

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Владимирский государственный университет  
имени Александра Григорьевича и Николая Григорьевича Столетовых»  
(ВлГУ)

Институт экономики и туризма

УТВЕРЖДАЮ:



Директор института

Козлов Д.А.

сентября 2023 года

**ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ (СРЕДСТВ) ПО ДИСЦИПЛИНЕ**

**Интеллектуальный анализ данных**

(наименование дисциплины)

**направление подготовки / специальность**

**01.03.05 Статистика**

(код и наименование направления подготовки (специальности))

**направленность (профиль) подготовки**

**«Бизнес – аналитика»**

(наименование направленности (профиля) подготовки)

Владимир, 2023

## 1. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине, в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	
ОПК – 3 Способен осознанно применять методы математической и дескриптивной статистики для анализа количественных данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ, содержательно интерпретировать полученные результаты, готовить статистические материалы для докладов, публикаций и других аналитических материалов	ОПК-3.1 Знает современный статистический и математический инструментарий для решения профессиональных задач	Знает современный статистический и математический инструментарий для решения профессиональных задач, используемы во многомерных статистических методах Умеет применять на практике статистический инструментарий Владеет методами многомерного статистического анализа для решения профессиональных задач	Тестовые вопросы Ситуационные задачи Практико-ориентированное задание Эссе
	ОПК-3.2 Умеет использовать информационно-коммуникационные технологии и программные средства для анализа количественных данных	Знает основные программные средства, необходимые для применения методов многомерного статистического анализа Умеет пользоваться информационными технологиями для анализа данных Владеет основными программными средствами для анализа количественных данных при использовании многомерных статистических методов	
	ОПК-3.3 Владеет навыками интерпретации полученных результатов анализа количественных данных и подготовки материалов для докладов, публикаций и других аналитических материалов	Знает способы интерпретации и представления полученных данных Умеет анализировать имеющиеся данные и интерпретировать их Владеет навыками интерпретации полученных результатов анализа количественных данных и подготовки материалов для докладов, публикаций и других аналитических материалов	

ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-4.1 Знает принципы работы современных информационных технологий	Знает основные информационные технологии для осуществления интеллектуального анализа данных Умеет использовать информационные технологии для интеллектуального анализа данных Владеет информационными технологиями для решения профессиональных задач	Тестовые вопросы Ситуационные задачи Практико-ориентированное задание Эссе
	ОПК-4.2 Умеет выбирать информационные технологии	Знает критерии выбора информационных технологий Умеет выбирать информационные технологии Владеет критериями отбора информационных технологий для проведения интеллектуального анализа данных	
	ОПК-4.3 Владеет навыками использования современных информационных технологий при решении задач профессиональной деятельности	Знает специальные и прикладные программы умеет применять информационные технологии на практике Владеет навыками использования современных информационных технологий при решении задач профессиональной деятельности	

## 2. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ ПО ДИСЦИПЛИНЕ

### Рейтинг-контроль №1

**1. Какую часть мирового рынка Data Mining занимают услуги или консультации по эффективному внедрению этой технологии для решения актуальных бизнес-задач:**

- а) более 75% ;
- б) около половины;
- в) менее 10% рынка.

**2. В основу программного продукта Cognos 4Thought положена технология:**

- а) множественной регрессии;
- б) нейронных сетей;
- в) деревьев решений.

**3. Достаточно высокая стоимость, невозможность добавлять свои функции, сложность подготовки данных, практическое отсутствие в интерфейсе терминов предметной области – это слабые стороны:**

- а) адаптация программного обеспечения под конкретную задачу;
- б) заказ готового решения у фирмы-разработчика;
- в) готового программного обеспечения.

**4. MAP-сплайны в системе STATISTICA – это:**

а) непараметрическая процедура, в работе которой не используется никаких предположений об общем виде функциональных связей между зависимыми и независимыми переменными;

б) процедура, опирающаяся на предположения о типе и накладывающая ограничения на класс зависимостей;

в) параметрическая процедура, основанная на предположениях о виде функциональных связей между зависимыми и независимыми переменными.

**5. Готовые алгоритмы, полная конфиденциальность информации, техническая поддержка производителя, общение с другими пользователями пакета — это преимущества использования:**

- а) адаптация программного обеспечения под конкретную задачу;
- б) заказ готового решения у фирмы-разработчика
- в) готового программного обеспечения.

**6. Cognos 4Thought предназначен для:**

- а) прогнозирования;
- б) моделирования;
- в) оба варианта верны;
- г) нет правильного ответа.

**7. Какое решение в большей мере требует наличия высококвалифицированных специалистов при внедрении и использования инструмента Data Mining:**

- а) адаптация программного обеспечения под конкретную задачу;
- б) заказ готового решения у фирмы-разработчика;
- в) использование готового программного обеспечения.

**8. Рабочее пространство STATISTICA Data Miner не включает такого элемента:**

- а) тестирование;
- б) анализ данных, моделирование;
- в) подготовка, преобразования и очистка данных.

**9. Вариант использования адаптированного программного обеспечения Data Mining:**

- а) имеет неоспоримые преимущества перед использованием готового программного обеспечения;
- б) всегда проигрывает перед использованием готового программного обеспечения;
- в) имеет как сильные, так и слабые стороны.

**10. Инструментальное средство для оперативного анализа данных и формирования отчетов по OLAP-технологии:**

- а) Cognos Impromptu;
- б) Cognos PowerPlay;
- в) Cognos Scenario.

**11. На этапе подготовки данных:**

- а) специалисты компании Разработчика подготавливают данные для их дальнейшего анализа;
- б) специалисты компании Разработчика и Заказчика подготавливают данные для их дальнейшего анализа;
- в) специалисты компании Заказчика подготавливают данные для их дальнейшего анализа.

**12. На каких этапах 4Thought поддерживает анализ данных:**

- а) сбор данных;
- б) преобразование данных;
- в) интерпретация модели;
- г) на всех этапах.

**13. Преимуществом использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является:**

- а) наличие терминов предметной области;
- б) сложность подготовки данных;
- в) полная конфиденциальность информации.

**14. Охарактеризуйте систему Cognos Scenario:**

- а) является средством оперативного анализа данных;
- б) формирует отчеты по OLAP-технологии;
- в) интеллектуальное инструментальное средство поиска данных.

**15. Преимуществом использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является:**

- а) адаптированность;
- б) не требуется дописывать программный код;
- в) сложность подготовки данных.

**16. Охарактеризуйте систему Cognos Scenario:**

- а) формирует отчеты по OLAP-технологии;
- б) позволяет руководителям выявлять скрытые тенденции и модели бизнеса;
- в) является средством оперативного анализа данных.

**17. На этапе первичного исследования данных:**

- а) всю работу осуществляет заказчик;
- б) со стороны заказчика может потребоваться максимальное участие;
- в) со стороны заказчика может потребоваться лишь минимальное участие.

18. Модуль Oracle Data Mining доступен из таких редакций:

- а) Personal Edition;
- б) Enterprise Edition;
- в) OneStandard Edition.

**19. Постановка бизнес-задачи – это этап, который:**

- а) формулирует конкретные бизнес-задачи, и они уже не могут быть изменены;
- б) формулирует конкретные бизнес-задачи, и они могут быть изменены в ходе прохождения именно этого цикла;
- в) формулирует конкретные бизнес-задачи, и они не могут быть изменены в ходе прохождения именно этого цикла.

**20. Архитектура хранилища типа «звезда» в Deductor называется:**

- а) сценарием;
- б) процессом;
- в) проектом.

#### Рейтинг-контроль №2

**1. «Извлечение полезных сведений невозможно без хорошего понимания сути данных», верно ли утверждение:**

- а) верно;
- б) неверно. Технологии не нужно понимание данных.
- в) неверно. Технологии Data Mining не нужен аналитик, поэтому понимание кем либо данных — излишне.

**2. Большинство аналитических методов, используемые в технологии Data mining**

– это:

- а) новейшие математические алгоритмы и методы;
- б) известные математические алгоритмы и методы;
- в) классические статистические методы.

**3. Выберите характеристику, наиболее подходящую для Data Mining:**

- а) подходит для понимания ретроспективных данных;
- б) подходит для обобщения ретроспективных данных;
- в) опирается на ретроспективные данные для получения ответов на вопросы о будущем.

**4. Частью какой из перечисленных стадий является валидация закономерностей:**

- а) свободный поиск;
- б) анализ исключений;
- в) прогностическое моделирование.

**5. Какая из перечисленных ниже групп методов достаточно часто использует для выявления взаимосвязей в данных концепцию усреднения по выборке:**

- а) Data Mining;
- б) OLAP;
- в) статистические методы.

**6. На стадии свободного поиска осуществляется:**

- а) использование выявленных закономерностей для предсказания неизвестных значений;
- б) выявление закономерностей;
- в) анализ исключений.

**7. В результате использования инструментов Data Mining пользователь может:**

- а) получить гипотезы о взаимосвязях в данных, самостоятельно выдвинутые инструментом Data Mining;
- б) получить подтверждение или опровержение гипотез, выдвинутых пользователем;
- в) оба варианта верны;
- г) нет верного ответа.

**8. Нейронные сети относятся к группам:**

- а) методов на основе уравнений;
- б) статистических методов;
- в) методов кросс-табуляции.

**9. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:**

- а) быть очевидными;
- б) чем больше найдено закономерностей, тем лучше;
- в) быть неочевидными.

**10. Какой из перечисленных ниже пунктов не является названием стадии Data Mining:**

- а) свободный поиск;
- б) индукция правил;
- в) анализ исключений.

**11. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:**

- а) быть объективными;
- б) быть очевидными;
- в) чем больше найдено закономерностей, тем лучше.

**12. Какие из перечисленных ниже пунктов являются названиями стадий Data Mining:**

- а) прогностическое моделирование;
- б) свободный поиск;
- в) оба варианта верны;
- г) нет верного ответа.

**13. Инструменты Data Mining:**

- а) могут самостоятельно строить гипотезы о взаимосвязях в данных;
- б) могут самостоятельно строить гипотезы о взаимосвязях в данных, которые обязательно подтверждаются;
- в) не могут самостоятельно строить гипотезы о взаимосвязях в данных.

**14. Для какой шкалы применимы только такие операции как равно, не равно, больше, меньше:**

- а) номинальная шкала;
- б) интервальная шкала;
- в) порядковая шкала.

**15. Если сравнивать Data Mining, машинное обучение и статистику, какая из дисциплин сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов:**

- а) Data Mining;
- б) статистика;
- в) машинное обучение.



**Рейтинг-контроль №3**

**1. К какой категории данных относится вес измеряемых объектов:**

- а) дискретным данным;
- б) непрерывным данным;
- в) оба варианта верны;
- г) нет верного ответа.

**2. Назовите фактор, обусловивший возникновение и развитие Data Mining:**

- а) совершенствование аппаратного и программного обеспечения;
- б) совершенствование технологий хранения и записи данных;
- в) оба варианта верны;
- г) нет верного ответа.

**3. Для какой шкалы применимы только такие операции как равно и не равно:**

- а) порядковая шкала;
- б) номинальная шкала;
- в) интервальная шкала.

**4. В процессе работы Data Mining программы пользователь может получить такие результаты:**

- а) только статистически достоверные результаты;
- б) только верные результаты, ложные выводы исключены;
- в) большой процент ложных, недостоверных или бессмысленных результатов.

**5. Такие данные как температура воздуха относятся к:**

- а) дискретным данным;
- б) непрерывным данным;
- в) оба варианта верны;
- г) нет верного ответа.

**6. Data Mining — это процесс обнаружения в сырых данных знаний, необходимых для:**

- а) принятия решений в различных сферах человеческой деятельности;
- б) увеличения стоимости анализа данных;
- в) замены аналитика в процессе принятия решений.

**7. Объект описывается как:**

- а) свойство, характеризующее объект;
- б) набор атрибутов;
- в) поле таблицы.

**8. Подготовка данных в процессе Data Mining является:**

- а) необязательным этапом работы;
- б) может вообще отсутствовать;
- в) существенным этапом работы.

**9. Свойство, характеризующее объект:**

- а) Данные;
- б) Атрибут;
- в) Инструменты Data Mining.

**10. Какая из перечисленных дисциплин более сосредоточена на теории проверки гипотез:**

- а) Data Mining;
- б) визуализация;
- в) статистика.

**11. Процесс обнаружения в сырых данных знаний, необходимых для принятия решений в различных сферах человеческой деятельности:**

- а) Данные;
- б) Data Mining;
- в) Атрибут.

**12. Data Mining — это процесс обнаружения в сырых данных:**

- а) практических закономерностей;
- б) большого количества закономерностей;
- в) ранее сформулированных гипотез.

**13. Номинальная шкала – это шкала:**

- а) содержащая категории, которые могут упорядочиваться;
- б) содержащая только две категории;
- в) содержащая только категории, которые не могут упорядочиваться.

**14. Data Mining — это процесс обнаружения в сырых данных:**

- а) объективных закономерностей;
- б) ранее сформулированных гипотез;
- в) большого количества закономерностей.

**15. Объектом не является:**

- а) строка таблицы;
- б) переменная;
- в) запись.

**Иные оценочные материалы для проведения текущего контроля успеваемости****Практические задачи**

**Задача 1.** Изучение опыта применения методов кластеризации данных.

**Задача 2.** Программирование методов кластеризации данных.

**Задача 3.** Лингвистическое резюмирование результатов кластеризации данных.

**Задача 4.** Прогнозирование на основе статистического подхода.

**Задача 5.** Прогнозирование на основе нечеткого подхода.

**Тематика докладов**

1. Дескриптивный анализ данных.
2. Полный статистический дескрипт вероятностной структуры и параметров данных.
3. Полиномиальные и стохастические модели. Оценка параметров движения.
4. Моделирование динамических стохастических процессов в среде Матлаб (R).
5. Регрессионная оценка параметров движения.
6. Метод статистических испытаний.
7. Имитационное моделирование случайных событий и процессов. Метод Монте-Карло.
8. Многомерные задачи классификации и распознавания.
9. Основные технологии статистической классификации многомерных случайных объектов средствами ИАД.
10. Прогнозирование на основе фильтра Калмана.
11. Прогнозирование состояния сложных динамических систем статистическими и другими средствами ИАД.
12. ИНС с обратным распространением ошибки.
13. Моделирование двухслойной ИНС с обратным распространением ошибки и применение в задачах распознавания.
14. Модели и прогнозирование хаотических процессов.
15. Моделирование хаотических процессов средствами ИАД.
16. Анализ возможности построения прогноза в хаотических средах.

**Тематика презентаций**

1. Пакеты NumPy, Scipy, математические операции в них.
2. Пакет Pandas, работа с данными в нем.
3. Линейные методы классификации и регрессии.

4. Метрики качества алгоритмов машинного обучения, кросс-валидация.
5. Деревья решений, их построение.
6. Композиции алгоритмов. Случайные леса.
7. Работа с реальными данными. Предобработка признаков.
8. Кластеризация реальных данных.
9. Поиск частых множеств и ассоциативных правил

### **3. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ**

#### **Вопросы к экзамену**

1. Введение в машинное обучение и анализ данных.
2. Анализ данных в различных прикладных областях.
3. Основные определения.
4. Этапы анализа данных.
5. Постановки задач машинного обучения.
6. Примеры прикладных задач и их типы.
7. Линейная алгебра и анализ данных.
8. Линейные пространства, их примеры из машинного обучения.
9. Коллинеарность и линейная независимость.
10. Скалярное произведение, косинус угла, примеры их применения.
11. Векторы и матрицы, операции над ними.
12. Матричное умножение.
13. Системы линейных уравнений.
14. Обратная матрица.
15. Математический анализ и анализ данных (на примере парной линейной регрессии и МНК).
16. Производная и градиент, их свойства и интерпретации.
16. Типы функций: непрерывные, разрывные, гладкие.
17. Градиентный спуск.
18. Выпуклые функции и их особое место в оптимизации.
19. Теория вероятностей и анализ данных.
20. Случайные величины.
21. Дискретные и непрерывные распределения, их свойства.
22. Примеры распределений и их важность в анализе данных: биномиальное, пуассоновское, нормальное, экспоненциальное.

23. Характеристики распределений: среднее, медиана, дисперсия, квантили.
24. Пример их использования при генерации признаков.
25. Центральная предельная теорема.
26. Математическая статистика и анализ данных.
27. Оценивание параметров распределений.
28. Метод максимального правдоподобия.
29. Доверительные интервалы и бутстрэппинг.
30. Линейная регрессия.
31. Квадратичная функция потерь и предположение о нормальном распределении шума.
32. Метод наименьших квадратов: аналитическое решение и оптимизационный подход.
33. Стохастический градиентный спуск.
34. Тонкости градиентного спуска: размер шага, начальное приближение, нормировка признаков.
35. Проблема переобучения.
36. Регуляризация.
37. Линейная классификация.
38. Аппроксимация дискретной функции потерь.
39. Логистическая регрессия.
40. Регрессия: квадратичные и абсолютные потери, абсолютные логарифмические отклонения.
41. Примеры использования.
42. Классификация: доля верных ответов, ее недостатки.
43. Точность и полнота, их объединение: арифметическое среднее, минимум, гармоническое среднее (F-мера).
44. Оценки принадлежности классам: площади под кривыми. AUC-ROC, AUC-PRC, их свойства.
45. Оценивание качества алгоритмов.
46. Отложенная выборка, ее недостатки.
47. Оценка полного скользящего контроля.
48. Кросс-валидация. Leave-one-out.
49. Практические особенности кросс-валидации.
50. Стратификация.
51. Потенциальные проблемы с разбиением зависимой или динамической выборки.
52. Логические методы и их интерпретируемость.

53. Деревья решений. Проблема построения оптимального дерева решений.
54. Жадный алгоритм, основные его параметры.
55. Построение деревьев решений. Критерий ветвления.
56. Выбор оптимального разбиения в задачах регрессии. Сложности выбора разбиения в задаче классификации.
57. Примеры критериев: энтропийный (прирост информации),
58. Джини и их модификации.
59. Критерии завершения построения.
60. Регуляризация и стрижка деревьев.
61. Уменьшение дисперсии при усреднении алгоритмов методом бутстреп.
62. Блендинг алгоритмов.
63. Понятие смещения и разброса (иллюстрация на примере линейных методов и решающих деревьев).
64. Уменьшение разброса с помощью усреднения.
65. Случайный лес.
66. Оценка out-of-bag.
67. Шумы и выбросы в данных.
68. Методы поиска выбросов.
69. Пропуски в данных, методы их восстановления.
70. Несбалансированные выборки: проблемы и методы борьбы.
71. Задача отбора признаков, примеры подходов.
72. Задача анализа потребительской корзины.
73. Поддержка и достоверность.
74. Частые, замкнутые и максимальные частые множества.
75. Алгоритм Априори.
76. Меры “интересности правил”.
77. Алгоритм K-Means.
78. Проблема устойчивости результатов и важность грамотной инициализации, алгоритм K-Means++.
79. Выбор числа кластеров.
80. Оценка качества кластеризации.
81. Нейронные сети. Типичные задачи.

82. Алгоритм обратного распространения ошибки.
83. Блоки нейронной сети.
84. Архитектуры современных нейронных сетей.
85. Типы нейронных сетей для различных видов данных.
86. Нейронные сети для анализа изображений и видео

### Критерии оценки тестирования студентов

Оценка выполнения тестов	Критерий оценки
0,5 балла за правильный ответ на 1 вопрос	Правильно выбранный вариант ответа (в случае закрытого теста), правильно вписанный ответ (в случае открытого теста)

#### Регламент проведения тестирования и оценивания

№	Вид работы	Продолжительность
1.	Предел длительности тестирования (20 вопросов)	35-40 мин.
2.	Внесение исправлений	до 5 мин.
	Итого (в расчете на тест)	до 45 мин.

### Критерии оценки выполнения заданий студентами

#### Регламент выполнения заданий

№	Вид работы	Продолжительность
1.	Предел длительности защиты задания	до 5-7 мин.
2.	Внесение исправлений в представленное решение	до 2 мин.
3.	Комментарии преподавателя	до 1 мин.
	Итого (в расчете на одно задание)	до 10 мин.

Оценка в баллах	Критерии оценивания задания
<b>15 баллов</b>	Задание выполнены полностью, все элементы и взаимосвязи модели (проекта) обоснованы.
<b>10 баллов</b>	Задание выполнены полностью, но нет достаточного обоснования взаимосвязей, элементов модели (проекта)
<b>5 баллов</b>	Модели (проекты) имеют незаконченную структуру. Обоснование модели (проекта) дано частично.
<b>0 баллов</b>	Задание не выполнено.

### Критерии оценки устных ответов студентов

#### Регламент проведения устного опроса

№	Вид работы	Продолжительность
1.	Предел длительности ответа на каждый вопрос	до 3 мин.
2.	Внесение студентами уточнений и дополнений	до 1 мин.
3.	Дискуссия с участием учебной группы по ответу на вопрос	до 2 мин.
4.	Комментарии преподавателя	до 1 мин.
	Итого продолжительность устного ответа (на один) вопрос)	до 7 мин.

Оценка в баллах	Критерии оценивания ответа
5	Ответ отличается последовательностью, полнотой, логикой изложения. Легко воспринимается аудиторией. При ответе на вопросы выступающий демонстрирует глубину владения материалом. Ответы формулируются аргументировано, обосновывается собственная позиция в проблемных ситуациях.
4	Ответ отличается последовательностью, логикой изложения. Но обоснование сделанных выводов не достаточно аргументировано. Неполно раскрыто содержание проблемы.
3	Ответ направлен на пересказ содержания проблемы, но не демонстрирует умение выделять главное, существенное. Выступающий не владеет пониманием сути излагаемой проблемы

### Критерии оценки участия в дискуссии

В целях закрепления практического материала и углубления теоретических знаний по разделам дисциплины предполагается проведение обсуждений в форме дискуссий по актуальным темам, вопросам, что позволяет углубить процесс познания, раскрыть понимание прикладной значимости осваиваемой дисциплины.

Критерии	Оценка в баллах
Демонстрирует полное понимание обсуждаемой проблемы, высказывает собственное суждение по вопросу, аргументировано отвечает на вопросы участников дискуссии, соблюдает регламент выступления.	1
Понимает суть рассматриваемой проблемы, может высказать типовое суждение по вопросу, отвечает на вопросы участников семинара, однако выступление носит затянутый или не аргументированный характер.	0,5
Принимает участие в обсуждении, однако собственного мнения по вопросу не высказывает, либо высказывает мнение, не отличающееся от мнения других докладчиков.	0,2
Не принимает участия в обсуждении	0



**Показатели, критерии и шкала оценивания компетенций промежуточной аттестации знаний по учебной дисциплине «Интеллектуальный анализ данных» на экзамене.**

<b>Оценка в баллах</b>	<b>Оценка за ответ на зачете</b>	<b>Критерии оценивания компетенций</b>	<b>Уровень освоения компетенций</b>
91 -100 Баллов	«Отлично»	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Свободно ориентируется в учебной и профессиональной литературе.	Высокий
76 – 90 баллов	«Хорошо»	Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе.	Хороший
61 – 75 баллов	«Удовлетворительно»	Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине.	Достаточный
0 – 60 баллов	«Неудовлетворительно»	Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает серьезные	Компетенции не сформированы

		затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине.	
--	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

#### 4. ИТОГОВЫЕ ТЕСТОВЫЕ ЗАДАНИЯ ПО ДИСЦИПЛИНЕ

№ п/п	Контролируемые разделы (темы)	Тестовые задания	Код контролируемой компетенции (или ее части)
1	Основные понятия анализа данных	<p><b>1. Data Mining</b> — это процесс обнаружения в сырых данных</p> <p>а) ранее сформулированных гипотез  б) неочевидных закономерностей  в) практических закономерностей  г) объективных закономерностей  д) большого количества закономерностей</p> <p><b>2. В методе опорных векторов для классификации используется ...</b></p> <p>а) не все множество образцов, а лишь их небольшая часть, которая находится на границах  б) все множество образцов  в) часть образцов, которая находится вне границ</p>	<b>ОПК - 3</b>
2	Математические объекты и методы в анализе данных	<p><b>3. Рассматривается модель линейной регрессии</b>  <math>Y_i = \phi(X_i, \theta_1, \dots, \theta_p) + \varepsilon_i, i = 1, \dots, n</math>, где <math>\varepsilon_i</math> - ненаблюдаемые центрированные погрешности, имеющие плотность распределения <math>f(x)</math>. Для оценивания неизвестных параметров <math>\theta_1, \dots, \theta_p</math> применен метод наименьших квадратов (МНК). От чего зависят величины дисперсий <math>D\theta_1, \dots, D\theta_p</math>, полученных МНК-оценок?</p> <p><b>4. Наблюдения</b> <math>X_{ij}, j = 1, 2, 3; i = 1, \dots, n_j</math> описываются моделью следующего вида  <math>X_{ij} = \mu + \tau_j + \varepsilon_{ij}</math>, где <math>\mu</math> - неизвестное общее среднее, <math>\tau_j</math> - отклонение от среднего, вызванное изменением уровня факторной переменной, <math>\varepsilon_{ij}</math> - погрешности с нулевым математическим ожиданием. Контраст <math>\theta</math> параметров <math>\tau</math> в этой модели задан</p> $\theta = \sum_{j=1}^3 c_j \tau_j$ <p>следующим образом, где <math>c_1 = 1, c_2 = -1, c_3 = 0</math>. Что характеризует определенный таким образом контраст?</p>	<b>ОПК - 4</b>
3	Линейная регрессия и классификация	<p><b>5. Известны парные коэффициенты корреляции случайных величин <math>X, Y, Z</math>:</b>  <math>\rho_{xy} = 0.3, \rho_{xz} = 0.4, \rho_{yz} = -0.5</math> Частный</p>	<b>ОПК - 4</b>

		<p>коэффициент корреляции случайных величин <math>X</math> и <math>Y</math> при фиксированном значении <math>Z</math> будет</p> <p>а) положительным б) отрицательным в) равным нулю</p> <p><b>6. Фактором в задаче однофакторного дисперсионного анализа называют</b></p> <p>а) значение измеряемого признака б) переменную, которая влияет на значение измеряемого признака в) погрешность измеряемого признака</p>	
4	Оценивание качества алгоритмов	<p><b>7. Задача однофакторного дисперсионного анализа является обобщением задачи проверки гипотезы об однородности двух выборок против альтернативы о том, что</b></p> <p><b>8. Переменная <math>A</math> измеряется в номинальной шкале и имеет 5 градаций, переменная <math>B</math> измеряется в номинальной шкале и имеет 2 градации. Для того чтобы выяснить, являются ли переменные <math>A</math> и <math>B</math> зависимыми, применяют критерий хи-квадрат. Какое число степеней свободы будет иметь статистика хи-квадрат в случае справедливости основной гипотезы?</b></p>	ОПК – 4
5	Логические методы	<p><b>9. Специалист по анализу данных, который имеет, как минимум, основы статистических знаний и способен применять технологии Data Mining, а также интерпретировать полученные результаты - это...</b></p> <p>а) специалист по добыче данных б) специалист предметной области в) администратор баз данных г) программист д) все ответы неверны</p>	ОПК – 3
6	Композиции алгоритмов	<p><b>10. Назовите метод, недостаток которого приведен ниже: "Существует сложность выбора меры "близости", от этой меры главным образом зависит объем множества записей, которые нужно хранить в памяти для достижения удовлетворительной классификации или прогноза"</b></p> <p><b>11. При ... для каждого обучающего входного примера требуется знание правильного ответа или функции оценки качества ответа</b></p>	ОПК – 3
7	Особенности реальных данных	<p><b>12. Продолжите фразу: "Прогнозирование будет иметь смысл, если горизонт прогнозирования ..."</b></p>	ОПК – 3
8	Анализ частых множеств признаков и ассоциативных правил	<p><b>13. Достоверность ассоциативного правила определяет...</b></p> <p>а) количество транзакций, содержащих определенный набор данных б) какая вероятность того, что из события <math>A</math> следует событие <math>B</math> в) процент транзакций, содержащих определенный набор данных</p>	ОПК – 3
9	Кластеризация данных	<p><b>14. Закончите фразу. При применении кластерного анализа переменные ...</b></p> <p><b>15. Закончите фразу. Спорный объект кластеризации — это объект, который по мере сходства ...</b></p> <p><b>16. При использовании какого метода необходимо задавать количество кластеров?</b></p> <p><b>17. Назовите характеристики кластерного анализа:</b></p> <p><b>18. Перечислите недостатки алгоритма быстрой кластеризации</b></p>	ОПК – 3
10	Нейронные сети	<p><b>19. Заполните пропуски в формулировке: "... - выходная связь нейрона, с которой сигнал (возбуждения или торможения) поступает на... следующих нейронов"</b></p>	ОПК – 3

		<p>20. Группа синапсов нейрона – это ...</p> <p>21. Как называются сети без обратных связей?</p> <p>22. Что необходимо сделать, если на данной обучающей паре символ персептрона не совпадает с нужным ответом?</p> <p>23. В алгоритме обратного распространения при "проходе вперед" ... (закончите фразу).</p> <p>24. В чем заключается метод аккредитации?</p>	
--	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

### Ключ

1	2	3	4	5	6
б), в), г)	а)	а) вид матрицы плана, б) дисперсии погрешностей	разность средних значений откликов, соответствующих первому и второму уровням фактора	а)	б)
7	8	9	10	11	12
Параметром сдвига	4	а)	метод "к-ближайших соседей"	"обучении с учителем"	не меньше, чем время, которое необходимо для реализации решения, принятого на основе прогноза
13	14	15	16	17	18
б)	должны измеряться в сравнимых шкалах	может быть отнесен к нескольким кластерам	метод к-средних	а) позволяет сокращать размерность данных, б) позволяет делать данные более наглядными, в) имеет в своем арсенале около сотни алгоритмов	а) чувствительность к выбросам, б) алгоритм может медленно работать на больших базах данных
19	20	21	22	23	24
аксон, синапсы	однаправленные входные связи, соединенные с выходами других нейронов	персептрон	нужно изменять все весовые значения между одновременно активными нейронами	а) вычисляется ошибка сети, б) корректируются весовые значения сети	активировании лишь одного нейрона Кохонена, имеющего наибольшее значение NET

### Критерии оценки

Оценка в баллах	Оценка за итоговый тест
65-80 баллов	«Отлично»
50-64 баллов	«Хорошо»
40-49 баллов	«Удовлетворительно»
Менее 40 баллов	«Неудовлетворительно»

Разработчик: к.ф-м.н., доц. Крылов В.Е.

Фонд оценочных материалов (средств) рассмотрен и одобрен на заседании кафедры «Бизнес-информатика и экономика»

Протокол № 1 от 30.08.2023 года

Заведующий кафедрой д.э.н., профессор Тесленко И.Б.

Фонд оценочных материалов (средств) рассмотрен и одобрен на заседании учебно-методической комиссии направления 01.03.05 Статистика

Протокол № 1 от 05.09.2023 года

Председатель комиссии к.э.н., доцент Ярьес О.Б.