

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Отделение менеджмента



подписано электронно-цифровой подписью

Программа дисциплины
Анализ данных Б2.Б.8

Направление подготовки: 080500.62 - Бизнес-информатика

Профиль подготовки: Информационно-аналитические системы в бизнесе

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Григорьева И.С.

Рецензент(ы):

Турилова Е.А.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Турилова Е. А.

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

Учебно-методическая комиссия Института управления, экономики и финансов (отделение менеджмента):

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

Регистрационный No 94993614

Казань
2014

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) Григорьева И.С. кафедра математической статистики отделение прикладной математики и информатики , Irina.Grigorieva@kpfu.ru

1. Цели освоения дисциплины

Целью освоения дисциплины "Анализ данных" является формирование у студентов фундаментальных теоретических знаний по вопросам методики и практики применения методов статистического анализа данных, а также обучение студентов современным программным средствам в которых реализованы модули, осуществляющие решение задач анализа данных.

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.

Дисциплина "Анализ данных" изучается на четвертом году обучения. Данная дисциплина является логическим продолжением ряда курсов, изученных студентами по программе бакалавриата направления "Бизнес-информатика", включая "Моделирование бизнес процессов", "Базы данных", "Информационные системы управления производственной компанией".

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:

- принципы работы программных средств, предназначенных для статистического анализа данных;
- наиболее распространенные статистические методы анализа данных и условия их применения.

2. должен уметь:

- строить адекватные адаптивные модели для прогнозирования реальных социально-экономических данных и проводить по ним прогнозирование;
- осуществлять проверку адекватности и точности построенных моделей;
- самостоятельно выбирать методы анализа данных в практических ситуациях.

3. должен владеть:

- навыками тестирования программных модулей на модельных данных;
- навыками поиска решений в условиях риска и неопределенности;
- навыками анализа данных различной природы.

4. должен демонстрировать способность и готовность:

- применения программных систем, предназначенных для статистического анализа данных, а также тестировании программных модулей на модельных данных.

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. Введение в интеллектуальный анализ данных.	7	1-3	2	0	2	
2.	Тема 2. Обработка данных.	7	4-6	2	0	4	
3.	Тема 3. Методы классификации.	7	7-9	2	0	4	контрольная точка
4.	Тема 4. Задача поиска ассоциативных правил.	7	10-12	2	0	4	
5.	Тема 5. Кластерный анализ.	7	13-15	2	0	4	
6.	Тема 6. Введение в нечеткую логику.	7	16-18	2	0	2	контрольная точка
	Тема . Итоговая форма контроля	7		0	0	0	зачет
	Итого			12	0	20	

4.2 Содержание дисциплины

Тема 1. Введение в интеллектуальный анализ данных.

лекционное занятие (2 часа(ов)):

Введение в интеллектуальный анализ данных. Области использования. Связь с другими дисциплинами. Компоненты систем Data Mining.

лабораторная работа (2 часа(ов)):

Основные этапы анализа данных.

Тема 2. Обработка данных.

лекционное занятие (2 часа(ов)):

Предварительная обработка данных. Очистка данных.

лабораторная работа (4 часа(ов)):

Интеграция и преобразование данных. Сокращение данных.

Тема 3. Методы классификации.

лекционное занятие (2 часа(ов)):

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

лабораторная работа (4 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.

лабораторная работа (4 часа(ов)):

Метод "A priori" генерации ассоциативных правил.

Тема 5. Кластерный анализ.

лекционное занятие (2 часа(ов)):

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

лабораторная работа (4 часа(ов)):

Агломеративные и дивизимные методы иерархической кластеризации.

Тема 6. Введение в нечеткую логику.

лекционное занятие (2 часа(ов)):

Определение нечетких множеств. Основные операции с нечеткими множествами.

лабораторная работа (2 часа(ов)):

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
3.	Тема 3. Методы классификации.	7	7-9	подготовка к контрольной точке	20	контрольная точка
6.	Тема 6. Введение в нечеткую логику.	7	16-18	подготовка к контрольной точке	20	контрольная точка
	Итого				40	

5. Образовательные технологии, включая интерактивные формы обучения

В соответствии с требованиями ФГОС удельный вес занятий, проводимых в активных и интерактивных формах, составляет не менее 40% аудиторных занятий. Так, в процессе изучения дисциплины "Анализ данных" студенты разбирают практические примеры в компьютерном классе, решают предлагаемые кейсы, выступают со стендовыми докладами. До 50% лекционных и практических занятий проходят с использованием презентаций MS PowerPoint.

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.

Тема 2. Обработка данных.

Тема 3. Методы классификации.

контрольная точка , примерные вопросы:

1. В чем состоят цели многомерного анализа данных? 2. Дайте формальное описание данных, используемое в многомерном статистическом анализе и опишите представление данных в электронном виде. 3. Определите основные шкалы измерений в многомерных выборках. 4. Как произвести преобразование данных из одной шкалы в другую?

Тема 4. Задача поиска ассоциативных правил.

Тема 5. Кластерный анализ.

Тема 6. Введение в нечеткую логику.

контрольная точка , примерные вопросы:

1. Определите понятие диапазон значений признака и способы определения диапазона значений. 2. Какие и как рассчитываются основные характеристики многомерной выборки в предварительном анализе данных? 3. Определите средства визуализации данных в предварительном анализе даны? 4. Для чего используются модельные данные при решении задач многомерного анализа?

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

1. В чем состоит идея метода отбраковки?
2. Определите способы моделирования случайных чисел выборок в
3. Определите принципы моделирования многомерного нормального распределения?
4. Определите проблему робастности в анализе данных.
5. Что понимается под грубыми ошибками и каковы последствия их проявлений в статистической совокупности?
6. Назовите известные вам критерии робастности.
7. Дайте характеристику алгоритмов определения "подозрительных" объектов выборки.
8. Каковы последствия отсутствия значений в многомерной выборке и как с ними бороться.
9. Какие существуют способы обработки грубых ошибок?
10. Что понимается под статистической гипотезой и какие характерные признаки для нее существуют?
11. Что означает: простая и сложная, одномерная и многомерная статистическая гипотезы?
12. Какого рода ошибки могут допускаться и чем определяется достоверность выводов при проверке статистических гипотез?
13. Определите предпосылки применения критерия Фишера.
14. Определите предпосылки применения критерия Стьюдента.
15. Приведите примеры ситуаций в которых целесообразно применении критерия Стьюдента.
16. Приведите примеры ситуаций в которых целесообразно применении критерия Фишера.

17. Определите предпосылки применения критерия U-критерий Манна-Уитни (Вилксона)?
18. Как подсчитывается сумма рангов в критерии Вилксона?
19. Какое распределение используется в критерии Вилксона?
20. Определите общую модель дисперсионного анализа.
21. Какие компоненты дисперсии рассматриваются в анализе?
22. Какая функция EXCEL используется для проверки статистического критерия в дисперсионном анализе?
23. Какой вывод следует из того, что статистическая гипотеза в дисперсионном анализе верна?
24. Как определяются степени свободы в дисперсионном анализе?
25. Как рассчитывается коэффициент Юла?
26. Как рассчитываются ранги по выборке?
27. Принципы разработки типизаций?
28. Опишите какие программные средства, используемые для сортировки в Excel.
29. В чем состоит метод Дельфи?
30. Назовите стадии экспертного опроса.
31. В чем состоит проблема подбора экспертов?

7.1. Основная литература:

Козлов А.Ю. Статистический анализ данных в MS Excel: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - М.: ИНФРА-М, 2014. - 320 с.// <http://www.znanium.com/bookread.php?book=429722>

Дайитбегов Д.М. Компьютерные технологии анализа данных в эконометрике: Монография / Д.М. Дайитбегов. - 3-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с.// <http://www.znanium.com/bookread.php?book=365692>

Ярушкина Н.Г. Интеллектуальный анализ временных рядов: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с.// <http://www.znanium.com/bookread.php?book=249314>

7.2. Дополнительная литература:

Экономическая статистика: Учебник / Московский Государственный Университет им. М.В.Ломоносова (МГУ); Под ред. Ю.Н.Иванова - 4 изд., перераб. и доп. - М.: ИНФРА-М, 2011. - 668 с.// <http://www.znanium.com/bookread.php?book=245351>

Ефимова М.Р. Общая теория статистики: Учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 416 с.// <http://www.znanium.com/bookread.php?book=251320>

Вуколов Э.А. Основы статистического анализа. Практ. по стат. мет. и исслед. операций с исп. пакетов STATISTICA и EXCEL: Уч. пос./ Э.А. Вуколов - 2 изд., испр. и доп. - М.: Форум: НИЦ Инфра-М, 2013. - 464 с.// <http://www.znanium.com/bookread.php?book=369689>

Журнал "Дискретная математика" // http://elibrary.ru/title_about.asp?id=7778

Журнал "Дискретный анализ и исследование операций" // http://elibrary.ru/title_about.asp?id=25528

Журнал "Дифференциальные уравнения" // http://elibrary.ru/title_about.asp?id=9677

Журнал "Математические заметки" // http://elibrary.ru/title_about.asp?id=7874

Журнал "Математические труды" // http://elibrary.ru/title_about.asp?id=7875

7.3. Интернет-ресурсы:

Борисенко В. Материалы по курсу: Алгоритмы и структуры данных? -

<http://mech.math.msu.su/~vvb/HSE/>

Официальный сайт Интернет-университета информационных технологий. - <http://www.intuit.ru>

Полнотекстовая база данных по общественным и гуманитарным наукам -

<http://www.ebiblioteka.ru/>

Электронная библиотека Elibrary - <http://elibrary.ru>

Электронно-библиотечная система - <http://www.knigafund.ru/>

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 "КнигаФонд", доступ к которой предоставлен студентам. Электронно-библиотечная система "КнигаФонд" реализует легальное хранение, распространение и защиту цифрового контента учебно-методической литературы для вузов с условием обязательного соблюдения авторских и смежных прав. КнигаФонд обеспечивает широкий законный доступ к необходимым для образовательного процесса изданиям с использованием инновационных технологий и соответствует всем требованиям новых ФГОС ВПО.

Компьютерный класс, для проведения лабораторных занятий. Проекционное оборудование, используемое для проведения презентаций.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 080500.62 "Бизнес-информатика" и профилю подготовки Информационно-аналитические системы в бизнесе .

Автор(ы):

Григорьева И.С. _____

"__" _____ 201__ г.

Рецензент(ы):

Турилова Е.А. _____

"__" _____ 201__ г.