

Umetna inteligenca

1. Seminarska raziskovalna naloga 2013-2014

Na učilnici sta podani datoteki "nba0809.txt" in "nba0910.txt" s podatki o odigranih tekmah rednega dela sezon 2008-2009 in 2009-2010 profesionalne ameriške košarkarske lige NBA. Vsaka vrstica v objavljenih datotekah predstavlja opis ene tekme.

Posamezna tekma je opisana z naslednjimi atributi:

DATE	datum odigrane tekme oblike YYYYMMDD (letnica, mesec, dan)
AWAY_NAME	kratica iz treh črk, ki predstavlja ime gostujoče ekipe
HOME_NAME	kratica iz treh črk, ki predstavlja ime domače ekipe
AWAY_2PA	število vseh metov za dve točki gostujoče ekipe
AWAY_2PM	število uspešnih metov za dve točki gostujoče ekipe
AWAY_3PA	število vseh metov za tri točke gostujoče ekipe
AWAY_3PM	število uspešnih metov za tri točke gostujoče ekipe
AWAY_FTA	število vseh prostih metov gostujoče ekipe
AWAY_FTM	število uspešnih prostih metov gostujoče ekipe
AWAY_TO	število izgubljenih žog gostujoče ekipe
AWAY_ORB	število skokov v napadu gostujoče ekipe
AWAY_DRB	število skokov v obrambi gostujoče ekipe
HOME_2PA	število vseh metov za dve točki domače ekipe
HOME_2PM	število uspešnih metov za dve točki domače ekipe
HOME_3PA	število vseh metov za tri točke domače ekipe
HOME_3PM	število uspešnih metov za tri točke domače ekipe
HOME_FTA	število vseh prostih metov domače ekipe
HOME_FTM	število uspešnih prostih metov domače ekipe
HOME_TO	število izgubljenih žog domače ekipe
HOME_ORB	število skokov v napadu domače ekipe
HOME_DRB	število skokov v obrambi domače ekipe
AWAY_PTS_Q1	število doseženih točk gostujoče ekipe na koncu prve četrtine
HOME_PTS_Q1	število doseženih točk domače ekipe na koncu prve četrtine
AWAY_PTS_Q2	število doseženih točk gostujoče ekipe na koncu druge četrtine
HOME_PTS_Q2	število doseženih točk domače ekipe na koncu druge četrtine
AWAY_PTS_Q3	število doseženih točk gostujoče ekipe na koncu tretje četrtine
HOME_PTS_Q3	število doseženih točk domače ekipe na koncu tretje četrtine
AWAY_PTS_Q4	število doseženih točk gostujoče ekipe na koncu četrte četrtine
HOME_PTS_Q4	število doseženih točk domače ekipe na koncu četrte četrtine
AWAY_PTS_FINAL	končno število točk gostujoče ekipe (po morebitnih podaljških)
HOME_PTS_FINAL	končno število točk domače ekipe (po morebitnih podaljških)

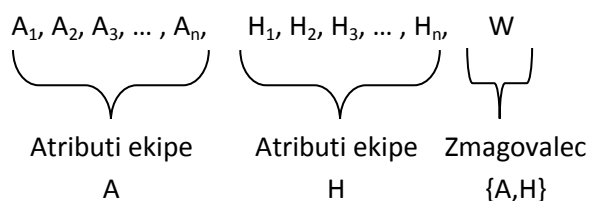
Naloge

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: gibanje rezultata s časom, najuspešnejše ekipe glede na razmerje zmag in porazov, povprečno število košev po posameznih četrtinah 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 druge (tretje) četrtine,
- napovedovanje končne razlike v rezultatu ob podani razliki v točkah na koncu druge (tretje) četrtine.

Pri učenju teh modelov je seveda potrebno ustrezno razširiti učno množico s podatki z rezultati na koncu druge (tretje) četrtine.

Ocenjevanje

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