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

подписано электронно-цифровой подписью

Программа дисциплины

Анализ данных Б2.Б.8

Направление подготовки: 080500.62 - Бизнес-информатика
Профиль подготовки: не предусмотрено
Квалификация выпускника: бакалавр
Форма обучения: очное
Язык обучения: русский

Автор(ы):

Миссаров М.Д.

Рецензент(ы):

Миссаров М.Д.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Миссаров М. Д.
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 914815

Казань
2014

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

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

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

Программу дисциплины разработал(а)(и) заведующий кафедрой, д.н. (доцент) Миссаров М.Д. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , Moukadas.Missarov@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 " Б2.Б.8 Общепрофессиональный" основной образовательной программы 080500.62 Бизнес-информатика и относится к базовой (общепрофессиональной) части. Осваивается на 4 курсе, 7 семестр.

Данный курс проводится на 4 курсе, в 7 семестре.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-12 (общекультурные компетенции)	осознает сущность и значение информации в развитии современного общества; владеет основными методами, способами и средствами получения, хранения, переработки информации;
ПК-20 (профессиональные компетенции)	использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования;
ПК-21 (профессиональные компетенции)	готовить научно-технические отчеты, презентации, научные публикации по результатам выполненных исследований.

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.

4. должен демонстрировать способность и готовность:

-

4. Структура и содержание дисциплины/ модуля

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

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

Форма промежуточного контроля дисциплины зачет в 7 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. Основные понятия интеллектуального анализа данных.	7	1-3	2	0	2	устный опрос домашнее задание
2.	Тема 2. Кластерный анализ .	7	4-6	3	0	4	домашнее задание устный опрос
3.	Тема 3. Байесовская классификация.	7	7-9	3	0	6	устный опрос домашнее задание
4.	Тема 4. Регрессионный анализ.	7	10-12	4	0	8	домашнее задание устный опрос
.	Тема . Итоговая форма контроля	7		0	0	0	зачет
	Итого			12	0	20	

4.2 Содержание дисциплины

Тема 1. Основные понятия интеллектуального анализа данных.

лекционное занятие (2 часа(ов)):

Классификация задач интеллектуального анализа данных. Расстояния, сходства, метрики для различных типов данных.

лабораторная работа (2 часа(ов)):

Разбор примеров и решение задач по теме Основные понятия интеллектуального анализа данных.

Тема 2. Кластерный анализ .

лекционное занятие (3 часа(ов)):

Эвристические алгоритмы кластерного анализа. Иерархическая кластеризация .Числовые характеристики кластерного разбиения.

лабораторная работа (4 часа(ов)):

Разбор примеров и решение задач по теме Кластерный анализ в среде R.

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

Тема 3. Байесовская классификация.

лекционное занятие (3 часа(ов)):

Байесовский подход к задаче классификации. Многомерное гауссовское распределение. Линейный и квадратичный дискриминантный анализ.

лабораторная работа (6 часа(ов)):

Разбор примеров и решение задач по теме Байесовская классификация в среде R.

Тема 4. Регрессионный анализ.

лекционное занятие (4 часа(ов)):

Однофакторная линейная регрессионный. Множественная линейная регрессия. Проверка гипотез о коэффициентах регрессии. Нелинейная регрессия. Логистическая регрессия.

лабораторная работа (8 часа(ов)):

Разбор примеров и решение задач по теме Регрессионный анализ в среде R.

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. Основные понятия интеллектуального анализа данных.	7	1-3	подготовка домашнего задания	4	домашнее задание
				подготовка к устному опросу	4	устный опрос
2.	Тема 2. Кластерный анализ .	7	4-6	подготовка домашнего задания	4	домашнее задание
				подготовка к устному опросу	4	устный опрос
3.	Тема 3. Байесовская классификация.	7	7-9	подготовка домашнего задания	6	домашнее задание
				подготовка к устному опросу	6	устный опрос
4.	Тема 4. Регрессионный анализ.	7	10-12	подготовка домашнего задания	6	домашнее задание
				подготовка к устному опросу	6	устный опрос
	Итого				40	

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.

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

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. Основные понятия интеллектуального анализа данных.

домашнее задание , примерные вопросы:

Решение задач на вычисление расстояний между объектами для различных типов данных изучение основных функций языка R для решения задач анализа данных

устный опрос , примерные вопросы:

Разбор задач по теме Основные понятия интеллектуального анализа данных.

Тема 2. Кластерный анализ .

домашнее задание , примерные вопросы:

Изучение функций языка R для решения задач кластерного анализа, решение задач по этой теме

устный опрос , примерные вопросы:

Разбор задач и обсуждение основных понятий по теме Кластерный анализ.

Тема 3. Байесовская классификация.

домашнее задание , примерные вопросы:

Изучение функций языка R для решения задач линейного и квадратичного дискриминантного анализа

устный опрос , примерные вопросы:

Разбор задач и обсуждение основных понятий по теме Байесовская классификация.

Тема 4. Регрессионный анализ.

домашнее задание , примерные вопросы:

Изучение функций языка R для решения задач однофакторной и множественной линейной регрессии.

устный опрос , примерные вопросы:

Разбор задач и обсуждение основных понятий по теме Регрессионный анализ.

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

По данному курсу предусмотрен зачет. Примерные вопросы к зачету - Приложение 1.

Вопросы к зачету

1. Классификация задач анализа данных.
2. Расстояния для числовых данных.
3. Корреляционные сходства и расстояния.
4. Расстояния на строках и перестановках.
5. Эвристические алгоритмы кластерного анализа.
6. Иерархическая кластеризация .
7. Числовые характеристики кластерного разбиения.
8. Байесовский подход к задаче классификации.
9. Многомерное гауссовское распределение.
10. Линейный дискриминантный анализ.

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

11. Квадратичный дискриминантный анализ.
12. Однофакторная линейная регрессионный.
13. Множественная линейная регрессия.
14. Проверка гипотез о коэффициентах регрессии.
15. Нелинейная регрессия.
16. Логистическая регрессия

7.1. Основная литература:

1. Степанов, Роман Григорьевич. Технология Data Mining: Интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т.?Казань: Казанский государственный университет, 2009.?110 с

2. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. ? 3-е изд., перераб. и доп. ? СПб.: БХВ-Петербург, 2009. ? 512 с.: ил. + CD-ROM ? (Учебная литература для вузов).

http://www.znanium.com/bookread.php?book=350638

3.Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузовский учебник: ИНФРА-М, 2010. - 578 с.: 70x100 1/16. - (Научная книга). (переплет) ISBN 978-5-9558-0191-9 http://www.znanium.com/bookread.php?book=251791

4.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. - (Университетская серия). - ISBN 978-5-4257-0092-6.

http://www.znanium.com/bookread.php?book=451186

5. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R", 2013

http://zilant.kpfu.ru/course/view.php?id=17341

6. Кашина О.А., Миссаров М.Д. Электронный курс

"Статистический анализ данных", 2013

http://zilant.kpfu.ru/course/view.php?id=17260

7.2. Дополнительная литература:

.Анализ данных : стат. и вычисл. методы для науч. работников и инженеров / З. Брандт ; Пер. с англ. О.И. Волковой ; Под ред. Е.В. Чепурина .? Москва : МИР : АСТ, 2003 .? 686 с. : ил. ; 24 + 1 опт. диск (CD-ROM) .? Загл. и авт. ориг.: Data Analysis / Siegmund Brandt .? Библиогр. в подстроч. примеч. ? Указ. компьютер. прогр.: с. 666-667 .? Предм. указ.: с. 668-674 .? ISBN 5-03-003478-1 (Мир) .? ISBN 5-17-019778-0 (АСТ) , 5000.

2.Геостатистический анализ данных в экологии и природопользовании (с применением пакета R) : учебное пособие / Казан. (Приволж.) федер. ун-т, Ин-т экологии и географии ; [авт.-сост.: д-р биол. наук, проф. А. А. Савельев и др.] .? Казань : Казанский университет, 2012 .? 120 с. : ил. ; 21 .? Библиогр.: с. 108-110 (22 назв.).

3.Анализ данных секвенирования транскриптома и метаболома [Текст: электронный ресурс] : учебно-методическое пособие / Н. И. Акберова .? Электронные данные (1 файл: 1,19 Мб) .? (Казань : Казанский федеральный университет, 2014) .? Загл. с экрана .? Для 3-го семестра .? Вых. дан. ориг. печ. изд.:.

Режим доступа: только для студентов и сотрудников КФУ .?

<URL:http://libweb.kpfu.ru/ebooks/01-IFMB/01_012_000740.pdf>

7.3. Интернет-ресурсы:

Data Mining Labs - http://dmlabs.org/

The R Project for Statistical Computing - http://www.r-project.org/

курс "Data mining" - http://www.intuit.ru/studies/courses/6/6/info

курс "Интеллектуальный анализ данных средствами MS SQL Server 2008" - http://www.intuit.ru/studies/courses/2312/612/info

программный комплекс RStudio для интеллектуального анализа данных - http://www.rstudio.com/

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Анализ данных" предполагает использование следующего материально-технического обеспечения:

Лекции и 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 080500.62 "Бизнес-информатика" и профилю подготовки не предусмотрено .

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

Автор(ы):
Миссаров М.Д. _____
"__" _____ 201 __ г.

Рецензент(ы):
Миссаров М.Д. _____
"__" _____ 201 __ г.

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ