Примерные оценочные материалы, применяемые при проведении промежуточной аттестации по дисциплине (модулю)

«Методы анализа больших данных в управлении качеством»

При проведении промежуточной аттестации обучающемуся предлагается дать ответы на вопросы из нижеприведённого списка.

**Примерный перечень вопросов (зачёт)**

1. Что такое Big Data и каковы основные характеристики больших данных (Volume, Velocity, Variety, Veracity)?
2. Опишите основные технологии хранения и обработки больших данных.
3. В чем разница между традиционными СУБД и NoSQL базами данных? Приведите примеры использования.
4. Что такое MapReduce и как он применяется для обработки больших данных?
5. Опишите архитектуру и основные компоненты Hadoop.
6. Какие существуют методы очистки и предобработки больших данных?
7. Каковы основные подходы к визуализации больших данных?
8. Опишите методы распределенных вычислений в контексте обработки больших данных.
9. В чем заключается принцип работы машинного обучения на больших данных?
10. Какие существуют алгоритмы кластеризации для больших данных?
11. Опишите методы прогнозирования временных рядов на больших данных.
12. Что такое ассоциативные правила и как они применяются в анализе больших данных?
13. Какие существуют подходы к обеспечению безопасности больших данных?
14. Опишите методы снижения размерности данных в контексте больших массивов информации.
15. В чем заключается принцип работы рекомендательных систем на больших данных?
16. Какие существуют методы обнаружения аномалий в больших данных?
17. Опишите основные подходы к параллельной обработке данных.
18. Что такое распределенные файловые системы и как они используются для хранения больших данных?
19. Какие существуют методы оценки качества моделей на больших данных?
20. Опишите основные принципы работы с потоковыми данными.
21. В чем заключается принцип работы алгоритмов классификации на больших данных?
22. Какие существуют методы интеграции разнородных данных?
23. Опишите основные подходы к обеспечению целостности данных при их обработке.
24. Какие существуют методы оптимизации запросов к большим данным?
25. Опишите основные принципы работы с графовыми данными в контексте больших данных.

**Примерный перечень задач для проведения текущего и промежуточного контроля (зачёт)**

Тест № 1

1. Что такое большие данные (Big Data)?

а) Набор данных, которые невозможно обработать традиционными методами.

б) Набор данных, которые можно обработать с помощью специализированного программного обеспечения.

+в) Набор данных, которые имеют большой объём, скорость изменения и разнообразие.

г) Набор данных, которые имеют высокую ценность для бизнеса.

1. Какие характеристики больших данных являются основными?

а) Объём, скорость изменения и разнообразие.

б) Ценность для бизнеса, объём, скорость изменения и разнообразие.

+в) Объём, скорость изменения, разнообразие и достоверность.

г) Объём, скорость изменения, разнообразие и ценность для бизнеса.

1. Какие технологии используются для хранения больших данных?

а) Реляционные базы данных и файловые системы.

б) Реляционные базы данных, NoSQL базы данных и файловые системы.

в) Реляционные базы данных, NoSQL базы данных, графовые базы данных и файловые системы.

+г) Реляционные базы данных, NoSQL базы данных, графовые базы данных и облачные хранилища.

1. Что такое MapReduce?

а) Метод обработки больших данных, основанный на распределённых вычислениях.

+б) Технология для работы с большими данными, использующая Hadoop.

в) Метод анализа временных рядов на основе машинного обучения.

г) Метод очистки и предобработки больших данных.

1. Что такое Hadoop?

а) Технология для работы с большими данными, использующая MapReduce.

+б) Архитектура для распределённого хранения и обработки больших данных.

в) Технология для работы с потоковыми данными.

г) Технология для работы с графовыми данными.

1. Какие методы используются для очистки и предобработки больших данных?

а) Фильтрация, нормализация и агрегация.

б) Фильтрация, нормализация, агрегация и обогащение.

+в) Фильтрация, нормализация, обогащение и дедупликация.

г) Фильтрация, нормализация, обогащение и кластеризация.

1. Какие подходы используются для визуализации больших данных?

а) Графики, диаграммы и таблицы.

б) Графики, диаграммы, таблицы и карты.

+в) Графики, диаграммы, таблицы, карты и тепловые карты.

г) Графики, диаграммы, таблицы, карты и деревья решений.

1. Что такое машинное обучение?

а) Метод анализа больших данных, основанный на статистических моделях.

+б) Метод анализа больших данных, основанный на искусственных нейронных сетях.

в) Метод анализа больших данных, основанный на деревьях решений.

г) Метод анализа больших данных, основанный на кластеризации.

1. Какие методы используются для классификации больших данных?

а) Методы опорных векторов и деревья решений.

+б) Методы опорных векторов, деревья решений и искусственные нейронные сети.

в) Методы опорных векторов, деревья решений, искусственные нейронные сети и кластеризация.

г) Методы опорных векторов, деревья решений, искусственные нейронные сети и метод k-средних.

1. Какие методы используются для интеграции разнородных данных?

а) Объединение, слияние и сопоставление.

б) Объединение, слияние, сопоставление и трансформация.

+в) Объединение, слияние, сопоставление, трансформация и обогащение.

г) Объединение, слияние, сопоставление, трансформация и дедупликация.

Тест № 2

1. Что такое база данных?

а) Набор структурированных данных

б) Система хранения информации

в) Компьютерная программа для работы с данными

+г) Всё вышеперечисленное

1. Какие существуют типы баз данных?

а) Реляционные и нереляционные

б) Иерархические, сетевые и реляционные

в) Реляционные, иерархические и сетевые

+г) Все вышеперечисленные типы

1. Что такое реляционная база данных?

+а) База данных, основанная на таблицах

б) База данных, использующая иерархический подход

в) База данных, использующая сетевой подход

г) База данных, основанная на графическом представлении данных

1. Что такое SQL?

а) Язык программирования для работы с базами данных

б) Стандарт для хранения и обработки данных

в) Программное обеспечение для управления базами данных

+г) Всё вышеперечисленное

1. Что такое нормализация базы данных?

а) Процесс разделения таблиц на несколько меньших таблиц

б) Метод оптимизации запросов к базе данных

в) Техника обеспечения безопасности данных

+г) Всё вышеперечисленное

1. Что такое индекс в базе данных?

+а) Структура данных, используемая для ускорения поиска информации

б) Инструмент для обеспечения безопасности данных

в) Метод оптимизации запросов к базе данных

г) Всё вышеперечисленное

1. Что такое OLTP (Online Transaction Processing)?

+а) Технология обработки транзакций в реальном времени

б) Метод оптимизации запросов к базе данных

в) Программное обеспечение для управления базами данных

г) Всё вышеперечисленное

1. Что такое OLAP (Online Analytical Processing)?

а) Технология обработки транзакций в реальном времени

б) Метод анализа данных в базах данных

в) Программное обеспечение для управления базами данных

+г) Всё вышеперечисленное

1. Что такое ETL (Extract, Transform, Load)?

а) Процесс извлечения данных из различных источников

б) Процесс преобразования данных для соответствия требованиям базы данных

в) Процесс загрузки данных в базу данных

+г) Всё вышеперечисленное

1. Что такое NoSQL?

а) Технология обработки транзакций в реальном времени

б) Метод анализа данных в базах данных

+в) Нереляционная база данных

г) Всё вышеперечисленное