



Федеральное агентство морского и речного транспорта
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Государственный университет морского и речного флота
имени адмирала С.О. Макарова»**
Воронежский филиал ФГБОУ ВО «ГУМРФ имени адмирала С.О. Макарова»

Кафедра математики, информационных систем и технологий

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине *«Анализ больших данных»*
(приложение к рабочей программе дисциплины)

Направление подготовки 09.03.02 Информационные системы и технологии

Направленность (профиль) Информационные системы на транспорте

Уровень высшего образования бакалавриат

Форма обучения заочная

г. Воронеж
2023

1. Перечень компетенций и этапы их формирования в процессе освоения дисциплины

В результате освоения ОПОП бакалавриата обучающийся должен овладеть следующими результатами обучения по дисциплине:

Таблица 1

Планируемые результаты обучения по дисциплине

Код и наименование компетенции	Код индикатора достижения компетенции	Планируемые результаты обучения по дисциплине
ПК-3: Способность проводить научные исследования при разработке, внедрении и сопровождении информационных систем и технологий с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	ПК-3.1: анализ исходных данных, оценку качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	Знать: методы анализа исходных данных, оценку качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Уметь: анализировать информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Владеть: навыками оптимизации информационной системы в области анализа больших данных
	ПК-3.2 Оптимизация информационных систем и технологий для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	Знать: Инструменты и методы оценки качества и эффективности ИС Уметь: оптимизировать информационные системы и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Владеть: навыками осуществления оптимизации ИС для достижения новых целевых показателей в области анализа больших данных
ОПК-8: Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем.	ОПК-8.1. Математическое моделирование сложных систем, анализ данных	Знать: требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных Уметь: использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов Владеть: способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям

2. Паспорт фонда оценочных средств для проведения текущей и промежуточной аттестации обучающихся

Таблица 2

Оценочные средства для проведения текущей и промежуточной аттестации обучающихся

№ п/п	Наименование раздела (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Тема 1. Постановка проблемы анализа данных.	ОПК-8	Тестирование, экзамен
2	Тема 2. Вероятностное моделирование условий неопределенности.	ПК-3 ОПК-8	Тестирование, экзамен
3	Тема 3. Оптимальный линейный средне-квадратический прогноз.	ПК-3 ОПК-8	Тестирование, экзамен
4	Тема 4. Целенаправленное проектирование и редукция размерности.	ПК-3 ОПК-8	Тестирование, экзамен
5	Тема 5. Классификация многомерных измерений. Кластерный анализ.	ПК-3 ОПК-8	Тестирование, экзамен

Таблица 3

Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине

Результат обучения по дисциплине	Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине				Процедура оценивания
	2	3	4	5	
ПК-3.1 Знать: проводить анализ исходных данных, оценку качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	<i>Отсутствие или фрагментарные представления об анализе исходных данных, оценке качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Неполные представления об анализе исходных данных, оценке качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Сформированные, но содержащие пробелы представления об анализе исходных данных, оценке качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Сформированные систематические представления об анализе исходных данных, оценке качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	Тестирование, экзамен
ПК-3.1 Уметь: анализировать	<i>Отсутствие умений или фрагментарные</i>	<i>В целом удовлетворительные, но не</i>	<i>В целом удовлетворительные, но</i>	<i>Сформированные умения проводить</i>	Тестирование, экзамен

Результат обучения по дисциплине	Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине				Процедура оценивания
	2	3	4	5	
информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	<i>умения анализировать информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>систематизированные умения анализировать информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>содержащие отдельные пробелы умения анализировать информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>анализировать информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	
ПК-3.1 Владеть: навыками оптимизации информационной системы	<i>Отсутствие владения или фрагментарное владение навыками оптимизации информационной системы</i>	<i>В целом удовлетворительные, но не систематизированные навыки оптимизации информационной системы</i>	<i>В целом удовлетворительные, но содержащие отдельные пробелы навыки оптимизации информационной системы</i>	<i>Сформированное владение навыками оптимизации информационной системы</i>	<i>Тестирование, экзамен</i>
ПК-3.2 Знать: Инструменты и методы эффективного использования и сопровождения ИС	<i>Отсутствие или фрагментарные представления об инструментах и методах эффективного использования и сопровождения ИС</i>	<i>Неполные представления об инструментах и методах эффективного использования и сопровождения ИС</i>	<i>Сформированные, но содержащие отдельные пробелы представления об инструментах и методах эффективного использования и сопровождения ИС</i>	<i>Сформированные систематические представления об инструментах и методах эффективного использования и сопровождения ИС</i>	<i>Тестирование, экзамен</i>
ПК-3.2 Уметь: оптимизировать информационные системы и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	<i>Отсутствие или фрагментарные представления об оптимизации информационных систем и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Неполные представления об оптимизации информационных систем и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Сформированные, но содержащие отдельные пробелы представления об оптимизации информационных систем и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Сформированные систематические представления об оптимизации информационных систем и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности</i>	<i>Тестирование, экзамен</i>
ПК-3.2 Владеть:	<i>Отсутствие или фрагментарные</i>	<i>Неполные представления о</i>	<i>Сформированные, но содержащие</i>	<i>Сформированные</i>	<i>Тестирование,</i>

Результат обучения по дисциплине	Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине				Процедура оценивания
	2	3	4	5	
навыками осуществления оптимизации ИС для достижения новых целевых показателей в области анализа больших данных	<i>представления о достижении новых целевых показателей в области анализа больших данных</i>	<i>достижении новых целевых показателей в области анализа больших данных</i>	<i>отдельные пробелы представления о достижении новых целевых показателей в области анализа больших данных</i>	<i>систематические представления о достижении новых целевых показателей в области анализа больших данных</i>	экзамен
ОПК-8.1 Знать: требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных	<i>Отсутствие или фрагментарные представления о требованиях, связанных с созданием и использованием SQL и NoSQL систем хранения и обработки данных</i>	<i>Неполные представления о требованиях, связанных с созданием и использованием SQL и NoSQL систем хранения и обработки данных</i>	<i>Сформированные, но содержащие отдельные пробелы представления о требованиях, связанных с созданием и использованием SQL и NoSQL систем хранения и обработки данных</i>	<i>Сформированные систематические представления о требованиях, связанных с созданием и использованием SQL и NoSQL систем хранения и обработки данных</i>	Тестирование, экзамен
ОПК-8.1 Уметь: использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов	<i>Отсутствие умений или фрагментарные умения использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов</i>	<i>В целом удовлетворительные, но не систематизированные умения использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов</i>	<i>В целом удовлетворительные, но содержащие отдельные пробелы умения использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов.</i>	<i>Сформированные умения использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов</i>	Тестирование, экзамен
ОПК-8.1 Владеть: способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по	<i>Отсутствие владения или фрагментарные способности собирать, обрабатывать и интерпретировать данные современных научных исследований,</i>	<i>В целом удовлетворительные, но не систематизированные владения способностью собирать, обрабатывать и интерпретировать данные современных научных исследований</i>	<i>В целом удовлетворительные, но содержащие отдельные пробелы владения способностью собирать, обрабатывать и интерпретировать данные</i>	<i>Сформированные владения способностью собирать, обрабатывать и интерпретировать данные современных научных исследований,</i>	Тестирование, экзамен

Результат обучения по дисциплине	Критерии оценивания результата обучения по дисциплине и шкала оценивания по дисциплине				Процедура оценивания
	2	3	4	5	
соответствующим научным исследованиям	<i>необходимые для формирования выводов по соответствующим или научным исследованиям</i>	<i>исследований, необходимые для формирования выводов по соответствующим научным исследованиям</i>	<i>современных научных исследований, необходимые для формирования выводов по соответствующим или научным исследованиям</i>	<i>для формирования выводов по соответствующим научным исследованиям</i>	

2. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ

Тестовые задания для проведения текущего контроля

Тестовые задания по теме 1

- Суммарная оперативная память IBM Watson составляет порядка:
 - 100 гигабайт
 - 5000 терабайт
 - 10 зетабайт
 - 15 терабайт
- Кто ввел термин Большие данные?
 - Клиффорд Линч
 - Алан Тьюринг
 - Бьерн Страуструп
 - Дональд Кнут
- Какие данные занимают больше мировой памяти относительно остальных?
 - Structured Data
 - Unstructured Data
 - Semi-Structured Data
 - Quasi-Structured Data
- BigData – это ...
 - Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации, или обработки.
 - Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.
 - Колоссальный объем данных, собранных человечеством.
 - Класс в Java, предназначенный для хранения данных от 100 Гб

5. Какая компания создала технологию MapReduce?

- А) Google
- Б) Yahoo
- В) EMC
- Г) Oracle

6. Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:

- А) Структурированными
- Б) Полуструктурированными
- В) Квазиструктурированными
- Г) Неструктурированными

7. Что означает термин «Big Data» в информационных технологиях?

- А) Комплексный набор методов для создания файлов большого объема
- Б) Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.
- В) Файлы с большим количеством данных.
- Г) Представление времени, дня, месяца и года в качестве значения количества миллисекунд, прошедших с начала нашей эры.

8. Данные имеющие определенный тип, формат и структуру (например, транзакционные данные) являются:

- А) Структурированными
- Б) Полуструктурированными
- В) Квазиструктурированными
- Г) Неструктурированными

9. Чему примерно равен объем всей существующей на земле информации (в байтах)?

- А) 10^{11}
- Б) 10^{21}
- В) $10^{1010101}$ Г) 10^{171}

10. В каком году впервые был введен термин Большие данные?

- А) 2002
- Б) 2004
- В) 2006
- Г) 2008

Тестовые задания по теме 2

11. Что является средством анализа в VI?

- А) Карты показателей;

- Б) Совместная работа и управление рабочими процессами;
- В) Информационные панели;
- Г) BI инфраструктура.

12. Основное умение исследователя данных?

- А) Умение находить наиболее важные элементы в хранимой информации
- Б) Уметь прогнозировать исход работы системы
- В) Находить скрытые логические связи в системе собранной информации
- Г) Отличать неструктурированные данные от структурированных

13. Какой язык программирования из перечисленных является наиболее важным для аналитика?

- А) C++
- Б) PHP
- В) F# Г) R

14. Что означает термин «Business Intelligence» в информационных технологиях?

- А) Комплексный набор методов для создания бизнес планов.
- Б) Методы и инструменты для перевода необработанной информации в осмысленную, удобную для восприятия форму.
- В) Файлы, содержащие информацию о бизнес плане.
- Г) Технологии, направленные на развитие бизнеса.

15. Языком, на котором был разработан RabbitMQ, является:

- А) Java
- Б) Python В) C++
- Г) Erlang

16. Что является главным результатом процесса Business Intelligence?

- А) Возможность принятия решений для бизнеса
- Б) Результаты интеллектуального анализа данных
- В) Возможность использования искусственного интеллекта
- Г) Получение структуризации данных после выполнения всех шагов процесса

17. Что из перечисленного не является средством анализа?

- А) Продвинутая визуализация
- Б) Reporting
- В) Predictive Modelling
- Г) Data Mining

18. Что относится к средствам предоставления информации в «Business Intelligence»?

- А) Генератор нерегламентированных запросов

- Б) Совместная работа и управление рабочими процессами
- В) Предиктивное моделирование и Data Mining
- Г) Карты показателей

19. Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является:

- А) OLAP
- Б) Data Mining
- В) Predictive Modelling
- Г) Data Science

20. Что не является целью процесса Business Intelligence?

- А) Интерпретация большого количества данных;
- Б) Моделирование исходов различных вариантов действий;
- В) Модификация существующего программного обеспечения;
- Г) Отслеживание результатов решений.

Тестовые задания по теме 3

21. Что из этого не является реализацией Hadoop?

- А) Google MapReduce
- Б) Phoenix
- В) GreenMint
- Г) Qizmt

22. Какие из перечисленных пунктов являются достоинствами MapReduce?

- А) Оптимальная производительность
- Б) Эффективное применение в маленьких кластерах с небольшим объемом данных
- В) Масштабируемость
- Г) Отказоустойчивость

23. Что такое Oozie?

- А) Распределенный координационный сервис
- Б) Нереляционная распределенная база данных
- В) Язык управления потоком данных и исполнительная среда для анализа больших объемов данных
- Г) Сервис для записи и планировки заданий Hadoop

24. Сколько уровней имеет лямбда-архитектура?

- А) 2
- Б) 3
- В) 4
- Г) 5

25. Какие компоненты являются частями MapReduce?

- А) Task Tracker
- Б) Name Node и Data Node
- В) Job Tracker и Task Tracker
- Г) Job Tracker, Task Tracker, Name Node и Data Node

26. Что такое Spark?

- А) Инструмент для кластерных вычислений
- Б) Графический движок
- В) Библиотека для работы с графами
- Г) Технология распределенных вычислений

27. Дайте определение Map Reduce...

- А) Модель распределенных вычислений, предназначенная для параллельных вычислений над очень большими (до нескольких петабайт) объемами данных
- Б) Набор компонентов и интерфейсов для распределенных файловых систем и общего ввода-вывода
- В) Распределенная файловая система, работающая на больших кластерах типовых машин
- Г) Распределенный сервис для коллекционирования, сбора, и перемещения больших массивов данных

28. Что из этого является недостатком MapReduce?

- А) Фиксированный алгоритм обработки данных
- Б) Масштабируемость
- В) Отказоустойчивость
- Г) Возможность автоматического распараллеливания

29. Какое API было добавлено в Hadoop v2.0?

- А) YAWN
- Б) YARN
- В) SARN
- Г) DARN

30. Какая цель у NameNode в HDFS?

- А) Хранить индекс того, какая часть данных находится в каком узле
- Б) Хранить имя файла, хранящегося в конкретном узле
- В) Хранить индекс узла, в котором хранится имя файла
- Г) Хранить имена узлов

Тестовые задания по теме 4

31. Вертикальное масштабирование...

- А) Требуется изменений в прикладных программах, работающих на таких системах
- Б) Не требуется никаких изменений в прикладных программах, работающих на таких системах
- В) Уменьшает производительность каждого компонента БД
- Г) Увеличивает скорость загрузки данных

32. Для достижения какого свойства в БД типа NoSQL нет JOIN операций?

- А) Intercepting
- Б) Concurrency
- В) Consistency
- Г) Capacity

33. Что, согласно теореме CAP (теореме Брюера), возможно обеспечить в любой реализации распределённых вычислений?

- А) Только согласованность данных
- Б) Только доступность данных
- В) Согласованность данных, доступность данных, устойчивость к разделению
- Г) Не более двух свойств из трёх вышеприведённых

34. Выберите верное определение понятия AP-система:

- А) Система, во всех узлах которой данные согласованы и обеспечена доступность, жертвует устойчивостью к распаду на секции
- Б) Распределённая система, в каждый момент обеспечивающая целостный результат и способная функционировать в условиях распада
- В) Распределённая система, отказывающаяся от целостности результата
- Г) Система, автоматически изменяющая данные алгоритма своего с целью сохранения оптимального состояния

35. Что означает термин NoSQL?

- А) Не SQL
- Б) Не только SQL
- В) Без SQL
- Г) SQL – плохо

36. Разбиение системы на более мелкие структурные компоненты и разнесение их по отдельным физическим машинам (или их группам), и (или) увеличение количества серверов, параллельно выполняющих одну и ту же функцию, это:

- А) Горизонтальное масштабирование
- Б) Вертикальное масштабирование
- В) Master- slave репликация
- Г) Peer-to-peer репликация

37. Что из перечисленного относится к графо-ориентированным хранилищам (Graph Store)?

- А) Neo4j
- Б) BaseX
- В) Elasticsearch
- Г) Ничего

38. Что поддерживает NoSQL?

- А) Операцию Insert
- Б) Полностью стандарт SQL
- В) Операцию Join
- Г) Операцию Group by

39. Какие три свойства фигурируют в определении теоремы CAP?

- А) Согласованность данных
- Б) Сложность
- В) Доступность
- Г) Устойчивость к разделению

40. Выделение таблицы или группы таблиц на отдельный сервер это...

- А) Горизонтальное масштабирование
- Б) Вертикальное масштабирование
- В) Горизонтальный шардинг
- Г) Вертикальный шардинг

Тестовые задания по теме 5

41. Какая из БД на 100% совместима с интерфейсом языка R?

- А) MySQL R
- Б) Oracle R
- В) PostgreSQL R
- Г) NoSQL R

42. Что из этого не является типом визуализации?

- А) График
- Б) Текст
- В) Круговая диаграмма
- Г) Гистограмма

43. Отображение зависимости значений одной величины от другой - это...

- А) Матрица
- Б) График
- В) Диаграмма
- Г) Карта

44. Функция округления до единиц вверх в языке «R»:

- A) Ceiling(x)
- Б) Floor(x)
- В) Trunc(x)
- Г) Round(x,2)

45. Что такое сингулярность?

- A) Точка, в которой функция равна нулю
- Б) Точка, в которой первая производная равна нулю
- В) Точка, в которой вторая производная равна нулю
- Г) Точка, в которой математическая функция стремится к бесконечности или имеет какие-либо иные нерегулярности поведения

46. Какой тип лицензии у языка R?

- A) Adware
- Б) Commercial CC
- В) Open source
- Г) Shareware

47. Какие достоинства у Amazon S3?

- A) Будет работать всегда
- Б) Нужно самостоятельно решать сложные задачи распределения файлов между серверами
- В) Внезапные всплески популярности не приведут к отказу железа
- Г) Все вышеперечисленное

48. Что из перечисленного помогает следить за эволюцией документа, над созданием

которого работает одновременно большое количество авторов?

- A) Пространственный поток
- Б) Исторический поток
- В) Визуальный поток
- Г) Интерактивный поток

49. Преподнесение какой-либо полезной информации в форме интересного рассказа – это...

- A) сторителлинг
- Б) инфографика
- В) бизнес аналитика
- Г) картограмма

50. Что хорошо подходит для дедупликации?

- A) Картинки, видео, музыка
- Б) Виртуальные машины
- В) Сжатые данные

Г) Резервные копии

51. Что является результатом решения задачи регрессии?

- А) множество допустимых ответов конечно и их называют метками классов
- Б) допустимым ответом является действительное число или числовой вектор
- В) множество допустимых ответов бесконечно
- Г) алгоритм, принимающий на входе описание объекта

52. Основная цель статистического анализа:

- А) Поиск генеральной совокупности
- Б) Выяснение свойств генеральной совокупности
- В) Сравнение генеральных совокупностей
- Г) Выявление последовательности входного набора

53. Определённое предположение о распределении вероятностей, лежащем в основе

наблюдаемой выборки данных, - это...

- А) Статистический критерий
- Б) Статистическая выборка
- В) Статистическая гипотеза
- Г) Задача кластеризации

54. К каким алгоритмам классификации относится метод ближайших соседей?

- А) Метрическим
- Б) Логическим
- В) Линейным
- Г) Нет верного ответа

55. Преимуществом метода ближайшего соседа является:

- А) Устойчивость к погрешностям
- Б) Наличие настраиваемых параметров
- В) Высокое качество классификации
- Г) Простота реализации

56. С помощью какого алгоритма можно найти ассоциативное правило?

- А) Алгоритм apriori
- Б) Алгоритм k-means
- В) Алгоритм c-means
- Г) Иерархический алгоритм

57. Технология машинного обучения, когда нет ответов и требуется искать зависимости

между объектами, называется ...

- А) Самостоятельное обучение
- Б) Обучение без учителя

- В) Обучение с учителем
- Г) Обучение по зависимостям

58. Критерий Пирсона является:

- А) Критерием значимости
- Б) Параметрическим критерием
- В) Критерием согласия
- Г) Непараметрическим критерием

59. Чем отличаются ошибки первого и второго рода при принятии решений?

- А) Ошибка первого рода значительнее, нежели второго
- Б) Ошибка второго рода не обнаруживает различия, которые есть, а первого обнаруживает, которых нет
- В) Ошибка второго рода значительнее, нежели первого
- Г) Ошибка первого рода не обнаруживает различия, которых нет, а второго обнаруживает

60. Графическая характеристика качества бинарного классификатора ROC-кривая показывает зависимость...

- А) Величины TPR (доля верных положительных классификаций) от величины FPR (доля ложных положительных классификаций)
- Б) Величины FPR (доля ложных положительных классификаций) от величины TPR (доля верных положительных классификаций)
- Г) Величины TNR (доля верных отрицательных классификаций) от величины FPR (доля ложных положительных классификаций)

Критерии оценки результатов тестирования

Оценка результатов тестирования. За каждый правильный ответ начисляется 1 балл. Для перевода баллов в оценку применяется универсальная шкала оценки образовательных достижений. Если обучающийся набирает

- от 90 до 100% от максимально возможной суммы баллов - выставляется оценка «отлично»;
- от 80 до 89% - оценка «хорошо»;
- от 51 до 79% - оценка «удовлетворительно»;
- менее 51% - оценка «неудовлетворительно».

3. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОГО КОНТРОЛЯ

Вопросы для подготовки к экзамену

1. Что такое большие данные. Характеристики больших данных (3v, 4v)
2. В чем отличие систем хранения больших данных (масштабируемость, отказоустойчивость)

3. Что такое Hadoop. Основные компоненты Hadoop (HDFS, MapReduce)
4. Файловая система распределенного хранения больших данных HDFS
5. Модель распределенных вычислений MapReduce
6. Узлы Hadoop кластера (DataNode, NameNode, TaskTracker, JobTracker)
7. Отказоустойчивость, что если упадет DataNode, NameNode, TaskTracker, JobTracker
8. Основные проблемы Hadoop 1. Основные отличия Hadoop 2 от Hadoop 1
9. Hadoop 2. Изменения в HDFS
10. Hadoop 2. Архитектура Yarn
11. Языки высокого уровня над Hadoop (Pig, Hive)
12. Что такое Spark, основная идея
13. Зачем нужны NoSQL базы данных
14. Типы NoSQL баз данных
15. Модель данных HBase
16. Основные этапы извлечения информации из текстов
17. Подходы к распознаванию сущностей из текстов
18. Язык AQL. Компоненты языка AQL

Критерии оценки ответов на экзамене

Таблица 4

Критерии оценки

Наименование показателя	Критерии оценки	Максимальное количество баллов	Количество баллов
I. КАЧЕСТВО ОТВЕТА			
1 Соответствие ответов, поставленным вопросам	- систематизированные, глубокие и полные знания по всем разделам учебной программы - полное и глубокое усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины - умение ориентироваться в теориях, концепциях и направлениях по изучаемой дисциплине	10	
2. Грамотность изложения	- владение терминологией и понятийным аппаратом проблемы; - научный стиль изложения.	5	
3. Самостоятельность выполнения работы, глубина проработки материала, использование рекомендованной и справочной литературы	- степень знакомства автора работы с актуальным состоянием изучаемой проблематики; - дополнительные знания, использованные при написании работы, которые получены помимо предложенной образовательной программы;	5	

Общая оценка за выполнение	20	
ОТВЕТЫ НА ДОПОЛНИТЕЛЬНЫЕ ВОПРОСЫ ПО СОДЕРЖАНИЮ РАБОТЫ		
Вопрос 1	5	
Вопрос 2	5	
Общая оценка за ответы на вопросы	10	
Итого	30	

Для перевода баллов критериально-шкалированной таблицы в оценку применяется универсальная шкала оценки образовательных достижений. Если студент набирает 27-30 баллов и выше - оценка «отлично», 26 -21 баллов и выше - оценка «хорошо», 18-21 баллов и выше - оценка «удовлетворительно», менее 18 - оценка «неудовлетворительно».

Примеры тем для курсовой работы

1. Перегрузка информацией и Data Mining
2. Модели информационно-развивающихся систем
3. Виды знаний и способы их представления
4. Классы систем Data Mining
5. Методы матричного анализа
6. Метод наименьших квадратов в линейной модели измерений.
7. Множественный регрессионный анализ.

Критерии оценивания:

- полнота и правильность ответа;
- степень осознанности, понимания изученного;
- языковое оформление ответа.

Таблица 5

Показатели и шкала оценивания выполнения курсовой работы

Шкала оценивания	Показатели
5	работа выполнена без ошибок, обучающийся представил оригинальное и грамотное решение, четко и грамотно оформляет пояснительную записку без отступлений от требований к её оформлению, подробно и безошибочно отвечает на все заданные ему вопросы, проявляет при работе достаточную самостоятельность
4	работа выполнена с незначительными ошибками, но при опросе обучающийся проявляет понимание ошибок и способов их исправления, не допускает существенных погрешностей в ответах на вопросы, аккуратно выполняет демонстрационный материал и пояснительную записку
3	работа выполнена без грубых ошибок, но при опросе обучающийся проявляет недостаточное понимание всех подробностей проделанной работы; допускает при ответах на вопросы неточности и неправильные формулировки; допускает небрежность в графической работе и в оформлении пояснительной записки.
2	принципиальные ошибки в представленной к защите работе и обучающийся при ответах на вопросы, не может устранить указанные недостатки, небрежно выполняет работу и представляет неполную и не

	соответствующую правилам оформления пояснительную записку, проявляет полное пренебрежение к срокам выполнения проекта.
--	--

Составитель: д.ф.-м.н., профессор Кузьменко Р. В.

Зав. кафедрой: к.ф.-м.н., доцент Черняева С. Н.