

dm 1

Data Mining. Metodologia prowadzenia projektów, przygotowywanie danych i wprowadzenie do modelowania

Kurs podstawowy z zakresu drążenia danych jest adresowany do osób, które chcą wykorzystywać techniki drążenia danych jako źródło wiedzy wspomagające proces podejmowania decyzji.

Kurs ten składa się z czterech elementów:

- wprowadzenia do drążenia danych,
- metodologii prowadzenia projektów drążenia danych,
- podstaw obsługi produktu firmy SPSS – CLEMENTINE,
- praktycznego wykorzystania technik drążenia danych w określonych zastosowaniach.

Wprowadzenie do tematyki drążenia danych obejmuje następujące zagadnienia:

- czym jest drążenie danych,
- użyteczność drążenia danych z perspektywy biznesowej,
- rola drążenia danych w systemach podejmowania decyzji,
- drążenie danych jako element budowy analitycznych systemów CRM,
- drążenie danych a infrastruktura informatyczna firmy.

Wstęp ten ma przedstawić drążenie danych w szerszym kontekście i pokazać jego rolę w codziennym funkcjonowaniu firmy jako elementu systemu wspomagania decyzji.

Drugim istotnym elementem będzie zapoznanie uczestników kursu z **metodologią prowadzenia projektów drążenia danych**. Projekty takie muszą być mocno zakorzenione w realiach biznesowych firmy i odpowiadać na ich konkretne potrzeby. Przyjęcie właściwej metodologii jest jednym z kluczowych elementów efektywnych rozwiązań drążenia danych. Ta część kursu bazuje na metodologii prowadzenia projektów drążenia danych – CRISP-DM. Prace nad tą metodologią zostały zainicjowane w 1996 roku przez trzy firmy – DAIMLERCHRYSLER (dawniej DAIMLER-BENZ), SPSS (dawniej ISL)

i NCR. Z czasem wokół projektu CRISP-DM została utworzona organizacja CRISP-DM SIG, skupiająca ponad dwustu członków. Jej celem jest dalszy rozwój metodologii drążenia danych. Prace nad metodologią są wspierane także przez Komisję Europejską.

W trakcie szkolenia omówione zostaną poszczególne **etapy procesu drążenia danych**: od definiowania celów biznesowych do wdrożenia wyników do pracy, jako niezbędnego warunku osiągnięcia tych celów.

Szczegółowo omówione zostaną kolejne etapy, tzn.:

- określanie celów biznesowych,
- audyt danych,
- przygotowanie danych,
- modelowanie,
- ewaluacja,
- wdrożenie.

Następnym elementem szkolenia będzie **wprowadzenie do obsługi programu CLEMENTINE** w stopniu umożliwiającym aktywny udział w części praktycznej. Omówione zostaną kwestie dostępu do danych. Sporo uwagi zostanie poświęcone procedurom wykorzystywanym do integracji danych z różnych źródeł, sprawdzania ich jakości i przygotowania ich do analiz. W tej części poruszone zostaną też zagadnienia eksploracji i graficznej prezentacji danych. W skrócie omówione zostaną techniki modelowania. Na zakończenie zajmiemy się prezentacją wyników i ich wdrażaniem do pracy.

Zasadniczą część kursu wypełnią **przykłady** praktycznego wykorzystania technik drążenia danych w określonych zastosowaniach:

- analiza migracji klientów (churn analysis),
- wykrywanie nadużyć (fraud analysis),
- web mining,
- analiza koszykowa (basket analysis),
- analiza ryzyka (risk analysis).

Omówione zostaną problemy, z którymi można się zetknąć w trakcie przygotowywania danych i możliwości radzenia sobie z nimi. Dużo uwagi poświęcone zostanie budowaniu modeli przewidujących zachowania klientów, określających ich skłonność do migracji, możliwości nadużyć czy preferencje do zakupu określonych produktów. Przedstawimy zwłaszcza budowanie takich modeli, które mogą być wykorzystywane w strukturze informatycznej firmy do poprawy relacji z klientami. Wszystkie analizy będą odbywać się na przykładowych danych z takich branż jak telekomunikacja, ubezpieczenia, bankowość, hipermarkety i e-sklepy.

wymagania wstępne dla uczestników kursu

Podstawowa znajomość obsługi komputera i środowiska MS WINDOWS.