

# Umetna inteligenca

## Seminarska naloga 1

**Naloga se izvaja v paru. Zagovori seminarских nalog bodo potekali v času laboratorijskih vaj v tednu med 5. in 9. decembrom. Na zagovoru je obvezna prisotnost obeh članov tima.**

Na učilnici je podana datoteka "nbadata.txt" s podatki o odigranih tekmah rednega dela šestih sezon (od 2012-13 do 2017-18) profesionalne ameriške košarkarske lige NBA. Vsaka vrstica v objavljenih datotekah predstavlja opis ene tekme.

Posamezna tekma je opisana z naslednjimi atributi:

Ime atributa	Pomen atributa
gmSeason	oznaka sezone
gmDate	datum odigrane tekme oblike YYYY-MM-DD (letnica, mesec, dan)
(away/home)Abbr	kratica iz treh črk, ki predstavlja ime gostujoče/domače ekipe
(away/home)DayOff	število dni od prejšnje tekme gostujoče/domače ekipe
(away/home)PTS	število doseženih točk gostujoče/domače ekipe
(away/home)AST	število asistenc gostujoče/domače ekipe
(away/home)TO	število izgubljenih žog gostujoče/domače ekipe
(away/home)STL	število ukradenih žog gostujoče/domače ekipe
(away/home)BLK	število blokad gostujoče/domače ekipe
(away/home)PF	število osebnih napak gostujoče/domače ekipe
(away/home)FGA	število vseh metov iz igre gostujoče/domače ekipe
(away/home)FGM	število uspešnih metov iz igre gostujoče/domače ekipe
(away/home)2PA	število vseh metov za dve točki gostujoče/domače ekipe
(away/home)2PM	število uspešnih metov za dve točki gostujoče/domače ekipe
(away/home)3PA	število vseh metov za tri točke gostujoče/domače ekipe
(away/home)3PM	število uspešnih metov za tri točke gostujoče/domače ekipe
(away/home)FTA	število vseh prostih metov gostujoče/domače ekipe
(away/home)FTM	število uspešnih prostih metov gostujoče/domače ekipe
(away/home)ORB	število skokov v napadu gostujoče/domače ekipe
(away/home)DRB	število skokov v obrambi gostujoče/domače ekipe
(away/home)TRB	število skokov gostujoče/domače ekipe
(away/home)PTS1	število doseženih točk v prvi četrtini gostujoče/domače ekipe
(away/home)PTS2	število doseženih točk v drugi četrtini gostujoče/domače ekipe
(away/home)PTS3	število doseženih točk v tretji četrtini gostujoče/domače ekipe
(away/home)PTS4	število doseženih točk v četrti četrtini gostujoče/domače ekipe
(away/home)PTSEx	število doseženih točk v podaljšku gostujoče/domače ekipe

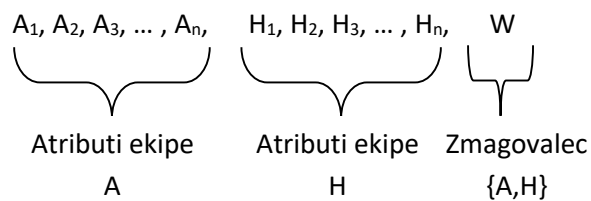
## Naloga

Namen seminarske naloge je uporaba metod strojnega učenja za analizo podane športne domene. V prvi vrsti nas zanima napovedovanje zmagovalca **pred začetkom** tekme.

Izvorni podatki opisujejo statistike **na koncu** tekme in so neprimerni za napovedovanje končnega izida, saj že vsebujejo informacijo, kaj se je zgodilo. Podatke je najprej potrebno prevesti v primerno obliko. **Potrebujemo attribute za opis karakteristik ekip, ki so znani (izračunljivi) pred začetkom tekme.**

Najpreprostejša možnost je, da posamezno ekipo predstavimo s povprečnimi statistikami, ki jih izračunamo iz **že odigranih** tekem te ekipe. Torej, sestavimo učno množico, kjer je vsaka tekma predstavljena v eni vrstici.

Na primer, pri izdelavi učne množice za klasifikacijski problem (napovedovanje zmagovalca), bomo tekmo med gostujočo ekipo A in domačo ekipo H predstavili v obliki:



kjer so  $A_1, A_2, A_3, \dots, A_n$  ( $H_1, H_2, H_3, \dots, H_n$ ) atributi izračunani iz **predhodnih tekem** ekipe A (H), W pa dejanski zmagovalec tekme. Pri izdelavi učne množice za regresijski problem (napovedovanje razlike v končnem rezultatu), bo W predstavljal razliko v doseženih točkah med domačo in gostujočo ekipo.

Seminarska naloga je raziskovalnega tipa, kar pomeni, da ni vnaprej začrtane poti do končne rešitve. Atributni prostor je možno poljubno razširiti: učenemu algoritmu lahko pomagamo tako, da med attribute dodamo odstotke uspešnih metov namesto navadnih števcov. Med attribute lahko dodamo rezultat prejšnje tekme istih tekmecev ali razmerje zmag vseh njunih medsebojnih tekem (do tega trenutka), lahko dodamo trenutno formo ekip (število zmag v zadnjih nekaj tekmah) in tako naprej. **Pozorni moramo biti le na to, da pri izračunu vrednosti atributov za določeno tekmo uporabimo samo podatke s tekem, ki so bile odigrane pred opazovano tekmo.**

Konkretne naloge, ki jih je potrebno opraviti:

1. Vizualizacija podatkov  
Izrišite nekaj zanimivih grafov, ki ilustrirajo podane podatke. Na primer: najuspešnejše ekipe glede na razmerje zmag in porazov, povprečno število metov za tri točke po posameznih sezonah in podobno.
2. Klasifikacija  
Z uporabo podanih podatkov zgradite vsaj tri klasifikacijske modele, ki bodo na podlagi konkretnega para GOSTUJOČE\_MOŠTVO - DOMAČE\_MOŠTVO napovedali verjetnost zmage domačinov.
3. Regresija  
Z uporabo podanih podatkov zgradite vsaj tri regresijske modele, ki bodo na podlagi konkretnega para GOSTUJOČE\_MOŠTVO - DOMAČE\_MOŠTVO napovedali razliko med točkami domačinov in gostov.
4. Evalvacija modelov  
Zgrajene modele ustrezno ovrednotite in predstavite dobljene rezultate.

### **Za višje ocene (9 in 10)**

Za višje ocene je potrebno dodatno zgraditi modele za:

- napovedovanje končnega zmagovalca ob podanem rezultatu (razliki v točkah med domačo in gostujočo ekipo) na koncu prve, druge oziroma tretje četrtine,
- napovedovanje končne razlike v rezultatu ob podani razliki v točkah na koncu prve, druge oziroma tretje četrtine.

Pri učenju teh modelov je seveda potrebno ustrezno razširiti učno množico s podatki z rezultati na koncu izbrane četrtine.

### **Ocenjevanje**

Na končno oceno seminarske naloge vplivajo kvaliteta zgrajenih modelov, inovativnost in elegantnost rešitve, argumentacija izbranih postopkov in razlaga dobljenih rezultatov.