

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего 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

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
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 продемонстрировать способность и готовность:

-

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	Практи- ческие занятия	Лабора- торные 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	Устный опрос Письменное домашнее задание
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	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

№	Раздел дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
4.	Тема 4. Регрессионный анализ.	7	10-12	подготовка домашнего задания	4	домашнее задание
				подготовка к устному опросу	4	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

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:

Прогнозирование на основе линейной регрессионной модели - изучение постановки задачи и подходов к её решению. Выполнение задания на поиск коэффициентов линейной регрессионной модели (на конкретных данных) в электронном курсе.

устный опрос , примерные вопросы:

Основные понятия регрессионного анализа.

Итоговая форма контроля

зачет (в 7 семестре)

Примерные вопросы к итоговой форме контроля

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. Мультиколлинеарность и способы решения связанных с ней проблем.
15. Метод наименьших квадратов.
16. Проверка гипотез о коэффициентах регрессии.
17. Характеристики качества регрессионной модели. Коэффициент детерминации.
18. Обобщения линейной регрессионной модели. Нелинейная регрессия.
19. Бинарная классификация.
20. Логистическая регрессия.

7.1. Основная литература:

Кашина О.А., Миссаров М.Д. Электронный курс 'Анализ данных в среде R', 2014. -Режим доступа: <https://edu.kpfu.ru/course/view.php?id=833>

Кашина О.А., Миссаров М.Д. Электронный курс' Статистический анализ данных', 2013. - Режим доступа: <https://edu.kpfu.ru/course/view.php?id=835>

Макшанов, А.В. Технологии интеллектуального анализа данных [Электронный ресурс] : учебное пособие / А.В. Макшанов, А.Е. Журавлев. - Электрон. дан. -

Санкт-Петербург : Лань, 2018. - 212 с. - Режим доступа: <https://e.lanbook.com/book/109617>

7.2. Дополнительная литература:

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). - Режим доступа: <http://znanium.com/catalog/product/350638>

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- Режим доступа: <http://znanium.com/catalog/product/249314>

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. - (Научная книга). ISBN 978-5-9558-0275-6 - Режим доступа: <http://znanium.com/catalog/product/365692>

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:

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.