

Analiza karakteristika

Testiranje hipoteza

Varijabla vision je kategorijalna varijabla koja opisuje viziju posla kojom se poduzeće bavi:

- clear vision – poduzeće ima jasnu viziju posla kojim se bavi, dobro napravljen poslovni plan
- no clear vision – poduzeće nema jasnu viziju posla kojim se bavi, nema dobro napravljen poslovni plan
- established business – poduzeće ima uhodani posao kojim se bavi već nekoliko godina

Elementarna statistika varijable s obzirom na urednost vraćanja kredita (dobar/loš):

```
Cell Contents
-----
                N
Chi-square contribution
    N / Row Total
    N / Col Total
    N / Table Total
-----
```

Total observations in Table: 195

sme.df\$vision	sme.df\$default		Row Total
	bad	good	
clear_vision	5 2.115 0.179 0.076 0.026	23 1.082 0.821 0.178 0.118	28 0.144
established_business	49 0.150 0.320 0.742 0.251	104 0.077 0.680 0.806 0.533	153 0.785
no_clear_vision	12 11.128 0.857 0.182 0.062	2 5.693 0.143 0.016 0.010	14 0.072
Column Total	66 0.338	129 0.662	195

Iz tablice se može uočiti da je od svih klijenata koji imaju jasnu viziju posla 82,1% je dobrih i 17,9% loših dok je kod onih koji nemaju jasnu viziju posla postotak onih koji su dobri 14,3%, a onih koji su loši 85,7%. Kod uhodanih poduzeća 85,7% je dobrih i 14,3% loših. Iz prikazanog se može zaključiti da su rizičniji klijenti oni koji nemaju jasnu viziju posla u odnosu na one koji imaju uhodani posao ili počinju s poslom i imaju jasnu viziju posla kojim se žele baviti.

Testiranje zavisnosti između vizije posla i urednosti vraćanja kredita upotrebom hi-kvadrat testa potvrdilo je postojanje zavisnosti ($\chi^2=20,245$; $p<0.001$).

```
> chisq.test(sme.df$vision, sme.df$default)
```

```
    Pearson's Chi-squared test
```

```
data: sme.df$vision and sme.df$default  
X-squared = 20.245, df = 2, p-value = 4.017e-05
```

WOE – weight of evidence

Dodatna analiza karakteristika napravljena je kroz WOE (weight of evidence) tablicu:

Vizija	Broj dobrih	Broj loših	%dobrih	%loših	i.odds	w.o.e.	i.v.
Nema viziju	2	12	1.55%	18.18%	0.085	-2.465	0.410
Ima viziju	23	5	17.83%	7.58%	2.352	0.855	0.088
Uhodan posao	104	49	80.62%	74.24%	1.086	0.083	0.005
Ukupno	129	66	100%	100%			0.503

Kako se izračunavaju i tumače elementi u woe tablici:

$$\%dobri = \frac{\text{broj dobrih u toj kategoriji}}{\text{ukupan broj dobrih}} = \frac{2}{129} = 1,55\%$$

$$\%loši = \frac{\text{broj loših u toj kategoriji}}{\text{ukupan broj loših}} = \frac{12}{66} = 18,18\%$$

i.odds = information odds

$$i. odds = \frac{\%dobrih}{\%loših} = \frac{17,83\%}{7,58\%} = 2,352$$

Koliko dobrih dolazi na 1 lošeg?

Na 1 lošeg klijenta koji ima viziju za svoj poslovni pothvat dolazi 2,352 dobrih klijenata s vizijom.

lli

Šanse su 2,352:1 da ćemo pronaći klijenta s vizijom u dobrima u odnosu na to da ćemo pronaći klijenta s vizijom u lošima.

Ako je $i.odds > 1$, ima više dobrih nego loših u tom razredu odnosno za taj atribut.

$$woe = \ln(i.odds)$$

$$woe = \ln(2,352) = 0,855$$

Ako je woe negativan, to pokazuje da je veća proporcija loših u odnosu na dobre i obrnuto.

i.v.= information value

$$i.v. = \sum_{i=1}^n (distribucija_dobri_i - distribucija_losi_i) \cdot \ln(i.odds)$$

Information odds pokazuje koliko informacija sadrži ta varijabla:

< 0,02	neprediktivno
0,02 do 0,1	slaba prediktivnost
0,1 do 0,3	srednja prediktivnost
>0,3	jako prediktivno

Testiranje hipoteza

Varijabla days je varijabla koja pokazuje koliki je broj dana naplate potraživanja poduzeća, a računa se kao omjer 365/koeficijent obrtaja potraživanja. Što je broj dana veći, to poduzeću treba više vremena da naplati svoja potraživanja.

Elementarna statistika varijable s obzirom na urednost vraćanja kredita (dobar/loš):

```
> describeBy(days, group=default,mat=TRUE)
  item group1 vars  n   mean   sd median trimmed  mad   min   max range  skew  kurtosis   se
X11   1   bad    1  66 57.27059 17.84744 53.7969 55.90927 12.40187 28.6154 120.1110 91.4956 1.033553 1.1187342 2.196868
X12   2   good   1 129 48.99940 14.81685 47.1204 47.70795 17.28623 23.9706  87.6296 63.6590 0.652308 -0.2276232 1.304551
```

Iz tablice se može uočiti da klijentima koji uredno vraćaju kredite u prosjeku treba manje dana (48,99) da naplate svoja potraživanja u odnosu na one koji kasne u plaćanju kredita (57,27).

t-test je pokazao da postoji statistički značajna razlika u očekivanju za trajanje naplate potraživanja između onih koji uredno vraćaju kredit i onih koji su loši (t-test=3.237, p=0.001588)

```
welch Two Sample t-test
```

```
data: days by default
t = 3.2372, df = 111.86, p-value = 0.001588
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.208695 13.333686
sample estimates:
mean in group bad mean in group good
 57.27059          48.99940
```

WOE – weight of evidence

Dodatna analiza karakteristika napravljena je kroz WOE (weight of evidence) tablicu:

Dani naplate potraživanja	Broj dobrih	Broj loših	%dobrih	%loših	i.odds	w.o.e.	i.v.
<=35	33	5	25,58	7,7	3,366	1,214	0,217
>35 <=50	48	22	37,21	33,33	1,116	0,11	0,0043
>50 <=80	43	29	33,33	43,95	0,758	-0,277	0,029
>80	5	10	3,8	15,15	0,257	-1,359	0,1542
	129	66					0,4045

Koliko dobrih dolazi na 1 lošeg?

Na 1 lošeg klijenta koji ima trajanje naplate potraživanja <=35 dana dolazi 3,366 dobrih klijenata s naplatom potraživanja <=35 dana.

Ili

Šanse su 3,366:1 da ćemo pronaći klijenta koji ima naplatu potraživanja <=35 u dobrima u odnosu na to da ćemo pronaći klijenta s naplatom potraživanja <=35 u lošima.

Iz tablice se može uočiti da je veća proporcija 'loših' u odnosu na 'dobre' u skupinama gdje je naplata potraživanja veća dok je situacija obrnuta kako dani naplate potraživanja padaju. Možemo zaključiti da se rizičnost klijenata smanjuje sa smanjenjem dana naplate potraživanja odnosno da su oni koji brže naplaćuju manje rizični. Primjerice, na 1 lošeg klijenta u skupini '<=35' dolazi 3,366 dobrih dok na 1 lošeg klijenta u skupini '>80' dolazi 0,257 dobrih klijenata.

Informacijska vrijednost varijable 'dob' iznosi 0.4045 što pokazuje jaku prediktivnost.

S obzirom na provedenu analizu, možemo zaključiti da trajanje naplate potraživanja ima smisla uključiti u razvoj scoring modela za procjenu kreditnog rizika.

Važno!

Kod woe vrijednosti za numeričke varijable, u pravilu se MORA postići MONOTONOST kako bi se uočilo raste li ili pada rizičnost s porastom numeričke varijable:

Iz gornje tablice se vidi da s porastom dana naplate potraživanja, raste rizičnost – obratite pažnju na woe vrijednosti.

Dakle, numeričke vrijednosti se trebaju rasporediti u kategorije tako da woe monotono pada ili monotono raste.