

Grzegorz Stolecki

Wprowadzenie do Data Mining



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



**WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI**

**UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY**



O mnie...

Życiorys

- Z komputerem – od 1986 roku
- Z technologią **AI+BI** – od 1993 roku
- Z SQL Server – od 1998 roku
- MCTS BI

Co robię?

- Projektowanie i wdrażanie rozwiązań Business Intelligence
- Szkolenia i warsztaty

Co mnie interesuje?

- BI, kontroling, sztuczna inteligencja.
- Historia i militaria

Gdzie mnie można spotkać?

- Polish SQL Server User Group
- Śląska Regionalna Grupa Microsoft
- wss.pl



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Agenda

Co to...?

- Data Mining – parę podstawowych pojęć
- Co jest potrzebne by zacząć kopanie?

Co może...?

- Do czego można zastosować Data Mining?
- Algorytmy, modele, struktury

Jak to zrobić...?

- Kilka prostych przykładów



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – skąd się wziął?

Aby podjąć dobrą decyzję potrzebných jest trochę informacji...

I generacja: bazy danych ewidencyjne



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – skąd się wziął?

Aby podjąć dobrą decyzję potrzebných jest trochę informacji...

II generacja: h



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



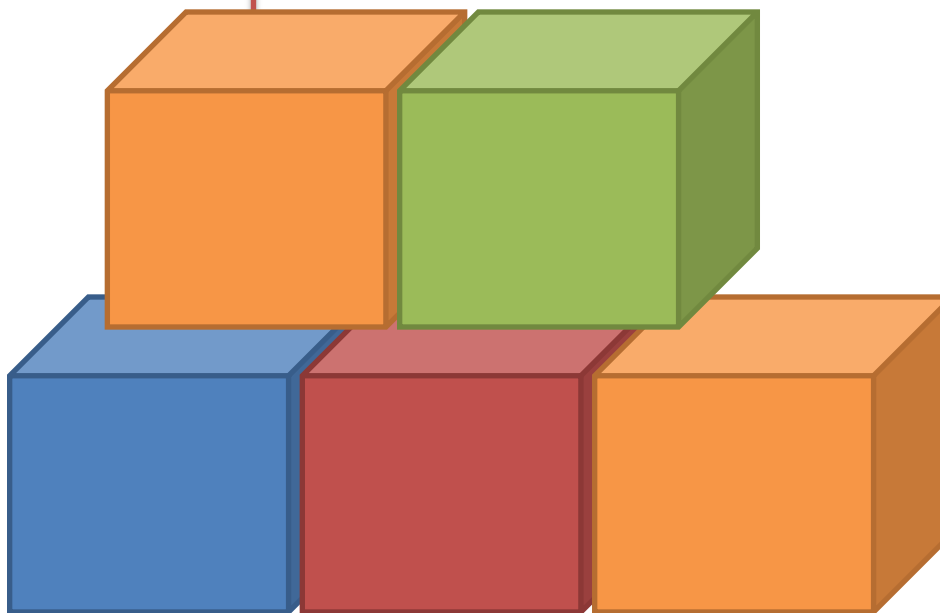
Data Mining – skąd się wziął?

Aby podjąć dobrą decyzję potrzebných jest trochę informacji...

III generacja: w

HELP !

OLAP



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



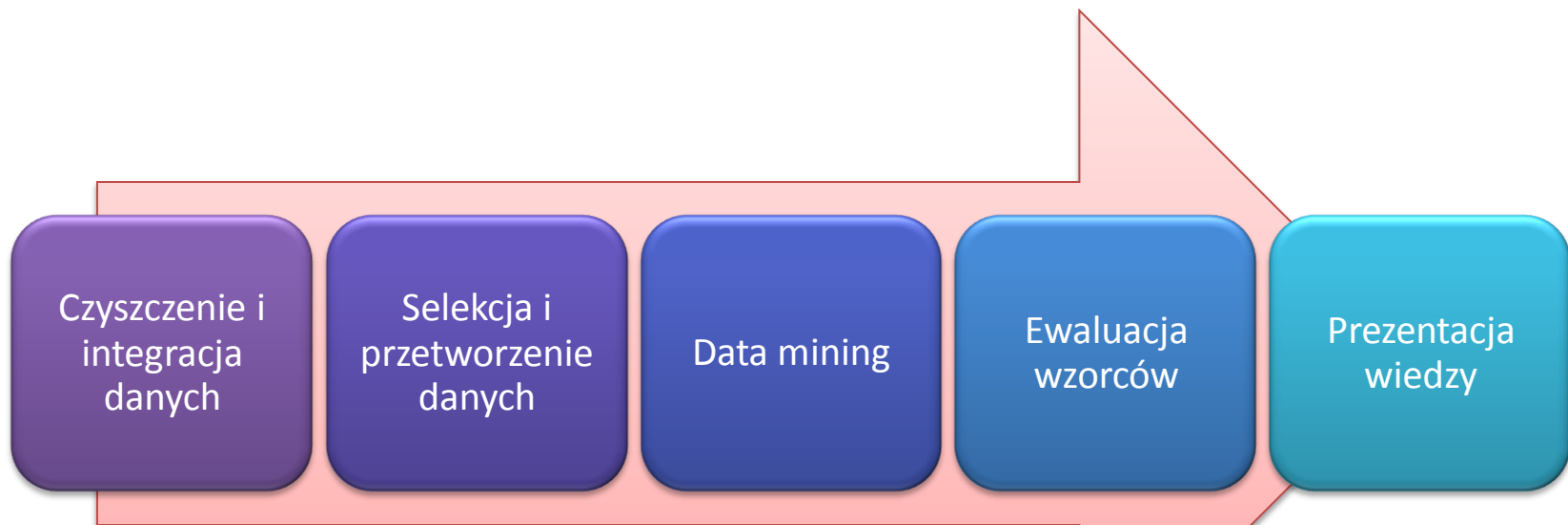
WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – co to jest?

Data mining to proces analizy, którego celem jest odkrycie cennej informacji, wzorców i wiedzy ukrytych w dużych ilościach danych zgromadzonych przez systemy informatyczne.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – zastosowania

- Analizy „churn”
- Koszyk towarów
- Wykrywanie oszustw i fałszerstw w transakcjach
- Ocena ryzyka (np. kredytowego)
- Segmentacja klientów
- Profilowane kampanie marketingowe
- Prognozowanie sprzedaży, kosztów
- Analiza odwiedzin serwisu WWW
- Wykrywanie wzorców zachowań
- Ocena, który z 500 milionów obiektów gwiazdnych jest wart zainteresowania
- Rekordowe wyniki sportowców – czynniki
- ... i wiele, wiele innych...



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

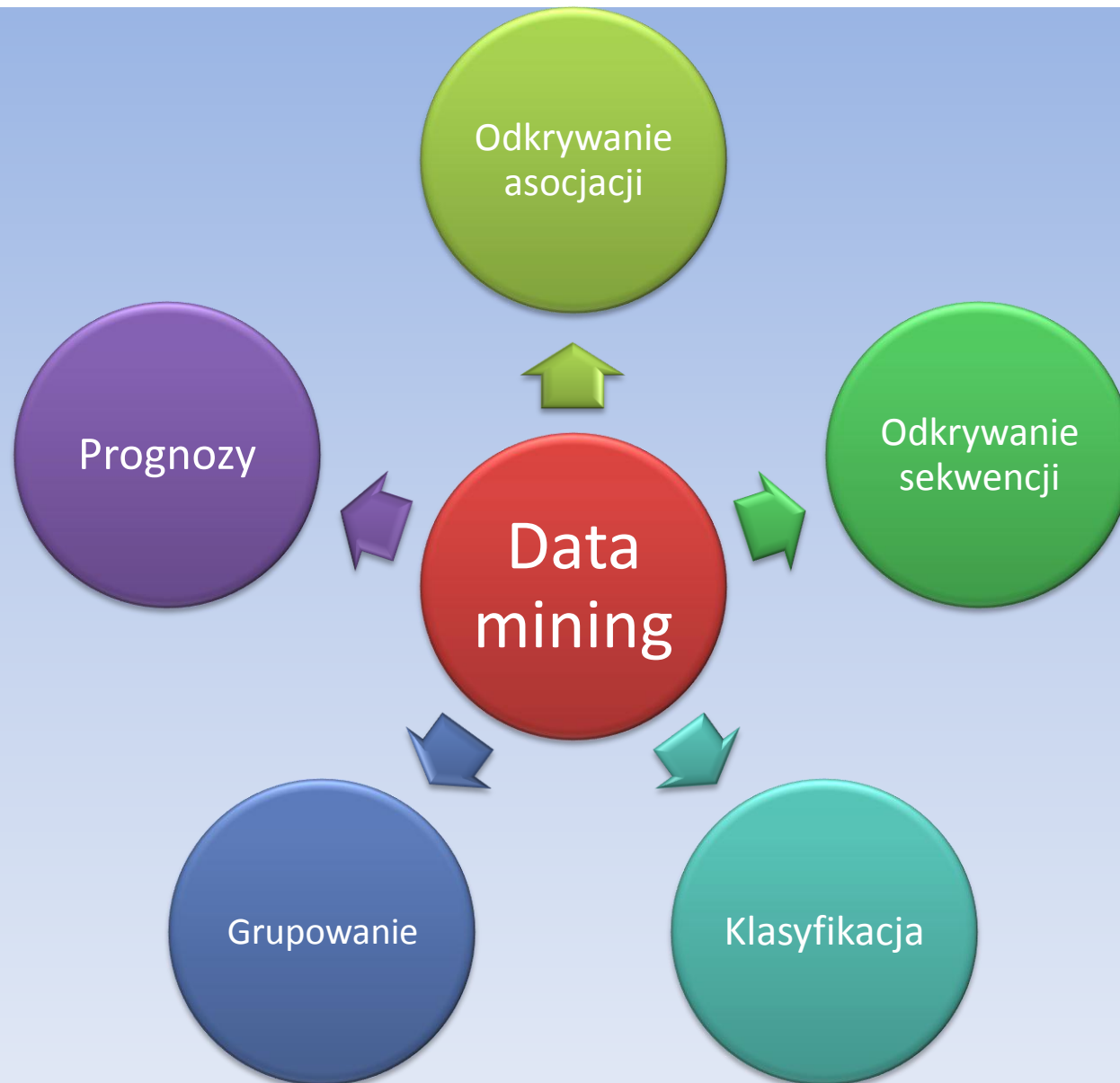


WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – techniki



Data Mining – jak się to robi (1)

Określenie problemu biznesowego

- Na jakie pytania ma odpowiadać model ?
- Kto będzie odbiorcą analiz ?

Zebranie danych

- Jakie informacje są potrzebne aby udzielić odpowiedzi na pytania ?
- Gdzie znajdują się wymagane dane ?

Konsolidacja i czyszczenie danych

- Utworzenie hurtowni danych
- Realizacja procedur ETL
- Ujednolicenie formatu i struktury danych



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

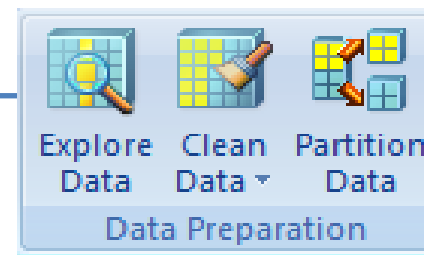
UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – jak się to robi (2)

Przygotowanie danych

- Analiza jakości danych
- Uzupełnienie danych brakujących
- Identyfikacja i eliminacja wartości ekstremalnych



Budowa modelu analitycznego

- Wybór rodzaju modelu
- Dobranie algorytmów data-mining
- Dobór optymalnych parametrów modelu

Trenowanie modelu

- Wybór testowej próbki danych
- Przetworzenie treningowej próbki danych w celu identyfikacji wzorców
- Analiza dopasowania wzorców do rzeczywistych danych



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – jak się to robi (3)

Interpretacja rezultatów

- Ocena merytorycznej poprawności odnalezionych wzorców i reguł
- Wykorzystanie wzorców do analizy całości dostępnych danych

Monitoring modelu

- Ocena jakości dopasowania modelu dla nowych danych



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – narzędzia Microsoft

- SQL Server Analysis Services (SQL Server)
- Business Intelligence Development Studio
- Data Mining Add-In dla Office 2007

- Third-party plugins



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



demo



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



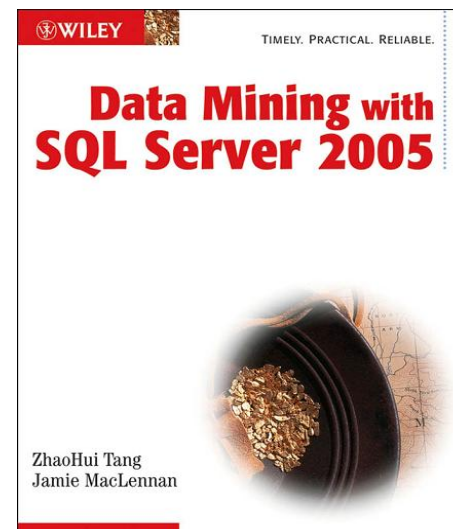
WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Data Mining – warto zobaczyć

- BOL – SQL Server Books Online
- <http://www.sqlserverdatamining.com/>
- Data Mining with SQL Server 2008
- RapidMiner (www.rapidminer.com)
- R Project, WEKA, Kmine



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Na koniec...

Q & A



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



To już naprawdę wszystko

Dziękuję za uwagę !

Grzegorz Stolecki

grzegorz.stolecki@plssug.org.pl



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

