

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
"Казанский (Приволжский) федеральный университет"  
Высшая школа информационных технологий и интеллектуальных систем



УТВЕРЖДАЮ  
Проректор по образовательной деятельности КФУ  
Проф. Д.А. Таюрский

\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г.

подписано электронно-цифровой подписью

**Программа дисциплины**  
Анализ данных Б1.В.ДВ.4

Направление подготовки: 09.03.03 - Прикладная информатика

Профиль подготовки: не предусмотрено

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Новиков А.А.

**Рецензент(ы):**

Замалиев Р.Р.

**СОГЛАСОВАНО:**

Заведующий(ая) кафедрой: Насыров С. Р.

Протокол заседания кафедры No \_\_\_\_\_ от "\_\_\_\_\_" \_\_\_\_\_ 201\_\_ г

Учебно-методическая комиссия Высшей школы информационных технологий и интеллектуальных систем:

Протокол заседания УМК No \_\_\_\_\_ от "\_\_\_\_\_" \_\_\_\_\_ 201\_\_ г

Регистрационный No 689552119

## 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) ассистент, к.н. Новиков А.А. Кафедра математического анализа отделение математики , A.Nobukob@gmail.com

### 1. Цели освоения дисциплины

Цель дисциплины - умение анализировать неструктурированные данные для

- а) дальнейшей структуризации для управления данными;
- б) создания прогностических моделей информации.

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

Данная учебная дисциплина включена в раздел "Б1.В.ДВ.4 Дисциплины (модули)" основной образовательной программы 09.03.03 Прикладная информатика и относится к дисциплинам по выбору. Осваивается на 3 курсе, 5 семестр.

Курс читается после курса, включающего в себя материал по алгоритмам и структурам данных и после курса по статистике.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-23 (профессиональные компетенции)	способностью применять системный подход и математические методы в формализации решения прикладных задач
ПК-7 (профессиональные компетенции)	способностью проводить описание прикладных процессов и информационного обеспечения решения прикладных задач

В результате освоения дисциплины студент:

3. должен владеть:

4. должен демонстрировать способность и готовность:

Основные понятия, терминологию анализа данных.

Основные методы исследования и примеры использования анализа данных..

Обучающийся должен знать теоретические основы анализа данных..

Уметь решать типовые задачи и формулировать прикладные задачи в терминах анализа данных..

Уметь строить модели естественных явлений, исследуемых методами анализа данных.

Владеть основными методами исследования, использующими анализа данных.

Владеть навыками интуитивного подхода к решению нестандартных задач с привлечением анализа данных.

Демонстрировать способность и готовность решать задачи, использующие методы анализа данных.

### 4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 5 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. Работа с датафреймами	5		6	6	0	Проверка практических навыков
2.	Тема 2. Методы кластеризации.	5		6	6	0	Письменное домашнее задание
3.	Тема 3. Методы классификации	5		6	6	0	Письменное домашнее задание
4.	Тема 4. Введение в регрессионный анализ	5		6	6	0	Письменное домашнее задание
5.	Тема 5. Разбор кейсов применения анализа данных	5		12	12	0	Ситуационная задача
.	Тема . Итоговая форма контроля	5		0	0	0	Зачет
	Итого			36	36	0	

#### 4.2 Содержание дисциплины

##### Тема 1. Работа с датафреймами

###### **лекционное занятие (6 часа(ов)):**

Библиотека pandas. Типы данных DataFrame и Series. Доступ по индексу в DataFrame. Фильтрация DataFrame. Группировка и агрегирование в Pandas. Чтение и запись файлов в Pandas, работа с реляционными базами данных. Визуализация данных в pandas, дашборды. Библиотека NumPy. Трехмерные и анимированные визуализации.

###### **практическое занятие (6 часа(ов)):**

Библиотека pandas. Типы данных DataFrame и Series. Доступ по индексу в DataFrame. Фильтрация DataFrame. Группировка и агрегирование в Pandas. Чтение и запись файлов в Pandas, работа с реляционными базами данных. Визуализация данных в pandas, дашборды. Библиотека NumPy. Трехмерные и анимированные визуализации.

## **Тема 2. Методы кластеризации.**

### ***лекционное занятие (6 часа(ов)):***

Метрические пространства, формальное определение и общие свойства метрических пространств. Метрические кластеризация. метод k-средних (k-means), их модификации, метод k-мод (k-mods), методы нечеткой кластеризации с примерами (c-means). Иерархическая кластеризация. Метод формального элемента, логические дендрограммы. Деревья решений, алгоритм случайного леса.

### ***практическое занятие (6 часа(ов)):***

Метрические пространства, формальное определение и общие свойства метрических пространств. Метрические кластеризация. метод k-средних (k-means), их модификации, метод k-мод (k-mods), методы нечеткой кластеризации с примерами (c-means). Иерархическая кластеризация. Метод формального элемента, логические дендрограммы. Деревья решений, алгоритм случайного леса.

## **Тема 3. Методы классификации**

### ***лекционное занятие (6 часа(ов)):***

Методы классификации. Методы наибольшего правдоподобия. Метод kNN. Наивный Байесовский классификатор. Решающие деревья и случайный лес. Нормированные векторные пространства. Нормы и нормированные векторные пространства. Метрические и нормированные пространства. Классификация в векторных пространствах. Логистическая регрессия. Метод опорных векторов.

### ***практическое занятие (6 часа(ов)):***

Методы классификации. Методы наибольшего правдоподобия. Метод kNN. Наивный Байесовский классификатор. Решающие деревья и случайный лес. Нормированные векторные пространства. Нормы и нормированные векторные пространства. Метрические и нормированные пространства. Классификация в векторных пространствах. Логистическая регрессия. Метод опорных векторов.

## **Тема 4. Введение в регрессионный анализ**

### ***лекционное занятие (6 часа(ов)):***

Метод наименьших квадратов. Модификации метода наименьших квадратов. Одномерная линейная регрессионная модель. Одномерная нелинейная регрессионная модель. Многомерная линейная регрессионная модель. Понижение размерностей. Сингулярное разложение. Метод главных компонент. Многомерная нелинейная регрессионная модель.

### ***практическое занятие (6 часа(ов)):***

Метод наименьших квадратов. Модификации метода наименьших квадратов. Одномерная линейная регрессионная модель. Одномерная нелинейная регрессионная модель. Многомерная линейная регрессионная модель. Понижение размерностей. Сингулярное разложение. Метод главных компонент. Многомерная нелинейная регрессионная модель.

## **Тема 5. Разбор кейсов применения анализа данных**

### ***лекционное занятие (12 часа(ов)):***

Библиотека scikit-learn и язык программирования R. Разбор примеров применения анализа данных в конкретных ситуациях, взятых из кейсов сайта Kaggle. Факторный анализ экономических показателей во влиянии на цену на недвижимость. Рассмотрение задачи скоринга в применении к дифференциальной диагностике болезней.

### ***практическое занятие (12 часа(ов)):***

Библиотека scikit-learn и язык программирования R. Разбор примеров применения анализа данных в конкретных ситуациях, взятых из кейсов сайта Kaggle. Факторный анализ экономических показателей во влиянии на цену на недвижимость. Рассмотрение задачи скоринга в применении к дифференциальной диагностике болезней.

## **4.3 Структура и содержание самостоятельной работы дисциплины (модуля)**

№	Раздел дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Работа с датафреймами	5		Используя библиотеки pandas и scikit-learn, создать программу, которая производит кластеризацию массива двумерных данных методом k-means на основе заданной метрики, которая производит классификацию	6	Проверка практических навыков
2.	Тема 2. Методы кластеризации.	5		подготовка домашнего задания	6	Письменное домашнее задание
3.	Тема 3. Методы классификации	5		подготовка домашнего задания	6	Письменное домашнее задание
4.	Тема 4. Введение в регрессионный анализ	5		подготовка домашнего задания	6	Письменное домашнее задание
5.	Тема 5. Разбор кейсов применения анализа данных	5		Оценивается состоятельность концепции, предложенной к заданному кейсу.	12	Ситуационная задача
	Итого				36	

### 5. Образовательные технологии, включая интерактивные формы обучения

деловых и ролевых игр,  
разбор конкретных ситуаций,

### 6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

#### Тема 1. Работа с датафреймами

Проверка практических навыков , примерные вопросы:

Используя библиотеки pandas произвести объединение информации из нескольких файлов.

#### Тема 2. Методы кластеризации.

Письменное домашнее задание , примерные вопросы:

Используя библиотеки pandas и scikit-learn, создать программу, которая производит кластеризацию массива двумерных данных методом k-means на основе заданной метрики

#### Тема 3. Методы классификации

Письменное домашнее задание , примерные вопросы:

Используя библиотеки pandas и scikit-learn, создать программу, которая производит классификацию методом логистической регрессии

#### **Тема 4. Введение в регрессионный анализ**

Письменное домашнее задание , примерные вопросы:

Используя библиотеки pandas и scikit-learn, создать программу, которая строит многомерную полиномиальную регрессионную модель по заданной выборке.

#### **Тема 5. Разбор кейсов применения анализа данных**

Ситуационная задача , примерные вопросы:

Оценивается состоятельность концепции, предложенной группой к заданному кейсу.

#### **Итоговая форма контроля**

зачет (в 5 семестре)

Примерные вопросы к итоговой форме контроля

Метрические пространства, формальное определение и общие свойства метрических пространств. Метрические кластеризация. метод k-средних (k-means), их модификации, метод k-мод (k-mods), методы нечеткой кластеризации с примерами (c-means). Иерархическая кластеризация. Метод формального элемента, логические дендрограммы. Деревья решений, алгоритм случайного леса.

Методы классификации. Методы наибольшего правдоподобия. Метод kNN. Наивный Байесовский классификатор. Решающие деревья и случайный лес. Нормированные векторные пространства. Нормы и нормированные векторные пространства. Метрические и нормированные пространства. Классификация в векторных пространствах. Логистическая регрессия. Метод опорных векторов.

Метод наименьших квадратов. Модификации метода наименьших квадратов. Одномерная линейная регрессионная модель. Одномерная нелинейная регрессионная модель. Многомерная линейная регрессионная модель. Понижение размерностей. Сингулярное разложение. Метод главных компонент. Многомерная нелинейная регрессионная модель.

#### **7.1. Основная литература:**

Теория вероятностей, Ч. 2. Случайные величины, , 2013г.

Теория вероятностей, Ч. 1. Элементарная теория вероятностей, , 2013г.

Буховец, А.Г. Алгоритмы вычислительной статистики в системе R [Электронный ресурс] : учеб. пособие / А.Г. Буховец, П.В. Москалев. - Электрон. дан. - Санкт-Петербург: Лань, 2015. - 160 с. - Режим доступа: <https://e.lanbook.com/book/68459>

Буре, В.М. Методы прикладной статистики в R и Excel [Электронный ресурс] : учеб. пособие / В.М. Буре, Е.М. Парилина, А.А. Седаков. - Электрон. дан. - Санкт-Петербург : Лань, 2016. - 152 с. - Режим доступа: <https://e.lanbook.com/book/81558>

Статистика : учебник / В.В. Глинский, В.Г. Ионин, Л.К. Серга [и др.] ; под ред. В.Г. Ионина. - 4-е изд., перераб. и доп.-М. : ИНФРА-М, 2018. - 355 с. - (Высшее образование: Бакалавриат) - Режим доступа: <http://znanium.com/bookread2.php?book=941774>

#### **7.2. Дополнительная литература:**

Лекции по теории вероятностей и математической статистике, Володин, Игорь Николаевич, 2006г.

Технология Data Mining: Интеллектуальный анализ данных, Степанов, Роман Григорьевич, 2009г.

Геостатистический анализ данных в экологии и природопользовании (с применением пакета R), Савельев, Анатолий Александрович, 2012г.

Теория вероятностей и математическая статистика: Учебное пособие / Бирюкова Л.Г., Бобрик Г.И., Матвеев В.И., - 2-е изд. - М.: НИЦ ИНФРА-М, 2017. - 289 с. - (Высшее образование: Бакалавриат) - Режим доступа: <http://znanium.com/bookread2.php?book=370899>

Теория вероятностей, математическая статистика, математическое программирование: Учебное пособие / Белько И.В., Морозова И.М., Криштапович Е.А. - М.:НИЦ ИНФРА-М, Нов. знание, 2016. - 299 с. - (Высшее образование: Бакалавриат) - Режим доступа: <http://znanium.com/bookread2.php?book=542521>

Колдаев В.Д. Численные методы и программирование : учеб. пособие / В.Д. Колдаев ; под ред. проф. Л.Г. Гагариной. -М. : ИД 'ФОРУМ' : ИНФРА-М, 2017. - 336 с. Режим доступа: <http://znanium.com/bookread2.php?book=672965>

### **7.3. Интернет-ресурсы:**

machinelearning.ru - <http://www.machinelearning.ru>

Stepik.org - <https://stepik.org/course/16560/>

Введение в машинное обучение -

<https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie?action=enroll>

Машинное обучение в бизнесе - <https://stepik.org/course/16560/>

Хабр - [habr.com](http://habr.com)

### **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 "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.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен студентам. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

мультимедийная аудитория, компьютерный класс

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 09.03.03 "Прикладная информатика" и профилю подготовки не предусмотрено .

Автор(ы):

Новиков А.А. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Замалиев Р.Р. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.