

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

Проф. Талорский Д.А.



_____ 20__ г.

подписано электронно-цифровой подписью

Программа дисциплины
Технология Data Mining Б1.В.ДВ.5

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Математические основы и программное обеспечение информационной безопасности и защиты информации

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Разинков Е.В.

Рецензент(ы):

Степанов Р.Г.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Латыпов Р. Х.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No 951216

Казань
2016

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) ассистент, к.н. Разинков Е.В. кафедра системного анализа и информационных технологий отделение фундаментальной информатики и информационных технологий , Evgenij.Razinkov@kpfu.ru

1. Цели освоения дисциплины

Курс охватывает следующие разделы:

- Предварительная обработка данных;
- Задача классификации с обучением;
- Поиск ассоциативных правил;
- Кластерный анализ;
- Нечеткая логика и нечеткие множества.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.5 Дисциплины (модули)" основной образовательной программы 02.04.02 Фундаментальная информатика и информационные технологии и относится к дисциплинам по выбору. Осваивается на 2 курсе, 3 семестр.

"Технологии Data mining" входит в состав профессиональных дисциплин, М2.ДВ8.1. Читается 2 курсе, в 3 семестре.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1 (общекультурные компетенции)	способность к абстрактному мышлению, анализу, синтезу
ОК-3 (общекультурные компетенции)	готовность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3 (профессиональные компетенции)	способность использовать и применять углубленные теоретические и практические знания в области фундаментальной информатики и информационных технологий
ОПК-4 (профессиональные компетенции)	способность самостоятельно приобретать и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять своё научное мировоззрение
ОПК-5 (профессиональные компетенции)	способность использовать углублённые знания правовых и этических норм при оценке последствий своей профессиональной деятельности, при разработке и осуществлении социально значимых проектов
ПК-1 (профессиональные компетенции)	способность проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива
ПК-13 (профессиональные компетенции)	способность разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологий

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-14 (профессиональные компетенции)	способность выполнять работу экспертов в ведомственных, отраслевых или государственных экспертных группах по экспертизе проектов, тематика которых соответствует направленности (профилю) программы магистратуры
ПК-16 (профессиональные компетенции)	способность участвовать в деятельности профессиональных сетевых сообществ по конкретным направлениям
ПК-17 (профессиональные компетенции)	способность осознавать корпоративную политику в области повышения социальной ответственности бизнеса перед обществом, принимать участие в ее развитии
ПК-2 (профессиональные компетенции)	способность использовать углубленные теоретические и практические знания в области информационных технологий и прикладной математики, фундаментальных концепций и системных методологий, международных и профессиональных стандартов в области информационных технологий
ПК-3 (профессиональные компетенции)	способность разрабатывать концептуальные и теоретические модели решаемых научных проблем и задач проектной и производственно-технологической деятельности

В результате освоения дисциплины студент:

1. должен знать:

о современных методах Data Mining.

2. должен уметь:

понимать основные проблемы, возникающие при анализе данных, и пути их решения.

3. должен владеть:

навыками анализа данных различной природы

применять в своей дальнейшей профессиональной деятельности полученные знания и навыки

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 72 часа(ов).

Форма промежуточного контроля дисциплины зачет в 3 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Введение в интеллектуальный анализ данных.	3	1-3	4	2	0	домашнее задание
2.	Тема 2. Предварительная обработка данных.	3	4-6	2	2	0	домашнее задание
3.	Тема 3. Задача классификации с обучением.	3	7-9	2	2	0	домашнее задание
4.	Тема 4. Задача поиска ассоциативных правил	3	10-12	2	2	0	домашнее задание
5.	Тема 5. Кластерный анализ.	3	13-15	2	2	0	домашнее задание
6.	Тема 6. Введение в нечеткую логику.	3	16-18	2	4	0	контрольная работа
	Тема . Итоговая форма контроля	3		0	0	0	зачет
	Итого			14	14	0	

4.2 Содержание дисциплины

Тема 1. Введение в интеллектуальный анализ данных.

лекционное занятие (4 часа(ов)):

Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining.

практическое занятие (2 часа(ов)):

Введение в интеллектуальный анализ данных. Основные этапы анализа данных.

Исторический обзор

Тема 2. Предварительная обработка данных.

лекционное занятие (2 часа(ов)):

Предварительная обработка данных. Очистка данных.

практическое занятие (2 часа(ов)):

Предварительная обработка данных. Интеграция и преобразование данных. Сокращение данных.

Тема 3. Задача классификации с обучением.

лекционное занятие (2 часа(ов)):

Задача классификации с обучением. Классификация с использованием деревьев решений, нейронных сетей.

практическое занятие (2 часа(ов)):

Наивная байесовская классификация, байесовские сети. Метод k ближайших соседей.

Тема 4. Задача поиска ассоциативных правил

лекционное занятие (2 часа(ов)):

Задача поиска ассоциативных правил на примере анализа рыночной корзины.

практическое занятие (2 часа(ов)):

Поиск ассоциативных правил. Метод генерации ассоциативных правил.

Тема 5. Кластерный анализ.

лекционное занятие (2 часа(ов)):

Кластерный анализ. Типы данных в кластерном анализе. Методы k средних, k медоидов.

практическое занятие (2 часа(ов)):

Кластерный анализ. Агломеративные и дивизимные методы иерархической кластеризации.

Тема 6. Введение в нечеткую логику.

лекционное занятие (2 часа(ов)):

Введение в нечеткую логику. Определение нечетких множеств.

практическое занятие (4 часа(ов)):

Основные операции с нечеткими множествами. Алгоритм нечеткой кластеризации.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Введение в интеллектуальный анализ данных.	3	1-3	подготовка домашнего задания	6	домашнее задание
2.	Тема 2. Предварительная обработка данных.	3	4-6	подготовка домашнего задания	6	домашнее задание
3.	Тема 3. Задача классификации с обучением.	3	7-9	подготовка домашнего задания	8	домашнее задание
4.	Тема 4. Задача поиска ассоциативных правил	3	10-12	подготовка домашнего задания	8	домашнее задание
5.	Тема 5. Кластерный анализ.	3	13-15	подготовка домашнего задания	8	домашнее задание
6.	Тема 6. Введение в нечеткую логику.	3	16-18	подготовка к контрольной работе	8	контрольная работа
	Итого				44	

5. Образовательные технологии, включая интерактивные формы обучения

Обучение происходит в форме лабораторных занятий и самостоятельной работы студентов. Список литературы разделен на две категории: необходимый для сдачи зачета минимум и дополнительная литература.

Изучение курса подразумевает не только овладение теоретическим материалом, но и получение практических навыков для более глубокого понимания разделов на основе решения задач и упражнений, иллюстрирующих доказываемые теоретические положения, а также развитие абстрактного мышления и способности самостоятельно доказывать утверждения.

Самостоятельная работа предполагает выполнение домашних работ. Практические задания, выполненные в аудитории, предназначены для указания общих методов решения задач определенного типа. Закрепить навыки можно лишь в результате самостоятельной работы.

Кроме того, самостоятельная работа включает подготовку к зачету. При подготовке к сдаче зачета весь объем работы рекомендуется распределять равномерно по дням, отведенным для подготовки к зачету, контролировать каждый день выполнения работы. Лучше, если можно перевыполнить план. Тогда будет резерв времени.

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

Тема 1. Введение в интеллектуальный анализ данных.

домашнее задание , примерные вопросы:

Подготовка к дискуссии. Примерные темы: - Субъектные и объектные системы. Первая модель управления (внешнее управление). - Сравнительный анализ подходов к понятию управления: как средство достижения цели, как процесс управляющих воздействий, как процесс обработки информации. - Управление как процесс выбора и принятия решений. Вторая модель управления - автоматическое управление

Тема 2. Предварительная обработка данных.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Обсуждение. Решение задач и примеров. Работа с персональными модифицированными данными и измененным сценарием Провести изучение структуры информации на сайте КФУ

Тема 3. Задача классификации с обучением.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Обсуждение. Решение задач и примеров. Задача фильтрации для данных сайта КФУ Задача сглаживания для данных сайта КФУ

Тема 4. Задача поиска ассоциативных правил

домашнее задание , примерные вопросы:

Контрольная работа Подготовить отчет по демопримеру Корреляционный анализ Два варианта исследование корреляций всего массива данных для выходных данных, соответственно цифрам вашего порядкового номера Парное исследование корреляций

Тема 5. Кластерный анализ.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Обсуждение. Решение задач и примеров. Самостоятельно разобрать Пример кластеризации с помощью k-means алгоритма. Кластеризация регионов

Тема 6. Введение в нечеткую логику.

контрольная работа , примерные вопросы:

онтрольная работа Реализовать разобранный пример кластеризации с помощью дерева решений для модифицированных данных Разобрать пример кластеризации с помощью дерева решений из документации "Руководство аналитика" и реализовать его практически Самостоятельно разобрать пример кластеризации с помощью карты Кохонена. (типовые и модифицированные данные)

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

1. Основные задачи, которые выполняют ИАС.
2. Роль и место анализа в принятии решений.
3. Проблемы анализа в свете использования информационных технологий.
4. Содержание аспекта сбора и хранения данных.
5. Содержание аспекта анализа данных и предоставления результатов анализа пользователям.
6. Классификация средств выполнения анализа с помощью ИТ.
7. Состав информационных технологий и информационных систем на предприятии и из

внешней среды - источников данных для сосредоточения в информационном хранилище или непосредственно для анализа.

Архитектура информационно-аналитических систем.

1. Основные принципы построения архитектуры Информационно-аналитических систем.
2. Структура средств сбора и доработки данных.
3. Способы и состав средств преобразования данных.
4. Среда разработки средств сбора, доработки данных и информационных хранилищ.
5. Основные принципы построения структуры информационных хранилищ.
6. Положения, заложенные в архитектуру средств оперативного (OLAP) анализа.
7. Архитектура комплекса средств интеллектуального анализа (Data mining).

Тема 3. Технологии сбора и хранения данных - концепция информационных хранилищ.

- 1 Принципы гибкой архитектуры данных и открытых систем, которыми руководствуются при создании ИАС.
2. Информационный обмен, связанный с аналитической работой.
3. Понятие о метаданных (МД), базе метаданных - репозитории, используемых в информационном хранилище.
4. Принципы построения информационных хранилищ.
5. Требования к качеству данных и способы его обеспечения при загрузке в информационное хранилище.
6. Проблемы, разрешаемые при приведении данных к единой структуре информационного хранилища.
7. Концепции построения структур хранилищ данных.
8. Назначение, состав и выполняемые функции базы метаданных - репозитория ИХ.
9. Принципы создания репозитория ИХ.
10. Элементы моделей данных ИХ (факт-таблица, таблицы измерений, кон-сольные таблицы).
11. Принципы создания репозитория ИХ.
12. Схемы представления - модели многомерных данных.
13. Типы многомерных OLAP-систем.

Тема 4. Признаки OLAP-систем, технологии оперативного анализа данных.

1. Признаки OLAP-систем по тесту FASMI.
2. Задачи и содержание оперативного (OLAP) анализа.
3. Классификация ИТ-анализа по режиму и темпу.
4. Функции, выполняемые процедурами сечение и поворот.
5. Функции, выполняемые процедурами свертка и развертка.
6. Функции, выполняемые процедурами проекция и построение трендов.
7. Содержание требований к OLAP-системе под рубрикой "быстрый".
8. Содержание требований к OLAP-системе под рубрикой "анализ".
9. Содержание требований к OLAP-системе под рубрикой "разделяемой".
10. Содержание требований к OLAP-системе под рубрикой "многомерной".
11. Что означают требования к OLAP-системе под рубрикой "информации"?
12. Сущность MOLAP-системы.
13. Сущность ROLAP-системы.

ТИПОВАЯ УЧЕБНАЯ ПРОГРАММА ДИСЦИПЛИНЫ ПС_ТУПД

Форма В Стр. 7 из 7

14. Сущность HOLAP-системы.

Тема 5. Технологии интеллектуального анализа данных.

1. Содержание понятия "знания", классификация видов знаний.
2. Интеллектуальный анализ данных (Data mining), цели и решаемые задачи.
3. Состав и содержание специфических задач интеллектуального анализа.
4. Особенности средств интеллектуального анализа данных.
5. Содержание методики нечёткая логика.
6. Сущность кластеризации данных, её отличие от классификации.
7. Области применения методов интеллектуального анализа.
8. Системы рассуждений на основе аналогичных случаев.
9. Классификационные и регрессионные деревья решений.
10. Байесовское обучение (ассоциации).
11. Генетические алгоритмы.
12. Эволюционное программирование и алгоритмы ограниченного перебора.

Тема 6. Основы управления информационно-аналитическими и интеллектуальными системами и их проектирования.

1. Состав программных инструментальных средств ИАС.
2. Средства сбора и доработки данных.
3. Средства оперативного OLAP- анализа.
4. Средства интеллектуального анализа данных.
5. Основы проектирования ИАС.
6. Содержание основной задачи проектирования ИАС.
7. Принципы проектирования ИАС.
8. Содержание этапов проектирования ИС применительно к проектированию ИАС.
9. Факторы, которые необходимо учесть при проектировании ИАС.
10. Управление информационно-аналитическими системами.
11. Задачи и средства администрирования ИАС.
12. Технологии загрузки данных в информационное хранилище.
13. Содержание планирования работы ИАС.
14. Принципы и этапы проектирования ИАС.
15. Рынок инструментальных средств ИАС.

7.1. Основная литература:

1. Статистический анализ данных в MS Excel [Электронный ресурс]: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - М.: ИНФРА-М, 2012. - 320 с. . - Режим доступа: <http://www.znanium.com/bookread.php?book=238654>
2. Компьютерные технологии анализа данных в эконометрике [Электронный ресурс]: Монография / Д.М. Дайитбегов. - 3-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с. . - Режим доступа: <http://www.znanium.com/bookread.php?book=365692>
3. Интеллектуальный анализ временных рядов [Электронный ресурс]: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с. . - Режим доступа: <http://www.znanium.com/bookread.php?book=249314>

7.2. Дополнительная литература:

1. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. URL: <http://znanium.com/bookread2.php?book=451186>

2. Информационные технологии: разработка информационных моделей и систем: Учеб. пос. / А.В.Затонский - М.: ИЦ РИОР: НИЦ ИНФРА-М, 2014 - 344с. URL: <http://znanium.com/bookread2.php?book=400563>

7.3. Интернет-ресурсы:

Интернет-портал по алгоритмике и программированию - <http://algotlist.manual.ru/>

Интернет-портал по технологии анализа данных - <http://www.olap.ru/>

Интернет-портал ресурсов по ИТ - <http://www.intuit.ru>

Интернет-портал ресурсов по математическим наукам - <http://www.math.ru/>

Электронная библиотека по техническим наукам - <http://techlibrary.ru>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Технология Data Mining" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

лабораторные занятия по дисциплине проводятся в аудитории, оснащенной доской и мелом (маркером)

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 02.04.02 "Фундаментальная информатика и информационные технологии" и магистерской программе Математические основы и программное обеспечение информационной безопасности и защиты информации .

Автор(ы):

Разинков Е.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Степанов Р.Г. _____

"__" _____ 201__ г.