

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Петровская Анна Викторовна
Должность: Директор
Дата подписания: 15.10.2024 11:34:46
Уникальный программный ключ:
798bda6555fbdebe827768f6f1710bd17a9070c31fdc1b6a6ac5a1f10c8c5199

Приложение 6
к основной профессиональной образовательной программе
по направлению подготовки 38.03.07 Товароведение
направленность (профиль) программы Товарная экспертиза,
оценочная деятельность и управление качеством

**Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский экономический университет имени Г.В. Плеханова»**

Факультет экономики, менеджмента и торговли

Кафедра бухгалтерского учета и анализа

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

по дисциплине «Основы работы с Большими данными»

Направление подготовки 38.03.07 Товароведение

направление (профиль) программы

Товарная экспертиза, оценочная деятельность и управление качеством

Уровень высшего образования Бакалавриат

Год начала подготовки 2022

Составитель:

к.п.н, доцент кафедры бухгалтерского
учета и анализа

В.В. Салий

Оценочные материалы одобрены на заседании кафедры бухгалтерского учета и
анализа протокол № 1 от 30.08.2021 г.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

По дисциплине «Основы работы с Большими данными»

ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ И ЭТАПОВ ИХ ФОРМИРОВАНИЯ ПО ДИСЦИПЛИНЕ

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций (код и наименование индикатора)	Результаты обучения (знания, умения)
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи	УК-1.1. З-1. Знает основные методы критического анализа и основы системного подхода как общенаучного метода.
		УК-1.1. У-1. Умеет анализировать задачу, используя основы критического анализа и системного подхода.
		УК-1.1. У-2. Умеет осуществлять поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации.

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ХАРАКТЕРИЗУЮЩИЕ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ

Перечень учебных заданий на аудиторных занятиях

Темы групповых дискуссий

Тема 1. Основы Больших данных

Индикаторы достижения: УК-1.1.

1. Что такое большие данные?
2. Методики анализа больших данных
3. Аналитический инструментарий
4. Рынки решений для управления большими данными
5. Теория и практика больших данных в отраслях

6. Как использовать аналитические данные для разработки качественных ИТ-сервисов

7. Большие данные (big data) в информационных технологиях.

8. Совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence.

9. В данную серию включают средства массово-параллельной обработки неопределённо структурированных данных, прежде всего, решениями категории NoSQL, алгоритмами MapReduce, программными каркасами и библиотеками проекта Hadoop.

10. В качестве определяющих характеристик для больших данных отмечают три V: объём (англ. volume, в смысле величины физического объёма), скорость (англ. velocity в смыслах как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (англ. variety, в смысле возможности одновременной обработки различных типов структурированных и полуструктурированных данных).

11. Информационно-аналитических систем.

12. Структура средств сбора и доработки данных.

13. Способы и состав средств преобразования данных.

14. Среды разработки средств сбора, доработки данных и информационных хранилищ.

15. Основные принципы построения структуры информационных хранилищ.

Тема 2. Обработка больших данных.

Индикаторы достижения: УК-1.1.

1. BigData и аналитика в реальном времени

2. BigData и ценовая политика в обработке данных клиентов

3. Обзор технологий хранения больших данных.

4. Базы данных. Системы управления базами данных. Модели данных.

5. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.

6. Основные понятия математической статистики.

7. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).

8. Принципы гибкой архитектуры данных и открытых систем, которыми руководствуются при создании ИАС.

9. Информационный обмен, связанный с аналитической работой.
10. Понятие о метаданных (МД), базе метаданных - репозитории, используемых в информационном хранилище.
11. Принципы построения информационных хранилищ.
12. Требования к качеству данных и способы его обеспечения при загрузке в информационное хранилище.
13. Проблемы, разрешаемые при приведении данных к единой структуре информационного хранилища.
14. Концепции построения структур хранилищ данных.
15. Назначение, состав и выполняемые функции базы метаданных - репозитория ИХ.
16. Принципы создания репозитория ИХ.
17. Элементы моделей данных ИХ (факт-таблица, таблицы измерений, консольные таблицы).

Тема 3. Анализ и извлечение данных

Индикаторы достижения: УК-1.1.

1. Цели факторного анализа.
2. Программные средства анализа данных: Statistica, SPSS, Excel и другие; их преимущества и недостатки.
3. Представление исходных данных в программе R-Studio.
4. SQL over BigData. Hive
5. Проблема смещения данных в обработке больших данных
6. Применение SQL в IT индустрии
7. Сравнение решений Hive и MapReduce на примере задач анализа логов
8. Beyond MapReduce. Spark
9. Машинное обучение на больших данных
10. Алгоритмы для работы с большими данными
11. Методы онлайн обучения
12. Градиентный спуск
13. Решение задач кластеризации на больших данных
14. Задача подсчета слов в датасете (WordCount)
15. Задачи и содержание оперативного (OLAP) анализа.
16. Классификация ИТ-анализа по режиму и темпу.
17. Функции, выполняемые процедурами сечение и поворот.
18. Функции, выполняемые процедурами свертка и развертка.
19. Функции, выполняемые процедурами проекция и построение трендов.

Тема 4. Технология Data Mining

Индикаторы достижения: УК-1.1.

1. Введение в интеллектуальный анализ данных
2. Субъектные и объектные системы.

3. Первая модель управления (внешнее управление).
4. Сравнительный анализ подходов к понятию управления: как средство достижения цели, как процесс управляющих воздействий, как процесс обработки информации.
5. Управление как процесс выбора и принятия решений.
6. Вторая модель управления - автоматическое управление
7. Предварительная обработка данных
8. Задача классификации с обучением
9. Основные принципы построения архитектуры
10. Области применения методов интеллектуального анализа.
11. Системы рассуждений на основе аналогичных случаев.
12. Классификационные и регрессионные деревья решений.
13. Байесовское обучение (ассоциации).
14. Генетические алгоритмы.
15. Эволюционное программирование и алгоритмы ограниченного перебора.

Критерии оценки (в баллах):

20 баллов выставляется обучающемуся, если он демонстрирует высокий уровень владения материалом по всем темам дискуссий, превосходное умение формулировать свою позицию, отстаивать её в споре, задавать вопросы, обсуждать дискуссионные положения, высокий уровень этики ведения дискуссии. Уровень сформированности компетенций соответствует продвинутому уровню;

15- баллов выставляется обучающемуся, если он демонстрирует владение материалом по всем темам дискуссий на уровне выше среднего, умение отстаивать её в споре, задавать вопросы, обсуждать дискуссионные положения, знание этики ведения дискуссии. Уровень сформированности компетенций соответствует повышенному уровню;

10 баллов выставляется обучающемуся, если он демонстрирует владение материалом по всем темам дискуссий не в полном объеме, умение задавать вопросы, обсуждать дискуссионные положения, знание этики ведения дискуссии. Уровень сформированности компетенций соответствует базовому уровню.

Контрольная работа

По теме 4. Технология Data Mining

Индикаторы достижения: УК-1.1.

1. Приведите формальную постановку основных задач Data Mining.
2. Перечислите и охарактеризуйте основные задачи Data Mining.
3. К основным задачам Data Mining относятся:

4. Постановка задачи классификации.
5. Постановка задачи регрессии.
6. Постановка задачи кластеризации.
7. Постановка задачи поиска ассоциативных правил.
8. Постановка задачи поиска последовательности.
9. Постановка задачи прогнозирования.
10. Чем отличается классификация от кластеризации?
11. Приведите варианты, сходство и отличия задач группировки и их графическую иллюстрацию.
12. Приведите отличия и сходство задачи прогнозирования от задачи поиска ассоциативных правил с привлечением общей системной модели решения задач и на основе их формальной постановки.
13. Опишите сущность формального концептуального анализа и его применение.
14. Сформулируйте постановку и приведите примеры задачи лингвистического резюмирования.
15. Охарактеризуйте задачу поиска аномалий и сопоставьте ее с кластеризацией данных на разных уровнях представления.
16. Подберите адекватные формальные постановки задач Data Mining для решения следующих проблем:
 - a. Разделить проекты на проекты, имеющие высокую степень успешности, среднюю и неуспешные, и определить их метрики.
 - b. Найти аналогичные проекты по набору требований в виде гетерогенных параметров.
 - c. Определить риски проектирования и реализации программного продукта по диаграмме выгорания.
 - d. Спрогнозировать время выполнения проекта исходя из набора требований, KPI персонала и объема финансирования.
 - e. Определить, какую квалификацию имеет «Большинство разработчиков».
17. В чем отличие технологии Data Mining от других технологий анализа больших объемов данных?
18. Назовите наиболее распространенные методы анализа структурированных данных с использованием технологии Data Mining. Какие типы скрытых зависимостей они выявляют?
19. Почему при создании человеко-машинных информационных систем большое внимание уделяется зависимостям, полученным в виде правил «если, то»?
20. Кратко опишите суть метода построения дерева решений. Какие ограничения имеются у этого метода?
21. Кратко опишите суть метода поиска ассоциативных правил. Какие ограничения имеются у этого метода?
21. В чем суть алгоритма Apriori, каковы его основные шаги? Каков смысл свойства анти-монотонности?

22. Что такое обобщенное ассоциативное правило? В чем отличие алгоритма вычисления обобщенных ассоциативных правил от алгоритма Apriori?

23. Какие улучшения требуются алгоритму поиска обобщенных ассоциативных правил для обеспечения приемлемых технических характеристик?

24. Какие прогрессивные методы Data Mining предполагается развивать в ближайшее время для получения скрытых зависимостей?

25. Какие программные продукты, реализующие технологию Data Mining, Вам известны? Чем вызвано такое их разнообразие?

26. Какие программные продукты позволяют получать скрытые зависимости в виде правил «если, то»?

27. Приведите примеры инцидентов и проблем, возникающих при управлении ИТ службами организаций или компаний.

28. Почему после выяснения и устранения проблемы необходимо еще некоторое время отслеживать ранее возникавшие инциденты?

29. Как объяснить разное число правил, найденных для одного массива данных методами построения деревьев и поиска ассоциативных зависимостей?

30. Почему вид правил, найденных различными методами при обработке одного и того же набора данных, различается и «условными» и «следственными» частями?

31. Определите главные отличия результатов, получаемых методами построения деревьев и поиска ассоциативных зависимостей?

Критерии оценки (в баллах):

10 баллов выставляется обучающемуся, если он правильно ответил на 80% вопросов (компетенция сформирована на продвинутом уровне);

9 баллов выставляется обучающемуся, если он правильно ответил на 70% вопросов (компетенция сформирована на продвинутом уровне);

8 баллов выставляется обучающемуся, если он правильно ответил на 60% вопросов теста (компетенция сформирована на повышенном уровне);

7 баллов выставляется обучающемуся, если он правильно ответил на 55% вопросов теста (компетенция сформирована на повышенном уровне);

6 баллов выставляется обучающемуся, если он правильно ответил на 50% вопросов теста (компетенция сформирована на базовом уровне)

Задания для текущего контроля

Комплект тестовых заданий по дисциплине

Индикаторы достижения: УК-1.1.

1. Какие науки включены в Data Mining?

- а. статистика, базы данных, искусственный интеллект;
- б. информатика, базы данных, статистика;
- в. искусственный интеллект, базы данных, базы знаний;
- г. информатика, базы данных, хранилища данных.

2. Какая подсистема СППР отвечает за интеллектуальный анализ данных?

- а. OLTP;
- б. хранилище данных;
- в. SQL;
- г. OLAP;
- д. Data Mining.

3. Репозиторий – это...

- а. словарь терминов;
- б. хранилище метаданных;
- в. каталог с файлами.

4. Если для реализации многомерной модели используют многомерные базы данных, то способ реализации гиперкуба называется

- а. MOLAP;
- б. ROLAP;
- в. HOLAP;
- г. DOLAP;
- д. JOLAP.

5. Задача классификации сводится к ...

- а. нахождению частых зависимостей между объектами или событиями;
- б. определению класса объекта по его характеристикам;
- в. определению по известным характеристикам объекта значения некоторого его параметра;
- г. поиску независимых групп и их характеристик во всем множестве анализируемых данных.

6. Целью поиска ассоциативных правил является ...

- а. нахождение частых зависимостей между объектами или событиями;
- б. определение класса объекта по его характеристикам;
- в. определение по известным характеристикам объекта значения некоторого его параметра;
- г. поиск независимых групп и их характеристик во всем множестве анализируемых данных.

7. Назовите основной этап работы с хранилищами данных?

- а. этап очистки данных;
- б. этап обновления;

в. этап нормализации.

8. Что называют кубом OLAP?

- а. структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице измерений;
- б. структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице фактов;
- в. таблицу размерностей.

9. Информационные хранилища созданы для удобства ...

- а. руководителей всех уровней для принятия решений;
- б. стратегического планирования;
- в. реорганизации бизнеса;
- г. предметных приложений;
- д. редактирования данных.

10. Информационные хранилища размещаются на ...

- а. библиотеках-автоматах;
- б. сетевых серверах;
- в. мейнфреймах;
- г. серверах и кластерах серверов;
- д. файл-серверах.

11. Ассоциация — ...

- а. установление зависимости непрерывной выходной переменной от входных переменных;
- б. группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
- в. выявление закономерностей между связанными событиями;
- г. установление зависимости дискретной выходной переменной от входных переменных.

12. Сумма расстояний между двумя точками, принадлежащими к одному кластеру, которые строго меньше, чем расстояние между двумя точками, не принадлежащими к одному кластеру определяет

- а. Качество кластеризации
- б. Робастность классификации
- в. Производительность кластеризации

13. Точность классификации объектов определяется точностью по кроссвалидации в виде

- а. Процента правильно классифицированных объектов
- б. Отношения количества правильно классифицированных объектов к количеству неправильно классифицированных объектов

в. Дисперсией значений признаков

14. Методы CART, C4.5, Randomforest относят к методам

- а. Кластеризации
- б. Классификации
- в. Ассоциации

15. Задача поиска редких и нетипичных объектов относится к задаче

- а. Поиска ассоциативных правил
- б. Поиска аномалий
- в. Кластеризации

16. Основные сквозные цифровые технологии Программы «Цифровая экономика Российской Федерации»

- а. робототехника; автономные системы; облачные вычисления
- б. искусственный интеллект; робототехника; анализ Больших данных; интернет вещей и индустриальный интернет
- в. робототехника; умные устройства; интернет вещей и индустриальный интернет

17. Большие данные характеризуются наиболее точно

- а. Большим объемом данных в хранилищах
- б. Большим объемом разнообразных данных, имеющих высокую скорость изменчивости
- в. Большим объемом слабоструктурированных данных

18. Для выявления групп схожих паттернов в данных по множеству признаков применяют

- а. Методы ассоциации
- б. Методы прогнозирования
- в. Методы классификации
- г. Методы кластеризации

19. Дано множество данных обучающих примеров, для каждого из которых известны классы принадлежности, требуется для новых данных, для которых неизвестны их классы, определить эти классы. Какие модели для решения указанной задачи применимы

- а. Модели обучения с учителем
- б. Модели обучения без учителя
- в. Модели смешанного обучения
- г. Модели кластерного обучения

20. Для решения задач предиктивной аналитики в экономике применяют

- а. Модели иерархической кластеризации

- б. Модель на основе дерева решений
- в. Модель CRISP-DM
- г. Модель Hadoop

21. Прогнозирование оттока клиентов в бизнесе относится к задачам
- а. Предиктивной аналитики
 - б. Диагностической аналитики
 - в. Дескриптивной аналитики
 - г. Когнитивной аналитики.

Критерии оценки (в баллах):

10 баллов выставляется обучающемуся, если он правильно ответил на 10 вопросов теста (компетенция сформирована на продвинутом уровне);

9 баллов выставляется обучающемуся, если он правильно ответил на 9 вопросов теста (компетенция сформирована на продвинутом уровне);

8 баллов выставляется обучающемуся, если он правильно ответил на 8 вопросов теста (компетенция сформирована на повышенном уровне);

7 баллов выставляется обучающемуся, если он правильно ответил на 7 вопросов теста (компетенция сформирована на повышенном уровне);

6 баллов выставляется обучающемуся, если он правильно ответил на 6 вопросов теста (компетенция сформирована на базовом уровне)

5 баллов выставляется обучающемуся, если он правильно ответил на 5 вопросов теста (компетенция сформирована на базовом уровне).

Типовые практические задания

Индикаторы достижения: УК-1.1.

Задание 1.

Поиск источников информации в сети Интернет: открытые и закрытые источники данных. Портал открытых данных РФ. Сохранение данных в программе Excel. Преобразование и первичная обработка данных.

Задание 2.

Представление исходных данных в программе R-Studio (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах Excel и R-Studio: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий

(регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы).

Задание 3.

Возможности графического представления информации в программе R-Studio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

Задание 4

Исходная база Больших данных содержит «сырые» многомерные числовые данные.

Требуется создать формальную постановку задачи, извлечь и описать данные, применить метод кластеризации для выборочных данных, сформировать группы похожих объектов и вычислить признаки типичного объекта для каждой группы. Подготовить отчет.

Критерии оценки (в баллах):

5 баллов выставляется обучающемуся, если он в полном объеме и правильно выполнил задание (компетенция сформирована на продвинутом уровне);

4 балла выставляется обучающемуся, если он в полном объеме и с незначительными замечаниями выполнил задание (компетенция сформирована на повышенном уровне);

3 балла выставляется обучающемуся, если он на базовом уровне, ошибками выполнил задание (компетенция сформирована на базовом уровне).

Задания для творческого рейтинга

Тематика докладов

Индикаторы достижения: УК-1.1.

1. Критерии оценки программных средств Data Mining.
2. Классификация программных средств Data Mining.
3. Программное обеспечение Data Mining для поиска ассоциативных правил.
4. Программное обеспечение для решения задач кластеризации и сегментации.
5. Программное обеспечение для решения задач классификации.
6. Программное обеспечение Data Mining для решения задач оценивания и прогнозирования.
7. Обзор программного средства RapidMiner.
8. Обзор Statistica.
9. Обзор Deductor.
10. Обзор KNIME.
11. Обзор Weka.

12. Обзор SAS Enterprise Miner.
13. Обзор IBM SPSS Modeler.
14. Обзор Oracle Data Mining ODM.
15. Поддержка принятия решений на основе анализа Больших данных
16. Большие данные и искусственный интеллект
17. Экосистема Hadoop
18. Архитектуры систем для обработки Больших данных
19. Методы TextMining
20. Применение технологий Большие данные для решения задач в микроэкономике
21. Применение технологий Большие данные для решения задач в макроэкономике
22. Ограничения технологии Большие данные
23. Большие данные и хранилища данных.
24. Применение Больших данных для извлечения новых знаний о клиентах
25. Этапы обработки Больших данных
26. Озера данных и Большие данные
27. Облачные решения Больших данных
28. Интеграция Больших данных с моделями машинного обучения
29. Большие данные и DataScience
30. Большие данные и интернет-вещей
31. Большие данные как дальнейшее развитие информационной экономики
32. Цифровая экономика и цифровая трансформация, взаимодействие с большими данными
33. Распределенные вычисления и хранилища данных (облачное хранение)
34. Проблема создания и размещения дата-центров
35. Большие данные и принятие решений.
36. Искусственный интеллект
37. Понятие bigdata. Новые подходы к накоплению и обработке данных в экономике и финансах на микро- и макроуровнях.
38. Открытые данные компьютерных поисковых систем и социальных сетей. GoogleTrends. YandexWorstat.
39. Цифровая трансформация модели потребления.
40. Тенденции в области цифрового контента.
41. Применение методов машинного обучения к большим данным
42. Применение методов Data Mining к большим данным
43. Применение нейронных сетей к анализу больших данных.
44. Поиск ассоциативных правил в больших данных.
45. Программные средства анализа «больших данных»

Критерии оценки (в баллах):

20 баллов выставляется обучающемуся, если он без ошибок подготовил доклад. Содержание и оформление доклада соответствует требованиям в полном

объеме. Уровень сформированности компетенций соответствует продвинутому уровню;

19-14 балла выставляется обучающемуся, если он с незначительными замечаниями по содержанию или оформлению подготовил отчет по реферату. Уровень сформированности компетенций соответствует повышенному уровню;

13-10 балла выставляется обучающемуся, если он с ошибками подготовил отчет по реферату. Содержание и оформление реферата соответствует требованиям не менее, чем на 50 процентов. Уровень сформированности компетенций соответствует базовому уровню.

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ХАРАКТЕРИЗУЮЩИЕ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ ВО ВРЕМЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Структура зачетного задания

<i>Наименование оценочного средства</i>	<i>Максимальное количество баллов</i>
<i>Вопрос 1</i>	<i>15</i>
<i>Вопрос 2</i>	<i>15</i>
<i>Практическое задание 1</i>	<i>10</i>

Задания, включаемые в зачетное задание

Перечень вопросов к зачету:

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 BigData в различных сферах деятельности.
12. Применение BigData 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 технологий 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. Понятие и характеристики больших данных: современные подходы к обработке и хранению.
27. Терминология, история появления. Технические сложности работы с

- большими данными.
28. Влияние больших данных.
 29. Роль и место больших данных в решении аналитических и исследовательских задач профессиональной деятельности.
 30. Источники больших данных.
 31. Платформа больших данных.
 32. Наука о данных.
 33. Процесс изучения данных.
 34. Методы исследования больших данных.
 35. Общая схема анализа.
 36. Извлечение и визуализация данных.
 37. Этапы моделирования.
 38. Процесс построения моделей.
 39. Формы представления данных, типы и виды данных.
 40. Представление наборов данных.
 41. Аналитические платформы: классификация и особенности применения.
 42. Языки визуального моделирования.
 43. Реляционные базы данных.
 44. Распределенные СУБД.
 45. Технологии GOOGLE BIGTABLE.
 46. Полнотекстовый поиск.
 47. Параллельные запросы.
 48. Технология поиска и интеграция.
 49. Процесс обработки отчетов.
 50. Экосистемы больших данных и его компоненты.
 51. Облачные решения BigData.
 52. Интеграция данных с использованием технологий BigData.
 53. Построение и загрузка озер и хранилищ данных.
 54. Обеспечение качества данных.
 55. Понятие о технологии Data Mining.
 56. Реализация в пакетах прикладных программ.
 57. Сетевые технологии Data Mining.
 58. Подготовка данных к анализу.
 59. Методика извлечения знаний.
 60. Актуальность технологий Data Mining как средств обработки больших объемов информации.
 61. Ассоциативные правила.
 62. Поиск и достоверность ассоциативных правил.
 63. Алгоритм генерации ассоциативных правил.
 64. Определение кластеризации.
 65. Цели кластеризации в Data Mining.
 66. Применение классификации и регрессии.
 67. Статистические методы.
 68. Основные понятия нейронных сетей.

69. Определение и структура дерева решений.
70. Выбор атрибута и разбиение в узле.

Практические задания к зачету:

1. Используя методологию DataMining, решите задачу кластеризации данных в объекте исследования «Финансовая компания».
2. Используя методологию DataMining, решите задачу классификации данных в объекте исследования «Финансовая компания».
3. Используя методологию DataMining, решите задачу прогнозирования данных в объекте исследования «Финансовая компания».
4. Используя методологию DataMining, решите задачу поиска аномалий в данных в объекте исследования «Финансовая компания».
5. Используя методологию DataMining, решите задачу кластеризации данных в объекте исследования «Интернет-магазин».
6. Используя методологию DataMining, решите задачу классификации данных в объекте исследования «Интернет-магазин».
7. Используя методологию DataMining, решите задачу прогнозирования данных в объекте исследования «Интернет-магазин».
8. Используя методологию DataMining, решите задачу поиска аномалий в данных в объекте исследования «Интернет-магазин».
9. Используя методологию DataMining, решите задачу кластеризации данных в объекте исследования «Рекламная компания».
10. Используя методологию DataMining, решите задачу классификации данных в объекте исследования «Рекламная компания».
11. Используя методологию DataMining, решите задачу прогнозирования данных в объекте исследования «Рекламная компания».
12. Используя методологию DataMining, решите задачу поиска аномалий в данных в объекте исследования «Рекламная компания».
13. Используя методологию DataMining, решите задачу кластеризации социально-экономических данных субъектов РФ.
14. Используя методологию DataMining, решите задачу классификации социально-экономических данных субъектов РФ.
15. Используя методологию DataMining, решите задачу прогнозирования данных социально-экономических данных региона.
16. Используя методологию DataMining, решите задачу поиска аномалий в данных социально-экономических данных региона.
17. Задача классификации сводится к ...
 - а. нахождению частых зависимостей между объектами или событиями;
 - б. определению класса объекта по его характеристикам;
 - в. определению по известным характеристикам объекта значение некоторого его параметра;
 - г. поиске независимых групп и их характеристик во всем множестве анализируемых данных.

18. Если несколько событий связаны друг с другом, то это...
- а. ассоциация;
 - б. последовательность;
 - в. классификация;
 - г. кластеризация.
19. Что называют хромосомами в генетических алгоритмах?
- а. кодировку исходных логических закономерностей в базе данных;
 - б. направление эволюционного программирования;
 - в. большой класс систем;
 - г. набор закономерностей.
20. Как называется класс систем, архитектура которых имеет аналогию с построением нервной ткани из нейронов?
- а. статистические пакеты;
 - б. деревья решений;
 - в. нейронные сети;
 - г. генетические алгоритмы.
21. Как переводится Data Mining?
- а. «добыча» или «раскопка данных»;
 - б. «значение»;
 - в. «хранение»;
 - г. «перечисление данных».
22. Аналитик это ...
- а. специалист в области анализа и моделирования
 - в. специалист в предметной области;
 - г. человек, решающий определенные задачи;
 - д. человек, который имеет опыт в программировании.
23. Задача кластеризации заключается в ...
- а. нахождения частых зависимостей между объектами или событиями;
 - б. определения класса объекта по его характеристикам;
 - в. определение по известным характеристикам объекта значение некоторого его параметра;
 - г. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
24. К предсказательным моделям относятся такие модели данных:
- а. модели классификации и последовательностей;
 - б. регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - в. классификации, кластеризации, исключений, итоговые и ассоциации;
 - г. модели классификации, последовательностей и исключений.
25. Модели классификации описывают ...
- а. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в. функциональные зависимости между зависимыми и независимыми

показателями и переменными в понятной человеку форме;

г. группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

26. Регрессивные модели описывают ...

а. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;

б. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;

в. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;

г. группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

27. Модели исключений описывают ...

а. исключительные ситуации в записях, которые резко отличаются чем-либо от основного множества записей;

б. ограничения на данные анализируемого массива;

в. закономерности между связанными событиями;

г. группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

28. Модели ассоциации проявляют ...

а. исключительные ситуации в записях, которые резко отличаются чем-либо от основного множества записей;

б. ограничения на данные анализируемого массива;

в. закономерности между связанными событиями;

г. группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

29. Очистка данных — ...

а. комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б. процесс дополнения данных некоторой информацией, позволяющей повысить эффективность решения аналитических задач;

в. объект, содержащий структурированные данные, которые могут оказаться полезными для решения аналитической задачи;

г. комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразование в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

30. Консолидация — ...

а. комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б. процесс дополнения данных некоторой информацией, позволяющей повысить эффективность решения аналитических задач;

в. объект, содержащий структурированные данные, которые могут оказаться полезными для решения аналитической задачи;

г. комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразование в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

31. Метаданные — ...

а. некоторый набор операций над базой данных, который рассматривается как единое завершённое, с точки зрения пользователя, действие над некоторой информацией, обычно связанное с обращением к базе данных;

б. разновидность систем хранения, ориентированная на поддержку процесса анализа данных, обеспечивающая целостность, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов;

в. высокоуровневые средства отражения информационной модели и описания структуры данных;

г. установление зависимости дискретной выходной переменной от входных переменных.

32. Регрессия — ...

а. установление зависимости непрерывной выходной переменной от входных переменных;

б. группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;

в. выявление закономерностей между связанными событиями;

г. установление зависимости дискретной выходной переменной от входных переменных.

33. Ассоциация — ...

а. установление зависимости непрерывной выходной переменной от входных переменных;

б. группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;

в. выявление закономерностей между связанными событиями;

г. установление зависимости дискретной выходной переменной от входных переменных.

34. Аналитическая платформа — ...

а. специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных;

б. группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;

в. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданные входные переменные и возвращающий правильный выходной результат;

г. раздел теории искусственного интеллекта, изучающий методы

построения алгоритмов, способных обучаться на данных.

35. Ошибка обучения — ...

- а. ошибка, допущенная моделью на обучающем множестве;
- б. неизвестная ошибка, допускаемая моделью на данных, не использовавшихся при обучении, называется ошибкой обобщения;
- в. имена, типы, метки и назначения полей исходной выборки данных;
- г. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданные входные переменные и возвращающий правильный выходной результат.

36. Исследование и обнаружение машиной (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые: ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком, называется

- а. OLTP;
- б. хранилищем данных;
- в. OLAP;
- г. Data Mining.

37. Какие науки включены в Data Mining:

- а. статистика, базы данных, искусственный интеллект;
- б. информатика, базы данных, статистика;
- в. искусственный интеллект, базы данных, базы знаний;
- г. информатика, базы данных, хранилища данных.

**Показатели и критерии оценивания планируемых результатов освоения компетенций и результатов обучения,
шкала оценивания**

Шкала оценивания		Формируемые компетенции	Индикатор достижения компетенции	Критерии оценивания	Уровень освоения компетенций
85 – 100 баллов	«отлично»	УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи	Знает верно и в полном объеме основные методы критического анализа и основы системного подхода как общенаучного метода.	Продвинутый
				Умеет верно и в полном объеме анализировать задачу, используя основы критического анализа и системного подхода.	
				Умеет верно и в полном объеме осуществлять поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации.	
70 – 84 баллов	«хорошо»	УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения	УК-1.1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи	Знает с незначительными замечаниями основные методы критического анализа и основы системного подхода как общенаучного метода.	

		поставленных задач		<p>Умеет с незначительными замечаниями анализировать задачу, используя основы критического анализа и системного подхода.</p> <p>Умеет с незначительными замечаниями осуществлять поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации.</p> <p>Умеет с незначительными замечаниями: использовать современный инструментарий и интеллектуальные информационно-аналитические системы</p>	Повышенный
50 – 69 баллов	«удовлетворительно»	УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи	<p>Знает на базовом уровне с ошибками основные методы критического анализа и основы системного подхода как общенаучного метода.</p> <p>Умеет с незначительными замечаниями анализировать задачу, используя основы критического анализа и системного подхода.</p> <p>Умеет с незначительными замечаниями осуществлять поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации.</p>	Базовый

				Умеет с незначительными замечаниями: использовать современный инструментарий и интеллектуальные информационно-аналитические системы	
менее 50 баллов	«неудовлетворительно»	УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи	Не знает на базовом уровне основные методы критического анализа и основы системного подхода как общенаучного метода.	Компетенции не сформированы
				Не умеет на базовом уровне анализировать задачу, используя основы критического анализа и системного подхода.	
				Не умеет с на базовом уровне осуществлять поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации.	
				Не умеет на базовом уровне: использовать современный инструментарий и интеллектуальные информационно-аналитические системы	