

# Napovedovanje izidov svetovnega prvenstva v nogometu z uporabo modela matričnega razcepa

Štefan Dobravec

Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška 25, 1000 Ljubljana, Slovenija  
E-pošta: stefan.dobravec@fe.uni-lj.si

**Povzetek.** Napovedovanje izidov nogometnih srečanj zaradi stohastične narave je velik izziv. Dodatne težave povzročata odsotnost standardnega nabora parametrov za napovedovanje in odsotnost strokovnega znanja. V prispevku opisana metoda temelji na uporabi modela matričnega razcepa, ki napoveduje rezultate izključno na podlagi izidov že odigranih srečanj brez uporabe strokovnega znanja, in sicer napoveduje število zadetkov moštva proti danemu nasprotniku. Uporabljen je model z upoštevanjem pristranskosti, kar omogoča učinkovitejšo karakterizacijo moštev. Metoda je testirana na primeru nedavno končanega svetovnega prvenstva. Ocenjena je z vidika uspešnosti pri napovedovanju števila zadetkov in z vidika uspešnosti pri napovedovanju izidov srečanj. Pri napovedovanju uspešnosti je bila z analizo krivulj ROC dosežena uspešnost  $AUC = 0,677$ .

**Ključne besede:** napovedovanje športnih izidov, matrični razcep

## Forecasting the football world cup results using a matrix-factorization model

Forecasting results of football matches poses a serious challenge due to the stochastic nature of the game. The absence of a standardized set of parameters and difficulty to incorporate the expert knowledge present additional problems. The forecasting method presented in the paper is based on the matrix factorization model, which uses only results of the already played matches as an input. More specifically, it forecasts the number of goals scored by a team against a given opponent. A model with biases is used to better characterize the teams. The method is tested on the case of the recently finished world cup championship. The success of forecasting the number of goals scored and the success of forecasting the match result are evaluated. The quality of the classifier is assessed by the ROC curve analysis, yielding the value of  $AUC = 0.677$ .

## 1 UVOD

Napovedovanje rezultatov športnih dogodkov je vedno zanimiva tema, tako z vidika zabave, kot tudi s popolnoma poslovnega vidika (npr. športne stave). Zato je nogomet kot eden svetovno najbolj priljubljenih športov pogosto predmet raziskav, saj po eni strani zaradi stohastične narave nudi velik raziskovalni izziv, po drugi strani pa ga običajno usmerja natančno izbrana igralna strategija [1].

Različne tehnike modeliranja nogometnega srečanja nas privedejo do različnih algoritmov za napovedovanje izidov. V literaturi [2] so postopki modeliranja razvrščeni v štiri splošne skupine: (i) empirični modeli, (ii) dinamični sistemi, (iii) statistični postopki in (iv) ekspertni sistemi. V skupini statističnih postopkov za najosnovnejši pristop velja uporaba Poissonove porazde-

litve pri napovedovanju števila doseženih zadetkov [3], izid srečanja je torej določen posredno (iz medsebojne napovedi za nastopajoči moštvi). Nasprotno pa večina drugih pristopov napoveduje izid srečanja neposredno, pri čemer pa rezultati v večini primerov ne odstopajo bistveno [4]. V skupini ekspertnih sistemov prevladujejo pristopi k modeliranju na podlagi Bayes-ovih mrež (ang. Bayesian Networks) [1] [5] [6]. Ti pristopi so praviloma kompleksni, temeljijo na številnih predpostavkah in zahtevajo veliko statističnih vzorcev [8], vendar omogočajo preprosto vključevanje znanja o domeni uporabe, zato so v napovedih praviloma natančnejši [5]. Na pristope iz preostalih dveh skupin pa v zadnjih letih tako rekoč sploh ne naletimo.

Skupna težava opisanih pristopov je nekompatibilnost nabora podatkov, uporabljenih za napovedovanje izidov. Standardnega nabora namreč ni, prav tako tudi ni enotne raziskovalne podatkovne zbirke. V literaturi [7] lahko najdemo analizo parametrov nogometnega srečanja, kljub temu opisani pristopi uporabljajo od primera do primera različne parametre, kar onemogoča pravo primerjavo njihove uspešnosti. Delno lahko vzroke teh težav pripišemo dinamičnosti nogometne igre, ki otežuje sistematično zbiranje statističnih podatkov, ki je poleg tega še omejeno le na izbrana ligaška in turnirska tekmovanja.

V prispevku je opisana metoda za napovedovanje izidov svetovnega nogometnega prvenstva, ki temelji na modelu matričnega razcepa. Tovrstni postopki so v zadnjih letih pridobili na veljavi na področju priporočilnih sistemov za multimedijske vsebine, kjer so izkazali svojo superiornost pri napovedi uporabniške izkušnje [9]. Pomembna prednost opisane metode je v tem, da model matričnega razcepa uporablja latentne parametre.

Večina do zdaj uveljavljenih pristopov namreč uporablja subjektivno določen (znanje strokovnjaka), od primera do primera različen nabor parametrov. V primeru, ko bodisi strokovnega znanja bodisi primerne nabora merljivih parametrov nimamo, so latentni parametri modela matričnega razcepa boljša izbira kot nabor parametrov, ki ga ne znamo ustrezno ovrednotiti.

Predlagana metoda temelji izključno na rezultatih (število doseženih zadetkov) že odigranih srečanj, torej brez dodatnih statističnih podatkov in strokovnega znanja. Rezultat srečanja je namreč univerzalen podatek in je vedno zabeležen. Metoda napoveduje število doseženih zadetkov posameznega moštva na določenem srečanju in posredno tudi izid srečanja.

V nadaljevanju je najprej predstavljena metodologija napovedovanja števila zadetkov, sledi opis postopka razvrščanja izidov nogometnih tekem, rezultati napovedovanja in razvrščanja za izbrani primer ter sklepne ugotovitve.

## 2 METODOLOGIJA IN PODATKI

### 2.1 Testni podatki

Algoritem je bil preizkušen na primeru nedavnega svetovnega prvenstva v nogometu v Braziliji 2014, ki se je odvijalo v turnirski obliki. Prvi del turnirja je skupinski del, kjer so sodelujoči razdeljeni v osem skupin po štiri ekipe. V skupini ekipe igrajo po sistemu vsak z vsakim, torej dve srečanja na krog, skupaj trije krogi, v osmih skupinah pa to pomeni skupaj 48 srečanj. Najuspešnejši ekipi iz vsake skupine napredujeta v izločilni del turnirja, kjer se igra po sistemu izločanja. V drugem delu je odigranih 16 tekem: 8 v osmini finala, 4 v četrtfinalu ter po dve v polfinalu in finalu).

Poleg srečanj svetovnega prvenstva smo v postopku učenja algoritma uporabili tudi prijateljska srečanja ekip udeleženk neposredno (en mesec) pred prvenstvom, in sicer skupaj 56 srečanj.

Vsi podatki, tako o že odigranih srečanjih kot tudi o srečanjih naslednjega kroga, so bili dostopni na uradni strani svetovnega nogometnega prvenstva (<http://www.fifa.com/worldcup/>). Ker v izločilnem delu turnirja neodločen rezultat ni več mogoč (igra se do končnega zmagovalca), smo uporabili rezultat po izteku rednega dela srečanja, prav tako tudi napoved velja za rezultat po izteku rednega dela srečanja.

### 2.2 Model matričnega razcepa

Modeli matričnega razcepa so se na področju priporočilnih sistemov multimedijskih vsebin izkazali kot učinkoviti pri napovedovanju uporabniške izkušnje [9]. Pri tem gre za poskus modeliranja interakcije med uporabniki in multimedijskimi vsebinami tako, da se model gradi na podlagi ocen, ki so jih uporabniki dodelili vsebinam. Model se nato uporabi za napoved uporabnikove izkušnje (zadovoljstva) s še neocenjeno vsebino.

V našem primeru je model prirejen tako, da temelji na številu doseženih zadetkov  $g_{en}$ . Opazujemo torej, koliko zadetkov izbrana ekipa  $e$  doseže proti nasprotni ekipi  $n$ . Z modelom napovemo pričakovano število zadetkov  $\hat{g}_{en}$ , torej koliko zadetkov bo izbrana ekipa dosegla proti določenemu nasprotniku. Odločitev za ta parameter utemeljujemo z dejstvom, da število zadetkov enega in drugega moštva neposredno odloča o končnem izidu srečanja. V model matričnega razcepa sicer lahko vključimo tudi druge parametre, ki vplivajo na napoved (v [10] je na primer dodana časovna komponenta), vendar smo se zadovoljili z osnovnim modelom z upoštevanjem pristranskosti (ang. bias):

$$\hat{g}_{en} = \mu + b_e + b_n + \mathbf{q}_n^T \mathbf{p}_e. \quad (1)$$

V modelu sta ekipi ( $e$  in  $n$ ) predstavljeni z vektorjema ( $\mathbf{p}_e$  in  $\mathbf{q}_n$ ) v prostoru latentnih parametrov tako, da njun produkt ponazarja tisti del napovedi, ki je posledica neposredne 'interakcije',  $\mu$  pomeni povprečje vseh doseženih zadetkov,  $b_e$  in  $b_n$  pa odstopanja izbrane in nasprotne ekipe od tega povprečja.

Prostor latentnih parametrov je v postopku učenja modela postavljen tako, da ustrezno povzame tiste vidike srečanj iz učne množice  $\kappa_T$ , ki najbolje pojasnijo odstopanja v številu doseženih zadetkov izbrane ekipe proti različnim nasprotnikom. Za učenje je uporabljen stohastični gradientni postopek (ang. Stochastic Gradient Descent), ki temelji na zmanjševanju kvadratične napake napovedi za znane rezultate iz  $\kappa_T$ :

$$\min \sum_{(e,n) \in \kappa_T} (g_{en} - \hat{g}_{en})^2 + \lambda (\|\mathbf{p}_e\|^2 + \|\mathbf{q}_n\|^2 + b_n^2 + b_e^2), \quad (2)$$

Če rezultate iz  $\kappa_T$  predstavimo s  $(i, j, G_i, G_j)$ , kjer sta  $i$  in  $j$  ekipi,  $G_i$  in  $G_j$  pa števili zadetkov teh ekip, potem za vsak rezultat izluščimo dve vrednosti, ki ju upoštevamo v enačbi:  $g_{ij} = G_i$  ter  $g_{ji} = G_j$ . Drugi del vsote je namenjen regularizaciji, ki jo nadzoruje parameter  $\lambda$ .

Testna množica algoritma so pari moštev ( $i$  in  $j$ ), ki igrajo srečanje v naslednjem krogu tekmovanja  $\kappa_E = \{(i, j)\}$ .

### 2.3 Napovedovanje izida

Uporabljeni model matričnega razcepa ne napoveduje neposredno izida, temveč pričakovano število zadetkov ( $\hat{g}_{en}$ ) izbrane ekipe ( $e$ ) proti nasprotniku ( $n$ ). Za vsako srečanje  $(i, j)$  iz  $\kappa_E$  tako dobimo dve napovedi:  $\hat{G}_i = \hat{g}_{ij}$  in  $\hat{G}_j = \hat{g}_{ji}$ . Za napovedovanje izida je treba načrtovati še razvrščevalnik ( $\mathcal{C}$ ), ki pare napovedi razvrsti v enega naslednjih treh razredov: zmaga prvega moštva v paru (1), zmaga drugega moštva v paru (2) in neodločen rezultat (0):

$$(i, j, \hat{G}_i, \hat{G}_j) \xrightarrow{\mathcal{C}} \{(1), (2), (0)\}. \quad (3)$$

Za navedene oznake razredov smo se odločili, ker so splošno uveljavljene tako pri statistični analizi nogometnih rezultatov kot na področju nogometnih stav.

Najpreprostejši način razvrščanja predstavlja primerjava na najbližje celo število zaokrožene napovedi števila zadetkov:

$$C: c = \begin{cases} (1) & ; \text{round}(\hat{G}_i) > \text{round}(\hat{G}_j) \\ (2) & ; \text{round}(\hat{G}_i) < \text{round}(\hat{G}_j) \\ (0) & ; \text{round}(\hat{G}_i) = \text{round}(\hat{G}_j). \end{cases} \quad (4)$$

Takšno razvrščanje ima težave, ko sta napovedani vrednosti  $\hat{G}_i$  in  $\hat{G}_j$  tik ob meji zaokroževanja, vendar vsak na svoji strani (npr. 1,51 in 1,49). V takem primeru je kljub zelo podobni napovedi števila zadetkov napovedan izid zmagaja in ne neodločen izid.

Več uspeha lahko pričakujemo, če za razvrščanje uporabimo katerega od uveljavljenih postopkov razvrščanja. V raziskavi je uporabljen Bayesov naivni razvrščevalnik (ang. Naive Bayes). Ta spada med bolj robustne in v večini primerov uspešne razvrščevalnike, čeprav temelji na domnevi o nekoreliranosti vhodnih značilnk (v tem primeru  $\hat{G}_i$  in  $\hat{G}_j$ ) [12].

Medtem ko v prvem primeru ne potrebujemo učne faze, pa je ta seveda potrebna pri razvrščanju z uporabo Bayesovega naivnega razvrščevalnika. Uspešnost razvrščanja je seveda odvisna tudi od velikosti učne množice, zato izidi v tem primeru niso napovedani 'sproti'. Uspešnost napovedovanja je namreč odvisna tudi od velikosti učne množice, ker pa nabor podatkov obsega vsega 64 izidov srečanj, je pri določanju učnega in testnega nabora uporabljeno navzkrižno pregibanje (ang. cross-folding) nabora podatkov. V tem primeru je uporabljenih 8 pregibov (8-folding), da je v vsakem enako število vzorcev, ki so po pregibih razporejeni naključno tako, da je porazdelitev glede na razred enaka kot v celotnem naboru.

#### 2.4 Ovrednotenje

Ocena uspešnosti napovedovanja števila doseženih zadetkov je podana s srednjo kvadratično napako (ang. Root Mean Square Error - RMSE) med napovedanim in dejanskim številom zadetkov parov ekip ( $e, n$ ) iz testnega niza  $\kappa_E$ . Za primerjavo med različnimi nabori podatkov, kjer lahko pride do večjih medsebojnih statističnih odstopanj, pa moramo uporabiti njeno normalizirano obliko (NRMSE) [15]:

$$\text{RMSE} = \sqrt{\frac{\sum_{\kappa_E} (g_{en} - \hat{g}_{en})^2}{N_{\kappa_E}}}, \quad (5)$$

$$\text{NRMSE} = \frac{\text{RMSE}}{\bar{g}}. \quad (6)$$

Uspešnost napovedi razvrščanja izidov srečanj je prikazana s pomočjo matrice pravih in napačnih razvrstitev (ang. Confusion Matrix). V našem primeru gre za razvrščanje v tri razrede, kot je ponazorjeno z enačbo 3. V matriki (glej tabelo 1) so s strešico (npr.  $\hat{1}$ ) označeni napovedani razredi, brez strešice pa dejanski (doseženi) razredi. Vrednosti  $p_{ij}$  pomenijo deleže izidov, ki smo jih razvrstili v  $i$ -ti razred, dejansko pa spadajo v  $j$ -ti razred.

Na podlagi matrice pravih in napačnih razvrstitev je izpeljana cela vrsta mer uspešnosti razvrščevalnika, ki so pogosto odvisne od področja uporabe. Smiselno je, da uporabimo tiste mere, ki jih je za dani primer preprosto interpretirati [16]. Najpogosteje se uporabljajo mere: skupna uspešnost razvrščanja (ang. Overall Success Rate - OSR), natančnost (ang. Precision) in priklic (ang. Recall). Uspešnost razvrščanja, ki poda odstotek pravilno napovedanih izidov, izračunamo kot delež pravilno napovedanih izidov:

$$\text{OSR} = p_{11} + p_{22} + p_{33}. \quad (7)$$

Za natančnejši vpogled v učinkovitost razvrščevalnika se uporabljata še natančnost in priklic, ki ovrednotita uspešnost razvrščanja v posamične razrede. Natančnost pove, koliko izidov izbranega tipa smo napovedali pravilno, priklic pa pove, koliko izidov, razvrščenih v izbrani razred, je bilo napovedanih pravilno:

$$P_i = \frac{p_{ii}}{p_{i1} + p_{i2} + p_{i3}}, R_i = \frac{p_{ii}}{p_{1i} + p_{2i} + p_{3i}}. \quad (8)$$

	(1)	(0)	(2)
$\hat{1}$	$p_{11}$	$p_{12}$	$p_{13}$
$\hat{0}$	$p_{21}$	$p_{22}$	$p_{23}$
$\hat{2}$	$p_{31}$	$p_{32}$	$p_{33}$

Tabela 1: Matrika pravih in napačnih razvrstitev

Tovrstne splošne ocene uspešnosti imajo nekaj težav, in sicer kažejo le na uspešnost razvrščanja pri izbranih pogojih (naučenih odločitvenih pragih) in so občutljive na neuravnoteženost razredov (ang. Class Skew) [11] [13]. Fawcett [11] zato predlaga uporabo krivulj ROC (ang. Receiver Operating Characteristics), ki prikazuje razmerje med dvema parametroma: deležem pravih razvrstitev v izbrani razred (ang. True Positive Rate) in deležem napačnih razvrstitev v izbrani razred (ang. False Positive Rate). Kot mero kakovosti razvrščevalnika pa isti avtor predlaga površino pod krivuljo ROC - AUC (ang. Area Under Curve). Za naključno razvrščanje, ki ga kaže diagonala v diagramu ROC tako velja, da je  $\text{AUC} = 0,5$ .

Ker pri Bayesovem naivnem razvrščevalniku uporabljamo navzkrižno pregibanje nabora podatkov z osmimi cikli (8-folding), kot rezultat dobimo 8 krivulj ROC, ki jih, kot je predlagano v [13], predstavimo s povprečno krivuljo in intervalom zaupanja, kakovost razvrščevalnika pa z AUC povprečne krivulje.

Krivulje ROC so v osnovi namenjene analizi razvrščanja v dva razreda, primer v prispevku pa uporablja tri razrede. Namesto krivulje bi tako morali uporabiti šestrazsežnostni politop ter s tem bolj zapletene postopke analize [11]. Preprostejša možnost je prikaz s tremi osnovnimi krivuljami ROC, po eno za vsak razred

(kjer 'pozitivnega' pomeni izbrani razred, 'negativnega' pa preostala dva). Pri tem zato dobimo tri vrednosti AUC, za skupno oceno pa uporabimo, kot predlaga [14], uteženo povprečje:

$$AUC = \sum_{c_i \in \mathcal{C}} AUC(c_i) * w(c_i), \quad (9)$$

kjer so uteži  $w(c_i)$  deleži vzorcev v razredih  $c_i$ .

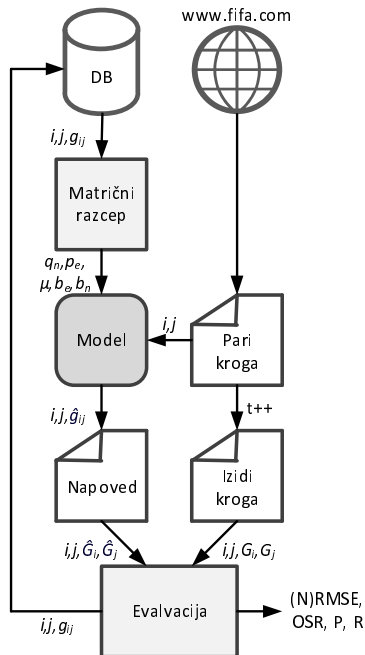
### 3 POSKUS

Potek poskusa je ponazorjen na sliki 1. Rezultate že odigranih srečanj ( $i, j, g_{ij}$ ), shranjene v lokalni podatkovni zbirki, smo uporabili v učni fazi, katere rezultat so parametri modela za napoved rezultatov naslednjega kroga srečanj ( $q_n, p_e, \mu, b_e, b_n$ ). Pare naslednjega kroga smo pridobili z uradne spletne strani, prav tako tudi rezultate po odigranem krogu. Ti so testna množica podatkov, na podlagi katerih je bilo opravljeno ovrednotenje napovedanih rezultatov, kot je to opisano v prejšnjem poglavju. Izide odigranih tekem smo nato dodali v lokalno podatkovno zbirko in ga uporabili v naslednji učni fazi.

Učna faza in faza ovrednotenja sta se ponovili po vsakem odigranem krogu. V začetnem stanju (pred začetkom prvenstva) so bili v učni fazi uporabljeni izidi pripravljalnih tekem sodelujočih ekip.

### 4 REZULTATI

Uspešnost napovedovanja števila doseženih zadelkov smo ugotavljali po vsakem odigranem krogu. Pri izračunu uspešnosti po določenem krogu so upoštewane



Slika 1: Izvedba poskusa

	(1)	(0)	(2)	P
( $\hat{1}$ )	0.219	0.188	0.109	0.424
( $\hat{0}$ )	0.078	0.047	0.078	0.230
( $\hat{2}$ )	0.078	0.031	0.172	0.611
R	0.583	0.176	0.478	OSR= 0.438

Tabela 2: Matrika pravilnih in napačnih razvrstitev pri razvrščanju z zaokroževanjem

napovedi in izidi vseh do tedaj odigranih srečanj. Ugotovitve so povzete na sliki 2.

Uspešnost napovedovanja izidov z uporabo razvrščanja z zaokroževanjem (glej enačbo 4) je prikazana v tabeli 2. Izračun je narejen po končanem prvenstvu, torej ob upoštevanju napovedi in izidov vseh srečanj.

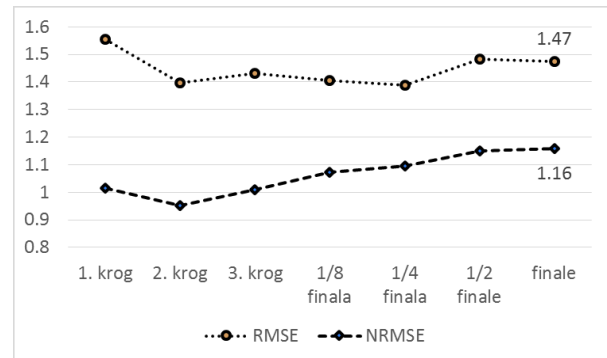
Uspešnost Bayesovega razvrščevalnika je prikazana v tabeli 3. Opazno je občutno izboljšanje natančnosti (P) razvrščanja za razred (0), priklica (R) za razred (2) ter posledično uspešnosti razvrščanja (OSR). Na sliki 3 so prikazane krivulje ROC za vsak razred posebej, kjer je senčen 95-odstotni interval zaupanja. Najslabše se obnese napovedovanje v razred (1), kjer je dobljena krivulja tudi najbližje diagonali, ki kaže naključno razvrščanje.

Za dani primer je bila ugotovljena naslednja porazdelitev razredov:  $p_{(1)} = 0,375, p_{(0)} = 0,266, p_{(2)} = 0,359$ . Po enačbi 9 tako dobimo za Bayesov razvrščevalnik

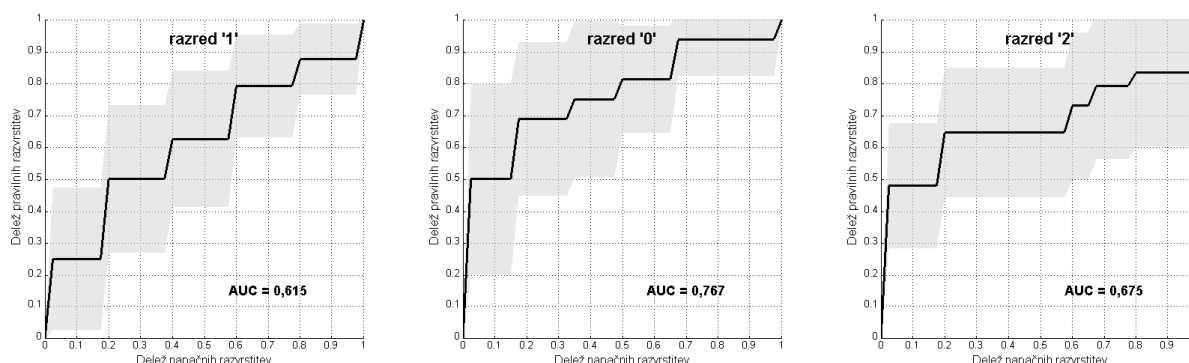
$$AUC = 0,677.$$

### 5 SKLEP

Opisana metoda temelji izključno na rezultatih (število doseženih zadelkov) že odigranih srečanj, torej brez dodatnih statističnih podatkov in strokovnega znanja. Tovrstna obravnava je še posebej primerna, ko statističnih podatkov in/ali strokovnega znanja ni na voljo oziroma ga ne znamo pravilno uporabiti.



Slika 2: Napaka pri napovedovanju števila doseženih zadelkov



Slika 3: Krivulje ROC za posamezne razrede Bayesovega razvrščevalnika

	(1)	(0)	(2)	P
$\hat{(1)}$	0.234	0.219	0.109	0.417
$\hat{(0)}$	0.031	0.031	0.016	0.400
$\hat{(2)}$	0.109	0.016	0.234	0.652
R	0.625	0.118	0.652	OSR=0.500

Tabela 3: Matrika pravih in napačnih razvrstitev pri Bayesovem razvrščevalniku

Pri napovedovanju števila doseženih zadetkov ni bilo pričakovati visoke uspešnosti, saj samo število zadetkov pri nogometu v večini primerov ni ključnega pomena, pomembnejši je namreč izid. To še posebej velja v srečanjih 'na izločanje', kjer ima zmagaz z minimalno razliko (1:0) za moštvi popolnoma enake posledice kot npr. visoka zmagaz s 7:0, kar seveda občutno vpliva na igralno taktiko.

Napovedovanje izidov s preprostim postopkom zakroževanja napovedanih vrednosti je iz že opisanih razlogov relativno neuspešno. Podrobnejši vpogled v matriko pravih in napačnih razvrstitev kaže na hude težave pri napovedovanju neodločenih rezultatov. Z razvrščanjem z uporabo Bayesovega razvrščevalnika se to izboljša, prav tako se izboljša delež pravilno napovedanih izidov (OSR).

Pomankljivost uporabljenega primera je v razmeroma majhni učni množici. Zaradi stohastične narave nogometne igre bi se namreč z večjo učno množico izboljšala učinkovitost napovedovanja. Večjo učno množico bi imeli na voljo, če bi podaljšali obdobje napovedovanja (npr. na več let), vendar bi se pri tem soočili s časovno komponento karakterizacije moštvev.

### LITERATURA

[1] B. Min, J. Kim, C. Choe, H. Eom, R. McKay, "A compound framework for sports results prediction: A football case study," *Knowledge-Based Systems*, vol. 21, n. 7, pages 551–562, 2008.  
 [2] M. Hughes, I. Franks, *The essentials of performance analysis: an introduction*, Routledge, 2007.

[3] D. Karlis, I. Ntzoufras, "Analysis of sports data by using bivariate poisson models," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, n. 3, pages 381–393, Wiley Online Library, 2003.  
 [4] J. Goddard, "Regression models for forecasting goals and match results in association football," *International Journal of Forecasting*, vol. 21, n. 2, pages 331–340, Elsevier, 2005.  
 [5] A. Joseph, N. Fenton, M. Neil, "Predicting football results using Bayesian nets and other machine learning techniques," *Knowledge-Based Systems*, vol. 19, n. 7, pages 544–553, Elsevier, 2006.  
 [6] A. Constantinou, N. Fenton, M. Neil, "pi-football: A Bayesian network model for forecasting Association Football match outcomes," *Knowledge-Based Systems*, vol. 36, pages 322–339, 2012.  
 [7] D. Bunker, R. Thorpe, "A model for teaching games in the secondary school," *Bulletin of Physical Education*, n. 10, pages 9–16, 1982.  
 [8] G. Kumar, *Machine Learning for Soccer Analytics*, Cambridge University Press, MSc thesis, KU Leuven, 2013.  
 [9] Y. Koren, R. Bell, C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, n. 8, pages 30–37, IEEE 2009.  
 [10] Y. Koren, "Collaborative Filtering with Temporal Dynamics," *Communications of the ACM*, vol. 53, n. 4, pages 89–97, 2010.  
 [11] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, n. 8, pages 861–874, Elsevier, 2006.  
 [12] I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, n. 22, pages 41–46, 2001.  
 [13] F. J. Provost, T. Fawcett, R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," *proceedings of ICML-98*, vol. 98, pages 445–453, 1998.  
 [14] F. J. Provost, P. Domingos, "Well-trained PETs: Improving probability estimation trees," *Citeseer*, 2000.  
 [15] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, n. 1, pages 5–53, ACM, 2004.  
 [16] V. Labatut, H. Cherifi, "Evaluation of Performance Measures for Classifiers Comparison," *Ubiquitous Computing and Communication Journal*, vol. 6, pages 21–34, 2011.

Štefan Dobravec je leta 2003 magistriral s področja elektrotehnike na Univerzi v Ljubljani. Je asistent na Fakulteti za elektrotehniko. Njegovo področje raziskovanja vključuje uporabniku prilagojene vmesnike, priporočilne sisteme za multimedijske vsebine ter obdelavo socialnih signalov.