

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Костромской государственной университет»

(КГУ)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

## **Прикладные задачи анализа данных**

Направление подготовки – **01.04.02 «Прикладная математика и информатика»**

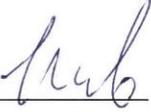
Направленность «**Математическое моделирование и программирование**»

Квалификация (степень) выпускника: магистр

**Кострома  
2024**

Рабочая программа дисциплины «Прикладные задачи анализа данных» разработана в соответствии с Федеральным(и) государственным(и) образовательным(и) стандартом(ами) высшего образования – магистратура по направлению подготовки 01.04.02 Прикладная математика и информатика, приказ №13 от 10 января 2018 года.

Разработал:  Ивков Владимир Анатольевич, доцент, к.э.н., доцент  
подпись

Рецензент:  Козырев Сергей Борисович, доцент, к.ф.-м.н., доцент  
подпись

**УТВЕРЖДЕНО:**

На заседании кафедры прикладной математики и информационных технологий

Протокол заседания кафедры №12 от 22 мая 2019 г.

Заведующий кафедрой прикладной математики и информационных технологий

 Секованов Валерий Сергеевич, профессор, д.п.н., к.ф.-м.н.

**ПРОГРАММА ПЕРЕУТВЕРЖДЕНА:**

На заседании кафедры прикладной математики и информационных технологий

Протокол заседания кафедры №\_6\_\_ от \_\_14.05.2024 г.

Заведующий кафедрой прикладной математики и информационных технологий

\_Ивков В.А.\_\_\_\_\_ к.э.н., доцент\_(ФИО), ученая степень, ученое звание

подпись

## 1. Цели и задачи освоения дисциплины

Изучение дисциплины «Прикладные задачи анализа данных» (блок 1, обязательная часть) важно для магистров направления подготовки 01.04.02 «Прикладная математика и информатика», поскольку они знакомятся с основными методами анализа больших данных и построения на основе анализа данных компьютерных и математических моделей.

При изучении дисциплины «Прикладные задачи анализа данных» формируется профессиональная компетенция (ПК4): (Способен к применению современных методологий разработки программного обеспечения) индикаторы

ПК-4.1. Знать: принципы, методы и средства разработки программного обеспечения с использованием современных технологий;

ПК-4.2. Уметь: разрабатывать прикладные программы в области профессиональной деятельности с применением современных средств разработки;

ПК-4.3. Иметь навыки: применения современных методологий разработки программного обеспечения.

**Цель изучения дисциплины:** «Прикладные задачи анализа данных» – знакомство с распространенными типами прикладных задач анализа данных и схемами их решения.

**Задачи дисциплины:**

– разобрать прикладные задачи из различных областей анализа данных: анализ текста и информационный поиск;

– научить составлять схемы решения задач анализа данных;

– познакомить с некоторыми алгоритмами машинного обучения и прогнозирования временных рядов.

Данный курс носит прикладной характер. В связи с этим для его успешного освоения магистру необходимо владеть одним из языков программирования высокого уровня и уметь работать с математическим пакетом. Выбор языка, системы программирования и математического пакета определяется преподавателем.

Дисциплина по выбору «Прикладные задачи анализа данных» относится к вариативной части основного цикла. В отношении технологического содержания она дополняет дисциплины «Распределенные вычисления на кластере» и «Теория игр и исследование операций». В отношении класса решаемых задач она находится в одном ряду с дисциплиной «Моделирование искусственного интеллекта».

**Магистры, завершившие изучение дисциплины «Прикладные задачи анализа данных» «должны знать:**

– основные идеи анализа данных;

- основные библиотеки языка программирования python;
- некоторые задачи анализа данных и методы их решения в пакете R и в python;
- понятие достоверности результатов модели, подходы к ее проверке.

**Магистры, завершившие изучение дисциплины «Прикладные задачи анализа данных» должны уметь:**

- строить модели анализа данных;
- реализовать модель с помощью различных средств программирования (R или python);
- некоторые задачи анализа данных и методы их решения в пакете R и в python;
- проверять достоверность модели.

**Магистры, завершившие изучение дисциплины «Прикладные задачи анализа данных» должны владеть:**

- математическими методами прикладной математики, включая статистические методы и др.;
- компьютерными технологиями, включая программирование и компьютерную математику;
- компьютерными технологиями создания математических моделей, пользуясь различными средствами программирования (R или python).

**Магистры, завершившие изучение дисциплины «Прикладные задачи анализа данных» должны освоить индикаторы компетенции ПК4: ПК4.1, ПК4.2, ПК4.3, с помощью которых формируется компетенция ПК4:**

ПК-4.1. Знать: принципы, методы и средства разработки программного обеспечения с использованием современных технологий

ПК-4.2. Уметь: разрабатывать прикладные программы в области профессиональной деятельности с применением современных средств разработки

ПК-4.3. Иметь навыки: применения современных методологий разработки программного обеспечения

При изучении дисциплины «Прикладные задачи анализа данных» планируется у обучаемого развивать гибкость, критичность мышления, интуицию и способность преодолевать стереотипы мышления.

## **2. Перечень планируемых результатов обучения по дисциплине**

## **«Прикладные задачи анализа данных»**

В результате освоения дисциплины обучающийся должен:

Освоить индикаторы ПК4:

ПК-4.1. Знать: принципы, методы и средства разработки программного обеспечения с использованием современных технологий

ПК-4.2. Уметь: разрабатывать прикладные программы в области профессиональной деятельности с применением современных средств разработки

ПК-4.3. Иметь навыки: применения современных методологий разработки программного обеспечения.

На базе ПК-4.1–ПК-4.3 обучающийся должен освоить компетенцию ПК4: «Способен к применению современных методологий разработки программного обеспечения».

### **3. Место дисциплины в структуре ОП ВО**

Дисциплина по выбору «Прикладные задачи анализа данных» изучается в третьем семестре. Для изучения дисциплины «Прикладные задачи анализа данных» необходимы знания, умения и навыки, формируемые изучаемыми магистрами, изучаемыми в первом семестре дисциплинами «Математика анализа данных», «Технологии программирования».

Изучение дисциплины «Прикладные задачи анализа данных» является основой для освоения последующих дисциплин: «Моделирование искусственного интеллекта», «Управление разработкой программных продуктов», практики «Научно-исследовательская работа».

Дисциплина «Прикладные модели анализа данных» интегрирует с дисциплинами «Распределенные вычисления на кластере», «Моделирование искусственного интеллекта», Данная интеграция включает в себя логическую и содержательную взаимосвязь, поскольку при ее изучении используются как математические методы, так и компьютерные технологии.

Компетенция ПК-4 дополнительно формируется практикой «Научно-исследовательская работа», защитой выпускной квалификационной работой.

### **4. Объем дисциплины**

#### **4.1. Объем дисциплины в зачетных единицах с указанием академических часов и виды учебной работы**

Виды учебной работы,	Очная форма	Очно-заочная	Заочная
Общая трудоемкость в зачетных единицах	2		
Общая трудоемкость в часах	72		
Аудиторные занятия в часах, в том числе:	16		
Лекции	8		
Практические занятия	8		
Лабораторные занятия	-		
Самостоятельная работа в часах	56		
Форма промежуточной аттестации	(зачет)		

## 4.2. Объем контактной работы на 1 обучающегося

Виды учебных занятий	Очная форма	Очно-заочная	Заочная
Лекции	8		
Практические занятия	8		
Лабораторные занятия	-		
Консультации	-		
Зачет/зачеты	-		
Экзамен/экзамены	зачет		
Курсовые работы			
Курсовые проекты			
Всего	16		

## 5 Содержание дисциплины, структурированное по темам (разделам), с указанием количества часов и видов занятий

### 5.1 Тематический план учебной дисциплины

№	Название раздела, темы	Всего з.е/час	Аудиторные занятия			Самостоятельная работа
			Лекц.	Практ.	Лаб.	
1	Схемы решения задач анализа данных	0,25/18	2	2	-	14
2	Прогнозирование временных рядов	0,25/18	2	2	-	14
3	Анализ текста и информационный поиск	0,25/8	2	2	-	14
4	Алгоритмы машинного обучения	0,25/18	2	2	-	14
	Итого:	2/72	8	8	-	56

### 5.2. Содержание:

**Тема 1. Схемы решения задач анализа данных.** История анализа данных. Формализация понятия «среднее». Среднее арифметическое, медиана, многомерная медиана как результат итерационного процесса. Среднее как решение оптимизационной задачи. Оценка вероятности. Весовые схемы.

**Тема 2. Прогнозирование временных рядов.** Построение временного ряда. Регрессионный анализ временного ряда. Сглаживание временных рядов. Компоненты временного ряда. Смыкание рядов.

**Тема 3. Анализ текста и информационный поиск.** Построение модели анализа текста. Методы обработки больших массивов неструктурированной информации, в том числе текста. Кластеризация, квалификация и фильтрация документов.

**Тема 4. Алгоритмы машинного обучения.** Алгоритмы анализа сложных наборов данных. Системы рекомендации. Построение логистической регрессионной модели, алгоритмы k ближайших соседей, дерево решений и информационный лес.

## 6. Методические материалы для обучающихся по освоению дисциплины

### 6.1. Самостоятельная работа обучающихся по дисциплине (модулю)

№ п/п	Раздел (тема) дисциплины	Задание	Часы	Методические рекомендации по выполнению задания	Форма контроля
1.	Схемы решения задач анализа данных	Изучение теоретического материала.	4	Используйте литературу [1], [3]	Устный опрос
2.	Прогнозирование временных рядов	Изучение литературы, составление компьютерных программ.	7	Используйте литературу [1], [3], [4]	Реферат
3.	Анализ текста и информационный поиск	Изучение литературы, составление компьютерных программ, решение задач.	15	Используйте литературу [3], [4]	Индивидуальное собеседование, проверка домашних заданий, контрольная работа
4.	Алгоритмы машинного обучения	Изучение литературы, разработка индивидуального проекта.	15	Используйте литературу [1], [4]	Индивидуальное собеседование, проверка домашних заданий

### 6.2. Тематика и задания для практических занятий

#### Тематика и задания для практических занятий

**Практическое занятие 1.** Обработка большого массива данных. Построение алгоритма поиска среднего значения и других статистических характеристик массива данных. Применение языка R. Библиотеки языка для обработки больших данных.

**Практическое занятие 2.** Обработка большого массива данных. Построение временного ряда. Построение тренда, выделение компонент временного ряда. Сглаживание ряда. Использование библиотек языка

программирования R.

**Практическое занятие 3.** Приемы обработки больших массивов данных для выполнения задач кластеризации документов. Квалификация документов по ряду признаков. Информационный поиск в массиве данных с помощью библиотек R.

**Практическое занятие 4.** Построение логистической регрессионной модели. Построение дерева решений.

## **7. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины**

*а) основная:*

1. *Горяинова Е.Р.* Прикладные методы анализа статистических данных: Учебное пособие. – М.: ИД ГУ ВШЭ, 2012.
2. *Кабаков Р.* Анализ и визуализация данных в R. – М.: ДМК, 2016.
3. *Кулаичев А.П.* Методы и средства комплексного анализа данных. Учебное пособие. – М.: Форум, 2013.

*б) дополнительная:*

4. *Тюрин Ю.Н.* Анализ данных на компьютере. – М.: МЦНМО, 2016.
5. *Айзек М.П.* Графика, формулы, анализ данных в Excel. Пошаговые примеры. – СПб.: Наука и техника, 2019.

## **8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

*Информационно-образовательные ресурсы:*

1. Библиотека ГОСТов. Все ГОСТы, [Электронный ресурс],

URL:<http://vsegost.com/>

*Электронные библиотечные системы:*

1. ЭБС Университетская библиотека онлайн - <http://biblioclub.ru>
2. ЭБС «Лань» <https://e.lanbook.com>
3. ЭБС «ZNANIUM.COM» <http://znanium.com>

## **9. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Аудитория 228Е для лекционных, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Оборудование: посадочные места 16, рабочее место преподавателя. Имеется мультимедиа – компьютер (переносной) с проектором. Установлено 16 компьютеров.

Аудитория 227Е для лекционных, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Оборудование: посадочные места 15, рабочее место преподавателя.

Имеется мультимедиа – компьютер (переносной) с проектором. Установлено 15 компьютеров.

Лицензионное программное обеспечение:

Windows 8 Pro лицензия 01802000875623 постоянная 1-шт.; LibreOffice 5.0, лицензия GNU LGPL; Microsoft VisualStudio 2013, лицензия; РТСMathCadPrime 1.0, лицензия; JavaSE 1.8, лицензия GNU LGPL;

Свободно распространяемое программное обеспечение:

Языки программирования R, python, офисный пакет.