

Analiza šuteva za jedan i dva poena na utakmicama Srpske ženske A lige u sezoni 2010/11 Data Mining metodama

Branko Markoski, Dejan Lacmanović, Miodrag Ivković, Dragica Radosav
University of Novi Sad, Technical Faculty "Mihajlo Pupin"
Zrenjanin, Republic of Serbia
markonins@yahoo.com, dlacman@yahoo.com,
misa.ivkovic@gmail.com, dolores023@open.telekom.rs

Jovan Šetrajčić
Department of Physics
University of Novi Sad, Faculty of Sciences
Novi Sad, Republic of Serbia
bora@df.uns.ac.rs

Sadržaj— Sport je u poslednjih par decenija postao više od igre; on danas predstavlja veliki biznis, u koji se ulažu ogromna novčana sredstva. Vrlo često, pravovremena informacija znači razliku između pobeđe i poraza. Sportska statistika je najznačajnija trenerima, ali je i čitava populacija ljubitelja i poklonika sporta danas, uz pomoć masovne upotrebe tehnike i medija poput televizije i interneta, u mogućnosti da prati učinak timova i pojedinaca. U radu je izvršena analiza najvažnijeg statističkog parametra, šuteva za jedan i dva poena na utakmicama "A" ženske lige u sezoni 2010-2011 u Srbiji.

Ključne reči- statistika; sport; košarka; data mining

I. UVOD

Košarka predstavlja kompleksnu igru između dva tima čiji je cilj da postignu veći broj poena i tako pobeđe. Tokom same igre dešava se veliki broj događaja koje je jako teško, ako ne i nemoguće sve zabeležiti. Pored osnovnih događaja kao što su šutevi ka košu, skokovi, osvojene i izgubljene lopte, postoji još dosta relevantnih događaja kao što su kretnje po terenu, vrste odbrana, gradnja lopte. Košarkaški skauting i analiza sopstvenog i protivničkog tima je postao neizostavan deo pripreme za sve utakmice u profesionalnim ligama. Korišćenje data mining-a u sportu omogućava velike prednosti svojim korisnicima. Ono upravo omogućava da se sagledaju svi bitni elementi košarkaške igre i da se iz prikupljenih podataka izvuče znanje. Na taj način timovi upoznaju sebe, jer su u mogućnosti da vide šta im je potrebno za pobeđu, gde najčešće greše, koje elemente igre treba da poprave, i dr. [1] Pored poznavanja sebe, upotrebom košarkaške analize i data mining-a kao najvišeg stepena analize, timovi mogu da upoznaju svoje protivnike i da sprema za sledećeg protivnika odnosno taktiku za narednu utakmicu. Data mining u sportu doživljava nagli rast poslednjih godina. Alati i tehnike koje se razvijaju imaju za cilj da što bolje izmere performanse kako igrača kao pojedinca, tako i celog tima. Ove nove metode merenja performansi privlače pažnju najvećih sportskih udruženja, jer se u sportu danas nalazi velika količina novca. Pre prednosti koje je doneo data mining, sportske organizacije su se gotovo isključivo oslanjale na stručnost i upućenost ljudi koji su bili zaduženi za skauting. Statistika igra veoma bitnu ulogu u samom skautingu, jer na osnovu nekih događaja koji su se desili u prošlosti skauti

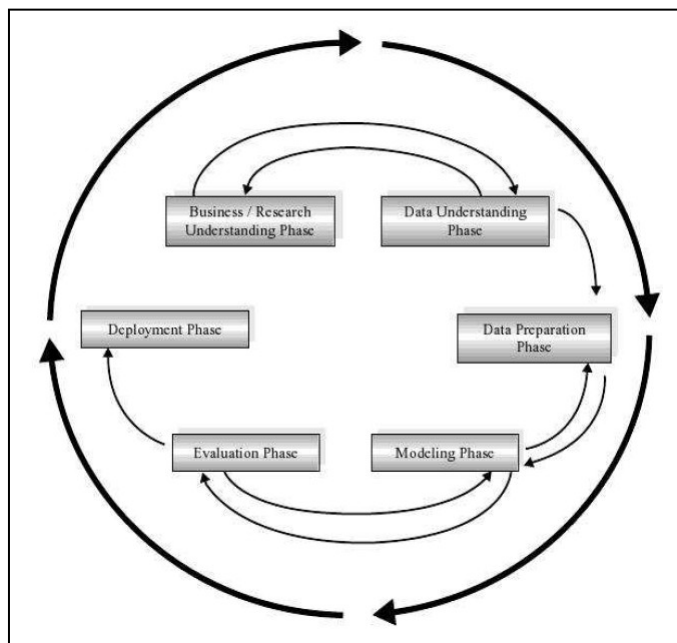
pokušavaju da precizno predvide buduće događaje. U prošlosti, košarkaška statistika je bila luksuz dostupan samo velikim profesionalnim klubovima. Za prosečnog trenera, statistika je predstavljala pravu noćnu moru, zahtevajući veliku količinu vremena i truda, najpre u samom prikupljanju statističkih podataka, a zatim i u ručnom izračunavanju različitih zbirnih statističkih parametara. Za većinu trenera statistika jednostavno nije bila vredna tolikog truda. [2].

Međutim, računari su sve to promenili. Oni su skinuli veliki teret i odgovornost za vođenje statistike sa trenera, istovremeno im pružajući pregršt informacija o kojima su oni do pre samo petnaestak godina mogli da sanjaju. I najbolje od svega, rasprostranjenost i relativno niska cena računara i softvera omogućuje dostupnost ovih informacija svima. Kako je rasla količina prikupljenih podataka, sportske organizacije su se okrenule ka pronalaženju praktičnijih metoda kako bi dobile znanje iz podataka. [8]-[9] Ovo je u početku vodilo ka angažovanju statističara koji bi za datu organizaciju omogućavali bolja merenja performansi i donošenje ispravnijih odluka. Drugi korak je bio da se pronadu još praktičnije metode kako bi se izvuklo dragoceno znanje odnosno da se krene sa upotrebom tehnika data mining-a. Pravilno primenjene tehnike data mining-a mogu rezultovati boljim performansama celog tima tako što će pripremiti igrače na određene situacije, identifikovati individualni doprinos pojedinih igrača, proceniti igru protivničkog tima i istražiti bilo kakve slabosti. Da bi mogao da se primeni data mining, pre svega su potrebni podaci. U zavisnosti od količine i bogatstva raspoloživih podataka moguće je izvući i odgovarajuće znanje. [3]-[6]

II. DATA MINING

Postoji nastojanje u nekim kompanijama, u zavisnosti od inertnosti određenog odseka, da pristupe data mining-u na svoj način. Međuindustrijski standard je očigledno bio neophodan i to standard koji je nezavistan od grane industrije, nezavistan od alata i nezavistan od primene. Međuindustrijski standardni proces za data mining (CRISP-DM) je razvijen 1996. godine od strane analitičara koji su predstavljali DaimlerChrysler, SPSS i NCR. CRISP predstavlja neprofitni standard koji je slobodan za korišćenje, a služi za uklapanje data mining-a u opštu strategiju rešavanja problema za poslovne i istraživačke

svrhe. Najčešći zadaci koji se postavljaju pred data mining su: Deskripcija, Estimacija, Predikcija, Klasifikacija, Klasterovanje i Asocijacija. Prema CRISP-DM data mining-u projekat ima životni ciklus koji se sastoji od šest faza kao što je prikazano na slici 1. Treba primetiti da je tok faza promenljiv. Odnosno, sledeća faza često zavisi od izlaza pridruženih prethodnoj fazi. Najvažnije zavisnosti između faza su označene strelicama. Npr. pretpostavimo da se nalazimo u fazi modelovanja. U zavisnosti od ponašanja i karakteristika modela, možda se moramo vratiti u fazu pripreme podataka za dalje precišćavanje pre prelaska na fazu ocene modela. [5].



Slika 1. CRISP-DM iterativni proces

Iterativna priroda CRISP-DM je predstavljena spoljnim krugom na slici broj 1. Često, rešenje određenog poslovnog ili istraživačkog problema vodi ka daljim pitanjima, koja mogu biti napadnuta koristeći isti opšti proces kao i pre. Lekcije naučene na prethodnim projektima treba koristiti kao ulaze u novim projektima. U nastavku su date konture svake faze. Prema Gartner Grupi, "Data mining predstavlja proces otkrivanja novih smislenih korelacija, šablona i smernica tako što posmatra velike količine podataka koje se nalaze u skladištima, a sve to koristeći tehnologije za prepoznavanje šablona kao i statističke i matematičke tehnike". [4] Postoje i drugačije definicije:

- "Data mining predstavlja analizu (najčešće velikog) posmatranog skupa podataka da bi pronašli nesumnjive veze i da bi sumirali podatke na nove načine koji su istovremeno razumljivi i korisni vlasniku podataka" (Hand).
- „Data mining predstavlja interdisciplinarnu granu koja spaja tehnike od mašinskog učenja, prepoznavanja šablona, statistike, baze podataka i vizuelizacije da bi zajedno rešili pitanje dolaska do informacija iz velikih baza podataka" (Simoudis).

Trenutni izuzetni rast na polju data mining-a i otkrivanja znanja je podstaknut skupom različitih faktora:

- eksplozivan rast u prikupljanju podataka, kao što je skener u supermarketu.
- odlaganje podataka u skladišta, tako da cela firma ima pristup ka aktuelnoj verziji baze.
- dodatni pristup podacima preko Web-a ili lokalnih mreža.
- takmičarski pritisak da bi se povećao udeo na tržištu u uslovima globalne ekonomije.
- razvoj komercijalnih data mining softvera.
- konstantan porast kompjuterske moci i skladišnih kapaciteta.

III. STANDARDNI STATISTIČKI ELEMENTI

Košarka je veoma dinamična igra i u toku utakmice moguće je pratiti veliki broj događaja na terenu. S obzirom da je deo zadatka ovog rada prikupljanje i obrada standardnih statističkih elemenata u košarci, potrebno je definisati koji su to standardni statistički elementi. U ovu svrhu korišćen je priručnik za statističare Svetske košarkaške federacije FIBA. [10]-[11] U njemu su izdvojeni sledeći elementi igre:

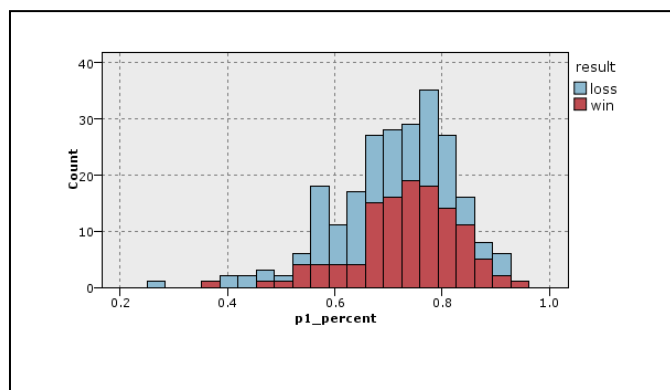
1. uspešno izvedeno slobodno bacanje
2. neuspešno izvedeno slobodno bacanje
3. uspešan šut za 2 poena
4. neuspešan šut za 2 poena
5. uspešan šut za 3 poena
6. neuspešan šut za 3 poena
7. skok u odbrani
8. skok u napadu
9. asistencija
10. ukradena (osvojena) lopta
11. izgubljena lopta
12. lična greška (faul)
13. tehnička greška
14. blokada
15. vreme provedeno u igri (minutaža)

Prikupljanjem podataka o navedenim elementima igre moguće je generisati standardne statističke izveštaje. Međutim, da bi statistika bila potpunija i da bi mogli da se realizuju dodatni statistički izveštaji, potrebno je pojedine elemente dopuniti. Mogućnosti za dopunu pojedinih elemenata ogledaju se u sledećem :

1. Prilikom šuta, uspešnog ili neuspešnog, za 2 ili 3 poena, moguće je akciju dopuniti informacijom sa koje pozicije na terenu je šut upućen.

2. Akciju izgubljene lopte moguće je dopuniti informacijom na koji način je lopta izgubljena (loše vođenje, loše dodavanje ...).
3. Prilikom lične greške moguće je evidentirati i nad kojim igračem je lična greška napravljena.
4. Prilikom blokade, kao i kod lične greške, moguće je evidentirati i nad kojim igračem je blokada napravljena.

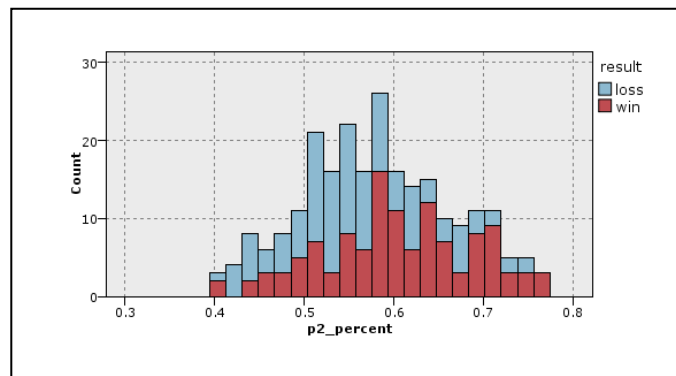
Tokom trajanja igre dešavaju se različiti prekršaji, koji se najčešće nazivaju "faulovi". Ometajući fizički kontakt (lični faul) je situacija u kojoj igrača na nedozvoljen način sprečava protivnički igrač da postigne koš. U ovakvoj situaciji se košarkašu, koji je bio u napadu, dozvoljava da sa linije za slobodna bacanja postigne poene. Ukoliko je šut bio za tri poena, igrač će dobiti tri slobodna bacanja, a ako je šut bio za dva poena, on će dobiti priliku da dva puta sa linije za slobodna bacanja uputi šut. Svaki pogodak sa ove linije donosi jedan poen ekipi. Prilikom šuta za dva poena, tri poena, kao i prilikom slobodnih bacanja dešava se da igrač postigne koš ili da promaši. Odnos postignutih koševa sa ukupnim brojem pokušaja se naziva procenat šuta. U košarkaškoj statistici se posebno određuje procenat šuta za jedan poen, procenat šuta za dva poena i procenat šuta za tri poena. Na slici 2 je prikazan uticaj šuta za jedan poen na krajnji ishod utakmice. Sa grafika se vidi da broj pobeda naglo skače kada procenat šuta za jedan poen pređe granicu od 62%. Ako je procenat šuta za jedan poen ispod 65% timovi u većini slučajeva gube utakmicu. Na osnovu dobijenih zavisnosti može se reći da šut za jedan poen ima značajan uticaj na konačan ishod utakmice. Sa grafika se može videti da se tokom sezone dešavalo da odnos procenta šuta i ishoda utakmice odstupa od opštih pravila. Naime, desilo se da je ekipa imala šut za jedan poen preko 85%, a da je ipak izgubila utakmicu. Takođe, odigrana je utakmica u kojoj je ekipa imala procenat šuta za jedan poen samo 32%, a da je zabeležila pobjedu. [7]



Slika 2. Uticaj procenta šuta za jedan poen na konačan ishod utakmice u Prvoj A ženskoj ligi Srbije

Statistički minimum za procenat za jedan poen iznosi 24%, a statistički maksimum iznosi 95.2%. Srednja vrednost za ovaj šut, kada se posmatraju sve utakmice u ligi, iznosi 71.4%. Standardna devijacija iznosi 0.11. Iz ovog se može zaključiti da u u Prvoj A ženskoj ligi Srbije procenat šuta za jedan poen ne utiče u tolikoj meri na krajnji rezultat utakmice. Takođe sa padom procenta, broj poraza se postepeno povećava.

Statistički minimum za procenat za jedan poen iznosi 0%, što znači da ekipa nije realizovala ni jedno slobodno bacanje, a statistički maksimum iznosi 95.2%. Srednja vrednost za ovaj šut, kada se posmatraju sve utakmice u ligi, iznosi 66.8%. Standardna devijacija iznosi 0.125. Za svakog trenera šut za jedan poen je veoma bitan. Oni u svakoj pauzi na treningu traže od igrača da vežbaju slobodna bacanja, kako bi taj segment košarkaške igre podigli na viši nivo.



Slika 3. Uticaj procenta šuta za dva poena na konačan ishod utakmice u Prvoj A ženskoj ligi Srbije

Sa grafika se vidi da je broj pobeda veći kada je procenat šuta za 2 poena preko 55%. Ispod procenta šuta za dva poena od 58%, broj poraza je veći. Iz ovoga se može zaključiti da procenat za dva poena utiče na ishod utakmice.

I kod procenta šuta za dva poena postoje određeni izuzeci u odnosu na opšte pravilo. Dešavalo se da ekipe imaju procenat šuta za dva poena preko 61%, a da ipak izgube utakmicu. Takođe se dešavalo da ekipa sa samo 39% šuta za dva poena dobije utakmicu. Statistički minimum za procenat za dva poena iznosi 39.5%, a statistički maksimum iznosi 83%. Srednja vrednost za ovaj šut, kada se posmatraju sve utakmice u ligi, iznosi 58.2%. Standardna devijacija iznosi 0.083. Procenat šuta za dva poena ima višu minimalnu vrednost, a manju maksimalnu vrednost u odnosu na šut za jedan poen.

Posmatrajući grafik za žensku ligu može se videti da je broj pobeda veći kada kada je procenat šuta za dva poena preko 50%. Ispod procenta šuta za dva poena od 40%, broj poraza je značajno veći. Iz grafika se može zaključiti da ovaj procenat značajno utiče na ishod utakmice.

IV. ZAKLJUČAK

Košarkaška igra napreduje velikom brzinom. Broj kvalitetnih igrača i timova nezadrživo raste. Na visokim nivoima takmičenja ne postoje timovi koji mogu pre utakmice računati na sigurnu pobjedu. Dobra priprema za utakmicu donosi prevagu između prosečnosti i vrhunskih rezultata. Danas živimo u svetu informacija. Osnovna obeležja informacije, koja definišu njen kvalitet, a samim tim i vrednost, su : tačnost, potpunost, razumljivost i pravovremenost. Bolje je nemati informacije uopšte nego imati netačne, nepotpune, nerazumljive ili nepravovremene informacije. Pregledajući izveštaje o utakmicama svoje ekipe kao i o utakmicama suparničke ekipe pred utakmicu, trenerima je omogućeno da postave odgovarajuću taktiku uzimajući u obzir vrline i mane

svoje kao i suparničke ekipe. U toku utakmice pregledanjem dostupnih izveštaja treneri analiziraju timove i igrače i uviđanjem grešaka momentalno skreću pažnju svojim igračima na ideju o taktici za dalji toku utakmice. Pomoću tehnika data mining-a moguće je uraditi analizu kako pojedinih igrača, tako i tima kao celine. Na osnovu modelovanja podataka iz Prve A ženske košarkaške lige Srbije došlo se do saznanja o načinu igre i odlučujućim elementima koji utiču na konačan ishod utakmice. Pored njih, važno je da se procenat šuta za dva poena održi na visokom nivou odnosno da se ne promašuju "ziceri".

ACKNOWLEDGEMENTS

This work was partially supported by the Serbian Ministry of Education and Sciences (Grant No: 171039).

LITERATURA

- [1] Lyons, K. "Data Mining and Knowledge Discovery", Australian Sports Commission Journals 2, 2005
- [2] O'Reilly, N., P. Knight "Knowledge Management Best Practices in National Sport Organizations". International Journal of Sport Management and Marketing 2(3), Pages: 264-280. 2007
- [3] Schumaker R., Solieman O., Chen H. "Sports Data Mining", Springer 1st edition, 2010
- [4] Stefani, R. "A Taxonomy of Sports Rating Systems". IEEE Transactions on Systems, Man, and Cybernetics - Part A 29(1): 116-120. 1999
- [5] Choo, C. W. "The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions". International Journal of Information Management 16(5): 329-340. 1996
- [6] Fieitz, L. & D. Scott. "Prediction of Physical Performance Using Data Mining". Research Quarterly for Exercise and Sport 74(1): 1-25. 2003

- [7] Dean Oliver, "Basketball on paper – Rules and tools for performance analysis", Brassey's, Washington DC, 2005
- [8] Scouting America's Top Basketball Programs, Volume 1, Jamie Angeli, 2003
- [9] Scouting America's Top Basketball Programs, Volume 2, Jamie Angeli, 2004
- [10] Basketball on paper – Rules and tools for performance analysis, Dean Oliver, Brassey's, Washington DC, 2004
- [11] Analiza i učenje košarkaške igre, Slavko Trninić, 1996

ABSTRACT

Sport, in the last few decades, has become more than a game, it is now a big business in which huge funds are invested. Very often, timely information means the difference between victory and defeat. Sports statistics is very important for coaches, but also the whole population of fans and sports enthusiasts, with the help of mass media and the usage of techniques such as television and the Internet, are able to monitor the performance of teams and individuals. This paper is an analysis of the most important statistical parameters, and one and two points shots from games from a women's "A" league in the 2010-2011 season in Serbia.

ANALYSIS OF ONE AND TWO POINT SHOTS FROM GAMES FROM A SERBIAN WOMEN'S A LEAGUE IN SEASON 2010/11 USING DATA MINING METHODS

Branko Markoski, Dejan Lacmanović, Miodrag Ivković,
Dragica Radosav, Predrag Pecev