

Uvod i obeležja za razmatranje u modelovanju

Osnovna ideja mog fuzzy sistema će biti da pokušam da procenim potencijal naših mladih fudbalera koji igraju na napadačkim pozicijama u veznom redu u zavisnosti od njihovih atributa koji su im pripisani u okviru igre "Football Manager 2016"[1]. Baza podataka je preuzeta sa "Kaggle" web sajta [2], i podaci su u potpunosti očišćeni od nevalidnih vrednosti[3].

Kako bi uspeo da pronađem najpogodnije varijable za modelovanje sistema, ideja mi je da koristim podatke vezane za naše fudbalere koji igraju na napadačkim pozicijama u veznom redu. Razlog zašto sam se odlučio za baš igrače na ovim pozicijama jeste što se igrači na ovim pozicijama smatraju za najveće majstore fudbala [4]. Takođe, u našem fudbalu igrači baš na tim pozicijama i u godinama koje su manje od 23 spadaju u red najtraženijih svetskih fudbalera. Neki od njih su Sergej Milinković-Savić, Nemanje Radonjić, Mijat Gaćinović, Filip Kostić itd.

Što se samih pozicija tiče, u žargonu "Football Manager"-a, to su takozvane napadačke pozicije u veznom redu (eng. "attacking midfielder"), koje mogu imati uz sebe i odredbu da li je u pitanju levi, desni ili centralni napadački vezni igrač. Igrači na ovim pozicijama u svom timu imaju različit niz roli koje pokrivaju. Te role se odnose na samo kreiranje igre, zadatke na bočnim pozicijama, zadatke u neposrednoj blizini protivničkog gola pri kreiranju šanse za pogodak [5].

Kako bi igrači mogli adekvatno da odgovore na zadatke u pomenutim rolama, neophodno je da imaju izražene određene attribute koji su im pripisani kroz varijable u data setu. Te varijable se odnose na dodavanje (eng. Passing), driblanje (eng. Dribbling), utrčavanje/probijanje (eng. Crossing), šut (eng. Shooting), završnicu (eng. Finishing), agresivnost (eng. Aggression), brzinu (eng. Speed), kreativnost (eng. Creativity) [6].

Izuzetno je zanimljiva činjenica da je data-set kreiran tako da za većinu obeležja važi normalna raspodela, pa se nakon filtriranja podataka dosta jednostavno nađe raspodela koja odgovara kojoj varijabli i njenoj pridruženoj vrednosti u zavisnosti od ekspertske preporuke.

Kako bi validaciju modela sprovedi na adekvatan način, neophodno je da napravimo izlaznu varijablu koja će nam tome služiti. Ako bi validaciju vršili samo na jednoj izlaznoj varijabli, inicijalnim istraživanjima smo došli do zaključka da ne postoji značajna korelacija između ulaznih varijabli i samo jedna od potencijalnih izlaznih, pa će nam biti ideja da pokušamo da napravimo izlaznu varijablu koja će biti u dovoljnoj meri korelsana sa ulaznim podacima da bi mogli da izvršimo validaciju modela.

Izlaznu varijablu ćemo formirati od sledećih podataka:

- vrednosti igrača na specijalizovanom sajtu "Transfermarkt" u milionima evra [7]

- učinak igrača u tekućoj sezoni (golovi i asistencije) korelisan sa odgovarajućim koeficijentom

Podatke o učinku ćemo kreirati na osnovu podataka na "Transfermarkt"-u, a koeficijente za korelaciju ćemo dobiti optimizacijom parametara.

Modelovanje

Analiza ulaznih i izlazne promenljive

Izlazna promenljiva

Kao što je već napomenuto, izlaznu varijablu koja će biti indeks TMP (eng. Total Market and Performance) ćemo kreirati na osnovu varijable cena igrača i varijabli vezanih za performanse igrača u tekućoj sezoni.

Cene igrača su vrlo ujednačena varijabla sa par outliera koji se odnose na cene baš na cene dva naša najbolja igrača na ovim pozicijama trenutno, Milinkovića-Savića i Kostića. Njih ćemo sniziti na vrednosti ispod gornjeg ograničivača, negde na 72.5% vrednosti statističkog skupa. Dakle ove dve vrednosti su značajno remetile sliku pri sagledavanju korelacije sa pomenutim paramterom.

Što se tiče druge komponente izlazne varijable, ona se odnosi na performanse igrača u tekućem prvenstvu, a u slučaju onih koji su trenutno povređeni ili nisu angžovani, kod njih su uzete vrednosti od prethodne sezone. Ova varijabla agregiran je od strane broja utakmica, broja golova i broja asistencija u ovoj godini. Kako bi mogli da sagledamo koja varijabla izaziva najveću korelisanost sa ulaznim varijablama, radili smo optimizaciju koeficijenata za svaku od ovih promenljivih. Formula za ovu izlaznu varijablu je sledeća:

$$\text{performanse} = x * \text{koeficijentZaGolove} + y * \text{koeficijentZaAsistencije} + z * \text{koeficijentZaUtakmice}$$

performanse - promenljiva koja oslikava performanse igrača u tekućoj sezoni kroz indeks

x - broj golova za igrača u sezoni

y - broj asistencija za igrača u sezoni

z - broj utakmica za igrača u sezoni

Parametri za optimizaciju

Za svaki od pomenuta tri paramtera korišćeni su identični koeficijenti za optimizaciju. To su vrednosti od 0.2 do 1.8 sa korakom od 0.1. To znači da smo sledeće vrednosti koristili za optimizaciju:

0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8

Nakon što je izvršena optimizacija, najveća suma koeficijenata korelacije se javila kod vrednosti: $(x,y,z) = (0.2, 1.8, 0.2)$.

Ako pokušama da interpretiramo rezultate optimizacije, doći ćemo do zaključka da broj asistencija ima neuporedivo veći značaj od druga dva paramtera na sliku o ispunjenosti potencijala pojedinaca. To je donekle i logično objašnjenje jer igrači na napadačkim pozicijama u veznom redu imaju dosta kontakta sa loptom na protivničkoj polovini i neretko su inicijatori gol šansi za svoje napdače i samim tim neretko su asistenti.

Takođe, nije minoran ni doprinos promenljivih kao što su broj golova i broj utakmica. Broj golova dodatno dopunjuje sliku o kvalitetu pojedinca pa je dosta nepovoljno isključiti ga. Broj utakmica ukazuje na niz nekih faktora koji se mogu kriti iz te promenljive. Neki od njih su to koliko je fudbaler sklon povredama, koliko ima dobru saradnju sa saigračima, trenerom, koliko se pošteno i profesionalno odnosi prema svojoj okolini, pa smatramo da i ovaj faktor ne bi trebalo isključiti iz daljih analiza.

Kreiranje TMP indeksa

Kada smo uspeali da kreiramo ove dve varijable, osnovna ideja je da kreiramo indeks koji će agregirati njihovo dejstvo. Pre nego što kreiramo taj indeks, normalizovaćemo obe promenljive na interval od 0 do 1. To činimo pomoću formule:

$$\frac{x - \min(x)}{\max(x) - \min(x)}$$

Tako normalizovane vrednosti koristimo za kreiranje TMP indeksa pomoću formule:

TMP indeks = $\frac{\text{cenaIgračaNaTržistu}}{\text{performanseIgračaUTekmućojGodini}}$ *

*- u pitanju su normalizovane vrednosti

Nakon ovoga, imamo odgovorajuću varijablu koja će nam služiti za validaciju našeg modela.

Ulazne varijable

Već smo spomenuli niz varijabli koje su pogodne za rad u našem modelu, ali je bilo poželjno isključiti nekoliko, kako zbog interpretacije rezultata tako i zbog mogućnosti da se model pretrenira. Koeficijenti korelacije koje ćemo uzeti u razmatranje kod selekcije promenljivih za kreiranje modela su upravo one koje odgovaraju paramterima koje smo uzeli za modeliranje promenljive koja oslikava performanse igrača u ovoj godini. Zapravo tu je suma svih koeficijenata i najveća.

U zavisnosti na vrednost koeficijenta korelacije, određivali smo jačinu veze [9]:

- mala jačina veze (0.10 do 0.29)
- srednja jačina veze (0.30 do 0.49)
- velika jačina veze (0.50 do 1)

Promenljiva koja je jako korelisana sa izlaznom promenljivom je varijabla Crossing koja je ekvivalentna oceni za utrčavanje igrača (probijanje). To je zapravo i jedina promenljiva koja je visoko korelisana sa TMP indeksom, mada su i promenljive Acceleration i Dribbling dosta blizu tog praga od 0.5 od kog kažemo da su promenljive visoko korelisane. Dakle ubrzanje (koeficijent korelacije 0.4915) i dribling (koeficijent korelacije 0.4741) su zamalo visoko korelisane promenljive, ali možemo za njih reći da su srednje korelisane veličine. U ovu grupu spadaju i promenljive Aggresion i Passing, odnosno promenljiva koja se odnosi na agresivnost igrača za koju koeficijent korelacije iznosi 0.3535, i ocena za dodavanja koja pruža igrač sa vrednosti 0.3168 za koeficijent korelacije. Ovo su dve osobine koje su svakako karakteristične za igrače na ovoj poziciji. Dodavanje koje se u većoj meri odnosi na igrače koji se nalaze na centralnim pozicijama, a agresivnost za igrače koji su više na početnim pozicijama, pa ovaj njihov atribut ima velikog uticaja kada vrše pritisak na poslednju liniju protivnika i kada utrčavaju na centaršut. Takođe postoji još jedna varijabla koja je slabo korelisana sa sa TMP a to su godine. To bi u principu značilo da nema previše uticaja koliko neki igrač ima godina u odnosu na naš indeks, što donekle ima smisla znajući da neki mlađi igrači pružaju slične partije kao starije kolege i imaju dosta veću vrednost na tržištu što znači da godine nisu presudan faktor kod procene potencijala.

Promenljiva	Crossing	Acceleration	Dribbling	Aggresion	Passing	Age
Koeficijent korelacije	0.6238	0.4915	0.4741	0.3535	0.3168	0.2683

Modelovanje FIS-a

Fazifikacija

Kako smo odredili ulazne promenljive koje su značajne, neophodno je da ih modelujemo kao ulaze u FIS. Gotovo sve promenljive podležu normalnoj raspodeli, tako da ćemo nju koristiti za modelovanje ulaznih promenljivih. Prva ulazna promenljiva koju ćemo analizirati su godine. Ova promenljiva je karakteristična jer za nju je dosta jednostavno odrediti kompoziciju ulazne promenljive, a to će biti prema uobičajenim starosnim kategorijama-uzrastima mladih fudbalera. Klasifikacija je data u sledećoj tabeli:

Starosna kategorija	Uzrast	Srednja vrednost	Standardna devijacija
mladi	do 17 godina	16.25	0.56
junior	od 18 do 20 godina	19.04	0.82
mladji senior	od 21 do 23	21.91	0.81

Naredna ulazna promenljiva koju ćemo modelovati je Crossing, tačnije utrčavanje igrača prema голу protivnika. I ova promenljiva ima normalnu raspodelu pa ćemo statistički skup podeliti na tri jedna dela koja će oslikavati gradacijski slabog, prosečnog i izvrsnog utrčivača.

Kategorija utračanje	Vrednosti ocene	Srednja vrednost	Standardna devijacija
slab	do 7	4.41	1.41
srednji	od 8 do 10	9.12	0.84
jak	od 11 do 16	12.59	0.9

Kao karakteristika koja ima dosta veze sa sa utrčavanjem, jeste karakteristika Acceleration, koja se tiče ubrzanja koje postiže fudbaler. Modelujemo sledeće raspodele koje se odnose na ulaznu varijablu:

Starosna kategorija	Vrednost ocene	Srednja vrednost	Standardna devijacija
sport	do 11	10.2	0.98
prosek	od 12 do 13	12.51	0.5
brz	od 14 do 15	14.2	0.41

Drbling spada u jednu od glavnih karakteristika igrača koji se nalaza bliže centralnom terenu igrača. Ovu karakteristiku ćemo kreirati na sledeći način:

Starosna kategorija	Vrednost ocene	Srednja vrednost	Standardna devijacija
loš	od 2 do 8	6.16	1.75

dobar	9 do 11	10.23	0.78
sjajan	od 12 do 16	12.76	0.88

Agresivnost je jedna jako poželjna osobina igrača na ovim pozicijama jer na taj način stavljaju od pritisak protivničke odbrambene igrača i uzrokuju njihove greške sa potencijalnim golom kao dobiti tog pritiska. Promenljivu koja opisuje agresivnost modelujemo na sledeći način:

Agresivnost	Vrednost ocene	Srednja vrednost	Standardna devijacija
hladan	od 6 do 7	6.36	0.48
srednje	8 do 10	8.96	0.7
vatren	od 11 do 12	11.72	0.45

Jedna od glavnih karakteristika igrača na ovim pozicijama se odnosi na dodavanje, i pregled ocena tog paramtera je dati sledećom tabelom:

Dodavanje	Vrednost ocene	Srednja vrednost	Standardna devijacija
loš	od 3 do 8	5.25	1.49
dobar	9 do 11	9.68	0.47
brilijantan	od 12 do 15	12.74	0.83

Takođe, neophodno je da i fazifikujemo i izlaznu varijablu TMP indeks. Pošto ovaj indeks nema normalnu rasodelu, ovaj indeks ćemo aproksimirati sa trougaonom raspodelom. Trouglovi su određeni sledećim tačkama:

- slab potencijal [0.06403 0.3949 0.8751]
- prosečan potencijal [0.3571 0.8477 1.275]
- visok potencijal [1.011 1.401 2]

Koordinate trouglova su određivane na osnovu podataka o 10 igrača iz kategorije sa najvećim indeksom, zatim retrospektivno i za naredne dve kategorije. Metrike koje su korišćene su minimalna i maksimalna vrednost, kao i medijana i srednja vrednost za položaj trećeg temena trougla.

Definisanje If-Then pravila

Kako bi odredili izlaze iz našeg fuzzy sistema, neophodno je da definišemo pravila po kojima će se ulazne promenljive transformisati u izlaznu fuzzy promenljivu koja oslikava TMP indeks. Pravila po kojima će se evaluirati naš sistem su data ispod:

godine	utrčavanj e	ubrzavanje	dribling	agresivnost	dodavanj e	TMP
mladji_senior	-	-	sjajan	vatren	brilijantan	visok
mladji_senior	jak	brz	-	vatren	-	visok
mladi	-	-	prosečan	vatren	brilijantan	visok
mladi	srednji	prosečan	prosečan	-	-	prosečan
juniori	spor	-	-	-	loš	slab
mladji_senior	-	spor	loš	-	loš	slab
junior	jak	prosečan	prosečan	-	-	slab
mladi	srednji	-	-	hladan	loš	slab
junior	jak	-	-	vatren	brilijantan	visok

Kako bi na osnovu prikazane tabele generisali odgovarajuća pravila, korišćemo veznik AND za sva pojedina pravila. Predpostavićemo takođe da su sva pravila jednake težine iako je moguće dodeljivati im odgovarajuće težine [10].

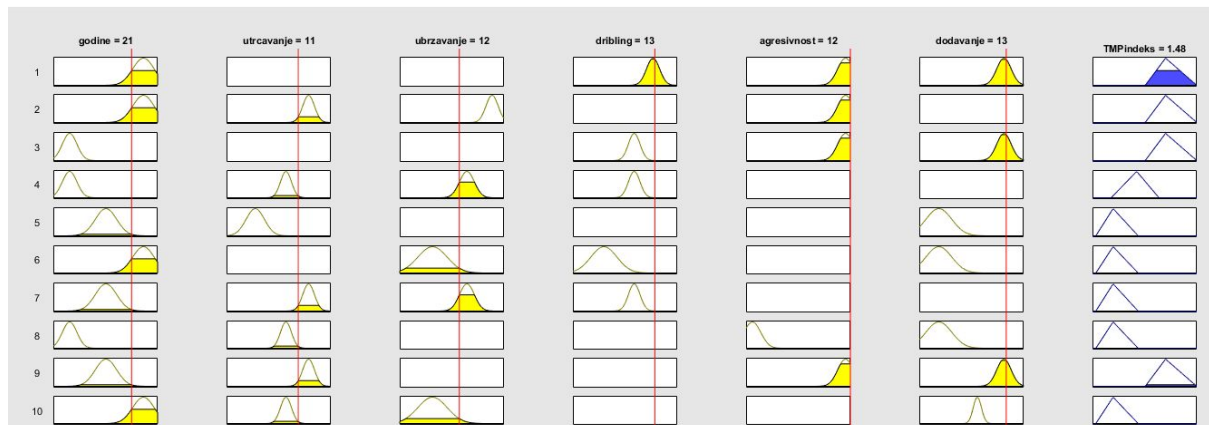
Što se tiče fazi operatora koje ćemo koristiti kod evaluacije našeg modela, korišćemo standardni presek kao fazi konjuktiju (t-norma), a kao fazi disjunktiju (t-konormu) korišćemo standardnu uniju. Matematički zapisano to bi izgledalo na sledeći način [10]:

- standardni presek za fazi konjuktiju $i(R1,R2)=\min(R1,R2)$
- standardna unija za fazi disjunktiju $u(R1,R2)=\max(R1,R2)$

Implikacija

Ulaz u ovaj proces je vrednost funkcije pripadnosti koji smo dobili iz prethodnih koraka primenom operatora AND/OR na fazifikovane ulaze. Izlaz iz ovog procesa je fazi skup dobijen presekom izlazne promenljive i vrednost funkcije pripadnosti koja predstavlja ulaz u ovaj proces [10].

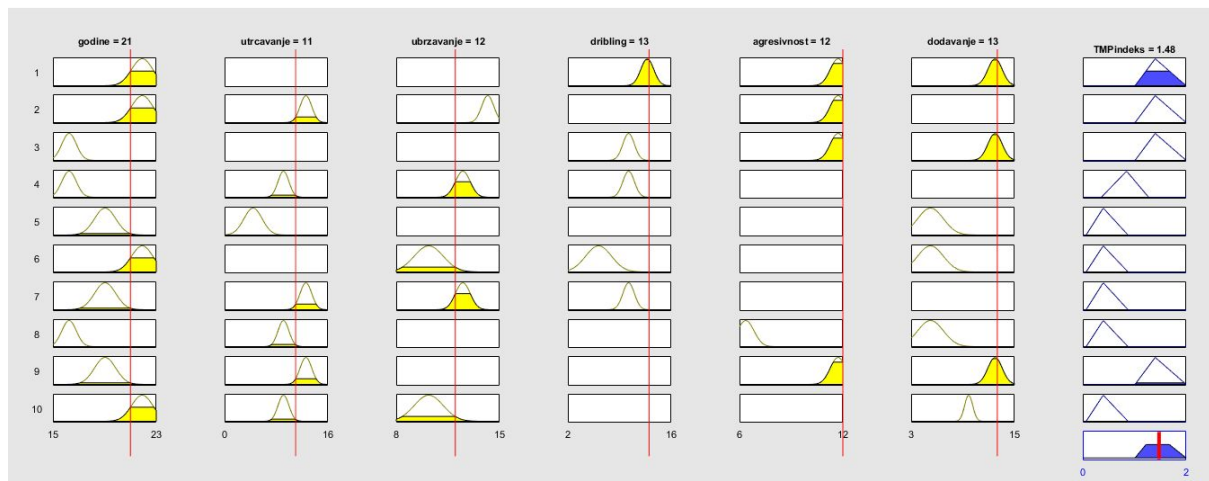
Primer implikacija na našem radu je dati ispod pri čemu svako pravilo ima isti značaj, odnosno 1. U ovom koraku je moguće definisati različite težine za pojedinačna pravila.



Agregacija posledica

Agregacija posledica je proces u kom se vrši agregacija fazi skupova koji predstavlja izlaze iz If-Then pravila ulazi u ovaj proces su fazi skupovi dobijeni u prethodnom koraku. Agregacijom ovih ulaznih skupova dobija se po jedan rezultujući skup za svaku od unapred definisanih izlaznih varijabli [10].

U našem primeru agregacija će se vršiti pomoću max pravila.



Evaluacija i defazifikacija modela

Defazifikacija predstavlja proces kojim se posmatrani fazi skup određenom metodom prevodi u jednu "crisp" vrednost. Osnovna ideja koja se krije iza ovog procesa je što donosiocu odluke vrednost iz agregacije posledica ništa ne znači, pa je potrebno konstruisati konkretnu vrednosti koja može da se poredi sa više različitih vrednosti [10].

Metoda koju koristimo za defazifikaciju je "centroid". Kod ove metode, vraća se težinski centar oblasti ispod krive, dakle ako bi posmatrali kao telo neki fazi skup, onda bi to bila ravnotežna tačka [10].

Ispod je data evaluaciona tabela našeg fisa. Konkretna vrednost se nalazi u koloni fis. Tabele "Name" i "Age" se odnose na ime i godine pojedinačnih igrača, a Normalna kolona se odnosi na TPM indeks koji je igrač stvarno ostvario. Evaluaciona kolona se odnosi na rang fudbalera procenjen našim fisom.

	Name	Age	Normalna	Evaluaciona	fis
1	Sergej Milinkovic Savic	21	A	A	1.480
2	Filip Kostic	23	A	A	1.500
3	Mijat Gacinovic	21	A	A	1.500
4	Nemanja Radonjic	20	A	A	1.490
5	Andrija Zivkovic	19	A	A	1.500
6	Milan Pavkov	22	A	A	1.500
7	Nikola Ninkovic	21	A	A	1.480
8	Mirko Ivanic	22	A	B	0.838
9	Danilo Pantic	19	A	A	1.500
10	Srdjan Plavsic	20	A	B	1.460
11	Marko Simic	22	B	C	0.470
12	Nikola Filipovic	16	B	C	0.767
13	Luka Ilic	17	B	B	0.847
14	Luka Adzic	17	B	B	0.844
15	Petar Micin	17	B	C	0.515
16	Armin Djerlek	15	B	C	0.818
17	Veljko Birmancevic	18	B	C	0.670
18	Vladimir Siladji	23	B	C	0.466
19	Aleksandar Lutovac	19	B	B	1.040
20	Stefan Mihajlovic	17	B	C	0.670
21	Filip Malbasic	23	C	A	1.500
22	Aleksandar Palocevic	22	C	A	1.490
23	Nemanja Mihajlovic	20	C	A	1.480
24	Stefan Hajdin	23	C	C	0.796
25	Dejan Drazic	20	C	A	1.480
26	Filip Djordjevic	21	C	C	0.467
27	Alen Masovic	21	C	C	0.467
28	Milos Pantovic	19	C	B	0.933
29	Damjan Gojkov	18	C	B	0.843
30	Ivan Markovic	22	C	A	1.480

Reference:

- [1] <https://www.sigames.com/games/football-manager-2016>
- [2] <https://www.kaggle.com/>
- [3] <https://www.kaggle.com/ajinkyablaze/football-manager-data>
- [4] <https://footballsgreatest.weebly.com/attacking-midfielders.html>

- [5] https://medium.com/@v_maedhros/understanding-roles-in-football-manager-and-real-life-part-1-73054cfbb303
- [6] <https://forum.topeleven.com/top-eleven-general-discussion/72092-positioning-important-st-amc-amr-aml.html>
- [7] <https://www.transfermarkt.com/>
- [8] <https://kassiesa.home.xs4all.nl/bert/uefa/data/method5/crank2020.html>
- [9] Pallant, J. (2001). SPSS Survival Manual. A&U Academic
- [10] Nastavni materijali na predmetu "Fuzzy logika"- Fakultet organizacionih nauka ,2020.