
Dobývání znalostí (zápočtový projekt)

Jaromír Šatánek

Petr Poštulka

Cíl projektu

- Cílem projektu bylo pro vygenerovaná vstupní data (seznam soukromých firem, jejich vlastnosti, požadavek na úvěr) pořídit reálné výstupy (dostane úvěr / nedostane úvěr). Na základě těchto dat vygenerovat rozhodovací strom a pravidla.
-

Popis důležitých atributů

- Obrat – množství peněz, které se ve firmě „protočí“.
 - Přímé náklady – např. nákup zboží, náklady vynaložené na výrobu výrobků apod.
 - Marže – poměr obrat / přímé náklady
 - Provozní náklady – celkové náklady na výrobu ve společnosti s výjimkou finančních nákladů
-

Popis důležitých atributů

- Provozní zisk – obrat – provozní náklady
 - Zadluženost – poměr závazky / vlastní jmění
Dělí se na 3 druhy (výrobní, služby, obchod)
 - Likvidita – poměr pohledávky / závazky
 - Poměr provozní zisk / finální náklady
-

Trénovací data

Rostoucí obrat	Rostoucí marže	Provozní zisk	Obrat	Zadluženost	Pohledávky / závazky	Provoz. zisk / fin. náklady	Úvěr poskytnut?
Ano	Ne	350 K	80 M	5:1	2:4	2:1	Ne
Ano	Ano	50K	1 M	3:1	1:9	4:3	Ano
Ne	Ano	50 M	500 M	1:2	1:3	6:1	Ne
Ano	Ano	3M	100 M	8:1	2:1	2:4	Ne

Trénovací data

Rostoucí obrat	Rostoucí marže	Provozní zisk	Obrat	Zadluženost	Pohledávky / závazky	Provoz. zisk / fin. náklady	Úvěr poskytnut?
Ano	Ano	- 2M	50 M	4:1	3:2	5:2	Ne
Ano	Ne	4M	320 M	2:1	1:1	2:4	Ne
Ne	Ano	6M	130 M	3:1	3:1	4:3	Ano
Ano	Ano	500K	2 M	4:1	2:1	4:1	Ano

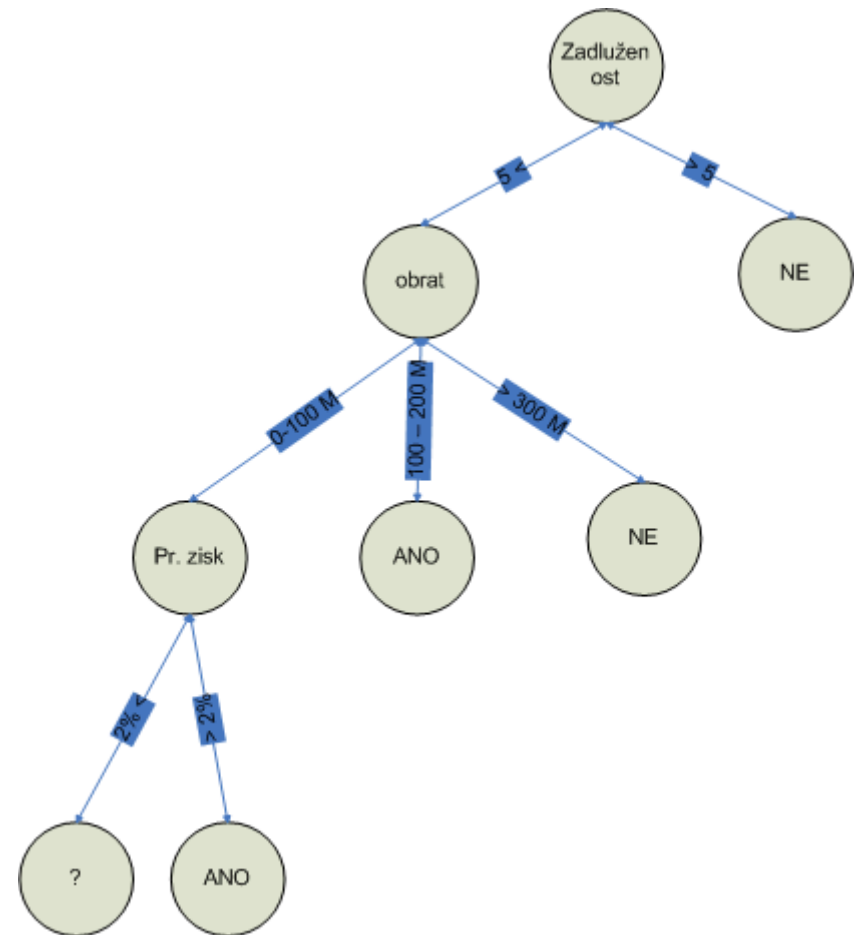
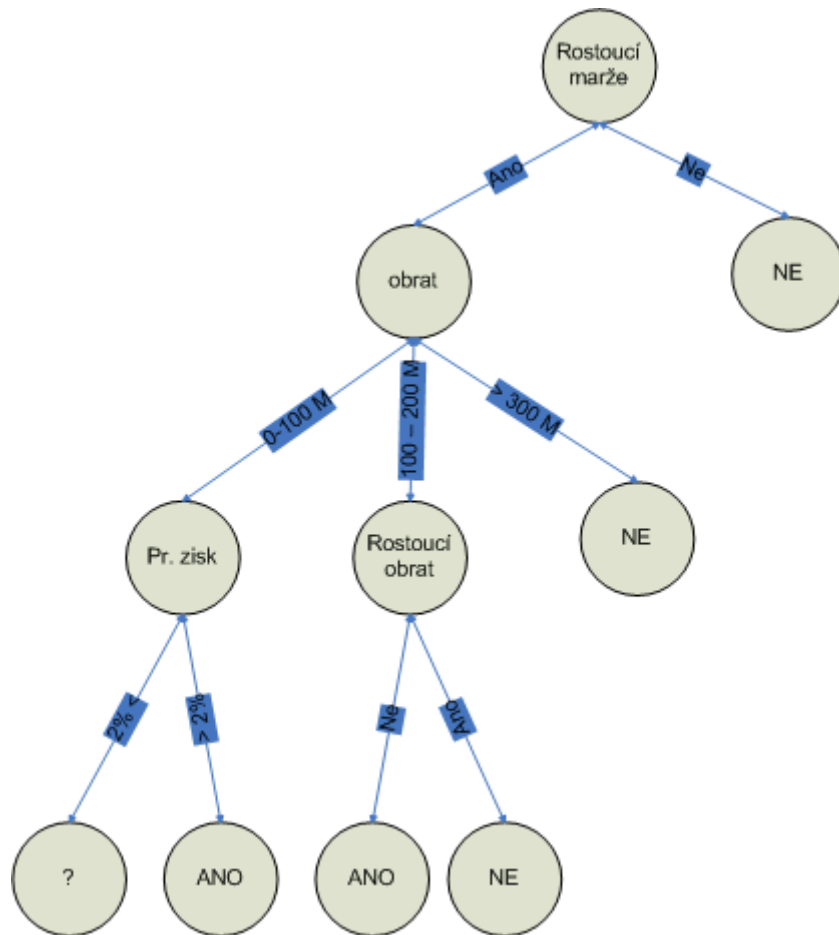
Trénovací data

Rostoucí obrat	Rostoucí marže	Provozní zisk	Obrat	Zadluženost	Pohledávky / závazky	Provoz. zisk / fin. náklady	Úvěr poskytnut?
Ano	Ano	110K	20 M	3:1	1:9	3:1	Ano
Ne	Ne	1.5 M	270 M	7:1	1:8	2:1	Ne

Výpočet

- Atributy se spojitými hodnotami jsme rozdělili do intervalů.
 - Aplikovali jsme algoritmus ID3
 - Výpočet entropií za pomoci naprogramovaných funkcí v matlabu
-

Rozhodovací stromy



Rozhodovací pravidla strom 1

- `if(!rostoucí marže) then NE`
 - `if(rostoucí marže && obrat>300M) then NE`
 - `if(rostoucí marže && obrat in (100M až 300M) && rostoucí obrat) then NE`
 - `if(rostoucí marže && obrat in (100M až 300M) && (!rostoucí obrat)) then ANO`
 - `if(rostoucí marže && obrat<100M && provozní zisk<2%) then ???`
 - `if(rostoucí marže && obrat<100M && provozní zisk>2%) then ANO`
-

Rozhodovací pravidla strom 2

- `if(zadluženost>5:1) then NE`
- `if(zadluženost<5:1 && obrat > 300M) then NE`
- `if(zadluženost<5:1 && obrat in (100M az 300M)) then ANO`
- `if(zadluženost<5:1 && obrat<100M && provozní zisk > 2%) then ANO`
- `if(zadluženost<5:1 && obrat<100M && provozní zisk < 2%) then ???`

Závěr

- Zvolený algoritmus neposkytl na trénovacích datech příliš dobré výsledky.
 - špatné rozdělení intervalů pro spojité atributy
 - špatně nagenеровaná data
- Roli hraje také náhodný výběr atributů se stejnou váženou entropií – viz rozdílné rozhodovací stromy
- Kladnou částí projektu bylo posouzení generovaných dat na základě reálných podmínek.