

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Костромской государственный университет»
(КГУ)

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
АНАЛИЗ ДАННЫХ**

Направление подготовки: 44.03.05 Педагогическое образование
(с двумя профилями)

Направленности: Информатика, 3D-технологии и робототехника

Квалификация выпускника: бакалавр

**Кострома
2023**

Рабочая программа дисциплины «Анализ данных» разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению 44.03.05 Педагогическое образование (с двумя профилями подготовки) (уровень бакалавриата), утвержденным приказом Министерства образования и науки Российской Федерации от 22.02.2018 № 125 (зарегистрировано Министерством юстиции Российской Федерации 15.03.2018 регистрационный № 50358), с изменениями, внесенными приказом Министерства науки и высшего образования Российской Федерации от 08.02.2021 № 83 (зарегистрировано Министерством юстиции Российской Федерации 12.03.2021 регистрационный № 62739); в соответствии с учебным планом направления подготовки 44.03.05 Педагогическое образование (с двумя профилями подготовки) (направленности Информатика, 3D-технологии и робототехника), год начала подготовки 2023, 2024.

Разработал: Ивков Владимир Анатольевич, доцент кафедры прикладной математики и информационных технологий, к.э.н., доцент

Рецензент: Меркурьева Н. В., руководитель центра цифрового образования IT-куб, к. т. н., доцент

УТВЕРЖДЕНО:

Заведующий кафедрой высшей математики:

Матыцина Т. Н., к. ф.-м. н., доцент

Протокол заседания кафедры № 8 от 04.07.2024 г.

1. Цели и задачи освоения дисциплины

Цель дисциплины: сформировать математический фундамент для освоения методов анализа больших данных.

Задачи дисциплины:

повторить основные концепции математического анализа, линейной алгебры, методов оптимизации, теории вероятностей и математической статистики;

– овладеть принципами и понятийным аппаратом, описывающими современные методы анализа данных;

– усвоить теоретические основы современных технологий и методов анализа данных

Кроме того, одной из задач изучения данного курса является научно-образовательное, профессионально-трудовое воспитание обучающихся посредством содержания дисциплины и актуальных воспитательных технологий.

2. Перечень планируемых результатов обучения по дисциплине

В результате освоения дисциплины обучающийся должен освоить компетенции:

ОПК-8 - Способен осуществлять педагогическую деятельность на основе специальных научных знаний.

Код и содержание индикаторов компетенции

ИОПК-8.1. Демонстрирует владение системой специальных научных знаний в предметной области;

ИОПК-8.2. Применяет специальные предметные знания в педагогической деятельности по направленности программы;

ОПК-9 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

Код и содержание индикаторов компетенции

ИОПК-9.1. Должен знать и понимать принципы работы современных информационных технологий;

ИОПК-9.2. Уметь использовать современные информационные системы и технологии в решении профессиональных задач;

ИОПК-9.3. Иметь навыки работы с современным общесистемным и офисным программным обеспечением, в т.ч. отечественного производства;

ИОПК-9.4. Иметь навыки обеспечения информационной безопасности при работе с современными информационными системами и технологиями.

В результате освоения дисциплины обучающийся должен

Знать:

– основные концепции математического анализа, линейной алгебры, математической статистики;

– основные литературные источники, в том числе интернет-ресурсы, отражающие современный уровень развития методов анализа данных;

Уметь:

– составить математическую модель анализа данных;

– подбирать необходимые программные инструменты для решения вычислительных задач.

Владеть:

– методами и навыками решения задач анализа данных.

3. Место дисциплины в структуре ОП ВО

Дисциплина относится к *обязательной* части учебного плана.

Изучается в 7 семестре на очной форме обучения, на 5 курсе заочной формы обучения.

Изучение дисциплины основывается на ранее освоенных дисциплинах/практиках: «Математика», «Языки и методы программирования», «Теория вероятностей и математическая статистика».

Изучение дисциплины является основой для освоения следующих дисциплин: «Машинное обучение и искусственный интеллект».

4. Объем дисциплины (модуля)

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

Виды учебной работы,	Очная форма	Заочная форма
Общая трудоемкость в зачетных единицах	4	4
Общая трудоемкость в часах	144	144
Аудиторные занятия в часах, в том числе:	18	6
Лекции	8	2
Практические занятия	0	0
Лабораторные занятия	10	4
Практическая подготовка	0	0
Самостоятельная работа в часах	87,65	126,65
Контроль	36 + 2,35	9 + 2,35
Форма промежуточной аттестации:	Экзамен в 7 семестре	Экзамен

4.2. Объем контактной работы на 1 обучающегося

Виды учебных занятий	Очная форма	Заочная форма
Лекции	8	2
Практические занятия	0	0
Лабораторные занятия	10	4
Консультации	2	2
Зачет/зачеты	0	0
Экзамен/экзамены	0,35	0,35
Курсовые работы	0	0
Курсовые проекты	0	0
Практическая подготовка	0	0
Всего	20,35	8,35

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

5.1 Тематический план учебной дисциплины

Очная форма обучения

№ п/п	Наименование темы	Всего часов	Аудиторные занятия			Сам. раб.
			Лекц.	Практ.	Лаб.	
	5 семестр					
1	Машинное обучение и анализ данных	22	2	-	2	18
2	Библиотеки python и	22	2	-	2	18

	линейная алгебра					
3	Оптимизация и матричные разложения	22	2	-	2	18
4	Теория вероятностей	22	2	-	2	18
5	Статистика	17,65	-	-	2	15,65
	ИКР	2,35	-	-	-	-
	Итого в 7 семестре	108	8	-	10	87,65 + 2,35
	Экзамен	36	-	-	-	36
	Итого:	144	8	-	10	87,65 + 2,35 + 36

Заочная форма обучения

№ п/п	Наименование темы	Всего часов	Аудиторные занятия			Сам. раб.
			Лекц.	Практ.	Лаб.	
	5 курс					
1	Машинное обучение и анализ данных	28	2	-	-	26
2	Библиотеки python и линейная алгебра	27	-	-	1	26
3	Оптимизация и матричные разложения	27	-	-	1	26
4	Теория вероятностей	26	-	-	1	25
5	Статистика	24,65	-	-	1	23,65
	ИКР	2,35	-	-	-	-
	Экзамен	9	-	-	-	9
Итого:		144	2	-	4	126,65 + 2,35 + 9

5.2. Содержание

Тема 1. Машинное обучение и анализ данных. Анализ данных и машинное обучение опираются на результаты из математического анализа, линейной алгебры, методов оптимизации, теории вероятностей. Без фундаментальных знаний по этим наукам невозможно понять, как строены методы анализа данных. Для успешного применения методов анализа данных нужно уметь программировать. Наиболее удобным языком анализа данных является Python.

Тема 2. Библиотеки python и линейная алгебра. Реализация операций с многомерными массивами данных. Pandas. Data Frame. Индексация и селекция. Знакомство с NumPy, SciPy и Matplotlib. Решение оптимизационных задач в SciPy. Векторные пространства. Матричные операции..

Тема 3. Оптимизация и матричные разложения. Нахождение наилучших значений параметров системы. Минимизация затрат и максимизация точности предсказаний. Матричные разложения при построении регрессионных моделей. Уменьшение размерности данных. Частные

производные и градиент. Применение градиента. Производные по направлению. Касательная плоскость и линейное приближение. Направление наискорейшего роста. Оптимизация негладких функций. Метод имитации отжига. Генетические алгоритмы и дифференциальная эволюция. Разложение матриц в произведение. Сингулярное разложение. Приближение матрицей меньшего ранга. Связь сингулярного разложения и приближения матрицей меньшего ранга.

Тема 4. Теория вероятностей. Базовые концепции теории вероятностей. Случайность в теории вероятностей и статистике. Свойства вероятности. Условная вероятность. Дискретные и случайные величины. Непрерывные случайные величины. Центральная предельная теорема.

Тема 5. Статистика. Популярные распределения. Статистики оценки параметров. Доверительные интервалы. Оценка распределения по выборке. Важные характеристики распределений.

6. Методические материалы для обучающихся по освоению дисциплины

6.1. Самостоятельная работа обучающихся по дисциплине (модулю)

№ п/п	Наименование темы	Задание	Кол-во часов		Формы текущего контроля
1	Машинное обучение и анализ данных	Основные методы анализа больших данных	18	26	Устный опрос
2	Библиотеки python и линейная алгебра	Подключение и использование статистических библиотек	18	26	Индивидуальное собеседование,
3	Оптимизация и матричные разложения	Решение задач линейной алгебры с помощью среды программирования	18	26	Устный опрос, контрольная работа
4	Теория вероятностей	Решение задач теории игр с помощью Python	18	25	Устный опрос
5	Статистика	Решение статистических задач с помощью Python	15,65	23,65	Индивидуальное собеседование,
	Экзамен	Подготовка	36	9	

6.2. Тематика и задания для лабораторных занятий

Тема 1. Машинное обучение и анализ данных. Основные задачи математического анализа. Нахождение производных. Задачи линейной алгебры для анализа данных. Матричные вычисления. Задачи методов оптимизации. Задачи теории вероятностей и математической статистики.

Тема 2. Библиотека Python и линейная алгебра. Операции с матрицами с помощью библиотек NumPy, SciPy и Matplotlib. Решение оптимизационных задач в SciPy. Векторные пространства. Разрешимость систем линейных уравнений и ранги.

Тема 3. Оптимизация и матричные разложения. Нахождение частных производных. Применение градиента. Повторение: гладкость и градиентный спуск. Методы оптимизации в негладких задачах. Линейная алгебра. Повторение. Матричные разложения.

Тема 4. Теория вероятностей. Работа со случайными величинами (ipython notebook).

Тема 5. Статистика. Анализ больших данных. Построение выборки. Выбор распределения. Оценка распределения по выборке (ipython notebook).

6.3. Тематика и задания для практических занятий

Практические занятия отсутствуют.

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

№	Литература	Кол-во книг
Основная литература		
1	Прерва А., Иванова В. Путь аналитика. Практическое руководство IT-специалиста. – СПб : Питер, 2015. - 304 с.	10
2	Тыртышников, Е. Е. Матричный анализ и линейная алгебра. - М. : Физматлит, 2007. - 480 с.	10
3	Гмурман, В.Е. Теория вероятностей и математическая статистика : [учеб. пособие для студ.]. - М. : Высшая школа, 2002. - 480 с.	10
4	Вентцель, Е. С. Исследование операций: задачи, принципы, методология : учеб. пособие для студ. вузов, напр. "Математика", "Компьютерные науки". - 4-е изд., стер. - Москва : Дрофа, 2006.	10
5	Васин, А. А. Исследование операций : [учеб. пособие для студ. высш. учеб. заведений] : рекомендовано Науч.-метод. советом. - М. : Академия, 2008. - 464 с.	
Дополнительная литература		
1	Волков, И. К. Исследование операций : Учеб. для студ. высш. техн. учеб. заведений / Под ред. В. С. Зарубина, А. П. Крищенко. 3 е изд., стер. М. : МГТУ, 2004. 440 с.	
2	Тюрин Ю. Н. Анализ данных на компьютере : учеб. пособие для вузов / под ред. В. Э. Фигурнова. - Москва : ИНФРА-М; Финансы и статистика, 1995. - 384 с.	

3	Брандт, З. Анализ данных : стат. и вычисл. методы для науч. работников и инженеров / [пер. с англ. О. И. Волковой ; под ред. Е. В. Чепурина]. - М. : Мир : АСТ, 2003. - 686 с.	
---	--	--

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

Для проведения занятий по дисциплине необходимы учебная аудитория – компьютерный класс, оснащенный проектором или электронной доской. Необходимое программное обеспечение – Офисный пакет; Среда программирования рuscharm.