

Wymagania minimalne na kolokwium końcowe:

1. Modele CRISP-DM i SEMMA
2. Wykorzystanie funkcji summary z pakietu dbms_stat_funcs
3. Zastosowanie klauzul Group By Rollup i Cube, funkcji Grouping, Grouping_Id, klauzuli Grouping Sets.
4. Funkcje analityczne i ich zastosowania: sum, count, min, max, lead, lag, ntile, width_bucket. Podział, sortowanie, okna itp.
5. Podstawy Algorytmów Genetycznych: elementy algorytmu, modyfikacje metod selekcji, krzyżowania i mutacji, rodzaje kodowania, schematy, algorytmy genetyczne dla zmiennych rzeczywistych.
6. Technik eksploracji danych: klasyfikacja, regresja. Zdefiniowanie pojęć, przykłady zastosowań, różnice.
7. Uczenie z nadzorem i bez nadzoru – pojęcia, zastosowania, różnice, przykłady głównych algorytmów.
8. Drzewa decyzyjne – ich zastosowania, parametry, budowa.
9. Naive Bayes – zastosowania, działanie.
10. Współczynniki oceny jakości modeli: zakładki Data Miner: Performance, Performance Matrix, ROC, Lift.
11. Budowa i zastosowanie Performance Matrix (Confusion Matrix). Współczynniki: Predictive confidence, Overall accuracy rate, Average accuracy rate, Sensitivity, Specificity, Error rate.
12. Obliczanie koszty błędów.
13. Zastosowanie i współczynniki dla Lift: Gain Cumulative, Target density, Records Cumulative.
14. Zastosowanie krzywych ROC. Opis wykresu, osie, sposób oceny algorytmów za pomocą ROC.
15. Jakość danych: cechy danych o wysokiej jakości, podstawowe metody poprawiania jakości danych, normalizacja min-max, standaryzacja, identyfikacja punktów oddalonych, rozstęp międzykwantylowy, odległość Levenshteina.