

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Поверинов Игорь Егорович

Должность: Проректор по учебной работе

Дата подписания: 17.10.2023 17:19:41

Уникальный программный ключ:

6d465b936eef331cede482bde6d12ab98216652f016465d53b72a2eab0de1b2

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«Чувашский государственный университет имени И. Н. Ульянова»
(ФГБОУ ВО «ЧГУ им. И.Н. Ульянова»)**

Факультет экономический

Кафедра бухгалтерского учета и электронного бизнеса

Утверждены
в составе образовательной программы
высшего образования

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
(ФОНД ОЦЕНОЧНЫХ СРЕДСТВ)**

«Анализ больших данных»

Направление подготовки / специальность 38.04.05 Бизнес-информатика

Квалификация выпускника магистр

Направленность (профиль) / специализация «Бизнес-информатика в цифровой экономике»

Год начала подготовки - 2023

Чебоксары - 2023

Составитель(и):

Доцент, кандидат экономических наук Л. А. Савинова

Согласовано

методической комиссией факультета экономический 17.04.2023, протокол № 3

Декан факультета Н. В. Морозова

1. Паспорт оценочных материалов по дисциплине (модулю) «Анализ больших данных»

1.1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОП ВО

<i>Код и наименование компетенции</i>	<i>Код и наименование индикатора достижения компетенции</i>	<i>Дескрипторы индикатора достижения компетенции (результаты обучения)</i>
ПК-5 Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации	ПК-5.1 Анализирует потребности заинтересованных лиц и/или подразделений организации в исследовании больших данных и разрабатывает и согласует техническое задание на создание методической и технологической инфраструктуры больших данных	Знать: методы выявления потребностей заинтересованных лиц организации к методической и технологической инфраструктуре больших данных Уметь: анализировать потребности заинтересованных лиц и или подразделений организации в исследовании больших данных Владеть: навыками разработки и согласования технического задания на создание методической и технологической инфраструктуры больших данных
ПК-5 Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации	ПК-5.2 Управляет получением, хранением, передачей, обработкой больших данных	Знать: методы и инструменты получения, извлечения, преобразования, загрузки и обработки больших данных Уметь: пользоваться методами и инструментами получения, хранения, передачи, обработки больших данных Владеть: навыками управления, получения, хранения, передачи и обработки больших данных

ПК-5 Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации	ПК-5.3 Стратегически управляет развитием методологической и технологической инфраструктуры анализа больших данных в организации	Знать: основные методы стратегического управления развитием методологической и технологической инфраструктуры анализа больших данных на цифровом предприятии Уметь: эффективно применять основные методы стратегического управления развитием методологической и технологической инфраструктуры анализа больших данных на цифровом предприятии Владеть: навыками и методиками стратегического управления развитием методологической и технологической инфраструктуры анализа больших данных на цифровом предприятии
---	---	--

1.2. Структура дисциплины (модуля)

<i>№ п/п</i>	<i>Контролируемые разделы (темы) дисциплины</i>	<i>Код контролируемых индикаторов достижения компетенций</i>	<i>Наименование оценочного средства</i>
1	Большие данные: современные подходы к обработке и хранению	ПК-5.1, ПК-5.2, ПК- 5.3	Контрольные вопросы
Тема 1. Основы больших данных			
Тема 2. Обработка больших данных			
2	Программное обеспечение в области анализа больших данных	ПК-5.1, ПК-5.2, ПК- 5.3	Тематика докладов, Типовые задания
Тема 1. Анализ и извлечение данных			
Тема 2. Технология Data Mining			

3	Индивидуальная контактная работа	ПК-5.1, ПК-5.2, ПК-5.3	Вопросы к экзамену
Тема 1. Индивидуальная контактная работа			

2. Объем дисциплины (модуля) и виды учебной работы

Общая трудоемкость дисциплины (ЗЕ): 4

Форма промежуточной аттестации: (курсовой проект, экзамен)

3. Критерии оценки успеваемости обучающихся

Формы и виды контроля знаний обучающихся, предусмотренные по данной дисциплине:

- текущий контроль;
- промежуточная аттестация.

Текущий контроль успеваемости проводится с целью проверки знаний обучающихся, приобретения и развития навыков самостоятельной работы, усиления связи между преподавателем и обучающимся, совершенствования работы кафедр по развитию навыков самостоятельной работы, по повышению академической активности обучающихся.

Промежуточная аттестация, как форма контроля успеваемости по дисциплинам (разделам дисциплин) и видам учебной деятельности, проводится для проверки степени усвоения обучающимися программного учебного материала и установления соответствия результатов проверки требованиям государственных образовательных стандартов к обязательному минимуму содержания или формирования компетенций, установленных федеральными государственными образовательными стандартами высшего образования.

В зависимости от видов контроля знаний обучающихся, предусмотренных учебным планом, для оценки успеваемости применяются следующие критерии.

Критерии оценивания на зачете:

- «зачтено» ставится, если обучающийся продемонстрировал наличие глубоких и исчерпывающих знаний в объеме пройденного программного материала, грамотное и логически стройное изложение материала при ответе, знание дополнительно рекомендованной литературы;
- «не зачтено» ставится, если обучающийся продемонстрировал наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Критерии оценивания на экзамене:

- для оценки «отлично» - наличие глубоких и исчерпывающих знаний в объеме пройденного программного материала, правильные и уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, знание дополнительно рекомендованной литературы;
- для оценки «хорошо» - наличие твердых и достаточно полных знаний программного материала, незначительные ошибки при освещении заданных вопросов, правильные действия по применению знаний на практике, четкое изложение материала;
- для оценки «удовлетворительно» - наличие твердых знаний пройденного материала, изложение ответов с ошибками, уверенно исправляемыми после дополнительных вопросов, необходимость наводящих вопросов, правильные действия по применению знаний на практике;

- для оценки «неудовлетворительно» - наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Критерии оценивания курсовой работы (проекта), расчетно-графической работы:

Оценка по курсовой работе (проекту), расчетно-графической работе выставляется на основании результатов защиты обучающимся своих работ при непосредственном участии преподавателей кафедры, руководителя курсовой работы (проекта), с возможным присутствием других обучающихся из учебной группы.

«Отлично» - работа выполнена в соответствии с утвержденным планом, полностью раскрыто содержание каждого вопроса, студентом сформулированы собственные аргументированные выводы по теме работы. Оформление работы соответствует предъявляемым требованиям. При защите работы обучающийся свободно владел материалом и отвечал на вопросы.

«Хорошо» - работа выполнена в соответствии с утвержденным планом, полностью раскрыто содержание каждого вопроса. Незначительные замечания к оформлению работы. При защите работы обучающийся владел материалом, но отвечал не на все вопросы.

«Удовлетворительно» - работа выполнена в соответствии с утвержденным планом, но не полностью раскрыто содержание каждого вопроса. Обучающимся не сделаны собственные выводы по теме работы. Грубые недостатки в оформлении работы. При защите работы обучающийся владел материалом, отвечал не на все вопросы.

«Неудовлетворительно» - если работа не выполнена в соответствии с утвержденным планом, не раскрыто содержание каждого вопроса, обучающимся не сделаны выводы по теме работы, имеются грубые недостатки в оформлении работы, при защите работы обучающийся не владел материалом, не отвечал на вопросы, то работа направляется на дальнейшую доработку.

4. Оценочные материалы текущего контроля успеваемости

1 Контрольные вопросы

1. Какие существуют подходы к построению и загрузке озер и хранилищ данных?
2. Приведите классификацию бизнес-задач в соответствии с большим типом данных.
3. Каковы основные принципы построения систем хранения и обработки больших данных?
4. Каковы принципы проектирования архитектуры систем больших данных?
5. В чем заключаются особенности сбора, хранения, обработки и анализа больших данных?
6. Каковы требования к распределенным информационным системам?
7. С помощью каких средств возможно построить распределенную информационную систему?
8. Для чего предназначена технология Map-Reduce?
9. Для чего предназначена система Apache Hadoop?
10. В чем заключается особенность баз данных NoSQL?
11. Каковы возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки информации?
12. Назовите способы репликации и кластеризации баз данных.
13. Назовите возможности СУБД MongoDB.
14. В чем принципиальное отличие концепции Big Data от традиционного подхода BI?

15. Сформулируйте понятие явной (выраженной) и скрытой (структурной) информации.

16. Дайте определение контент-анализа.

17. Каковы основные понятия контент-анализа?

18. Какие существуют виды контент-анализа?

19. Каковы этапы контент-анализа?

20. Каковы основные признаки, характеризующие «Большие данные»?

Критерии оценивания:

– оценка «отлично» выставляется, если материал изложен грамотно, в определенной логической последовательности; продемонстрировано системное и глубокое знание программного материала, продемонстрированы знания, умения и навыки, достаточные для дальнейшего усвоения материала;

– оценка «хорошо» выставляется, если материал изложен грамотно, в определенной логической последовательности; продемонстрировано достаточно полное знание программного материала, или, по крайней мере, показано общее понимание вопроса и продемонстрированы знания, умения и навыки, достаточные для дальнейшего усвоения материала;

– оценка «удовлетворительно» выставляется, если показано общее понимание вопроса и продемонстрированы знания, умения и навыки, достаточные для дальнейшего усвоения материала;

– оценка «неудовлетворительно» выставляется, если обнаружено незнание или непонимание большей, или наиболее важной части учебного материала; показано непонимание вопроса и не продемонстрированы знания, умения и навыки, достаточные для дальнейшего усвоения материала.

2 Типовые задания

Задание 1

1. Найти потоковые данные, формирующие базу больших данных.

2. Провести процедуру структуризации и записи данных.

3. Написать функцию обращения к данным.

4. Провести визуализацию и первичный статистический анализ больших данных.

Задание 2

Исследование современных проблем и перспектив решения обработки больших данных, получаемых или сохраняемых в сети Интернет (веб-данных), а так же возможность практической реализации технологии Data Mining для больших веб-данных на практическом примере.

1. Выполнить обзор библиографических источников по проблемам анализа больших данных.

2. Применить технологию Data Mining для анализа больших веб-данных.

3. Выполнить компьютерное моделирование практической задачи с помощью языка программирования C#.

4. Создать структуру базы данных на языке описания данных DDL для накопления веб- данных.

Задание 3

Подобрать данные для таблицы [Годы; Численность населения; Рост ВВП; Динамика безработицы; Динамика промышленного производства; Прирост людей, получивших очное высшее образование; Развитие науки (высокотехнологичных отраслей); Динамика доходов на душу населения; Динамика преступности] и проанализировать их взаимное влияние, отобразить корреляцию:

1. Роста ВВП на прирост населения
2. Прироста населения на динамику безработицы
3. Прирост людей с высшим образованием на рост промышленного производства
4. Прирост людей с высшим образованием на развитие науки
5. Прирост людей с высшим образованием на динамику доходов на душу населения
6. Динамику безработицы на динамику преступности
7. С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.
8. С помощью функции `predict()` постройте прогноз по столбцу, соответствующему варианту.

Критерии оценивания:

– оценка «отлично» выставляется обучающемуся, если он имеет знания в пределах области задания с пониманием границ применимости, имеет диапазон знаний, требуемых для развития творческих решений;

– оценка «хорошо» выставляется обучающемуся, если он имеет достаточно полные знания в пределах области задания с пониманием границ применимости, имеет достаточный диапазон знаний, требуемых для развития творческих решений, имеются незначительные ошибки при освещении заданных вопросов;

– оценка «удовлетворительно» выставляется обучающемуся, если он имеет определенные знания в пределах области задания с пониманием границ применимости, имеет ограниченный диапазон знаний, требуемых для развития творческих решений, имеются незначительные ошибки при освещении заданных вопросов;

– оценка «неудовлетворительно» выставляется обучающемуся, если он не владеет знаниями в пределах области задания, не имеет достаточный диапазон знаний, требуемых для развития творческих решений.

3 Тематика докладов

1. Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных.

2. Роль аналитика по данным (Data Scientist). Ключевые компетенции аналитика. Отличия BI от Data Science.

3. Корреляция и регрессионный анализ. Коэффициент корреляции. Графическое представление. Постановка задачи регрессионного анализа. Линейная регрессия. Метод наименьших квадратов. Привести примеры использования регрессионного анализа.

4. Классификация. Признаковое описание объекта и таблица объект-свойства. Постановка задачи. Отличия задачи классификации от задачи регрессии. Определение модели и алгоритма. Процесс обучения. Проблема переобучения. Регуляризация. Cross validation. Привести примеры использования алгоритмов классификации. Дополнительный вопрос: привести модель в линейной регрессии.

5. Кластеризация. Метрики. Матрица парных расстояний. Постановка задачи кластеризации. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.

6. Ассоциативные правила. Определение. Достоверность и поддержка. Отличия построения ассоциативного правила от решающего правила задачи классификации. Привести примеры использования ассоциативных правил.

7. Парадигма Map Reduce. Описать принцип работы. Нарисовать диаграмму. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.

8. Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры использования визуализации.

9. «Жизненный цикл» проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы над проектом. Перечислить основные роли исполнителей проекта.

10. Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии.

Критерии оценивания:

– оценка «отлично» выставляется, если работа раскрывает содержание вопроса, обнаружено знание положений темы, знание и понимание ответа на вопросы;

– оценка «хорошо» выставляется, если работа раскрывает основное содержание вопроса, обнаружено знание основных положений темы, знание и понимание большей, или наиболее важной части ответа на вопросы;

– оценка «удовлетворительно» выставляется, если работа в основном раскрывает содержание вопроса, демонстрируется знание и понимание большей части ответа на вопросы, обнаружено неполное знание основных положений темы.

– оценка «неудовлетворительно» выставляется, если не раскрыто основное содержание материала, обнаружено незнание основных положений темы, обнаружено незнание и непонимание большей, или наиболее важной части ответа на вопросы.

5. Оценочные материалы промежуточной аттестации обучающихся

Перечень вопросов к зачету

Не предусмотрено.

Перечень вопросов к экзамену

1. Понятие и назначение Big Data.
2. Технологии обработки больших объемов данных.
3. Особенности высокоскоростной обработки данных.
4. Типы слабоструктурированных данных и особенности их обработки.
5. Основные принципы работы Hadoop.
6. Средства сбора и доработки данных.
7. Визуализация больших данных.
8. Задачи администрирования в системах обработки больших данных.
9. Принципы построения систем хранения и обработки больших данных.
10. Методы обеспечения качества данных.
11. Применение Big Data в различных сферах деятельности.
12. Применение Big Data в решении бизнес-задач.
13. Источники больших данных.
14. Платформы больших данных.
15. Процесс изучения данных.
16. Экосистемы больших данных.
17. Программы с открытым исходным кодом.

18. Облачные решения BigData.
19. Применение Spark.
20. Интеграция данных с использованием технологии BigData.
21. Построение и загрузка озер и хранилищ данных.
22. Классификация бизнес-задач в соответствии с большим типом данных.
23. Использование большого типа данных для классификации больших характеристик данных.
24. Основные принципы построения систем хранения и обработки больших данных.
25. Принципы проектирования архитектуры систем больших данных.
26. Особенности сбора, хранения, обработки и анализа больших данных.
27. Требования к распределенным информационным системам. Средства построения распределенных информационных систем.
28. Технология Map-Reduce.
29. Система Apache Hadoop.
30. Базы данных NoSQL. Особенности, классификация.
31. Возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки информации.
32. Способы репликации и кластеризации баз данных.
33. Документо-ориентированные базы данных. Работа с документо-ориентированными БД на языке JSON.
34. Возможности СУБД MongoDB.
35. В чем принципиальное отличие концепции Big Data от традиционного подхода BI?
36. Понятие явной (выраженной) и скрытой (структурной) информации.
37. Определение контент-анализа.
38. Каковы основные понятия контент-анализа?
39. Какие существуют виды контент-анализа?
40. Каковы этапы контент-анализа?
41. Каковы основные признаки, характеризующие «Большие данные»?
42. Модели развертывания облачных хранилищ.
43. Модели обслуживания облачных хранилищ.
44. Постановка и описание проблемы «последней мили».
45. Безопасность, производительность и надежность при работе с облачными данными.
46. Экономическая составляющая облачных подходов.
47. Способы машинного обучения.
48. Основные фазы обработки «больших данных».

Примерная тематика курсовых работ

Не предусмотрено.

Примерная тематика курсовых проектов

1. Распознавание жестов сурдоязыка австралийских аборигенов.
2. Распознавание автора по произведениям живописи.
3. Предсказание исхода шахматной партии.
4. Сортировка ядер грецких орехов на видеоизображении.
5. Распознавание рукописных цифр с помощью наивного байесовского классификатора.
6. Распознавание спамовых писем с помощью деревьев решений.

7. Понижение признакового пространства методом главных компонент при диагностировании клеток опухоли по изображениям мазка крови.
8. Предсказание октанового числа бензина по инфракрасному спектру с помощью регрессии на главные компоненты и проекции на латентные структуры.
9. Сегментация базы данных клиентов методами кластеризации и предсказание реакции клиента.
10. Распознавание лиц методом собственных лиц.
11. Активное обучение на основе разноточных гауссовских процессов.
12. Идентификация и верификация речевых сигналов на основе нейронных сетей.
13. Прогнозирование редких событий на основе рекуррентных нейронных сетей.
14. Глубинные гауссовские процессы.
15. Восстановление многообразий, описывающих область значений многомерной зависимости.
16. Прогнозирование временных рядов на основе восстановления многообразий.
17. Обучение в распределенной среде.
18. Алгоритмы предсказания с использованием экспертных стратегий.
19. Группировка рёбер в динамической визуализации графов.
20. Метрики схожести текстов, основанные на колмогоровской сложности.

Примерная тематика расчетно-графических работ

Не предусмотрено.