

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ
В.Ф. УТКИНА»

Кафедра «Электронные вычислительные машины»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

Б1.В.05 «Интеллектуальные системы и технологии»

Направление подготовки
02.03.03 Математическое обеспечение и администрирование
информационных систем

Профиль
«Программное обеспечение компьютерных технологий и систем
искусственного интеллекта»

Уровень подготовки
Бакалавриат

Квалификация выпускника – бакалавр

Форма обучения – очная

Рязань 2025

1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов (контрольных заданий, описаний форм и процедур проверки), предназначенных для оценки качества освоения обучающимися данной дисциплины как части ОПОП.

Цель – оценить соответствие знаний, умений и владений, приобретенных обучающимся в процессе изучения дисциплины, целям и требованиям ОПОП в ходе проведения промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций, закрепленных за дисциплиной.

Промежуточный контроль по дисциплине осуществляется путем проведения зачета. Форма проведения зачета – билеты с письменным ответом на два теоретических вопроса. При необходимости, проводится устная беседа с обучаемым для уточнения уровня владения материалом.

Выполнение заданий на практических занятиях в течение семестра и заданий на самостоятельную работу является обязательным условием для допуска к зачету.

2. ПАСПОРТ ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Контролируемые разделы (темы) дисциплины (результаты по разделам)	Код контролируемой компетенции (или её части)	Наименование оценочного средства
Раздел 1. Теоретико-математические основы искусственного интеллекта.	ПК-10, ПК-13, ПК-21	Зачёт
Раздел 2. Практики сбора, очистки и предварительной обработки данных.	ПК-10, ПК-13, ПК-21	Зачёт
Раздел 3. Технологии и инструменты анализа данных.	ПК-10, ПК-13, ПК-21	Зачёт

3. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ

Сформированность компетенции (или ее части) в рамках освоения данной дисциплины оценивается по трехуровневой шкале:

- 1) пороговый уровень является обязательным для всех обучающихся по завершении освоения дисциплины;
- 2) продвинутый уровень характеризуется превышением минимальных характеристик сформированности компетенций по завершении освоения дисциплины;
- 3) эталонный уровень характеризуется максимально возможной выраженностью компетенций и является важным качественным ориентиром для самосовершенствования.

Уровень освоения компетенций, формируемых дисциплиной:

Описание критериев и шкалы оценивания экзаменационного билета:

Шкала оценивания	Критерий
Зачтено (эталонный уровень)	выставляется студенту, который дал полные ответы на вопросы, показал глубокие систематизированные знания, смог привести примеры, ответил на дополнительные вопросы преподавателя
Зачтено (продвинутый уровень)	выставляется студенту, который дал полные ответы на вопросы, но на некоторые дополнительные вопросы преподавателя ответил только с помощью наводящих вопросов
Зачтено (пороговый уровень)	выставляется студенту, который дал неполные ответы на вопросы в билете и смог ответить на дополнительные вопросы с помощью преподавателя
Не зачтено	выставляется студенту, который не смог ответить на вопросы

4. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ

4.1. Промежуточная аттестация в форме зачёта

Коды компетенций	Результаты освоения ОПОП Содержание компетенций (код и содержание индикатора)
ПК-10	Способен осуществлять поиск сбор очистку и предварительный анализ данных ПК-10.1. Обосновывает способы и варианты применения методов предварительного анализа данных в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. ПК-10.2. Применяет методы анализа данных для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ.
ПК-13	Способен применять различные модели и (или) технологии обработки данных ПК-13.1. Осуществляет выбор технологий обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями. ПК-13.2. Разрабатывает и отлаживает прикладные решения с элементами ИИ с применением различных технологий обработки данных.
ПК-21	Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ ПК-21.1. Обосновывает способы и варианты применения методов и моделей в задачах искусственного интеллекта, включая их модификацию и адаптацию к специфике задачи. ПК-21.2. Применяет аппарат теории вероятностей, матстатистики и теории информации для формулирования и анализа задач искусственного интеллекта.

Тестовые вопросы для оценки уровня освоения компетенций (закрытый тип):

Раздел I «Теоретико-математические основы искусственного интеллекта»

Вопрос № 1

Что такое градиентный спуск?

- a) Алгоритм линейной регрессии.
- b) Метод оптимизации весов нейронной сети путем минимизации функции потерь.
- c) Способ увеличения размерности признаков.
- d) Техника предобработки данных перед подачей в алгоритм.
- e) Подход к обработке выбросов в данных.

Вопрос № 2

Какой метод машинного обучения основан на максимальном правдоподобии?

- a) Байесовский вывод.
- b) Генеративные состязательные сети (GAN).
- c) Кластерный анализ методом k-means.
- d) Методы опорных векторов (SVM).
- e) Максимальное апостериорное оценивание (MAP).

Вопрос № 3

Какие типы связей существуют в байесовской сети?

- a) Циклические и циклические связи.
- b) Направленные и ненаправленные связи.
- c) Прямоугольные и квадратичные связи.
- d) Линейные и экспоненциальные связи.
- e) Глобальные и локальные связи.

Вопрос № 4

Что представляет собой автокодировщик (autoencoder)?

- a) Нейросеть, предназначенная исключительно для предсказания меток классов.
- b) Архитектура нейросети, состоящая из двух частей: кодировщика и декодировщика.
- c) Классический алгоритм кластеризации.
- d) Графовая модель случайных величин.
- e) Модель глубокого обучения для временного ряда.

Вопрос № 5

В чём суть метода главных компонент (PCA)?

- a) Уменьшение числа признаков путём выявления наиболее значимых комбинаций исходных переменных.
- b) Увеличение количества наблюдений для повышения точности модели.
- c) Обнаружение скрытых факторов, влияющих на зависимую переменную.
- d) Оценка влияния каждого признака на итоговую классификацию.
- e) Выявление временной зависимости в последовательности данных.

Вопрос № 6

Какие признаки характеризуют временную последовательность?

- a) Постоянная дисперсия и среднее значение.
- b) Независимые наблюдения друг от друга.
- c) Наличие структуры в данных, зависящей от порядка следования элементов.

- d) Равномерное распределение значений по всей выборке.
- e) Отсутствие коррелированных точек данных.

Вопрос № 7

Что означает термин «нормализация данных»?

- a) Преобразование всех признаков таким образом, чтобы среднее равнялось нулю, а стандартное отклонение единице.
- b) Добавление новых признаков, повышающих точность модели.
- c) Исключение дублирующихся записей из набора данных.
- d) Преобразование категориальных признаков в числовые значения.
- e) Изменение масштаба отдельных признаков для улучшения производительности алгоритма.

Вопрос № 8

Какой метод используется для оценки качества классификации в задачах бинарной классификации?

- a) Коэффициент Джини.
- b) Среднее абсолютное отклонение.
- c) Матрица ошибок (confusion matrix).
- d) Дисперсионный анализ ANOVA.
- e) Коэффициент детерминации R^2 .

Вопрос № 9

Что такое ансамблевые методы (ensemble methods)?

- a) Использование множества слабых моделей для формирования сильной общей модели.
- b) Применение одной мощной модели для решения сложных задач.
- c) Объединение разных типов признаков для достижения наилучших результатов.
- d) Повторение эксперимента с разными наборами начальных условий.
- e) Создание гибридных моделей из комбинации искусственных нейронов и традиционных статистических методов.

Вопрос № 10

Что называется латентным пространством в глубоком обучении?

- a) Пространство, которое отражает структуру признаков в изначальной форме.
- b) Низкоразмерное пространство, представляющее скрытые характеристики объектов.
- c) Геометрическое представление признаков, доступное для визуализации человеком.
- d) Распределённое пространство признаков для каждой категории объекта.
- e) Область пространства признаков, содержащая выбросы.

Вопрос № 11

Какова основная цель подхода к обучению с учителем (supervised learning)?

- a) Нахождение скрытой структуры данных.
- b) Прогнозирование целевых меток на основе заданных признаков.
- c) Разделение данных на группы без априорной информации о классах.
- d) Генерация новых примеров данных, похожих на исходные.
- e) Поиск общих закономерностей в большом объёме необработанных данных.

Вопрос № 12

Какой тип алгоритмов чаще всего применяется для обнаружения аномалий в больших объемах данных?

- a) Деревья решений.
- b) Неуправляемое обучение (unsupervised learning).
- c) Регрессия.
- d) Глубокое обучение (deep learning).
- e) Логистическая регрессия.

Вопрос № 13

Каково основное назначение кросс-валидации (cross validation)?

- a) Определение оптимального размера обучающей выборки.
- b) Проверка устойчивости модели к переобучению.
- c) Минимизация вычислительных затрат на обработку данных.
- d) Повышение скорости вычислений в процессе обучения.
- e) Оптимизация гиперпараметров вручную.

Вопрос № 14

Что представляет собой ансамбль деревьев решений Random Forest?

- a) Набор случайно инициализированных нейронных сетей.
- b) Совокупность слабо взаимосвязанных моделей SVM.
- c) Случайный подбор подмножества признаков и экземпляров данных для построения дерева решений.
- d) Последовательность регрессоров, применяемых последовательно.
- e) Метод разделения данных на классы с использованием условных вероятностей.

Вопрос № 15

Какой вид регрессии является линейным приближением отношения вероятности события к её отрицанию?

- a) Полиномиальная регрессия.
- b) Логистическая регрессия.
- c) Бинарная логистическая регрессия.
- d) Пуассоновская регрессия.
- e) Lasso-регрессия.

Вопрос № 16

Что обозначают сигмOIDные активаторы в нейронных сетях?

- a) Нормализацию выходного сигнала в диапазоне $[-1, 1]$.
- b) Сигнал активации, преобразованный в биполярный сигнал $(-1$ или $+1)$.
- c) Приведение выходного сигнала к диапазону $[0, 1]$, имитируя вероятность принадлежности классу.
- d) Ограничение выходной величины веса нейрона до фиксированного значения.
- e) Возможность представления нейроном непрерывных функций высокого порядка.

Вопрос № 17

Какую проблему решает регуляризация Ridge (L2)?

- a) Предотвращение переобучения за счёт ограничения суммы квадратов коэффициентов.
- b) Автоматическое исключение несущественных признаков из модели.

- с) Улучшение интерпретируемости модели за счёт снижения коллинеарности признаков.
- d) Решение проблемы мультиколлинеарности среди входных признаков.
- е) Упрощение архитектуры нейронной сети за счёт уменьшения количества слоёв.

Вопрос № 18

Какой принцип лежит в основе методов стохастического градиентного спуска (SGD)?

- а) Постепенное уменьшение шага обновления весов.
- б) Послойное обновление весов, начиная с последнего слоя.
- с) Обучение модели на небольших порциях данных, позволяющее ускорить процесс обучения.
- d) Параллельное обновление весов всеми экземплярами данных одновременно.
- е) Регуляризация коэффициентов модели для предотвращения переобучения.

Вопрос № 19

Что значит гиперпараметр «learning rate» в обучении нейронных сетей?

- а) Скорость изменения весов при обновлении модели.
- б) Количество эпох обучения нейронной сети.
- с) Вероятность правильного распознавания объекта моделью.
- d) Уровень шума в обучающих данных.
- е) Величина штрафующего члена регуляризации.

Вопрос № 20

Какой алгоритм предназначен для автоматической группировки сходных объектов без наличия меток?

- а) KNN-классификатор.
- б) K-means-кластеризация.
- с) Decision Tree Classifier.
- d) Support Vector Machine.
- е) Logistic Regression.

Вопрос № 21

Как называют модель, способную воспроизводить новые образцы данных, похожие на тренировочные?

- а) Трансферная модель.
- б) Гибридная модель.
- с) Рекомбинаторная модель.
- d) Генеративная модель.
- е) Представительная модель.

Вопрос № 22

Какова роль softmax-функции в многослойных перцептронах?

- а) Расчёт скалярного произведения между признаками и весами.
- б) Выбор наилучшего параметра гиперплоскости.
- с) Преобразование выходного вектора нейросети в распределение вероятностей.
- d) Сокращение размерности входных данных.
- е) Генерация случайных чисел для начального распределения весов.

Вопрос № 23

Какой класс алгоритмов помогает восстанавливать пропущенные значения в данных?

- a) Методы кластеризации.
- b) Алгоритмы импутирования (imputation algorithms).
- c) Ансамбли решающих деревьев.
- d) Модели нейронных сетей прямого распространения.
- e) Методы уменьшения размерности PCA.

Вопрос № 24

Какой показатель измеряет долю правильно классифицированных положительных случаев относительно общего числа реальных положительных примеров?

- a) Точность (precision).
- b) Чувствительность (recall).
- c) Специфичность (specificity).
- d) F-мера (F-score).
- e) Ошибка первого рода (Type-I error).

Вопрос № 25

Почему используются рекуррентные нейронные сети (RNNs) для обработки временных рядов?

- a) Они позволяют обрабатывать только статические наборы данных.
- b) Эти сети способны учитывать зависимость текущего элемента от предыдущих элементов последовательности.
- c) RNN оптимальны для изображений и других двумерных структур.
- d) Такие сети ускоряют работу с большими объемами данных.
- e) Рекуррентные сети устойчивы к шумовым данным.

Вопрос № 26

Какой подход обеспечивает получение вложенного описания данных в низкой размерности?

- a) Автоэнкодеры.
- b) Самообучающиеся машины (self-training machines).
- c) Полносвязные нейронные сети.
- d) Ядерные методы (kernel methods).
- e) Энтропийный подход.

Вопрос № 27

Какой метод позволяет выявлять потенциальные факторы риска в медицинских исследованиях?

- a) Анализ главных компонент (PCA).
- b) Классификация на основе ближайших соседей (KNN).
- c) Лассо-регрессия (Lasso regression).
- d) Байесовская классификация.
- e) Генеративные состязательные сети (GAN).

Вопрос № 28

Что определяет качество восстановления данных при работе с автоэнкодерами?

- a) Размер скрытых слоев.
- b) Тип используемых активаторов.
- c) Кодировующее расстояние (reconstruction loss).

- d) Глубина нейронной сети.
- e) Масштаб нормализованных признаков.

Вопрос № 29

Что представляет собой метрику AUC-ROC?

- a) Площадь под кривой, отражающую производительность модели на уровне различимости классов.
- b) Критерий качества оценки предсказательной способности модели.
- c) Измерение близости между фактическими значениями и предсказанными моделями.
- d) Показатель стабильности нейронной сети при изменении гиперпараметров.
- e) Индекс разброса данных вокруг среднего значения.

Вопрос № 30

Какой тип задач лучше всего подходит для подходов reinforcement learning?

- a) Постановка целей, решение которых требует пошагового принятия решений.
- b) Быстрое выявление групп схожих объектов.
- c) Одновременная обработка большого объёма текстовых данных.
- d) Задача линейной регрессии.
- e) Проблемы, связанные с обработкой изображений.

Раздел II «Практики сбора, очистки и предварительной обработки данных»

Вопрос № 1

Что такое стандартизация данных?

- a) Процедура масштабирования признаков таким образом, чтобы они имели нулевое среднее и единичную дисперсию.
- b) Процесс приведения всех значений признаков к единому масштабу.
- c) Метод устранения повторяющихся записей в наборе данных.
- d) Удаление отсутствующих значений и замена их средним значением.
- e) Манипуляция над набором данных для выделения редких категорий.

Вопрос № 2

Каковы основные причины появления выбросов в данных?

- a) Неправильная сборка данных, человеческий фактор, сбой оборудования.
- b) Недостаточная предварительная подготовка данных.
- c) Высокая степень нормальности распределённых данных.
- d) Скрытая корреляция между признаками.
- e) Только редкие случаи, обусловленные природой явления.

Вопрос № 3

Какой инструмент используют для быстрой проверки однородности распределений данных?

- a) Тест Колмогорова-Смирнова.
- b) Шкала Лайкерта.
- c) Т-тест Стьюдента.
- d) Анализ главных компонент (РСА).
- e) Многомерный дискриминантный анализ.

Вопрос № 4

Какой метод часто применяют для преобразования асимметрично распределённых данных?

- a) Стандартизация Z-нормировкой.
- b) Мин-Макс масштабирование.
- c) Логарифмическое преобразование.
- d) Ортогонализация признаков.
- e) Импутация средних значений.

Вопрос № 5

Что подразумевает удаление выбросов?

- a) Замена экстремально высоких или низких значений на средние значения выборки.
- b) Перевод значений в нормальное распределение.
- c) Удаление данных, значительно отличающихся от основного массива данных.
- d) Масштабирование всех признаков до единого диапазона.
- e) Переопределение критериев измерения данных.

Вопрос № 6

Как называются методы удаления выбросов на основе статистики?

- a) Квазирегулярные процедуры.
- b) Эмпирические правила.
- c) Детерминированные подходы.
- d) Статистически обоснованные критерии.
- e) Хирургические вмешательства.

Вопрос № 7

Что означает эффект мультиколлинеарности в данных?

- a) Высококоррелируемые признаки оказывают сильное влияние на коэффициент регрессии.
- b) Данные имеют высокую дисперсию.
- c) Значительное количество пустых полей в таблице данных.
- d) Признаки содержат много уникальных значений.
- e) Присутствие сильно сгруппированных значений.

Вопрос № 8

Какую технику обычно применяют для обнаружения выбросов в одномерных данных?

- a) Интерквартильное правило (IQR rule).
- b) Корреляционный анализ Пирсона.
- c) Один-против-всех (one-vs-all).
- d) K-fold кросс-валидация.
- e) Разложение по сингулярным значениям (SVD).

Вопрос № 9

Какой из методов относится к процедурам нормализации данных?

- a) PCA-анализ.
- b) Чистка пропусков.
- c) MinMaxScaler.
- d) Кластеризация k-means.
- e) Байесовские сети.

Вопрос № 10

Что включает этап предварительного исследования данных (EDA)?

- a) Визуализацию данных, изучение распределений, проверку гипотез.
- b) Настройку гиперпараметров модели.
- c) Импорт библиотек Python.
- d) Запуск глубокого обучения на GPU.
- e) Форматирование CSV файлов.

Вопрос № 11

Что понимают под качеством данных?

- a) Степень соответствия данных требованиям конкретной задачи.
- b) Объём хранимых данных.
- c) Скорость загрузки данных из хранилища.
- d) Доступность открытых баз данных.
- e) Частота обновления базы данных.

Вопрос № 12

Что такое корреляционная матрица?

- a) Таблица, показывающая силу и направление взаимосвязи между признаками.
- b) Диаграмма рассеяния данных.
- c) График плотности распределения данных.
- d) Фильтрация нерелевантных признаков.
- e) Временная диаграмма развития процесса.

Вопрос № 13

Какой подход применяется для заполнения пропусков в данных?

- a) One-hot encoding.
- b) Импутация средней, медианы или ближайшего соседа.
- c) Логистическая регрессия.
- d) Карта Q-Q.
- e) Альфа-преобразование.

Вопрос № 14

Как называется проблема выбора подходящего объема данных для эффективного обучения модели?

- a) Проблема перекоса выборки.
- b) Перекос классов (class imbalance).
- c) Проблема малого объёма данных (small sample size problem).
- d) Катастрофическое забывание.
- e) Локальный минимум.

Вопрос № 15

Что подразумевается под балансировкой классов?

- a) Программирование равенства классов.
- b) Интуитивное деление классов на большие и малые.
- c) Согласованное соотношение представителей классов в выборке.
- d) Построение иерархической системы классов.
- e) Добавление виртуальных классов.

Вопрос № 16

Что показывает коэффициент VIF (Variance Inflation Factor)?

- a) Средний уровень взаимодействия признаков.
- b) Суммарную величину отклонения результата от ожидаемого.

- c) Влияние мультиколлинеарности на стандартные ошибки коэффициентов регрессии.
- d) Сложность настройки гиперпараметров модели.
- e) Необходимость увеличить размер выборки.

Вопрос № 17

Что понимается под концепцией "Big Data"?

- a) Объем данных настолько велик, что традиционные инструменты хранения и обработки становятся неэффективными.
- b) Большие объемы научных публикаций.
- c) Маленькие наборы данных, обрабатываемые современными компьютерами.
- d) Экономия места на жестком диске.
- e) Легкость восприятия людьми большого объема информации.

Вопрос № 18

Какой индикатор сигнализирует о наличии сильных выбросов в данных?

- a) Медианное значение меньше среднего арифметического.
- b) Большое значение коэффициента асимметрии.
- c) Высокий коэффициент корреляции Пирсона.
- d) Очень маленькое значение p-value в t-тесте.
- e) Широкий диапазон квартилей.

Вопрос № 19

Какая техника применяется для устранения мультиколлинеарности?

- a) Градиентный спуск.
- b) K-folds cross-validation.
- c) PLS-regression (Partial Least Squares).
- d) Обычная регрессия.
- e) Clustering with k-means.

Вопрос № 20

Какой инструмент позволяет быстро выявить наличие корреляций между признаками?

- a) Тепловая карта корреляций.
- b) Корреляционное дерево.
- c) Квадратичный дискриминантный анализ.
- d) Multivariate adaptive regression splines.
- e) Гомоскедастичность.

Вопрос № 21

Как называется ситуация, когда одни классы представлены гораздо больше других в наборе данных?

- a) Смещение классов (Class Imbalance).
- b) Редкая категория (Sparse Category).
- c) Смещённая выборка (Biased Sample).
- d) Мониторинг предвзятости (Bias Monitoring).
- e) Адекватность выборки (Sample Adequacy).

Вопрос № 22

Что подразумевает термин "подгонка данных" (data fitting)?

- a) Поиск оптимальной функциональной зависимости между признаками и

целевой переменной.

- b) Внесение изменений в данные для лучшего понимания.
- c) Исправление систематических ошибок в сборе данных.
- d) Исследование выбросов и устранение шоков.
- e) Устранение проблем многократного обращения к данным.

Вопрос № 23

Какой подход рекомендуется для обработки значительных объемов данных в ограниченном объеме памяти?

- a) Mini-batch training.
- b) Gradient descent.
- c) Cross-validation.
- d) Overfitting prevention.
- e) Batch normalization.

Вопрос № 24

Как определяется понятие "редкий класс" (rare class)?

- a) Класс, имеющий незначительную долю в общем наборе данных.
- b) Наиболее важный класс для анализа.
- c) Самый распространенный класс в выборке.
- d) Контрольный класс для сравнения с остальными.
- e) Высокоинформативный класс.

Вопрос № 25

Что такое импутация (imputation)?

- a) Восстановление недостающих данных на основе имеющихся сведений.
- b) Стандартизация данных.
- c) Удаление дубликатов строк.
- d) Кластеризация данных.
- e) Метод коррекции смещения.

Вопрос № 26

Какой критерий считается важным показателем качества модели?

- a) Среднеквадратичное отклонение (MSE).
- b) Количество использованных признаков.
- c) Размер обучающего набора данных.
- d) Длительность тренировки модели.
- e) Расстояние Махаланобиса.

Вопрос № 27

Какой алгоритм служит для быстрого выявления аномалий в больших массивах данных?

- a) DBSCAN.
- b) Linear Discriminant Analysis.
- c) K-means clustering.
- d) Kernel density estimation.
- e) Regularized logistic regression.

Вопрос № 28

Какая задача решается методом масштабирования (scaling)?

- a) Приведение данных к одному масштабу для дальнейшего анализа.

- b) Преобразование качественных признаков в количественные.
- c) Увеличение объема выборки.
- d) Определение доминирующих характеристик в данных.
- e) Удаление ненужных признаков.

Вопрос № 29

Какой аспект важен при оценке качества данных?

- a) Чистота данных (cleanliness of data).
- b) Периодичность пополнения данных.
- c) Вид используемой операционной системы.
- d) Цена хранения данных.
- e) Количество копий данных.

Вопрос № 30

Что делают с корневыми источниками данных перед построением модели?

- a) Предварительно исследуют и очищают.
- b) Применяют случайные вращения.
- c) Используют исключительно сырые данные.
- d) Перемешивают данные.
- e) Формируют промежуточные таблицы.

Раздел III «Технологии и инструменты анализа данных»

Вопрос № 1

Что такое Named Entity Recognition (NER) в Natural Language Processing (NLP)?

- a) Идентификация сущностей, таких как имена собственные, географические названия, организации и т.п., в тексте.
- b) Конвертация текста в изображение.
- c) Методология перевода естественного языка в формализованную речь роботов.
- d) Процедуры генерации речи по заданному шаблону.
- e) Метод сжатия больших документов в короткие резюме.

Вопрос № 2

Какой инструмент широко применяется для анализа настроений (sentiment analysis) в социальных медиа?

- a) Спектральный анализ.
- b) Word Embeddings.
- c) Bag-of-Words model.
- d) Convolutional Neural Networks (CNN).
- e) Long Short-Term Memory networks (LSTM).

Вопрос № 3

Что представляет собой технология Computer Vision?

- a) Изучение способов интерпретации и понимания зрительной информации компьютером.
- b) Метод обработки звукового сигнала.
- c) Технология шифрования и дешифровки данных.
- d) Обработка крупных массивов численных данных.
- e) Изучение законов робототехники.

Вопрос № 4

Какой слой нейронной сети особенно полезен для обработки изображений?

- a) Pooling layer.
- b) Dense layer.
- c) Recurrent layer.
- d) Activation function.
- e) Output layer.

Вопрос № 5

Какая архитектура нейронной сети эффективно работает с последовательностью символов (например, текстом)?

- a) Autoencoders.
- b) Restricted Boltzmann Machines.
- c) Feedforward neural network.
- d) Recurrent Neural Network (RNN).
- e) Generative Adversarial Networks (GAN).

Вопрос № 6

Какой алгоритм применяется для выделения ключевых слов в документе?

- a) TF-IDF (Term Frequency-Inverse Document Frequency).
- b) Gradient Descent.
- c) Principal Component Analysis (PCA).
- d) Support Vector Machines (SVM).
- e) Classification and Regression Trees (CART).

Вопрос № 7

Что такое векторное представление слов (word embeddings)?

- a) Метод задания каждому слову документа конкретного номера.
- b) Векторное отображение слова в многомерном пространстве, учитывающее семантические связи.
- c) Простая бинарная система кодировки слов.
- d) Алгоритм перевода текста на разные языки.
- e) Принцип категоризации текстов.

Вопрос № 8

Какой тип рекомендательных систем основывается на анализе поведения пользователей и выборе аналогичных товаров?

- a) Content-based filtering.
- b) Collaborative filtering.
- c) Hybrid recommendation systems.
- d) Rule-based recommendations.
- e) Personalization engines.

Вопрос № 9

Как называется подход к извлечению смысла из коротких фрагментов текста (например, заголовков новостей)?

- a) Summarization.
- b) Text classification.
- c) Sentence parsing.
- d) Semantic role labeling.
- e) Information extraction.

Вопрос № 10

Что такое эмбединги изображений (image embeddings)?

- a) Двухэтапный процесс компиляции изображений.
- b) Особый тип фильтров изображений.
- c) Векторы, содержащие важные признаки изображения, используемые для дальнейших операций.
- d) Библиотека Python для обработки изображений.
- e) Шаблон обработки пикселей.

Вопрос № 11

Какой метод активно используется для генерации изображений на основе текстовых запросов?

- a) Transfer Learning.
- b) Variational Autoencoders (VAE).
- c) Transformer models.
- d) CycleGAN.
- e) DeepDream.

Вопрос № 12

Какой подход используется для сегментации изображений (разделения изображения на области)?

- a) Pixel-level classification.
- b) Object detection.
- c) Image captioning.
- d) Style transfer.
- e) Super-resolution techniques.

Вопрос № 13

Какой элемент необходим для успешного функционирования системы персональных рекомендаций?

- a) История покупок пользователя.
- b) Профиль предпочтений другого пользователя.
- c) Статистика поисковых запросов.
- d) Информационно-поисковая система.
- e) Уникальность товара.

Вопрос № 14

Что описывает концепция object detection в computer vision?

- a) Распознавание лиц на изображениях.
- b) Расположение объектов на картинке и определение их границ.
- c) Создание миниатюр изображений.
- d) Компьютерное зрение в мобильных приложениях.
- e) Генерация искусственно созданных картинок.

Вопрос № 15

Какой механизм реализует deepfake технологии?

- a) Simple feature matching.
- b) Generative adversarial networks (GAN).
- c) Binary decision trees.
- d) Reinforcement learning.

e) Convolutional autoencoders.

Вопрос № 16

Что такое обратный перевод (back translation) в NLP?

- a) Техника повышения разнообразия данных путем перевода предложений на иностранный язык и обратно.
- b) Метод для автоматического извлечения ключевых слов.
- c) Машинный перевод внутри одного языка.
- d) Улучшенный метод транслитерации.
- e) Восстановления оригинального текста после искажения.

Вопрос № 17

Что такое позиционное кодирование (positional encoding) в трансформерах?

- a) Специальный механизм для передачи положения токенов в предложении.
- b) Система подсчета частоты встречаемости слов.
- c) Форма синтаксического анализа текста.
- d) Простой метод сортировки слов по алфавиту.
- e) Типичная схема сокращения словаря.

Вопрос № 18

Какая мера часто используется для оценки эффективности классификационных моделей в NLP?

- a) Mean squared error (MSE).
- b) Accuracy.
- c) Bias-variance tradeoff.
- d) Root mean square deviation (RMSD).
- e) Silhouette coefficient.

Вопрос № 19

Какой алгоритм используется для обнаружения фреймов и рамок вокруг объектов на изображении?

- a) Histogram equalization.
- b) Edge detection filters.
- c) Region proposal networks (RPNs).
- d) Fourier transform.
- e) Median filter.

Вопрос № 20

Какой подход используется для автоматической генерации подписей к изображениям?

- a) Single-layer perceptron.
- b) Hierarchical clustering.
- c) Encoder-decoder architecture.
- d) Bayesian inference.
- e) Hidden Markov Models (HMM).

Вопрос № 21

Что такое masking в attention механизмах трансформеров?

- a) Подавление части входных данных для предотвращения утечки информации.
- b) Усечение длинных предложений.
- c) Агрегация соседних слов в одно целое.

- d) Заполнение пробелов в структуре предложения.
- e) Генерация дополнительных признаков.

Вопрос № 22

Какой термин описывает явление потери информативности глубоких уровней нейронной сети?

- a) Vanishing gradients.
- b) Overfitting.
- c) Exploding gradients.
- d) Feature engineering.
- e) Weight initialization.

Вопрос № 23

Какой тип данных характерен для задач computer vision?

- a) Текстовые файлы.
- b) Видео и изображения.
- c) Звуковые сигналы.
- d) Числа с плавающей точкой.
- e) Биометрические показатели.

Вопрос № 24

Какой инструмент используется для анализа эмоциональной окраски текста?

- a) POS tagging.
- b) Dependency parsing.
- c) Emotion recognition through sentiment analysis.
- d) Optical character recognition (OCR).
- e) Speech-to-text conversion.

Вопрос № 25

Что такое морфологический разбор (morphological parsing) в NLP?

- a) Вычисление грамматических свойств слов (род, число, падеж и др.).
- b) Анализ частотности употребления слов.
- c) Генерация абстрактного содержания текста.
- d) Автоматическое создание художественных произведений.
- e) Распознавание образов на фотографиях.

Вопрос № 26

Какой параметр характеризует качество распознавания лица на изображении?

- a) Цвет фона.
- b) Время суток съемки.
- c) Ассигасу (точность распознавания).
- d) Угол наклона камеры.
- e) Качество освещения.

Вопрос № 27

Что такое обучение с подкреплением (reinforcement learning)?

- a) Метод обучения, направленный на улучшение качества текста.
- b) Стратегия обучения, где агент получает вознаграждение за принятие правильных действий.
- c) Анализ лингвистических особенностей текста.
- d) Структура для сохранения данных.

е) Процесс переноса знаний с одной задачи на другую.

Вопрос № 28

Какой алгоритм используется для фильтрации стоп-слов в документах?

- a) CountVectorizer.
- b) Latent Dirichlet Allocation (LDA).
- c) Stop words removal.
- d) FastText embedding.
- e) Multi-class classification.

Вопрос № 29

Что такое batch normalization в нейронных сетях?

- a) Нормализация признаков перед вводом в сеть.
- b) Улучшение обучения сети путем нормализации внутренних представлений.
- c) Стандартизация размеров пакетов данных.
- d) Замена устаревших слоев новыми.
- e) Умножение весов слоев на постоянный коэффициент.

Вопрос № 30

Что такое эхо-граммы (echograms) в системах синтеза речи?

- a) Характеристика звучания голосовых команд.
- b) Особенность звуковой волны.
- c) Метрика для измерения качества синтезированной речи.
- d) Механизм обратной связи при распознавании голоса.
- e) Репрезентация акустических сигналов, помогающая анализировать спектр звука.

Тестовые вопросы для оценки уровня освоения компетенций (открытый тип):

Раздел I «Теоретико-математические основы искусственного интеллекта»

Вопрос № 1

Название классического метода кластеризации, который формирует заранее указанное количество кластеров путём постепенного перемещения центров масс.

Вопрос № 2

Основной недостаток метода kNN (ближайших соседей): необходимость хранить весь обучающий _____.

Вопрос № 3

Основное преимущество использования генеративных моделей — способность создавать _____.

Вопрос № 4

Автокодировщик состоит из двух частей: _____ и декодировщика.

Вопрос № 5

Байесовская сеть — это графовая модель, где вершины соответствуют случайным величинам, а рёбра указывают на _____.

Вопрос № 6

При какой технике из временного ряда исключают низкочастотные колебания, оставляя высокочастотные детали?

Вопрос № 7

ARMA-модель комбинирует две составляющие: авторегрессивную (AR) и _____ среднюю (MA).

Вопрос № 8

Глубокая нейронная сеть состоит из множества _____, обеспечивающих возможность сложного преобразования признаков.

Вопрос № 9

Подбор лучшей архитектуры нейронной сети и гиперпараметров называется _____.

Вопрос № 10

Основной задачей этапа препроцессинга данных является устранение неполадок, заполнение пропусков и _____ данных.

Вопрос № 11

Процесс поиска границы раздела между классами в задаче классификации носит название _____.

Вопрос № 12

Временные ряды характеризуются наличием трёх основных составляющих: тренда, _____ и остаточной составляющей.

Вопрос № 13

Один из распространённых методов борьбы с переобучением — введение регуляризатора на сумму _____ весов.

Вопрос № 14

Свёрточные слои в нейронных сетях выполняют операцию _____, извлекая характерные особенности изображения.

Вопрос № 15

Для расчёта показателей качества классификации часто используют таблицу, называемую матрицей _____.

Вопрос № 16

Стандартный подход к оценке модели в условиях недостатка данных — проведение _____ валидации.

Вопрос № 17

Логистическая регрессия используется преимущественно для задач _____ классификации.

Вопрос № 18

Подавление эффектов малых амплитуд и усиление больших значений характерно для операции _____ данных.

Вопрос № 19

Генеративные состязательные сети состоят из двух конкурирующих модулей: генератора и _____.

Вопрос № 20

Что получают на выходе обучённого автокодировщика, проходящего через процедуру сжатия и восстановления?

Вопрос № 21

Что такое pooling в свёрточных нейронных сетях?

Вопрос № 22

Вероятностная графа, выражающая совместное распределение случайных величин, называемая _____ сетью.

Вопрос № 23

Вариационные автоэнкодеры отличаются от обычных добавлением ограничений на _____ пространство.

Вопрос № 24

Ключевая идея метода k ближайших соседей (kNN) основана на определении

_____.

Вопрос № 25

Методом ARIMA можно эффективно обработать временные ряды, обладающие характеристиками стационарности и _____.

Вопрос № 26

Модель LSTM обладает специальной структурой с ячейками памяти, позволяющей справляться с проблемой _____ градиента.

Вопрос № 27

Основная задача алгоритма ЕМ (Expectation Maximization) — найти параметры смеси распределений, соответствующие наибольшему значению _____.

Вопрос № 28

Одно из преимуществ обучения с учителем (Supervised Learning) — наличие чётких _____ для оцениваемых моделей.

Вопрос № 29

Отсутствие ярко выраженного тренда в данных свидетельствует о свойствах временного ряда, называемых _____.

Вопрос № 30

Поддерживающие векторные машины (SVM) строятся на поиске оптимальной _____ плоскости.

Раздел II «Практики сбора, очистки и предварительной обработки данных»

Вопрос № 1

Термин, означающий замену отсутствующих значений средними или медианными значениями соответствующих столбцов.

Вопрос № 2

Метод, применяемый для удаления крайних значений в наборе данных.

Вопрос № 3

Преобразование данных, при котором все значения попадают в интервал от 0 до 1.

Вопрос № 4

Статистический метод, позволяющий определить однородность выборки по критерию межквартильного промежутка.

Вопрос № 5

Процедура, необходимая для подготовки больших наборов данных к хранению и обработке.

Вопрос № 6

Тип данных, который подлежит удалению из-за некорректности или невозможности обработки.

Вопрос № 7

Метод, предназначенный для обнаружения аномалий в одномерных данных.

Вопрос № 8

Показатель, рассчитывающий средний процент ошибки прогноза относительно реального значения.

Вопрос № 9

Процедура, выполняемая перед анализом данных для исключения ошибок измерений и артефактов.

Вопрос № 10

Преобразование признаков, сводящее каждое значение к стандартному отклонению от среднего.

Вопрос № 11

Наиболее эффективный метод для обработки больших наборов данных в

ограниченных ресурсных условиях.

Вопрос № 12

Характеристика, показывающая, насколько близко поведение реальных данных соответствует предположенному распределению.

Вопрос № 13

Алгоритм, используемый для автоматического разделения объектов на группы.

Вопрос № 14

Набор техник, направленных на повышение качества данных путём упрощения и унификации.

Вопрос № 15

Вид данных, которые требуют особого внимания из-за своей редкости или важности.

Вопрос № 16

Способ предварительной обработки, при котором выполняется замена определенных значений более общими категориями.

Вопрос № 17

Процедура, позволяющая оценить значимость признаков и исключить малозначимые атрибуты.

Вопрос № 18

Простой способ выявления аномалий, основанный на сравнении текущих значений с историческими показателями.

Вопрос № 19

Тип анализа, проверяющий стабильность данных во времени.

Вопрос № 20

Метод, используемый для улучшения точности предсказаний в рамках ансамбля моделей.

Вопрос № 21

Этапы процесса ETL, необходимые для подготовки больших наборов данных.

Вопрос № 22

Класс методов, предназначенных для изучения структуры данных без наличия заранее определённой маркировки.

Вопрос № 23

Что применяется для расчета доверительного интервала при проверке нормального распределения данных?

Вопрос № 24

Способ интеграции большого объёма разнородных данных в единую систему.

Вопрос № 25

Термин, обозначающий невозможность точно сопоставить реальные данные теоретическому распределению.

Вопрос № 26

Какая статистика помогает обнаружить аномалии в выборке с помощью метода стандартных отклонений?

Вопрос № 27

Совокупность приёмов, направленных на улучшение точности и надёжности анализа данных.

Вопрос № 28

Механизм, служащий для преодоления эффекта доминирования одних признаков над другими.

Вопрос № 29

Процедура, устраняющая проблему гетероскедастичности в регрессионных моделях.

Вопрос № 30

Критерий, позволяющий выбрать оптимальный набор признаков для построения модели.

Раздел III «Технологии и инструменты анализа данных»

Вопрос № 1

Архитектура нейросети, используемая для обработки последовательных данных, таких как текст или звук.

Вопрос № 2

Специализированная архитектура нейросети, применяемая для задач компьютерного зрения, таких как распознавание изображений.

Вопрос № 3

Базовая единица в обработке естественных языков, соответствующая словам или лексическим конструкциям.

Вопрос № 4

Нейросетевая архитектура, известная своей эффективностью в создании фотореалистичных изображений.

Вопрос № 5

Имя архитектуры, ставшей популярной благодаря своим успехам в задачах NLP, особенно известной своими возможностями в понимании контекста.

Вопрос № 6

Подход, при котором рекомендация формируется исходя из прошлого поведения пользователя.

Вопрос № 7

Фундаментальный принцип обработки текстов, при котором слова преобразуются в числовые векторы.

Вопрос № 8

Принцип, лежащий в основе современных моделей рекомендаций, сочетающий контентные и поведенческие подходы.

Вопрос № 9

Известная библиотека Python для реализации продвинутых задач NLP.

Вопрос № 10

Алгоритм, часто используемый для сегментирования изображений в задачах компьютерного зрения.

Вопрос № 11

Одна из популярных архитектур трансформеров, разработанная Google для задач машинного перевода.

Вопрос № 12

Модель, объединяющая принципы самообучения и генеративных моделей, используемая для генерации убедительных текстовых сообщений.

Вопрос № 13

Что означает аббревиатура OCR в контексте технологий обработки данных?

Вопрос № 14

Подход к обучению нейросети, при котором модель учится представлять семантические смыслы слов в виде низкоразмерных векторов.

Вопрос № 15

Главная причина использования сверточных слоев в CNN для задач компьютерного зрения.

Вопрос № 16

Тип ядра, часто используемый в задачах обработки изображений для выделения краев.

Вопрос № 17

Что называется центральной частью архитектуры transformer'a, позволяющей ему учитывать позицию токенов в последовательности?

Вопрос № 18

Процесс, при котором нейросеть самостоятельно выделяет ключевые признаки и учится на собственных ошибках.

Вопрос № 19

Самый известный метод извлечения признаков изображений, реализуемый свёрточными нейросетями.

Вопрос № 20

Специализированная область обработки изображений, нацеленная на восстановление четкого изображения из размытого оригинала.

Вопрос № 21

Метрика, используемая для оценки качества перевода текстов и автоматически сравнивающая переводы с эталонными образцами.

Вопрос № 22

Термин, обозначающий комплекс инструментов и методик для управления потоком данных от поступления до анализа и выдачи рекомендаций.

Вопрос № 23

Выберите название библиотеки Python, специализированной для обработки и анализа текстов.

Вопрос № 24

Тип рекомендаций, основанных на содержании самих продуктов или материалов, представленных пользователю.

Вопрос № 25

Термин, обозначающий формирование выводов или реакций на основании ранее накопленного опыта и обученных моделей.

Вопрос № 26

Система, автоматизирующая выполнение повседневных действий на основе истории активности пользователя.

Вопрос № 27

Алгоритм, предназначенный для точного локализации объектов на изображении и вывода прямоугольников вокруг обнаруженных объектов.

Вопрос № 28

Самая популярная библиотека для обработки естественного языка на платформе Python.

Вопрос № 29

Основная характеристика, позволяющая рекомендательным системам предлагать пользователям персонализированные товары или услуги.

Вопрос № 30

Алгоритм, используемый для перевода естественного языка в машиночитаемый формат.