Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Борисова Виктория Валерьевна
Должность: Ректор
Дата подписания: 18.06.2025 19:43:13
Уникальный программный ключ:
8d665791f4048370b679b22cf26583a2f341522e

Негосударственное образовательное частное учреждение высшего образования
«МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ
ПСИХОЛОГО-ПЕДАГОГИЧЕСКИХ ИННОВАЦИЙ»

УТВЕРЖДАЮ:
Ректор
НОЧУ ВО «МУППИ»
_____ В.В. Борисова
подпись
«13» апреля 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Введение в технологии Big Data

направление подготовки
38.03.02 Менеджмент

Профиль подготовки:
Менеджмент цифровых технологий

Квалификация (степень) выпускника – *бакалавр*

Форма обучения
очно-заочная

Москва 2023 г.

1. Перечень планируемых результатов изучения дисциплины, соотнесенных с планируемыми результатами освоения образовательной программы

В рамках освоения основной профессиональной образовательной программы бакалавриата обучающийся должен овладеть следующими результатами обучения по дисциплине «Введение в технологии Big Data»

Код и наименование компетенций	Индикаторы достижения компетенции
ПК-2. Способен обосновывать организационно-управленческие решения в области цифровых технологий	ИПК-2.1. Способен формировать альтернативные решения на основе аналитических данных. ИПК-2.2. Способен проводить анализ и обосновывать выбор решения в области цифровых технологий. ИПК-2.3. Способен проводить оценку ресурсов, необходимых для реализации цифровых решений.

2. Место дисциплины в структуре ОПОП

Учебная дисциплина Б1.В.05 «Введение в технологии Big Data» относится к части, формируемой участниками образовательных отношений цикла Б.1 «Дисциплины (модули)».

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся

Общая трудоемкость (объем) дисциплины «Введение в технологии Big Data» составляет 3 зачетные единицы.

Объём дисциплины по видам учебных занятий (в часах) – очно-заочная форма обучения

Вид учебной работы	Всего часов	Семестры	
		7	-
Аудиторные занятия (всего)	**24**	**24**	-
В том числе:	-	-	-
Лекции	12	12	-
Практические занятия (ПЗ)	12	12	-
Семинары (С)	-	-	-
Лабораторные работы (ЛР)	-	-	-
Самостоятельная работа (всего)	**84**	**84**	-
В том числе:	-	-	-
Курсовой проект (работа)	-	-	-
Расчетно-графические работы	-	-	-
Реферат	-	-	-
Подготовка к практическим занятиям	66	66	-
Тестирование	18	18	-
Вид промежуточной аттестации – зачет			-
Общая трудоемкость час / зач. ед.	**108/3**	**108/3**	-

4. Содержание дисциплины

4.1. Разделы дисциплины и трудоемкость по видам учебных занятий

Очно-заочная форма обучения

№ п/п	Раздел/тема Дисциплины	Общая трудоёмкость	Виды учебных занятий, включая самостоятельную работу обучающихся, час		
			Контактная работа		Самостоятельная работа обучающихся
		Всего	лекции	практические занятия	
1.	**Раздел 1. Технологии анализа данных**	54	6	6	42
2.	**Раздел 2. Интеллектуальный анализ данных**	54	6	6	42
Всего		**108**	**12**	**12**	**84**
Зачет		**-**	**-**	**-**	**-**
Итого		**108**	**12**	**12**	**84**

4.2. Содержание разделов дисциплины

Раздел 1. Технологии анализа данных

Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.

Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.

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.

Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования.

Начало работы. Понятие сценария и узла обработки. Консолидация данных. Трансформация данных. Визуализация данных.

Раздел 2. Интеллектуальный анализ данных

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.

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.

Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов.

Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура.

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.

Ассоциативные правила. Поиск ассоциативных правил.

Кластеризация. Алгоритм кластеризации k-means.

Прогнозирование с помощью линейной регрессии.

Классификация с помощью нейросети.

Классификация с помощью деревьев решений.

4.3. Практические занятия / лабораторные занятия

Очно-заочная форма обучения

Занятия 1-2. Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. Начало работы. Понятие сценария и узла обработки. Консолидация данных. Трансформация данных. Визуализация данных.

Занятия 3-5. Ассоциативные правила. Поиск ассоциативных правил. Кластеризация. Алгоритм кластеризации k-means. Прогнозирование с помощью линейной регрессии. Классификация с помощью нейросети. Классификация с помощью деревьев решений.

5. Учебно-методическое и информационное обеспечение дисциплины

5.1. Основная литература

1. Радченко, И. А. Технологии и инфраструктура Big Data : учебное пособие / И. А. Радченко, И. Н. Николаев. — Санкт-Петербург : НИУ ИТМО, 2018. — 52 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/136430

2. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452

3. Управление данными : учебник / Ю. Ю. Громов, О. Г. Иванова, А. В. Яковлев, В. Г. Однолько ; Тамбовский государственный технический университет. – Тамбов : Тамбовский государственный технический университет (ТГТУ), 2015. – 192 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=444642

5.2. Дополнительная литература

1. Березовская, Е. А. Работа с сервисом бизнес-аналитики Yandex DataLens : учебное пособие : [16+] / Е. А. Березовская, С. В. Крюков ; Южный федеральный университет, Экономический факультет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2022. – 94 с. : ил., табл. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=698669

2. Борисова, Л. Р. Математика и анализ данных с поддержкой MS Excel и языка R : учебное пособие : [16+] / Л. Р. Борисова, Н. И. Светлова, И. Ю. Седых ; под ред. И. Ю. Седых ; Финансовый университет при Правительстве Российской Федерации. – Москва : Прометей, 2023. – 728 с. : ил., табл., схем., граф. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=701041

5.3. Лицензионное программное обеспечение

1. Microsoft Windows 10 Pro
2. Microsoft Office 2007

5.4. Современные профессиональные базы данных и информационные справочные системы

1. ЭБС «Университетская библиотека онлайн» https://biblioclub.ru/index.php
2.Научная электронная библиотека http://www.elibrary.ru
3. Федеральный образовательный портал «Экономика. Социология. Менеджмент» http://ecsocman.hse.ru
4. Административно-управленческий портал http://www.aup.ru/
5. Официальный интернет-портал правовой информации http://pravo.gov.ru.
6. Компьютерные информационно-правовые системы «Консультант» http://www.consultant.ru , «Гарант» http://www.garant.ru.
7. Научная электронная библиотека «КиберЛенинка» https://cyberleninka.ru/
8. Электронно-библиотечная система «Лань» https://e.lanbook.com/

6. Материально-техническое обеспечение дисциплины

1. Лекционная аудитория, аудитория для групповых и индивидуальных консультаций, оснащенная комплектом мебели для учебного процесса, учебной доской, персональным компьютером, плазменной панелью.
2. Аудитория информационных технологий, оснащенная комплектом мебели для учебного процесса, учебной доской, персональными компьютерами с возможностью подключения к сети «Интернет».
3. Аудитория для самостоятельной работы студентов, оснащенная комплектом мебели для учебного процесса, учебной доской, персональными компьютерами с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде Университета.

7. Методические рекомендации по организации изучения дисциплины

7.1. Методические рекомендации преподавателю

Данный раздел настоящей рабочей программы предназначен для начинающих преподавателей и специалистов-практиков, не имеющих опыта преподавательской работы.

Дисциплина «Введение в технологии Big Data» является дисциплиной, формирующей у обучающихся частично компетенцию ПК-2. В условиях конструирования образовательных систем на принципах компетентностного подхода произошло концептуальное изменение роли преподавателя, который, наряду с традиционной ролью носителя знаний, выполняет функцию организатора научно-поисковой работы обучающегося, консультанта в процедурах выбора, обработки и интерпретации информации, необходимой для практического действия и дальнейшего развития, что должно обязательно учитываться при проведении лекционных и практических занятий по дисциплине «Введение в технологии Big Data».

Преподавание теоретического (лекционного) материала по дисциплине «Введение в технологии Big Data» осуществляется на основе междисциплинарной интеграции и четких междисциплинарных связей в рамках образовательной программы и учебного плана по направлению 38.03.02 Менеджмент.

Подробное содержание отдельных разделов дисциплины «Введение в технологии Big Data» рассматривается в п.5 рабочей программы.

Методика определения итогового семестрового рейтинга обучающегося по дисциплине «Введение в технологии Big Data» представлена в составе ФОС по дисциплине в п 8 рабочей программы.

Примерные варианты тестовых заданий для текущего контроля и перечень вопросов к зачету по дисциплине также представлены в п 8 рабочей программы.

Перечень основной и дополнительной литературы и нормативных документов, необходимых в ходе преподавания дисциплины «Введение в технологии Big Data», приведен в п.8 настоящей рабочей программы. Преподавателю следует ориентировать обучающихся на использование при подготовке к промежуточной аттестации оригинальной версии нормативных документов, действующих в настоящее время.

7.2. Методические указания обучающимся

Получение углубленных знаний по дисциплине достигается за счет активной самостоятельной работы обучающихся. Выделяемые часы целесообразно использовать для знакомства с учебной и научной литературой по проблемам дисциплины, анализа научных концепций.

В рамках дисциплины предусмотрены различные формы контроля уровня достижения обучающимися заявленных индикаторов освоения компетенций. Форма текущего контроля – активная работа на практических занятиях, подготовка и прохождение тестирования. Формой промежуточного контроля по данной дисциплине является зачет, в ходе которого оценивается уровень достижения обучающимися заявленных индикаторов освоения компетенций.

Методические указания по освоению дисциплины.

Лекционные занятия проводятся в соответствии с содержанием настоящей рабочей программы и представляют собой изложение теоретических основ дисциплины.

Посещение лекционных занятий является обязательным.

Конспектирование лекционного материала допускается как письменным, так и компьютерным способом.

Регулярное повторение материала конспектов лекций по каждому разделу в рамках подготовки к текущим формам аттестации по дисциплине является одним из важнейших видов самостоятельной работы студента в течение семестра, необходимой для качественной подготовки к промежуточной аттестации по дисциплине.

Проведение практических занятий по дисциплине «Введение в технологии Big Data» осуществляется в следующих формах:

- анализ правовой базы, регламентирующей деятельность организаций различных организационно-правовых форм;
- опрос по материалам, рассмотренным на лекциях и изученным самостоятельно по рекомендованной литературе;
- решение типовых расчетных задач по темам;
- анализ и обсуждение практических ситуаций по темам.

Посещение практических занятий и активное участие в них является обязательным.

Подготовка к практическим занятиям обязательно включает в себя изучение конспектов лекционного материала и рекомендованной литературы для адекватного понимания условия и способа решения заданий, запланированных преподавателем на конкретное практическое занятие.

Методические указания по выполнению различных форм внеаудиторной самостоятельной работы

Изучение основной и дополнительной литературы, а также нормативно-правовых документов по дисциплине проводится на регулярной основе в разрезе каждого раздела в соответствии с приведенными в п.5 рабочей программы рекомендациями для подготовки к промежуточной аттестации по дисциплине «Введение в технологии Big Data». Список основной и дополнительной литературы и обязательных к изучению нормативно-правовых документов по дисциплине приведен в п.7 настоящей рабочей программы. Следует отдавать предпочтение изучению нормативных документов по соответствующим разделам дисциплины по сравнению с их адаптированной интерпретацией в учебной литературе.

Решение задач в разрезе разделов дисциплины «Введение в технологии Big Data» является самостоятельной работой обучающегося в форме домашнего задания в случаях

недостатка аудиторного времени на практических занятиях для решения всех задач, запланированных преподавателем, проводящим практические занятия по дисциплине.

Методические указания по подготовке к промежуточной аттестации

Промежуточная аттестация по дисциплине «Введение в технологии Big Data» проходит в форме зачета. Примерный перечень вопросов к зачету по дисциплине «Введение в технологии Big Data» и критерии оценки ответа обучающегося на зачете для целей оценки достижения заявленных индикаторов сформированности компетенций приведены в составе ФОС по дисциплине в п 8 рабочей программы.

Обучающийся допускается к промежуточной аттестации по дисциплине независимо от результатов текущего контроля.

8. Фонд оценочных средств по дисциплине

8.1. Перечень компетенций с указанием этапов их формирования в процессе освоения дисциплины. Формы контроля формирования компетенций

Код и наименование компетенций	Индикаторы достижения компетенции	Форма контроля	Этапы формирования (разделы дисциплины)
ПК-2. Способен обосновывать организационно-управленческие решения в области цифровых технологий	ИПК-2.1. Способен формировать альтернативные решения на основе аналитических данных. ИПК-2.2. Способен проводить анализ и обосновывать выбор решения в области цифровых технологий. ИПК-2.3. Способен проводить оценку ресурсов, необходимых для реализации цифровых решений.	Промежуточный контроль: зачет Текущий контроль: опрос на практических занятиях; тестирование	Разделы 1-9

8.2. Показатели и критерии оценивания компетенций при изучении дисциплины, описание шкал оценивания

8.2.1 Критерии оценки ответа на зачете

(формирование компетенции ПК-2 индикаторы ИПК-2.1, ИПК-2.2, ИПК-2.3)

«зачтено»

обучающийся демонстрирует системные теоретические знания, практические навыки, владеет терминами, делает аргументированные выводы и обобщения, приводит примеры, показывает свободное владение монологической речью и способность быстро реагировать на уточняющие вопросы.

обучающийся демонстрирует прочные теоретические знания, практические навыки, владеет терминами, делает аргументированные выводы и обобщения, приводит примеры, показывает свободное владение монологической речью, но при этом делает несущественные ошибки, которые быстро исправляет самостоятельно или при незначительной коррекции преподавателем.

обучающийся демонстрирует неглубокие теоретические знания, проявляет слабо сформированные навыки анализа явлений и процессов, недостаточное умение делать аргументированные выводы и приводить примеры, показывает недостаточно свободное владение монологической речью, терминами, логичностью и последовательностью изложения, делает ошибки, которые может исправить только при коррекции преподавателем.

«не зачтено»

обучающийся демонстрирует незнание теоретических основ предмета, отсутствие практических навыков, не умеет делать аргументированные выводы и приводить примеры, показывает слабое владение монологической речью, не владеет терминами, проявляет отсутствие логичности и последовательности изложения, делает ошибки, которые не может исправить даже при коррекции преподавателем, отказывается отвечать на дополнительные вопросы.

8.2.2 Критерии оценки работы обучающегося на практических занятиях

(формирование компетенции ПК-2 индикаторы ИПК-2.1, ИПК-2.2, ИПК-2.3)

«5» (отлично): выполнены все практические задания, предусмотренные практическими занятиями, обучающийся четко и без ошибок ответил на все контрольные вопросы, активно работал на практических занятиях.

«4» (хорошо): выполнены все практические задания, предусмотренные практическими занятиями, обучающийся с корректирующими замечаниями преподавателя ответил на все контрольные вопросы, достаточно активно работал на практических занятиях.

«3» (удовлетворительно): выполнены все практические задания, предусмотренные практическими занятиями с замечаниями преподавателя; обучающийся ответил на все контрольные вопросы с замечаниями.

«2» (неудовлетворительно): обучающийся не выполнил или выполнил неправильно практические задания, предусмотренные практическими занятиями; обучающийся ответил на контрольные вопросы с ошибками или не ответил на контрольные вопросы.

8.2.3 Критерии оценки тестирования

(формирование компетенции ПК-2 индикаторы ИПК-2.1, ИПК-2.2, ИПК-2.3)

Тестирование оценивается в соответствии с процентом правильных ответов, данных обучающимся на вопросы теста.

Стандартная шкала соответствия результатов тестирования выставляемой балльной оценке:

- «отлично» - свыше 85% правильных ответов;
- «хорошо» - от 70,1% до 85% правильных ответов;
- «удовлетворительно» - от 55,1% до 70% правильных ответов;
- от 0 до 55% правильных ответов – «неудовлетворительно»

«5» (отлично): тестируемый демонстрирует системные теоретические знания, владеет терминами и обладает способностью быстро реагировать на вопросы теста.

«4» (хорошо): тестируемый в целом демонстрирует системные теоретические знания, владеет большинством терминов и обладает способностью быстро реагировать на вопросы теста.

«3» (удовлетворительно): системные теоретические знания у тестируемого отсутствуют, он владеет некоторыми терминами и на вопросы теста реагирует достаточно медленно.

«2» (неудовлетворительно): системные теоретические знания у тестируемого отсутствуют, терминологией он не владеет и на вопросы теста реагирует медленно.

8.2.4. Итоговое соответствие балльной шкалы оценок и уровней сформированности компетенций по дисциплине:

Уровень сформированности компетенции	Оценка	Пояснение

Высокий	«зачтено»	теоретическое содержание и практические навыки по дисциплине освоены полностью; все предусмотренные программой обучения учебные задания выполнены на высоком уровне; компетенции сформированы
Средний	«зачтено»	теоретическое содержание и практические навыки по дисциплине освоены полностью; все предусмотренные программой обучения учебные задания выполнены с незначительными замечаниями; компетенции в целом сформированы
Удовлетвор-тельный	«зачтено»	теоретическое содержание и практические навыки по дисциплине освоены частично, но пробелы не носят существенного характера; большинство предусмотренных программой обучения учебных задач выполнено, но в них имеются ошибки; компетенции сформированы частично
Неудовлетворительный	«не зачтено»	теоретическое содержание и практические навыки по дисциплине не освоены; большинство предусмотренных программой обучения учебных заданий либо не выполнено, либо содержит грубые ошибки; дополнительная самостоятельная работа над материалом не приводит к какому-либо значимому повышению качества выполнения учебных заданий; компетенции не сформированы

8.3. Методические материалы (типовые контрольные задания), определяющие результаты обучения по дисциплине, соотнесенные с индикаторами достижения

Контрольные задания, применяемые в рамках текущего контроля и промежуточной аттестации по дисциплине, носят универсальный характер и предусматривают возможность комплексной оценки всего набора заявленных по данной дисциплине индикаторов сформированности компетенций.

8.3.1. Текущий контроль (работа на практических занятиях)

(формирование компетенции ПК-2 индикаторы ИПК-2.1, ИПК-2.2, ИПК-2.3)

Примеры заданий для практических занятий

Кейс 1. Проверить гипотезу о значимом отличии среднего балла за экзамены десятом и одиннадцатом классах, используя критерий Стьюдента и критерий Манна Уитни.
Построить диаграммы «ящик с усами» для школьников, имеющих разные хобби.
Построить диаграмму «дерево-листья». Данные находятся в файле тесты Школа.txt.
Задачу решить в R и в SPSS.
Построить задачу классификации хобби в зависимости от результатов тестирования.
Задачу классификации решить с помощью деревьев решений в R.

Кейс 2. Создать две случайные последовательности двух случайных величинах, размером в 200 наблюдений, полученных с помощью генераторов нормально распределенных случайных чисел, имеющих одинаковое математическое ожидание, равное 5 и ско, соответственно 1 и 2.
- Проверить гипотезу о равенстве математических ожиданий и дисперсий данных величин.
- Изменить генератор, добавив в первый генератор смещение математического

ожидания. Вновь проверить статистическую гипотезу.
- Проверить гипотезы о нормальном законе распределения.
- Найти сумму пяти случайных величин, равномерно распределенных на интервале 0, 2. Проверить гипотезу о нормальном законе распределения суммы.
Задачу решить с помощью статистических критериев в R. Построить вероятностные и квантиль-квантиль графики.

Кейс 3. Решить задачу кластерного анализа для файла Семейное положение.txt. при решении задачи кластерного анализа:
- определить склонность к кластеризации;
- определить лучшую метрику иерархической кластеризации;
- выполнить иерархическую кластеризацию;
- определить состав и центроиды кластеров;
- Решить задачу кластеризации методом k-средних;
- Выполнить интерпретацию полученных кластеров;
- Визуализировать полученную кластеризацию;
- Задачу решить в RStudio и в SPSS.

Кейс 4. В наборе Animals библиотеки cluster имеются данные о 20 животных. Заданы 6 бинарных признаков: теплокровные/нетеплокровные; летают/не летают; позвоночный/беспозвоночный; находящихся под угрозой вымирания; живущих в группах. Решить задачу кластерного анализа наблюдений в SPSS и в R. Использовать иерархическую кластеризацию и кластеризацию методом k-средних.

8.3.2. Текущий контроль (тестирование)
(формирование компетенции ПК-2 индикаторы ИПК-2.1, ИПК-2.2, ИПК-2.3)

Примерные варианты тестовых заданий

Задание 1. Сколько Петабайт в Зеттабайте?
Задание 2. Укажите фактор, способствовавший появлению тренда больших данных
- Маркетинговые кампании крупных корпораций
- Снижение издержек на хранение данных
- Появление новых технологий обработки потоковых данных
- Выпуск баз данных с обработкой данных в памяти
Задание 3. Определение больших данных: 1. Выберите верный ответ
- Большие данные – это обработка или хранение более 1 Тб информации.
- Проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки сущностная обработка данных затруднена или невозможна.
- Большие данные – это огромная PR-акция крупных вендоров и не более того.
- Большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект.
Задание 4. Выберите неверный ответ:
- Большие данные – это данные объёма свыше 1 Тб
- Проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки сущностная обработка данных затруднена или невозможна.
- Большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров.
- Большие данные как правило не структурированы.
Задание 5. Отметьте те из вариантов, в которых данные структурированы:
- Данные о продажах компании, представленные в виде помесячных отчётов в формате

MS Word. ii. Таблица с ежедневными показаниями температуры помещения за год в файле формата csv.

- Текст педагогической поэмы А.С. Макаренко, представленный в формате PDF.
- Библиотека фильмов, представленных в формате mpeg4 на одном жестком диске.

Задание 6. Выберите неверное высказывание:

- Большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных.
- Увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации.
- Удешевление систем хранения на единицу информации привело к росту рынка больших данных.
- Большое разнообразие источников данных

Задание 7. Отметьте неверное понимание Variety в контексте характеристик Big Data:

- Высокая скорость генерирования данных.
- Разные типы данных в колонках таблиц реляционных СУБД.
- Разнообразие отраслей, являющихся источниками данных.
- Разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные.

12

ЗАДАНИЕ 8. Коэффициент парной корреляции характеризует тесноту ____ связи между _____

переменными.

ВАРИАНТЫ ОТВЕТОВ:

1) линейной … несколькими 2) нелинейной … несколькими

3) линейной … двумя 4) нелинейной … двумя

ЗАДАНИЕ 9. Установите соответствие между наименованиями элементов уравнения Y=b0+b1X+e и их буквенными обозначениями:

1. параметры регрессии
2. объясняющая переменная
3. объясняемая переменная
4. случайные отклонения

ВАРИАНТЫ ОТВЕТОВ:

A) Y B) b0, b1

C) X D) e

ЗАДАНИЕ 10. Для линейного уравнения регрессии метод наименьших квадратов используется при оценивании параметров…

ВАРИАНТЫ ОТВЕТОВ:

1) B 2) y

3) X 4) a

ЗАДАНИЕ 11.Сколько параметров содержит парное линейное уравнение регрессии?

ВАРИАНТЫ ОТВЕТОВ:

А) 1

Б) 2

В) 3

Г) 4

ЗАДАНИЕ 12. При выполнении предпосылок МНК оценки параметров регрессии обладают

свойствами:

ВАРИАНТЫ ОТВЕТОВ:

1) достоверность 2) эффективность

3) несмещенность 4) несостоятельность

ЗАДАНИЕ 13. Как влияет увеличение объема выборки на величину остаточной дисперсии

случайной величины?

ВАРИАНТЫ ОТВЕТОВ:
13
А) Никак.
Б) Остаточная дисперсия увеличивается.
В) Остаточная дисперсия уменьшается.
Г) Результат зависит от конкретного вида случайной величины.
ЗАДАНИЕ 14. В кластерном анализе используются методы объединения …
ВАРИАНТЫ ОТВЕТОВ:
1) Ближнего соседа 2) Дальнего соседа
3) Среднего соседа 4) центроидный метод
ЗАДАНИЕ 15. В кластерном анализе для определения близости между кластерами используются метрики …
ВАРИАНТЫ ОТВЕТОВ:
1) Эвклидово расстояние 2) Куб Эвклидова расстояния
3) Взвешенное эвклидово расстояние 4) Квадрат Эвклидова расстояния
ЗАДАНИЕ 16.В дискриминантном анализе обучающая выборка используется для …
ВАРИАНТЫ ОТВЕТОВ:
1) Выявления значимых признаков 2) Выявления аномального измерения
3) Разделения объектов на классы 4) Выбора вида модели

8.3.3. Промежуточный контроль (вопросы к зачету)
(формирование компетенции ПК-2 индикаторы ИПК-2.1, ИПК-2.2, ИПК-2.3)

Примерные вопросы к зачету

1. Дать определение понятия «большие данные». Перечислить свойства больших данных. Перечислить основные сквозные цифровые технологии.

2. Указать роль аналитика по данным (Data Scientist). Сформулировать ключевые компетенции аналитика. Выделить отличия BI от Data Science.

3. Сформулировать парадигма Map Reduce. Описать принцип работы. Нарисовать диаграмму. Перечислить слабые и сильные стороны. Обозначить области применимости.

Привести примеры использования.

4. Перечислить задачи систем поддержки принятия решений. OLTP и OLAP-системы.

Принципы построения информационных хранилищ. Модели информационных хранилищ.

Многомерная модель данных. Правила Кодда.

5. Дать характеристику размерностных моделей, моделей MOLAP, ROLAP, HOLAPсистемы.

6. Характеризовать витрины данных.

7. Сделать обзор ETL (Extract Transform Load) технологии.

8. Характеризовать стандарты Data Mining, стандарты CWM, CRISP, PMML.

9. Рассмотреть содержание и этапы методология KDD. Дать характеристику задачи предобработки данных.

10. Рассмотреть модель распределенных вычислений. Map Reduce. Реализацию Map Reduce на Python

11. Сделать обзор языков Python. Описать основы языка Python. Привести примеры среды разработки. Сделать обзор платформы Anaconda.

12. Продемонстрировать организацию использования пакета Anaconda.

13. Характеризовать основы синтаксиса языка, переменных, ключевых слов языка, основ программирования на языке Python.

14. Характеризовать типы данных языка Python, функциональности для работы с данными. Описать организацию установки пакетов научных вычислений на Python, установки Keras.

15. Сделать обзор организации решении задач анализа данных. Дать определение понятия тензора, скаляров, векторов, матриц, тензоров третьего и высшего рангов, временных рядов или последовательностей. Рассмотреть операции над тензорами.

16. Характеризовать фреймворк Keras, библиотеки Google TensorFlow

17. Определить содержание задач кластерного анализа, понятие кластера, параметров кластера. Объяснить меры близости между кластерами, используемые метрики кластерного анализа.

18. Характеризовать базовые алгоритмы кластеризации. Рассмотреть иерархическую кластеризацию, понятие дендрограммы, организацию ее использования.

19. Характеризовать метод К-средних. Раскрыть понятие центроида, профиля кластеров. Связать понятия кластерного и регрессионного анализа. Описать организацию решения задач кластерного анализ на Python.

20. Дать определение задачи классификации. Перечислить методы классификации. Дать общую характеристику классификационного анализа с обучением. Продемонстрировать решение задачи классификации в Orange.

21. Объяснить основное содержание метода деревьев решений. Привести примеры деревьев решений. Характеризовать алгоритмы построения деревьев решений. Показать организацию решения задачи классификации в Python.

22. Характеризовать метод логистической регрессии, организацию ее решения в Python. Привести пример решения задачи различными методами классификации с помощью Orange. Исследовать результаты классификации различными методами.

23. Объяснить содержание метода k-ближайших соседей. Продемонстрировать решение задачи и сравнить с результатами решения другими методами.

24. Сделать обзор инструментов, используемых при оценке качества задач классификации. Таблица сопряженности (матрица ошибок). Дать определение понятий чувствительности и специфичности, ROC-кривой, критериев оценки качества классификации.

25. Дать определение понятия нейронной сети. Характеризовать архитектуру нейронной сети.

26. Сделать обзор основ машинного обучения. Характеризовать организацию оценки моделей машинного обучения, предназначение тренировочных, проверочных и контрольных данных, понятие функции потерь.

27. Характеризовать механизм нейронных сетей на основе обучения, метод обратного распространение ошибки.

28. Продемонстрировать решение задач классификации, регрессии, прогнозирования с помощью нейронных сетей в Python, Orange.

29. Дать определения понятий поверхностного и глубокого обучения, глубокого обучения в технологиях компьютерного зрения.

30. Характеризовать содержание глубокого обучения для текста и последовательностей. Сделать обзор рекуррентных нейронных сетей, слоев LSTM, GRU.