



### **3. cjelina: Upotreba i primjena kredit scoring modela**

# Principi izgradnje (osnovni koraci)

- 1. Studija provedivosti
- 2. Definicija uzorka
- 3. Prikupljanje podataka
- 4. Analiza karakteristika
- 5. Zaključivanje o odbijenima
- 6. Modeliranje skor-kartice
- 7. Validacija skor-kartice
- 8. Postavljanje strategije i implementacija

# 7. Validacija scoring modela

- validirati zapravo znači testirati scoring model primjenom različitih kvalitativnih i kvantitativnih testova
- kvantitativni testovi na out-of-sample i out-of-time
- termin ‘validation’ dolazi iz IRB minimalnih zahtjeva:
- *“The institution shall have a regular cycle of model validation that includes monitoring of model performance and stability; review of model relationship; and testing of model outputs against outcomes.”*

# 7. Validacija - zajednički zahtjevi (1/2)

1. Ciljana vrijednost je PD
  - rezultat modela mora biti prezentiran kao PD
2. Potpunost
  - prilikom modeliranja odnosno ocjenjivanja kreditne sposobnosti treba uzeti u obzir sve dostupne podatke
3. Objektivnost
  - isti rezultat modela mora se reproducirati od strane različitih subjekata

# 7. Validacija - zajednički zahtjevi (1/2)

## 4. Prihvaćenost

- model mora biti percipiran, od strane korisnika, kao onaj koji točno procjenjuje kreditnu sposobnost njihovih klijenata

## 5. Konzistentnost

- rating/scoring model ne smije biti u kontradikciji s prihvaćenim teorijama i metodama

# 7. Validacija – opći zahtjevi za pojedinu metodu

- svaka metoda ima svoje zahtjeve
- generalno: modeli dobiveni statističkim metodama su, u smislu rezultata koje daju, najbolji pa se uvijek, kada se to može, ide na primjenu tih metoda u razvijanju scoring/rating modela
- više o tome u: Guidelines on Credit Risk Management – Rating Models and Validation; ONB and FMA; 2004

# 7. Validacija – kvalitativna validacija

- dizajn modela
  - treba pregledati kompletnu dokumentaciju vezanu uz metodu
- kvaliteta podataka
  - potrebni podaci, izvor podataka, reprezentativnost uzorka podataka, čišćenje, definicija uzorka
- upotreba modela
  - kako je scoring/rating model integriran u cijeli risk management i reporting sistem banke

# 7. Validacija – kvantitativna validacija

- preciznost klasifikacije modela
    - kolika je sposobnost modela da razlikuje dobre i loše
  - postupak pregledavanja kalibracije modela
    - koliko se default rate na master skali razlikuje od realizirane default rate
  - ispitivanje stabilnosti modela
  - benchmarking
  - stress testing
- 
- više o tome u: Guidelines on Credit Risk Management – Rating Models and Validation; ONB and FMA; 2004.



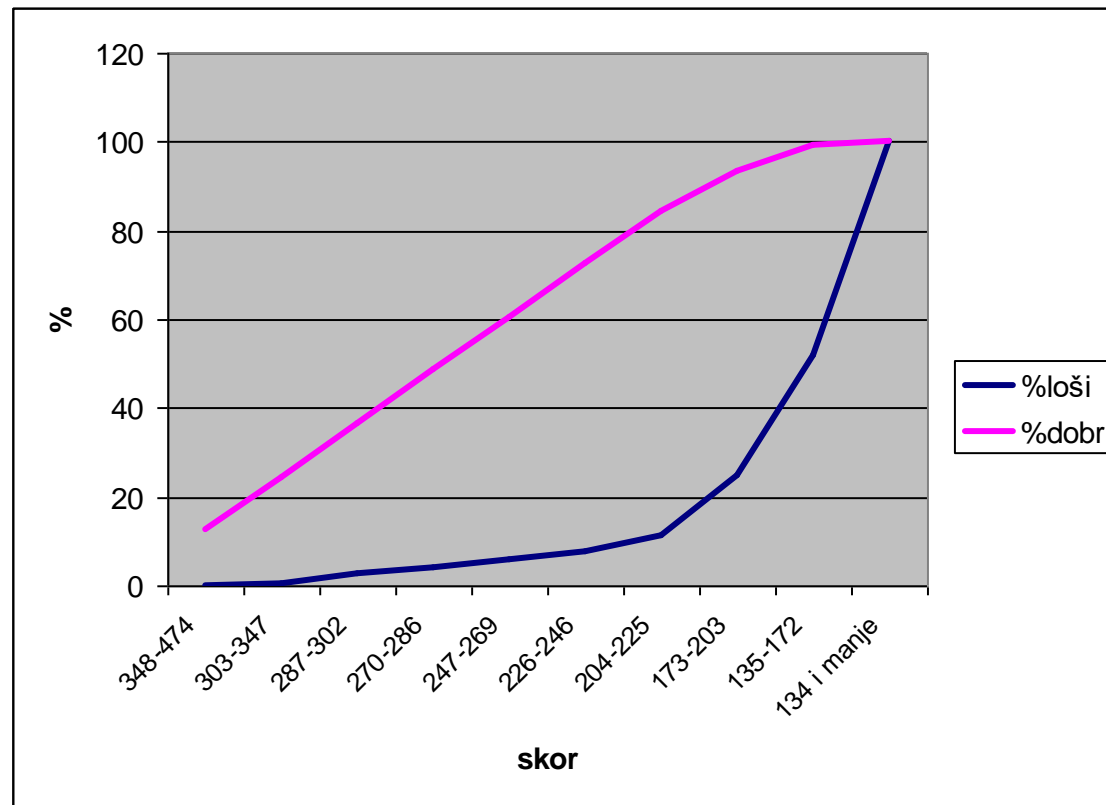
# 7. Kvantitativna validacija – preciznost klasifikacije modela

- distribucija frekvencija dobrih i loših s obzirom na score/rating klasu (tablično, grafički)
- postotak uspješne klasifikacije (ukupno, dobri, loši)
- greška tipa I i greška tipa II
- confusion matrix
- ROC (receiver operating characteristic) krivulja
- AUC
- KS statistika
- GINI koeficijent

# - Distribucija dobrih i loših

<b>Skor vrijednosti</b>	<b>Postotak loših</b>	<b>Postotak loših (kumulativno)</b>	<b>Postotak dobrih</b>	<b>Postotak dobrih (kumulativno)</b>
<b>348-474</b>	<b>0</b>	<b>0</b>	<b>12.767</b>	<b>12.767</b>
<b>303-347</b>	<b>0.675</b>	<b>0.675</b>	<b>11.756</b>	<b>24.523</b>
<b>287-302</b>	<b>1.913</b>	<b>2.588</b>	<b>12.033</b>	<b>36.556</b>
<b>270-286</b>	<b>1.262</b>	<b>3.850</b>	<b>12.058</b>	<b>48.614</b>
<b>247-269</b>	<b>1.892</b>	<b>5.742</b>	<b>12.059</b>	<b>60.673</b>
<b>226-246</b>	<b>1.893</b>	<b>7.635</b>	<b>12.059</b>	<b>72.732</b>
<b>204-225</b>	<b>3.445</b>	<b>11.080</b>	<b>11.547</b>	<b>84.279</b>
<b>173-203</b>	<b>13.955</b>	<b>25.035</b>	<b>8.940</b>	<b>93.219</b>
<b>135-172</b>	<b>26.862</b>	<b>51.897</b>	<b>6.008</b>	<b>99.227</b>
<b>134 i manje</b>	<b>48.103</b>	<b>100</b>	<b>0.773</b>	<b>100</b>
<b>ukupno</b>	<b>100</b>		<b>100</b>	

# -Grafički prikaz distribucije dobrih i loših



# - Postotak uspješne klasifikacije

- **total hit rate**
  - omjer ukupnog broja klijenata ispravno klasificiranih (prema modelu) i ukupnog broja klijenata
- **good hit rate**
  - omjer dobrih klijenata ispravno klasificiranih (prema modelu) i stvarno dobrih klijenata
- **bad hit rate**
  - omjer loših klijenata ispravno klasificiranih (prema modelu) i stvarno loših klijenata

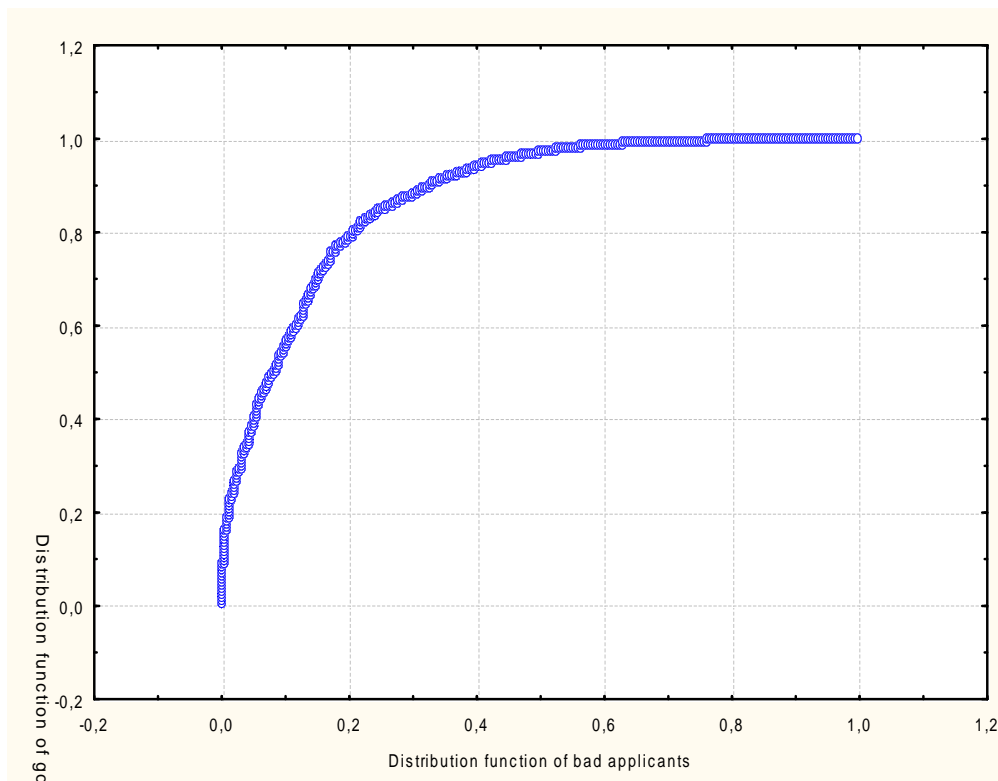
## - Greška tipa I i II

- greška tipa I ( $\alpha$ )
  - odobravanje kredita lošem klijentu
  - $\alpha = c/(a+c)$  (iz confusion matrix)
- greška tipa II ( $\beta$ )
  - neodobravanje kredita dobrom klijentu
  - $\beta = b/(b+d)$  (iz confusion matrix)

## - Confusion matrix

Actual	Predicted	
	Bad	Good
Bad	a (true negative)	c (false positive)
Good	b (false negative)	d (true positive)

# - ROC krivulja



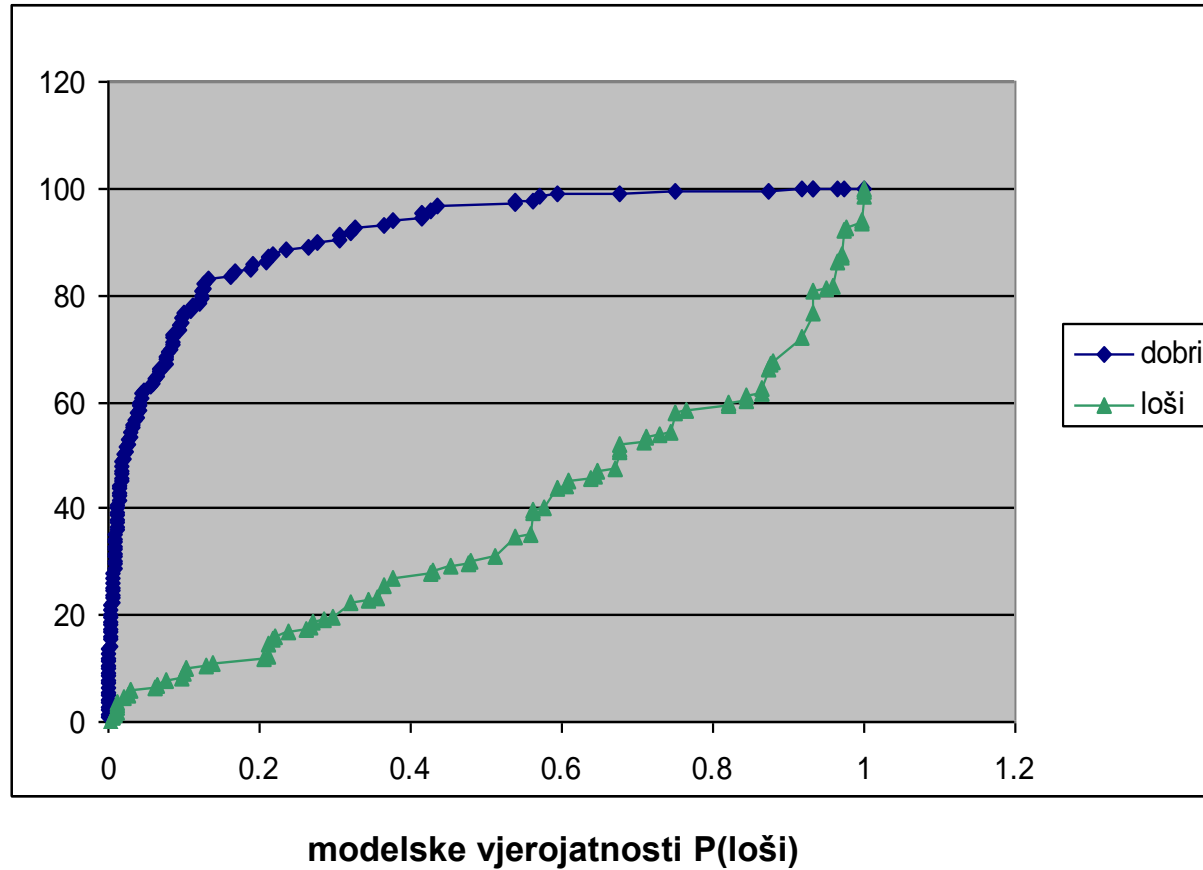
# - KS statistika - izračun

Score	Loši		Kumulativ loših %	Dobri		Kumulativ dobrih %	Razlika   KS stat.
	#	%		#	%		
< 190	20	0.35	0.35	283	0.12	0.12	0.23
190-194	94	1.64	1.99	2195	0.93	1.05	0.94
195-199	143	2.49	4.47	2578	1.09	2.15	2.32
200-209	935	16.30	20.78	16577	7.04	9.19	11.59
210-219	1104	19.25	40.03	21074	8.95	18.13	21.90
220-229	909	15.85	55.88	23128	9.82	27.96	27.92
230-239	782	13.64	69.52	23625	10.03	37.99	<b>31.53</b>
240-249	565	9.85	79.38	25741	10.93	48.92	30.46
250-259	467	8.14	87.52	27940	11.86	60.78	26.74
260-269	285	4.97	92.48	28147	11.95	72.73	19.75
270-279	191	3.33	95.82	25070	10.65	83.38	12.44
280-289	93	1.62	97.44	17526	7.44	90.82	6.62
290-299	57	0.99	98.43	11405	4.84	95.67	2.76
300>	90	1.57	100.00	10208	4.33	100.00	0.00
Ukupno	5735	100.00		235497	100.00		



# KS statistika - grafički

kumulativne rel.frekv.poznatih i procijenjenih komitenata



## - KS statistika

- $KS < 20$ : kartica vjerojatno neće raditi
- 20 – 40: fair (osrednje?)
- 41 – 50: good
- 51 – 60: very good
- 61 – 75: awesome
- $KS > 75$ : too good to be true

## - GINI koeficijent

- $2 * AUC - 1$
- statistika koja mjeri sposobnost scoring modela da rangira rizik
- 0% model ne razlikuje dobre od loših
- 100% savršeno ih razlikuje
- prosječan scoring model ima vrijednost GINI između 40 i 60%

# Zadatak

- Za sljedeće podatke napravite:
- distribuciju frekvencija dobrih i loših (tablicu i graf)
- izračunajte postotak uspješne klasifikacije
- confusion matrix
- grešku tipa I i II
- nacrtajte ROC krivulju
- izračunajte AUC
- izračunajte Gini
- izračunajte KS statistiku

100	B	230	G
110	B	240	G
120	G	250	B
130	B	260	G
140	B	270	G
150	G	280	G
160	B	290	B
170	G	300	G
180	G	310	B
190	B	320	G
200	G	330	G
210	G	340	G
220	B		