1. **Seminarska naloga**

***UMETNA INTELIGENCA***

1. Chart, bar chart

   Description automatically generatedA picture containing graphical user interface

   Description automatically generated**Vizualizacija podatkov**

Iz zgornjih grafov vidimo razmerje zmag med domacini in gosti.

Chart, line chart

Description automatically generated

Chart, histogram

Description automatically generatedGraf prikazuje stevilo zmag ekipe Dallas po sezonah od 2013-14 do 2017-18.

Graf prikazuje porzdelitev najboljsih ekip po razmerje zmag in porazov.

1. **Priprava podatkov**

Naprej sem atribute vseh metov in uspesnih metov zamenjal z atributi razmerjem uspesnih metov. Nato sem dodal atribut razmerje upesnosti napada. Izracunal sem ga iz razmerje stevila uspesnih zadetkov in vseh napadov. Kot napad sem stel zadetek ali zgresen met, izgubljena zoga ali ukradena zoga. Vse napade sem izracunal iz stevila vseh metov + izgubljene zoge + ukradene zoge.

Za klasifikacijo sem dodal ciljno spremnljicko ‘win’ – zmaga domacih.

Za regresijo sem dodal ciljno spremnljicko ‘diff’ – razlika rezultata.

Ko napovedujem zmagovalca po koncani cetrtini uporabim se atribut ‘diffScore’, ki ima vrednost razlike rezltata po 1/4, 2/4, 3/4.

Dodal sem se atribut forme ekipe, ki nam pove v kasni formi je ekipa po razmerje zmag zadnjih 5 odigranih iger.

Vsaka tekma v ucni mnozici predstavlja povprecje atributov zadnjih 5 iger te ekipe.

Text, letter

Description automatically generatedText

Description automatically generated

1. **Klasifikacija**

Za ucno mnozico pri klasifikacije sem vzel tekme 4 sezon, ekipe za katero napovedujemo.

Za gradnjo ucne mnozice sem napisal metode:

* makeDf (data set z vsemi atributi)
* makeTeamMean (izracuna povprecne vrednosti zadnjih 5 odigranih iger)
* makeSeasonMean (za vse odigrane tekme doda novo vrstico, ki predstavlja povprecja domace in gostujoce ekipe)

Napovedovanje za (Domaci: ekipa A, gostje: ekipa B)

Vrstice ucne mnozice predstavljajo tekme, kjer:

* Domaci == ekipa A ali
* Gostje == ekipa B

Napovedoval sem za razlicne pare ekip v sezoni 5 (2016-17).

Napovedal sem rezultat za 200 ekip.

Metode klasifikacije in napovedovanje tekem so v datoteki classification.R in prediction.R

Text

Description automatically generated

Najboljso natancnost ima model nakljucnega gozda, s tocnostjo 0.585.

Kot osnovo imamo vecniski razred, ki ima tocnost 0.565.

Nakljucni gozd ima vecjo natancnost kot vecinski razred, malo, ampak je vseeno uporaben.

Medtem, ko sta modela najblizjih sosedov (knn) in dreves neuporbna, ker imata manjso tocnost, kot vecinski klasifikator.

Text

Description automatically generated

V tem primeru smo upostevali atribut ‘diffScore;, ki nam je povedal razliko v tockah po odigrani 1/4 igre. Vidimo da je natancnost pri vseh modelih razen vecinskega klasifikatorja zelo narasla.

Se vedno je zmagal klasifikator nakljucnega gozda, ki je od vecinskega klasifikatorja boljsi kar za 25%.

1. **Regresija**

Pri regresiji sem uporabil metode drevesa, nakljucnih gozdov, knn in svm.

Kot oceno sem uporabil mae )srednja absolutna napaka) in rmae (relativna srednja napaka).

1. Drevo

Graphical user interface, text

Description automatically generated

1. Nakljucni gozd

Graphical user interface, text

Description automatically generated

1. KNN

Graphical user interface, text

Description automatically generated

1. Svm

Text

Description automatically generated

Opazimo, da se je najbolje odrezal drevo z najmanjso relativno napako 0.65, sledi mu nakljucni gozd, z napako 0.72, knn z napako 0.83 in najslabse svm z napko 0.84.