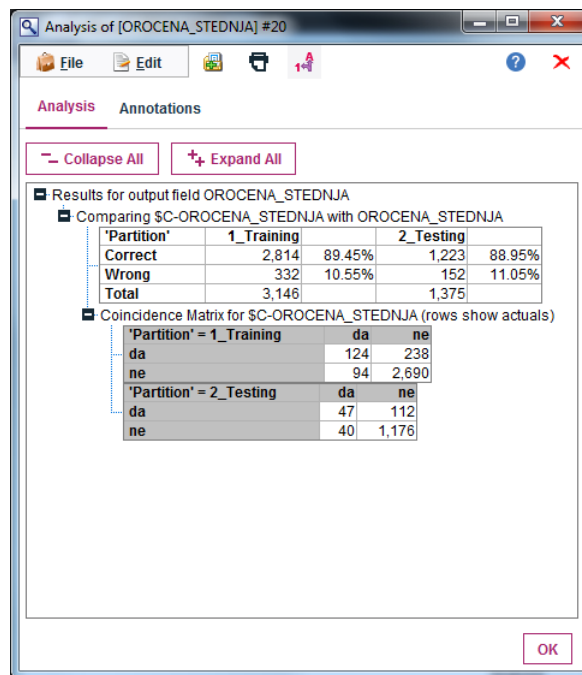


Primer:

- U alatu IBM SPSS Modeler primeniti klasifikaciju nad skupom *bank.csv* korišćenjem C5.0. Ciljni atribut je oročena štednja.
 - Koji atributi su korišćeni pri pravljenju modela?
 - Komentarisati dobijen model.

Komentari

Napravljena su dva modela. Prvo je napravljen model bez matrice cena, dalje u tekstu *model 1*. Promenom vrednosti za parametre dobijaju se slični rezultati. Za model sa izabranim vrednostima (videti radni tok **bank_class.str** kako su definisani parametri) preciznost i matrica konfuzije za test i trening skup je prikazana na slici 1



Analysis of [OROCENA_STEDNJA] #20

Analysis Annotations

[- Collapse All] [+ Expand All]

Results for output field OROCENA_STEDNJA

Comparing \$C-OROCENA_STEDNJA with OROCENA_STEDNJA

'Partition'	1_Training		2_Testing	
Correct	2,814	89.45%	1,223	88.95%
Wrong	332	10.55%	152	11.05%
Total	3,146		1,375	

Coincidence Matrix for \$C-OROCENA_STEDNJA (rows show actuals)

'Partition' = 1_Training		da	ne
da		124	238
ne		94	2,690

'Partition' = 2_Testing		da	ne
da		47	112
ne		40	1,176

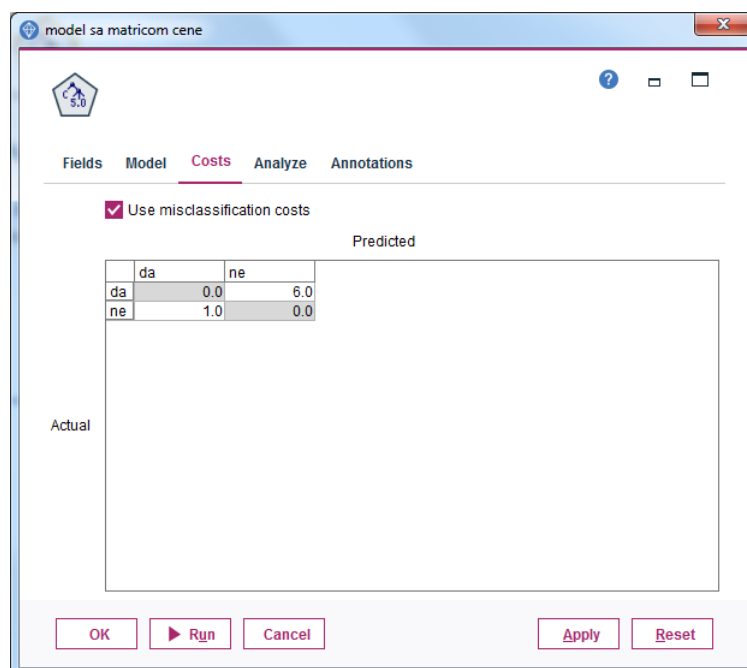
OK

Slika 1: Rezultat primene *modela 1* na trening i test skup

Pošto je ukupna preciznost za trening i test skup slična (89,45% i 88,95%) zaključujemo da nije došlo do preprilagođavanja. Ako pogledamo matricu konfuzije, primećujemo da je predviđanje za klasu *ne* odlično dok to ne važi za klasu *da*. Preciznije u trening skupu 124 instance klase *da* su dobro klasifikovane (34,25% instanci klase *da*), dok je za 238 (65,75% instanci klase *da*) instanci klase *da* dodeljena klasa *ne*. U test skupu je 47 instanci klase *da* dobro klasifikovano (29,56% instanci klase *da*), dok je za 112 (70,44% instanci klase *da*) instanci klase *da* dodeljena klasa *ne*. Iz ugla banke, koja ne želi da ispusti klijenta koji bi bio zainteresovan za njihovu ponudu - oročenje štednje, ovaj model nije dovoljno dobar. Korišćenjem matrice cena ovakva informacija se može uključiti u tok istraživanja i koristiti pri pravljenju modela.

Pri pravljenju drugog modela korišćene su iste vrednosti parametra kao i za prvi model, ali je definisana i matrica cena u kojoj je definisano da je veća greška ukoliko se klijent koji bi bio zainteresovan za oročenje štednje klasifikuje kao klijent koji nije zainteresovan za oročenje štednje, nego ako se klijent koji nije zainteresovan za oročenje štednje klasifikuje

kao klijent koji je zainteresovan za oročenje štednje. Ovaj model je dalje u tekstu označen sa *model 2*. Matrica cena za pravljenje modela 2 je prikazana na slici 2.

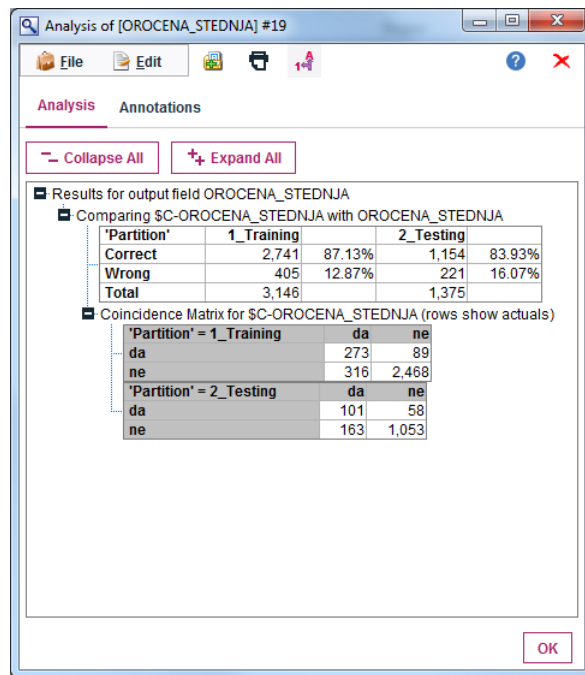


Slika 2: Matrica cena definisana za pravljenje *modela 2*

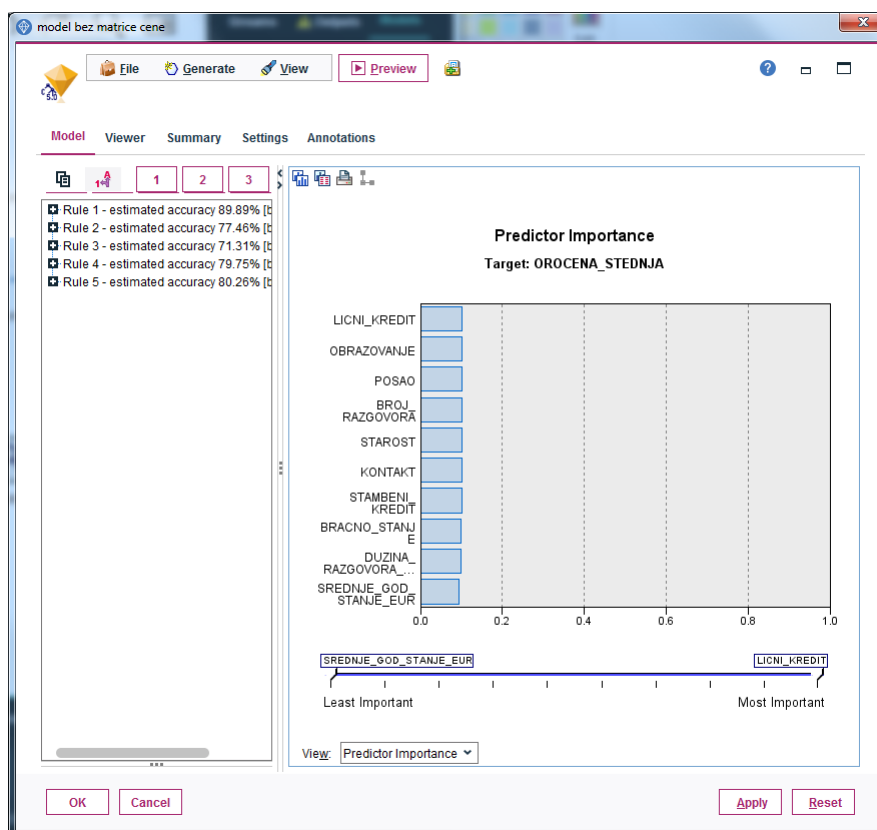
Pošto razlika između ukupne preciznost za trening i test skupu za *model 2* (87,13% i 83,93%) nije velika zaključujemo da nije došlo do prilagođavanja (slika 3). Ako pogledamo matricu kofuzije, primećujemo da je predviđanje za klasu *da* bolje nego za *model 1*. Preciznije, u trening skupu 273 instance klase *da* su dobro klasifikovane (75,41% instanci klase *da*), dok je za 89 (24,59% instanci klase *da*) instanci klase *da* dodeljena klasa *ne*. U test skupu je 101 instanca klase *da* dobro klasifikovana (63,52% instanci klase *da*), dok je za 58 (36,48% instanci klase *da*) instanci klase *da* dodeljena klasa *ne*, što je značajno poboljšanje u odnosu na rezultate za *modela 1*.

Očekivano, *model 2* lošije klasifikuje instance klase *ne* u odnosu na *model 1*. Na osnovu namene modela u praksi - da se za nove (neobrađene) klijente odredi da li su zainteresovani za oročenu štednju, bolje je ne izostaviti potencijalnog klijenta po cenu dodavanja određenog broja nezainteresovanih klijenata.

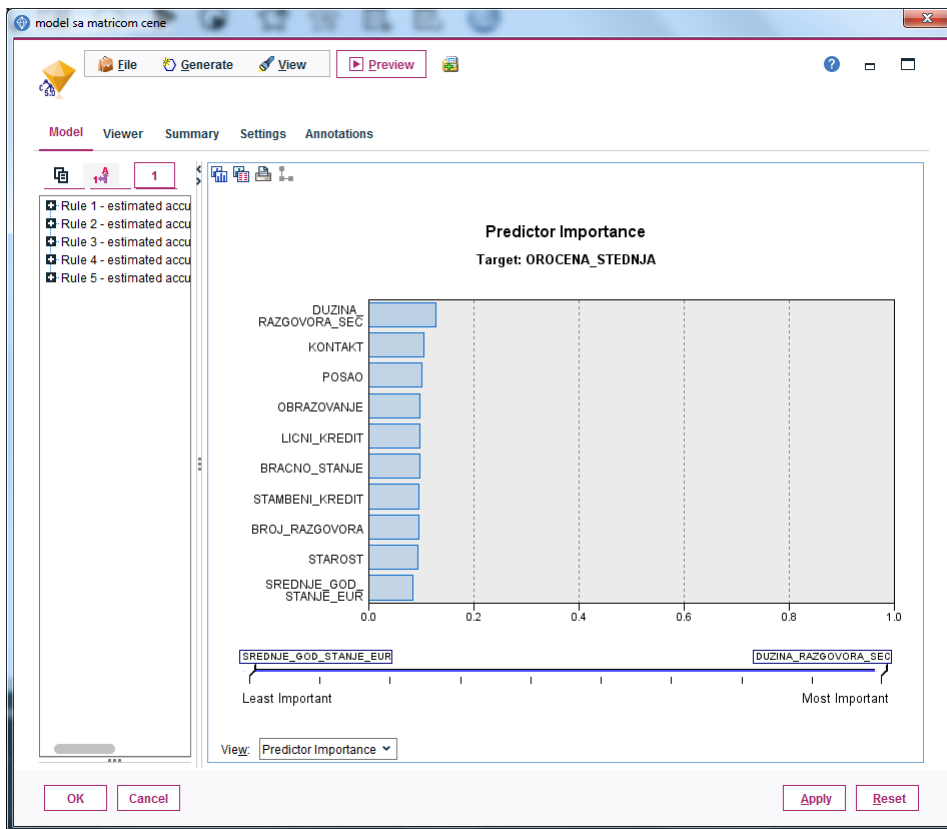
Atributi koji su korišćeni za pravljenje *modela 1* i njihova značajnost su prikazani na slici 4, a atributi koji su korišćeni za pravljenje *modela 2* i njihova značajnost su prikazani na slici 5. Zaključujemo da su svi zadati atributi za predviđanje korišćeni pri pravljenju modela i da im je značajnost skoro ista (u opsegu od 0,9-0,13).



Slika 3: Rezultat primene *modela 2* na trening i test skup



Slika 4: Značajnost atributa u *modelu 1*



Slika 5: Značajnost atributa u modelu 2