

UNIVERZA V MARIBORU
FAKULTETA ZA NARAVOSLOVJE IN MATEMATIKO
Oddelek za matematiko in računalništvo

MAGISTRSKO DELO

Jasmina Kolar

Maribor, 2014

UNIVERZA V MARIBORU
FAKULTETA ZA NARAVOSLOVJE IN MATEMATIKO
Oddelek za matematiko in računalništvo

Magistrsko delo

IRB BONITETNI MODEL ZA SLOVENSKO PREBIVALSTVO

na študijskem programu 2. stopnje Matematika

Mentor:

Drago Bokal

Kandidatka:

Jasmina Kolar

Maribor, 2014

ZAHVALA

Šele ko zmoremo, česar nismo pričakovali,
odkrijemo sebe in premagamo dvom, ki je namišljeno strašilo.
Odpre se tisoč novih poti in samo od nas je odvisno,
katero si bomo izbrali.

Zahvaljujem se mentorju Dragu Bokalu za njegov trud in pomoč ter strokovno vodenje pri nastajanju magistrskega dela. Prav tako bi se rada zahvalila vsem delavcem banke X, ki so mi na kakršenkoli način pomagali.

Posebna zahvala velja tudi staršem in fantu Borisu, ki so mi v času študija stali ob strani in me podpirali. Hvala tudi vsem ostalim, ki verjamejo vame!

Vsem iskreno hvala!

UNIVERZA V MARIBORU
FAKULTETA ZA NARAVOSLOVJE IN MATEMATIKO

IZJAVA

Podpisana Jasmina Kolar, rojena 05. novembra 1989, študentka Fakultete za naravoslovje in matematiko Univerze v Mariboru, študijskega programa 2. stopnje Matematika, izjavljam, da je magistrsko delo z naslovom

IRB BONITETNI MODEL ZA SLOVENSKO PREBIVALSTVO

pri mentorju Dragu Bokalu avtorsko delo. V magistrskem delu so uporabljeni viri in literatura korektno navedeni; teksti niso uporabljeni brez navedbe avtorjev.

Maribor, 25. oktober 2014

Jasmina Kolar

IRB bonitetni model za slovensko prebivalstvo

program magistrskega dela

Namen magistrskega dela je seznanitev z IRB pristopom ter razvoj poskusnega modela za ocenjevanje kreditnega tveganja, prilagojenega slovenskemu prebivalstvu. Navedeni pristop omogoča podrobnejše razvrščanje komitentov, kar zagotavlja učinkovitejšo pomoč pri spremeljanju celotnega portfelja. Cilj magistrskega dela je ugotoviti, katere spremenljivke oziroma kazalniki v največji meri vplivajo na višino verjetnosti neplačila, ter pridobiti praktične izkušnje pri urejanju in pripravi baze podatkov za namen statističnega modeliranja.

Osnovni viri:

1. J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann Publishers, San Francisco, 2006.
2. J. N. Crook, D. B. Edelman, L. C. Thomas, *Credit Scoring and Its Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 2002.
3. B. Baesens, T. Van Gestel, *Credit Risk Management: Basic Concepts*, Oxford University Press, New York, 2009.
4. M. Šušterič, E. Zavodnik, Priprava slovenskih bank na spremembe kapitalskega sporazuma Basel II, 2003.

Drago Bokal

KOLAR, J.: IRB bonitetni model za slovensko prebivalstvo.

Magistrsko delo, Univerza v Mariboru, Fakulteta za naravoslovje in matematiko, Oddelek za matematiko in računalništvo, 2014.

IZVLEČEK

V Sloveniji je kredit najbolj razširjen finančni produkt kreditiranja prebivalstva. Ker pa je banka ob izdaji kredita izpostavljena določenemu tveganju finančne izgube, je pomembno, da ima zgrajen učinkovit sistem upravljanja s kreditnim tveganjem. Učinkovito pomoč pri spremljanju kreditov in celotnega portfelja omogoča IRB pristop.

V magistrskem delu z uporabo IRB pristopa razvijemo nov bonitetni model za ocenjevanje kreditnega tveganja, prilagojenega slovenskemu prebivalstvu. Pri tem uporabimo zgodovinske podatke banke X, na podlagi katerih razvijemo model, ki napoveduje, ali je nek komitent zmožen odplačati kredit (ostane plačnik) ali ne (postane neplačnik).

Ker gre za delo iz finančno-matematičnega področja, je razdeljeno na tri dele. V prvem delu je na kratko predstavljen finančni oz. bančni vidik dela. Drugi del predstavlja matematični vidik ter jedro magistrskega dela, saj v njem razvijem bonitetni model za slovensko prebivalstvo. V tretjem delu je predstavljen konkreten primer uporabe razvitega bonitetnega modela.

Ključne besede: IRB, kreditno tveganje, bonitetni model, modeliranje, ROC analiza, odločitveno drevo, binarna logistična regresija.

Math. Subj. Class. (2010): 62-07 podatkovna analiza,
91B06 teorija odločanja,
62M10 časovne vrste, avtokorelacija, regresija, itd.,
62J02 splošna nelinearna egresija.

KOLAR, J.: IRB credit rating model for the Slovenian population.

Master Thesis, University of Maribor, Faculty of Natural Sciences and Mathematics, Department of Mathematics and Computer Science, 2014.

ABSTRACT

Credit is the most widely used financial product of retail lending in Slovenia. Since the bank is exposed to a certain risk of financial loss, it is important to have an efficient system of credit risk management. The IRB approach is one such tool recommended by Basel regulations, which provides effective assistance in monitoring of loans and the entire bank credit portfolio.

In this thesis, we develop a new credit rating model for evaluating credit risk that is adjusted to the Slovenian population by using the IRB approach. By using historical data of the Bank X, we develop a model that predicts whether a customer is able to repay the loan (he remains non-defaulted) or is unable to repay the loan (he defaults).

Since the thesis refers to a financial-mathematical field, it is divided into three parts. The first part is a brief overview of the financial ie. banking aspect of the thesis. The second part presents the mathematical aspect and the core of the master thesis, because it is the part where we develop the credit rating model for the Slovenian population. Finally, the third part presents a concrete example of the application of the developed credit rating model.

Keywords: IRB, credit risk, credit rating, modeling, ROC analysis, decision tree, binary logistic regression.

Math. Subj. Class. (2010): 62-07 data analysis,
91B06 decision theory,
62M10 time series, auto-correlation, regression, etc.,
62J02 general nonlinear regression.

Kazalo

Seznam uporabljenih kratic in simbolov	xii
Uvod	1
1 Kreditiranje prebivalstva	3
1.1 Kreditni proces	4
1.1.1 Proces odobritve kredita	4
1.1.2 Proces spremljave kredita	4
1.1.3 Proces zgodnjega odkrivanja kreditnega tveganja	5
1.1.4 Proces razvrščanja dolžnika in/ali izpostavljenosti	5
1.2 Tveganja pri kreditiranju prebivalstva	6
1.2.1 Načela upravljanja s kreditnim tveganjem	7
1.2.2 Modeli ocenjevanja kreditnega tveganja	8
2 IRB pristop	10
2.1 Razvoj	10
2.2 Lastnosti in prednosti	13
2.3 Minimalne zahteve za uvedbo in nadaljnjo uporabo IRB pristopa	13
2.3.1 Zasnova in postavitev bonitetnega sistema	14
2.3.2 Delovanje bonitetnega sistema	15
2.3.3 Upravljanje bank in nadzorstvo	16
2.3.4 Uporaba internih bonitetnih ocen	16

2.4	Ključni elementi	17
2.4.1	Razvrstitev izpostavljenosti	17
2.4.2	Parametri tveganja	18
2.4.3	Funkcija tveganja	26
2.4.4	Nadzorniški pregled skladnosti z minimalnimi zahtevami	28
3	Matematične osnove	29
3.1	ROC analiza	31
3.1.1	ROC krivulja	33
3.1.2	Območje pod ROC krivuljo	37
3.1.3	Gini koeficient	39
3.1.4	Večrazredni problemi	40
3.2	Binarna logistična regresija	41
3.2.1	Regresijska funkcija in model	42
3.2.2	Določitev optimalne mejne vrednosti	45
3.3	Klasifikacijska drevesa	46
3.3.1	Razvrščanje vzorcev	47
3.3.2	Entropija	48
3.3.3	Informacijski prispevek	51
3.4	Korelacijska analiza	53
3.4.1	Spearmanov koeficient korelacije	54
4	Bonitetni model za slovensko prebivalstvo	55
4.1	Tehnologije izdelave modela	56
4.1.1	SPSS	56
4.1.2	Rapid Miner	57
4.2	Definicija dogodka neplačila	58
4.3	Portfelj oziroma populacija	58

4.3.1	Finančni kazalniki	60
4.4	Razvoj modela	62
4.5	Bonitetni model	66
4.6	Kalibracija modela	69
4.7	Kapitalska zahteva	69
4.8	Pomanjkljivosti	70
5	Konkreten primer: Občina Makole	72
5.1	Zakaj Makole potrebujejo kredit?	72
5.2	Kam sodijo občani?	73
6	Zaključek	75
	Literatura	77
	Priloge	80

Seznam uporabljenih kratic in simbolov

<i>BDP</i>	bruto domači proizvod
<i>CCF</i>	konverzijski faktor (<i>ang. credit conversion factor</i>)
<i>CF</i>	denarni tok (<i>ang. cash flow</i>)
<i>CL</i>	kreditni limit (<i>ang. credit limit</i>)
<i>DA</i>	črpani znesek (<i>ang. drawn amount</i>)
<i>EAD</i>	izpostavljenost v primeru neplačila (<i>ang. exposure at default</i>)
<i>EL</i>	pričakovana izguba (<i>ang. expected loss</i>)
<i>EL_a</i>	vrednost pričakovane izgube (<i>ang. expected loss amount</i>)
<i>EL_{BE}</i>	najboljša ocena povprečne izgube (<i>ang. expected loss best</i>)
<i>FN</i>	lažno negativna vrednost (<i>ang. false negative</i>)
<i>FP</i>	lažno pozitivna vrednost (<i>ang. false positive</i>)
<i>G</i>	Ginijev koeficient
IRB pristop	pristop na podlagi notranjih bonitetnih sistemov
<i>K</i>	kapitalska zahteva
<i>LGD</i>	izguba v primeru neplačila (<i>ang. loss given default</i>)
<i>M</i>	zapadlost (<i>ang. maturity</i>)
<i>N</i>	množica naravnih števil
<i>PD</i>	verjetnost neplačila (<i>ang. probability of default</i>)
<i>RR</i>	delež poplačila/izterjatve (<i>ang. recovery rate</i>)
<i>RW</i>	utež tveganja (<i>ang. risk weight</i>)
<i>RWA</i>	tveganju prilagojena sredstva (<i>ang. risk weighted assets</i>)
<i>TN</i>	pravilno negativna vrednost (<i>ang. true negative</i>)
<i>TP</i>	pravilno pozitivna vrednost (<i>ang. true positive</i>)
<i>TV</i>	testni vzorec
<i>UA</i>	nečrpani znesek (<i>ang. undrawn amount</i>)
ρ	korelacijski koeficient
\S	poglavje

Uvod

V Sloveniji je kredit najbolj razširjen finančni produkt kreditiranja prebivalstva. Posledično bančni kredit posamezniku predstavlja osnovni produkt, s katerim se primerjajo ostale oblike kreditiranja. Ker je banka ob izdaji kredita izpostavljena določenemu tveganju finančne izgube, je pomembno, da ima zgrajen učinkovit sistem upravljanja s kreditnim tveganjem. Zaradi negotovosti v ekonomskem okolju je ta naloga postala še zahtevnejša. Tako je finančna kriza, kateri še ni videti konca, poslovanje bank postavila pred veliko preizkušnjo.

Nov kapitalski sporazum, imenovan Basel II, je na področju kreditnega tveganja prinesel povsem nove možnosti za učinkovitejše upravljanje z njim. Po Baslu II mora ocena kreditnega tveganja temeljiti na podlagi portfeljskega pristopa in je lahko ocenjena z uporabo standardiziranega ali IRB (*ang. Internal Rate-Based*) pristopa. Ker IRB pristop omogoča podrobnejše razvrščanje komitentov, kar zagotavlja učinkovitejšo pomoč pri spremeljanju kreditov, trendov in celotnega portfelja ter ima poleg tega tudi številne druge prednosti, banke težijo k uporabi slednjega.

Namen magistrskega dela je seznanitev z IRB pristopom ter razviti poskusni model za ocenjevanje kreditnega tveganja, prilagojenega slovenskemu prebivalstvu. Pri tem bomo uporabili zgodovinske podatke ene izmed poslovnih bank, na podlagi katerih bomo razvili model, ki bo napovedal, s kakšno verjetnostjo bo nek komitent zmožen odplačati kredit (bo ostal plačnik) ali ne (bo postal neplačnik).

Cilj magistrskega dela je ugotoviti, katere spremenljivke oziroma kazalniki v največji meri vplivajo na višino verjetnosti neplačila, ter pridobiti praktične izkušnje pri urejanju in pripravi baze podatkov za namen statističnega modeliranja.

Ker gre za delo iz finačno-matematičnega področja, je razdeljeno na tri dele. V prvem delu je na kratko predstavljen finančni oziroma bančni vidik dela. Tako prvo poglavje na kratko opiše, kaj je kredit, kako poteka sam proces kreditiranja ter kakšna tveganja se pri tem pojavljajo. Pri tem le na kratko opišemo modele ocenjevanja kreditnega tveganja, saj bi obravnava vsakega od teh v podrobnosti presegala namen naloge; v nalogi namreč razvijemo lasten model. V drugem poglavju se podrobneje posvetimo IRB pristopu, ki predstavlja

osnovo nadaljnega dela. Tako izvemo, kako je potekal sam razvoj tega pristopa ter kakšni so njegovi ključni elementi, lastnosti in prednosti. Poleg natančnosti napovedovanja je zelo pomembno, da je model transparenten ter logičen in posledično tudi "lahko" preverljiv.

Drugi del magistrskega dela predstavlja matematični vidik ter jedro magistrskega dela. V četrtem poglavju tako spoznamo statistične metode analize podatkov, kot so ROC analiza, binarna logistična regresija in odločitvena drevesa. Te metode so uporabljeni v poglavju, ki sledi, kjer na podlagi teoretičnih izhodišč, matematičnega znanja in zgodovinskih podatkov banke X razvijemo nov bonitetni model za slovensko prebivalstvo, ki temelji na statističnem ocenjevanju verjetnosti neplačila. Da lahko te verjetnosti ustrezno ocenimo, moramo imeti postavljene jasne kriterije za določitev statusa neplačnika.

V tretjem delu je predstavljen konkretni primer uporabe razvitega bonitetnega modela. V zadnjem poglavju tako ugotovimo, v kateri bonitetni razred zgoraj razvitega modela bi se uvrstili občani občine Makole. Na podlagi pridobljenih rezultatov lahko banka podrobnejše oceni, kakšnemu tveganju bi se izpostavila, če bi tem komitentom odobrila posojilo.

Pri razvoju modela si pomagamo s statističnim programskim paketom SPSS ter s programom za podatkovno rudarjenje imenovanim Rapid Miner. Ker pri razvoju modela uporabljamo zaupne podatke ene izmed slovenskih poslovnih bank, je prikaz rezultatov v nalogi nekoliko omejen in prilagojen varovanju zaupnih podatkov.

Poglavlje 1

Kreditiranje prebivalstva

Zaradi načina življenja in družbenih ter gospodarskih razmer v katerih živimo, potrebujemo denar, če želimo imeti nek življenjski standard. Velikokrat se zgodi, da za ohranitev našega standarda nimamo na voljo dovolj denarja. Če želimo pridobljeni standard ohraniti, si moramo denar nekje izposoditi. Najpogostejša in najzanesljivejša oblika izposoje denarja je kredit. Za kreditiranje prebivalstva je značilno, da je posamezen kredit, v primerjavi z ostalimi naložbami banke, enemu komitentu relativno majhen. Ne glede na to pa vsi krediti prebivalstvu pomenijo pomemben del aktivne strani bilance bank.

Pri aktivnih poslih oziroma v našem primeru pri kreditiranju prebivalstva, banka nastopa kot upnik (*ang. creditor*) oziroma kreditodajalec, posameznik, ki si denar izposodi, pa kot dolžnik (*ang. debtor*) oziroma kreditojemalec. Običajno imajo kreditojemalci, ki jim banka odobri kredit, enake obrestne mere. To pomeni, da enaka obrestna mera določa ceno kredita pri premožnem individualnem kreditojemalcu, ki si denar izposodi na primer za nakup nepremičnine višjega cenovnega razreda, kakor tudi pri manj premožnem individualnemu kreditojemalcu, ki denar potrebuje za nakup na primer majhne nepremičnine najnižjega cenovnega razreda. Tako se v navedenem primeru kreditojemalca razlikujeta le v višini kredita, ki ga lahko najameta ter posledično tudi v obrokih odplačevanja oziroma v višini mesečne obveznosti, ki sta jo sposobna odplačevati. Pri kreditiranju fizičnih oseb lahko tako rečemo, da se posamezniki razlikujejo na podlagi količinskih omejitev in ne na podlagi razlik v obrestnih merah.

Storitve, s pomočjo katerih banke plasirajo svoj denar, imenujemo kreditne storitve. Tako v bančni poslovni praksi kakor tudi v strokovni literaturi poznamo različne vrste kreditnih storitev, ki temeljijo na v naprej opredeljenih potrebah kreditojemalcev po finančnih sredstvih na eni strani ter na potrebah bank glede realnega zavarovanja kreditov na drugi strani. Ker banke nenehno iščejo nove poti za pridobitev ter ohranitev komitentov, se oblike kreditov v

bankah vedno znova spreminjajo ter izpopolnjujejo. Tako poslovne banke ponujajo različne kredite, ki se razlikujejo po tem, komu so namenjeni ter v kakšni obliki so dani. Prebivalstvo oziroma fizične osebe izposojena sredstva najpogosteje porabijo za osebno potrošnjo ali investicije v nepremičnine in premičnine.

1.1 Kreditni proces

Zakon o bančništvu določa, da mora vsaka banka vzpostaviti kreditni proces, ki je ustrezen le, če vključuje naslednje postopke:

- proces odobritve kredita,
- proces spremljave kredita,
- proces zgodnjega odkrivanja povečanega kreditnega tveganja,
- proces razvrščanja dolžnika in/ali izpostavljenosti.

1.1.1 Proces odobritve kredita

Pred odobritvijo kredita mora vsaka banka oceniti in analizirati vse pomembne dejavnike, ki vplivajo na oceno tveganja dolžnika in/ali izpostavljenosti. Pri tem mora metodologija dodelitve ocene tveganja dolžnika in/ali izpostavljenosti upoštevati zahtevnost in vrednost posameznega kredita. Prav tako morajo banke kot primaren vir poplačila kredita upoštevati plačilno sposobnost kreditojemalca.

Sprejeta zavarovanja za posamezen kredit predstavlja sekundarni vir poplačila kredita. Pred odobritvijo kredita morajo banke oceniti vrednost in pravno veljavnost zavarovanja. V kolikor je vrednost zavarovanja v veliki meri odvisna od finančnega stanja dajalca osebnega zavarovanja (ki je tretja oseba), morajo banke oceniti tudi tveganje te osebe. Tako morajo banke zagotoviti ustrezno politiko sprejemljivih vrst zavarovanj in določiti metodologijo ocenjevanja vrednosti zavarovanj.

1.1.2 Proces spremljave kredita

Proces spremljave kredita vključuje spremljavo posameznega kredita, dolžnika in njegovih zavarovanj. Tako morajo banke po odobritvi kredita zagotoviti redno spremljanje izpolnjevanja pogojev, ki izhajajo iz kreditne pogodbe. Kadar gre za namenski kredit, morajo banke zagotoviti tudi spremljanje porabe odobrenih sredstev v dogovorjene namene.

Banke morajo skozi celotno obdobje trajanja kredita spremljati poslovanje dolžnika ter pri tem upoštevati tveganost zadevnega dolžnika in/ali izpostavljenosti. Prav tako morajo zagotoviti spremljavo in ocenjevanje pravne veljavnosti ter vrednosti zavarovanja v ustreznih časovnih razmikih, ki so odvisni od vrste zavarovanja. V kolikor banke pridobijo notranje ali zunanje informacije, ki nakazujejo na znatno povečanje tveganja dolžnika in/ali izpostavljenosti ter zavarovanja, jih morajo nemudoma ponovno oceniti. Vse enote, ki so vključene v proces upravljanja kreditnega tveganja, morajo biti o zadevnih informacijah nemudoma obveščene.

Zelo pomemben del spremljave kredita predstavlja analiza kreditnega portfelja. Skupaj z ocenami prihodnjih trendov morajo banke redno izvajati analize kreditnega portfelja. Rezultati teh analiz se morajo upoštevati pri oblikovanju politik ter strategij prevzemanja in upravljanja s kreditnim tveganjem. Prav tako je potrebno ugotavljati njihovo ustreznost.

Vsaka banka mora opredeliti tudi merila za obravnavo dolžnika in/ali izpostavljenosti, ki zahteva podrobnejšo spremljavo. Takšnim kreditom pravimo problematični krediti. Zaradi določitve načina njihove nadaljnje obravnave morajo banke takšne izpostavljenosti in/ali dolžnike redno pregledovati. Prav tako morajo banke opredeliti merila za predajo dolžnika in/ali izpostavljenosti enoti za izterjavo problematičnih kreditov. Če so ta merila izpolnjena, mora banka oceniti, ali je prestrukturiranje izpostavljenosti dolžnika smiselno ali ne. V primeru, ko je prestrukturiranje izpostavljenosti zadevnega dolžnika smiselno, mora banka oblikovati ustrezen načrt prestrukturiranja in spremljati njegovo izvajanje ter učinke. Če je dolžnik v stečajnem postopku, in če je izpostavljenost zavarovana, mora banka v proces unovčenja zavarovanja vključiti ustrezne službe in/ali zunanje izvajalce.

1.1.3 Proces zgodnjega odkrivanja kreditnega tveganja

Proces zgodnjega odkrivanja povečanega kreditnega tveganja omogoča pravočasno ugotavljanje dolžnikov, ki izkazujejo povečano tveganje. Pri tem morajo banke določiti ustrezne kvalitativne in kvantitativne kazalnike zgodnjega ugotavljanja povečanega kreditnega tveganja. Kjer je primerno, lahko banke določene kredite iz tega procesa tudi izvzamejo.

1.1.4 Proces razvrščanja dolžnika in/ali izpostavljenosti

Za namen ocenjevanja kreditnega tveganja mora vsaka banka vzpostaviti ustrezen proces razvrščanja dolžnikov in/ali izpostavljenosti v bonitetne razrede. Ta proces je ustrezen le, če temelji na kvantitativnih in, kjer je mogoče, kvalitativnih merilih ter upošteva bistvene značilnosti posameznega dolžnika in/ali izpostavljenosti. Navedena merila morajo zagotavljati

jasno razvrščanje tveganj v ustrezne bonitetne razrede. Podobno kot pri ostalih procesih morajo banke zagotoviti redno spremjanje in ugotavljanje primernosti procesa razvrščanja ter v skladu z mednarodnimi standardi računovodskega poročanja oblikovati rezervacije oz. oslabitve.

V nalogi se bomo podrobneje posvetili ravno temu procesu, ki zajema razvoj ustreznega in kvalitetnega bonitetnega modela za razvrščanje fizičnih oseb v bonitetne razrede.

1.2 Tveganja pri kreditiranju prebivalstva

Upravljanje tveganj je za banke ključnega pomena, saj se banka srečuje s tveganji praktično pri vsakem poslu. Tveganje je povezano z negotovostjo, ki se odraža v nepričakovanih spremembah dogodkov. Za banke kot kreditodajalce negotovost izhaja iz toka depozitov, nepričakovanih sprememb obrestnih mer in sposobnosti kreditojemalca, da lahko vrne kredit.

Obstajajo različni možni načini delitev bančnih tveganj ter različni pristopi upravljanja z njimi. Eden izmed načinov delitve bančnih tveganj je delitev na:

- kreditno tveganje,
- obrestno tveganje,
- likvidnostno tveganje,
- tržno tveganje,
- operativno tveganje in
- ostala tveganja.

Obravnava vsakega od teh tveganj v podrobnosti bi presegala namen naloge, zato se bomo podrobneje posvetili kreditnemu tveganju in njegovemu upravljanju.

Kreditno tveganje, ki je poznano tudi pod izrazom riziko izpada terjatev¹, nedvomno zavzema osrednje mesto med tveganji. Kreditno tveganje pomeni, da tretja oseba (kreditojemalec) ne bo sposobna ali ne bo pripravljena poravnati svojih obveznosti. Povezano je s skoraj vsemi bančnimi transakcijami in zanj velja splošno pravilo, da čim dlje terjatev traja, tem večje je tveganje (Bobek, [7], str. 71).

Pri vsakem kreditiranju prevzame banka potencialno tveganje, da ji kreditojemalec kredita ne bo vrnil v dogovorjenem roku ali pa sploh ne. Torej mora banka računati z možnostjo

¹S tem je mišljena nevarnost delne ali popolne izgube pri terjatvah.

izpada pri vračanju kredita ter pri plačilu aktivnih obresti. Bančni posel je namreč zaključen šele, ko banki posojilojemalc vrne dolg brez kakršnekoli izgube. Banke omejujejo slaba posojila tako, da pred odobritvijo analizirajo boniteto potencialnega kreditojemalca, pozorno spremljajo vračanje kredita in razpršujejo svoj kreditni portfelj, s čimer želijo znižati kreditno tveganje. Za preprečevanje tveganj sprejemajo banke posebne ukrepe, ki so znani pod skupnim imenom zavarovanje kredita (Slak, [28], str. 51).

1.2.1 Načela upravljanja s kreditnim tveganjem

Upravljanje s tveganji je eden izmed ključnih elementov za uspešno poslovanje banke, saj je banka po naravi poslovanja "proizvajalec tveganj". Banka namreč tveganje sprejema, ga transformira in vgrajuje v različne bančne produkte in storitve. Tveganje potencialno povečuje izgubo oziroma zmanjšuje dobiček, zato postaja upravljanje s tveganji ključna bančna funkcija. Po besedah Bessisa ima banka, ki aktivno upravlja s tveganji, pred konkurenco odločilno prednost.

Preventivni ukrepi bank naj bi preprečili neugodne posledice tveganj, ki niso predvidljive in jih ni mogoče preprečiti. Razlikujemo preventivne ukrepe s politiko lastnega kapitala na eni strani in preventivne ukrepe bilančne politike na drugi strani (politika tihih rezerv); postavke premoženja v naložbah bilanciramo po blažjem načelu najnižje vrednosti, obremenitve pa predvidimo z najvišjo vrednostjo (Glogovšek, [16], str. 16).

Merjenje kreditnega tveganja je ključno za oblikovanje pravilne cene kredita, določitev vrste in načina zavarovanja ter postavitev ustreznih limitov zadolžitve za posameznega komitenta (Filipan, [13], str. 31). Ocenjevanje kreditnega tveganja sestoji iz kvantitativne in kvalitativne analize (Dimovski, Gregorič, [11], str. 82 - 83):

- **kvantitativna analiza** zajema zbiranje podatkov o finančni odgovornosti kreditojemalca in o naravi njegovih finančnih potreb. Temelji na matematičnih in statističnih metodah in daje objektivno oceno kreditnega tveganja;
- **kvalitativna analiza** je subjektivnega značaja in se osnuje na finančni analizi predloženih računovodskih podatkov s strani kreditojemalca, njegovih planiranih finančnih rezultatov in denarnih tokov, da bi ocenili kreditojemalčovo sposobnost "servisiranja" dolga. Tveganje ocenjuje z vidika dejavnikov, ki so značilni za posamezne posojilojemalce in z vidika tržnih dejavnikov, ki vplivajo na vse komitente.

Najpomembnejši del postopka merjenja kreditnega tveganja predstavlja izračun verjetnosti neplačila, na podlagi katere kreditojemalca razvrstimo v ustrezen razred tveganosti. Poleg tega pa merjenje kreditnega tveganja zajema še izračun izpostavljenosti banke in izračun dejanske izgube v odstotkih, ki jo banka realizira, ko je že pričela s postopkom izterjave.

1.2.2 Modeli ocenjevanja kreditnega tveganja

Baselski odbor za bančni nadzor v svoji analizi stanja na področju modeliranja kreditnega tveganja ([3]) navaja, da mora biti interni model za kreditno tveganje:

- transparenten in logičen,
- primerljiv z modeli drugih finančnih institucij,
- testiran na realnih podatkih in
- v praksi pri upravljanju s tveganji dejansko uporabljen (ne sme le zadostiti zahtevam bančnih regulatorjev).

Obstaja veliko različnih modelov ocenjevanja kreditnega tveganja. V splošnem ločimo tradicionalne in sodobne modele, ki jih včasih težko ločimo, saj so v novih modelih uporabljene mnoge zamisli tradicionalnih modelov, pri katerih osnovo modeliranja predstavlja izračun pričakovane in nepričakovane izgube kreditnega portfelja banke.

Med tradicionalne modele ocenjevanja kreditnega tveganja med drugim sodi tudi kreditno točkovanje, ki predstavlja tehniko upravljanja s kreditnim tveganjem, ki analizira tveganja posojiljemalcu na podlagi dodeljevanja točk, ki so dodeljene stranki, z namenom določanja njihove stopnje tveganja oziroma odobritve/neodobritve posojila. Za dober model kreditnega točkovanja mora veljati, da je diskriminatoren (*ang. discriminative*), kar pomeni, da visoke točke odražajo skoraj ničelno tveganje, nizke točke pa ustrezajo zelo visokemu tveganju (ali obratno). Kreditne točke so pogosto razdeljene v homogene skupine. Segmentirane točke predstavljajo diskretne ocene tveganja, ki so poznane kot razredi in ratingi (*ang. ratings*) tveganja.

Model kreditnega točkovanja pri majhnih dolžnikih se osredotoča na različne kreditojemalčeve lastnosti kot so na primer karakter, jamstva in kapital. Tako so ocene posledica temeljite analize tržnih (javnih) in zasebnih informacij vseh relevantnih virov. Proses ocenjevanja vključuje kvantitativno analizo, ki se osredotoča na strukturo dolga, računovodske izkaz, bilančne podatke in informacijski sektor.

Ko točke zberemo v homogene segmente točk ali razrede tveganja, kot rezultat dobimo bonitetno oceno. Tako kreditne točke kot tudi bonitetne ocene podajo oceno kreditnega tveganja, vendar pa se terminologija točkovanja uporablja predvsem v okolju majhnih dolžnikov, kjer večina statističnih sistemov točkovanja samodejno točkuje velike podatkovne baze kupcev.

Povečanje finančnih trgov in tveganj zunajbilančnih postavk ter zmanjševanje obrestnih prihodkov so le nekateri izmed mnogih dejavnikov, ki so vplivali na razvoj sodobnih kreditnih modelov in tehnik. Velik napredek v razvoju so prinesle tudi izboljšave v obdelavi

statističnih podatkov, ekonometrični simulaciji ter vpeljavi teorije iger in matematičnega programiranja. Tako vrsta akademskih raziskav in bonitetnih metodologij verjetnost neplačila ocenjuje s t. i. kreditnimi točkovalnimi modeli, ki pravzaprav predstavljajo statistična orodja za multivariatno analizo, kjer se pojasnjuje vpliv nabora neodvisnih spremenljivk na odvisno spremenljivko. Novejši modeli torej uporabljajo dognanja finančne teorije in temeljijo na matematičnih in statističnih metodah.

Satistična analiza se deli na:

- **univariatno:**

kazalniki se analizirajo eden za drugim, kar pomeni, da se ne upošteva njihova medsebojna odvisnost. Sem spadajo ROC analiza, merjenje korelacije z odvisno spremenljivko ter grafična analiza. Naštete metode so opisane v §3 in §4.3.1;

- **multivariatno:**

sem sodijo diskriminantna analiza, model linearne verjetnosti ter logistična in probit regresija. Pri diskriminantni analizi razvijemo funkcijo, ki maksimizira razmerje med varianco med skupinami proti varianci znotraj skupin. Pomembna predpostavka pri tej analizi je, da morajo biti spremenljivke normalno porazdeljene, porazdelitve znotraj skupin pa morajo biti enake. Model linearne verjetnosti predstavlja linearno funkcijo neodvisnih spremenljivk na podlagi preteklega poslovanja. Logistična regresija je podrobnejše opisana v §3.2. Probit oz. normit regresija je zelo podobna logistični, pomembna razlika je predpostavka, da kumulativna vrednost ni porazdeljena logistično, ampak normalno.

Tudi v tej nalogi se zavedamo pomena in učinkovitosti podatkovnega rudarjenja ter statistične analize. Iz matematičnega vidika smo se odločili, da bomo za razvoj bonitetnega modela uporabili metodologijo statističnega ocenjevanja verjetnosti neplačila z uporabo ROC analize, odločitvenih dreves in binarne logistične regresije. Omenjene metode so podrobnejše opisane v §3.

Poglavlje 2

IRB pristop

Pristop na podlagi notranjih bonitetnih sistemov je eden izmed dveh¹ možnih načinov ocenjevanja kreditnega tveganja. Njegov cilj je zajeti resnično ekonomsko tveganje bančnega portfelja ter natančno oceniti tveganja, kar omogoča natančnejši izračun in posledično nižje kapitalske regulative.

Čeprav morda nekatere banke lahko uporabijo lastne izkušnje, bo vsaka banka, ki prevzame IRB pristop, skoraj zagotovo morala upoštevati skupne značilnosti izgub in izterjav, opredeljenih na podlagi širokega nabora akademskih in gospodarskih raziskav. Tudi nadzorniki se morajo za oceno ustreznosti banke prav tako zavedati, katere so te skupne značilnosti.

2.1 Razvoj

Zamisel o IRB pristopu je bila prvič objavljena v prvem posvetovalnem dokumentu [3] Baselskega odbora za bančni nadzor, izdanem v juniju 1999. Od tistega časa je odbor vložil veliko truda in dela v razvoj začetnih idej. Sestavni del tega dela je bilo posvetovanje z gospodarskimi združenji in posameznimi bankami v obliki anket, izmenjave podatkov in predstavitev. Tako so prizadevanja odbora in povratne informacije sodelujočih pomagale oblikovati začetne predloge. Pri nadaljnjem razvoju IRB pristopa je Baselski odbor želel razviti okvir, ki je kredibilen, v bonitetnem smislu zanesljiv in odraža zanesljivo prakso upravljanja s kreditnimi tveganji. Poleg zagotavljanja spodbude za individualne banke je odbor upal, da se bo istočasno pristop ustalil in zagotovil spodbudo za nenehno izboljšanje praks upravljanja s tveganji na širši ravni bančništva. Odbor je bil mnenja, da je najboljši način za zagotavljanja teh ciljev uvedba evolucijskega pristopa k IRB okvirju, ki odraža stalni razvoj samega upravljanja s kreditnim tveganjem.

¹ Drugi je standardizirani pristop, o katerem si lahko več preberete v [4], [8], [21].

Banke so notranje bonitetne sisteme zelo dolgo časa uporabljale kot sredstvo za kategorizacijo svojih izpostavljenosti v široke, kvalitativne in diferencirane plasti tveganja. Mnoge banke so, skozi širitev svojih zmogljivosti za kvantifikacijo kreditnega tveganja povezanega z njihovimi izpostavljenostmi, v zadnjih letih dosegle velik napredek pri krepitevi teh tradicionalnih ter kakovostno usmerjenih notranjih ocen kreditnega tveganja. IRB pristop bo kot nadgradnja teh sposobnosti, za vsako kategorijo izpostavljenosti (npr. podjetja, prebivalstvo ipd.) določil en sam okvir, ki bo dani niz komponent tveganja ali "vhodne podatke" pretvoril v minimalne kapitalske zahteve. Iz ozira do nekaterih od teh komponent tveganja je odbor predstavil dve metodologiji ocenjevanja: osnovni in napredni pristop.

V svojih raziskavah bančne prakse in razpravah z industrijo je odbor odkril, da se mnoge banke pri vzpostavljanju verodostojnih in zanesljivih ocen nekaterih dejavnikov tveganja, ki bi jih lahko ustrezzo potrdila tako banka kot tudi njeni nadzorniki, soočajo s težavami. Kljub temu pa mnoge od teh bank lahko in tudi zagotavljajo pomembne in izmerljive ocene izmed najbolj temeljnih komponent kreditnega tveganja - tveganja neplačila dolžnika. Tako je za te banke, za nekatere kategorije izpostavljenosti, odbor predlagal osnovni pristop, po katerem banke podajo svoje ocene tveganja neplačila dolžnika, ocene dodatnih dejavnikov tveganja pa so pridobljene z uporabo standardiziranih pravil. Osnovni pristop lahko uporablja banke, ki lahko nadzornikom dokažejo, da izpolnjujejo določene minimalne zahteve, ki se nanašajo na notranje bonitetne sisteme, procese obvladovanja tveganja in sposobnost ocenjevanja potrebnih komponent tveganja.

Odbor je odkril tudi, da nekatere banke lahko (ali pa bodo kmalu zmožne) zagotovijo zanesljive in dosledne ocene dodatnih komponent tveganja, ki so:

- verjetna izguba ob nastopu neplačila kreditojemalca,
- verjetna stopnja izpostavljenosti do kreditojemalca v času neplačila in
- učinek osebnih jamstev in kreditnih izvedenih finančnih instrumentov na tveganje izpostavljenosti.

Tako je poleg osnovnega pristopa na voljo tudi nabor naprednih metodologij, ki omogočajo bankam, da uporabljajo svoje interne ocene komponent tveganja. Odbor je bil mnenja, da je širša uporaba teh ocen pomemben del dinamičnega in za tveganja senzibilnega IRB pristopa, ki lahko prepozna in razlikuje tiste banke, ki lahko zagotovijo dovolj robustne in izmerljive ocene tveganja. Poleg tega uporaba naprednega pristopa natančneje usklajuje kapitalske zahteve z internimi praksami merjenja in upravljanja tveganja bank ter je v skladu s filozofijo zagotavljanja spodbude za banke, da izboljšajo te prakse. Banke, ki želijo uporabiti napredni pristop, bodo morale izpolnjevati minimalne zahteve osnovnega pristopa, kot tudi dodaten niz specifičnih zahtev, značilnih za komponente tveganja, ki se ocenjujejo.

Na začetku je bil bankam dovoljen izračun svojih kapitalskih zahtev na podlagi lastnih ali t. i. vendor portfolio modelov kreditnega tveganja. Leta 1999 je odbor uporabo in izvajanje takšnih modelov, raziskal v svojem poročilu z naslovom Modeliranje kreditnega tveganja: Sedanje prakse in uporaba [4]. Poročilo se konča z ugotovitvijo, da je v tistem času uporaba rezultatov takšnih modelov kot podlaga za določitev minimalnih kapitalskih zahtev prezgodnjega. Glavni pomanjkljivosti uporabe modelov kreditnega tveganja kot podlaga za minimalne regulatorne kapitalske zahteve sta bili kakovost podatkov ter sposobnost bank in nadzornikov, da bi preverili vzorčne rezultate². Vendar pa, ker morajo banke izpolnjevati stroge zahteve prakse, v smislu »vhodov in izhodov«³ notranjega bonitetnega sistema, in najpomembnejše, na tej stopnji, izločili lastne presoje portfeljskih učinkov, kot sta koncentracija in diverzifikacija, je bil odbor mnenja, da je te pomanjkljivosti mogoče odpraviti v okviru IRB pristopa. Celo napredni pristop naj ne bi dovolil, da bi prilagoditve posameznih bank do ukrepov kreditnega tveganja odražale tveganje korelacije med različnimi posojiljemalcem (v bistvu je to meja kompleksnosti, ki je IRB pristop, kot je bil oblikovan v tistem času, ne bi dosegel oz. presegel). Vendar pa je bil, pri oblikovanju okvira, ki predvideva in zagotavlja ustrezne spodbude za razvoj praks upravljanja s tveganji, odbor mnenja, da bi pristop na podlagi notranjih ocen v prihodnosti utrl pot do prehoda na modeliranje kreditnega tveganja, na podlagi celotnega portfelja.

Pri nadaljnjem oblikovanju IRB pristopa je odbor kot izhodišče izbral pristop za kategorije izpostavljenosti do enot centralne ravni držav ter centralnih bank do institucij in do podjetij. Odbor je prav tako razvil predloge za kategorijo izpostavljenosti do majhnih dolžnikov in kasneje IRB pristop razširil še na ostale vrste izpostavljenosti. Bančna skupina, ki izpolnjuje minimalne zahteve in uporablja IRB pristop za nekatere izpostavljenosti, mora v razumno kratkem času ta pristop sprejeti tudi po vseh ostalih kategorijah izpostavljenosti ter pomembnih poslovnih enotah (skupine, podružnice in hčerinska podjetja). Banke se morajo skupaj z domačim nadzornikom strinjati o agresivnem ter spretinem začetnem načrtu. Nekatere izpostavljenosti v nesignifikantnih poslovnih enotah, ki so nepomembne glede na velikost in zaznani profil tveganja, so lahko, v skladu z nacionalno diskrecijo, iz zgoraj omenjenega pravila izvzete.

Kot lahko torej vidimo, se IRB pristop nenehno spreminja in dopolnjuje. Kljub temu pa njegove osnovne lastnosti in prednosti ostaja jo ves čas enake.

²Notranji bonitetni sistemi igrajo ključno vlogo v številnih modelih kreditnega tveganja, kar pomeni, da sta kakovost in validacija pomembni tako za IRB pristop kot tudi za modeliranje kreditnega tveganja.

³Kot vhodi in izhodi so mišljeni podatki, ki jih banka uporablja pri modeliranju kreditnega tveganja ter rezultati, ki jih pri tem dobti.

2.2 Lastnosti in prednosti

IRB pristop omogoča podrobnejše razvrščanje komitentov, kar zagotavlja učinkovito pomoč pri spremljanju kreditov, trendov in celotnega portfelja. Pri tem daje prednost lastni bančni oceni tveganja, saj banka sama najbolje pozna tveganje, ki mu je izpostavljena. Tako ta pristop temelji na lastnem sistemu internih bonitetnih ocen. Lahko bi tudi rekli, da temelji na sistemu razvrščanja komitentov, ki omogoča razlikovanje kreditnega tveganja za različne vrste portfelja, kot na primer za podjetja, fizične osebe, kapitalske naložbe in podobno. Zaradi naštetih lastnosti ima IRB pristop seveda tudi številne prednosti:

- namesto dveh sistemov za merjenje tveganj bodo banke lahko uporabljale enega: tako za interno upravljanje s tveganji kot tudi za regulatorne namene;
- banke, ki bodo uporabljale pristop, ki temelji na notranjih ocenah, bodo veljale za bolj kredibilne, saj uporaba naprednejših pristopov merjenja tveganja zahteva izpolnjevanje številnih pogojev ter validacijo s strani nadzorstva;
- uveljavitev IRB pristopa omogoča nižje regulatorne kapitalske zahteve;
- doslednost ocenjevanja in kakovost pridobivanja podatkov nista več odvisni od objektivnosti in neodvisnosti zunanjih agencij, saj IRB pristop omogoča lastni nadzor nad kakovostjo in validacijo modela;
- omogoča boljše razlikovanje med razredi tveganja ter pokriva veliko večje število dolžnikov;
- upošteva dodatne faktorje tveganja⁴.

2.3 Minimalne zahteve za uvedbo in nadaljnjo uporabo IRB pristopa

Banke, ki bodo želele pridobiti soglasje za uvedbo in nadaljnjo uporabo svojega sistema internih bonitetnih ocen, bodo morale zadostiti obsežnemu naboru minimalnih zahtev, in to ne le na začetku, torej v trenutku odobritve, temveč ves čas svojega poslovanja. Tako bodo morale sistematično razvrščati ter meriti kreditno tveganje, tako da bo lahko nadzornik potrdil konsistentnost, zanesljivost in veljavnost njihovega bonitetnega sistema (Svet Banke Slovenije, [30]).

⁴IRB pristop upošteva informacije, ki jih zunanje agencije poznajo zelo slabo ali pa z njimi sploh niso seznanjene.

Ker se v nalogi osredotočamo na kreditiranje prebivalstva, bomo podrobneje opisali minimalne zahteve, ki se nanašajo na kategorijo izpostavljenosti do majhnih dolžnikov (glej §2.4.1).

2.3.1 Zasnova in postavitev bonitetnega sistema

Termin ‐bonitetni sistem‐ obsega metode, procese, kontrolo, zbiranje podatkov ter potrebne informacijske tehnologije, ki služijo za ocenjevanje kreditnega tveganja, dodelitev internih bonitetnih ocen posameznim dolžnikom ali terjatvam ter določitev ocen neplačil in izgub.

V okviru vsake posamezne kategorije izpostavljenosti lahko banka uporablja različne metodologije razvrščanja. Pri tem lahko bonitetne sisteme prilagodi za posebne panože ali tržne segmente, vendar pa dolžnikov ne sme razporejati na način, s katerim bi minimizirala regulatorne kapitalske zahteve. Pri opredelitvi bonitetnega razreda ali skupine pa se morajo upoštevati kvalitativna in kvantitativna merila, saj so bonitetni razredi z nejasnimi ali samo kvantitativnimi merili neustrezni.

Za vse banke, ki uporabljajo statistične modele in druge mehanske metode za razvrščanje izpostavljenosti v bonitetne razrede dolžnikov ali skupine izpostavljenosti, Zakon o bančništvu določa, da morajo:

- dokazati, da ima model dobro napovedno moč in da kapitalska zahteva zaradi njegove uporabe ni izkrivljena. Prav tako mora model biti v čim večji meri nepristranski;
- vzpostaviti proces vrednotenja vhodnih podatkov, ki vključuje oceno popolnosti, natančnosti in primernosti podatkov;
- dokazati, da uporabljeni podatki za izgradnjo modela predstavljajo populacijo dejanskih dolžnikov ali izpostavljenosti banke;
- vzpostaviti redne postopke preverjanja veljavnosti modela, kot na primer pregled specifikacij modela in testiranje pridobljenih rezultatov modela glede na dejanske izide ter spremljanje delovanja in stabilnosti modela;
- dopolnjevati statistični model s človeško presojo in človeškim nadzorom ter tako garantirati pravilno uporabo modelov.

Bonitetni sistemi za portfelje majhnih dolžnikov morajo obravnavati tako tveganje kreditojemalca ter posla kot tudi njune pomembne značilnosti. Prav tako morajo banke vsako izpostavljenost, ki v okviru IRB pristopa po definiciji sodi med majhne dolžnike, razporediti v določeno skupino (*ang. pool*). Dodeljevanje mora omogočati grupiranje izpostavljenosti

na primeren in homogen način, smiselno diferenciacijo tveganja, ter natančno in dosledno ocenjevanje izgube v okviru skupine. Razporejanje v skupine pa mora med drugim temeljiti na sledečih dejavnikih tveganja:

- značilnosti tveganja kreditojemalca (npr. vrsta/tip kreditojemalca, starost, zaposlitev ...);
- velikost posojila;
- zapadlost (10 let, 30 let ...);
- značilnosti terjatve, upoštevajoč vrsto naložbe in/ali zavarovanja;
- izpostavljenosti do problematičnih kreditov (pričakuje se, da bodo banke ločeno od ostalih izpostavljenosti izkazovale zapadle terjatve oz. izpostavljenost do dolžnikov v zamudi).

Vsaka banka mora prav tako vse podrobnosti v zvezi s sestavo in z delovanjem svojih bonitetnih procesov zapisati oziroma dokumentirati v pisni obliki. Dokumentacija mora zajemati lastne opredelitve neplačila in izgube, metodologijo statističnih sistemov (če so uporabljeni) ter razloge za izbor kvalitativnih in kvantitativnih meril za razvrščanje v bonitetne razrede ali skupine ter analizo, ki ta izbor utemeljuje. Tako dokumentacija dokazuje skladnost z minimalnimi zahtevami in obravnava razločevanje kategorij izpostavljenosti in/ali poslovnih enot, bonitetna merila, odgovornost oseb, ki ocenjujejo dolžnike in izpostavljenosti, pogostost preverjanja razvrstitev in nadzor vodstva nad bonitetnimi procesi (Svet Banke Slovenije, [30], str. 15 - 25).

2.3.2 Delovanje bonitetnega sistema

Delovanje bonitetnega sistema zajema obseg bonitetnih ocen (*ang. coverage of ratings*), celovitost dodeljevanja bonitetne ocene, razveljavitev (*ang. overrides*), vzdrževanje baz podatkov ter teste izjemnih situacij za ocenjevanje kapitalske ustreznosti. Pri izpostavljenosti do majhnih dolžnikov morajo banke za vsako identificirano skupino vsaj enkrat letno pregledati karakteristike stanja neplačil (*ang. delinquency status*) in izgube ter, vse od prve dodelitve bonitetne ocene dalje, zbirati in hraniti podatke o kreditojemalcih in priznanih porokih. Še posebej morajo zbirati informacije o neplačnikih in terjtvah, ki niso bile poravnane. Banke, ki uporabljajo IRB pristop, morajo izdelati učinkovite teste izjemnih situacij za ocenitev kapitalske ustreznosti. Ti testi bi naj prepoznali oziroma napovedali tako verjetne, pričakovane dogodke in spremembe v prihodnosti kot tudi bodoči ekonomski položaj, ki bi glede na strukturo izpostavljenosti banke in oceno sposobnosti banke, da se zoperstavi tem spremembam, lahko imel neugodne posledice (Svet Banke Slovenije, [30], str. 15-17).

2.3.3 Upravljanje bank in nadzorstvo

Vsi bistveni vidiki, ki zadevajo bonitetne ocene in procese določitve le-teh, morajo biti potrjeni s strani uprave ali ustreznega izvršnega telesa in s strani višjega menedžmenta. Dobro poznavanje sistema in procesov delovanja bonitetnega sistema je ena osnovnih nalog višjega menedžmenta.

Zakon o bančništву določa, da mora nadzorstvo kreditnega tveganja biti neodvisno od osebja in enot, pristojnih za odobravanje ali obnavljanje izpostavljenosti ter mora poročati neposredno upravi in višjemu menedžmentu. Pристојности nadzorstva pa so:

- testiranje in spremljanje bonitetnih razredov in skupin;
- izdelava analiz in poročil o bonitetnih sistemih;
- izvajanje postopkov za preverjanje doslednosti opredelitve bonitetnih razredov in skupin;
- preverjanje in dokumentiranje vseh sprememb znotraj bonitetnih sistemov, vključno z razlogi za spremembe;
- aktivno sodelovanje pri oblikovanju ali izbiranju, uvedbi in potrditvi primernosti modelov, uporabljenih v bonitetnih sistemih;
- spremjava in pregled ter nenehno pregledovanje in dograjevanje modelov, uporabljenih v bonitetnih sistemih.

2.3.4 Uporaba internih bonitetnih ocen

Interne bonitetne ocene ter ocene neplačil in izgub morajo igrati glavno vlogo znotraj procesa odobritve kreditov, upravljanja s tveganji, notranje alokacije kapitala in vodenja banke, ki uporablja IRB pristop. Bonitetni sistemi ter ocene, izdelani in uvedeni izključno za pridobitev odobritve IRB pristopa ter za zagotovitev IRB vhodnih podatkov, ne bodo sprejemljivi. Lastne ocene parametrov tveganja (glej §2.4.2) morajo tako temeljiti na vseh relevantnih in dosegljivih podatkih, informacijah in metodah. Ocene morajo biti osnovane na izkušnjah iz preteklosti ter empiričnih evidencah in ne le na osnovi subjektivne ali strokovne presoje. Zakon o bančništву tako določa, da morajo banke verodostojno beležiti uporabo podatkov o internih bonitetnih ocenah, saj morajo dokazati, da že vsaj tri leta uporabljajo bonitetni sistem, ki v veliki meri izpolnjuje minimalne zahteve opredeljene po Baslu II.

2.4 Ključni elementi

IRB pristop v namen določanja potrebnega kapitala za pokrivanje kreditnega tveganja, poleg že omenjenih minimalnih zahtev, vsebuje štiri ključne elemente:

- razvrstitev izpostavljenosti;
- parametre tveganja;
- funkcijo tveganja in
- nadzorniški pregled skladnosti z minimalnimi zahtevami.

2.4.1 Razvrstitev izpostavljenosti

Zakon o bančništву določa, da morajo banke vsako svojo izpostavljenost, za katero ne izračunajo kapitalske zahteve za tržna tveganja, za namen izračuna tveganjem prilagojenih sredstev izpostavljenosti, razvrstiti v enega od naslednjih razredov izpostavljenosti:

- do enot centralne ravni držav in centralnih bank;
- do institucij;
- do podjetij;
- do majhnih dolžnikov (bančništvo na drobno);
- iz naslova lastniških instrumentov;
- do pozicije v listinjenju;
- do drugih sredstev iz naslova nekreditnih obveznosti.

Pri razvrščanju izpostavljenosti v različne razrede morajo banke uporabiti metodologijo, ki zagotavlja pravilno in dosledno oblikovanje kategorij izpostavljenosti.

Ker se v nalogi osredotočimo na kreditiranje prebivalstva, bomo podrobnejše opisali le razred izpostavljenosti do majhnih dolžnikov.

Po Zakonu o bančništву je izpostavljenost kategorizirana kot izpostavljenost do majhnih dolžnikov, če izpolnjuje vsa naslednja merila, glede na:

- a) lastnosti posojilojemalca ali nizke vrednosti posameznih izpostavljenosti:

- izpostavljenost do fizičnih oseb⁵ v splošnem spadajo med izpostavljenosti do majhnih dolžnikov, ne glede na velikost izpostavljenosti;
- stanovanjska hipotekarna posojila⁶ (*ang. residential mortgage loans*) spadajo v razred izpostavljenosti do majhnih dolžnikov, ne glede na velikost izpostavljenosti tako dolgo, dokler se kredit ne podaljša na posameznika, ki je lastnik in stnovalec nepremičnine;
- Krediti malim podjetjem, pri katerih je skupna izpostavljenost manjša od enega milijona evrov, ter manjša posojila podjetjem, ki jih najame ali pa zajamči posameznik, spadajo v kategorijo izpostavljenosti do majhnih dolžnikov;

b) veliko število izpostavljenosti:

- izpostavljenost mora biti ena iz velike skupine izpostavljenosti, ki jih upravljajo banke na ravni skupin. Nadzorniki lahko, znotraj skupine izpostavljenosti, določijo minimalno število izpostavljenosti, ki jih je treba obravnavati kot izpostavljenosti do majhnih dolžnikov;
- izpostavljenosti do majhnih podjetij, nižje od enega milijona evrov, se lahko obravnavajo kot izpostavljenosti do majhnih dolžnikov, če banka takšne izpostavljenosti v svojem notranjem sistemu upravljanja s tveganji ves čas obravnava dosledno in na enak način kot druge izpostavljenosti do majhnih dolžnikov. Takšnih izpostavljenosti ni več potrebno upravljati individualno ter primerljivo z izpostavljenostmi do podjetij, ampak kot del segmenta portfelja ali skupine izpostavljenosti s podobnimi značilnostmi tveganja. Vendar pa to ne izključuje možnosti, da bodo izpostavljenosti do majhnih dolžnikov v neki fazi procesa upravljanja s tveganji obravnavane individualno.

2.4.2 Parametri tveganja

Kapitalska zahteva za izpostavljenost znotraj vsakega od prej omenjenih šestih razredov izpostavljenosti, je odvisna od specifičnega niza komponent tveganja, ki predstavlja ključne vhodne podatke za metode v okviru IRB pristopa. Te komponente so:

- verjetnost neplačila (*PD – ang. probability of default*);
- izguba v primeru neplačila (*LGD – ang. loss given default*);

⁵Na primer obnavljajoči se krediti in kreditne linije (npr. kreditne kartice in limiti), kot tudi osebna ročna posojila in najemi (npr. obročna posojila, študentska in izobraževalna posojila, osebne finance in druge izpostavljenosti s podobnimi lastnostmi).

⁶Vključno s prvo in nadaljnji zastavnimi pravicami in ročnimi posojili.

- izpostavljenost ob neplačilu (*EAD – ang. exposure at default*);
- zapadlost (*M – ang. effective maturity*);
- velikost dolžnika.

Pomembno komponento tveganja predstavlja tudi korelacija (ρ - *ang. correlation*) med komponentami, zlasti v okviru portfeljskih modelov za merjenje kreditnega tveganja (Fitch Ratings, [14], str. 4).

Na podlagi ocenjevane komponente tveganja ločimo:

1. osnovni pristop: (*ang. foundation IRB*)

za oceno kreditnega tveganja lahko banke uporabijo interne ratinge ter lastne modele le za ocenjevanje *PD*, vrednosti za *LGD*, *EAD* in *M* pa so določene s strani nadzornika oz. zunanje agencije;

2. napredni pristop: (*ang. advanced IRB*)

banke lahko uporabijo lastne ocene za vse komponente tveganja (*PD*, *LGD*, *EAD*, *M*). Ta pristop uporabljajo banke, ki imajo najboljše znanje ter najbolj razvite metode in modele za merjenje in upravljanje kreditnega tveganja.

Posebnost pri računanju kapitalske zahteve predstavlja razred izpostavljenost do majhnih dolžnikov, kjer morajo banke tako pri osnovnem kot tudi naprednem pristopu same izračunati ocene *PD*, *LGD* in *EAD*. To pomeni, da ta razred izpostavljenosti nima osnovnega pristopa.

Verjetnost neplačila

Verjetnost neplačila je finančni izraz, ki opisuje verjetnost, da dolžnik svoje obveznosti v določenem časovnem obdobju ne bo poravnal. Z drugimi besedami lahko rečemo tudi, da zagotavlja oceno verjetnosti, da stranka finančne ustanove ne bo mogla izpolniti svoje dolžniške obveznosti. *PD* je tako ključni parameter pri izračunu regulatornega⁷ kapitala bančnih institucij v okviru Basla II. Vsaka ocena *PD* mora predstavljati konzervativen pogled dolgoročnega (časovni horizont za oceno *PD* je eno leto) povprečja *PD* in mora biti utemeljena na podlagi zgodovinskih izkušenj ter empiričnih dokazov. Prav tako mora, za izpolnjevanje pogojev IRB pristopa, pripravo ocen in procesov upravljanja s tveganji in bonitetnih nalog, odražati popolno skladnost z minimalnimi zahtevami (vključno z interno uporabo in zahtevami po razkritju, povezanimi z ocenami).

⁷Kapital potreben za pokrivanje kreditnega tveganja.

Če želimo primerjati *PD* med različnimi institucijami potrebujemo enotno definicijo neplačila. Preden je kapitalski sporazum Basel II določil mednarodno sprejemljivo definicijo neplačila, so banke imele svoje opredelitev tega pojma, ki pa niso bile vedno skladne. Tako primerjava ni bila možna, saj so banke kot neizpolnitev obveznosti smatralo različno število dni zaostajanja plačila in različno dovoljeno mejo vrednosti neizterjane terjatve. Enotna definicija pravi, da neizpolnitev dolžniških obveznosti nastopi takrat:

- kadar je malo verjetno, da bo dolžnik svoj dolg banki sposoben odplačati, brez da bi zastavil zavarovanje;
- kadar ima dolžnik več kot 90 dnevno zamudo pri izpolnitvi bistvenih kreditnih obveznostih⁸. Pri izpostavljenosti do majhnih dolžnikov/prebivalstva in do oseb javnega prava lahko nacionalni nadzornik, v kolikor je mnenja, da je to glede na lokalne razmere primerneje, to obdobje podaljša na 180 dni.

Ugotovili smo torej, da *PD* predstavlja ocena verjetnosti, da se bo neizpolnitev dolžnosti zgodila po določenem časovnem obdobju ocenjevanja. Običajno je to obdobje eno leto. Ocenimo ga lahko za posameznega dolžnika, kar je običajna praksa pri t. i. wholesale banking-u ali pa za segment dolžnikov, ki si delijo podobne značilnosti kreditnega tveganja, kar pa je običajna praksa pri poslovanju s prebivalstvom. Pri standardiziranem pristopu znotraj kategorije izpostavljenosti do majhnih dolžnikov znaša *PD* najmanj 0,03 %, *PD* neplačnikov ali neplačanih izpostavljenosti pa znaša 100 %.

Za ocenjevanje verjetnosti neplačila obstaja veliko alternativ. En način je ocena, pridobljena iz zgodovinske baze podatkov dejanskih neplačil, pri čemer uporabimo sodobne metode, kot na primer logistično regresijo. Drug način je določanje ocene na podlagi tržnih informacij o kreditnih razmikih produktov, ki vsebujejo kreditno tveganje (kreditni izvedeni finančni instrumenti, podjetniške obveznice) ali pa implicitno preko cen lastniških vrednostnih papirjev. Najenostavnnejši pristop je uporaba zunanjih bonitetnih agencij, kot so Standard and Poor's, Fitch in Moody's Investors Service za ocenjevanje *PD* iz zgodovinskih izkušenj z neplačili.

Pri oceni *PD* moramo upoštevati, da ta ni odvisna samo od značilnosti tveganja nekega dolžnika, ampak tudi od gospodarskega okolja in stopnje, s katero ta vpliva na dolžnika. Tako lahko razpoložljive informacije za oceno *PD* razdelimo v dve večji skupini:

- makroekonomski podatki⁹, kot so indeksi cen stanovanj, brezposelnost, stopnja rasti BDP itd.;

⁸Prekoračitev se smatra kot zapadla, ko dolžnik preseže odobren limit ali ko mu je ta nižji od vsote tekočih obveznosti.

⁹Ti podatki so skupni večim dolžnikom.

- dolžniku specifični podatki, kot so rast prihodkov, število neplačil v zadnjih šestih mesecih (majhni dolžniki) itd.

Glede na vrsto uporabljenih informacij za izračun ocene PD ločimo neobremenjeni *PD* (*ang. unstressed PD*) in obremenjeni *PD* (*ang. stressed PD*). Neobremenjeni *PD* upošteva tako makroekonomske kot tudi dolžniku specifične podatke. To pomeni, da se bo v primeru poslabšanja makroekonomskih razmer *PD* dolžnika vedno bolj povečeval, če pa se gospodarske razmere izboljšajo, pa se bo *PD* zmanjšal. Obremenjeni *PD* pa upošteva trenutne dolžniku specifične informacije ter t. i. stressed makroekonomske podatke, ki so neodvisni od trenutnega gospodarskega stanja. Ta *PD* se glede na značilnosti tveganja dolžnika sčasoma spreminja, vendar pa spremembe v gospodarskem ciklu nanj nimajo večjega vpliva, saj so neugodne gospodarske razmere že vključeno v oceno.

Izguba v primeru neplačila

Izguba v primeru neplačila se izraža v odstotku od kreditne izpostavljenosti, za vsako posamezno bonitetno oceno ter prav tako predstavlja pomemben parameter pri izračunu regulatornega kapitala bančnih institucij v okviru Basla II. Lahko tudi rečemo, da v primeru neizpolnitve dolžnikovih obveznosti *LGD* predstavlja odstotek izgube vseh izpostavljenosti. Ker v splošnem nanjo vplivajo ključne značilnosti poslovanja, kot sta jamstvo ter stopnja podrejenosti, je *LGD* objektno specifična. Za lažje razumevanje sledi primer enostavnega izračuna.

Zgled. Dolžnik ne more plačati dolga v višini 400.000 € (izpostavljenost - *EAD*), banka ali zavarovalnica pa je njegovo zavarovanje (npr. stanovanje), zastavila le za 320.000 € neto (vključno s stroški, povezanimi s ponovnim odkupom). To pomeni, da je 80.000 € oziroma 20 % izgubljenih. V tem primeru je torej *LGD* enaka 20 %.

Teoretično se *LGD* lahko računa na različne načine, vendar pa je najbolj priljubljen bruto¹⁰ *LGD*, kjer se skupne izgube delijo z izpostavljenostjo ob neplačilu. Drugi način je delitev izgub z nezajamčenim deležem kreditne linije (kjer jamstvo zajema del *EAD*). To je znano kot blanco¹¹ *LGD*. Če je v zadnjem primeru vrednost zavarovanja enaka nič, je blanco *LGD* enaka bruto *LGD*.

¹⁰Bruto *LGD* je med akademiki najbolj priljubljena zaradi svoje preprostosti in njihove dostopnosti do podatkov na trgu obveznic, kjer so vrednosti zavarovanja pogosto neizračunane, nepomembne ali pa sploh niso znane.

¹¹Blanco *LGD* je priljubljena med nekaterimi bankami, ki imajo veliko zastavljenih objektov in bi zaradi amortizacije kakovosti jamstva radi uravnali izgube med nezajamčenimi in zajamčenimi odseki. Slednje je tudi diskretna zahteva Basla II, vendar večina bank v tem trenutku za izpolnitev ni dovolj sofisticiranih.

Za določitev potrebnega kapitala za banke ali finančne institucije v okviru Basla II, mora vsaka institucija izračunati tveganju prilagojena sredstva (*ang. risk-weighted assets*). To zahteva oceno *LGD* za vsako od posameznih izpostavljenosti. Kot je bilo že omenjeno, obstajata dva načina za izpeljavo te ocene: osnovni in napredni pristop.

Pri osnovnem pristopu je *LGD* ocenjena preko standardnih nadzornih pravil, ki na podlagi značilnosti danega posla, vključno s prisotnostjo in vrsto jamstva, razlikujejo raven *LGD*. Pri tem so nadzorna pravila in obravnave konservativne. Kot izhodišče je Baselski odbor, za večino nezajamčenih poslov, predlagal uporabo 45 % *LGD*. Pri podrejenih izpostavljenostih pa se uporablja 75-odstotna *LGD*. Pri transakcijah z določenimi finančnimi jamstvi se *LGD* zmanjša ali poveča na stopnjo, do katere je posel zajamčen¹². Za posle z določenimi jamstvi v obliki komercialnih in stanovanjskih nepremičnin se uporablja poseben niz nadzornih vrednosti *LGD* in pravil za priznanje. Vsi drugi posli se za ta regulatoren namen obravnavajo kot nezajamčeni.

Pri naprednem pristopu banka, na podlagi robustnih podatkov in analiz, ki jih je mogoče potrditi tako interna kot tudi preko nadzornika, sama določi primerno oceno za *LGD*. Tako je, na podlagi širšega nabora značilnosti posla (npr. tip proizvoda, širši nabor oblik jamstva) in značilnosti posojiljemalca, banka sposobna razlikovati vrednosti *LGD*. Podobno kot pri ocenah *PD* se pričakuje, da te vrednosti predstavljajo konservativen pogled dolgoročnega (najmanj 5 let) povprečja. Banke bi lahko prosto uporabljale tudi bolj konservativne ocene. Banke, ki želijo uporabiti svoje lastne ocene *LGD*, bodo morale nadzornikom pokazati, da lahko izpolnijo dodatne minimalne zahteve, ki zadevajo celovitost in zanesljivost teh ocen.

LGD model določata vrednost in/ali kakovosti bančnega jamstva za zagotovitev posojila. Jamstva so lahko bodisi stroji, kot so avtomobili, tovornjaki ali gradbeni stroji bodisi hipoteke, skrbniški računi ali pa blago. Višja kot je vrednost jamstva, nižja je *LGD* in s tem tudi potencialna izguba s katero se banka ali zavarovalnica sooča v primeru neplačila. Banke, ki uporabljajo napredni pristop, morajo določiti vrednosti *LGD*, medtem ko morajo tiste, ki uporabljajo osnovni pristop, vrednosti *LGD* določiti le za portfelj majhnih dolžnikov¹³.

LGD se meri na osnovi kreditnega produkta. Običajno zavzema vrednosti med 0 in 100 % in je bodisi zastopana na kontinuirani način ali preko izgube kakovosti. V primeru, da izgube ni, je *LGD* enaka nič. Kadar pa je izguba enaka celotnemu znesku izpostavljenosti, je *LGD* enaka 100 %. Negativna *LGD* pa označuje dobiček (npr. zaradi obrestne mere in plačila zamudnine). Izguba v primeru neplačila ni fiksen, ampak je dinamičen parameter,

¹²V ta namen se uporablja tako imenovana metodologija striženja (*ang. haircut methodology*), ki je prilagojena metodologiji v standardiziranem pristopu.

¹³Leta 2013 je bilo v Združenem kraljestvu z lastnimi modeli hipotekarnih *LGD* devet podjetij. V Švici sta bili dve banki. V Nemčiji ima tržni vodja Bausparkasse Schwäbisch Hall svoje modele hipotekarnih *LGD*. Znotraj razreda premoženja podjetij pa mnogo nemških bank še vedno uporablja samo vrednosti, podane s strani regulatorjev v okviru osnovnega IRB pristopa.

saj se njegove vrednosti od posla do posla spreminjajo. Nekateri dolžniki si celo opomorejo in poplačajo ves dolg ter odložena plačila. V nekaterih primerih se lahko pogajanja med dolžnikom, ki ne izpolnjuje obveznosti in vsemi upniki, konča v dogovoru, kjer vse udeležene stranke nosijo del izgube. V najslabšem primeru lahko neizpolnitev obveznosti vodi v stečajni postopek z visokimi izgubami ter v zaključek sodelovanja med banko in stranko. Način neizpolnitve obveznosti ima lahko velik vpliv na dejansko izgubo, ki pa v trenutku neplačila in naložbe ni poznana. V primeru neplačila imajo banke pravico sprejeti pravne ukrepe. Pri tem lahko čas in vrsta ukrepov vplivata na dejansko vračilo dolga.

Znotraj razreda izpostavljenosti do majhnih dolžnikov ločimo dve možnosti za izračun ocene *LGD*. Prvi pristop predstavlja neposredno merjenje izgube na portfelju. Drugi pristop pa o oceni *LGD* sklepa na podlagi izgube in *PD*. Pri obeh pristopih mora biti ocena podana na podlagi podatkov, pridobljenih iz najmanj petletnega časovnega obdobja. V praksi se vrednosti *LGD* zelo spreminjajo in so odvisne od vrste neplačila in njegove razrešitve:

- **poplačilo dolga:** kmalu po nastopu neplačila je dolžnik poravnal svoje obveznosti, na primer zaradi dodatnega prihodka ali delničarjevega posredovanja. Tako kreditojemalec še naprej izpolnjuje svoje pogodbene obveznosti. V tem primeru banka nima znatne škode in odnos s stranko se ne spremeni;
- **prestrukturiranje:** dolžnik je svoj dolg po prestrukturiraju zmožen poplačati, na primer po tem, ko so pogajanja o dolgu vodila do podaljšanja zapadlosti ter delne oprostitve dolga. V tem primeru banka raje sprejme takojšnjo manjšo izgubo, kot pa veliko izgubo v kasnejšem časovnem obdobju;
- **likvidacija:** banka predlaga likvidacijo ali stečaj dolžnika in uveljavi garancije oz. zaseže zastavo, pri čemer lahko likvidacijski postopki vključujejo visoke pravne stroške ter še višje bančne izgube.

Vrsto razrešitve pred nastopom neplačila težko predvidimo. V povprečju likvidacija najpogosteje nastopi pri strankah, v katere si banke in investorji želijo manj investirati. V primerih visokega neplačila in tveganja izgube, bo banka, z zahtevanjem zastave ali jamstva, poskušala zmanjšati tveganje izgube.

Nekatere banke imajo poleg *PD* bonitetne lestvice, ločeno *LGD* ocenjevalno levcivo. Druge banke združujejo informacije o *LGD* in *PD* v bonitetno levcivo pričakovane izgube (*EL* - angl. *expected loss*), ki predstavlja produkt *PD* in *LGD*:

$$EL = PD \cdot LGD. \quad (2.1)$$

Pred časom so tudi zunanje bonitetne agencije začele izračunavati *LGD*, posebej v smislu spremljanja deleža poplačila (*RR* - angl. *recovery rate*), ki poleg verjetnosti neplačila pred-

stavlja najpomembnejšo informacijo za odločitev glede kreditne investicije (Fitch Ratings, [14], str. 14). Delež poplačila terjatve je torej komplementarna vrednost izgube v primeru neplačila:

$$RR = 1 - LGD. \quad (2.2)$$

Izpostavljenost v primeru neplačila

Izpostavljenost v primeru neplačila je parameter, ki se v okviru Basla II uporablja pri izračunu regulatornega kapitala bančnih institucij in je, tako kot LGD , objektno specifičen. V splošnem EAD predstavlja oceno obsega izpostavljenosti banke do dolžnika v primeru neizpolnitve obveznosti in je v primeru fiksnih izpostavljenosti (npr. ročna posojila¹⁴ - *ang. term loans*) enaka trenutnemu znesku neporavnane dolga oz. nominalnemu znesku. V primeru odprtih kreditnih linij je potrebno upoštevati tudi pričakovane prihodnje odlive iz naslova neizkoriščenega dela kreditne linije.

Podobno kot pri LGD se izračun ocene EAD pri osnovnem pristopu razlikuje od tistega pri naprednjem pristopu. Osnovni pristop upošteva osnovna sredstva, vrednotenje terminskih pogodb, vrsto naložbe in podrobnosti obvez. Ne upošteva pa garancij, jamstev ali zavarovanj (ignorira tehnike za zmanjšanje kreditnega tveganja). Samo oceno EAD določi zunanjja agencija. Pri naprednjem pristopu imajo banke pri izračunu ocene EAD več svobode, saj lahko same določijo, kakšna vrednost EAD je primerna za posamezno izpostavljenost. Tako lahko, na podlagi širšega nabora značilnosti posla (npr. vrste kreditnega produkta) ter značilnosti posojiljemalca, razlikujejo vrednosti EAD . Za te vrednosti se pričakuje, da predstavljajo konservativen pogled dolgoročnega (najmanj 5 let) povprečja. V idealnem primeru je EAD ocenjena na podlagi časovnega obdobja, ki v celoti zajema nek ekonomski cikel¹⁵. Banke, ki želijo uporabiti svoje lastne ocene EAD , bodo morale nadzornikom pokazati, da lahko izpolnijo dodatne minimalne zahteve glede celovitosti in zanesljivosti teh ocen.

EAD je ugotovljena v trenutku neplačila (lahko tudi kasneje) in običajno niha med črpanim zneskom (*DA* - *ang. drawn amount*) na napovedani datum in deležem neizkoriščenega zneska (*UA* - *ang. undrawn amount*), ki je na voljo in je seveda odvisen od danega kreditnega limita (*CL* - *ang. credit limit*). Velja naslednja zveza:

$$EAD = DA_t + CCF \cdot (CL_t - DA_t), \quad (2.3)$$

kjer CCF predstavlja konverzijski faktor (*ang. credit conversion factor*). Napoved je obi-

¹⁴Vsa posojila, ki imajo rok zapadlosti.

¹⁵Podobno kot pri LGD bi tudi tukaj lahko prosto uporabljale tudi bolj konservativne ocene.

čajno narejena v relevantni časovni točki t , v letu pred nastankom neplačila. CCF tako niha med 0 in 1, odvisno od višine neizkoriščenega zneska ($UA = CL - DA$). Vrednosti pod 0 se lahko pojavijo, če je banka pred nastopom neplačila zmanjšala izpostavljenost. Vrednosti CCF so nad 1, kadar so pred nastopom neplačila dovoljene dodatne kreditne linije.

Vsako iz bonitetnega sistema izhaja jočo utež tveganja, ki se spremeni v tveganju prilagojena sredstva, je potrebno povezati z višino izpostavljenosti. Tako bo vsaka napaka v izračunu EAD neposredno vplivala na tveganju prilagojena sredstva in s tem tudi na kapitalske zahteve. Torej lahko vidimo, da je natančna ocena vrednosti EAD zelo pomembna.

Zapadlost

Zapadlost je omejeno časovno obdobje, znotraj katerega finančni posel ostaja neporavnani. Na koncu tega časovnega obdobja je dolžnik dolžan poravnati vse svoje obveznosti do upnika. Najpogosteje se ta izraz uporablja v okviru naložb s fiksnim donosom, kot so obveznice in depoziti. V okviru IRB pristopa se zapadlost obravnava kot posredna komponenta tveganja, saj daljša kot je zapadlost, večja je negotovost in posledično tudi tveganje, ki se mu banka izpostavlja. Tako bodo banke nadzornikom morale predložiti efektivne pogodbene zapadlosti njihovih izpostavljenosti.

Vpliv dejanske zapadlosti znotraj osnovnega pristopa je zelo omejen. Napredni pristop omogoča večjo občutljivost tveganja. V večini primerov so meritve zapadlosti tehnično preproste, vendar lahko kljub temu zahtevajo veliko poznavanja informacijske tehnologije. V okviru osnovnega pristopa se znotraj funkcije tveganja uporablja dva in polletna zapadlost ($M = 2,5$). Za nekatere kratkoročne (*ang. repo-style*) posle se uporablja le polletna zapadlost ($M = 0,5$). Nacionalni nadzorni organi lahko določijo tudi druge stopnje zapadlosti. Pri naprednjem pristopu morajo banke meriti zapadlosti sklenjenih poslov na sledeč način:

$$M = \frac{\sum_k t_k CF_k}{\sum_k CF_k}, \quad (2.4)$$

kjer CF_k označuje denarne tokove, ki jih po pogodbi plačuje dolžnik v času t_k ali časovnem obdobju od sklenitve pogodbe dalje ($t = 0$). Denarni tokovi so v principu lahko plačila obresti, glavnice in odpravnin. Kadar finančne institucije ne morejo izračunati zapadlosti, se za njeno vrednost določi maksimalen čas, ki ga ima dolžnik še na voljo, da v celoti poravna svoje pogodbene obveznosti. Običajno je to kar preostala nominalna zapadlost.

Velikost dolžnika

Velikost dolžnika predstavlja vrednost sklenjenega posla v dani denarni valuti oz. vsoto, ki jo mora dolžnik v določenem časovnem obdobju vrniti upniku. Večja kot je velikost dolžnika,

večja je morebitna izguba banke in obratno.

2.4.3 Funkcija tveganja

Uteži za tehtanje kreditnega tveganja se v okviru IRB pristopa izračunajo kot funkcija vhodnih podatkov PD , LGD , EAD in v nekaterih primerih tudi M . Tako komponente tveganja s pomočjo zvezne funkcije uteženega tveganja (*ang. risk weight function*) pretvorimo v uteži, ki določajo kapitalske zahteve za posamezne kategorije izpostavljenosti bančnega portfelja. Ta pristop omogoča večjo diferenciacijo tveganja ter prilaganje različnim bonitetnim strukturam bank, saj opušča enoten nabor uteži, ki ga uporablja standardizirani pristop, kjer je posojilojemalcem na podlagi nadzornih standardnih obravnav ali ocen zunanjih bonitetnih institucij dodeljena ena od petih uteži tveganja (0 %, 20 %, 50 %, 100 %, 150 %). Glede na dodatno občutljivost IRB pristopa uteži tveganja (*RW - ang. risk weights*) odražajo celoten spekter kreditne kvalitete, saj je namesto petih posrednih skupin tveganja standardiziranega pristopa uporabljena zvezna funkcija uteži. Tako v okviru IRB pristopa različna skupina vhodnih podatkov v splošnem ustvari različne uteži tveganja. Na ta način lahko imajo izpostavljenosti do dolžnikov, kjer ocene komponent tveganja podajo zelo nizko stopnjo tveganja, uteži tveganja, ki so pod vrednostjo njihovih ekvivalentov standardiziranega pristopa in obratno.

Kapitalska zahteva (K - *ang. capital requirement*) za kreditno tveganje za kategorijo izpostavljenosti do majhnih dolžnikov znaša najmanj 8 % tveganjem prilagojenih zneskov (*RWA* - *ang. risk-weighted assets*) izpostavljenosti:

$$K \geq 8\% \cdot RWA \quad (2.5)$$

in se izračuna po formuli:

$$K = LGD \cdot N((1 - \rho)^{-0,5} \cdot G(PD) + (\rho/(1 - \rho))^{0,5} \cdot G(0, 999)) - PD \cdot LGD, \quad (2.6)$$

kjer je:

- $N(x)$ kumulativna porazdelitvena funkcija za standardizirano normalno slučajno spremenljivko z , ki predstavlja verjetnost, da je normalna slučajna spremenljivka s povprečjem nič in varianco ena, manjša ali enaka x ;
- ρ korelacija med PD in LGD :

$$\rho = 0,03 \cdot (1 - e^{-35 \cdot PD}) / (1 - e^{-35}) + 0,16 \cdot (1 - (1 - e^{-35 \cdot PD}) / (1 - e^{-35})); \quad (2.7)$$

- $G(z) = N^{-1}(z) = x$ je inverzna kumulativna porazdelitvena funkcija za standardizirano normalno slučajno spremenljivko z (vrne vrednost x za katero velja $N(x) = z$).

Za izračun tveganju prilagojenih sredstev se uteži tveganja pomnožijo z mero izpostavljenosti oz. z oceno EAD , nato se prištejejo izračunane količine celotnega portfelja, nazadnje pa se vsemu doda še korekcijski faktor, v obliki standardnega nadzornega indeksa:

$$\begin{aligned} RWA &= f(PD, LGD, EAD, M) \\ &= RW \cdot EAD \\ &= K \cdot 12,5 \cdot EAD. \end{aligned}$$

Uteži tveganja za izpostavljenosti do majhnih dolžnikov se izračunajo v skladu z naslednjo formulo:

$$RW = K \cdot 12,5 \cdot 1,06. \quad (2.8)$$

Konstanti 12,5 in 1,06 sta določeni s strani bančnih regulatorjev.

V primeru, ko je PD enak 100 %, velja:

$$RW = \max\{0, 12.5 \cdot (LGD - ELBE)\}, \quad (2.9)$$

kjer je $ELBE$ najboljša ocena povprečne izgube banke za neplačane izpostavljenosti.

Za izpostavljenosti do majhnih dolžnikov, zavarovane z nepremičninami, se namesto vrednosti korelacije, izračunane s korelacijsko formulo 2.7, uporabi $\rho = 0,15$.

Za kvalificirane obnavljajoče se izpostavljenosti do majhnih dolžnikov pa se namesto vrednosti korelacije, izračunane s korelacijsko formulo 2.7, uporabi $\rho = 0,04$.

Zneski pričakovanih izgub (EL_a - ang. *expected loss amounts*) za izpostavljenosti do majhnih dolžnikov se izračunajo v skladu s formulama 2.1 in:

$$EL_a = EL \cdot EAD. \quad (2.10)$$

EL za neplačane izpostavljenosti ($PD = 100\%$) je vsota najboljših ocen pričakovane izgube banke za vsako izpostavljenost glede na trenutne gospodarske razmere ter druge okoliščine in mogočih dodatnih nepričakovanih izgub v obdobju razreševanja teh izpostavljenosti.

2.4.4 Nadzorniški pregled skladnosti z minimalnimi zahtevami

Banke ter druge institucije, ki bodo uporabljale lastne modele za izračun regulatornega kapitala, morajo vzpostaviti zanesljive metode za potrditev in presojo natančnosti in konistentnosti modela ter njegovih vhodnih podatkov. To pomeni, da morajo biti vsi koraki modeliranja jasno in razumljivo zapisani ter v celoti dokumentirani, da lahko nadzorniki po potrebi preverijo ustreznost razvitega modela.

Če bodo banke že lele uporabljati lastne modele, bodo morale nadzornikom dokazati, da njihov model izpolnjuje naslednje pogoje:

- zaradi uporabe modela ne sme priti do izkrivljanja regulatornih kapitalskih zahtev;
- spremenljivke modela morajo tvoriti smiseln sklop napovednih kazalcev;
- model kot celota mora biti nepristranski in v povprečju točen za vse vrste dolžnikov;
- podatki, uporabljeni za postavitev modela, morajo biti reprezentativni za dejanski portfelj kreditojemalcev in terjatev banke;
- lastno pridobljene ocene morajo biti celovite in zanesljive.

Poglavlje 3

Matematične osnove

Pri razvoju modela se bomo naslonili na statistično terminologijo, zato jo najprej obnovimo. Statistična enota je posamezni proučevani element. Množica vseh proučevanih statističnih enot se imenuje statistična populacija. Ta množica je lahko zelo velika (včasih tudi neskončna), zato je ne moremo obravnavati v celoti in izberemo le del, ki ga proučujemo. Podmnožico populacije, na osnovi katere sklepamo o lastnostih celotne populacije, imenujemo vzorec.

Lastnostim statističnih enot, ki jih opazujemo, pravimo statistični znaki oz. statistične spremenljivke (npr. spol, starost). Statistične spremenljivke se glede na tip izražanja vrednosti delijo na opisne ali atributivne, ki jih lahko opišemo le z besedami (npr. poklic) in na številske ali numerične, katerih vrednost lahko izrazimo s števili (npr. starost).

V našem primeru statistično populacijo tvorijo vsi potencialni kreditojemalci, ki so fizične osebe. Posamezen komitent predstavlja statistično enoto, njegova lastnost oz. atribut predstavlja statistični znak.

Naj bo X vzorec, ki zaobjema statistične enote x_1, x_2, \dots, x_n . Naj bo $A : X \rightarrow Z_A$ statistični znak in $a_j = A(x_j)$ vrednost statističnega znaka A na statistični enoti x_j . Naj $A(x_j)$ predstavlja slučajno spremenljivko. Zaloga $Z_A = \{a_1, \dots, T, \dots, a_l\}$ je zaloga vrednosti atributa oz. statističnega znaka A , kjer:

- $l = |Z_A|$ predstavlja število vrednosti atributa A in se od atributa do atributa spreminja;
- T je element zaloge vrednosti atributa A in je spremenljiv (ni fiksen) parameter;

V primeru zveznih vrednosti atributa A je njegova zaloga vrednosti nek interval.

Naj bo $Y : X \rightarrow Z_Y$ posebej izbran statistični atribut enot X , ki ga imenujemo klasifikacijski atribut na podlagi katerega vzorec razdelimo na podvzorce oziroma klasifikacijske razrede. Naj $Y(x_j)$ predstavlja slučajno spremenljivko. Zaloga $Z_Y = \{v_1, \dots, v_k\}$, kjer je $k \in \mathbb{N}$, je zaloga vrednosti klasifikacijskega atributa Y . V primeru binarnega klasifikacijskega modela je $k = 2$, kar pomeni, da ima klasifikacijski atribut le 2 klasifikacijski vrednosti in tako lahko vzorec X razdelimo na dva podvzorca. V prvem vzorcu imajo podatki klasifikacijsko lastnost v_1 , v drugem pa klasifikacijsko lastnost v_2 .

Verjetnost, da ima podatek klasifikacijsko lastnost $v \in Z_Y$, izračunamo po formuli

$$P_v(X) = \frac{|X_v|}{|X|}, \quad (3.1)$$

kjer je X_v podvzorec podatkov s klasifikacijsko lastnostjo v . Verjetnost, da ima podatek, z lastnostjo $a \in Z_A$, klasifikacijsko lastnost $v \in Z_Y$, pa izračunamo po formuli

$$P_v(X_a) = \frac{|X_{v,a}|}{|X_a|}, \quad (3.2)$$

kjer je $X_{v,a}$ podvzorec podatkov s klasifikacijsko lastnostjo v in lastnostjo a .

Zgled. Dan imamo portfelj komitentov, za katere vemo, koliko so stari, kakšno izobrazbo so zaključili ter ali so plačniki oziroma neplačniki. Dani portfelj tako predstavlja vzorec X , posamezen komitent predstavlja podatek, starost in izobrazba sta atributa oziroma lastnosti komitenta, klasifikacijski atribut Y pa ima zalogo vrednosti $Z_Y = \{0, 1\}$, kjer 0 predstavlja plačnike, 1 pa neplačnike.

Recimo, da imata atributa izobrazba in starost sledeči zalogi vrednosti:

$$Z_{izobrazba} = \{osnovnosolska, srednjesolska, visjesolska\};$$

$$Z_{starost} = \{18, 19, 20, \dots, 60\}.$$

Pri atributu izobrazba je $l = 3$, pri starosti pa je $l = 43$.

Naj bo P število vseh plačnikov, N število vseh neplačnikov, P_a oz. N_a pa število vseh plačnikov oz. neplačnikov z lastnostjo a . Potem je verjetnost, da je komitent plačnik enaka

$$P_0(X) = \frac{|X_0|}{|X|} = \frac{P}{P + N},$$

verjetnost, da je komitent neplačnik je enaka

$$P_1(X) = \frac{|X_1|}{|X|} = \frac{N}{P + N},$$

in verjetnost, da je komitent z osnovnošolsko (os) izobrazbo plačnik, je enaka

$$P_0(X_{os}) = \frac{|X_{0,os}|}{|X_{os}|} = \frac{P_{os}}{P_{os} + N_{os}}.$$

3.1 ROC analiza

ROC (*ang. receiver operating characteristics*) analiza zagotavlja orodja za izbiro optimalnih modelov in zavrnitev suboptimalnih ter je neposredno in na naraven način povezana z analizo stroškov/prednosti izdelave diagnostičnih modelov. Prav tako zagotavlja koristno sredstvo za primerjavo uspešnosti več kot enega kazalnika za isti rezultat ter ocenjevanje diagnostične natančnosti kazalnika¹, ki temelji na tem, kako dobro kazalnik loči naš vzorec v dve skupini².

Za primer si poglejmo binarni klasifikacijski model, kjer klasifikacijski atribut K zavzame 2 vrednosti: pozitivno (1) ali negativno (0). Pri tem ločimo dejanske vrednosti in napovedane vrednosti, ki so v obeh primerih lahko pozitivne kot tudi negativne. Tako obstajajo štirje možni izidi tega modela:

- če je izid napovedi 1 in je dejanska vrednost prav tako 1, se napovedana vrednost imenuje pravilno uvrščena pozitivna vrednost oz. **pravilno pozitivna vrednost** (TP – *ang. true positive*), če pa je dejanska vrednost 0, potem napovedano vrednost imenujemo nepravilno uvrščena pozitivna vrednost oz. **lažno pozitivna vrednost** (FP – *ang. false positive*);
- nasprotno dobimo pravilno uvrščeno negativno vrednost oz. **pravilno negativno vrednost** (TN – *true negative*), ko sta napoved izida in dejanska vrednost obe enaki 0, in nepravilno uvrščeno negativno vrednost oz. **lažno negativna vrednost** (FN – *ang. false negative*), ko je izid napovedi 0, dejanska vrednost pa 1.

V našem primeru želimo izdelati bonitetni model, ki bo napovedal, ali bo nek komitent v prihodnosti postal neplačnik, ali bo ostal plačnik. Lažna pozitivna vrednost se v tem primeru pojavi, kadar model napove, da bo komitent postal neplačnik, v resnici pa je ostal plačnik. Podobno dobimo lažno negativno vrednost, kadar model napove, da bo komitent ostal plačnik, v resnici pa je postal neplačnik.

Poglejmo si primer, kjer imamo N plačnikov in P neplačnikov. Potem lahko rezultate modela prikažemo v kontingenčni tabeli oz. matriki razvrstitev (*ang. confusion matrix*)

Kontingenčna tabela lahko poda več različnih metrik vrednotenja:

¹Kazalnik je atribut oz. v našem primeru posamezna lastnost komitenta kot npr. starost ali zakonski status.

²V našem primeru sta to skupina plačnikov in skupina neplačnikov.

		Dejanske vrednosti	1	0
Napovedane vrednosti	1	TP	FP	
	0	FN	TN	

Slika 3.1: Kontingenčna tabela

- **senzitivnost** (TPR - ang. *true positive rate*) predstavlja delež vseh pravilno pozitivnih med dejansko pozitivnimi (P) vrednostmi:

$$TPR = TP/P = TP/(TP + FN); \quad (3.3)$$

- **specifičnost** (SPC - ang. *specificity*) predstavlja delež pravilno negativnih med vsemi dejansko negativnimi (N) vrednostmi:

$$SPC = TN/N = TN/(FP + TN); \quad (3.4)$$

- **vrednost pozitivno napovedanih** (PPV - ang. *positive predictive value*) predstavlja delež pravilno pozitivnih med vsemi napovedano pozitivnimi vrednostmi:

$$PPV = TP/(TP + FP); \quad (3.5)$$

- **vrednost negativno napovedanih** (NPV - ang. *negative predictive value*) predstavlja delež pravilno negativnih med vsemi negativno napovedanimi vrednostmi:

$$NPV = TN/(TN + FN); \quad (3.6)$$

- **stopnja lažno pozitivnih vrednosti** (FPR - ang. *false positive rate*) predstavlja delež lažno pozitivnih med vsemi dejansko negativnimi vrednostmi:

$$FPR = FP/N = FP/(FP + TN) = 1 - SPC; \quad (3.7)$$

- **stopnja napačnega odkritja** (FDR - ang. *false discovery rate*) predstavlja delež

lažno pozitivnih med vsemi pozitivno napovedanimi vrednostmi:

$$FDR = FP/(FP + TP) = 1 - PPV; \quad (3.8)$$

- **stopnja lažno negativnih vrednosti** (*FNR* - *ang. false negative rate*) predstavlja delež lažno negativnih med vsemi dejansko pozitivnimi vrednostmi:

$$FNR = FN/P = FN/(FN + TP); \quad (3.9)$$

- **natančnost** (*ACC* - *ang. accuracy*) predstavlja delež vseh pravilno napovedanih vrednosti med vsemi dejanskimi vrednostmi:

$$ACC = (TP + TN)/(P + N). \quad (3.10)$$

3.1.1 ROC krivulja

V teoriji zaznavanja signalov ROC krivulja predstavlja grafični prikaz, ki ponazarja delovanje binarnega klasifikacijskega sistema v odvisnosti od odločitvenega pravila klasifikacije. Koordinate točk na krivulji tvorita senzitivnost ($y = TPR$) in stopnja lažno pozitivnih vrednosti ($x = FPR$), pri čemer TPR določa, koliko pravilno pozitivnih izidov se pojavi v primerjavi z vsemi dejansko pozitivnimi vrednostmi, FPR pa določa, koliko lažno pozitivnih izidov se pojavi v primerjavi z vsemi dejansko negativnimi vrednostmi. Tako krivulja podaja razmerje med senzitivnostjo in specifičnostjo določenega klasifikatorja. Točka, ki na ROC grafu ponazarja situacijo z največjim deležem pravilno uvrščenih negativnih primerov in najmanjšim deležem pravilno uvrščenih pozitivnih primerov ima koordinato $(0,0)$.

Najboljša možna metoda napovedovanja bi podala točko v zgornjem levem kotu grafa oz. koordinato $(0,1)$, ki predstavlja 100 % senzitivnost³ in 100 % specifičnost⁴. Točka $(0,1)$ se imenuje tudi popolna klasifikacija. Povsem naključno ugibanje bi dalo točko vzdolž diagonalne daljice (tako imenovane daljice brez diskriminacije), ki poteka iz spodnjega levega v zgornji desni kot⁵. Intuitiven primer naključnega ugibanja predstavlja odločanje na podlagi metanja kovanca (glava ali grb). S povečanjem velikosti vzorca se ROC točka naključnega klasifikatorja pomakne proti točki $(0.5, 0.5)$.

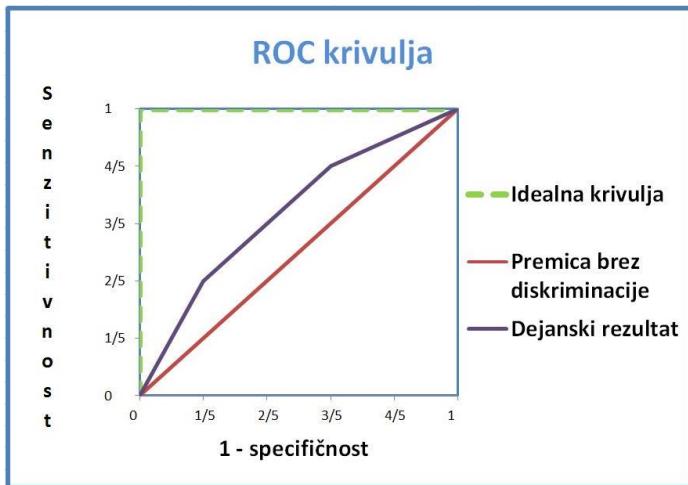
Diagonala razdeli ROC območje na dva enako velika dela. Točke nad in pod⁶ diagonalo

³To pomeni, da ni prisotnih nobenih lažno negativnih vrednosti.

⁴Ni lažno pozitivnih vrednosti.

⁵Ne glede na pozitivne in negativne bazne vrednosti.

⁶Pri tem je potrebno upoštevati, da se rezultati dosledno slabega prediktorja lahko preprosto obrnejo, da se pridobi dober prediktor.



Slika 3.2: Prikaz ROC krivulje

predstavljajo dobre⁷ klasifikacijske rezultate, točke blizu diagonale pa slabe⁸.

Krivulja ROC parametrično prikazuje $TPR(T)$ v primerjavi z $FPR(T)$, kjer je T spremenljiv (ni fiksen) parameter in predstavlja vrednost atributa A . Kadar klasifikacija temelji na diskretnih vrednostih atributa A , sta $TPR(T)$ in $FPR(T)$ definirani na sledeči način:

$$FPR(T) = \sum_{a=T}^{a_l} P_0(a) \quad \text{in} \quad TPR(T) = \sum_{a=T}^{a_l} P_1(a).$$

Pri tem je

$$P_0(a) = \frac{P_a}{P} \quad \text{in} \quad P_1(a) = \frac{N_a}{N},$$

kjer je

- P_a število vseh plačnikov z lastnostjo a ,
- N_a število vseh neplačnikov z lastnostjo a ,
- $P_0(a)$ verjetnost, da ima plačnik vrednost atributa A enako a in
- $P_1(a)$ verjetnost, da ima neplačnik vrednost atributa A enako a .

Začetna točka krivulje je vedno točka $T_0(0,0)$, končna točka pa je točka $T_l(1,1)$.

Zgled. Dan imamo portfelj komitentov, ki opisuje, ali so komitenti plačniki ali neplačniki.

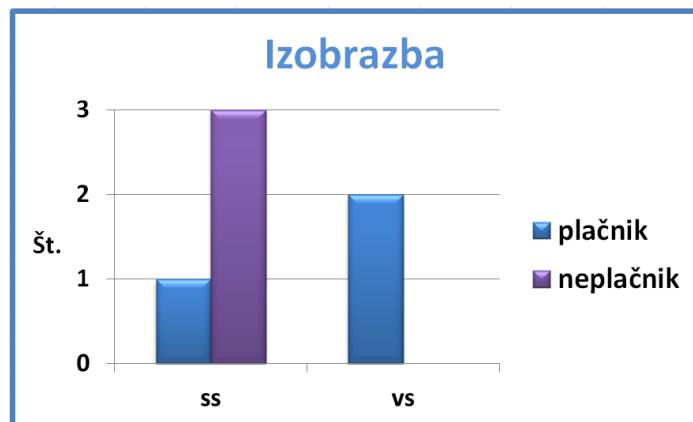
⁷To pomeni, da so boljši od naključne izbire.

⁸Dobimo podobne rezultate kot pri naključni izbiri.

Stanovanjski status	Status osebe	Izobrazba	Plačnik
lastnik	študent	višješolska	da
najemnik	zaposlen	srednješolska	da
lastnik	zaposlen	višješolska	da
pri starših	študent	srednješolska	ne
lastnik	študent	srednješolska	ne
najemnik	študent	srednješolska	ne

Slika 3.3: Primer portfelja komitentov

Poglejmo si porazdelitvi plačnikov in neplačnikov glede na izobrazbo. Ker ima komitent lahko bodisi srednješolsko (ss) bodisi višješolsko (vs) izobrazbo, je njena zaloga vrednosti enaka $Z_{izobrazba} = \{ss, vs\}$. To pomeni, da bodo ROC krivuljo določale le 3 točke.



Slika 3.4: Porazdelitvi komitentov glede na izobrazbo in status neplačnika oz. plačnika

Najprej izračunamo:

$$P_0(ss) = \frac{3}{3} = 1 \quad P_1(ss) = \frac{1}{3}$$

$$P_0(vs) = \frac{0}{3} = 0 \quad P_1(vs) = \frac{2}{3}.$$

Nato sledi izračun koordinat za prikaz ROC krivulje. Prva je točka $T_0(0,0)$. Sledi izračun koordinat druge točke:

$$FPR(T = vs) = \sum_{a \in \{vs\}} P_0(a) = 0$$

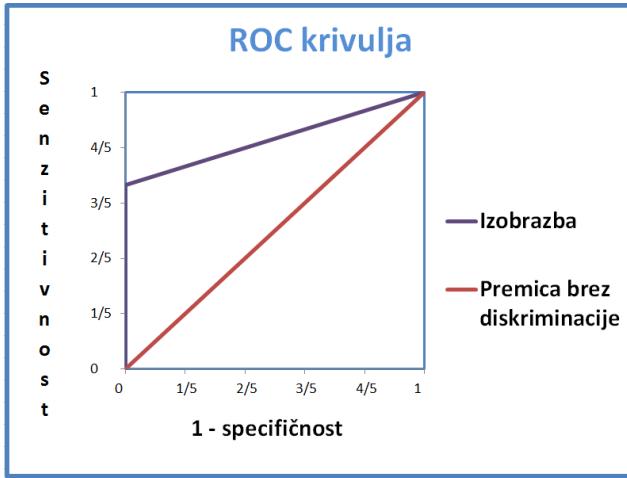
$$TPR(T = vs) = \sum_{a \in \{vs\}} P_1(a) = \frac{2}{3}.$$

Tako dobimo točko $T_1(0, \frac{2}{3})$.

$$FPR(T = ss) = \sum_{a \in \{ss, vs\}} P_0(a) = 1 + 0 = 1$$

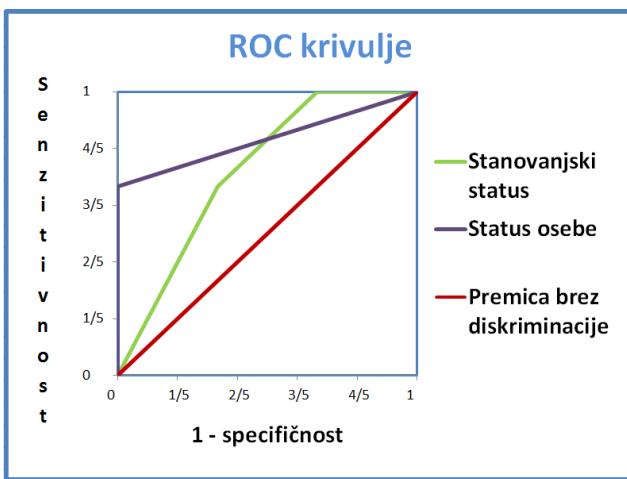
$$TPR(T = ss) = \sum_{a \in \{ss, vs\}} P_1(a) = \frac{2}{3} + \frac{2}{3} = 1$$

Tretja in zadnje pa je točka $T_2(1, 1)$. ROC krivulja za atribut izobrazba je tako sledeča:



Slika 3.5: ROC krivulja za kazalnik izobrazba

Če zgoraj opisani postopek ponovimo še za atributa status osebe in stanovanjski status, dobimo sledeči ROC krivulji na sliki 3.6.



Slika 3.6: Prikaz ROC krivulj za atributa status osebe in stanovanjski status

Opazimo lahko, da ROC krivulji atributov izobrazba in status osebe sovpadata.

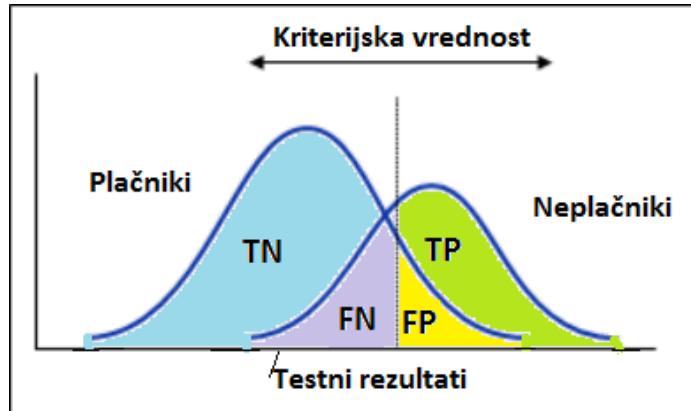
Klasifikacije pogosto temeljijo tudi na zveznih atributih. V tem primeru se verjetnost za pripadnost razredu zapiše kot funkcija odločilnega/mejnega parametra T kot $P_1(T)$ in verjetnost ne pripadnosti kot $P_0(T)$. Pri tem je lažno pozitivna stopnja FPR podana z

$$FPR(T) = \int_T^{\infty} P_0(t)dt \quad (3.11)$$

in pravilno pozitivna stopnja z

$$TPR(T) = \int_T^{\infty} P_1(t)dt. \quad (3.12)$$

Zgled. Naj bodo starosti komitentov pri plačnikih in neplačnikih porazdeljene normalno s povprečji x in y let. Bonitetni model bi lahko meril starosti komitentov in klasificiral število rezultatov, ki so nad neko določeno mejo in kažejo na neplačnika. Mejno vrednost lahko spremenimo in posledično se spremeni tudi FPR . Dvig meje bi povzročil zmanjšanje števila FP , povečanje števila FN ter levi pomik na ROC krivulji. Dejanska oblika krivulje pa je odvisna od tega koliko se posamezni porazdelitvi dejansko prekrivata oz. sovpadata, kar prikazuje slika 3.7.



Slika 3.7: Simbolični prikaz porazdelitev plačnikov in neplačnikov

3.1.2 Območje pod ROC krivuljo

Območje pod ROC krivuljo (*AUC - ang. area under the curve*) je na x osi določeno z FPR (1-specifičnost), na y osi pa z TPR (senzitivnost) ter prikazuje relativne kompromise med pravilno in lažno pozitivnimi. Vsak rezultat napovedi oz. instanca kontingenčne tabele predstavlja eno točko v ROC območju, ki predstavlja pomembno merilo točnosti kazalnika. Če je ploščina območja enaka 1, ROC krivulja sestoji iz dveh ravnih črt: navpične med točkama $(0, 0)$ in $(0, 1)$ ter horizontalne med točkama $(0, 1)$ in $(1, 1)$. V tem primeru je

kazalnik 100 % natančen, saj sta tako njegova občutljivost kot specifičnost enaki 1. To pomeni, da ni lažno pozitivnih in lažno negativnih napovedi. Po drugi strani pa kazalniku, ki ne more razlikovati med plačniki in neplačniki, ustreza ROC krivulja, ki je ekvivalentna diagonalni daljici med (0, 0) in (1, 1). Ploščina ROC območja za to krivuljo pa znaša 0,5.

V splošnem se ploščina območij pod ROC krivuljo giblje med 0,5 in 1, kot je prikazano na sliki 1 in je podana z enačbo⁹:

$$\begin{aligned} AUC &= \int_{-\infty}^{-\infty} y(T)x'(T)dT \\ &= \int_{-\infty}^{-\infty} TPR(T)FPR'(T)dT \\ &= \int_{-\infty}^{-\infty} TPR(T)P_0(T)dT \\ &= \langle TPR \rangle, \end{aligned}$$

kjer oglati oklepaji določajo povprečje porazdelitve plačnikov.

Kadar imamo diskrette vrednosti pa velja

$$AUC = \frac{1}{2} \sum_{k=2}^m (x_k - x_{k-1})(y_k + y_{k-1}). \quad (3.13)$$

Na ta način je AUC mogoče izračunati s povprečnim številom trapezoidnih približkov.

Zgled. Nadaljevali bomo prvi zgled iz poglavja o ROC krivuljah, kjer smo narisali 3 ROC krivulje. Tako bomo sedaj izračunali njihove ploščine pod krivuljo.

Ploščina pod ROC krivuljo atributa izobrazba je tako enaka

$$\begin{aligned} AUC &= \frac{1}{2} \sum_{k=1}^2 (x_k - x_{k-1})(y_k + y_{k-1}) \\ &= \frac{1}{2} [(0 - 0)(2/3 + 0) + (1 - 0)(1 + 2/3)] \\ &= 5/6. \end{aligned}$$

Ker ROC krivulja statusa osebe svopada s krivuljo izobrazbe, sta tudi ploščini pod krivuljama

⁹Meje integrala se obrnejo, ko ima T nižje vrednosti na x osi.

enaki. Ploščina pod ROC krivuljo atributa stanovanjski status pa je enaka

$$\begin{aligned} AUC &= \frac{1}{2} \sum_{k=1}^3 (x_k - x_{k-1})(y_k + y_{k-1}) \\ &= \frac{1}{2} [(1/3 - 0)(2/3 + 0) + (2/3 - 1/3)(1 + 2/3) + (1 - 2/3)(1 + 1)] \\ &= 13/18. \end{aligned}$$

Dva ali več kazalnikov lahko primerjamo s statistično primerjavo območja pod krivuljo ROC, ki se nanaša na posamezen kazalnik. Kazalniki so lahko korelirani, kadar nastopijo pri večkratnih meritvah na istem komitentu, lahko pa so nekorelirani, kadar izhajajo iz meritev na različnih komitentih. Pri tem je potrebno preveriti, ali je razlika med AUC vrednostmi različnih kazalnikov signifikantna. Stopnja signifikantnosti je verjetnost, da je pri naključni izbiri vzorca vrednost parametra na tem naključnem vzorcu v kritičnem območju glede na vrednost parametra na populaciji. Torej, če je populacijska vrednost parametra enaka 0, potem je pri normalni porazdelitvi spremenljivke verjetnost, da je pri naključnem vzorcu vrednost parametra vsaj 2σ stran od 0, enaka približno 5%. Torej je stopnja signifikantnosti za $x = 2\sigma$ enaka 0,05.

Zanesljivo in veljavno AUC oceno lahko interpretiramo kot verjetnost, da bo klasifikator dodelil višji rezultat naključno izbranemu pozitivnemu primeru, kot pa naključno izbranemu negativnemu. Vendar pa kritična raziskava kaže pogoste napake pri pridobivanju zanesljivih in veljavnih AUC ocen. Tako je praktična vrednost mera AUC vprašljiva, kar povečuje verjetnost, da AUC namesto rešitev primerjave klasifikacijske natančnosti strojnega učenja (*ang. machine learning*) dejansko uvaja le negotovost.

3.1.3 Gini koeficient

Gini koeficient (G - *ang. Gini coefficient/index*) meri neenakosti med vrednostmi frekvenčne porazdelitve. Vrednost nič pomeni popolno enakost, kjer so vse vrednosti enake (na primer, kjer ima vsakdo enake dohodke). Večji kot je koeficient, večja je razlika med plačniki in neplačniki. Kazalnik z visokim G tako bolje loči plačnike od neplačnikov.

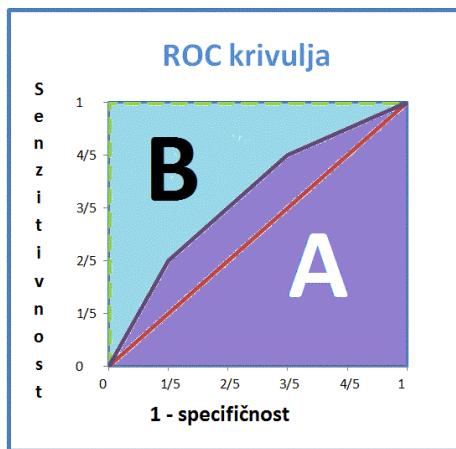
Koeficient je definiran kot razmerje površin območij na ROC grafu. Če je površina pod krivuljo ROC enaka A in površina nad ROC krivuljo enaka B , velja $G = (A - 1/2)/(A + B - 1/2)$, in ker je $A + B = 1$, velja $G = 2 \cdot A - 1$. Torej je mera AUC z Gini koeficientom povezana na naslednji način:

- če je $AUC > 0,5$, velja

$$G = 2 \cdot AUC - 1, \quad (3.14)$$

- sicer velja

$$G = 1 - 2 \cdot AUC. \quad (3.15)$$



Slika 3.8: Prikaz območij na ROC grafu

Zgled. Spomnimo se prvega zgleda iz poglavja o ROC krivuljah. Ponovno bomo nadaljevali ta zgled. Izračunali bomo Gini koeficient za vsako izmed ROC krivulj.

Podobno, kot pri ploščini pod krivuljo, velja tudi tukaj, da sta Gini koeficiente za atributa izobrazba in status osebe enaka:

$$G = 2 \cdot 5/6 - 1 \approx 0,67.$$

Gini koeficient za atribut stanovanjski status pa je enak

$$G = 2 \cdot 13/18 - 1 \approx 0,44.$$

Opazimo torej, da atributa izobrazba in status osebe bolje ločita plačnike od neplačnikov, kot pa atribut stanovanjski status.

3.1.4 Večrazredni problemi

Razširitev ROC krivulj za klasifikacijske probleme z več kot dvema razredoma je bila vedno okorna, saj stopnje prostosti kvadratično naraščajo s številom razredov in je ROC prostor

dimenzijsi $c(c - 1)$, kjer je c število razredov (glej Srinivasan, [29]). Nekateri pristopi so bili razviti posebej za primer s tremi razredi (trirazredni ROC) (glej Mossman, [23]). Kot metrika uspešnosti za več razredne probleme je bil analiziran in preučevan tudi izračun prostornine pod ROC površjem (*VUS – ang. volume under the surface*, glej Ferri, Hernandez-Orallo in Salido, [12]). Zaradi kompleksnosti aproksimacije resničnih vrednosti *VUS* so, kot metrika vrednotenja, bolj priljubljeni nekateri drugi pristopi (glej Till in Hand, [36]), ki temeljijo na razširitvi mere *AUC*.

3.2 Binarna logistična regresija

Regresijska analiza proučuje odnos med odvisno in eno ali več neodvisnimi (pojasnjevalnimi) spremenljivkami. Za preučevanje tega odnosa se uporablja regresijski model, s pomočjo katerega napovedujemo vrednosti odvisne spremenljivke na podlagi ene ali več neodvisnih spremenljivk. Pri tem gre za prilaganje ustrezne matematične funkcije empiričnim podatkom.

Glede na število spremenljivk ločimo:

- **bivariatna regresija:**

regresijski model vsebuje samo dve neodvisni spremenljivki;

- **multivariatna regresija:**

regresijski model vsebuje več kot dve neodvisni spremenljivki;

Glede na vrsto odvisne spremenljivke ločimo:

- **navadna regresija:**

odvisna spremenljivka je numerična in normalno porazdeljena;

- **logistična regresija;**

odvisna spremenljivka je kategorialna in binomsko porazdeljena.

Glede na obliko ločimo:

- **linearna regresija:**

regresijski model ima obliko linearne funkcije;

- **nelinearna regresija:**

regresijski model ima obliko nelinearne funkcije (kvadratna, kubična, eksponentna ...).

Logistična regresija je oblika regresijske analize, ki jo lahko uporabimo, če je odvisna spremenljivka kategorialna (npr. poročen, samski, izvenzakonska skupnost ...). Pri binarni logistični regresiji je odvisna spremenljivka binarna oz. dihotomna, kar pomeni, da zavzema 2 različni vrednosti.

Pri uporabi binarne logistične regresije je potrebno upoštevati osnovne predpostavke in poznati parametre, ki omogočajo evalvacijo modela. Če torej želimo na podatkih izvesti binarno logistično regresijo, morajo ti zadostiti štirim predpostavkam:

1. dihotomija:

odvisna spremenljivka modela mora biti dihotomna. To pomeni, da lahko zavzame le 2 različni vrednosti (v našem primeru 0 in 1 oz. plačnik in neplačnik);

2. neodvisne spremenljivke:

model lahko vključuje eno ali več neodvisnih spremenljivk, ki so lahko bodisi zvezne (intervalne ali takšne, ki podajo neko razmerje (*ang. ratio variables*)) bodisi kategorialne (nominalne ali ordinalne oz. vrstilne);

3. neodvisnost:

podatki v modelu morajo biti neodvisni in odvisna spremenljivka mora imeti enolične kategorije, ki pokrivajo vse možne vrednosti in se med seboj ne prekrivajo. To pomeni, da morajo vrednosti odvisne spremenljivke podatke razdeliti na particije množic;

4. linearnost:

med zvezno neodvisnimi spremenljivkami ter logaritemsko transformirano odvisno spremenljivko mora obstajati linearna zveza. Pri preoblikovanju odvisne spremenljivke lahko uporabimo naravno logaritemsko transformacijo.

Opazimo, da je ena izmed prednosti logistične regresije ta, da ne zahteva homoskedastičnosti in normalne porazdeljenosti za odvisno spremenljivko.

3.2.1 Regresijska funkcija in model

Regresijski model izpeljemo iz logistične funkcije

$$F(t) = \frac{e^t}{1 + e^t} = \frac{1}{1 + e^{-t}}, \quad (3.16)$$

kjer parameter $t \in \mathbb{R}$ izrazimo kot linearno kombinacijo pojasnjevalnih spremenljivk x_i in dobimo

$$F(X_i) = \frac{1}{1 + e^{-g(X_i)}}, \quad (3.17)$$

kjer je

$$g(X_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_n X_{n,i}, \quad (3.18)$$

kjer je β_i (parcialni) regresijski koeficient i -te pojasnjevalne spremenljivke. Dobljena funkcija predstavlja pogojno verjetnost, da bo odvisna spremenljivka glede na dane pojasnjevalne spremenljivke zavzela vrednost $i = 0, 1$: $F(X_i) = P(Y = i | X \leq x) = p_i$.

Inverz logistične funkcije predstavlja populacijski regresijski model

$$g(X_i) = \ln \frac{F(X_i)}{1 - F(X_i)} = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_n X_{n,i}, \quad (3.19)$$

iz katerega izpeljemo tako imenovane obete

$$O_i = \frac{F(X_i)}{1 - F(X_i)} = \frac{p_i}{1 - p_i} = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_n X_{n,i}}. \quad (3.20)$$

Regresijski koeficienti se v logistični regresiji ocenjujejo neposredno z metodo največjega verjetja, ki maksimizira rezultirajočo pogojno porazdelitveno funkcijo oz. funkcijo verjetja za zbir vzorčnih vrednosti y_1, \dots, y_n . Ker je odvisna spremenljivka Y dihotomna, ima Bernoulijevo (binomsko) porazdelitev pri vsakem kovariatnem vzorcu s povprečjem p_i in varianco $p_i(1 - p_i)$. Tako je funkcija verjetja enaka

$$L = \prod_{i=1}^n F(X_i)^{y_i} (1 - F(X_i))^{1-y_i}. \quad (3.21)$$

Naravni logaritem funkcije verjetja spremeni produkt v vsoto in tako dobimo

$$\begin{aligned} \ln L &= \sum_{i=1}^n y_i \ln F(X_i) + (1 - y_i) \ln(1 - F(X_i)) \\ &= \sum_{i=1}^n \ln(1 - F(X_i)) + \sum_{i=1}^n y_i \frac{F(X_i)}{1 - F(X_i)} \\ &= \sum_{i=1}^n \ln(1 - F(X_i)) + \sum_{i=1}^n y_i (\beta_0 + \beta_1 X_{1,i} + \dots + \beta_n X_{n,i}) \\ &= \sum_{i=1}^n -\ln(1 + e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_n X_{n,i}}) + \sum_{i=1}^n y_i (\beta_0 + \beta_1 X_{1,i} + \dots + \beta_n X_{n,i}). \end{aligned}$$

Če torej želimo dobiti ceneilke regresijskih koeficientov, moramo najprej logaritmirano funk-

cijo verjetja odvajati

$$\begin{aligned}\frac{\partial \ln L}{\partial \beta_j} &= -\sum_{i=1}^n \frac{1}{1+e^{g(X_i)}} e^{g(X_i)} X_{i,j} + \sum_{i=1}^n y_i X_{i,j} \\ &= \sum_{i=1}^n (y_i - F(X_i)) X_{i,j}\end{aligned}$$

in parcialne odvode enačiti z 0

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_{i=1}^n (y_i - F(X_i)) X_{i,j} = 0. \quad (3.22)$$

Nato rešimo sistem dobljenih $n + 1$ linearnih enačb z $n + 1$ neznankami in tako dobimo cenilke b_0, \dots, b_n za regresijske koeficiente.

Kot že omenjeno, logistična regresija kot enega izmed rezultatov poda oceno verjetnosti neplačila kreditojemalca oz. neko vrednost med 0 in 1. Katerega komitenta bomo na podlagi dobljene verjetnosti uvrstili med neplačnike, je odvisno od izbire mejne vrednosti M . Tako vse komitente, za katere velja $F(X_i) \geq M$, uvrstimo med bodoče neplačnike, vse ostale komitente pa uvrstimo med bodoče plačnike. S spremjanjem mejne vrednosti se spremeni tudi število bodočih plačnikov ter neplačnikov. Večji kot je M , manj komitentov bomo uvrstili med neplačnike ter več med plačnike in obratno. Posledično se spreminja tudi natančnost napovedovanja.

S spremjanjem mejne vrednosti lahko tako vplivamo na natančnost napovedi modela. Pri tem pa se moramo vprašati, ali nam je pomembnejša natančnost napovedi neplačnikov oz. senzitivnost ali natančnost napovedi plačnikov oz. specifičnost, saj v praksi zelo redko dobimo 100 % natančnost obeh napovedi. Odgovor na to vprašanje je seveda senzitivnost. V praksi se namreč v bančnem portfelju nahaja le majhen¹⁰ odstotek komitentov, ki so neplačniki in tako zlahka dosežemo zelo visoko specifičnost. Namreč, če imamo v portfelju le 2 % neplačnikov, in če vse komitente razglasimo za plačnike, dobimo 100 % specifičnost in 98 % natančnost modela. Ker pa model ne prepozna niti enega izmed neplačnikov, seveda nikakor ni ustrezен, saj ima banka večjo izgubo, če komitent ne more odplačati posojila, kot pa, če ne prejme dobička od sklenjene pogodbe.

Zgled. S pomočjo binarne logistične regresije na zgodovinskih podatkih izračunamo, kakšna je verjetnost, da bo nek komitent postal neplačnik. Ker gre za že pretekle dogodke, imamo v tabeli podane tudi dejanske razvrstitve komitentov med plačnike (0) in neplačnike (1). Rezultati regresije so podani v spodnji tabeli.

¹⁰Če bi odstotek bil velik, bi to pomenilo, da banka nima ustrezne metodologije dodeljevanja posojil komitentom.

ID	$F(X_i)$	Status
1	0,4	0
2	0,6	1
3	0,5	0
4	0,1	0
5	0,8	1

Slika 3.9: Ocena verjetnosti neplačila kreditojemalcev

1. $M = 0,5$

Na podlagi binarne logistične regresije med neplačnike uvrstimo tri od petih komitentov. Tako s 100 % natančnostjo pravilno napovemo neplačnika in s 66,7 % plačnika. Skupna natančnost napovedi modela je enaka 80 %.

2. $M = 0,6$

Med neplačnike uvrstimo dva od petih komitentov. Tako s 100 % natančnostjo pravilno napovemo neplačnika in s 100 % plačnika. Skupna natančnost napovedi modela pa je enaka 100 %.

3. $M = 0,4$

Med neplačnike uvrstimo štiri od petih komitentov. Tako s 100 % natančnostjo pravilno napovemo neplačnika in s 33,3 % plačnika. Skupna natančnost napovedi modela pa je enaka 60 %.

Kot lahko vidimo, je v tem primeru najboljša izbira za mejno vrednost $M = 0,6$. Ker pa v praksi v portfelju nimamo le 5 komitentov, ampak mnogo več, izbira idealne mejne vrednosti ni tako očitna.

3.2.2 Določitev optimalne mejne vrednosti

Na natančnost napovedi modela lahko vplivamo s spremenjanjem mejne vrednosti M . Ker pa imamo v praksi opravka z velikim številom komitentov v portfelju, optimalna vrednost ni razvidna na prvi pogled. Zelo redko se tudi zgodi, da lahko določimo takšno mejno vrednost, da bi dobili 100 % natančnost napovedi modela, saj se zaradi (ne)natančnosti modela verjetnosti lahko pojavi situacija, da ima kak plačnik, ki je zelo podoben skupini neplačnikov, lahko višjo verjetnost neplačila od kakega neplačnika, katerega lastnosti so zelo podobne skupini plačnikov. Če torej želimo določiti optimalen M , moramo najprej razumeti, kaj se dogaja z izgubami oz. dobičkom bančnih posojil ter kakšen kriterij bomo upoštevali.

V bančni praksi je običajno tako, da banka izgube, ki jih utrpi zaradi neplačevanja obveznosti, pokrije z dobičkom, ki ga ustvari s posli, kjer komitent poravna ves dolg. Poleg pokritja izgub pa ima banka še običajno nek dobiček. Maksimalen dobiček bi banka imela v primeru, ko bi vsi komitenti poravnali svoj dolg. Če bi torej želeli maksimirati dobiček, bi mejo nastavili vedno tako visoko, da bi bili vsi komitenti razglašeni za plačnike. To pa v praksi seveda ni mogoče, saj imajo banke v svojem portfelju običajno vedno nek majhen odstotek komitentov, ki so neplačniki.

Kot vidimo, maksimiranje dobička ni ustrezan pristop za določitev mejne vrednosti. Ker želimo z bonitetnim modelom čim bolje napovedati, kateri komitent bo postal plačnik in kateri ne, je najboljši kriterij za določitev M maksimiranje natančnosti modela, pri pogoju, da je senzitivnost višja od specifičnosti, saj model razvijamo ravno iz tega razloga, da smo zmožni prepoznati, kdo svojega dolga ne bo zmožen poravnati.

Postopek izračuna

Naj bo K množica vseh komitentov. Ker imamo opravka z zgodovinskimi podatki za vsakega komitenta, vemo tudi, ali je plačnik ali neplačnik. Ta podatek označimo z y_i . Nato izvedemo logistično regresijo in za vsakega komitenta dobimo verjetnost neplačila p_i . Sedaj določimo neko vrednost za M (na primer $M = 0,5$), ki bo določila, kateri komitent bo klasificiran kot plačnik in kateri kot neplačnik (s_i).

Iščemo takšen M , da bo natančnost napovedi modela maksimalna, pri pogoju, da je senzitivnost večja ali enaka specifičnosti: $\max(\frac{TP+TN}{P+N})$ p.p. $\frac{TP}{P} \geq \frac{TN}{N}$. Spomnimo se, da TP predstavlja število pravilno napovedanih neplčanikov, TN število pravilno napovedanih plačnikov, P število vseh pozitivnih vrednosti (v našem primeru št. vseh neplačnikov) in N število vseh negativnih vrednosti (v našem primeru št. vseh plačnikov).

3.3 Klasifikacijska drevesa

Klasifikacijsko ali odločitveno drevo (*ang. decision tree*) predstavlja klasifikacijski model, ki je zelo enostaven za razumevanje in ima, kot že pove samo ime, strukturo drevesa. Sestavlja ga:

- **notranja vozlišča:**

med notranja vozlišča sodijo vsa vozlišča, ki se razdelijo na dva ali več drugih vozlišč na podlagi nekih atributov. Sem spada tudi koren, ki predstavlja prvo oz. začetno vozlišče v drevesu in se v njem nahaja celoten portfelj oz. vzorec, ki ga želimo klasificirati. V vsakem notranjem vozlišču se opravi preizkus, ki preveri vrednost enega ali več

neodvisnih spremenljivk (atributov oz. argumentov funkcije). V prvem primeru tako govorimo o enorazsežnih, v drugem pa o večrazsežnih drevesih. Preizkus pa ima lahko dva (binarna drevesa) ali več izidov;

- **veje:**

veje predstavljajo vrednosti atributov (npr. plačnik, neplačnik);

- **listi:**

listi predstavljajo posamezen razred klasifikacije. Tako vsak list vsebuje predpis, ki odvisni spremenljivki priredi razred oz. vrednost iz bodisi množice kod (nominalne vrednosti) bodisi iz množice realnih števil. V prvem primeru govorimo o razvrščevalnih, v drugem pa o regresijskih ali modelskih drevesih.

3.3.1 Razvrščanje vzorcev

Pot iz korena proti listom predstavlja zaporedje vmesnih odločitev, ki jih sprejemamo pri razvrščanju vzorcev. Za attribute, podane z numeričnim zapisom, vsaka delna odločitev temelji na tem, ali je vrednost atributa večja ali manjša od naprej podane mejne vrednosti. Takšno razvrščanje vzorcev v razrede ustvari ločilne meje v prostoru statističnih enot, ki so vzporedne z osmi prostora.

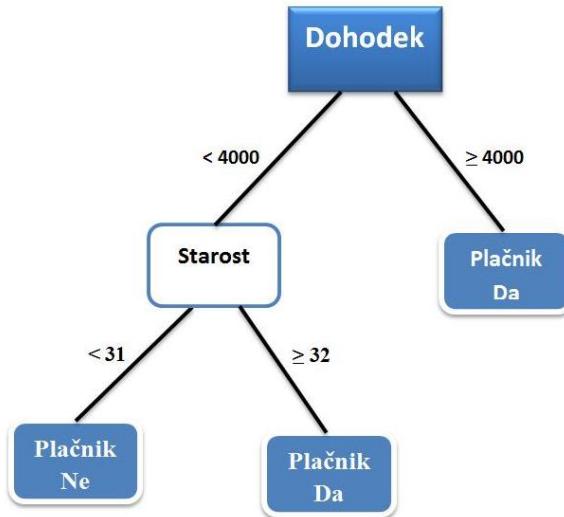
Zgled. Na sliki 3.10 imamo podan vzorec, kjer je starost podana v letih in dohodek v evrih.

ID	Spol	Starost	Dohodek	Plačnik
1	Ž	32	1000	Da
2	Ž	52	10000	Da
3	Ž	26	1500	Ne
4	Ž	54	1500	Da
5	M	25	4000	Da
6	M	49	7000	Da

Slika 3.10: Primer vzorca X

Poiskati želimo pravilo, ki bo ločilo plačnike od neplačnikov glede na vrednosti danih atributov. Izdelamo lahko razvrščevalno odločitveno drevo, ki ga prikazuje slika 3.11.

Tako dobimo naslednje pravilo za razvrščanje:



Slika 3.11: Primer odločitvenega drevesa

- če komitent prejema več kot 4000 € dohodka, je plačnik;
- če komitent prejema manj kot 4000 € dohodka in je star 31 let ali več, je plačnik;
- če komitent prejema manj kot 4000 € dohodka in je star 30 let ali manj, ni plačnik.

Pri gradnji odločitvenih dreves je pomembno, kako dobro posamezni atributi ločujejo primere iz različnih razredov. Odgovor na to vprašanje nam podata naslednji meri:

- mera za dobro ločevanje primerov, ki ji pravimo entropija;
- mera kvalitete atributa, ki ji pravimo informacijski prispevek (*ang. information gain*).

3.3.2 Entropija

Entropija pomeni stopnjo degeneriranosti sistema v smeri kaosa. V matematiki predstavlja količino, ki meri negotovost izida poskusa, povezanega s slučajno spremenljivko. Ta vrsta entropije določa za pričakovano vrednost količino informacije, ki jo pridobimo takrat, ko izvedemo poskus, in dobimo njeno vrednost. Torej je merilo za količino informacije, ki jo dobimo s poznavanjem vrednosti slučajne spremenljivke. Označimo jo s H .

Če imamo klasifikacijski atribut Y , ki lahko zavzame diskretne vrednosti v_1, v_2, \dots, v_k , je entropija zanj enaka:

$$H(Y) = \sum_{v \in Z_Y} P_v(X) I(X) = - \sum_{v \in Z_Y} P_v(X) \log_b P_v(X), \quad (3.23)$$

kjer

- se $P_v(X)$ izračuna po formuli 3.1;
- je $I(X) = -\log_b P_v(X)$;
- je b osnova logaritma. Kadar je enota za entropijo bit, je b enak 2.

Problem pri računanju entropije nastopi, kadar je $p = 0$. V tem primeru definiramo, da je entropija enaka nič.

Zgled. Dan imamo vzorec oz. portfelj, v katerem je polovica komitentov plačnikov in polovica neplačnikov. Pri naključni izbiri enega komitenta iz portfelja velja, da je verjetnost, da bo izbrani komitent plačnik (p_0), enaka verjetnosti, da bo izbrani komitent neplačnik (p_1), ki je enaka 0,5. Torej velja $p_0 = p_1 = 0,5$.

$$\begin{aligned} H &= -(p_0 \cdot \log_2 p_0 + p_1 \cdot \log_2 p_1) \\ &= -2 \cdot p_0 \cdot \log_2 p_0 \\ &= -2 \cdot 0,5 \cdot \log_2 2^{-1} \\ &= \log_2 2 \\ &= 1. \end{aligned}$$

To pomeni, da nam vsaka naključna izbira komitenta da 1 bit informacije.

Če naključno izbiramo dva komitenta, imamo štiri možne izide:

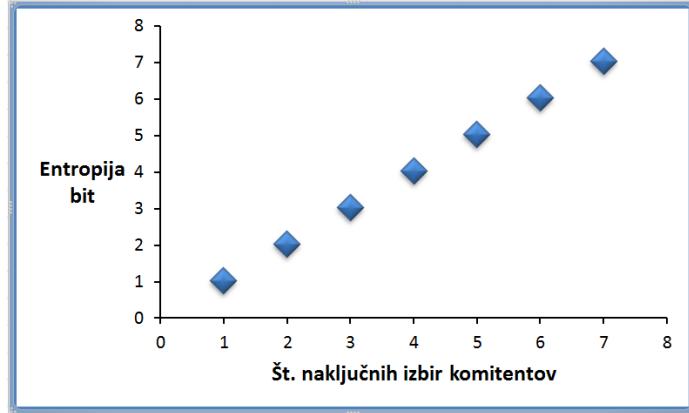
- izbrali smo dva plačnika;
- izbrali smo dva neplačnika;
- prvi izbrani je plačnik, drugi pa neplačnik;
- prvi izbrani je neplačnik drugi pa plačnik.

Verjetnost vsakega izida je 0,25. V tem primeru ima entropija vrednost 2 bita. Pri 20 zaporednih izbirah je verjetnost vsakega izida enaka 2^{-20} , entropija pa 20 bitov.

Primer je popolnoma drugačen, če razmerje med plačniki in neplačniki ni enako 1:1 oz. če p_0 in p_1 nista enaki (vsak izid izbire ni enako verjeten). Recimo, da je verjetnost naključne izbire plačnika 0,6 in verjetnost naključne izbire neplačnika 0,4. Velja $p_0 + p_1 = 1$. Tako dobimo entropijo

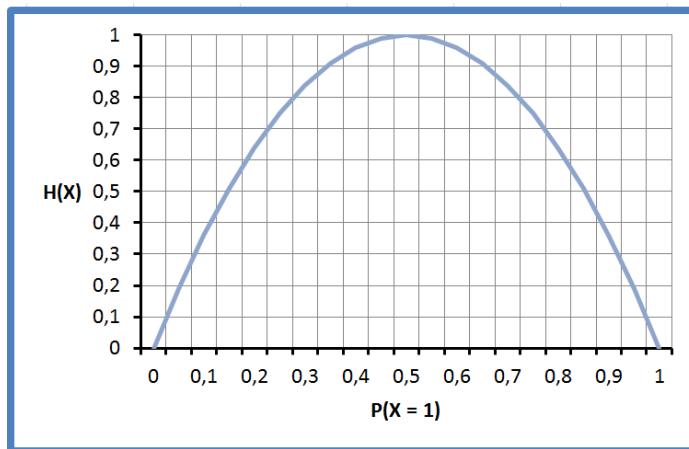
$$H = -(p \cdot \log_2 p + q \cdot \log_2 q) = 0,97$$

kjer je q enak $1 - p$. V tem primeru imamo večjo negotovost, s tem pa tudi zmanjšano entropijo oziroma vsaka naključna izbira komitenta nam da manj kot 1 bit informacije.



Slika 3.12: Povezava med entropijo in št. naključnih izbir komitentov iz portfelja, kjer je razmerje med plačniki in neplačniki enako 1:1

Če bi v portfelju imeli le plačnike (ali neplačnike), bi bila entropija enaka nič. Naključna izbira komitenta nam ne bi dala nobene informacije, saj je izid izbire vedno znan oziroma enak, kar lahko vidimo na spodnji sliki, ki prikazuje entropijo pri naključni izbiri komitenta v odvisnosti od razmerja med plačniki in neplačniki oz. verjetnostjo izbire plačnika (ali neplačnika). Vidimo, da ima funkcija maksimum in sicer je entropija največja pri verjetnosti 0,5, ko imamo v portfelju enako število plačnikov in neplačnikov.



Slika 3.13: Entropija v odvisnosti od verjetnosti izbire neplačnika

3.3.3 Informacijski prispevek

Informacijski prispevek atributa nam pove, koliko informacije o razredu nosi nek atribut. Uporablja se pri gradnji odločitvenih dreves, kjer množico primerov razdelimo glede na vrednosti nekega atributa in s pomočjo informacijskega prispevka iz množice vseh atributov izberemo tistega, ki najbolj zmanjša entropijo.

Definiran je z

$$Gain(Y, A) = H(Y) - \sum_{a \in Z_A} \frac{|X_a|}{|X|} \cdot H(a), \quad (3.24)$$

$$H(a) = - \sum_{v \in Z_Y} P_v(X_a) \log_b P_v(X_a), \quad (3.25)$$

kjer se $P_v(X_a)$ izračuna po formuli 3.2.

Zgled. Na sliki 3.14 imamo dan portfelj komitentov, ki opisuje, ali so komitenti plačniki ali neplačniki.

Stanovanjski status	Status osebe	Izobrazba	Plačnik
lastnik	študent	višješolska	da
najemnik	zaposlen	srednješolska	da
lastnik	zaposlen	višješolska	da
pri starših	študent	srednješolska	ne
lastnik	študent	srednješolska	ne
najemnik	študent	srednješolska	ne

Slika 3.14: Primer portfelja komitentov

Klasifikacijski atribut je v tem primeru atribut plačnik, ki ima dve klasifikacijski vrednosti: da in ne. Zanima nas, kateri atribut je najbolj informativen oz. kateri nosi največ informacije o razredu neplačnikov oz. plačnikov. Torej moramo izračunati informacijski prispevek vsakega atributa in izberemo tistega, ki bo imel največjega.

Najprej izračunamo verjetnosti, da je komitent plačnik oziroma neplačnik:

$$P(da) = \frac{3}{6} = \frac{1}{2} \quad \text{in} \quad P(ne) = \frac{3}{6} = \frac{1}{2},$$

nato pa entropijo naših podatkov

$$\begin{aligned} H(Y) &= -[P(da) \cdot \log_2(P(da)) + P(ne) \cdot \log_2(P(ne))] \\ &= -\left(\frac{1}{2} \cdot \log_2 \frac{1}{2} + \frac{1}{2} \cdot \log_2 \frac{1}{2}\right) \\ &= 1. \end{aligned}$$

Sedaj lahko izračunamo informacijske prispevke vseh atributov. Najprej poglejmo atribut izobrazba, ki ima dve različni vrednosti: srednješolska (*ss*) in višješolska (*vs*).

$$Gain(Y, izobrazba) = H(Y) - \sum_{a \in \{ss, vs\}} \frac{|X_a|}{|X|} \cdot H(a)$$

Od tega imamo s srednješolsko izobrazbo 1 plačnika in 3 neplačnike

$$H(ss) = -\left(\frac{1}{4} \cdot \log_2 \frac{1}{4} + \frac{3}{4} \cdot \log_2 \frac{3}{4}\right) \approx 0,811$$

z višješolsko izobrazbo pa 2 plačnika in 0 neplačnikov

$$H(vs) = -\left(\frac{2}{2} \cdot \log_2 \frac{2}{2} + \frac{0}{2} \cdot \log_2 \frac{0}{2}\right) = 0.$$

Tako dobimo informacijski prispevek

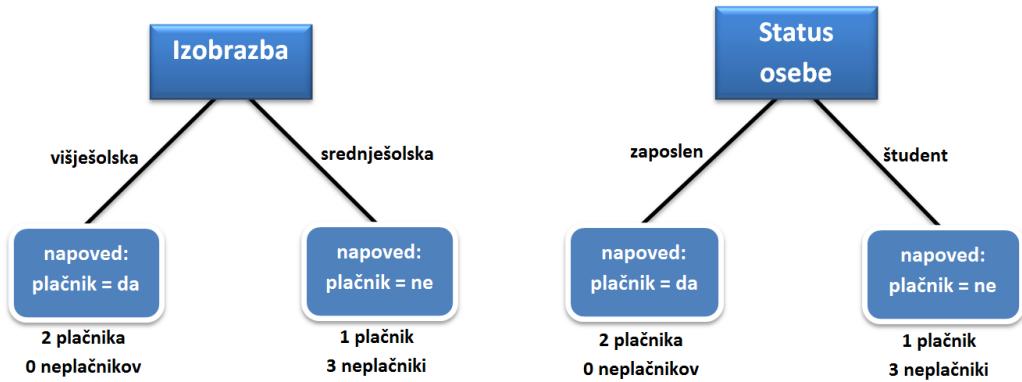
$$Gain(Y, izobrazba) = 1 - \left(\frac{2}{6} \cdot 0 + \frac{4}{6} \cdot 0,811\right) \approx 0,46$$

Če postopek izračuna ponovimo še za atributa status osebe (*statO*) in stanovanjski status (*stanS*), dobimo:

$$Gain(Y, statO) \approx 0,46 \quad \text{in} \quad Gain(Y, stanS) \approx 0,21.$$

Ker imata atributa izobrazba in status osebe enak informacijski prispevek, lahko napravimo dve enakovredni odločitveni drevesi.

Opazimo lahko, da kazalnika izobraba in status osebe tako na podlagi ROC analize kot tudi na podlagi izračuna informacijskega prispevka najbolje ločita plačnike od neplačnikov.



Slika 3.15: Prvi nivo enakovrednih odločitvenih dreves

3.4 Korelacijska analiza

Korelacijska analiza proučuje soodvisnost (oz. povezanost ali stopnjo usklajenosti) med dvema (ali več) statističnima spremenljivkama, ne da bi katero od njiju smatrali za neodvisno ali odvisno¹¹. Pri osnovni korelacijski analizi oziroma analizi kovariance preizkušamo, kako močno sta dve statistični spremenljivki X in Y na populaciji med seboj povezani ali korelirani.

Mera za medsebojno povezanost spremenlivk je kovarianca

$$K(X, Y) = E(XY) - E(X)E(Y),$$

jakost linearne povezanosti pa merimo z brezdimenzijskim korelacijskim koeficientom $\rho(X, Y)$. V splošnem ima korelacijski koeficient lastnost, da sta statistični spremenljivki X in Y nekorelirani natanko tedaj, ko je $\rho(X, Y) = 0$ oz. $E(XY) = E(X)E(Y)$. Kadar pa je $\rho(X, Y) \neq 0$, sta spremenljivki X in Y korelirani.

Korelacijsko je moč meriti z več različnimi koeficienti, prilagojenimi za različne tipe podatkov, ki so na voljo. Izmed korelacijskih koeficientov je najbolj znan Pearsonov korelacijski koeficient, ki je računan na podlagi kovariance in standardnih odklonov (σ) serij obeh spremenljivk:

$$\rho(X, Y) = \frac{K(X, Y)}{\sigma(X)\sigma(Y)},$$

¹¹Soodvisnost med dvema statističnima spremenljivkama lahko povzroča tudi kakšna tretja (ali več) spremenljivka.

Pogoj za računanje tega koeficenta je linearna odvisnost obeh vpletenih spremenljivk. Enačbo, ki najbolje opisuje linearno odvisnost obeh spremenljivk, je moč izračunati z linearno regresijo, ki je najbolj natančna ob visokih vrednostih korelacijskega koeficenta. Za določanje povezanosti spremenljivk, ki niso povezane linearno, pa se uporablja Spearmanov koeficient korelacije.

3.4.1 Spearmanov koeficient korelacije

Spearmanov koeficient korelacije je statistični kazalec, ki prikazuje neparametrske stopnje povezanosti dveh spremenljivk oziroma predstavlja kakovost opisa povezanosti med spremenljivkama, ki jo napravi določena funkcija. Pri tem za razliko od Pearsonovega koeficenta korelacije ne predpostavlja linearne povezanosti spremenljivk in enakomernosti frekvenčne porazdelitve, prav tako pa je z njim moč računati stopnjo povezanosti neveznih spremenljivk. Zaradi slednjega je Spearmanov koeficient primeren za izračun odvisnosti ordinalnih spremenljivk.

Spearmanov koeficient korelacije je v bistvu posebna oblika Pearsonovega koeficenta, kjer so podatki pred izračunom koeficientov preoblikovani v range¹². Za testiranje neodvisnosti ga lahko uporabimo, če sta X in Y vsaj ordinalni spremenljivki.

Naj imata statistični spremenljivki X in Y na vzorcu velikosti n vrednosti

$$X_1, X_2, \dots, X_n \quad \text{in} \quad Y_1, Y_2, \dots, Y_n.$$

Dane vrednosti rangiramo tako, da I_k označuje rang X_k in J_k rang Y_k , $D_k = I_k - J_k$ pa naj bo razlika rangov. Spearmanov korelacijski koeficient je število

$$R_S = 1 - \frac{6}{n^3 - n} \sum_{k=1}^n D_k^2,$$

ki zavzame vrednosti z intervala $[-1, 1]$. Podobno kot pri Pearsonovem korelacijskem koeficientu tudi tukaj velja, da sta statistični spremenljivki X in Y korelirani, če je $R_S \neq 0$ ter nista korelirani, če $R_S = 0$. Večja, kot je absolutna vrednost koeficenta, bolj močno sta spremenljivki korelirani.

Bolj zahtevna oblika enačbe nastane pri računanju zveznih rangov, a je v večini primerov napaka zaradi neupoštevanja izravnava majhna.

¹²V preprostejši obliki računanja koeficenta so absolutne vrednosti oblikovane v range že poprej, nato pa se za končni izračun uporabi razlika med rangi različnih vrednosti, katerih medsebojna odvisnost se išče.

Poglavlje 4

Bonitetni model za slovensko prebivalstvo

Z uporabo IRB pristopa bomo razvili poskusni bonitetni model za ocenjevanje kreditnega tveganja, prilagojenega slovenskemu prebivalstvu. Namen modela je študijske narave, pri njegovem razvoju pa bomo uporabili enoletne zgodovinske podatke ene izmed slovenskih poslovnih bank, na podlagi katerih bomo razvili model, ki bo napovedal, ali bo nek komitent zmožen odplačati kredit (bo ostal plačnik) ali ne (bo postal neplačnik).

Pri gradnji modela moramo upoštevati smernice oz. zahteve kapitalskega sporazuma Basel II. Tako mora biti model predvsem razumljiv, učinkovit, kalibriran na verjetnost neplačila in empirično preverjen. Model, ki ga bomo razvili v tej nalogi, je kombinacija dveh statističnih metod, in sicer razvoja odločitvenega drevesa in uporabe binarne logistične regresije v listih drevesa.

Osnovo bonitetnega modela tako tvori statistično ocenjevanje verjetnosti neplačila, ki je sestavljeno iz kvantitativnega in kvalitativnega dela. Prvi in osnovni korak, ki vpliva na vse nadaljnje, pa je opredelitev dogodka neplačila, ki mora biti skladna z Basлом II. Šele, ko imamo določene jasne kriterije za določitev statusa neplačnika, lahko opredelimo finančne kazalnike in njihovo relacijo do neplačila¹.

Naslednji korak predstavlja izbira vzorca, na katerem se bo preizkušal model, ter ureditev, redukcija in po potrebi transformacija kazalnikov, kar običajno bistveno prispeva k izboljšavi modela. V nadaljevanju opredelimo najboljši model in izračunamo verjetnosti neplačila za bonitetne razrede. Na podlagi dobljenih verjetnosti neplačila lahko potem izračunamo še ostale parametre in zahtevano kapitalsko rezervo.

¹S tem zagotovimo intuitivnost vsakega kazalnika.

4.1 Tehnologije izdelave modela

Kot je bilo že omenjeno, si pri razvoju modela pomagamo s statističnim programskim paketom SPSS ter s programom za podatkovno rudarjenje imenovanim Rapid Miner.

4.1.1 SPSS

Statistični programski paket SPSS je namenjen predvsem za statistično obdelavo podatkov z osebnimi računalniki in temelji na operacijskem sistemu Windows. S pomočjo programa lahko izdelamo razne grafe (npr. histogram, stolpični/razsevni diagram), ROC krivulje ter izvedemo številne analize podatkov kot na primer različne oblike regresije (linearna, multi-kolinearna, logistična ...), korelacijsko analizo, univariatno analizo, multivariatno analizo ter testiranje statističnih hipotez.

Delo z SPSS-om poteka istočasno z večimi okni:

- okno s podatki (Data Editor),
- okno z rezultati analiz (Viewer),
- okno z navodili za izvajanje analiz (Syntax Editor).

V oknu Data Editor urejamo datoteko s podatki in njihovimi opisi. Za analize lahko uporabljamo bodisi podatke, ki so že pripravljeni za takšne analize (tabele s končnico sav) bodisi podatke prenesemo iz drugih programskih paketov (npr. Excel) ali jih vnesemo ročno. Okno Data Editor ima dva pogleda:

- pogled Data View je namenjen vstavljanju in urejanju podatkov;
- pogled Variable View je namenjen vstavljanju in urejanju spremenljivk.

V tabeli s podatki vsaka enota vedno zaseda eno vrstico in vsaka spremenljivka en stolpec. Vsaki spremenljivki je potrebno določiti:

- **ime spremenljivke:** ime spremenljivke je sestavljeno iz enega do osem znakov, pri čemer mora biti prvi znak velika ali mala črka, zadnji pa ne sme biti pika ali podčrtaj. V imenu spremenljivke prav tako ne sme biti posebnih znakov, kot je npr. presledek. SPSS ne pozna razlike med velikimi in malimi črkami;
- **tip spremenljivke:** spremenljivka je lahko numerična ali opisna;

- **število mest in število decimalnih mest**, ki jih podatki zasedajo;
- **labelo**: labela je dodatno pojasnilo o pomenu imena in vrednosti spremenljivke;
- **labele vrednosti** (če jih potrebujemo): opisnim spremenljivkam včasih priredimo številske vrednosti zaradi lažje obdelave;
- **manjkajoče vrednosti**: če za nekatere enote nimamo podatka, namesto neznane manjkajoče vrednosti vnesemo neko vnaprej določeno izbrano vrednost, ki jo v SPSS-u definiramo kot uporabniško manjkajoče vrednost. Ta vrednost mora biti različna od vseh vrednosti, ki jih spremenljivka lahko zavzame. Če ne določimo manjkajoče vrednosti, SPSS določi svojo manjkajoče vrednost, ki se imenuje sistemska manjkajoča vrednost;
- **mersko lestvico**: ordinalna (*ang. ordinal*), nominalna (*ang. nominal*), intervalna in razmerostna (*ang. scale*). SPSS ne pozna razlike med intervalnimi in razmerostnimi spremenljivkami, oba tipa pozna kot scale.

4.1.2 Rapid Miner

RapidMiner je programska platforma, ki zagotavlja integrirano okolje za strojno učenje, podatkovno in besedilno rudarjenje ter napovedne in poslovne analitike. Uporablja se tako za poslovne in industrijske namene, kot tudi za raziskave, izobraževanje, usposabljanje, hitro izdelavo prototipov in razvoj aplikacij. Podpira vse korake procesa podatkovnega rudarjenja, vključno z vizualno predstavitevijo podatkov, validacijo in optimizacijo.

Rapid Miner je napisan v programskem jeziku Java in je opremljen z grafičnim uporabniškim vmesnikom (GUI - *ang. graphical user interface*) za načrtovanje in izvedbo analitičnih delovnih tokov (*ang. workflows*) oz. procesov. Vsak proces sestavlja eden ali več operatorjev. Vsak operator izvaja specifično nalogo v procesu in izhod vsakega operatorja tvori vnos naslednjega. Ustvarjen proces se lahko uporablja tudi v drugih programih ali kot API vmesniki (*ang. application programming interface*) oziroma programske knjižnice (v visokonivojskih programskeh jezikih), ki jim omogočajo enostavnejšo in hitrejšo implementacijo grafičnih uporabniških vmesnikov.

Pri izdelavi odločitvenega drevesa v Rapid Minerju kot mero za gradnjo odločitvenega drevesa uporabimo informacijski prispevek, saj dobimo tako na preizkušenih vzorcih najboljše rezultate pri natančnosti napovedi modela. Poleg te mere so na voljo še mere relativen informacijski prispevek (*ang. gain ratio*), Gini indeks in natančnost (*ang. accuracy*), ki pa so se v večini primerov izkazale kot neuporabne, saj program ni izdelal odločitvenega drevesa.

4.2 Definicija dogodka neplačila

Kot je bilo že omenjeno, se mora opredelitev dogodka neplačila skladati z opredelitvijo kapitalskega sporazuma Basel II, ki pravi, da neizpolnitev dolžniških obveznosti nastopi takrat:

- kadar je malo verjetno, da bo dolžnik banki sposoben odplačati svoj dolg, brez da bi zastavil zavarovanje;
- kadar ima dolžnik več kot 90-dnevno zamudo pri izpolnitvi bistvenih kreditnih obveznosti.

Poleg enotne definicije lahko vsaka banka uporablja tudi dodatne, bolj konservativne kriterije. Tako smo v magistrski nalogi uporabili definicijo neplačila banke X, ki pravi, da neplačilo nastopi, kadar se zgodi vsaj eden izmed naslednjih dogodkov:

1. Obstaja majhna verjetnost, da bo dolžnik poravnal svoje kreditne obveznosti, ne da bi bilo za poplačilo potrebno uporabiti ukrepe, kot je unovčenje zavarovanja. Slednje nastopi v naslednjih primerih:
 - izpostavljenost se obravnava kot nedonosna (*ang. non-accrued loan*);
 - izvedba oslabitev, ki so posledica zaznave občutnega poslabšanja kreditne kvalitete dolžnika;
 - prodaja izpostavljenosti s pomembno ekonomsko izgubo;
 - odobreno prestrukturiranje problematične izpostavljenosti, pri čemer se učinki pomembne opustitve ali odloga plačila odražajo v zmanjšani finančni obveznosti dolžnika;
 - vložitev predloga za uvedbo insolvenčnega postopka nad dolžnikom, ki lahko vodi k zmanjšanju ali odložitvi plačila obveznosti banki.
2. Dolžnik več kot 90 dni zamuja s plačilom pomembne kreditne obveznosti do banke;
3. Odpoved pogodbe zaradi neizpolnjevanja finančnih obveznosti.

4.3 Portfelj oziroma populacija

Prebivalstvo oz. fizične osebe, ki so komitenti banke X, predstavljajo razpršeni portfelj z množico majhnih kreditov, posamezna lastnost komitenta pa predstavlja t. i. finančni kazalnik. Ker je v celotnem portfelju zelo veliko število komitentov ter kazalnikov in tako

statistična obdelava, zaradi tehnoloških omejitev, ni možna oz. bi zahtevala preveliko časa, moramo za razvoj modela najprej izbrati reprezentativen vzorec, ki dobro predstavlja celotno populacijo. Pri tem je temeljno načelo slučajnosti oz. slučajno izbiranje elementov iz populacije v vzorec. Izbira vzorca vpliva na nadaljnje delo in dobljene rezultate. Velja, da lahko že najmanjša sprememba v vzorcu signifikantno vpliva na končni rezultat oz. obliko modela. Iz tega razloga je izbira vzorca oziroma vhodnih podatkov, ki tvorijo podatkovno bazo, zelo pomemben dejavnik za zagotavljanje kvalitete končnega modela. Vzorec mora biti tudi dovolj velik, da lahko omogoča natančno napovedovanje verjetnosti neplačila.

Ker v primeru kreditnega portfelja banke X lahko rečemo, da gre za celotno populacijo te banke, se načeloma lahko uporablja zgolj enostavno slučajno vzorčenje (SRS - *ang. simple random sampling*), ki je tudi najpreprostejša metoda. V praksi pa se uporabljajo tudi kompleksni vzorčni načrti (na primer proporcionalni stratificirani vzorci ipd.). Do določene mere lahko tudi vzorec naknadno popravimo z različnimi tehnikami uteževanja (na primer napihovanje).

Zaradi relativno majhnega števila neplačnikov (le 1 % v primerjavi s celotno populacijo), v vzorec vključimo vse neplačnike, dobre komitente oz. plačnike pa izberemo slučajno ali proporcionalno. Bolj kot je razvojni vzorec reprezentativen, bolj veljaven bo tudi model za celotno populacijo. Pri generiranju razvojnega vzorca za model uporabimo večje število vzorcev (*ang. multiple sampling*, glej Stephens, [33], str. 113), ki bi naj omogočili lažjo odločitev o izboru končnega modela. Tako pripravimo sledeče vzorce:

- **V1:** izberemo vse neplačnike ter naključno toliko plačnikov, da dobimo razmerje 1:9 (neplačniki : plačniki). Tako dobimo vzorec s 7890 komitenti. Z njim se želimo čim bolj približati razmerju v populaciji. Če bi želeli doseči enako razmerje, kot je v populaciji, bi to pomenilo, da bi imeli vzorec s 78900 komitenti, kar pa presega naše tehnološke omejitve². Če bi za osnovo izbrali manj neplačnikov, smo izpostavljeni večjemu tveganju, da vzorec ni reprezentativen za celotno populacijo oz. za vse neplačnike v populaciji.
- **V2:** izberemo vse neplačnike ter naključno izberemo toliko plačnikov, da dobimo razmerje 1:3 (neplačniki : plačniki). Tako dobimo vzorec s 3156 komitenti.
- **V3:** izberemo vse neplačnike ter naključno izberemo toliko plačnikov, da dobimo razmerje 1:2 (neplačniki : plačniki). Tako dobimo vzorec s 2367 komitenti.
- **TV:** za osnovo vzamemo enoletni bančni portfelj prejšnjega leta. Nato iz portfelja izločimo vse mladoletne osebe in komitente, ki nimajo zapolnjenih vseh podatkov pri

²Pri obdelavi velikih količin podatkov računalniški programi bodisi zamrznejo bodisi podatkov ne moremo niti naložiti v program.

kazalnikih, ki se uporabijo za razvoj modela. Tako dobimo testni vzorec z 85104 komitenti, od katerih je 1836 neplačnikov.

Najprej izvedemo pregled kazalnikov, kot je opisano v naslednjem poglavju. Nato na vseh vzorcih izdelamo odločitveno drevo in na podlagi najboljše napovedne moči izberemo razvojni vzorec, ki ga uporabimo za nadaljnjo analizo.

Statistične enote oz. komitenti, ki niso bili vključeni v razvojni vzorec, se načeloma uporabijo za testiranje robustnosti modela, t. i. prileganje modela enotam (komitentom) izven razvojnega vzorca (*ang. out-of-sample fit*). Ker pa smo vse neplačnike vključili v razvojni vzorec, to ni možno. Iz tega razloga bomo uporabili t. i. prileganje modela izven časa (*ang. out-of-time fit*).

4.3.1 Finančni kazalniki

Kot že omenjeno finančni kazalnik predstavlja posamezno lastnost komitenta (starost, stanovanjski status, dohodek ipd.), v statistiki pa kazalniki predstavljajo slučajne spremenljivke. Da bomo lahko razvili ustrezni bonitetni model, moramo najprej izbrati oz. določiti zanesljive in teoretično utemeljene spremenljivke. Kazalniki, ki vstopajo v model, morajo biti intuitivni oz. smiselni z vidika njihovih pričakovanih vplivov na neplačilo. To pomeni, da morajo biti napovedovalni, kar pomeni, da vsak kazalnik v modelu samostojno dobro razlikuje (diskriminira) med plačniki in neplačniki. Kazalniki morajo imeti tudi dovolj opazovanj (vrednosti) oz. malo manjkajočih vrednosti.

Za razvoj modela najprej, na podlagi kvalitativne analize (subjektivno, logično mišljenje) in literature, iz več kot 100 kazalnikov, ki so bodisi numerični bodisi kategorični, izberemo tiste, ki se pogosto pojavljajo v literaturi, ter za katere smo mnenja, da vplivajo na komitentovo zmožnost odplačevanja dolga. Tako dobimo nabor 23 kazalnikov (K_1, K_2, \dots, K_{23}), ki jih uporabimo za nadaljnjo analizo.

Zmanjšanje števila kazalnikov ima poleg praktičnih razlogov (boljša preglednost modela) tudi statistične, saj se pri velikem številu kazalnikov običajno pojavi multikolinearnost³, kar vpliva na nestabilnost modela.

Prvi korak obdelave podatkov predstavlja njihova anonimizacija, kar pomeni, da izbrišemo stolpec z imeni in priimki komitentov, ki pri razvoju modela ne igra nobene ključne vloge, saj je vsak komitent identificiran z identifikacijskim številom. Nato iz populacije izbrišemo

³Visoka korelacija med kazalniki. Dva visoko korelirana kazalnika ne doprineseta dodatne informacije za napovedovanje neplačila, koeficienti modela pa so nenatančni zaradi povečanja njihove standardne napake, kar implicira slabo out-of-sample prileganje modela.

vse komitente, ki so mlajši od 18 let, saj se tem komitentom kredit ne dodeljuje. Ker je v portfelju veliko praznih polj, moramo ta ustrezno zapolniti, saj bi na koncu, če bi upoštevali le komitente, ki imajo zapolnjena vsa polja, imeli prazen portfelj. Tako pri kazalnikih, ki imajo prazna polja in je to smiselno oz. ekvivalentno, prazna polja zapolnimo z vrednostjo 0.

Sledi ocenjevanje kvalitete kazalnikov in njihova redukcija. V ta namen uporabimo sledeča analitična orodja:

- **količina podatkov:** preden se lotimo vzorčenja, si najprej pogledamo, kolikšen delež podatkov posameznega kazalnika je oz. ni zapoljen. To pomeni, da izračunamo odstotek praznih polj za posamezen stolpec;
- **grafična analiza in diskretizacija:** s pomočjo grafov, ki ponazarjajo relacijo med kazalnikom in stopnjo neplačila, dobimo prvi vtis o tem, ali nek kazalnik dobro razločuje med plačniki in neplačniki ali ne. Tako posamezen kazalnik, če je to potrebno, razdelimo v smiselne kategorije oz. razrede⁴ in za vsak razred izračunamo njegovo stopnjo neplačila (*DR - ang. default rate*):

$$DR = \text{št. neplačnikov v razredu} / \text{št. vseh komitentov v razredu}.$$

Lahko bi izračunali tudi stopnje neplačila glede na izpostavljenost komitentov:

$$DR = \text{vsota izpostavljenosti neplačnikov v razredu} / \text{vsota izpostavljenosti vseh komitentov v razredu}.$$

Kazalniki, pri katerih stopnja neplačila po razredih bodisi narašča bodisi pada, kažejo na dobro napovedno moč. Ker pa gre le za prvi vtis, je za natančnejšo oceno potrebna še dodatna analiza;

- **merjenje korelacije:** neparametrična korelacija pokaže povezanost med posameznim kazalnikom in odvisno spremenljivko. Za izračun korelacije smo uporabili Spearmanov koreacijski koeficient. Bolj kot je kazalnik koreliran z odvisno spremenljivko, večja je verjetnost, da bo dobro razločeval med plačniki in neplačniki;
- **ROC analiza:** v SPSS-u izvedemo že opisano ROC analizo in tako za vsak kazalnik dobimo območje pod ROC krivuljo ter Gini koeficient. Kazalniki z višjim Gini koeficientom bolje razlikujejo med plačniki in neplačniki.

Rezultati in njihov opis so predstavljeni v naslednjem poglavju.

Poleg naštetih metod ocenjevanja kazalnikov bi lahko uporabili tudi:

⁴V razrede moramo razdeliti le numerične kazalnike.

- neparametrični test mediane, ki prikaže razliko v porazdelitvi in lokaciji med plačniki in neplačniki;
- analiza glavnih komponent, faktorska analiza;
- dodatno podatkovno rudarjenje (*ang. data mining*) kot na primer izdelava asociacijskih pravil ali gručenje.

4.4 Razvoj modela

Kot smo že omenili, bomo razvili model, ki je kombinacija dveh statističnih metod: odločitvenega drevesa in binarne logistične regresije v listih odločitvenega drevesa. Najprej ocenimo kvaliteto kazalnikov, kot je opisano v prejšnjem poglavju. Ocene so vidne v spodnji tabeli, grafi pa se nahajajo v prilogi.

Kazalnik	Tip	Delež manjkajočih vrednosti	Vrednost korelacijskega koeficienta		AUC	Gini index
			Spearmanov rho	st. značilnosti		
K1	kategorični	0%	0,578	0,000	0,730	46%
K2	numerični	0%	-0,124	0,000	0,380	24%
K3	kategorični	0%	-0,282	0,000	0,245	51%
K4	kategorični	85%	0,041	0,004	0,525	5%
K5	kategorični	86%	0,025	0,075	0,515	3%
K6	kategorični	80%	0,024	0,094	0,516	3%
K7	kategorični	33%	-0,037	0,009	0,467	7%
K8	kategorični	39%	-0,027	0,056	0,475	5%
K9	numerični	0%	0,588	0,000	0,810	62%
K10	numerični	0%	0,384	0,000	0,703	41%
K11	kategorični	0%	-0,076	0,000	0,487	3%
K12	numerični	0%	0,210	0,000	0,532	6%
K13	numerični	0%	-0,217	0,000	0,292	42%
K14	numerični	0%	-0,217	0,000	0,292	42%
K15	numerični	0%	0,151	0,000	0,646	29%
K16	numerični	0%	0,222	0,000	0,714	43%
K17	numerični	0%	-0,202	0,000	0,306	39%
K18	numerični	0%	-0,369	0,000	0,285	43%
K19	numerični	0%	-0,096	0,000	0,419	16%
K20	numerični	0%	0,242	0,000	0,713	43%
K21	numerični	0%	-0,266	0,000	0,244	51%
K22	numerični	0%	0,187	0,000	0,676	35%
K23	numerični	0%	-0,233	0,000	0,276	45%

Slika 4.1: Tabela z ocenami kvalitete kazalnikov

Na podlagi pridobljenih ocen pridemo do sledečega sklepa:

- na podlagi količine zapolnjenih podatkov kazalnike $K4, K5, K6$ zaradi visokega odstotka manjkajočih podatkov izločimo. Ker gre za kategorične kazalnike, zapolnjevanje praznih polj ne pride v poštev, saj lahko pristransko vplivamo na rezultat nadaljnje analize. Prav tako tudi rezultati nadaljnje ocene kvalitete teh kazalnikov niso zanesljivi. V vzorce tako vključimo le tiste komitente, ki imajo pri vseh neizločenih kazalnikih zapolnjena vsa polja;
- na podlagi grafične analize opazimo, da bi lahko kazalniki $K1, K2, K3, K4, K7, K9, K14, K16, K20, K23$ imeli dobro napovedno moč, saj pri njih stopnja neplačila skozi vse razrede bodisi narašča bodisi pada. Ker ima kazalnik $K4$ velik odstotek manjkajočih podatkov, bi lahko bil graf z vsemi podatki popolnoma drugačen. Podobno velja tudi za kazalnik $K7$, ki ima prav tako relativno veliko manjkajočih podatkov;
- na podlagi korelacijske analize ugotovimo, da vsi kazalniki statistično značilno korelirajo z odvisno spremenljivko. Kot že omenjeno pa zaradi visokega odstotka manjkajočih podatkov pri kazalnikih $K4, K5$ in $K6$ rezultat ni zanesljiv;
- na podlagi ROC analize ugotovimo, da med neplačniki in plačniki najbolje razlikuje kazalnik $K9$. Dobro napovedno moč kažejo tudi kazalniki $K1, K3, K21$ in $K23$.

Na podlagi univariatne analize bi tako poleg kazalnikov z visokim deležem manjkajočih podatkov (nad 40 %) izločili še 11 kazalnikov, ki imajo Gini indeks manjši od 45 % ter hkrati njihove stopnje neplačila ne naraščajo (ali padajo) skozi vse razrede. Ker se v praksi lahko zgodi, da kombinacija dveh ali več kazalnikov dobro razločuje med plačniki in neplačniki, kljub temu, da posamezen kazalnik sam po sebi slabo vpliva na status neplačila, ne izločimo nobenega.

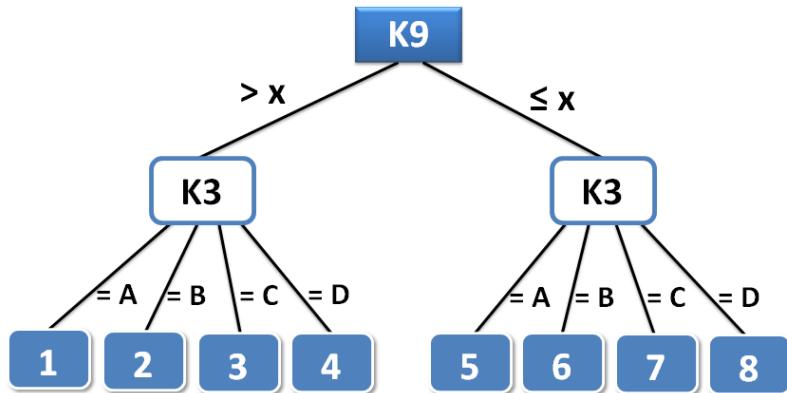
Da bi izbrali vzorec, ki najbolje predstavlja populacijo in daje najboljše rezultate, na treh različnih vzorcih izdelamo odločitveno drevo in na podlagi najboljše napovedne moči izberemo enega, ki predstavlja naš razvojni vzorec za nadaljnjo analizo. Vzorci so opisani v prejšnjem poglavju, rezultat izdelave odločitvenega drevesa pa v spodnji tabeli.

Vzorec Lastnost	V1	V2	V3
AUC	0,798	0,908	0,92
Gini index	60%	82%	84%
Natančnost	93%	90%	87%
Senzitivnost	57%	82%	86%
Specifičnost	97%	92%	87%
Razmerje	1:9	1:3	1:2

Slika 4.2: Primerjave lastnosti modela med različnimi vzorci

Opazimo, da se v skladu z našimi pričakovanji z manjšanjem razmerja napovedna moč modela veča. To je posledica dejstva, da če je v vzorcu več neplačnikov, jih model lažje prepozna. Pri vseh treh modelih odločitveno drevo komitente najprej razdeli na dva dela glede na kazalnik $K9$. Nato ostane pri prvem vzorcu en del nerazdeljen, drugi del pa se razdeli na 4 razrede glede na kazalnik $K3$. Pri 2. in 3. vzorcu pa oba dela razdelimo na 4 razrede glede na kazalnik $K3$. Tako ima odločitveno drevo za prvi vzorec 5 listov, za drugi in tretji pa 8 listov.

Na podlagi dobljenih rezultatov se odločimo, da bomo nadaljnjo analizo izvedli na vzorcu V2, saj ima glede na razmerje med neplačniki in plačniki najboljše napovedne lastnosti.



Slika 4.3: Odločitveno drevo za vzorec V2 (in V3)

Sedaj vzorec V2 na podlagi dobljenega odločitvenega drevesa razdelimo na 8 razredov oz. podvzorcev in na vsakem izvedemo binarno logistično regresijo. Pri tem mejno vrednost za določitev statusa plačnika oz. neplačnika določimo, kot je zapisano v §3.2.2. Nato preverimo natančnost dobljenega modela na testnem vzorcu VT. Rezultati so prikazani v spodnji tabeli.

Lastnost	Vzorec	V2	TV
AUC	0,759	0,786	
Gini index	52%	57%	
Natančnost	72%	65%	
Senzitivnost	91%	88%	
Specifičnost	66%	64%	

Slika 4.4: Primerjava rezultatov modela

Ker model nima visoke natančnosti, ga bomo poskusili izboljšati z metodo napihovanja

podatkov. To pomeni, da bomo v podvzorcih, kjer je to potrebno (R7 in R8), izravnali število neplačnikov in plačnikov, da dobimo razmerje 1:3, ki se je pri izbiri razvojnega vzorca izkazalo kot najboljše. Ker je v prvih štirih razredih več neplačnikov kot plačnikov, napihanje podatkov v njih ni potrebno. Prav tako napihanje ni potrebno v razredih 5 in 6, saj že ustrezata želenemu razmerju. Tako napihnemo le razreda 7 in 8.

Napihanje podatkov izvedemo po sledečih korakih:

1. prekopiramo vse podatke prvega neplačnika v razredu;
2. njegove numerične vrednosti kazalnikov spremenimo na sledeči način:

$$\hat{K} = K \cdot (1 + (\text{rand}() - 0,5)/100),$$

kjer je

- \hat{K} nova vrednost kazalnika,
- K izhodiščna vrednost kazalnika in
- $\text{rand}()$ funkcija, ki vrne naključno realno število iz intervala $[0, 1)$;

3. postopek ponovimo za toliko neplačnikov, da dobimo želeno razmerje⁵.

Tako le na napihnjenih podvzorcih ponovno izvedemo binarno logistično regresijo in primerjamo natančnost dobljenega modela na nenapihnjenih razredih ter na razredih iz testnega vzorca. Dobljeni rezultati so prikazani v spodnji tabeli.

Lastnost Vzorec	Natančnost		Senzitivnost		Specifičnost	
	napihnjen model	nenapihnjen model	napihnjen model	nenapihnjen model	napihnjen model	nenapihnjen model
Razred 7	22%	63%	90%	86%	16%	61%
Razred 8	77%	75%	89%	89%	76%	74%
V2	61%	72%	89%	91%	50%	66%
Testni razred 7	22%	63%	94%	86%	22%	63%
Testni razred 8	0%	61%	100%	87%	0%	61%
TV	54%	65%	90%	88%	54%	64%

Slika 4.5: Primerjava rezultatov napovedi modelov

Kot lahko vidimo, se je napihanje v tem primeru izkazalo kot neučinkovito, saj se napovedna moč modela v razredih razvojnega vzorca sicer nekoliko izboljša, v razredih testnega vzorca pa drastično zmanjša. Morda bi dobili boljše rezultate, če bi uporabili drugo razmerje, vendar pa bi bilo metodo napihanja potrebno podrobnejše preučiti ter preizkusiti, da bi na to vprašanje znali zanesljivo odgovoriti.

⁵Če je potrebno ponovno začnemo pri prvem neplačniku in postopek ponavljamo dalje.

4.5 Bonitetni model

Model na podlagi odločitvenega drevesa komitente najprej razdeli v 8 razredov in nato v vsakem razredu z logistično regresijo oceni funkcijo, ki napoveduje prihodnje verjetnosti neplačila za vsakega komitenta. Ker v razredu 3 z regresijo ne izboljšamo napovedi odločitvenega drevesa, je verjetnost neplačila za vse komitente v tem razredu enaka. Vsi komitenti, ne glede na razred, ki imajo zabeležen najmanj en dogodek iz definicije neplačila (glej §4.2), pa imajo verjetnost neplačila enako 1.

Na podlagi opravljenih analiz dobimo sledeči model za ocenjevanje verjetnosti neplačila:

$$PD = \begin{cases} 1/(1 + e^{-a}) & K9 > x \wedge K3 = "A" \\ 1/(1 + e^{-b}); & K9 > x \wedge K3 = "B" \\ 0,84; & K9 > x \wedge K3 = "C" \\ 1/(1 + e^{-c}); & K9 > x \wedge K3 = "D" \\ 1/(1 + e^{-d}); & K9 \leq x \wedge K3 = "A" \\ 1/(1 + e^{-e}); & K9 \leq x \wedge K3 = "B" \\ 1/(1 + e^{-f}); & K9 \leq x \wedge K3 = "C" \\ 1/(1 + e^{-g}); & K9 \leq x \wedge K3 = "D" \\ 1; & izpolnjen vsaj en pogoj za status neplačnika \end{cases}$$

kjer je

$$\begin{aligned} a = & -0,15878 + 0,04431 \cdot K9 + 0,00801 \cdot K10 - 0,00154 \cdot K13 - 0,00029 \cdot K15 \\ & + 0,0003 \cdot K16 - 1,49519 \cdot K19 + 2,40803 \cdot K20, \end{aligned}$$

$$b = 0,037823 + 0,050091 \cdot K9 - 2,98193619 + 5,205342 \cdot K20 + 0,0007 \cdot K22,$$

$$c = -2,15105 + 0,083069 \cdot K9,$$

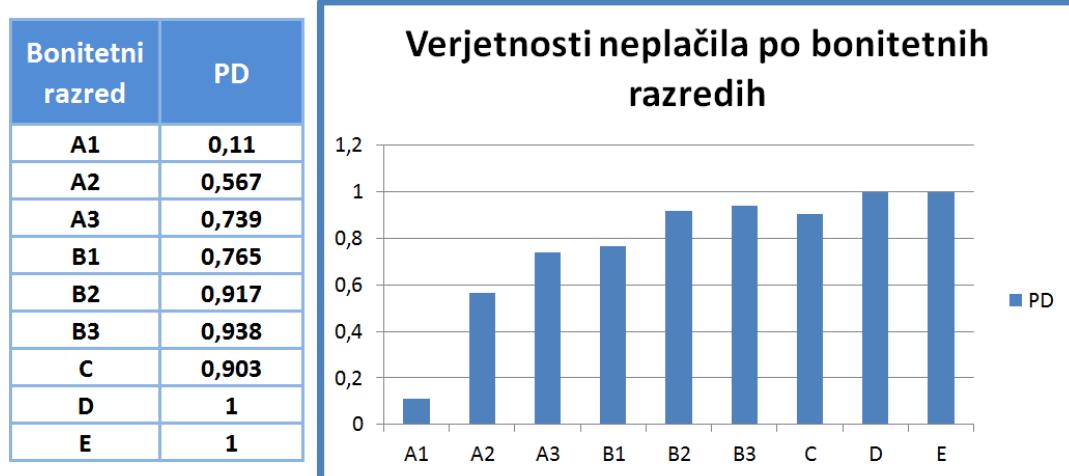
$$d = -2,313948 + 1,138887 \cdot K9 + 0,003079 \cdot K10 + 2,426322 \cdot K20 - 0,000154 \cdot K22,$$

$$e = -0,254621 - 0,035795 \cdot K2 - 0,000816 \cdot K22,$$

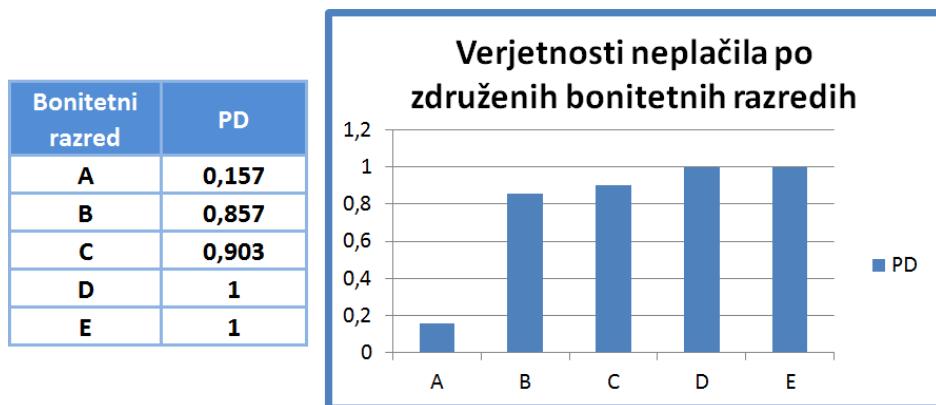
$$f = -4,86692 + 0,012747 \cdot K10 + 3,277529 \cdot K20,$$

$$g = -2,42358 - 0,06683 \cdot K2 + 0,034159 \cdot K10 - 0,004053 \cdot K18 + 2,950749 \cdot K20.$$

Na podlagi razvitega modela, dobimo sledeče ocene verjetnosti neplačila



Slika 4.6: Ocene verjetnosti neplačila po bonitetnih razredih



Slika 4.7: Ocene verjetnosti neplačila po združenih bonitetnih razredih

Povprečna verjetnost neplačila razvojnega vzorca je enaka 0,25, kar pomeni, da bi se v povprečju komitenti uvrstili v bonitetni razred A2.

Sedaj si poglejmo še natančnost končnega modela, preverjeno na testnem vzorcu. Tako kot do sedaj slednje storimo z izvedbo ROC analize v programu SPSS. Pri tem dobimo sledeče rezultate:

Specifičnost	Senzitivnost	Natančnost	Gini index
64%	88%	65%	81%

Slika 4.8: Mere natančnosti končnega modela

		Napovedane vrednosti	
		0	1
Dejanske vrednosti	0	53352	29918
	1	227	1609

Slika 4.9: Matrika razvrstitev

Area Under the Curve

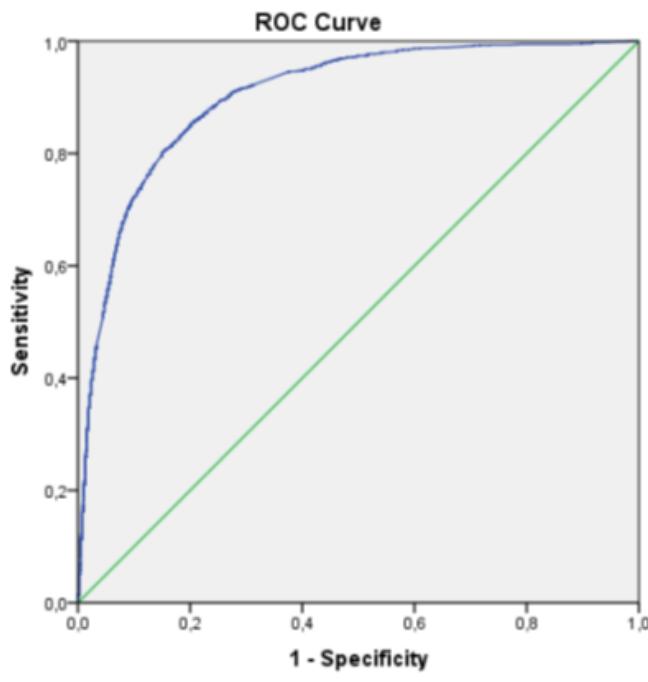
Test Result Variable(s): PD

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,903	,003	,000	,897	,910

The test result variable(s): PD has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5



Diagonal segments are produced by ties.

Slika 4.10: Rezultat ROC analize

Ker smo dali prednost senzitivnosti, je specifičnost modela nekoliko nižja in posledično tudi sama natančnost modela ni ravno visoka. Kljub temu ne moremo trditi, da je dobljeni rezultat slab. Model res ni optimalen, vendar pa se moramo zavedati, da je v praksi skoraj nemogoče doseči 100 % natančnost.

4.6 Kalibracija modela

Ker se delež neplačnikov v razvojnem vzorcu razlikuje od deleža v celotnem portfelju, so z logistično regresijo ocenjene verjetnosti neplačila bistveno previsoke. Iz tega razloga moramo dobljene ocene kalibrirati. Pri tem vrednosti v vzorcu niso neposredno povezane s tistimi v portfelju, ampak so posredno povezane z uporabo relativnih frekvenc neplačnikov (*RDF*-ang. *relative default frequencies*): $RDF = PD/(1 - PD)$.

Verjetnosti neplačnika ustrezno popravimo po naslednji formuli:

$$\widetilde{PD}_{razreda} = PD_{razreda} * \frac{PD_{portfelja}}{PD_{vzorca}},$$

kjer je

- $\widetilde{PD}_{razreda}$ kalibrirana verjetnost neplačila za bonitetni razred,
- $PD_{razreda}$ ocenjena verjetnost neplačila za bonitetni razred,
- $PD_{portfelja}$ ocenjeno dolgoletno povprečje neplačila in
- PD_{vzorca} vzorčna povprečna ocena verjetnosti neplačila.

Verjetnosti neplačila za bonitetna razreda D in E ne spremojamo, saj komitent, ki je že neplačnik, ne more imeti druge verjetnosti kot 1. Ker se vse ostale vrednosti spremenijo za enak faktor, oblika grafa na slikah 4.6 in 4.7 ostane enaka, le vrednosti se nekoliko znižajo.

4.7 Kapitalska zahteva

Kot že vemo, izračun verjetnosti neplačila predstavlja najpomembnejši del postopka merjenja kreditnega tveganja. Poleg tega merjenje kreditnega tveganja zajema še izračun izpostavljenosti banke in izračun dejanske izgube v odstotkih, ki jo banka realizira, ko je že pričela s postopkom izterjave.

LGD izračunamo tako, da skupne izgube delimo z izpostavljenostjo portfelja. Ker bi za ta izračun morali upoštevati najmanj petletno časovno obdobje, ga ne moremo izvesti, ker

imamo na razpolago le dvoletno obdobje. Prav tako se moramo zavedati, da je LGD dinamičen parameter in se njegove vrednosti v praksi spreminjajo ter so odvisne od vrste neplačila in njegove razrešitve.

EAD v splošnem predstavlja oceno obsega izpostavljenosti banke do dolžnika v primeru neizpolnitve obveznosti in je, v primeru fiksnih izpostavljenosti, enaka trenutnemu znesku neporavnane dolga oz. nominalnemu znesku. V primeru odprtih kreditnih linij je potrebno upoštevati tudi pričakovane prihodnje odlive iz naslova neizkoriščenega dela kreditne linije. Ker tudi za izračun EAD potrebujemo najmanj petletno časovno obdobje, ga ne moremo izvesti.

Kot vemo, se uteži za tehtanje kreditnega tveganja v okviru IRB pristopa izračunajo kot funkcija vhodnih podatkov PD , LGD , EAD . Tako komponente tveganja s pomočjo kontinuirane funkcije 2.8 pretvorimo v uteži, ki določajo kapitalske zahteve za posamezne kategorije izpostavljenosti bančnega portfelja. Če bi torej želeli določiti kapitalsko zahtevo, bi najprej morali izračunati še vrednosti za parametra LGD in EAD ter uporabiti formulo 2.6.

4.8 Pomanjkljivosti

Pri zgoraj razvitem modelu gre le za prvi poskus uporabe IRB pristopa, s katerim smo dobili prvi vtis in nekatera temeljna znanja, ki jih banke potrebujejo za razvoj lastnega modela. Tako se je potrebno zavedati, da ima model številne pomanjkljivosti in omejitve, ki bi jih bilo potrebno posebej preučiti in odpraviti:

- pri nadaljnjem razvoju bonitetnega modela bi bilo potrebno razmisljiti o smiselnosti gradnje ločenega modela za različne vrste sklenjenih poslov (na primer limiti in krediti ter dolgoročni in kratkoročni krediti). Vsaka vrsta posla ima namreč svoje specifike, ki vplivajo na verjetnost neplačila. Kot že vemo, je pomemben dejavnik tudi zapadlost posla, ki je v zgoraj razvitem modelu nismo upoštevali. To pomanjkljivost lahko odpravimo tako, da razvijemo ločen model za različne zapadlosti, ki se lahko od posla do posla precej razlikuje;
- v razvoj modela nismo vključili nekaterih kazalnikov, ki bi lahko dobro razločevali med plačniki in neplačniki, saj teh podatkov nismo imeli na voljo. Boljšo napovedno moč kot imajo kazalniki v modelu, bolj natančen je tudi sam model. V nadaljevanju bi bilo tako potrebno vključiti tudi te podatke. Prav tako nekaterih kazalnikov v model nismo mogli vključiti zaradi velikega števila manjkajočih podatkov. Kot vidimo, je zelo pomembno, da imajo banke zgrajene dobre baze podatkov, ki so v celoti zapolnjene;

- v razvoj modela so vključeni le podatki, s katerimi razpolaga banka X. Ker lahko komitenti sklepajo posle tudi pri drugih bankah, bi moral optimalen model upoštevati tudi te podatke. V praksi se namreč lahko zgodi, da nek komitent pri banki X ni zamujal s plačili in nima zabeleženega nobenega kritičnega dogodka, pri neki drugi banki pa ni bil zmožen poravnati svojega dolga. Delna rešitev tega problema bi bila vključitev bonitetne ocene bonitetnega modela SISBON Credit Scoring. Ta ocena temelji na podatkih iz informacijskega sistema SISBON, ki se nanašajo na dejansko in potencialno zadolženost ter korektnost izpolnjevanja pogodbenih obveznosti fizičnih oseb pri vseh bankah v Sloveniji. Še bolje bi bilo, če bi banka lahko razpolagala z vsemi podatki iz SISBON-a, vendar so ti dostopni le posamično in ne skupinsko. Prav tako se lahko do podatkov dostopa le preko vpogleda ali pdf izpisa, ne morejo pa se prenesti v bančno bazo podatkov. Tako bi morali vsak podatek v bazo prenesti ročno, kar bi za celoten portfelj zahtevalo ogromno časa;
- za razvoj modela se pogosto uporablja tudi transformirani kazalniki. Tako bi bilo potrebno podrobno preučiti in preizkusti možne transformacije kazalnikov (npr. Winsorizacija, kategorizacija, standardizacija ...) ter jih po potrebi v modelu tudi uporabiti;
- tako kot v večini statističnih modelov za določanje verjetnosti neplačila ima tudi ta model pomanjkljivost, da so vanj vključeni le t. i. "trdi", ne pa tudi "mehki" finančni kazalniki. Pri poskusu teoretične utemeljitve mehkih dejavnikov⁶ se je namreč izkazalo, da je to področje relativno slabo definirano ter raziskano;
- za razvoj modela bi lahko uporabili tudi več različnih vzorcev z istim razmerjem, izmed katerih bi izbirali razvojni vzorec. Morda bi se kateri drug vzorec izkazal za boljšega in bi tako imeli povsem drugačno osnovo za razvoj modela.

Opisane pomanjkljivosti so vidne tudi v relativno slabi napovedni moči oz. natančnosti samega modela. Najverjetneje bi natančnost lahko precej izboljšali, če bi opravili še omenjene dodatne analize in odpravili pomanjkljivosti modela.

⁶Povezani so s subtilnim in neoprijemljivim znanjem, predpostavkami in načinom razmišljanja, ki vplivajo na obnašanje in odločitve posameznika.

Poglavlje 5

Konkreten primer: Občina Makole

Makole so ena izmed občin v Republiki Sloveniji, ki je nastala 1. marca 2006 z izločitvijo iz občine Slovenska Bistrica. Nahaja se v Dravinjski dolini ob izteku Jelovškega potoka, na zahodnem robu Haloz. Samo območje občine je statistično uvrščeno na mejo med podravsko in celjsko regijo ter meji z občinami Rogaška Slatina, Majšperk, Poljčane in Slovenska Bistrica.

Makole so bile prvič kot vas omenjene leta 1375, kasneje pa so dobine pravice trga. Danes so Makole turistično zanimiva občina z velikimi možnostmi razvoja in se ponašajo z bogato kulturno in naravno dediščino: cerkev sv. Andreja in poznogotska cerkev sv. Lenarta, dvorec Štatenberg, slap Šošterca, jama Belojača, rudnik Šega, vinogradi ter še in še.

Turizem je ena izmed gospodarskih panog, ki je v Makolah v stalnem razvoju in rasti. V zadnjem času se je vidno razmahnila tudi rekreacijska dejavnost, za možnost trženja v turizmu pa obstajajo še lovski in ribiški turizem, konjeništvo ter mnoge dopolnilne dejavnosti na kmetijah.

Občina Makole ima nekaj več kot 2000 prebivalcev. Od tega je slabih 40 odstotkov delovno aktivnih. Nizek odstotek zaposlenosti je posledica velikega števila upokojencev in ukvarjanja s samooskrbnim kmetijstvom.

5.1 Zakaj Makole potrebujejo kredit?

Makole so eden izmed krajev v Sloveniji brez urejenega javnega kanalizacijskega omrežja in tako morajo za odpadne vode iz gospodinjstva poskrbeti lastniki hiš sami. Nekateri jih imajo še vedno speljane v pretočne greznice, ki močno obremenjujejo in onesnažujejo okolje, zato bodo morali do leta 2017 vanje vgraditi ustrezno čistilno napravo oziroma z njo v

celoti nadomestiti greznicu. Lastniki novogradenj pa morajo že zdaj v skladu z zakonodajo v projektu predvideti gradnjo nepretočne triprekatne greznice ali postavitev male čistilne naprave ustrezne velikosti.

Po državnem operativnem programu mora tako občina Makole do 31. 12. 2017 naselja, kjer je to izvedljivo oziroma smiselno, opremiti z javno kanalizacijo, zaključeno s čistilno napravo. V ostalih naseljih v občini, kjer izgradnja javnega kanalizacijskega sistema ni predvidena, pa je potrebno individualno za vsak objekt posebej ali več objektov skupaj urediti odvajanje odpadne komunalne vode v male komunalne čistilne naprave (MKČN) ali nepretočne greznice. Enako velja za posamezne stavbe v navedenih aglomeracijah, ki jih ne bo mogoče priključiti na sistem javne kanalizacije.

Cene MKČN se gibljejo nekje od 2500 € (za gospodinjstva z dvema osebama) dalje, glede na to, da pa večina ljudi (tako v Makolah kot tudi drugod po Sloveniji) MKČN še nima, bo v naslednjih letih njihovo povpraševanje zagotovo naraslo in predvidoma bodo narastle tudi njihove cene. Tako bo večina gospodinjstev za njihov nakup najverjetneje potrebovala manjši kredit.

5.2 Kam sodijo občani?

Z uporabo zgoraj razvitega modela bomo preverili, v kateri bonitetni razred bi se uvrstili občani občine Makole. V ta namen med vsemi komitenti poslovne banke izberemo le tiste, ki stanujejo v občini Makole ter v skladu z modelom določimo njihove verjetnosti neplačila.

Na podlagi dobljenih verjetnosti neplačila, občane razdelimo v bonitetne razrede.

Bonitetni razred	% občanov
A	78,69%
B	15,87%
C	0,80%
D	1,60%
E	1,44%

Slika 5.1: Razporeditev občanov po bonitetnih razredih

- V vzorcu imamo 3 % neplačnikov, ki so razdeljeni v bonitetna razreda D in E. Tako je med Makolčani 1 % več neplačnikov kot v celotnem portfelju banke.

- 1,28 % komitentov ima verjetnost neplačila enako 0,99 %, kar pomeni, da so tik pred tem, da postanejo neplačniki.
- 79 % komitentov bi se uvrstilo v bonitetni razred A, kar je več kot v primerjavi s celotnim portfeljem banke.
- Vzorčna povprečna ocena verjetnosti neplačila občine Makole znaša 0,1, kar pomeni, da bi se v povprečju Makolčan uvrstil v bonitetni razred A.

Kot lahko vidimo je med občani občine Makole v primerjavi s celotnim portfeljem banke nekoliko več neplačnikov. Ostali, komitenti, pa so se v povprečju izkazali kot dobri komitenti z nizkimi ocenami verjetnosti neplačila. Kar 78 % občanov ima namreč verjetnost neplačila manjšo ali enako 0,1 in le 3,5 % ima verjetnost neplačila višjo kot 0,5.



Slika 5.2: Razporeditev občanov občine Makole po bonitetnih razredih.

Občani občine Makole so se v povprečju torej izkazali kot dobri komitenti z nizkimi verjetnostmi neplačila. Na podlagi dobljenih rezultatov lahko banka sprejme ustrezne ukrepe za učinkovito upravljanje s tveganjem kreditiranja občanov občine Makole, ki je relativno nizko. Prav tako bi lahko z dodatno analizo razvili ustrezno ponudbo, prilagojeno občanom s katero bi pritegnili nove posojilnjemalce, ki bodo potrebovali kredit za nakup male komunalne čistilne naprave.

Poglavlje 6

Zaključek

Vsaka banka je pri svojem poslovanju izpostavljena različnim vrstam tveganj, ki se jih ne more izogniti, lahko jih le znižuje oziroma obvladuje z učinkovitim upravljanjem. Nov kapitalski sporazum, imenovan Basel II, je na področju kreditnega tveganja prinesel povsem nove možnosti za učinkovitejše upravljanje z njim. Tako smo se najprej seznanili z osnovnimi pojmi in značilnostmi kreditiranja prebivalstva. Pri tem smo se podrobneje posvetili kreditnemu tveganju, ki zavzema osrednje mesto med vsemi tveganji. Sledila je predstavitev IRB pristopa, kjer smo spoznali njegove lastnosti, prednosti ter glavne parametre. Prvi del magistrskega dela nam poda neko širšo sliko teme, ki je obravnavana. Spoznamo, zakaj je upravljanje kreditnega tveganja tako pomembno in zakaj bi banke morale iz standardiziranega prestopiti na IRB pristop.

V nadaljevanju smo preučili statistične metode, ki smo jih uporabili pri gradnji modela. Spoznali smo, kako poteka ROC analiza, kaj je binarna logistična regresija, kako zgradimo odločitveno drevo in kaj je korelacija. Podrobni opis omenjenih metod nam omogoča, da razumemo, kaj in kako sta izvedla oba programa, ki smo ju uporabili pri razvoju modela. Pri tem smo ugotovili, da imajo kazalniki, ki na podlagi ROC analize najbolje razločujejo med plačniki in neplačniki, prav tako tudi najvišji informacijski prispevek. Pri izvedbi binarne logistične regresije so se za najbolj signifikante izkazali kazalniki, ki bi jih ob upoštevanju zgolj ROC analize zaradi nizkega Gini indeksa izločili.

Sledil je razvoj bonitetnega modela za slovensko prebivalstvo. Prvi korak pri izdelavi modela je bilo zbiranje podatkov, ki so ga izvedli informatiki poslovne banke. Sledil je pregled in čiščenje podatkov, kjer smo izločili komitente z nesmiselnimi oz. neustreznimi podatki (npr. negativna starost). Nato smo, v skladu z definicijo Basla II, vsakemu komitentu pripisali bodisi status plačnika bodisi status neplačnika. Tako smo dobili portfelj komitentov, na osnovi katerega smo za nadaljnjo analizo iz več kot sto atributov izbrali 23 kazalnikov. Po

analizi in pregledu kazalnikov smo 3 izločili, ostale pa uporabili za razvoj modela. Sledila je izdelava odločitvenega drevesa na različnih vzorcih. Za razvojni vzorec smo izbrali tiste, ki je glede ne razmerje neplačnikov in plačnikov podal najboljši rezultat. Nato smo v listih odločitvenega drevesa na izbranem vzorcu uporabili široko uporabljen metodo za ocenjevanje binarnih odvisnih spremenljivk, imenovano binarna logistična regresija. Tako dobimo model, ki komitente najprej razdeli v 8 razredov in nato v vsakem razredu oceni funkcijo, ki napoveduje prihodnje verjetnosti neplačila. Tako dobimo za vsakega komitenta njegovo verjetnost neplačila. Dobljene vrednosti smo nato združili v bonitetne razrede in dobili verjetnosti neplačila za posamezen bonitetni razred. Ker sta razmerji med plačniki in neplačniki v vzorcu in portfelju različni, smo dobljene vrednosti popravili in tako dobili osnovo za mapiranje verjetnosti neplačila po bonitetnih skupinah.

V osrednjem delu magistrskega dela smo z namenom razvoja bonitetnega modela po IRB pristopu, analizirali vpliv različnih kazalnikov poslovanja na verjetnost neplačila. Razviti model temelji na razvoju odločitvenega drevesa ter binarni logistični regresiji, ki smo jo izvedli v listih drevesa, in je izdelan v skladu z Baslom II. Spoznali smo, da razvoj takšnega modela zahteva veliko časa in truda, saj obstajajo različne metode in koraki modeliranja, ki jih je potrebno dobro preučiti in tudi preizkusiti. Ugotovimo tudi, da ima poskusni model še veliko pomanjkljivosti, ki bi jih bilo potrebno odpraviti in tako predstavlja le osnovo za nadaljnje delo. Šele, ko bi model ustrezno popravili in dopolnili, bi ga banka lahko vgradila v svoj sistem in bi tako zaživel tudi v praksi.

V zadnjem poglavju dela razviti model uporabimo na portfelju občanov občine Makole. Tako ugotovimo, kakšno verjetnost neplačila imajo posamezni komitenti ter v kateri bonitetni razred bi se uvrstili. Na podlagi dobljenih rezultatov lahko banka sprejme ustrezne ukrepe za učinkovito upravljanje s tveganjem kreditiranja občanov občine Makole.

Pri razvitem modelu gre le za prvi poskus uporabe IRB pristopa, s katerim smo dobili prvi vtis in nekatera temeljna znanja, ki jih banke potrebujejo za razvoj lastnega modela. Spoznali smo, da je modeliranje dolgotrajen proces, ki zahteva veliko časa, truda in predvsem natančnosti.

Literatura

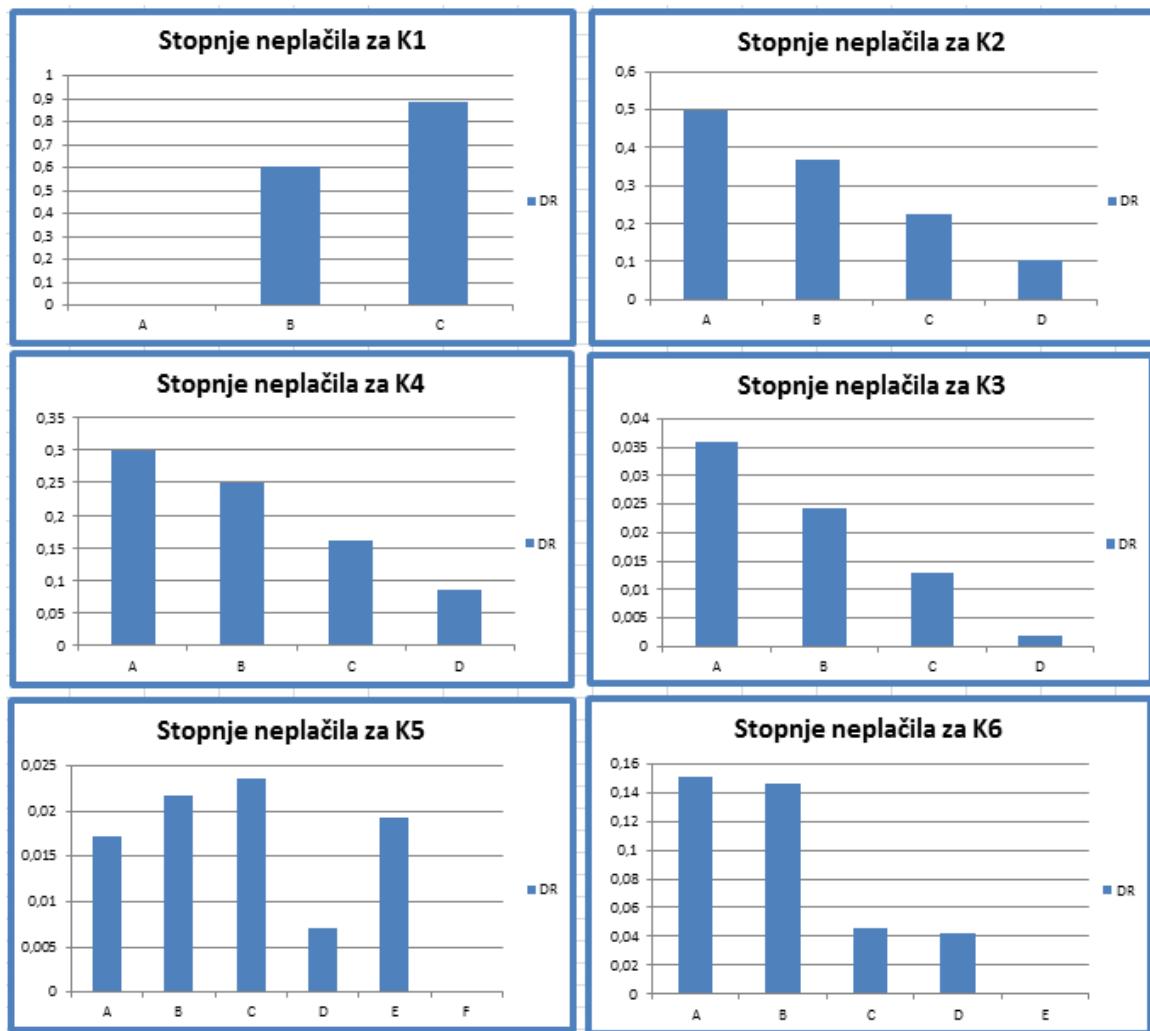
- [1] Andrejc, D.: *Izdelava modela verjetnosti neplačila v poslovni banki*. Diplomsko delo, Ljubljana, 2009.
- [2] Baesens, B.; Van Gestel, T.: *Credit Risk Management: Basic Concepts*. Oxford University Press, New York, 2009.
- [3] Basel Committee on Banking Supervision: *Principles for the Management of Credit Risk* - Consultative Document. Basel, julij 1999.
- [4] Basel Committee on Banking Supervision: *Credit Risk Modelling: Current Practices and Applications*. Basel, april 1999.
- [5] Basel Committee on Banking Supervision: *The Internal Ratings-Based Approach* - Consultative Document. Supporting Document to the New Basel Capital Accord. A Revised Framework, Basel, januar 2001.
- [6] Bessis, J.: *Risk Management in banking*. John Wiley & Sons, Ltd, Chichester, 2002.
- [7] Bobek, D.: *Organiziranje in poslovanje banke*. Ekonomsko poslovna fakulteta, Maribor, 1992.
- [8] Coklin, T.: *Upravljanje in obvladovanje kreditnega tveganja v bankah*. Diplomsko delo, Gornja Radgona, 2009.
- [9] Crook, J. N.; Edelman, D. B.; Thomas, L. C.: *Credit Scoring and Its applications*. Society for Industrial and Applied Mathematics, Philadelphia, 2002.
- [10] Čargo, A.; Štajner, M.: *Minimalne zahteve za uvedbo IRB pristopa*. Banka Slovenije.
- [11] Dimovski, V.; Gregorič, A.: *Temelji bančništva*. Ekonomsko fakulteta, Univerza v Ljubljani, Ljubljana, 2000.
- [12] Ferri, C.; Hernandez-Orallo, J.; Salido, M. A.: *Volume under the ROC Surface for Multi-class Problems*. Machine Learning: ECML 2003. pp. 108–120, 2003.

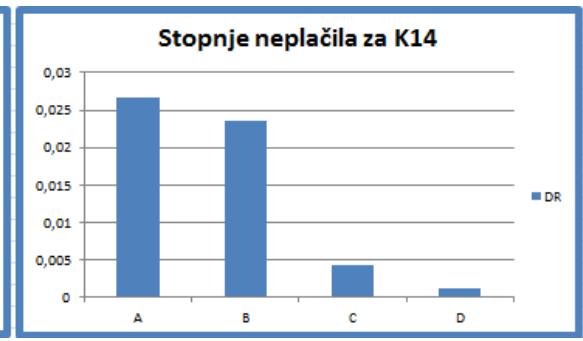
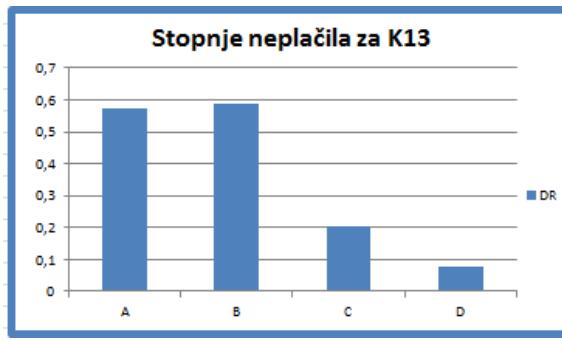
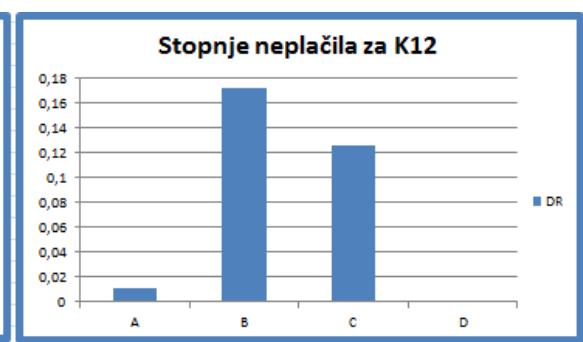
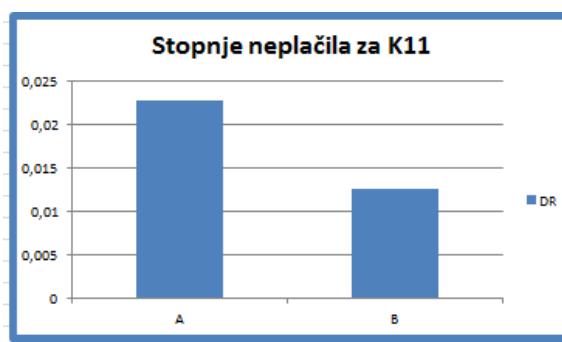
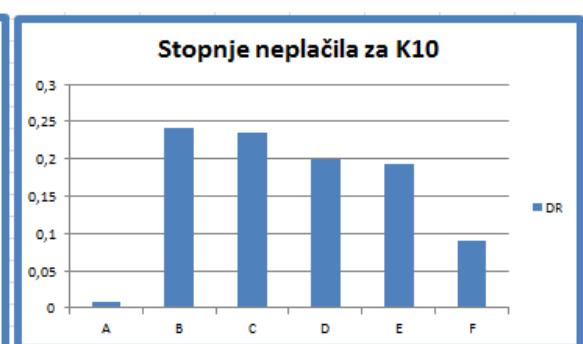
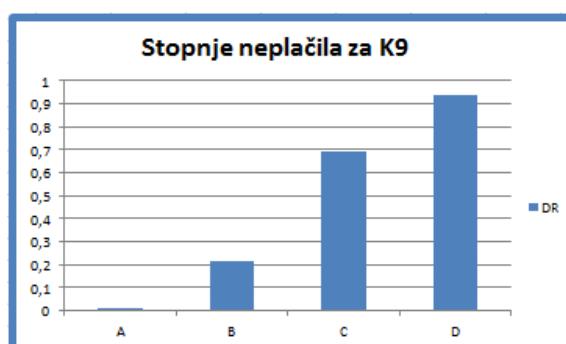
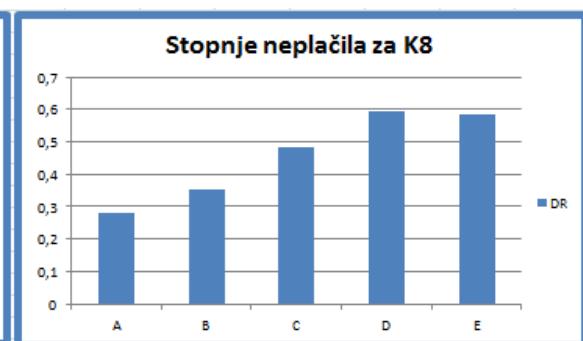
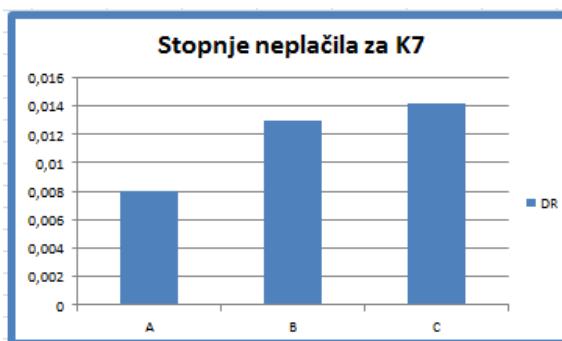
- [13] Filipan, T.: *Upravljanje kreditnega portfelja v poslovni banki*. Magistrsko delo, Ljubljana, 2003.
- [14] Fitch Ratings: *How much Credit in Credit Risk Models?* New York, 2007.
- [15] Frank, E.; Witten, I. H.: *Data Mining. Practical Machine learning Tools and Techniques*. Second Edition. Morgan Kaufmann Publishers, San Francisco, 2005.
- [16] Glogovšek, J.: *Organizacijski in drugi vidiki bančnega poslovanja*. Ekonomsko poslovna fakulteta, Maribor, 1996.
- [17] Han, J.; Kamber, M.: *Data Mining: Concepts and Techniques*. Second Edition. Morgan Kaufmann Publishers, San Francisco, 2006.
- [18] Hojs, M.: *Vplivi novega kapitalskega sporazuma pri kreditiranju malih in srednjih podjetij v Novi KBM*. Diplomsko delo, Maribor, 2009.
- [19] Jovan, M.; Šušterič, M.: *Statistično ocenjevanje verjetnosti neplačila za slovenska podjetja*. Članek, 2004.
- [20] Korenjak, A.: *Regresijska analiza*. Diplomsko delo, Maribor, 2010.
- [21] Kovač, V.: *Kreditno tveganje v bančnem poslovanju*. Diplomsko delo, Maribor, 2009.
- [22] Križnik, T.: *Model obvladovanja kreditnega tveganja za slovenske banke*. Magistrsko delo, Ljubljana, 2005.
- [23] Mossman, D.: *Three-way ROCs*. Medical Decision Making 19: 78–89. doi:10.1177/0272989x9901900110, 1999.
- [24] Peršuh, G.: *Kreditni portfelj in njegova ocena po Baslu II*. Specialistično delo, Maribor, 2006.
- [25] Research Gate (online). (citirano 25. 10. 2014). Dostopno na naslovu: http://www.researchgate.net/post/How_to_interpret_a_Constant_in_a_Regression_Result.
- [26] Ružič, T.: *Obvladovanje kreditnega tveganja*. Delo diplomskega seminarja, Maribor, 2010.
- [27] SAS (online). (citirano 25. 10. 2014). Dostopno na: <http://support.sas.com/kb/23/136.html>.
- [28] Slak, L.: *Obvladovanje tveganj v bančnem poslovanju po novem kapitalskem sporazumu Basel II*. Magistrsko delo, Maribor, 2005.

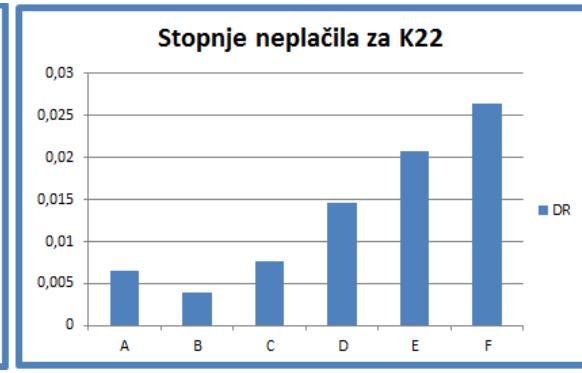
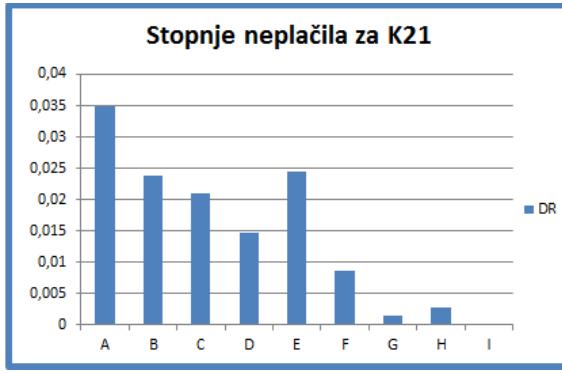
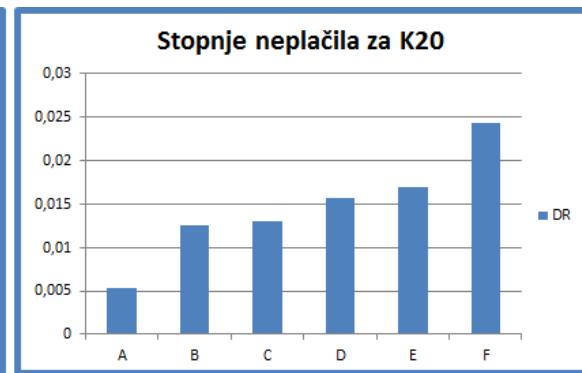
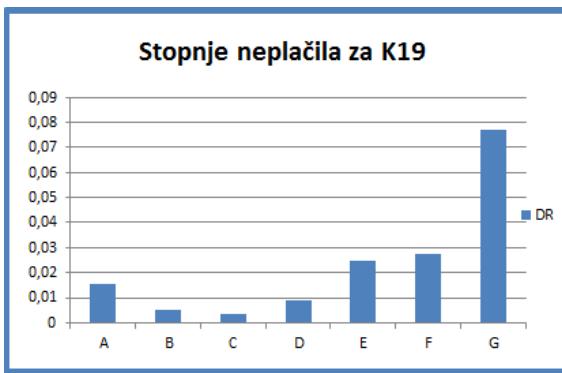
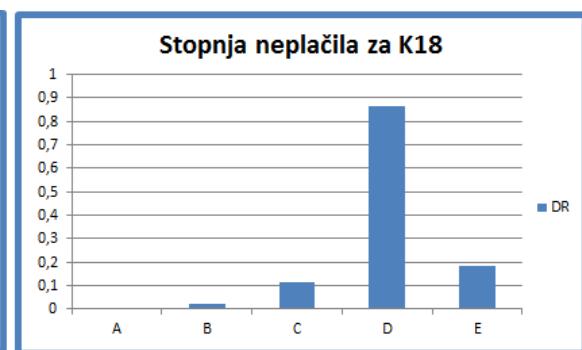
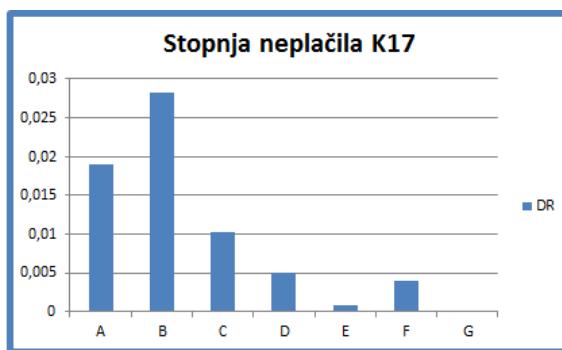
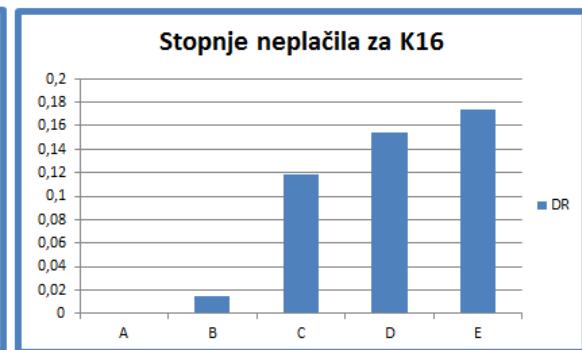
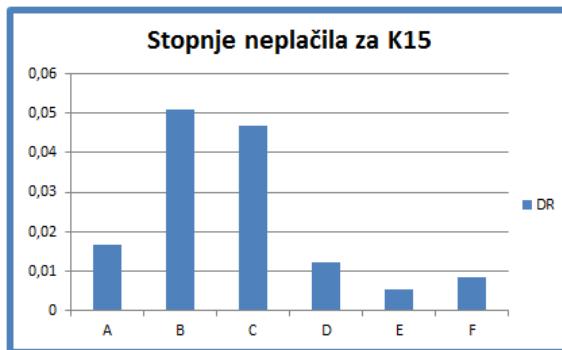
- [29] Srinivasan, A.: *Note on the Location of Optimal Classifiers in N-dimensional ROC Space*. Technical Report PRG-TR-2-99, Oxford University Computing Laboratory, Oxford, 1999.
- [30] Svet Banke Slovenije: *Sklep o izračunu kapitalske zahteve za kreditno tveganje po pristopu na podlagi notranjih bonitetnih sistemov za banke in hranilnice*. Uradni list RS, št. 135/06, december 2006.
- [31] Svet Banke Slovenije: *Sklep o ocenjevanju izgub iz kreditnega tveganja bank in hranilnic*. Uradni list RS, št. 12/13, februar 2013.
- [32] Svet Banke Slovenije: *Sklep o upravljanju s tveganji in izvajanjem procesa ocenjevanja ustreznega notranjega kapitala za banke in hranilnice*. Uradni list RS, št. 12/14, februar 2014.
- [33] Stephens; K. S.: *The Handbook of Applied Acceptance Sampling: Plans, Procedures and Principles*. ASQ Quality press, Milwaukee, 2001.
- [34] Šušterič, M.; Zavodnik, E.: *Priprava slovenskih bank na spremembe kapitalskega sporazuma Basel II*. Članek, 2003.
- [35] The R Project for Statistical Computing (online). (citirano 25. 10. 2014). Dostopno na <http://www.linkedin.com/groups/what-does-it-mean-when-77616.S.214626107>.
- [36] Till, D. J.; Hand, R. J.: *A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems*. Machine Learning, 2012.

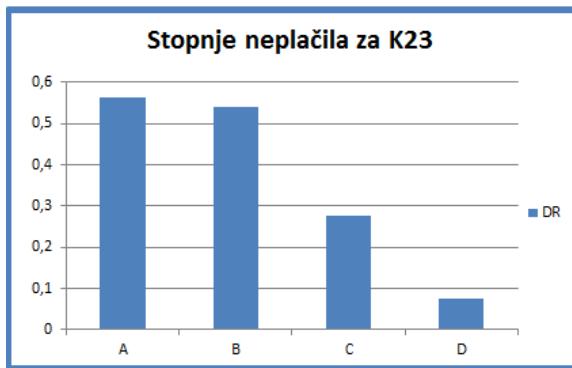
Priloge

Stopnje neplačila po kazalnikih









Prikaz rezultatov binarne logistične regresije po razredih

Razred 1

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	193,648 ^a	,160	,323

a. Estimation terminated at iteration

		Predicted		Correct
		0	1	
Observed	0	21	20	51,0%
	1	25	314	93,0%
Overall Percentage				88,0%

a. The cut value is ,745

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
K9	,04431	,010	18,872	1	,000	1,045
K10	,00801	,003	6,707	1	,010	1,008
K13	-,00154	,001	3,994	1	,046	,998
K19	-1,49519	,749	3,986	1	,046	,224
K20	2,40803	,563	18,309	1	,000	11,112
K15	-,00029	,000	2,765	1	,096	1,000
K16	,00030	,000	2,808	1	,094	1,000
Constant	-,15878	,550	,083	1	,773	,853

a. Variable(s) entered on step 1: K9, K10, K13, K19, K20, K15, K16.

Razred 2**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	38,063 ^a	,201	,380

a. Estimation terminated at iteration number 6
because parameter estimates changed by less
than ,001.

Classification Table^a

Observed	Predicted		ge Correct	
	Y			
	0	1		
Step 1 Y 0	6	3	66,6	
1	3	60	95,2	
Overall Percentage			91,9	

a. The cut value is ,67

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a K9	,050091	,022	5,391	1	,020	1,051
K19	-2,981936	1,479	4,065	1	,044	,051
K20	5,205342	2,131	5,965	1	,015	182,243
K22	,000700	,000	4,463	1	,035	1,001
Constant	,037823	,770	,002	1	,961	1,039

a. Variable(s) entered on step 1: K9, K19, K20, K22.

Razred 3**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	54,571 ^a	,186	,316

a. Estimation terminated at iteration
number 7 because parameter estimates

Classification Table^a

Observed	Predicted		Percentage Correct	
	Y			
	0	1		
Step 1 Y 0	0	13	0,0	
1	0	67	1,0	
Overall Percentage			83,8	

a. The cut value is ,600

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a K10	,026	,013	4,193	1	,041	1,026
Constant	,615	,376	2,677	1	,102	1,849

a. Variable(s) entered on step 1: K10.

Razred 4

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	87,706	,124	,171

a. Estimation terminated at iteration number 7 because parameter estimates changed by

Classification Table^a

Observed	Y	Predicted		ge Correct
		0	1	
Step 1	Y	0	32	14
		1	6	19
Overall Percentage				71,8

a. The cut value is ,600

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	K9	,083069	,024	11,846	1	,001
	Constant	-2,151050	,497	18,740	1	,000

a. Variable(s) entered on step 1: K9.

Razred 5

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	568,443 ^a	,230	,330

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Classification Table^a

Observed	Y	Predicted		ge Correct
		0	1	
Step 1	Y	0	237	221
		1	20	153
Overall Percentage				61,8

a. The cut value is ,0922

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	K9	1,138887	,433	6,932	1	,008
	K10	,003079	,001	7,179	1	,007
	K20	2,426322	,338	51,577	1	,000
	K22	-,000154	,000	3,232	1	,072
	Constant	-2,313948	,205	127,649	1	,000

a. Variable(s) entered on step 1: K9, K10, K20, K22.

Razred 6

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	182,403 ^a	,277	,416

a. Estimation terminated at iteration number 7
because parameter estimates changed by less than ,001.

Classification Table^a

Observed	Predicted		ge Correct	
	Y			
	0	1		
Step 1 Y	105	76	58,0	
1	6	50	89,3	
Overall Percentage			65,4	

a. The cut value is ,153

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
K2	-,035795	,020	3,220	1	,073	,965
K22	-,000816	,000	34,060	1	,000	,999
Constant	-,254621	,780	,107	1	,744	,775

a. Variable(s)
entered on
step 1: K2,
K22.

Razred 7

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	146,017 ^a	,132	,322

a. Estimation terminated at iteration number 7
because parameter estimates changed by less than ,001.

Classification Table^a

Observed	Predicted		ge Correct	
	Y			
	0	1		
Step 1 Y	213	137	60,9	
1	4	24	85,7	
Overall Percentage			62,7	

a. The cut value is ,0435

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
K10	,012747	,003	22,944	1	,000	1,013
K20	3,277529	,759	18,644	1	,000	26,510
Constant	-4,866922	,656	55,033	1	,000	,008

a. Variable(s) entered on step 1: K10, K20.

Razred 8

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	247,512 ^a	,071	,309

a. Estimation terminated at iteration number 8
because parameter estimates changed by less
than ,001.

Classification Table^a

Observed	Predicted		ge Correct	
	Y			
	0	1		
Step 1 Y 0	959	330	74,4	
1	4	34	89,5	
Overall Percentage			74,8	

a. The cut value is ,01869

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
K2	-,0666830	,019	12,692	1	,000	,935
K10	,034159	,009	14,174	1	,000	1,035
K18	-,004053	,002	3,826	1	,050	,996
K20	2,950749	,548	28,956	1	,000	19,120
Constant	-2,423580	,844	8,254	1	,004	,089

a. Variable(s) entered on step 1: K2, K10, K18, K20.

Opazimo, da imamo pri nekaterih primerih v model vključeno nesignifikantno konstanto. Po nekaterih mnenjih bi takšno konstanto iz modela morali izključiti, vendar običajno strokovnjaki, ki se že več let ukvarjajo s takšnim modeliranjem na gospodarskem in finančnem področju, odsvetujejo izključitev nesignifikantne konstante iz modela (glej [25], [27], [35]).

Z modeliranjem dobimo krivuljo, ki z linearnim členom odvisnosti aproksimira podatke, ki smo jih uporabili za razvoj modela. Po besedah Johna Ridinga, z vključitvijo konstante v model zagotovimo, da je vpliv členov višjega reda v okolini vzorčnih podatkov vključen v model [25]. Na tej intuitivni osnovi smo se odločili, da bomo v modelih obdržali tudi nesignifikantne konstantne člene.