МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

ВВЕДЕНИЕ В АНАЛИЗ БОЛЬШИХ ДАННЫХ

Направление и направленность (профиль)
09.03.04 Программная инженерия. Программная инженерия

Год набора на ОПОП
2024

Форма обучения
очная

Владивосток 2024

Рабочая программа дисциплины (модуля) «Введение в анализ больших данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.03.04 Программная инженерия (утв. приказом Минобрнауки России от 19.09.2017г. №920) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Ермолицкая М.З., кандидат биологических наук, доцент, Научно-образовательный центр "Искусственный интеллект", Marina.Ermolitskaya@vvsu.ru

Утверждена на заседании научно-образовательный центр "искусственный интеллект" от 19.06.2024 , протокол № 1

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)
Кригер А.Б.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1582918206
Номер транзакции	0000000000D19514
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью освоения дисциплин ы «Введение в анализ больших данных» является теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику для решения различных задач практической и научно-исследовательской деятельности.

Задачи освоения дисциплины:

· приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;

· применение статистических и математических методов для анализа больших объемов информации;

· приобретение практических навыков работы с программой RStudio.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине		
			Код результата	Формулировка результата	
09.03.04 «Программная инженерия» (Б-ИН)	ОПК-8 : Способен осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий	ОПК-8.1к : Применяет методы поиска и хранения информации с использованием современных информационных технологий	РД1	Знание	основных методов обработки и анализа больших данных
			РД2	Умение	проводить сравнительный анализ и выбор статистических методов для анализа конкретных данных
			РД3	Навык	применения статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio
		ОПК-8.2к : Использует современные информационные технологии для обработки и анализа информации	РД1	Знание	основных методов обработки и анализа больших данных
			РД2	Умение	проводить сравнительный анализ и выбор статистических методов для анализа конкретных данных
			РД3	Навык	применения статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio

2 Место дисциплины (модуля) в структуре ОПОП

Освоение дисциплины формирует у обучающихся компетенции, необходимые для подготовки в соответствии с требованиями ФГОС ВО в области использования современных технологий для обработки, анализа и визуализации больших данных.

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудоемкость	Объем контактной работы (час)						СРС	Форма аттестации
				(З.Е.)	Всего	Аудиторная			Внеаудиторная			
						лек.	прак.	лаб.	ПА	КСР		
09.03.04 Программная инженерия	ОФО	Б1.Б	7	3	49	16	32	0	1	0	59	Э

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Код результата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Введение в анализ больших данных. Обзор источников информации.	РД1	2	0	0	10	собеседование
2	Технологии хранения и обработки больших данных.	РД1	2	0	0	6	собеседование
3	Современные программные средства анализа больших объемов информации.	РД1	2	0	0	6	собеседование
4	Методы обработки и анализа больших данных.	РД1	10	0	0	10	собеседование
5	Сбор и хранение больших данных.	РД2	0	6	0	10	отчет по выполненным практическим работам
6	Методы обработки, анализа и визуализации больших данных в программе RStudio.	РД2, РД3	0	26	0	17	отчет по выполненным практическим работам
Итого по таблице			**16**	**32**	**0**	**59**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Введение в анализ больших данных. Обзор источников информации.

Содержание темы: Основные определения, термины, задачи анализа больших данных. Вопросы безопасности. Понятие Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных.

Формы и методы проведения занятий по теме, применяемые образовательные

технологии: лекции, на которых дается основной систематизированный материал по темам.

Виды самостоятельной подготовки студентов по теме: чтение предлагаемой литературы, подготовка к собеседованию, итоговому тесту.

Тема 2 Технологии хранения и обработки больших данных.

Содержание темы: Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, на которых дается основной систематизированный материал по темам.

Виды самостоятельной подготовки студентов по теме: чтение предлагаемой литературы, подготовка к собеседованию, итоговому тесту.

Тема 3 Современные программные средства анализа больших объемов информации.

Содержание темы: Обзор современных популярных программных средств анализа данных. Платные, бесплатные программные средства: Statistica, SPSS, Excel, R-Studio и другие; их преимущества и недостатки.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, на которых дается основной систематизированный материал по темам.

Виды самостоятельной подготовки студентов по теме: чтение предлагаемой литературы, подготовка к собеседованию, итоговому тесту.

Тема 4 Методы обработки и анализа больших данных.

Содержание темы: Основные понятия математической статистики. Методы анализа данных: дескриптивная статистика, критерии для проверки на нормальность распределения; параметрические, непараметрические, номинальные методы (критерии для определения значимости различий в выборках, определение зависимости между переменными, построение регрессионных моделей, дисперсионный, кластерный, дискриминантный, факторный анализы).

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, на которых дается основной систематизированный материал по темам.

Виды самостоятельной подготовки студентов по теме: чтение предлагаемой литературы, подготовка к собеседованию, итоговому тесту.

Тема 5 Сбор и хранение больших данных.

Содержание темы: Поиск источников информации в сети Интернет: открытые и закрытые источники данных. Портал открытых данных РФ. Сохранение данных в программе MS Excel. Преобразование и первичная обработка данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практические занятия проводятся в компьютерном классе с использованием программ RStudio (преподаватель излагает тему, приводит примеры и дает задание для самостоятельного выполнения, при необходимости консультирует студентов).

Виды самостоятельной подготовки студентов по теме: подготовка отчета по практическим работам, подготовка к итоговому тесту.

Тема 6 Методы обработки, анализа и визуализации больших данных в программе RStudio.

Содержание темы: Представление исходных данных в программе RStudio (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах MS Excel и RStudio: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа. Применение различных методов выделения, извлечения и группировки данных, которые

позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы). Возможности графического представления информации в программе RStudio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практические занятия проводятся в компьютерном классе с использованием программы RStudio (преподаватель излагает тему, приводит примеры и дает задание для самостоятельного выполнения, при необходимости консультирует студентов).

Виды самостоятельной подготовки студентов по теме: подготовка отчета по практическим работам, подготовка к итоговому тесту.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

На лекциях дается основной систематизированный материал по темам.

Практические занятия проводятся в компьютерном классе с использованием программ MS Excel и RStudio. Преподаватель излагает тему, приводит примеры и дает задание для самостоятельного выполнения. При необходимости консультирует студентов.

Самостоятельная работа студентов подразумевает чтение предлагаемой преподавателем литературы и использование интернет-ресурсов для подготовки к занятиям, текущей и промежуточной аттестации.

Промежуточная аттестация - итоговый тест.

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и

характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Гусева Е. Н. Теория вероятностей и математическая статистика : Учебники [Электронный ресурс] - Москва : Флинта , 2016 - 220 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=83543

2. Гутова С. Г., Алтемерова О. А. Теория вероятностей и математическая статистика : Учебники [Электронный ресурс] - Кемерово : Кемеровский государственный университет , 2016 - 216 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=481538

3. Миркин Б. Г. ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ. Учебник и практикум [Электронный ресурс] : М.:Издательство Юрайт , 2019 - 174 - Режим доступа: https://biblio-online.ru/book/vvedenie-v-analiz-dannyh-432851

7.2 Дополнительная литература

1. Белько, И. В. Теория вероятностей, математическая статистика, математическое программирование : учебное пособие / И. В. Белько, И. М. Морозова,Е. А. Криштапович. — Москва : ИНФРА-М, 2022. — 299 с. : ил. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-011748-5. - Текст : электронный. - URL: https://znanium.com/catalog/product/1862599 (дата обращения: 06.09.2023).

2. Волкова Полина Андреевна. Статистическая обработка данных в учебно-исследовательских работах : Учебное пособие [Электронный ресурс] : Форум , 2019 - 96 - Режим доступа: http://znanium.com/go.php?id=1030246

3. Непомнящая Наталья Васильевна. Статистика: общая теория статистики, экономическая статистика. Практикум : Учебное пособие [Электронный ресурс] , 2015 - 376 - Режим доступа: http://znanium.com/go.php?id=549841

4. Основы теории вероятностей и математической статистики : учеб.-метод. пособие / Т.И. Волынкина; Г.В. Воронина .— Орёл : Изд-во Орел ГАУ, 2018 .— 90 с. — URL: https://lib.rucont.ru/efd/684477 (дата обращения: 30.09.2024)

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. - Электронная книга, адрес доступа: http://r-analytics.blogspot.com
2. Электронная библиотечная система «Университетская библиотека онлайн» - Режим доступа: http://biblioclub.ru/
3. Электронная библиотечная система ZNANIUM.COM - Режим доступа: http://znanium.com/
4. Электронно-библиотечная система "ZNANIUM.COM"
5. Электронно-библиотечная система "РУКОНТ"
6. Электронно-библиотечная система издательства "Юрайт" - Режим доступа: https://biblio-online.ru/
7. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/
8. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/
9. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

· Компьютеры

· Монитор облачный 23" LG23CAV42K/мышь Genius Optical Wheel проводная/клавиатура Genius KB110 проводная

· Мультимедийный проектор CASIO (Япония)

· Облачный монитор LG Electronics черный +клавиатура+мышь

· Уст-во бесп.пит.SmartUPS 3000

Программное обеспечение:

· RStudio

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

ВВЕДЕНИЕ В АНАЛИЗ БОЛЬШИХ ДАННЫХ

Направление и направленность (профиль)

09.03.04 Программная инженерия. Программная инженерия

Год набора на ОПОП
2024

Форма обучения
очная

Владивосток 2024

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
09.03.04 «Программная инженерия» (Б-ИН)	ОПК-8 : Способен осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий	ОПК-8.1к : Применяет методы поиска и хранения информации с использованием современных информационных технологий
		ОПК-8.2к : Использует современные информационные технологии для обработки и анализа информации

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Компетенция ОПК-8 «Способен осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий»

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ОПК-8.1к : Применяет методы поиска и хранения информации с использованием современных информационных технологий	РД 1	Знание	основных методов обработки и анализа больших данных	знание методов обработки и анализа больших данных, используемых при решении профессиональных задач
	РД 2	Умение	проводить сравнительный анализ и выбор статистических методов для анализа конкретных данных	проведение сравнительного анализа и выбора статистических методов для анализа конкретных данных
	РД 3	Навык	применения статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio	применение статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio
ОПК-8.2к : Использует современные информационные технологии для обработки и анализа информации	РД 1	Знание	основных методов обработки и анализа больших данных	знание методов обработки и анализа больших данных, используемых при решении профессиональных задач
	РД 2	Умение	проводить сравнительный анализ и выбор статистических методов для анализа конкретных данных	проведение сравнительного анализа и выбора статистических методов для анализа конкретных данных

	РД3	Навык	применения статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio	применение статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очная форма обучения				
РД1	Знание : основных методов обработки и анализа больших данных	1.1. Введение в анализ больших данных. Обзор источников информации.	Собеседование	Тест
		1.2. Технологии хранения и обработки больших данных.	Собеседование	Тест
		1.3. Современные программные средства анализа больших объемов информации.	Собеседование	Тест
		1.4. Методы обработки и анализа больших данных.	Собеседование	Тест
РД2	Умение : проводить сравнительный анализ и выбор статистических методов для анализа конкретных данных	1.5. Сбор и хранение больших данных.	Практическая работа	Тест
		1.6. Методы обработки, анализа и визуализации больших данных в программе RStudio.	Практическая работа	Тест
РД3	Навык : применения статистических методов для обработки и анализа больших объемов информации с использованием программы RStudio	1.6. Методы обработки, анализа и визуализации больших данных в программе RStudio.	Практическая работа	Тест

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Вид учебной деятельности	Оценочное средство			
	Собеседование	Отчет по выполненным практическим работам	Итоговый тест	Итого

Лекции	10			10
Практические занятия		60		60
Самостоятельная работа	10			10
Итоговая аттестация			20	20
Итого	20	60	20	100

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, об наруживает всестороннее, систематическое и глубокое знание учебного матер иала, усвоил основную литературу и знаком с дополнительной литературой, ре комендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниям и, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: ос новные знания, умения освоены, но допускаются незначительные ошибки, нет очности, затруднения при аналитических операциях, переносе знаний и умени й на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительн о»	Студент демонстрирует сформированность дисциплинарных компетенций: в х оде контрольных мероприятий допускаются значительные ошибки, проявляетс я отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарн ым компетенциям, студент испытывает значительные затруднения при оперир овании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворитель но»	У студента не сформированы дисциплинарные компетенции, проявляется недо статочность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворитель но»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Примерный перечень вопросов по темам

Контрольные вопросы для собеседования по темам

Тема 1. Введение в анализ больших данных. Обзор источников информации.

1. Дайте определение понятию «информационные ресурсы».
2. Что означает «информационный поиск»?
3. Информационно-коммуникационные технологии, что это?
4. Перечислите основные компоненты процесса поиска информации.
5. Определите понятие «информационные системы».
6. Охарактеризуйте портал открытых данных РФ.
7. Определите сущность понятия «большие данные».
8. Определите понятие Data Mining.
9. Каковы главные проблемы безопасности «больших данных»?
10. Дайте характеристику принципу безопасности «intelligence».

Тема 2. Технологии хранения и обработки больших данных.

1. Перечислите технологии хранения больших данных.
2. Характеристики системы хранения данных RCS.

3. Анализ больших данных в QlikView.
4. Характеристики системы хранения данных Полибайт.
5. Какие модели данных вы знаете?
6. Что включает первичная обработка данных?

Тема 3. Современные программные средства анализа больших объемов информации.

1. Перечислите программные средства анализа данных: платные и бесплатные.
2. Преимущества работа с данными в программе R-Studio.
3. Каковы возможности представления данных в программе R-Studio?

Тема 4. Статистические методы анализа данных.

1. Опишите свойства нормального распределения.
2. Определите различия между параметрическими, непараметрическими и номинальными методами.
3. Критерии для определения различий в выборках.
4. Опишите основную идею корреляционного анализа.
5. Что показывает коэффициент корреляции Пирсона?
6. Коэффициенты связи между переменными, не подчиняющимися нормальному закону распределения.
7. Для чего применяют регрессионный анализ?
8. Типы регрессионных моделей.
9. Как проверить адекватность построенной регрессионной модели?
10. Основная идея дисперсионного анализа.
11. Сущность кластерного анализа.
12. Для чего используют дискриминантный анализ?
13. Цели применения факторного анализа.

Краткие методические указания

Собеседование проводится после изучения соответствующей темы. Преподаватель в устной форме задает вопросы студентам на лекционных занятиях.

Шкала оценки

№	Баллы	Описание
5	16–20	Процент правильных и обоснованных ответов от 95% до 100%
4	11–15	Процент правильных и обоснованных ответов от 80 до 94%
3	6–10	Процент правильных ответов с помощью наводящих вопросов от 65 до 79%
2	0–5	Процент правильных ответов от 45 до 64%

5.2 Примеры заданий для выполнения практических работ

Тема 1. Сбор данных из различных источников в сети Интернет. Портал открытых данных РФ. Хранение данных в программе MS Excel.

Тема 2. Первичный анализ данных и визуализация данных в программе Excel.

Тема 3. Знакомство с программой RStudio. Синтаксис. Представление исходных данных в программе RStudio (векторы, массивы, матрицы, списки, таблицы).

Тема 4. Выборка и преобразование исходных данных в программе RStudio. Удаление пропущенных значений.

Тема 5. Статистическая обработка данных в программе RStudio: подсчет описательных статистик.

Тема 6. Возможности графического представления информации в программе RStudio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

Тема 7. Законы распределения вероятностей, реализованные в R. Проверка данных на нормальность распределения: критерии Шапиро-Уилка, Колмогорова-Смирнова и др. Уровень статистической достоверности.

Тема 8. Сравнение выборок. Критерий Стьюдента, Критерий согласия хи-квадрат

Пирсона, Критерий Колмогорова-Смирнова.

Тема 9. Непараметрические методы сравнения для зависимых и независимых выборок: критерий Уилкоксона, критерий Краскела-Уоллиса.

Тема 10. Корреляционный анализ. Расчет коэффициентов корреляции Пирсона, Спирмена, Кендалла

Тема 11. Регрессионный анализ (линейная зависимость). Построение линейной модели. Проверка адекватности построенной модели.

Тема 12. Регрессионный анализ (нелинейная зависимость). Определение вида зависимости. Построение модели. Проверка адекватности построенной модели.

Тема 13. Однофакторный дисперсионный анализ.

Тема 14. Многофакторный дисперсионный анализ.

Тема 15. Факторный анализ.

Тема 16. Кластерный анализ.

Тема 17. Возможности графического представления информации в программе RStudio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

Краткие методические указания

На выполнение одной практической работы отводится не более одного двухчасового занятия. После выполнения каждой практической работы студент должен представить отчет в виде скрипта с описанием полученных результатов, а также, ответить на сопутствующие вопросы по теме.

Шкала оценки

№	Баллы	Описание
5	47–60	Студент демонстрирует умения на итоговом уровне: умеет свободно выполнять практически е задания, предусмотренные программой, свободно оперирует приобретенными знаниями и у мениями, применяет их в ситуациях повышенной сложности.
4	32–46	Студент демонстрирует умения на среднем уровне: освоил основные умения, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе у мений на новые, нестандартные ситуации.
3	26–31	Студент демонстрирует умения и навыки на базовом уровне: в ходе контрольных мероприяти й допускаются значительные ошибки, проявляется отсутствие отдельных умений, навыков по дисциплинарным компетенциям, испытываются значительные затруднения при оперировани и умениями и при их переносе на новые ситуации.
2	0–25	Студент демонстрирует умения и навыки на уровне ниже базового: проявляется недостаточно сть умений и навыков.

5.3 Итоговый тест

Данные - это

a) факты, характеризующие объекты, процессы, явления предметной области

b) данные, рассматриваемые в каком-либо контексте, из которого пользователь может составить собственное мнение

c) закономерности проблемной области, полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области

d) сведения, передаваемые людьми устным, письменным или другим способом

Каким признаком не обладают большие данные?

a) многообразие данных;

b) достоверность данных;

c) скорость накопления данных;

d) типизацией исходных данных.

Слабо структурированные данные могут быть записаны в форме

a) таблиц;

b) отдельных векторов (строк);

c) записей произвольной последовательности;

d) таблиц и отношений.

Неструктурированные данные могут быть записаны в форме

a) таблиц;

b) записей произвольной последовательности;

c) таблиц и отношений;

d) записей произвольной длины.

Аналитик- это

a) специалист, занимающийся анализом в различных сферах деятельности и разработкой моделей ля проведения анализа;

b) специалист в выбранной предметной области;

c) сотрудник выполняющих узкий круг задач в выбранной области;

d) сотрудник, который имеет опыт в программировании.

Эксперт - это

a) специалист, занимающийся анализом в различных областях деятельности и разработкой моделей ля проведения анализа;

b) специалист в выбранной предметной области;

c) сотрудник выполняющих узкий круг задач в выбранной области;

d) сотрудник, который имеет опыт в программировании.

Классификация -

a) некоторый набор операций над базой данных, который рассматривается, как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных;

b) разновидность систем хранения, ориентирована на поддержку процесса анализа данных и их целостность;

c) высокоуровневые средства отражения информационной модели и описания структуры данных;

d) это установление зависимости дискретной выходной переменной от входных переменных.

Обучающая выборка -

a) эта группировка объектов (наблюдений) на основе данных, описывающих свойства объектов;

b) набор данных, каждая запись которого представляет собой учебный пример, содержащий заданные входы, и соответствующий правильный выходной результат;

c) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Ошибка обучения -

a) это ошибка, допущенная моделью на учебном множестве;

b) это ошибка, полученная на тестовых примерах;

c) имена, типы и назначения полей исходной выборки данных;

d) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданные входы, и соответствующий правильный выходной результат.

Данные, рассматриваемые в каком-либо контексте, из которого пользователь может составить собственное мнение - это

a) данные;

b) знания;

c) информация.

Множество примеров, используемое для проверки работы сконструированной модели, называется

a) тестовым множеством;

b) множеством входных переменных;

c) обучающим множеством;

d) множеством выходных переменных.

Модель (алгоритм) называют обучаемым если

a) модель осуществляет интерактивное взаимодействие с экспертом;

b) модель (алгоритм) самостоятельно обнаруживает в данных присутствующие в них закономерности;

c) модель (алгоритм) самостоятельно использует известные закономерности.

Таблицы в базах данных предназначены для

a) хранения данных базы;

b) отбора и обработки данных базы;

c) автоматического выполнения группы команд;

d) визуализации данных.

Data Mining включает в себя

a) один базовый метод обнаружения знаний;

b) только статистические методы обработки данных для извлечения знаний;

c) большое число различных методов извлечения знаний;

d) большое число статистических методов извлечения знаний.

Целью построения модели регрессии можно назвать

a) прогнозирование числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных;

b) объединение объектов или наблюдений, на основе близости значений их атрибутов (признаков);

c) исследование взаимной связи между объектами и/или событиями;

d) разбиения множества объектов или наблюдений на априорно заданные группы.

Целью задачи классификации можно назвать

a) числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных;

b) объединение объектов или наблюдений, на основе близости значений их атрибутов (признаков);

c) исследование взаимной связи между объектами и/или событиями;

d) разбиения множества объектов или наблюдений на априорно заданные группы.

Целью задачи кластеризации можно назвать

a) числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных;

b) объединение объектов или наблюдений, на основе близости значений их атрибутов (признаков);

c) исследование взаимной связи между объектами и/или событиями;

d) разбиения множества объектов или наблюдений на априорно заданные группы.

Целью формирования ассоциативных правил можно назвать

a) числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных;

b) объединение объектов или наблюдений, на основе близости значений их атрибутов (признаков);

c) установление взаимной связи между объектами и/или событиями;

d) разбиения множества объектов или наблюдений на априорно заданные группы.

К классу прогнозирующих задач Data Mining относится

a) кластеризация;

b) поиск ассоциативных правил;

c) регрессия;

d) Классификация.

Два основных типа данных в статистике

a) качественные и количественные;

b) количественные и символьные;

c) текстовые и числовые;

d) векторы и массивы.

Краткие методические указания

Итоговый тест проводится в электронной форме во время последнего в учебном периоде практического занятия. Тест состоит из 20 тестовых заданий. На выполнение теста отводится 20 минут. Во время проведения теста использование литературы и других информационных ресурсов допускается только по предварительному согласованию с преподавателем.

Шкала оценки

№	Баллы	Описание
5	16–20	Процент правильных ответов от 91% до 100%
4	11–15	Процент правильных ответов от 80 до 90%
3	6–10	Процент правильных ответов от 65 до 79%
2	0–5	Процент правильных ответов от 40 до 64%