



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования

КГЭУ

**«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГЭУ»)**

АКТУАЛИЗИРОВАНО

Решением Ученого совета ИЦТЭ КГЭУ

Протокол №7 от 19.03.2024

«УТВЕРЖДАЮ»

Директор института Цифровых технологий и
экономики

Торкунова Ю.В.

«26»_октября 2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Методы и модели анализа данных

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность(и) (профиль(и)) 09.03.01 Автоматизированное управление бизнес-процессами и финансами

Квалификация

бакалавр

г. Казань, 2020

Рабочая программа дисциплины разработана в соответствии с ФГОС ВО - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника (приказ Минобрнауки России от 19.09.2017 г. № 929)

Программу разработал(и):

доцент, к.т.н. _____ Гадильшина В.Р.

Рабочая программа рассмотрена и одобрена на заседании кафедры Инженерная кибернетика, протокол №11 от 26.10.2020

Зав. кафедрой _____ Смирнов Ю.Н.

Программа рассмотрена и одобрена на заседании выпускающей кафедры Инженерная кибернетика, протокол № 11 от 26.10.2020

Зав. кафедрой _____ Смирнов Ю.Н.

Программа одобрена на заседании методического совета института Цифровых технологий и экономики, протокол № 2 от 26.10.2020

Зам. директора института Цифровых технологий и экономики

_____ Косулин В.В.

Программа принята решением Ученого совета института Цифровых технологий и экономики
протокол №2 от 26.10.2020

Согласовано:

Руководитель ОПОП _____ Смирнов Ю.Н.

1. Цель, задачи и планируемые результаты обучения по дисциплине

Целью освоения дисциплины является формирование у обучающихся системного представления о технологиях многомерного анализа данных, интеллектуального анализа данных, развитие навыков практического применения моделей и методов анализа данных в различных экономических деловых ситуациях и бизнес-процессах.

- формирование представлений о способах реализации и инструментах многомерного анализа данных;
- изучение областей применения, методов и средств Data Mining;
- формирование теоретических знаний и практических навыков анализа данных при решении типовых экономических задач;
- изучение структуры и функций хранилищ данных.

Компетенции, формируемые у обучающихся, запланированные результаты обучения по дисциплине, соотнесенные с индикаторами достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине (знать, уметь, владеть)
ПК-1 Способен провести реинжиниринг бизнес-процессов	ПК-1.1 Анализирует бизнес-процессы заказчика	<i>Знать:</i> основные понятия теории анализа данных бизнес-процессов (31) методы сбора, обработки и контроля качества данных для анализа бизнес-процессов (32) <i>Уметь:</i> применять методы сбора и систематизации данных о бизнес-процессах (У1) анализировать полученные данные о бизнес-процессах (У2) проводить интерпретацию результатов анализа данных о бизнес-процессах (У3) <i>Владеть:</i> навыками сбора исходных данных для описания бизнес-процессов (В1) методикой анализа исходных данных для описания бизнес-процессов (В2)

ПК-1 Способен провести реинжиниринг бизнес-процессов	ПК-1.2 Применяет инструменты и методы реинжиниринга бизнес-процессов	<p><i>Знать:</i> основные методы и инструменты многомерного интеллектуального анализа данных бизнес-процессов (З1)</p> <p><i>Уметь:</i> осуществлять выбор методов анализа данных бизнес-процессов (У1) использовать инструменты для практической реализации методов реинжиниринга бизнес-процессов (У2)</p> <p><i>Владеть:</i> методами прикладного многомерного анализа данных бизнес-процессов (В1) методами реинжиниринга и управления бизнес-процессами с применением современных инструментальных средств (В2)</p>
--	--	--

2. Место дисциплины в структуре ОПОП

Дисциплина Методы и модели анализа данных относится к части, формируемой участниками образовательных отношений учебного плана по направлению подготовки 09.03.01 Информатика и вычислительная техника.

Код компетенции	Предшествующие дисциплины (модули), практики, НИР, др.	Последующие дисциплины (модули), практики, НИР, др.
УК-1		Выполнение и защита выпускной квалификационной работы
УК-2		Выполнение и защита выпускной квалификационной работы
УК-3		Выполнение и защита выпускной квалификационной работы
УК-4		Выполнение и защита выпускной квалификационной работы
УК-5		Выполнение и защита выпускной квалификационной работы
УК-6		Выполнение и защита выпускной квалификационной работы
УК-7		Выполнение и защита выпускной квалификационной работы
УК-8		Выполнение и защита выпускной квалификационной работы
ОПК-1	Теория вероятностей и математическая статистика	
ОПК-1		Выполнение и защита выпускной квалификационной работы
ОПК-2		Выполнение и защита выпускной квалификационной работы
ОПК-2	Информационные технологии	
ОПК-3		Выполнение и защита выпускной квалификационной работы
ОПК-4		Выполнение и защита выпускной квалификационной работы

ОПК-5		Выполнение и защита выпускной квалификационной работы
ОПК-6		Выполнение и защита выпускной квалификационной работы
ОПК-7		Выполнение и защита выпускной квалификационной работы
ОПК-8		Выполнение и защита выпускной квалификационной работы
ОПК-9		Выполнение и защита выпускной квалификационной работы
ПК-1		Выполнение и защита выпускной квалификационной работы
ПК-2		Информационно-математическое моделирование бизнес-процессов Выполнение и защита выпускной квалификационной работы
ПК-3		Выполнение и защита выпускной квалификационной работы
ПК-4		Выполнение и защита выпускной квалификационной работы
ПК-5		Выполнение и защита выпускной квалификационной работы
ПК-6		Выполнение и защита выпускной квалификационной работы
ПК-7		Выполнение и защита выпускной квалификационной работы

Для освоения дисциплины обучающийся должен:

знать: основные понятия математического моделирования;

уметь: выбирать математические модели и методы описания бизнес-процессов;

владеть: навыками работы с компьютером как средством управления информацией, математическими пакетами программ.

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) (ЗЕ), всего 108 часов, из которых 42 часов составляет контактная работа обучающегося с преподавателем (занятия лекционного типа 16 час., занятия семинарского типа (практические, семинарские занятия, лабораторные работы и т.п.) 24 час., групповые и индивидуальные консультации 0 час., прием экзамена (КПА), зачета с оценкой - 1 час., самостоятельная работа обучающегося 66 час, контроль самостоятельной работы (КСР) - 2 час.

Вид учебной работы	Всего часов	Семестр
		7
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	108	108

КОНТАКТНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ С ПРЕПОДАВАТЕЛЕМ, в том числе:	42	42
Лекционные занятия (Лек)	16	16
Практические занятия (Пр)	24	24
Контроль самостоятельной работы и иная контактная работа (КСР)*	2	2
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ (СРС):	66	66
Подготовка к промежуточной аттестации в форме: (зачет)		
ФОРМА ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ	За	За

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	Семестр	Распределение трудоемкости (в часах) по видам учебной работы, включая СРС									Формируемые результаты обучения (знания, умения, навыки)	Литература	Формы текущего контроля успеваемости	Формы промежуточной аттестации	Максимальное количество баллов по балльно - рейтинговой системе	
		Занятия лекционного типа	Занятия практического / семинарского типа	Лабораторные работы	Групповые консультации	Самостоятельная работа студента, в т.ч.	Контроль самостоятельной работы (КСР)	подготовка к промежуточной аттестации	Сдача зачета / экзамена	Итого						
Раздел 1. Хранилище данных																
1. Хранилище данных	7	4	4			14	0,5			22,5						
Раздел 2. OLAP-системы																
2. OLAP- системы	7	4	6			14	0,5			24,5						
Раздел 3. Интеллектуальный анализ данных (Data Mining)																
3. Интеллектуальный анализ данных (Data Mining)	7	2				14	0,5			16,5						
Раздел 4. Задачи и методы Data Mining																
4. Задачи и методы Data Mining	7	6	14			24	0,5			44,5						
ИТОГО		16	24			66	2			108						

3.3. Тематический план лекционных занятий

Номер раздела дисциплины	Темы лекционных занятий	Трудоемкость, час.
--------------------------	-------------------------	--------------------

1	Концепция хранилищ данных, их виды и свойства. Проблемы создания и организации хранилищ данных. Виды данных, предварительная обработка данных, их очистка, интеграция и преобразование.	4
2	Многомерная модель данных. Определение OLAP-систем, их архитектура. Многомерный анализ данных.	4
3	Классификация задач Data Mining. Области применения Data Mining. Процесс Data Mining.	2
4	Методы классификации, прогнозирования: регрессионные методы, метод деревьев решений, опорных векторов, метод Naïve Bayes, нейронные сети. Поиск ассоциативных правил. Методы кластерного анализа.	6
Всего		16

3.4. Тематический план практических занятий

Номер раздела дисциплины	Темы практических занятий	Трудоемкость, час.
1	Типы наборов данных. Шкалы измерений. Форматы хранения данных	4
2	Базы данных, основные положения. Требования по данным для OLTR и OLAP систем, их сравнение	6
3	Методы регрессионного анализа	2
4	Методы построения правил классификации, деревьев решений	2
5	Метод опорных векторов, "ближайших соседей", классификатор Байеса	2
6	Нейронные сети: обучение, модели	2
7	Методы поиска ассоциативных правил	2
8	Методы кластерного анализа	2
9	Анализ и обобщение данных в Excel	2
Всего		24

3.5. Тематический план лабораторных работ

Данный вид работы не предусмотрен учебным планом

3.6. Самостоятельная работа студента

Номер раздела дисциплины	Вид СРС	Содержание СРС	Трудоемкость, час.
1			14
2			14
3			14
4			24
Всего			66

4. Образовательные технологии

При проведении учебных занятий используются традиционные образовательные технологии, такие как лекций, практические занятия, самостоятельная работа, консультации. Кроме них используются: работа с пакетами прикладных программ, работа в интерактивной образовательной среде LMS Moodle, компьютерное тестирование, индивидуальные задания и контроль их поэтапного выполнения.

5. Оценивание результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля успеваемости, проводимого по балльно-рейтинговой системе (БРС), и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в течение семестра, включает: защиты отчетов по практическим работам, компьютерное тестирование.

Результат (зачтено/не зачтено) промежуточной аттестации в форме зачета определяется по совокупности результатов текущего контроля успеваемости по дисциплине.

Обобщенные критерии и шкала оценивания уровня сформированности компетенции (индикатора достижения компетенции) по итогам освоения дисциплины:

Планируемые результаты обучения	Обобщенные критерии и шкала оценивания результатов обучения			
	неудовлетворительно	удовлетворительно	хорошо	отлично
	не зачтено	зачтено		
Полнота знаний	Уровень знаний ниже минимальных требований, имеют место грубые ошибки	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок
Наличие умений	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме
Наличие навыков (владение опытом)	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов

Характеристика сформированности компетенции (индикатора достижения компетенции)	Компетенция в полной мере сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач	Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству практических задач	Сформированность компетенции в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных практических (профессиональных) задач	Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных практических (профессиональных) задач
Уровень сформированности компетенции (индикатора достижения компетенции)	Низкий	Ниже среднего	Средний	Высокий

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора достижения компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции (индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено			не зачтено
ПК-1	ПК-1.1	Знать				
		основные понятия теории анализа данных бизнес-процессов (31)	В полном объеме знает основные понятия теории анализа данных бизнес-процессов	Достаточно полно знает основные понятия теории анализа данных бизнес-процессов, допускает мелкие неточности	Уровень знаний по теме минимальный	Уровень знаний ниже минимальных требований

		методы сбора, обработки и контроля качества данных для анализа бизнес-процессов (У2)	В полном объеме знает методы сбора, обработки и контроля качества данных для анализа бизнес-процессов	Достаточно полно знает методы сбора, обработки и контроля качества данных для анализа бизнес-процессов, допускает мелкие неточности	Уровень знаний по теме минимальный	Уровень знаний ниже минимальных требований
Уметь						
		применять методы сбора и систематизации данных о бизнес-процессах (У1)	Уверенно показывает умения применять методы сбора и систематизации данных о бизнес-процессах	Показывает все основные умения применять методы сбора и систематизации данных о бизнес-процессах, допускает мелкие недочеты	Допускает много недочетов при применении методов сбора и систематизации данных о бизнес-процессах	Не умеет применять методы сбора и систематизации данных о бизнес-процессах
		анализировать полученные данные о бизнес-процессах (У2)	Уверенно показывает умения анализировать полученные данные о бизнес-процессах	Показывает все основные умения анализировать полученные данные о бизнес-процессах, допускает мелкие недочеты	Допускает много недочетов при анализе полученных данных о бизнес-процессах	Не умеет анализировать полученные данные о бизнес-процессах
		проводить интерпретацию результатов анализа данных о бизнес-процессах (У3)	Уверенно показывает умения проводить интерпретацию результатов анализа данных о бизнес-процессах	Показывает все основные умения проводить интерпретацию результатов анализа данных о бизнес-процессах, допускает мелкие недочеты	Допускает много недочетов при проведении интерпретации результатов анализа данных о бизнес-процессах	Не умеет проводить интерпретацию результатов анализа данных о бизнес-процессах
Владеть						

		навыками сбора исходных данных для описания бизнес-процессов (B1)	В полном объеме владеет навыками сбора исходных данных для описания бизнес-процессов	Демонстрирует базовые навыки сбора исходных данных для описания бизнес-процессов, допускает некоторые недочеты	Имеет минимальные навыки сбора исходных данных для описания бизнес-процессов, допускает некоторые недочеты	Отсутствуют базовые навыки сбора исходных данных для описания бизнес-процессов, допускает грубые ошибки
		методикой анализа исходных данных для описания бизнес-процессов (B2)	В полном объеме владеет методикой анализа исходных данных для описания бизнес-процессов	Демонстрирует владение методикой анализа исходных данных для описания бизнес-процессов, допускает некоторые недочеты	Владеет на минимальном уровне методикой анализа исходных данных для описания бизнес-процессов, допускает некоторые недочеты	Не владеет методикой анализа исходных данных для описания бизнес-процессов, имеют место грубые ошибки
ПК-1.2	Знать					
		основные методы и инструменты многомерного интеллектуального анализа данных бизнес-процессов (31)	В полном объеме знает основные методы и инструменты многомерного интеллектуального анализа данных бизнес-процессов	Достаточно полно знает основные методы и инструменты многомерного интеллектуального анализа данных бизнес-процессов, допускает мелкие неточности	Уровень знаний по теме минимальный	Уровень знаний ниже минимальных требований
	Уметь					
		осуществлять выбор методов анализа данных бизнес-процессов (У1)	Уверенно показывает умения осуществлять выбор методов анализа данных бизнес-процессов	Показывает все основные умения осуществлять выбор методов анализа данных бизнес-процессов, допускает мелкие недочеты	Допускает много недочетов при выборе методов анализа данных бизнес-процессов	Не умеет осуществлять выбор методов анализа данных бизнес-процессов

		использовать инструменты для практической реализации методов реинжиниринга бизнес-процессов (У2)				
	Владеть					
	методами прикладного многомерного анализа данных бизнес-процессов (В1)	В полном объеме владеет методами прикладного многомерного анализа данных бизнес-процессов	Демонстрирует владение методами прикладного многомерного анализа данных бизнес-процессов, допускает некоторые недочеты	Владеет на минимальном уровне методами прикладного многомерного анализа данных бизнес-процессов, допускает некоторые недочеты	Не владеет методами прикладного многомерного анализа данных бизнес-процессов, имеют место грубые ошибки	
	методами реинжиниринга и управления бизнес-процессами с применением современных инструментальных средств (В2)	В полном объеме владеет методами реинжиниринга и управления бизнес-процессами с применением современных инструментальных средств	Демонстрирует владение методами реинжиниринга и управления бизнес-процессами с применением современных инструментальных средств, допускает некоторые недочеты	Владеет на минимальном уровне методами реинжиниринга и управления бизнес-процессами с применением современных инструментальных средств, допускает некоторые недочеты	Не владеет методами реинжиниринга и управления бизнес-процессами с применением современных инструментальных средств, имеют место грубые ошибки	

Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины. Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре-разработчике в бумажном и электронном виде.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

Основная литература

№ п/п	Автор(ы)	Наименование	Вид издания (учебник, учебное пособие, др.)	Место издания, издательство	Год издания	Адрес электронного ресурса	Кол-во экземпляров в библиотеке КГЭУ
-------	----------	--------------	---	-----------------------------	-------------	----------------------------	--------------------------------------

1	Макшанов А. В., Журавлев А. Е.	Технологии интеллектуального анализа данных	учебное пособие	СПб.: Лань	2018	https://e.lanbook.com/book/109617	1
2	Гулаков В. К., Трубаков А. О., Трубаков Е. О.	Структуры и алгоритмы обработки многомерных данных	монография	СПб.: Лань	2018	https://e.lanbook.com/book/107305	1

Дополнительная литература

№ п/п	Автор(ы)	Наименование	Вид издания (учебник, учебное пособие, др.)	Место издания, издательство	Год издания	Адрес электронного ресурса	Кол-во экземпляров в библиотеке КГЭУ
1	Ростовцев В. С.	Искусственные нейронные сети	учебник	СПб.: Лань	2019	https://e.lanbook.com/book/122180	1
2	Москвитин А. А.	Данные, информация, знания: методология, теория, технологии	монография	СПб.: Лань	2019	https://e.lanbook.com/book/113937	1
3	Чураков Е. П.	Введение в многомерные статистические методы	учебное пособие	СПб.: Лань	2016	https://e.lanbook.com/book/87598	1

6.2. Информационное обеспечение

6.2.1. Электронные и интернет-ресурсы

№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	Электронно-библиотечная система «Лань»	https://e.lanbook.com/
2	Электронно-библиотечная система «ibooks.ru»	https://ibooks.ru/
3	Электронно-библиотечная система «book.ru»	https://www.book.ru/
4	Энциклопедии, словари, справочники	http://www.rubricon.com
5	Портал "Открытое образование"	http://npoed.ru
6	Единое окно доступа к образовательным ресурсам	http://window.edu.ru

6.2.2. Профессиональные базы данных

№ п/п		Адрес	Режим доступа
1	Официальный интернет-портал правовой	http://pravo.gov.ru	http://pravo.gov.ru
2	Справочная правовая система	http://consultant.ru	http://consultant.ru
3	Справочно-правовая система	http://garant.ru	http://garant.ru

6.2.3. Информационно-справочные системы

№ п/п		Адрес	Режим доступа
1	Научная электронная библиотека	http://elibrary.ru	http://elibrary.ru
2	Российская государственная библиотека	http://www.rsl.ru	http://www.rsl.ru
3	Международная реферативная база данных	http://www.zbmath.org	http://www.zbmath.org
4	Международная реферативная база данных	http://link.springer.com	http://
5	Образовательный портал	http://www.uceba.com	http://www.uceba.com

6.2.4. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Способ распространения (лицензионное/свободно)	Реквизиты подтверждающих документов
1	Visual Studio Professional 2013 Russian OLP NL AcademicEdition	Программный продукт содержащий в себе инструменты и службы для разработки	ЗАО "СофтЛайнТрейд" №2014.1610 от 05.11.2014 Неискл. право. Бессрочно
2	Office Professional Plus 2007 Windows32 Russian DiskKit MVL CD	Пакет программных продуктов содержащий в себе необходимые офисные программы	ЗАО "СофтЛайнТрейд" №225/10 от 28.01.2010 Неискл. право. Бессрочно
3	Windows 7 Профессиональная (SevenPro_Check)	Пользовательская операционная система	"ЗАО ""ТаксНет-Сервис"" №ПО-ЛИЦ 0000/2014 от 27.05.2014 Неискл. право. Бессрочно
4	Браузер Chrome	Система поиска информации в сети интернет	Свободная лицензия Неискл. право. Бессрочно
5	LMS Moodle	ПО для эффективного онлайн-взаимодействия преподавателя и студента	Свободная лицензия Неискл. право. Бессрочно
6	MATLAB Academic new Product From 10 to 24 Group Licenses (per License)	Пакет прикладных программ для решения задач технических вычислений.	ЗАО "СофтЛайнТрейд" №2013.39442 Неискл. право. Бессрочно

7. Материально-техническое обеспечение дисциплины

№ п/п	Вид учебной работы	Наименование специальных помещений и помещений для СРС	Оснащенность специальных помещений и помещений для СРС
1	Самостоятельная работа обучающегося	Компьютерный класс с выходом в Интернет В-600а	моноблок (30 шт.), система видеонаблюдения (6 видеокамер), проектор, экран
		Читальный зал библиотеки	проектор, переносной экран, тонкие клиенты (13 шт.), компьютеры (5 шт.)
2	Лекционные занятия	Учебная аудитория для проведения занятий лекционного типа	доска интерактивная, моноблок (16 шт.)
3	Практические занятия	Учебная аудитория для проведения практических занятий	интерактивная доска, моноблок (25 шт.)

8. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей

психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета www/kgeu.ru. Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);
- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);
- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;
- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;
- обеспечения возможности для обучающегося получить адресную консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;
- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;
- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;
- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;
- обеспечивается необходимый уровень освещенности помещений;
- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ, инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на

компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

Лист регистрации изменений

Дополнения и изменения в рабочей программе дисциплины на 20__ /20__
учебный год

В программу вносятся следующие изменения:

1. _____

2. _____

3. _____

*Указываются номера страниц, на которых
внесены изменения,
и кратко дается характеристика этих
изменений*

Программа одобрена на заседании кафедры –разработчика «__» _____ 20__ г.,
протокол № _____

Зав. кафедрой _____ Смирнов Ю.Н.

Программа одобрена методическим советом института _____
«__» _____ 20__ г., протокол № _____

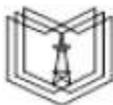
Зам. директора по УМР _____ / _____ /

Подпись, дата

Согласовано:

Руководитель ОПОП _____ / _____ /

Подпись, дата



КГЭУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования

«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГЭУ»)

«УТВЕРЖДАЮ»

Директор института Цифровых технологий и
экономики

_____ Торкунова Ю.В.

«__» _____ 2020 г.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ по дисциплине

Методы и модели анализа данных

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность (профиль) 09.03.01 Автоматизированное управление бизнес-процессами
и финансами

Квалификация

бакалавр

г. Казань, 2020

Оценочные материалы по дисциплине «Методы и модели анализа данных»- комплект контрольно-измерительных материалов, предназначенных для оценивания результатов обучения на соответствие индикаторам достижения компетенции(й):

ПК-1 Способен провести реинжиниринг бизнес-процессов

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля успеваемости, проводимого по балльно-рейтинговой системе (БРС), и промежуточной аттестации.

Текущий контроль успеваемости обеспечивает оценивание процесса обучения по дисциплине. При текущем контроле успеваемости используются следующие оценочные средства: практическое задание, тестирование, устный опрос.

Промежуточная аттестация имеет целью определить уровень достижения запланированных результатов обучения по дисциплине за 7 семестр. Форма промежуточной аттестации зачёт.

Оценочные материалы включают задания для проведения текущего контроля успеваемости, разработанные в соответствии с рабочей программой дисциплины.

1. Технологическая карта

Семестр 7

Номер раздела/ темы дисциплины	Вид СРС	Наименование оценочного средства	Код индикатора достижения компетенций	Уровень освоения дисциплины, баллы				
				неудов-но	удов-но	хорошо	отлично	
				не зачтено	зачтено			
				низкий	ниже среднего	средний	высокий	
Текущий контроль успеваемости								
1	Изучение теоретического материала, подготовка к устному опросу, тестированию	УО, Т	ПК-1.1 ПК-1.2	менее 11	11 - 15	15 - 17	17 - 20	
2	Изучение теоретического материала, подготовка к устному опросу, тестированию	УО, Т	ПК-1.1 ПК-1.2	менее 11	11 - 15	15 - 17	17 - 20	
3	Изучение теоретического материала, подготовка к устному опросу, тестированию	УО, Т	ПК-1.1 ПК-1.2	менее 11	11 - 15	15 - 17	17 - 20	
4	Изучение теоретического материала, подготовка отчетов по	ПЗ	ПК-1.1 ПК-1.2	менее 22	22 - 24	25 - 33	34 - 40	

	практическим занятиям						
Итого баллов				0-54	55-69	70-84	85-100

2. Перечень оценочных средств

Краткая характеристика оценочных средств, используемых при текущем контроле успеваемости и промежуточной аттестации обучающегося по дисциплине:

Наименование оценочного средства	Краткая характеристика оценочного средства	Оценочные материалы
Практическое задание (ПЗ)	Средство оценки умения применять полученные теоретические знания для решения практических задач; содержит четкий алгоритм действий	Перечень индивидуальных заданий
Тестирование (Т)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося	Тестовые задания на ресурсах LMS Moodle
Устный опрос (УО)	Средство контроля, рассчитанное на выяснение объема знаний обучающегося по определенному разделу	Вопросы по разделу дисциплины

3. Оценочные материалы текущего контроля успеваемости обучающихся

Наименование оценочного средства	Устный опрос по разделу 1 «Хранилище данных»
Представление и содержание оценочных материалов	Оценочный материал содержит 15 вопросов по разделу <i>Примеры вопросов по разделу дисциплины</i> 1. Причины возникновения хранилищ данных. 2. Концепция хранилищ данных. 3. Показатели и документы. 4. Виды данных: метаданные. 5. Измерения и классы. 6. Интегрированное хранилище данных.
Критерии оценки и шкала оценивания в баллах	При оценке устного опроса учитываются следующие критерии: 1. Знание материала 2. Последовательность изложения 3. Владение речью и терминологией 4. Применение конкретных примеров Шкала оценивания: <i>Высокий уровень:</i> - содержание материала раскрыто в полном объеме, предусмотренном программой дисциплины. содержание материала раскрыто последовательно, достаточно хорошо продумано, материал изложен грамотным языком, с точным использованием терминологии, показано умение иллюстрировать материал конкретными примерами – 10 баллов; <i>Средний уровень:</i> - содержание материала раскрыто достаточно полно, показано общее понимание вопроса, последовательность изложения материала достаточно продумана, приведение примеров с незначительными ошибками – 7 баллов; <i>Ниже среднего уровень:</i> - содержание материала раскрыто неполно, показано общее понимание вопроса, достаточное для дальнейшего изучения программного материала, последовательность изложения материала недостаточно продумана, приведение примеров вызывает затруднение – 4 балла;

	<p><i>Низкий уровень:</i></p> <p>- не раскрыто основное содержание учебного материала, путаница в изложении материала, допущены ошибки в определении понятий, неумение приводить примеры при объяснении материала – 0 баллов. Максимальное количество баллов - 10</p>
Наименование оценочного средства	Тестирование по разделу 1 «Хранилище данных»
Представление и содержание оценочных материалов	<p>Тестирование проводится в системе LMS Moodle. Содержание тестов соответствует тематике пройденного материала. База вопросов более 100, которая регулярно обновляется и случайным образом формируются в тест. Тест содержит 20 вопросов с заданиями разных типов.</p> <p><i>Примеры тестовых заданий</i></p> <p>1. Классификации метаданных по их функциональному назначению не включает следующий класс:</p> <p>А. предметно-ориентированные метаданные В. уникальные метаданные С. структурные метаданные</p> <p>2. Укажите правильное соответствие между этапами бизнес - модели создания хранилища данных и временными затратами, которые имеют место быть на каждом этапе:</p> <p>Этап:</p> <p>1)формулирование требований 2)создание вычислительной среды для хранилища данных 3)моделирование данных 4)определение процедур извлечения, преобразования и загрузки данных 5)проектирование аналитических отчетов 6)разработка приложений хранилища данных 7)настройка производительности 8)проверка качества</p> <p>Временные затраты:</p> <p>А) от двух недель до полутора месяцев В) от двух недель до двух месяцев С) от одной недели до полутора месяцев D) от двух недель до двух месяцев Е) от одной недели до месяца F) зависит от числа разрабатываемых отчетов G) от одной до четырех недель H) не должно превышать одну-две недели</p> <p>3. Многомерные "информационные кубы" ...</p> <p>А инструменты OLAP анализа В собирают и хранят всю информацию о деятельности предприятия С совокупность таблиц базы данных</p> <p>4. Укажите существующие типы иерархий в измерениях:</p> <p>А. сбалансированные В. несбалансированные С. неровные D. ровные</p> <p>5. Подсчитываемый показатель, возвращающий процент дохода, не может быть агрегирован из значений процентов своих дочерних ячеек других размерностей. Как называется такой показатель?</p> <p>А аддитивный В полуаддитивный</p>

	С. неаддитивный
Критерии оценки и шкала оценивания в баллах	<p>Результат тестирования автоматически оценивает программа LMS Moodle в баллах по предварительной настройке. Настройка теста позволяет оценивать результат в автоматическом режиме. Знания обучающегося в результате промежуточной аттестации в формате модульного тестирования оцениваются в зависимости от количества правильных ответов следующим образом:</p> <p>10 баллов – 100% правильного ответа 7 баллов – от 70 % до 84 % правильных ответов 4 балла – от 50% до 69% правильных ответов 0 баллов – меньше 50% правильных ответов</p> <p>Максимальное количество баллов - 10</p>
Наименование оценочного средства	Устный опрос по разделу 2 «OLAP- системы»
Представление и содержание оценочных материалов	<p>Оценочный материал содержит 15 вопросов по разделу <i>Примеры вопросов по разделу дисциплины</i></p> <ol style="list-style-type: none"> 1. Двенадцать правил Кодда. 2. Тест FASMI. 3. Классификация OLAP-продуктов по способу хранения данных: MOLAP, ROLAP и HOLAP. 4. Многомерная модель данных. 5. Архитектура OLAP приложений. 6. Интегрированное хранилище данных.
Критерии оценки и шкала оценивания в баллах	<p>При оценке устного опроса учитываются следующие критерии:</p> <ol style="list-style-type: none"> 1. Знание материала 2. Последовательность изложения 3. Владение речью и терминологией 4. Применение конкретных примеров <p>Шкала оценивания:</p> <p><i>Высокий уровень:</i> - содержание материала раскрыто в полном объеме, предусмотренном программой дисциплины. содержание материала раскрыто последовательно, достаточно хорошо продумано, материал изложен грамотным языком, с точным использованием терминологии, показано умение иллюстрировать материал конкретными примерами – 10 баллов;</p> <p><i>Средний уровень:</i> - содержание материала раскрыто достаточно полно, показано общее понимание вопроса, последовательность изложения материала достаточно продумана, приведение примеров с незначительными ошибками – 7 баллов;</p> <p><i>Ниже среднего уровень:</i> - содержание материала раскрыто неполно, показано общее понимание вопроса, достаточное для дальнейшего изучения программного материала, последовательность изложения материала недостаточно продумана, приведение примеров вызывает затруднение – 4 балла;</p> <p><i>Низкий уровень:</i> - не раскрыто основное содержание учебного материала, путаница в изложении материала, допущены ошибки в определении понятий, неумение приводить примеры при объяснении материала – 0 баллов.</p> <p>Максимальное количество баллов - 10</p>
Наименование оценочного средства	Тестирование по разделу 2 «OLAP- системы»
Представление и	Тестирование проводится в системе LMS Moodle. Содержание тестов соответствует тематике пройденного материала. База вопросов более 100,

<p>содержание оценочных материалов</p>	<p>которая регулярно обновляется и случайным образом формируются в тест. Тест содержит 20 вопросов с заданиями разных типов.</p> <p><i>Примеры тестовых заданий</i></p> <p>1. Тип отчета, наиболее соответствующий MOLAP и многомерным базам данных, носит название:</p> <p>А. диагональный В. табличный С. матричный</p> <p>2. Перечислите преимущества формата MOLAP:</p> <p>А. превосходные свойства индексации В. высокая эффективность использования дискового пространства С. обеспечивает значительно более высокий уровень защиты данных и хорошие возможности разграничения прав доступа по сравнению с другими форматами D. высокая производительность</p> <p>3. На основе каких понятий выполняется многомерное моделирование?</p> <p>А. атрибуты, измерения, параметры (метрики), иерархия, гранулированность В. факты, атрибуты, измерения, параметры (метрики), иерархия, гранулированность С. факты, атрибуты, измерения, иерархия, гранулированность</p> <p>4. Установите соответствие между понятиями, связанными с многомерным моделированием, и их определениями</p> <p>Понятие:</p> <p>1) факт 2) измерение 3) атрибут 4) параметр 5) иерархия 6) гранулированность</p> <p>Определение:</p> <p>А) это числовая характеристика факта, который определяет эффективность деятельности или бизнес - действия организации с точки зрения измерения В) это интерпретация факта с некоторой точки зрения в реальном мире С) это описание характеристики реального объекта предметной области D) это набор связанных элементов данных, содержащих метрики и описательные данные Е) это уровень детализации данных, сохраняемых в ХД F) это отношение часть-целое или родитель – потомок</p> <p>5. Сбалансированная иерархия – это иерархия</p> <p>А. в которой все ветви измерения имеют одно и то же количество уровней В. в которой все ветви измерения имеют приблизительно одно и то же количество уровней С. в которой все ветви измерения отличаются на определенное количество уровней</p>
<p>Критерии оценки и шкала оценивания в баллах</p>	<p>Результат тестирования автоматически оценивает программа LMS Moodle в баллах по предварительной настройке. Настройка теста позволяет оценивать результат в автоматическом режиме. Знания обучающегося в результате промежуточной аттестации в формате модульного тестирования оцениваются в зависимости от количества правильных ответов следующим образом:</p> <p>10 баллов – 100% правильного ответа 7 баллов – от 70 % до 84 % правильных ответов 4 балла – от 50% до 69% правильных ответов 0 баллов – меньше 50% правильных ответов</p> <p>Максимальное количество баллов - 10</p>
<p>Наименование оценочного</p>	<p>Устный опрос по разделу 3 «Интеллектуальный анализ данных (Data Mining)»</p>

<p align="center">средства</p>	
<p>Представление и содержание оценочных материалов</p>	<p>Оценочный материал содержит 15 вопросов по разделу <i>Примеры вопросов по разделу дисциплины</i></p> <ol style="list-style-type: none"> 1. Классификация задач Data Mining. 2. Подготовка исходных данных. 3. Предсказательные модели Data Mining. 4. Практическое применение Data Mining в бизнес-сфере. 5. Базовые методы Data Mining. 6. Процесс обнаружения знаний.
<p>Критерии оценки и шкала оценивания в баллах</p>	<p>При оценке устного опроса учитываются следующие критерии:</p> <ol style="list-style-type: none"> 1. Знание материала 2. Последовательность изложения 3. Владение речью и терминологией 4. Применение конкретных примеров <p>Шкала оценивания:</p> <p><i>Высокий уровень:</i></p> <p>- содержание материала раскрыто в полном объеме, предусмотренном программой дисциплины. содержание материала раскрыто последовательно, достаточно хорошо продумано, материал изложен грамотным языком, с точным использованием терминологии, показано умение иллюстрировать материал конкретными примерами – 10 баллов;</p> <p><i>Средний уровень:</i></p> <p>- содержание материала раскрыто достаточно полно, показано общее понимание вопроса, последовательность изложения материала достаточно продумана, приведение примеров с незначительными ошибками – 7 баллов;</p> <p><i>Ниже среднего уровень:</i></p> <p>- содержание материала раскрыто неполно, показано общее понимание вопроса, достаточное для дальнейшего изучения программного материала, последовательность изложения материала недостаточно продумана, приведение примеров вызывает затруднение – 4 балла;</p> <p><i>Низкий уровень:</i></p> <p>- не раскрыто основное содержание учебного материала, путаница в изложении материала, допущены ошибки в определении понятий, неумение приводить примеры при объяснении материала – 0 баллов.</p> <p>Максимальное количество баллов - 10</p>
<p>Наименование оценочного средства</p>	<p align="center">Тестирование по разделу 3 «Интеллектуальный анализ данных (Data Mining)»</p>
<p>Представление и содержание оценочных материалов</p>	<p>Тестирование проводится в системе LMS Moodle. Содержание тестов соответствует тематике пройденного материала. База вопросов более 100, которая регулярно обновляется и случайным образом формируются в тест. Тест содержит 20 вопросов с заданиями разных типов.</p> <p><i>Примеры тестовых заданий</i></p> <ol style="list-style-type: none"> 1. Назовите факторы, обусловившие возникновение и развитие Data Mining: <ol style="list-style-type: none"> A. совершенствование аппаратного и программного обеспечения B. совершенствование технологий хранения и записи данных табличный C. накопление большого количества ретроспективных данных D. совершенствование алгоритмов обработки информации E. необходимость замены аналитика информационной технологией 2. В результате использования инструментов Data Mining пользователь может: <ol style="list-style-type: none"> A. получить гипотезы о взаимосвязях в данных, самостоятельно выдвинутые инструментом Data Mining B. получить подтверждение или опровержение гипотез, выдвинутых пользователем C. проверить гипотезы о взаимосвязях в данных, самостоятельно выдвинутые пользователем инструмента Data Mining

	<p>D. все ответы верны</p> <p>3. Данные – это...</p> <p>A. необработанный материал, предоставляемый поставщиками данных и используемый потребителями для формирования информации на основе данных</p> <p>B. готовый материал для формирования информации</p> <p>C. синоним информации</p> <p>4. Какая из перечисленных ниже стадий может считаться дополнительной или частью одной из основных стадий Data mining:</p> <p>A. выявление закономерностей (свободный поиск)</p> <p>B. использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование)</p> <p>C. валидация</p> <p>5. Согласно классификации по стратегиям, задачи Data Mining подразделяются на:</p> <p>A. обучение с учителем</p> <p>B. обучение без учителя</p> <p>C. дескриптивные</p> <p>D. прогнозирующие</p>
<p>Критерии оценки и шкала оценивания в баллах</p>	<p>Результат тестирования автоматически оценивает программа LMS Moodle в баллах по предварительной настройке. Настройка теста позволяет оценивать результат в автоматическом режиме. Знания обучающегося в результате промежуточной аттестации в формате модульного тестирования оцениваются в зависимости от количества правильных ответов следующим образом:</p> <p>10 баллов – 100% правильного ответа</p> <p>7 баллов – от 70 % до 84 % правильных ответов</p> <p>4 балла – от 50% до 69% правильных ответов</p> <p>0 баллов – меньше 50% правильных ответов</p> <p>Максимальное количество баллов - 10</p>
<p>Наименование оценочного средства</p>	<p>Отчет по практическим занятиям по разделу 4 «Задачи и методы Data Mining»</p>
<p>Представление и содержание оценочных материалов</p>	<p><i>Практическая работа включает:</i></p> <ol style="list-style-type: none"> 1) план работы 2) краткие теоретические сведения по теме 3) примеры программ 4) методику выполнения и задания для самостоятельной работы 5) выполнение задания путем написания кода программы и последующего оформления отчета о выполненной работе 6) контрольные вопросы <p style="text-align: center;">Практическое занятие по теме «Методы регрессионного анализа»</p> <p><i>Примеры заданий</i></p> <p>Задание 1</p> <p>Вариант 1.</p> <p>На основе выборки объемом в 38 ед. из двумерной генеральной совокупности были получены выборочные характеристики: $b_{xy}=1.8$; $S_x^2=9$; $S_y^2=36$. Проверить значимость генерального коэффициента корреляции с $\alpha=0.1$.</p> <p>Вариант 2.</p> <p>На основании 30 выборочных данных о выручке (y) и массе реализованной продукции (x) выявлено, что выборочная доля дисперсии выручки, вызванная вариацией массы реализованной продукции, составила 89%. При уровне значимости 0.05 проверить значимость коэффициента</p>

корреляции.

Задание 2

В таблице заданы значения наблюдений x_t и y_t .

Вариант 1

t	1	2	3	4	5	6	7	8	9	10
x_t	9	11	15	6	17	13	14	16	7	11
y_t	31	31	35	28	33	31	34	34	28	31

Вариант 2

t	1	2	3	4	5	6	7	8	9	10
x_t	14	18	16	8	19	6	16	13	8	6
y_t	110	136	125	84	140	77	120	100	84	75

Пусть $z_t = \sqrt{x_t}$. Предполагается, что зависимую переменную y и факторы связывает уравнение множественной линейной регрессии

$$y_t = a + b \cdot x_t + c \cdot z_t + \varepsilon_t,$$

где a , b и c неизвестные параметры уравнения, ε_t – случайные отклонения.

1. Определите МНК оценки параметров уравнения.
2. С уровнем значимости 0,05 проверьте гипотезу $b=0$ (о влиянии фактора x на результат) и $c=0$ (о влиянии фактора z на результат).
3. Определите коэффициент детерминации и скорректированный коэффициент детерминации.
4. По критерию Фишера F с уровнем значимости 0,05 оцените качество модели в целом
5. Составьте корреляционную таблицу наблюдений и вычислите частные коэффициенты корреляции.

Задание 3

В таблице заданы значения наблюдений x_t и y_t .

Вариант 1

t	1	2	3	4	5	6	7	8	9	10
x_t	9	11	15	6	17	13	14	16	7	11
y_t	31	31	35	28	33	31	34	34	28	31

Вариант 2

t	1	2	3	4	5	6	7	8	9	10
x_t	14	18	16	8	19	6	16	13	8	6
y_t	110	136	125	84	140	77	120	100	84	75

Предполагается, что зависимую переменную y и независимую x связывает нелинейное регрессионное уравнение:

Вариант 1

$$y_t = a + b \cdot \frac{1}{x_t} + \varepsilon_t$$

Вариант 2

$$y_t = a \cdot e^{b \cdot x_t} \cdot \varepsilon_t$$

1. Проведите линеаризацию модели, определите оценки параметров нелинейной модели.
2. Оцените качество модели с помощью коэффициента детерминации и дисперсионного отношения F .
3. Определите прогнозное значение \hat{y}_{11} при $x_{11}=N$, где N – номер Вашего варианта. Постройте 95% доверительный интервал для прогноза.
4. Оцените с помощью эластичности силу влияния фактора на результат в

точке x_{11} .

5. На диаграмме рассеяния постройте график прогнозных значений. Определите сумму квадратов отклонений наблюдений от нелинейного прогноза.

Практическое занятие по теме «Факторный анализ: метод главных компонент»

Пример задания

Задание. Исходные данные для обработки находятся в файле Data_2.txt, в котором представлен массив следующих показателей деятельности предприятий:

Y_1 - производительность труда; Y_2 - индекс снижения себестоимости продукции; Y_3 - рентабельность; X_4 - трудоемкость единицы продукции; X_5 - удельный вес рабочих в составе ППП; X_6 - удельный вес покупных изделий; X_7 - коэффициент сменности оборудования; X_8 - премии и вознаграждения на одного работника; X_9 - удельный вес потерь от брака; X_{10} - фондоотдача; X_{11} - среднегодовая численность ППП; X_{12} - среднегодовая стоимость ОПФ; X_{13} - среднегодовой фонд заработной платы ППП; X_{14} - фондовооруженность труда; X_{15} - оборачиваемость нормируемых оборотных средств; X_{16} - оборачиваемость ненормируемых оборотных средств; X_{17} - непроизводственные расходы.

Вариант 1: результативный признак - Y_1 , факторные признаки: $X_6, X_8, X_{11}, X_{12}, X_{17}$

Вариант 2: результативный признак - Y_2 , факторные признаки: $X_4, X_5, X_7, X_9, X_{17}$

По данным варианта провести факторный анализ. Для факторных признаков X найти оценку матрицы R парных коэффициентов корреляции. На основании матрицы собственных значений определить вклад компонент в суммарную дисперсию. Отобрать и указать вклад первых главных компонент. Используя матрицу факторных нагрузок A , дать экономическую интерпретацию полученным главным компонентам. По матрице значений главных компонент F провести классификацию объектов по двум первым главным компонентам. Дать интерпретацию полученным результатам. Используя вектор значений результативного признака Y и матрицу F , построить уравнение регрессии на главные компоненты с максимальным числом значимых коэффициентов регрессии.

Практическое занятие по теме «Методы деревьев решений, «ближайших соседей», логистическая регрессия, классификатор Байеса»

Примеры заданий

Задание 1.

1. Найдите реальные данные для задачи классификации.
2. Перечислите факторы (и укажите единицы их измерения) и поясните смысл переменной отклика.
3. Если число факторов $m \leq 2$, то визуализируйте данные (постройте облако точек, где разным классам соответствуют разные цвета).
4. Реализуйте метод kNN (метод ближайших соседей).
5. Оцените точность полученного решения с помощью метода кросс-валидации.
6. Задайте несколько новых данных. Если $m \leq 2$, то покажите соответствующие точки на графике (выделите их другим цветом).
7. Определите значение переменной отклика (номер класса) для новых данных.

Задание 2

1. Примените метод деревьев решений для задачи классификации (для использованных ранее или новых данных).
2. Если число признаков $m = 2$, визуализуйте данные (постройте облако точек).
3. Исследуйте дерево решений; если позволяет размерность, постройте его график.
4. Проанализируйте точность полученных решений для тестовых данных (с известным значением переменной отклика), сравните результаты с ранее полученными.
5. Задайте нескольких новых данных, покажите соответствующие точки на графике (при $m = 2$) (выделите их другим цветом).
6. Определите класс для новых данных (т.е. для данных, где известны только значения факторов, но не переменной отклика).

Задание 3

Используя данные о заработной плате и платёжеспособности клиентов банка из файла Clients.txt), построить логистическую регрессионную модель:

1. найти и коэффициенты логистической регрессионной модели;
2. постройте на одном рисунке облако точек, соответствующее исходным данным (точки, соответствующие благонадёжным клиентам, сделайте синими, неблагонадёжным - красными) и логистическую кривую, соответствующую найденным значениям коэффициентов;
3. оцените благонадёжность нескольких новых клиентов согласно построенной логистической регрессионной модели, сделайте выводы.

Задание 4

1. Выберите данные для задачи классификации (для использованных ранее или новых данных).
2. Решите задачу с помощью наивного Байесовского классификатора; если число признаков равно 2, то визуализируйте данные.
3. Постройте кросс-валидационную таблицу, сделайте вывод о точности решения задачи классификации.
4. Задайте несколько новых данных, покажите соответствующие точки на графике (выделите их другим цветом).
5. Определите класс для новых данных.

Практическое занятие по теме «Нейронные сети: обучение, модели»

Пример задания

Задание

Разработать программу «Распознавание образов», отвечающую следующим требованиям.

1. Распознавание образов должно выполняться искусственной нейронной сетью, обучаемой по алгоритму с обратным распространением ошибки. Общие требования к сети и программе:

- связи между нейронами – прямые;
- количество скрытых слоев – 1;
- количество нейронов выходного слоя (классов образов) – не менее 4;
- количество обучающих образов – не менее 1 на каждый класс образов;
- функция активации – сигмоидальная
- распознаваемые (тестовые) образы – формируются пользователем;
- норма обучения и количество эпох обучения – задаются пользователем;
- режим обучения или распознавания – задается пользователем

Вариант 1

Тип образа – арифметические операции (+, -, /, *, $\sqrt{\quad}$, % и т.д.), количество нейронов скрытого слоя – 4

Вариант 2

	<p>Тип образа – геометрические фигуры, количество нейронов скрытого слоя - 7</p> <p>Практическое занятие по теме «методы поиска ассоциативных правил» <i>Пример задания</i> Задание. Имеется выборка из базы данных компании, которая занимается продажей электронной и бытовой техники (файл Data_apriori.txt). В данных содержится информация о покупках 1000 клиентов по 10 наименованиям. Применить априорный метод поиска ассоциативных правил для данных и пояснить обнаруженные зависимости между переменными.</p> <p>Практическое занятие по теме «методы кластерного анализа» <i>Пример задания</i> Задание. Исходные данные для обработки находятся в файле Data_2.txt, в котором представлен массив следующих показателей деятельности предприятий: Y1 - производительность труда; Y2 - индекс снижения себестоимости продукции; Y3 - рентабельность; X4 - трудоемкость единицы продукции; X5 - удельный вес рабочих в составе ППП; X6 - удельный вес покупных изделий; X7- коэффициент сменности оборудования; X8 - премии и вознаграждения на одного работника; X9 - удельный вес потерь от брака; X10- фондоотдача; X11 - среднегодовая численность ППП; X12 - среднегодовая стоимость ОПФ; X13 - среднегодовой фонд заработной платы ППП; X14- фондовооруженность труда; X15 - оборачиваемость нормируемых оборотных средств; X16 - оборачиваемость ненормируемых оборотных средств; X17 - непроизводственные расходы. Вариант 1: результативный признак - Y1, факторные признаки: X6, X8, X11, X12, X17 Вариант 2: результативный признак – Y2, факторные признаки: X4, X5, X7, X9, X17</p> <p>К данным варианта применить метод К-средних для классификации предприятий и построить дендрограмму. По дендрограмме выбрать окончательный вариант разбиения предприятий на кластеры и дать интерпретацию полученным результатам, используя статистические характеристики кластеров (векторы средних и средних квадратических отклонений).</p>
<p>Критерии оценки и шкала оценивания в баллах</p>	<p><i>При оценке отчетов по лабораторным заданиям учитываются следующие критерии:</i></p> <ol style="list-style-type: none"> 1. Знание теоретического материала 2. Выполнение самостоятельных заданий 3. Отчет о выполненной работе 4. Ответы на контрольные вопросы <p>Шкала оценивания <i>Высокий уровень знаний</i> - теоретический материал знает, правильно выполнены все задания в соответствии с требованиями, своевременно предоставлен отчет о выполнении работы - 40 баллов. <i>Средний уровень знаний:</i> - теоретический материал знает, правильно выполнены все задания, предоставлен отчет о выполнении работы, либо в случае несвоевременного предоставления отчета или с наличием несущественных ошибок в выполнении заданий -30 баллов. <i>Ниже среднего уровня:</i> Выполнено не все, но более 50% заданий, несвоеременно предоставлен отчет</p>

о выполнении работы - **20** баллов.

Низкий уровень:

- выполнено менее 50% заданий, отчет о выполнении работы не предоставлен

– **0** баллов

Максимальное количество баллов – **40**.