

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет - ЭКОНОМИЧЕСКИЙ

УТВЕРЖДАЮ:
Проректор по учебной работе,
качеству образования - первый
проректор
Г.А. Хагуров
(подпись)
« 31 » мая 2024 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

**Б1.В.04 АНАЛИЗ ДАННЫХ В ПРОФЕССИОНАЛЬНОЙ
СФЕРЕ**

Направление подготовки/специальность 38.03.05 Бизнес-информатика

Направленность (профиль) / специализация Бизнес в цифровой экономике

Форма обучения очная

Квалификация бакалавр

Краснодар 2024

Рабочая программа дисциплины "Анализ данных в профессиональной сфере" составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки/ специальности 38.03.05 – Бизнес-информатика

Программу составил(и):

И.В. Ариничев

доцент кафедры теоретической экономики ФГБОУ ВО "КубГУ",
кандидат экономических наук

подпись

Рабочая программа дисциплины "Анализ данных в профессиональной сфере" утверждена на заседании кафедры теоретической экономики протокол № 9 «02» апреля 2024 г.

Заведующий кафедрой теоретической экономики Сидоров В.А.

подпись

Утверждена на заседании учебно-методической комиссии экономического факультета протокол № 9 «14» мая 2024 г.

Председатель УМК факультета Дробышевская Л.Н.

подпись

Рецензенты:

Ложников М.М., директор ООО "Бейсик систем"

Калайдина Г.В., к. ф-м. н,

доцент кафедры анализа данных и искусственного интеллекта
ФГБОУ ВО "КубГУ"

1 Цели и задачи изучения дисциплины (модуля)

1.1 Цель освоения дисциплины

Цель освоения дисциплины Анализ данных в профессиональной сфере состоит в формировании знаний, умений и навыков (компетенций) по одному из приоритетных в современных информационных технологиях направлению - интеллектуальной обработке данных.

1.2 Задачи дисциплины

1. ознакомление бакалавров с основными принципами интеллектуального анализа данных - а именно, видами задач анализа данных, классами моделей (линейные, логические, нейросетевые), метриками качества и подходами к предварительной обработке данных;

2. формирование у бакалавров практических навыков сбора и обработки данных для решения социально-экономических задач;

3. формирование у бакалавров представления о технических и методологических средствах анализа данных, обеспечивающих хранение и управление больших объемов данных.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Б1.В.04 «Анализ данных в профессиональной сфере» относится к части, формируемой участниками образовательных отношений Блока 1 "Дисциплины (модули)" учебного плана. В соответствии с рабочим учебным планом дисциплина изучается на 3 курсе по очной форме обучения. Вид промежуточной аттестации: экзамен.

Перечень предшествующих дисциплин, необходимых для ее изучения:

- Математика;
- Базы данных;
- Информационно-коммуникационные технологии в профессиональной деятельности;
- Методы оптимальных решений;

Перечень последующих дисциплин, для которых данная дисциплина является предшествующей в соответствии с учебным планом:

- Эконометрика;
- Инструменты и технологии бизнес-аналитики;

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

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора* достижения компетенции	Результаты обучения по дисциплине
ПК-3 Способен обосновывать решения в профессиональной деятельности	
ИПК-3.2. Анализирует данные для целей обоснования принятия решений в профессиональной сфере	<i>Знает:</i> основные типы задач анализа данных и подходы к их решению; <i>Знает:</i> методы визуализации и представления результатов анализа данных. <i>Знает:</i> современные методы и инструментальные средства анализа данных;
	<i>Умеет:</i> проводить сравнительный анализ методов и инструментальных средств анализа данных; <i>Умеет:</i> планировать аналитические работы с использованием технологий анализа данных;

Код и наименование индикатора* достижения компетенции	Результаты обучения по дисциплине
	<p><i>Трудовое действие:</i> выбор методов и инструментальных средств анализа данных для проведения аналитических работ;</p> <p><i>Трудовое действие:</i> Сбор, очистка и предварительная обработка массивов данных для последующего анализа.</p>

Результаты обучения по дисциплине достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зачетные единицы (144 часа), их распределение по видам работ представлено в таблице

Виды работ	Всего часов	Форма обучения			
		очная		очно-заочная	заочная
		5 семестр (часы)	5 семестр (часы)	X семестр (часы)	X курс (часы)
Контактная работа, в том числе:		58,3	58,3		
Аудиторные занятия (всего):		52	52		
занятия лекционного типа		18	18		
лабораторные занятия		34	34		
практические занятия					
семинарские занятия					
Иная контактная работа:		6,3	6,3		
Контроль самостоятельной работы (КСР)		6	6		
Промежуточная аттестация (ИКР)		0,3	0,3		
Самостоятельная работа, в том числе:		59	59		
Самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)		59	59		
Контроль:		26,7	26,7		
Подготовка к экзамену		26,7	26,7		
Общая трудоёмкость	час.	144	144		
	в том числе контактная работа	58,3	58,3		
	зач. ед	4	4		

2.2 Содержание дисциплины

Распределение видов учебной работы и их трудоёмкости по разделам дисциплины. Разделы (темы) дисциплины, изучаемые в 5 семестре (*очная форма обучения*)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1.	Введение в аналитику данных		2		4	8
2.	Подготовка данных для анализа		2		4	8
3.	Визуализация данных		2		4	8
4.	Введение в машинное обучение		2		4	8
5.	Задача классификации. Метрические методы. Логические методы.		4		6	8
6.	Задачи регрессии. Линейные модели. Введение в нейронные сети.		4		6	8
7.	Обучение без учителя.		2		6	11
	ИТОГО по разделам дисциплины	111	18		34	59
	Контроль самостоятельной работы (КСР)	6				
	Промежуточная аттестация (ИКР)	0,3				
	Подготовка к текущему контролю	26,7				
	Общая трудоемкость по дисциплине	144				

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1.	Введение в аналитику данных	Аналитика данных как процесс. Структурированные и неструктурированные данные. Технологии аналитики данных. Прикладные задачи анализа данных. Инструменты анализа данных. Понятие больших данных.	Т, О
2.	Подготовка данных для анализа	Группировка и преобразование данных. Обогащение данных. Квантование и скользящее окно. Транспонирование данных.	Т, О
3.	Визуализация данных	Введение в визуализацию. Визуализаторы общего назначения. Таблицы. Графики. Диаграммы. Круговые диаграммы. Гистограммы. Сложные визуализаторы общего назначения. OLAP-анализ. Географические карты. Тепловые карты. Плоское дерево. Диаграмма связей. Облако тегов. Пузырьковая диаграмма. Диаграмма рассеяния.	Т, О
4.	Введение в машинное обучение	Введение. Постановки задач в машинном обучении. Обучение с учителем и без. Классификация, регрессия, ранжирование, кластеризация. Обучающая и тестовая выборки. Проблема переобучения. Кросс-валидация.	Т, О
5.	Задача классификации. Метрические методы. Логические методы.	Мультикласс и мультилейбл классификация. Обучение классификаторов. Метод ближайших соседей. Решающие деревья.	Т, О
6.	Задачи регрессии. Линейные модели. Введение в нейронные сети.	Обучение регрессионной модели. МНК. Предпосылки регрессионного анализа. Линейная регрессионная модель. Полносвязная нейросеть.	Т, О
7.	Обучение без учителя.	Кластеризация данных, задачи обобщения, обнаружения аномалий, сокращения размерности.	Т, О

2.3.2 Занятия семинарского типа (практические / семинарские занятия/ лабораторные работы)

№	Наименование раздела (темы)	Тематика занятий/работ	Форма текущего контроля
1.	Введение в аналитику данных	Общие сведения о платформе Loginom. Интерфейс платформы, основные понятия. Компоненты и узлы. Сценарий Loginom. Компонент "Калькулятор" и визуализатор "Таблица". Автосинхронизация. Работа с переменными в Loginom. Компоненты "Условие" и "Замена". Подмодели. Компоненты "Узел-ссылка" и "Выполнение узла". Внешние компоненты и библиотеки. Пример решения задачи: возвраты.	ЛР
2.	Подготовка данных для анализа	ETL-процесс. Преобразование данных как часть ETL-операций. Основные методы преобразования данных. Группировка данных. Функции агрегации. Преимущества группировки. Операции с датой и временем. Оработка теории на практике в Loginom. Общие сведения об обогащении данных. Слияние (Внутреннее соединение. Левое и правое соединение. Полное соединение. Разность). Соединение. Дополнение данных. Объединение. Оработка теории на практике в Loginom.	ЛР
3.	Визуализация данных	Визуализатор "Статистика" в Loginom. Визуализатор "Диаграмма" в Loginom.	ЛР
4.	Введение в машинное обучение	Первичный Анализ данных в профессиональной сфере с Loginom. Основные методы очистки. Использование справочников и таблиц замены. Анализ строк. Регулярные выражения. Частотный анализ. Контрольные числа. Оработка теории на практике в Loginom.	ЛР
5.	Задача классификации. Метрические методы. Логические методы.	Мультикласс и мультилейбл классификация. Обучение классификаторов. Метод ближайших соседей. Решающие деревья.	ЛР
6.	Задачи регрессии. Линейные модели. Введение в нейронные сети.	Обучение регрессионной модели. МНК. Предпосылки регрессионного анализа. Линейная регрессионная модель. Полносвязная нейросеть.	ЛР
7.	Обучение без учителя.	Обучение без учителя: РСА, кластеризация.	ЛР

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

2.3.3 Примерная тематика курсовых работ (проектов)

Не предусмотрено

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	Занятия лекционного и семинарского типа	Методические указания для подготовки к занятиям лекционного и семинарского типа. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года.. Режим доступа: https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya
2	Выполнение самостоятельной работы обучающихся	Методические указания по выполнению самостоятельной работы обучающихся. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года.. Режим доступа: https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya
3	Выполнение лабораторных работ	Методические указания по выполнению лабораторных работ. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года.. Режим доступа: https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya
4	Интерактивные методы обучения	Методические указания по интерактивным методам обучения. Утверждены на заседании Совета экономического факультета ФГБОУ

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

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии, применяемые при освоении дисциплины (модуля)

В ходе изучения дисциплины предусмотрено использование следующих образовательных технологий: лекции, практические занятия, подготовка письменных аналитических работ, самостоятельная работа студентов.

Компетентностный подход в рамках преподавания дисциплины реализуется в использовании интерактивных технологий и активных методов (проектных методик, мозгового штурма, разбора конкретных ситуаций) в сочетании с внеаудиторной работой.

Информационные технологии, применяемые при изучении дисциплины: использование информационных ресурсов, доступных в информационно-телекоммуникационной сети Интернет.

Адаптивные образовательные технологии, применяемые при изучении дисциплины – для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Анализ данных в профессиональной сфере».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме *тестовых заданий* и **промежуточной аттестации** в форме вопросов и заданий к экзамену.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Код и наименование индикатора (в соответствии с п. 1.4)	Результаты обучения (в соответствии с п. 1.4)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация

1	ИПК-3.2 Анализирует данные для целей обоснования принятия решений в профессиональной сфере	<p><i>Знает: основные типы задач анализа данных и подходы к их решению;</i></p> <p><i>Знает: методы визуализации и представления результатов анализа данных.</i></p> <p><i>Знает: современные методы и инструментальные средства анализа данных;</i></p> <p><i>Умеет: проводить сравнительный анализ методов и инструментальных средств анализа больших данных;</i></p> <p><i>Умеет: планировать аналитические работы с использованием технологий анализа данных;</i></p> <p><i>Трудовое действие: выбор методов и инструментальных средств анализа данных для проведения аналитических работ;</i></p> <p><i>Трудовое действие: Сбор, очистка и предварительная обработка массивов данных для последующего анализа.</i></p>	<i>Тест по теме, защита лабораторной работы</i>	Вопрос на экзамене: 1-38
---	--	---	---	--------------------------

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Тесты для актуализации и проверки знаний.

Тест №1 по дисциплине «Анализ данных в профессиональной сфере»

1. Под анализом данных понимается
 - а) Метод хранения и организации данных
 - б) Процесс обнаружения и извлечения полезной информации из наборов данных
 - в) Инструмент для визуализации данных
 - г) Система для обработки текстов

2. Структурированные данные это

- а) Данные упорядоченные определенным образом и организованные с целью обеспечения возможности применения к ним некоторых действий
- б) Данные произвольные по форме, включающие тексты и графику, мультимедиа (видео, речь, графика)
- в) Данные, для которых определены некоторые правила и форматы, но в самом общем виде
- г) Разнообразные данные, поступающие с высокой скоростью, объем которых постоянно растет

3. Какие из следующих типов данных относятся к текстовыми? (Выберите все правильные ответы)

- а) Числа
- б) Строки
- в) Даты
- г) Булевы значения

4. К качественным дискретным шкалам относятся... (Выберите все правильные ответы)

- а) Номинальная шкала
- б) Шкала отношений
- г) Интервальная шкала
- д) Порядковая (ординальная) шкала

5. Отметьте особенности бизнес-данных (Выберите все правильные ответы)

- а) Объемы хранимых данных очень велики
- б) Бизнес-данные содержат ошибки, выбросы, противоречия и пропуски
- в) Бизнес-данные накапливаются специально для решения аналитических задач
- г) Основные потребители данных – лица принимающие решения

6. Выборочное распределение (набор данных, выборка) это

- а) Множество всех возможных значений случайной величины
- б) Случайная величина, описывающая интересующее явление
- в) Подмножество данных, извлеченных из генеральной совокупности
- г) Числовая характеристика случайной величины

7. Под очисткой данных понимается:

- а) Удаление всех данных из базы данных
- б) Процесс удаления некорректных или поврежденных данных из набора данных
- в) Преобразование данных в формат, пригодный для анализа или обработки
- г) Шифрование данных для обеспечения безопасности

8. Почему отсутствие стандартов хранения данных вызывает сложности при анализе данных и ухудшает его результат? (Выберите все верные ответы)

- а) Невозможно собрать данные из разных форматов
- б) Увеличивается количество ошибок в данных
- в) Часть данных может быть потеряна из-за использования недопустимых символов
- г) Одни и те же записи могут повторяться многократно с незначительными отличиями и анализироваться как разные

9. Под регулярным выражением понимается

- а) Математическое выражение, используемое для выполнения сложных вычислений
- б) Последовательность символов, которая определяет шаблон поиска в тексте
- в) Метод работы с массивами в программировании
- г) Алгоритм для сортировки данных в определенном порядке

10. Чему равно расстояние Левенштейна между словами «строка» и «собака» (Введите число в поле ввода)

11. Что означает True Positive (TP) в матрице ошибок?

- а) Количество правильно классифицированных отрицательных объектов
- б) Количество ошибочно классифицированных положительных объектов
- в) Количество правильно классифицированных положительных объектов
- г) Количество ошибочно классифицированных отрицательных объектов

12. Какие из перечисленных этапов анализа выделил в своей концепции анализа Дж. Тьюки? (Выберите все правильные ответы)

- а) Итоговый
- б) Разведочный
- в) Модельный

г) Итоговый

13. Low-code аналитика это

- а) Метод анализа данных, основанный на использовании низкоуровневых языков программирования
- б) Подход к разработке информационных систем, который позволяет создавать приложения с минимальным использованием кода.
- в) Технология анализа данных, использующая искусственный интеллект для автоматизации процесса.
- г) Методология, основанная на использовании больших данных для принятия бизнес-решений.

14. Какое из следующих утверждений характеризует эффект переобучения модели интеллектуального анализа данных?

- а) Модель слишком сильно настроилась на обучающую выборку и плохо работает на новых данных
- б) Модель недостаточно обучена и не может сделать точные прогнозы
- в) Модель имеет высокую обобщающую способность
- г) Модель слишком проста и не учитывает особенности данных

15. Какая метрика используется для оценки качества моделей классификации?

- а) MSE (Mean Squared Error)
- б) Accuracy
- в) R-квадрат
- г) Коэффициент корреляции

КЛЮЧИ К ТЕСТАМ

№ вопроса	1	2	3	4	5	6	7	8	9	10	11	12
Ответ	б	а	б	а, г	а, б, г	в	б	б, в, г	б	3	в	а,б,г
	13	14	15									
	б	а	б									

Тест №2 по дисциплине «Анализ данных в профессиональной сфере»

1. Задача классификации в контексте интеллектуального анализа данных это

- а) Процесс оценки качества данных
- б) Процесс группировки данных по схожим характеристикам
- в) Процесс прогнозирования категории или метки объекта
- г) Процесс определения корреляции между переменными

2. Неструктурированные данные это

- а) Данные упорядоченные определенным образом и организованные с целью обеспечения возможности применения к ним некоторых действий
- б) Данные произвольные по форме, включающие тексты и графику, мультимедиа (видео, речь, графика)
- в) Данные, для которых определены некоторые правила и форматы, но в самом общем виде
- г) Данные, которые организованы в четкую и предопределенную структуру, обычно в виде таблиц или баз данных.

3. Какие из следующих типов данных относятся к числовым? (Выберите все правильные ответы)

- а) Целые числа
- б) Вещественные числа
- в) Даты
- г) Булевы значения

4. К количественным непрерывным шкалам относятся... (Выберите все правильные ответы)

- а) Номинальная шкала
- б) Шкала отношений
- в) Интервальная шкала
- г) Порядковая (ординальная) шкала

5. Алгоритм глубокого обучения это

- а) Алгоритм для прогнозирования числовых значений
- б) Алгоритм для уменьшения размерности данных
- в) Алгоритм для обучения нейронных сетей с большим количеством слоев
- г) Алгоритм для кластеризации данных

6. Что такое перекрестная проверка?

- а) Метод для оценки качества модели на тестовых данных
- б) Метод для построения ансамбля моделей
- в) Метод для группировки данных по схожим характеристикам
- г) Метод для обработки пропущенных значений

7. Какой метод используется для понижения размерности данных, сохраняя при этом наибольшую долю информации?

- а) Метод кластеризации
- б) Метод наименьших квадратов
- в) Метод анализа главных компонент
- г) Метод наибольшего правдоподобия

8. Что такое выбросы (аномалии) в данных?

- а) Метод для оценки качества модели на тестовых данных
- б) Специфические значения данных, отличающиеся от общего тренда
- в) Метод для построения ансамбля моделей
- г) Метод обработки пропущенных значений

9. Какой метод используется для кластеризации данных?

- а) Логистическая регрессия
- б) Дерево принятия решений
- в) Метод наименьших квадратов
- г) К-средних

10. В каких из перечисленных случаев требуется очистка данных

- а) Данные содержат пропущенные значения
- б) Данные содержат дубликаты объектов
- в) Данные содержат выбросы
- г) Все вышеперечисленное

11. Нормализация данных это

- а) Процесс приведения данных к диапазону от 0 до 1

- б) Процесс удаления всех данных
- в) Процесс проверки и исправления ошибок и неточностей в данных
- г) Процесс шифрования данных

12. Какой метод используется для заполнения пропущенных значений в данных?

- а) Удаление строк или столбцов с пропущенными значениями
- б) Заполнение средним или медианой
- в) Заполнение на основе соседних значений
- г) Все вышеперечисленное

13. Расстояние Левенштейна между словами «Австрия» и «Австралия» равно ...?

14. Какая метрика используется для оценки качества моделей регрессии? (Выберите все правильные ответы)

- а) Средняя ошибка аппроксимации
- б) Ассигасу
- в) R-квадрат
- г) Коэффициент корреляции

15. Матрица ошибок в задаче классификации это

- а) Матрица, показывающая долю правильно классифицированных объектов
- б) Матрица, показывающая долю ошибочно классифицированных объектов
- в) Матрица, показывающая различные комбинации правильных и неправильных классификаций
- г) Матрица, показывающая вероятность принадлежности к разным классам

Тест №2 по дисциплине «Анализ данных в профессиональной сфере»

1. Задача кластеризации интеллектуального анализа данных это

- а) Задача оценки качества данных
- б) Задача группировки данных по схожим характеристикам
- в) Задача прогнозирования категории или метки объекта
- г) Задача определения корреляции между переменными

2. Слабоструктурированные данные это

- а) Данные упорядоченные определенным образом и организованные с целью обеспечения возможности применения к ним некоторых действий
- б) Данные произвольные по форме, включающие тексты и графику, мультимедиа (видео, речь, графика)
- в) Данные, для которых определены некоторые правила и форматы, но в самом общем виде
- г) Данные, которые организованы в четкую и предопределенную структуру, обычно в виде таблиц или баз данных.

3. В какой шкале может быть задан признак «цвет глаз» человека

- а) Номинальная шкала
- б) Шкала отношений
- г) Интервальная шкала
- д) Порядковая (ординальная) шкала

4. Если качественный признак имеет 5 уровней по номинальной шкале, какое количество фиктивных переменных понадобится для его кодирования? (Введите число в поле ввода)

5. Какие из следующих типов данных используются для представления иерархической структуры? (Выберите все правильные ответы)

- а) Строки
- б) Целые числа
- в) Списки
- г) Словари

6. Какой подход в анализе данных используется для прогнозирования числовых значений?

- а) Кластеризация
- б) Регрессия
- в) Классификация
- г) Анализ временных рядов

7. Какая метрика используется для оценки тесноты линейной связи между признаками?

- а) Средняя ошибка аппроксимации
- б) Accuracy

- в) Коэффициент детерминации
- г) Коэффициент корреляции

8. Какая метрика позволяет оценить баланс между полнотой (recall) и точностью (precision)?

- а) Accuracy
- б) F-score
- в) R-квадрат
- г) True Positive (TP)

9. Под моделью анализа данных понимается:

- а) Физическое устройство для хранения и обработки данных
- б) Программное обеспечение для визуализации данных
- в) Математическая или статистическая конструкция, используемая для представления и анализа данных
- г) Методология для сбора данных из различных источников

10. К пропущенным значениям в данных относятся

- а) Значения, которые являются ошибочными и требуют удаления
- б) Значения, которые отсутствуют или не заполнены
- в) Значения, которые находятся за пределами допустимого диапазона
- г) Значения, которые имеют неправильный формат

11. Регуляризация модели анализа данных это

- а) Процесс увеличения сложности модели для лучшей аппроксимации данных
- б) Процесс уменьшения сложности модели для улучшения обобщающей способности
- в) Процесс выбора оптимальных гиперпараметров модели
- г) Процесс обработки выбросов в данных

12. Выберите неправильные утверждения о настольных статистических пакетах

- а) Они слабо интегрированы с промышленными источниками данных
- б) Обработка больших объемов данных статистическими пакетами затруднена
- в) Они содержат богатые средства предобработки и очистки
- г) Определенный пакет ориентирован на решение задач одного класса

13. Какой компонент аналитической платформы анализа данных является опциональным?

- а) Интеграционный сервер
- б) Репозиторий моделей
- в) Хранилище данных
- г) Аналитический сервер

14. Выберите корректное определение термина «прикладная задача» в контексте интеллектуального анализа данных

- а) Это задача, которая решается с помощью построения моделей Data Mining
- б) Это задача, которая сформулирована в терминологии, понятной менеджменту среднего и высшего звена
- в) Это задача, которая сформулирована в терминологии, понятной аналитику
- г) Это задача исследования математических свойств случайных процессов

15. Какие из приведенных задач актуальны и для банков, и для ритейла, и для телекома

- а) Построение скоринговых моделей
- б) Сегментация клиентов
- в) Оптимизация запасов
- г) Управление инкассациями

Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен)

1) Анализ данных в профессиональной сфере: история, области применения, технологии, задачи и инструменты. Аналитический и информационный подходы к моделированию. Концепция разведочного анализа Дж.Тьюки.

2) Формы представления данных: структурированные, неструктурированные, слабоструктурированные. Типы данных и шкалы. Непрерывные и дискретные данные. Особенности бизнес-данных.

3) Информативность данных. Метод сбора данных. Принципы формализации данных. Требования к данным.

4) Аналитика данных как процесс. Принципы анализа данных. Извлечение и визуализация данных. Построение и применение моделей. Процесс анализа данных. Тиражирование знаний.

- 5) Понятия, связанные с аналитикой данных (бизнес-аналитика (BI), Data Mining, Data Warehousing, Big Data). MapReduce, Hadoop, No SQL. Data Science.
- 6) KDD. Этапы KDD. Методология CRISP-DM.
- 7) Методы и задачи Data Mining. Дескриптивные и предсказательные. Кластеризация, ассоциативные правила, сокращение описания.
- 8) Методы и задачи Data Mining. Классификация и регрессия. Машинное обучение (ML).
- 9) Инструментарий аналитики данных. Программное обеспечение аналитики данных. Компоненты аналитической платформы. Обзор вендоров.
- 10) Сбор и подготовка данных. Процесс ETL (DataPrep). Основные методы преобразования данных.
- 11) Группировка данных. Функции агрегации. Преобразование упорядоченных данных. Скользящее окно (интервал прогноза, горизонт прогноза, глубина истории), операции с датой и временем.
- 12) Обогащение данных. Слияние (внутреннее соединение, левое и правое соединения, разность, полное соединение), соединение, дополнение, объединение.
- 13) Качество данных. Аудит и профайлинг данных. Этапы аудита данных. Отчет по результатам профайлинга данных. Отчет по статистике. Индикаторы качества данных. Стратегии борьбы с пропусками в данных. Выбросы и экстремальные значения.
- 14) Биннинг и оптимальное квантование. Конечные классы. WoE-анализ и IVиндекс. WoE-диаграммы. Монотонный тренд.
- 15) Бинарная логистическая регрессия. Логит-преобразование и логистическая кривая. Порог отсечения. Критерии отбора факторов (отношение правдоподобия; информационный критерий Акаике; информационный критерий Байеса; информационный критерий Ханнана-Куина). Отбор факторов и защита от переобучения (enter, forward, backward, stepwise, ridge, LASSO, elastic-net).
- 16) Метрики качества моделей бинарной классификации. Матрица ошибок (Confusion matrix). ACC (меткость). PPV (точность). TPR (полнота, отзыв, чувствительность). TNR (специфичность). F1-мера.
- 17) Метрики качества моделей бинарной классификации. ROC-анализ. AUCROC. Ошибки классификации I и II рода. Визуализатор «качество бинарной классификации» в Loginom.
- 18) Классификация с учетом издержек. Установление оптимального порога отсечения. Правило Байеса.
- 19) Логистическая регрессия. Визуализаторы «выход регрессии», «отчет по регрессии», «коэффициенты логистической регрессии». Отношение шансов (OR).

- 20) Общие сведения о платформе Loginom. Интерфейс платформы, основные понятия. Компоненты и узлы. Сценарий Loginom.
- 21) Компонент «Калькулятор» и визуализатор «Таблица». Автосинхронизация. Работа с переменными в Loginom. Компоненты «Условие» и «Замена». Подмодели.
- 22) Компоненты «Узел-ссылка» и «Выполнение узла». Внешние компоненты и библиотеки. Производные компоненты и узлы. Обучение и переобучение узлов. Предобработка и очистка данных. Заполнение пропусков и редактирование выбросов. Компонент «Дубликаты и противоречия» в Loginom. Удаление дубликатов в Loginom.
- 23) Группировка и преобразование даты: ETL-процесс. Преобразование данных как часть ETL-операций. Основные методы преобразования данных. Группировка данных. Преимущества группировки. Функции агрегации.
- 24) Общие сведения об обогащении данных. Реализация в Loginom процессов слияния, соединения, дополнения и объединения данных.
- 25) Квантование. Цели квантования. Результат квантования. Выбор числа интервалов в квантовании. Методы квантования. Преобразование упорядоченных данных. Скользящее окно.
- 26) Транспонирование данных: структура данных. Представление данных. Некорректная структура. Понятие транспонирования. Транспонирование на этапе ETL.
- 27) Понятие, цели и задачи визуализации. Этапы визуализации. Методы визуализации. Визуализатор «Статистика» и «Качество данных» в Loginom. Простые визуализаторы общего назначения: таблицы, диаграммы.
- 28) Сложные визуализаторы общего назначения: OLAP-анализ. Визуализатор «Куб» в Loginom.
- 29) Компонент «Цикл» в Loginom.
- 30) Корреляционный анализ в Loginom. Коэффициенты корреляции Пирсона и Спирмена. Интерпретация значений корреляционной матрицы.
- 31) Факторный анализ. Задачи факторного анализа. Факторные нагрузки. Собственное значение (критерий Кайзера) и объем объясняемой дисперсии в %. Ортогональное вращение (варимакс и квартимакс). Оптимальное число факторов (метод «каменистой осыпи»).
- 32) Методы очистки данных в Loginom. Сверка со справочником. Анализ строк. Частотный анализ. Расчет контрольного числа. Регулярные выражения Perl. Регулярные выражения в Loginom.
- 33) Data Mining: задача ассоциации. Ассоциативные правила в Loginom. Применение ассоциативных правил при анализе рыночной корзины. Поддержка. Достоверность. Лифт.

- 34) Data Mining: кластеризация в Logiom. Задача кластеризации. Гипотеза компактности. Этапы кластеризации. Проблема определения оптимального числа кластеров. K-Means и G-Means кластеризация. Кластеризация транзакций. EM-кластеризация.
- 35) Data Mining: нейросеть (регрессия) и линейная регрессия в Logiom. Статистические методы. Сравнение моделей.
- 36) Data Mining: нейросеть (классификация) и логистическая регрессия в Logiom. Машинное обучение. Сравнение моделей.
- 37) Анализ и прогнозирование временных рядов. Автокорреляция. ARIMAX.
- 38) Разработка бизнес-приложений в low-code платформе Logiom (ABC-XYZ анализ, RFM-анализ, CVP-анализ, анализ чувствительности, точки заказа, коэффициенты чувствительности и оборачиваемости)

Критерии оценивания результатов обучения

<i>Оценка</i>	<i>Критерии оценивания по экзамену</i>
<i>Высокий уровень «5» (отлично)</i>	<i>оценку «отлично» заслуживает студент, освоивший знания, умения, компетенции и теоретический материал без пробелов; выполнивший все задания, предусмотренные учебным планом на высоком качественном уровне; практические навыки профессионального применения освоенных знаний сформированы.</i>
<i>Средний уровень «4» (хорошо)</i>	<i>оценку «хорошо» заслуживает студент, практически полностью освоивший знания, умения, компетенции и теоретический материал, учебные задания не оценены максимальным числом баллов, в основном сформировал практические навыки.</i>
<i>Пороговый уровень «3» (удовлетворительно)</i>	<i>оценку «удовлетворительно» заслуживает студент, частично с пробелами освоивший знания, умения, компетенции и теоретический материал, многие учебные задания либо не выполнил, либо они оценены числом баллов близким к минимальному, некоторые практические навыки не сформированы.</i>
<i>Минимальный уровень «2» (неудовлетворительно)</i>	<i>оценку «неудовлетворительно» заслуживает студент, не освоивший знания, умения, компетенции и теоретический материал, учебные задания не выполнил, практические навыки не сформированы.</i>

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

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

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

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

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление

информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

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

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. Перечень учебной литературы, информационных ресурсов и технологий

5.1. Учебная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511121>

2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511020>

5.2. Периодическая литература

Указываются печатные периодические издания из «Перечня печатных периодических изданий, хранящихся в фонде Научной библиотеки КубГУ» <https://www.kubsu.ru/ru/node/15554>, и/или электронные периодические издания, с указанием адреса сайта электронной версии журнала, из баз данных, доступ к которым имеет КубГУ:

1. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
2. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» www.biblioclub.ru
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных:

1. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
2. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
3. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>
4. Президентская библиотека им. Б.Н. Ельцина <https://www.prlib.ru/>
5. Электронная коллекция Оксфордского Российского Фонда <https://ebookcentral.proquest.com/lib/kubanstate/home.action>

6. Springer Journals <https://link.springer.com/>
7. Nature Journals <https://www.nature.com/siteindex/index.html>
8. Springer Nature Protocols and Methods
<https://experiments.springernature.com/sources/springer-protocols>
9. Springer Materials <http://materials.springer.com/>
10. zbMath <https://zbmath.org/>
11. Nano Database <https://nano.nature.com/>
12. Springer eBooks: <https://link.springer.com/>
13. "Лекториум ТВ" <http://www.lektorium.tv/>
14. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы:

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа:

1. Американская патентная база данных <http://www.uspto.gov/patft/>
2. Полные тексты канадских диссертаций <http://www.nlc-bnc.ca/thesescanada/>
3. КиберЛенинка (<http://cyberleninka.ru/>);
4. Министерство науки и высшего образования Российской Федерации
<https://www.minobrnauki.gov.ru/>;
5. Федеральный портал "Российское образование" <http://www.edu.ru/>;
6. Информационная система "Единое окно доступа к образовательным ресурсам"
<http://window.edu.ru/>;
7. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
8. Федеральный центр информационно-образовательных ресурсов
(<http://fcior.edu.ru/>);
9. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
10. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
11. Служба тематических толковых словарей <http://www.glossary.ru/>;
12. Словари и энциклопедии <http://dic.academic.ru/>;
13. Образовательный портал "Учеба" <http://www.uceba.com/>;
14. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы
http://xn--273--84d1f.xn--plai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы

КубГУ:

1. Среда модульного динамического обучения <http://moodle.kubsu.ru>
2. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://mschool.kubsu.ru/>
3. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru/>;
4. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
5. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ"
<http://icdau.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины (модуля)

Самостоятельная работа студентов является неотъемлемой частью процесса подготовки. Дисциплину рекомендуется изучать путем систематической проработки

лекционного материала, самостоятельной проработки рекомендуемой литературы, руководств и методических указаний к выполнению практических занятий. Цель самостоятельной работы – расширение кругозора и углубление знаний в области финансового инструментария.

Контроль за выполнением самостоятельной работы проводится при изучении каждой темы дисциплины на семинарских занятиях. Это текущий опрос, тестовые задания, контрольная работа.

В часы, отведенные для самостоятельной работы, студенты под руководством преподавателя обязаны выполнять индивидуальные практические задания, полученные на практических занятиях. При выполнении этих заданий необходимо использовать теоретический материал, делать ссылки на соответствующие формулы, проверять выполнимость предпосылок, необходимых для применения того или иного метода.

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

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

7. Материально-техническое обеспечение по дисциплине (модулю)

По всем видам учебной деятельности в рамках дисциплины используются аудитории, кабинеты и лаборатории, оснащенные необходимым специализированным и лабораторным оборудованием.

Наименование специальных помещений	Оснащенность специальных помещений	Перечень лицензионного программного обеспечения
Учебные аудитории для проведения занятий лекционного типа	Мебель: учебная мебель Технические средства обучения: экран, проектор, ноутбук	Microsoft Windows 8, 10, Microsoft Office Professional Plus
Учебные аудитории для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Мебель: учебная мебель Технические средства обучения: экран, проектор, ноутбук	Microsoft Windows 8, 10, Microsoft Office Professional Plus

Для самостоятельной работы обучающихся предусмотрены помещения, укомплектованные специализированной мебелью, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование помещений для самостоятельной работы обучающихся	Оснащенность помещений для самостоятельной работы обучающихся	Перечень лицензионного программного обеспечения
Помещение для самостоятельной работы обучающихся (читальный зал Научной библиотеки)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее	Microsoft Windows 8, 10, Microsoft Office Professional Plus

	доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	
Помещение для самостоятельной работы обучающихся (ауд.213 А, 218 А)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	Microsoft Windows 8, 10, Microsoft Office Professional Plus