

Umetna inteligenca

Seminarska naloga 1

Naloga se izvaja v paru. Zagovori seminarskih nalog bodo potekali v času laboratorijskih vaj v tednu med 26. in 30. novembrom. Na zagovoru je obvezna prisotnost obeh članov tima.

Na učilnici je podana datoteka "regular.txt" s podatki o odigranih tekmah rednega dela sezon 2014-15, 2015-16 in 2016-17 profesionalne ameriške košarkarske lige NBA. Vsaka vrstica v objavljenih datotekah predstavlja opis ene tekme.

Posamezna tekma je opisana z naslednjimi atributi:

SEASON	Oznaka sezone
DATE	datum odigrane tekme oblike YYYYMMDD (letnica, mesec, dan)
HOME	kratica iz treh črk, ki predstavlja ime domače ekipe
HPTS	število doseženih točk domače ekipe
H2PM	število uspešnih metov za dve točki domače ekipe
H2PA	število vseh metov za dve točki domače ekipe
H3PM	število uspešnih metov za tri točke domače ekipe
H3PA	število vseh metov za tri točke domače ekipe
HFTM	število uspešnih prostih metov domače ekipe
HFTA	število vseh prostih metov domače ekipe
HORB	število skokov v napadu domače ekipe
HDRB	število skokov v obrambi domače ekipe
HAST	število asistenc domače ekipe
HSTL	število ukradenih žog domače ekipe
HTOV	število izgubljenih žog domače ekipe
HBLK	število blokad domače ekipe
HPF	število osebnih napak domače ekipe
AWAY	kratica iz treh črk, ki predstavlja ime gostujoče ekipe
APTS	število doseženih točk gostujoče ekipe
A2PM	število uspešnih metov za dve točki gostujoče ekipe
A2PA	število vseh metov za dve točki gostujoče ekipe
A3PM	število uspešnih metov za tri točke gostujoče ekipe
A3PA	število vseh metov za tri točke gostujoče ekipe
AFTM	število uspešnih prostih metov gostujoče ekipe
AFTA	število vseh prostih metov gostujoče ekipe
AORB	število skokov v napadu gostujoče ekipe
ADRB	število skokov v obrambi gostujoče ekipe
AAST	število asistenc gostujoče ekipe
ASTL	število ukradenih žog gostujoče ekipe
ATOV	število izgubljenih žog gostujoče ekipe
ABLK	število blokad gostujoče ekipe
APF	število osebnih napak gostujoče ekipe

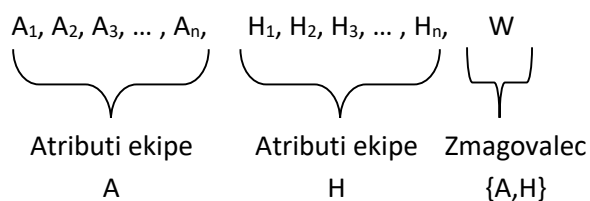
Naloge

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: učnemu 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)

V datoteki "playoff.txt" so zbrani podatki o odigranih tekmah končnice sezon 2014-15, 2015-16 in 2016-17 lige NBA. Za višje ocene je potrebno:

- ovrednotiti točnost napovedi tekem v končnici z uporabo modelov iz rednega dela lige,
- s pomočjo naučenega modela velikokrat simulirati možne razplete končnice in za vsakega udeleženca določiti verjetnost (pred začetkom playoffa), da bo končni zmagovalec lige.

Ocenjevanje

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