



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ВВЕДЕНИЕ В DATA MINING
Кафедра прикладной математики

Образовательная программа бакалавриата
01.03.05 - Статистика

Направленность (профиль) программы
Анализ больших данных

Форма обучения
Очная

Статус дисциплины: входит в часть,
формируемую участниками образовательных отношений,
модуль профильной направленности

Махачкала, 2023

Рабочая программа дисциплины «Введение в Data Mining» составлена в 2023 году в соответствии с требованиями ФГОС ВО бакалавриата по направлению подготовки 01.03.05 - Статистика от 14.08.2020 г. № 1032

Разработчики:

1. кафедра прикладной математики, Магомедова Е.С. к. ф.-м. н., доцент;

Рабочая программа дисциплины одобрена:

на заседании кафедры прикладной математики от «20» 01 2023 г., протокол № 5

Зав. кафедрой К Кадиев Р.И.

на заседании Методической комиссии факультета математики и компьютерных наук от «25» января 2023 г., протокол № 4.

Председатель М Ризаев М.К.

Рабочая программа дисциплины согласована с учебно-методическим управлением «20» февраля 2023 г. А

(подпись)

Аннотация рабочей программы дисциплины

Дисциплина «Введение в Data Mining» входит в часть, формируемую участниками образовательных отношений образовательной программы *бакалавриата* по направлению – **01.03.05 – Статистика**.

Дисциплина реализуется на факультете математики и компьютерных наук *кафедрой прикладной математики*.

Содержание дисциплины охватывает круг вопросов, связанных с ознакомлением с базовыми математическими моделями и освоением вычислительными методами решения задач математического анализа, линейной алгебры и статистики, пакеты прикладных программ, информатика, а также знакомством с современными направлениями развития вычислительной математики и информатики.

Дисциплина нацелена на формирование следующих компетенций выпускника: профессиональных – ОПК-3, ПК-1, ПК-5.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: *лекции, практические занятия, лабораторные занятия, самостоятельная работа*.

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости в форме контрольных работ, коллоквиума, защиты лабораторных работ. И промежуточный контроль в форме зачета.

Объем дисциплины 5 зачетных единиц, в том числе в академических часах по видам учебных занятий 180 часов.

Семес тр	Учебные занятия						Форма промежуточной аттестации (зачет, дифференцирован ный зачет, экзамен	
	в том числе							
	Контактная работа обучающихся с преподавателем					СРС, в том числе экза мен		
	Всег о	из них						
Лекц ии		Лабораторн ые занятия	Практиче ские занятия	КСР	консульта ции			
5	180	30	16	30			104	зачет
Итого:	180	30	16	30			104	

1. Цели освоения дисциплины

Цель изучения курса «Введение в Data Mining» - формирование у обучающихся системного представления о технологиях *интеллектуального анализа данных* (Data Mining), их применении и инструментах, изучение основных методов прикладного анализа данных, развитие навыков практического применения методов Data Mining для решения различных научных и технических задач. Целью является овладение студентами умений применять аппарат математики и информатики при решении задач математической статистики, разработки алгоритмов и программ численного решения различных задач, встречающиеся в статистике и задачах прикладного характера.

Задачи дисциплины :

Сформировать у обучающихся следующие способности:

- проектировать информационные процессы и системы с использованием инновационных инструментальных средств
- принимать эффективные проектные решения в условиях неопределенности и риска.

Процесс изучения дисциплины «Data mining» направлен на формирование:

- Способности проектировать информационные процессы и системы с использованием инновационных инструментальных средств

- Способности принимать эффективные проектные решения в условиях неопределенности и риска

В результате изучения дисциплины обучающийся должен знать:

- задачи, модели и методы Data Mining, области применения Data Mining
- этапы и стадии процесса обнаружения знаний
- понятие классификации, регрессии, прогнозирования
- методы построения правил классификации, деревьев решений, математических функций
- понятие кластеризации, базовые и адаптивные методы кластеризации, формальные критерии качества кластеризации
- понятия визуализации, визуального анализа данных
- характеристики средств визуализации данных
- концепцию хранилища данных, принципы организации хранилища данных
- многомерную модель данных, определение OLAP-систем, концептуальное многомерное представление, архитектуру OLAP-систем

А также уметь:

- отбирать эффективные средства интеллектуального анализа данных
- выполнять постановку задачи классификации и прогнозирования
- проводить оценку и выбирать оптимальный алгоритм классификации для решения конкретной задачи
- проводить оценку и выбирать оптимальный алгоритм прогнозирования для решения конкретной задачи
- выполнять постановку задачи поиска ассоциативных правил
- проводить оценку и выбирать оптимальный алгоритм поиска ассоциативных правил для решения конкретной задачи
- выполнять постановку задачи кластеризации

2. Место дисциплины в структуре ОПОП бакалавриата

Дисциплина «Введение в Data Mining» входит в часть, формируемую участниками образовательных отношений образовательной программы *бакалавриата* по направлению подготовки **01.03.05 – Статистика**.

Курс «Введение в Data Mining» вводится после изучения дисциплин алгебра, информатика, математический анализ, математическая статистика, так как для успешного усвоения этого курса студентам необходимы знания по указанным дисциплинам.

Изученные в курсе методы могут применяться при решении различных математических моделей в естествознании. А также могут найти применение при выполнении индивидуальных заданий, курсовом и дипломном проектировании.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения)

Код и наименование компетенции из ОПОП (в соответствии с ОПОП)	Код и наименование индикатора достижения компетенций (в соответствии с ОПОП)	Планируемые результаты обучения	Процедура освоения
ОПК-3. Способен осознанно применять методы математической и дескриптивной статистики для анализа количественных данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ, содержательно интерпретировать полученные результаты, готовить статистические материалы для докладов, публикаций и других аналитических материалов.	ОПК 3.1: общую методику статистического исследования и способы количественной формализации объекта наблюдений	<i>Знает:</i> общую методику статистического исследования и способы количественной формализации объекта наблюдений. <i>Умеет:</i> применить общую методику статистического исследования и способы количественной формализации объекта наблюдений при решении профессиональных задач. <i>Владеет:</i> навыками применения общей методики статистического исследования и способы количественной формализации объекта наблюдений при решении прикладных задач.	Контрольные работы, лабораторные работы, экзамен

	ОПК-3.2. Уметь применять математический эконометрический инструментальный для анализа количественных данных, в том числе с применением информационных систем и технологий	<i>Знает:</i> как применить математический и эконометрический инструментальный для анализа количественных данных, в том числе с применением информационных систем и технологий. <i>Умеет:</i> применять математический и эконометрический инструментальный для анализа количественных данных, в том числе с применением информационных систем и технологий. <i>Владеет:</i> математическим и эконометрическим инструментальными для анализа количественных данных, в том числе с применением информационных систем и технологий вычислительной техникой.	
--	---	---	--

	ОПК-3.3. Владеет навыками выбора применения инструментальных средств для обработки количественных данных, навыками интерпретации результатов формулирования выводов рекомендаций для подготовки аналитических материалов.	<i>Знает:</i> как применить математические и статистические инструментария и современную вычислительную технику для решения прикладных задач. <i>Умеет:</i> применить математические и статистические инструментария и современную вычислительную технику для решения прикладных задач. <i>Владеет:</i> навыками применения математического и статистического инструментария для решения прикладных задач, методами работы с современной вычислительной техникой.	
ПК-1. Способен собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям	ПК-1.1. <i>Знает</i> методы сбора и обработки данных, полученными в области математических и естественных наук, программирования и информационных технологий для формирования выводов по соответствующим научным исследованиям	<i>Знает:</i> стандартные методы и технические средства для статистических наблюдений. <i>Умеет:</i> применить стандартные методы и технические средства при статистических наблюдениях. <i>Владеет:</i> методами и техническими средствами для статистических наблюдений.	Контрольные работы, лабораторные работы, экзамен

	ПК-1.2. <i>Умеет:</i> собирать данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации. <i>Умеет:</i> собирать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации. <i>Владеет:</i> методами сбора данных об объекте исследования и выбора соответствующий инструментарий для обработки информации.	
	ПК-1.3. <i>Владеет:</i> навыками сбора и обработки информации, в том числе с применением полученных в информационно-математических и (или) естественных наук, программирования и информационных технологий для формирования выводов по соответствующим научным исследованиям.	<i>Знает:</i> статистические методы обработки информации, в том числе с применением информационно-коммуникационных технологий. <i>Умеет:</i> применять статистические методы для обработки информации, в том числе с применением информационно-коммуникационных технологий. <i>Владеет:</i> статистическими методами обработки информации, в том числе применением информационно-коммуникационных технологий

ПК-5. Способен разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	ПК-5.1. Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Контрольные работы, лабораторные работы, экзамен
	ПК-5.2. Умеет разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	

	ПК-5.3. Владеет навыками разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи	Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	
--	---	--	--

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 5 зачетных единиц, 180 академических часов.

4.2. 1. Структура дисциплины.

№	Название разделов и тем дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)						Формы текущего контроля успеваемости (по неделям семестра)	Форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лаб. Раб.	Сам. раб	Подг. к экз.	Общ. тр		
	Модуль 1. Теоретические и методологические основы интеллектуального анализа данных			6	6	4	20		36		

1	Базовые понятия и основные задачи интеллектуального анализа данных (ИАД). Этапы ИАД. Основные методы и области применения ИАД. Эволюция методов обработки и анализа данных.	5	1	2	2		6		10	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
2	Алгоритмы предварительной данных. Шумы и выбросы в данных. Пропуски данных и методы их восстановления	5	2	2	2	2	6		12	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
3	Основные методы «DataMinin». Регрессия. Задача классификации. Прогнозирование и временные ряды.	5	3	2	2	2	8		14	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
Модуль 2. Инструментальные средства интеллектуального анализа данных				8	8	2	18		36	
4	Нейросетевые методы анализа данных. Экспертные системы. Этапы проектирования: идентификация, концептуализация, формализация, реализация, тестирование, опытная эксплуатация.	5	4	2	2	2	2		8	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа

5	Рынок инструментов Data Mining. Классификация инструментов Data Mining. Табличные процессоры как инструмент анализа.	5	5	4	4		8		16	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
6	Средства извлечения и очистки данных. Средства извлечения данных: методы и возможности. Очистка данных. Инструменты очистки данных. Построение и использование модели. Стандарты Ра Мише.	5	6	4	4	2	10		20	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
Модуль 3. Современные подходы к осуществлению ИАД средствами информационных технологий				6	6	4	20		36	
7	Аналитические платформы Data Mining: основные возможности, сфера применения, этапы решения задач. Импорт/экспорт данных. Визуализация результатов.	5	9	6	6	4	20		36	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
Модуль 4. Практика применения интеллектуальных технологий				8	8	4	16		36	
8	Применение ИАД «Data Mining» для бизнес-задач. Отбор данных для анализа. Предварительная подготовка данных. Понижение размерности, сглаживание	5	13	2	4		10		16	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа

	аномалий, фильтрация, группировка.									
9	Применение ИАД «Data Mining» в маркетинговой деятельности и в сфере логистики.	5	14	4	2	2	4		12	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
10	Анализ разрыва. Портфолио-анализ. Анализ маржинальной прибыли.									
11	Анализ потоков платежей. Финансовая "паутина".	5	15	2	2	2	2		8	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа
	Модуль 5. Финансовый и инвестиционный «Data Mining».							36	36	
	Анализ потоков платежей. Финансовая "паутина".	5	15	2	2	2	2		8	Индивидуальный фронтальный опрос, тестирование. Контрольная, лабораторная работа

4.3. Содержание дисциплины, структурированное по темам (разделам).

4.3.1. Содержание лекционных занятий по дисциплине

Модуль 1. Теоретические и методологические основы «Введение в Data Mining» (интеллектуального анализа данных ИАД 30 ч.)

Тема 1.

Лекция №1. Базовые понятия и основные задачи «Введение в Data Mining» интеллектуального анализа данных (ИАД). Этапы ИАД. Основные методы и области применения ИАД. Эволюция методов обработки и анализа данных.

Тема 2.

Лекция N2. Алгоритмы предварительной подготовки данных. Понятие "данные".

Набор данных и их атрибутов. Типы наборов данных. Измерения и шкалы. Методы и способы получения первичной информации.

Шумы и выбросы в данных. Пропуски данных и методы их восстановления. Задачи отбора признаков. Несбалансированные выборки. Базы и витрины данных.

Тема 3.

Лекция N 3. Основные методы ИАД «Введение в Data Mining».

Регрессия. Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации. Точность классификации: оценка уровня ошибок. Оценивание классификационных методов. Деревья решений. Процесс конструирования дерева решений. Метод опорных векторов. Метод «ближайшего соседа». Байесова классификация. Задача прогнозирования. Сравнение задач прогнозирования и классификации. Прогнозирование и временные ряды. Решение задачи прогнозирования. Задача кластеризации. Применение кластерного анализа. Иерархические методы. Итеративные методы. Методы поиска ассоциативных правил. Методы визуализации. Качество визуализации. Представление пространственных характеристик. Основные тенденции в визуализации.

Тема 4.

Лекция N4. Нейросетевые методы анализа данных. Принципы функционирования искусственных нейронных сетей. Основные области их применения. Многослойные нейросети. Экспертные системы. Составные части экспертной системы.

Статические и динамические экспертные системы. Приобретение знаний. Извлечение знаний из данных. Механизмы приобретения и объяснения знаний. Этапы проектирования экспертной системы: идентификация, концептуализация, формализация, реализация, тестирование, опытная эксплуатация.

Участники процесса проектирования: эксперты, инженеры по знаниям, конечные пользователи.

Модуль 2. Инструментальные средства интеллектуального анализа данных**Тема 5.**

Лекция N 5. Рынок инструментов Data Mining. Классификация инструментов Data Mining. Табличные процессоры как инструмент анализа. Эволюция инструментов: от Excel до SPSS, Statistica, E-views.

Лекция N 6. Этапы выполнения многомерного статистического анализа и интерпретация результатов Программное обеспечение для решения задач классификации. Программное обеспечения для решения задач кластеризации и сегментации.

<i>Лекция N 7.</i> Программное обеспечение Data Mining для поиска ассоциативных правил, Программное обеспечение для решения задач оценивания и прогнозирования. Статистический анализ данных.		
Тема 6. <i>Лекция N 8.</i> Средства извлечения и очистки данных. Средства извлечения данных: методы и возможности. Очистка данных. Инструменты очистки данных. Построение и использование модели. Стандарты Ра Мише.		
Модуль 3. Современные подходы к осуществлению ИАД средствами информационных технологий		
Тема 7. Лекция N 9. Аналитические платформы Data Mining: основные возможности, сфера применения, этапы решения задач. <i>Лекция N 10.</i> Импорт/экспорт данных. Визуализация результатов. <i>Лекция N 11.</i> Возможности автоматизации отдельных этапов анализа данных.		
Модуль 4. Практика применения интеллектуальных технологий		
Тема 8. <i>Лекция N12.</i> Применение ИАД для бизнес-задач. Отбор данных для анализа. Предварительная подготовка данных. Понижение размерности, сглаживание аномалий, фильтрация, группировка. Анализ обеспечения ресурсами. Анализ возможных прерываний бизнес-процессов. Определение верхних границ цен. Анализ стратегической позиции предприятия		
Тема 9. <i>Лекция N 13.</i> Применение ИАД в маркетинговой деятельности и в сфере логистики. Анализ разрыва. Портфолио-анализ. Анализ маржинальной прибыли. <i>Лекция N14.</i> сравнительные расчеты. Анализ цепочек и издержек логистических процессов.		
Тема 10. <i>Лекция N 15.</i> Финансовый и инвестиционный ИАД. Анализ потоков платежей. Финансовая "паутина". Функционально-стоимостной анализ. Анализ ставки дисконтирования с поправкой на риск.		

4.3.2. Содержание практических занятий по дисциплине

№ п/п	Тема	Аудиторные часы (30ч)
	Модуль 1. Теоретические и методологические основы «Введение в Data Mining» интеллектуального анализа данных	8

1	Алгоритмы предварительной подготовки данных. Понятие "данные". Набор данных и их атрибутов. Типы наборов данных. Измерения и шкалы. Методы и способы получения первичной информации.	2
2	Регрессия. Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации. Точность классификации: оценка уровня ошибок.	2
3	Нейросетевые методы анализа данных. Принципы функционирования искусственных нейронных сетей.	2
4	Приобретение знаний. Извлечение знаний из данных. Механизмы приобретения и объяснения знаний	2
	Модуль 2. Инструментальные средства интеллектуального анализа данных	8
5	Этапы выполнения многомерного статистического анализа и интерпретация результатов Программное обеспечение для решения задач классификации	4
6	Программное обеспечение Data Mining для поиска ассоциативных правил, Программное обеспечение для решения задач оценивания и прогнозирования. Статистический анализ	2
7	Очистка данных. Инструменты очистки данных. Построение и использование модели	2
8	Стандарты Ра Мише.	
	Модуль 3. Современные подходы к осуществлению ИАД средствами информационных технологий	6
9	Аналитические платформы Data Mining: основные возможности, сфера применения, этапы решения задач.. Возможности автоматизации отдельных этапов анализа данных.	4
10	Импорт/экспорт данных. Визуализация результатов. Возможности автоматизации отдельных этапов анализа данных.	
11	Интерпретация процедуры интеллектуального анализа временных рядов и прогнозирования с помощью различных инструментальных средств	2
	Модуль 4. Практика применения интеллектуальных технологий	12
12	Применение ИАД для бизнес-задач. Отбор данных для анализа. Предварительная подготовка данных. Понижение размерности,	2

	сглаживание аномалий, фильтрация, группировка.	
13	Анализ обеспечения ресурсами. Анализ возможных прерываний бизнес-процессов. Определение верхних границ цен. Анализ стратегической позиции предприятия	2
14	Применение ИАД в маркетинговой деятельности и в сфере логистики. Анализ разрыва. Портфолио-анализ.	4
15	Финансовый и инвестиционный ИАД. Анализ потоков платежей. Финансовая "паутина". Функционально-стоимостной анализ. Анализ ставки дисконтирования с поправкой на риск.	4

4.3.3. Содержание лабораторных занятий по дисциплине

№ п/п	Тема (16ч.)
	Модуль 1. Теоретические и методологические основы интеллектуального анализа данных «Введение в Data Mining»
1.	Обзор основных открытых баз и витрин данных в целях осуществления ИАД.
2.	Извлечение и анализ данных из открытых источников. Методы графического представления и визуализации данных.
3.	Осуществление корреляционного и регрессионного анализа с помощью различных инструментальных средств (MS Office, E-views 6.0 ит.д.)
	Модуль 2.
4	Классификация и кластеризация данных Методы графического представления и визуализации данных
	Модуль 3
5	Этапы проектирования экспертной системы
6	Реализация и интерпретация процедуры интеллектуального анализа временных рядов и прогнозирования с помощью различных инструментальных средств (MS Office, E-views 6.0 и т.д.)
	Модуль 4.
7	. Реализация и интерпретация результатов многомерного статистического анализа

8	Реализация и интерпретация результатов интеллектуального анализа данных для решения основных бизнес задач в маркетинговой деятельности и в сфере логистики
---	--

5. Образовательные технологии

С целью формирования и развития профессиональных навыков обучающихся в соответствии с требованиями ФГОС ВО по направлению подготовки предусматривается широкое использование в учебном процессе активных и интерактивных форм проведения занятий.

В основе преподавания дисциплины «Введение в Data Mining» лежит лекционно-семинарская система обучения в сочетании с лабораторными занятиями. Это связано с необходимостью активного продумывания теоретического материала и дальнейшего приложения его к прикладным задачам. Индивидуальные особенности обучающихся учитываются подбором заданий разного уровня сложности для самостоятельной работы студентов.

При проведении практических занятий по данной дисциплине используются такие методы обучения, как тестирование, фронтальный опрос, индивидуальный опрос, метод малых групп и т.п. Лекции при этом проводятся с использованием средств визуализации лекционного материала (мультимедийных презентаций) и применением таких методов и технологий, как дискуссия, проблемная лекция и т.п. При проведении семинаров и практических занятий в интерактивной форме используются следующие методы: дебаты, метод проектов, мини-конференция, решение ситуационных задач и т.п.

Кроме того, в процессе изучения дисциплины с целью повышения качества обучения предполагается использование научно-исследовательской работы студентов.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

6.1. Виды и порядок выполнения самостоятельной работы.

Самостоятельная работа по данной дисциплине включает такие формы работы как: работа с базовым конспектом; работа с литературой, чтение дополнительной литературы; лабораторная работа, подготовка отчета; выполнение контрольной работы; подготовка к экзамену.

Самостоятельная работа студентов (СРС) включает контролируемую и внеаудиторную самостоятельную работу, направлена на повышение качества обучения, углубление и закрепление знаний студента, развитие аналитических навыков по проблематике учебной дисциплины, активизацию учебно-познавательной деятельности студентов и снижение аудиторной нагрузки. Часть программного материала выносится для самостоятельного внеаудиторного изучения с последующим текущим или итоговым контролем знаний на занятиях или экзамене. Контроль СРС и оценка ее результатов организуется как самоконтроль (самооценка) студента, а также как контроль и оценка со стороны преподавателя, например в ходе собеседования. Баллы, полученные по СРС студентом, обязательно учитываются при итоговой аттестации по курсу.

Формы контроля СРС включают: *тестирование; устную беседу по теме с преподавателем; выполнение индивидуального задания и др.* Роль студента в СРС -

самостоятельно организовывать свою учебную работу по предложенному преподавателем, методически обеспеченному плану. СРС по курсу учитывает индивидуальные особенности слушателей и включает не только задания, связанные с решением типовых задач, но также творческие задания, требующие самостоятельно «добывать» знания из разных областей, группировать и концентрировать их в контексте конкретной решаемой задачи. Технология обучения предусматривает выработку навыков презентации результатов выполненного индивидуального задания и создание условий для командной работы над комплексной темой с распределением функций и ответственности между членами коллектива. Оценка результатов выполнения индивидуального задания осуществляется по критериям, известным студентам, отражающим наиболее значимые аспекты контроля за выполнением этого вида работ.

6.2.Содержание самостоятельной работы студентов по дисциплине (модулю)

Виды и порядок выполнения самостоятельной работы

1. Изучение рекомендованной литературы.
2. Подготовка к отчетам по лабораторным занятиям.
3. Решение задач.
4. Подготовка к коллоквиуму.
5. Подготовка к экзамену.

№	Виды самостоятельной работы	Вид контроля	Учебно-методич. обеспечения
1	Изучение рекомендованной литературы	Устный опрос по разделам дисциплины	[1,2,3,4] раздел 8
2	Подготовка к тестовому опросу	Устный опрос, либо компьютерное тестирование	[1,2,3,4] раздел 8
3	Решение задач	Проверка домашнего задания	[4] раздел 8
4	Подготовка к коллоквиуму	Промежуточная аттестация в форме контрольной работы	[1,2,3,4] раздел 8
5	Подготовка к экзамену	Устный опрос, либо компьютерное тестирование	[1,2,3,4] раздел 8

Содержание самостоятельной работы студентов по дисциплине «Введение в Data Mining»

№	Наименование тем и вопросы, выносимые на самостоятельную работу	Форма СР
---	---	----------

1	Тема 1 Введение в Data Mining. Данные. Основные вопросы: Технологии выбора признаков	лабораторная работа, подготовка отчета; работа
2	Тема 2 Методы и стадии Data Mining. Информация и знания Основные вопросы: Прогнозирование количественных переменных	лабораторная работа, отчет; работа с доп. литературой
3	Тема 3 Задачи Data Mining. Классификация и кластеризация. Основные вопросы: Метод главных компонент.	лабораторная работа, подготовка отчета;
4	Тема 4 Методы классификации и прогнозирования. Основные вопросы: Экспертные системы	лабораторная работа, подготовка отчета
5	Тема 5 Методы кластерного анализа. Иерархические методы. Основные вопросы: Методы представления знаний	лабораторная работа, подготовка отчета
6	Тема 6 Процесс Data Mining. Основные вопросы: Деревья классификаций.	лабораторная работа, подготовка отчета
7	Тема 7 Инструменты Data Mining. Основные вопросы: Визуализация данных	лабораторная работа, подготовка
8	Тема 8 Data Mining консалтинг Основные вопросы: Ассоциативные правила.	лабораторная работа, отчет; выполнение контрольной

6.3. *Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.*

Вопросы для самостоятельного изучения по конкретным разделам (модулям) приведены в п. 7.2 настоящей программы. Там же приведены темы рефератов и типовые контрольные работы по дисциплине.

Задания для проверочной работы, самостоятельной работы, домашние задания содержатся в пособиях, указанных в списке учебной литературы.

Методические разработки для выполнения работ имеются на кафедре ПМ и выдаются студентам методистом кафедры. Учебная литература (учебники, учебные пособия) и информационные ресурсы приведены в п. 8 настоящей "Программы".

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

7.1. Типовые контрольные задания

Контрольная работа 1.

задание 1

1. Основные методы и области применения ИАД. Эволюция методов обработки и анализа данных.
2. Программное обеспечение для решения задач кластеризации и сегментации.

задание 2

1. Понятие "данные". Набор данных и их атрибутов. Типы наборов данных. Измерения и шкалы.
2. Табличные процессоры как инструмент анализа. Эволюция инструментов: от Excel до SPSS, Statistica, E-views.

задание 3

1. Задачи отбора признаков. Несбалансированные выборки. Базы и витрины данных
2. Иерархичные методы. Итеративные методы. — Методы поиска ассоциативных правил.

задание 4

1. Методы и способы получения первичной информации. Шумы и выбросы в данных. Пропуски данных и методы их восстановления
2. Этапы проектирования экспертной системы: идентификация, концептуализация, формализация, реализация, тестирование, опытная эксплуатация.

Контрольная работа 2.

задание 5

1. Методы визуализации. Качество визуализации. Представление пространственных характеристик
2. Этапы выполнения многомерного статистического анализа и интерпретация результатов.

задание 6

1. Базовые понятия и основные задачи интеллектуального анализа данных(ИАД).

Этапы ИАД.

2. Иерархичные — методы. Итеративные методы. — Методы поиска ассоциативных правил.

Зачетное задание 7

1. Задача прогнозирования. Сравнение задач прогнозирования и классификации
2. Средства извлечения данных: методы и возможности

Зачетное задание 8

1. Составные части экспертной системы. Статические и динамические экспертные системы.
2. Процесс конструирования дерева решений.

Контрольная работа 3.

Зачетное задание 9

Е Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации.

2. Этапы осуществления многомерного статистического анализа

Зачетное задание 10

1. Стандарты Data Mining.
2. Этапы осуществления корреляционно-регрессионного анализа

Зачетное задание 11

1. Рынок инструментов Data Mining. Классификация инструментов Data Mining.
2. Очистка данных. Инструменты очистки данных.

Зачетное задание 12

1. Принципы функционирования искусственных нейронных сетей. Основные области их применения. Многослойные нейросети.
2. Программное обеспечение для решения задач оценивания и прогнозирования.

Контрольная работа 4.

Зачетное задание 13

1. Приобретение знаний. Извлечение знаний из данных. Механизмы приобретения и объяснения знаний.
2. Программное обеспечение для решения задач классификации.

Зачетное задание 14

1. Участники процесса проектирования: эксперты, инженеры по знаниям, конечные пользователи.
2. Этапы и основные методики статистического анализа больших данных.

Зачетное задание 15

Программное обеспечение Рая Мите для поиска ассоциативных правил.

2. Построение и использование эконометрической модели для анализа больших данных

7.2. Вопросы для устного опроса, тестирования (собеседования) по дисциплине «Введение в Data Mining»

1. Назовите основную цель интеллектуального анализа данных.
2. Приведите наиболее распространенные определения информации.

3. Приведите формализованное описание факта.
 4. Что называется фактографическими данными?
 5. В чем специфика методов интеллектуального анализа?
 6. Какие методы интеллектуального анализа Вы знаете?
 7. В чем состоит содержание методов нечеткой логики, системы рассуждений на основе аналогичных случаев? В чем состоит содержание методов нейронных сетей и генетических алгоритмов?
 8. В чем заключаются методы ассоциаций, кластеризации и классификации?
 9. Назовите процессы, явления, закономерности, при исследовании которых используются методы интеллектуального анализа.
 10. В каких предметных областях и для каких целей используются методы интеллектуального анализа?
- И. Какие средства интеллектуального анализа представлены на рынке, каковы их возможности?
12. Что рекомендуется для успешного комплексирования подсистемы интеллектуального анализа ИАС и достижения успеха при его выполнении? Перечислите основные составляющие экспертной системы?
 13. Назовите основные задачи, решаемые средствами ИАД?
 14. Назовите три уровня анализа информации. Чем они отличаются?
 15. Чем отличается классификация от кластеризации?
 16. Назовите различие между данными и знаниями.
 17. Что называется моделью?
 18. Перечислите основные виды моделирования.
 19. Что такое математическое моделирование?
 20. Назовите основные этапы моделирования.
 21. Что изучает математическая статистика?
 22. Основные задачи математической статистики?
 23. Что называется генеральной совокупностью? Выборкой?
 24. Какие бывают виды выборок?
 25. Какая выборка является репрезентативной?
 26. Что называется наблюдением? Статистическим рядом? Вариационным рядом? Рангом?
 27. Дайте определение выборочной (эмпирической) функцией распределения.
 28. Перечислите основные свойства статистических оценок.
- Перечислите основные типы гипотез, проверяемых в ходе статистической обработки измерений;
29. В чем сущность гипотезы согласия? Гипотезы об однородности выборок наблюдений? Гипотезы о числовых значениях параметров исследуемой генеральной совокупности? Гипотезы о типе зависимости между компонентами исследуемого признака?
 31. Что определяют уровень значимости и мощность статистических критериев?
 32. В чем состоит МНК?
 33. Какие метрики могут использоваться для построения альтернатив к МНК?

34. Опишите общую схему восстановления зависимости между двумя переменными.
35. Что называется линейной регрессией?
36. Как описывается линейная регрессия с несколькими переменными в скалярной форме?
37. Назовите свойства оценок по МНК.
38. Приведите различные определения системного анализа.
39. Назовите цель системного анализа.
40. Дайте классификацию СА в зависимости от уровня структуризации данных.
41. Назовите основные свойства системы.
42. Инструменты многомерного статистического анализа. Статистический анализ данных. Основные возможности, спектр решаемых задач.
43. Этапы выполнения многомерного статистического анализа и интерпретация результатов
44. Что называется прогнозом? Чем прогноз отличается от предсказания?
45. В чем состоит экстраполяционный прогноз? Экспертный прогноз?
46. Перечислите проблемы построения качественного, достоверного прогноза.
47. Приведите вариант классификации задач, решаемых на основе данной Data Mining.
48. Назовите особенности хранения информации в аналитических системах.
49. Приведите примеры программных продуктов на основе DM.
50. Дерево решений: принципы построения, использование, интерпретация результатов.
51. Метод поиска ассоциативных правил, области его применения.
52. Опишите модель нейрона.
53. Приведите пример искусственной нейронной сети.
54. Чем отличаются обучение с учителем и без учителя?
55. Нейросетевые методы анализа данных.
56. Принципы функционирования искусственных нейронных сетей. Основные области их применения.
57. Многослойные нейросети. Аппроксимация функциональных зависимостей. Прогнозирование временных рядов.
58. Что называется кластеризацией?
59. Когда используется термин таксономия?
60. Что называется признаками?
61. Приведите формализованную постановку задачи классификации?
62. Что такое признаковое пространство?
63. Что называется решающим правилом?
64. Приведите алгоритм типовой системы классификации.
65. Как реализуются логические методы классификации?
66. Перечислите методы оценивания информативности признаков
67. Приведите формализованную постановку задачи кластеризации.
68. Где используется кластер-анализ?
69. Приведите общее правило объединения двух кластеров.

70. Что называется разбиением? Иерархией?
71. В чем состоит метод «ближайшего соседа»? «Дальнего соседа»?
72. Что называется методом средней связи?
73. Назовите достоинства алгоритмов иерархической кластеризации,
74. Что называется внутрикластерным расстоянием? Межкластерным расстоянием?
75. стоянием?
76. Как определяется стабильность кластерной структуры?
77. Что называется главной компонентой?
78. Чему равна дисперсия главной компоненты?
79. Что называется функцией Лагранжа?
80. Как определяется физический смысл главных компонент?
81. Чем отличается факторный анализ от метода главных компонент?
82. Какими оптимальными свойствами обладают линейные главные компоненты?
83. Какая величина используется в качестве меры искажения матрицы попарных расстояний?

7.3. Задания к лабораторным работам

Задания к лабораторной работе 1. Обзор основных открытых баз и витрин данных в целях осуществления DATA VINING.

1. Рассмотреть основные базы открытых данных: Федеральной службы государственной статистики (Росстат), органов исполнительной власти на разных уровнях территориальной агрегации, Единой межведомственной информационно-статистической системы, Всероссийского центра исследования общественного мнения (ВЦИОМ), Центрального банка РФ, Национального агентства финансовых исследований, Аналитического центра Юрия Левады, Фонда общественного мнения, портала открытых данных РФ (Data.gov.ru), World Bank open data, Statista, International Monetary Fund data, Opec data, WTO statistics, kaggle.com и т.п.
2. Рассмотреть структуру и наполнение отдельных видов баз данных;
- 3: Выявить преимущества отдельных видов баз данных для осуществления интеллектуального анализа данных конкретных профессиональных задач.
4. Подготовить отчет по лабораторной работе 1.

Задание к лабораторной работе 2. Извлечение и анализ данных из открытых источников.

1. Осуществить выбор темы исследования, по которой будет производиться извлечение и анализ данных из открытых источников;
2. Произвести поиск необходимой информации в локальных поисковых системах;
3. Осуществить поиск необходимой информации в открытых базах данных;
4. Произвести обработку первичной информации, полученной на этапах 2 и 3 лабораторной работы;
5. Осуществить визуализацию данных, по рассматриваемой тематике собранных и обработанных на предыдущих этапах;

6. Произвести анализ и оценку данных по рассматриваемой тематике, полученных на предыдущих этапах лабораторной работы.

7. Подготовить отчет по лабораторной работе 2.

Задание к лабораторной работе 3. Осуществление корреляционного и регрессионного анализа с помощью различных инструментальных средств.

1. Осуществить выбор темы исследования, по которой будет производиться корреляционно-регрессионный анализ

2. Произвести первичную подготовку и обработку данных для осуществления корреляционного и регрессионного анализа с помощью различных инструментальных средств;

3. Определить парные и частные коэффициенты корреляции.

Проанализировать и объяснить их значения. Оценить значимость парных и частных коэффициентов корреляции. Сравнить соответствующие парные и частные коэффициенты корреляции. Сделать выводы.

4. Получить уравнение множественной регрессии в EXCEL, применив соответствующие процедуры. Пояснить смысл полученного уравнения, коэффициентов регрессии и константы уравнения.

5. Оценить значимость параметров уравнения множественной регрессии (2 способами). Проверку осуществить на 5% уровне значимости. Объяснить полученные результаты.

6. Вычислить значения $R_{yx_1x_2}$ $R^2_{yx_1x_2}$. Объяснить полученные результаты.

7. Проверить гипотезу о статистической значимости $R^2_{yx_1x_2}$. Сделать выводы.

8. Найти значение исправленного коэффициента детерминации. Объяснить его смысл.

9. Проверить значимость уравнения регрессии. Сделать выводы.

10. Выявить наличие (отсутствие) мультиколлинеарности в построенной модели, применив соответствующие тесты. Объяснить полученные результаты.

11. Выявить наличие (отсутствие) гетероскедастичности в построенной модели, применив соответствующие тесты. Объяснить полученные результаты.

12. Выявить наличие (отсутствие) автокорреляции в построенной модели (критерий Дарбина — Уотсона; критерий знаков). Объяснить полученные результаты.

13. Рассчитать средние коэффициенты эластичности. Дать их интерпретацию.

14. Определить среднюю ошибку аппроксимации. Сделать выводы.

15. Подготовка отчета по лабораторной работе 3.

Задание к лабораторной работе 4. Классификация и кластеризация данных

1. Осуществить выбор темы исследования, по которой будет производиться классификация и кластеризация данных

2. Осуществить выбор данных для осуществления классификации и кластерного анализа

3. Рассмотреть основные методы кластеризации данных

4. Осуществить классификацию выбранных данных

5. Выполнить кластеризацию объектов с помощью иерархических

агломеративных методов, используя различные меры расстояния между объектами:

6. Выполнить кластеризацию объектов на основе использования качественных признаков.
7. Провести кластеризацию объектов, используя метод К-средних и количественные показатели в качестве исходных данных.
8. Построить правило отнесения объектов к одному из выделенных классов (кластеров) на основе методов дискриминантного анализа.
9. Сформулировать выводы по проведенному кластерному анализу данных
10. — Подготовить отчета по лабораторной работе 4.

Задание к лабораторной работе 5. Методы графического представления и визуализации данных

1. Осуществить выбор темы исследования, по которой будет производиться графическое представление и визуализация данных
2. Произвести выбор данных для их дальнейшего графического представления и визуализации
3. Рассмотреть основные способы графического представления и визуализации данных.
4. Осуществить выбор способов графического представления и визуализации данных для решения конкретной профессиональной задачи на основе выбранных данных
5. Осуществить графическое представление и визуализацию выбранных данных
6. Подготовить отчет по лабораторной работе 5.

Задание к лабораторной работе 6. Этапы проектирования экспертной системы

1. Осуществить постановку цели, задач исследования; выбор объекта исследования.
2. Определить участников и их роли в процессе создания и эксплуатации экспертной системы
3. Произвести концептуализацию экспертной системы
4. Осуществить формализация экспертной системы
5. Произвести реализацию прототипной версии экспертной системы
6. Осуществить тестирование экспертной системы
7. Определить и проанализировать результаты проектирования экспертной системы.
8. Подготовить отчет по лабораторной работе 6.

Задание к лабораторной работе 7. Реализация и интерпретация результатов многомерного статистического анализа.

1. Осуществить постановку цели, задач исследования; выбор объекта исследования.
2. Произвести выбор данных для реализации многомерного статистического анализа
3. Осуществить корреляционно-регрессионный анализ на основе выбранных данных
4. Осуществить дискриминантный анализ с учетом цели и задач исследования

5. Произвести факторный анализ на основе выбранных данных
6. Сформулировать основные выводы на основе проведенного комплексного многомерного анализа
7. Подготовить отчет по лабораторной работе 7.

7.4. Экзаменационные билеты

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1

1. Основные термины и понятия интеллектуального анализа данных.
2. Алгоритм осуществления корреляционно-регрессионного анализа.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №2

1. Возможности автоматизации отдельных этапов анализа данных.
2. Инструментальные методы и основные этапы проведения анализа обеспечения ресурсами предприятия и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №3

1. Задачи классификации данных - постановки задачи, примеры практических приложений.
2. Инструментальные методы и этапы осуществления функционально-стоимостной анализа и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №4

1. Методы и способы получения первичной информации
2. Возможности применения ИАД в маркетинговой деятельности и в сфере логистики.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №5

1. Визуализация результатов интеллектуального анализа данных.
2. Основные этапы проведения анализа стратегической позиции предприятия и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №6

1. Понятие "данные". Набор данных и их атрибутов. Типы наборов данных. Измерения и шкалы.
2. Предварительная подготовка данных. Понижение размерности, сглаживание аномалий, фильтрация, группировка. .

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №7

1. Аналитические платформы Data Mining: основные возможности, сфера применения.
2. Инструментальные методы и основные этапы определения верхних границ цен и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №8

1. Эволюция методов обработки и анализа данных.
2. Инструментальные методы и основные этапы проведения анализа цепочек и издержек логистических процессов и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №9

1. Возможности автоматизации отдельных этапов анализа данных.
2. Инструментальные методы и основные этапы осуществления функционально-стоимостной анализа и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №10

1. Шумы и выбросы в данных. Пропуски данных и методы их восстановления. Задачи отбора признаков.
2. Инструментальные методы, основные этапы и особенности проведения анализа ставки дисконтирования с поправкой на риск и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №11

1. Базовые понятия и основные задачи интеллектуального анализа данных (ИАД). Этапы ИАД.
2. Особенности осуществления процедуры ИАД «Финансовая “паутина”» и интерпретация ее результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №12

1. Базы и витрины данных
2. Основные этапы анализа потоков платежей и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №13

1. Этапы решения задач с помощью аналитических платформ Data Mining. Импорт/экспорт данных.
2. Возможности применения методов ИАД для осуществления финансового и инвестиционного анализа.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №14

1. Понятие «несбалансированные выборки» и основные методики осуществления ИАД на их основе
2. Основные этапы проведения анализа маржинальной прибыли и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №15

1. Применение ИАД для бизнес-задач. Отбор данных для анализа бизнес-задач.
2. Основные этапы проведения портфолио-анализа и интерпретация его результатов.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №16

1. Эволюция методов обработки и анализа данных
2. Основные этапы проведения анализа возможных прерываний бизнес-процессов и интерпретация его результатов.

7.5. Вопросы для устного опроса, коллоквиума

1. Что такое информационно-аналитическая система?
2. Что вызвало появление и широкое распространение информационно аналитических систем?
3. В чем заключаются аспекты сбора и хранения информации? В чем состоит содержание аспектов анализа данных и предоставления результатов анализа пользователям?
4. Какие типы инструментальных средств для реализации информационно-аналитических систем вы можете назвать?
5. Какие информационные технологии и информационные системы на предприятии и из внешней среды являются источником данных для сосредоточения в информационном хранилище или непосредственно для анализа?
6. В каких видах информационных систем используются результаты анализа?
7. Для каких целей предназначена подсистема интеллектуального анализа данных?

- 8 Какие направления и методы поддерживают средства подсистемы интеллектуального анализа ИАС?
9. Какое значение имеет аналитическая работа для успеха предприятия?
10. Какие факторы влияют на деятельность предприятия?
11. Какие информационные ресурсы используются для подготовки принятия решений?
12. Что является предпосылкой для принятия правильных решений?
13. Какие объемы данных используются в процессе анализа?
14. Назовите требования к информации, которая используется для принятия решений.
15. Как организуется аналитическая работа на предприятии?
16. Из каких элементов состоит цепочка движения связанной с аналитической работой информации?
17. Перечислите источники, потребители информации, формы ее представления. В чем состоит содержание экономического анализа?
18. Что является предметом текущего и стратегического анализа?
19. Перечислите признаки, по которым методики анализа можно разделить на группы.
20. В чем состоит содержание методов анализа в маркетинговой деятельности?
21. Какие методы анализа применяются в областях обеспечения ресурсами и логистики? Каким образом поддерживаются методики анализа финансового состояния, инвестиций и инноваций средствами ИАС?
22. В чем сущность методик стратегического анализа?
23. Какие средства ИАС могут быть использованы при выполнении стратегического анализа?
24. Какие методики оценки текущего состояния предприятия Вы знаете?
25. В чем состоят методики анализа ситуации по слабым сигналам, оценки рисков и управления ими?
26. В чем состоит анализ отклонений?
27. Специфика анализа данных в менеджменте. Отбор данных для анализа. Предварительная подготовка данных.

7.6. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Экзамен проводится по расписанию промежуточной аттестации в письменном виде. В Экзаменационном задании — 2 теоретических вопроса. Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в экзаменационную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

Критерии оценивания отдельного задания к лабораторной работе:

Каждое задание к лабораторной работе оценивается отдельно по 10 баллов каждое. Максимальная общая оценка — 70 баллов.

- 9-10 баллов выставляется, если обучающийся: выполнил работу в полном объеме, самостоятельно, с соблюдением необходимой последовательности; грамотно оформил представленный отчет;
- 7-8 баллов выставляется, если обучающийся: выполнил работу в полном объеме, самостоятельно, с соблюдением необходимой последовательности; грамотно оформил представленный отчет; дана содержательная интерпретация полученных при решении задач результатов; материал изложен четко; допускаются отдельные логические и стилистические погрешности, уверенно исправленные после дополнительных вопросов;
- 5-6 баллов выставляется, если обучающийся: выполнил работу в полном объеме с соблюдением необходимой последовательности; грамотно оформил представленный отчет: дана содержательная интерпретация полученных результатов; допускаются отдельные логические и стилистические погрешности; обучающийся может испытывать некоторые затруднения в формулировке суждений;
- 0-4,9 баллов выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

Критерии оценивания экзамена:

Максимальное количество баллов 100. Каждый ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ содержит 2 вопроса. Ответ на каждый вопрос оценивается отдельно, максимально 50 баллов каждый.

Критерии оценивания устного опроса, собеседования

Максимальный балл — 20.

В течение семестра обучающийся отвечает максимум на 20 вопросов. Ответ на каждый вопрос оценивается максимум в 1 балл.

Критерии оценивания Заданий к лабораторным работам:

Каждое задание к лабораторной работе оценивается отдельно по 20 баллов каждая.

Максимальная общая оценка — 80 баллов.

Методические рекомендации по написанию, требования к оформлению отчета по лабораторной работе

Отчет должен содержать следующее: титульный лист; содержание отчета; исходные данные для осуществления расчетов и анализа; основные этапы осуществления лабораторной работы в соответствии с программой; выводы, предложения и критические замечания; приложения.

Требования, предъявляемые к содержанию основных разделов отчета:

- четкость и логическая последовательность изложения материала;
- Убедительность аргументации;
- краткость и четкость формулировок, исключающих возможность неоднозначного толкования.

Отчет должен содержать описание действий с иллюстрациями, формулами, результаты выполнения лабораторной работы в тех или иных программных средствах.

Отчет по лабораторной работе выполняется на стандартных листах белой бумаге, на одной стороне, формата А4 (210х297 мм). Текст печатается шрифтом Times New Roman № 12 через 1

интервал.

Текст работы должен быть аккуратно оформлен и экономически грамотно изложен с учетом требований современной орфографии.

Все листы работы (текстовые, табличные) должны быть выполнены с соблюдением следующих минимальных размеров полей: с левой стороны - 30 мм; правой - не менее 10 мм; сверху

- и снизу - 20 мм. При этом текст рамкой не очерчивается.

Страницы в отчете должны иметь сквозную нумерацию.

оценки "*отлично*" заслуживает студент, обнаруживший всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоивший основную и знакомый с дополнительной литературой, рекомендованной программой. Как правило, оценка "*отлично*" выставляется студентам, усвоившим взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявившим творческие способности в понимании, изложении и использовании учебно-программного материала.

- оценки "*хорошо*" заслуживает студент, обнаруживший полное знание учебно-программного материала, успешно выполняющий предусмотренные в программе задания, усвоивший основную литературу, рекомендованную в программе. Как правило, оценка "*хорошо*" выставляется студентам, показавшим систематический характер знаний по дисциплине и способным к их самостоятельному пополнению и обновлению в ходе дальнейшей учебной работы и профессиональной деятельности.

-оценки "*удовлетворительно*" заслуживает студент, обнаруживший знания основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по специальности, справляющийся с выполнением заданий, предусмотренных программой, знакомый с основной литературой, рекомендованной программой. Как правило, оценка "*удовлетворительно*" выставляется студентам, допустившим погрешности в ответе на экзамене и при выполнении экзаменационных заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя.

-оценка "*неудовлетворительно*" выставляется студенту, обнаружившему пробелы в знаниях основного учебно-программного материала, допустившему принципиальные ошибки в выполнении предусмотренных программой заданий. Как правило, оценка "*неудовлетворительно*" ставится студентам, которые не могут продолжить обучение или приступить к профессиональной деятельности по окончании вуза без дополнительных занятий по соответствующей дисциплине.

Рекомендуемые границы оценок:

- «отлично» - не менее 86% правильных ответов,
- «хорошо» - 66-85% правильных ответов,
- «удовлетворительно» - 51-65% правильных ответов,
- «неудовлетворительно» - менее 50% правильных ответов.

Конечный результат складывается как средневзвешенная оценка текущего и промежуточного контролей соответственно с весами 50% .

Текущий контроль по дисциплине включает:

—посещение занятий – 10 баллов,

- участие на практических занятиях – 10 баллов
- выполнение лабораторных заданий – 10 баллов,
- коллоквиум – 30 баллов,
- выполнение аудиторных контрольных работ – 40 баллов

Промежуточный контроль по дисциплине включает:

- защита лабораторных работ – 60 баллов;
- письменная контрольная работа – 40 баллов.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.

а). основная литература:

1. Чубукова И. А. Data Mining: Учебное пособие Москва: Интернет- Университет Информационных Технологий (ИНТУИТ) Бином. Лаборатория знаний, 2008
2. Нестеров С. А.Интеллектуальный анализ данных средствами MS SQL Server 2008 Москва: Национальный Открытый Университет «ИНТУИТ», 2016
3. Жуковский О. И. Информационные технологии и анализ данных: учебное пособие.-Томск: Эль Контент, 2014
4. Байдаков А. Н., Звягинцева О. С., Назаренко А. В., Запорожец Д. В., Бабкина О. Н. Моделирование бизнес-процессов: учебное пособие, Ставрополь: Ставропольский государственный аграрный университет, 2017
5. Кремер, Н. Ш., Путко Б. А., Кремер, Н. Ш. Эконометрика: учебник для студентов вузов Москва: ЮНИТИ-ДАНА, 2017
6. Мыльников Л. А., Краузе Б., Кютц М., Баде К., Шмидт И. А. Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография Москва: Библио-Глобус, 2017

Дополнительная литература

1. Елисеева И. И. Эконометрика: учеб. для бакалавриата и магистратуры. М.: Юрайт, 2016
2. Журнал "Вопросы статистики",
3. Мещихина Е. Д., Иванов О. Е. Информационные системы и технологии в экономике: учебное пособие Йошкар-Ола: Марийский государственный технический университет, 2012
4. Data Mining Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016
5. Пальмов, С. В. Интеллектуальный анализ данных: учебное пособие Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017
6. Воронов, В. И., Воронова, Л. И., Усачев, В. А. Data Mining - технологии обработки больших данных: учебное пособие Москва: Московский технический университет связи и информатики, 2018
7. Анализ данных. Часть 2. Инструменты Data Mining: Учебное пособие Москва: Московский городской педагогический университет, 2012

Профессиональные базы данных и информационные справочные системы

Федеральная служба государственной статистики (Росстат)- http://www.rosstat.gov.ru
--

Единая межведомственная информационно-статистическая система- http://www.fedstat.ru
Всероссийский центр исследования общественного мнения (ВЦИОМ)- http://www.wciom.ru
Центральный банк РФ- http://www.cbr.ru
Национальное агентство финансовых исследований- http://www.nafi.ru
Аналитический центр Юрия- http://www.levada.ru
Фонд общественного мнения- http://www.fom.ru
Портала открытых данных РФ- http://www.data.gov.ru
World Bank open data- http://www.worldbank.org
Statistic portal for market data- http://www.statista.com
International Monetary Fund data- http://www.imf.org
WTO statistics- http://www.wto.ru
Machine learning and data science community- http://www.kaggle.com
Консультант +

Перечень программного обеспечения

MS Office
E-Views 6.0

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

1. Федеральный портал российское образование <http://edu.ru>;
2. Электронные каталоги Научной библиотеки Даггосуниверситета <http://elib.dgu.ru/?q=node/256>;
3. Образовательные ресурсы сети Интернет <http://catalog.iot.ru/index.php>;
4. Электронная библиотека <http://elib.kuzstu.ru>.
5. ЭБС «Юрайт»: <http://www.biblio-online.ru/>

10. Методические указания для обучающихся по освоению дисциплины.

Лекционный курс. Лекция является основной формой обучения в высшем учебном заведении. В ходе лекционного курса «Data Mining» проводится систематическое изложение научных и познавательных материалов, освещение основных понятий дисциплины и закрепление теоретического материала.

В тетради для конспектирования лекций необходимо иметь поля, где по ходу конспектирования студент делает необходимые пометки. В конспектах рекомендуется применять сокращения слов, что ускоряет запись.

Студенту необходимо активно работать с конспектом лекции: после окончания лекции рекомендуется перечитать свои записи, внести поправки и дополнения на полях. Конспекты лекций можно использовать при подготовке к экзамену, контрольным тестам, при выполнении самостоятельных заданий.

Практические занятия. Практические занятия по «Data Mining» имеют цель закрепить теоретические знания по дисциплине, изложенные на лекции, решая практические задачи. На практическом занятии студент должен иметь тетрадь для практических занятий, в которую записываются все задачи решенные в аудитории и дома самостоятельно.

Важное место в самостоятельной работе студентов должна занимать работа в

образовательной среде ИНТЕРНЕТ. Такие ресурсы указаны в разделе «Программное обеспечение и интернет ресурсы» данной рабочей программы.

Лабораторные занятия. На лабораторных занятиях студент должен научиться решать с помощью ЭВМ практические задачи по информационным системам и технологиям в статистике. При этом главное – научиться пользоваться информационными технологиями при решении задач статистики. Рейтинг направлен на повышение ритмичности и эффективности самостоятельной работы студентов. Он основывается на широком использовании тестов и заинтересованности каждого студента в получении более высокой оценки знаний по дисциплине.

Принципы рейтинга: непрерывный контроль (в идеале на каждом из аудиторных занятий) и получение более высокой оценки за работу, выполненную в срок. При проведении практических занятий необходимо предусматривать широкое использование активных и интерактивных форм (компьютерных симуляций, деловых и ролевых игр).

Рейтинг включает в себя два вида контроля: текущий, промежуточный и итоговый по дисциплине.

Текущий контроль (ТК) - основная часть рейтинговой системы, основанная на беглом опросе раз в неделю или в две недели. Формы: тестовые оценки в ходе практических занятий, оценки за выполнение индивидуальных заданий и лабораторных работ. Важнейшей формой ТК, позволяющей опросить всех студентов на одном занятии являются короткие тесты из 2-3 тестовых заданий.

Основная цель ТК: своевременная оценка успеваемости студентов, побуждающая их работать равномерно, исключая малые загрузки или перегрузки в течение семестра.

Лекционные занятия желательно проводить в режиме презентаций с демонстрацией применения основных методов анализа и синтеза, что улучшает динамику лекций. Основное время лекции лучше тратить на подробные аналитические комментарии и особенности применения рассматриваемого материала в профессиональной деятельности студента.

Практические занятия следует проводить, используя профессиональные программы.

Лабораторный практикум проводится фронтальным методом в специально оборудованных классах. Так как используется компьютерное моделирование, то следует проводить занятия в компьютерном классе либо самостоятельно на домашнем компьютере. При этом и коллоквиум, и защита результатов исследований проводятся по традиционной методике в классе.

Промежуточный контроль (ПК) - это проверка знаний студентов по разделу программы. Формы: контрольная работа из 3-5 заданий.

Цель ПК: побудить студентов отчитаться за усвоение раздела дисциплины накопительным образом, т.е. сначала за первый, затем за второй, затем за третий разделы каждого семестра.

Итоговый контроль по дисциплине Data Mining - это проверка уровня учебных достижений студентов по всей дисциплине за семестр. Формы контроля: экзамен в 5-ом семестре. Цель итогового контроля: проверка базовых знаний по дисциплине, полученных при изучении всех модулей семестра, после которого можно рассчитывать на то, что процесс обучения по дисциплине завершен и в дальнейшем студент может сам при

необходимости совершенствовать свои знания.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

Для успешного освоения дисциплины, обучающийся использует следующие программные средства: пакеты для решения задач математического программирования: Mathcad, Delphi, Matlab, Pethon, C++.

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

Специальные помещения представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, для самостоятельной работы.

Все лекционные аудитории укомплектованы мультимедийными и техническими средствами обучения. В каждой аудитории 35 рабочих мест. Аудитории для семинарских занятий оснащены доской, рабочими местами для студентов в объеме 25-30. Лабораторные занятия проводятся по подгруппам в компьютерных классах. Компьютерные классы оснащены необходимым числом компьютеров и мультимедийным оборудованием. На компьютерах установлено необходимое программное обеспечение.

Электронно-библиотечные системы (электронная библиотека) и электронная информационно-образовательная среда обеспечивает одновременный доступ обучающихся, включая удаленный доступ к современным профессиональным базам данных и информационным справочным системам.