

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
АНАЛИЗ БОЛЬШИХ ДАННЫХ

Направление подготовки:
38.03.05 Бизнес-информатика

Направленность:
Бизнес-аналитика

Квалификация выпускника: бакалавр

Кострома
2023

Рабочая программа дисциплины «Анализ больших данных» разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению подготовки 38.03.05 Бизнес-информатика (приказ Минобрнауки России от 29 июля 2020 г. № 838).

Разработал: Илюхина Анна Святославовна, зав. кафедрой бизнес-информатики и сервиса КГУ,
к.э.н. доцент

Рецензенты:

Игнатъев Сергей Николаевич

заместитель генерального директора ИТ -
компании ООО «Гелиос-С»

Румянцев Дмитрий Сергеевич

Генеральный директор ООО «НПП
Ювелирсофт»

ПРОГРАММА УТВЕРЖДЕНА

на заседании кафедры бизнес-информатики и сервиса
(протокол от 20 апреля 2023 г. № 8)

ПРОГРАММА ПЕРЕУТВЕРЖДЕНА

на заседании кафедры бизнес-информатики и сервиса
(протокол от _____ 202__ г. № _____)

1. Цели и задачи освоения дисциплины

Цель дисциплины: научить студентов осваивать и применять информационные технологии анализа данных в выбранной предметной области.

Задачи дисциплины: Изучить:

- основные принципы анализа данных
- виды и типы данных
- методы анализа данных
- статистические методы анализа данных
- понятие интеллектуального анализа данных
- методы datamining
- программные средства интеллектуального анализа данных.

2. Перечень планируемых результатов обучения по дисциплине

В результате освоения дисциплины обучающийся должен:

освоить компетенции:

ПК-3. Способен осуществлять разработку и исследование математических и компьютерных моделей поддержки принятия решений в экономике и бизнесе

Код и содержание индикаторов компетенции:

ИК.ПК-3.4. Владеет инструментами анализа больших данных

1) Знать:

- методы анализа данных.
- информационные технологии, применяемые для анализа данных.
- концепцию интеллектуального анализа данных.
- информационные технологии организации, хранения представления и обработки данных.

2) Уметь:

- использовать информационные технологии и методы анализа данных с

Применением ЭВМ.

- самостоятельно осваивать обновленные и вновь появившиеся информационные технологии анализа данных.

3) Владеть:

- навыками применения и использования методов и программных средств анализа данных для решения задач в предметной области.

3. Место дисциплины в структуре ОП ВО

Дисциплина относится к вариативной части учебного плана. Изучается в 4 семестре в соответствии с учебным планом на очной форме .

Изучение дисциплины основывается на ранее освоенных дисциплинах 1 курса: экономико-математическое моделирование, Информационные системы в экономике и управлении.

Изучение дисциплины является основой для освоения дисциплины методы оптимизации и прогнозирования и прохождения преддипломной практики.

4. Объем дисциплины

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

Виды учебной работы,	Очная форма
Общая трудоемкость в зачетных единицах	3
Общая трудоемкость в часах	108
Аудиторные занятия в часах, в том числе:	50
Лекции	16
Практические занятия	34
Лабораторные занятия	
Практическая подготовка	
Самостоятельная работа в часах	57,75
Форма промежуточной аттестации	Зачет

4.2. Объем контактной работы на 1 обучающегося

Виды учебных занятий	Очная форма
Лекции	16
Практические занятия	34
Лабораторные занятий	
Консультации	
Зачет/зачеты	0,25
Экзамен/экзамены	
Курсовые работы	
Курсовые проекты	
Практическая подготовка	
Всего	50,25

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

5.1 Тематический план учебной дисциплины

Очная форма

№	Название раздела, темы	Все-го з.е./ час	Аудиторные занятия			Самостоятельная работа
			лекции	лабор.	практические	
1	Первичная обработка данных и формы выражения показателей, показатели структуры данных	8	2		2	4
2	Методы формирования выборок.	10	2		2	6
3	Регрессионный анализ	12	2		4	6
4	Байесовская классификация	14	2		4	8
5	Введение в аффинитивный анализ.	16	2		6	8
6	Кластерный анализ.	18	2		6	10
7	Классификация и регрессия.	18	2		6	10
8	Введение в ансамбли моделей.	11,75	2		4	5,75
	всего	107,75	16		34	57,75

5.2. Содержание:

ТЕМА 1. ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ И ФОРМЫ ВЫРАЖЕНИЯ ПОКАЗАТЕЛЕЙ, ПОКАЗАТЕЛИ СТРУКТУРЫ ДАННЫХ

Типы данных (качественные и количественные). Методы сводки и группировки. Шкалы измерения данных. Виды статистических наблюдений (охват, систематичность и источник).

Принципы группировки, формула Стержесса.

Графическое представление данных: знак Варзара, круговые, квадратные и столбчатые диаграммы; секторные и радиальные диаграммы.

Описательные статистики признака: показатели центра распределения для сгруппированных и несгруппированных данных, порядковые характеристики. Применение при изучении социально-экономических процессов. Частные показатели структурных сдвигов. Обобщающие показатели структурных сдвигов. Показатели концентрации и централизации при изучении признаков.

ТЕМА 2. МЕТОДЫ ФОРМИРОВАНИЯ ВЫБОРОК

Методы формирования выборок и определение их необходимого объема в зависимости от цели проводимого исследования.

Простая случайная, стратификационная и серийная выборки.

ТЕМА 3. РЕГРЕССИОННЫЙ АНАЛИЗ

Линейная множественная и однофакторная регрессия, проверка качества регрессионных уравнений, проверка достоверности оцененных параметров регрессии (несмещенность, состоятельность; эффективность).

Устранение эффекта гетероскедастичности и автокорреляции в остатках регрессионных моделей.

Интерпретация полученных результатов моделирования: приростной и маржинальный анализ. Нелинейная регрессия. Методы подбора вида функции нелинейной регрессии.

Итеративные процедуры оценки параметров нелинейной регрессии. Производственные функции, как пример нелинейной регрессии (Кобба-Дугласа, Солоу, Тинбергена).

Показатели предельной нормы замены, средние и маржинальные величины. Коэффициенты эластичности

ТЕМА 4. БАЙЕСОВСКАЯ КЛАССИФИКАЦИЯ

Модель наивного байесовского классификатора. Оценка параметров для наивных байесовских моделей методом максимального правдоподобия. Построение классификатора по вероятностной модели. Апостериорное правило принятия решения (MAP).

Применение наивного байесовского классификатора для задач фильтрации.

ТЕМА 5. ВВЕДЕНИЕ В АФФИНТИВНЫЙ АНАЛИЗ

Поиск последовательных шаблонов. Алгоритм a priori. Иерархические ассоциативные правила.

ТЕМА 6. КЛАСТЕРНЫЙ АНАЛИЗ

Введение в кластерный анализ. Классификация методов кластерного анализа. Алгоритм кластеризации – k-means/. Сети Кохонена. Карты Кохонена. Проблемы алгоритмов кластеризации.

ТЕМА 7. КЛАССИФИКАЦИЯ И РЕГРЕССИЯ

Машинное обучение. Деревья решений. Алгоритмы построения деревьев решений. Алгоритмы ID3 и C4.5. Введение в нейронные сети

ТЕМА 8. АНСАМБЛИ МОДЕЛЕЙ

Обзор альтернатив ансамблей моделей.

6. Методические материалы для обучающихся по освоению дисциплины

6.1. Самостоятельная работа обучающихся по дисциплине (модулю)

Очная форма

№ п/п	Раздел (тема)	Задание	Часы	Методические рекомендации по выполнению задания	Форма контроля
1	Первичная обработка данных и формы выражения показателей, показатели структуры данных	Подбор и обзор литературы по проблематике	4	Работа с учебной и научной литературой Сбор, систематизация, классификация и обобщение информации по обсуждаемым вопросам	Представление эссе (доклада) с последующим обсуждением
2	Методы формирования выборок.	Привести примеры из практики и обосновать целесообразность работы с простой случайной, стратификационной и серийной выборками	6	Рассмотрение источников экономической литературы: экономическое обоснование основных понятий, категорий и законов; формулы расчётов основных показателей?	Проверка конспекта
3	Регрессионный анализ	Оформление практической-расчетной работы по теме	6	Использовать рассмотренные примеры [2]	Защита работы
4	Байесовская классификация	Оформление практической-	8	Использовать рассмотренные примеры [2]	Практикум по решению типовых

		расчетной работы по теме			заданий
5	Введение в аффинитивный анализ.	Творческое задание на применение Алгоритм a priori	8	Использовать рассмотренные примеры [2]	Проверка представленной работы
6	Кластерный анализ.	Оформление практической работы по теме	10	Использовать рассмотренные примеры [2]	Защита работы
7	Классификация и регрессия.	Введение в нейронные сети	10	Использовать рассмотренные примеры [2]	Представление доклада с последующим обсуждением
8	Введение в ансамбли моделей.	Изучить альтернативные рассмотренным на занятии ансамбли моделей	5,75	Работа с учебной и научной литературой Сбор, систематизация, классификация и обобщение информации по обсуждаемым вопросам	Представление доклада с последующим обсуждением

6.2. Тематика и задания для практических занятий

ТЕМА 1. ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ И ФОРМЫ ВЫРАЖЕНИЯ ПОКАЗАТЕЛЕЙ, ПОКАЗАТЕЛИ СТРУКТУРЫ ДАННЫХ

Проведение процедуры группировки, расчет необходимого числа групп на основе формулы Стержесса. Графическое представление данных

ТЕМА 2. МЕТОДЫ ФОРМИРОВАНИЯ ВЫБОРОК

Формирование выборок и определение их необходимого объема в зависимости от цели проводимого исследования

Расчет средних показателей рядов динамики (средних хронологических), частных и обобщающих показателей структурных сдвигов, показателей концентрации (Лоренца и Джини) и централизации при изучении признаков.

ТЕМА 3. РЕГРЕССИОННЫЙ АНАЛИЗ

Расчет коэффициентов эластичности, маржинальных и средних величин, показателей предельной технологической нормы замены для уравнений нелинейной регрессии.

ТЕМА 4. БАЙЕСОВСКАЯ КЛАССИФИКАЦИЯ

Применение модели наивного байесовского классификатора. Оценка параметров для наивных байесовских моделей методом максимального правдоподобия. Построение классификатора по вероятностной модели. Апостериорное правило принятия решения (MAP).

ТЕМА 5. ВВЕДЕНИЕ В АФФИНТИВНЫЙ АНАЛИЗ

Реализация алгоритм a priori. Иерархические ассоциативные правила.

ТЕМА 6. КЛАСТЕРНЫЙ АНАЛИЗ

Реализация алгоритма кластеризации – k-means.

ТЕМА 7. КЛАССИФИКАЦИЯ И РЕГРЕССИЯ

Реализация алгоритмов построения деревьев решений. Алгоритмы ID3 и C4.5.

ТЕМА 8. БАЙЕСОВСКАЯ КЛАССИФИКАЦИЯ

Обзор альтернативных ансамблей моделей.

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

а) Основная литература:

1. Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2014. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426> (08.12.2018).
2. Анализ данных и процессов: Учебное пособие / Барсемян А.А., Куприянов М.С., Холод И.И. - СПб:БХВ-Петербург, 2009. - 512 с. ISBN 978-5-9775-0368-6 - Режим доступа: <http://znanium.com/catalog/product/35063>

б) Дополнительная литература:

3. Статистический анализ данных в MS Excel: учеб. пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. –М. : ИНФРА-М, 2017. –320 с. –(Высшее образование: бакалавриат). — www.dx.doi.org/10.12737/2842.
<http://znanium.com/catalog.php?bookinfo=858510>

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

Электронные библиотечные системы:

1. ЭБС «Лань»
2. ЭБС «Университетская библиотека online» <http://www.biblioclud.ru>
3. ЭБС «Znaniium»

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

1. Мультимедийное оборудование.
2. Компьютерный класс.

Программное обеспечение:

- LibreOffice [GNU LGPL v3+](#) ([свободно распространяемый офисный пакет с открытым исходным кодом](#));
- OpenOffice [Apache License 2.0](#) ([свободный пакет офисных приложений](#));
- Adobe Acrobat Reader (Бесплатная программа для просмотра документов в формате PDF)