**Predikcija ishoda fudbalskih utakmica**

**Tatjana Zdravković SW/38-2014**

1. Motivation

Motivacija za ovaj projekat je bila da pokušam da odradim zadatak vezan za temu o kojoj ne znam skoro ništa, fokusirajući se na programiranje i razumevanje statistike koja stoji iza rešavanja problema.

2. Research questions

Ono što sam htela da postignem ovim projektom je da relativno uspešno pokušam da predvidim ishod jedne fudbalske utakmice, da li će jedan tim pobediti, izgubiti ili će biti nerešeno, na osnovu raznih podataka prikupljenih od 2005-2017 godine engleske premijer lige.

3. Related work

Uglavnom je praksa da se odredi određena jačina nekog kluba na osnovu prethodnih rezultata koje su postizali, datih, golova, kornera, kartona...Takođe uzima se u obzir da li se igra na domaćem terenu ili ne jer je statistički pokazano da ako je tim na domaćem terenu ima veću šansu za pobedu.

4. Methodology

Prvo sam parsirala dataset koji sam našla. Spojila sam podatke iz svih godina u jedan csv fajl i izbacila sam podatke koje sam smatrala da su mi nepotrebni, kao što su kvote klađenja i statistike vezane za klađenje. Zatim sam taj fajl učitala i dalje nastavila da ga parsiram i izbacujem ono što sam smatrala da neće doprineti poboljšanju mog projekta. Nakon toga sam uzela da računam jačinu timova na osnovu datih golova u kolu i prosečnog broja datih golova i golova primljenih u kolu I prosečnog broja primljenih golova. To sam računala i za domaći tim i za gostujući tim. Nakon toga trebalo je da se izračuna kako su prolazili na prethodnim utakmicama. Gledala sam kako su prolazili na poslednjih k utakmica jer sam po tome mogla da vidim spremnost tima te sezone, jer kao što je to uvek u sportu postoji mogućnost da tim poklekne ili se poboljša u narednoj sezone usled različitih fakltora. To je izračunato i za domaći i za gostujući tim. Sada kada sam sredila i pribavila sve podatke trebalo je da se ubace u algoritme za klasifikaciju.

5. Discussion

Opredelila sam se da isprobam RandomForest, KNN, SVM i NaiveBayes. Nakon isprobavanja svih tih algoritama utvrdila sam da najbolje radi SVM sa tačnošću 0,53, gde sam koristila accuracy\_score za tačnost, jer sam na osnovu nje mogla da vidim koliki je procenat klasa koje pogodim. Takođe radila sam i validaciju kako bi još bolje pronašla optimalniji model. Takođe sam isprobavala razne parametre u klasifikatorima kako bi ih optimizovala. Model je sačuvan u fajl kako bi se mogao koristiti za buduće predikcije.

6. References

<https://www.researchgate.net/publication/282287529_Using_Machine_Learning_to_Predict_Winners_of_Football_League_for_Bookies>

<https://www.sciencedirect.com/science/article/pii/S2210832717301485>

<http://publications.lib.chalmers.se/records/fulltext/250411/250411.pdf>

<https://pdfs.semanticscholar.org/7d1f/8ff04a87b29eddc8eb84300d98d7dd3ffe30.pdf>

<http://referaat.cs.utwente.nl/conference/14/paper/7226/predicting-sports-events-from-past-results.pdf>