

Data Mining

Питання до заліку

1. Зміст терміна "Data Mining".
2. Поняття статистики. Поняття машинного навчання. Поняття штучного інтелекту.
3. Розвиток технології баз даних.
4. Data Mining як частина ринку інформаційних технологій: класифікація аналітичних систем, думки експертів про Data Mining.
5. Проблеми технології Data Mining.
6. Відмінності Data Mining від інших методів аналізу даних. Існуючі підходи до аналізу.
7. Поняття про дані у широкому розумінні. Набір даних та їх атрибутів.
8. Змінна. Значення. Генеральна сукупність. Вибірка. Параметри. Статистики. Гіпотези.
9. Шкали: номінальна, порядкова, інтервальна, відносна, діхотомічна.
10. Типи наборів даних.
11. Табличні дані. Графічні дані. Формати зберігання даних.
12. Основні положення баз даних.
13. Системи управління базами даних (СУБД). Вимоги до СУБД.
14. Класифікація видів даних. Метадані.
15. Класифікація стадій Data Mining.
16. Класифікація методів Data Mining.
17. Статистичні методи Data Mining.
18. Кібернетичні методи Data Mining.
19. Властивості методів Data Mining.
20. Завдання Data mining: класифікація, кластеризація, асоціація, послідовна асоціація, прогнозування, оцінювання, аналіз зв'язків, візуалізація.
21. Класифікація завдань Data mining: за стратегіями: навчання з вчителем, навчання без вчителя, інші.
22. Класифікація завдань Data mining в залежності від моделей, що використовуються: описові, прогнозуючі.
23. Розподіл завдань Data mining: автоматичне дослідження та відкриття (вільний пошук), пояснення та опис, зв'язок понять.
24. Поток "від даних до рішень". Поток "від завдання до додатку".
25. Інформація. Властивості інформації. Знання та їх властивості.
26. Сфери застосування асоціативних правил.
27. Завдання аналізу ринкового кошика. Транзакційна база даних.
28. Визначення та характеристики асоціативних правил.
29. Межі підтримки та достовірності асоціативного правила.
30. Алгоритм Apriori та його різновиди.

31. Програмні засоби реалізації методу пошуку асоціативних правил.
32. Реалізація методу пошуку асоціативних правил у пакеті `arules` середовища R.
33. Поняття про кластерний аналіз.
34. Завдання кластерного аналізу.
35. Методи кластерного аналізу: ієрархічні та неієрархічні.
36. Міри подібності: квадрат евклідової відстані, Манхеттінська відстань, відстань Чебишева, відсоток незгоди.
37. Методи об'єднання або зв'язку.
38. Алгоритм k-середніх.
39. Перевірка якості кластеризації.
40. Застосування методу дерева рішень для завдань класифікації та прогнозування.
41. Переваги дерев рішень.
42. Процес конструювання дерева рішень.
43. Критерій розщеплення. Зупинка побудови дерева.
44. Скорочення дерева або відсікання гілок.
45. Алгоритми, що реалізують дерева рішень.
46. Завдання Data mining, що вирішуються з допомогою штучних нейронних мереж.
47. Елементи нейронних мереж. Архітектура нейронних мереж.
48. Навчання нейронних мереж.
49. Моделі нейронних мереж: одношаровий та багатошаровий перцептрон.
50. Програмне забезпечення для роботи з нейронними мережами.
51. Карти Кохонена, що самоорганізуються.
52. Проблеми дискримінантного аналізу.
53. Дискримінантний факторний аналіз.
54. Геометричний прогнозний дискримінантний аналіз.
55. Ймовірнісний дискримінантний аналіз.
56. Вимірювання якості моделі: лямбда Уїлкса, коефіцієнт детермінації, скорегований коефіцієнт детермінації.
57. Дискримінантний аналіз якісних змінних (метод DISQUAL).
58. Переваги та недоліки дискримінантного аналізу.
59. Послідовність етапів регресійного аналізу.
60. Завдання регресійного аналізу.
61. Тренд, сезонність та цикл.
62. Точність прогнозу часових рядів.
63. Види прогнозів часових рядів.
64. Методи прогнозування часових рядів.
65. Однофакторна лінійна регресія в середовищі R.
66. Множинна лінійна регресія у середовищі R.