

# Data Mining - systémy pre podporu rozhodovania

## **Stručný syllabus:**

1. Úvod do hĺbkovej analýzy dát (data mining) – základné pojmy a definície, SEMMA metodológia, SAS Enterprise Miner – klasifikácia uzlov (nodes),
2. Prediktívne modelovanie pomocou rozhodovacích stromov – úvod do práce v SAS Enterprise Miner (tvorba projektov, resp. procesných diagramov), príprava dát pre modelovanie – základné zásady a ťažkosti, úvod do rozhodovacích stromov.
3. Prediktívne modelovanie pomocou rozhodovacích stromov – druhy rozhodovacích stromov, tvorba a interpretácia rozhodovacích stromov, porovnávanie a hodnotenie modelov.
4. Prediktívne modelovanie pomocou rozhodovacích stromov – porovnávanie a hodnotenie modelov (úvod): grafy navýšenia (Lift Chart), zmätočná matica (Confusion Matrix), vhodná prahová hodnota pre klasifikáciu pozorovaní (threshold), náklady na chybnú klasifikáciu (cost of misclassification), profitová matica (Profit Matrix),
5. Prediktívne modelovanie pomocou regresných modelov – porovnanie modelov lineárnej a logistickej regresie , príklad na tvorbu logistického modelu – príprava dát (nahradzovanie chýbajúcich hodnôt (Imputation, Replacement)), distribučná analýza premenných (Insight), transformácia premenných pre vstup do modelov.
6. Prediktívne modelovanie pomocou regresných modelov - odhad parametrov logistického a lineárneho modelu, metódy výberu premenných – Forward, Backward, Stepwise.
7. Výber premenných do prediktívnych modelov – metódy selekcie premenných a Enterprise Miner (Variable Selection Node)
8. Prediktívne modelovanie pomocou neurónových sietí – úvod do problematiky, základné pojmy, neurón, Multilayer Perceptron.
9. Prediktívne modelovanie pomocou neurónových sietí – tréovanie neurónovej siete, vizualizácia modelov.
10. Hodnotenie a implementácia modelov – metódy porovnávania modelov, lift chart, profit chart, ROC chart.
11. Hodnotenie a implementácia modelov – kombinovanie modelov, generovanie a použitie skórovacieho kódu (score code).
12. Zhuková analýza – metóda k-priemerov (k-means clustering), segmentácia zákazníkov.
13. Práca na semestrálnych projektoch v tímoch pod vedením učiteľa.
14. Práca na semestrálnych projektoch v tímoch pod vedením učiteľa.