Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 24.10.2023 11:54:08
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Декан факультета информационных технологий

/Д. Г. Демидов/

«28» 04 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Технологии обработки больших данных»

Направление подготовки
09.03.01 Информатика и вычислительная техника
Образовательная программа (профиль подготовки)
«Программное обеспечение информационных систем»

Квалификация (степень) выпускника
бакалавр

Форма обучения
заочная

Москва 2022

Программа дисциплины «Технологии обработки больших данных» составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению **09.03.01 Информатика и вычислительная техника** и профилю подготовки **«Программное обеспечение информационных систем».**

Программу составил

_____/А.Е. Рабинвич /

Программа дисциплины утверждена на заседании кафедры «Прикладная информатика»
«____» августа 2022 г. протокол № _____

Заведующий кафедрой
доцент, к.э.н. /С. В. Суворов/

Программа согласована с руководителем образовательной программы по направлению подготовки **09.03.01 Информатика и вычислительная техника** по профилю подготовки **«Программное обеспечение информационных систем».**

_____ /С. В. Суворов/
«____» августа 2022 г.

Программа утверждена на заседании учебно-методической комиссии факультета Информационных технологий

Председатель комиссии _____ / Д. Г. Демидов/

«____»_____2022 г. Протокол:

1. Цель освоения дисциплины

Целью освоения дисциплины «Технологии обработки больших данных» является ознакомление обучающихся с теоретическими аспектами анализа данных, методами, возможностью их применения, предоставление практических навыков по использованию инструментальных средств.

Задачами дисциплины является изучение процессов обработки и методов анализа данных, приобретение навыков работы с Data Mining в современных ИС.

2. Место дисциплины в структуре ООП

Дисциплина «Технологии обработки больших данных» является дисциплиной базовой части ОП подготовки обучающихся по направлению 09.03.01 Информатика и вычислительная техника.

Для изучения дисциплины «Технологии обработки больших данных» обучающимся необходимы знания по предыдущим (смежным) дисциплинам:

- Современные информационные системы
- Современные компьютерные технологии
- Современные проблемы прикладной математики и информатики

3. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы.

В результате освоения дисциплины у обучающихся формируются следующие компетенции и должны быть достигнуты следующие результаты обучения как этап формирования соответствующих компетенций:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать	Перечень планируемых результатов обучения по дисциплине
ПК-2	способностью понимать, совершенствовать и применять современный математический аппарат	**знать:** определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задач. **уметь:** использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных **владеть:** методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.

Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единиц, т.е. 108 академических часов (из них 92 часов – самостоятельная работа обучающихся).

Разделы дисциплины «Технологии обработки больших данных» изучаются на третьем курсе в пятом семестре: лекции – 4 часов, лабораторные работы – 12 часов, форма контроля – зачет.

Содержание разделов дисциплины

Тема 1. *Данные, многомерные данные, базы данных.*

Тема 2. *Основные понятия интеллектуального анализа данных.*

Тема 3. *Методы и стадии ИАД.*

Тема 4. *Задачи ИАД. Информация и знания. Классификация и кластеризация. Прогнозирование и визуализация.*

Тема 5. *Основы анализа данных.*

Тема 6. *Методы классификации и прогнозирования.*

Тема 7. *Методы кластерного анализа. Иерархические методы.*

Тема 8. *Способы визуального представления данных.*

Тема 9. *Комплексный подход к внедрению ИАД*

Тема 10. *Построение и использование модели.*

Тема 11. *Организационные и человеческие факторы в ИАД.*

Тема 12. *Рынок инструментов анализа данных*

5. Образовательные технологии

Методика преподавания дисциплины «Технологии обработки больших данных» и реализация компетентностного подхода в изложении и восприятии материала предусматривает использование следующих активных и интерактивных форм проведения групповых, индивидуальных, аудиторных занятий в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся:

– подготовка к выполнению лабораторных работ в лабораториях университета;

– проведение интерактивных занятий по процедуре подготовки к интернет-тестированию на сайтах: *i-exam.ru, fepo.ru;*

– использование интерактивных форм текущего контроля в форме аудиторного и внеаудиторного интернет-тестирования;

Удельный вес занятий, проводимых в интерактивных формах, определен главной целью образовательной программы, особенностью контингента обучающихся и содержанием дисциплины «Технологии обработки больших данных» и в целом по дисциплине составляет 50% аудиторных занятий.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы обучающихся

В процессе обучения используются следующие оценочные формы самостоятельной работы обучающихся, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

В третьем семестре

- подготовка к выполнению лабораторных работ и их защита.

Оценочные средства текущего контроля успеваемости включают контрольные вопросы и задания в форме бланкового и (или) компьютерного тестирования, для контроля освоения обучающимися разделов дисциплины.
Образцы тестовых заданий, контрольных вопросов и заданий для проведения текущего контроля, вопросов к зачету, приведены в приложении.

В четвертом семестре

- подготовка к выполнению лабораторных работ и их защита.

Оценочные средства текущего контроля успеваемости включают контрольные вопросы и задания в форме бланкового и (или) компьютерного тестирования, для контроля освоения обучающимися разделов дисциплины.
Образцы тестовых заданий, контрольных вопросов и заданий для проведения текущего контроля, вопросов к экзамену, приведены в приложении.

Образцы тестовых заданий, тем лабораторных работ, контрольных вопросов и заданий для проведения текущего контроля, экзаменационных вопросов, приведены в приложении.

6.1. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (модулю)

6.1.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.

В результате освоения дисциплины (модуля) формируются следующие компетенции:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать
ПК-2	способностью понимать, совершенствовать и применять современный математический аппарат

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплин (модулей), практик в соответствии с учебным планом и календарным графиком учебного процесса.

6.1.2. Описание показателей и критериев оценивания компетенций, формируемых по итогам освоения дисциплины (модуля), описание шкал оценивания

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине (модулю).

Показатель	Критерии оценивания			
	2	3	4	5
ПК-2 – способностью понимать, совершенствовать и применять современный математический аппарат				
Знать: определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задач	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие следующих знаний: определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задач	Обучающийся демонстрирует неполное соответствие следующих знаний: определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задачй, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний: определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задач, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний: определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задач

Показатель	Критерии оценивания			
	2	3	4	5
уметь: использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных	Обучающийся не умеет или в недостаточной степени умеет: использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных	Обучающийся демонстрирует неполное соответствие следующих умений: использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании умениями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих умений: использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных .Умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.	Обучающийся демонстрирует полное соответствие следующих умений: использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных. Свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.

Показатель	Критерии оценивания			
	2	3	4	5
владеть: методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.	Обучающийся не владеет или в недостаточной степени владеет методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.	Обучающийся владеет методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.	Обучающийся частично владеет методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.	Обучающийся в полном объеме владеет методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.

Шкалы оценивания результатов промежуточной аттестации и их описание:

Форма промежуточной аттестации: экзамен.

Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине «Интеллектуальные информационные системы» (указывается что именно – прошли промежуточный контроль, выполнили лабораторные работы, выступили с докладом и т.д.)

Шкала оценивания	***Описание***
Отлично	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.*
Хорошо	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует неполное, правильное соответствие знаний, умений, навыков приведенным в таблицах показателей, либо если при этом были допущены 2-3 несущественные ошибки.*
Удовлетворительно	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, в котором освещена основная, наиболее важная часть материала, но при этом допущена одна значительная ошибка или неточность.*
Неудовлетворительно	*Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.*

Фонды оценочных средств представлены в приложении 1 к рабочей программе.

Приложение 1 к
рабочей программе

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Направление подготовки: **09.03.01 Информатика и вычислительная техника**

Форма обучения: очная

Вид профессиональной деятельности: (В соответствии с ФГОС ВО)

Кафедра: Прикладная информатика

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине «Технологии

обработки больших данных» Состав: 1. Паспорт

фонда оценочных средств
2. Описание оценочных средств:

Вопросы

Тесты

Москва, 2022 год

Таблица 1

ПОКАЗАТЕЛЬ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

Технологии обработки больших данных					
ФГОС ВО **09.03.01 Информатика и вычислительная техника**					
В процессе освоения данной дисциплины обучающиеся формирует и демонстрирует следующие **общекультурные компетенции**:					
КОМПЕТЕНЦИИ		**Перечень компонентов**	**Технология формирования компетенций**	**Форма оценочного средства****	**Степени уровней освоения компетенций**
ИНДЕКС	**ФОРМУЛИРОВКА**				

ПК-2	способностью понимать, совершенствовать и применять современный математический аппарат	**знать:** определения и свойства интегралов Римана и Лебега, признаки сходимости функциональных рядов, свойства степенных рядов, преобразование Фурье и его свойства, основные типы обыкновенных уравнений, метод Эйлера, метод вариации произвольных постоянных, основные понятия теории устойчивости, формулы Крамера, жорданову форму матрицы, уравнения кривых второго порядка и их свойства, поверхности второго порядка, основные математические модели дискретного характера и методы их использования для решения типовых задач. **уметь:** использовать основные понятия и методы математического анализа, дифференциальных уравнений, дискретной математики, алгебры, геометрии и информатики при обработке и интерпретации собранных данных **владеть:** методами математического анализа и навыками их практического применения: навыками дифференцирования функций, методами решения линейных	самостоятельная работа, лабораторные работы	УО, Т, П	**Базовый уровень** - способен к абстрактному мышлению, анализу, синтезу при решении общих вопросов **Повышенный уровень** - способен к абстрактному мышлению, анализу, синтезу профессиональных задач, предлагать варианты совершенствования исследуемых процессов

		дифференциальных уравнений, методами решения систем линейных алгебраических уравнений.навыками формирования потребительской аудитории и осуществления взаимодействия с потребителями»			

**- Сокращения форм оценочных средств см. в приложении 2 к РП.

Приложение 2
к рабочей программе

Перечень оценочных средств по дисциплине _____

№ ОС	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
12	Устный опрос собеседование, (УО)	Средство контроля, организованное как специальная беседа педагогического работника с обучающимся на темы, связанные с изучаемой дисциплиной, и рассчитанное на выяснение объема знаний обучающегося по определенному разделу, теме, проблеме и т.п.	Вопросы по темам/разделам дисциплины
14	Тест (Т)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося.	Фонд тестовых заданий

Примерные тестовые задания

Основные понятия технологии проектирования информационных систем

1. Какие науки включены в Data Mining?

a. статистика, базы данных, искусственный интеллект;
b. информатика, базы данных, статистика;
c. искусственный интеллект, базы данных, базы знаний;
d. информатика, базы данных, хранилища данных.

2. Каким образом технология Data Mining используется в интернет?

a. для создания сайтов;
b. для организации поисковых систем;
c. для отображения web-страниц.

3. Какие задачи решаются в СППР?

a. ввод данных, преобразование данных, вывод данных;
b. ввод данных, модификация данных, передача данных;
c. ввод данных, хранение данных, анализ данных.

4. Какой класс задач анализа отвечает за выполнение запросов?

a. информационно-поисковый;
b. оперативно-аналитический;
c. интеллектуальный.

5. Какой класс задач анализа отвечает за обобщение данных?

a. информационно-поисковый;
b. оперативно-аналитический;
c. интеллектуальный.

6. Какой класс задач анализа отвечает за построение моделей?

a. информационно-поисковый;
b. оперативно-аналитический;
c. интеллектуальный.

7. Какой класс задач анализа отвечает за поиск закономерностей в данных?

a. информационно-поисковый;

b. оперативно-аналитический;
c. интеллектуальный.

8. Какая подсистема СППР отвечает за ввод данных?

a. OLTP;
b. хранилище данных;
c. SQL;
d. OLAP;
e. Data Mining.

9. Какая подсистема СППР отвечает за хранение данных?

a. OLTP;
b. хранилище данных;
c. SQL;
d. OLAP;
e. Data Mining.

10. 1 Какая подсистема СППР отвечает за информационно-поисковый анализ данных?

a. OLTP;
b. хранилище данных;
c. SQL;
d. OLAP;
e. Data Mining.

11. Какая подсистема СППР отвечает за оперативный анализ данных?

a. OLTP;
b. хранилище данных;
c. SQL;
d. OLAP;
e. Data Mining.

12. Какая подсистема СППР отвечает за Технологии обработки больших данных?

a. OLTP;
b. хранилище данных;
c. SQL;
d. OLAP;
e. Data Mining.

13. Как реализуется подсистема ввода данных?

a. с помощью технологии Data Mining;
b. с помощью базы данных;
c. с помощью СУБД;
d. с помощью хранилища данных;
e. с помощью витрины данных.

14. Какие данные могут храниться в системе анализа?

a. детализированные;
b. обобщенные;
c. детализированные и обобщенные.

15. Могут ли в системе анализа храниться данные в разных форматах?

a. могут;
b. не могут;
c. могут только данные в текстовых форматах.

16. Какие данные могут храниться в системе анализа?

a. нормализованные;
b. избыточные (денормализованные);
c. частично нормализованные.

17. Как называются структуры данных, предназначенные для ввода, модификации и поиска?

a. оперативные источники данных;
b. хранилища данных;
c. базы данных;
d. файлы данных.

18. Как называются структуры данных, предназначенные для анализа?

a. оперативные источники данных;
b. хранилища данных;
c. базы данных;
d. витрины данных.

19. Предметно-ориентированный, интегрированный, неизменчивый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений – это

a. оперативный источник данных;
b. хранилище данных;
c. база данных;
d. файл данных.

20. Что является главным недостатком виртуального хранилища данных?

a. большое время обработки запросов;
b. значительные ресурсы компьютера;
c. разные форматы и кодировки данных в разных ОИД;
d. практическая невозможность получения данных за длительный период времени.

21. Как называются данные, непосредственно переносимые из ОИД?

a. метаданные;
b. агрегированные данные;
c. детальные.

22. Какие данные отражают сущность события?

a. измерения;
b. метаданные;
c. факты.

23. Какие данные описывают события?

a. измерения;
b. метаданные;
c. факты.

24. Как называются числовые фактические данные, которые могут быть просуммированы по всем измерениям?

a. аддитивные;
b. полуаддитивные;
c. неаддитивные.

25. Как называются числовые фактические данные, которые могут быть просуммированы только по определенным измерениям?

a. аддитивные;
b. полуаддитивные;
c. неаддитивные.

26. Как называются фактические данные, которые не могут быть просуммированы ни по одному измерению?

a. аддитивные;
b. полуаддитивные;
c. неаддитивные.

27. На какие вопросы должны отвечать метаданные?

a. что, кто, где, как, когда, почему;
b. что, кто, зачем, как, когда, почему;
c. что, кто, где, как, по какой причине, почему;
d. что, кто, где, как, когда, сколько.

28. Репозиторий – это

a. словарь терминов;
b. хранилище метаданных;
c. каталог с файлами.

29. Если поток образуется данными, копируемыми из ОИД, то он называется

a. входным потоком;
b. потоком обобщения;
c. архивным потоком;
d. потоком метаданных;
e. обратным потоком.

30. Если поток образуется агрегированными данными, то он называется

a. входным потоком;
b. потоком обобщения;
c. архивным потоком;
d. потоком метаданных;
e. обратным потоком.

31. Если поток образуется детальными данными, количество обращений к которым снизи-лось, то он называется

a. входным потоком;
b. потоком обобщения;
c. архивным потоком;
d. потоком метаданных;
e. обратным потоком.

32. Если поток образуется данными, переносимыми в репозиторий, то он называется

a. входным потоком;
b. потоком обобщения;
c. архивным потоком;
d. потоком метаданных;
e. обратным потоком.

33. Если поток образуется очищенными данными, записываемыми в ОИД, то он называется

a. входным потоком;
b. потоком обобщения;
c. архивным потоком;
d. потоком метаданных;
e. обратным потоком.

34. Многомерный анализ – это:

a. одновременный анализ по нескольким измерениям;
b. одновременный анализ по нескольким параметрам;
c. одновременный анализ по нескольким данным.

35. Формирование подмножества многомерного массива данных, соответствующего един-ственному значению одного или нескольких элементов измерений, не входящих в это подмножество, называется

a. гиперкубом;
b. срезом гиперкуба;
c. базой данных;
d. витриной данных.

36. Технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений, называется

a. OLTP;
b. OLAP;
c. Data Mining.

37. Если для реализации многомерной модели используют многомерные базы данных, то способ реализации гиперкуба называется

a. MOLAP;
b. ROLAP;
c. HOLAP;
d. DOLAP;
e. JOLAP.

38. Если для реализации многомерной модели используют реляционные базы данных, то способ реализации гиперкуба называется

a. MOLAP;
b. ROLAP;
c. HOLAP;
d. DOLAP;
e. JOLAP.

39. Если для реализации многомерной модели используют и многомерные, и реляционные базы данных, то способ реализации гиперкуба называется

a. MOLAP;
b. ROLAP;
c. HOLAP;
d. DOLAP;
e. JOLAP.

40. Настольная OLAP, предназначенная для локального анализа и представления данных, называется

a. MOLAP;
b. ROLAP;
c. HOLAP;
d. DOLAP;
e. JOLAP.

41. OLAP, предназначенная для создания и управления данными и метаданными, называется

a. MOLAP;
b. ROLAP;

c. HOLAP;
d. DOLAP;
e. JOLAP.

42. В каком отношении находятся таблица фактов и таблица измерений?

a. «один-к-одному»;
b. «один-ко-многим»;
c. «многие-ко-многим».

43. Исследование и обнаружение машиной (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые: ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком, называется

a. OLTP;
b. хранилищем данных;
c. OLAP;
d. Data Mining.

44. Какие операции над данными включены в ETL-процесс?

a. ввод, модификация, вывод;
b. чтение, изменение, запись;
c. извлечение, преобразование, загрузка;
d. получение, хранение, анализ.

45. Гиперкуб – это …

a. объект, все измерения которого имеют одинаковую размерность;
b. поликуб;
c. объект, все измерения которого имеют разную размерность;
d. многомерный куб;
e. многомерная база данных.

46. Многомерный просмотр данных основан на …

a. многомерной базе данных;
b. технологии мультимедиа;
c. многослойной базе;
d. сетевой технологии;
e. гипертекстовой технологии.

47. Какую технологию используют большинство хранилищ данных?

a. концептуальную БД;
b. реляционную БД;
c. иерархическую БД;
d. физическую БД.

48. Какое конструирование у хранилищ данных?

a. физическое;
b. логическое;
c. логическое и физическое;
d. иерархическое.

49. Концептуальную модель хранилища данных можно представить в виде…

a. таблицы;
b. графического рисунка;
c. геометрических фигур;
d. схемы.

50. К классу описательных задач относятся:

a. кластеризация и классификация;
b. кластеризация и поиск ассоциативных правил;
c. классификация и регрессия;
d. классификация и поиск ассоциативных правил.

51. К классу предсказательных задач относятся:

a. кластеризация и классификация;
b. кластеризация и поиск ассоциативных правил;
c. классификация и регрессия;
d. классификация и поиск ассоциативных правил.

52. К классу задач supervised learning (обучение с учителем) относятся:

a. кластеризация и классификация;
b. кластеризация и поиск ассоциативных правил;
c. классификация и регрессия;
d. классификация и поиск ассоциативных правил.

53. К классу задач unsupervised learning (обучение без учителя) относятся:

a. кластеризация и классификация;
b. кластеризация и поиск ассоциативных правил;
c. классификация и регрессия;
d. классификация и поиск ассоциативных правил.

54. Задача классификации сводится к …

a. нахождению частых зависимостей между объектами или событиями;
b. определению класса объекта по его характеристикам;
c. определению по известным характеристикам объекта значения некоторого его параметра;
d. поиску независимых групп и их характеристик во всем множестве анализируемых данных.

55. Задача регрессии сводится к …

a. нахождению частых зависимостей между объектами или событиями;
b. определению класса объекта по его характеристикам;
c. определению по известным характеристикам объекта значения некоторого его параметра;
d. поиску независимых групп и их характеристик во всем множестве анализируемых данных.

56. Задача кластеризации заключается в …

a. нахождении частых зависимостей между объектами или событиями;
b. определении класса объекта по его характеристикам;
c. определении по известным характеристикам объекта значения некоторого его параметра;
d. поиске независимых групп и их характеристик во всем множестве анализируемых данных.

57. Целью поиска ассоциативных правил является …

a. нахождение частых зависимостей между объектами или событиями;
b. определение класса объекта по его характеристикам;
c. определение по известным характеристикам объекта значения некоторого его параметра;
d. поиск независимых групп и их характеристик во всем множестве анализируемых данных.

58. К описательным моделям относятся следующие модели данных:

a. модели классификации и последовательностей;
b. регрессивные, кластеризации, исключений, итоговые и ассоциации;
c. модели классификации, кластеризации, исключений, итоговые и ассоциации;
d. модели классификации, последовательностей и исключений.

59. Модели классификации описывают …

a. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
b. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
c. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
d. группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

60. Модели последовательностей описывают …

a. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
b. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
c. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
d. группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

61. Регрессивные модели описывают …

a. правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
b. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
c. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
d. группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

62. Какова основная цель создания и использования хранилищ данных:

a. анализ данных для принятия управленческих решений;
b. надежное хранение, накопленных данных;
c. резервное копирование данных.

63. OLAP - это:

a. технология онлайновой быстрой аналитической обработки сложных запросов к базе данных;
b. технология онлайновой обработки небольших по размерам, но идущих большим потоком транзакций, требующих максимально быстрого ответа.

64. Какие из вариантов ответов являются характерными требованиями к хранению данных для принятия решений в хранилищах данных?

a. данные ориентированы на приложения;
b. данные управляются транзакциями;
c. данные обобщены либо очищены.

65. Назовите основной этап работы с хранилищами данных?

a. этап очистки данных;
b. этап обновления;

c. этап нормализации.

66. Что называют кубом OLAP?

a. структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице измерений;
b. структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице фактов;
c. таблицу размерностей.

67. Информационные хранилища созданы для удобства …

a. руководителей всех уровней для принятия решений;
b. стратегического планирования;
c. реорганизации бизнеса;
d. предметных приложений;
e. редактирования данных.

68. Информационные хранилища размещаются на …

a. библиотеках-автоматах;
b. сетевых серверах;
c. мейнфреймах;
d. серверах и кластерах серверов;
e. файл-серверах.

69. При слиянии данных в информационное хранилище из внутренних и внешних источников обеспечивается …

a. предметная ориентация данных;
b. выбор требуемых сведений из предметных приложений по наименованиям;
c. гипертекстовый просмотр данных;
d. согласование данных по наименованию;
e. хранение данных по предметным областям.

70. Интеллектуальный выбор данных из информационного хранилища – это …

a. реализация методов искусственного интеллекта;
b. выбор по заданному алгоритму;
c. реализация самообучающихся систем;
d. реализация экономико-статистических методов.

71. В процессе погружения в информационное хранилище данные …

a. очищаются от ненужной для анализа информации;
b. агрегируются;
c. преобразуются из разных типов данных предметных приложений в единую структуру хранения;
d. индексируются;
e. синхронизируются.

72. Классификация — …

a. некоторый набор операций над базой данных, который рассматривается как единственное завершенное, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных;
b. разновидность систем хранения, ориентированная на поддержку процесса анализа данных, обеспечивающая непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов;
c. высокоуровневые средства отражения информационной модели и описания структуры данных;

d. это установление зависимости дискретной выходной переменной от входных переменных.

73. Регрессия — …

a. это установление зависимости непрерывной выходной переменной от входных переменных;
b. это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
c. выявление закономерностей между связанными событиями;
d. это установление зависимости дискретной выходной переменной от входных переменных.

74. Кластеризация — …

a. это установление зависимости непрерывной выходной переменной от входных переменных;
b. это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
c. выявление закономерностей между связанными событиями;
d. это установление зависимости дискретной выходной переменной от входных переменных.

75. Ассоциация — …

a. это установление зависимости непрерывной выходной переменной от входных переменных;
b. это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
c. выявление закономерностей между связанными событиями;
d. это установление зависимости дискретной выходной переменной от входных переменных.

76. Машинное обучение — …

a. специализированное программное решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных;
b. это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
c. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
d. подразделение искусственного интеллекта, изучающее методы построения алгоритмов, способных обучаться на данных.

77. Аналитическая платформа — …

a. специализированное программное решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных;
b. это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
c. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
d. подразделение искусственного интеллекта, изучающее методы построения алгоритмов, способных обучаться на данных.

78. Обучающая выборка — …

a. это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
b. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр и соответствующий ему правильный выходной результат;
c. выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

79. Ошибка обучения — …

a. это ошибка, допущенная моделью на учебном множестве;
b. это ошибка, полученная на тестовых примерах, то есть вычисляется по тем же формулам, но для тестового множества;
c. имена, типы, метки и назначения полей исходной выборки данных;
d. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат.

80. Ошибка обобщения — …

a. это ошибка, допущенная моделью на учебном множестве;
b. это ошибка, полученная на тестовых примерах, то есть вычисляется по тем же формулам, но для тестового множества;
c. имена, типы, метки и назначения полей исходной выборки данных;
d. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат.

Вопросы для подготовки к экзамену

1. Организация хранилищ данных.
2. Очистка данных.
3. Задачи Data Mining.
4. Модели Data Mining
5. Методы построения правил классификации
6. Методы построения деревьев решений
7. Поиск ассоциативных правил
8. Иерархические алгоритмы кластеризации
9. Неиерархические алгоритмы кластеризации
10. Адаптивные методы кластеризации
11. Область применения хранилищ данных
12. Основные требования к данным, вводимым в хранилище данных
13. Возможности создания хранилищ данных в реляционных БД
14. Понятие о витринах данных
15. Понятие многомерного представления при описании структур данных
16. Алгоритм добычи данных - метод деревьев решений
17. Алгоритм добычи данных – кластеризация
18. Технология аналитической обработки данных
19. Варианты реализации OLAP
20. Распределенные корпоративные хранилища данных
21. Зависимые и независимые киоски данных

22. Характеристика виртуальных кубов
23. Архитектура хранилища HOLAP
24. Архитектура хранилища ROLAP
25. Архитектура хранилища MOLAP
26. Особенности ETL-процесса
27. Методы и этапы процесса очистки данных
28. Задачи Data Mining.
29. Классификация задач DataMining
30. Задача классификации и регрессии
31. Задача поиска ассоциативных правил
32. Задача кластеризации
33. Модели DataMining
34. Методы DataMining
35. Возможности современных программ, используемых для создания хранилищ данных Характеристика информационной технологии принятия решений
36. Роль хранилища данных в информационной технологии принятия решений
37. Задача анализа данных. Технологии обработки больших данных. Отличия от других видов анализа данных.
38. Постановка и порядок решение задачи интеллектуального анализа данных на примере информационной модели торговой компании.
39. Закономерности в данных: неочевидные, объективные, полезные.
40. Области применения интеллектуального анализа данных.
41. Подготовка данных для анализа. Обработка пропущенных и недостающих данных. Анализ исключений.
42. Основные концепции баз данных. Модель данных. СУБД.
43. Использование баз данных в интеллектуальном анализе данных.
44. Постановка и порядок решения задачи интеллектуального анализа данных на примере хранилища слабоструктурированных текстовых документов.
45. Введение в многомерные базы данных. Особенности базы данных для оперативной аналитической обработки данных. Порядок и особенности построения хранилищ данных.
46. Виды и особенности шкал измерений данных.
47. Введение в OLAP. Типовая модель данных для OLAP. Особенности приложений для оперативной аналитической обработки данных.
48. Постановка и порядок решения задачи интеллектуального анализа данных на примере документной базы данных.
49. Особенности использования данных при интеллектуальном анализе данных. Данные, информация и знания.
50. Основные задачи интеллектуального анализа данных.
51. Стадии интеллектуального анализа данных. Начальные этапы.
52. Вычисление и анализ показателей описательной статистики.
53. Характеристики временных рядов, используемые в ходе интеллектуального анализа данных.
54. Построение и использование моделей для интеллектуального анализа данных.
55. Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.
56. Преимущественные особенности методов анализа данных: статистических, машинного обучения, оперативного анализа, интеллектуального анализа.

57. Решение задач кластеризация в ходе интеллектуального анализа данных: цель и основные методы.
58. Сущность и методы решения задачи прогнозирования при интеллектуальном анализе данных.
59. Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных в СППР.
60. Назначение, основные этапы развития и виды систем ИАД.
61. Основные формы представления и способы объяснения (интерпретации) результатов ИАД.
62. Построение, применение и интерпретации карт категориальных данных.
63. Применение методов и средств картографии для анализа данных и интерпретации его результатов.
64. Применение методов классификации и кластеризации для сегментации данных.
65. Методы анализа разнотипных экспериментальных данных.
66. Принципы и общая характеристика методов кластерного анализа.
67. Деревья решений, принципы их построения, практического применения и интерпретации результатов (характеристика деревьев).
68. Влияние размерности и объемов исходных данных на выбор и применение средств ИАД.
69. Способы визуализации информации в системах ИАД.
70. Системы для визуализации многомерных данных.
71. Графические способы представления результатов ИАД.
72. Табличные способы представления результатов ИАД.
73. Компоненты и модели временных рядов (тренда, сезонности, цикла, случайной компоненты).
74. Простая линейная и многомерная регрессия.
75. Нелинейная регрессия.
76. Бинарная и множественная логистическая регрессия.
77. Применение генетических алгоритмов для анализа данных.
78. Измерительные шкалы, представление переменных, ввод и редактирование данных.
79. Объединение и агрегирование (файлов) данных в информационно-аналитических системах с применением средств ИАД.
80. Трансформируемые таблицы и интерактивные диаграммы (на примере «сводных таблиц» MS Office).
81. Подготовка отчетов и экспорт результатов анализа из систем ИАД.
82. Классические методы многомерного статистического анализа.
83. Роль и место методов классического статистического анализа для решения основных задач ИАД.
84. Задачи выявления и восстановление зависимостей в ИАД.
85. Задачи и методы классификации в ИАД.
86. Методы решения задач прогнозирования
87. Построение прогнозов на основе различных моделей.
88. Анализ и прогнозирование временных рядов: цели, задачи, методы (временной и частотный подходы к анализу временных рядов).
89. Технологии обработки больших данных: основные принципы организации, построения и применения.
90. Сравнительный анализ средств (систем) анализа и прогнозирования на основе деревьев решений (на примере не менее трех систем).

Вопросы для подготовки к зачету:

1. 12 правил Кодда оценки средств для OLAP
2. Построение кубов данных. SQL Server Analysis Services
3. Инструменты разработки и администрирования, клиентские инструменты
4. Построение кубов данных.
5. Использование Business Intelligence Development Studio для создания проекта
6. Построение кубов данных. Создание источника данных
7. Построение кубов данных. Создание представления источника данных
8. Построение кубов данных. Создание куба с помощью мастера
9. Построение кубов данных. Развертывание и просмотр куба
10. Построение реляционной витрины данных
11. Заполнение реляционной витрины данных
12. Настройка витрины данных
13. Построение OLAP-срезов
14. OLAP-средства программы MS Excel
15. Создание OLAP-куба в среде MS Excel
16. Средства анализа OLAP-куба
17. Создание OLAP-куба с помощью Deductor
18. Обработка OLAP-куба в Deductor. Создание отчетов
19. Применение систем ИАД для информационно-аналитического обеспечения принятия решений (на основе выбранного примера).
20. Применение кластерного анализа для сокращения количества переменных при моделировании (на выбранном примере).
21. Построение и интерпретации карт визуализации зависимостей (на выбранном примере).
22. Применение нейронных сетей к задачам регрессионного анализа (на выбранном примере).
23. Применение нейронных сетей к задачам прогнозирования временных рядов (на выбранном примере).
24. Использование деревьев решений в задачах классификации (на выбранном примере).
25. Способы декомпозиции временных рядов: выявления тренда, сезонной, циклической и случайных составляющих (на выбранном примере).
26. Возможности программного обеспечения и практическое применение инструментов классического статистического анализа данных для решения задач ИАД (на примере выбранной системы или пакета прикладных программ).
27. Использование моделей временных рядов для анализа данных и прогнозирования (на выбранном примере).
28. Решение задач классификации нейронными сетями (на выбранном примере).
29. Создание отчетов и экспорт результатов из систем ИАД в различные приложения (на примере выбранной системы)

7. Учебно-методическое и информационное обеспечение дисциплины.

а) основная литература:

1. Чубукова И. А. Data Mining. Интернет-Университет Информационных Технологий 2008. - 383 с. http://www.knigafund.ru/books/177992
2. Нестеров С. А. Интеллектуальный анализ средствами MS SQL Server 2008 / Национальный Открытый Университет «ИНТУИТ», 2016. - 338 с. http://www.knigafund.ru/books/177094

б) дополнительная литература:

1. Броневич А.Г., Каркищенко А.Н., Лепский А.Е. Анализ неопределенности выделения информативных признаков и представлений изображений: монография. Физматлит 2013. - 309 с. http://www.knigafund.ru/books/207497
2. Семенов А., Соловьев Н., Чернопрудова Е., Цыганков А. Интеллектуальные системы: учебное пособие. ОГУ 2013. - 236 с. http://www.knigafund.ru/books/181693

в) программное обеспечение и интернет-ресурсы:

Пакет деловой графики Microsoft Office Visio 2010 (2013, 2016).

Табличный процессор Microsoft Office Excel 2010 (2013, 2016).

Интернет-ресурсы включают учебно-методические материалы в электронном виде, представленные на сайте XXXXXXXX.ru в разделе «Библиотека» (*http://lib.mami.ru/ebooks/*).

Варианты контрольных заданий по дисциплине представлены на сайтах: *http://i-exam.ru*, *http:// fepo.ru.*

8. Материально-техническое обеспечение дисциплины.

- Лекционные аудитории с компьютерным и видеопроекционным оборудованием для презентаций с выходом в Интернет, средствами звуковоспроизведения
- Компьютерные классы с соответствующим программным обеспечением и видеопроекционным оборудованием для презентаций с выходом в Интернет, средствами звуковоспроизведения ауд. ав4805, ав4809, ав4810, ав4811. Программное обеспечение Редактор деловой графики CASE-средство Microsoft Visio 2010 (2007, 2013, 2016) и Табличный процессор Microsoft Office Excel 2010 (2013, 2016).

9. Методические рекомендации для самостоятельной работы обучающихся

10. Методические рекомендации для преподавателя

Программа составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки Бакалавров 09.03.01 Информатика и вычислительная техника

Приложение 3.

Структура и содержание дисциплины «Технологии обработки больших данных» по направлению подготовки 09.03.01 Информатика и вычислительная техника (Бакалавр)

Раздел	Семестр	Неделя													
			Л	П/С	Лаб	СРС	КСР	К.Р.	К.П.	РГР	Реферат	К/р	Э	З
Тема 1. *Данные, многомерные данные, базы данных.*	9	1	1		1	6								
Тема 2. *Основные понятия интеллектуального анализа данных.*	9	2	1		1	6								
Тема 2. *Основные понятия интеллектуального анализа данных.*	9	3			1	5								
Тема 3. *Методы и стадии ИАД.*	9	4			1	5								
Тема 3. *Методы и стадии ИАД.*	9	5				5								
Тема 4. *Задачи ИАД. Информация и знания. Классификация и кластеризация. Прогнозирование и визуализация.*	9	6	1		1	5								
Тема 4. *Задачи ИАД. Информация и знания. Классификация и кластеризация. Прогнозирование и визуализация.*	9	7				5								
Тема 5. *Основы анализа данных.*	9	8			1	5								
Тема 6. *Методы классификации и прогнозирования.*	9	9				5								
Тема 6. *Методы классификации и прогнозирования.*	9	10				5								
Тема 7. *Методы кластерного анализа. Иерархические методы.*	9	11	1		1	5								
Тема 7. *Методы кластерного анализа. Иерархические методы.*	9	12				5								
Тема 8. *Способы визуального представления данных.*	9	13			1	5								
Тема 9. *Комплексный подход к внедрению ИАД*	9	14			1	5								
Тема 9. *Комплексный подход к внедрению ИАД*	9	15				5								
Тема 10. *Построение и использование модели.*	9	16			1	5								
Тема 11. *Организационные и человеческие факторы в ИАД.*	9	17			1	5								
Тема 12. *Рынок инструментов анализа данных*	9	18			1	5								
Форма аттестации														+
Всего часов по дисциплине			4		12	92								