

## ЗМІСТ

|            |   |
|------------|---|
| ВСТУП..... | 6 |
|------------|---|

### Частина 1. Моделі, методи й алгоритми DATA MINING

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| <b>1. Розв’язання задач класифікації.....</b>                                                      | <b>9</b>  |
| 1.1. Стислі теоретичні відомості.....                                                              | 9         |
| 1.2. Алгоритми класифікації.....                                                                   | 9         |
| 1.2.1. Байєсівська класифікація.....                                                               | 9         |
| 1.2.2. Алгоритм покриття.....                                                                      | 11        |
| 1.3. Приклад реалізації алгоритму покриття мовою JAVA.....                                         | 13        |
| 1.4. Контрольні запитання до роботи 1.....                                                         | 16        |
| 1.5. Завдання до роботи 1.....                                                                     | 17        |
| <b>2. Розв’язання задач кластеризації.....</b>                                                     | <b>20</b> |
| 2.1. Стислі теоретичні відомості.....                                                              | 20        |
| 2.2. Алгоритми кластеризації.....                                                                  | 21        |
| 2.3. Приклад програми кластеризації алгоритмом Fuzzy C Means<br>(нечітка розбивка) мовою Java..... | 25        |
| 2.4. Контрольні запитання до роботи 2.....                                                         | 28        |
| 2.5. Завдання до роботи 2.....                                                                     | 28        |
| <b>3. Пошук асоціативних груп об’єктів.....</b>                                                    | <b>31</b> |
| 3.1. Стислі теоретичні відомості.....                                                              | 31        |
| 3.1.1. Задача пошуку асоціативних груп.....                                                        | 31        |
| 3.1.2. Сиквенціальний аналіз.....                                                                  | 32        |
| 3.1.3. Представлення результатів.....                                                              | 32        |
| 3.2. Алгоритми пошуку асоціативних правил.....                                                     | 33        |
| 3.3. Приклад програми пошуку асоціативних груп мовою Java.....                                     | 36        |
| 3.4. Контрольні запитання до роботи 3.....                                                         | 39        |
| 3.5. Завдання до роботи 3.....                                                                     | 39        |
| <b>4. Довгострокове прогнозування часових рядів.....</b>                                           | <b>40</b> |
| 4.1. Стислі теоретичні відомості.....                                                              | 40        |
| 4.1.1. Часовий ряд та його компоненти.....                                                         | 40        |
| 4.1.2. Характеристики випадкового процесу.....                                                     | 42        |
| 4.2. Алгоритм аналізу часових рядів.....                                                           | 43        |
| 4.3. Приклад демонстраційної програми декомпозиції часового ряду<br>тренд-аналізом мовою VBA.....  | 49        |
| 4.4. Контрольні запитання до роботи 4.....                                                         | 60        |
| 4.5. Завдання до роботи 4.....                                                                     | 60        |

|                                                                             |           |
|-----------------------------------------------------------------------------|-----------|
| <b>5. Адаптивне прогнозування часових рядів.....</b>                        | <b>61</b> |
| 5.1. Стислі теоретичні відомості.....                                       | 61        |
| 5.1.1. Модель Брауна.....                                                   | 61        |
| 5.1.2. Модель Трігга-Ліча-Шоуна.....                                        | 63        |
| 5.1.3. Модель адаптивного фільтру.....                                      | 64        |
| 5.2. Алгоритми адаптивного прогнозування часових рядів.....                 | 65        |
| 5.2.1. Алгоритм адаптивного прогнозування методом Брауна....                | 65        |
| 5.2.2. Алгоритм адаптивного прогнозування моделлю<br>Трігга-Ліча-Шоуна..... | 66        |
| 5.2.3. Алгоритм прогнозування моделлю адаптивного фільтра                   | 68        |
| 5.3. Приклад програми прогнозування адаптивним фільтром мовою<br>AVA.....   | 72        |
| 5.4. Контрольні запитання до практичної роботи 5.....                       | 74        |
| 5.5. Завдання до практичної роботи 5.....                                   | 74        |
| Література.....                                                             | 75        |

## **Частина 2. Застосування алгоритмів Datamining у середовищі MS SQL Server 2008**

|                                                                                                                           |           |
|---------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1. Лабораторна робота №1. Доступ до вхідних даних.....</b>                                                             | <b>77</b> |
| 1.1. Встановлення зв'язку з сервером за допомогою SSMS.....                                                               | 77        |
| 1.2. Створення нового проекту Analysis Services.....                                                                      | 78        |
| 1.3. Налаштування проекту.....                                                                                            | 80        |
| 1.4. Створення джерела даних (data source) та представлення<br>джерела даних.....                                         | 82        |
| 1.5. Модифікація представлення джерела даних.....                                                                         | 87        |
| 1.6. Створення іменованих запитів (named queries) .....                                                                   | 90        |
| <b>2. Лабораторна робота №2. Аналіз поштової розсилки за допомогою<br/>алгоритмів інтелектуального аналізу даних.....</b> | <b>96</b> |
| 2.1. Доступ до вхідних даних.....                                                                                         | 96        |
| 2.2. Створення структури DM.....                                                                                          | 97        |
| 2.3. Додавання нових моделей до структури DM.....                                                                         | 104       |
| 2.4. Аналіз дерева рішень.....                                                                                            | 107       |
| 2.5. Аналіз кластеризації.....                                                                                            | 110       |
| 2.6. Аналіз спрощеного алгоритму Баєса.....                                                                               | 112       |
| 2.7. Порівняння точності моделей.....                                                                                     | 114       |
| 2.8. Перевірка моделі з фільтром.....                                                                                     | 118       |
| 2.9. Створення прогнозів та робота з ними.....                                                                            | 120       |

|                                                                     |            |
|---------------------------------------------------------------------|------------|
| <b>3. Лабораторна робота №3.Алгоритм Microsoft Time Series.....</b> | <b>128</b> |
| 3.1. Аналіз тенденцій у часі.....                                   | 128        |
| 3.2. Аналіз вхідних даних.....                                      | 129        |
| 3.3. Створення структури та моделі DM.....                          | 132        |
| 3.4. Аналіз моделі DM.....                                          | 134        |
| 3.5. Прогнозування.....                                             | 136        |
| 3.6. Прогноз загальних обсягів продажів.....                        | 137        |
| <b>4. Лабораторна робота №4.Асоціативні правила.....</b>            | <b>140</b> |
| 4.1. Налаштування даних для аналізу покупок.....                    | 140        |
| 4.2. Створення структури DM та налаштування моделі.....             | 142        |
| 4.3. Аналіз взаємозв'язків.....                                     | 144        |
| 4.4. Створення моделі з фільтром.....                               | 147        |
| 4.5. Прогнозування кошику покупця.....                              | 150        |
| <b>5. Приклади контрольних завдань.....</b>                         | <b>155</b> |
| <b>6. Література.....</b>                                           | <b>159</b> |