

Cel projektu:

1. Zapoznanie się z możliwościami praktycznego zastosowania środowiska serwera baz danych oraz zaawansowanych procedur analizy danych realizowanymi z użyciem metod uczenia maszynowego.
2. Nabycie umiejętności wykorzystania opcji Oracle Data Miner w środowisku SQL Developer do rozwiązywania praktycznych problemów analizy danych.
3. Doskonalenie umiejętności pracy zespołowej.

Środki:

1. Serwer baz danych Oracle Enterprise Edition 11g.
2. Aplikacja SQL Developer skonfigurowana do pracy z Data Miner:
 - a. <http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r2/prod/bidw/datamining/ODM11gR2-SetUp.htm>
3. UCI Machine Learning Repository:
 - a. <http://archive.ics.uci.edu/ml/datasets.html>

Zakres projektu:

1. Wybrać dwa (można więcej) dowolne zbiory danych ze strony UCI Machine Learning Repository.
2. Dokonać analizy tych zbiorów:
 - a. Zapoznać się z poszczególnymi atrybutami.
 - b. Poznać format zapisu danych, typy danych itp.
 - c. Przeanalizować dotychczasowe zastosowania tych zbiorów (artykuły naukowe itp.).
3. Wykonać:
 - a. Zimportować dane do bazy Oracle wykorzystując dowolną technikę (np. tabele zewnętrzne: <http://oracle-base.com/articles/9i/external-tables-9i.php>, SQL Loader itp.).
 - b. Wykorzystać Oracle SQL Developer oraz serwer Oracle Database Enterprise Edition 11g (lub 12c) do przygotowania wybranego modelu (np. Classification) dla tych danych.
 - c. Sprawdzić różne algorytmy (np. Decision Tree) i ich skuteczność.
 - d. Wykonać eksperymenty z różnymi wartościami parametrów charakteryzujących poszczególne algorytmy – sprawdzić ich wpływ na wyniki działania modelu.
4. Przygotować dokumentację do projektu (podzieloną na dwa rozdziały, każdy dla jednego zbioru danych), obejmującą:
 - a. Dokładny opis konkretnego zbioru danych, źródła pochodzenia, jego atrybutów, dotychczasowych zastosowań itp.
 - b. Opisane wymagania jakie postawiliśmy sobie w budowaniu modelu, tj. np.: co zamierzamy klasyfikować (jeżeli wybraliśmy klasyfikację), który atrybut jest atrybutem decyzyjny itd.
 - c. Opis przebiegu procesu budowy modelu (w SQL Developer) wraz ze screenami.
 - d. Opis uzyskanych rezultatów wraz z komentarzami.
 - e. Wnioski wynikające z realizacji projektu – bardzo ważne! Próba oceny, jakie czynniki wpływają na jakość uzyskanych rezultatów. Co należałoby zrobić, aby podnieść jakość?

Termin i forma oddania projektu:

1. Termin: 2016.05.29. Przekroczenie terminu może wpływać na ocenę. Można oddawać wcześniej.
2. Forma oddania projektu:
 - a. Dokumentacja w formacie PDF oraz pliki bazy (export) przesłane mailem na adres kc@pk.edu.pl.
 - b. Ustna obrona projektu w grupach, wg ustalonych terminów (zostaną podane później).