

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра вычислительной техники

Паспорт экзамена

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

1. Методика оценки

Экзамен проводится в устной форме, по билетам. Билет формируется по следующему правилу: выбирается два вопроса из диапазона вопросов 1-23 (список вопросов приведен ниже). В ходе экзамена преподаватель вправе задавать студенту 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. Основные понятия АД, типы задач АД, примеры задач. Математический аппарат АД.
2. Анализ и прогнозирование временных рядов. Этапы решения задачи прогнозирования временного ряда на примере.

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

2. Критерии оценки

- Ответ на экзаменационный билет считается **неудовлетворительным**, если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений, оценка составляет ниже 20 баллов.
- Ответ на экзаменационный билет засчитывается на **пороговом** уровне, если студент отвечает на два вопроса, но с серьезными замечаниями, оценка составляет 20-25 баллов.
- Ответ на экзаменационный билет засчитывается на **базовом** уровне, если студент отвечает на два вопроса, но с незначительными замечаниями, оценка составляет 26-34 баллов.
- Ответ на экзаменационный билет засчитывается на **продвинутом** уровне, если студент при ответе на вопросы проводит сравнительный анализ подходов к решению задачи статистического анализа данных, выявляет проблемы, предлагает механизмы решения, приводит конкретные примеры из практики, не допускает ошибок и способен обосновать выбор метода решения задачи анализа данных, оценка составляет 35-40 баллов.

3. Шкала оценки

В общей оценке по дисциплине экзаменационные баллы учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины.

Экзамен считается сданным, если средняя сумма баллов по всем заданиям составляет не менее 20 баллов (максимальное количество баллов за зачет – 40).

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

Вопросы к экзамену

1. Основные понятия АД, типы задач АД, примеры задач. Математический аппарат АД.
2. Основные этапы компьютерной технологии анализа данных, особенности.
3. Графический разведочный анализ данных. Методы визуализации данных на примере прикладного пакета Statistica. Достоинства и недостатки визуальных методов анализа данных.
4. Первичный разведочный анализ данных (ПРАД). Основные числовые характеристики данных.
5. Первичный разведочный анализ данных (ПРАД). Исследование закона распределения данных. Критерии согласия.
6. Основные методы поиска ассоциативных правил. Характеристики качества правил ассоциации.
7. Кластерный анализ. Постановка задачи. Примеры задач АД. Методологические этапы проведения кластерного анализа. Используемые меры расстояния между объектами и выбор меры. Достоинства и недостатки различных мер расстояний.
8. Кластерный анализ. Классификация методов кластерного анализа данных. Иерархические агломеративные методы (метод одиночной связи, метод полной связи, метод Варда), метод К-средних. Достоинства и недостатки методов кластеризации.
9. Кластерный анализ. Оценка адекватности полученных решений: кофенетическая корреляция; тесты значимости признаков на основе дисперсионного анализа; метод повторных выборок; процедура Монте-Карло.
10. Деревья классификации (деревья решений). Примеры задач АД. Основные понятия, связанные с деревьями классификации, показатели сложности и точности деревьев решений. Методологические этапы построения дерева решений.
11. Деревья классификации (деревья решений). Выбор критерия точности прогноза. Метод CART построения дерева решений.
12. Деревья классификации (деревья решений). Определение момента прекращения ветвления, правила останова. Процедуры проверки прогнозной способности построенного дерева решений.
13. Деревья классификации (деревья решений). Выбор размера дерева, кросс-проверочное отсечение по минимальной цене сложности.
14. Анализ и прогнозирование временных рядов. Примеры задач АД. Определение временного ряда. Классификация факторов, влияющих на формирование значений временного ряда, структурная модель временного ряда.
15. Анализ и прогнозирование временных рядов. Этапы решения задачи прогнозирования временного ряда на примере.
16. Анализ и прогнозирование временных рядов. Методы определения вида функции тренда.

17. Анализ и прогнозирование временных рядов. Гармонический анализ для определения вида сезонной составляющей временного ряда.
18. Анализ и прогнозирование временных рядов. Анализ случайных остатков. Определение адекватности построенной модели (верификация модели) данным наблюдения.
19. Генетические алгоритмы. Определение, задачи АД, решаемые на основе использования генетических алгоритмов. Основные понятия и термины, принятые в области генетических алгоритмов. Характеристики генетических алгоритмов.
20. Эволюционное моделирование. Основные определения. Применение алгоритмов эволюционного моделирования для решения задач структурной идентификации модели.
21. Компьютерные технологии решения разных типов задач АД в современном статистическом программном обеспечении на примере реальных пакетов.
22. Технологии Data Mining, основные этапы, особенности.
23. Направления развития методов, технологий и средств решения задач АД.