

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. Минзарипов Р.Г.

"\_\_" \_\_\_\_\_ 20\_\_ г.

**Программа дисциплины**

Введение в анализ данных М1.В.1

Направление подготовки: 010400.68 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

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  
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) 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

Целью курса Введение в анализ данных является формирование у студента представлений об основных задачах интеллектуального анализа данных и методах их решения.

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

Данная учебная дисциплина включена в раздел " М1.В.1 Общенаучный" основной образовательной программы 010400.68 Прикладная математика и информатика и относится к вариативной части. Осваивается на 1 курсе, 1 семестр.

Цикл: Цикл: М1.В. 1

Дисциплина "Управление рисками" изучается на первом году обучения.

### 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)	способностью владеть культурой мышления, умение аргументированно и ясно строить устную и письменную речь (ОК-1); способностью уважительно и бережно относиться к историческому наследию и культурным традициям, толерантность в восприятии социальных и культурных различий (ОК-2); способностью понимать движущие силы и закономерности исторического процесса; роль насилия и ненасилия в истории, место человека в историческом процессе, политической организации общества (ОК-3); способностью понимать и анализировать мировоззренческие, социально и лично значимые философские проблемы (ОК-4); способностью понимать сущность и значение информации в развитии современного информационного общества, сознавать опасности и угрозы, возникающие в этом процессе, соблюдать основные требования информационной безопасности, в том числе защиты государственной тайны (ОК-5); способностью демонстрации общенаучных базовых знаний естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой (ПК-1); способностью приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии (ПК-2); способностью понимать и применять в исследовательской и прикладной деятельности современный математический аппарат (ПК-3);

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-2 (общекультурные компетенции)	способностью уважительно и бережно относиться к историческому наследию и культурным традициям, толерантность в восприятии социальных и культурных различий
ПК-1 (профессиональные компетенции)	способностью демонстрации общенаучных базовых знаний естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой
ПК-2 (профессиональные компетенции)	способностью приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии
ПК-3 (профессиональные компетенции)	способностью понимать и применять в исследовательской и прикладной деятельности современный математический аппарат

В результате освоения дисциплины студент:

4. должен демонстрировать способность и готовность:

В результате освоения дисциплины обучающийся должен:

- Знать: основные задачи и методы анализа данных

- Уметь: формулировать задачи анализа данных и находить алгоритмы их решения, интерпретировать полученные результаты

- Владеть: математическими методами. используемыми при решении задач анализа данных

#### 4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины зачет в 1 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. Представление и преобразование данных.	1	1-2	2	2	0	домашнее задание устный опрос

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
2.	Тема 2. Кластерный анализ .	1	3-4	2	2	0	домашнее задание
3.	Тема 3. Снижение размерности.	1	5-6	2	2	0	домашнее задание
4.	Тема 4. Байесовская классификация.	1	7-9	3	3	0	домашнее задание
5.	Тема 5. Деревья решений .	1	10-11	2	2	0	контрольная работа
6.	Тема 6. Регрессионный анализ.	1	12-16	4	4	0	домашнее задание
	Тема . Итоговая форма контроля	1		0	0	0	зачет
	Итого			15	15	0	

#### 4.2 Содержание дисциплины

##### Тема 1. Представление и преобразование данных.

###### **лекционное занятие (2 часа(ов)):**

Типы данных, обработка данных. Меры сходства, метрики, ультраметрики. Классификация задач анализа данных.

###### **практическое занятие (2 часа(ов)):**

Разбор примеров и решение задач по теме "Представление и преобразование данных".

##### Тема 2. Кластерный анализ .

###### **лекционное занятие (2 часа(ов)):**

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

###### **практическое занятие (2 часа(ов)):**

Разбор примеров и решение задач по теме "Кластерный анализ".

##### Тема 3. Снижение размерности.

###### **лекционное занятие (2 часа(ов)):**

Матрица ковариаций. Метод главных компонент.

###### **практическое занятие (2 часа(ов)):**

Разбор примеров и решение задач по теме "Снижение размерности".

##### Тема 4. Байесовская классификация.

###### **лекционное занятие (3 часа(ов)):**

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

###### **практическое занятие (3 часа(ов)):**

Разбор примеров и решение задач по теме "Байесовская классификация".

##### Тема 5. Деревья решений .

###### **лекционное занятие (2 часа(ов)):**

Алгоритм CART. Энтропия как мера информации

**практическое занятие (2 часа(ов)):**

Разбор примеров и решение задач по теме "Деревья решений".

**Тема 6. Регрессионный анализ.**

**лекционное занятие (4 часа(ов)):**

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

**практическое занятие (4 часа(ов)):**

Разбор примеров и решение задач по теме "Регрессионный анализ".

### 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. Представление и преобразование данных.	1	1-2	подготовка домашнего задания	4	домашнее задание
				подготовка к устному опросу	4	устный опрос
2.	Тема 2. Кластерный анализ .	1	3-4	подготовка домашнего задания	8	домашнее задание
3.	Тема 3. Снижение размерности.	1	5-6	подготовка домашнего задания	14	домашнее задание
4.	Тема 4. Байесовская классификация.	1	7-9	подготовка домашнего задания	16	домашнее задание
5.	Тема 5. Деревья решений .	1	10-11	подготовка к контрольной работе	16	контрольная работа
6.	Тема 6. Регрессионный анализ.	1	12-16	подготовка домашнего задания	16	домашнее задание
	Итого				78	

### 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

#### Тема 1. Представление и преобразование данных.

домашнее задание , примерные вопросы:

-подготовка к практическим занятиям -доработка заданий, выполняемых на практических занятиях

устный опрос , примерные вопросы:

-подготовка к устному опросу

## **Тема 2. Кластерный анализ .**

домашнее задание , примерные вопросы:

-подготовка к практическим занятиям -доработка заданий, выполняемых на практических занятиях

## **Тема 3. Снижение размерности.**

домашнее задание , примерные вопросы:

-подготовка к практическим занятиям -доработка заданий, выполняемых на практических занятиях

## **Тема 4. Байесовская классификация.**

домашнее задание , примерные вопросы:

-подготовка к практическим занятиям -доработка заданий, выполняемых на практических занятиях

## **Тема 5. Деревья решений .**

контрольная работа , примерные вопросы:

-подготовка к контрольной работе

## **Тема 6. Регрессионный анализ.**

домашнее задание , примерные вопросы:

-подготовка к практическим занятиям -доработка заданий, выполняемых на практических занятиях

## **Тема . Итоговая форма контроля**

Примерные вопросы к зачету:

-

### **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.znaniyum.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.znaniyum.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.znaniyum.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:**

1. 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. Многомерный анализ данных методами прикладной статистики : учебное пособие / С. С. Барковский, В. М. Захаров, А. М. Лукашов [и др.] ; М-во образования и науки Рос. Федерации, Федер. агентство по образованию, Гос. образоват. учреждение высш. проф. образования "Казан. гос. техн. ун-т" .? Казань : [Изд-во Казанского государственного технического университета], 2010 .? 120, [1] с. : ил. ; 20 .? Библиогр.: с. 118-119 (12 назв.) .? ISBN 978-5-7579-1526-5 ((в обл.)) , 135.
3. Математика и загадочный генетический код : (к 10-летию завершения программы "Геном человека") / В. М. Гупал .? Москва : РИОР : ИНФРА-М, [2014] .? 286, [1] с., [1] л. ил. : ил. ; 21 .? (Научная мысль) .? На тит. л.: Электронно-Библиотечная Система znanium. com .? Библиогр.: с. 277-285 (190 назв.) .? ISBN 978-5-369-01276-5 ((РИОР)) .? ISBN 978-5-16-009122-8 ((ИНФРА-М)) , 100.

### 7.3. Интернет-ресурсы:

курс - <http://www.intuit.ru/studies/courses/6/6/info>

курс - <http://www.intuit.ru/studies/courses/2312/612/info>

Data Mining Labs программный комплекс RStudio для интеллектуального анализа данных - <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/>

### 8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Введение в анализ данных" предполагает использование следующего материально-технического обеспечения:

-

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 010400.68 "Прикладная математика и информатика" и магистерской программе Анализ данных и его приложения .

Автор(ы):

Миссаров М.Д. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Пинягина О.В. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.