

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ
Директор Таганрогского института
имени А. П. Чехова (филиала)
РГЭУ (РИНХ)
_____ С. А. Петрушенко
«20» мая 2025 г.

**Рабочая программа дисциплины
Интеллектуальный анализ данных**

Направление подготовки
09.04.03 Прикладная информатика

Направленность (профиль) программы магистратуры
09.04.03.02 Информационные системы и анализ больших данных

Для набора 2025 года

Квалификация
Магистр

КАФЕДРА информатики**Распределение часов дисциплины по семестрам / курсам**

Семестр (<Курс>.<Семестр на курсе>)	3 (2.1)		Итого	
	14 1/6			
Неделя	14 1/6			
Вид занятий	УП	РП	УП	РП
Лекции	14	14	14	14
Лабораторные	28	28	28	28
Итого ауд.	42	42	42	42
Контактная работа	42	42	42	42
Сам. работа	66	66	66	66
Часы на контроль	36	36	36	36
Итого	144	144	144	144

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 28.02.2025 протокол № 9.

Программу составил(и): канд. экон. наук, Доц., Тюшняков В.Н.

Зав. кафедрой: Тюшнякова И.А.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	формирование у обучающихся системного представления о технологии интеллектуального анализа данных, основных методах, моделях и средствах Data Mining, их практическом применении для решения прикладных задач, развитие навыков анализа данных различной природы, разработки, адаптирования и тестирования компонентов прикладного программного обеспечения информационных систем
-----	---

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ОПК-2:	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач;
ОПК-2.1:	Знать современные интеллектуальные технологии для решения профессиональных задач;
ОПК-2.2:	Уметь обосновывать выбор современных интеллектуальных технологий и программной среды при разработке оригинальных программных средств для решения профессиональных задач
УК-1:	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий
УК-1.1:	Знать процедуры критического анализа, методики анализа результатов исследования и разработки стратегий проведения исследований, организации процесса принятия решения.
УК-1.2:	Уметь принимать конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий.
УК-1.3:	Владеть методами установления причинно-следственных связей и определения наиболее значимых среди них; методиками постановки цели и определения способов ее достижения; методиками разработки стратегий действий при проблемных ситуациях.

В результате освоения дисциплины обучающийся должен:

Знать:

методы системного анализа и математического моделирования, применяемые в исследовании организационно-технических и экономических процессов; основы построения моделей интеллектуального анализа данных с использованием различных методов и алгоритмов для обнаружения в наборах данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний (соотнесено с индикатором УК-1.1)
современные интеллектуальные технологии для решения профессиональных задач (соотнесено с индикатором ОПК-2.1)

Уметь:

осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (соотнесено с индикатором УК-1.2)
строить модели Data Mining в соответствии с установленными требованиями, визуализировать и интерпретировать результаты анализа для поддержки принятия решений в различных сферах человеческой деятельности (соотнесено с индикатором ОПК-2.2)

Владеть:

навыками применения методов системного анализа и математического моделирования в процессе исследования организационно-технических и экономических систем; навыками проверки полученных на этапе построения моделей правил и зависимостей в наборах данных (соотнесено с индикатором УК-1.3)

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Интеллектуальный анализ данных

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
1.1	Введение в интеллектуальный анализ данных. Базовые понятия	Лекционные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.2	Введение в интеллектуальный анализ данных. Базовые понятия и основные задачи интеллектуального анализа данных	Самостоятельная работа	3	6	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3

1.3	Основные этапы проведения интеллектуального анализа данных. Подготовка исходных данных	Лекционные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.4	Этапы проведения интеллектуального анализа данных	Самостоятельная работа	3	10	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.5	Этапы проведения интеллектуального анализа данных	Лабораторные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.6	Инструментальные средства интеллектуального анализа данных	Лекционные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.7	Интеллектуальный анализ данных в СУБД на примере Microsoft SQL Server	Лабораторные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.8	Настройки интеллектуального анализа данных для Microsoft Office	Лабораторные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.9	Анализ точности прогноза и использование модели интеллектуального анализа	Лабораторные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.10	Алгоритмы интеллектуального анализа данных	Самостоятельная работа	3	10	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.11	Построение модели кластеризации, трассировка и перекрестная проверка	Лабораторные занятия	3	4	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.12	Классификация методов кластеризации	Лекционные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.13	Структуры интеллектуального анализа данных.	Самостоятельная работа	3	12	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2

					УК-1.3
1.14	Создание структуры и модели интеллектуального анализа. Задача кластеризации	Лабораторные занятия	3	4	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.15	Практика применения интеллектуального анализа данных	Лекционные занятия	3	4	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.16	Модели интеллектуального анализа данных	Самостоятельная работа	3	10	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.17	Процесс обнаружения знаний. Инструментальные средства интеллектуального анализа данных	Самостоятельная работа	3	10	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.18	Задача классификации. Создание структуры и моделей интеллектуального анализа. Сравнение точности моделей	Лабораторные занятия	3	4	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.19	Классификация инструментов Data Mining. Области применения Data Mining	Лекционные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.20	Data Mining, OLAP и хранилища данных	Самостоятельная работа	3	8	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.21	Просмотр моделей интеллектуального анализа (дерева решений, упрощенный алгоритма Байеса, нейронные сети). Написание «одноэлементных» прогнозирующих запросов	Лабораторные занятия	3	4	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.22	Интеллектуальный анализ текста. Web mining	Лабораторные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.23	Рекомендательные системы. Алгоритмы коллаборативной фильтрации	Лабораторные занятия	3	2	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1 УК-1.2 УК-1.3
1.24	Подготовка к промежуточной аттестации	Экзамен	3	36	ОПК-2 УК-1 ОПК-2.1 ОПК-2.2 УК-1.1

					УК-1.2 УК-1.3
--	--	--	--	--	------------------

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущего контроля и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Учебные, научные и методические издания

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
1	Нестеров С. А.	Интеллектуальный анализ данных средствами MS SQL Server 2008	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	http://biblioclub.ru/index.php?page=book&id=429083
2	Барсегян А., Куприянов М., Степаненко В., Холод И.	Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP. 2 изд.	Санкт-Петербург: БХВ-Петербург, 2008	https://ibooks.ru/reading.php?short=1&productid=335156
3	Пальмов, С. В.	Интеллектуальный анализ данных: учебное пособие	Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017	http://www.iprbookshop.ru/75376.html

5.1. Учебные, научные и методические издания

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
1	Чубукова И. А.	Data Mining: учебное пособие	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ) Бином. Лаборатория знаний, 2008	http://biblioclub.ru/index.php?page=book&id=233055
2	Воронов, В. И., Воронова, Л. И., Усачев, В. А.	Data Mining - технологии обработки больших данных: учебное пособие	Москва: Московский технический университет связи и информатики, 2018	http://www.iprbookshop.ru/81324.html
3	Воронова, Л. И., Воронов, В. И.	Machine Learning: регрессионные методы интеллектуального анализа данных: учебное пособие	Москва: Московский технический университет связи и информатики, 2018	http://www.iprbookshop.ru/81325.html

5.2. Профессиональные базы данных и информационные справочные системы

Научная электронная библиотека: <https://www.elibrary.ru>

Лекционный материал "Наука о данных" (Data Science Lectures): <https://www3.cs.stonybrook.edu/~skiena/data-manual/lectures/>

Документация и видеоуроки по Orange data mining: <https://orangedatamining.com/docs/>

5.3. Перечень программного обеспечения

OpenOffice

5.4. Учебно-методические материалы для обучающихся с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор;
- экран / интерактивная доска.

Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет, и/или в специализированных лабораториях, предусмотренных образовательной программой.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1 Показатели и критерии оценивания компетенций:

1.2

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ОПК-2: Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач			
<i>Знать:</i> способы решения прикладных задач с использованием современных информационно-коммуникационных технологий	Осуществление поиска и сбора необходимой литературы, изучение основной и дополнительной литературы, подготовка доклада	соответствие проблеме исследования; полнота и содержательность ответа; умение приводить примеры; умение отстаивать свою позицию; умение пользоваться дополнительной литературой; соответствие представленной информации материалам учебной литературы, сведениям из информационных ресурсов Интернет	ВЭ – вопросы к экзамену (вопросы 1-30), Д – доклад (1-25)
<i>Уметь:</i> анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности	Выполнение индивидуального домашнего задания, подготовка доклада, использование современных инструментальных средств	достоверность решения заданий с помощью программных средств, умение пользоваться инструментальными средствами и электронными ресурсами	ЛЗ- лабораторные задания (1-8), ИЗ- индивидуальное задание(1-4), Д – доклад (1-25)
<i>Владеть:</i> навыками научного поиска и практической работы с информационными источниками; методами принятия решений	Выполнение индивидуального домашнего задания, подготовка доклада, использование современных инструментальных средств	достоверность решения заданий с помощью программных средств, соответствие проблеме исследования, умение пользоваться инструментальными средствами и электронными ресурсами	ЛЗ- лабораторные задания (1-8), ИЗ- индивидуальное задание(1-4), Д – доклад (1-25)

УК-1: Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий			
<p><i>Знать:</i> методы системного анализа и математического моделирования, применяемые в исследовании организационно-технических и экономических процессов; основы построения моделей интеллектуального анализа данных с использованием различных методов</p>	<p>Осуществление поиска и сбора необходимой литературы, изучение основной и дополнительной литературы, подготовка доклада</p>	<p>соответствие проблеме исследования; полнота и содержательность ответа; умение приводить примеры; умение отстаивать свою позицию; умение пользоваться дополнительной литературой; соответствие представленной информации материалам учебной литературы, сведениям из информационных ресурсов Интернет</p>	<p>ВЭ – вопросы к экзамену (вопросы 1-30), Д – доклад (1-25)</p>
<p><i>Уметь:</i> осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач; строить модели Data Mining в соответствии с установленными требованиями, визуализировать и интерпретировать результаты анализа для поддержки принятия решений в различных сферах человеческой деятельности</p>	<p>Выполнение индивидуального домашнего задания, подготовка доклада, использование современных инструментальных средств</p>	<p>достоверность решения заданий с помощью программных средств, умение пользоваться инструментальными средствами и электронными ресурсами</p>	<p>ЛЗ- лабораторные задания (1-8), ИЗ- индивидуальное задание(1-4), Д – доклад (1-25)</p>
<p><i>Владеть:</i> навыками применения методов системного анализа и математического моделирования в процессе исследования организационно-технических и экономических систем; навыками проверки полученных на этапе построения моделей правил и зависимостей в наборах данных</p>	<p>Выполнение индивидуального домашнего задания, подготовка доклада, использование современных инструментальных средств</p>	<p>достоверность решения заданий с помощью программных средств, соответствие проблеме исследования, умение пользоваться инструментальными средствами и электронными ресурсами</p>	<p>ЛЗ- лабораторные задания (1-8), ИЗ- индивидуальное задание(1-4), Д – доклад (1-25)</p>

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

- 84-100 баллов (оценка «отлично»);
- 67-83 баллов (оценка «хорошо»);
- 50-66 баллов (оценка «удовлетворительно»);
- 0-49 баллов (оценка «неудовлетворительно»).

2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Вопросы к экзамену

по дисциплине «Интеллектуальный анализ данных»

1. Базовые понятия и основные задачи интеллектуального анализа данных.
2. Эволюция задач сбора и обработки информации.
3. Архитектуры данных: Базы данных и модели данных.
4. Свойства и виды данных. Источники данных.
5. Этапы проведения интеллектуального анализа данных.
6. Сбор данных. Визуализация данных.
7. Очистка данных. Обнаружение вбросов.
8. Подготовка исходных данных в процессе интеллектуального анализа данных.
9. Аналитика больших данных.
10. Понятия Data Mining и Big Data.
11. DataMining-технологии. Области применения Data Mining.
12. Требования, предъявляемые к новым знаниям. Задачи DataMining.
13. Этапы CRISP-DM.
14. Модели DataMining: классификация, кластеризация, поиск ассоциативных правил, итоговые модели, регрессионные модели.
15. Алгоритмы интеллектуального анализа данных.
16. Структуры интеллектуального анализа данных.
17. Модели интеллектуального анализа данных.
18. Процесс обнаружения знаний.
19. Инструментальные средства интеллектуального анализа данных.
20. Построение модели кластеризации в Orange data mining.
21. Построение модели линейной регрессии в Orange data mining.
22. Интеллектуальный анализ текста. Web mining.
23. Анализ данных в G N-grams.
24. Методика обнаружения нового знания в хранилищах данных.
25. Классификация методов кластеризации.
26. Визуальный анализ данных.
27. Характеристики средств визуализации данных.
28. Сферы применения технологии интеллектуального анализа данных.
29. OLAP и хранилища данных.
30. Машинное обучение и анализ данных.

Экзаменационное задание (билет) включает 2 теоретических вопроса (формируются из представленных вопросов к экзамену) и 1 практическое задание (формируются из перечня лабораторных заданий, представленных в разделе «Лабораторные задания»).

Критерии оценивания:

Максимальное количество баллов за экзаменационное задание – 100 (50 баллов максимально за теоретические вопросы, 50 баллов максимально за практическое задание).

Критерии оценивания одного теоретического вопроса.

Критерии оценивания теоретического вопроса	Баллы
Изложенный материал фактически верен, наличие глубоких исчерпывающих знаний; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе	21-25
Наличие твердых и достаточно полных знаний, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, неуверенность и неточность ответов на дополнительные и наводящие вопросы	17-20
Неполный ответ на вопросы; затрудняется ответить на дополнительные вопросы	1-16
Ответ не связан с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы	0
<i>Максимальный балл за ответ на теоретический вопрос</i>	<i>25</i>

Критерии оценивания одного практического задания.

Критерии оценивания практического задания	Баллы
Задание выполнено в полном объеме, в представленном решении обоснованно получены правильные ответы, проведен анализ, дана грамотная интерпретация полученных результатов, сделаны выводы	41-50
Задание выполнено в полном объеме, но при анализе и интерпретации полученных результатов допущены незначительные ошибки, выводы – достаточно обоснованы, но неполны	26-40
Задание выполнено не в полном объеме, при анализе и интерпретации полученных результатов допущены ошибки, выводы – но неполные или отсутствуют	1-25
Задание выполнено полностью неверно или отсутствует решение	0
<i>Максимальный балл за выполнение практического задания</i>	<i>50</i>

Итоговая оценка формируется из суммы набранных баллов за выполнение экзаменационного задания (2 теоретических вопроса и 1 практическое задание) и соответствует шкале:

- 84-100 баллов (оценка «отлично»);

- 67-83 баллов (оценка «хорошо»);
- 50-66 баллов (оценка «удовлетворительно»);
- 0-49 баллов (оценка «неудовлетворительно»).

Лабораторные задания

по дисциплине «Интеллектуальный анализ данных»

Тематика лабораторных работ

1. Этапы проведения интеллектуального анализа данных
2. Orange data mining. Анализ данных о ценах на недвижимость. Анализ датасета «Housing»
3. Графики частотности языковых единиц на массиве печатных источников (N-grams)
4. Работа с различными источниками данных в Deductor Academic (Basegroup)
5. Предобработка, трансформация, исследование наборов данных в Loginom Community
6. Анализ набора данных «wine.csv» для построения регрессивной модели предсказания бедующей цены
7. Интеллектуальный анализ текста. Анализ текстовых сообщений в социальных сетях
8. Интеллектуальный анализ изображений. Работа с библиотекой Image Analytics в Orange data mining
9. Анализ точности прогноза и использование модели интеллектуального анализа
10. Создание структуры и модели интеллектуального анализа. Задача кластеризации. Кластеризация в Orange

Критерии оценивания:

Максимальное количество баллов – 50 (за 10 лабораторных заданий).

Для каждого лабораторного задания:

Критерий оценивания	Баллы
Задание выполнено в полном объеме, в представленном решении обоснованно получены правильные ответы, проведен анализ, дана грамотная интерпретация полученных результатов, сделаны выводы	5
Задание выполнено в полном объеме, но при анализе и интерпретации полученных результатов допущены незначительные ошибки, выводы – достаточно обоснованы, но неполны	3-4
Задание выполнено частично, отсутствует анализ и интерпретация полученных результатов допущены значительные ошибки, отсутствуют выводы	1-2
Задание выполнено полностью неверно или отсутствует решение	0
<i>Максимальный балл за одно лабораторное задание</i>	5

Перечень тем для докладов

по дисциплине «Интеллектуальный анализ данных»

1. Специфика работы с информацией. Свойства информации. Информация и данные.
2. Развитие систем хранения и обработки данных. Системы оперативной обработки информации.
3. Системы консолидации и аналитической обработки информации.
4. Предварительная обработка данных.
5. Виды и методы анализа данных. Роль анализа данных в сфере бизнеса.
6. Иерархическая модель данных, условия целостности иерархической модели данных.
7. Сетевая модель данных, условия целостности сетевой модели данных.
8. Реляционная модель данных, реляционные базы данных.
9. Реляционные, многомерные, гибридные, виртуальные хранилища данных, их преимущества и недостатки.
10. Концепция многомерного представления данных.
11. Методы обработки агрегированных данных.
12. Визуализация данных: визуализаторы OLAP-анализа. Сферы применения OLAP-технологии.
13. Стадии DataMining.
14. Сферы применения технологии интеллектуального анализа данных.
15. Модели интеллектуального анализа данных.
16. Деревья принятия решений.
17. Модели и инструменты реализации высокопроизводительных вычислений.
18. Инструментальные средства интеллектуального анализа данных.
19. Классификация методов интеллектуального анализа данных.
20. Качество кластеризации и выбор наилучшего решения.
21. Формализация многомерного представления данных: метки, иерархии, ячейки, меры.
22. Операции над многомерными данными. Методы обработки агрегированных данных.
23. Интеллектуальный анализ данных в СУБД Microsoft SQL Server.
24. Иерархические алгоритмы. Неиерархические алгоритмы.
25. Индуктивные и дедуктивные подходы Data Mining.

Критерии оценивания:

- 15-25 баллов - выставляется обучающемуся, если тема соответствует содержанию доклада; основные понятия проблемы изложены верно; сделаны обобщения и сопоставления различных точек зрения по рассматриваемому вопросу; сделаны и аргументированы основные выводы, доклад сопровождается разработанной мультимедийной презентацией;

- 0-14 баллов - выставляется обучающемуся, если содержание не соответствует теме; нет ссылок на использованные источники; тема не полностью раскрыта; нет выводов.

Темы индивидуального задания

по дисциплине Интеллектуальный анализ данных

Из 1.

Классификация на основе деревьев решений

Цель работы: освоить алгоритмы классификации интеллектуального анализа данных, научиться строить деревья решений и классификационные правила в среде Deductor Studio.

Пример задания:

Классифицировать заемщиков путем построения дерева решений и оценить их кредитоспособность.

Последовательность выполнения задания:

1. Импортировать данные из текстового файла в Deductor Studio.
2. Вызвать Мастер обработки и выбрать метод обработки "Дерево решений". Настроить назначения столбцов. Сделать поле "Выдача кредита" – выходным, "Код клиента" и "Номер паспорта" – информационными, остальные показатели – входными. Настроить параметры процесса обучения. Выбрать способы визуализации результатов: Дерево решений, Правила, Что-если, Таблица сопряженности, Значимость атрибутов.
3. Проанализировать данные при помощи выбранных визуализаторов. Определить, сколько примеров было распознано неверно. Оценить качество дерева решений. Определить, какие факторы являются более важными, какие второстепенными, насколько сильно каждый входной фактор влияет на результат. Просмотреть все правила, согласно которым можно отнести заемщиков к тому или иному классу. Определить, что именно оказывает влияние на принятие решения о выдаче кредита, каковы поддержка и достоверность правил.
4. Оценить кредитоспособность заемщиков пенсионного возраста. Для этого предварительно отобрать соответствующих клиентов из исходного набора данных, используя инструмент «Фильтрация». Затем построить дерево решений. Просмотреть полученные результаты. Сформулировать заключение о достоинствах деревьев решений для проведения подобного анализа.

ИЗ 2.

Поиск ассоциативных правил на примере анализа рыночной корзины

Цель работы: освоить алгоритмы поиска ассоциативных правил интеллектуального анализа данных, научиться строить ассоциативные правила в среде Deductor Studio.

Пример задания:

Выполнить поиск ассоциативных зависимостей на основе данных о покупках продуктов нескольких групп в магазине с целью последующего применения результатов для стимулирования продаж.

Последовательность выполнения задания:

1. Импортировать данные из текстового файла в Deductor Studio.
2. Вызвать Мастер обработки и выбрать метод обработки "Ассоциативные правила". Указать столбец для идентификации транзакции: "Номер операции". В качестве элемента транзакции определить: "Наименование товара".
3. Настроить параметры построения ассоциативных правил: минимальную и максимальную поддержку 20% и 80% и достоверность 60% и 90%.
4. После завершения процесса поиска просмотреть полученные результаты, используя специальные визуализаторы: "Популярные наборы", "Правила", "Дерево правил", "Что-если".
5. Интерпретировать все правила. Определить поддержку и достоверность правил.
6. Проанализировать потребительскую корзину. Ответить на вопрос: что, возможно, забыл приобрести покупатель, если он уже приобрел печенье и мармелад. Проанализировать предпочтения покупателей магазина. Определить, какие товары пользуются в данной товарной точке наибольшим спросом. Составить списки товаров, которые следует размещать на полках магазина рядом друг с другом.

ИЗ 3.

Кластеризация с помощью самоорганизующейся карты

Цель работы: освоить алгоритмы кластеризации интеллектуального анализа данных с использованием самоорганизующейся карты в среде Deductor Studio.

Пример задания:

Выполнить кластеризацию кредитных организаций путем построения самоорганизующейся карты Кохонена.

Последовательность выполнения задания:

1. Импортировать данные из текстового файла в Deductor Studio.
2. Вызвать Мастер обработки и выбрать из списка метод обработки "Карта Кохонена". Задать назначения столбцов. Столбцу "Прибыль" указать назначение «Выходной». Столбцы: "Активы", "Собственные активы", "Банковские активы", "Денежные средства", "Количество филиалов" определить как входные. Остальные столбцы сделать информационными. Указать количество кластеров: 3.
3. В списке визуализаторов выбрать Карту Кохонена, а также "Что-если". Указать поля, которые необходимы для отображения (все входные столбцы, прибыль и кластеры).
4. Получить карту Кохонена. Проследить зависимости между различными показателями. Выявить скрытые закономерности. Ответить на следующие вопросы. В какие кластеры попали наиболее прибыльные организации? Какие черты характерны для этих организаций? Какие факторы практически не оказывают влияние на прибыльность?
5. Используя диаграмму "Что-если", решить задачу прогнозирования. Ввести данные для прогноза и определить, к какому кластеру принадлежит данная точка. Подсчитать среднее по результирующему полю всех точек кластера.

Из 4.

Прогнозирование временных рядов с помощью нейронной сети и линейной регрессии

Цель работы: освоить аналитические алгоритмы прогнозирования временных рядов с использованием нейросети и линейной регрессии в среде Deductor Studio.

Пример задания:

Выполнить прогнозирование объема производства продукции на два года вперед с помощью нейросети и линейной регрессии.

Последовательность выполнения задания:

1. Импортировать данные из текстового файла в Deductor Studio.
2. Запустить Мастер обработки и выбрать в нем инструмент "Редактирование выбросов". Для удаления аномалий из исходного набора данных выбрать поле для обработки "Объем".
3. Осуществить спектральную обработку временного ряда. Для сглаживания данных выбрать поле "Объем" и указать ему тип обработки "Вычитание шума", степень подавления "Малая". Выбрать способы визуализации: таблицу и диаграмму. Просмотреть полученный результат на диаграмме. Сделать заключение об общей тенденции динамического ряда.
4. Преобразовать данные к скользящему окну. Назначить в качестве используемого столбца поле "Объем" и задать глубину погружения – 3 (считаем, что объем производства в следующем году зависит от объема в предыдущие годы). Просмотреть полученный результат трансформации данных.
5. Выполнить прогнозирование временного ряда на основе нейронной сети. Установить в качестве входных полей: "Объем-3", "Объем-2" и "Объем-1" данные по объему производства за три предыдущих года, а в качестве выходного "Объем". Поле "Годы" сделать информационным. После построения модели отобразить нейронную сеть в виде графа, диаграммы и диаграммы рассеяния (для просмотра качества обучения).
6. Просмотреть полученный граф нейросети, диаграмму и диаграмму рассеяния. Сделать выводы о качестве обучения нейросети.
7. Для построения требуемого прогноза выбрать инструмент «Прогнозирование» мастер обработки. Поля "Объем-3", "Объем-2" и "Объем-1" сделать входными, а поле

"Объем" – выходным. Горизонт прогноза – 2. В качестве визуализатора выбрать диаграмму прогноза.

8. Осуществить прогнозирование временного ряда на основе линейной регрессии. Указать в качестве входных полей: "Объем-3", "Объем-2" и "Объем-1", а в качестве выходного "Объем". После построения модели выбрать способы отображения данных: коэффициенты регрессии, диаграмма, диаграмма рассеяния. Просмотреть полученные результаты и оценить точность построения модели.

9. Выполнить прогнозирование. Настроить связи столбцов для прогнозирования. Поля "Объем-3", "Объем-2" и "Объем-1" – входные, поле "Объем" – выходное. Горизонт прогноза – 2. Просмотреть диаграмму прогноза.

10. Сравнить прогноз, полученный с помощью линейной регрессии, с прогнозом на основе нейронной сети. Сделать соответствующие выводы.

Критерии оценивания:

- 21-25 баллов - выставляется обучающему, если все пункты задания выполнены безошибочно;

- 17-20 баллов - выставляется обучающему, если задание, большей частью, выполнено или выполнено с ошибками.

- 0-16 баллов - выставляется обучающему, если задание выполнено с существенными ошибками, выполнено наполовину или не выполнено содержание не соответствует теме; нет ссылок на использованные источники; тема не полностью раскрыта; нет выводов.

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения обучающихся до промежуточной аттестации.

Промежуточная аттестация проводится в форме экзамена, это аттестация в период сессии, которая проводится в соответствии с действующим в РГЭУ (РИНХ) Положением о курсовых, экзаменах и зачётах.

Экзамен проводится по расписанию промежуточной аттестации в компьютерном классе. Количество вопросов в экзаменационном задании – 3 (2 теоретических вопроса и 1 практическое задание). Объявление результатов производится в день экзамена. Результаты аттестации заносятся в ведомость и зачетную книжку обучающегося.

Обучающиеся, не прошедшие промежуточную аттестацию по графику промежуточной аттестации, должны ликвидировать задолженность в установленном порядке.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- лабораторные работы.

Важным условием успешного освоения дисциплины «Интеллектуальный анализ данных» является создание системы правильной организации труда, позволяющей распределить учебную нагрузку равномерно в соответствии с графиком образовательного процесса. В ходе лекционных занятий рассматриваются теоретические основы, понятия, методы и проблемы интеллектуального анализа данных. Знакомство с дисциплиной происходит уже на первой лекции, где от обучающегося требуется не просто внимание, но и самостоятельное оформление конспекта. При работе с конспектом лекций необходимо учитывать тот фактор, что одни лекции дают ответы на конкретные вопросы темы, другие – лишь выявляют взаимосвязи между явлениями, помогая обучающемуся понять глубинные процессы развития изучаемого предмета.

Изучение дисциплины проходит с акцентом на лабораторные работы. В ходе лабораторных занятий развиваются умения и навыки использования различных инструментов интеллектуального анализа данных. Все задания к лабораторным занятиям рекомендуется выполнять непосредственно после соответствующей темы лекционного курса, что способствует лучшему усвоению материала, позволяет своевременно выявить и устранить «пробелы» в знаниях, систематизировать ранее пройденный материал, на его основе приступить к овладению новыми знаниями и навыками.

В процессе подготовки к лабораторным занятиям, обучающимся необходимо обратить особое внимание на самостоятельное изучение рекомендованной литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами периодических изданий и Интернета является наиболее эффективным методом получения дополнительных знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме.

Лабораторные занятия проводятся в компьютерных классах с применением специально разработанных учебно-методических материалов, в которых изложены подробные методические рекомендации по изучению каждой темы и выполнению заданий. Наличие таких учебно-методических и дидактических материалов позволяет каждому

обучающемуся работать в своем индивидуальном темпе, а также дополнительно прорабатывать изучаемый материал во время самостоятельных занятий.

Перед выполнением лабораторной работы требуется получить вариант задания. Далее необходимо ознакомиться с заданием. Выполнение лабораторной работы следует начать с изучения теоретических сведений, которые приводятся в соответствующих методических указаниях. Результаты работы необходимо оформить в виде отчета. Лабораторная работа считается выполненной, если предоставлен отчет о результатах выполнения задания; проведена защита проделанной работы. Защита проводится в два этапа: демонстрация результатов выполнения задания; ответы на вопросы из перечня контрольных вопросов, который приводится в задании на лабораторную работу.

Для успешного овладения предлагаемым курсом обучающийся должен обладать определённой информационной культурой: навыками работы с литературой, умением определять и находить информационные ресурсы, соответствующие целям и задачам образовательного процесса, получать к ним доступ и использовать в целях повышения эффективности своей профессиональной деятельности. При изучении данного курса необходимо использовать компьютер, изучать дополнительные информационные ресурсы.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации обучающиеся могут воспользоваться электронно-библиотечными системами.