

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Кубанский государственный университет»
(ФГБОУ ВО «КубГУ»)

Факультет компьютерных технологий и прикладной математики
Кафедра математического моделирования

УТВЕРЖДАЮ

Проректор по учебной работе,
качеству образования – первый
проректор

Иванов А.Г.

подпись

«30»

06

2017 г.



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

**Б1.В.02 « МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПРЕДСТАВЛЕНИЯ И
АНАЛИЗА МОДЕЛЕЙ»**

Направление подготовки 01.04.02 Прикладная математика и информатика
Профиль Математическое моделирование

Квалификация (степень) выпускника – магистр
Форма обучения: очная

Краснодар 2017

Рабочая программа дисциплины Математические методы представления и анализа моделей составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 01.04.02 Прикладная математика и информатика, направленность (профиль): Математическое моделирование.

Программу составили:

Кармазин В.Н. – кандидат физико-математических наук, доцент



ПОДПИСЬ

Заведующий кафедрой прикладной математики
д.ф.-м.н., профессор М.Х. Уртенов
«29» июня 2017 г.



ПОДПИСЬ

Рабочая программа обсуждена на заседании кафедры
математического моделирования от 21 июня 2017 г.,
протокол № 16.

Заведующий кафедрой математического
моделирования Бабешко В.А.



Утверждена на заседании учебно-методической комиссии факультета
компьютерных технологий и прикладной математики от 29 июня 2017 г.,
протокол № 4.

Председатель УМК факультета компьютерных технологий и прикладной

математики к.ф.-м.н., доцент К.В. Малыхин
Рецензенты:



Шапошникова Татьяна Леонидовна.

Доктор педагогических наук, кандидат физико-математических наук,
профессор. Почетный работник высшего профессионального образования
РФ. Директор института фундаментальных наук (ИФН) ФГБОУ ВО
«КубГТУ».

Марков Виталий Николаевич.

Доктор технических наук. Профессор кафедры информационных систем и
программирования института компьютерных систем и информационной
безопасности (ИКСИБ) ФГБОУ ВО «КубГТУ».

1. Цели и задачи учебной дисциплины

1.1 Цели изучения дисциплины определены государственным образовательным стандартом высшего образования и соотнесены с общими целями ООП ВО по направлению подготовки «Прикладная математика и информатика», в рамках которой преподается дисциплина.

Целью освоения учебной дисциплины «Математические методы представления и анализа моделей» является овладение знаниями и навыками интеллектуального анализа данных при решении ряда прикладных задач производственной и научно-исследовательской деятельности.

1.2 Задачи дисциплины:

- актуализация и развитие знаний в области математических моделей;
- изучение существующих технологий подготовки данных к анализу;
- овладение практическими умениями и навыками реализации технологий интеллектуального анализа данных, формирования и проверки гипотез об их природе и структуре, варьирования применяемыми моделями;
- формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

1.3 Место учебной дисциплины в структуре ООП ВО

Курсы обязательные для предварительного изучения: теория вероятностей и математическая статистика, концепции современного естествознания, численные методы, методы оптимизации, программирование.

Дисциплина направлена на формирование знаний, практических умений и навыков по применению современных методов интеллектуального анализа данных в различных сферах человеческой деятельности, формирует у обучающихся способность к теоретико-методологическому анализу проблем поиска новых, нетривиальных закономерностей с помощью интеллектуального анализа данных; формирование компетенций в анализе методов и процедур интеллектуального анализа данных. В совокупности изучение этой дисциплины готовит обучаемых как к различным видам практической экспертной и аналитической деятельности, так и к научно-теоретической, исследовательской деятельности.

Дисциплины, в которых используется материал данной дисциплины: исследование операций и системный анализ, математическое моделирование экономических систем, при прохождении практик и подготовке магистерской диссертации.

1.4 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Компетенции обучающегося, формируемые в результате освоения курса «Математические методы представления и анализа моделей»:

Код компетенции	Формулировка компетенции
ОК-1	способностью к абстрактному мышлению, анализу, синтезу
Знать	– тенденции разработки универсальных программных средств и аналитических платформ предназначенных для анализа данных, построения прогнозов и аналитических сценариев
Уметь	– проводить научные исследования и получать новые научные и прикладные результаты на основе интеллектуального анализа данных
Владеть	– навыками практического применения методов консолидации, трансформации, визуализации, оценки качества, очистки и предобработки

	данных для качественной подготовки данных к анализу
--	---

ОПК-3	способностью самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение
-------	---

Знать	– возможности отечественных и зарубежных универсальных программных средств и аналитических платформ, применяемых для анализа данных
Уметь	– использовать возможности отечественных и зарубежных универсальных программных средств и аналитических платформ для поиска закономерностей и связей, правил и знаний в электронных массивах данных
Владеть	– способностью применять методы анализа прикладной области на различных уровнях

ПК-1	способностью проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива
Знать	– возможности отечественных и зарубежных универсальных программных средств и аналитических платформ, применяемых для анализа данных
Уметь	– способностью проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе, участвовать в реинжиниринге прикладных и информационных процессов
Владеть	– применять системный подход и математические методы исследования операций к решению задач математического моделирования.

ПК-4	способностью разрабатывать и анализировать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности
Знать	– вопросы визуализации данных, их очистки и предобработки
Уметь	– ориентироваться в технологиях анализа данных
Владеть	– навыками сравнения моделей.

ПК-9	способностью к преподаванию математических дисциплин и информатики в общеобразовательных организациях, профессиональных образовательных организациях и образовательных организациях высшего образования
------	---

Знать	– основные информационные ресурсы для получения новых знаний
Уметь	– получать новые знания и умения с помощью информационных технологий
Владеть	– навыками работы с различными источниками информации

ПК-11	способностью разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологий
Знать	– средства получения, переработки и представления информации с помощью информационно-коммуникационных технологий
Уметь	– применять полученные знания для использования в практической деятельности анализа данных
Владеть	– способностью организовывать процессы анализа моделей и систем на основе технологий развития корпоративных баз знаний

2. Структура и содержание дисциплины

2.1 Распределение трудоемкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зач.ед. (72 часа), их распределение по видам работ представлено в таблице.

Вид учебной работы		Всего часов	Семестры (часы)			
			9	—		
Контактная работа, в том числе:						
Аудиторные занятия (всего):		32	32			
Занятия лекционного типа		16	16	-	-	-
Лабораторные занятия		16	16	-	-	-
Занятия семинарского типа (семинары, практические занятия)		-	-	-	-	-
		-	-	-	-	-
Иная контактная работа:						
Контроль самостоятельной работы (КСР)		-	-			
Промежуточная аттестация (ИКР)		0,3	0,3			
Самостоятельная работа, в том числе:						
<i>Курсовая работа</i>		-	-	-	-	-
<i>Проработка учебного (теоретического) материала</i>		4	4	-	-	-
<i>Выполнение индивидуальных заданий (подготовка сообщений, презентаций)</i>		3	3	-	-	-
<i>Реферат</i>		3	3	-	-	-
Подготовка к текущему контролю		3	3	-	-	-
Контроль:						
Подготовка к экзамену		26,7	26,7			
Общая трудоемкость	час.	72	72	-	-	-
	в том числе контактная работа	32,3	32,3			
	зач. ед	2	2			

2.2 Структура учебной дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы дисциплины, изучаемые в семестре 9

Учебно-тематический план очной формы обучения

№	Наименование разделов	Количество часов			
		Всего	Аудиторная работа		Самостоятельная работа
			Л	ЛР	
Введение в анализ данных					
1	Технологии анализа данных.	3	2	–	1
2	Консолидация и трансформация данных.	4	1	2	1
3	Визуализация данных. Очистка и предобработка данных.	4	1	2	1
Технология анализа данных Data Mining					

№	Наименование разделов	Количество часов			
		Всего	Аудиторная работа		Самост. работа
			Л	ЛР	
4	Data Mining: задача ассоциации.	7	2	4	1
5	Data Mining: кластеризация.	7	2	4	1
6	Data Mining: классификация и регрессия. Статистические методы.	6	2	2	2
7	Data Mining: классификация и регрессия. Машинное обучение.	6	2	2	2
Анализ моделей					
8	Ансамбли моделей.	4	2	–	2
9	Сравнение моделей.	4	2	–	2
	Всего по разделам дисциплины:	45	16	16	13
	ИКР	0,3			
	Контроль	26,7			
Итого:		72	16	16	13

2.3 Содержание разделов дисциплины:

№ раздела	Наименование раздела/модуля	Форма текущего контроля
1	2	4
1	Введение в анализ данных.	1. Опрос по результатам индивидуального задания. 2. Отчет по бизнес-задаче. 3. Проверка выполнения лабораторных работ № 1, 2.
2	Технология анализа данных Data Mining.	1. Опрос по результатам индивидуального задания. 2. Отчет по бизнес-задаче. 3. Проверка выполнения лабораторных работ № 3-6.
3	Анализ моделей.	1. Подготовка рефератов, презентаций, выступлений. 2. Отчет по бизнес-задаче.

защита лабораторной работы (ЛР), выполнение курсового проекта (КП), курсовой работы (КР), расчетно-графического задания (РГЗ), написание реферата (Р), эссе (Э), коллоквиум (К), тестирование (Т)

2.3.1 Занятия лекционного типа

№ раздела	Наименование раздела/модуля	Содержание раздела/модуля	Форма текущего контроля
-----------	-----------------------------	---------------------------	-------------------------

	ля		
1	2	3	4
1	Введение в анализ данных.	Тема 1. Технологии анализа данных. Принципы анализа данных. Структурированные данные. Подготовка данных к анализу. Технологии KDD и Data Mining. Аналитические платформы. Введение в алгоритмы Data Mining. Тема 2. Консолидация и трансформация данных. Задача консолидации. Основные концепции хранилищ данных. Многомерные, реляционные, гибридные, виртуальные хранилища данных. Нечеткие срезы. Извлечение, очистка и преобразование данных в ETL. Загрузка данных в хранилище и из локальных источников. Обогащение данных. Введение в трансформацию данных. Трансформация упорядоченных данных. Группировка данных. Слияние данных. Квантование. Нормализация и кодирование данных. Тема 3. Визуализация данных. Очистка и предобработка данных. Введение в визуализацию. Визуализаторы общего назначения. OLAP -анализ. Визуализаторы для оценки качества моделей. Визуализаторы, применяемые для интерпретации результатов анализа. Технологии и методы оценки качества данных. Очистка и предобработка. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений. Введение в сокращение размерности. Сокращение числа признаков. Сокращение числа значений признаков и записей. Сэмплинг.	4. Опрос по результатам индивидуального задания. 5. Отчет по бизнес-задаче.
2	Технология анализа данных Data Mining.	Тема 4. Data Mining : задача ассоциации. Ассоциативные правила. Значимость ассоциативных правил. Поиск ассоциативных правил. Алгоритм Apriori. Генерация ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил. Последовательные шаблоны. Поиск последовательных шаблонов. Алгоритм AprioriAll. Тема 5. Data Mining : кластеризация. Введение в кластеризацию. Алгоритм кластеризации k-means. Нечеткая кластеризация. Сети Кохонена. Обучение сети Кохонена. Карты Кохонена. Методика построения карты. Выбор числа нейронов карты. Недостатки карт Кохонена. Проблемы алгоритмов кластеризации. Неопределенность в выборе критерия качества кластеризации. Выбор числа кластеров.	1. Опрос по результатам индивидуального задания. 2. Отчет по бизнес-задаче.

		Тема 6. Data Mining : классификация и регрессия. Статистические методы. Введение в классификацию и регрессию. Простая линейная регрессия. Оценка соответствия простой линейной регрессии реальным данным. Простая регрессионная модель. Модель множественной линейной регрессии. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Основы логистической регрессии. Интерпретация модели логистической регрессии. Множественная логистическая регрессия. Простой байесовский классификатор. Тема 7. Data Mining : классификация и регрессия. Машинное обучение. Введение в деревья решений. Алгоритмы построения деревьев решений. Упрощение деревьев решений. Введение в нейронные сети. Искусственный нейрон. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей. Классификация с помощью нейронных сетей. Прогнозирование с помощью нейронных сетей.	
3	Анализ моделей.	Тема 8. Ансамбли моделей. Введение в ансамбли моделей. Виды ансамблей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей. Тема 9. Сравнение моделей. Оценка эффективности и сравнение моделей. Оценка ошибки модели. Издержки ошибочной классификации. Lift- и Profit- кривые. ROC- анализ. Обучение в условиях несбалансированности классов.	1. Опрос по результатам индивидуального задания. 2. Отчет по бизнес-задаче.

2.3.2 Семинарские занятия – не предусмотрены

2.3.3 Лабораторные занятия

№ п/п	Наименование раздела	Наименование лабораторных работ	Форма текущего контроля
1.	Введение в анализ данных.	Консолидация и трансформация данных в аналитической платформе Deductor 5.0.	Проверка выполнения лабораторных работ № 1
		Визуализация, очистка и предобработка данных в	Проверка выполнения

		аналитической платформе Deductor 5.0.	лабораторных работ № 2
2.	Технология анализа данных Data Mining.	Data Mining : задача ассоциации.	Проверка выполнения лабораторных работ № 3
		Data Mining : кластеризация..	Проверка выполнения лабораторных работ № 4
		Data Mining : классификация и регрессия. Статистические методы.	Проверка выполнения лабораторных работ № 5
		Data Mining : классификация и регрессия. Машинное обучение.	Проверка выполнения лабораторных работ №6

2.3.4 Курсовые работы – не предусмотрены

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающегося по дисциплине

Целью самостоятельной работы студента является углубление знаний, полученных в результате аудиторных занятий. Вырабатываются навыки самостоятельной работы. Закрепляются опыт и знания, полученные во время лабораторных занятий. Помещения для самостоятельной работы студентов – аудитория № 102-А и читальный зал.

№	Вид самостоятельной работы	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к семинарским занятиям	Методические указания для подготовки к лекционным и семинарским занятиям, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г. Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г.
2	Подготовка к лабораторным занятиям	Методические указания по выполнению лабораторных работ, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г.
3	Подготовка к решению задач и тестов	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г.
4	Подготовка докладов	Методические указания для подготовки эссе, рефератов, курсовых работ, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г.
5	Подготовка к решению расчетно-графических заданий	Методические указания по выполнению расчетно-графических заданий, утвержденные на заседании кафедры прикладной

	заданий (РГЗ)	математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г. Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г.
6	Подготовка к текущему контролю	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры прикладной математики факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №22 от 29.06.2017 г.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

С точки зрения применяемых методов используются как традиционные информационно-объяснительные лекции, так и интерактивная подача материала с мультимедийной системой. Компьютерные технологии в данном случае обеспечивают возможность разнопланового отображения алгоритмов и демонстрационного материала. Такое сочетание позволяет оптимально использовать отведенное время и раскрывать логику и содержание дисциплины.

Лекции представляют собой систематические обзоры математических методов представления и анализа моделей с подачей материала в виде презентаций.

Лабораторное занятие позволяет научить студента применять теоретические знания при решении и исследовании конкретных задач. Лабораторные занятия проводятся в компьютерных классах, при этом практикуется работа в группах. Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы. Это обусловлено тем, что в процессе исследования часто встречаются задачи, для которых единых подходов не существует. Каждая конкретная задача при своем исследовании имеет множество подходов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Занятия, проводимые с использованием интерактивных технологий

№	Наименование разделов (тем)	Количество часов	
		всего ауд. часов	интерактивные часы
1	2	3	4
1.	Введение в анализ данных.	8	3
2.	Технология анализа данных Data Mining.	20	6
3.	Анализ моделей.	4	1
	Итого по дисциплине:	32	10

4. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

4.1 Фонд оценочных средств для проведения текущей аттестации

Учебная деятельность проходит в соответствии с графиком учебного процесса. Процесс самостоятельной работы контролируется во время аудиторных занятий и индивидуальных консультаций. Самостоятельная работа студентов проводится в форме изучения отдельных теоретических вопросов по предлагаемой литературе.

Фонд оценочных средств дисциплины состоит из средств текущего контроля (см. список лабораторных работ, задач и вопросов) и итоговой аттестации (зачета).

В качестве оценочных средств, используемых для текущего контроля успеваемости, предлагается перечень вопросов, которые прорабатываются в процессе освоения курса. Данный перечень охватывает все основные разделы курса, включая знания, получаемые во время самостоятельной работы. Кроме того, важным элементом технологии является самостоятельное решение студентами и сдача заданий. Это полностью индивидуальная форма обучения. Студент рассказывает свое решение преподавателю, отвечает на дополнительные вопросы.

Примерные задания на лабораторные работы

Лабораторная работа 1: Консолидация и трансформация данных в аналитической платформе Deductor 5.0

Цель работы:

1. Ознакомиться с архитектурой, основными частями и пользовательским интерфейсом Deductor.
2. Получить навыки создания сценариев обработки и визуализации данных, создания и наполнения хранилища данных.

Этапы выполнения лабораторной работы

1. Назначение и компоненты аналитической платформы Deductor.
2. Хранилище данных - Deductor Warehouse.
3. Создание нового хранилища данных.
4. Проектирование структуры хранилища данных.
5. Загрузка информации в хранилище.
6. Извлечение информации из хранилища.

Варианты заданий:

1. Создайте пустое хранилище данных **Фармация**, спроектируйте его структуру и загрузите в него информацию из следующих текстовых файлов: **groups.txt**, **produces.txt**, **stores.txt**, **sales.txt**. Результатом работы должен стать сценарий загрузки **load.txt**.
2. Убедитесь, что в хранилище загружена вся информация о продажах.
3. Импортируйте информацию о продажах из **ХД Фармация**, включая атрибуты товара.
4. Установите следующие срезы: а) кроме последнего периода в 1 месяц от имеющихся данных; б) по какой-либо одной товарной группе.

Вопросы для самоконтроля

1. Для каких целей используют хранилища данных?
2. Из каких частей состоит такое хранилище?
3. Что представляет собой семантический слой и для чего он нужен?
4. Опишите структуру данных «снежинка».
5. Приведите примеры объектов хранилища данных.

Лабораторная работа 2: Визуализация, очистка и предобработка данных в аналитической платформе Deductor 5.0

Цель работы:

1. Освоить навыки извлечения информации из хранилища данных.
2. Освоить навыки построения многомерных отчетов и их анализа.

Этапы выполнения лабораторной работы

1. Оперативная аналитическая обработка данных (OLAP), многомерный анализ данных.
2. OLAP в DEDUCTOR.
3. Разработка системы аналитической отчетности.

Варианты заданий:

Разработайте систему аналитической отчетности для созданного в лабораторной работе 1 хранилища данных **Фармация**. Для этого напишите в Deductor Studio сценарий обработки данных, сохраните его под именем **olap.ded** и выберите любые 5-7 пунктов из списка, приведенного ниже. Кроме кросс-таблицы, самостоятельно изучите кросс-диаграмму.

1. Постройте кросс-таблицу и кросс-диаграмму по трем измерениям (отдел, месяц года, товарная групп), в ячейках которой отображается стоимость (сумма) и объем (количество единиц) проданной продукции за все периоды, данные по которым имеются в ХД. Определите, какая торговая точка и какая товарная группа дает наибольшую сумму продаж. Постройте кросс-диаграмму суммы продаж, включающую общую сумму продаж, продажи по торговым точкам, продажи по товарным группам.
2. То же за последние 3 месяца, данные по которым имеются в хранилище.
3. То же за последние 3 недели по имеющимся данным.
4. Найдите сумму максимальной и средней стоимости покупки за последний месяц по имеющимся данным.

5. Сформируйте многомерный отчет и график загруженности торговых точек по времени суток и торговым точкам. Определите, на какие часы приходятся пики продаж.
6. То же за последние 3 месяца по имеющимся данным.
7. Сформируйте многомерный отчет и график загруженности торговых точек по дням недели.
8. То же за последний месяц по имеющимся данным.
9. Сформируйте многомерный отчет и график загруженности торговых точек по дням месяца.
10. То же за последние 3 месяца по имеющимся данным.
11. Определите 20 товаров, дающих самый большой суммарный объем продаж за все периоды.
12. То же за последние 3 недели по имеющимся данным.
13. Определите 10 товаров, дающих самый большой объем продаж по воскресеньям.
14. Определите товары, дающие 50% объема продаж.
15. То же, но за последние 3 месяца по имеющимся данным.
16. То же, но за последнюю неделю по имеющимся данным.
17. Определите 10 товаров, дающих самый большой объем продаж с 18 до 21 ч.
18. Определите товары, дающие летом 50% объема продаж.

Вопросы для самоконтроля

1. Что представляет собой OLAP-анализ, в чем состоит его назначение?
2. Опишите технологию OLAP-анализа.
3. Какие преимущества дает использование кросс-таблиц и кросс-диаграмм?
4. Как осуществляется фильтрация данных в кросс-таблицах?
5. Для чего нужна аналитическая отчетность и как она реализована в системе Deductor?

Лабораторная работа 3: Data Mining : задача ассоциации

Цель работы:

1. Изучить основные свойства ассоциативных правил.
2. Освоить технику применения ассоциативных правил на примере задачи стимулирования розничных продаж.

Этапы выполнения лабораторной работы

- Вычисление поддержки и достоверности правил.
- Значимость ассоциативных правил.
- Поиск ассоциативных правил.
- Построение набора ассоциативных правил.
- Интерпретация полученных правил.

Варианты заданий:

Небольшая сеть из трех магазинов, продающих мелкие штучные товары, желает провести исследование связанных покупок. По мнению специалистов компании, знание того, какие товары покупаются совместно, поможет правильно расположить их на витринах. С этой целью были собраны все чеки за последние три месяца (около 20 тыс.). В них присутствуют 17 товаров (анальгетик, дезодорант, журнал, зубная паста, карандаши и др.). Таблица данных содержится в файле **transactions.txt** и включает два столбца:

Транзакция и Товар.

Опираясь на имеющиеся данные, выполните следующие действия.

1. Решите задачу поиска ассоциаций в Deductor.
2. Выделите непонятные, на ваш взгляд, ассоциативные правила, а также правила, представляющие интерес. Сколько правил попало в эти категории?
3. Найдите правило, имеющее максимальный лифт.
4. Заказчика данного исследования интересует, какие товары покупают с поздравительной открыткой. Сколько таких товаров оказалось в выбранном перечне. Какая из ассоциаций представляет в этом плане наибольший интерес (имеет максимальный лифт).

Вопросы для самоконтроля

1. В чем состоит цель аффинитного анализа и какие практические задачи решаются с его помощью?
2. Дайте определение понятия «транзакция».
3. Из каких элементов состоит ассоциативное правило?
4. Как рассчитывается поддержка и достоверность ассоциативного правила и какое значение имеют эти показатели в процессе их поиска?
5. В чем состоит смысл показателя «лифт»?
6. Какие параметры необходимо указать при настройке алгоритма *a priori*?
7. На какие группы подразделяются ассоциативные правила при их содержательной интерпретации?
8. Перечислите рекомендации, которых следует придерживаться при анализе ассоциаций в программе Deductor и объясните их смысл.

Лабораторная работа 4: Data Mining : кластеризация

Цель работы:

1. Ознакомиться с возможностями выделения в имеющихся данных однородных групп и их изучения с помощью средств кластерного анализа.
2. Освоить технику применения алгоритма кластеризации k-means.
3. Освоить технику применения самоорганизующихся карт Кохонена для решения задачи кластеризации.

Этапы выполнения лабораторной работы

1. Нормализация исходных данных.
2. Выбор меры близости объектов и кластеров.

3. Кластеризация с помощью k-means.
4. Построение карты Кохонена.
5. Интерпретация и сравнение полученных результатов.

Варианты заданий:

Руководство филиала региональной телекоммуникационной компании, предоставляющей услуги мобильной связи, поставило задачу сегментации абонентской базы. Ее целями являются:

1. Построение профилей абонентов путем выявления их схожего поведения в плане частоты, длительности и времени звонков, а также ежемесячных расходов.
2. Оценка наиболее и наименее доходных сегментов.
3. Разработка маркетинговых акций, направленных на определенные группы клиентов.
4. Разработка новых тарифных планов.
5. Оптимизация расходов на адресную SMS-рассылку о новых услугах и тарифах.
6. Предотвращение оттока клиентов в другие компании.

Исходные данные. Были отобраны только активные абоненты, которые регулярно пользовались услугами сотовой связи в течение последних нескольких месяцев. Данные находятся в файле `mobile.txt`.

Вопросы для самоконтроля

1. Для решения каких задач предназначены методы кластерного анализа?
2. В чем заключается отличие методов классификации с учителем и без учителя?
3. Перечислите наиболее известные меры сходства и различия объектов и классов.
4. Опишите процедуры классификации, чаще всего применяемые в кластерном анализе.

Лабораторная работа 5: Data Mining : классификация и регрессия. Статистические методы

Цель работы:

1. Изучить основы метода логистической регрессии.
2. Изучить основы метода деревьев решений.
3. Освоить технику практического применения указанных методов в Deductor на примере задачи кредитного скоринга.

Этапы выполнения лабораторной работы

1. Кредитный скоринг с помощью логистической регрессии.
2. Кредитный скоринг с помощью деревьев решений.
3. ROC-анализ.

Варианты заданий:

Используя предложенный набор кредитных историй, хранящихся в текстовом файле **potreb.txt** решить в Deductor следующие задачи.

1. Постройте скоринговую модель на основе логистической регрессии.
2. Определите оптимальный порог отсечения и порог, при котором чувствительность модели равна 90%.
3. Постройте несколько моделей деревьев решений при различных настройках алгоритма, а также отдельные деревья решений для заемщиков, состоящих и не состоящих в браке. Выберите модель, которая чаще отказывает в выдаче кредита.
4. Сравните качество полученных моделей и сформулируйте рекомендации по их выбору при различной кредитной политике.

Вопросы для самоконтроля

1. Что представляет собой скоринг?
2. Почему скоринг имеет важное значение в розничном кредитовании?
3. Какие алгоритмы используются при построении скоринговых моделей?
4. Как строится ROC-кривая?
5. Как рассчитывают показатели чувствительности и специфичности модели и в чем состоит их смысл?
6. Как можно определить оптимальный порог отсечения?
7. Какие преимущества имеет алгоритм C4.5?
8. Какие достоинства и недостатки имеют модели логистической регрессии и деревьев решений применительно к задаче кредитного скоринга?

Лабораторная работа 6: Data Mining : классификация и регрессия. Машинное обучение

Цель работы:

1. Изучить принципы функционирования искусственных нейронных сетей.
2. Освоить методы их построения на примере задачи повышения эффективности массовой рассылки клиентам.

Этапы выполнения лабораторной работы

1. Выбор структуры нейронной сети.
2. Выбор функции активации.
3. Выбор алгоритма обучения нейронной сети.
4. Тестирование нейронной сети.

Варианты заданий:

Торговая компания, осуществляющая продажу товаров, располагает информацией о своих клиентах и их покупках. Компания провела рекламную рассылку 13504 клиентам и получила отклик в 14,5% случаев.

1. Необходимо построить модели отклика и проанализировать результаты.
2. Предложить способы минимизации издержек на новые почтовые рассылки .
3. Поставленные задачи решить при помощи многослойного персептрона.

Вопросы для самоконтроля

1. Какие задачи можно решать при помощи многослойного персептрона?
2. Как формируется обучающая выборка для решения задачи классификации?
3. Как нормируются обучающие данные?
4. Сформулируйте эмпирические правила подбора количества скрытых слоев, количества нейронов, объема обучающей выборки, коэффициента обучения.
5. Укажите недостатки алгоритма обратного распространения ошибки.

4.2 Фонд оценочных средств для проведения промежуточной аттестации

Задания по бизнес-задачам

Вариант 1. Консолидация данных и аналитическая отчетность аптечной сети в Deductor

Вариант 2. Ассоциативные правила в стимулировании розничных продаж в Deductor

Вариант 3. Сегментация клиентов телекоммуникационной компании в Deductor

Вариант 4. Скоринговые модели для оценки кредитоспособности заемщиков в Deductor

Вариант 5. Прогнозирование продаж товаров в оптовой компании с помощью Deductor

Вариант 6. Повышение эффективности массовой рассылки клиентам с помощью Deductor

Примерный перечень вопросов к экзамену.

1. Принципы анализа данных. Структурированные данные. Подготовка данных к анализу.
2. Технологии KDD и Data Mining. Аналитические платформы. Введение в алгоритмы Data Mining.
3. Задача консолидации. Основные концепции хранилищ данных.
4. Многомерные, реляционные, гибридные, виртуальные хранилища данных.
5. Нечеткие срезы. Извлечение, очистка и преобразование данных в ETL.
6. Загрузка данных в хранилище из локальных источников. Обогащение данных.
7. Введение в трансформацию данных. Трансформация упорядоченных данных.
8. Группировка данных. Слияние данных.
9. Квантование. Нормализация и кодирование данных.
10. Введение в визуализацию данных. Визуализаторы общего назначения.
11. OLAP -анализ.
12. Визуализаторы для оценки качества моделей.
13. Визуализаторы, применяемые для интерпретации результатов анализа.
14. Технологии и методы оценки качества данных. Очистка и предобработка данных.
15. Фильтрация данных. Обработка дубликатов и противоречий.
16. Выявление аномальных значений. Восстановление пропущенных значений.

17. Введение в сокращение размерности. Сокращение числа признаков. Сокращение числа значений признаков и записей. Сэмплинг.
18. Ассоциативные правила. Значимость ассоциативных правил.
19. Поиск ассоциативных правил. Алгоритм Apriori. Генерация ассоциативных правил.
20. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.
21. Последовательные шаблоны. Поиск последовательных шаблонов. Алгоритм AprioriAll.
22. Введение в кластеризацию. Алгоритм кластеризации k-means. Нечеткая кластеризация.
23. Сети Кохонена. Обучение сети Кохонена. Карты Кохонена.
24. Методика построения карты. Выбор числа нейронов карты. Недостатки карт Кохонена.
25. Неопределенность в выборе критерия качества кластеризации. Выбор числа кластеров.
26. Введение в классификацию и регрессию. Простая линейная регрессия. Оценка соответствия простой линейной регрессии реальным данным. Простая регрессионная модель.
27. Модель множественной линейной регрессии. Регрессия с категориальными входными переменными.
28. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей.
29. Основы логистической регрессии. Интерпретация модели логистической регрессии.
30. Множественная логистическая регрессия. Простой байесовский классификатор.
31. Введение в деревья решений. Алгоритмы построения деревьев решений. Упрощение деревьев решений.
32. Введение в нейронные сети. Искусственный нейрон. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей.
33. Классификация с помощью нейронных сетей
34. Прогнозирование с помощью нейронных сетей.
35. Введение в ансамбли моделей. Виды ансамблей.
36. Бэггинг. Бустинг. Альтернативные методы построения ансамблей.
37. Оценка эффективности и сравнение моделей. Оценка ошибки модели.
38. Издержки ошибочной классификации. Lift- и Profit- кривые.
39. ROC- анализ.
40. Обучение в условиях несбалансированности классов.

Критерии выставления оценок.

Оценка «отлично»:

- систематизированные, глубокие и полные знания по всем разделам дисциплины, а также по основным вопросам, выходящим за пределы учебной программы;
- точное использование научной терминологии систематически грамотное и логически правильное изложение ответа на вопросы;
- безупречное владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке научных и практических задач;
- выраженная способность самостоятельно и творчески решать сложные проблемы и нестандартные ситуации;
- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной учебной программой по дисциплине;

- умение ориентироваться в теориях, концепциях и направлениях дисциплины и давать им критическую оценку, используя научные достижения других дисциплин;

- творческая самостоятельная работа на практических/семинарских/лабораторных занятиях, активное участие в групповых обсуждениях, высокий уровень культуры исполнения заданий;

- высокий уровень сформированности заявленных в рабочей программе компетенций.

Оценка «хорошо»:

- достаточно полные и систематизированные знания по дисциплине;

- умение ориентироваться в основном теориях, концепциях и направлениях дисциплины и давать им критическую оценку;

- использование научной терминологии, лингвистически и логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;

- владение инструментарием по дисциплине, умение его использовать в постановке и решении научных и профессиональных задач;

- усвоение основной и дополнительной литературы, рекомендованной учебной программой по дисциплине;

- самостоятельная работа на практических занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий;

- средний уровень сформированности заявленных в рабочей программе компетенций.

Оценка «удовлетворительно»:

- достаточный минимальный объем знаний по дисциплине;

- усвоение основной литературы, рекомендованной учебной программой;

- умение ориентироваться в основных теориях, концепциях и направлениях по дисциплине и давать им оценку;

- использование научной терминологии, стилистическое и логическое изложение ответа на вопросы, умение делать выводы без существенных ошибок;

- владение инструментарием учебной дисциплины, умение его использовать в решении типовых задач;

- умение под руководством преподавателя решать стандартные задачи;

- работа под руководством преподавателя на практических занятиях, допустимый уровень культуры исполнения заданий;

- достаточный минимальный уровень сформированности заявленных в рабочей программе компетенций.

Оценка «неудовлетворительно»:

- фрагментарные знания по дисциплине;

- отказ от ответа (выполнения письменной работы);

- знание отдельных источников, рекомендованных учебной программой по дисциплине;

- неумение использовать научную терминологию;

- наличие грубых ошибок;

- низкий уровень культуры исполнения заданий;

- низкий уровень сформированности заявленных в рабочей программе компетенций.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

- при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

- при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

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

5.1 Основная литература:

1. Петров, А.В. Моделирование процессов и систем [Электронный ресурс] : учеб. пособие — Электрон. дан. — Санкт-Петербург : Лань, 2015. — 288 с. — Режим доступа: <https://e.lanbook.com/book/68472>.
2. Ганичева, А.В. Математические модели и методы оценки событий, ситуаций и процессов [Электронный ресурс] : учеб. пособие / А.В. Ганичева. — Электрон. дан. — Санкт-Петербург : Лань, 2017. — 188 с. — Режим доступа: <https://e.lanbook.com/book/91891>.
3. Пегат, А. Нечеткое моделирование и управление [Электронный ресурс] — Электрон. дан. — Москва : Издательство "Лаборатория знаний", 2015. — 801 с. — Режим доступа: <https://e.lanbook.com/book/84106>.
4. Буховец А.Г. Алгоритмы вычислительной статистики в системе R/ А.Г. Буховец, П.В. Москалев. СПб.: Лань, 2015. 160 с. [Электронный ресурс]. - Режим доступа: <https://e.lanbook.com/book/68459>.
5. Гаврилова Т.А. Инженерия знаний. Модели и методы: / Т.А. Гаврилова, Д.В. Кудрявцев, Д.И. Муромцев. СПб.: Лань, 2016. 324 с. [Электронный ресурс]. - Режим доступа: <https://e.lanbook.com/book/81565>.
6. Кохонен Т. Самоорганизующиеся карты. М.: Изд-во "Лаборатория знаний", 2017. 660 с. [Электронный ресурс]. - Режим доступа: <https://e.lanbook.com/book/94143>.
7. Жданов А.А. Автономный искусственный интеллект. М.: Изд-во "Лаборатория знаний", 2015. 362 с. [Электронный ресурс]. - Режим доступа: <https://e.lanbook.com/book/70761>.

5.2 Дополнительная литература:

1. Юдович В.И. Математические модели естественных наук: учебное пособие. СПб: Лань, 2011. 336 с. [Электронный ресурс]. - Режим доступа: <https://e.lanbook.com/book/689>.
2. Колокольцов, В.Н. Математическое моделирование многоагентных систем конкуренции и кооперации (Теория игр для всех) [Электронный ресурс] : учеб. пособие / В.Н. Колокольцов, О.А. Малафеев. — Электрон. дан. — Санкт-Петербург : Лань, 2012. — 624 с. — Режим доступа: <https://e.lanbook.com/book/3551>.

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

1. Википедия, свободная энциклопедия. [Электронный ресурс]. – Wikipedia <http://ru.wikipedia.org>
2. Проектирование систем управления \ Fuzzy Logic Toolbox С.Д.Штовба "Введение в теорию нечетких множеств и нечеткую логику" http://matlab.exponenta.ru/fuzzylogic/book1/1_7.php
3. Электронно-библиотечная система Издательство «Лань». <http://e.lanbook.com>

7. Методические рекомендации по организации изучения дисциплины

Контрольная работа представляет собой самостоятельную реферативную работу студентов. Каждый студент выполняет работу по одной теме.

Для написания реферата необходимо подобрать литературу. Общее количество литературных источников, включая тексты из Интернета, (публикации в журналах), должно составлять не менее 10 наименований. Учебники, как правило, в литературные источники не входят.

Рефераты выполняют на листах формата А4. Страницы текста, рисунки, формулы нумеруют, рисунки снабжают порисуночными надписями. Текст следует печатать шрифтом №14 с интервалом между строками в 1,5 интервала, без недопустимых сокращений. В конце реферата должны быть сделаны выводы.

В конце работы приводят список использованных источников.

Реферат должен быть подписан студентом с указанием даты ее оформления.

Работы, выполненные без соблюдения перечисленных требований, возвращаются на доработку.

Выполненная студентом работа определяется на проверку преподавателю в установленные сроки. Если у преподавателя есть замечания, работа возвращается и после исправлений либо вновь отправляется на проверку, если исправления существенные, либо предъявляется на зачете, где происходит ее защита.

Для разъяснения непонятных вопросов лектором и ассистентом еженедельно проводятся консультации, о времени которых группы извещаются заранее.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

Творческие задания (проекты), способствующие формированию компетенций базовой части ООП

Контрольная работа представляет собой самостоятельную реферативную работу студентов. Каждый студент выполняет работу по одной теме.

Для написания реферата необходимо подобрать литературу. Общее количество литературных источников, включая тексты из Интернета, (публикации в журналах), должно составлять не менее 10 наименований. Учебники, как правило, в литературные источники не входят.

Рефераты выполняют на листах формата А4. Страницы текста, рисунки, формулы нумеруют, рисунки снабжают порисуночными надписями. Текст следует печатать шрифтом №14 с интервалом между строками в 1,5 интервала, без недопустимых сокращений. В конце реферата должны быть сделаны выводы.

В конце работы приводят список использованных источников.

Реферат должен быть подписан студентом с указанием даты ее оформления.

Работы, выполненные без соблюдения перечисленных требований, возвращаются на доработку.

Выполненная студентом работа определяется на проверку преподавателю в установленные сроки. Если у преподавателя есть замечания, работа возвращается и после исправлений либо вновь отправляется на проверку, если исправления существенные, либо предъявляется на зачете, где происходит ее защита.

Творческие задания (проекты), способствующие формированию компетенций базовой части ООП

Интеллектуальные технологии сегодня рассматриваются как одно из главных новых направлений при решении задач маркетинга.

Основные направления интеллектуальных технологий для анализа реальных ситуаций:

- Ассоциативные правила;
- Классификация;
- Кластеризация.

Проведите анализ по одной из выбранных вами тематик (не менее 10 слайдов и 20 листов текста). Возможно использование звукового сопровождения, анимации (аудио-, и видеоматериала).

На первой странице слайда обязательно укажите Ф.И.О. автора, курс. Оценивается работа по следующим критериям:

- полнота представленного материала;
- оформление;
- представление и защита.

Темы презентаций и докладов

1. Консолидация данных и аналитическая отчетность аптечной сети в Deductor
2. Ассоциативные правила в стимулировании розничных продаж в Deductor
3. Сегментация клиентов телекоммуникационной компании в Deductor
4. Скоринговые модели для оценки кредитоспособности заемщиков в Deductor
5. Прогнозирование продаж товаров в оптовой компании с помощью Deductor
6. Повышение эффективности массовой рассылки клиентам с помощью Deductor

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

8.1 Перечень необходимого программного обеспечения

1. Операционная система MS Windows (разделы 1, 2, 3 дисциплины).
2. Интегрированное офисное приложение MS Office (разделы 1, 2, 3 дисциплины).

3. Программное обеспечение для организации управляемого коллективного и безопасного доступа в Интернет (разделы 2, 3 дисциплины).
4. Statistica Neural Network (раздел 2 дисциплины).
5. Matlab Neural Networks Toolbox, Fuzzy Logic Toolbox, ANFIS (разделы 1 и 3 дисциплины).

8.2 Перечень информационных справочных систем

Обучающимся должен быть обеспечен доступ к современным профессиональным базам данных, профессиональным справочным и поисковым системам:

Электронная библиотечная система (ЭБС) «Университетская библиотека ONLINE» (<http://www.biblioclub.ru>)

Электронная библиотечная система издательства «Лань» (<https://e.lanbook.com>)

Электронная библиотечная система «Юрайт» (<http://www.biblio-online.ru>)

Электронная библиотечная система «ZNANIUM.COM» (<https://znanium.com>)

Электронно-библиотечная система (ЭБС) BOOK.ru (<http://www.book.ru>)

Электронная библиотечная система eLIBRARY.RU (<http://www.elibrary.ru>)

Справочно-правовая система «Консультант Плюс» (<http://www.consultant.ru>)

Справочно-правовая система «Гарант» (<http://www.garant.ru>)

«Консультант студента» (www.studentlibrary.ru)

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

№	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащенность
1.	Занятия лекционного типа	Аудитории, оснащенные презентационной техникой (проектор, экран, ноутбук) и соответствующим программным обеспечением (ауд. 129, 131, 133, А305, А307).
2.	Групповые и индивидуальные консультации	Аудитории для семинарских занятий, групповых и индивидуальных консультаций, укомплектованные необходимой мебелью (доска, столы, стулья): (ауд. 129, 131).
3.	Лабораторные занятия	Учебная мебель, компьютерная техника (компьютерные классы: 101, 105/1, 107(2), 107(3), 107(5)); учебная мебель, компьютерная техника с возможностью подключения к сети «Интернет» и доступом в электронную информационно-образовательную среду университета (компьютерные классы: 106, 106а, А301).
4.	Текущий контроль, промежуточная аттестация	Аудитории для семинарских занятий, текущего контроля и промежуточной аттестации, укомплектованные необходимой мебелью (доска, столы, стулья) (ауд. 129, 131, 133, А305, А307, 147, 148, 149, 150, 100С, А301б, А512); аудитории с компьютерами с лицензионным программным обеспечением и выходом в «Интернет» (106, 106а, А301).
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета (аудитории: 102а, читальный зал).