

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Агабекян Раиса Левоновна
Должность: ректор
Дата подписания: 07.02.2024 08:44:30
Уникальный программный ключ:
4237c7ccb9b9e111bbaf1f4fcd9201d015c4dbaa123456789012345678901234

Негосударственное аккредитованное некоммерческое частное образовательное учреждение высшего образования «Академия маркетинга и социально-информационных технологий – ИМСИТ» (г. Краснодар)
(НАН ЧОУ ВО Академия ИМСИТ)

УТВЕРЖДАЮ

Проректор по учебной работе, доцент

Н.И. Севрюгина

20.11.2023

Б1.В.08

**Системы сбора и анализа больших данных
рабочая программа дисциплины (модуля)**

Закреплена за кафедрой	Кафедра математики и вычислительной техники	
Учебный план	45.03.04 Интеллектуальные системы в гуманитарной сфере	
Квалификация	бакалавр	
Форма обучения	очная	
Общая трудоемкость	6 ЗЕТ	
Часов по учебному плану	216	Виды контроля в семестрах:
в том числе:		экзамены 7
аудиторные занятия	96	
самостоятельная работа	84	
контактная работа во время промежуточной аттестации (ИКР)	0	
часов на контроль	34,7	

Распределение часов дисциплины по семестрам

Семестр (<Курс>. <Семестр на курсе>)	7 (4.1)		Итого	
	УП	РП		
Неделя	15 5/6			
Вид занятий	УП	РП	УП	РП
Лекции	32	32	32	32
Практические	64	64	64	64
Контактная работа на аттестации (в период экз. сессий)	0,3	0,3	0,3	0,3
Консультации перед экзаменом	1	1	1	1
В том числе в форме практ. подготовки	8		8	
Итого ауд.	96	96	96	96
Контактная работа	97,3	97,3	97,3	97,3
Сам. работа	84	84	84	84
Часы на контроль	34,7	34,7	34,7	34,7
Итого	216	216	216	216

Программу составил(и):

к.т.н., доцент, Сорокина Виктория Владимировна

Рецензент(ы):

директор АО «ЮГ-СИСТЕМА», Глебов О.В.; д.т.н., профессор, профессор кафедры информационных систем и программирования КубГТУ, Видовский Л.А

Рабочая программа дисциплины

Системы сбора и анализа больших данных

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 45.03.04 Интеллектуальные системы в гуманитарной сфере (приказ Минобрнауки России от 24.04.2018 г. № 324)

составлена на основании учебного плана:

45.03.04 Интеллектуальные системы в гуманитарной сфере

утвержденного учёным советом вуза от 20.11.2023 протокол № 3.

Рабочая программа одобрена на заседании кафедры

Кафедра математики и вычислительной техники

Протокол от 13.10.2023 г. № 3

Зав. кафедрой Исикова Наталья Павловна

Согласовано с представителями работодателей на заседании НМС, протокол № 3 от 20.11.2023.

Председатель НМС проф. Павелко Н.Н.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
1.1	заключается в получении обучающимися знаний и навыков, необходимых для эффективного использования больших объемов данных. Изучение систем сбора больших данных поможет обучающимся понять, как эффективно собирать, хранить и обрабатывать большие объемы данных, а также как анализировать эти данные для получения ценной информации при принятии обоснованных решений.
Задачи: - изучение основных принципов работы систем сбора больших данных - развитие навыков работы с данными и статистическим анализом - понимание современных тенденций и направлений в области данных и аналитики - умение работать с инструментами и технологиями, используемыми для обработки и анализа больших данных	

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	
Цикл (раздел) ОП:	Б1.В
2.1	Требования к предварительной подготовке обучающегося:
2.1.1	Теория вероятностей и математическая статистика
2.1.2	Язык программирования Python
2.1.3	Интеллектуальный анализ данных и основы машинного обучения
2.1.4	Обработка больших данных
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:
2.2.1	Производственная практика: Преддипломная практика
2.2.2	Производственная практика: Технологическая (проектно-технологическая) практика
2.2.3	Подготовка к процедуре защиты и защита выпускной квалификационной работы
2.2.4	Производственная практика: научно-исследовательская работа

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ, ИНДИКАТОРЫ ИХ ДОСТИЖЕНИЯ и планируемые результаты обучения	
ПК-7: Способен проводить анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	
ПК-7.1: Знает современные методы и инструментальные средства анализа больших данных	
Знать	
Уровень 1	Минимальный необходимый уровень знаний современных методов и инструментальных средств анализа больших данных
Уровень 2	Уровень знаний современных методов и инструментальных средств анализа больших данных в объеме, соответствующем программе подготовки, допущено несколько негрубых ошибок
Уровень 3	Уровень знаний современных методов и инструментальных средств анализа больших данных в объеме, соответствующем программе подготовки, без ошибок
ПК-7.2: Умеет проводить анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования, разрабатывать и оценивать модели больших данных	
Уметь	
Уровень 1	Продемонстрированы основные умения в проведении анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования, разрабатывать и оценивать модели больших данных в рамках избранных видов профессиональной деятельности, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме
Уровень 2	Продемонстрированы все основные умения в проведении анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования, разрабатывать и оценивать модели больших данных в рамках избранных видов профессиональной деятельности, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочётами
Уровень 3	Продемонстрированы все основные умения в проведении анализ больших данных в соответствии с утвержденными требованиями к результатам аналитического исследования, разрабатывать и оценивать модели больших данных в рамках избранных видов профессиональной деятельности, решены все основные задачи с отдельными несущественными недочётами, выполнены все задания в полном объеме
ПК-7.3: Владеет навыками выбора методов и инструментальных средств анализа больших данных для проведения аналитических работ	
Владеть	
Уровень 1	Имеется минимальный набор навыков выбора методов и инструментальных средств анализа больших данных для проведения аналитических работ с негрубыми ошибками и некоторыми недочётами
Уровень 2	Продемонстрированы базовые навыки выбора методов и инструментальных средств анализа больших данных для проведения аналитических работ с некоторыми недочётами
Уровень 3	Продемонстрированы навыки выбора методов и инструментальных средств анализа больших данных для проведения аналитических работ без ошибок и недочётов

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)						
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература и эл. ресурсы	Практ. подг.
	Раздел 1. Принципы работы систем сбора больших данных					
1.1	Современные тенденции и направления в области данных и аналитики /Лек/	7	2	ПК-7.1	Л1.1 Л1.2 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
1.2	Инструменты и технологии обработки и анализа больших данных /Лек/	7	2	ПК-7.1	Л1.1 Л1.2 Л1.6 Л1.8Л2.2 Л2.3	
1.3	Основные принципы работы систем сбора больших данных /Лек/	7	2	ПК-7.1	Л1.1 Л1.2 Л1.6 Л1.8Л2.2 Л2.3	
1.4	Средства Data Mining /Пр/	7	8	ПК-7.1 ПК-7.2	Л1.1 Л1.2 Л1.4 Л1.6 Л1.8Л2.2 Л2.3	
1.5	Преимущества технологии Big Data и примеры применения /Лек/	7	2	ПК-7.1 ПК-7.2	Л1.1 Л1.2 Л1.6 Л1.8Л2.2 Л2.3	
1.6	Принципы работы систем сбора больших данных /Ср/	7	10	ПК-7.1 ПК-7.2	Л1.1 Л1.2 Л1.6 Л1.8Л2.2 Л2.3	
1.7	Методы извлечения из данных полезных инсайтов /Ср/	7	14	ПК-7.1	Л1.1 Л1.2 Л1.6 Л1.8Л2.2 Л2.3	
	Раздел 2. Основы анализа данных					
2.1	Технология Big Data: сбор, хранение, обработка данных /Лек/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.2 Л2.3	
2.2	Алгоритмы машинного обучения и статистического анализа /Лек/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.3	Основные концепции современной архитектуры данных /Лек/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.4	Статистический анализ в машинном обучении /Пр/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.5	BI-системы /Лек/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.6	Yandex DataLens облачный BI сервис /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.7	Подключение к БД с данными /Пр/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.8	Описание датасета и логики работы с ним /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.9	Создание чартов /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.7Л2.1 Л2.4	
2.10	Создание дашборда из чартов и селекторов /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.11	Визуализация динамики рекламных расходов в Power BI /Пр/	7	6	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.12	Excel расчет и визуализация данных /Пр/	7	6	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.13	Теория вероятности /Ср/	7	10	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.14	Линейные модели /Ср/	7	10	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.15	SQL — язык запросов /Ср/	7	5	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
2.16	Python — язык программирования /Ср/	7	5	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
	Раздел 3. Экосистема Hadoop					
3.1	Распределенные файловые системы /Лек/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
3.2	Введение в потоковую обработку данных - Spark /Лек/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4	
3.3	DataFrame/SQL, метаданные, форматы файлов, источники данных, RDD /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5 Л1.6 Л1.7 Л1.8	

3.4	Zeppelin, Spark UI: чтение и запись из БД (JDBC), CSV, Parquet, партиционирование, SQL-запросы с агрегацией и соединениями, планы выполнения запросов, мониторинг /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.5	Spark Streaming, Spark Structured Streaming, Flink. Чтение/обработка/запись потоков между Kafka, реляционной БД и файловой системой /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.6	Инструмент для обработки и анализа данных - Hive /Лек/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.7	Архитектура, метаданные таблиц, форматы файлов, язык запросов HiveQL /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.8	Hue, hive, beeline, Tez UI: создание таблиц, чтение и запись CSV, Parquet, ORC, партиционирование, SQL-запросы с агрегацией и соединениями /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.9	Система управления рабочими процессами - Oozie /Лек/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.10	Инструмент для быстрого доступа к большим объемам структурированных и полуструктурированных данных - HBase /Лек/	7	2	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.11	HBase shell: запись и чтение данных /Пр/	7	4	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
3.12	Распределенные файловые системы /Ср/	7	15	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.4Л2.1 Л2.2 Л2.3 Л2.4
3.13	Data Warehouse — хранилище данных /Ср/	7	15	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.4Л2.1 Л2.2 Л2.3 Л2.4
Раздел 4. Промежуточная аттестация					
4.1	Консультация /Консл/	7	1	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4
4.2	Контактная работа /КАЭ/	7	0,3	ПК-7.1 ПК-7.2 ПК-7.3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5 Л1.6 Л1.7 Л1.8Л2.1 Л2.2 Л2.3 Л2.4

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

5.1. Контрольные вопросы и задания

1. Современные тенденции и направления в области данных и аналитики
2. Инструменты и технологии обработки и анализа больших данных
3. Основные принципы работы систем сбора больших данных
4. Data Warehouse — хранилище данных
5. Средства Data Mining
6. SQL — язык запросов
7. Преимущества технологии Big Data и примеры применения
8. Принципы работы систем сбора больших данных
9. Методы извлечения из данных полезных инсайтов
10. Технология Big Data: сбор, хранение, обработка данных
11. Алгоритмы машинного обучения и статистического анализа
12. Основные концепции современной архитектуры данных
13. Статистический анализ в машинном обучении
14. BI-системы
15. Yandex DataLens облачный BI сервис
16. Yandex DataLens. Подключение к БД с данными
17. Yandex DataLens. Описание датасета и логики работы с ним
18. Yandex DataLens. Создание чартов
19. Yandex DataLens. Создание дашборда из чартов и селекторов
20. Визуализация динамики рекламных расходов в Power BI
21. Excel расчет и визуализация данных
22. Python — язык программирования

23.	Jupyter Notebook
24.	Библиотека Pandas
25.	Библиотека NumPy
26.	Распределенные файловые системы
27.	Введение в потоковую обработку данных - Spark
28.	DataFrame/SQL, метаданные, форматы файлов, источники данных, RDD
29.	Zerppelin, Spark UI: чтение и запись из БД (JDBC), CSV, Parquet, партиционирование
30.	SQL-запросы с агрегацией и соединениями, планы выполнения запросов, мониторинг
31.	Spark Streaming, Spark Structured Streaming, Flink. Чтение/обработка/запись потоков между Kafka, реляционной БД и файловой системой
32.	Инструмент для обработки и анализа данных - Hive
33.	Архитектура, метаданные таблиц, форматы файлов, язык запросов HiveQL
34.	Hue, hive, beeline, Tez UI: создание таблиц, чтение и запись CSV, Parquet, ORC, партиционирование, SQL-запросы с агрегацией и соединениями
35.	Система управления рабочими процессами - Oozie
36.	Инструмент для быстрого доступа к большим объемам структурированных и полуструктурированных данных – Hbase
37.	HBase shell: запись и чтение данных

5.2. Темы письменных работ

- 1) Понятие Больших данных.
- 2) Особенности сбора, хранения, обработки и анализа Больших данных
- 3) Требования к распределенным информационным системам
- 4) Средства построения распределенных информационных систем
- 5) Технология Map-Reduce
- 6) Система Apache Hadoop
- 7) Базы данных NoSQL. Особенности, классификация
- 8) Возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки информации. CAP-теорема.
- 9) Способы репликации и кластеризации баз данных
- 10) Документно-ориентированные базы данных
- 11) Возможности СУБД MongoDB
- 12) Работа с документно-ориентированными БД на языке JSON.

5.3. Фонд оценочных средств

- 1 Базисом в линейном пространстве размерности 4 могут быть наборы векторов:
- 1) $(1,1,0,0), (0,0,1,1), (1,1,-1,-1), (-1,-1,1,1)$
 - 2) $(12,0,0,0), (0,13,0,0), (0,0,14,0), (0,0,0,-100)$
 - 3) $(1,0,0), (0,1,0), (0,0,1)$
 - 4) $(1,0,0,0), (1,1,0,0), (1,1,1,0), (1,1,1)$
- 2 Выберите наборы линейно НЕ зависимых векторов (несколько вариантов ответа):
- 1) $(5,0,0), (0,100,0), (0,0,-2000)$
 - 2) $(1,2,3), (2,4,6), (15,0,0)$
 - 3) $(2,-5,7), (3,3,3), (5,-2,10)$
 - 4) $(1,0,0), (1,1,0), (1,1,1)$
- 3 Вычислите скалярное произведение двух векторов: $(1,-1,1,-1)$ и $(7,10,-15,14)$
- 1) - 32
 - 2) 30
 - 3) 28
 - 4) - 20
- 4 Найдите верное утверждение:
- 1) Пространство элементарных исходов содержит оцифрованные вероятности возникновения некоторого явления
 - 2) Вероятностная мера измеряется в процентах и принимает значения от 0 до 100
 - 3) Тройка, образующая вероятностное пространство -- множество элементарных исходов, сигма-алгебра над этим множеством и вероятностная мера
 - 4) Нет верных утверждений
- 5 Вычислите векторное выражение: $5 * (1,-1,3) - (12 * (0,0,1)) / (3 * (2,2,2))$
- 1) $(3,6,12)$
 - 2) $(6,2,1)$
 - 3) $(-17,8,4)$
 - 4) $(5,-5,13)$
- 6 Снаряд пытаются бросить на расстояние 25 метров. Но снаряд может с одинаковой вероятностью упасть на любом расстоянии от бросающего в промежутке от 1 до 25 метров. Какова вероятность, что снаряд упадет на расстоянии от 5 до 9 метров включительно или на расстоянии от 11 до 13 метров включительно?
- 1) 0,36
 - 2) 0,50
 - 3) 0,24
 - 4) 1

- 7 Какая область значений у функции $F(x) = \sqrt{-x-1}$
- 1) Вещественные числа R в промежутке $[0, \text{inf}]$
 - 2) Вещественные числа R
 - 3) Вещественные числа R в промежутке $[-\text{inf}, 0]$
 - 4) $[-\text{inf}, -1]$
 - 5) $[1, \text{inf}]$
- 8 Вычислите производную функции $F = 3x^3 - 6x^2 + 1$ в точке 2
- 1) 12
 - 2) 10
 - 3) 18
 - 4) 36
- 9 Можно утверждать, что точка x дифференцируемой функции F является точкой экстремума, если:
- 1) $F'(x) = 0$
 - 2) $F'(x) \neq 0$
 - 3) $F'(x) = 0$ и $F''(x) > 0$
 - 4) $F'(x) = 0$ и $F''(x) < 0$
- 10 В numpy скалярное произведение векторов можно посчитать с помощью (несколько вариантов ответа):
- 1) `np.product()`
 - 2) `np.multiply()`
 - 3) `np.dot()`
 - 4) `@`
 - 5) `&`
- 11 Какая область определения у функции $F(x) = \sqrt{-x-1}$:
- 1) Вещественные числа R
 - 2) $[-\text{inf}, -1]$
 - 3) Вещественные числа R в промежутке $[-\text{inf}, 0]$
 - 4) Вещественные числа R в промежутке $[0, \text{inf}]$
 - 5) $[1, \text{inf}]$
- 12 Какая формула шага градиентного спуска правильная?
- 1) $x = x + \epsilon \cdot \nabla_x$
 - 2) $x = x - \epsilon \cdot \nabla_x$
 - 3) $x = x + \epsilon \cdot |\nabla_x|$
 - 4) $x = x - \epsilon \cdot |\nabla_x|$
- 13 Петя подбрасывает симметричную монетку 7 раз. В предыдущие разы 1 раз выпал орел и 5 раз выпала решка. Какова вероятность, что на седьмой раз выпадет орел?
- 1) 0,50
 - 2) 1
 - 3) 0,25
 - 4) 0,75
- 14 Какому распределению соответствует функция `random.choice()` из пакета `random` языка Python?
- 1) Нормальному
 - 2) Равномерному
 - 3) Парето
- 15 Какое описание характеризует современное состояние технологий искусственного интеллекта?
- 1) Создание программ, повторяющих выполнение человеком интеллектуальных задач
 - 2) Создание узкоспециализированных интеллектуальных систем на основе обработки больших данных
 - 3) Существование машины, умеющей мыслить и решать разнообразные творческие задачи
- 16 Какую задачу решает GPT-3?
- 1) Игра в шахматы
 - 2) Постановка медицинского диагноза
 - 3) Генерация текстов
 - 4) Биологические задачи
- 17 Какое из перечисленных понятий НЕ входит в понятие искусственного интеллекта?
- 1) Машинное обучение
 - 2) Глубинное обучение
 - 3) Экспертные системы
 - 4) Аналитика данных
- 18 Выберите верное утверждение
- 1) Искусственный интеллект - это четко определенное понятие, означающее создание машины, повторяющей умственные процессы человека
 - 2) Искусственный интеллект - это сложное понятие, не имеющее четкого определения и включающее различные области математики, информационных технологий и др.
 - 3) Искусственный интеллект - это робот для общения с людьми посредством текстового интерфейса, разработанный Аланом Тьюрингом в 1950 году
- 19 В чем состоит тест Тьюринга?
- 1) Человеку показывают серию картин, и он должен выделить те, которые созданы машиной
 - 2) Человек получает ответы на вопросы от другого человека и от компьютера "вслепую" и должен определить, кто из собеседников -- компьютер

- 3) Машина должна “выжить” в сложной, искусственно заданной среде, с которой она взаимодействует посредством некоторого механизма
- 20 В 50-х годах 20-го века Фрэнк Розенблатт разработал перцептрон Розенблатта. Что это такое?
- 1) Модель восприятия информации мозгом человека
 - 2) Модель восприятия информации глазом человека
 - 3) Модель восприятия информации мозгом мыши
 - 4) Модель человеческой руки
- 21 Выберите верное утверждение: Современный искусственный интеллект основывается на...
- 1) изучении и компьютерном повторении структуры человеческого мозга
 - 2) создании искусственного мозга на основе биотехнологий
 - 3) обучении алгоритмов, способных решать задачи, аналогичные тем, что решает человек
- 22 Что такое DeepBlue?
- 1) Компьютер, применявшийся для диагностики заболеваний в 1970-х годах
 - 2) Компьютер, победивший чемпиона мира по шахматам в 1997 году
 - 3) Компьютер, победивший чемпиона мира по игре в го в 2015 году
 - 4) Компьютер, прошедший тест Тьюринга в 2014 году
- 23 Что такое GPT-3?
- 1) Нейронная сеть, распознающая объекты на изображениях точнее человека
 - 2) Нейронная сеть, ставящая медицинские диагнозы с 99% точностью
 - 3) Нейронная сеть, генерирующая тексты на естественном языке с уровнем качества, близким к человеческому
 - 4) Нейронная сеть, победившая чемпиона мира по игре в го
- 24 Какая метрика регрессии уделяет большое внимание выбросам?
- 1) MSE
 - 2) MAE
 - 3) Quantile loss
 - 4) MAPE
- 25 Какая метрика классификации не зависит от порога?
- 1) Precision
 - 2) Recall
 - 3) AUC-ROC
 - 4) Accuracy
- 26 Что из перечисленного НЕ является метрикой качества классификации?
- 1) точность
 - 2) средняя квадратичная ошибка
 - 3) площадь под ROC-кривой
 - 4) полнота
- 27 Выберите верное утверждение про метрики качества:
- 1) Главная цель решения задачи машинного обучения -- получить идеальное значение метрики качества (например, среднюю квадратичную ошибку 0) на тестовой выборке.
 - 2) Главная цель решения задачи машинного обучения -- получить идеальное значение метрики качества (например, среднюю квадратичную ошибку 0) на обучающей выборке.
 - 3) Главная цель решения задачи машинного обучения -- получить как можно более хорошее достижимое значение метрики качества на тестовой выборке.
 - 4) Главная цель решения задачи машинного обучения -- получить как можно более хорошее достижимое значение метрики качества на обучающей выборке.
- 28 В чем состоит задача классификации?
- 1) Предсказать признак для объекта по классам
 - 2) Предсказать класс для признака по объектам
 - 3) Предсказать признак для класса по объектам
 - 4) Предсказать класс для объекта по признакам
- 29 Вы - директор крупного супермаркета, желающий сократить количество краж покупателями с помощью установки системы видеослежения. Какой из перечисленных подходов основан на применении машинного обучения?
- 1) Нанять сотрудника, который будет следить за камерами
 - 2) Нанять сотрудника, который спросит у нескольких охранников, какими приемами они пользуются при определении краж, и запрограммирует эти приемы
 - 3) Собрать набор видеозаписей, на которых есть кражи и на которых их нет, и нанять сотрудника, который на основе этих данных построит алгоритм определения краж
 - 4) Купить робота, который будет ездить по залам супермаркета и пугать покупателей
- 30 Данные в машинном обучении должны быть представлены в стандартном виде - в виде таблицы. Что задают строки и столбцы этой таблицы?
- 1) Объекты и классы
 - 2) Объекты и признаки
 - 3) Признаки и классы
 - 4) Клиенты и их признаки
- 31 Рассмотрим пользователя социальной сети как объект в задаче машинного обучения. Что из перечисленного является задачей классификации?
- 1) Предсказание заработной платы пользователя
 - 2) Предсказание пола пользователя

3)	Предсказание профессии пользователя
4)	Предсказание, какой пост пользователь сделает следующим
32 Рассмотрим клиента банка как объект в задаче машинного обучения. Что из перечисленного является задачей бинарной классификации?	
1)	Предсказание, вернет ли клиент кредит
2)	Предсказание, когда клиент вернет кредит
3)	Предсказание, воспользуется ли клиент услугой досрочного погашения
4)	Предсказание, подключит ли клиент мобильный банк
33 Метод главных компонент...	
1)	Строит новые признаки как сложные нелинейные функции от исходных признаков
2)	Строит новые признаки как линейные функции от исходных признаков
3)	Отбирает самые важные признаки из исходных
4)	Предсказывает класс объекта на основе главных компонент
34 Выберите метод отбора признаков:	
1)	Метод главных компонент
2)	Фильтрация
3)	Автокодировщик
4)	MDS
35 Какую задачу решает метод t-SNE?	
1)	Моделирует объекты в двух- или трехмерном пространстве
2)	Отбирает признаки
3)	Группирует объекты в кластеры
4)	Выделяет сотни признаков

5.4. Перечень видов оценочных средств

Задания со свободно конструируемым ответом (СКО) предполагает составление развернутого ответа на теоретический вопрос. Задание с выбором одного варианта ответа (ОВ, в задании данного типа предлагается несколько вариантов ответа, среди которых один верный). Задания со свободно конструируемым ответом (СКО) предполагает составление развернутого ответа, включающего полное решение задачи с пояснениями.

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

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год
Л1.1	Кондрашов Ю. Н.	Язык SQL. Сборник ситуационных задач по дисциплине «Базы данных»: Учебно-практическое пособие	Москва: Русайнс, 2021, URL: https://book.ru/book/942020
Л1.2	Мартишин С.А., Симонов В.Л.	Базы данных. Практическое применение СУБД SQL- и NoSQL-типа для применения проектирования информационных систем: Учебное пособие	Москва: Издательский Дом "ФОРУМ", 2021, URL: http://znanium.com/catalog/document?id=374126
Л1.3	Козлов А.Ю., Мхитарян В. С.	Статистический анализ данных в MS Excel: Учебник	Москва: ООО "Научно-издательский центр ИНФРА-М", 2021, URL: http://znanium.com/catalog/document?id=378179
Л1.4	Мартишин С.А., Симонов В.Л., Храпченко М.В.	Базы данных. Практическое применение СУБД SQL- и NoSQL-типа для применения проектирования информационных систем: Учебное пособие	Москва: Издательский Дом "ФОРУМ", 2023, URL: https://znanium.com/catalog/document?id=419783
Л1.5	Козлов А.Ю., Мхитарян В. С., Шишов В.Ф.	Статистический анализ данных в MS Excel: Учебник	Москва: ООО "Научно-издательский центр ИНФРА-М", 2023, URL: https://znanium.com/catalog/document?id=423653
Л1.6	Чернышев С. А.	Алгоритмы и структуры данных на Python: Учебное пособие	Москва: КноРус, 2024, URL: https://book.ru/book/949701
Л1.7	Сидоркина И. Г.	Системы искусственного интеллекта: Учебное пособие	Москва: КноРус, 2022, URL: https://book.ru/book/944621
Л1.8	Криволапов С. Я.	Статистические вычисления на платформе Jupyter Notebook с использованием Python: Учебник	Москва: КноРус, 2022, URL: https://book.ru/book/943660

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год
Л2.1	Трофимов В.Б., Кулаков С.М.	Интеллектуальные автоматизированные системы управления технологическими объектами: Учебное пособие	Вологда: Инфра-Инженерия, 2020, URL: http://znanium.com/catalog/document?id=361646

	Авторы, составители	Заглавие	Издательство, год
Л2.2	Жуков Р.А.	Язык программирования Python. Практикум: Учебное пособие	Москва: ООО "Научно-издательский центр ИНФРА-М", 2021, URL: http://znanium.com/catalog/document?id=365208
Л2.3	Протождьяконов А.В., Пылов П.А., Садовников В.Е.	Алгоритмы Data Science и их практическая реализация на Python: Учебное пособие	Вологда: Инфра-Инженерия, 2022, URL: https://znanium.com/catalog/document?id=417222
Л2.4	Андрейчиков А.В., Андрейчикова О.Н.	Интеллектуальные информационные системы и методы искусственного интеллекта: Учебник	Москва: ООО "Научно-издательский центр ИНФРА-М", 2023, URL: https://znanium.com/catalog/document?id=417737
6.3.1. Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства			
6.3.1.1	Windows 10 Pro RUS Операционная система – Windows 10 Pro RUS Подписка Microsoft Imagine Premium – Order №143659 от 12.07.2021		
6.3.1.2	7-Zip Архиватор 7-Zip Программное обеспечение по лицензии GNU GPL		
6.3.1.3	Яндекс Браузер Браузер Яндекс Браузер Лицензионное соглашение на использование программ Яндекс Браузер https://yandex.ru/legal/browser_agreement/		
6.3.1.4	LibreOffice Офисный пакет LibreOffice Программное обеспечение по лицензии GNU GPL		
6.3.1.5	MS SQL Server 2019 СУБД Microsoft SQL Server 2019 Подписка Microsoft Imagine Premium – Order №143659 от 12.07.2021		
6.3.1.6	Visual Studio Code Редактор исходного кода, разработанный Microsoft для Windows, Linux и macOS Программное обеспечение по лицензии GNU GPL		
6.3.1.7	MySQL Server Community СУБД MySQL Программное обеспечение по лицензии GNU GPL		
6.3.2. Перечень профессиональных баз данных и информационных справочных систем			
6.3.2.1	Консультант Плюс http://www.consultant.ru		
6.3.2.2	ИСО Международная организация по стандартизации https://www.iso.org/ru/home.html		
6.3.2.3	Кодекс – Профессиональные справочные системы https://kodeks.ru		

7. МТО (оборудование и технические средства обучения)			
Ауд	Наименование	ПО	Оснащение
208	Лаборатория Интеллектуальные системы и технологии (Research Laboratory of Intelligent Systems and Technologies)	Windows 10 Pro RUS 7-Zip Яндекс Браузер Mozilla Firefox LibreOffice LibreCAD Inkscape Notepad++. Kaspersky Endpoint Security MS Access 2016 MS Project Pro 2016 MS SQL Server 2019 MS SQL Server Management Studio 18.8 MS Visio Pro 2016 MS Visual Studio Community Edition Visual Studio Code Blender Gimp Maxima Oracle VM VirtualBox StarUML V1 PostgreSQL IntelliJ IDEA PyCharm Community Edition Eclips Adobe Reader DC Diptrace Embarcadero RAD Studio XE8	Стол - 10 шт., стул - 21 шт., рабочее место преподавателя - 1 шт., персональных компьютеров с выходом в интернет - 20 шт., доска учебная – 1 шт., многофункциональное устройство – 1 шт., мультимедийный проектор – 1 шт., проекционный экран – 1 шт., междисциплинарная лабораторная станция NI ELVIS II и ПО Circuit Design Suit - 1 шт, соответствующее программное обеспечение

		Arduino Software (IDE) NetBeans IDE ZEAL ARIS Express Archimate Klite Mega Codec Pack Ramus Educational Micro-Cap Evaluation gvSIG Desktop Python	
108	Лаборатория искусственного интеллекта	Яндекс Браузер LibreOffice Astra Linux PostgreSQL	Стол - 20 шт., кресло офисное - 21 шт., рабочее место преподавателя - 1 шт., доска учебная – 1 шт., персональный компьютер с выходом в интернет – 21 шт., телевизор - 1 шт., многофункциональное устройство – 1 шт., соответствующее программное обеспечение
Читальный зал	Информационно-библиотечный центр (помещение для самостоятельной работы обучающихся)	7-Zip Яндекс Браузер Mozilla Firefox LibreOffice LibreCAD Inkscape Notepad++ Kaspersky Endpoint Security MS Access 2016 MS Project Pro 2016 MS Visio Pro 2016 Visual Studio Code Blender Gimp Maxima IntelliJ IDEA PyCharm Community Edition Adobe Reader DC MS Office Standart 2007 Windows 10 Pro	Стол - 20 шт., стул - 20 шт., рабочее место сотрудника - 2 шт., персональный компьютер с выходом в интернет и обеспечением доступа в электронную информационно-образовательную среду академии – 17 шт., многофункциональное устройство – 2 шт.

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

В соответствии с требованиями ФГОС ВО по направлению подготовки реализация компетентностного подхода предусматривает использование в учебном процессе активных и интерактивных форм проведения занятий (разбор конкретных задач, проведение блиц-опросов, исследовательские работы) в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся.

Лекционные занятия дополняются ПЗ и различными формами СРС с учебной и научной литературой. В процессе такой работы студенты приобретают навыки «глубокого чтения» - анализа и интерпретации текстов по методологии и методике дисциплины.

Учебный материал по дисциплине «Системы сбора и анализа больших данных». разделен на логически завершённые части (модули), после изучения, которых предусматривается аттестация в форме письменных тестов, контрольных работ. Работы оцениваются в баллах, сумма которых дает рейтинг каждого обучающегося. В баллах оцениваются не только знания и навыки обучающихся, но и их творческие возможности: активность, неординарность решений поставленных проблем. Каждый модуль учебной дисциплины включает обязательные виды работ – лекции, ПЗ, различные виды СРС (выполнение домашних заданий по решению задач, подготовка к лекциям и практическим занятиям).

Форма текущего контроля знаний – работа студента на практическом занятии, опрос. Форма промежуточных аттестаций – контрольная работа в аудитории, домашняя работа. Итоговая форма контроля знаний по модулям – контрольная работа с задачами по материалу модуля.

Методические указания по выполнению всех видов учебной работы размещены в электронной образовательной среде академии.

Методические указания и материалы по видам учебных занятий по дисциплине:

Вид учебных занятий, работ - Организация деятельности обучающегося

Лекция - Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения, отмечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, попытаться найти ответ в рекомендуемой литературе, если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.

Практические занятия - Конспектирование источников. Работа с конспектом лекций, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы, работа с текстом. Выполнение практических задач в инструментальных средах. Выполнение проектов. Решение расчётно-графических заданий, решение задач по алгоритму и др.

Самостоятельная работа - Знакомство с основной и дополнительной литературой, включая справочные издания,

зарубежные источники, конспект основных положений, терминов, сведений, требующихся для запоминания и являющихся основополагающими в этой теме. Составление аннотаций к прочитанным литературным источникам и др.

9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ОБУЧАЮЩИМСЯ ПО ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Самостоятельная работа студентов в ходе семестра является важной составной частью учебного процесса и необходима для закрепления и углубления знаний, полученных в период сессии на лекциях, практических и интерактивных занятиях, а также для индивидуального изучения дисциплины «Системы сбора и анализа больших данных» в соответствии с программой и рекомендованной литературой.

Самостоятельная работа выполняется в виде подготовки домашнего задания или сообщения по отдельным вопросам.

Контроль качества выполнения самостоятельной (домашней) работы может осуществляться с помощью устного опроса на лекциях или практических занятиях, обсуждения подготовленных научно-исследовательских проектов, проведения тестирования.

Устные формы контроля помогут оценить владение студентами жанрами научной речи (дискуссия, диспут, сообщение, доклад и др.), в которых раскрывается умение студентов передать нужную информацию, грамотно использовать языковые средства, а также ораторские приемы для контакта с аудиторией.

Письменные работы позволяют оценить владение источниками, научным стилем изложения, для которого характерны: логичность, точность терминологии, обобщенность и отвлеченность, насыщенность фактической информацией.