1.Видобування даних – дослідження і виявлення машиною в сирих даних прихованих знань, які раніше не були відомі, нетривіальні, практично корисні, доступні для інтерпретації людиною.

2.Базові методи – основною їх перевагою є їх простота, як щодо розуміння, так і реалізації. До недоліків можна віднести відсутність формальної теорії, на підставі якої будуються такі алгоритми, а отже, складнощі пов’язані з їх дослідженням і розвитком. Нечітка логіка – перевагою є те, що найяскравіше виявляються на прикладі загальної задачі нечіткого керування, дозволяється чітко і вдало подати мислення людини. Правила прийняття рішень є умовними висловами і реалізуються за допомогою механізму логічного виведення. Замість чіткого узагальненого правила нечітка нечітка логіка оперує з множиною приватних правил. Правила умовного типу дозволяють вирішувати задачі класифікації у режимі діалогу з оператором, що сприяє підвищенню якості класифікації вже в процесі експлуатації. Генетичні алгоритми - інтеграція ГА і нейронних мереж дозволяє вирішувати проблеми пошуку оптимальних значень ваг входів нейронів, а інтеграція ГА і нечіткої логіки дозволяє оптимзувати систему продукційних правилд, які можуть бути використані для керування операторами ГА. Нейронні мережі –перевагою нейронних мереж є те, що вони можуть апрксимувати будь-яку інформацію будь-яку неперервну функцію і тому досліднику немає необхідності наперед приймати якісь гіпотези стосовно моделі , але недоліком є те, що остаточне рішення залежить від початкових налаштувань мережі.

3.На ринку існує безліч інструментів для видобування даних – Analysis Services, SAS Enterprise Miner, PolyAnalyst, Cognos, STATISTICA Data Miner, Oracle, Data Mining, Oracle Deductor, KXEN.

4. Кластеризація на відміну від класифікації має неконтрольоване навчання, стратегією є начання без вчителя, має невідомі мітки класу навчальної множини , підставою для визначення класу в класифікації є підстава навчальної множини, а в кластеризації в початковій множині даних встановлюється існування класів або кластерів даних.

5. Асоціативна залежність – залежність між наявними однорідними сутностями яка подається у вигляду правил які можуть бути використані як для кращого розуміння природи аналізованих даних, так і для прогнозу подій.

6. Нейронні мережі - це клас моделей, заснованих на біологічній аналогії з мозком людини і призначених після проходження етапу так званого навчання на наявних даних для розв’язування різноманітних задач аналізу даних. За допомогою нейронних мереж можна апроксимувати будь-яку неперервну функцію.

7. Генетичні алгоритми (ГА) належать до універсальних методів оптимізації, що дозволяють вирішувати завдання різних типів (комбінаторні, загальні завдання з обмеженнями і без обмежень) і різного ступеня складності.

При цьому ГА характеризуються можливістю як однокритерійного, так і багатокритерійного пошуку у великому просторі, ландшафт якого є нерівним. Інтеґрація ГА і нейронних мереж дозволяє вирішувати проблеми пошуку оптимальних значень ваг входів нейронів, а інтеґрація ГА і нечіткої логіки дозволяє оптимізувати систему продукційних правил, які можуть бути використані для керування операторами ГА (двонапрямлена інтеґрація).

8. На ринку програмного забезпечення видобування даних існує величезна різноманітність продуктів, що відносяться до категорії видобування даних: Sql Server Analysis Services, SAS Enterprise Miner, PolyAnalyst, Cognos, STATISTICA Data Miner, Oracle Data Mining, Oracle Deductor, KXEN.

Тести

1. a) в)

2. а) б) г)

3. б)

4.1 а) ?

4.2 а)

5. б)

6. в)

7. г)

8. а) б)

9. в) г) д)

10. а) б) в)