

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. Таюрский Д.А.

"__" _____ 20__ г.

Программа дисциплины

Анализ данных Б1.Б.8

Направление подготовки: 38.03.05 - Бизнес-информатика

Профиль подготовки: не предусмотрено

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Кашина О.А. , Миссаров М.Д.

Рецензент(ы):

Пинягина О.В.

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

Казань
2017

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) доцент, к.н. (доцент) Кашина О.А. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , olga.kashina@mail.ru ; 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:

- Представление и преобразование данных;
- Кластерный анализ;
- Байесовская классификация;
- Регрессионный анализ.

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

Данная учебная дисциплина включена в раздел " Б1.Б.8 Дисциплины (модули)" основной образовательной программы 38.03.05 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
ПК-17 (профессиональные компетенции)	Способность использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования
ПК-18 (профессиональные компетенции)	Способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования

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 продемонстрировать способность и готовность:
 - использовать базовые теоретические знания, практические навыки и умения для проведения научных и научно-прикладных исследований, аналитической деятельности;
 - использовать современные инструментальные и вычислительные средства (пакет R) для визуализации данных, нахождения и анализа численных решений.

4. Структура и содержание дисциплины/ модуля

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

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	4	0	2	Устный опрос Письменное домашнее задание
2.	Тема 2. Кластерный анализ .	7	4-6	4	0	4	Письменное домашнее задание Устный опрос
3.	Тема 3. Байесовская классификация.	7	7-9	5	0	6	Устный опрос Письменное домашнее задание
4.	Тема 4. Регрессионный анализ.	7	10-12	5	0	6	Письменное домашнее задание Устный опрос
	Тема . Итоговая форма контроля	7		0	0	0	Зачет
	Итого			18	0	18	

4.2 Содержание дисциплины

Тема 1. Основные понятия интеллектуального анализа данных.

лекционное занятие (4 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 (4 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.

Тема 3. Байесовская классификация.

лекционное занятие (5 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 (5 часа(ов)):

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

лабораторная работа (6 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	2	Письменное домашнее задание
				подготовка домашнего задания	2	домашнее задание
				подготовка к устному опросу	2	Устный опрос
				подготовка к устному опросу	2	устный опрос
2.	Тема 2. Кластерный анализ .	7	4-6	подготовка домашнего задания	2	Письменное домашнее задание
				подготовка домашнего задания	2	домашнее задание
				подготовка к устному опросу	2	Устный опрос
				подготовка к устному опросу	2	устный опрос

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
3.	Тема 3. Байесовская классификация.	7	7-9	подготовка домашнего задания	3	Письменное домашнее задание
				подготовка домашнего задания	3	домашнее задание
				подготовка к устному опросу	3	Устный опрос
				подготовка к устному опросу	3	устный опрос
4.	Тема 4. Регрессионный анализ.	7	10-12	подготовка домашнего задания	2	Письменное домашнее задание
				подготовка домашнего задания	2	домашнее задание
				подготовка к устному опросу	2	Устный опрос
				подготовка к устному опросу	2	устный опрос
	Итого				36	

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. 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

Письменное домашнее задание , примерные вопросы:

Отчёт о выполненных заданиях в электронном курсе

Устный опрос , примерные вопросы:

Терминология анализа данных

устный опрос , примерные вопросы:

Анализ данных и смежные дисциплины. Приложения анализа данных.

Тема 2. Кластерный анализ .

домашнее задание , примерные вопросы:

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

Письменное домашнее задание , примерные вопросы:

Отчёт о выполненных заданиях по теме "Кластерный анализ" в электронном курсе

Устный опрос , примерные вопросы:

Постановка задачи кластерного анализа, подходы к решению задачи.

устный опрос , примерные вопросы:

Приложения кластерного анализа.

Тема 3. Байесовская классификация.

домашнее задание , примерные вопросы:

Проработка теоретического материала по теме "Байесовская классификация".

Письменное домашнее задание , примерные вопросы:

Отчёт о выполненных заданиях по теме "Байесовская классификация" в электронном курсе

Устный опрос , примерные вопросы:

Основные понятия классификации: постановка задачи, отличие от задачи кластеризации, подходы к решению, метод Байеса.

устный опрос , примерные вопросы:

Практические приложения задач классификации.

Тема 4. Регрессионный анализ.

домашнее задание , примерные вопросы:

Проработка теоретического материала по теме "Регрессионный анализ"

Письменное домашнее задание , примерные вопросы:

Отчёт о результатах решения задач регрессионного анализа в пакете R

Устный опрос , примерные вопросы:

Основные понятия регрессионного анализа: корреляция, регрессия, прямая регрессии, коэффициенты регрессии, гомоскедастичность, гетероскедастичность, прогноз

устный опрос , примерные вопросы:

Приложения регрессионного анализа.

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.
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. Интеллектуальный анализ временных рядов: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с.. - (Высшее образование). ISBN 978-5-8199-0496-1
<http://www.znanium.com/bookread.php?book=249314>
4. 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>
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:

1. Ризаев, Ильдус Султанович. Интеллектуальный анализ данных для поддержки принятия решений / И.С. Ризаев, Я. Рахал. ? Казань : [Изд-во МОиН РТ], 2011. ? 168 с. : ил. ; 21. ? (Серия 'Современная прикладная математика и информатика') .? Библиогр.: с. 140-146 (92 назв.) .? ISBN 978-5-4233-0017-3 ((в пер.)) , 100.
2. Наглядная статистика. Используем R! / А. Б. Шипунов, Е. М. Балдин, П. А. Волкова [и др.] .? Москва : ДМК Пресс, 2014. ? 296, [1] с. : ил. ; 21. ? Библиогр.: с. 291-292. ? Предм. указ.: с.293-296. ? ISBN 978-5-97060-094-8 ((в обл.)) , 100.
3. Сидняев, Николай Иванович.

Теория планирования эксперимента и анализа статистических данных : учебное пособие для магистров : для студентов и аспирантов вузов, обучающихся по специальности 'Прикладная математика' : учебник для студентов высших учебных заведений, обучающихся по физико-математическим направлениям подготовки и специальностям / Н. И. Сидняев .? 2-е изд., перераб. и доп. ? Москва : Юрайт, 2014 .? 495 с. : ил. ; 21 .? (УМО рекомендует, Учебное пособие) (Магистр) .? На обороте тит. л. авт.: Сидняев Н. И. - д.т.н., проф. ? По пред. изд. ? Библиогр.: с. 492-495 (50 назв.) .? ISBN 978-5-9916-2925-6 ((в пер.)) , 500.

4. 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.).

7.3. Интернет-ресурсы:

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

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

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

Электронный курс - <http://edu.kpfu.ru/course/view.php?id=833>

Электронный ресурс: R: Анализ и визуализация данных -

http://r-analytics.blogspot.ru/p/blog-page_20.html#.WbaQfrJJauI

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Анализ данных" 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 "БиблиоРоссика", доступ к которой предоставлен студентам. В ЭБС "БиблиоРоссика" представлены коллекции актуальной научной и учебной литературы по гуманитарным наукам, включающие в себя публикации ведущих российских издательств гуманитарной литературы, издания на английском языке ведущих американских и европейских издательств, а также редкие и малотиражные издания российских региональных вузов. ЭБС "БиблиоРоссика" обеспечивает широкий законный доступ к необходимым для образовательного процесса изданиям с использованием инновационных технологий и соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

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.

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 38.03.05 "Бизнес-информатика" и профилю подготовки не предусмотрено .

Автор(ы):

Кашина О.А. _____

Миссаров М.Д. _____

"__" _____ 201__ г.

Рецензент(ы):

Пинягина О.В. _____

"__" _____ 201__ г.