильной последовательности.

1. Вычисление остатков (ошибок) текущего ансамбля на обучающей выборке.
2. Инициализация ансамбля простой базовой моделью (например, константным предсказанием).
3. Добавление новой модели в ансамбль с некоторым весом (шагом обучения).
4. Обучение новой модели предсказывать вычисленные остатки предыдущего ансамбля.

15. Расположите этапы построения ансамбля методом стекинга в правильной последовательности.

1. Обучение мета-алгоритма на предсказаниях базовых моделей как на новых признаках.
2. Разбиение обучающей выборки для получения непереобученных предсказаний базовых моделей (например, методом кросс-валидации).
3. Выбор и обучение нескольких разнородных базовых моделей.
4. Получение предсказаний базовых моделей на валидационных частях разбиения.
5. Формирование итогового предсказания путём применения обученного мета-алгоритма к предсказаниям базовых моделей на новых данных.
6. Сбор предсказаний базовых моделей в новую таблицу признаков (мета-признаков).

16. Установите соответствие между метрикой качества модели и её определением.

MSE (среднеквадратичная ошибка) - Доля верно предсказанных положительных объектов среди всех объектов, предсказанных как положительные
Precision (точность) - Интегральная характеристика качества ранжирования объектов по вероятности принадлежности положительному классу
Recall (полнота) - Средний квадрат отклонения предсказанных значений от истинных, применяется для задачи регрессии
AUC (площадь под ROC-кривой) - Доля верно найденных положительных объектов среди всех истинно положительных объектов

17. Установите соответствие между алгоритмом машинного обучения и его ключевым механизмом работы.

Random Forest - Последовательное обучение моделей, каждая из которых исправляет остаточные ошибки предыдущего ансамбля
Градиентный бустинг - Последовательное разбиение пространства признаков по одному признаку за раз на основе критерия информативности разбиения
SVM - Вычисление апостериорной вероятности класса по теореме Байеса в предположении независимости признаков друг от друга
Наивный Байес - Усреднение предсказаний множества деревьев, обученных на случайных подвыборках объектов и признаков (бэггинг)
Решающее дерево - Поиск разделяющей гиперплоскости с максимальным отступом (зазором) от ближайших объектов классов

18. Установите соответствие между типом ядра метода опорных векторов и его особенностью.

RBF-ядро (радиальная базисная функция) - Строит разделяющую границу в виде гиперплоскости непосредственно в исходном пространстве признаков без их преобразования
Линейное ядро - Строит нелинейную границу через комбинации признаков заданной степени, менее гибко и вычислительно затратнее на высоких степенях, чем RBF
Полиномиальное ядро - Позволяет строить нелинейную разделяющую границу произвольной формы, отображая объекты в пространство более высокой (потенциально бесконечной) размерности

19. Установите соответствие между библиотекой Python и её основным назначением в разработке ML-решений.

NumPy - Построение интерактивных и анимированных визуализаций данных
Pandas - Разработка графического веб-интерфейса для демонстрации и использования обученной модели
Plotly - Загрузка, хранение и предобработка табличных данных в структурах Series и

DataFrame
Streamlit - Численные вычисления над многомерными массивами, включая векторизованные математические и линейно-алгебраические операции

20. Установите соответствие между гиперпараметром решающего дерева и тем, что он ограничивает.

Максимальная глубина дерева - Запрещает создание листовых узлов, содержащих меньше заданного числа обучающих объектов
Минимальное количество объектов в листе - Запрещает выполнять разбиение узла, если оно не улучшает критерий разбиения хотя бы на заданную величину
Минимальное улучшение критерия разбиения - Ограничивает число последовательных разбиений от корня до листа

9. Материально-техническое и учебно-методическое обеспечение дисциплины

9.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Титов А. Н. Обработка данных в Python. Основы работы с библиотекой Pandas: учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. - Казань: КНИТУ, 2022. - 116 с. - 978-5-7882-3164-8. - Текст: электронный // Издательство Лань: [сайт]. - URL: https://e.lanbook.com/book/331013 - Режим доступа: по подписке

Дополнительная литература

1. Пальмов С. В. Интеллектуальный анализ данных: учебное пособие / С. В. Пальмов. - Интеллектуальный анализ данных - Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. - 127 c. - 2227-8397. - Текст: электронный // IPR SMART: [сайт]. - URL: https://www.iprbookshop.ru/75376.html - Режим доступа: по подписке

9.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся

Профессиональные базы данных

Не используются.

Ресурсы «Интернет»

1. https://docs.python.org/3/index.html - Техническая документация по языку программирования Python
2. https://sdo.miigaik.ru/ - Система дистанционного обучения МИИГАиК

9.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения
(обновление производится по мере появления новых версий программы)

1. Anaconda;
2. Microsoft Visual Studio;

Перечень информационно-справочных систем
(обновление выполняется еженедельно)

1. 1С Предприятие;

9.4. Оборудование и технические средства обучения

Компьютерный класс
Маркерная доска - 1 шт.
Моноблок - 96 шт.
Проектор - 8 шт.
Автоматизированное рабочее место (АРМ) - 40 шт.
Плоттер - 1 шт.
Сканер - 1 шт.
Интерактивная панель - 1 шт.

10. Методические указания по освоению дисциплины (модуля)

С целью успешного освоения данного предмета студенту необходимо выполнять домашние работы по указанным преподавателям темам и представленным на занятиях материалам. Для более успешного выполнения работы студенту предлагается использование учебно-методических материалов, соответствующих указанной теме.

Самостоятельная работа студентов описывается и регулируется методическими рекомендациями по дисциплине.

Перечень вопросов для промежуточной аттестации:
1. Как в NumPy называется набор правил для выполнения математических операций над массивами разной формы?
2. Какие две основные структуры данных предоставляет библиотека Pandas?
3. Как называется процесс создания новых признаков из существующих данных при работе с Pandas?
4. Какой тип визуализации НЕ упоминается в качестве базового для библиотеки Matplotlib в предоставленном содержании?
5. Что из перечисленного НЕ является основной задачей машинного обучения?
6. Как называется проблема, когда модель отлично работает на обучающих данных, но плохо справляется с новыми, неизвестными данными?
7. Какая из этих метрик обычно используется для оценки качества моделей регрессии?
8. Какая пара метрик является ключевой для оценки моделей классификации и часто используется при построении матрицы ошибок?
9. Какова основная цель k-fold кросс-валидации?
10. Что означает параметр 'k' в алгоритме k-ближайших соседей (KNN)?
11. Какой метод оптимизации гиперпараметров перебирает все возможные комбинации из заданного набора значений?
12. Какой тип регуляризации в линейной регрессии способен полностью обнулять коэффициенты у наименее важных признаков?
13. Какой из перечисленных критериев используется для разделения узлов в дереве решений для задачи классификации?
14. Как называется процесс удаления ветвей из дерева решений для борьбы с переобучением?
15. В чем заключается основной принцип работы ансамблевого метода "Случайный лес" (Random Forest)?
16. Что из перечисленного является популярной реализацией алгоритма градиентного бустинга?
17. Какую роль выполняет "мета-алгоритм" в методе стекинга?
18. Для какой задачи, согласно содержанию, особенно хорошо подходит наивный байесовский классификатор?
19. В чем заключается "трюк с ядром" (kernel trick) в методе опорных векторов (SVM)?
20. Как расшифровывается аббревиатура TF-IDF, используемая при векторизации текстов?
21. Какая библиотека упоминается для сериализации (сохранения) обученных моделей

машинного обучения?
22. Какая технология упоминается для создания контейнеров, позволяющих легко развертывать ML-модели?
23. Какая ключевая особенность библиотеки Plotly по сравнению с Matplotlib отмечена в программе?
24. Какой алгоритм оптимизации упоминается как фундаментальный для обучения линейных моделей?
25. В чем заключается основная идея "бустинга"?
26. Как называется процесс приведения слов к их базовой или словарной форме?
27. Компромисс между способностью модели минимизировать ошибку на обучающем наборе и ее способностью обобщаться на новые данные известен как:
28. Что из перечисленного является типом ядра, используемого в методе опорных векторов (SVM)?
29. Какова цель создания REST API для модели машинного обучения?
30. Что является основной структурой данных в библиотеке NumPy?

Описание возможностей изучения дисциплины лицами с ОВЗ и инвалидами

При необходимости обучения студентов-инвалидов и лиц с ограниченными возможностями здоровья аудиторные занятия могут быть заменены или дополнены изучением полнотекстовых лекций, презентаций, видео- и аудиоматериалов в электронной информационно-образовательной среде (ЭИОС) университета. Индивидуальные задания подбираются в адаптированных к ограничениям здоровья формах (письменно или устно, в форме презентаций). Выбор методов обучения зависит от их доступности для инвалидов и лиц с ограниченными возможностями здоровья.

Рабочая программа дисциплины для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.