Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра автоматизированных систем управления

Паспорт экзамена

по дисциплине «Компьютерные технологии анализа и обработки данных», 1 семестр

1. Методика оценки

Экзамен проводится в устной форме, по билетам. Билет состоит из 2 вопросов и формируется по следующему правилу:

- первый вопрос выбирается из диапазона вопросов 1-18;
- второй вопрос из диапазона вопросов 19-36.

Таким образом, проверяются результаты обучения по дисциплине, соотнесенные с установленными в программе индикаторами достижения компетенций.

На экзамене преподаватель вправе задавать студенту уточняющие и дополнительные вопросы из общего перечня (п. 4).

Форма экзаменационного билета

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет АВТФ

Билет № _____
к экзамену по дисциплине «Компьютерные технологии анализа и обработки данных»

---

1. Оцифровки аналоговых данных и специфика обработки числовых и графических форматов
2. Методы работы с неполными данными и выбросами.

Утверждаю: зав. кафедрой ____ _____ должность, ФИО
(подпись)
(дата)

2. Критерии оценки результатов обучения, соотнесенных с уровнями освоения индикаторов достижения компетенций

Ответ на экзаменационный билет засчитывается на **продвинутом** уровне, если студент проводит сравнительный комплексный анализ материала, выявляет проблемы, предлагает механизмы их решения, представляет количественные характеристики определенных процессов, приводит конкретные примеры, не допускает ошибок и способен обосновать выбор метода решения задачи. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций не содержит

пробелов. Установленные в программе компетенции сформированы в полном объеме. Оценка составляет *от 35 до 40 баллов.*

Ответ на экзаменационный билет засчитывается на **базовом** уровне, если студент при ответе на вопросы формулирует основные понятия, дает характеристику процессов, явлений, проводит анализ причин, условий, способен представить количественные и качественные характеристики процессов, не допускает существенных ошибок при решении задачи. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит несущественные пробелы. Установленные в программе компетенции сформированы на базовом уровне. Оценка составляет *от 34 до* 27 *баллов.*

Ответ на экзаменационный билет засчитывается на **пороговом** уровне, если студент при ответе на вопросы дает определение основных понятий, может показать причинно-следственные связи явлений, при решении задачи допускает непринципиальные ошибки, например, вычислительные. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит пробелы. Установленные в программе компетенции сформированы на пороговом уровне. Оценка составляет *от* 26 *до* 20 *баллов.*

Ответ на экзаменационный билет считается **неудовлетворительным** (ниже порогового уровня), если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений, при решении задачи допускает принципиальные ошибки. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит существенные пробелы. Установленные в программе компетенции не сформированы. Оценка составляет *менее* 20 *баллов.*

3. Шкала оценки

Экзамен считается сданным, если сумма баллов по всем заданиям билета составляет от 20 до 40 баллов включительно. Сумма менее 20 баллов признается неудовлетворительным результатом промежуточной аттестации по дисциплине.

В общей оценке по дисциплине экзаменационные баллы учитываются в соответствии с правилами балльно-рейтинговой системы, установленными в НГТУ.

4. Вопросы к экзамену по дисциплине «Компьютерные технологии анализа и обработки данных»

1. Понятия «информация» и «данные»: терминологическая граница. Характеристики данных согласно ГОСТ Р ИСО/МЭК 20546-2021.
2. Оцифровки аналоговых данных и специфика обработки числовых и графических форматов.
3. Концептуальные различия между информационными (ИТ) и компьютерными технологиями (КТ). Роль вычислительной техники в процессах обработки данных.
4. Жизненный цикл обработки данных: этапы от сбора и оцифровки до представления знаний.
5. Эволюция подходов к организации данных и методам их анализа.
6. Отличие анализа данных от интеллектуального анализа данных (ИАД): роль человека, методы, степень организованности данных и ожидаемый результат.

7. Принципы оцифровки аналоговых сигналов (дискретизация, квантование) и их влияние на достоверность данных при последующем анализе.
8. Особенности представления и обработки числовых данных в задачах машинного обучения: проблемы масштабирования, нормализации и влияния выбросов на качество моделей.
9. Специфика графических данных как слабо- или неструктурированной информации. Особенности извлечения признаков из изображений для задач интеллектуального анализа (на примере датасета MNIST).
10. Общие принципы кодирования и сжатия информации (с потерями и без потерь) в контексте оптимизации ресурсов хранения и скорости обработки больших массивов данных.
11. Разреженное хеширование как специфический метод «кодирования» данных: принцип расширения размерности пространства для выявления скрытых сходств (на примере алгоритм FlyHash).
12. Процесс обнаружения знаний в базах данных (KDD): содержание и задачи пяти основных этапов.
13. Методология SEMMA: назначение и содержание этапов выборки, исследования, модификации, моделирования и оценки.
14. Методология CRISP-DM: жизненный цикл проекта интеллектуального анализа данных.
15. Концептуальное соответствие этапов методологий KDD, SEMMA и CRISP-DM.
16. Определение и характеристики больших данных (Big Data) согласно ГОСТ Р ИСО/МЭК 20546-2021. Модель «3V» Дугласа Лэйни.
17. Сравнительный анализ классического интеллектуального анализа данных (ИАД) и интеллектуального анализа больших данных (ИАБД).
18. Основные методы ИАД: классификация, кластеризация, регрессия, обнаружение аномалий, ассоциативные правила. Их назначение и примеры применения.
19. Определения алгоритма и модели машинного обучения согласно ГОСТ Р 71476-2024.
20. Разграничение этапов обучения (training) и применения (inference) в жизненном цикле систем машинного обучения.
21. Фундаментальные парадигмы машинного обучения (обучение с учителем, обучение без учителя, обучение с подкреплением).
22. Очистка и предварительная обработка данных.
23. Цели и задачи исследовательского анализа данных (EDA).
24. Методы работы с неполными данными и выбросами.
25. Трансформация и кодирование признаков.
26. Место инжиниринга признаков в машинном обучении. Методы отбора признаков.
27. Проблема проклятия размерности и методы снижения размерности.
28. Обучение, оценка и настройка модели машинного обучения. Понятие и роль baseline-модели в оценке результатов машинного обучения.
29. Метрики качества моделей машинного обучения. Диагностика переобучения и недообучения.

30. Специфика неструктурированных текстовых данных и этапы конвейера лингвистической предобработки (NLP-пайплайн: токенизация, лемматизация, стемминг, удаление стоп-слов).
31. Эволюция моделей анализа текста: от Bag-of-Words и TF-IDF к распределенным векторным представлениям (эмбеддингам).
32. Архитектурные принципы больших языковых моделей на основе механизма внимания и парадигма «предобучение и тонкая настройка» (pre-train and fine-tune).
33. Эволюция подходов к взаимодействию с большими языковыми моделями: от точечного инжиниринга промптов к системному инжинирингу контекста.
34. Анатомия контекстного окна LLM: системные инструкции, динамическая память, интеграция результатов RAG и вызовы внешних инструментов.
35. Архитектурные особенности и методы оптимизации компактных языковых моделей (SLM) на примере семейства Qwen (7B параметров) для периферийных вычислений.
36. Архитектуры гибридного интеллектуального анализа: интеграция рассуждений больших языковых моделей с выполнением структурированных запросов к реляционным базам данных (на примере технологии BlendSQL).