

Predviđanje broja poena Euroleague igrača

Projekat u okviru kursa Mašinsko učenje
Matematički fakultet

Nikola Belaković 1023/2023

27. jun 2024.

Sažetak

Učinak Euroleague igrača varira od sezone do sezone, čak i od utakmice do utakmice. Jasno je da bi metoda koja uspešno predviđa i analizira buduće performanse ovih igrača na terenu značajno pomogla trenerima pri odabiru tima.

U ovom radu istražen je problem predviđanja broja poena koje će igrač Euroleague postići na osnovu statistika sa prethodnih utakmica. Koristeći metode mašinskog učenja, razvijen je model koji analizira različite varijable, kao što su broj poena, skokova, asistencija i druge relevantne statistike iz prošlih mečeva, kako bi predvideo učinak igrača u narednim utakmicama.

Pristup uključuje prikupljanje i obradu podataka, izbor relevantnih karakteristika, obuku modela i evaluaciju performansi modela na testnim podacima. Posebna pažnja posvećena je primeni ovog modela u kontekstu klađenja, gde je cilj precizno tipovati da li će igrač postići veći ili manji broj poena od granice koju postavljaju bukmejkere.

Rezultati istraživanja pokazuju da model može pružiti značajne uvide i biti koristan alat za ljubitelje sportskog klađenja, povećavajući šanse za donošenje boljih odluka. Dalji rad će se fokusirati na unapređenje tačnosti modela i istraživanje dodatnih faktora koji mogu uticati na performanse igrača.

Sadržaj

1	Uvod	3
2	Prikupljanje, obrada i analiza podataka	3
2.1	Opis korišćenog skupa podataka	3
2.2	Obrada podataka	4
2.3	Analiza podataka	5
3	Treniranje, evaluacija i testiranje modela	6
3.1	Priprema podataka	6
3.2	Korišćeni modeli i izbor najboljeg	6
3.3	Treniranje i evaluacija	7
3.4	Testiranje modela i distribucija predviđenih poena	7
3.5	Zaključak	7
4	Analiza efikasnosti modela u sportskom klađenju	8
4.1	Regresija	8
4.2	Klasifikacija	8
4.3	Upoređivanje modela	8
4.4	Prikaz rada modela	8
5	Zaključak	9
	Literatura	10

1 Uvod

Košarka je dinamičan sport koji zahteva brze odluke, visoku fizičku spremnost i taktičku preciznost. Performanse igrača na terenu često zavise od mnogobrojnih faktora kao što su fizička pripremljenost, trenutna forma, protivnici, taktike trenera i mnogi drugi. U svetu profesionalne košarke, kao što je Euroleague, precizna analiza i predviđanje učinka igrača mogu biti ključni za uspeh timova i trenera [3].

U modernom sportu, napredna analitika i tehnologija igraju sve važniju ulogu. Metode mašinskog učenja postaju nezamenljiv alat za razumevanje kompleksnih obrazaca i predviđanje budućih performansi. Zamislite alat koji može precizno predvideti koliko će poena igrač postići u narednoj utakmici na osnovu njegovih prethodnih statistika. Takva tehnologija može značajno pomoći trenerima u donošenju informisanih odluka, unapređenju strategija tima, kao i zaludenicima u košarku i sportsko kladenje.

Cilj ovog rada je da istraži kako mašinsko učenje može biti primenjeno za predviđanje broja poena Euroleague igrača. Kroz analizu podataka iz prethodnih utakmica, razvijen je model koji koristi različite statističke pokazatelje kako bi pružio precizne prognoze. Ovaj pristup ne samo da nudi vredne uvide za trenere, već ima i praktičnu primenu u industriji sportskog kladenja, omogućavajući korisnicima da donose informisanije odluke.

U narednim delovima prezentacije, detaljnije će biti objašnjene metode prikupljanja podataka, postupak razvoja modela, kao i rezultati istraživanja i njihov značaj za košarkaški svet i šire.

2 Prikupljanje, obrada i analiza podataka

2.1 Opis korišćenog skupa podataka

Za izradu i testiranje ovog prediktivnog modela koristio sam nekoliko skupova podataka:

- **Euroliga skupovi podataka sa Kaggle-a:**
 - **euroleague_alltime.csv**: Tabela koja sadrži informacije o imenima igrača i njihovim pozicijama u timu.
 - **euroleague_box_score.csv**: Tabela pruža detaljne statistike učinka pojedinačnih igrača na svakom meču.
 - **euroleague_header.csv**: Tabela uključuje sveobuhvatne statistike učinka timova na svakom meču.
- **Prilagođeni skup podataka:**
 - **granice_igraci.csv**: Tabela koju sam sam kreirao, a koja sadrži imena igrača sa određenih mečeva i njihove granice poena određene od strane bukmejкера. Ovaj skup podataka služi kao dodatna karakteristika za naš model, potencijalno poboljšavajući njegovu tačnost i relevantnost za kladenje.

2.2 Obrada podataka

1. Filtriranje sezona:

- Prikupljene statistike obuhvataju samo sezone od 2018. godine i kasnije. Razlog za ovo je što veliki broj igrača iz ranijih perioda ima vrlo mali broj odigranih utakmica zbog drugačijeg formata takmičenja, što može dovesti do nepouzdanosti rezultata. Takođe, korišćenjem samo sezona od 2018. godine obezbeđuje se dovoljan broj instanci za obuku modela.

2. Odbacivanje nepotrebnih atributa:

- Uklonjeni su atributi koji nisu relevantni za predviđanje broja poena, čime je smanjena kompleksnost modela.

3. Čišćenje podataka:

- Igrači koji nisu odigrali ni minut na meču su izbačeni iz skupa podataka kako bi se osigurala tačnost modela.

4. Dodavanje naprednih statistika:

- Uvedeni su novi atributi kao što je true shooting percentage (TS%) za poboljšanje preciznosti modela.
- $TS\% = \frac{PTS}{2 * (FGA + 0.44 * FTA)}$
 - **PTS**: Broj postignutih poena (Points).
 - **FGA**: Broj šuteva iz igre (Field Goal Attempts).
 - **FTA**: Broj izvedenih slobodnih bacanja (Free Throw Attempts) [2].

5. Pravljenje tabela sa statistikom igrača:

- **Na poslednjih n (5) mečeva**: Kreirana tabela koja sadrži prosečnu statistiku igrača u poslednjih pet utakmica.
- **Na poslednjih n (2) meča protiv datog protivnika**: Kreirana tabela sa prosečnom statistikom igrača iz poslednja dva meča protiv određenog protivnika.
- **U dosadašnjem toku sezone**: Kreirana tabela sa prosečnom statistikom učinka igrača u tekućoj sezoni.
- **Broj poena koje ekipa prima sa date pozicije**: Kreirana tabela koja pokazuje broj poena koje tim prima sa određene pozicije u dosadašnjem toku sezone.

6. Dodavanje atributa rangiranja igrača:

- Uveden je atribut koji predstavlja rang igrača u timu na osnovu prosečnog broja poena. Ovaj atribut uzima u obzir odsustvo ključnih igrača zbog povrede ili odluke trenera, što može uticati na broj poena koje igrač postiže.

7. Defanzivni rejting protivničkog tima:

- Kreirana tabela sa defanzivnim rejtingom protivničkog tima izračunatim po formuli:
- Defanzivni rejting = $\frac{OPP_PTS}{FGA + \frac{FTA * 0.9}{2} + TO}$
 - **OPP_PTS**: Broj postignutih poena protivnika (Opponent Points).
 - **FGA**: Broj protivničkih šuteva iz igre (Field Goal Attempts).
 - **FTA**: Broj protivničkih izvedenih slobodnih bacanja (Free Throw Attempts).

- **TO:** Broj izgubljenih lopti protivnika (Turnovers) [1].

8. Ofanzivni rejting tima:

- Kreirana tabela sa ofanzivnim rejtingom tima, računatim po istoj formuli kao i defanzivni rejting, ali za tim u kojem igrač igra.

9. Procenat pobeda tima:

- Kreirana tabela koja prikazuje procenat pobeda tima u dosadašnjem delu sezone.

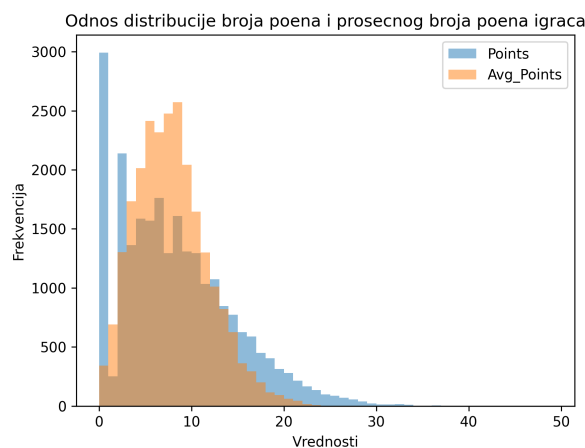
10. Konačni skup podataka:

- Sve gore navedene tabele i statistike su spojene u jedan konačan skup podataka koji se koristi za obuku i testiranje modela.

2.3 Analiza podataka

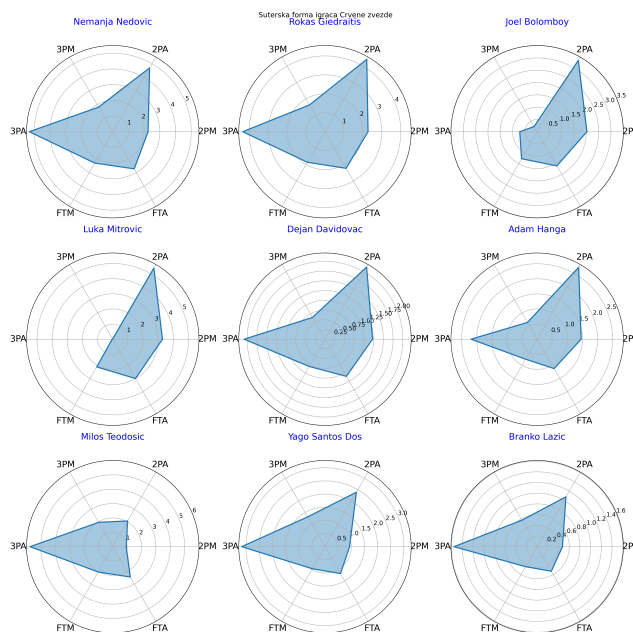
U ovoj sekciji će biti prikazani neki od mnogih grafika koji mogu biti izvučeni iz konačnog skupa podataka, a koji mogu pomoći treneru u unapređenju svog tima, kao i sportskim kladioničarima u boljoj informisanosti, koja im može pomoći u predviđanju broja poena.

Grafik na slici 1 prikazuje raspodelu broja poena koje igrači postižu na meču i prosečnog broj poena igrača u dosadašnjem toku sezone pre svakog meča. Analiza ovog grafika ukazuje na značajne varijacije u performansama igrača, što može biti izazov za predviđanje broja poena.



Slika 1: Raspodela broja poena i prosečnog broja poena.

Naredni grafik šuterske forme trenerima pruža dragocen uvid u efikasnost i šutersku sposobnost igrača tima. Analiza broja realizacija u odnosu na broj pokušaja pomaže u identifikaciji igrača koji više doprinose timu i manje troše lopte. Ovakav pogled omogućava trenerima da donose informisane odluke o rotaciji igrača i prilagođavanju strategije kako bi poboljšali efikasnost tima na terenu.



Slika 2: Šuterska forma igrača Crvene zvezde.

3 Treniranje, evaluacija i testiranje modela

U ovom poglavlju će biti govoreno o različitim konfiguracijama koje su testirane kako bi pronašli model koji najbolje uči na podacima i postiže najmanju srednju apsolutnu grešku (MAE). Fokus je stavljen na regresione modele, jer priroda problema zahteva predviđanje numeričkih vrednosti.

3.1 Priprema podataka

Prvi korak je bio razdvajanje skupa podataka na trening i test skup. Zatim je primenjena standardizacija koristeći *StandardScaler* kako bi se normalizovale vrednosti atributa. Različite konfiguracije su trenirane i evaluirane na odvojenom skupu podataka za trening, a zatim najbolja konfiguracija testirana na test skupu.

3.2 Korišćeni modeli i izbor najboljeg

Odabrana su dva modela za testiranje: *AdaBoostRegressor* i *BaggingRegressor*. Korišćena je funkcija *GridSearchCV* iz biblioteke *sklearn* za isprobavanje različitih kombinacija hiperparametara i pronalazak one koja minimizira negativnu srednju apsolutnu grešku (*neg_mean_absolute_error*). Ovaj kriterijum je korišćen jer je cilj naći model koji pravi najmanju razliku između predviđenih i stvarnih poena igrača.

3.3 Treniranje i evaluacija

GridSearchCV funkcija omogućava unakrsnu validaciju, što je pomoglo da se dobiju stabilne procene performansi modela. Za svaki model, prosleđeni su različiti parametri i njihove kombinacije, a *GridSearchCV* je vraćao konfiguracije koje su davale najbolji rezultat u smislu MAE.

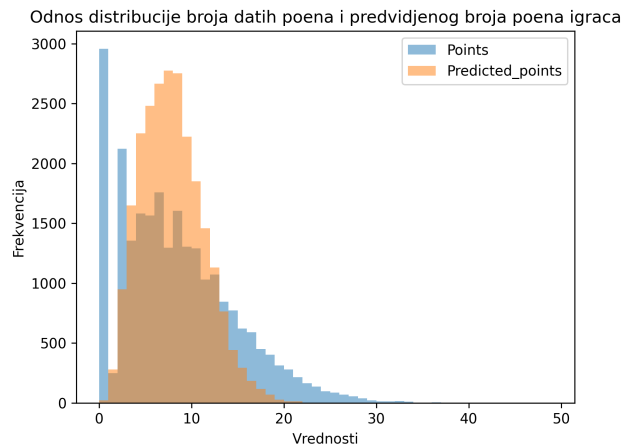
3.4 Testiranje modela i distribucija predviđenih poena

Rezultati testiranja pokazuju da model postiže zadovoljavajuće rezultate, s obzirom na složenost problema.

- *AdaBoostRegressor*:
 - R2 score: 0.300695001197817
 - Mean absolute error: 4.046131856942441
- *BaggingRegressor*:
 - R2 score: 0.30132591012485066
 - Mean absolute error: 4.037432347637131

3.5 Zaključak

Na osnovu rezultata testiranja, možemo zaključiti da je naš model sposoban da relativno precizno predvidi broj poena koje će igrači postići, uzimajući u obzir različite faktore i statistike. Grafik 3 distribucije predviđenih i datih poena pokazuje koliko je naš model blizak stvarnim vrednostima, što nam daje osnovu za dalju optimizaciju i prilagođavanje modela.



Slika 3: Dstribucija broja datih i predviđenih poena igrača.

4 Analiza efikasnosti modela u sportskom kladenju

U ovoj sekciji su prikazani rezultati testiranja klasifikacionih i regresionih modela za potrebe sportskog tipovanja na broj poena igrača.

4.1 Regresija

Prvi pristup je bio korišćenje regresionog modela za direktno predviđanje tačnog broja poena igrača. Nakon što su modeli istrenirani i testirani na test skupu koji uključuje granice poena, analizirani su rezultati kako bi se procenila tačnost predviđanja. Dobijeni rezultati su upoređeni sa stvarnim granicama i evaluirani u kontekstu profitabilnosti kladenja. Kladioničarsko testiranje regresionog modela je pokazalo da model postiže tačnost od 57.83% u predviđanju poena, što bi u praksi rezultiralo profitom od 10% uloženog novca. Ovaj rezultat pruža uvid u praktičnu primenu modela u kontekstu sportskog kladenja.

4.2 Klasifikacija

Drugi pristup kladioničarskom testiranju uključivao je klasifikaciju. Svaki igrač je klasifikovan kao "premašuje granicu" ili "ne premašuje granicu" na osnovu predviđenog broja poena u odnosu na kladioničarsku granicu. Rezultati su analizirani kako bi se izračunala tačnost klasifikacije i ukupan profit ostvaren na osnovu kladenja. Korišćen je SVC model sa raznim parametrima koji su optimizovani korišćenjem GridSearchCV. Veličina podataka je smanjena tako da su korišćeni samo podaci za koje postoji granica poena.

Matrica konfuzije na test skupu:

	Predviđeno negativno	Predviđeno pozitivno
Stvarno negativno	222	136
Stvarno pozitivno	163	150

Kladioničarsko testiranje klasifikacionog modela je pokazalo da model postiže tačnost od 55.44% u predviđanju poena, što bi u praksi rezultiralo profitom od 5% uloženog novca.

4.3 Upoređivanje modela

U poređenju ova dva pristupa, regresioni model je pokazao bolje performanse u smislu preciznosti i srednje apsolutne greške. Niži procenat tačnosti klasifikatora je delimično posledica manjeg skupa podataka koji je bio dostupan za klasifikaciju u poređenju sa skupom podataka za regresiju. Regresioni model je takođe pokazao profitabilnost od 10% u analizi sportskog kladenja, što ga čini efikasnijim za ovu specifičnu primenu.

4.4 Prikaz rada modela

U nastavku je prikazana tabela sa detaljima o svakom igraču, uključujući granicu, datum meča, stvarni broj datih poena, predviđeni broj poena i

preporuku modela da li treba igrati da će igrač postići više ili manje poena od zadate granice.

Ime igrača	Granica poena	Datum meča	Broj datih poena	Predviđeni broj poena	Tip
Timothe Luwawu-Cabarrot	10.5	2024-03-01	7	13.12	+
Joffrey Lauvergne	11.5	2024-03-01	17	10.92	-
Paris Lee	11.5	2024-03-01	5	12.23	+
Mike Scott	10.5	2024-03-01	12	10.26	-
Shavon Shields	16.5	2024-03-01	9	14.86	-
Thomas Walkup	8.5	2022-11-24	4	8.43	-
Kostas Papanikolau	8.5	2022-11-24	19	7.62	-
Alec Peters	13.5	2022-11-24	0	6.31	-
Isaiah Canaan	11.5	2022-11-24	3	6.82	-
Nemanja Nedović	13.5	2022-11-24	12	14.81	+

5 Zaključak

Predviđanje broja poena igrača je kompleksan problem koji zahteva pažljiv pristup. Koristeći napredne tehnike mašinskog učenja, kao što su ansambli modela, postigli smo značajne rezultate u tačnosti predviđanja. Važno je uzeti u obzir faktore kao što su prethodni rezultati, forma igrača, kao i timski kontekst prilikom pravljenja modela za predviđanje. Dalji rad može uključivati uvođenje dodatnih karakteristika igrača ili primenu naprednijih modela za bolje uklapanje u različite situacije. Međutim, važno je napomenuti da idealan model za predviđanje broja poena igrača ne postoji, jer postoje faktori poput sreće koji se ne mogu predvideti i kvantifikovati.

Literatura

- [1] Dr. James Gels. Calculating basketball oer and der. <https://www.coachesclipboard.net/calculating-basketball-oer-der.html>.
- [2] Basketball reference. <https://www.basketball-reference.com/about/glossary.html>.
- [3] Wikipedia. <https://en.wikipedia.org/wiki/Basketball>.