

Wstęp (9)

- Proces eksploracji danych (10)
- Instalacja i konfiguracja narzędzi (12)
 - o Serwer SQL (12)
 - o Arkusz kalkulacyjny Excel (15)
 - o Dodatek Data Mining do pakietu Office (15)
 - o Przykłady (16)
- Konwencje i oznaczenia (20)

Część I: Modelowanie (23)

Rozdział 1. Eksploracja danych jako technika wspomagania decyzji (25)

- Modelowanie świata (25)
 - o Obiekty, zdarzenia i reguły (26)
 - o Dane (27)
 - o Informacje (27)
 - o Wiedza (29)
 - o Decyzje (31)
- Eksploracja danych (32)
 - o Hipotezy (32)
 - o Kłopoty ze sformułowaniem problemu (33)

Rozdział 2. Analiza biznesowa (35)

- Cele modelowania i eksploracji danych (35)
 - o Opisywanie danych czy wspieranie decyzji? (36)
 - o Decydenci (38)
- Zakres projektu eksploracji danych (39)
 - o Dane źródłowe (40)
 - o Kontekst (40)
- Sprecyzowanie spodziewanych wyników (42)
 - o Modele deskrypcyjne (43)
 - o Modele predykcyjne (43)
- Prawdopodobieństwo sukcesu projektu eksploracji danych (44)
- Ocena ryzyka (45)

Rozdział 3. Ocena danych (49)

- Dane źródłowe (49)
 - o Błędy pomiaru (50)
- Przypadki, czyli to, co badamy (51)
- Profilowanie danych za pomocą usługi SQL Server Integration Services (54)
- Atrybuty i ich stany (57)
 - o Atrybuty jednowartościowe i wielowartościowe (57)
 - o Atrybuty monotoniczne (59)
 - o Rozkład wartości (59)
- Integralność danych (62)
 - o Duplikaty (62)
 - o Zakres wartości (63)

- o Zgodność ze wzorcem (63)
- Próbkowanie i reprezentatywność danych (64)
 - o Próbkowanie danych (64)
 - o Zbieżność do rzeczywistego rozkładu (65)
 - o Odchylenie standardowe (67)
 - o Zmienność atrybutów tekstowych (68)
- Brakujące dane (69)
 - o Model brakujących danych (70)
- Zależności pomiędzy atrybutami (73)
 - o Niezależne atrybuty (74)
 - o Nadmiarowe atrybuty (75)
 - o Anachronizmy (76)
- Mierzenie informacji (76)
 - o Bity (77)
 - o Zaskoczenie (77)
 - o Kontekst (78)

Rozdział 4. Przygotowanie danych (79)

- Przestrzeń stanów (79)
- Atrybuty dyskretne (81)
 - o Grupowanie (81)
 - o Numerowanie stanów (84)
 - o Atrybuty porządkowe (85)
 - o Atrybuty okresowe (86)
- Atrybuty ciągłe (86)
 - o Wartości skrajne (87)
 - o Normalizacja zakresu (87)
 - o Dyskretyzacja (90)
- Serie danych (92)
 - o Trend (96)
 - o Okresowość i sezonowość (96)
 - o Szum (97)

Rozdział 5. Poprawa jakości danych (99)

- Uzupełnienie wartości (99)
- Wzbogacenie danych (103)
- Redukcja wymiarów (105)
 - o Korelacje (106)
- Dane dla modeli deskrypcyjnych (108)
- Dane dla modeli predykcyjnych (109)
 - o Zmiana proporcji (109)
- Dane na potrzeby analizy wariantowej (111)
 - o Analiza wariantowa (111)
- Wydzielenie danych testowych (113)

Część II: Eksploracja (117)

Rozdział 6. Techniki eksploracji danych (119)

- Zastosowania (119)
- Dodatek Data Mining do pakietu Office (121)
 - o Ocena i przygotowanie danych źródłowych (121)
- Techniki eksploracji danych (126)
 - o Klasyfikacja (126)
 - o Szacowanie (136)
 - o Asocjacja (141)
 - o Grupowanie (145)
 - o Analiza sekwencyjna (151)
 - o Analiza wariantowa (152)
 - o Prognozowanie (156)

Rozdział 7. Serwer SQL jako platforma eksploracji danych (161)

- Excel jako klient SQL Server Analysis Services (162)
 - o Narzędzia eksploracji zewnętrznych danych (162)
 - o Praca z modelami eksploracji danych (184)
 - o Formuły arkusza Excel (191)
- Projekty eksploracji danych (192)
 - o Business Intelligence Development Studio (192)
 - o Źródła danych (195)
 - o Widoki danych źródłowych (196)
 - o Struktury eksploracji danych (199)
 - o Modele eksploracji danych (206)
 - o Zapytania predykcyjne (210)
 - o Zagnieżdżanie przypadków (213)
- Zarządzanie serwerem SSAS i modelami eksploracji danych poprzez SQL Server Management Studio (216)
- Usługi eksploracji danych serwera SQL (218)
 - o Architektura (219)
 - o Bezpieczeństwo (221)
 - o Integracja z pozostałymi usługami Business Intelligence (223)

Rozdział 8. DMX (227)

- Terminologia (227)
 - o Atrybut (227)
 - o Wartość i stan (229)
 - o Przypadek (229)
 - o Klucze (230)
 - o Struktury eksploracji danych (231)
 - o Modele eksploracji danych (232)
- Składnia języka DMX (232)
 - o Tworzenie struktur eksploracji danych (233)
 - o Tworzenie modeli eksploracji danych (235)
 - o Przetwarzanie struktur i modeli eksploracji danych (239)
 - o Odczytywanie zawartości struktur i modeli eksploracji danych (243)
 - o Zapytania predykcyjne (245)
 - o Funkcje predykcyjne (251)

Rozdział 9. Naiwny klasyfikator Bayesa firmy Microsoft (253)

- Omówienie (253)
 - Ograniczenia (255)
 - Parametry (256)
- Zastosowania naiwnego klasyfikatora Bayesa (258)
 - Badanie zależności pomiędzy atrybutami (258)
 - Klasyfikacja dokumentów (260)

Rozdział 10. Drzewa decyzyjne firmy Microsoft i algorytm regresji liniowej firmy Microsoft (267)

- Omówienie (268)
 - Ograniczenia (272)
 - Parametry (273)
- Zastosowania drzew decyzyjnych (275)
 - Klasyfikacja klientów (275)
 - Szacowanie potencjalnych zysków (277)
 - Asocjacja klientów i wypożyczanych przez nich filmów (279)

Rozdział 11. Szeregi czasowe firmy Microsoft (281)

- Omówienie (281)
 - Ograniczenia (285)
 - Parametry (286)
- Zastosowania szeregów czasowych firmy Microsoft (288)
 - Prognozowanie sprzedaży (289)
 - Prognozowanie sprzedaży na podstawie przeplatanych serii danych (291)
 - Prognozowanie sprzedaży na podstawie danych odczytanych z kostki wielowymiarowej (292)
 - Prognozowanie sprzedaży na podstawie krótkich serii danych (293)
 - Analiza wariantowa (295)

Rozdział 12. Algorytm klastrowania firmy Microsoft (297)

- Omówienie (297)
 - Ograniczenia (302)
 - Parametry (303)
- Zastosowania algorytmu klastrowania (305)
 - Analiza skupień komórek (305)
 - Klasyfikacja komórek (309)
 - Przygotowanie danych do dalszej eksploracji (312)
 - Wykrywanie anomalii (314)

Rozdział 13. Algorytm klastrowania sekwencyjnego firmy Microsoft (319)

- Omówienie (320)
 - Ograniczenia (323)
 - Parametry (323)
- Zastosowania algorytmu klastrowania sekwencyjnego (324)

- o Analiza sekwencji odwiedzanych stron WWW (324)
- o Klasyfikacja klientów na podstawie kolejności kupowanych przez nich towarów (327)
- o Przewidywanie kolejnych zdarzeń (329)
- o Wykrywanie nietypowych sekwencji zdarzeń (332)

Rozdział 14. Algorytm odkrywania reguł asocjacyjnych firmy Microsoft (335)

- Omówienie (336)
 - o Ograniczenia (340)
 - o Parametry (341)
- Zastosowania reguł asocjacyjnych (341)
 - o Badanie zależności pomiędzy wartościami atrybutów (342)
 - o Analiza koszykowa (343)
 - o Analiza typu cross-selling (347)

Rozdział 15. Sieci neuronowe firmy Microsoft i algorytm regresji logistycznej firmy Microsoft (351)

- Omówienie (352)
 - o Ograniczenia (358)
 - o Parametry (360)
- Zastosowania sieci neuronowych i regresji logistycznej (361)
 - o Szacowanie potencjalnych zysków (362)
 - o Klasyfikacja dokumentów (366)

Rozdział 16. Ocena i poprawa modeli (369)

- Powrót do średniej (369)
- Kryteria porównawcze (371)
 - o Łatwość interpretacji (373)
 - o Dokładność predykcji (374)
 - o Wiarygodność predykcji (374)
 - o Wydajność i skalowalność (375)
 - o Przydatność (375)
- Metody oceniania modeli eksploracji danych (376)
 - o Wykresy podniesienia i zysku (376)
 - o Macierz klasyfikacji (384)
 - o Ocena dokładności modeli algorytmu szeregów czasowych firmy Microsoft (386)
 - o Walidacja krzyżowa (387)
 - o Odchylenie wewnątrz- i międzyklastrowe (390)
- Problemy (391)
 - o Niewłaściwie postawione zadania (391)
 - o Niewłaściwe dane źródłowe (392)
 - o Nieprzygotowane dane źródłowe (393)
 - o Niewłaściwe lub źle sparametryzowane algorytmy eksploracji danych (394)

Rozdział 17. Programowanie predykcyjne (397)

- Narzędzia programistyczne (397)
- Wizualizatory modeli eksploracji danych (398)
- Raporty usługi SSRS (399)
- Inteligentne aplikacje (401)
 - o Kontrola poprawności danych (401)
 - o Uzupełnianie brakujących danych (404)
 - o Adaptacyjny interfejs (406)

Skorowidz (415)