

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Луганский государственный университет имени Владимира Даля»

Экономический факультет
Кафедра экономической кибернетики и прикладной статистики

УТВЕРЖДАЮ:
Декан экономического факультета
Тхор Е.С.
(подпись)
« 24 » 2023 года



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ»

По направлению подготовки 38.04.05 Бизнес-информатика
Программа магистратуры: «Бизнес-аналитика»

Луганск - 2023

Лист согласования РПУД

Рабочая программа учебной дисциплины «Технологии анализа данных и машинное обучение» по направлению подготовки 38.04.05 Бизнес-информатика. – 47 с.

Рабочая программа учебной дисциплины «Технологии анализа данных и машинное обучение» составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.04.05 Бизнес-информатика, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 12 августа 2020 года № 990.

СОСТАВИТЕЛЬ (СОСТАВИТЕЛИ):

к.т.н., доцент Велигура А.В.

Рабочая программа дисциплины утверждена на заседании кафедры экономической кибернетики и прикладной статистики «18» 04 20 23 г., протокол № 26

Заведующий кафедрой экономической кибернетики
и прикладной статистики А.В. Велигура

Переутверждена: « » _____ 20 г., протокол № _____

Согласована (для обеспечивающей кафедры)
Декан экономического факультета Тхор Е.С.

Переутверждена: « » _____ 20 года, протокол № _____

Рекомендована на заседании учебно-методической комиссии экономического факультета «21» апреле 20 23 г., протокол № 4.

Председатель учебно-методической
комиссии экономического факультета Е.Н. Шаповалова

Структура и содержание дисциплины

1. Цели и задачи дисциплины, ее место в учебном процессе

Целью дисциплины «Технологии анализа данных и машинное обучение» является формирование у студентов комплекса теоретических знаний и методологических основ в области интеллектуального анализа данных, а также практических навыков, необходимых для внедрения и практического использования интеллектуальных алгоритмов анализа и обработки данных.

Задачи: изучение существующих технологий подготовки данных к анализу; изучение основных методов поиска в данных внутренних закономерностей, взаимосвязей, тенденций; овладение практическими умениями и навыками реализации технологий аналитической обработки данных, формирования и проверки гипотез о их природе и структуре, варьирования применяемыми моделями; формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина «Технологии анализа данных и машинное обучение» относится к обязательной части цикла дисциплин.

Содержание дисциплины является логическим продолжением содержания дисциплин «Методология и методы научных исследований в отрасли», «Информационные технологии в управлении проектами», «Бизнес-анализ», «Цифровые платформы и экосистемы современного бизнеса».

Содержание дисциплины служит основой для преддипломной практики и выпускной квалификационной работы.

3. Требования к результатам освоения содержания дисциплины

Код и наименование компетенции	Индикаторы достижений компетенции реализуемой дисциплине) (по	Перечень планируемых результатов
ОПК-3. Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.3. Выявляет возможности, создаваемые информационными и цифровыми технологиями	Знать: Методы и алгоритмы интеллектуального анализа данных
		Уметь: Применять методы и алгоритмы интеллектуального анализа данных для поиска, выработки и применения новых решений в профессиональной деятельности

		Владеть: инструментальными средствами реализации методов и алгоритмов интеллектуального анализа данных
ОПК-5. Способен проводить исследования, организовывать самостоятельную и коллективную научно-исследовательскую, проектную и учебно-профессиональную деятельность для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий	ОПК-5.3. Формирует научные отчеты, публикации, аналитические отчеты, презентации по результатам выполненной деятельности	Знать: Методы визуализации результатов исследований
		Уметь: Разрабатывать научные отчеты, публикации, аналитические отчеты, презентации по результатам выполненной деятельности
		Владеть: Инструментами создания аналитических отчетов, презентаций по результатам выполненной деятельности

4. Структура и содержание дисциплины

4.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Объем часов (зач. ед.)	
	Очная форма	Заочная форма
Общая учебная нагрузка (всего)	144 (4 зач. ед)	144 (4 зач. ед)
Обязательная аудиторная учебная нагрузка (всего) в том числе:	56	28
Лекции	28	14
Семинарские занятия	-	-
Практические занятия	28	14
Лабораторные работы	-	-
Курсовая работа (курсовой проект)	-	-
Другие формы и методы организации образовательного процесса (расчетно-графические работы, групповые дискуссии, ролевые игры, тренинг, компьютерные симуляции, интерактивные лекции, семинары, анализ деловых ситуаций и т.п.)		-
Самостоятельная работа студента (всего)	88	116
Форма аттестации	экзамен	экзамен

4.2. Содержание разделов дисциплины

Тема 1. ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ

Введение. Основные понятия. Пирамида знаний. Постановка задачи анализа данных. История развития технологий анализа данных. Методика анализа данных. Методы сбора и подготовки исходного набора данных.

Моделирование неопределённости. Вероятностный подход. Условные Байесовские сети.

Тема 2. ВВЕДЕНИЕ В ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Предмет и задачи интеллектуального анализа данных. Примеры практических приложений. KDD (Knowledge discovery in databases). Data Mining. Задачи классификации, кластеризации и регрессии. Оценка качества алгоритмов Data Mining. Метод скользящего контроля (кросс-валидация).

Тема 3. ВИЗУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Визуальный анализ данных. Применение различных средств визуализации.

Тема 4. АФФИНТИВНЫЙ АНАЛИЗ. ПОИСК ПОСЛЕДОВАТЕЛЬНЫХ ШАБЛОНОВ

Введение в аффинитивный анализ (affinity analysis). Алгоритм a priori. Иерархические ассоциативные правила.

Тема 5. КЛАСТЕРНЫЙ АНАЛИЗ

Введение в кластеризацию. Классификация методов кластеризации. Алгоритм кластеризации k-means. Сети Кохонена (KCN – Kohonen network). Карты Кохонена (SOM – self organizing map). Проблемы алгоритмов кластеризации.

Тема 6. КЛАССИФИКАЦИЯ И РЕГРЕССИЯ. СТАТИСТИЧЕСКИЕ МЕТОДЫ

Введение в классификацию и регрессию. Простая линейная регрессия. Оценка соответствия простой линейной регрессии реальным данным. Простая регрессионная модель. Множественная линейная регрессия. Модель множественной линейной регрессии. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Основы логистической регрессии. Интерпретация модели логистической регрессии.

Тема 7. КЛАССИФИКАЦИЯ И РЕГРЕССИЯ. МАШИННОЕ ОБУЧЕНИЕ

Введение в деревья решений. Алгоритмы построения деревьев решений. Алгоритмы ID3 и C4.5. Алгоритм CART. Упрощение деревьев решений. Введение в нейронные сети. Искусственный нейрон. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей. Алгоритм обратного распространения ошибки.

Тема 8. НЕЙРОСЕТЕВЫЕ АЛГОРИТМЫ КЛАССИФИКАЦИИ

Понятия искусственных нейронных сетей. Перцептрон Розенблатта, правило обучения. Многослойные сети прямого распространения. Алгоритм

обратного распространения ошибки. Машины опорных векторов. Вероятностные нейронные сети. Дилемма смещения и дисперсии.

Тема 9. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

Временной ряд и его компоненты. Модели прогнозирования. Прогнозирование в торговле и логистике.

Тема 10. АНСАМБЛИ МОДЕЛЕЙ

Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей.

Тема 11. ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ПЕРСПЕКТИВНЫХ НАПРАВЛЕНИЯХ РАЗВИТИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИИ

Обзор наиболее успешных продуктов, библиотек, toolkit, framework в выбранном направлении из области анализа данных. Вопросы: история развития, степень развития, назначение, функции, конкретные примеры использования, самостоятельное применение, лицензии, стандарты, применяемые средства анализа данных, предоставляемые API, совместимые технологии и т.д.

4.3. Лекции

№ п/п	Название темы	Объем часов	
		Очная форма	Заочная форма
1	Введение в анализ данных	2	1
2	Введение в интеллектуальный анализ данных.	2	1
3	Визуальный анализ данных	2	1
4	Аффинитивный анализ. Поиск последовательных шаблонов	2	1
5	Кластерный анализ	2	1
6	Классификация и регрессия. Статистические методы	2	1
7	Классификация и регрессия. Машинное обучение	4	2
8	Нейросетевые алгоритмы классификации	4	2
9	Анализ и прогнозирование временных рядов	4	2
10	Ансамбли моделей	2	1
11	Использование методов интеллектуального анализа данных в перспективных направлениях развития информационных технологий	2	1
Итого:		28	14

4.4. Практические (семинарские) занятия

Учебным планом не предусмотрены

4.4. Лабораторные работы

№ п/п	Название темы	Объем часов	
		Очная форма	Заочная форма
1	Введение в анализ данных	2	1
2	Введение в интеллектуальный анализ данных.	2	1
3	Визуальный анализ данных	2	1
4	Аффинитивный анализ. Поиск последовательных шаблонов	2	1
5	Кластерный анализ	2	1
6	Классификация и регрессия. Статистические методы	2	1
7	Классификация и регрессия. Машинное обучение	4	2
8	Нейросетевые алгоритмы классификации	4	2
9	Анализ и прогнозирование временных рядов	4	2
10	Ансамбли моделей	2	1
11	Использование методов интеллектуального анализа данных в перспективных направлениях развития информационных технологии	2	1
Итого:		28	14

4.6. Самостоятельная работа студентов

№ п/п	Название темы	Вид СРС	Объем часов	
			Очная форма	Заочная форма
1	Введение в анализ данных	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	7
2	Введение в интеллектуальный анализ данных.	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	7
3	Визуальный анализ данных	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	7
4	Аффинитивный анализ. Поиск последовательных шаблонов	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	7
5	Кластерный анализ	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	7
6	Классификация и регрессия. Статистические методы	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	8
7	Классификация и регрессия. Машинное обучение	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	8

8	Нейросетевые алгоритмы классификации	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	5	8
9	Анализ и прогнозирование временных рядов	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	4	7
10	Ансамбли моделей	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	4	7
11	Использование методов интеллектуального анализа данных в перспективных направлениях развития информационных технологии	Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний и умений.	4	7
12	Экзамен	Подготовка к зачету	36	36
Итого:			88	116

4.7. Курсовые работы/проекты.

Курсовая работа рабочим учебным планом не предусмотрена.

5. Образовательные технологии

Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

Информационные технологии: использование электронных образовательных ресурсов (электронный конспект, размещенный на сайте кафедры cpsdahl.ru) при подготовке к лекциям и лабораторным работам.

Мультимедийные технологии: презентации к лекциям.

6. Формы контроля освоения дисциплины

Текущая аттестация студентов производится в дискретные временные интервалы лектором и преподавателем(ями), ведущими практические занятия по дисциплине в следующих формах:

собеседование (устный или письменный опрос);

контрольная работа;

тесты.

Фонды оценочных средств, включающие типовые задания, контрольные работы, тесты и методы контроля, позволяющие оценить результаты текущей и промежуточной аттестации обучающихся по данной дисциплине, помещаются в приложении к рабочей программе в соответствии с «Положением о фонде оценочных средств».

Форма аттестации по результатам освоения дисциплины проходит в форме письменного/устного экзамена (включает в себя ответ на теоретические

вопросы и решение задач) либо в сочетании различных форм (компьютерного тестирования, решения задач и пр.). Студенты, выполнившие 75% текущих и контрольных мероприятий на «отлично», а остальные 25 % на «хорошо», имеют право на получение итоговой отличной оценки.

В экзаменационную ведомость и зачетную книжку выставляются оценки по шкале, приведенной в таблице.

Шкала оценивания	Характеристика знания предмета и ответов
отлично (5)	Студент глубоко и в полном объеме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
хорошо (4)	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
удовлетворительно (3)	Студент знает только основной программный материал, допускает неточности, недостаточно четкие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.
неудовлетворительно (2)	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.

7. Учебно-методическое и информационное обеспечение дисциплины:

а) основная литература:

1. Чубукова И.А., Data Mining / Чубукова И.А. - М.: Национальный Открытый Университет "ИНТУИТ", 2016. (Основы информационных технологий) - ISBN 978-5-94774-819-2 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785947748192.html>.

2. Мельниченко А.С., Математическая статистика и анализ данных: учеб. пособие / А.С. Мельниченко - М.: МИСиС, 2018. - 45 с. - ISBN 978-5-906953-62-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785906953629.html>.

3. Кожаринов А.С., Моделирование и анализ информационных и бизнес-процессов в информационных системах: метод. указ. к выполнению курсовых работ / А.С. Кожаринов. - М.: МИСиС, 2017. - 27 с. - Текст:

электронный // ЭБС "Консультант студента": [сайт]. - URL: http://www.studentlibrary.ru/book/Misis_362.html.

б) дополнительная литература:

1. Кук Д., Машинное обучение с использованием библиотеки H2O / Кук Д. - М.: ДМК Пресс, 2018. - 250 с. - ISBN 978-5-97060-508-0 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785970605080.html>.

2. Джеймс Г., Введение в статистическое обучение с примерами на языке R / Джеймс Г., Уиттон Д., Хасты Т., Тибширани Р. - М.: ДМК Пресс, 2017. - 456 с. - ISBN 978-5-97060-495-3 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785970604953.html>.

3. Рашка С., Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - М.: ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785970604090.html>.

4. Афонин П.Н., Статистический анализ с применением современных программных средств: учебное пособие / Афонин П.Н., Афонин Д.Н. - СПб.: ИЦ Интермедия, 2017. - 100 с. - ISBN 978-4383-0080-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN978438300809.html>.

5. Тагиева Р.Ф., Обработка экспериментальных данных. Ч.1: учебное пособие: в 2 ч. / Р.Ф. Тагиева, А.Н. Титов - Казань: Издательство КНИТУ, 2017. - 96 с. - ISBN 978-5-7882-2261-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785788222615.html>.

6. Александровская Ю.П., Многомерный статистический анализ в экономике: учебное пособие / Александровская Ю. П. - Казань: Издательство КНИТУ, 2017. - 96 с. - ISBN 978-5-7882-2191-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785788221915.html>.

7. Шелудько В.М., Язык программирования высокого уровня Python. Функции, структуры данных, дополнительные модули: учебное пособие / Шелудько В. М. - Ростов н/Д: Изд-во ЮФУ, 2017. - 107 с. - ISBN 978-5-9275-2648-2 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785927526482.html>.

в) методические указания:

1. Методические указания к лабораторным работам по дисциплине «Технологии анализа данных и машинное обучение» для студентов направления подготовки 38.04.05 – Бизнес-информатика [Электронный ресурс] / сост. А.В. Велигура. – Луганск: ЛНУ им. В. Даля, 2022. – 75 с.

2. Методические указания к самостоятельной работе по дисциплине «Технологии анализа данных и машинное обучение» для студентов направления подготовки 38.04.05 – Бизнес-информатика [Электронный ресурс] / сост. А.В. Велигура. – Луганск: ЛНУ им. В. Даля, 2022. – 26 с.

3. Методические указания к практическим занятиям по дисциплине «Технологии анализа данных и машинное обучение». Занятие 1. Проектирование и наполнение хранилища данных [Электронный ресурс] / сост.: А.В. Велигура, Н.Н. Попова. – Луганск: ЛНУ им. В. Даля, 2023. – 18 с.

г) Интернет-ресурсы:

1. Министерство образования и науки Российской Федерации – <http://минобрнауки.рф/>

2. Федеральная служба по надзору в сфере образования и науки – <http://obrnadzor.gov.ru/>

3. Министерство образования и науки Луганской Народной Республики – <https://minobr.su>

4. Министерство промышленности и торговли Луганской Народной Республики – <https://www.minpromlnr.su/main.php/>

5. Министерство экономического развития Луганской Народной Республики – <https://merlnr.su/>

6. Министерство финансов Луганской Народной Республики – <https://minfinlnr.su/>

7. Народный совет Луганской Народной Республики – <https://nslnr.su>

8. Государственный комитет статистики Луганской Народной Республики – <https://www.gkslnr.su/>

9. Портал Федеральных государственных образовательных стандартов высшего образования – <http://fgosvo.ru/>

10. Федеральный портал «Российское образование» – <http://www.edu.ru/>

11. Информационная система «Единое окно доступа к образовательным ресурсам» – <http://window.edu.ru/>

12. Федеральный центр информационно-образовательных ресурсов – <http://fcior.edu.ru/>

13. Статистические сборники ВШЭ – <https://www.hse.ru/org/hse/primarydata/>

14. OpenOffice.org: Теория и практика – <https://www.altlinux.org/Books:Openoffice>

15. Базовый курс по OpenOffice - <https://4creates.com/training/49-bazovyy-kurs-po-openoffice.html>

Электронные библиотечные системы и ресурсы

16. Электронно-библиотечная система «Консультант студента» – <http://www.studentlibrary.ru/>

17. Электронно-библиотечная система «StudMed.ru» – <https://www.studmed.ru/>

Информационный ресурс библиотеки образовательной организации

18. Научная библиотека имени А. Н. Коняева – <http://biblio.dahluniver.ru/>

8. Материально-техническое и программное обеспечение дисциплины

Для проведения лекционных занятий требуется аудитория, оборудованная мультимедийным проектором с экраном. Для проведения лабораторных и практических занятий необходим компьютерный класс и презентационная техника (проектор, экран, компьютер/ноутбук)

Программное обеспечение:

Функциональное назначение	Бесплатное программное обеспечение	Ссылки
Офисный пакет	Libre Office 6.3.1	https://www.libreoffice.org/ https://ru.wikipedia.org/wiki/LibreOffice
Операционная система	UBUNTU 19.04	https://ubuntu.com/
Браузер	Firefox Mozilla	http://www.mozilla.org/ru/firefox/fx
Графический редактор	GIMP (GNU Image Manipulation Program)	http://www.gimp.org/ http://gimp.ru/viewpage.php?page_id=8 http://ru.wikipedia.org/wiki/GIMP
Редактор PDF	Adobe Acrobat Reader	https://get.adobe.com/ru/reader/
Аудиоплеер	VLC	http://www.videolan.org/vlc/

9. Оценочные средства по дисциплине

Паспорт оценочных средств по учебной дисциплине «Технологии анализа данных и машинное обучение»

Перечень компетенций (элементов компетенций), формируемых в результате освоения учебной дисциплины (модуля) или практики

№ п/п	Код контролируемой компетенции	Формулировка контролируемой компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Контролируемые темы учебной дисциплины, практики	Этапы формирования (семестр изучения)
1	ОПК-3	Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального	ОПК-3.3	Тема 1 Тема 2 Тема 3 Тема 4 Тема 5 Тема 6 Тема 7 Тема 8 Тема 9 Тема 10 Тема 11	3
2	ОПК-5.	Способен проводить исследования, организовывать самостоятельную и коллективную научно-исследовательскую, проектную и учебно-профессиональную деятельность для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий	ОПК-5.3	Тема 1 Тема 2 Тема 3 Тема 4 Тема 5 Тема 6 Тема 7 Тема 8 Тема 9 Тема 10 Тема 11	3

Показатели и критерии оценивания компетенций, описание шкал оценивания

№ п/п	Код контролируемой компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Перечень планируемых результатов	Контролируемые разделы (темы) учебной дисциплины (модуля), практики	Наименование оценочного средства ²
1	ОПК-3.	ОПК-3.3	Знать: Методы и алгоритмы интеллектуального анализа данных Уметь: Применять методы и алгоритмы интеллектуального анализа данных для поиска, выработки и применения новых решений в профессиональной деятельности Владеть: инструментальными средствами реализации методов и алгоритмов интеллектуального анализа данных	Тема 1 Тема 2 Тема 3 Тема 4 Тема 5 Тема 6 Тема 7 Тема 8 Тема 9 Тема 10 Тема 11	Собеседование (устный или письменный опрос), контрольная работа, тесты
2	ОПК-5	ОПК-5.3	Знать: Методы визуализации результатов исследований Уметь: Разрабатывать научные отчеты, публикации, аналитические отчеты, презентации по результатам выполненной деятельности Владеть: Инструментами создания аналитических отчетов, презентаций по результатам выполненной деятельности	Тема 1 Тема 2 Тема 3 Тема 4 Тема 5 Тема 6 Тема 7 Тема 8 Тема 9 Тема 10 Тема 11	Собеседование (устный или письменный опрос), контрольная работа, тесты

Оценочные средства по дисциплине «Технологии анализа данных и машинное обучение»

Перечень вопросов (для проведения собеседования (устный или письменный опрос))

1. Область применения хранилищ данных.
2. Основные требования к данным, вводимым в хранилище данных.
3. Возможности создания хранилищ данных в реляционных БД.
4. Понятие о витринах данных.
5. Понятие многомерного представления при описании структур данных.
6. Алгоритм добычи данных - метод деревьев решений.
7. Алгоритм добычи данных – кластеризация.
8. Технология аналитической обработки данных (OLAP).
9. Варианты реализации OLAP.
10. Распределенные корпоративные хранилища данных.
11. Зависимые и независимые киоски данных.
12. Характеристика виртуальных кубов.
13. Архитектура хранилища HOLAP.
14. Архитектура хранилища ROLAP.
15. Архитектура хранилища MOLAP.
16. Особенности ETL-процесса.
17. Методы и этапы процесса очистки данных.
18. Задачи Data Mining. Классификация задач DataMining.
19. Задача классификации и регрессии.
20. Задача поиска ассоциативных правил.
21. Задача кластеризации.
22. Модели DataMining.
23. Методы DataMining.
24. Возможности современных программ, используемых для создания хранилищ данных.
25. Характеристика информационной технологии принятия решений.
26. Роль хранилища данных в информационной технологии принятия решений.
27. 12 правил Кодда оценки средств для OLAP.
28. Построение кубов данных. SQL Server Analysis Services 2005. Инструменты разработки и администрирования, клиентские инструменты.
29. Построение кубов данных. Использование Business Intelligence Development Studio для создания проекта.
30. Построение кубов данных. Создание источника данных.
31. Построение кубов данных. Создание представления источника данных (DSV).
32. Построение кубов данных. Создание куба с помощью мастера.
33. Построение кубов данных. Развертывание и просмотр куба.

34. Построение реляционной витрины данных.
35. Заполнение реляционной витрины данных.
36. Настройка витрины данных.
37. Построение OLAP-срезов.
38. OLAP-средства программы MS Excel.
39. Создание OLAP-куба в среде MS Excel.
40. Средства анализа OLAP-куба.
41. Создание OLAP-куба с помощью Deductor.
42. Обработка OLAP-куба в Deductor. Создание отчетов.
43. Классификация ПО Data Mining.
44. Обзор и анализ ПО по бизнес-характеристикам и техническим характеристикам.
45. Подбор ПО для решения задач оценивания и прогнозирования.
46. Подбор ПО для решения задач классификации.
47. Подбор ПО для решения задач кластеризации и сегментации.
48. Подбор ПО для поиска и ассоциативных правил.
49. Из каких частей состоит Deductor?
50. Какие варианты поставки Deductor существуют?
51. Чем отличается версия Professional от Academic?
52. Имеются ли ограничения по количеству обрабатываемых записей в версии Deductor Academic?
53. Сколько категорий пользователей Deductor можно выделить?
54. Перечислите функции аналитика.
55. Кто обычно занимается проектированием и наполнением хранилища данных?
56. Каким образом лицензируется Deductor?
57. Что такое ассоциативные правила? Приведите пример.
58. В чём отличие поиска ассоциативных правил и секвенциального анализа?
59. Как можно модифицировать алгоритм Apriori?
60. Для чего используются ассоциативные правила при анализе данных?
61. Что такое достоверность правила?
62. Что такое поддержка правила?
63. Какой алгоритм генерации ассоциативных правил имеется в Deductor?
64. Какие входные поля набора данных необходимы для запуска обработчика Ассоциативные правила в Deductor?
65. Какие специализированные визуализаторы предлагаются к узлу обработчику Ассоциативные правила?
66. Какой визуализатор используется для формирования предложений клиенту?
67. С помощью, каких кнопок можно отсортировать правила по убыванию лифта?
68. Дайте определение понятия «кластер».
69. Дайте определение понятия «кластеризация».

70. В чём заключается задача кластеризации? Приведите пример.
71. Как представляются результаты при кластеризации?
72. На какие два больших класса можно поделить алгоритмы кластеризации.
73. Охарактеризуйте два любых алгоритма кластеризации.
74. Назовите входные данные алгоритма k-средних.
75. Назовите выходные данные алгоритма k - средних.
76. Приведите последовательность шагов в алгоритме k - средних.
77. Приведите порядок кластерного анализа с помощью ПО kmeans.

**Критерии и шкала оценивания по оценочному средству собеседование
(устный или письменный опрос)**

Шкала оценивания (интервал баллов)	Критерий оценивания
5	собеседование (устный или письменный опрос) прошел на высоком уровне (студент в полном объеме осветил рассматриваемый вопрос, владеет профильным понятийным (категориальным) аппаратом и т.п.)
4	собеседование (устный или письменный опрос) прошел на среднем уровне (студент в целом осветил рассматриваемый вопрос, владеет профильным понятийным (категориальным) аппаратом и т.п.)
3	собеседование (устный или письменный опрос) на низком уровне (студент допустил существенные неточности, изложил материал с ошибками, не владеет в достаточной степени профильным категориальным аппаратом и т.п.)
2	собеседование (устный или письменный опрос) прошел на неудовлетворительном уровне или не представлен (студент не готов, не выполнил задание и т.п.)

Контрольная работа

Задание 1.

Сгенерируйте вектор длины $N=1000$, элементами которого являются реализации нормально распределенной случайной величины с математическим ожиданием, равным 1, и стандартным отклонением, равным 0.3.

Подсчитайте статистическое мат. ожидание и стандартную ошибку, не используя встроенные функции и проверьте правильность результата.

Подсчитайте .95, .99-квантили. Исследуйте отклонение статистического мат. ожидания от 1 при росте N ($N=1000, 2000, 4000, 8000$).

Задание 2.

Создайте фрейм данных из $N=20$ записей со следующими полями: `Nrow` — номер записи, `Name` — имя пользователя, `BirthYear` — год рождения, `EmployYear` — год приема на работу, `Salary` — зарплата, где `Nrow` изменяется от 1 до N , `Name` задается произвольно, `BirthYear` распределен равномерно на отрезке $[1960, 1985]$, `EmployYear` распределен равномерно на отрезке $[\text{BirthYear}+18, 2006]$, `Salary` для работников младше 1975 г.р. определяется по формуле $\text{Salary} = (\ln(2007 - \text{EmployYear}) + 1) * 8000$, для остальных $\text{Salary} = (\log_2(2007 - \text{EmployYear}) + 1) * 8000$.

Подсчитайте число сотрудников с зарплатой, большей 15000. Добавьте в таблицу поле, соответствующее суммарному подоходному налогу (ставка 13%), выплаченному сотрудником за время работы в организации, если его зарплата за каждый код начислялась согласно формулам для `Salary`, где вместо 2007 следует последовательно подставить каждый год работы сотрудника в организации.

Задание 3.

Напишите функцию, которая принимает на вход числовой вектор x и число разбиений интервала k (по умолчанию равно числу элементов вектора, разделенному на 10) и выполняет следующее: находит минимальное и максимальное значение элементов вектора x_{\min} и x_{\max} , разделяет полученный отрезок $[x_{\min}; x_{\max}]$ на k равных интервалов и подсчитывает число элементов вектора, принадлежащих каждому интервалу. Далее должен строиться график, где по оси абсцисс — середины интервалов, по оси ординат — число элементов вектора, принадлежащих интервалу, разделенное на общее число точек. Проведите эксперимент на данной функции, где x — вектор длины 5000, сгенерированный из экспоненциально распределенной случайной величины, $k=500$. Приближение какого графика мы получаем в итоге при большом числе точек и числе разбиений?

Задание 4.

Спроектируйте и реализуйте метод наименьших квадратов.

Задание 5.

Используя тест Шапиро-Уилка, проверьте, являются ли нормально распределенными характеристики цветов ириса (фрейм данных `iris`). Уровень значимости $\alpha = 0.05$.

Задание 6.

Для $k=10, 15, 20, 25, 30$ сгенерируйте 200 реализаций нормальной распределенной случайной величины с мат. ожиданием, равным k , и стандартным отклонением, равным $\sqrt{2k}$, и 200 реализаций случайной величины, распределенной по закону χ^2 с k степенями свободы. Используя тест Колмогорова-Смирнова, проверьте гипотезу о том, что данные выборки относятся к одному непрерывному распределению. Уровень значимости $\alpha = 0.05$.

Задание 7.

Загрузите таблицу из файла `allcountries.txt`, содержащую информацию о населении, площади и ряде других характеристик современных государств. Выберите из таблицы те страны, для которых доступна информация о

населении и площади (нет отсутствующих значений NA) и площадь больше 10.0. Пусть $\text{area_log} = \log_{10}(\log_{10}(\text{area}))$, $\text{population_log} = \log_{10}(\log_{10}(\text{population}))$. Постройте линейную регрессию (используя функцию `lm`) для зависимости population_log от area_log . Используя тест Колмогорова-Смирнова, проверьте гипотезу о том, что population_log и $f(\text{area_log})$, где $f()$ — построенная регрессионная функция, относятся к одному непрерывному распределению.

Уровень значимости $\alpha = 0.05$.

Задание 8.

Используя критерий χ^2 проверьте нуль-гипотезу, состоящую в том, что цвет глаз женщин не зависит от цвета волос (на фрейме данных `HairEyeColor`).

Задание 9.

Загрузите таблицу из файла `readingspeed.txt`, которая содержит информацию о скорости чтения у детей в зависимости от применяемой методики обучения (DRA — direct reading activities, SC — standart curriculum).

Используя t-тест, проверьте гипотезу о том, что среднее время чтения для обеих методик совпадает (используйте разные альтернативные гипотезы). Объясните полученные результаты.

Задание 10.

Загрузите данные из файла `reglab1.txt`. Используя функцию `lm`, постройте регрессию (используйте разные модели). Выберите наиболее подходящую модель, объясните свой выбор.

Задание 11.

Реализуйте следующий алгоритм для уменьшения количества признаков, используемых для построения регрессии: для каждого $k \in \{0, 1, \dots, p\}$ выбрать подмножество признаков мощности k , минимизирующее $\text{RSS}(\beta)$.

Используя полученный алгоритм, выберите оптимальное подмножество признаков для данных из файла `reglab2.txt`. Объясните свой выбор.

Дайте интерпретацию вычисленным значениям t-статистики и p-value для коэффициентов $\hat{\beta}_i$.

Задание 12.

Загрузите данные из файла `sygage.txt`. Постройте регрессию, выражающую зависимость возраста исследуемых отложений от глубины залегания, используя веса наблюдений. Оцените качество построенной модели.

Задание 13.

Загрузите данные из файла `alligators.txt`. Выберите лучшую регрессионную модель (возможно нелинейную), отражающую зависимость веса аллигатора от его длины.

Задание 13.

Загрузите библиотеку MASS и набор данных `longley`. Исключите из набора данных переменную `Population`. Разделите данные на тестовую и обучающую выборки равных размеров случайным образом. Постройте ridge regression для значений $\text{lambda} = 10^{\text{seq}(-3, 2, \text{by}=0.2)}$, подсчитайте ошибку на

тестовой и обучающей выборке для данных значений λ , постройте графики. Объясните полученные результаты

Задание 14.

Загрузите набор данных `monica` из пакета `DAAG` (http://www.stats.uwo.ca/DAAG/DAAG_0.95.zip).

Постройте дерево классификации для модели, задаваемой следующей формулой: $\text{outcome} \sim .$, дайте интерпретацию полученным результатам. Является ли построенное дерево избыточным? Если да, то выполните операцию «snip off» над соответствующими узлами.

Задание 15.

Загрузите набор данных `sram7` из пакета `DAAG`. Постройте дерево классификации для модели, задаваемой следующей формулой: $\text{yesno} \sim .$, дайте интерпретацию полученным результатам. Запустите процедуру «costcomplexity pruning» с выбором параметра k по умолчанию, `method = 'misclass'`, выведите полученную последовательность деревьев. Какое из полученных деревьев, на Ваш взгляд, является оптимальным? Объясните свой выбор.

Задание 16.

Загрузите набор данных `nsw74psid1` из пакета `DAAG`. Постройте регрессионное дерево для модели, задаваемой следующей формулой: $\text{re78} \sim .$

Постройте регрессионную модель и SVM-регрессию для данной формулы. Сравните качество построенных моделей, выберите оптимальную и объясните свой выбор.

Задание 17.

Для $N=10, 100, 1000$ сгенерируйте вектор длины N , состоящий из реализаций нормально распределенной случайной величины $N(0,1)$. Вычислите стандартную ошибку для статистического мат. ожидания, используя `bootstrap` с числом повторений $B=200$, исследуйте ее зависимость от N , сравните с теоретической оценкой.

Задание 18.

Загрузите набор данных `mouse.t`, вычислите статистическое мат. ожидание и медиану и оцените стандартную ошибку, используя `bootstrap` ($B=50, 100, 250, 500, 1000$).

Задание 19.

Загрузите набор данных `law` и вычислите корреляцию между средним баллом, набранным учениками одной школы на тесте по правоведению (LSAT), и их средним баллом по всем предметам (GPA). Оцените стандартную ошибку коэффициента корреляции, используя `bootstrap`. Вычислите коэффициент корреляции на наборе данных `law82` и сравните полученные результаты.

Задание 20.

Загрузите набор данных `spatial`. Вычислите 90%-доверительные интервалы для статистической дисперсии для A , используя 1) `bootstrap-t`, 2) $BC\alpha$, 3) ABC , 4) `bootstrap percentile`, и сравните полученные интервалы.

Задание 21.

Разбейте множество объектов из набора данных `pluton` на 3 кластера методом центров тяжести (`kmeans`). Сравните качество разбиения в зависимости от максимального числа итераций алгоритма.

Задание 22.

Сгенерируйте набор данных в двумерном пространстве, состоящий из 3 кластеров, каждый из которых сильно «вытянут» вдоль одной из осей. Исследуйте качество кластеризации методом `clara` в зависимости от 1) использования стандартизации; 2) типа метрики. Объясните полученные результаты.

Задание 23.

Постройте дендрограмму для набора данных `votes.repub` (число голосов, поданных за республиканцев на выборах с 1856 по 1976 год). Проинтерпретируйте полученный результат.

Критерии и шкала оценивания по оценочному средству контрольная работа

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Контрольная работа выполнена на высоком уровне (правильные ответы даны на 90-100% вопросов/задач)
4	Контрольная работа выполнена на среднем уровне (правильные ответы даны на 75-89% вопросов/задач)
3	Контрольная работа выполнена на низком уровне (правильные ответы даны на 50-74% вопросов/задач)
2	Контрольная работа выполнена на неудовлетворительном уровне (правильные ответы даны менее чем на 50%)

Тесты

- Какие науки включены в Data Mining?
 - статистика, базы данных, искусственный интеллект;
 - информатика, базы данных, статистика;
 - искусственный интеллект, базы данных, базы знаний;
 - информатика, базы данных, хранилища данных.
- Каким образом технология Data Mining используется в интернет?
 - для создания сайтов;
 - для организации поисковых систем;
 - для отображения web-страниц.
- Какие задачи решаются в СППР?
 - ввод данных, преобразование данных, вывод данных;
 - ввод данных, модификация данных, передача данных;
 - 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. 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 интеллектуальный анализ 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?
- метаданные;
 - агрегированные данные;
 - детальные.
22. Какие данные отражают сущность события?
- измерения;
 - метаданные;
 - факты.
23. Какие данные описывают события?
- измерения;
 - метаданные;
 - факты.
24. Как называются числовые фактические данные, которые могут быть просуммированы по всем измерениям?
- аддитивные;
 - полуаддитивные;
 - неаддитивные.
25. Как называются числовые фактические данные, которые могут быть просуммированы только по определенным измерениям?
- аддитивные;
 - полуаддитивные;
 - неаддитивные.
26. Как называются фактические данные, которые не могут быть просуммированы ни по одному измерению?
- аддитивные;
 - полуаддитивные;
 - неаддитивные.
27. На какие вопросы должны отвечать метаданные?
- что, кто, где, как, когда, почему;
 - что, кто, зачем, как, когда, почему;
 - что, кто, где, как, по какой причине, почему;
 - что, кто, где, как, когда, сколько.
28. Репозиторий – это
- словарь терминов;
 - хранилище метадаанных;
 - каталог с файлами.
29. Если поток образуется данными, копируемыми из ОИД, то он называется
- входным потоком;
 - потоком обобщения;
 - архивным потоком;
 - потоком метадаанных;
 - обратным потоком.

30. Если поток образуется агрегированными данными, то он называется
- а. входным потоком;
 - б. потоком обобщения;
 - в. архивным потоком;
 - г. потоком метаданных;
 - д. обратным потоком.
31. Если поток образуется детальными данными, количество обращений к которым снизилось, то он называется
- а. входным потоком;
 - б. потоком обобщения;
 - в. архивным потоком;
 - г. потоком метаданных;
 - д. обратным потоком.
32. Если поток образуется данными, переносимыми в репозиторий, то он называется
- а. входным потоком;
 - б. потоком обобщения;
 - в. архивным потоком;
 - г. потоком метаданных;
 - д. обратным потоком.
34. Если поток образуется очищенными данными, записываемыми в ОИД, то он называется
- а. входным потоком;
 - б. потоком обобщения;
 - в. архивным потоком;
 - г. потоком метаданных;
 - д. обратным потоком.
35. Многомерный анализ – это:
- а. одновременный анализ по нескольким измерениям;
 - б. одновременный анализ по нескольким параметрам;
 - в. одновременный анализ по нескольким данным
36. Формирование подмножества многомерного массива данных, соответствующего единственному значению одного или нескольких элементов измерений, не входящих в это подмножество, называется
- а. гиперкубом;
 - б. срезом гиперкуба;
 - в. базой данных;
 - г. витриной данных.
37. Технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений, называется
- а. OLTP;
 - б. OLAP;
 - в. Data Mining.

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

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

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

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

40. 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.

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

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

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

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

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

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

44. 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.
45. Какие операции над данными включены в ETL-процесс?
- a. ввод, модификация, вывод;
 - b. чтение, изменение, запись;
 - c. извлечение, преобразование, загрузка;
 - d. получение, хранение, анализ.
46. 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.
- 47 Многомерный просмотр данных основан на ...
- a. многомерной базе данных;
 - b. технологии мультимедиа;
 - c. многослойной базе;
 - d. сетевой технологии;
 - e. гипертекстовой технологии.
48. Какую технологию используют большинство хранилища данных?
- a. концептуальную БД;
 - b. реляционную БД;
 - c. иерархическую БД;
 - d. физическую БД.
49. Какое конструирование у хранилищ данных?
- a. физическое;
 - b. логическое;
 - c. логическое и физическое;
 - d. иерархическое.
50. Концептуальную модель хранилища данных можно представить в виде...
- a. таблицы;
 - b. графического рисунка;
 - c. геометрических фигур;
 - d. схемы.
51. 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.
52. 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.
53. 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.
54. 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.
55. Задача классификации сводится к ...
- a. нахождения частых зависимостей между объектами или событиями;
 - b. определения класса объекта по его характеристиками;
 - c. определение по известным характеристиками объекта значение некоторого его параметра;
 - d. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
56. Задача регрессии сводится к ...
- a. нахождения частых зависимостей между объектами или событиями;
 - b. определению класса объекта по его характеристиками;
 - c. определению по известным характеристиками объекта значения некоторого его параметра;
 - d. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
57. Задача кластеризации заключается в ...
- a. нахождения частых зависимостей между объектами или событиями;
 - b. определения класса объекта по его характеристиками;
 - c. определение по известным характеристиками объекта значение некоторого его параметра;
 - d. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
8. Целью поиска ассоциативных правил является ...
- a. нахождения частых зависимостей между объектами или событиями;
 - b. определения класса объекта по его характеристиками;
 - c. определение по известным характеристиками объекта значение некоторого его параметра;
 - d. поиска независимых групп и их характеристик в всем множестве анализируемых данных.

59. В описательных моделях относятся следующие модели данных:
- a. модели классификации и последовательностей;
 - b. регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - c. классификации, кластеризации, исключений, итоговые и ассоциации;
 - d. модели классификации, последовательностей и исключений.
60. 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.
61. 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 анализу.
62. 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 анализу.
63. Какова основная цель создания и использования хранилищ данных:
- a. анализ данных для принятия управленческих решений;
 - b. надежное хранение, накопленных данных;
 - c. резервное копирование данных.
64. OLAP - это:
- a. технология онлайн-быстрой аналитической обработки сложных запросов к базе данных;

в. технология онлайн-обработки небольших по размерам, но идущих большим потоком транзакций, требующих максимально быстрого ответа.

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

- а. данные ориентированы на приложения;
- б. данные управляются транзакциями;
- с. данные обобщены либо очищены.

66. Перечислите основные этапы работы с хранилищами данных?

- а. этап очистки данных;
- б. этап обновления;
- с. этап нормализации.

67. Что называют кубом OLAP?

- а. структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице измерений;
- б. структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице фактов;
- с. таблицу размерностей.

68. Информационные хранилища созданы для удобства ...

- а. руководителей всех уровней для принятия решений;
- б. стратегического планирования;
- с. реорганизации бизнеса;
- д. предметных приложений;
- е. редактирования данных.

69. Информационные хранилища размещаются на ...

- а. библиотеках-автоматах;
- б. сетевых серверах;
- с. мейнфреймах;
- д. серверах и кластерах серверов;
- е. файл-серверах.

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

- а. предметная ориентация данных;
- б. выбор требуемых сведений из предметных приложений по наименованиям;
- с. гипертекстовый просмотр данных;
- д. согласование данных по наименованию;
- е. хранение данных по предметным областям.

71. Интеллектуальный выбор данных из информационного хранилища – это ...

- а. реализация методов искусственного интеллекта;
- б. выбор по заданному алгоритму;
- с. реализация самообучающихся систем;

- d. реализация экономико-статистических методов.
72. 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.
73. Какова основная цель создания и использования хранилищ данных:
- a. анализ данных для принятия управленческих решений;
 - b. надежное хранение накопленных данных;
 - c. резервное копирование данных.
74. Классификация — ...
- a. некоторый набор операций над базой данных, который рассматривается как единственное завершённое, 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.
75. 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.
76. 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.
77. 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.

78. 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.

79. 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.

80. 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. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
- d. выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

81. 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.

82. 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.

Критерии и шкала оценивания по оценочному средству тесты

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Тесты выполнены на высоком уровне (правильные ответы даны на 90-100% тестов)
4	Тесты выполнены на среднем уровне (правильные ответы даны на 75-89% тестов)
3	Тесты выполнены на низком уровне (правильные ответы даны на 50-74% тестов)
2	Тесты выполнены на неудовлетворительном уровне (правильные ответы даны менее чем на 50% тестов)

Оценочные средства для промежуточной аттестации (экзамен)

Вопросы к экзамену

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. Какой класс задач анализа отвечает за обобщение данных?
 - a. информационно-поисковый;
 - b. оперативно-аналитический;
 - c. интеллектуальный.
3. Какая подсистема СППР отвечает за хранение данных?
 - a. OLTP;
 - b. хранилище данных;

- c. SQL;
 - d. OLAP;
 - e. Data Mining.
4. 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.
5. Как называются структуры данных, предназначенные для ввода, модификации и поиска?
- a. оперативные источники данных;
 - b. хранилища данных;
 - c. базы данных;
 - d. файлы данных.
6. Как называются данные, непосредственно переносимые из ОИД?
- a. метаданные;
 - b. агрегированные данные;
 - c. детальные.
7. Как называются числовые фактические данные, которые могут быть просуммированы только по определенным измерениям?
- d. аддитивные;
 - e. полуаддитивные;
 - f. неаддитивные.
8. 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.
9. Многомерный анализ – это:
- a. одновременный анализ по нескольким измерениям;
 - b. одновременный анализ по нескольким параметрам;
 - c. одновременный анализ по нескольким данным.
10. Если для реализации многомерной модели используют реляционные базы данных, то способ реализации гиперкуба называется
- a. MOLAP;
 - b. ROLAP;
 - c. HOLAP;
 - d. DOLAP;
 - e. JOLAP.

11. В каком отношении находятся таблица фактов и таблица измерений?
- «один-к-одному»;
 - «один-ко-многим»;
 - «многие-ко-многим».
12. Многомерный просмотр данных основан на ...
- многомерной базе данных;
 - технологии мультимедиа;
 - многослойной базе;
 - сетевой технологии;
 - гипертекстовой технологии.
13. К классу описательных задач относятся:
- кластеризация и классификация;
 - кластеризация и поиск ассоциативных правил;
 - классификация и регрессия;
 - классификация и поиск ассоциативных правил.
14. Задача классификации сводится к ...
- нахождения частых зависимостей между объектами или событиями;
 - определения класса объекта по его характеристиками;
 - определение по известным характеристиками объекта значение некоторого его параметра;
 - поиска независимых групп и их характеристик в всем множестве анализируемых данных.
15. В описательных моделях относятся следующие модели данных:
- модели классификации и последовательностей;
 - регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - классификации, кластеризации, исключений, итоговые и ассоциации;
 - модели классификации, последовательностей и исключений.
16. Какова основная цель создания и использования хранилищ данных:
- анализ данных для принятия управленческих решений;
 - надежное хранение, накопленных данных;
 - резервное копирование данных.
17. Что называют кубом OLAP?
- структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице измерений;
 - структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице фактов;
 - таблицу размерностей.

18. 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.

19. 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.

20. 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.

21. Каким образом технология Data Mining используется в интернет?

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

22. Какой класс задач анализа отвечает за построение моделей?

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

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

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

24. Какие данные могут храниться в системе анализа?

- a. детализированные;

- b. обобщенные;
 - c. детализированные и обобщенные.
25. Как называются структуры данных, предназначенные для анализа?
- a. оперативные источники данных;
 - b. хранилища данных;
 - c. базы данных;
 - d. витрины данных.
26. Какие данные отражают сущность события?
- a. измерения;
 - b. метаданные;
 - c. факты.
27. Как называются фактические данные, которые не могут быть просуммированы ни по одному измерению?
- a. аддитивные;
 - b. полуаддитивные;
 - c. неаддитивные.
28. Если поток образуется агрегированными данными, то он называется
- a. входным потоком;
 - b. потоком обобщения;
 - c. архивным потоком;
 - d. потоком метаданных;
 - e. обратным потоком.
29. Формирование подмножества многомерного массива данных, соответствующего единственному значению одного или нескольких элементов измерений, не входящих в это подмножество, называется
- a. гиперкубом;
 - b. срезом гиперкуба;
 - c. базой данных;
 - d. витриной данных.
30. 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.
31. 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.
32. Какую технологию используют большинство хранилища данных?
- a. концептуальную БД;
 - b. реляционную БД;
 - c. иерархическую БД;
 - d. физическую БД.
33. 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.
34. Задача регрессии сводится к ...
- a. нахождения частых зависимостей между объектами или событиями;
 - b. определения класса объекта по его характеристиками;
 - c. определение по известным характеристиками объекта значение некоторого его параметра;
 - d. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
35. 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 анализа.
36. OLAP - это:
- a. технология онлайн-быстрой аналитической обработки сложных запросов к базе данных;
 - b. технология онлайн-обработки небольших по размерам, но идущих большим потоком транзакций, требующих максимально быстрого ответа.
37. Информационные хранилища созданы для удобства ...
- a. руководителей всех уровней для принятия решений;
 - b. стратегического планирования;
 - c. реорганизации бизнеса;
 - d. предметных приложений;

- e. редактирования данных.
38. 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.
39. 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.
40. 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. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
 - d. выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.
41. 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.
42. Какой класс задач анализа отвечает за поиск закономерностей в данных?
- a. информационно-поисковый;
 - b. оперативно-аналитический;
 - c. интеллектуальный.
43. Какая подсистема СППР отвечает за оперативный анализ данных?
- a. OLTP;
 - b. хранилище данных;
 - c. SQL;

- d. OLAP;
 - e. Data Mining.
44. Могут ли в системе анализа храниться данные в разных форматах?
- a. могут;
 - b. не могут;
 - c. могут только данные в текстовых форматах.
45. 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.
46. Какие данные описывают события?
- a. измерения;
 - b. метаданные;
 - c. факты.
47. 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.
48. Если поток образуется детальными данными, количество обращений к которым снизилось, то он называется
- a. входным потоком;
 - b. потоком обобщения;
 - c. архивным потоком;
 - d. потоком метаданных;
 - e. обратным потоком.
49. 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.
50. Настольная OLAP, предназначенная для локального анализа и представления данных, называется
- a. MOLAP;
 - b. ROLAP;
 - c. HOLAP;
 - d. DOLAP;
 - e. JOLAP.

51. Какие операции над данными включены в ETL-процесс?
- ввод, модификация, вывод;
 - чтение, изменение, запись;
 - извлечение, преобразование, загрузка;
 - получение, хранение, анализ.
52. Какое конструирование у хранилищ данных?
- физическое;
 - логическое;
 - логическое и физическое;
 - иерархическое.
53. К классу задач supervised learning (обучение с учителем) относятся:
- кластеризация и классификация;
 - кластеризация и поиск ассоциативных правил;
 - классификация и регрессия;
 - классификация и поиск ассоциативных правил.
54. Задача кластеризации заключается в ...
- нахождения частых зависимостей между объектами или событиями;
 - определения класса объекта по его характеристиками;
 - определение по известным характеристиками объекта значение некоторого его параметра;
 - поиска независимых групп и их характеристик в всем множестве анализируемых данных.
55. Модели последовательностей описывают ...
- правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - группы, на которые можно разделить объекты, данные о которых подвергаются анализу.
56. Какие из вариантов ответов являются характерными требованиями к хранению данных для принятия решений в хранилищах данных?
- данные ориентированы на приложения;
 - данные управляются транзакциями;
 - данные обобщены либо очищены.
57. Информационные хранилища размещаются на ...
- библиотеках-автоматах;
 - сетевых серверах;
 - мейнфреймах;

- d. серверах и кластерах серверов;
 - e. файл-серверах.
58. Какова основная цель создания и использования хранилищ данных:
- a. анализ данных для принятия управленческих решений;
 - b. надежное хранение накопленных данных;
 - c. резервное копирование данных.
59. 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.
60. 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.
61. Какой класс задач анализа отвечает за выполнение запросов?
- a. информационно-поисковый;
 - b. оперативно-аналитический;
 - c. интеллектуальный.
62. Какая подсистема СППР отвечает за ввод данных?
- a. OLTP;
 - b. хранилище данных;
 - c. SQL;
 - d. OLAP;
 - e. Data Mining.
63. Какая подсистема СППР отвечает за интеллектуальный анализ данных?
- a. OLTP;
 - b. хранилище данных;
 - c. SQL;
 - d. OLAP;
 - e. Data Mining.
64. Какие данные могут храниться в системе анализа?
- a. нормализованные;

- b. избыточные (денормализованные);
 - c. частично нормализованные.
65. 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.
66. Как называются числовые фактические данные, которые могут быть просуммированы по всем измерениям?
- a. аддитивные;
 - b. полуаддитивные;
 - c. неаддитивные.
67. Репозиторий – это
- a. словарь терминов;
 - b. хранилище метаданных;
 - c. каталог с файлами.
68. 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.
69. Если для реализации многомерной модели используют многомерные базы данных, то способ реализации гиперкуба называется
- a. MOLAP;
 - b. ROLAP;
 - c. HOLAP;
 - d. DOLAP;
 - e. JOLAP.
70. 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.
71. 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.

72. Концептуальную модель хранилища данных можно представить в виде...

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

73. 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.

74. Целью поиска ассоциативных правил является ...

- a. нахождения частых зависимостей между объектами или событиями;
- b. определения класса объекта по его характеристиками;
- c. определение по известным характеристиками объекта значение некоторого его параметра;
- d. поиска независимых групп и их характеристик в всем множестве анализируемых данных.

75. 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 анализу.

76. Перечислите основные этапы работы с хранилищами данных?

- a. этап очистки данных;
- b. этап обновления;
- c. этап нормализации.

77. 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.

78. Классификация — ...

a. некоторый набор операций над базой данных, который рассматривается как единственное завершённое, 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.

79. 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.

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.

Критерии и шкала оценивания итоговой аттестации

Шкала оценивания	Характеристика знания предмета и ответов
отлично (5)	Студент глубоко и в полном объёме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
хорошо (4)	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
удовлетворительно (3)	Студент знает только основной программный материал, допускает неточности, недостаточно чёткие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.
неудовлетворительно (2)	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.

Лист изменений и дополнений

№ п/п	Виды дополнений и изменений	Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения	Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами)