

Документ подписан Министерством науки и высшего образования Российской Федерации
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 02.04.2024 14:22:08
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

УТВЕРЖДАЮ
Директор Института магистратуры
Иванова Е.А.
«01» июня 2023г.

**Рабочая программа дисциплины
Системы аналитики больших данных**

Направление 09.04.03 Прикладная информатика
магистерская программа 09.04.03.01 "Информационные системы и технологии в
бизнесе"

Для набора 2023 года

Квалификация
магистр

КАФЕДРА Информационных систем и прикладной информатики**Распределение часов дисциплины по семестрам**

Семестр (<Курс>.<Семестр на курсе>)	4 (2.2)		Итого	
	13 4/6			
Неделя	УП	РП	УП	РП
Лекции	18	18	18	18
Практические	36	36	36	36
Итого ауд.	54	54	54	54
Контактная работа	54	54	54	54
Сам. работа	117	117	117	117
Часы на контроль	9	9	9	9
Итого	180	180	180	180

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 28.03.2023 протокол № 9.

Программу составил(и): д.э.н., доцент, Щербаков С.М.

Зав. кафедрой: д.э.н., доцент Щербаков С.М.

Методическим советом направления: д.э.н., доцент, Щербаков С.М.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	формирование у обучающихся способности осуществлять руководство проектами по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных с применением новых методов и алгоритмов машинного обучения.
-----	---

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-2:Способен руководить проектами по созданию систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения со стороны заказчика

ПК-4:Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика

В результате освоения дисциплины обучающийся должен:

Знать:
новые методы и алгоритмы машинного обучения (соотнесено с индикатором ПК-2.1) основы аналитики больших данных (соотнесено с индикатором ПК-4.1)
Уметь:
руководить разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика (соотнесено с индикатором ПК-2.2) создавать комплексные системы на основе аналитики больших данных (соотнесено с индикатором ПК-4.2)
Владеть:
навыками руководства созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения (соотнесено с индикатором ПК-2.3) навыками руководства проектами по построению комплексных систем на основе аналитики больших данных в различных сферах (соотнесено с индикатором ПК-4.3)

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Кварт	Часов	Компетенции	Литература
	Раздел 1. Основы построения и работы с системами аналитики больших данных				
1.1	Тема 1. Разработка систем аналитики больших данных. /Лек/	4	2	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.2	Тема 2. Хранение больших данных в облаке. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище. Реляционные базы данных. Нереляционные базы данных. Сервисы нереляционных баз данных. Корпоративные хранилища данных (DWH). /Лек/	4	4	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.3	Тема 2. Хранение больших данных в облаке. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище. Реляционные базы данных. Нереляционные базы данных. Сервисы нереляционных баз данных. Корпоративные хранилища данных (DWH). Выполнение заданий с использованием Yandex Data Lens, ClickHouse. /Пр/	4	12	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.4	Тема 1. Разработка систем аналитики больших данных. /Ср/	4	26	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.5	Тема 2. Хранение больших данных в облаке. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище. Реляционные базы данных. Нереляционные базы данных. Сервисы нереляционных баз данных. Корпоративные хранилища данных (DWH). /Ср/	4	28	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
	Раздел 2. Разработка систем аналитики больших данных				

2.1	Тема 3. Доставка больших данных в облако. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения. Доставка данных в реляционные БД и хранилища. Доставка данных в нереляционные базы данных. Доставка данных в HDFS-совместимые хранилища. Прямая загрузка потоковых данных. Облачные сервисы развертывания кластерных систем. Облачные сервисы копирования и трансформации данных. /Лек/	4	6	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.2	Тема 3. Доставка больших данных в облако. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения. Доставка данных в реляционные БД и хранилища. Доставка данных в нереляционные базы данных. Доставка данных в HDFS-совместимые хранилища. Прямая загрузка потоковых данных. Облачные сервисы развертывания кластерных систем. Облачные сервисы копирования и трансформации данных. Выполнение заданий с использованием Yandex Data Lens, ClickHouse. /Пр/	4	12	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.3	Тема 4. Аналитика больших данных в облаке. Интерактивный анализ данных. Анализ реляционных данных. Встроенные редакторы запросов сервиса. Потоковый анализ данных. Общие сведения. Пакетный анализ данных. /Лек/	4	6	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.4	Тема 4. Аналитика больших данных в облаке. Интерактивный анализ данных. Анализ реляционных данных. Встроенные редакторы запросов сервиса. Потоковый анализ данных. Общие сведения. Пакетный анализ данных. Выполнение заданий с использованием Yandex Data Lens, ClickHouse. /Пр/	4	12	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.5	Тема 3. Доставка больших данных в облако. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения. Доставка данных в реляционные БД и хранилища. Доставка данных в нереляционные базы данных. Доставка данных в HDFS-совместимые хранилища. Прямая загрузка потоковых данных. Облачные сервисы развертывания кластерных систем. Облачные сервисы копирования и трансформации данных. /Ср/	4	28	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.6	Тема 4. Аналитика больших данных в облаке. Интерактивный анализ данных. Анализ реляционных данных. Встроенные редакторы запросов сервиса. Потоковый анализ данных. Общие сведения. Пакетный анализ данных. /Ср/	4	35	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.7	/Экзамен/	4	9	ПК-2 ПК-4	Л1.1 Л1.2Л2.1 Л2.2 Л2.3

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1	Адлер, Ю. П., Черных, Е. А.	Статистическое управление процессами. «Большие данные»: учебное пособие	Москва: Издательский Дом МИСиС, 2016	http://www.iprbookshop.ru/64199.html неограниченный доступ для зарегистрированных пользователей

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.2	Зыков Роман	Роман с Data Science. Как монетизировать большие данные. — (Серия «IT для бизнеса»)	Санкт-Петербург: Питер, 2021	https://ibooks.ru/reading.php?short=1&productid=3770 33 неограниченный доступ для зарегистрированных пользователей

5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1	Емельянов А. А.	Прикладная информатика: журнал	Москва: Синергия ПРЕСС, 2010	https://biblioclub.ru/index.php?page=book&id=120321 неограниченный доступ для зарегистрированных пользователей
Л2.2	Пол Дейтел, Харви Дейтел	Python: Искусственный интеллект, большие данные и облачные вычисления	Санкт-Петербург: Питер, 2021	https://ibooks.ru/reading.php?short=1&productid=371701 неограниченный доступ для зарегистрированных пользователей
Л2.3	Кадырова, Н. О., Павлова, Л. В.	Статистический анализ больших данных: подход на основе машин опорных векторов: учебное пособие	Санкт-Петербург: Санкт-Петербургский политехнический университет Петра Великого, 2022	https://www.iprbookshop.ru/128651.html неограниченный доступ для зарегистрированных пользователей

5.3 Профессиональные базы данных и информационные справочные системы

Национальная электронная библиотека (НЭБ) - <https://rusneb.ru/>

ИСС «КонсультантПлюс»

ИСС «Гарант» <http://www.internet.garant.ru/>

5.4. Перечень программного обеспечения

Yandex Data Lens

ClickHouse

5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор, экран / интерактивная доска.

Практические занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет.

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

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1. Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-2. Способен руководить проектами по созданию систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения со стороны заказчика			
З. новые методы и алгоритмы машинного обучения	знает функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), Э – вопросы к экзамену (1-8)
У. руководить разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика	применяет современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (1-2)
В. навыками руководства созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения	руководит выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (1-2)
ПК-4. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика			
З. основы аналитики больших данных	знает методологию и принципы руководства проектами по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), Э – вопросы к экзамену (9-17)
У. создавать комплексные системы на основе аналитики больших данных	решает задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (3)
В. навыками руководства проектами по построению комплексных систем на основе аналитики больших данных в различных сферах	выявляет области деловой деятельности, которые потенциально могут получить отдачу от аналитики	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (3)

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

84-100 баллов (оценка «отлично»),

67-83 баллов (оценка «хорошо»),

50-66 баллов (оценка «удовлетворительно»),

0-49 баллов (оценка «неудовлетворительно»).

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

Вопросы к экзамену

1. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище.
2. Реляционные базы данных.
3. Нереляционные базы данных. Сервисы нереляционных баз данных.
4. Корпоративные хранилища данных (DWH).
5. Хранилища данных типа Data Lake.
6. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения.
7. Прямая загрузка данных. Доставка данных в реляционные БД и хранилища.
8. Прямая загрузка данных. Доставка данных в нереляционные базы данных.
9. Прямая загрузка данных. Доставка данных в HDFS-совместимые хранилища.
10. Прямая загрузка потоковых данных.
11. Облачные сервисы развертывания кластерных систем.
12. Облачные сервисы копирования и трансформации данных.
13. Интерактивный анализ данных.
14. Анализ реляционных данных.
15. Встроенные редакторы запросов сервиса.
16. Потоковый анализ данных. Общие сведения.
17. Пакетный анализ данных.

Экзаменационное задание включает два вопроса – два теоретических вопроса и одно практико-ориентированное задание из числа приведенных ниже практических заданий.

Критерии оценивания:

- 84-100 баллов (оценка «отлично») – изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;

- 67-83 баллов (оценка «хорошо») – наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, обучающийся усвоил основную литературу, рекомендованную в рабочей программе дисциплины;

- 50-66 баллов (оценка «удовлетворительно») – наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике;

- 0-49 баллов (оценка «неудовлетворительно») – ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Тестовые задания

1. Большинство данных в мире в 2011 году содержалось:
 - а) в цифровом виде
 - б) в аналоговом виде
2. Объём накопленных человечеством цифровых данных на 2012 год измеряется:
 - а) петабайтами
 - б) зеттабайтами
 - в) эксабайтами
 - г) йоттабайтами
3. Укажите фактор, способствовавший появлению тренда больших данных
 - а) маркетинговые кампании крупных корпораций
 - б) снижение издержек на хранение данных

- в) появление новых технологий обработки потоковых данных
г) выпуск баз данных с обработкой данных в памяти
4. Какие вероятные разочарования тренда больших данных?
а) из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных
б) из-за угрозы безопасности личной жизни (privacy) граждан будут упрощены процедуры сбора данных, что приведёт к падению ценности больших данных
в) нет
5. Отметьте значимые события, повлиявшие на формирование тренда больших данных:
а) разработка Hadoop
б) изобретение принципа MapReduce
в) разработка языка Python
г) победа Deepblue в матче с Г.Каспаровым
6. Выберите верный ответ
а) большие данные – это обработка или хранение более 1 Тб информации
б) проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна
в) большие данные – это огромная PR-акция крупных вендоров и не более того
г) большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект
7. Выберите неверный ответ:
А) большие данные – это данные объёма свыше 1 Тб
Б) проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна
В) большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров
Г) большие данные как правило не структурированы
9. Отметьте те из вариантов, в которых данные структурированы:
а) данные о продажах компании, представленные в виде помесечных отчётов в текстовом формате
б) таблица с ежедневными показаниями температуры помещения за год в файле формата csv
в) текст педагогической поэмы А.С. Макаренки, представленный в формате PDF
г) библиотека фильмов, представленных в формате mpeg4 на одном жестком диске
10. Перечислите четыре основных характеристики Big Data:
а) Virtualization, Volume, Variability, Velocity
б) Variety, Velocity, Volume, Value
в) Verification, Volume, Velocity, Visualization
г) Video, Value, Variety, Volume
11. Выберите неверное высказывание:
а) большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных
б) увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации
в) удешевление систем хранения на единицу информации привело к росту рынка больших данных
12. Отметьте неверное понимание Variety в контексте характеристик Big Data:
а) высокая скорость генерирования данных
б) разные типы данных в колонках таблиц реляционных СУБД
в) разнообразие отраслей, являющихся источниками данных
г) разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные

Критерии оценивания:

Из имеющегося банка тестов формируется тестовое задание, содержащее 10 тестовых вопросов для одного обучающегося. Каждый тестовый вопрос содержит 4 варианта ответов, один или несколько из которых – верные.

Правильный ответ на один тестовый вопрос – 1 балл, неправильный – 0 баллов.

Максимальное количество баллов за тестовые задания – 10.

Практические задания

Практическое задание №1

Развертывание облачного хранилища общего назначения

Облачное хранилище Непосредственное хранение информации. Доступ к панели мониторинга и управления. Работа с хранилищем на веб-портале. Создание бакетов и настройка права доступа к ним. Сравнение возможностей и ограничений облачных хранилищ общего назначения для решения прикладных задач.

Практическое задание №2

Развертывание облачных сервисов копирования и трансформации данных

Сервисы трансформации и копирования играют важную роль в построении информационных систем анализа больших данных. Требования к сервисам хранения и анализа противоречивы и не могут быть удовлетворены в рамках одного хранилища, особенно когда данных много. Чтобы удовлетворить эти требования, необходимо использовать хранилища разных типов и сервис, который связывает их информационно. Средство каталогизации, трансформации и копирования данных.

Практическое задание №3

Развертывание облачных сервисов для потокового анализа данных

Потоковый анализ состоит в анализе потока в виде последовательности сообщений, то есть в применении алгоритма анализа к каждому сообщению потока. Этот анализ может состоять в выделении из всего потока сообщений, соответствующих определенным признакам, например, сообщений об ошибках, которые должны быть направлены в отдельное хранилище или сервис, занимающийся обработкой ошибок. Пример прикладной задачи – анализ банковских транзакций и определение в этом потоке потенциально подозрительных (выдача большой суммы в нетипичной для плательщика стране или выдача мелких сумм в разных банкоматах в течение малого интервала времени).

Критерии оценивания (для каждого задания):

27-30 б. – задание выполнено верно;

19-26 б. – при выполнении задания были допущены неточности, не влияющие на результат;

9-18 б. – при выполнении задания были допущены ошибки;

1-8 б. – при выполнении задания были допущены существенные ошибки.

Максимальное количество баллов за все практические задания – 90 (3 задания по 30 баллов).

3. Методические материалы, определяющие процедуры оценивания результатов освоения образовательной программы

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме экзамена.

Экзамен проводится по расписанию промежуточной аттестации. Количество вопросов в экзаменационном задании – 3 (два теоретических вопроса и одно практико-ориентированное задание). Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в экзаменационную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- практические занятия.

В ходе лекционных занятий рассматриваются основные теоретические вопросы, даются рекомендации для самостоятельной работы и подготовке к практическим занятиям.

В ходе практических занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются навыки практической работы.

При подготовке к практическим занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;
- изучить конспекты лекций;
- подготовить ответы на все вопросы по изучаемой теме.

В процессе подготовки к практическим занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях и практических занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий методом выполнения тестовых и практических заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме, дополнить конспекты лекций недостающим материалом, выписками из рекомендованных первоисточников, выделить непонятные термины, найти их значение в энциклопедических словарях.

Студент должен готовиться к предстоящему практическому занятию по всем обозначенным в рабочей программе дисциплины вопросам.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.