

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Костина Лариса Николаевна
Должность: проректор
Дата подписания: 29.07.2024 16:03:30
Уникальный программный ключ:
1800f7d89cf4ea7507265ba593fe87537eb15a6c

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
"ДОНЕЦКАЯ АКАДЕМИЯ УПРАВЛЕНИЯ И ГОСУДАРСТВЕННОЙ СЛУЖБЫ"

Факультет

Факультет государственной службы и управления

Кафедра

Информационных технологий

"УТВЕРЖДАЮ"

Проректор

_____ Л.Н. Костина

27.04.2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.11

"Интеллектуальный анализ данных"

Направление подготовки 09.03.03 Прикладная информатика

Профиль "Прикладная информатика в управлении корпоративными информационными системами"

Квалификация

Бакалавр

Форма обучения

очная

Общая трудоемкость

4 ЗЕТ

Год начала подготовки по учебному плану

2024

Составитель(и):

канд. техн. наук, доцент

_____ И.В. Тарабаева

Рецензент(ы):

канд. физ.-мат. наук, доцент

_____ Н.В. Брадул

Рабочая программа дисциплины (модуля) "Интеллектуальный анализ данных" разработана в соответствии с:

Федеральным государственным образовательным стандартом высшего образования - бакалавриат по направлению подготовки 09.03.03 Прикладная информатика (приказ Минобрнауки России от 19.09.2017 г. № 922).

Рабочая программа дисциплины (модуля) составлена на основании учебного плана Направление подготовки 09.03.03 Прикладная информатика Профиль "Прикладная информатика в управлении корпоративными информационными системами", утвержденного Ученым советом ФГБОУ ВО "ДОНАУИГС" от 27.04.2024 протокол № 12.

Рабочая программа рассмотрена и одобрена на заседании кафедры Информационных технологий

Протокол от 16.04.2024 № 9

Заведующий кафедрой:

канд. физ.-мат. наук, доцент, Брадул Н.В.

(подпись)

Визирование РПД для исполнения в очередном учебном году**"УТВЕРЖДАЮ"**

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2025 - 2026 учебном году на заседании кафедры Информационных технологий

Протокол от "____" _____ 2025 г. №____

Зав. кафедрой канд.физ.-мат.наук, доцент, Брадул Н.В.

(подпись)

Визирование РПД для исполнения в очередном учебном году**"УТВЕРЖДАЮ"**

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2026 - 2027 учебном году на заседании кафедры Информационных технологий

Протокол от "____" _____ 2026 г. №____

Зав. кафедрой канд.физ.-мат.наук, доцент, Брадул Н.В.

(подпись)

Визирование РПД для исполнения в очередном учебном году**"УТВЕРЖДАЮ"**

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2027 - 2028 учебном году на заседании кафедры Информационных технологий

Протокол от "____" _____ 2027 г. №____

Зав. кафедрой канд.физ.-мат.наук, доцент, Брадул Н.В.

(подпись)

Визирование РПД для исполнения в очередном учебном году**"УТВЕРЖДАЮ"**

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2028 - 2029 учебном году на заседании кафедры Информационных технологий

Протокол от "____" _____ 2028 г. №____

Зав. кафедрой канд.физ.-мат.наук, доцент, Брадул Н.В.

(подпись)

РАЗДЕЛ 1. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ

1.1. ЦЕЛИ ДИСЦИПЛИНЫ

Изучение теоретических основ, методов и технологий интеллектуального анализа данных (Data Mining), а также их практического применения для обнаружения скрытых закономерностей, прогнозирования и поддержки принятия решений.

1.2. УЧЕБНЫЕ ЗАДАЧИ ДИСЦИПЛИНЫ

изучить базовые понятия Data Mining и его место в процессе извлечения знаний;
освоить ключевые методы интеллектуального анализа данных;
применять методы интеллектуального анализа данных для решения реальных задач;
изучить стадии интеллектуального анализа данных;
научиться работать с инструментами Data Mining: Deductor, RapidMiner.

1.3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Цикл (раздел) ОПОП ВО:	Б1.В
------------------------	------

1.3.1. Дисциплина "Интеллектуальный анализ данных" опирается на следующие элементы ОПОП ВО:

Теория вероятностей и математическая статистика

Аналитика данных на Python

1.3.2. Дисциплина "Интеллектуальный анализ данных" выступает опорой для следующих элементов:

Преддипломная практика

Разработка программных приложений

1.4. РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ:

ПКс-4.1: Применяет современные информационные технологии при проектировании, реализации, оценке качества и анализа эффективности программного обеспечения для решения задач в различных предметных областях

Знать:

Уровень 1	методы и инструментальные средства интеллектуального анализа больших данных
Уровень 2	программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных стандарты проведения анализа данных
Уровень 3	особенности методов обработки и анализа данных - современный опыт использования анализа больших данных

Уметь:

Уровень 1	применять информационные технологии для решения профессиональных задач
Уровень 2	структурировать данные для дальнейшей обработки
Уровень 3	решать прикладные задачи с помощью методов интеллектуального анализа данных и оценивать модели больших данных

Владеть:

Уровень 1	навыками применения информационных технологий для реализации профессиональных задач
Уровень 2	принципами управления крупными массивами данных и их интеллектуальным анализом при решении профессиональных задач
Уровень 3	программными средствами для разработки алгоритмов машинного обучения, алгоритмами построения искусственных нейронных сетей

В результате освоения дисциплины "Интеллектуальный анализ данных" обучающийся

3.1	Знать:
	основные методы сбора, подготовки, исследования и моделирования данных;
	возможности различных методов интеллектуального анализа данных;
	классы задач, для которых целесообразно использовать методы интеллектуального анализа данных (классификация, кластеризация, регрессия, прогнозирование и др.).
3.2	Уметь:
	выбирать методы интеллектуального анализа данных, адекватные решаемой бизнес-задаче;
	применять методы интеллектуального анализа данных для решения практических задач, таких как классификация, регрессия, кластеризация и прогнозирование;

	строить и оценивать модели машинного обучения, используя различные алгоритмы (k-ближайших соседей, логистическая регрессия, метод опорных векторов и др.;
	работать с библиотеками Python для анализа данных и машинного обучения.
3.3 Владеть:	
	навыками работы с инструментами операционной аналитики для автоматизации принятия решений;
	навыками программирования на Python для реализации алгоритмов интеллектуального анализа данных.
1.5. ФОРМЫ КОНТРОЛЯ	
Текущий контроль успеваемости позволяет оценить уровень сформированности элементов компетенций (знаний, умений и приобретенных навыков), компетенций с последующим объединением оценок и проводится в форме: устного опроса на лекционных и семинарских/практических занятиях (фронтальный, индивидуальный, комплексный), письменной проверки (тестовые задания, контроль знаний по разделу, ситуационных заданий и т.п.), оценки активности работы обучающегося на занятии, включая задания для самостоятельной работы.	
Промежуточная аттестация	
Результаты текущего контроля и промежуточной аттестации формируют рейтинговую оценку работы студента. Распределение баллов при формировании рейтинговой оценки работы студента осуществляется в соответствии с действующим локальным нормативным актом. По дисциплине "Интеллектуальный анализ данных" видом промежуточной аттестации является Зачет с оценкой	

РАЗДЕЛ 2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1. ТРУДОЕМКОСТЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ						
Общая трудоёмкость дисциплины "Интеллектуальный анализ данных" составляет 4 зачётные единицы, 144 часов.						
Количество часов, выделяемых на контактную работу с преподавателем и самостоятельную работу обучающегося, определяется учебным планом.						
2.2. СОДЕРЖАНИЕ РАЗДЕЛОВ ДИСЦИПЛИНЫ						
Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература	Инте ракт.	Примечание
Раздел 1. Раздел 1						
Тема 1. Введение в интеллектуальный анализ данных /Лек/	7	4	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 1. Введение в интеллектуальный анализ данных /Пр/	7	8	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 1. Введение в интеллектуальный анализ данных /Ср/	7	18	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	

Тема 2. Стадии интеллектуального анализа данных /Лек/	7	2	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 2. Стадии интеллектуального анализа данных /Пр/	7	4	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 2. Стадии интеллектуального анализа данных /Ср/	7	12	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 3. Методы интеллектуального анализа данных /Лек/	7	4	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 3. Методы интеллектуального анализа данных /Пр/	7	16	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 3. Методы интеллектуального анализа данных /Ср/	7	16	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 4. Этапы проведения интеллектуального анализа данных /Лек/	7	2	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 4. Этапы проведения интеллектуального анализа данных /Пр/	7	6	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 4. Этапы проведения интеллектуального анализа данных /Ср/	7	12	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	

				Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4		
Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner) /Лек/	7	4	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner) /Пр/	7	14	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner) /Ср/	7	16	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
Консультации по курсу /Конс/	7	2	ПКс-4.1	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	

РАЗДЕЛ 3. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В процессе освоения дисциплины используются следующие образовательные технологии: лекции (Л), практические занятия (ПР), самостоятельная работа студентов (СР) по выполнению различных видов заданий.

1. В процессе освоения дисциплины используются следующие интерактивные образовательные технологии: Лекционный материал представлен в виде слайд-презентации в формате «Power Point». Для наглядности используются материалы различных справочных материалов, научных статей т.д. В ходе лекции предусмотрена обратная связь со студентами, активизирующие и поощрительные вопросы, обращение к аудитории за «подсказкой», проверка готовности студентов к лекции. При проведении лекций используется проблемно-ориентированный междисциплинарный подход, предполагающий творческие вопросы и создание дискуссионных ситуаций.

2. При изложении теоретического материала используются такие методы:

- монологический;
- показательный;
- диалогический;
- эвристический;
- исследовательский.

3. Используются следующие принципы дидактики высшей школы:

- последовательность обучения;
- систематичность обучения;
- доступность обучения;
- принцип научности;
- принципы взаимосвязи теории и практики;
- принцип наглядности и др.

В конце каждой лекции предусмотрено время для ответов на проблемные вопросы.

4. Самостоятельная работа предназначена для внеаудиторной работы студентов, связанной с изучением дополнительной литературы по дисциплине, подготовкой к текущему и семестровому контролю, а также выполнением индивидуального задания за компьютером с использованием необходимого программного обеспечения

РАЗДЕЛ 4. УЧЕБНО-МЕТОДИЧЕСКОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

4.1. Рекомендуемая литература

1. Основная литература

	Авторы,	Заглавие	Издательство, год
Л1.1	А. А. Алетдинова, М. Ш. Муртазина.	Интеллектуальный анализ больших данных : учебное пособие (66с.)	Новосибирск : НГТУ, 2023
Л1.2	Л. В. Жукова	Интеллектуальный анализ данных на платформе Loginom : методические указания (45с.)	Москва : Издательский Дом НИТУ «МИСиС», 2023
Л1.3	Л. А. Демидова	Интеллектуальный анализ данных на языке Python : учебно-методическое пособие - Текст : электронный. - URL: https://e.lanbook.com/book/218693 (92с.)	Москва : РТУ МИРЭА, 2021

2. Дополнительная литература

	Авторы,	Заглавие	Издательство, год
Л2.1	Ю. В. Ланских, В. Г. Ланских	Интеллектуальный анализ данных : учебное пособие (240с.)	Киров : ВятГУ, 2023
Л2.2	А. В. Замятин	Интеллектуальный анализ данных : учебное пособие (123с.)	Москва : РТУ МИРЭА, 2023
Л2.3	М. В. Хачумов	Введение в интеллектуальный анализ данных : учебное пособие (196сю)	Москва : РТУ МИРЭА, 2023

3. Методические разработки

	Авторы,	Заглавие	Издательство, год
ЛЗ.1	И.В. Тарабаева	Конспект лекций по учебной дисциплине по учебной дисциплине «Интеллектуальный анализ данных» для обучающихся 4 курса образовательной программы бакалавриата направления подготовки 9.03.03 Прикладная информатика (профиль "Прикладная информатика в управлении корпоративными информационными системами") очной формы обучения (96с.)	Донецк : ФГБОУ ВО "ДОНАУИГС", 2024
ЛЗ.2	И.В. Тарабаева	Методические рекомендации для проведения практических занятий по учебной дисциплине «Интеллектуальный анализ данных» для обучающихся 4 курса образовательной программы бакалавриата направления подготовки 9.03.03 Прикладная информатика (профиль "Прикладная информатика в управлении корпоративными информационными системами") очной формы обучения (60с.)	Донецк : ФГБОУ ВО "ДОНАУИГС", 2024

4.2. Перечень ресурсов

информационно-телекоммуникационной сети "Интернет"

Э1	ЭБС «ЗНАНИУМ»	https://znanium.ru
Э2	Научная электронная библиотека «КиберЛенинка»	https://cyberleninka.ru/
Э3	ЭБС «ЛАНЬ»	https://e.lanbook.com
Э4	ЭБС «SOCHUM»	https://sochum.ru

4.3. Перечень программного обеспечения

Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства:

Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства:

- Libre Office (лицензия Mozilla Public License v2.0.)
- 7-Zip (лицензия GNU Lesser General Public License)
- AIMP (лицензия LGPL v.2.1)
- STDU Viewer (freeware for private non-commercial or educational use)
- GIMP (лицензия GNU General Public License)
- Inkscape (лицензия GNU General Public License)

4.4. Профессиональные базы данных и информационные справочные системы

Не используются

4.5. Материально-техническое обеспечение дисциплины

Для проведения учебных занятий, предусмотренных образовательной программой, закреплены аудитории согласно расписанию учебных занятий:

рабочее место преподавателя, посадочные места по количеству обучающихся, доска меловая, персональный компьютер с лицензированным программным обеспечением общего назначения, мультимедийный проектор, экран, интерактивная панель.

РАЗДЕЛ 5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Контрольные вопросы и задания

ВОПРОСЫ К ЗАЧЕТУ

1. Дайте определение Data Mining согласно Григорию Пятецкому-Шапиро.
2. В чем заключается ключевая цель интеллектуального анализа данных?
3. Что подразумевается под "скрытыми знаниями" в контексте Data Mining?
4. Чем отличается Data Mining от Knowledge Discovery in Databases (KDD)?
5. Перечислите и кратко опишите 5 этапов процесса KDD.
6. Почему этап очистки данных критически важен для анализа?
7. Какие задачи решаются на этапе трансформации данных?
8. Какую роль играет интерпретация результатов в KDD?
9. Какие существуют типы данных (числовые, символьные, транзакционные)?
10. Чем отличаются дискретные и непрерывные данные? Приведите примеры.
11. Что такое транзакционные данные и где они применяются?
12. Почему Data Mining стал активно развиваться в 90-х годах XX века?
13. Какие научные дисциплины легли в основу методов Data Mining?
14. Как шкала измерения влияет на выбор методов Data Mining?
15. Какие форматы данных поддерживают инструменты Data Mining?
16. Назовите три стадии ИАД (рис. 2) и дайте их краткую характеристику.
17. Что такое свободный поиск (Discovery) и какие методы он включает?
18. Как работает прогностическое моделирование на основе найденных закономерностей?
19. Приведите пример использования анализа исключений (Forensic Analysis).
20. Почему свободный поиск считается индуктивным, а прогнозирование – дедуктивным?
21. На какие две группы делятся методы ИАД по принципу работы с данными?
22. Какие алгоритмы относятся к методам рассуждений на прецедентах (k-NN, Lazy-Learning)?
23. В чем преимущество деревьев решений перед нейронными сетями?
24. Как ассоциативные правила (Apriori) применяются в ритейле?
25. Какие методы используются для прогнозирования временных рядов?
26. На какие две группы делятся методы ИАД по принципу работы с данными?
27. Какие алгоритмы относятся к методам рассуждений на прецедентах (k-NN, Lazy-Learning)?
28. В чем преимущество деревьев решений перед нейронными сетями?
29. Как ассоциативные правила (Apriori) применяются в ритейле?
30. Какие методы используются для прогнозирования временных рядов?
31. В чем принципиальное отличие логистической регрессии от линейной регрессии? В каких задачах Data Mining применяется каждая из них?
32. Как работает метод опорных векторов (SVM) для классификации? Какие ядра (kernel) наиболее часто используются и почему?
33. Какие метрики расстояния (евклидово, манхэттенское, косинусное) применяются в алгоритмах кластеризации? Как выбор метрики влияет на результат?
34. В чем преимущество иерархической кластеризации перед k-means? Когда ее стоит использовать?
35. Как алгоритм Apriori находит частые наборы itemset'ов? Какие параметры (support, confidence, lift) критичны для интерпретации правил?
36. Почему ансамблевые методы (Random Forest, Gradient Boosting) часто outperform одиночные

модели? Как они сочетаются с деревьями решений?

37. Почему понимание задачи – первый и важнейший этап анализа?

38. Какие шаги включает подготовка данных (препроцессинг)?

39. Как происходит проверка качества моделей (валидация)?

40. Почему интерпретация результатов требует участия аналитика?

41. Приведите пример практического применения Data Mining в бизнесе (например, прогнозирование оттока клиентов).

42. В чем заключаются ключевые различия между Deductor и RapidMiner с точки зрения функционала для Data Mining?

43. Какие типы данных (табличные, транзакционные, временные ряды) поддерживают обе платформы?

44. Какие методы трансформации данных (нормализация, дискретизация) доступны в этих платформах?

45. Как в этих платформах реализованы методы кластеризации (k-means, DBSCAN)?

46. Какие инструменты визуализации результатов (графики, тепловые карты) предлагают Deductor и RapidMiner?

47. С какими внешними системами (1C, SQL, Excel) интегрируются Deductor и RapidMiner?

48. Как в Deductor и RapidMiner реализована оптимизация глубины дерева (pruning) для предотвращения переобучения модели? Какие параметры настройки доступны для этого?

49. Как в Deductor и RapidMiner реализованы инструменты для очистки данных (обработка пропусков, выбросов)?

50. Какие алгоритмы классификации (деревья решений, SVM, нейросети) есть в Deductor и RapidMiner?

51. Как платформы помогают интерпретировать модели (например, объяснение правил ассоциации или важности признаков)?

52. Приведите пример реальной задачи (например, прогнозирование продаж), где Deductor или RapidMiner дают преимущество перед другими инструментами.

ТИПОВЫЕ ТЕСТОВЫЕ ЗАДАНИЯ

ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ

ЗАДАНИЯ ЗАКРЫТОГО ТИПА

ВЫБЕРИТЕ ОДИН ВЕРНЫЙ ОТВЕТ / ВЫБЕРИТЕ НЕСКОЛЬКО ВЕРНЫХ ОТВЕТОВ*

Задание 1

Прочитайте текст, выберите один правильный ответ и запишите аргументы, обосновывающие выбор.

ИАД (интеллектуальный анализ данных) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей, то есть извлечения информации, которая может быть охарактеризована как знания. Технология Knowledge Discovery in Databases (KDD) описывает не конкретный алгоритм или математический аппарат, а последовательность действий, которую необходимо выполнить для построения модели с целью извлечения знания. KDD включает в себя этапы подготовки данных, предобработки (очистки) данных, трансформации данных, построения моделей (применения методов Data Mining) и интерпретации полученных результатов.

Какой этап процесса KDD является ключевым для применения методов Data Mining?

- А. Выборка данных из источников.
- Б. Очистка и трансформация данных.
- В. Построение моделей с помощью алгоритмов Data Mining.
- Г. Интерпретация полученных результатов.

Ответ:

Обоснование выбора:

Задание 2

Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор.

ИАД (интеллектуальный анализ данных) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей, то есть извлечения информации, которая может быть охарактеризована как знания.

Все методы ИАД подразделяются на две большие группы по принципу работы с исходными обучающими данными: непосредственное использование обучающих данных; выявление и использование формализованных закономерностей.

Какие два метода из перечисленных относятся к группе «Выявление и использование формализованных закономерностей» в классификации методов ИАД?

- А. Метод ближайшего соседа (k-NN).
- Б. Деревья решений.
- В. Байесовские сети.

Г. Нейронные сети.

Ответ:

Обоснование выбора:

*(ответ – все ответы верны – быть не может)

ЗАДАНИЯ ОТКРЫТОГО ТИПА**Задание 1**

Прочитайте текст и запишите развернутый обоснованный ответ.

ИАД (интеллектуальный анализ данных) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей, то есть извлечения информации, которая может быть охарактеризована как знания.

Стадии процесса интеллектуального анализа данных:

- Свободный поиск определяется как процесс исследования исходной БД на предмет поиска скрытых закономерностей без предварительного определения гипотез относительно вида этих закономерностей.
- Прогностическое моделирование (Predictive Modeling). Здесь, на второй стадии ИАД, используются плоды работы первой, то есть найденные в БД закономерности применяются для предсказания неизвестных значений.
- Анализ исключений (Forensic Analysis). Предметом данного анализа являются аномалии в раскрытых закономерностях, то есть необъясненные исключения.

Задание:

Почему свободный поиск (Discovery) считается важным этапом для анализа больших данных?

Как прогностическое моделирование использует результаты свободного поиска?

Приведите пример анализа исключений (Forensic Analysis) из профессиональной сферы.

Ответ:

Задание 2

Прочитайте текст и запишите развернутый обоснованный ответ.

Развитие информационно-коммуникационных технологий способствует значительному увеличению объемов, генерируемых данных. Эффективный анализ этих массивов позволяет извлекать ценные инсайты, способствующие повышению эффективности управленческих решений в коммерческом секторе, здравоохранении и государственном управлении. Данная область исследований и практического применения получила название интеллектуальный анализ данных. Его цель заключается в обнаружении ранее неизвестных, нетривиальных и практически полезных закономерностей в больших наборах данных.

Задание:

1. Дайте определение интеллектуального анализа данных (ИАД) согласно Григорию Пятецкому-Шапиро. Какие ключевые характеристики скрытых знаний он выделяет?
2. Объясните, почему в 90-х годах XX века возникла потребность в методах ИАД. Какие дисциплины легли в его основу?
3. Приведите пример нетривиальной закономерности, которую можно обнаружить с помощью ИАД, и объясните её практическую пользу.

5.2. Темы письменных работ

Письменные работы не предусмотрены

5.3. Фонд оценочных средств

Фонд оценочных средств дисциплины "Геоинформационные системы" разработан в соответствии с локальным нормативным актом ФГБОУ ВО "ДОНАУИГС". Фонд оценочных средств дисциплины "Геоинформационные системы" в полном объеме представлен в виде приложения к данному РПД

5.4. Перечень видов оценочных средств

Устный опрос (проводится на практических занятиях; контроль знаний раздела учебной дисциплины)

Индивидуальные задания

Собеседование

РАЗДЕЛ 6. СРЕДСТВА АДАПТАЦИИ ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ К

ПОТРЕБНОСТЯМ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться одни из следующих вариантов восприятия информации с учетом их индивидуальных психофизических особенностей:

- 1) с применением электронного обучения и дистанционных технологий.
- 2) с применением специального оборудования (техники) и программного обеспечения, имеющихся в ФГБОУ ВО "ДОНАУИГС".

В процессе обучения при необходимости для лиц с нарушениями зрения, слуха и опорно-двигательного аппарата предоставляются следующие условия:

- для лиц с нарушениями зрения: учебно-методические материалы в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные задания и консультации.
- для лиц с нарушениями слуха: учебно-методические материалы в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.
- для лиц с нарушениями опорно-двигательного аппарата: учебно-методические материалы в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

РАЗДЕЛ 7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО УСВОЕНИЮ ДИСЦИПЛИНЫ

Аудиторные занятия по дисциплине "Интеллектуальный анализ данных" проводятся в форме лекционных и практических занятий.

На лекционных занятиях, согласно учебному плану дисциплины, обучающимся предлагается рассмотреть основные темы курса. Студенту предлагается участвовать в диалоге с преподавателем, в ходе которого могут обсуждаться моменты, актуальные для его будущей практической деятельности; он может высказать свое мнение после сопоставления разных фактов и разнообразных точек зрения на них.

К числу важнейших умений, являющихся неотъемлемой частью успешного учебного процесса, относится умение работать с различными литературными источниками, содержание которых так или иначе связано с изучаемой дисциплиной.

Подготовку к любой теме курса рекомендуется начинать с изучения презентационных материалов или учебной литературы, в которых дается систематизированное изложение материала, разъясняется смысл разных терминов и сообщается об изменениях в подходах к изучению тех или иных проблем данного курса.

Методические указания по организации самостоятельной работы

Самостоятельная работа по дисциплине организована в следующих видах:

1. изучить теоретический материал по заданной теме;
2. выбрать методы решения поставленной задачи;
3. выполнить индивидуальные задания;
4. проанализировать полученные результаты;
5. отчитаться перед преподавателем по теоретической и практической части индивидуальной работы.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ДОНЕЦКАЯ АКАДЕМИЯ УПРАВЛЕНИЯ И ГОСУДАРСТВЕННОЙ СЛУЖБЫ»**

**Факультет государственной службы и управления
Кафедра информационных технологий**

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

«Интеллектуальный анализ данных»

Направление подготовки	09.03.03 Прикладная информатика
Профиль	«Прикладная информатика в управлении корпоративными информационными системами»
Квалификация	бакалавр
Форма обучения	очная

Донецк
2024

Фонд оценочных средств по дисциплине «Интеллектуальный анализ данных» для обучающихся 4 курса образовательной программы бакалавриата направления подготовки 09.03.03 Прикладная информатика (профиль: «Прикладная информатика в управлении корпоративными информационными системами») очной формы обучения

Автор,

разработчик:

доцент, канд. техн. наук, доцент Тарабаева И.В.

ФОС рассмотрен на
заседании кафедры

информационных технологий

Протокол заседания кафедры от

16.04.2024 г.

№

№ 9

Заведующий кафедрой

Н.В. Брадул

РАЗДЕЛ 1.
ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине «Интеллектуальный анализ данных»

1.1. Основные сведения о дисциплине

Таблица 1

Характеристика дисциплины
(сведения соответствуют разделу РПД)

Образовательная программа	бакалавриат
Направление подготовки Профиль	09.03.03 Прикладная информатика «Прикладная информатика в управлении корпоративными информационными системами»
Количество разделов дисциплины	1
Часть образовательной программы	Б1.В.11
Формы текущего контроля	индивидуальные задания, устный опрос, собеседование
<i>Показатели</i>	Очная форма обучения
Количество зачетных единиц (кредитов)	4
Семестр	7
Общая трудоемкость (академ. часов)	144
Аудиторная контактная работа:	66
Лекционные занятия	16
Практические занятия	48
Консультации	2
Самостоятельная работа	74
Контроль	4
Форма промежуточной аттестации	Зачет с оценкой

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

Таблица 2

Перечень компетенций и их элементов

Компетенция	Индикатор компетенции и его формулировка	Элементы индикатора компетенции	Индекс элемента
ПКс-4: Способность моделировать прикладные (бизнес) процессы и предметную область	ПКс-4.1: Применяет современные информационные технологии при проектировании, реализации, оценке качества и анализа эффективности программного обеспечения для решения задач в различных предметных областях	Знать:	
		1. методы и инструментальные средства интеллектуального анализа больших данных	ПКс-4.1 З-1
		2. программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных стандарты проведения анализа данных	ПКс-4.1 З-2
		3. особенности методов обработки и анализа данных - современный опыт использования анализа больших данных	ПКс-4.1 З-3
		Уметь:	
		1. применять информационные технологии для решения профессиональных задач	ПКс-4.1 У-1
		2. структурировать данные для дальнейшей обработки	ПКс-4.1 У-2
		3. решать прикладные задачи с	ПКс-4.1 У-3

Компетенция	Индикатор компетенции и его формулировка	Элементы индикатора компетенции	Индекс элемента
		помощью методов интеллектуального анализа данных и оценивать модели больших данных	
		Владеть:	
		1. навыками применения информационных технологий для реализации профессиональных задач	ПКс-4.1 В-1
		2. принципами управления крупными массивами данных и их интеллектуальным анализом при решении профессиональных задач	ПКс-4.1 В-2
		3 программными средствами для разработки алгоритмов машинного обучения, алгоритмами построения искусственных нейронных сетей	ПКс-4.1 В-3

Таблица 3

Этапы формирования компетенций в процессе освоения основной образовательной программы

№ п/п	Контролируемые разделы (темы) дисциплины	Этапы формирования компетенций (номер семестра)	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
-------	--	---	---	----------------------------------

Раздел 1.				
1.	Тема 1. Введение в интеллектуальный анализ данных	7	ПКс-4.1 3-1 ПКс-4.1 3-3 ПКс-4.1 У-2 ПКс-4.1 В-1 ПКс-4.1 В-2	Устный опрос Индивидуальное задание №1 Собеседование
2.	Тема 2. Стадии интеллектуального анализа данных	7	ПКс-4.1 3-3 ПКс-4.1 У-2 ПКс-4.1 В-2	Устный опрос Собеседование
3.	Тема 3. Методы интеллектуального анализа данных	7	ПКс-4.1 3-1 ПКс-4.1 3-2 ПКс-4.1 У-1 ПКс-4.1 У-2 ПКс-4.1 У-3 ПКс-4.1 В-2 ПКс-4.1 В-3	Устный опрос Индивидуальное задание №2 Индивидуальное задание №3 Собеседование
4.	Тема 4. Этапы проведения интеллектуального анализа данных	7	ПКс-4.1 3-2 ПКс-4.1 3-3 ПКс-4.1 У-2 ПКс-4.1 В-2	Собеседование Устный опрос
5.	Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner)	7	ПКс-4.1 3-1 ПКс-4.1 3-2 ПКс-4.1 У-1 ПКс-4.1 У-2 ПКс-4.1 У-3 ПКс-4.1 В-1 ПКс-4.1 В-2	Устный опрос Индивидуальное задание №4 Индивидуальное задание №5 Собеседование

РАЗДЕЛ 2 ТЕКУЩИЙ КОНТРОЛЬ ПО ДИСЦИПЛИНЕ

Текущий контроль знаний используется для оперативного и регулярного управления учебной деятельностью (в том числе самостоятельной работой) обучающихся. В условиях балльно-рейтинговой системы контроля результаты текущего оценивания обучающегося используются как показатель его текущего рейтинга. Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы по индивидуальной инициативе преподавателя. Данный вид контроля стимулирует у обучающегося стремление к систематической самостоятельной работе по изучению дисциплины.

Таблица 2.1.

Распределение баллов по видам учебной деятельности

(балльно-рейтинговая система)

Наименование Раздела/Темы	Вид задания				
	ПЗ		Всего за тему	КЗР (С)	ИЗ
	УО	РЗ			
P.1.T.1	2	4	6		10
P.1.T.2	2	4	6	10	
P.1.T.3	2	4	6		20
P.1.T.4	2	4	6	10	
P.1.T.5	2	4	6		20
Итого: 100б	10	20	30	20	50

УО – устный опрос;

С – собеседование;

РЗ – разноуровневые задания;

ПЗ – практическое занятие;

КЗР – контроль знаний по Разделу;

СР – самостоятельная работа обучающегося

ИЗ – индивидуальное задание

2.1. Рекомендации по оцениванию индивидуальных заданий обучающихся

Максимальное количество баллов*	Критерии
Отлично	Выставляется обучающемуся: если выполнены все пункты работы самостоятельно, без ошибок, если предложен более рациональный алгоритм решения задачи.
Хорошо	Выставляется обучающемуся: если самостоятельно выполнены все пункты работы, допущены незначительные ошибки, если предложен более рациональный алгоритм решения задачи.
Удовлетворительно	Выставляется обучающемуся: если самостоятельно (или с помощью преподавателя) выполнены все пункты работы, допущены грубые ошибки.
Неудовлетворительно	Выставляется обучающемуся: если с помощью преподавателя выполнены не все пункты работы, допущены грубые ошибки.

* Представлено в таблице 2.1.

ТИПОВЫЕ ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ УРОВНЯ КОМПЕТЕНЦИИ

Тема 1. Введение в интеллектуальный анализ данных

Индивидуальное задание 1

Задание

Необходимо организовать подготовку данных для построения модели классификации. В качестве данных выбран набор данных об ирисах Фишера. Данный набор данных предназначен для построения модели классификации. Набор данных о 150 экземплярах ириса, по 50 экземпляров из трёх видов – Ирис щетинистый (*Iris setosa*), Ирис виргинский (*Iris virginica*) и Ирис разноцветный (*Iris versicolor*) необходимо скачать из репозитория Center for Machine Learning and Intelligent Systems (необходим только один текстовый файл с данными измерений *Iris.data*): <http://archive.ics.uci.edu/ml/datasets/Iris..> Для каждого экземпляра измерялись четыре характеристики (в сантиметрах): 1) длина наружной доли околоцветника (*sepal length*); 2) ширина наружной доли околоцветника (*sepal width*); 3) длина внутренней доли околоцветника (*petal length*); 4) ширина внутренней доли околоцветника (*petal width*). На основании этого набора данных требуется построить правило классификации, определяющее вид растения по данным измерений. Это задача многоклассовой классификации, так как имеется три класса – три вида ириса.

Тема 3. Методы интеллектуального анализа данных

Индивидуальное задание 3

Задание

На основе предоставленного набора данных *Mall_customers.csv* реализовать модель кластеризации методом К-средних. Для реализации кластерного анализа использовать язык Python.

Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner)

Индивидуальное задание 5

Задание

В коммерческом банке имеется продукт "Нецелевой потребительский кредит": кредиты предоставляются на любые цели с принятием решения в течение нескольких часов. В банке накоплена статистическая информация

о заемщиках и качестве обслуживания ими долга за несколько месяцев. Руководство банка, поставило перед отделом кредитования задачу разработать модель, позволяющую минимизировать риски невозврата кредита. На основе этой модели должно приниматься решение - выдать клиенту, обратившемуся в банк, кредит или отказать.

Показать последовательность решения этой бизнес-задачи классификации заемщика с помощью подхода, который основан на алгоритме дерева решений. Использованию инструментарий аналитической платформы Deductor.

2.2. Рекомендации по оцениванию устных ответов обучающихся

С целью контроля усвоения пройденного материала и определения уровня подготовленности обучающихся к изучению новой темы в начале практического занятия преподавателем проводится индивидуальный устный опрос по выполненным заданиям предыдущей темы.

Критерии оценки.

Оценка «отлично» ставится, если обучающийся:

- 1) полно и аргументировано отвечает по содержанию вопроса;
- 2) обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры;
- 3) излагает материал последовательно и правильно, с соблюдением исторической и хронологической последовательности;

Оценка «хорошо» – ставится, если обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает одна-две ошибки, которые сам же исправляет.

Оценка «удовлетворительно» – ставится, если обучающийся обнаруживает знание и понимание основных положений данного задания, но:

- 1) излагает материал неполно и допускает неточности в определении понятий или формулировке правил;
- 2) не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры;
- 3) излагает материал непоследовательно и допускает ошибки.

ТИПОВЫЕ ВОПРОСЫ ДЛЯ САМОПОДГОТОВКИ ДЛЯ ПРОВЕРКИ УРОВНЯ КОМПЕТЕНЦИИ

Контролируемые разделы (темы) учебной дисциплины	Вопросы для контроля знаний по разделам дисциплины
Раздел 1.	
Тема 1. Введение в интеллектуальный анализ данных	1. Дайте определение Data Mining согласно Григорию Пятецкому-Шапиро. 2. В чем заключается ключевая цель интеллектуального анализа данных? 3. Что подразумевается под "скрытыми знаниями" в контексте Data Mining? 4. Чем отличается Data Mining от Knowledge Discovery in Databases (KDD)? 5. Перечислите и кратко опишите 5 этапов процесса KDD. 6. Почему этап очистки данных критически важен для анализа? 7. Какие задачи решаются на этапе трансформации данных? 8. Какую роль играет интерпретация результатов в KDD? 9. Какие существуют типы данных (числовые, символьные, транзакционные)? 10. Чем отличаются дискретные и непрерывные данные? Приведите примеры. 11. Что такое транзакционные данные и где они применяются?
Тема 2. Стадии интеллектуального анализа данных	12. Назовите три стадии ИАД и дайте их краткую характеристику. 13. Что такое свободный поиск (Discovery) и какие методы он включает? 14. Как работает прогностическое моделирование на основе найденных закономерностей? 15. Приведите пример использования анализа исключений (Forensic Analysis).
Тема 3. Методы интеллектуального анализа данных	16. На какие две группы делятся методы ИАД по принципу работы с данными? 17. Какие алгоритмы относятся к методам рассуждений на прецедентах (k-NN, Lazy-Learning)?

Контролируемые разделы (темы) учебной дисциплины	Вопросы для контроля знаний по разделам дисциплины
	18. В чем преимущество деревьев решений перед нейронными сетями? 19. Как ассоциативные правила (Apriori) применяются в ритейле? 20. Какие методы используются для прогнозирования временных рядов?
Тема 4. Этапы проведения интеллектуального анализа данных	21. Почему понимание задачи – первый и важнейший этап анализа? 22. Какие шаги включает подготовка данных (препроцессинг)? 23. Как происходит проверка качества моделей (валидация)? 24. Почему интерпретация результатов требует участия аналитика?
Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner)	25. В чем заключаются ключевые различия между Deductor и RapidMiner с точки зрения функционала для Data Mining? 26. Какие типы данных (табличные, транзакционные, временные ряды) поддерживают обе платформы? 27. Какие методы трансформации данных (нормализация, дискретизация) доступны в этих платформах? 28. Как в этих платформах реализованы методы кластеризации (k-means, DBSCAN)? 29. Какие инструменты визуализации результатов (графики, тепловые карты) предлагают Deductor и RapidMiner? 30. С какими внешними системами (1C, SQL, Excel) интегрируются Deductor и RapidMiner? 31. Как в Deductor и RapidMiner реализована оптимизация глубины дерева (pruning) для предотвращения переобучения модели? Какие параметры настройки доступны для этого?

Контролируемые разделы (темы) учебной дисциплины	Вопросы, выносимые на самостоятельное обучение по разделам дисциплины (собеседование)
Раздел 1.	
Тема 1. Введение в интеллектуальный анализ данных	1. Почему Data Mining стал активно развиваться в 90-х годах XX века? 2. Какие научные дисциплины легли в основу методов Data Mining? 3. Как шкала измерения влияет на выбор методов Data Mining? 4. Какие форматы данных поддерживают инструменты Data Mining?
Тема 2. Стадии интеллектуального анализа данных	5. Почему свободный поиск считается индуктивным, а прогнозирование – дедуктивным?
Тема 3. Методы интеллектуального анализа данных	6. В чем принципиальное отличие логистической регрессии от линейной регрессии? В каких задачах Data Mining применяется каждая из них? 7. Как работает метод опорных векторов (SVM) для классификации? Какие ядра (kernel) наиболее часто используются и почему? 8. Какие метрики расстояния (евклидово, манхэттенское, косинусное) применяются в алгоритмах кластеризации? Как выбор метрики влияет на результат? 9. В чем преимущество иерархической кластеризации перед k-means? Когда ее стоит использовать? 10. Как алгоритм Apriori находит частые наборы itemset'ов? Какие параметры (support, confidence, lift) критичны для интерпретации правил? 11. Почему ансамблевые методы (Random Forest, Gradient Boosting) часто outperform одиночные модели? Как они сочетаются с деревьями решений?
Тема 4. Этапы проведения интеллектуального анализа данных	12. Приведите пример практического применения Data Mining в бизнесе (например, прогнозирование оттока клиентов).

Контролируемые разделы (темы) учебной дисциплины	Вопросы, выносимые на самостоятельное обучение по разделам дисциплины (собеседование)
Тема 5. Использование инструментов интеллектуального анализа данных (Deductor, RapidMiner)	13. Как в Deductor и RapidMiner реализованы инструменты для очистки данных (обработка пропусков, выбросов)? 14. Какие алгоритмы классификации (деревья решений, SVM, нейросети) есть в Deductor и RapidMiner? 15. Как платформы помогают интерпретировать модели (например, объяснение правил ассоциации или важности признаков)? 16. Приведите пример реальной задачи (например, прогнозирование продаж), где Deductor или RapidMiner дают преимущество перед другими инструментами.

2.3. Рекомендации по оцениванию результатов тестовых заданий обучающихся

В завершении изучения каждого раздела дисциплины проводится тестирование (контроль знаний по разделу).

Критерии оценивания. Уровень выполнения текущих тестовых заданий оценивается в баллах. Максимальное количество баллов по тестовым заданиям представлено в таблице 2.1.

Тестовые задания представлены в виде оценочных средств и в полном объеме представлены в банке тестовых заданий в электронном виде. В фонде оценочных средств представлены типовые тестовые задания, разработанные для изучения дисциплины «Интеграция информационных систем».

ТИПОВЫЕ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ ЗАДАНИЯ ЗАКРЫТОГО ТИПА

ВЫБЕРИТЕ ОДИН ВЕРНЫЙ ОТВЕТ / ВЫБЕРИТЕ НЕСКОЛЬКО ВЕРНЫХ ОТВЕТОВ*

Задание 1

Прочитайте текст, выберите один правильный ответ и запишите аргументы, обосновывающие выбор.

ИАД (интеллектуальный анализ данных) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей, то есть извлечения информации, которая может быть охарактеризована как знания. Технология Knowledge Discovery in Databases (KDD) описывает не конкретный алгоритм или математический аппарат, а

последовательность действий, которую необходимо выполнить для построения модели с целью извлечения знания. KDD включает в себя этапы подготовки данных, предобработки (очистки) данных, трансформации данных, построения моделей (применения методов Data Mining) и интерпретации полученных результатов.

Какой этап процесса KDD является ключевым для применения методов Data Mining?

- А. Выборка данных из источников.
- Б. Очистка и трансформация данных.
- В. Построение моделей с помощью алгоритмов Data Mining.
- Г. Интерпретация полученных результатов.

Ответ:

Обоснование выбора:

Задание 2

Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор.

ИАД (интеллектуальный анализ данных) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей, то есть извлечения информации, которая может быть охарактеризована как знания.

Все методы ИАД подразделяются на две большие группы по принципу работы с исходными обучающими данными: непосредственное использование обучающих данных; выявление и использование формализованных закономерностей.

Какие два метода из перечисленных относятся к группе «Выявление и использование формализованных закономерностей» в классификации методов ИАД?

- А. Метод ближайшего соседа (k-NN).
- Б. Деревья решений.
- В. Байесовские сети.
- Г. Нейронные сети.

Ответ:

Обоснование выбора:

****(ответ – все ответы верны – быть не может)***

ЗАДАНИЯ ОТКРЫТОГО ТИПА

Задание 1

Прочитайте текст и запишите развернутый обоснованный ответ.

ИАД (интеллектуальный анализ данных) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых

закономерностей, то есть извлечения информации, которая может быть охарактеризована как знания.

Стадии процесса интеллектуального анализа данных:

- Свободный поиск определяется как процесс исследования исходной БД на предмет поиска скрытых закономерностей без предварительного определения гипотез относительно вида этих закономерностей.
- Прогностическое моделирование (Predictive Modeling). Здесь, на второй стадии ИАД, используются плоды работы первой, то есть найденные в БД закономерности применяются для предсказания неизвестных значений.
- Анализ исключений (Forensic Analysis). Предметом данного анализа являются аномалии в раскрытых закономерностях, то есть необъясненные исключения.

Задание:

Почему свободный поиск (Discovery) считается важным этапом для анализа больших данных?

Как прогностическое моделирование использует результаты свободного поиска?

Приведите пример анализа исключений (Forensic Analysis) из профессиональной сферы.

Ответ:

Задание 2

Прочитайте текст и запишите развернутый обоснованный ответ.

Развитие информационно-коммуникационных технологий способствует значительному увеличению объемов, генерируемых данных. Эффективный анализ этих массивов позволяет извлекать ценные инсайты, способствующие повышению эффективности управленческих решений в коммерческом секторе, здравоохранении и государственном управлении. Данная область исследований и практического применения получила название интеллектуальный анализ данных. Его цель заключается в обнаружении ранее неизвестных, нетривиальных и практически полезных закономерностей в больших наборах данных.

Задание:

1. Дайте определение интеллектуального анализа данных (ИАД) согласно Григорию Пятецкому-Шапиро. Какие ключевые характеристики скрытых знаний он выделяет?
2. Объясните, почему в 90-х годах XX века возникла потребность в методах ИАД. Какие дисциплины легли в его основу?
3. Приведите пример нетривиальной закономерности, которую можно обнаружить с помощью ИАД, и объясните её практическую пользу.

Ответ:

ВОПРОСЫ К ЗАЧЕТУ

1. Дайте определение Data Mining согласно Григорию Пятецкому-Шапиро.
2. В чем заключается ключевая цель интеллектуального анализа данных?
3. Что подразумевается под "скрытыми знаниями" в контексте Data Mining?
4. Чем отличается Data Mining от Knowledge Discovery in Databases (KDD)?
5. Перечислите и кратко опишите 5 этапов процесса KDD.
6. Почему этап очистки данных критически важен для анализа?
7. Какие задачи решаются на этапе трансформации данных?
8. Какую роль играет интерпретация результатов в KDD?
9. Какие существуют типы данных (числовые, символьные, транзакционные)?
10. Чем отличаются дискретные и непрерывные данные? Приведите примеры.
11. Что такое транзакционные данные и где они применяются?
12. Почему Data Mining стал активно развиваться в 90-х годах XX века?
13. Какие научные дисциплины легли в основу методов Data Mining?
14. Как шкала измерения влияет на выбор методов Data Mining?
15. Какие форматы данных поддерживают инструменты Data Mining?
16. Назовите три стадии ИАД (рис. 2) и дайте их краткую характеристику.
17. Что такое свободный поиск (Discovery) и какие методы он включает?
18. Как работает прогностическое моделирование на основе найденных закономерностей?
19. Приведите пример использования анализа исключений (Forensic Analysis).
20. Почему свободный поиск считается индуктивным, а прогнозирование – дедуктивным?
21. На какие две группы делятся методы ИАД по принципу работы с данными?
22. Какие алгоритмы относятся к методам рассуждений на прецедентах (k-NN, Lazy-Learning)?
23. В чем преимущество деревьев решений перед нейронными сетями?
24. Как ассоциативные правила (Apriori) применяются в ритейле?
25. Какие методы используются для прогнозирования временных рядов?
26. На какие две группы делятся методы ИАД по принципу работы с данными?
27. Какие алгоритмы относятся к методам рассуждений на прецедентах (k-NN, Lazy-Learning)?
28. В чем преимущество деревьев решений перед нейронными сетями?
29. Как ассоциативные правила (Apriori) применяются в ритейле?
30. Какие методы используются для прогнозирования временных рядов?
31. В чем принципиальное отличие логистической регрессии от линейной регрессии? В каких задачах Data Mining применяется каждая из них?
32. Как работает метод опорных векторов (SVM) для классификации? Какие ядра (kernel) наиболее часто используются и почему?

33. Какие метрики расстояния (евклидово, манхэттенское, косинусное) применяются в алгоритмах кластеризации? Как выбор метрики влияет на результат?
34. В чем преимущество иерархической кластеризации перед k-means? Когда ее стоит использовать?
35. Как алгоритм Apriori находит частые наборы itemset'ов? Какие параметры (support, confidence, lift) критичны для интерпретации правил?
36. Почему ансамблевые методы (Random Forest, Gradient Boosting) часто outperform одиночные модели? Как они сочетаются с деревьями решений?
37. Почему понимание задачи – первый и важнейший этап анализа?
38. Какие шаги включает подготовка данных (препроцессинг)?
39. Как происходит проверка качества моделей (валидация)?
40. Почему интерпретация результатов требует участия аналитика?
41. Приведите пример практического применения Data Mining в бизнесе (например, прогнозирование оттока клиентов).
42. В чем заключаются ключевые различия между Deductor и RapidMiner с точки зрения функционала для Data Mining?
43. Какие типы данных (табличные, транзакционные, временные ряды) поддерживают обе платформы?
44. Какие методы трансформации данных (нормализация, дискретизация) доступны в этих платформах?
45. Как в этих платформах реализованы методы кластеризации (k-means, DBSCAN)?
46. Какие инструменты визуализации результатов (графики, тепловые карты) предлагают Deductor и RapidMiner?
47. С какими внешними системами (1C, SQL, Excel) интегрируются Deductor и RapidMiner?
48. Как в Deductor и RapidMiner реализована оптимизация глубины дерева (pruning) для предотвращения переобучения модели? Какие параметры настройки доступны для этого?
49. Как в Deductor и RapidMiner реализованы инструменты для очистки данных (обработка пропусков, выбросов)?
50. Какие алгоритмы классификации (деревья решений, SVM, нейросети) есть в Deductor и RapidMiner?
51. Как платформы помогают интерпретировать модели (например, объяснение правил ассоциации или важности признаков)?
52. Приведите пример реальной задачи (например, прогнозирование продаж), где Deductor или RapidMiner дают преимущество перед другими инструментами.