

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное учреждение высшего образования «Горно-Алтайский государственный университет»
(ФГБОУ ВО ГАГУ, ГАГУ, Горно-Алтайский государственный университет)**

Инструменты анализа данных в экономике
рабочая программа дисциплины (модуля)

Закреплена за кафедрой **кафедра экономики, туризма и прикладной информатики**

Учебный план 38.03.01_2025_815.plx
38.03.01 Экономика
Бизнес-аналитика и цифровая экономика

Квалификация **бакалавр**

Форма обучения **очная**

Общая трудоемкость **5 ЗЕТ**

Часов по учебному плану	180	Виды контроля в семестрах: экзамены 5
в том числе:		
аудиторные занятия	56	
самостоятельная работа	87,2	
часов на контроль	34,75	

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	5 (3.1)		Итого	
Неделя	17 1/6			
Вид занятий	уп	рп	уп	рп
Лекции	16	16	16	16
Лабораторные	40	40	40	40
Консультации (для студента)	0,8	0,8	0,8	0,8
Контроль самостоятельной работы при проведении аттестации	0,25	0,25	0,25	0,25
Консультации перед экзаменом	1	1	1	1
Итого ауд.	56	56	56	56
Контактная работа	58,05	58,05	58,05	58,05
Сам. работа	87,2	87,2	87,2	87,2
Часы на контроль	34,75	34,75	34,75	34,75
Итого	180	180	180	180

Программу составил(и):

к.ф.-м.н., доцент, Губкина Елена Владимировна

Рабочая программа дисциплины

Инструменты анализа данных в экономике

разработана в соответствии с ФГОС:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 38.03.01 Экономика (приказ Минобрнауки России от 12.08.2020 г. № 954)

составлена на основании учебного плана:

38.03.01 Экономика

утвержденного учёным советом вуза от 30.01.2025 протокол № 2.

Рабочая программа утверждена на заседании кафедры

кафедра экономики, туризма и прикладной информатики

Протокол от 10.04.2025 протокол № 9

Зав. кафедрой Газукина Юлия Геннадьевна

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2026-2027 учебном году на заседании кафедры
кафедра экономики, туризма и прикладной информатики

Протокол от _____ 2026 г. № ____
Зав. кафедрой Газукина Юлия Геннадьевна

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2027-2028 учебном году на заседании кафедры
кафедра экономики, туризма и прикладной информатики

Протокол от _____ 2027 г. № ____
Зав. кафедрой Газукина Юлия Геннадьевна

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2028-2029 учебном году на заседании кафедры
кафедра экономики, туризма и прикладной информатики

Протокол от _____ 2028 г. № ____
Зав. кафедрой Газукина Юлия Геннадьевна

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2029-2030 учебном году на заседании кафедры
кафедра экономики, туризма и прикладной информатики

Протокол от _____ 2029 г. № ____
Зав. кафедрой Газукина Юлия Геннадьевна

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Цели: Сформировать компетенции обучающегося в области теоретических и практических основ технологий анализа больших данных, получение знаний об особенностях анализа данных для решения разнообразных бизнес-аналитических задач, возникающих в профессиональной деятельности.
1.2	Задачи: Освоить методологии обработки больших данных; Изучить методами и технологиями работы с данными и знаниями; Сформировать представление о структуризации данных и технологиях Knowledge Discovery in Databases и Data Mining; Освоить основы анализа данных

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:		Б1.В
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Эконометрика	
2.1.2	Статистика	
2.1.3	Цифровая экономика	
2.1.4	Теория вероятности и математическая статистика	
2.1.5	Методы оптимальных решений	
2.1.6	Математика для экономистов	
2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Технологическая (проектно-технологическая) практика	
2.2.2	Цифровая трансформация среды и бизнеса	
2.2.3	Выполнение и защита выпускной квалификационной работы	
2.2.4	Преддипломная практика	
2.2.5	Электронная коммерция и интернет-маркетинг	

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ПК-1: Способность рассчитывать, анализировать и интерпретировать информацию, необходимую для выявления тенденций в функционировании и развитии хозяйствующих субъектов	
ИД-1.ПК-1: Способен анализировать финансовую информацию организаций, рассчитывать экономические и финансовые показатели, используя современные методы и технологии	
Знает основы финансового анализа Умеет анализировать информацию для проведения расчетов экономического и финансового показателей, используя ИКТ Владеет навыками использования ИКТ для проведения расчетов финансово-экономических показателей, с применением современных ИТ и ИКТ	
ПК-2: Способен осуществлять сбор и анализ информации для бизнес-анализа для формирования возможных решений	

ИД-1.ПК-2: Выявляет, регистрирует, анализирует и классифицирует риски, разрабатывает комплекс мероприятий по управлению ими							
Знает методы анализа и классификации рисков Умеет разрабатывать комплексы предотвращения рисков с использованием ИКТ и информационных ресурсов Владеет навыками : выявления, анализа и регистрации рисков методами и технологиями разработки комплексов мероприятий по уменьшению рисков и их последствий							
ИД-2.ПК-2: Анализирует внутренние/внешние факторы и условия, влияющие на деятельность организации							
знает внутренние/внешние факторы и условия, влияющие на деятельность организации умеет анализировать факторы и условия, влияющие на деятельность организации владеет навыками применения ИКТ и ПО для проведения анализа мероприятий по деятельности организации							
ИД-3.ПК-2: Применяет IT-инструменты для обеспечения работ по бизнес-анализу							
знает основные инструменты ИТ, которые применяются для бизнес-анализа умеет применять ИКТ для проведения бизнес-анализа владеет навыками применения современных ИКТ для проведения бизнес-анализа							
ИД-4.ПК-2: Отбирает, применяет и адаптирует соответствующие методы, инструменты и техники анализа бизнес-ситуации и предметной области, включая методы и инструменты анализа данных							
знает методы, инструменты и техники бизнес-анализа умеет применять методы, инструменты и техники бизнес-анализа							
владеет навыками отбора, применения и адаптации соответствующих методов, инструментов и техник анализа бизнес-ситуации и предметной области, включая методы и инструменты анализа данных							

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)							
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература	Инте ракт.	Примечание
	Раздел 1. Технологии анализа данных						
1.1	Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных. /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
1.2	Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных. /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
1.3	Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
1.4	Начало работы. Понятие сценария и узла обработки. Консолидация данных. Трансформация данных. Визуализация данных /Лаб/	5	8		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
1.5	Регрессионный анализ /Лаб/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	

1.6	Корреляционный и дисперсионный анализ /Лаб/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
1.7	Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.(Подготовка к занятию и выполнение правтического занятия) /Ср/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной ратуры и подготовка к промежуточно й аттестации
1.8	Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных. /Ср/	5	3		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной ратуры и подготовка к промежуточно й аттестации
1.9	Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. Подготовка к занятиям и выполнение практического задания. /Ср/	5	4		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной ратуры и подготовка к промежуточно й аттестации
	Раздел 2. Интеллектуальный анализ данных						
2.1	Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил. /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.2	Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации. /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.3	Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.4	Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура. /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	

2.5	Определение дерева решений. Причины популярности и условия применимости. Структура дерева решений. Выбор атрибута разбиения в узле. Алгоритм ID3, критерий выбора атрибута разбиения ID3, пример работы алгоритма. Проблема переобучения, Неизвестные значения атрибутов, алгоритм C4.5. /Лек/	5	2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.6	Ассоциативные правила. Поиск ассоциативных правил. /Лаб/	5	6		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.7	Кластеризация. Алгоритм кластеризации k-means. /Лаб/	5	6		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.8	Прогнозирование с помощью линейной регрессии. /Лаб/	5	6		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.9	Классификация с помощью нейросети. /Лаб/	5	6		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.10	Классификация с помощью деревьев решений. /Лаб/	5	4		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	
2.11	Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и леввередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.. /Ср/	5	6		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и подготовка к промежуточной аттестации
2.12	Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации. /Ср/	5	4		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и подготовка к промежуточной аттестации
2.13	Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов /Ср/	5	4		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и подготовка к промежуточной аттестации
2.14	Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура. /Ср/	5	4		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и подготовка к промежуточной аттестации
2.15	Определение дерева решений. Причины популярности и условия применимости. Структура дерева решений. Выбор атрибута разбиения в узле. Алгоритм ID3, критерий выбора атрибута разбиения ID3, пример работы алгоритма. Проблема переобучения, Неизвестные значения атрибутов, алгоритм C4.5. /Ср/	5	4		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и подготовка к промежуточной аттестации

2.16	Ассоциативные правила. Поиск ассоциативных правил. /Ср/	5	10		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и
2.17	Кластеризация. Алгоритм кластеризации k-means. /Ср/	5	10		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и
2.18	Прогнозирование с помощью линейной регрессии. /Ср/	5	12		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и
2.19	Классификация с помощью нейросети. /Ср/	5	12		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и
2.20	Классификация с помощью деревьев решений. /Ср/	5	12,2		Л1.1 Л1.2 Л1.3Л2.1 Л2.2	0	Изучение дополнительной литературы и
	Раздел 3. Консультации						
3.1	Консультация по дисциплине /Конс/	5	0,8	ИД-1.ПК-1 ИД-1.ПК-2 ИД-2.ПК-2 ИД-3.ПК-2 ИД-4.ПК-2		0	
	Раздел 4. Промежуточная аттестация (экзамен)						
4.1	Подготовка к экзамену /Экзамен/	5	34,75	ИД-1.ПК-1 ИД-1.ПК-2 ИД-2.ПК-2 ИД-3.ПК-2 ИД-4.ПК-2		0	
4.2	Контроль СР /КСРАтт/	5	0,25	ИД-1.ПК-1 ИД-1.ПК-2 ИД-2.ПК-2 ИД-3.ПК-2 ИД-4.ПК-2		0	
4.3	Контактная работа /КонсЭк/	5	1	ИД-1.ПК-1 ИД-1.ПК-2 ИД-2.ПК-2 ИД-3.ПК-2 ИД-4.ПК-2		0	

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Пояснительная записка

1. Назначение фонда оценочных средств. Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины Инструменты анализа данных в экономике
2. Фонд оценочных средств включает контрольные материалы для проведения текущего контроля и промежуточной аттестации вопросов к промежуточной аттестации

5.2. Оценочные средства для текущего контроля

Контрольные тесты и задания

Название вопроса*: 1 (ПК-1)

Формулировка вопроса: Что такое анализ риска?

Варианты ответов

- 1) систематизация множества рисков на основании каких-либо признаков и критериев, позволяющих объединить подмножества рисков в более общие понятия
- 2) систематическое научное исследование степени риска, которому подвержены конкретные объекты, виды деятельности и проекты
- 3) начальный этап системы мероприятий по управлению рисками, состоящий в систематическом выявлении рисков, характерных для определенного вида деятельности, и определении их характеристик
- 4) такого понятия нет

Ключ: 2) систематическое научное исследование степени риска, которому подвержены конкретные объекты, виды деятельности и проекты

Название вопроса*: 2 (ПК-1)

Формулировка вопроса: Как правило, одного лишь суждения будет недостаточно для принятия решения:

Варианты ответов

- 1) когда ситуация не дает достаточно информации
- 2) когда ситуация уникальна или очень сложна
- 3) когда ситуация довольно проста
- 4) такой ситуации не существует

Ключ: 2) когда ситуация уникальна или очень сложна

Название вопроса*: 3 (ПК-1)

Формулировка вопроса: Способность системы, сети или процесса справляться с увеличением рабочей нагрузки при добавлении ресурсов, называется

Ключ: масштабируемость

Название вопроса*: 4 (ПК-1)

Формулировка вопроса: следствие действия либо бездействия, в результате которого существует реальная возможность получения неопределенных результатов различного характера это

Ключ: риск

Название вопроса*: 1 (ПК-2)

Формулировка вопроса: 2. Основные источники больших данных: ...

Варианты ответов

- 1) статистическая информация
- 2) реляционные базы данных
- 3) показания считывающих устройств
- 4) интернет

Ключ: 3) показания считывающих устройств

Название вопроса*: 2 (ПК-2) Какой из следующих инструментов является основным для обработки и анализа больших данных в экономике?

Формулировка вопроса:

Варианты ответов

- 1) hadoop
- 2) Microsoft Excel
- 3) Adobe Photoshop
- 4) Google Chrome

Ключ: 1) hadoop

Название вопроса*: 3 (ПК-2)

Формулировка вопроса: Как называется выбор в электронной таблице данных, соответствующих определенным условиям

Ключ: фильтрация

Название вопроса*: 4 (ПК-2)

Формулировка вопроса: Какой процесс отвечает за объединение данных из разных источников для создания общего набора данных для анализа?

Ключ: Data Integration

Критерии оценки для всех аттестаций, проходящих в форме компьютерного тестирования.

менее 60% - неудовлетворительно

60%-74 %- удовлетворительно

75%-89%- хорошо

90% и более - отлично

1. К проблемам обработки и анализа данных относятся: ...
 - формат, отличный от традиционных баз данных
 - четкая структурированность данных, не позволяющая обработать разноформатные данные
 - отсутствие нужных инструментов, чтобы связать данные между собой
 - хранение больших вычислительных мощностей
2. Тест Тьюринга умышленно исключает прямое когнитивное взаимодействие между тестирующим и компьютером ...
 - Верно
 - Неверно
3. Ключевые характеристики Big Data с условным наименованием «5V»: ...
 - Value
 - Voice
 - Volume
 - Veracity
 - Variety
4. На этапе внедрения изучается: ...
 - Каким образом можно развернуть выбранные модели в бизнес-среде
 - Соответствует ли модель целям процесса?
 - Как интегрировать модели в бизнес-процессы организации
 - Как интегрировать модели в техническую инфраструктуру
5. Соответствие распределения времени между основными задачами обработки больших данных
 - построение обучающих моделей
 - очистка и организация данных
 - анализ данных для выявления закономерностей
 - сбор данных
 - уточнение алгоритмов
 - другие задачи
1. Особенности Электронной библиотеки Всемирного Банка: ...
 - содержит Обширный набор экономических, социальных и экологических показателей
 - включает показатели по странам за период с 1930 года
 - предоставляет свободный и открытый доступ к данным о развитии в странах по всему миру
 - является полнотекстовой базой статистических, аналитических анализов
- Всемирного Банка
2. Основные источники больших данных: ...
 - статистик
 - реляционные базы данных
 - показания считывающих устройств
 - интернет
3. Верные заключения о больших данных: ...
 - интеллектуальные сети обеспечивают возможность коммунальным предприятиям гораздо лучше управлять их энергосистемами
 - текстовые данные не являются типами данных, применимыми в больших данных
 - со многими источниками больших данных связаны проблемы соблюдения конфиденциальности
 - одни и те же базовые технологии могут быть использованы в различных отраслях для решения различных задач
4. Особенности данных, генерируемых интеллектуальными сетями: ...
 - предполагает наличие сложных систем мониторинга, связи и генерации энергии
 - уступают по надежности традиционным линиям электропередач
 - обеспечивают более надежное обслуживание и восстановление после отключения питания
 - домовладелец может проверить, какую мощность потребляют приборы, включив их по очереди
1. Особенности хранения и управления Big Data: ...
 - %-100 Big Data всегда хранятся и организуются в централизованных файловых системах
 - информация хранится на нескольких (иногда тысячах) жестких дисках, на стандартных компьютерах
 - для обеспечения отказоустойчивости и надежности, каждую часть информации

сохраняют несколько раз

- Big Data обычно хранятся и организуются в распределенных файловых системах

2. Три основные задачи, связанные с большими данными: ...

- Хранение и управление

- Совершенствование методов анализа Big Data

- Обеспечение атомарности RDBMS

- Обработка неструктурированной информации

3. Недостаток неструктурированных данных в Big Data: ...

- сохранение “всех данных”, независимо от того, какая часть данных актуальна для последующего принятия решения

- сохранение “всех данных”, независимо от того, какая часть данных актуальна для последующего принятия решения

- для извлечения полезной информации требуется последующая обработка этих огромных массивов данных

4. Метод классификации больших данных, основанный на таблице частот, при котором алгоритм разбивает датасет на все меньшие куски данных, формируя тем самым дерево, называется: ...

- ZeroR

- Decision Tree

- OneR

- Naive Bayesian

1. Библиотека, используемая при работе с большим данными для повышения скорости выполнения кода, компилирующая код непосредственно перед выполнением (JIT компиляция), называется ...

- Numexpr

- Hadoopy

- NumPy

- Numba

2. Особенности баз данных на древовидных структурах: ...

- Индексы строятся на базе деревьев и хеш-таблицах

- 100 просматривают содержимое таблиц от первой строки до последней

- используют индекс для ускорения поиска

3. Типы алгоритмов, при работе с большими данными, оптимизирующие процесс обработки и анализ больших данных: ...

- блочные алгоритмы

- алгоритмы MapReduce

- онлайн-алгоритмы

- структурные алгоритмы

4. Особенности «узких мест» и простоя при работе с большими данными: ...

- SSD работают медленнее, чем HDD

- Некоторые системы простаивают, т.к. компоненты компьютера создают «узкие места»,

- SSD работают быстрее центрального процессора

- Некоторые программы не успевают быстро поставлять данные процессору, т.к. читают данные с HDD

5. Примеры онлайн-алгоритмов: ...

- Линейный метод наименьших квадратов

- Онлайн-выпуклая оптимизация

- Регрессивное обучение

- Метод стохастического градиентного спуска

- Инкрементальный стохастический градиентный спуск

1. Примеры распределенных и облачных хранилищ данных: ...

- APACHE HADOOP

- MICROSOFT ACCESS

- GOOGLE CLOUD STORAGE

- AMAZON S3

2. Особенности распределенных хранилищ данных: ...

- предусмотрено централизованное хранение данных

- данные расположены на нескольких компьютерах в разных местах

- компьютерная сеть, в которой информация реплицируемым образом

- информация хранится в распределенной базе данных

3. Для подключения специализированных расширений для языков командной строки к облачным ресурсам: ...

- не требуется дополнительных действий

- необходимо импортировать ключи

- необходимо выполнить вход в аккаунт через форму ввода логина/пароля

4. Системы, применяющие шаблон для конфигурирования различных инфраструктур: ...

- Ansible
 - Chef
 - GitHub
 - Puppet
1. Компоненты Hadoop: ...
- YARN
 - MapReduce
 - Dask
 - HDFS
2. Особенности HDFS: ...
- Отказоустойчивая
 - Файловая система на основе JavaScript
 - Надежная и экономичная
 - Масштабируемая
3. Hadoop состоит из двух основных компонентов: ...
- узла каталога
 - узла данных
 - формата узла
 - имени узла
4. Особенности HIVE: ...
- 30%имеет два компонента - драйверы и командную строку
 - не позволяет обрабатывать данные реальном времени
 - 40%осуществляет пакетную обработку
 - 30%поддерживает все типы данных SQL
1. Основные принципы BASE баз данных NOSQL: ...
- Soft state
 - Basically Available
 - Eventual consistency
 - Atomicity
2. Принцип при котором: если блок данных включается в базу, то он либо включается полностью, либо не включается вообще, называется: ...
- Atomicity
 - Durability
 - Consistency
 - Isolation
3. Задачи, которые решает NoSQL: ...
- приводит все БД таблицы к нормальной форме
 - анализ и хранение больших объемов данных в распределенной среде
 - Задачи, связанные с доступом к большим объемам разнородных данных
 - Задачи, связанные с обработкой больших объемов разнородных данных
4. Характерные свойства ACID для реляционной базы данных: ...
- атомарность
 - горизонтальное масштабирование
 - изолированность
 - непротиворечивость
5. Способность системы, сети или процесса справляться с увеличением рабочей нагрузки при добавлении ресурсов, называется ...
- Масштабируемость
 - Согласованность данных
 - Изолированность
 - Атомарная операция
1. Технология MapReduce может использоваться для: ...
- счётчиков частоты обращений к заданному адресу
 - индексации веб-контента
 - подсчета слов в большом файле
 - вычисления объема памяти
2. Функция Reduce: ...
- каждой итерации заданной функции передаются новый элемент списка
 - принимает на вход список значений
 - применяет к каждому элементу списка некую функцию и возвращает новый список
 - преобразует список к единственному атомарному значению
3. При возможности использования MapReduce и Spark наилучшим считается: ...
- отдать предпочтение MapReduce
 - отдать предпочтение одному из них в зависимости от задачи
 - совместное использование
 - отдать предпочтение Spark
4. Особенности Spark

- разработан компанией Apache
- занимается хранением файлов и управлением ресурсами
- является взаимодополняющей системой к Hadoop
- инфраструктура анализа больших данных и кластерных вычислений
- 1. Принципы написания кода интерфейсов: ...
- Отсутствие записи состояния клиента
- Отделение клиента от сервера
- Кэшируемость
- Одноуровневость системы
- 2. Графовая база данных использует модель графа, чтобы представить данные в виде: ...
- ребер
- атомов
- вершин
- узлов
- связей
- 3. Принцип REST API, в котором отражено, что сервера могут располагаться на разных уровнях, при этом каждый сервер взаимодействует только с ближайшими уровнями и не связан запросами с другими, называется ...
- Stateless
- Layered System
- Uniform Interface
- Client-Server

5.3. Темы письменных работ (эссе, рефераты, курсовые работы и др.)

Примерная тематика рефератов

1. Определение и проблемы языков программирования.
2. Парадигмы языков программирования.
3. Императивные языки программирования. Функциональные языки программирования
- Объектно-ориентированные языки программирования. Логические языки программирования. Скриптовые языки программирования
6. Аппаратная организация компьютеров. Принцип программного управления. Структура виртуальной машины. Порядок функционирования виртуальной машины.
7. Трансляция и интерпретация. Этапы трансляции.
8. Нотации выражений. Порядок вычислений. Присваивание.
9. Базовые операторы. Операторы перехода. Поток управления. Составные операторы.
10. Условные операторы. Операторы циклов.
1. Качество синтаксиса. Синтаксические элементы.
2. Абстрактные синтаксические деревья. Способы обхода деревьев для различных нотаций арифметических выражений. Связь АСД с деревьями разбора.
3. Граматики. КС-грамматика. Форма Бэкуса-Наура.
4. Деревья разбора. Выводы. Синтаксическая неоднозначность.
5. Списки. Расширенная форма Бэкуса-Наура. Синтаксические схемы.
6. Синтезируемые атрибуты. Атрибутные грамматики.
7. Операционная и аксиоматическая семантика.
8. Правило упрощения. Правила вывода для последовательности, условного оператора и цикла while.
9. Определение инварианта цикла по индукции. Общая схема определения инварианта цикла.
10. Денотационная семантика.

Критерии оценки реферата

Отметка «отлично» за письменную работу, реферат, сообщение ставится, если изложенный в докладе материал:

- отличается глубиной и содержательностью, соответствует заявленной теме;
- четко структурирован, с выделением основных моментов;
- доклад сделан кратко, четко, с выделением основных данных;
- на вопросы по теме доклада получены полные исчерпывающие ответы.

Отметка «хорошо» ставится, если изложенный в докладе материал:

- характеризуется достаточным содержательным уровнем, но отличается недостаточной структурированностью;
- доклад длинный, не вполне четкий;
- на вопросы по теме доклада получены полные исчерпывающие ответы только после наводящих вопросов, или не на все вопросы.

Отметка «удовлетворительно» ставится, если изложенный в докладе материал:

- недостаточно раскрыт, носит фрагментарный характер, слабо структурирован;
- докладчик слабо ориентируется в излагаемом материале;
- на вопросы по теме доклада не были получены ответы или они не были правильными.

Отметка «неудовлетворительно» ставится, если:

- доклад не сделан;
- докладчик не ориентируется в излагаемом материале;
- на вопросы по выполненной работе не были получены ответы или они не были правильными.

Тематика практико-ориентированных заданий

1. Распознавание (классификация) изображений методами машинного обучения
2. Прогнозирование исходов спортивных матчей с помощью машинного обучения (футбол, хоккей, баскетбол...)
3. Анализ прогнозирования стихийных бедствий с помощью методов машинного обучения
4. Классификация клиентов компании методами машинного обучения

5.4. Оценочные средства для промежуточной аттестации

Критерии оценки

5 (отлично)

Устный ответ: Полный, исчерпывающий ответ, явно демонстрирующий глубокое понимание предмета и широкую эрудицию в оцениваемой области. Критический, оригинальный подход к материалу. Учитываются баллы, накопленные в течение семестра.

Письменный ответ: Качество исполнения всех элементов практико-ориентированного задания полностью соответствует всем требованиям. Учитываются баллы, накопленные в течение семестра.

4 (хорошо)

Устный ответ: Ответ полный, основанный на проработке всех обязательных источников информации. Подход к материалу ответственный, но стандартный. Учитываются баллы, накопленные в течение семестра.

Письменный ответ: Практико-ориентированное задание выполнено в достаточном объеме, но ограничивается только основными подходами. Учитываются баллы, накопленные в течение семестра.

3 (удовлетворительно)

Устный ответ: Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации.

Присутствуют небольшие пробелы в знаниях или несущественные ошибки. Учитываются баллы, накопленные в течение семестра

Письменный ответ: Практико-ориентированное задание выполнено в соответствии с заданием. Имеются отдельные несущественные ошибки или отступления от правил оформления работы. Учитываются баллы, накопленные в течение семестра.

2 (неудовлетворительно)

Устный ответ: Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные грубые ошибки.

Письменный ответ: Не учитываются баллы, накопленные в течение семестра. Отсутствие одного или нескольких обязательных элементов практико-ориентированного задания, либо многочисленные грубые ошибки в работе, либо грубое нарушение правил оформления или сроков представления работы. Не учитываются баллы, накопленные в течение семестра

Вопросы для промежуточной формы онтроля

1. Основные принципы анализа данных.
2. Извлечение и визуализация данных.
3. Этапы моделирования и формы представления данных.
4. Этапы подготовки данных к анализу.
5. Технологии KDD.
6. Технологии Data Mining.
7. Корреляционный анализ.
8. Ассоциативные правила. Значимость и поиск ассоциативных правил.
9. Кластеризация. Алгоритм кластеризации k-means.
10. Меры расстояний.
11. Сети Кохонена. Карты Кохонена.
12. Классификация данных.
13. Линейная и логистическая регрессия.
14. Статистические методы обработки данных.
15. Нейронные сети. Искусственный нейрон.
16. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей.
17. Анализ и прогнозирование временных рядов.
18. Модели прогнозирования.
19. Аналитические платформы.
20. Программное обеспечение в области анализа данных.
21. Возможности языка R для обработки данных.
22. Аналитика для неструктурированных данных - MapReduce и Hadoop.

Примерный тест для проведения промежуточной аттестации в форме тестирования

Критерии оценки для всех аттестаций, проходящих в форме компьютерного тестирования.

менее 60% - неудовлетворительно

60%-74 %- удовлетворительно

75%-89%- хорошо

90% и более - отлично

1. Базовый принцип реляционных БД, при котором если блок данных включается в базу, то он либо включается полностью, либо не включается вообще, называется: ...

- Долгосрочность
- Атомарность
- Изолированность
- Согласованность

2. Принципы BASE баз данных NOSQL: ...

- Согласованность в конечном счете
- Атомарность
- Неустойчивое состояние
- Базовая доступность

3. Задачи хранения и управления больших данных связаны с тем, что: ...

- работа с Big Data не позволяет использовать NoSQL
- Big Data работает только со структурированной информацией
- Big Data работает не может работать со структурированной информацией
- большой объем данных не позволяет легко хранить и управлять ими с помощью реляционных БД

4. При отсутствии функции GPS сотовые телефоны: ...

- не смогут определить местоположение
- достаточно точно определяют местоположение
- определяют местоположение с большой погрешностью
- используют сигналы базовых станций операторов мобильной связи для определения местоположения

5. Особенности Hadoop: ...

- популярный фреймворк
- содержит платный набор утилит
- разработан на Java
- предназначен для разработки и выполнения централизованных программ

6. Сравнение MapReduce и Spark: ...

- Spark обладает более широким спектром возможностей для работы с данными
- MapReduce имеет более статическую архитектуру
- Преимущество Spark – линейная обработка огромных наборов данных
- Spark имеет более динамическую архитектуру

7. Целостность и внутренняя непротиворечивость данных, называется ...

- Атомарная операция
- Масштабируемость
- Изолированность
- Согласованность данных

8. Процесс изучения наборов данных с целью получения выводов о содержащейся в них информации, все чаще с помощью специализированных систем и программного обеспечения, называется: ...

- Data Science
- Expert system
- Data Analytics
- Анализ данных

9. Для хранения и построения запросов графовых данных используются: ...

- SQL
- SPARQL
- графовые базы данных
- реляционные базы данных

10. К принципам REST API относятся: ...

- Casheable
- Starting with the Null Style
- Stateless
- Interface Laurence Null

11. Фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов ...

- Hadoop
- Reduce
- YARN
- HDFS

12. Примеры неструктурированных данных: ...

- посты в соц. сетях

- электронный справочник
 - сообщения электронной почты
 - реляционная база данных
13. Особенности графовых баз данных: ...
- эффективно работают с данными, имеющими сложную структуру
 - полезны в областях, связанных с социальными сетями и логистикой
 - предоставляют эффективный способ хранения и обработки связанных данных
 - обычно применяются для работы с данными, имеющими простую структуру
14. Компоненты, которые образуют экосистему Hadoop: ...
- HDFS
 - YARN, MapReduce
 - SSD
 - HBase
15. Решение задачи прогнозирования сводится к решению подзадач: ...
- анализ точности прогноза
 - анализ гомогенности
 - выбор модели прогнозирования
 - анализ адекватности прогноза
16. Типы данных для больших данных в: ...
- структурированные
 - неструктурированные
 - квантовые
 - на естественном языке
17. Базовый принцип Acid, при котором: когда в базе данных что-то изменяется, ничего не может происходить точно с одними и теми же данными точно в один момент ...
- Согласованность
 - Долгосрочность
 - Изолированность
 - Атомарность
18. xVIEW – это: ...
- обезличенные медицинские данные пациентов
 - датасеты о финансах и ценах на товары
 - содержание изображения сложных сцен со всего мира, аннотированные с помощью ограничительных рамок
 - набор воздушных снимков Земли с аннотациями
19. В REST API: ...
- код запросов остается на стороне клиента
 - код запросов и код для доступа к данным всегда на стороне сервера
 - код запросов и код для доступа к данным всегда на стороне сервера
 - код для доступа к данным на стороне сервера
20. К основным методам кластеризации относятся: ...
- Полиномиальная кластеризация
 - Послойная кластеризация
 - Минимальное покрывающее дерево
 - Выделение связанных компонент
21. Основные методы кластеризации: ...
- Минимальное покрывающее дерево
 - Гребневая
 - Иерархический
 - k-средних
22. Ценность информации, как ключевой параметр для оценки эффективности вложений в ее обработку, включает ответы на вопросы: ...
- Дают ли собираемые данные ответы на поставленные вопросы?
 - Какой объем данных нужно собрать за единицу времени?
 - Оправдываются ли затраты на внедрение аналитических механизмов и систем?
 - Способна ли компания извлекать пользу из собираемых данных?
23. Информационная иерархия, в которой каждый следующий уровень характеризуется большим уровнем зрелости (пригодностью к продолжению жизни) и кратно меньшим объемом сведений, называется ...
- CRISP-DM
 - JSON
 - Data Science
 - DIKW
24. При работе с большими данными нехватка оперативной памяти выражается в том, что ОС начинает выгружать блоки памяти на диск и: ...
- скорость обработки данных не меняется

- скорость работы с данными резко растет
 - ОС резко виснет
 - скорость работы с данными резко падает
25. Динамично развивающееся направление облачных вычислений, ориентированное прежде всего на веб-разработчиков, называется ...
- SaaS
 - IaaS
 - ZaaS
 - PaaS
26. Режимы работы онлайн-алгоритмов: ...
- мини-пакетное обучение
 - полнопакетное обучение
 - полупакетное обучение
 - онлайн-обучение
27. Специализированные расширения для языков командной строки: ...
- shell
 - Python
 - CMD
 - AWS CLI
28. Не относится к основным методам анализа больших данных ...
- прогнозирование
 - лемматизация
 - классификация
 - кластеризация
29. Примеры использования поиска ассоциаций: ...
- анализ посещений веб-страниц
 - регрессивный анализ
 - анализ ДНК живых организмов
 - анализ рыночной корзины
30. Особенности использования программных библиотек ...
- программное манипулирование ресурсами облака выполняется принципом
 - SDK должна содержать ключи учетной записи, которая будет иметь доступ к облаку
 - SDK представляет собой набор классов и методов для работы с ресурсами облака
 - невозможность создания программ, которые сами себе создают облачные ресурсы
31. Распределенные и облачные хранилища данных: ...
- Яндекс.Диск
 - Dropbox
 - MySQL
 - OneDrive
32. Библиотека, используемая при работе с большими данными для повышения скорости выполнения кода, компилирующая код непосредственно перед выполнением (JIT компиляция), называется ...
- Numba
 - Numexpr
 - NumPy
 - Hadoopy
33. Учитывая "Разнородность" данных самым важным считается: ...
- объединить данные в структуру, поддающуюся анализу
 - добиться однообразия данных, переводом их в буквенные знаки или цифры
 - добиться взаимосвязанной структуры данных
 - объединить разнородные данные в общую структуру
34. Программы, использующие MapReduce: ...
- выполняются на распределенных узлах кластера
 - выполняются на центральном узле последовательно
 - автоматически централизуются
 - автоматически распараллеливаются
35. Библиотека, используемая при работе с большими данными, помогающая решить проблемы нехватки памяти, которые могут возникнуть при использовании NumPy, позволяющая сохранять массивы и работать с ними в оптимальной сжатой форме, называется ...
- Blaze
 - Numba
 - Bcolz
 - Numexpr
36. Онлайн-алгоритмы: ...
- алгоритмы, не работающие с внешней памятью

- в которых данные образуются как функция от времени
 - являются общей техникой, когда невозможна тренировка по всему набору данных
 - противоположны пакетной технике обучения
37. На первых двух этапах – понимания бизнес-целей и начального изучения данных – специалист: ...
- создает наборы данных, которые можно использовать для анализа
 - знакомится с данными
 - осуществляет интеграцию источников из нескольких баз данных
 - пытается сформулировать цели проекта с точки зрения бизнеса
38. Теорема CAP. Если БД распределена по разным серверам, то она может обеспечить не более двух из трёх следующих свойств (найдите все 3 свойства): ...
- изолированность
 - согласованность
 - доступность
 - долгосрочность
 - устойчивость к разделению
39. Функция MapReduce, выполняющая сортировку и фильтрацию данных, организуя их в виде группы, генерирующая результат на основе пары ключ-значение, называется ...
- Oozie
 - Reduce
 - HBase
 - Map
40. Область, которая относится к коллективным процессам, теориям, концепциям, инструментам и технологиям, которые позволяют просматривать, анализировать и извлекать ценные знания и информацию из необработанных данных, называется: ...
- Expert system
 - Data Analytics
 - Data Science
 - Наука о данных
41. Архитектура Spark состоит из следующих основных компонентов: ...
- Spark RDD
 - Spark Duo
 - Spark Core
 - Spark SQL
42. К основным методам решения задачи регрессии относятся: ...
- Ридж
 - Дискриминационная
 - Линейная
 - Полиномиальная
43. Большая часть собранной информации Big Data в распределенной файловой системе состоит из: ...
- неструктурированных данных
 - текста, изображений, фотографий или видео
 - реляционных таблиц
 - структурированных данных
44. Лучшими, считаются модели: ...
- которые плавно вписываются в существующую практику
 - ориентированные на широкий круг пользователей, столкнувшихся разными проблемами
 - ориентированные на конкретных пользователей, столкнувшихся с четко обозначенной проблемой

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л1.1	Федин Ф.О., Федин Ф.Ф.	Анализ данных. Часть 1. Подготовка данных к анализу: учебное пособие	Москва: Московский городской педагогический университет, 2012	http://www.iprbookshop.ru/26444.html

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л1.2	Федин Ф.О., Федин Ф.Ф.	Анализ данных. Часть 2. Инструменты Data Mining: учебное пособие	Москва: Московский городской педагогический университет, 2012	http://www.iprbookshop.ru/26445.html
Л1.3	Васильев Е. П., Орешков В. И.	Интеллектуальный анализ данных в технологиях принятия решений: учебное пособие	Рязань: Рязанский государственный радиотехнический университет, 2023	https://www.iprbookshop.ru/134854.html
6.1.2. Дополнительная литература				
	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л2.1	Целых А. Н., Целых А. А., Котов Э. М.	Современные методы прикладной информатики в задачах анализа данных: учебное пособие по курсу «Методы интеллектуального анализа данных»	Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2021	https://www.iprbookshop.ru/117165.html
Л2.2	Александровская Ю. П.	Информационные технологии статистического анализа данных: учебно-методическое пособие	Казань: Казанский национальный исследовательский технологический университет, 2019	https://www.iprbookshop.ru/100535.html

6.3.1 Перечень программного обеспечения	
6.3.1.1	Kaspersky Endpoint Security для бизнеса СТАНДАРТНЫЙ
6.3.1.2	MS Office
6.3.1.3	MS WINDOWS
6.3.1.4	Яндекс.Браузер
6.3.1.5	Astra Linux
6.3.1.6	LibreOffice
6.3.1.7	MS Access
6.3.1.8	NVDA
6.3.1.9	Python
6.3.1.10	R for Windows
6.3.1.11	СППР Выбор
6.3.1.12	РЕД ОС
6.3.1.13	Deductor Academic
6.3.1.14	Loginom
6.3.2 Перечень информационных справочных систем	
6.3.2.1	Межвузовская электронная библиотека
6.3.2.2	Электронно-библиотечная система «Издательство Лань»
6.3.2.3	Электронно-библиотечная система IPRbooks
6.3.2.4	База данных «Электронная библиотека Горно-Алтайского государственного университета»
6.3.2.5	Гарант
6.3.2.6	КонсультантПлюс

7. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ	
	кейс-метод

	метод проектов	
--	----------------	--

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Номер аудитории	Назначение	Основное оснащение
322 А2	Компьютерный класс. Лаборатория информатики и информационно-коммуникативных технологий. Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Помещение для самостоятельной работы	Рабочее место преподавателя. Посадочные места обучающихся (по количеству обучающихся). Компьютеры, ученическая доска, подключение к сети Интернет

9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

лекции

В ходе лекционных занятий вести конспектирование учебного материала. Обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации, положительный опыт в ораторском искусстве. После каждой лекции студенту следует внимательно прочитать и разобрать конспект, при этом:

- Понять и запомнить все новые определения.
- Понять все математические выкладки и лежащие в их основе физические положения и допущения; воспроизвести все выкладки самостоятельно, не глядя в конспект.
- Выполнить или доделать выкладки, которые лектор предписал сделать самостоятельно (если таковые имеются).
- Если лектор предписал разобрать часть материала более подробно самостоятельно по доступным письменным или электронным источникам, то необходимо своевременно это сделать.
- При возникновении каких-либо трудностей с пониманием материала рекомендуется попросить помощи у своих одногруппников или сокурсников. Также можно обратиться за помощью к лектору.

практические/ лабораторные занятия

Подготовку к практическим занятиям следует начинать с глубокого усвоения пройденного материала и учебной литературы. Все задания к практическому занятию необходимо выполнять в соответствии с инструкцией, анализировать полученные в ходе занятия результаты по приведенной методике. Приступая к решению задачи, студент должен хорошо уяснить ее условие и исходя из уже полученных им знаний, установить, какие вопросы вытекают из содержания задачи. Решение задач должно быть мотивированным и обоснованным теоретически.

самостоятельная работа

Большое значение имеет самостоятельная работа с литературой, выработка рациональных приемов самообразования. Овладение рациональными приемами работы с литературными источниками является признаком сформированности у личности культуры умственного труда как одной из важнейших предпосылок дальнейшего самообразования. Если часть учебного материала отведена на самостоятельное изучение, то необходимо приступить к этому незамедлительно после указания преподавателя и освоить материал в отведенные им сроки. Материал следует изучить по доступным письменным и электронным источникам, о которых сообщит преподаватель.