

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Курганский государственный университет»
(КГУ)

Кафедра «Программное обеспечение автоматизированных систем»



УТВЕРЖДАЮ:
Первый проректор
/ Т.Р. Змызгова/
Т.Р. Змызгова 2022 г.

Рабочая программа учебной дисциплины

МЕТОДЫ И АЛГОРИТМЫ АНАЛИЗА ДАННЫХ

образовательной программы высшего образования –
программы бакалавриата

09.03.04 Программная инженерия

Направленность:

Программное обеспечение автоматизированных систем

Форма обучения: очная, заочная

Курган 2022

Рабочая программа дисциплины «Методы и алгоритмы анализа данных» составлена в соответствии с учебным планом по программе бакалавриата «Программная инженерия» (Программное обеспечение автоматизированных систем), утвержденными для очной формы обучения «30» августа 2022 года, для заочной формы обучения «30» августа 2022 года.

Рабочая программа дисциплины одобрена на заседании кафедры «Программное обеспечение автоматизированных систем» 30 августа 2022, протокол № 1.

Рабочую программу составил:

к.п.н., доцент



А.А. Медведев

Согласовано:

Заведующий
кафедрой ПОАС



В.К. Волк

Начальник управления
образовательной
деятельности



И.В. Григоренко

Специалист
по учебно-методической работе
учебно-методического отдела



Г.В. Казанкова

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Интеллектуальный анализ данных» к обязательной части модуля «Математические и естественно-научные дисциплины» блока 1.

Изучение дисциплины базируется на результатах обучения, сформированных при изучении следующих дисциплин: «Информатика», «Основы программирования», «Теория вероятностей и математическая статистика», «Основы теории информации».

Результаты обучения по дисциплине необходимы при выполнении научно-исследовательской работы, а также при подготовке выпускной квалификационной работы.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Цель освоения дисциплины: изучение современных математических методов и получение практических навыков анализа исходной информации, применительно к задачам, возникающим при исследовании сложных систем.

Задачи дисциплины: используя ресурсы образовательной программы, университетского образовательного пространства, профессионального сообщества способствовать формированию у студентов навыков решения прикладных задач, связанных с анализом данных, обработкой результатов эксперимента с использованием современных программных средств и технологий.

Компетенции, формируемые в результате освоения дисциплины.

- Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1);
- Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности (ОПК-1);
- Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности. (ОПК-2);
- Способен применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой (ОПК-7).

В результате изучения дисциплины обучающийся должен

знать:

- современные методы и программные средства в профессиональной деятельности (для ОПК-7);

уметь:

- осуществлять поиск, применять анализ, синтез и системный подход для решения поставленных задач (для УК-1);
- определять круг задач, требуемых решения для достижения поставленной цели (для ОПК-1);

владеть:

1. ОБЪЕМ ДИСЦИПЛИНЫ

Всего: 3 зачетные единицы трудоемкости (108 академических часа)

Очная форма обучения

Вид учебной работы	На всю дисциплину	Семестр
		5
Аудиторные занятия (контактная работа с преподавателем), всего часов	64	64
в том числе:		
Лекции	32	32
Лабораторные работы	32	32
Практические работы	-	-
Аудиторные занятия в интерактивной форме, часов		
Самостоятельная работа, всего часов	44	44
в том числе:		
Подготовка к экзамену	27	27
Подготовка к зачету	-	-
Подготовка контрольной работы	8	8
Подготовка курсовой работы	-	-
Другие виды самостоятельной работы (самостоятельное изучение тем (разделов) дисциплины)	9	9
Вид промежуточной аттестации	Экзамен	Экзамен
Общая трудоемкость дисциплины и трудоемкость по семестрам, часов	108	108

Заочная форма обучения

Вид учебной работы	На всю дисциплину	Семестр
		7
Аудиторные занятия (контактная работа с преподавателем), всего часов	12	12
в том числе:		
Лекции	6	6
Лабораторные работы	6	6
Практические работы	-	-
Аудиторные занятия в интерактивной форме, часов		
Самостоятельная работа, всего часов	96	96
в том числе:		
Подготовка к экзамену	27	27
Подготовка к зачету	-	-
Подготовка контрольной работы	18	18
Подготовка курсовой работы	-	-
Другие виды самостоятельной работы (самостоятельное изучение тем (разделов) дисциплины)	51	51
Вид промежуточной аттестации	Экзамен	Экзамен
Общая трудоемкость дисциплины и трудоемкость по семестрам, часов	108	108

- современными методами и программными средствами для разработки приложений и создания программных прототипов решения прикладных задач (для ОПК-2);

- средствами организации поиска, анализа, синтеза, системного подхода при решении поставленных задач (для УК-1).

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Учебно-тематический план

Очная форма обучения

Рубеж	Номер раздела, темы	Наименование раздела, темы	Количество часов контактной работы с преподавателем	
			Лекции	Лабораторные работы
<i>5 семестр</i>				
Рубеж 1	1.	Интеллектуальный анализ данных	2	
	2.	Классификация и регрессия	4	4
	3.	Поиск ассоциативных правил	4	4
	4.	Кластеризация	4	4
		Рубежный контроль № 1		2
Рубеж 2	5.	Визуальный анализ данных (Visual Mining)	2	4
	6.	Анализ текстовой информации (Text Mining)	4	4
	7.	Стандарты Data Mining	2	
	8.	Data Mining в реальном времени	4	
	9.	Извлечение знаний из Web (Web Maining)	4	4
	10.	Средства анализа процессов (Process Maining)	2	4
		Рубежный контроль № 2		2
Всего			32	32
Итого			32	32

Заочная форма обучения

Номер раздела, темы	Наименование раздела, темы	Количество часов контактной работы с преподавателем	
		Лекции	Лабораторные работы
<i>7 семестр</i>			
2	Классификация и регрессия	2	2
3	Поиск ассоциативных правил	2	2
4	Кластеризация	2	2
Всего		6	6
Итого		6	6

4.2. Содержание лекционных занятий

5 семестр (очная форма обучения), 7 семестр (заочная форма обучения)

Тема 1 Интеллектуальный анализ данных

Добыча данных (Data Mining). Задачи и практическое применение Data Mining. Модели и методы Data Mining. Процесс обнаружения знаний. Управление знаниями. Средства Data Mining.

Тема 2 Классификация и регрессия

Постановка задачи. Представление результатов: правила классификации, деревья решений, математические функции. Методы построения правил классификации: алгоритм построения 1-правил, метод Naive Bayes. Методы построения деревьев решений: методика "разделяй и властвуй", алгоритм покрытия. Методы построения математических функций: метод наименьших квадратов, нелинейные методы, метод опорных векторов (SVM). Прогнозирование временных рядов.

Тема 3 Поиск ассоциативных правил

Постановка задачи. Секвенциальный анализ. Представление результатов. Алгоритмы: алгоритм Apriori, разновидности алгоритма Apriori.

Тема 4 Кластеризация

Постановка задачи кластеризации. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации. Представление результатов. Базовые алгоритмы кластеризации. Классификация алгоритмов. Алгоритм k-means. Адаптивные методы кластеризации.

Тема 5 Визуальный анализ данных (Visual Mining)

Выполнение визуального анализа данных. Характеристики средств визуализации данных. Методы визуализации: методы геометрических преобразований, отображение иконок, методы, ориентированные на пиксели, иерархические образы.

Тема 6 Анализ текстовой информации (Text Mining)

Задача анализа текстов: этапы анализа текстов, предварительная обработка текста, задачи Text Mining. Извлечение ключевых понятий из текста: общее описание процесса извлечения понятий из текста, стадия локального анализа, стадия интеграции и вывода понятий. Классификация текстовых документов. Методы классификации текстовых документов: представление текстовых документов, иерархические методы кластеризации текстов, бинарные методы кластеризации текстов. Задача аннотирования текстов. Средства анализа текстовой информации

Тема 7 Стандарты Data Mining

Области стандартизации Data Mining. Стандарт CWM. Структура и состав CWM. Пакет Data Mining. Стандарт CRISP. Структура стандарта CRISP. Стандарт PMML. Другие стандарты Data Mining: стандарт SQL/MM, стандарт Microsoft Data Mining extensions (DMX), стандарт Java Data Mining.

Тема 8 Data Mining в реальном времени

Использование Data Mining в реальном времени. Адаптивная добыча данных. Применение Data Mining в реальном времени. Рекомендательные машины: классификация рекомендательных машин, подход на основе содержания, совместное фильтрование. Анализ рыночной корзины и секвенциальный анализ. Усиление обучения и агенты. Инструменты Data Mining в реальном времени.

Тема 9 Извлечение знаний из Web (Web Mining)

Проблемы анализа информации из Web. Этапы Web Mining. Web Mining и другие интернет-технологии. Категории Web Mining. Методы извлечения Web-контента. Извлечение Web-контента для формирования баз данных. Извлечение Web-структур. Поиск Web-документов с учетом гиперссылок. Кластеризация Web-структур. Исследование использования Web-ресурсов.

Тема 10 Средства анализа процессов (Process Mining)

Автоматизация выполнения бизнес-процессов: понятие бизнес-процесса, формализация бизнес-процессов, Workflow-системы, проектирование бизнес-процессов. Анализ процессов: технология Process Mining, анализ протоколов, стандарт записи протоколов MXML, задачи Process Mining. Методы Process Mining. Библиотека алгоритмов Process Mining – ProM.

4.3. Лабораторные занятия

Номер раздела, темы	Наименование раздела, темы	Наименование лабораторной работы	Норматив времени, час.	
			Очная форма обучения	Заочная форма обучения
2	Классификация и регрессия	Лабораторная работа № 1. Решение задач классификации и регрессии	4	2
3	Поиск ассоциативных правил	Лабораторная работа № 2. Нахождение взаимосвязей в массивах данных.	4	2
4	Кластеризация	Лабораторная работа № 3. Кластеризация исходных данных	4	2
	Рубежный контроль 1.		2	
5	Визуальный анализ данных (Visual Mining)	Лабораторная работа № 4. Элементы визуального анализа данных	4	

6	Анализ текстовой информации (Text Mining)	Лабораторная работа № 5. Элементы анализа текстовой информации	4	
9	Извлечение знаний из Web (Web Mining)	Лабораторная работа № 6. Работа с Web	4	
10	Средства анализа процессов (Process Mining)	Лабораторная работа № 7. Сети Петри	4	
	Рубежный контроль 2.		2	
Всего			32	6
Итого			32	6

4.5. Контрольная работа

Контрольная работа (для очной и заочной формы обучения) используется для углубления знаний по особенностям обработки экспериментальных данных. Полный перечень вариантов контрольной работы приведен в учебно-методическом комплексе дисциплины.

4.5.1 Требования к оформлению контрольной работы

Работа должна быть выполнена на одной стороне листа белой бумаги формата А4 (210x297 мм). Интервал межстрочный - полуторный. Цвет шрифта - черный. Гарнитура шрифта основного текста — «Times New Roman» или аналогичная. Кегль (размер) от 12 до 14 пунктов. Размеры полей страницы (не менее): правое — 30 мм, верхнее, и нижнее, левое — 20 мм. Формат абзаца: полное выравнивание («по ширине»). Отступ красной строки одинаковый по всему тексту.

Страницы должны быть пронумерованы с учётом титульного листа, который не обозначается цифрой. Интервал между строками текста — 1,5. Размер шрифта для названия главы — 16 (полужирный), подзаголовок — 14 (полужирный), текста работы — 14. Точка в конце заголовка, располагаемого посередине листа, не ставится. Заголовки не подчёркиваются. Абзацы начинаются с новой строки и печатаются с отступом в 1,25 сантиметра. Оглавление (содержание) должно быть помещено в начале работы, сформировано автоматически.

Заголовки разделов и подразделов следует печатать на отдельной строке с прописной буквы без точки в конце, не подчеркивая, например: ВВЕДЕНИЕ, ЗАКЛЮЧЕНИЕ. Выравнивание по центру. Отбивка: перед заголовком — 12 пунктов, после — 6 пунктов. Расстояние между названием главы и последующим текстом должно быть равно двум междустрочным интервалам. Такое же расстояние выдерживается между заголовками главы и параграфа. Расстояния между строками заголовка принимают таким же, как и в тексте. Подчеркивать заголовки и переносить слова в заголовке не допускается.

При **цитировании** необходимо соблюдать следующие правила:

- текст цитаты заключается в кавычки и приводится без изменений, без произвольного сокращения цитируемого фрагмента (пропуск слов, предложений или абзацев допускается, если не влечет искажения всего фрагмента, и обозначается многоточием, которое ставится на месте пропуска) и без искажения смысла;

- каждая цитата должна сопровождаться ссылкой на источник, библиографическое описание которого должно приводиться в соответствии с требованиями библиографических стандартов.

Обучающиеся должны провести самостоятельный поиск в сети Интернет необходимой информации по выбранной теме, провести аналитический обзор и сравнительный анализ, подготовить выступление на 5-7 мин. с докладом и презентацией.

Нумерация. Страницы следует нумеровать арабскими цифрами, соблюдая сквозную нумерацию по всему тексту (титальный лист и оглавление включают в общую нумерацию). На титульном листе номер не проставляют. Номер страницы проставляют в центре нижней части листа без точки.

Титульный лист. В верхней части титульного листа пишется, в какой организации выполняется работа, далее буквами увеличенного кегля указывается тип («Курсовая работа») и вариант работы, ниже в правой половине листа — информация, кто выполнил и кто проверяет работу. В центре нижней части титульного листа пишется город и год выполнения.

Библиография

Библиографические ссылки в тексте курсовой работы оформляются в виде номера источника в квадратных скобках. Библиографическое описание (в списке источников) состоит из следующих элементов:

основного заглавия;

обозначения материала, заключенного в квадратные скобки;

сведений, относящихся к заглавию, отделенных двоеточием;

сведений об ответственности, отделенных наклонной чертой;

при ссылке на статью из сборника или периодического издания — сведений о документе, в котором помещена составная часть, отделенных двумя наклонными чертами с пробелами до и после них;

места издания, отделенного точкой и тире;

имени издателя, отделенного двоеточием;

даты издания, отделенной запятой.

Примеры (см. Примечание).

ПРИМЕЧАНИЕ

Список элементов библиографической записи сокращен

Книга, имеющая не более трех авторов:

Максимов, Н. В. Архитектура ЭВМ и вычислительных систем [Текст]: учеб. для вузов / Н. В. Максимов, Т. Л. Партыка, И. И. Попов. — М.: Инфра, 2005.

Книга с четырьмя и более авторами, сборник и т. п.:

Мировая художественная культура [Текст]: в 2-х т. / Б. А. Эренграсс [и др.]. — М.: Высшая школа, 2005. — Т. 2.

Статья из сборника:

Цивилизация Запада в 20 веке [Текст] / Н. В. Шишова [и др.] // История и культурология: учеб. пособие для студентов. — М, 2000. — Гл. 13. — С. 347-366.

Статья из журнала:

Мартышин, О. В. Нравственные основы теории государства и права [Текст] / О. В. Мартышин // Государство и право. — 2005. — № 7. — С. 5-12.

Электронное издание:

Сидыганов, Владимир Устинович. Модель Москвы [Электронный ресурс]: электронная карта Москвы и Подмосковья / Сидыганов В. У., Толмачев С. Ю., Цыганков Ю. Э. — Версия 2.0. — М.: Formoza, 1998.

Интернет-ресурс:

Бычкова, Л. С. Конструктивизм / Л. С. Бычкова // Культурология 20 век. — (<http://www.philosophy.ru/edu/ref/enc/k.html>).

4.5.2 Примеры тем контрольных работ

Варианты заданий №1

В таблице приведены данные о среднедушевом прожиточном минимуме в день на одного работающего x (в рублях) и данные о средней заработной плате за один рабочий день y (в рублях) в 15-ти регионах.

1. Постройте уравнение парной регрессии y от x .
2. Рассчитайте коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оцените статистическую значимость параметров регрессии и уравнения регрессии с помощью F -критерия Фишера и t -критерия Стьюдента.
4. Найдите доверительные интервалы для коэффициентов регрессии и уравнения регрессии на уровне значимости $\alpha = 0,05$.
5. Найдите и удалите из выборки две точки, наиболее удалённые от линии регрессии. Постройте линию регрессии для этой выборки. Сравните результаты.

Вариант 1		Вариант 2		Вариант 3	
Прожиточный минимум	Зарплата	Прожиточный минимум	Зарплата	Прожиточный минимум	Зарплата
234	445	215	486	206	369
246	484	226	531	216	410
261	518	239	569	229	441
237	457	217	500	208	383
267	524	245	574	235	443
318	623	292	682	279	526
201	396	184	433	177	335

264	517	242	566	232	436
219	434	201	476	192	369
261	517	239	567	229	439
228	449	209	492	200	380
345	685,	316	752	303	583
207	419	190	460	182	360
252	526	231	579	221	459
276	553	253	607	242	472

Варианты заданий №2

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (смотри таблицу своего варианта).

Требуется:

1. Построить линейную модель множественной регрессии и выполнить анализ результатов.

Вариант 1

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	11	9	6,3	21
2	6	3,6	12	12	11	6,4	22
3	6	3,9	14	13	11	7	24
4	7	4Д	17	14	12	7,5	25
5	7	3,9	18	15	12	7,9	28
6	7	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8	30
8	8	5,3	19	18	13	8,6	31
9	9	5,6	20	19	14	9,5	33
10	10	6,8	21	20	14	9	36

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Лекционный курс базируется на пассивном методе обучения, реализующем традиционную объяснительно-иллюстративную образовательную технологию, в рамках которой студенты выступают в роли слушателей, воспринимающих учебный материал, и участвующих в дискуссиях и экспресс-опросах.

При прослушивании лекций рекомендуется в конспекте отмечать все важные моменты, на которых заостряет внимание преподаватель, в частности те, которые направлены на качественное выполнение соответствующей лабораторной работы.

Преподавателем запланировано использование при чтении лекций технологии учебной дискуссии. Поэтому рекомендуется фиксировать для себя интересные моменты с целью их активного обсуждения на дискуссии в конце лекции.

Залогом качественного выполнения лабораторных работ является самостоятельная подготовка путем повторения материалов лекций. Рекомендуется подготовить вопросы по неясным моментам и обсудить их с преподавателем в начале каждого занятия.

Преподавателем запланировано применение на лабораторных занятиях технологий развивающейся кооперации, коллективного взаимодействия, разбора конкретных ситуаций.

Для текущего контроля успеваемости по очной форме обучения преподавателем используется балльно-рейтинговая система контроля и оценки академической активности. Поэтому настоятельно рекомендуется тщательно прорабатывать материал дисциплины при самостоятельной работе, участвовать во всех формах обсуждения и взаимодействия, как на лекциях, так и на лабораторных работах в целях лучшего освоения материала и получения высокой оценки по результатам освоения дисциплины.

Выполнение самостоятельной работы подразумевает самостоятельное изучение разделов дисциплины, подготовку к лабораторным работам, к рубежным контролям (для очной формы обучения), контрольной работы, подготовку к экзамену.

Рекомендуемая трудоемкость самостоятельной работы представлена в таблице.

Наименование вида самостоятельной работы	Рекомендуемая трудоемкость, акад. час.	
	Очная форма обучения	Заочная форма обучения
<i>5 семестр (очная форма обучения) 7 семестр (заочная форма обучения)</i>		
Самостоятельное изучение тем дисциплины:	-	45
1. Визуальный анализ данных (Visual Mining). Характеристики средств визуализации данных. Методы визуализации: методы геометрических преобразований, отображение иконок, методы, ориентированные на пиксели, иерархические образы.		12
2 Анализ текстовой информации (Text Mining). Этапы анализа текстов, предварительная обработка текста. Извлечение ключевых понятий из текста.		11
3. Извлечение знаний из Web. Этапы Web Mining. Категории Web Mining. Методы извлечения Web-контента.		11
4. Автоматизация выполнения бизнес-процессов: понятие бизнес-процесса, формализация бизнес-процессов, Workflow-системы, проектирование бизнес-процессов. Анализ процессов		11
Подготовка к лабораторным занятиям (по 1 часу на каждое занятие (очная форма обучения) и по 2 часа (заочная форма обучения))	7	6

Подготовка к рубежным контролям (по 1 часу на каждый рубеж)	2	-
Выполнение контрольной работы	8	18
Подготовка к экзамену	27	27
Всего:	44	96

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

6.1. Перечень оценочных средств

1. Балльно-рейтинговая система контроля и оценки академической активности студентов в КГУ (для очной формы обучения).
2. Отчеты студентов по лабораторным работам.
3. Контрольная работа.
4. Банк тестовых заданий к рубежным контролям № 1, 2 (для очной формы обучения).
5. Контрольная работа (для очной и заочной формы обучения)
6. Список вопросов к экзамену.

6.2. Система балльно-рейтинговой оценки работы студентов по дисциплине

5 семестр

№	Наименование	Содержание						
		Распределение баллов						
1	Распределение баллов за семестры по видам учебной работы, сроки сдачи учебной работы (доводятся до сведения студентов на первом учебном занятии)	Посещение лекций	Посещение лабораторных занятий	Выполнение лабораторных работ	Выполнение контрольной работы	Рубежный контроль № 1	Рубежный контроль № 2	Экзамен
		До 16 баллов (32 * 0,5 балла = 16 б.)	До 8 баллов (32 * 0,25 балла = 8 б.)	До 28 баллов (7 * 4 балла = 28 б.)	До 10 баллов	До 4 баллов	До 4 баллов	До 30 баллов
2	Критерий пересчета баллов в традиционную оценку по итогам работы в семестре и зачета	60 и менее баллов – неудовлетворительно; 61...73 – удовлетворительно; 74... 90 – хорошо; 91...100 – отлично						

№	Наименование	Содержание
3	Критерии допуска к промежуточной аттестации, возможности получения автоматического зачета (экзаменационной оценки) по дисциплине, возможность получения бонусных баллов	<p>Для допуска к промежуточной аттестации (экзамену) студент должен набрать по итогам текущего и рубежного контроля не менее 50 баллов, выполнить рубежный контроль № 1 и 2, выполнить и защитить 7 лабораторных работ, а также выполнить контрольную работу.</p> <p>Для получения экзамена автоматом студенту необходимо набрать за семестр минимум 68 баллов, оценка «удовлетворительно».</p> <p>По согласованию с преподавателем студенту, набравшему 68 баллов, могут быть добавлены дополнительные (бонусные) баллы за активное участие на консультациях, оригинальность принятых решений в ходе выполнения лабораторных работ, за участие в значимых учебных и внеучебных мероприятиях кафедры и выставлена за экзамен «автоматически» оценка «хорошо» или «отлично».</p>
4	Формы и виды учебной работы для неуспевающих (восстановившихся на курсе обучения) студентов для получения недостающих баллов в конце семестра	<p>В случае если к промежуточной аттестации набрана сумма менее 50 баллов, не выполнены все задания, то студенту необходимо выполнить дополнительные задания, до конца последней (зачетной) недели семестра. При этом необходимо проработать материал всех пропущенных лабораторных занятий.</p> <p>Формы дополнительных заданий (назначаются преподавателем):</p> <ul style="list-style-type: none"> - выполнение и защита невыполненных студентом заданий лабораторных занятий – до 2 баллов; - прохождение рубежного контроля – до 6 баллов; - выполнение письменных работ по теме, предложенной преподавателем – до 10 баллов. <p>Ликвидация академических задолженностей, возникших из-за разности в учебных планах при переводе или восстановлении, проводится путем выполнения дополнительных заданий, форма и объем которых определяется преподавателем.</p>

6.3. Процедура оценивания результатов освоения дисциплины

Рубежные контроли проводятся в форме письменных ответов на вопросы. Каждому студенту предлагается по 2 вопроса, за ответ на который студент может получить до 2 баллов (рубежи №1, 2).

Перед проведением каждого рубежного контроля преподаватель прорабатывает со студентами основной материал соответствующих разделов дисциплины в форме краткой лекции-дискуссии. Примерные варианты заданий для 1, 2 рубежного контроля приведены ниже. На каждый рубежный контроль студенту отводится 2 академических часа.

Преподаватель оценивает в баллах результаты каждого студента и заносит в ведомость учета текущей успеваемости.

Билеты на экзамен состоят из 2 вопросов и практического задания. Ответы на каждый вопрос оцениваются до 10 баллов, выполнение практического задания оценивается до 10 баллов. Время, отводимое студенту на подготовку к ответу на экзаменационный билет, составляет 1 астрономический час.

Результаты текущего контроля успеваемости и экзамена заносятся преподавателем в экзаменационную ведомость, которая сдается в организационный отдел института в день экзамена, а также выставляются в зачетную книжку студента.

6.4. Примеры оценочных средств для рубежных контролей и экзамена

Примерные вопросы для рубежных контролей

Рубежный контроль 1: Список вопросов

1. Основные задачи Data Mining
2. Модели и методы Data Mining
3. Средства Data Mining.
4. Классификация и регрессия. Краткая характеристика. Общее и различие.
5. Классификация и регрессия. Правила классификации
6. Классификация и регрессия. Деревья решений
7. Классификация и регрессия. Методы построения правил классификации: алгоритм построения 1-правил.
8. Классификация и регрессия. Методы построения правил классификации: метод Naive Bayes.
9. Классификация и регрессия. Методы построения математических функций: метод наименьших квадратов.
10. Методы построения математических функций: нелинейные методы.
11. Методы построения математических функций: метод опорных векторов (SVM).
12. Поиск ассоциативных правил, общие положения.
13. Поиск ассоциативных правил. Секвенциальный анализ.
14. Поиск ассоциативных правил Алгоритм Apriori.
15. Постановка задачи кластеризации.
16. Кластеризация. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации.
17. Кластеризация. Базовые алгоритмы кластеризации.
18. Кластеризация. Алгоритм k-means.
19. Кластеризация. Адаптивные методы кластеризации.

Рубежный контроль 2: Список вопросов

1. Визуальный анализ данных. Характеристики средств визуализации данных.
2. Визуальный анализ данных. Методы визуализации: методы геометрических преобразований
3. Визуальный анализ данных. Методы визуализации: отображение иконок.
4. Визуальный анализ данных. Методы визуализации: методы, ориентированные на пиксели, иерархические образы.
5. Задача анализа текстов: этапы анализа текстов, предварительная обработка текста, задачи Text Mining.

6. Извлечение ключевых понятий из текста: общее описание процесса извлечения понятий из текста, стадия локального анализа, стадия интеграции и вывода понятий.
7. Классификация текстовых документов.
8. Методы классификации текстовых документов: представление текстовых документов.
9. Методы классификации текстовых документов: иерархические методы кластеризации текстов.
10. Методы классификации текстовых документов: бинарные методы кластеризации текстов.
11. Задача аннотирования текстов.
12. Средства анализа текстовой информации.
13. Стандарты Data Mining. Стандарт CWM. Структура и состав CWM.
14. Стандарты Data Mining. Стандарт CRISP. Структура стандарта CRISP.
15. Использование Data Mining в реальном времени.
16. Рекомендательные машины.
17. Проблемы анализа информации из Web. Этапы Web Mining.
18. Категории Web Mining.
19. Методы извлечения Web-контента.
20. Извлечение Web-структур.
21. Поиск Web-документов с учетом гиперссылок.
22. Кластеризация Web-структур.
23. Автоматизация выполнения бизнес-процессов: понятие бизнес-процесса, формализация бизнес-процессов
24. Workflow-системы, проектирование бизнес-процессов.
25. Анализ процессов: технология Process Mining, анализ протоколов.
26. Методы Process Mining.

Примерный список вопросов к экзамену.

1. Добыча данных (Data Mining). Задачи и практическое применение Data Mining.
2. Модели и методы Data Mining. Процесс обнаружения знаний.
3. Постановка задачи классификации и регрессии. Представление результатов (правила классификации, деревья решений, математические функции).
4. Методы построения правил классификации (алгоритм построения 1-правил, метод Naïve Bayes).
5. Методы построения деревьев решений (методика "разделяй и властвуй", алгоритм покрытия).
6. Методы построения математических функций (метод наименьших квадратов, нелинейные методы, метод опорных векторов (SVM)).
7. Прогнозирование временных рядов.
8. Управление знаниями. Средства Data Mining.
9. Постановка задачи поиска ассоциативных правил. Секвенциальный анализ.
10. Алгоритм Apriori, его разновидности.

11. Постановка задачи кластеризации.
12. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации. Представление результатов.
13. Базовые алгоритмы кластеризации.
14. Классификация алгоритмов. Алгоритм k-means.
15. Адаптивные методы кластеризации.
16. Выполнение визуального анализа данных. Характеристики средств визуализации данных.
17. Методы визуализации (методы геометрических преобразований, отображение иконок, методы, ориентированные на пиксели, иерархические образы)
18. Задача анализа текстов: этапы анализа текстов, предварительная обработка текста, задачи Text Mining.
19. Извлечение ключевых понятий из текста: общее описание процесса извлечения понятий из текста, стадия локального анализа, стадия интеграции и вывода понятий.
20. Классификация текстовых документов. Методы классификации текстовых документов: представление текстовых документов, иерархические методы кластеризации текстов, бинарные методы кластеризации текстов.
21. Задача аннотирования текстов. Средства анализа текстовой информации
22. Области стандартизации Data Mining. Стандарт CWM. Структура и состав CWM.
23. Области стандартизации Data Mining. Стандарт CRISP. Структура стандарта CRISP.
24. Области стандартизации Data Mining. Стандарт PMML.
25. Области стандартизации Data Mining. Другие стандарты Data Mining: стандарт SQL/MM, стандарт Microsoft Data Mining extensions (DMX), стандарт Java Data Mining.
26. Использование Data Mining в реальном времени. Адаптивная добыча данных.
27. Применение Data Mining в реальном времени. Рекомендательные машины: классификация рекомендательных машин, подход на основе содержания, совместное фильтрование.
28. Применение Data Mining в реальном времени. Анализ рыночной корзины и секвенциальный анализ.
29. Применение Data Mining в реальном времени. Усиление обучения и агенты. Инструменты Data Mining в реальном времени.
30. Проблемы анализа информации из Web. Этапы Web Mining. Web Mining и другие интернет-технологии.
31. Категории Web Mining. Методы извлечения Web-контента. Извлечение Web-контента для формирования баз данных.
32. Извлечение Web-структур. Поиск Web-документов с учетом гиперссылок.
33. Кластеризация Web-структур. Исследование использования Web-ресурсов.
34. Автоматизация выполнения бизнес-процессов: понятие бизнес-процесса, формализация бизнес-процессов.
35. Workflow-системы, проектирование бизнес-процессов.
36. Анализ процессов: технология Process Mining, анализ протоколов, стандарт записи протоколов MXML.

37. Задачи и методы Process Mining.

Примерный список заданий к экзамену.

Реализовать метод k-средних и выполнить кластеризацию образов.

Вариант 1

Образы: четырехугольники.

Признаки: четыре длины сторон четырехугольника.

Класс «маленький четырехугольник»: (2,2,2,2), (2,2,3,2), (2,3,2,2), (3,2,2,2), (3,3,2,2), (2,3,3,2), (3,2,3,2), (3,3,3,2), (2,2,2,3), (2,2,3,3), (2,3,2,3), (3,2,2,3), (3,3,2,3), (2,3,3,3), (3,2,3,3), (3,3,3,3).

Класс «большой четырехугольник»: (9,9,9,9), (9,9,11,9), (9,11,9,9), (11,9,9,9), (11,11,9,9), (11,9,11,9), (9,11,11,9), (11,11,11,9), (9,9,9,11), (9,9,11,11), (9,11,9,11), (11,9,9,11), (11,11,9,11), (11,9,11,11), (9,11,11,11).

Класс «очень маленький четырехугольник»: (0.1,0.1,0.1,0.1).

Вариант 2

Образы: книги.

Признаки: количество страниц, количество картинок, количество таблиц, средняя длина слова.

Класс «детская книга»: (200,50,0,4), (150,37,0,5).

Класс «учебник»: (400,15,90,8), (500,3,105,7).

Класс «журнал»: (80,200,5,6), (100,198,4,7).

Вариант 3

Образы: погода на сутки (летом).

Признаки: количество часов с дождем, количество солнечных часов, средняя температура воздуха, количество детей во дворе днем, скорость ветра.

Класс «хорошая погода»: (0,17,25,15,2), (1,16,24,13,3), (1,17,23,15,2).

Класс «плохая погода»: (20,0,18,1,7), (17,2,18,0,8), (18,1,19,2,7).

Вариант 4

Образы: погода на сутки (зимой).

Признаки: количество часов со снегом, количество солнечных часов, средняя температура воздуха, количество детей во дворе днем, скорость ветра.

Класс «хорошая погода»: (0,13,-10,15,2), (1,12,-12,13,3), (1,13,-13,15,2).

Класс «плохая погода»: (20,0,-21,1,8), (17,2,-24,0,9), (18,1,-23,2,8).

Вариант 5

Образы: животные.

Признаки: количество полосок, рост (см), длина хвоста (см).

Класс «пчела»: (7,1,0), (8,2,0), (6,1,0).

Класс «зебра»: (30,150,50), (35,160,60), (33,155,50).

Класс «тигр»: (60,100,100), (70, 110,110), (55,100,110).

Класс «муравей»: (0,0.5,0), (0,0.6,0).

Вариант 6

Образы: одежда.

Признаки: длина рукава, количество пуговиц, длина изделия.

Класс «футболка»: (15,0,60), (20,0,70).

Класс «рубашка»: (50,8,60), (60,9,60), (55,8,70).

Класс «летняя рубашка»: (15,6,60), (20,8,70), (20,7,65).

Класс «халат»: (15,9,120), (20,10,130), (25,9,115).

Вариант 7

Образы: школьники.

Признаки: оценка по алгебре, оценка по геометрии, оценка по физике, оценка по истории, оценка по литературе, оценка по иностранному языку, оценка по физкультуре.

Класс «отличник»: (5,5,5,5,5,5,5), (4,5,5,5,5,5,5), (5,4,5,5,5,5,5), (5,5,5,4,5,5,5), (5,5,5,5,4,5,5), (5,5,5,5,5,4,5), (5,5,5,5,5,5,4).

Класс «троечник»: (3,3,3,3,3,3,3), (4,3,3,3,3,3,3), (3,4,3,3,3,3,3), (3,3,3,4,3,3,3), (3,3,3,3,4,3,3), (3,3,3,3,3,4,3), (3,3,3,3,3,3,4).

Класс «математик»: (5,5,5,3,3,3,4), (5,5,5,3,4,3,3), (5,5,5,4,3,3,4), (5,5,5,3,4,4,3), (5,5,5,4,4,3,4), (5,5,5,4,3,4,3).

Класс «гуманитарий»: (3,3,3,5,5,5,4), (3,4,3,5,5,5,3), (4,3,3,5,5,5,4), (3,4,4,5,5,5,3), (4,4,3,5,5,5,4), (4,3,4,5,5,5,3).

Вариант 8

Образы: ягоды.

Признаки: количество семечек, радиус сечения (см), количество на одном растении.

Класс «вишня»: (1,2,300), (1,1.5,400), (1,1,350).

Класс «арбуз»: (50,20,1), (60,25,1), (40,21,1).

Класс «облепиха»: (6,0.3,500), (8,0.2,550), (8,0.3,600).

Класс «виноград»: (6,1,1500), (8,1.5,1400), (8,0.7,1700).

Вариант 9

Образы: пятиугольники.

Признаки: пять длин сторон пятиугольника.

Класс «маленький пятиугольник»: (2,2,2,2,2), (2,2,3,2,2), (2,3,2,2,3), (3,3,2,2,2), (2,3,3,2,3), (3,2,3,2,2), (3,3,3,2,3), (2,2,2,3,2), (3,3,3,3,3).

Класс «большой пятиугольник»: (9,9,9,9,9), (9,9,11,9,11), (9,11,9,9,9), (11,11,9,9,9), (11,9,11,9,11), (9,11,11,9,9), (11,11,11,9,9).

Класс «очень маленький пятиугольник»: (0.1,0.1,0.1,0.1,0.1).

Вариант 10

Образы: треугольники.

Признаки: три длины сторон треугольника.

Класс «маленький треугольник»: (2,2,2), (2,2,3), (2,3,2), (3,2,2), (3,3,2), (3,2,3), (3,3,3).

Класс «большой треугольник»: (9,9,9), (9,9,11), (9,11,9), (11,9,9), (11,11,9), (11,9,11), (9,11,11), (11,11,11).

Класс «очень маленький треугольник»: (0.1,0.1,0.1).

6.5. Фонд оценочных средств

Полный банк заданий для текущего, рубежных контролей и промежуточной аттестации по дисциплине, показатели, критерии, шкалы оценивания компетенций, методические материалы, определяющие процедуры оценивания образовательных результатов, приведены в учебно-методическом комплексе дисциплины.

7. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ УЧЕБНАЯ ЛИТЕРАТУРА

7.1. Основная учебная литература

1 Богданов, Е. П. Интеллектуальный анализ данных : практикум для магистрантов направления 09.04.03 «Прикладная информатика» профиль подготовки «Информационные системы и технологии корпоративного управления» / Е. П. Богданов. - Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. - 112 с. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1087885>. – Режим доступа: по подписке.

2 Целых, А. Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу "Методы интеллектуального анализа данных" / А. Н. Целых, А. А. Целых, Э. М. Котов ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2021. - 130 с. - ISBN 978-5-9275-3783-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1894428>. – Режим доступа: по подписке.

3 Зарова, Е. В. Методы Data mining в обработке и анализе статистических данных (решения в R) : монография / Е.В. Зарова. — Москва : ИНФРА-М, 2021. — 232 с. : ил. - ISBN 978-5-16-016814-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1240276>. – Режим доступа: по подписке.

7.2 Дополнительная литература

1. Барсегян А.А., Куприянов М.С., Холод И.И., Тесс М.Д., Елизаров С.И. Анализ данных и процессов 3-е издание. Учебная литература для вузов. — БХВ-Петербург, 2009 — 512 с. : ил.

2. Замятин А. В. Интеллектуальный анализ данных : учеб. пособие / А. В. Замятин. – Томск : Томск. гос. ун-т, 2016. – 119 с.

3. Журавлёв, Ю. И. Распознавание. Математические методы. Программная система. Практические применения / Ю. И. Журавлёв, В. В. Рязанов, О. В. Сенько. — М. : Фазис, 2006. — 176 с.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

1. Кластерный анализ. StatSoft / [Электронный ресурс]. — URL: <http://statsoft.ru/home/textbook/modules/stcluan.html#k>.
2. Дюк В., Самойленко А. Data Mining: учебный курс (+CD). – СПб.: Изд. Питер, 2001. – 368 с.
3. Чубукова, И. А. Data Mining / Чубукова И. А. - Москва: Национальный Открытый Университет "ИНТУИТ", 2016. (Основы информационных технологий) - ISBN 978-5-94774-819-2. - Текст: электронный // ЭБС "Консультант студента" : [сайт]. - URL: <https://www.studentlibrary.ru/book/ISBN9785947748192.html>. - Режим доступа: по подписке.

9. РЕСУРСЫ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫЕ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1. Сайт дистанционного обучения в НОУ (Национальный Открытый Университет) «ИНТУИТ» содержит бесплатные курсы, программы повышения квалификации и профессиональной переподготовки, интересные доклады и другую полезную информацию <http://www.intuit.ru>.
2. Федеральный портал «Российское образование» <http://www.edu.ru/>
3. Информационный сайт, содержащий справочные материалы по информатике, которые включают в себя курс лекций, схемы, презентации, рефераты и др. informatikaplus.narod.ru.
4. Материалы сайта об интеллектуальном анализе данных: <http://www.rdatamining.com/>
5. База данных временных рядов <http://irafim.osu.cz/cif/main.php>
6. Сайт кафедры ПОАС КГУ «Информатика и программирование: шаг за шагом» <http://it.kgsu.ru/>.

10. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ

При чтении лекций используются слайдовые презентации.
Минимальные требования к программному обеспечению компьютера, используемого при показе слайдовых презентаций: Wondershare PDF Reader (свободно распространяемое программное обеспечение).

11. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально-техническое обеспечение дисциплины включает в себя учебные лаборатории и классы, оснащенные современными компьютерами (все

– в стандартной комплектации для практических занятий и самостоятельной работы), объединенными локальными вычислительными сетями с выходом в Интернет, мультимедийное оборудование (переносной персональный компьютер, мультимедийный проектор, мультимедийный экран).

Программные средства обеспечения учебного процесса должны включать: Python 3.8 или выше (свободно распространяемое программное обеспечение), Visual Studio Code 1.5 или выше (это свободно распространяемое программное обеспечение, которое может использоваться для разработки приложений на языке программирования C++); вспомогательные LibreOffice (программы презентационной графики; текстовые редакторы; графические редакторы – все программы свободно распространяемые).

12. ДЛЯ СТУДЕНТОВ, ОБУЧАЮЩИХСЯ С ИСПОЛЬЗОВАНИЕМ ДИСТАНЦИОННЫХ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

При использовании электронного обучения и дистанционных образовательных технологий (далее ЭО и ДОТ) занятия полностью или частично проводятся в режиме онлайн. Объем дисциплины и распределение нагрузки по видам работ соответствует п. 4.1. Распределение баллов соответствует п. 6.2, либо может быть изменено в соответствии с решением кафедры, в случае перехода на ЭО и ДОТ в процессе обучения. Решение кафедры об используемых технологиях и системе оценивания достижений обучающихся принимается с учетом мнения ведущего преподавателя и доводится до сведения обучающихся.

Аннотация к рабочей программе дисциплины
«МЕТОДЫ И АЛГОРИТМЫ АНАЛИЗА ДАННЫХ»

образовательной программы высшего образования –
программы бакалавриата

09.03.04 Программная инженерия

Направленность:

Программное обеспечение автоматизированных систем

Трудоемкость дисциплины: 3 з.е. (108 академических часов)

Семестр: 5 (очная), 7 (заочная)

Форма промежуточной аттестации: экзамен

Содержание дисциплины

Понятие Data Mining. Задачи и практическое применение Data Mining. Модели и методы Data Mining.

Классификация и регрессия. Поиск ассоциативных правил. Кластеризация. Базовые алгоритмы кластеризации.

Визуальный анализ данных (Visual Mining). Методы визуализации: методы геометрических преобразований, отображение иконок, методы, ориентированные на пикселы, иерархические образы.

Анализ текстовой информации (Text Mining). Этапы анализа текстов. Извлечение ключевых понятий из текста. Классификация текстовых документов.

Стандарты Data Mining.

Data Mining в реальном времени.

Извлечение знаний из Web. Этапы Web Mining. Категории Web Mining. Методы извлечения Web-контента. Извлечение Web-структур. Поиск Web-документов с учетом гиперссылок.