

## **Semestrální práce - report**

### **Bankroty - Klasifikace**

**/Y336VD/**

Tomáš Dlouhý

## Zadání

Použijte metodu k nejbližších sousedů nebo metodu rozhodovacích stromů ke klasifikaci cílového atributu zvoleného datasetu. Zvolte si jednu třídu jako cílovou - pozitivní. Naleznete nejlepší klasifikátor, který má false positive rate  $FPr < 0.3$ . Spočtete přesnost (accuracy) a true positive rate tohoto klasifikátoru.

## Data

K tvorbě semestrální práce byl zvolen dataset „Bankrotů“, který obsahuje 44 ekonomických atributů v součtu 2050 firem. Některé firmy přežijí a jiné zbankrotují. Mezi nedostatky tohoto datasetu patří neexistence popisků, čímž je znemožněno získat relevanci těchto dat. Data je též nutné před zahájením práce normalizovat.

## Zpracování dat

Ke zpracování dat byl použit program OpenOffice Calc, pomocí kterého byly data normalizována na interval  $<0, 1>$ , metodou MinMax dle vzorce:

$$y = \frac{x - \min\{x_0, x_n\}}{\max\{x_0, x_n\} - \min\{x_0, x_n\}}$$

Tím byla zajištěna stejná váha atributů pro klasifikaci. Dataset byl následně rozdělen na dvě stejné části (trénovací a testovací), aby byla zajištěna dostatečná kvalita modelu.

Následně jsem musel nalést vhodné  $k$ , které jsem získal pomocí iterace z intervalu  $<1; 30>$  viz obrázek 1, kdy jsem se zaměřil na jejich přesnost s pomocí parametru 2. Jako nejlepší  $k$  se ukázalo  $k=10$  s přesností accuracy=0,9844 (graf 1). Pro tuto část byla použita trénovací data.

Pak jsem pro získání hodnot TN, TP, FP, FN použil funkci programu RapidMiner BinomicalClassification Performance (obrázek 2), pro kterou byla použita testovací data,

a dosadil do vzorce:  $FPr = \frac{FP}{FP + TN}$  a výsledek byl 0,0315. Dále jsem spočítal

$$Tpr = \frac{TP}{TP + FN} = 0,989 \quad \text{a} \quad Acc = \frac{TP + TN}{TP + TN + FP + FN} = 0,9805$$

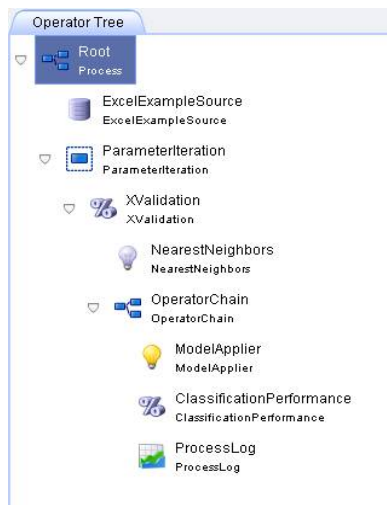
## Závěr

Nejlepší klasifikátor byl s parametrem  $K=10$ , pro něj je  $Fpr = 0,0315$ ,  $Acc=0,9805$  a  $Tpr=0,989$ .

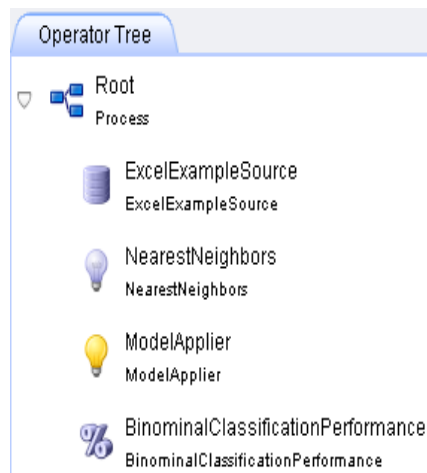
## Použitá literatura

[1] Stránky předmětu Y336VD - <http://ida.felk.cvut.cz/moodle>

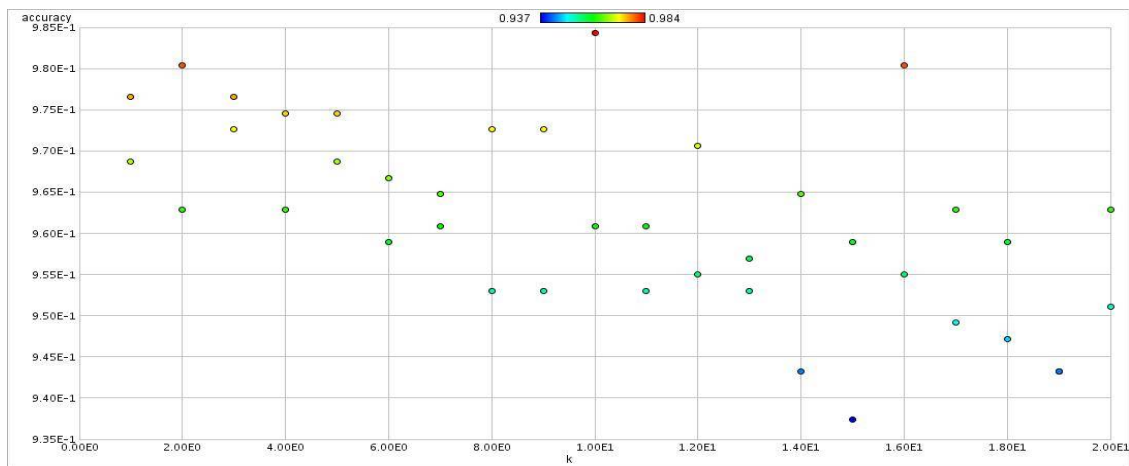
Přílohy



Obrázek 1



Obrázek 2



Graf 1