

Документ подписан посредством электронной подписи
Информация о владельце:
ФИО: Шамрай-Курбатова Лидия Викторовна
Должность: Ректор
Дата подписания: 11.06.2026 14:05:44
Уникальный программный ключ:
b1e4399771b07e18f31755456972d73b2ccfc531

Автономная некоммерческая организация высшего образования
«Волгоградский институт бизнеса»

Рабочая программа учебной дисциплины

Машинное обучение

(Наименование дисциплины)

54.03.01 Дизайн, направленность (профиль) «Цифровой дизайн»

(Направление подготовки / Профиль)

Бакалавр

(Квалификация)

Кафедра разработчик

Экономики и управления

Год набора

2026

Вид учебной деятельности	Трудоемкость (объем) дисциплины	
	Очная форма	Очно-заочная форма
	Д	В
Зачетные единицы	6	6
Общее количество часов	216	216
Аудиторные часы контактной работы обучающегося с преподавателями:	64	48
– Лекционные (Л)	32	24
– Практические (ПЗ)	32	24
– Лабораторные (ЛЗ)		
– Семинарские (СЗ)		
Самостоятельная работа обучающихся (СРО)	116	132
К (Р-Г) Р (П) (+;-)		
Тестирование (+;-)		
ДКР (+;-)		
Зачет (+;-)		
Зачет с оценкой (+;- (Кол-во часов))		
Экзамен (+;- (Кол-во часов))	+(36)	+(36)

Волгоград 2026

Содержание

Раздел 1. Организационно-методический раздел	3
Раздел 2. Тематический план.....	5
Раздел 3. Содержание дисциплины.....	7
Раздел 4. Организация самостоятельной работы обучающихся.....	17
Раздел 5. Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся.....	19
Раздел 6. Оценочные средства промежуточной аттестации (с ключами)	19
Раздел 7. Перечень учебной литературы, необходимой для освоения дисциплины	21
Раздел 8. Материально-техническая база и информационные технологии.....	25
Раздел 9. Методические указания для обучающихся по освоению дисциплины	27

Раздел 1. Организационно-методический раздел

1.1. Цели освоения дисциплины

Дисциплина «Машинное обучение» входит в перечень **Обязательных дисциплин** подготовки обучающихся по направлению **Дизайн, направленность (профиль) «Цифровой дизайн»**.

Целью дисциплины является формирование **компетенций** (в соответствии с ФГОС ВО и требованиями к результатам освоения основной профессиональной образовательной программы высшего образования (ОПОП ВО)):

ОПК-2 Способен работать с научной литературой; собирать, анализировать и обобщать результаты научных исследований; оценивать полученную информацию; самостоятельно проводить научно-исследовательскую работу; участвовать в научно-практических конференциях.

Дескрипторы общепрофессиональных компетенций:

ОПК-2.1 Способен собирать, анализировать и обобщать результаты исследований в области графического и цифрового дизайна, включая данные пользовательской аналитики и А/В-тестирования, включая анализ целесообразности применения методов искусственного интеллекта, формирование требований к данным и метрикам качества моделей.

ОПК-2.2 Способен проводить прикладные исследования (аудит пользовательских сценариев, конкурентный анализ), выполнять постановку целей, разработку концепции системы, разработку технического задания на создание цифрового продукта, в том числе с помощью ИИ-систем; участвовать в научных и профессиональных конференциях по дизайну и цифровым инновациям.

Перечисленные компетенции формируются в процессе достижения **индикаторов компетенций**:

Обобщенная трудовая функция/ трудовая функция	Код и наименование дескриптора компетенций	Код и наименование индикатора достижения компетенций (из ПС)
ОПК-2 Способен работать с научной литературой; собирать, анализировать и обобщать результаты научных исследований; оценивать полученную информацию; самостоятельно проводить научно-исследовательскую работу; участвовать в научно-практических конференциях	ОПК-2.1 Способен собирать, анализировать и обобщать результаты исследований в области графического и цифрового дизайна, включая данные пользовательской аналитики и А/В-тестирования, включая анализ целесообразности применения методов искусственного интеллекта, формирование требований к данным и метрикам качества моделей ОПК-2.2 Способен проводить прикладные исследования (аудит пользовательских сценариев, конкурентный анализ), выполнять постановку целей, разработку концепции системы, разработку технического задания на создание цифрового продукта, в том числе с помощью ИИ-	Знание: ИД-1 ОПК-2.1 Методика поиска, сбора и анализа информации, необходимой для разработки проектного задания на создание объектов визуальной информации, идентификации и коммуникации В/01.6 ИД-2 ОПК-2.2 Подходы к постановке целей, разработке концепции системы и технического задания на создание цифрового продукта, в том числе с помощью ИИ-систем (без привязки к профессиональному стандарту) Умения: ИД-3 ОПК-2.1 Использовать специальные компьютерные программы для проектирования объектов визуальной информации, идентификации и коммуникации В/02.6 ИД-4 ОПК-2.2 Использовать подходы к постановке целей, разработке концепции системы и технического задания на создание цифрового продукта, в том числе с помощью ИИ-систем (без привязки к профессиональному стандарту) Навыки и (или)опыт деятельности: ИД-5 ОПК-2.1 Изучение информации, необходимой для работы над дизайн-

	систем; участвовать в научных и профессиональных конференциях по дизайну и цифровым инновациям	проектом объекта визуальной информации, идентификации и коммуникации В/02.6 ИД-6 ОПК-2.2 Владение навыками постановки целей, разработки концепции системы и технического задания на создание цифрового продукта, в том числе с помощью ИИ-систем (без привязки к профессиональному стандарту)
--	--	--

**1.2. Место дисциплины в структуре ОПОП ВО
направления подготовки 54.03.01 Дизайн, направленность (профиль) «Цифровой дизайн»**

№	Предшествующие дисциплины (дисциплины, изучаемые параллельно)	Последующие дисциплины
1	2	3
1	Математический анализ и моделирование	Производственная практика (Проектно-технологическая практика)
2	Python: основные библиотеки для визуализации данных	Производственная практика (Преддипломная практика)
3		Подготовка к процедуре защиты и защита выпускной квалификационной работы

Последовательность формирования компетенций в указанных дисциплинах может быть изменена в зависимости от формы и срока обучения, а также преподавания с использованием дистанционных технологий обучения.

1.3. Нормативная документация

Рабочая программа учебной дисциплины составлена на основе:

- Федерального государственного образовательного стандарта высшего образования по направлению подготовки **54.03.01 Дизайн**;
- Учебного плана направления подготовки **54.03.01 Дизайн, направленность (профиль) «Цифровой дизайн»** 2026 годов набора;
- Образца рабочей программы учебной дисциплины (приказ № 113-О от 01.09.2021 г.).

Раздел 2. Тематический план

Очная форма обучения (полный срок)

№	Тема дисциплины	Трудоемкость				Код индикатора и дескриптора достижения компетенций
		Всего	Аудиторные занятия		СРО	
			Л	ПЗ (ЛЗ, СЗ)		
1	2	3	4	5	6	7
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	22	4	4	14	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	22	4	4	14	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	22	4	4	14	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	22	4	4	14	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	22	4	4	14	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
6	Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение	22	4	4	14	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	24	4	4	16	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей	24	4	4	16	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
Вид промежуточной аттестации (Экзамен)		+(36)				
Итого		216	32	32	116	

Очно-заочная форма обучения (полный срок)

№	Тема дисциплины	Трудоемкость				Код индикатора и дескриптора достижения компетенций
		Всего	Аудиторные занятия		СРО	
			Л	ПЗ (ЛЗ, СЗ)		
1	2	3	4	5	6	7
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	22	4	4	14	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2

2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	22	4	4	14	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	22	4	4	14	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	22	4	4	14	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	22	2	2	18	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
6	Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение	22	2	2	18	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	24	2	2	20	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей	24	2	2	20	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
Вид промежуточной аттестации (Экзамен)		+(36)				
Итого		216	24	24	132	

Раздел 3. Содержание дисциплины

3.1. Содержание дисциплины

Тема 1. Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей

Определение машинного обучения как подсистемы искусственного интеллекта, способной обучаться на данных без явного программирования правил. Ключевые понятия: признак (feature), целевая переменная (target), объект, выборка (обучающая, валидационная, тестовая). Типы задач: регрессия (прогнозирование непрерывной величины), классификация (отнесение к категориям), кластеризация (группировка без учителя). Дополнительные типы: задачи ранжирования, обнаружения аномалий, понижения размерности. Этапы построения модели: постановка задачи, сбор и предобработка данных, выбор модели, обучение, оценка качества, настройка гиперпараметров, эксплуатация и мониторинг. Роль инженерного подхода: экспериментирование, воспроизводимость, итеративное улучшение. Основные вызовы: переобучение (overfitting), недообучение (underfitting), дисбаланс классов, дрейф данных. История развития машинного обучения: от перцептрона до трансформеров и больших языковых моделей.

Тема 2. Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков

Качество данных как основной фактор успеха моделирования. Принцип GIGO (Garbage In, Garbage Out). Очистка данных: удаление дубликатов, исправление опечаток, фильтрация выбросов (методы IQR, Z-оценка). Типы пропусков: MCAR (случайные), MAR (условно-случайные), NMAR (неслучайные). Методы работы с пропусками: удаление строк/столбцов, заполнение константой, средним/медианой/модой, интерполяция, предсказание пропусков с помощью моделей. Нормализация и масштабирование: необходимость для градиентных методов. StandardScaler (Z-нормализация), MinMaxScaler (приведение к отрезку), RobustScaler (устойчив к выбросам), Normalizer (нормировка по длине вектора). Кодирование категориальных признаков: LabelEncoder (порядковое кодирование), OneHotEncoder (бинарные столбцы), OrdinalEncoder (для порядковых переменных), целевое кодирование (mean encoding). Создание новых признаков (feature engineering): полиномиальные признаки, взаимодействия, агрегирование, признаки на основе даты и времени. Обработка текстовых и изображений данных. Балансировка классов: методы ресемплинга (oversampling, undersampling, SMOTE).

Тема 3. Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)

Линейная регрессия: модель вида $y = w * x + b$. Метод наименьших квадратов (MSE — Mean Squared Error). Аналитическое решение (нормальные уравнения) и градиентный спуск. Предположения линейной регрессии (линейность, нормальность остатков, гомоскедастичность, независимость ошибок). Оценка качества: R^2 , скорректированный R^2 , MAE, RMSE, MAPE. Логистическая регрессия: модель для бинарной классификации. Сигмоидная функция (логистическая функция) для преобразования линейной комбинации в вероятность. Функция потерь: Log Loss (бинарная кросс-энтропия). Граница принятия решений. Мультиклассовая логистическая регрессия: one-vs-rest (OvR) и softmax (мультиномиальная логистическая регрессия). Проблема переобучения линейных моделей. Регуляризация: L1 (Lasso) — приводит к разреженным коэффициентам, L2 (Ridge) — штрафует квадраты коэффициентов, ElasticNet — комбинация L1 и L2. Выбор силы регуляризации (гиперпараметр C или alpha). Интерпретация коэффициентов: вклад каждого признака в предсказание. Достоинства линейных моделей: простота, интерпретируемость, быстрота обучения и инференса. Недостатки: неспособность моделировать нелинейные зависимости без преобразования признаков.

Тема 4. Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения

Деревья решений: принцип работы (рекурсивное разбиение пространства признаков). Крите-

рии разбиения: энтропия, Gini impurity (классификация), дисперсия (регрессия). Параметры дерева: максимальная глубина, минимальное количество объектов в листе, минимальное количество объектов для разбиения. Преимущества: интерпретируемость, работа с разными типами признаков, отсутствие необходимости масштабирования. Недостатки: склонность к переобучению, нестабильность (малые изменения данных меняют структуру). Ансамблевые методы: объединение нескольких моделей для улучшения качества. Бэггинг (Bagging): случайный лес — построение множества деревьев на случайных подвыборках данных и случайных подмножествах признаков. Усреднение предсказаний, снижение дисперсии. Оценка важности признаков в случайном лесу. Бустинг (Boosting): последовательное обучение моделей, каждая следующая исправляет ошибки предыдущих. Градиентный бустинг (GBM, XGBoost, LightGBM, CatBoost): градиентный спуск в пространстве функций. Преимущества градиентного бустинга: высокая точность, обработка пропусков, встроенная регуляризация. Борьба с переобучением в деревьях: ограничение глубины, увеличение минимального числа объектов в листе, early stopping, субдискретизация (subsampling), регуляризация. Сравнение случайного леса и градиентного бустинга: скорость обучения, интерпретируемость, устойчивость к выбросам.

Тема 5. Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров

Кластеризация как задача обучения без учителя: группировка объектов на основе их сходства. Алгоритм k-средних (k-means): инициализация центроидов (случайный выбор, k-means++), шаг E (назначение объектов ближайшему центроиду), шаг M (пересчёт центроидов). Критерий остановки (стабилизация центроидов или минимизация WCSS — Within-Cluster Sum of Squares). Проблема выбора числа кластеров k: метод локтя (elbow method), силуэтный коэффициент (silhouette score), индекс Калински-Харабаса. Преимущества k-средних: простота, масштабируемость. Недостатки: чувствительность к инициализации и выбросам, предположение о сферичности кластеров, необходимость задавать k. Иерархическая кластеризация: агломеративные (снизу вверх) и дивизивные (сверху вниз) методы. Метрики расстояния: евклидово, манхэттенское, косинусное. Правила объединения кластеров: single linkage (ближайший сосед), complete linkage (дальний сосед), average linkage (среднее), метод Уорда (минимизация дисперсии). Визуализация: дендрограмма, выбор числа кластеров по высоте разреза. Оценка качества кластеризации без меток: силуэтный коэффициент, индекс Данна, индекс Дэвиса-Болдина. Оценка с метками (если есть): adjusted Rand index (ARI), adjusted mutual information (AMI), homogeneity, completeness, V-measure. Применение кластеризации: сегментация клиентов, анализ документов, сжатие изображений, обнаружение аномалий.

Тема 6. Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение

Понятие обобщающей способности модели (generalization). Необходимость оценки качества на данных, не участвовавших в обучении. Разделение выборки: тренировочная, валидационная, тестовая. Метрики для классификации: матрица ошибок (confusion matrix) — TP, TN, FP, FN. Accuracy (доля правильных ответов) — проблема дисбаланса классов. Precision (точность) — доля релевантных среди предсказанных положительных. Recall (полнота) — доля найденных положительных объектов. F1-мера — гармоническое среднее precision и recall. Специфичность (specificity) — доля верно отнесённых отрицательных объектов. ROC-кривая и AUC-ROC (площадь под кривой) для оценки разделяющей способности. PR-кривая (precision-recall) для сильно несбалансированных задач. Метрики для регрессии: MAE (средняя абсолютная ошибка), MSE (средняя квадратичная ошибка), RMSE (корень из MSE), MAPE (средняя абсолютная процентная ошибка), R^2 (коэффициент детерминации). Метрики для кластеризации (внутренние и внешние). Кросс-валидация: k-fold, stratified k-fold (для сохранения пропорции классов), leave-one-out, временная кросс-валидация. Оценка стабильности модели. Переобучение (overfitting): модель слишком сложная, хорошо запоминает обучающую выборку, плохо обобщает. Признаки: высокое качество на обучении, низкое — на тесте. Недообучение (underfitting): модель слишком простая, не выявляет закономерности. Признаки: низкое качество и на обучении, и на тесте. Диагностика с помощью кривых обучения (learning curves) и кривых валидации (validation curves). Стратегии

борьбы с переобучением: упрощение модели, регуляризация, увеличение данных, аугментация, early stopping. Стратегии борьбы с недообучением: усложнение модели, добавление признаков, уменьшение регуляризации.

Тема 7. Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки

Искусственный нейрон как вычислительная единица: взвешенная сумма входов, добавление смещения (bias), применение функции активации. Многослойный перцептрон (MLP): входной слой, скрытые слои, выходной слой. Полносвязные (dense) слои. Функции активации: сигмоида (Sigmoid) — сжимает выход в (0,1), гиперболический тангенс (Tanh) — в (-1,1), ReLU (Rectified Linear Unit) — устраняет проблему затухающего градиента, Leaky ReLU, ELU, Swish, Softmax (для многоклассовой классификации, превращает выходы в вероятности). Функции потерь: MSE (регрессия), бинарная кросс-энтропия (бинарная классификация), категориальная кросс-энтропия (многоклассовая классификация). Обучение нейронных сетей: градиентный спуск (SGD — Stochastic Gradient Descent), оптимизаторы: Momentum, Adam, RMSprop, AdaGrad. Backpropagation (обратное распространение ошибки): вычисление градиентов с помощью цепного правила, распространение ошибки от выхода ко входам. Проблема затухающего градиента (vanishing gradient) и взрывного градиента (exploding gradient). Инициализация весов (Xavier, He). Регуляризация в нейронных сетях: L1, L2 (weight decay), Dropout (случайное отключение нейронов), Batch Normalization (нормализация активаций). Early stopping как способ предотвращения переобучения. Понятие эпохи, батча, learning rate, планировщик скорости обучения. Выбор архитектуры: количество слоёв, количество нейронов на слое. Применение нейронных сетей: компьютерное зрение (CNN), обработка последовательностей (RNN, LSTM, Transformer), генеративные модели (GAN, VAE, диффузионные модели). Ограничения нейронных сетей: необходимость больших объёмов данных, вычислительные ресурсы, неинтерпретируемость.

Тема 8. Применение машинного обучения: работа с библиотеками, построение и внедрение моделей

Экосистема Python для машинного обучения: NumPy (численные вычисления), Pandas (работа с данными), Matplotlib, Seaborn (визуализация), Scikit-learn (классические алгоритмы), XGBoost, LightGBM, CatBoost (градиентный бустинг), TensorFlow, Keras, PyTorch (глубокое обучение). Пайплайн обучения модели: импорт библиотек, загрузка и предобработка данных, разделение выборки, обучение, оценка, настройка гиперпараметров. Использование Pipeline из Scikit-learn для объединения этапов предобработки и моделирования. Настройка гиперпараметров: GridSearchCV (полный перебор), RandomizedSearchCV (случайный поиск), Bayesian Optimization (Optuna, Hyperopt). Сохранение и загрузка моделей: pickle, joblib, onnx, сохранение весов нейронных сетей (h5, SavedModel). Внедрение модели в продуктивную среду: упаковка в контейнер (Docker), развертывание REST API (Flask, FastAPI, TorchServe, TF Serving), пакетный (batch) инференс, использование в облачных платформах (AWS SageMaker, Google Vertex AI, Yandex DataSphere). Мониторинг моделей в продакшене: детекция дрейфа данных (Data Drift) и концепта (Concept Drift), отслеживание метрик качества, логирование предсказаний, алертинг. Экспериментирование и версионирование: MLflow, Weights & Biases, DVC. MLOps: непрерывная интеграция и доставка моделей. Реальные кейсы применения: кредитный скоринг, рекомендательные системы, обнаружение мошенничества, прогнозирование спроса, медицинская диагностика, обработка естественного языка. Этические аспекты внедрения ML: предвзятость (bias), справедливость (fairness), объяснимость (explainability), приватность данных. Оценка экономической эффективности ML-решений.

3.2. Содержание практического блока дисциплины

Очная форма обучения (полный срок)

№	Тема практического (семинарского, лабораторного) занятия
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, класси-

	фикация, кластеризация), этапы построения моделей
ПЗ 1	Знакомство со средой разработки: настройка Jupyter Notebook или VS Code, установка необходимых библиотек (numpy, pandas, matplotlib, seaborn, scikit-learn). Загрузка и первичное исследование датасетов (например, Iris, Boston Housing, Digits). Разделение данных на обучающую и тестовую выборки с помощью train_test_split. Реализация простейших моделей: линейная регрессия для предсказания стоимости жилья, логистическая регрессия для классификации ирисов, KMeans для кластеризации цифр. Сравнение качества моделей на обучающей и тестовой выборках. Визуализация результатов классификации и кластеризации. Обсуждение этапов построения моделей на практике: постановка задачи, подготовка данных, выбор модели, обучение, оценка. Анализ ошибок при неправильном разделении выборки или отсутствии предобработки.
ПЗ 2	Загрузка «грязного» датасета с пропусками, выбросами и категориальными признаками (например, Titanic, Adult Income). Анализ структуры данных: info, describe, isnull, unique. Очистка данных: удаление дубликатов, исправление типов данных. Визуализация выбросов с помощью boxplot и гистограмм, применение методов IQR и Z-оценки для их обнаружения, принятие решений об удалении или трансформации. Работа с пропусками: сравнение стратегий удаление строк (dropna), заполнение средним/медианой/модой (fillna), интерполяция, использование SimpleImputer из scikit-learn. Масштабирование признаков: применение StandardScaler, MinMaxScaler, RobustScaler, визуализация распределений до и после масштабирования. Кодирование категориальных признаков: LabelEncoder (порядковое кодирование для ordinal-переменных), OneHotEncoder (создание dummy-переменных), pd.get_dummies. Создание новых признаков: извлечение из даты (день недели, месяц, год), бинаризация непрерывных признаков, создание полиномиальных признаков с PolynomialFeatures. Оценка влияния предобработки на качество простой модели (логистическая регрессия). Оформление пайплайна предобработки с использованием ColumnTransformer.
ПЗ 3	Реализация линейной регрессии с использованием LinearRegression из scikit-learn на датасете с несколькими признаками. Вычисление коэффициентов модели и их интерпретация. Визуализация предсказаний vs реальных значений. Оценка качества с помощью MSE, MAE, R ² . Реализация логистической регрессии с LogisticRegression для бинарной классификации. Визуализация разделяющей поверхности для двух признаков. Интерпретация вероятностей (predict_proba) и границы принятия решений. Функции потерь: реализация MSE для линейной регрессии и Log Loss для логистической регрессии вручную (без использования готовых библиотек). Применение градиентного спуска: реализация простой версии с настройкой learning rate и числа итераций, сравнение с аналитическим решением. Регуляризация L1 (Lasso) и L2 (Ridge): обучение моделей с различными значениями параметра alpha, визуализация изменения коэффициентов (path plot). Сравнение разреженности коэффициентов при L1 и L2. Подбор оптимального alpha с помощью кросс-валидации (LassoCV, RidgeCV). Сравнение ElasticNet как комбинации L1 и L2. Анализ влияния масштабирования признаков на качество линейных моделей. Применение полиномиальных признаков (PolynomialFeatures) для моделирования нелинейных зависимостей, сравнение качества с обычной линейной регрессией.
ПЗ 4	Реализация дерева решений (DecisionTreeClassifier, DecisionTreeRegressor) на датасете с визуализацией структуры дерева (plot_tree, export_graphviz). Анализ влияния параметров (max_depth, min_samples_split, min_samples_leaf) на переобучение: построение кривых валидации. Реализация случайного леса (RandomForestClassifier, RandomForestRegressor). Оценка важности признаков (feature_importances_), визуализация в виде барплота. Сравнение качества случайного леса с одним деревом решений. Реализация градиентного бустинга: сравнение GradientBoostingClassifier, XGBoost, LightGBM, CatBoost. Подбор гиперпараметров (learning_rate, n_estimators,

	<p>max_depth, subsample) с помощью GridSearchCV и RandomizedSearchCV. Использование early stopping для предотвращения переобучения. Визуализация кривых обучения (learning curves) для диагностики переобучения и недообучения. Применение бэггинга и бустинга к одному и тому же датасету, сравнение времени обучения и качества. Использование кросс-валидации для оценки стабильности ансамблевых методов. Обработка дисбаланса классов с помощью class_weight и параметров scale_pos_weight в XGBoost.</p>
ПЗ 5	<p>Реализация алгоритма k-средних (KMeans) на синтетическом датасете (make_blobs) с визуализацией кластеров. Изучение проблемы инициализации: сравнение random и k-means++. Анализ влияния выбросов на центроиды. Выбор числа кластеров: метод локтя (инерция vs число кластеров), силуэтный коэффициент (silhouette_score), индекс Калински-Харабаса. Визуализация силуэтных диаграмм (silhouette_plot). Реализация иерархической кластеризации (AgglomerativeClustering). Построение дендрограммы с помощью scipy.cluster.hierarchy.dendrogram. Выбор числа кластеров по высоте разреза дендрограммы. Сравнение различных правил объединения (linkage): single, complete, average, ward. Применение кластеризации к реальному датасету (например, сегментация клиентов по поведенческим признакам). Нормализация признаков перед кластеризацией, сравнение результатов с и без масштабирования. Визуализация кластеров в 2D с помощью PCA и t-SNE. Оценка качества кластеризации с помощью внешних метрик (adjusted_rand_score, adjusted_mutual_info_score) если есть истинные метки, или внутренних (silhouette_score, davies_bouldin_score) если меток нет. Обсуждение интерпретации полученных кластеров и их практического применения.</p>
ПЗ 6	<p>Реализация задач классификации на датасете с дисбалансом классов (например, кредитный скоринг, обнаружение мошенничества). Вычисление матрицы ошибок (confusion_matrix) и её визуализация с помощью seaborn.heatmap. Расчёт метрик: accuracy, precision, recall, f1_score вручную и с использованием scikit-learn. Анализ ситуации, когда accuracy неинформативна, сравнение precision и recall. Построение ROC-кривой (roc_curve) и вычисление AUC-ROC (roc_auc_score). Сравнение ROC-AUC и PR-AUC (precision-recall curve) для несбалансированных данных. Кросс-валидация: реализация k-fold (KFold), stratified k-fold (StratifiedKFold) с сохранением пропорции классов. Сравнение результатов кросс-валидации с простым разбиением train/test. Использование cross_val_score, cross_validate. Диагностика переобучения и недообучения: построение кривых обучения (learning_curve) — график зависимости качества от размера обучающей выборки. Построение кривых валидации (validation_curve) для выбора гиперпараметра (например, max_depth дерева). Анализ графиков: зоны недообучения (низкое качество на трейне и тесте), переобучения (разрыв между трейном и тестом). Применение GridSearchCV и RandomizedSearchCV для подбора гиперпараметров с кросс-валидацией. Сравнение времени выполнения и качества различных методов поиска.</p>
ПЗ 7	<p>Знакомство с библиотеками глубокого обучения: установка TensorFlow и PyTorch. Создание простейшей нейронной сети (многослойный перцептрон) для классификации рукописных цифр MNIST. Определение архитектуры: входной слой ($28 \times 28 = 784$ нейрона), скрытые слои (например, 128 нейронов с ReLU), выходной слой (10 нейронов с Softmax). Компиляция модели: выбор оптимизатора (Adam, SGD), функции потерь (categorical_crossentropy), метрик (accuracy). Обучение модели: метод fit, определение эпох и батчей. Визуализация кривых обучения (потери и точность на обучающей и валидационной выборках). Эксперименты с различными функциями активации: ReLU, Sigmoid, Tanh, сравнение скорости сходимости и итогового качества. Изучение проблемы затухающего градиента на глубоких сетях. Реализация Dropout и BatchNormalization для борьбы с переобучением, сравнение результатов. Настройка скорости обучения: использование learning rate scheduling (ReduceLROnPlateau, ExponentialDecay). Реализация обратного распространения ошибки вручную (для простой сети с одним скрытым слоем) с использованием NumPy. Визуализация весов</p>

	и градиентов. Применение ранней остановки (EarlyStopping) для предотвращения переобучения. Сохранение и загрузка обученной модели (model.save, load_model). Загрузка предобученных моделей (например, VGG16, ResNet) из Keras Applications и их применение для извлечения признаков (transfer learning). Сравнение производительности нейронной сети с классическими алгоритмами (случайный лес, градиентный бустинг) на одной и той же задаче.
ПЗ 8	Построение полного пайплайна машинного обучения на примере реальной задачи (например, прогнозирование оттока клиентов или классификация текстов отзывов). Использование Pipeline из scikit-learn для объединения этапов предобработки (масштабирование, кодирование) и моделирования. Применение ColumnTransformer для отдельной обработки числовых и категориальных признаков. Оптимизация гиперпараметров пайплайна с помощью GridSearchCV. Сохранение лучшей модели в файл с помощью joblib или pickle. Создание простого REST API для модели с использованием FastAPI: загрузка сохранённой модели, определение эндпоинта predict, получение запроса в формате JSON, возврат предсказания. Контейнеризация приложения: написание Dockerfile, сборка образа, запуск контейнера. Тестирование API с помощью Postman или curl. Развёртывание модели в облачной среде (Google Colab с ngrok или локальный сервер). Логирование экспериментов с MLflow: фиксация параметров, метрик, артефактов. Настройка мониторинга: логирование предсказаний, вычисление метрик на новых данных, детекция дрейфа (вычисление PSI между обучающими и новыми данными). Оформление отчёта о проекте: постановка задачи, описание данных, выбранные модели, результаты, выводы. Презентация результатов с демонстрацией работы API. Обсуждение лучших практик MLOps и CI/CD для машинного обучения.

Очно-заочная форма обучения (полный срок)

№	Тема практического (семинарского, лабораторного) занятия
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей
ПЗ 1	Знакомство со средой разработки: настройка Jupyter Notebook или VS Code, установка необходимых библиотек (numpy, pandas, matplotlib, seaborn, scikit-learn). Загрузка и первичное исследование датасетов (например, Iris, Boston Housing, Digits). Разделение данных на обучающую и тестовую выборки с помощью train_test_split. Реализация простейших моделей: линейная регрессия для предсказания стоимости жилья, логистическая регрессия для классификации ирисов, KMeans для кластеризации цифр. Сравнение качества моделей на обучающей и тестовой выборках. Визуализация результатов классификации и кластеризации. Обсуждение этапов построения моделей на практике: постановка задачи, подготовка данных, выбор модели, обучение, оценка. Анализ ошибок при неправильном разделении выборки или отсутствии предобработки.
ПЗ 2	Загрузка «грязного» датасета с пропусками, выбросами и категориальными признаками (например, Titanic, Adult Income). Анализ структуры данных: info, describe, isnull, unique. Очистка данных: удаление дубликатов, исправление типов данных. Визуализация выбросов с помощью boxplot и гистограмм, применение методов IQR и Z-оценки для их обнаружения, принятие решений об удалении или трансформации. Работа с пропусками: сравнение стратегий удаление строк (dropna), заполнение средним/медианой/модой (fillna), интерполяция, использование SimpleImputer из scikit-learn. Масштабирование признаков: применение StandardScaler, MinMaxScaler, RobustScaler, визуализация распределений до и после масштабирования. Кодирование категориальных признаков: LabelEncoder (порядковое кодирование для ordinal-переменных), OneHotEncoder (создание dummy-переменных), pd.get_dummies. Создание новых признаков: извлечение из даты (день недели, месяц, год), бинаризация непрерывных признаков, создание полиномиальных признаков с PolynomialFeatures. Оценка влияния предобработки на качество простой модели (логистическая регрес-

	сия). Оформление пайплайна предобработки с использованием ColumnTransformer.
ПЗ 3	Реализация линейной регрессии с использованием LinearRegression из scikit-learn на датасете с несколькими признаками. Вычисление коэффициентов модели и их интерпретация. Визуализация предсказаний vs реальных значений. Оценка качества с помощью MSE, MAE, R ² . Реализация логистической регрессии с LogisticRegression для бинарной классификации. Визуализация разделяющей поверхности для двух признаков. Интерпретация вероятностей (predict_proba) и границы принятия решений. Функции потерь: реализация MSE для линейной регрессии и Log Loss для логистической регрессии вручную (без использования готовых библиотек). Применение градиентного спуска: реализация простой версии с настройкой learning rate и числа итераций, сравнение с аналитическим решением. Регуляризация L1 (Lasso) и L2 (Ridge): обучение моделей с различными значениями параметра alpha, визуализация изменения коэффициентов (path plot). Сравнение разреженности коэффициентов при L1 и L2. Подбор оптимального alpha с помощью кросс-валидации (LassoCV, RidgeCV). Сравнение ElasticNet как комбинации L1 и L2. Анализ влияния масштабирования признаков на качество линейных моделей. Применение полиномиальных признаков (PolynomialFeatures) для моделирования нелинейных зависимостей, сравнение качества с обычной линейной регрессией.
ПЗ 4	Реализация дерева решений (DecisionTreeClassifier, DecisionTreeRegressor) на датасете с визуализацией структуры дерева (plot_tree, export_graphviz). Анализ влияния параметров (max_depth, min_samples_split, min_samples_leaf) на переобучение: построение кривых валидации. Реализация случайного леса (RandomForestClassifier, RandomForestRegressor). Оценка важности признаков (feature_importances_), визуализация в виде барплота. Сравнение качества случайного леса с одним деревом решений. Реализация градиентного бустинга: сравнение GradientBoostingClassifier, XGBoost, LightGBM, CatBoost. Подбор гиперпараметров (learning_rate, n_estimators, max_depth, subsample) с помощью GridSearchCV и RandomizedSearchCV. Использование early stopping для предотвращения переобучения. Визуализация кривых обучения (learning curves) для диагностики переобучения и недообучения. Применение бэггинга и бустинга к одному и тому же датасету, сравнение времени обучения и качества. Использование кросс-валидации для оценки стабильности ансамблевых методов. Обработка дисбаланса классов с помощью class_weight и параметров scale_pos_weight в XGBoost.
ПЗ 5	Реализация алгоритма k-средних (KMeans) на синтетическом датасете (make_blobs) с визуализацией кластеров. Изучение проблемы инициализации: сравнение random и k-means++. Анализ влияния выбросов на центроиды. Выбор числа кластеров: метод локтя (инерция vs число кластеров), силуэтный коэффициент (silhouette_score), индекс Калински-Харабаса. Визуализация силуэтных диаграмм (silhouette plot). Реализация иерархической кластеризации (AgglomerativeClustering). Построение дендрограммы с помощью scipy.cluster.hierarchy.dendrogram. Выбор числа кластеров по высоте разреза дендрограммы. Сравнение различных правил объединения (linkage): single, complete, average, ward. Применение кластеризации к реальному датасету (например, сегментация клиентов по поведенческим признакам). Нормализация признаков перед кластеризацией, сравнение результатов с и без масштабирования. Визуализация кластеров в 2D с помощью PCA и t-SNE. Оценка качества кластеризации с помощью внешних метрик (adjusted_rand_score, adjusted_mutual_info_score) если есть истинные метки, или внутренних (silhouette_score, davies_bouldin_score) если меток нет. Обсуждение интерпретации полученных кластеров и их практического применения.
ПЗ 6	Реализация задач классификации на датасете с дисбалансом классов (например, кредитный скоринг, обнаружение мошенничества). Вычисление матрицы ошибок

	<p>(<code>confusion_matrix</code>) и её визуализация с помощью <code>seaborn.heatmap</code>. Расчёт метрик: <code>accuracy</code>, <code>precision</code>, <code>recall</code>, <code>f1_score</code> вручную и с использованием <code>scikit-learn</code>. Анализ ситуации, когда <code>accuracy</code> неинформативна, сравнение <code>precision</code> и <code>recall</code>. Построение ROC-кривой (<code>roc_curve</code>) и вычисление AUC-ROC (<code>roc_auc_score</code>). Сравнение ROC-AUC и PR-AUC (<code>precision-recall curve</code>) для несбалансированных данных. Кросс-валидация: реализация <code>k-fold</code> (<code>KFold</code>), <code>stratified k-fold</code> (<code>StratifiedKFold</code>) с сохранением пропорции классов. Сравнение результатов кросс-валидации с простым разбиением <code>train/test</code>. Использование <code>cross_val_score</code>, <code>cross_validate</code>. Диагностика переобучения и недообучения: построение кривых обучения (<code>learning_curve</code>) — график зависимости качества от размера обучающей выборки. Построение кривых валидации (<code>validation_curve</code>) для выбора гиперпараметра (например, <code>max_depth</code> дерева). Анализ графиков: зоны недообучения (низкое качество на трейне и тесте), переобучения (разрыв между трейном и тестом). Применение <code>GridSearchCV</code> и <code>RandomizedSearchCV</code> для подбора гиперпараметров с кросс-валидацией. Сравнение времени выполнения и качества различных методов поиска.</p>
ПЗ 7	<p>Знакомство с библиотеками глубокого обучения: установка <code>TensorFlow</code> и <code>PyTorch</code>. Создание простейшей нейронной сети (многослойный перцептрон) для классификации рукописных цифр <code>MNIST</code>. Определение архитектуры: входной слой ($28 \times 28 = 784$ нейрона), скрытые слои (например, 128 нейронов с <code>ReLU</code>), выходной слой (10 нейронов с <code>Softmax</code>). Компиляция модели: выбор оптимизатора (<code>Adam</code>, <code>SGD</code>), функции потерь (<code>categorical_crossentropy</code>), метрик (<code>accuracy</code>). Обучение модели: метод <code>fit</code>, определение эпох и батчей. Визуализация кривых обучения (потери и точность на обучающей и валидационной выборках). Эксперименты с различными функциями активации: <code>ReLU</code>, <code>Sigmoid</code>, <code>Tanh</code>, сравнение скорости сходимости и итогового качества. Изучение проблемы затухающего градиента на глубоких сетях. Реализация <code>Dropout</code> и <code>BatchNormalization</code> для борьбы с переобучением, сравнение результатов. Настройка скорости обучения: использование <code>learning rate scheduling</code> (<code>ReduceLROnPlateau</code>, <code>ExponentialDecay</code>). Реализация обратного распространения ошибки вручную (для простой сети с одним скрытым слоем) с использованием <code>NumPy</code>. Визуализация весов и градиентов. Применение ранней остановки (<code>EarlyStopping</code>) для предотвращения переобучения. Сохранение и загрузка обученной модели (<code>model.save</code>, <code>load_model</code>). Загрузка предобученных моделей (например, <code>VGG16</code>, <code>ResNet</code>) из <code>Keras Applications</code> и их применение для извлечения признаков (<code>transfer learning</code>). Сравнение производительности нейронной сети с классическими алгоритмами (случайный лес, градиентный бустинг) на одной и той же задаче.</p>
ПЗ 8	<p>Построение полного пайплайна машинного обучения на примере реальной задачи (например, прогнозирование оттока клиентов или классификация текстов отзывов). Использование <code>Pipeline</code> из <code>scikit-learn</code> для объединения этапов предобработки (масштабирование, кодирование) и моделирования. Применение <code>ColumnTransformer</code> для раздельной обработки числовых и категориальных признаков. Оптимизация гиперпараметров пайплайна с помощью <code>GridSearchCV</code>. Сохранение лучшей модели в файл с помощью <code>joblib</code> или <code>pickle</code>. Создание простого <code>REST API</code> для модели с использованием <code>FastAPI</code>: загрузка сохранённой модели, определение эндпоинта <code>predict</code>, получение запроса в формате <code>JSON</code>, возврат предсказания. Контейнеризация приложения: написание <code>Dockerfile</code>, сборка образа, запуск контейнера. Тестирование <code>API</code> с помощью <code>Postman</code> или <code>curl</code>. Развёртывание модели в облачной среде (<code>Google Colab</code> с <code>ngrok</code> или локальный сервер). Логирование экспериментов с <code>MLflow</code>: фиксация параметров, метрик, артефактов. Настройка мониторинга: логирование предсказаний, вычисление метрик на новых данных, детекция дрейфа (вычисление <code>PSI</code> между обучающими и новыми данными). Оформление отчёта о проекте: постановка задачи, описание данных, выбранные модели, результаты, выводы. Презентация результатов с демонстрацией работы <code>API</code>. Обсуждение лучших практик <code>MLOps</code> и <code>CI/CD</code> для машинного обучения.</p>

3.3. Образовательные технологии
Очная форма обучения (полный срок)

№	Тема занятия	Вид учебного занятия	Форма / Методы интерактивного обучения	% учебного времени
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	ПЗ	Дискуссионные технологии, Работа в малых группах, Взаимопроверка, Мозговой штурм, Интерактивные тренажёры	25
2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	ПЗ	Дискуссионные технологии, Групповое решение проблемных задач, Кейс-стади, Взаимообучение, Проектно-ориентированное обучение	25
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	ПЗ	Работа в парах, Деловая игра, Дискуссионные технологии, Мозговой штурм, Интерактивная визуализация с коллективным обсуждением	25
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	ПЗ	Семинар-дискуссия, Работа в группах с презентацией, Ролевая игра, Интерактивная доска, Кейс-стади	25
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	ПЗ	Групповое решение задач, Конкурс, Интерактивная визуализация, Мозговой штурм, Проектно-ориентированное обучение	25
6	Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение	ПЗ	Дискуссионные технологии, Работа в малых группах, Кейс-стади, Взаимопроверка, Анализ практических ситуаций	25
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	ПЗ	Работа в парах, Проектная деятельность, Мозговой штурм, Интерактивные тренажёры, Групповое обсуждение результатов	25
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей	ПЗ	Проектно-ориентированное обучение, Работа в группах, Кейс-стади, Практическая работа, Презентация результатов	25
Итого				25%

Очно-заочная форма обучения (полный срок)

№	Тема занятия	Вид учебного занятия	Форма / Методы интерактивного обучения	% учебного времени
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>

1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	ПЗ	Дискуссионные технологии, Работа в малых группах, Взаимопроверка, Мозговой штурм, Интерактивные тренажёры	25
2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	ПЗ	Дискуссионные технологии, Групповое решение проблемных задач, Кейс-стади, Взаимообучение, Проектно-ориентированное обучение	25
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	ПЗ	Работа в парах, Деловая игра, Дискуссионные технологии, Мозговой штурм, Интерактивная визуализация с коллективным обсуждением	25
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	ПЗ	Семинар-дискуссия, Работа в группах с презентацией, Ролевая игра, Интерактивная доска, Кейс-стади	25
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	ПЗ	Групповое решение задач, Конкурс, Интерактивная визуализация, Мозговой штурм, Проектно-ориентированное обучение	25
6	Методы оценки моделей: метрики качества (ассигасу, precision, recall, F1), кросс-валидация, переобучение и недообучение	ПЗ	Дискуссионные технологии, Работа в малых группах, Кейс-стади, Взаимопроверка, Анализ практических ситуаций	25
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	ПЗ	Работа в парах, Проектная деятельность, Мозговой штурм, Интерактивные тренажёры, Групповое обсуждение результатов	25
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей	ПЗ	Проектно-ориентированное обучение, Работа в группах, Кейс-стади, Практическая работа, Презентация результатов	25
Итого				25%

Раздел 4. Организация самостоятельной работы обучающихся

4.1. Организация самостоятельной работы обучающихся

№	Тема дисциплины	№ вопро- сов	№ рекоменду- емой литерату- ры
1	2	3	4
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	1-5	1, 2, 3, 12
2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	6-10	1, 2, 4, 13
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	11-15	3, 4, 7, 8
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	16-20	3, 4, 9, 8
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	21-25	2, 7, 8, 9
6	Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение	26-30	2, 3, 7, 12
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	31-35	6, 7, 3, 12
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей	36-40	1, 3, 4, 13, 14, 15

Перечень вопросов, выносимых на самостоятельную работу обучающихся

1. Что такое машинное обучение и где оно применяется?
2. Какие существуют типы задач машинного обучения?
3. В чём различие между обучением с учителем и без учителя?
4. Какие этапы включает процесс построения модели?
5. Что такое признаки и целевая переменная?
6. Зачем необходима предобработка данных?
7. Какие методы используются для обработки пропущенных значений?
8. Что такое нормализация и стандартизация данных?
9. Какие существуют способы кодирования категориальных признаков?
10. Что такое выбросы и как их обнаружить?
11. Что такое линейная регрессия?
12. В чём заключается метод наименьших квадратов?
13. Что такое логистическая регрессия и где она применяется?
14. Что такое функция потерь?
15. Зачем нужна регуляризация (L1, L2)?
16. Как устроено дерево решений?
17. Какие критерии используются для разбиения в деревьях?
18. Что такое переобучение?
19. Как работает случайный лес?
20. Что такое градиентный бустинг?
21. Что такое кластеризация?
22. Как работает алгоритм k-средних?
23. Как выбрать число кластеров?
24. Что такое иерархическая кластеризация?

25. Какие метрики используются для оценки кластеризации?
26. Что такое accuracy?
27. В чём разница между precision и recall?
28. Что такое F1-score?
29. Что такое матрица ошибок?
30. Что такое кросс-валидация?
31. Что такое нейронная сеть?
32. Какие функции активации существуют?
33. Что такое обратное распространение ошибки?
34. Что такое градиентный спуск?
35. Из каких слоёв состоит нейронная сеть?
36. Как происходит обучение модели на практике?
37. Какие библиотеки используются для машинного обучения?
38. Как разделяются данные на обучающую и тестовую выборки?
39. Что такое переобучение и как его избежать?
40. Как происходит внедрение модели в реальное приложение?

4.2. Перечень учебно-методического обеспечения самостоятельной работы обучающихся

Самостоятельная работа обучающихся обеспечивается следующими учебно-методическими материалами:

1. Указаниями в рабочей программе по дисциплине (п.4.1.)
2. Лекционные материалы в составе учебно-методического комплекса по дисциплине
3. Заданиями и методическими рекомендациями по организации самостоятельной работы обучающихся в составе учебно-методического комплекса по дисциплине.
4. Глоссарием по дисциплине в составе учебно-методического комплекса по дисциплине.

Раздел 5. Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся

Фонд оценочных средств по дисциплине представляет собой совокупность контролируемых материалов, предназначенных для измерения уровня достижения обучающимися установленных результатов образовательной программы. ФОС по дисциплине используется при проведении оперативного контроля и промежуточной аттестации обучающихся. Требования к структуре и содержанию ФОС дисциплины регламентируются Положением о фонде оценочных материалов по программам высшего образования – программам бакалавриата, магистратуры.

5.1. Паспорт фонда оценочных средств

Очная форма обучения (полный срок)

№	Контролируемые разделы (темы) дисциплины	Оценочные средства			
		Л	ПЗ (ЛЗ, СЗ)	СРО	Код индикатора и дескриптора достижения компетенций
1	2	3	4	5	6
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	УО	ЗЗ, МШ	ПРВ	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	УО	ЗЗ, МШ	ПРВ	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	УО	ЗЗ, Д, МШ	ПРВ	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	УО	ЗЗ, Д, МП	ПРВ	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	УО	ЗЗ, МШ	ПРВ	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
6	Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение	УО	ЗЗ, Д	ПРВ	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	УО	ЗЗ, МШ	ПРВ	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей	УО	ЗЗ, Д	ПРВ	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2

Очно-заочная форма обучения (полный срок)

№	Контролируемые разделы (темы) дисциплины	Оценочные средства			
		Л	ПЗ (ЛЗ, СЗ)	СРО	Код индикатора и дескриптора достижения компетенций

1	2	3	4	5	6
1	Введение в машинное обучение: основные понятия, типы задач (регрессия, классификация, кластеризация), этапы построения моделей	УО	ЗЗ, МШ	ПРВ	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
2	Предобработка данных: очистка данных, работа с пропусками, нормализация и масштабирование, кодирование категориальных признаков	УО	ЗЗ, МШ	ПРВ	ИД-1 ОПК- 2.1 ИД-2 ОПК- 2.2
3	Линейные модели: линейная и логистическая регрессия, функция потерь, регуляризация (L1, L2)	УО	ЗЗ, Д, МШ	ПРВ	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
4	Деревья решений и ансамблевые методы: случайный лес, градиентный бустинг, переобучение и способы его устранения	УО	ЗЗ, Д, МП	ПРВ	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
5	Методы кластеризации: k-средних, иерархическая кластеризация, оценка качества кластеров	УО	ЗЗ, МШ	ПРВ	ИД-3 ОПК- 2.1 ИД-4 ОПК- 2.2
6	Методы оценки моделей: метрики качества (accuracy, precision, recall, F1), кросс-валидация, переобучение и недообучение	УО	ЗЗ, Д	ПРВ	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
7	Нейронные сети: базовая архитектура, функции активации, обучение с учителем, обратное распространение ошибки	УО	ЗЗ, МШ	ПРВ	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2
8	Применение машинного обучения: работа с библиотеками, построение и внедрение моделей			ПРВ	ИД-5 ОПК- 2.1 ИД-6 ОПК- 2.2

Условные обозначения оценочных средств (Столбцы 3, 4, 5):

ЗЗ – защита выполненных заданий (творческих, расчетных и т.д.), представление презентаций;

ПРВ – проверка рефератов, отчетов, рецензий, аннотаций, конспектов, графического материала, эссе, переводов, решений заданий, выполненных заданий в электронном виде и т.д.;

МШ – Метод мозгового штурма;

Д – Дискуссия, полемика, диспут, дебаты;

МП – Метод проектов.

5.2. Тематика письменных работ обучающихся

1. Основные этапы разработки моделей машинного обучения
2. Сравнение обучения с учителем и без учителя
3. Методы предобработки данных и их влияние на качество модели
4. Обработка пропущенных данных и выбросов в датасетах
5. Линейная регрессия: теория и практическое применение
6. Логистическая регрессия в задачах классификации
7. Регуляризация моделей и борьба с переобучением
8. Деревья решений: принципы построения и применения
9. Ансамблевые методы: случайный лес и градиентный бустинг
10. Методы кластеризации и их применение на практике
11. Алгоритм k-средних: преимущества и ограничения
12. Оценка качества моделей машинного обучения
13. Метрики классификации: анализ и сравнение
14. Проблема переобучения и методы её решения

15. Основы нейронных сетей и их применение
16. Алгоритм обратного распространения ошибки
17. Градиентный спуск и его разновидности
18. Использование библиотек машинного обучения (например, Scikit-learn)
19. Применение машинного обучения в реальных задачах
20. Развертывание моделей машинного обучения и основы MLOps

5.3. Перечень вопросов промежуточной аттестации по дисциплине

Вопросы к экзамену:

1. Понятие машинного обучения, его цели и области применения
2. Основные типы задач машинного обучения (классификация, регрессия, кластеризация)
3. Обучение с учителем и без учителя: различия и примеры
4. Основные этапы построения модели машинного обучения
5. Понятие признаков и целевой переменной
6. Предобработка данных: цели и основные методы
7. Обработка пропущенных значений и выбросов
8. Нормализация и стандартизация данных
9. Кодирование категориальных признаков
10. Линейная регрессия: модель и интерпретация
11. Метод наименьших квадратов
12. Логистическая регрессия и её применение
13. Функции потерь в машинном обучении
14. Регуляризация (L1, L2) и её роль
15. Деревья решений: структура и принцип работы
16. Критерии разбиения в деревьях решений (Gini, энтропия)
17. Переобучение: причины и способы предотвращения
18. Ансамблевые методы: случайный лес
19. Градиентный бустинг: принцип работы
20. Кластеризация: понятие и задачи
21. Алгоритм k-средних
22. Иерархическая кластеризация
23. Метрики оценки качества кластеризации
24. Метрики оценки моделей классификации (accuracy, precision, recall, F1)
25. Матрица ошибок и её анализ
26. Кросс-валидация и её значение
27. Основы нейронных сетей: структура и элементы
28. Функции активации и их роль
29. Обратное распространение ошибки
30. Развертывание моделей и применение машинного обучения на практике

Раздел 6. Оценочные средства промежуточной аттестации (с ключами)

1. Прочитайте текст и выберите один правильный ответ. Что является основной целью машинного обучения?
 - а) ручное программирование всех возможных правил поведения системы
 - б) автоматическое извлечение закономерностей из данных для построения прогнозных моделей
 - в) визуализация статистических данных в виде графиков и диаграмм
 - г) хранение больших объёмов неструктурированной информации
 - д) обеспечение физической безопасности серверного оборудования
 Правильный ответ: б)
2. Прочитайте текст и выберите один правильный ответ. Какой тип машинного обучения ис-

пользуется, когда обучающая выборка содержит размеченные данные с известными целевыми переменными?

- а) обучение с подкреплением
- б) обучение без учителя
- в) обучение с учителем
- г) полуобучение
- д) активное обучение

Правильный ответ: в)

3. Прочитайте текст и выберите один правильный ответ. Что такое переобучение модели машинного обучения?

- а) ситуация, когда модель показывает низкую точность на обучающих и тестовых данных
- б) процесс увеличения количества слоёв в нейронной сети
- в) явление, при котором модель слишком точно подстраивается под обучающие данные, включая шум, и плохо обобщает новые данные
- г) метод увеличения объёма обучающей выборки искусственным путём
- д) алгоритм автоматического подбора гиперпараметров

Правильный ответ: в)

4. Прочитайте текст и выберите один правильный ответ. Какая метрика чаще всего используется для оценки качества моделей в задачах бинарной классификации?

- а) средняя абсолютная ошибка (MAE)
- б) коэффициент детерминации (R^2)
- в) точность (Accuracy) и F1-мера
- г) энтропия
- д) дисперсия

Правильный ответ: в)

5. Прочитайте текст и выберите два правильных ответа. Какие из перечисленных алгоритмов относятся к методам обучения без учителя?

- а) линейная регрессия
- б) метод k-средних
- в) деревья решений
- г) метод главных компонент

Правильный ответ: б), г)

6. Прочитайте текст и выберите два правильных ответа. Какие техники применяются для снижения переобучения моделей?

- а) регуляризация
- б) неограниченное увеличение глубины дерева
- в) отсев нейронов (dropout)
- г) удаление тестовой выборки

Правильный ответ: а), в)

7. Прочитайте текст и выберите два правильных ответа. Какие метрики используются для оценки задач регрессии?

- а) средняя квадратичная ошибка (MSE)
- б) точность (Accuracy)
- в) средняя абсолютная ошибка (MAE)
- г) полнота (Recall)

Правильный ответ: а), в)

8. Укажите правильную последовательность этапов подготовки данных для машинного обучения:

- 1) Разделение выборки на обучающую и тестовую части
 - 2) Сбор сырых данных из источников
 - 3) Масштабирование и нормализация признаков
 - 4) Обработка пропущенных значений и выбросов
- Правильный ответ: 2-4-3-1

9. Укажите правильную последовательность действий при применении метода градиентного спуска:

- 1) Вычисление градиента функции потерь
 - 2) Инициализация начальных значений параметров модели
 - 3) Обновление параметров в направлении антиградиента
 - 4) Проверка условия сходимости или достижения лимита итераций
- Правильный ответ: 2-1-3-4

10. Укажите правильную последовательность шагов кросс-валидации по схеме k-fold:

- 1) Обучение модели на объединённых k-1 частях
 - 2) Разбиение данных на k равных блоков
 - 3) Оценка качества на оставшемся блоке
 - 4) Усреднение результатов по всем итерациям
- Правильный ответ: 2-1-3-4

11. Установите соответствие между типами задач машинного обучения и их описанием:

- А) классификация
 - Б) регрессия
 - В) кластеризация
- 1) прогнозирование непрерывной числовой величины
 - 2) отнесение объекта к одному из заранее известных классов
 - 3) группировка объектов по схожести без предварительной разметки
 - 4) генерация нового контента на основе шума
- Правильный ответ: А – 2, Б – 1, В – 3

12. Установите соответствие между алгоритмами и их семейством методов:

- А) SVM (метод опорных векторов)
 - Б) Random Forest
 - В) K-Nearest Neighbors
- 1) алгоритмы, основанные на расстояниях до ближайших соседей
 - 2) ансамблевый метод, строящий множество решающих деревьев
 - 3) метод, ищущий разделяющую гиперплоскость с максимальным зазором
 - 4) метод, использующий градиентный бустинг
- Правильный ответ: А – 3, Б – 2, В – 1

13. Прочитайте текст и запишите правильный ответ. Ответ следует записать с маленькой буквы. Процесс искусственного увеличения объёма обучающей выборки путём применения случайных преобразований к исходным данным называется:

Правильный ответ: аугментация

14. Прочитайте текст и запишите правильный ответ. Ответ следует записать с маленькой буквы. Ситуация, когда модель демонстрирует высокую точность на обучающих данных, но низкую на новых данных, называется:

Правильный ответ: переобучение

15. Прочитайте текст и запишите развернутый обоснованный ответ. Ответ следует записать с маленькой буквы:

Дайте определение понятию функция потерь в машинном обучении – это ...

Правильный ответ: математическая функция, которая измеряет степень отклонения предсказаний модели от истинных значений, используемая для оптимизации параметров модели в процессе обучения (ответ студента может быть представлен в интерпретации, эквивалентной приведенному правильному ответу)

16. Прочитайте текст и запишите развернутый обоснованный ответ. Ответ следует записать с маленькой буквы:

Дайте определение понятию кросс-валидация – это ...

Правильный ответ: метод оценки обобщающей способности модели, при котором исходная выборка многократно разбивается на обучающую и проверочную части, а результаты усредняются для получения надёжной оценки качества (ответ студента может быть представлен в интерпретации, эквивалентной приведенному правильному ответу)

Раздел 7. Перечень учебной литературы, необходимой для освоения дисциплины

7.1. Основная литература

1. Маккинни У. Python и анализ данных. 3-е изд. Москва : ДМК Пресс, 2020. 540 с.
2. Вандер Плас Дж. Python для сложных задач: наука о данных и машинное обучение. Санкт-Петербург : Питер, 2018. 576 с.
3. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. 3-е изд. Москва : Диалектика, 2021. 848 с.
4. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Москва : Вильямс, 2017. 480 с.
5. Бишоп К. Распознавание образов и машинное обучение. Москва : Вильямс, 2018. 744 с.

7.2. Дополнительная литература

1. Гудфеллоу И., Бенджио Й., Курвилль А. Глубокое обучение. Москва : ДМК Пресс, 2018. 652 с.
2. Мёрфи К. Машинное обучение: вероятностный подход. Москва : Вильямс, 2020. 1100 с.
3. Джеймс Г., Уиттен Д., Хасты Т., Тибширани Р. Введение в статистическое обучение. Москва : ДМК Пресс, 2019. 456 с.
4. Мохри М., Ростамизаде А., Талвалькар А. Основы машинного обучения. Москва : ДМК Пресс, 2018. 432 с.
5. Харрисон М. Машинное обучение на практике. Москва : Питер, 2020. 384 с.

7.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Онлайн-курс «Machine Learning» (Coursera, Andrew Ng). Режим доступа: <https://www.coursera.org/learn/machine-learning> (дата обращения: 10.03.2026).
2. Видеолекции по машинному обучению (Stanford University). Режим доступа: <https://www.youtube.com/playlist?list=PLoROMvodv4rNHU1-iPeDRH-J0cL-CrIda> (дата обращения: 10.03.2026).
3. Документация библиотеки Scikit-learn. Режим доступа: <https://scikit-learn.org/stable/> (дата обращения: 10.03.2026).
4. Электронная библиотечная система «Лань». Режим доступа: <https://e.lanbook.com> (дата обращения: 10.03.2026).
5. Образовательная платформа «Юрайт». Режим доступа: <https://urait.ru> (дата обращения: 10.03.2026).

Раздел 8. Материально-техническая база и информационные технологии

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине:

Материально-техническое обеспечение дисциплины «**Машинное обучение**» включает в себя учебные аудитории для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, самостоятельной работы обучающихся.

Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения. Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети Интернет.

Дисциплина может реализовываться с применением дистанционных технологий обучения. Специфика реализации дисциплины с применением дистанционных технологий обучения устанавливается дополнением к рабочей программе. В части не противоречащей специфике, изложенной в дополнении к программе, применяется настоящая рабочая программа.

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине с применением дистанционных образовательных технологий включает в себя:

Компьютерная техника, расположенная в учебном корпусе Института (ул. Качинцев, 63, кабинет Центра дистанционного обучения):

1. Intel i 3 3.4Ghz\ОЗУ 4Gb\500GB\RadeonHD5450
2. Intel PENTIUM 2.9GHz\ОЗУ 4GB\500GB

3. личные электронные устройства (компьютеры, ноутбуки, планшеты и иное), а также средства связи преподавателей и студентов.

Информационные технологии, необходимые для осуществления образовательного процесса по дисциплине с применением дистанционных образовательных технологий включают в себя:

- система дистанционного обучения (СДО) (Learning Management System) (LMS) Moodle (Modular Object-Oriented Dynamic Learning Environment);

- электронная почта;

- система компьютерного тестирования;

- Цифровой образовательный ресурс IPR SMART;

- система интернет-связи skype;

- телефонная связь;

- ПО для организации конференций.

Обучение обучающихся инвалидов и обучающихся с ограниченными возможностями здоровья осуществляется посредством применения специальных технических средств в зависимости от вида нозологии.

При проведении учебных занятий по дисциплине используются мультимедийные комплексы, электронные учебники и учебные пособия, адаптированные к ограничениям здоровья обучающихся.

Лекционные аудитории оборудованы мультимедийными кафедрами, подключенными к звуковым колонкам, позволяющими усилить звук для категории слабослышащих обучающихся, а также проекционными экранами, которые увеличивают изображение в несколько раз и позволяют воспринимать учебную информацию обучающимся с нарушениями зрения.

При обучении лиц с нарушениями слуха используется усилитель слуха для слабослышащих людей Cyber Ear модель NAP-40, помогающий обучаемым лучше воспринимать учебную информацию.

Обучающиеся с ограниченными возможностями здоровья, обеспечены печатными и электронными образовательными ресурсами (программы, учебники, учебные пособия, материалы для самостоятельной работы и т.д.) в формах, адаптированных к ограничениям их здоровья и восприятия информации:

для лиц с нарушениями зрения:

- в форме электронного документа;

- в форме аудиофайла;

для лиц с нарушениями слуха:

- в печатной форме;
 - в форме электронного документа;
- для лиц с нарушениями опорно-двигательного аппарата:**
- в печатной форме;
 - в форме электронного документа;
 - в форме аудиофайла.

Раздел 9. Методические указания для обучающихся по освоению дисциплины

Дисциплина включает практические занятия, самостоятельную работу обучающегося.

В ходе изучения дисциплины «Машинное обучение» перед обучающимися стоит задача не только закрепить знания о сложных информационных явлениях, о чем свидетельствует содержание тематического плана, глубоко разобраться в объемном учебном материале, но и сформировать у себя на основе полученных компьютерных знаний соответствующие профессионально важные качества.

Практические занятия – один из самых эффективных видов учебных занятий, на которых обучающиеся учатся творчески работать с различной информацией, являются также действенной формой активизации самостоятельной работы обучающихся.

Целью практических занятий является закрепление полученных в ходе лекций, а также в ходе самостоятельной работы над учебной и специальной литературой, знаний, умений и навыков. На практических занятиях особо обращается внимание на умение обучающихся проявлять элементы творчества в процессе самостоятельной работы, применять полученные знания на практике.

Практические занятия занимают центральное место в учебном процессе, так как позволяют на завершающем этапе усвоения материала, после прослушанной лекции и самостоятельного поиска дополнительных сведений по рассматриваемой проблематике, окончательно уточнить, сформировать свои позиции в ходе работы в составе учебной группы.

Основное в подготовке и проведении практикума – это самостоятельная работа обучающегося над изучением темы лекционного материала. Практические занятия проводятся по специальным планам – заданиям, которые содержатся в материалах, подготовленных на кафедре. Обучающийся обязан точно знать план занятия либо конкретное задание к нему.

При подготовке к практическим занятиям следует чаще обращаться к справочной литературе, полнее использовать консультации (групповые и индивидуальные, устные и письменные) с преподавателями, которые читают лекции и проводят практикумы.

Таким образом, в процессе подготовке к практическому занятию рекомендуется:

- ознакомиться с вопросами плана;
- прочитать конспект лекции по изучаемой теме;
- прочитать соответствующие главы учебников, статьи;
- просмотреть перечень научных источников, предлагаемых в рабочей программе, выбрав несколько из них для углубленного изучения данной темы.

По каждому практическому заданию обучающиеся отчитываются преподавателю, оформляя письменный отчет, в котором сохраняют результаты своей работы в виде файлов. Результаты выполнения практических заданий оцениваются с учетом теоретических знаний по соответствующим вопросам дисциплины и уровнем владения практическими навыками при работе на компьютере.

Для углубленного изучения и освоения материала целесообразно выполнение практических работ, наряду с другими различными формами обучения обучающихся: тесты, задачи, упражнения, которые используются при проведении практических занятий, выполнении контрольных и аудиторных работ, а также при самостоятельном изучении данной дисциплины.

Одним из наиболее интенсивных способов изучения дисциплины является самостоятельное выполнение практических работ, на которых вырабатываются навыки по дисциплине «Машинное обучение».

СРО позволяет глубже освоить теоретические и практические вопросы, понять принципы дисциплины «Машинное обучение».

Основными задачами организации процесса самостоятельной работы по дисциплине являются:

- приобретение знаний по теоретическим основам дисциплины «Машинное обучение», являющихся дополнением к материалу лекционных аудиторных занятий;
- приобретение практических навыков по дисциплине «Машинное обучение».

Основные формы реализации СРО – изучение учебно-методической литературы по дисциплине «Машинное обучение». В качестве базовой литературы можно использовать учебники и учебные пособия, согласно приведенному списку в разделе 6 рабочей программы, а также любые

другие источники информации, такие как электронные учебники, обучающие и энциклопедические сайты, публикации журналов и конференций.

Обучающийся допускается к зачетному занятию по результатам успешного выполнения всех практических заданий и самостоятельной работы.

Учебно-методическое издание

Рабочая программа учебной дисциплины

Машинное обучение

(Наименование дисциплины в соответствии с учебным планом)

Скоробогатченко Дмитрий Анатольевич

(Фамилия, Имя, Отчество составителя)
