

Приложение 2 к РПД Анализ данных
01.03.02 Прикладная математика и информатика
направленность (профиль)
Системное программирование
и компьютерные технологии
Форма обучения – очная
Год набора – 2022

ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

1. Общие сведения

1.	Кафедра	Математики, физики и информационных технологий
2.	Направление подготовки	01.03.02 Прикладная математика и информатика
3.	Направленность (профиль)	Системное программирование и компьютерные технологии
4.	Дисциплина (модуль)	Б1.В.ДВ.06.02 Анализ данных
5.	Форма обучения	Очная
6.	Год набора	2022

2. Перечень компетенций

- | |
|--|
| – УК-1: Способен осуществлять поиск, критический анализ информации, применять системный подход для решения поставленных задач |
|--|

3. Критерии и показатели оценивания компетенций на различных этапах их формирования

Этапы формирования компетенций (разделы, темы дисциплины)	Формируемая компетенция	Критерии и показатели оценивания компетенций			Формы контроля сформированности компетенций
		Знать:	Уметь:	Владеть:	
Раздел 1. Основные понятия технологий сбора и систематизации данных	УК-1	<ul style="list-style-type: none"> – типы шкал, в которых могут быть представлены данные, и ограничения на задачи, которые могут быть решены с учётом этих шкал; – основные понятия Data Science (науки о данных); – принципы мониторинга как информационной технологии получения данных 	<ul style="list-style-type: none"> – выбирать средства анализа, наиболее эффективные для конкретных данных с учётом их природы, погрешности, пространственного и временного разрешения, а также задач исследования; – уметь критически оценивать возможности и ограничения используемых методов 	<ul style="list-style-type: none"> – навыками сбора первичной информации, организации и хранения данных для конкретного исследования; – технологиями поиска данных в сети Интернет и оценки их качества 	Участие в учебной дискуссии, заполнение глоссария
Раздел 2. Предобработка данных, визуализация, первичный статистический анализ	УК-1	<ul style="list-style-type: none"> – основные способы графического и табличного представления данных; – критерии определения аномальных значений в выборочных данных; – основные описательные статистики выборки 	<ul style="list-style-type: none"> – осуществлять дискретизацию непрерывных данных с учётом решаемой задачи; – выбирать наиболее подходящий способ табличного или графического представления данных, исходя из целей исследования; – применять методы первичной обработки данных; – правильно понимать и интерпретировать полученные результаты исследования 	<ul style="list-style-type: none"> – навыками самостоятельного проведения исследований; – основными категориями, понятиями, методами современной описательной статистики; – навыками работы с программными средствами, позволяющими осуществлять статистический анализ; 	Отчет о выполнении лабораторных работ, защита раздела
Раздел 3. Корреляционный и регрессионный анализы	УК-1	<ul style="list-style-type: none"> – основные понятия корреляционного и регрессионного анализа; – причины возникновения ложной корреляции и способы её выявления; – основные правила проверки значимости и интервального оценивания уравнения и коэффициентов регрессии; 	<ul style="list-style-type: none"> – проверять наличие статистически значимой линейной связи между переменными; – правильно понимать и интерпретировать полученные результаты исследования 	<ul style="list-style-type: none"> – методами корреляционного анализа; – навыками получения уравнения регрессии и проверки его статистической значимости; – навыками работы с программными средствами, позволяющими осуществлять корреляционный и регрессионный анализ 	Отчет о выполнении лабораторных работ, защита раздела
Раздел 4. Классификация	УК-1	<ul style="list-style-type: none"> – понятия и методы кластерного и классификационного анализа, – общие свойства и особенности методов машинного обучения с учителем и без учителя 	<ul style="list-style-type: none"> – правильно понимать и интерпретировать полученные результаты исследования; – выбирать метод классификации или кластеризации в зависимости от цели исследования и характера имеющихся данных; – проводить сравнительный анализ различных способов классификации и кластеризации множества объектов с использованием функционалов качества 	<ul style="list-style-type: none"> – основными приёмами и методами классификации и кластерного анализа в зависимости от характера используемой информации; – навыками работы с программными средствами, позволяющими осуществлять кластерный и классификационный анализы 	Отчет о выполнении лабораторных работ, защита раздела
Раздел 5. Кластерный анализ	УК-1	<ul style="list-style-type: none"> – понятия и методы кластерного и классификационного анализа, 	<ul style="list-style-type: none"> – правильно понимать и интерпретировать полученные результаты 	<ul style="list-style-type: none"> – основными приёмами и методами классификации и 	Отчет о выполнении лабораторных работ,

Этапы формирования компетенций (разделы, темы)	Формируемая компетенция	Критерии и показатели оценивания компетенций			Формы контроля сформированности компетенций
		Знать:	Уметь:	Владеть:	
		<ul style="list-style-type: none"> – общие свойства и особенности методов машинного обучения с учителем и без учителя – основные виды расстояний между объектами (метрик пространства) и кластерами, их особенности и возможности применения; – 	<ul style="list-style-type: none"> исследования; – выбирать метод классификации или кластеризации в зависимости от цели исследования и характера имеющихся данных; – проводить сравнительный анализ различных способов классификации и кластеризации множества объектов с использованием функционалов качества 	<ul style="list-style-type: none"> кластерного анализа в зависимости от характера используемой информации; – навыками работы с программными средствами, позволяющими осуществлять кластерный и классификационный, анализы 	защита раздела
Раздел 6. Метод главных компонент	УК-1	<ul style="list-style-type: none"> – эквивалентные формулировки задачи о построении базиса признакового пространства из главных компонент 	<ul style="list-style-type: none"> – интерпретировать главные компоненты в терминах предметной области; – использовать сингулярный анализ для получения главных компонент и оценки объяснённой дисперсии каждой компоненты 	<ul style="list-style-type: none"> – навыками идентификации и интерпретации связи признаков переменных и главных компонент; – – технологиями разложения средствами языка программирования (Python, R и т.д.) 	Отчет о выполнении лабораторных работ, защита раздела
Раздел 7. Анализ и прогнозирование временных рядов	УК-1	<ul style="list-style-type: none"> – основные методы анализа и прогнозирования временных рядов 	<ul style="list-style-type: none"> – использовать модели временных рядов, выполнять их параметрическую идентификацию, оценивать качество аппроксимации реальных данных выбранной моделью; – выполнять декомпозицию временных рядов в рамках аддитивной модели; – выделять гармонические и квазигармонические аддитивные компоненты временных рядов с помощью Фурье-анализа, вейвлет-анализа, сингулярного спектрального анализа и декомпозиции Хуанга; – выполнять анализ и прогнозирование временных рядов с помощью статистических моделей, сингулярного спектрального анализа и нейронных сетей 	<ul style="list-style-type: none"> – навыками работы с программными средствами, позволяющими осуществлять декомпозицию и прогнозирование временных рядов 	Отчет о выполнении лабораторных работ, защита раздела
Раздел 8. Нейросетевое моделирование и глубокое обучение	УК-1	<ul style="list-style-type: none"> – основные архитектуры нейронных сетей – особенности анализа речи, текстов на естественном языке и компьютерного зрения методами глубокого обучения. 	<ul style="list-style-type: none"> – правильно понимать и интерпретировать полученные результаты исследования 	<ul style="list-style-type: none"> – навыками работы с библиотеками, предоставляющими доступ к нейросетевым моделям (TensorFlow, Keras и др.) 	Отчет о выполнении лабораторных работ, защита раздела

Шкала оценивания в рамках балльно-рейтинговой системы:

«неудовлетворительно» – 60 баллов и менее; «удовлетворительно» – 61-80 баллов; «хорошо» – 81-90 баллов; «отлично» – 91-100 баллов

4. Критерии и шкалы оценивания

4.1 Тестирование по разделу дисциплины

Процент правильных ответов	До 60	61-80	81-100
Количество баллов за решенный тест	0	1-3	4-6

4.2 Выполнение и защита лабораторной работы

- 7-8 баллов выставляется, если студент вовремя и полностью выполнил задание на лабораторную работу, правильно и полностью описал и изложил необходимые результаты в отчете, аргументировав их на защите лабораторной работы.
- 5-6 балла выставляется, если студент выполнил задание на лабораторную работу, правильно описал и изложил необходимые результаты в отчете, аргументировав их на защите лабораторной работы, но задержал сдачу работы на одну неделю.
- 3-4 балла выставляется, если студент выполнил задание на лабораторную работу, правильно описал и изложил необходимые результаты в отчете, аргументировав их на защите лабораторной работы, но задержал сдачу работы на две недели.
- 1-2 балла выставляется, если студент выполнил задание на лабораторную работу, описал и изложил необходимые результаты в отчете, аргументировав их на защите лабораторной работы, но задержал сдачу работы более чем три недели.
- 0 баллов - если студент не выполнил задания и/или предоставил отчет.

4.3 Выступление с презентацией (доклад, реферат)

Характеристика выступления с презентацией	количество баллов
Содержание	
Сформулирована цель работы	0,5
Понятны задачи и ход работы	0,5
Информация изложена полно и четко	0,5
Иллюстрации усиливают эффект восприятия текстовой части информации	0,5
Сделаны выводы	0,5
Оформление презентации	
Единый стиль оформления	0,5
Текст легко читается, фон сочетается с текстом и графикой	0,5
Все параметры шрифта хорошо подобраны, размер шрифта оптимальный и одинаковый на всех слайдах	0,5
Ключевые слова в тексте выделены	0,5
Эффект презентации	
Общее впечатление от просмотра презентации	0,5
Мах количество баллов	5

4.4 Разработка и защита проекта

Характеристики работы студента	количество баллов
- студент глубоко и всесторонне усвоил проблему; - уверенно, логично, последовательно и грамотно его излагает; - опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью; - умело обосновывает и аргументирует выдвигаемые им идеи; - делает выводы и обобщения; - свободно владеет понятиями	10
- студент твердо усвоил тему, грамотно и по существу излагает ее, опираясь на знания основной литературы; - не допускает существенных неточностей; - увязывает усвоенные знания с практической деятельностью; - аргументирует научные положения; - делает выводы и обобщения; - владеет системой основных понятий	7

Характеристики работы студента	количество баллов
<ul style="list-style-type: none"> - тема раскрыта недостаточно четко и полно, то есть студент освоил проблему, по существу излагает ее, опираясь на знания только основной литературы; - допускает несущественные ошибки и неточности; - испытывает затруднения в практическом применении знаний; - слабо аргументирует научные положения; - затрудняется в формулировании выводов и обобщений; - частично владеет системой понятий 	3
<ul style="list-style-type: none"> - студент не усвоил значительной части проблемы; - допускает существенные ошибки и неточности при рассмотрении ее; - испытывает трудности в практическом применении знаний; - не может аргументировать научные положения; - не формулирует выводов и обобщений; - не владеет понятийным аппаратом 	0

4.5 Контрольное (экзаменационное) тестирование: балл рассчитывается пропорционально количеству верно решенных дидактически единиц (модулей):

Количество верно решенных ДЕ	0-5	6	7
Количество баллов	По 5 баллов за каждую ДЕ	32	40

5. Типовые контрольные задания и методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

5.1 Типовое контрольное тестовое задание:

1) Необработанный материал, предоставляемый источником и используемый потребителями для формирования на его основе полезного результата:

- a) информация,
- b) данные,
- c) знания.

2) Совокупность фактов, закономерностей и эвристических правил, с помощью которых решается поставленная задача

- a) информация,
- b) данные,
- c) знания.

3) Свойство информации, характеризующее степень ее соответствия настоящему моменту времени

- a) полнота
- b) достоверность
- c) ценность
- d) адекватность
- e) актуальность
- f) доступность

4) Для многомерных данных, предполагающих учёт пространственных координат, признаков переменных и времени часто используется термин «... данных»:

- a) гиперквадрат,
- b) гиперкуб,
- c) гиперсфера,
- d) гипербола.

5) Что из перечисленного не является машинным обучением?

- a) обучение по прецедентам,
- b) обучение с учителем,
- c) обучение без учителя,
- d) глубокое обучение,
- e) обучение с подкреплением,
- f) обучение по контрпримерам.

6) Наиболее распространённым плотностным методом кластеризации является...

- a) метод К-средних,
- b) метод EM,
- c) метод DBSCAN,
- d) метод Кохонена.

7) Какой вид анализа многомерных данных не позволяет перейти к пространству меньшей размерности?

- a) корреляционный анализ,
- b) анализ главных компонент,
- c) факторный анализ,
- d) анализ с использованием карт Кохонена.

8) Какой вид анализа временного ряда использует понятие «мгновенная частота»?

- a) Фурье-анализ,
- b) вейвлет-анализ,
- c) сингулярный спектральный анализ,
- d) анализ эмпирических мод Хуанга.

9) Какая из перечисленных логических функций двух переменных не может быть смоделирована нейроном МакКаллока-Питтса?

- a) OR,
- b) AND,
- c) XOR,
- d) все перечисленные могут быть смоделированы.

10) Какие математические объекты могут быть найдены с помощью аффинитивного анализа?

- a) ассоциативные правила,
- b) кластеры наибольшей мощности,
- c) главные компоненты,
- d) непериодическая последовательность максимальной длины.

Ключ: 1 - b 2 - c 3 - e 4 - b 5 - f 6 - c 7 - a 8 - d 9 - c 10 - a

5.2 Требования к докладу

Требования к оформлению доклада:

- Объем доклада – 5 страниц (без титульного листа и списка источников).
- Титульный лист должен быть оформлен по образцу (имеется файл с образцом).
- Основной текст работы оформлен в соответствии с требованиями, указанными ниже.
- В случае использования в тексте таблиц и/или рисунков на каждый объект должна быть ссылка в тексте работы. Например, «... основные виды программных средств представлены ниже (см. Таблица 1)» или «... схему передачи информации можно увидеть на рис. 1».
- Количество источников должно быть не менее трех, на все должны быть ссылки внутри текста.
- Список используемых источников должен быть оформлен в соответствии с требованиями, указанными ниже.

Для оформления основного текста работы:

- Шрифт – TimesNewRoman, размер – 14 пт.
- Абзац: междустрочный интервал – 1,5; выравнивание – «по ширине»; абзацный отступ – 1,25 см.
- Оформление рисунков (при необходимости): выравнивание рисунка – «по центру», подпись рисунка – «Рис. №. Название рисунка»; шрифт для подписи рисунка – TimesNewRoman, размер – 12 пт.
- Оформление таблиц (при необходимости): выравнивание таблицы – «по центру»; шрифт внутри таблицы – TimesNewRoman, размер – 11-12 пт.; выравнивание текста внутри таблицы – на усмотрение пользователя; подпись таблицы располагается над таблицей и состоит из двух частей: «Таблица №» – выравнивание по правому краю и «Название таблицы» – выравнивание по правому краю или по центру.

Для оформления источников (в соответствии с ГОСТ 2008):

- Источники должны быть расположены в алфавитном порядке и пронумерованы.
- В тексте доклада ссылка на источник выполняется в виде: [№], где № – номер источника в общем списке.

- Если в тексте используется дословная цитата, то она должна быть взята в кавычки, а в ссылке на источник указана страница: [5, с.15].

Примерные темы докладов:

1. Методы сбора данных.
2. Способы хранения данных.
3. Технология OLAP.
4. Данные, представление данных.
5. Информация и знание.
6. Виды и способы измерений.
7. Виды и способы использования шкал.
8. Свойства информации.
9. Свойства знаний.
10. Примеры использования аффинитивного анализа.
11. Примеры использования кластерного анализа.
12. Примеры использования методов классификации.
13. Прикладные задачи классификации.
14. Примеры использования метода построения деревьев решений.
15. Области применения двухслойных персептронов.
16. Глубокое обучение в задачах компьютерного зрения.
17. Глубокое обучение в задачах распознавания речи.
18. Глубокое обучение в задачах анализа текста на естественном языке.
19. Глубокое обучение в задачах прогнозирования валютных и социально-экономических показателей.
20. Задачи анализа регулярных пространственно распределённых данных и методы их решения.

5.3 Вопросы к зачету:

1. Анализ данных: понятия «информация», «данные», «знания».
2. Мониторинг природных систем: определения мониторинга, примеры. Данные мониторинга: классификация по форме представления, классификация по способу получения, этапы контроля, виды научно-технической продукции.
3. Основные принципы построения информационных систем мониторинга.
4. Виды шкал данных. Примеры процедур шкалирования данных предметной области.
5. Представление пространственно-временных данных: гиперкуб данных; суть многомерности данных; временные ряды, карты пространственных распределений, векторы признаков и состояний.
6. Свойства временных рядов: аксиоматические и проверяемые. Многомерный временной ряд гридированных данных в прямоугольной области: формальное описание, примеры.
7. Задачи исследования временных рядов.
8. Аддитивная модель временного ряда. Декомпозиция одномерного временного ряда на аддитивные составляющие: Фурье-анализ и вейвлет-анализ.
9. Аддитивная модель временного ряда. Декомпозиция временного ряда (одномерный и многомерный случаи) на аддитивные составляющие: сингулярный спектральный анализ и декомпозиция на эмпирические моды Хуанга.
10. Прогнозирование временных рядов по результатам сингулярного спектрального анализа (метод «Гусеница»).
11. Статистические модели одномерных временных рядов: AR, MA, ARIMA, ARX, TARX, GARCH.
12. Представление данных векторами линейного пространства признаков. Основные понятия Data Science: open data, big data, data mining, machine learning, supervised learning, unsupervised learning, pattern recognition, text mining.
13. Задача классификации: дерево решений.
14. Задача кластеризации: формулировка, основные понятия.
15. Метод кластеризации K-средних и его модификации.
16. Метод кластеризации Expectation-Maximization. Формула Байеса. Расчёт вероятности для различных законов распределения (Бернулли, биномиального, нормального).
17. Использование нейронных сетей Кохонена для распознавания образов.
18. Иерархические агломеративные и дивизивные методы кластеризации. Дендрограмма.
19. Методы кластеризации на основе нейронных сетей Кохонена: слоя, простой прямоугольной карты, растущей иерархической карты. Способы инициализации весов.
20. Метод кластеризации DBSCAN.
21. Краткая сравнительная характеристика методов и моделей, используемых для кластеризации: K-средние, Expectation-Maximization, карты Кохонена, DBSCAN.

22. Нейронные сети прямого распространения: модель МакКаллока-Питтса, персептрон Розенблатта, многослойный персептрон, функция активации, инициализация Нгуен-Видроу, правило обучения Хебба.
23. Модель МакКаллока-Питтса как простейший линейный классификатор. Моделирование логических функций AND, OR, XOR с помощью нейронов МакКаллока-Питтса. Однослойный и многослойный персептроны как классификаторы.
24. Многослойный персептрон с непрерывной функцией активации. Примеры функций активации. Обучение многослойного персептрона методом обратного распространения ошибки (вывод формул для двухслойного персептрона).
25. Обучение многослойного персептрона: метод Левенберга-Марквардта.
26. Прогнозирование одномерного временного ряда с использованием двухслойного персептрона: план решения задачи.
27. Анализ главных компонент: эквивалентные формулировки задачи, процедура формирования векторов нового базиса, приложения. Подготовка матрицы данных: центрирование, нормирование, стандартизация. Матрицы нагрузок и счётов.
28. Связь метода главных компонент с корреляционным анализом, задачей о собственных числах и собственных векторах матрицы, сингулярным спектральным анализом, картами Кохонена и многослойным персептроном.
29. Общие черты и отличия корреляционного, регрессионного, дискриминантного, факторного и дисперсионного анализов.
30. Методы анализа текстовой информации – Text Mining.