

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 18.09.2023 11:08:58
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации

ФГБОУ ВО «Тверской государственный университет»

Утверждаю:

Руководитель ООП:


Г.М. Соломаха
«25» 09 2023 г.

Рабочая программа дисциплины (с аннотацией)

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Направление подготовки

09.04.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

Профиль

Прикладная информатика в аналитической экономике

Для студентов 1 курса, очная форма обучения

Составитель: к.ф.-м.н. Гришина Е.Н.

Тверь 2023

I. Аннотация

1. Цель и задачи дисциплины

Целью освоения дисциплины является рассмотрение вопросов обработки и анализа данных.

Задачами освоения дисциплины являются изучение способов хранения и представления информации, изучение современных методов обработки и анализа данных, в том числе методов и моделей интеллектуального анализа данных.

2. Место дисциплины в структуре ООП

Дисциплина «Интеллектуальный анализ данных» относится к части учебного плана, формируемой участниками образовательных отношений.

Для успешного освоения дисциплины обучающиеся должны обладать следующими входными знаниями: хранение и представление информации, проектирование баз данных, проектирование хранилищ данных, владение базовыми методами математической статистики, методами многомерного статистического анализа, подходами к построению и проектированию информационных систем, систем поддержки принятия решений.

Освоение данной дисциплины необходимо как предшествующее для следующих дисциплин: «Интеллектуальные информационные технологии», «Моделирование неопределенности в задачах оптимизации и принятия решений».

3. Объем дисциплины: 3 зачетных единиц, 108 академических часов, в том числе:

контактная аудиторная работа: лекции 15 часов, практические занятия 15 часов;

контактная внеаудиторная работа: контроль самостоятельной работы 0, в том числе курсовая работа 0;

самостоятельная работа: 78 часов, в том числе контроль 0.

4. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
<p>ПК-1 Способен использовать и развивать методы научных исследований и инструментарий в области проектирования и управления информационными системами в экономике</p>	<p>ПК-1.1 Выявляет и формализует требования к информационным системам конкретного назначения, обеспечивающим информатизацию экономических процессов организаций и предприятий различного уровня</p> <p>ПК-1.2 Проводит формальную постановку прикладных задач макро и микроэкономики и задач разработки прикладного алгоритмического обеспечения информационных систем в экономике</p> <p>ПК-1.3 Предлагает и реализует метод решения задач в области проектирования и управления информационными системами в экономике</p> <p>ПК-1.4 Осуществляет аттестацию результатов научных исследований с использованием математического моделирования исследуемых задач</p>
<p>ПК-3 Способен адаптировать и развивать современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения прикладных задач различных классов и создания ИС</p>	<p>ПК-3.1 Проводит декомпозицию процесса автоматизации и информатизации прикладной задачи</p> <p>ПК-3.2 Решает отдельные подзадачи декомпозированного процесса автоматизации и информатизации</p> <p>ПК-3.3 Проектирует архитектуру программного обеспечения прикладной ИС</p>

5. Форма промежуточной аттестации и семестр прохождения зачет.

6. Язык преподавания русский.

II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)			Самостоятельная работа, в том числе Контроль (час.)
		Лекции	Практические занятия	Контроль самостоятельной работы (в том числе курсовая работа)	
1. Интеллектуальный анализ данных	8	1	1	—	6
2. Объекты и их атрибуты	12	2	2	—	8
3. Подготовка данных	12	2	2	—	8
4. Хранилище данных	12	2	2	—	8
5. Задача классификации и регрессии	16	2	2	—	12
6. Задача поиска ассоциативных правил	16	2	2	—	12
7. Задача кластеризации	16	2	2	—	12
8. Анализ текстовой информации	16	2	2	—	12
ИТОГО	108	15	15	—	78

III. Образовательные технологии

Учебная программа – наименование разделов и тем <i>(в строгом соответствии с разделом II РПД)</i>	Вид занятия	Образовательные технологии
1. Интеллектуальный анализ данных	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
2. Объекты и их атрибуты	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
3. Подготовка данных	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
4. Хранилище данных	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
5. Задача классификации и регрессии	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
6. Задача поиска ассоциативных правил	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
7. Задача кластеризации	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.
8. Анализ текстовой информации	Лекции, практические занятия	1. Изложение теоретического материала; 2. Решение задач.

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

1. Типовые контрольные задания для проверки уровня сформированности компетенции ПК-1: способность использовать и развивать методы научных исследований и инструментарий в области проектирования и управления информационными системами в экономике

Этап формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков (примеры)	Показатели и критерии оценивания компетенции, шкала оценивания
ПК-1.1 Выявляет и формализует требования к информационным системам конкретного назначения, обеспечивающим информатизацию экономических процессов организаций и предприятий различного уровня	Подготовить доклад на тему: 1. Алгоритм ID3 и его практическое применение; 2. Алгоритм k-means и его практическое применение.	Доклад содержит полную, исчерпывающую информацию – 20 баллов. Тема доклада раскрыта не полностью – 10-18 баллов. Доклад содержит грубые ошибки, говорящие о непонимании сути- 0 баллов.
ПК-1.2 Проводит формальную постановку прикладных задач макро и микроэкономики и задач разработки прикладного алгоритмического обеспечения информационных систем в экономике	Подготовить доклад на тему: 1. Нейронные сети и их практическое применение; 2. Методы визуализации в Data Mining.	Доклад содержит полную, исчерпывающую информацию – 20 баллов. Тема доклада раскрыта не полностью – 10-18 баллов. Доклад содержит грубые ошибки, говорящие о непонимании сути- 0 баллов.
ПК-1.3 Предлагает и реализует метод решения задач в области проектирования и управления информационными системами в экономике	Опишите и реализуйте: 1. Алгоритм Naïve Bayes для задачи с 4 независимыми переменными и 1 зависимой переменной; 2. Алгоритм покрытия.	Работа выполнена в соответствии с методическими рекомендациями – 12 баллов. Студент ответил на 2-3 дополнительных вопроса – 4 балла.

Этап формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков (примеры)	Показатели и критерии оценивания компетенции, шкала оценивания
ПК-1.4 Осуществляет аттестацию результатов научных исследований с использованием математического моделирования исследуемых задач	Опишите и реализуйте: 1. Алгоритм Apriori ; 2. Один из иерархических алгоритмов кластеризации на выбор.	Работа выполнена в соответствии с методическими рекомендациями – 12 баллов. Студент ответил на 2-3 дополнительных вопроса – 4 балла.

2. Типовые контрольные задания для проверки уровня сформированности компетенции ПК-3: способность адаптировать и развивать современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения прикладных задач различных классов и создания ИС

Этап формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков (примеры)	Показатели и критерии оценивания компетенции, шкала оценивания
ПК-3.1 Знает основы проектирования и элементы архитектурных решений информационных систем	Подготовить доклад на тему: 1. Алгоритм ID3 и его практическое применение; 2. Алгоритм k-means и его практическое применение.	Доклад содержит полную, исчерпывающую информацию – 20 баллов. Тема доклада раскрыта не полностью – 10-18 баллов. Доклад содержит грубые ошибки, говорящие о непонимании сути- 0 баллов.

Этап формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков (примеры)	Показатели и критерии оценивания компетенции, шкала оценивания
ПК-3.2 Применяет в практической деятельности профессиональные стандарты в области информационных технологий	Опишите и реализуйте: 1. Алгоритм Naïve Bayes для задачи с 4 независимыми переменными и 1 зависимой переменной; 2. Алгоритм покрытия.	Работа выполнена в соответствии с методическими рекомендациями – 12 баллов. Студент ответил на 2-3 дополнительных вопроса – 4 балла.
ПК-3.3 Составляет техническое задание на разработку информационной системы	Опишите и реализуйте: 1. Алгоритм Apriori ; 2. Один из иерархических алгоритмов кластеризации на выбор.	Работа выполнена в соответствии с методическими рекомендациями – 12 баллов. Студент ответил на 2-3 дополнительных вопроса – 4 балла.

V. Учебно-методическое и информационное обеспечение дисциплины

1) Рекомендуемая литература

а) Основная литература

1. Нестеров, С.А. Интеллектуальный анализ данных средствами MS SQL Server 2008 / С.А. Нестеров. - 2-е изд., испр. - М.: Национальный Открытый Университет «ИНТУИТ», 2016. - 338 с.: ил. - Библиогр. в кн.; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=429083>
2. Интеллектуальные системы: учебное пособие / А. Семенов, Н. Соловьев, Е. Чернопрудова, А. Цыганков; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург: ОГУ, 2013. - 236 с.; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=259148>
3. Ярушкина Н. Г. Интеллектуальный анализ временных рядов: учебное пособие /Н.Г. Ярушкина, И.Г. Перфильева, Т.В. Афанасьева. - Москва; Москва: Издательский Дом "ФОРУМ": ООО "Научно-издательский центр ИНФРА-М", 2012. - 160 с. - ISBN 9785819904961. -Режим доступа: <http://znanium.com/go.php?id=249314>

4. Ясницкий, Л.Н. Интеллектуальные системы [Электронный ресурс]: учеб. — Электрон. дан. — Москва: Издательство "Лаборатория знаний", 2016. — 224 с. — Режим доступа: <https://e.lanbook.com/book/90254>

б) Дополнительная литература:

1. Богданова, С.В. Информационные технологии: учебное пособие для студентов высших учебных заведений / С.В. Богданова, А.Н. Ермакова; Министерство сельского хозяйства РФ, ФГБОУ ВПО Ставропольский государственный аграрный университет. - Ставрополь: Сервисшкола, 2014. - 211 с.: ил. - Библиогр. в кн.; [Электронный ресурс]. - Режим доступа с.: ил. - Библиогр. в кн.; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=277476>
2. Грошев, А.С. Информационные технологии: лабораторный практикум / А.С. Грошев. - 2-е изд. - М.; Берлин: Директ-Медиа, 2015. - 285 с.: ил., табл. - Библиогр. в кн. - ISBN 978-5-4475-5065-3; То же [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=434666>

2) Программное обеспечение

а) Лицензионное программное обеспечение:

<p>Помещение для самостоятельной работы обучающихся: Компьютерный класс факультета ПМиК № 4б (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)</p>	<p>Google Chrome – бесплатное ПО; JetBrains PyCharm Community Edition 4.5.3 – бесплатное ПО; JetBrains PyCharm Edu 3.0 – бесплатное ПО; Microsoft Office профессиональный плюс 2013 – Акт приема-передачи № 369 от 21 июля 2017; Microsoft SQL Server 2014 Express LocalDB - бесплатное ПО; Microsoft Visio Professional 2013 - Акт приема-передачи № 369 от 21 июля 2017; MS Visual Studio Ultimate 2013 с обновлением 4 - Акт предоставления прав № Tr035055 от 19.06.2017; MiKTeX 2.9 – бесплатное ПО; Notepad++ - бесплатное ПО; Python 3.1 pygame-1.9.1 - бесплатное ПО; Python 3.4 numpy-1.9.2 - бесплатное ПО; Python 3.4.3 - бесплатное ПО; Python 3.5.1 (Anaconda3 2.5.0 64-bit) - бесплатное ПО; MS Windows 10 Enterprise – Акт приема-передачи № 369 от 21 июля 2017.</p>
---	--

3) Современные профессиональные базы данных и информационные справочные системы

1. ЭБС «ZNANIUM.COM» www.znanium.com;
2. ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/>;
3. ЭБС «Лань» <http://e.lanbook.com>.

4) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

<http://www.intuit.ru/department/database/datamining/>

[http://www.machinelearning.ru/wiki/index.php?title=Машинное обучение %28курс лекций%2C К.В.Воронцов%29](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_%28курс_лекций%2C_К.В.Воронцов%29)

VI. Методические материалы для обучающихся по освоению дисциплины

1. Вопросы (тест для промежуточной аттестации)

1. Какие задачи решаются с помощью систем поддержки принятия решений:
 1. ввод данных;
 2. хранение данных;
 3. анализ данных.
2. Какая задача, решаемая с помощью систем поддержки принятия решений, является основной:
 1. ввод данных;
 2. хранение данных;
 3. анализ данных.
3. С помощью какого класса задач анализа осуществляется группирование и обобщение необходимых аналитику данных:
 1. информационно-поисковый;
 2. оперативно-аналитический;
 3. интеллектуальный.
4. С помощью какого класса задач анализа осуществляется поиск данных на основе заранее определенных запросов:
 1. информационно-поисковый;
 2. оперативно-аналитический;
 3. интеллектуальный.

5. С помощью какого класса задач анализа осуществляется поиск функциональных закономерностей в накопленных данных, построение моделей и правил:
 1. информационно-поисковый;
 2. оперативно-аналитический;
 3. интеллектуальный.
6. Какие подсистемы входят в обобщенную архитектуру системы поддержки принятия решений:
 1. информационно-поисковая система;
 2. подсистема ввода данных;
 3. подсистема хранения;
 4. подсистема анализа;
 5. подсистема отчетов.
7. В какой подсистеме осуществляется операционная обработка данных на основе использования СУБД:
 1. информационно-поисковая система;
 2. подсистема ввода данных;
 3. подсистема хранения;
 4. подсистема анализа;
 5. подсистема отчетов.
8. В какой подсистеме подсистемы анализа реализуются методы и алгоритмы добычи данных:
 1. подсистема информационно-поискового анализа;
 2. подсистема оперативного анализа;
 3. подсистема интеллектуального анализа.
9. В каких системах присутствие человеческого фактора при вводе данных повышает вероятность ошибочных данных и может создать локальные проблемы в системах:
 1. система поддержки принятия решений;
 2. OLTP – система;

3. система управления базами данных;
 4. экономическая информационная система;
 5. информационно-поисковая система.
10. К каким системам предъявляется требование обеспечения формирования произвольных запросов к базе данных:
1. система поддержки принятия решений;
 2. OLTP – система;
 3. система управления базами данных;
 4. экономическая информационная система;
 5. информационно-поисковая система.
11. Для каких систем является приоритетных высокая производительность и доступность данных:
1. система поддержки принятия решений;
 2. OLTP – система;
 3. система управления базами данных;
 4. экономическая информационная система;
 5. информационно-поисковая система.
12. Хранилище данных – это
1. предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений;
 2. модель некоторой предметной области, состоящая из связанных между собой данных об объектах, их свойствах и характеристиках;
 3. система, обладающая свойствами ввода, хранения и анализа данных, относящихся к некоторой предметной области, с целью поиска решений.
13. Выделите свойства хранилища данных:
1. предметная ориентация;
 2. интеграция;
 3. поддержка хронологии;

4. неизменяемость;
 5. допущение избыточности данных;
 6. приоритетность характеристик системы.
14. Основное назначение OLAP-систем:
1. обнаружение скрытых знаний и закономерностей;
 2. поддержка аналитической деятельности;
 3. поддержка запросов пользователей-аналитиков;
 4. ввод и хранение данных.
15. Интеллектуальный анализ данных – это
1. исследование и обнаружение в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны;
 2. технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений;
 3. предметно-ориентированный, интегрированный, изменчивый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений;
 4. модель некоторой предметной области, состоящая из связанных между собой данных об объектах, их свойствах и характеристиках.

2. Лабораторный практикум

1. Для указанных данных проведите классификацию с помощью:
 - a. алгоритма 1R;
 - b. алгоритма Naive Bayes;
 - c. алгоритма покрытия (Sequential Covering);
 - d. алгоритма ID3.

Пример данных для классификации:

ID	Транзакции
1	{ картофель, хлеб }
2	{ хлеб, помидоры, масло }

3	{ картофель, помидоры, масло, мясо }
4	{ картофель, масло, мясо }
5	{ картофель, хлеб, помидоры }
6	{ картофель, хлеб, помидоры, масло }
7	{ картофель, мясо }
8	{ картофель, хлеб, масло }
9	{ картофель, помидоры, масло }
10	{ картофель, хлеб, помидоры, масло, мясо }

2. Для указанных данных найдите все ассоциативные правила с помощью:
- алгоритма Apriori;
 - алгоритма FP-Growth.

Пример данных для поиска ассоциативных правил:

Расстояние	Стоимость	Время в пути	Транспорт
Близко	Недорого	Небольшое	Автобус
Близко	Недорого	Большое	Поезд
Близко	Средне	Небольшое	Машина
Близко	Дорого	Небольшое	Машина
Близко	Дорого	Большое	Поезд
Далеко	Недорого	Небольшое	Автобус
Далеко	Средне	Небольшое	Поезд
Далеко	Средне	Большое	Автобус
Далеко	Дорого	Небольшое	Машина
Далеко	Дорого	Большое	Поезд

3. Разделите указанные данные на кластеры с помощью:
- одного из иерархических алгоритмов;
 - алгоритма k-means.

Пример данных для кластеризации:

Точка	x	y
A	4	4
B	3	2
C	5	3
D	4	1
E	1	2
F	1	1
G	5	4
H	2	4

3. Темы докладов

1. Практическое применение интеллектуального анализа данных.
2. Информационное хранилище (витрины данных, информационное хранилище двухуровневой и трехуровневой архитектуры).
3. Модели данных (реляционная, сетевая, иерархическая модели данных).
4. Концепция многомерного представления данных.
5. Алгоритм ID3 и его практическое применение.
6. Алгоритм ECLAT и его практическое применение.
7. Алгоритм k-means и его практическое применение.
8. Анализ текстовой информации (Text Mining).
9. Визуальный анализ данных (Visual Mining).
10. Нейронные сети.

4. Вопросы для подготовки к зачету

1. Системы поддержки принятия решений. Основные понятия. Задачи СППР.
2. База данных как основа СППР.

3. Архитектура ССПР.
4. Хранилище данных. Концепция хранилища данных.
5. Организация хранилища данных. Концепция хранилища данных и анализ.
6. OLAP-системы. Основные понятия.
7. Концептуальное многомерное представление.
8. Архитектура OLAP-систем.
9. Интеллектуальный анализ данных. Основные понятия ИАД.
10. Задачи ИАД. Область практического применения ИАД.
11. Модели и методы ИАД.
12. Задача классификации и регрессии. Постановка задачи.
13. Методы (алгоритмы) решения задачи классификации и регрессии.
14. Представление результатов в задаче классификации и регрессии.
15. Задача поиска ассоциативных правил. Постановка задачи.
16. Методы (алгоритмы) решения задачи поиска ассоциативных правил.
17. Представление результатов в задаче поиска ассоциативных правил.
18. Задача кластеризации. Постановка задачи.
19. Методы (алгоритмы) решения задачи кластеризации.
20. Представление результатов в задаче кластеризации.
21. Визуальный анализ данных.
22. Методы визуализации.
23. Анализ текстов, классификация текстовых документов.
24. Методы кластеризации текстовых документов.
25. Средства анализа текстовой информации.

VII. Материально-техническое обеспечение

<p>Учебная аудитория № 205 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)</p>	<p>Набор учебной мебели, меловая доска.</p>
---	---

VIII. Сведения об обновлении рабочей программы дисциплины

№ п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Дата и протокол заседания кафедры, утвердившего изменения
1.	3. Объем дисциплины	Выделение часов на практическую подготовку	От 29.10.2020 года, протокол № 3 ученого совета факультета
2.	II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	Выделение часов на практическую подготовку	От 29.10.2020 года, протокол № 3 ученого совета факультета
3	V. Учебно-методическое и информационное обеспечение, необходимое для проведения практики 2) Программное обеспечение	Внесены изменения в программное обеспечение	От 29.09.2022 года, протокол № 2 ученого совета факультета
4	VII. Материально-техническое обеспечение	Внесены изменения в материально-техническое обеспечение аудиторий	От 29.09.2022 года, протокол № 2 ученого совета факультета