



Федеральное агентство морского и речного транспорта
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«Государственный университет морского и речного флота
имени адмирала С.О. Макарова»**

Воронежский филиал ФГБОУ ВО «ГУМРФ имени адмирала С.О. Макарова»

Кафедра математики, информационных систем и технологий

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине «*Основы Data Mining*»
(приложение к рабочей программе дисциплины)

Направление подготовки 09.03.02 Информационные системы и технологии

Направленность (профиль) Информационные системы на транспорте

Уровень высшего образования бакалавриат

Форма обучения очная, очно-заочная

г. Воронеж
2021

1. Перечень компетенций и этапы их формирования в процессе освоения дисциплины

Рабочей программой дисциплины Основы Data Mining предусмотрено формирование следующих компетенций.

Таблица 1

Перечень компетенций и этапы их формирования в процессе освоения дисциплины

Код и наименование компетенции	Код индикатора достижения компетенции	Планируемые результаты обучения по дисциплине
ПКР-1: Способность проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла	ПКР-1.1	Знать: методы проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла
	ПКР-1.2	Уметь: проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.
	ПКР-1.3	Иметь навыки: проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла

2. Паспорт фонда оценочных средств для проведения текущей и промежуточной аттестации обучающихся

Таблица 2

Оценочные средства для проведения текущей и промежуточной аттестации обучающихся

№ п/п	Наименование раздела (темы) дисциплины	Формируемая компетенция	Наименование оценочного средства
1	OLTP-системы (On-Line Transaction Processing).	ПКР-1	<i>тестирование зачет</i>
2	Неэффективность использования OLTP-систем для решения задач анализа данных.	ПКР-1	<i>тестирование зачет</i>
3	Концепция хранилища данных (ХД).	ПКР-1	<i>тестирование зачет</i>
4	Организация ХД.	ПКР-1	<i>тестирование зачет</i>
5	Очистка данных.	ПКР-1	<i>тестирование зачет</i>
6	Многомерная модель данных. OLAP-системы.	ПКР-1	<i>тестирование зачет</i>
7	Архитектура OLAP-систем.	ПКР-1	<i>тестирование зачет</i>

№ п/п	Наименование раздела (темы) дисциплины	Формируемая компетенция	Наименование оценочного средства
8	Data Mining – интеллектуальный анализ данных.	ПКР-1	тестирование зачет
9	Клиентские OLAP-средства.	ПКР-1	тестирование зачет
10	Серверные OLAP-средства.	ПКР-1	тестирование зачет
11	Инструменты управления службой SSAS.	ПКР-1	тестирование зачет

Таблица 3

Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине

Результат обучения по дисциплине	Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине				Процедура оценивания
	2	3	4	5	
	Не зачтено	Зачтено			
ПКР-1.1 Знать: методы проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Отсутствие или фрагментарные представления о методах проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Неполные представления о методах проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Сформированные, но содержащие отдельные пробелы представления о методах проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Сформированные систематические представления о методах проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	тестирование зачет
ПКР-1.2 Уметь: проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Отсутствие умений или фрагментарные умения проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	В целом удовлетворительные, но не систематизированные умения проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	В целом удовлетворительные, но содержащие отдельные пробелы умения проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Сформированные умения проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	тестирование зачет
ПКР-1.3 Иметь навыки: проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Отсутствие владения или фрагментарные владения навыками проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	В целом удовлетворительные, но не систематизированные владения навыками проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	В целом удовлетворительные, но содержащие отдельные пробелы владения навыками проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	Сформированные владения навыками проведения научных исследований при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла.	тестирование зачет

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ

Тестовые задания для проведения текущего контроля

Тестовые задания по теме 1

1. «Извлечение полезных сведений невозможно без хорошего понимания сути данных», верно ли утверждение:

- а) верно
- б) неверно. Технологии не нужно понимание данных
- в) неверно. Технологии Data Mining не нужен аналитик, поэтому понимание кем-либо данных — излишне

2. Большинство аналитических методов, используемые в технологии Data mining – это:

- а) новейшие математические алгоритмы и методы
- б) известные математические алгоритмы и методы
- в) классические статистические методы

3. Выберите характеристику, наиболее подходящую для Data Mining:

- а) подходит для понимания ретроспективных данных
- б) подходит для обобщения ретроспективных данных
- в) опирается на ретроспективные данные для получения ответов на вопросы о будущем

4. Частью какой из перечисленных стадий является валидация закономерностей:

- а) свободный поиск
- б) анализ исключений
- в) прогностическое моделирование

5. Какая из перечисленных ниже групп методов достаточно часто использует для выявления взаимосвязей в данных концепцию усреднения по выборке:

- а) Data Mining
- б) OLAP
- в) статистические методы

6. На стадии свободного поиска осуществляется:

- а) использование выявленных закономерностей для предсказания неизвестных значений
- б) выявление закономерностей
- в) анализ исключений

7. В результате использования инструментов Data Mining пользователь может:

- а) получить гипотезы о взаимосвязях в данных, самостоятельно выдвинутые инструментом Data Mining
- б) получить подтверждение или опровержение гипотез, выдвинутых пользователем
- в) оба варианта верны
- г) нет верного ответа

8. Нейронные сети относятся к группам:

- а) методов на основе уравнений
- б) статистических методов
- в) методов кросс-табуляции

9. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:

- а) быть очевидными
- б) чем больше найдено закономерностей, тем лучше
- в) быть неочевидными

10. Какой из перечисленных ниже пунктов не является названием стадии Data Mining:

- а) свободный поиск
- б) индукция правил
- в) анализ исключений

11. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:

- а) быть объективными
- б) быть очевидными
- в) чем больше найдено закономерностей, тем лучше

12. Какие из перечисленных ниже пунктов являются названиями стадий Data Mining:

- а) прогностическое моделирование
- б) свободный поиск
- в) оба варианта верны
- г) нет верного ответа

13. Инструменты Data Mining:

- а) могут самостоятельно строить гипотезы о взаимосвязях в данных
- б) могут самостоятельно строить гипотезы о взаимосвязях в данных, которые обязательно подтверждаются
- в) не могут самостоятельно строить гипотезы о взаимосвязях в данных

14. Для какой шкалы применимы только такие операции как равно, не равно, больше, меньше:

- а) номинальная шкала
- б) интервальная шкала
- в) порядковая шкала

15. Если сравнивать Data Mining, машинное обучение и статистику, какая из дисциплин сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов:

- а) Data Mining
- б) статистика
- в) машинное обучение

16. К какой категории данных относится вес измеряемых объектов:

- а) дискретным данным
- б) непрерывным данным
- в) оба варианта верны
- г) нет верного ответа

17. Назовите фактор, обусловивший возникновение и развитие Data Mining:

- а) совершенствование аппаратного и программного обеспечения
- б) совершенствование технологий хранения и записи данных
- в) оба варианта верны
- г) нет верного ответа

18. Для какой шкалы применимы только такие операции как равно и не равно:

- а) порядковая шкала
- б) номинальная шкала
- в) интервальная шкала

19. В процессе работы Data Mining программы пользователь может получить такие результаты:

- а) только статистически достоверные результаты
- б) только верные результаты, ложные выводы исключены
- в) большой процент ложных, недостоверных или бессмысленных результатов

20. Такие данные как температура воздуха относятся к:

- а) дискретным данным
- б) непрерывным данным
- в) оба варианта верны
- г) нет верного ответа

21. Data Mining — это процесс обнаружения в сырых данных знаний, необходимых для:

- а) принятия решений в различных сферах человеческой деятельности
- б) увеличения стоимости анализа данных
- в) замены аналитика в процессе принятия решений

22. Объект описывается как:

- а) свойство, характеризующее объект
- б) набор атрибутов
- в) поле таблицы

23. Подготовка данных в процессе Data Mining является:

- а) необязательным этапом работы
- б) может вообще отсутствовать
- в) существенным этапом работы

24. Свойство, характеризующее объект:

- а) Данные
- б) Атрибут
- в) Инструменты Data Mining

25. Какая из перечисленных дисциплин более сосредоточена на теории проверки гипотез:

- а) Data Mining
- б) визуализация
- в) статистика

26. Процесс обнаружения в сырых данных знаний, необходимых для принятия решений в различных сферах человеческой деятельности:

- а) Данные
- б) Data Mining
- в) Атрибут

27. Data Mining — это процесс обнаружения в сырых данных:

- а) практических закономерностей
- б) большого количества закономерностей
- в) ранее сформулированных гипотез

28. Номинальная шкала – это шкала:

- а) содержащая категории, которые могут упорядочиваться
- б) содержащая только две категории
- в) содержащая только категории, которые не могут упорядочиваться

29. Data Mining — это процесс обнаружения в сырых данных:

- а) объективных закономерностей

- б) ранее сформулированных гипотез
- в) большого количества закономерностей

30. Объектом не является:

- а) строка таблицы
- б) переменная
- в) запись

Тестовые задания по теме 2

1. Какие тенденции в области ИТ-технологий способствовали появлению СППР?:

- а) Потребности топ менеджмента РФ;
- б) Широкое распространение персональных компьютеров и средств телекоммуникаций;
- в) Отсутствие стандартизированных пакетов прикладных программ;
- г) Достижения в области искусственного интеллекта.

2. Ключевой особенностью информационной технологии поддержки принятия решения является:

- а) Наличие разработанных методов принятия решений;
- б) Широкое распространение средств вычислительной техники и предметно-ориентированного программного обеспечения;
- в) Качественно новый метод организации взаимодействия ЛПР и прикладного программного обеспечения;
- г) Профессиональный рост в плане наличия компетенций в области ИТ-технологий у менеджеров.

3. Сбор и хранение информации, а также решение информационно поисковых задач средствами систем управления базами данных (СУБД) осуществляется в рамках:

- а) Хранилищ данных;
- б) OLTP (Online Transaction Processing) подсистем, реализующих транзакционную обработку данных;
- в) Витрин (киосков) данных;
- г) OLAP – систем.

4. В основе концепции хранилищ данных (ХД) лежит идея:

- а) Использования базы данных метаданных;
- б) Разделения данных, используемых для оперативной обработки и для решения задач анализа;
- в) Создания единой структуры хранения данных;

г) Иерархического построения хранимых наборов данных.

5. ХД интегрирует ранее разъединенные данные, содержащиеся в архивах, накапливаемых OLTP-системами из внешних источников, в единую базу данных, осуществляя их:

- а) Сортировку;
- б) Группировку;
- в) Предварительное согласование и агрегацию;
- г) Выборку и фильтрацию.

6. Подсистема анализа может быть построена на основе подсистемы:

- а) информационно-поискового анализа данных;
- б) информационно-поискового анализа на базе реляционных СУБД и статических запросов с использованием языка SQL;
- в) подсистемы оперативного анализа. Для реализации таких подсистем применяется технология оперативной аналитической обработки данных OLAP, использующая концепцию многомерного представления данных;
- г) подсистемы интеллектуального анализа, реализующие методы и алгоритмы Data Mining.

7. Обобщенная архитектура СППР может состоять из следующих элементов:

- а) Система управления данными (the data management system, DBMS);
- б) Система управления моделями (the model management system, MBMS),
- в) Машина знаний (the knowledge engine, KE),
- г) Интерфейс пользователя (the user interface) – лица, принимающего решения (ЛПР).

8. Выработка решения в рамках информационной технологии поддержки принятия решения происходит в рамках итерационного процесса, в котором участвуют:

- а) Традиционная транзакционная АИС;
- б) Информационно-справочная система;
- в) ЛПР;
- г) Сотрудники ИТ-подразделения предприятия.

9. Отличительными характеристиками информационной технологии поддержки принятия решений являются:

- а) Ориентация на решение слабо структурированных задач;
- б) Сочетание традиционных методов обработки данных с возможностями математических моделей;
- в) Направленность на профессионального пользователя;
- г) Низкая адаптивность (в плане учета особенностей технического и программного обеспечения, а также потребностей пользователей).

10. Без каких компонентов СППР не сможет решать в полном объеме возложенные на нее задачи:

- а) База данных, база моделей;
- б) Универсальные системные утилиты обеспечения совместимости с операционной системой;
- в) СУБД, СУБМ, система управления интерфейсом системы;
- г) Антивирусное программное обеспечение.

11. К какому классу программных средств в большей степени можно отнести СППР?

- а) Традиционные транзакционные системы;
- б) Информационно-справочные системы;
- в) Информационно-аналитические системы;
- г) Интеллектуальные информационные системы.

Тестовые задания по теме 3-4

1. Сбор и хранение информации, а также решение информационно поисковых задач средствами систем управления базами данных (СУБД) осуществляется в рамках:

- а) Хранилищ данных;
- б) OLTP (Online Transaction Processing) подсистеме, реализующей транзакционную обработку данных;
- в) Витрин данных;
- г) OLAP – систем.

2. В основе концепции ХД лежит идея:

- а) Использования базы данных метаданных;
- б) Разделения данных, используемых для оперативной обработки и для решения задач анализа;
- в) Создания единой структуры хранения данных;
- г) Иерархического построения хранимых наборов данных;

3. ХД интегрирует ранее разъединенные данные, содержащиеся в архивах, накапливаемых OLTP-системами из внешних источников, в единую базу данных, осуществляя их:

- а) Сортировку;
- б) Группировку;
- в) Предварительное согласование и агрегацию;
- г) Выборку и фильтрацию.

4. Подсистема анализа может быть построена на основе подсистемы:

- а) информационно-поискового анализа данных;
- б) информационно-поискового анализа на базе реляционных СУБД и статических запросов с использованием языка SQL;

- в) подсистемы оперативного анализа. Для реализации таких подсистем применяется технология оперативной аналитической обработки данных OLAP, использующая концепцию многомерного представления данных;
- г) подсистемы интеллектуального анализа, реализующие методы и алгоритмы Data Mining.

5. Что Вы понимаете под Хранилищем данных?

- а) Объектно-ориентированная база данных;
- б) Предметно-ориентированная корпоративная база данных, предназначенная для подготовки отчетов, анализа бизнес-процессов и поддержки принятия решений.
- в) Субъектно-ориентированная информационная совокупность файлов;
- г) База знаний СППР.

6. Что характерно для хранилищ данных:

- а) интеграция разнородных данных;
- б) интегрированная единая структура хранения данных, позволяющая эффективно применять модели реляционных баз данных;
- в) эффективное хранение и обработка больших объемов данных;
- г) единая распределенная многоплановая структура хранения данных, позволяющая эффективно применять сетевые модели баз данных;

7. Что еще характерно для хранилищ данных:

- а) Широкое использование нормативно-справочной информации, классификаторов и кодификаторов;
- б) Организация многоуровневых справочников метаданных;
- в) Обеспечение информационной безопасности ХД;
- г) Тесная привязка к системам оказания государственных услуг.

8. Сокращение затрат на разработку ХД может быть достигнуто путем создания:

- а) Баз данных;
- б) Витрин данных (ВД);
- в) Баз знаний;
- г) Выборок данных.

9. Виртуальное (распределенное) ХД. Выберите из приведенных ниже, правильные утверждения:

- а) В такой системе данные из OLTP-системы не копируются в единое хранилище;
- б) Данные из распределенного ХД извлекаются, преобразуются и интегрируются непосредственно при выполнении аналитических запросов в режиме реального времени;
- в) Интеграция данных производится автоматически в процессе сбора исходных данных;

г) Распределенное ХД формируется исходя из предпочтений ЛПР.

10. Достоинства виртуального ХД:

- а) Минимизация затрат на приобретение ПО;
- б) Повышение оперативности извлечения и обработки данных
- в) Возможность онтологического поиска требуемых данных в процессе подготовки решения;
- г) Минимизация объема хранимых данных (работа с текущими, актуальными данными);

11. Недостатки виртуального ХД, выберите верные ответы:

- а) более высокое, по сравнению с физическим ХД, время обработки запросов;
- б) необходимость постоянной доступности всех OLTP-источников;
- в) снижение быстродействия OLTP-систем;
- г) OLTP-системы не ориентированы на хранение данных за длительный период времени, по мере необходимости данные выгружаются в архивные, поэтому не всегда имеется физическая возможность получения полного набора данных.

12. Витрины данных. Выберите верные утверждения:

- а) Сокращение затрат на разработку ХД может быть достигнуто путем создания витрин данных (ВД);
- б) ВД упрощенный вариант ХД, содержащий самую разнообразную информацию как по структуре, так и по содержанию;
- в) ВД – неотъемлемый элемент ХД;
- г) ВД подмножество ХД, содержащее проблемно-ориентированные данные.

13. Данные в ХД делятся на категории, выберите верные:

- а) детальные данные (данные соответствующие элементарным событиям, фиксируемым в OLTP-системах. Подразделяются на: измерения наборы данных, необходимые для описания событий (товар, продавец, покупатель, магазин, ...); факты данные, отражающие сущность события (количество проданного товара, сумма продаж, ...);
- б) агрегированные (обобщенные) данные, получаемые на основании детальных путем суммирования по определенным измерениям;
- в) метаданные данные о данных, содержащихся в ХД;
- г) ключевые данные, определяющие выбор того или иного решения.

14. Основными таблицами ХД являются:

- а) таблицы фактов;
- б) таблицы измерений;
- в) реляционные таблицы;
- г) иерархические таблицы.

15. Таблица фактов содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться. Обычно выделяют такие типы фактов, как:

- а) Факты, связанные с транзакциями (Transaction facts);
- б) Факты, связанные с «моментальными снимками» (Snapshot facts);
- в) Факты, связанные с элементами документа (Line-item facts);
- г) Факты, связанные с событиями или состоянием объекта (Event or state facts).

16. Выберите правильные утверждения:

- а) OLAP (Online Analytical Processing) технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений.
- б) Цель OLAP-анализа – принятие решения аналитиком.
- в) Полномасштабная OLAP-система должна выполнять сложные и разнообразные функции, включающие сбор данных из различных источников, их согласование, преобразование и загрузку в хранилище, хранение аналитической информации, регламентную отчетность, поддержку произвольных запросов, многомерный анализ и др.
- г) OLAP (Online Analytical Processing) система оперативной обработки данных, основанная на применении реляционной модели данных.

17. Аналитические CRM-системы, выберите правильные утверждения:

- а) Имеет место сокращение затрат на проведение маркетинговой деятельности за счет использования широкого спектра аппаратно- программных средств, базирующихся на технологиях реляционных баз данных;
- б) Внедрение такого решения позволяет оптимизировать цепочки работы с клиентами, провести персонализацию обслуживания клиентов, повысить доходы от продаж, а также позволяют разрабатывать стратегии расширения рынка за счет привлечения клиентов на основе индивидуального подхода;
- в) CRM – неотъемлемый элемент автоматизации любого бизнес- процесса;
- г) CRM – при любых экономических условиях дает максимальный эффект для менеджмента компании.

18. Аналитические SRM (Supply Relationship Management) системы, выберите правильные утверждения:

- а) Занимаются управлением взаимоотношениями с поставщиками;
- б) Обеспечивают снижение затрат (от 50% и выше), потока сырья, планирования, исполнения и контроля прохождения;
- в) В процессе обучения персонала не возникают сложности;
- г) Система гарантированно дает увеличение прибыли на 30-40%.

19. Аналитические SCM - системы, выберите правильные утверждения:

- а) Занимаются управлением взаимоотношениями с поставщиками;
- б) Максимизация издержек сети сбыта;

- в) Увеличение времени внедрения новых производственных технологий;
- г) Представляют собой информационные системы для решения задач анализа и оптимизации в управлении жизненным циклом продукции.

20. Виртуальные предприятия, выберите правильные утверждения:

- а) Занимаются управлением систем продаж предприятия;
- б) Максимизация издержек сети сбыта;
- в) Увеличение времени внедрения новых производственных технологий;
- г) Представляют собой информационные системы для решения задач анализа и оптимизации в управлении жизненным циклом продукции.

21. Мультимедийные хранилища данных, выберите правильные утверждения:

- а) Внутренний формат для представления таких данных может быть представлен простым типом данных реляционных СУБД;
- б) Поиск данных прост при использовании стандартных средств традиционных СУБД;
- с) Примерами мультимедийных ХД являются разрабатываемые во всем мире электронные хранилища музейных данных;
- д) Представляют собой информационные хранилища для обработки больших данных и оптимизации в управлении.

22. Какие прикладные задачи решаются в рамках СУБД NoSQL при обработке временных рядов:

- а) Оказание финансовых услуг;
- б) Управление запасами;
- в) Предоставление коммерческой информации;
- г) Оптимизационные задачи.

23. Программные инструментальные средства информационных аналитических систем включают в себя:

- а) Средства системного сопровождения;
- б) Средства оперативного и интеллектуального анализа данных;
- в) Средства обеспечения информационной безопасности;
- г) Средства сбора, доработки и преобразования данных.

Тестовые задания по теме 6-11

1. Что понимается под термином OLAP (On-Line Analytical Process)?

- а) Оперативная обработка данных для управления;
- б) Обработка данных в режиме on-line;
- в) Интерактивная аналитическая обработка данных;
- г) Процесс фильтрации интегрированных данных.

2. OLAP наилучшего эффекта можно достичь с использованием:

- а) Реляционных баз данных;
- б) Хранилищ данных (Data Warehouse);
- в) Языка структурированных запросов SQL;
- г) Узкоспециализированных витрин данных.

3. Основной задачей хранилища является предоставление данных для анализа в простой и понятной форме:

- а) Предоставление данных для анализа в простой и понятной форме;
- б) Обеспечение сохранности данных;
- в) Обеспечение безопасности хранимых данных;
- г) Оперативная обработка данных.

4. Для эффективной работы аналитику требуется (правильных ответов несколько):

- а) База данных метаданных;
- б) Язык структурированных запросов SQL;
- в) Централизация всех данных и структурирование информации;
- г) Удобные инструменты для просмотра и визуализации информации.

5. OLAP организует данные в виде:

- а) Многомерных кубов (cubes);
- б) Сетевых или иерархических БД;
- в) Реляционных баз данных.

6. Многомерный анализ определяется как:

- а) информационно-поисковый анализ данных;
- б) информационно-поисковый анализ на базе реляционных СУБД и статических запросов с использованием языка SQL;
- в) Одновременный анализ по нескольким измерениям с последующей консолидацией;
- г) интеллектуальный анализ данных заложенной в них семантики.

7. Выберите имеющие место на практике методы извлечения информации из кубов данных:

- а) Манипуляционная операция «Вращение»;
- б) Манипуляционная операция управления моделями;
- в) Манипуляционная операция использования машины знаний (the knowledge engine, KE);
- г) Интерфейс пользователя (the user interface).

8. Выберите имеющие место на практике методы извлечения информации из кубов данных:

- а) Традиционная фильтрация данных;

- б) Выборка данных по некоторым полям реляционной базы данных; в) Отбор данных исходя из предпочтений ЛПР;
- г) Отношения и Иерархические Отношения.

9. Выберите имеющие место на практике методы извлечения информации из кубов данных:

- а) Операция «Агрегации»;
- б) Операция «Детализации»;
- в) Формирование «Среза»;
- г) Операция «Традиционный отчет».

10. Что Вы понимаете под OLAP-отчетом?

- а) Трехмерное представление требуемых данных;
- б) Многомерное представление выбранного набора данных;
- в) БД, обработанная с помощью выбранной СУБД;
- г) Управляемая динамическая OLAP-таблица, которая сопровождается синхронной диаграммой (графиком).

11. В чем заключаются требования Кодда к средствам оперативной аналитической обработки – OLAP -системам? (выберите несколько)

- а) Кодд сформулировал концепцию комплексного многомерного анализа данных, накопленных в хранилище;
- б) Невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений»;
- в) OLAP-системы должны предоставлять возможность выполнения оперативной аналитической обработки ЛПР;
- г) Отсутствие поддержки многопользовательского режима.

Оценка результатов тестирования.

За каждый правильный ответ начисляется 1 балл. Для перевода баллов в оценку применяется универсальная шкала оценки образовательных достижений. Если обучающийся набирает

- от 90 до 100% от максимально возможной суммы баллов - выставляется оценка «отлично» (зачтено);
- от 80 до 89% - оценка «хорошо» (зачтено);
- от 51 до 79% - оценка «удовлетворительно» (зачтено);
- менее 51% - оценка «неудовлетворительно» (не зачтено).

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОГО КОНТРОЛЯ

Промежуточная аттестация – зачет

Вопросы к зачету

1. Понятие data mining, компоненты data mining.
2. Цель применения сантимерт анализа (три типа моделей, привести пример по каждой модели).
3. Что такое препроцессинг документов для целей тематического моделирования и сантимерт анализа.
4. Цели и задачи кластерного анализа.
5. Направления в кластерном анализе.
6. Меры близости в кластерном анализе.
7. Алгоритм K-means
8. Алгоритм иерархической кластеризации
9. Достоинство и недостатки кластерного подхода.
10. Особенности тематического моделирования. Цель применения тематического моделирования.
11. Достоинства и недостатки тематического моделирования.
12. Суть метода наименьших квадратов (МНК). Цель применения метода МНК.
13. Суть модели скользящего среднего (МСС). Цели применения СС.
14. Суть модели Брауна (МБ). Цели применения МБ.
15. Принцип выбора адекватной модели в прогнозировании.
16. Цель применения регрессионной модели.
17. Суть регрессионной модели (смысл коэффициентов регрессии).
18. Виды регрессионных моделей.
19. Достоинства и недостатки регрессионных моделей.
20. Особенности сетевого анализа.
21. Суть применения сетевого анализа.
22. Понятие о нейронной сети.
23. Особенности нейронного моделирования.
24. Достоинства и недостатки нейронного моделирования.

Критерии оценки ответов на зачете

Таблица 5

Критерии оценки

Наименование показателя	Критерии оценки	Максимальное количество баллов	Количество баллов
I. КАЧЕСТВО ОТВЕТА			
1 Соответствие ответов, поставленным вопросам	- систематизированные, глубокие и полные знания по всем разделам учебной программы - полное и глубокое усвоение	10	

	основной и дополнительной литературы, рекомендованной рабочей программой дисциплины - умение ориентироваться в теориях, концепциях и направлениях по изучаемой дисциплине		
2. Грамотность изложения	- владение терминологией и понятийным аппаратом проблемы; - научный стиль изложения.	5	
3. Самостоятельность выполнения работы, глубина проработки материала, использование рекомендованной и справочной литературы	- степень знакомства автора работы с актуальным состоянием изучаемой проблематики; - дополнительные знания, использованные при написании работы, которые получены помимо предложенной образовательной программы;	5	
Общая оценка за выполнение		20	
ОТВЕТЫ НА ДОПОЛНИТЕЛЬНЫЕ ВОПРОСЫ ПО СОДЕРЖАНИЮ РАБОТЫ			
Вопрос 1		5	
Вопрос 2		5	
Общая оценка за ответы на вопросы		10	
Итого		30	

Для перевода баллов критериально-шкалированной таблицы в оценку применяется универсальная шкала оценки образовательных достижений. Если студент набирает 18-30 баллов и выше - оценка «зачтено», менее 18 - оценка «не зачтено».

Составитель: к.т.н., доцент Лапшин Д. Д.

Зав. кафедрой: д.т.н., профессор Лапшина М. Л.

Рабочая программа рассмотрена на заседании кафедры математики, информационных систем и технологий и утверждена на 2021/2022 учебный год.
Протокол № 10 от 22 июня 2021 г.

Лист актуализации фонда оценочных средств

«Основы Data Mining»

шифр по учебному плану, наименование

для подготовки бакалавров

Направление: (шифр – название) 09.03.02 Информационные системы и технологии

Профиль: Информационные системы на транспорте

Форма обучения очная, очно-заочная

Год начала подготовки: 2021

Курс 4

Семестр 7, 8

а) в фонд оценочных средств не вносятся изменения. ФОС актуализирован на 2024 / 2025 г. учебный год.

б) в фонд оценочных средств вносятся следующие изменения:

- 1) _____;
- 2) _____;
- 3) _____.

Разработчик (и): к.т.н., доцент Лапшин Д. Д.

(ФИО, ученая степень, ученое звание)

Фонд оценочных средств пересмотрен и одобрен на заседании кафедры математики, информационных систем и технологий протокол № 12 от «28» июня 2024 г.

Заведующий кафедрой: Черняева С. Н., к. ф.-м. н., доцент /

(ФИО, ученая степень, ученое звание)


(подпись)