**Predolog projekta iz numeričkih algoritama i numeričkog softvera**

Naziv teme

Predikcija rezultata fudbalskog meča na osnovu verovatnoća ishoda meča

Definicija problema

Cilj projekta jeste da se na osnovu kvota na ishod fudbaleske utakmice (kvota jeste koeficijent koji je u negativnoj korelaciji sa verovatnoćom mogućeg ishoda) višestrukom linearnom regresijom prognozira tačan rezultat meča između 2 ekipe, kao i minimalni i maksimalni broj golova koje koje će obe ekipe postići.

Metodologija

Zavisne varijable

Broj postignutih golova

Nezavisne varijable

*Predviđanje*: kvota na pobedu tima, kvota na remi, kvota na poraz tima

*Statistika tima*: domaći teren, posed lope, procenat dosadašnjih pobeda, procenat dosadašnjih remija, procenat dosadašnjih poraza u ligi

*Dodavanja*: broj dodavanja, preciznost dodavanja, broj ključnih dodavanja (dodavanje koja direktno dovode do šuta), broj progresivnih dodavanja (uspešna dodavanja koja pomeraju loptu ka protivničkoj gol liniji najmanje 10 metara od njene najudaljenije tačke u poslednjih 6 dodavanja)

*Šutevi*: broj primljenih golova, šutevi na gol, šutevi u okvir gola, broj penala, očekivani broj golova

Treniranje modela

Za koeficijent kvaliteta regresionog modela koristiće se prilagođena r^2 mera jer r^2 nije prikladna mera za model višestruke regresije. Proverom T testa i kolinearnosti nezavisnih promenljivih izbacivaće se sve varijable koje u teoriji nisu direktno vezane za mogući broj postignutih golova.

Varijable koje se neće uklanjati su: kvote na ishode meča (neophodne su za predikciju ishoda), procenti dosadašnjih pobeda, poraza i remija (predstavljaju formu tima), posed lopte (najbitniji parametar taktike), broj šuteva na gol i u okvir gola, te očekivani broj golova. Uzimaće se preciznost od 85% dok će odnos podataka za traniranje i validaciju biti 4 na prema 1 (80% podataka je za trenitranje).

Skup podataka

Za uzorak podataka za treniranje uzeće se podaci sa poslednjih 10 ligaških utakmica svakog tima iz Premijer Lige za sezonu 2023/2024. Uzima se poslednjih 10 utakmica kako bi se mogao odrediti određeni trend i forma svakog tima. Nophodno je ispitati svaki tim iz lige kako bi se pokrile različite taktike i stilovi igre, te da postoje parametri statistike igre kako boljih i dominantnijih, tako i slabijih i inferiornijih tima. Na taj način se obezbeđuje reprezentativan uzorak podataka za treniranje modela.

Za uzorak podataka za testiranje uzeće se podaci sa poslednjih 10 ligaških utakmica svakog tima iz Bundes Lige za sezonu 2023/2024. Uzima se ista sezona kao za podatke za testiranje kako bi podaci bili uniformni.

Podaci vezani za statistiku, dodvanja i šuteve će se sakupljati web scraping-om koristeći biblioteku selenium. Za treniranje modela će se učitani csv fajlovi mergovati koristeći biblioteku pandas po nazivima timova te će se tako dobiti 1 data frame koji sadrži sve neophodne podatke za model.

Način evaluacije

Korisnik će kroz konzolni meni uneti nazive domaćeg i gostujućeg tima, te kvote na remi i pobede domaće i gostujuće ekipe. Unose se 3 kvote: prva je vezana za pobedu domaćeg tima, druga za remi, te treća za pobedu gostujućeg tima. Prave se 2 različita regresiona modela, 1 za domaći i 1 za gostujući tim, gde će redosled kvota na ishod meča biti prosleđen u zavisnosti od toga da li je tim domaći ili gostujući (primer: ako su kvote 2.7, 4 i 2.3, domaćem timu će biti prosleđene kao 2.7, 4 i 2.3, dok će gostujućem timu biti prosleđene u obrnutom redosledu). Kako će se modelima kvote prosleđivati u obrnutom redosledu očekuje se i da brojevi postignutih i primljenih golova oba tima budu isti, samo u suprotnom redosledu. Vrednosti zavisnih promenljivih ova 2 regresiona modela će zajedno predstavljati prediktovani rezultat meča. Korisniku se dalje pruža mogućnost da kroz konzolni meni izabere mogućnost da vidi minimalni ili maksimalni prediktovani broj golova za klub koji izabere.

Tehnologije

Programski jezik: python

Potrebne biblioteke: statsmodels, pandas, matplotlib, seaborn, numpy, sklearn, selenium

Materijali

<https://fbref.com/en/>

Sajt sa kojeg će se preuzimati potrebni parametri za svaki odigrani meč određenog kluba.

<https://www.sofascore.com/>

Sajt sa kojeg će se preuzimati kvote za svaki odigrani meč određenog kluba.