

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1275

**WEB APLIKACIJA ZA POTPORU PRI DONOŠENJU ODLUKA
U VIRTUALNOJ MENADŽERSKOJ NOGOMETNOJ LIGI
PRIMJENOM METODA STROJNOG UČENJA**

Luka Raić

Zagreb, lipanj 2026.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1275

**WEB APLIKACIJA ZA POTPORU PRI DONOŠENJU ODLUKA
U VIRTUALNOJ MENADŽERSKOJ NOGOMETNOJ LIGI
PRIMJENOM METODA STROJNOG UČENJA**

Luka Raić

Zagreb, lipanj 2026.

DIPLOMSKI ZADATAK br. 1275

Pristupnik: **Luka Raić (0036537547)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: prof. dr. sc. Alan Jović

Zadatak: **Web aplikacija za potporu pri donošenju odluka u virtualnoj menadžerskoj nogometnoj ligi primjenom metoda strojnog učenja**

Opis zadatka:

Fantasy Premier League (FPL) službena je virtualna nogometna igra engleske Premier lige u kojoj milijuni korisnika diljem svijeta svake sezone sastavljaju virtualne timove od stvarnih nogometnih igrača. Korisnici tjedno donose strateške odluke o postavi, zamjenama, transferima i korištenju posebnih pogodnosti (tzv. čipova), a bodovi se dodjeljuju temeljem stvarnih nastupa igrača na terenu. Unatoč velikoj količini dostupnih podataka o igračima i timovima, većina korisnika odluke donosi temeljem intuicije, bez sustavne podatkovne analize. U ovom diplomskom radu potrebno je razviti web aplikaciju koja korisnicima nudi podatkovnu podršku pri donošenju strateških odluka u FPL-u. Aplikacija treba sadržavati dva glavna dijela. Prvi dio odnosi se na inteligentnog asistenta koji koristi metode strojnog učenja za predviđanje bodovnih učinaka igrača te na temelju tih predviđanja pruža korisnicima personalizirane preporuke za upravljanje timom. Za izradu prediktivnih modela potrebno je koristiti podatke dostupne putem službenog FPL API-ja i, po mogućnosti, iz drugih izvora. Drugi dio aplikacije odnosi se na prikaz statistika igrača i timova po sezonama, čime se korisniku omogućuje samostalna analiza i donošenje informiranih odluka. U radu je potrebno istražiti i primijeniti jednu ili više metoda strojnog učenja, poput pojačavanja gradijenta, neuronskih mreža ili ansambala, s ciljem postizanja što točnijih predviđanja. Razvijenu web aplikaciju potrebno je vrednovati simulacijom na povijesnim podacima te usporediti s osnovnim heurističkim pristupima.

Rok za predaju rada: 26. lipnja 2026.

Zahvaljujem se svom mentoru, prof. dr. sc. Alanu Joviću, na pomoći pruženoj pri izradi ovog rada te tijekom cijelog trajanja studija. Posebno se još zahvaljujem svojoj majci Jelici, ocu Robertu i sestri Karli na bezuvjetnoj podršci: bez njih ništa od ovog ne bi bilo moguće.

Sažetak

Web aplikacija za potporu pri donošenju odluka u virtualnoj menadžerskoj nogometnoj ligi primjenom metoda strojnog učenja

Luka Raić

Fantasy Premier League virtualna je menadžerska igra u kojoj sudionik prije svakoga kola bira transfere, početnu postavu i kapetana. U ovome je radu izrađen sustav koji takve odluke temelji na podacima. Objedinjeni podatkovni skup, sastavljen iz triju javno dostupnih izvora, obuhvaća sedam sezona na razini para igrač–susret. Na njemu su u kontroliranim uvjetima uspoređena dva pristupa predviđanju bodovnog učinka: izravna regresija po pozicijama i dekompozicija bodova po komponentama, pri kojoj zaseban model procjenjuje svaku sastavnicu bodovanja. Oba pristupa prema mjerama pogreške i korelaciji rangova nadmašuju tri heuristička referentna modela u svim kolima ispitne sezone i na svakoj poziciji, a dekompozicija je odabrana kao konačni model. Predviđanja se potom cjelobrojnim linearnim programiranjem pretvaraju u odluke o sastavu momčadi, nakon čega je cijela sezona ponovno odigrana u povijesnoj simulaciji. U njoj je najbolji model izjednačen s najboljom heuristikom, premda donosi bolje pojedinačne odluke, jer se momčad bira iz samoga vrha poretka, gdje je njegova prednost najmanja. Sve su sastavnice integrirane u web aplikaciju *Gaffer*, koja korisniku nudi inteligentnog asistenta s preporukama za svako kolo i statistički pregled za samostalnu analizu.

Ključne riječi: Fantasy Premier League; strojno učenje; predviđanje bodova igrača; gradijentno pojačavanje; cjelobrojno linearno programiranje; povijesna simulacija; sustav za potporu odlučivanju; web aplikacija

Abstract

Web application for decision support in fantasy football league using machine learning methods

Luka Raić

Fantasy Premier League is a fantasy football game in which participants select transfers, a starting lineup, and a captain before each gameweek. This work develops a system that supports such decisions based on data. The unified dataset, compiled from three publicly available sources, covers seven seasons at the player-match pair level. Two approaches to predicting point returns were compared under controlled conditions: direct regression by position and a decomposition of points into components, in which a separate model estimates each scoring component. According to error metrics and rank correlation, both approaches outperform three heuristic baseline models in every gameweek of the test season and for every position, with the decomposition selected as the final model. The predictions are then transformed into team-management decisions using integer linear programming, after which the entire season is replayed through historical simulation. In this simulation, the best model ties with the best heuristic despite making better individual decisions, because the team is chosen from the very top of the ranking, where its advantage is smallest. All components are integrated into the *Gaffer* web application, which offers users an intelligent assistant with recommendations for each gameweek and a statistical dashboard for independent analysis.

Keywords: Fantasy Premier League; machine learning; player points prediction; gradient boosting; integer linear programming; historical simulation; decision support system; web application

Sadržaj

Sažetak	1
Abstract	2
1. Uvod	6
2. Fantasy Premier League	7
2.1. Povijest igre	7
2.2. Pravila igre i sustav bodovanja	9
3. Srodni radovi	14
3.1. Predviđanje učinka igrača u FPL-u	14
3.2. Optimizacija odabira momčadi u FPL-u	17
3.3. Doprinosi ovoga rada	18
4. Prikupljanje, integracija i eksploratorna analiza podataka	20
4.1. Izvori podataka	20
4.1.1. Službeni podatci Fantasy Premier Leaguea	20
4.1.2. Podatci platforme Understat	21
4.1.3. Podatci platforme Sofascore	22
4.2. Objedinjeni podatkovni skup	23
4.2.1. Vremenska pokrivenost i struktura podatkovnog skupa	23
4.2.2. Čišćenje i standardizacija pri učitavanju	24
4.3. Eksploratorna analiza	24
4.3.1. Ciljna varijabla	24
4.3.2. Razlike po pozicijama	26
4.3.3. Raščlamba visokih učinaka	28

4.4.	Kronološka podjela i problem hladnog starta	29
4.5.	Od svojstava podataka do odluka pri modeliranju	30
5.	Modeli za predviđanje bodova igrača	31
5.1.	Formulacija problema	32
5.2.	Odabir i učenje modela	32
5.3.	Zajednički metodološki okvir	33
5.3.1.	Izgradnja značajki bez curenja informacija	33
5.3.2.	Dvostupanjska struktura predviđanja	34
5.3.3.	Podjela podataka	35
5.3.4.	Korišteni skup metrika	36
5.3.5.	Referentni modeli	37
5.4.	Pristup 1: izravna regresija po pozicijama	37
5.4.1.	Arhitektura i učenje	37
5.4.2.	Ugađanje hiperparametara	38
5.5.	Pristup 2: dekompozicija bodova po komponentama	39
5.5.1.	Arhitektura i učenje	39
5.5.2.	Hiperparametri	41
5.5.3.	Sastavljanje predviđanja	41
5.6.	Rezultati	42
5.6.1.	Ukupna usporedba s referentnim modelima	42
5.6.2.	Usporedba dvaju pristupa	45
5.6.3.	Model nastupa	46
5.6.4.	Visoki učinci i rep predviđanja	47
5.6.5.	Važnost značajki	48
6.	Optimizacija upravljanja momčadi i povijesna simulacija	50
6.1.	Problem odlučivanja	50
6.2.	Pregled mehanizma	51
6.3.	Protokol simulacije	53
6.4.	Rezultati	54
7.	Web aplikacija za potporu pri donošenju odluka	57
7.1.	Pregled funkcionalnosti	57

7.2. Arhitektura sustava	58
7.3. Inteligentni asistent	60
7.3.1. Alat za slaganje momčadi	60
7.3.2. Preporuke	61
7.4. Statistički modul	63
7.5. Simulacijski modul	65
7.6. Sigurnost	66
8. Zaključak	68
Literatura	70
Izjava o korištenju umjetne inteligencije	73

1. Uvod

Fantasy Premier League (FPL) virtualna je menadžerska igra engleskog natjecanja *Premier League* i, s više od 11 milijuna sudionika, najraširenija je igra te vrste na svijetu. Sudionik unutar zadanog proračuna sastavlja momčad od 15 stvarnih nogometaša te prije svakoga kola odlučuje o transferima, početnoj postavi i kapetanu, a bodove osvaja prema njihovim učincima. Iako su podatci o igračima i klubovima javno dostupni, raspršeni su po različitim izvorima i zahtijevaju vrijeme te predznanje za prikupljanje i tumačenje. Stoga se odluke većinom donose intuitivno, bez sustavne analize podataka.

Podatkovna potpora pri upravljanju FPL momčadi obuhvaća dva povezana problema. Potrebno je predvidjeti broj bodova igrača u sljedećem kolu, što otežavaju česti izostanci i znatan utjecaj slučajnosti na učinke, a zatim ta predviđanja pretvoriti u odluke uz ograničenja proračuna, sastava momčadi i cijene transfera. Drugi je problem zadatak kombinatorne optimizacije, tj. odabira momčadi, početne postavke i kapetana među stotinama igrača, dok je cilj rada razvoj web aplikacije koja korisniku pruža potporu pri takvu odlučivanju.

Aplikaciju čine inteligentni asistent i statistički pregled za samostalnu analizu. Asistent metodama strojnog učenja predviđa učinke igrača i iz njih izvodi preporuke za upravljanje momčadi, dok statistički pregled omogućuje analizu relevantnih podataka. Dva se pristupa predviđanju vrednuju metrikama na razini pojedinačnih susreta i povijesnom simulacijom cijele sezone, uz usporedbu s naivnim heuristikama. Poglavlja 2. i 3. redom prikazuju pravila i sustav bodovanja igre te srodne radove i doprinose rada. Slijede poglavlja o podatkovnom skupu (4.), modelima predviđanja (5.), optimizacijskom mehanizmu i simulaciji (6.), aplikaciji (7.) te zaključku (8.).

2. Fantasy Premier League

Fantasy Premier League (FPL) službena je virtualna menadžerska igra engleskog natjecanja *Premier League*, najvišeg stupnja engleskog ligaškog nogometnog sustava. Kao i u drugim virtualnim menadžerskim igrama (engl. *fantasy sports*), sudionici igre sastavljaju zamišljenu momčad sačinjenu od stvarnih igrača — u ovom slučaju nogometaša — te osvajaju bodove na temelju njihovih stvarnih statističkih učinaka te procijenjene uloge i doprinosa na terenu. S preko 11 milijuna sudionika, FPL je trenutno najpopularnija nogometna virtualna menadžerska igra na svijetu.

U potpoglavlju 2.1. će biti riječ o samoj povijesti igre, dok će u potpoglavlju 2.2. biti detaljnije opisana njezina pravila te sustav bodovanja.

2.1. Povijest igre

Počeci FPL-a sežu sve do devedesetih godina prošlog stoljeća, kada se igranje nogometnih virtualnih menadžerskih igara počinje ubrzano širiti iz Italije na ostale europske države: naime, nogometnu inačicu virtualne menadžerske igre je, potaknut time što je takva igra već postojala u bejzbolu, 1990. godine izumio talijanski novinar Riccardo Albini [1]. Mnoge su novine imale svoju vlastitu verziju igre, no od 1994. godine pa nadalje najpopularnija je bila ona objavljivana u britanskim novinama *The Daily Telegraph*. Tu je verziju *The Daily Telegraph* pokrenuo u suradnji s Andrewom Wainsteinom, programerom i nogometnim zaljubljenikom koji je još 1991. godine (godinu dana prije osnutka natjecanja *Premier League*) pokrenuo prvu nogometnu inačicu virtualne menadžerske igre u Ujedinjenom Kraljevstvu, i to putem svoje tvrtke *Fantasy League Ltd*, isprva kao malen, ručno vođen projekt s nekoliko stotina sudionika raspoređenih u osamdesetak liga [2]. Suradnjom s već spomenutim novinama *The Daily Telegraph* igra je 1994. godine postala prva masovna inačica na britanskom tržištu, a njezinu je popularnost dodatno

potaknula istoimena televizijska humoristična emisija *Fantasy Football League* s Frankom Skinnerom i Davidom Baddielom. Već je prvo izdanje natjecanja u sezoni 1994./95. privuklo oko 342 000 sudionika [3].

U doba prije interneta cijela se igra odvijala putem tiska, pošte i telefona. Sudionik bi iz novina izrezao prijavni obrazac te odabrao početnu momčad ili napravio transfer određenog igrača, pri čemu je morao ostati unutar svojega proračuna. Primjer odsječka takve novinske stranice prikazan je na slici 2.1. Nakon popunjavanja obrasca sudionik bi ga poštom poslao organizatoru, a ažurno stanje vrha tablice nakon odigranoga kola mogao je saznati sljedeće srijede u *The Daily Telegraphu*, gdje bi bili objavljeni ukupni bodovi sto najboljih momčadi [3]. Obrada rezultata pritom je bila poprilično mukotrpna: iako je Wainstein napisao program za izračun bodova te time uspio automatizirati dio posla, cjelokupni proces je trajao više od dvanaest sati, pri čemu je organizator noću ustajao svaka tri sata kako bi printeru dopunjavao potreban papir [4].

Slika 2.1. Stranica *The Daily Telegrapha* (2. kolovoza 2006.) s prijavnim obrascem igre *Telegraph Fantasy Football*. Preuzeto iz: [5]

Početak 2000-ih internet postaje sve rašireniji i probija se u razne sfere svakodnevnog života, a promjene koje je donio nisu zaobišle ni virtualne menadžerske igre. Godine 2002., u povodu desete sezone natjecanja, *Premier League* pokreće svoju službenu web stranicu, a unutar nje — u odjeljku *Games And Fun*, uz nekoliko drugih igara — korisnicima prvi put nudi *Fantasy Premier League*, službenu virtualnu menadžersku igru natjecanja [6]. Igru je razvio vanjski partner *ISM Fantasy Games*, britanska tvrtka koja je već

tada iza sebe imala nekoliko naslova: za izdanje UEFA Lige prvaka u sezoni 1999./2000. izradila je vlastitu virtualnu menadžersku igru, a 2000. godine i igru *CYKI: Can You Kick It?*, prvu njihovu takvu igru vezanu uz *Premier League* [6]. To je iskustvo zacijelo bilo presudno pri dodjeli službenog ugovora, pa je prvo izdanje FPL-a uvelike predstavljalo doradenu inačicu igre CYKI.

Iako su temeljni obrisi igre prepoznatljivi i danas, prva se sezona u nekoliko pojedinosti razlikovala od današnje. Najuočljivije, sastojala se od svega 32 kola umjesto današnjih 38: prvih šest kola smatralo se predsezonom tijekom koje su igrači mogli neograničeno mijenjati sastav i upoznati se s igrom, a natjecanje je “ozbiljno” počinjalo tek od sedmoga kola [6]. Prvo je izdanje privuklo 76 200 sudionika, a pobjednikom sezone 2002./03. okrunio se Graeme Haddow, koji je u spomenuta 32 kola skupio 1 940 bodova. Izgled tadašnjega sučelja, koje je u velikoj mjeri naslijedilo dizajn igre CYKI, prikazan je na slici 2.2.



Slika 2.2. Sučelje prvog izdanja igre *Fantasy Premier League*. Preuzeto iz: [6]

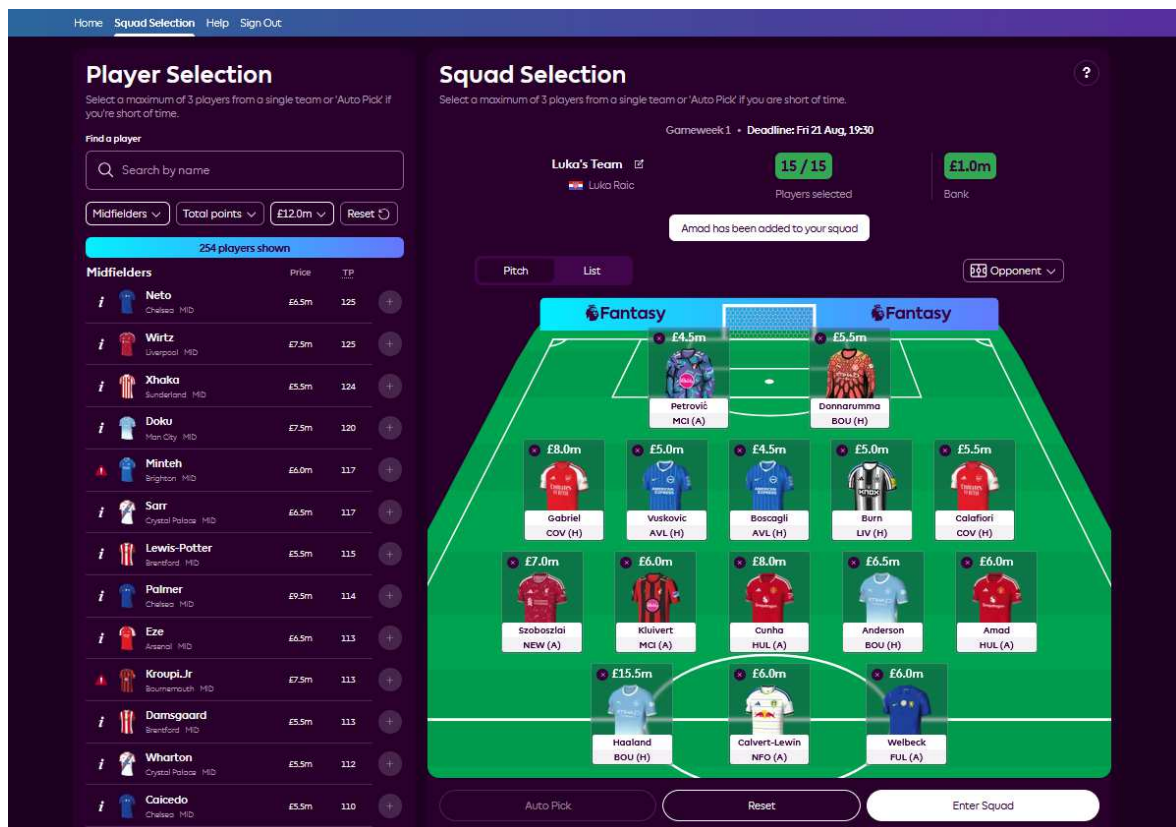
2.2. Pravila igre i sustav bodovanja

Tijek igranja FPL-a, kao i sâm cilj igre, poprilično je jednostavan te sličan onima u drugim virtualnim menadžerskim igrama. Naime, nekoliko tjedana prije početka prvoga kola nove sezone natjecanja *Premier League* počinje nova sezona FPL-a te se sudionicima omogućuje slaganje momčadi. Svaki sudionik na početku ima na raspolaganju proračun

u iznosu od £100 milijuna te s njim mora složiti početnu momčad od 15 igrača, pri čemu ta momčad mora sadržavati:

- 2 vratara
- 5 obrambenih igrača
- 5 veznih igrača
- 3 napadača

Igrači su raspoređeni u različite cjenovne rangove, pri čemu se njihove cijene određuju prema procijenjenoj vrijednosti i očekivanom učinku tijekom sezone. Na primjer, igrači s najvećom početnom cijenom za sezonu 2025./26. u FPL-u bili su Mohamed Salah, koji je početno vrijedio £14,5 milijuna, te Erling Haaland, koji je vrijedio £14 milijuna. Za usporedbu, najmanja početna cijena nekog igrača iznosila je £4 milijuna, što znači da nije moguće samo odabrati najbolje moguće igrače, nego je potrebna doza kalkulacije i razmišljanja unaprijed. Uz to, sudionicima nije dopušteno izabrati više od tri igrača iz iste momčadi. Sučelje za slaganje početne momčadi prikazano je na slici 2.3.



Slika 2.3. Sučelje za slaganje početne momčadi na početku sezone 2025./26. Izvor: [7]

Nakon što slože početnu momčad, sudionici moraju do kraja roka za trenutno kolo (90 minuta prije početka prvog susreta u kolu) od 15 dostupnih igrača odabrati njih 11 za početni sastav, pri čemu je odabrana formacija proizvoljna dok god su u njoj prisutni 1 vratar, barem 3 obrambena igrača te barem 1 napadač. Za svakog odabranog igrača sudionik dobije određeni broj bodova nakon završetka svih utakmica trenutnog kola, pri čemu broj dobivenih bodova za jednog igrača ovisi o njegovoj poziciji te statističkim učincima u danom kolu. Sustav bodovanja objašnjen je u tablici 2.1.

Tablica 2.1. Sustav bodovanja u FPL-u (sezona 2024./25.)

Radnja	Bodovi
<i>Nastup</i>	
Nastup do 60 minuta	1
Nastup od 60 ili više minuta (bez nadoknade)	2
<i>Postignuti golovi i asistencije</i>	
Gol vrataru	6
Gol obrambenog igrača	6
Gol veznog igrača	5
Gol napadača	4
Asistencija	3
<i>Sačuvana mreža</i>	
Sačuvana mreža (vratar ili obrambeni igrač)	4
Sačuvana mreža (vezni igrač)	1
<i>Vratarske akcije</i>	
Svake 3 obrane vrataru	1
Obranjeni jedanaesterac	5
<i>Bonus</i>	
Bonus bodovi za najbolje igrače utakmice	1–3
<i>Negativni bodovi</i>	
Promašeni jedanaesterac	–2
Svaka 2 primljena gola (vratar ili obrambeni igrač)	–1
Žuti karton	–1
Crveni karton	–3
Autogol	–2

Osim odabira početnog sastava za pojedino kolo, sudionici ujedno određuju kapetana i vicekapetana momčadi te redoslijed zamjena na klupi prema prioritetu. Igrač izabran za kapetana u danome kolu ostvaruje dvostruki broj bodova, a ako kapetan iz nekog razloga ne nastupi (primjerice zbog ozljede, odgode utakmice ili neuvršćavanja u sastav od strane trenera), dvostruke bodove umjesto njega ostvaruje vicekapetan. Ako pak ni vicekapetan ne odigra nijednu minutu, nijedan igrač ne ostvaruje dvostruke bodove. Slično načelo prioriteta primjenjuje se i pri automatskim zamjenama. Naime, na klupi se nalaze igrači koji nisu odabrani za početni sastav: zamjenski vratar i tri igrača koji nisu vratari,

poredana prema prioritetu za ulazak u sastav. U slučaju da neki igrač iz početne postavice ne odigra nijednu minutu u kolu, sustav ga automatski nastoji zamijeniti igračem s klupe koji je nastupio, slijedeći zadani redoslijed prioriteta. Pritom zamjena mora rezultirati pravilnim sastavom, koji sadržava točno jednoga vratara te najmanje tri obrambena igrača i jednoga napadača. Ako bi zamjena narušila te uvjete, sustav preskače toga igrača i razmatra sljedećega na klupi. Budući da se automatske zamjene i dodjela dvostrukih kapetanskih bodova provode nakon završetka kola, sudionik treba samo unaprijed pravilno odrediti redoslijed zamjena te kapetana i vicekapetana.

Kako bi se igra s taktičkog aspekta učinila zanimljivijom sudionicima, u nju je dodano nekoliko zanimljivih alata koje sudionici imaju na raspolaganju tijekom sezone, a među njima su i transferi. Slično kao i u stvarnome svijetu, virtualnim je menadžerima omogućeno da tijekom sezone prodaju i kupuju igrače te tako, biranjem igrača u boljoj formi ili s povoljnijim rasporedom utakmica, nastoje ostvariti više bodova od ostalih sudionika. Pritom je važno napomenuti da ipak postoje određena ograničenja kod provođenja transfera: naime, sudionik u svakome kolu na raspolaganju ima jedan besplatni transfer, a neiskorištene besplatne transfere može prenositi u iduća kola i skupiti ih najviše pet. Ako sudionik u trenutnom kolu ipak želi provesti veći broj transfera iznad broja raspoloživih, onda to može napraviti, ali po cijeni od četiri kaznena boda po svakom dodatnom transferu, pa menadžer pri svakoj izmjeni mora procijeniti hoće li mu očekivani dobitak u bodovima nadoknaditi taj trošak. Upravo u toj ravnoteži, između želje za stalnim poboljšavanjem momčadi i cijene koju prečesti transferi nose, leži velik dio taktičke dubine igre. Dodatnu dimenziju transferima daje činjenica da cijene igrača nisu statične, nego se mijenjaju tijekom sezone: promjene se temelje na neto broju transfera pojedinog igrača, odnosno na razlici između broja menadžera koji ga kupuju i onih koji ga prodaju, mjerenoj u odnosu na prag koji ovisi o njegovoj raširenosti u igri. Kada dovoljan broj menadžera kupi nekog igrača, njegova cijena preko noći raste za £0,1 milijun, a kada ga dovoljno njih proda, jednako toliko pada. Pritom menadžer pri prodaji igrača kojem je cijena porasla ne ostvaruje cijeli dobitak, nego mu se priznaje samo polovica ostvarene zarade. Time praćenje i predviđanje promjena cijena postaje dodatani taktički sloj: pravodobnom kupnjom igrača prije rasta cijene menadžer može povećati ukupnu vrijednost svoje momčadi i tako si priuštiti skuplje igrače.

Uz transfere, virtualnim menadžerima na raspolaganju stoje i tzv. čipovi — posebni potezi koji se tijekom sezone mogu iskoristiti ograničen broj puta, a koji privremeno mijenjaju pravila igre u korist menadžera. U sezoni 2025./26. na raspolaganju su sudionici imali četiri različite vrste čipa:

- **Wildcard** omogućuje neograničen broj besplatnih transfera u jednom kolu te trajnu promjenu momčadi.
- **Free Hit** također omogućuje neograničen broj transfera, ali samo za jedno kolo, nakon kojega se momčad vraća u prethodno stanje.
- **Triple Captain** kapetanu u odabranom kolu donosi trostruke bodove umjesto dvostrukih.
- **Bench Boost** omogućuje da u jednom kolu bodove ostvaruju i sva četiri igrača s klupe.

Svaki čip se može iskoristiti samo jednom unutar polusezone, pri čemu se na polovici sezone skup čipova obnavlja, pa menadžer tijekom cijele sezone na raspolaganju ima ukupno osam čipova.

Iako su pojedina pravila i sustav bodovanja na prvi pogled jednostavni, njihova međusobna isprepletenost stvara iznenađujuće velik i složen prostor odluka. Gotovo svaka od tih odluka u konačnici počiva na dvama temeljnim pitanjima: koliko će bodova pojedini igrač ostvariti u nadolazećim kolima te kako, uz sva navedena ograničenja, od dostupnih igrača sastaviti momčad koja maksimizira očekivani broj bodova. Prvo je pitanje problem predviđanja, a drugo problem optimizacije koji se prirodno može formulirati zadatkom linearnoga programiranja. Oba su pitanja privukla znatnu istraživačku pozornost, pa se pregled pristupa predviđanju učinka igrača i optimizaciji sastava momčadi te srodnih radova donosi u nastavku.

3. Srodni radovi

Inteligentni asistent razvijen u ovom radu objedinjuje dva prethodno naznačena komplementarna problema: predviđanje bodovnih učinaka pojedinih igrača u nadolazećem kolu te optimizaciju sastava momčadi. Sastoji se od modela za predviđanje bodova i optimizacijskog mehanizma koji, koristeći dobivena predviđanja i poštujući propisana pravila igre, određuje konačni sastav momčadi te ga u obliku prijedloga prosljeđuje krajnjem korisniku. Sukladno tome, i ovo je poglavlje podijeljeno na tri potpoglavlja. U prvome se potpoglavlju donosi pregled radova koji različitim metodama i pristupima — od statističkih i vjerojatnosnih modela do metoda strojnog i dubokog učenja — nastoje riješiti problem predviđanja bodovnih učinaka igrača za nadolazeće kolo u igri FPL. Slično tomu, drugo potpoglavlje daje pregled radova o optimizaciji odabira FPL momčadi. Razmatraju se egzaktna metode poput mješovitog cjelobrojnog linearnog programiranja, metaheuristike te pristupi utemeljeni na učenju. Naposljetku, treće potpoglavlje na temelju toga pregleda izdvaja doprinose ovoga rada.

3.1. Predviđanje učinka igrača u FPL-u

Matthews i suradnici [8] problemu predviđanja bodova u FPL-u pristupaju generativnim Bayesovim modelom. Vjerojatnosni model svakog igrača čine tri komponente: kategorička distribucija statusa u utakmici (počinje utakmicu, ulazi kao zamjena ili ne ulazi uopće), Bernoullijeva distribucija postizanja gola te Bernoullijeva distribucija postizanja asistencije, pri čemu su potonje dvije uvjetovane time da je igrač na terenu. Nad parametrima tih distribucija postavljene su konjugatne apriorne distribucije (Dirichletova nad kategoričkom distribucijom te beta-distribucija nad dvjema Bernoullijevima) koje se Bayesovim pravilom ažuriraju nakon svakog kola, a podatci prethodne sezone mogu poslužiti za određivanje informativnih apriornih distribucija. Bodovi se ne predviđaju

izravno, nego se procjenjuju kao prosjek velikog broja Monte Carlo simulacija utakmice nadograđenih na postojeći momčadski model nogometne utakmice.

Eilertsen i suradnici [9] predviđanje bodova provode trima zasebnim prognostičkim metodama. Prva je metoda tzv. modificirani prosjek: polazi od prosjeka bodova osvojenih u nekoliko posljednjih kola te ga korigira s obzirom na jakost protivnika (procijenjenu Elo ocjenama), prednost domaćeg terena i formu, dok se izglednost nastupa uzima u obzir naknadnom korekcijom prognoza podacima o ozljedama i suspenzijama. Druga je metoda linearna regresija prilagođena zasebno za svaku poziciju, uz odabir značajki LASSO regresijom. Treća metoda kladioničarske kvote (i timske i individualne) pretvara u vjerojatnosti, a zatim u očekivane bodove. Autori zaključuju da je treća metoda osobito korisna na početku sezone, kada za ostale metode još nema dovoljno podataka o dotadašnjim nastupima.

Bonello i suradnici [10] problem predviđanja formuliraju kao binarnu klasifikaciju: procjenjuje se hoće li igrač u pojedinom kolu osvojiti više od šest bodova. Budući da su podatci izrazito neuravnoteženi, jer većina igrača osvaja 0–2 boda po kolu, primjenjuju gradijentno pojačavanje (engl. *gradient boosting*). Metoda je odabrana nakon usporedbe sa strojem potpornih vektora i slučajnim šumama, a izrađeni su zasebni modeli za svaku poziciju. Glavni doprinos rada jest objedinjavanje različitih izvora podataka: uz povijesnu FPL statistiku, model koristi kladioničarske kvote te sentiment iz blogova i novinskih članaka. Sentiment se dobiva prepoznavanjem imenovanih entiteta i analizom sentimenta na razini entiteta. Objave s Twittera nisu poboljšale rezultate zbog pravopisnih pogrešaka, emotikona i nogometnog žargona, koje standardni alati za analizu sentimenta ne interpretiraju pouzdano.

Frees i suradnici [11] istražuju primjenu dubokog učenja, ponajprije konvolucijskih neuronskih mreža (engl. *convolutional neural networks*, CNN), za predviđanje bodova. Njihovu uspješnost uspoređuju s dvama referentnim modelima: hrbatnom regresijom (engl. *ridge regression*) i LightGBM-om. Primjenom zasebnih modela za svaku poziciju, prilagođena jednodimenzijaska CNN mreža (1D-CNN) postiže najbolji prosječni rezultat, uz MSE od 5,83. Time nadmašuje oba referentna modela za približno 13 %, a dotad najbolji model iz literature, LSTM, za 30 %. Najveći doprinos predviđanju pritom daju bodovi iz posljednjih nekoliko kola i težina sljedećeg protivnika. Autori također ispi-

tuju prijenosno učenje (engl. *transfer learning*) primjenom modela Longformer na novinskom korpusu, ali ne pronalaze koristan prediktivni signal. Taj rezultat povezuju s ograničenom količinom podataka i računalnih resursa, a ne nužno s neupotrebljivošću tekstualnih podataka.

Tamimi i Tran [12] izravno se nadovezuju na rad Freesa i suradnika te ga koriste kao referentni model. Statističke podatke obogaćuju sentimentom iz vijesti i podacima o ozljedama igrača (Transfermarkt): sentiment proizvode prednaučeni transformerski klasifikatori (pysentimiento i DeBERTa) primijenjeni na članke o svakom igraču (iz Google Newsa, *The Guardian*a i GDELT-a), a dobivene se oznake glasovanjem objedinjuju u značajku sentimenta svakog igrača. Predstavljaju dvije metode: u prvoj se na skup od oko 70 značajki primjenjuju metode pojačavanja (CatBoost, XGBoost i GB) za regresiju bodova i klasifikaciju nastupa (predviđanje hoće li igrač zaigrati u idućem kolu), pri čemu je CatBoost najbolji među njima. U drugoj metodi slijede Freesovu metodologiju i značajke za hrbatnu regresiju, LightGBM i 1D-CNN, dodaju vlastiti LSTM te uvode tri nove značajke izvedene iz sentimenta vijesti i podataka o ozljedama, vrednujući svaki model s njima i bez njih; konvolucijska mreža s dodatnim značajkama daje najbolji prosječni rezultat i nadmašuje referentni model, smanjujući njegov prosječni MSE sa 6,27 na 5,63 (u vlastitoj reprodukciji). Autori uočavaju da sentiment iz vijesti više doprinosi klasifikaciji nastupa nego regresiji bodova.

Groos [13] u svom radu predstavlja OpenFPL, metodu koja po pozicijama gradi ansamble kombiniranjem modela XGBoost i slučajnih šuma. Za razliku od vodećih komercijalnih servisa, koji koriste i vlasničke izvore, OpenFPL se temelji isključivo na javno dostupnim podacima: FPL API-ju i naprednim statistikama s Understata, poput očekivanih golova i asistencija te PPDA-a. Značajke o igraču, momčadi i protivniku izračunavaju se kao prosjeci u više vremenskih prozora — posljednjih 1, 3, 5, 10 i 38 utakmica. Hiperparametri se ugađaju automatskom pretragom, a uzorci s visokim bodovnim učinkom dobivaju veću težinu pri učenju. Model bodove predviđa izravno, a u evaluaciji na kolima 32–38 sezone 2024./25. dostiže točnost vodećega komercijalnog servisa te ga na igračima s visokim učinkom i nadmašuje.

Za razliku od Groosove metode OpenFPL, Notelid i Östlund [14] ne predviđaju ukupan broj bodova izravno: umjesto toga, razvijaju zasebne modele za pojedine sastavnice

FPL bodovanja kao što su golovi, asistencije i primljeni golovi, dok logističkom regresijom procjenjuju hoće li igrač nastupiti u početnoj postavi. Dobivene procjene zatim zbrajaju u ukupni broj bodova. U svom radu uspoređuju linearne i logističke regresijske modele s XGBoostom te pritom ostvaruju slične rezultate. U simulaciji sezone 2024./25., koja obuhvaća kola 1–21, najbolji rezultat postižu linearni modeli s 1 293 boda, u usporedbi s 1 256 bodova za XGBoost. Takav bi rezultat približno odgovarao plasmanu među najboljih 12 % menadžera.

3.2. Optimizacija odabira momčadi u FPL-u

Matthews i suradnici [8] sekvencijalni problem sastavljanja momčadi tijekom sezone postavljaju kao Markovljev proces odlučivanja (engl. *Markov decision process*, MDP), a budući da su prave sposobnosti igrača nepoznate, koriste njegovu inačicu nad prostorom vjerovanja (engl. *belief-state Markov decision process*), ekvivalentnu djelomično vidljivom Markovljevom procesu odlučivanja (engl. *partially observable Markov decision process*, POMDP), u kojoj stanje obuhvaća vjerovanje o sposobnostima igrača izraženo vjerojatnosnom distribucijom. Najbolje rezultate postižu Bayesovim Q-učenjem, kod kojeg se istraživanje prostora poteza usmjerava vrijednošću savršene informacije (engl. *value of perfect information*, VPI). Q-vrijednost svakog poteza procjenjuje se gledanjem nekoliko kola unaprijed, uz diskontni faktor (engl. *discount factor*) koji umanjuje težinu udaljenijih kola, čime se u obzir uzima i utjecaj na buduća kola, a ne samo na neposredno sljedeće. Generiranje pojedinog poteza, odnosno odabir 15 igrača, formulira se kao višedimenzijski problem naprtnjače i rješava cjelobrojnim programiranjem (rješavač CPLEX), pri čemu se mogućnost dodatnih (kaznenih) transfera modelira fiktivnim predmetima (engl. *dummy items*) negativne vrijednosti (-4) u funkciji cilja. Time rad objedinjuje predviđanje i optimizaciju u jedinstvenu petlju odlučivanja. Automatizirani igrač u prosjeku se plasira unutar gornjeg percentila (otprilike najboljih 1 %) među 2,5 milijuna ljudskih sudionika.

Eilertsen i suradnici [9] cjelovito upravljanje FPL momčadi tijekom sezone formaliziraju kao stohastički mješoviti cjelobrojni linearni program koji nazivaju problemom odlučivanja u FPL-u (engl. *Fantasy Premier League Decision Problem*, FPLDP). Budući da je rješavanje potpunog stohastičkog modela za cijelu sezonu računalno neizvedivo

(broj scenarija raste eksponencijalno, što dovodi do kombinatorne eksplozije), problem rješavaju deterministički, i to heuristikom kliznog horizonta s točkastim procjenama bodova: odluke donose kolo po kolo, uz razmatranje nekoliko kola unaprijed. U evaluaciji na sezoni 2017./18. najbolja metoda predviđanja (modificirani prosjek) doseže približno gornjih 8 % plasmana, dok preostale dvije metode padaju na 13–30 %; tek uvođenje ograničenja rizika rezultat podiže na otprilike gornjih 1,5 %. Upravo se time nesigurnost ishoda izravno uključuje u optimizaciju.

Notelid i Östlund [14] inicijalnu momčad od 15 igrača te tjednu početnu postavu određuju cjelobrojnim linearnim programiranjem s binarnim varijablama, uz rješavač CBC. Transfere odabiru pohlepnom heuristikom, odnosno biraju onaj s najvećom neposrednom koristi bez planiranja za više kola unaprijed. Posebnost rada jest povezivanje optimizatora s velikim jezičnim modelom ChatGPT, koji predviđanja i SHAP-vrijednosti pretvara u razumljiva objašnjenja i preporuke putem razgovornoga sučelja. Budući da je sustav izveden kao razgovorni asistent, ovaj se rad konceptualno najviše približava ideji interaktivnog FPL asistenta za korisnika.

Uz opisane egzaktnu pristupe, u literaturi su zastupljene i metaheuristike. Aribo [15] problem sastavljanja momčadi rješava genetskim algoritmom, pri čemu se ograničenja igre (proračun i struktura momčadi po pozicijama) u funkciju dobrote uključuju kao kaznene komponente, a igrači se, umjesto predviđanjem budućih bodova, vrednuju prosječnim brojem bodova po utakmici iz povijesnih podataka sezone 2023./24. Autor uspoređuje deset konfiguracija operatora genetskog algoritma (tri strategije selekcije, tri operatora križanja i tri operatora mutacije) te utvrđuje da najbolja rješenja dosljedno daje kombinacija turnirske selekcije, uniformnog križanja i mutacije premetanjem (engl. *scramble mutation*).

3.3. Doprinosi ovoga rada

Literatura o predviđanju bodova uglavnom se oslanja na dva načina oblikovanja problema: ukupni broj bodova može se predviđati izravno, jednim modelom, ili se najprije može rastaviti na pojedine sastavnice bodovanja, nakon čega se svaka od njih procjenjuje zasebnim modelom. Iako se oba pristupa često pojavljuju u postojećim radovima, oni se u pravilu ne vrednuju jedan uz drugi, nego se svaki rad temelji na vlastitoj arhitekturi,

vlastitom skupu podataka i značajki, vlastitoj podjeli podataka te vlastitim metrikama. Zbog toga se razlike u dobivenim rezultatima ne mogu pouzdano povezati samo s odabranom arhitekturom, jer u njima istodobno sudjeluju i sve ostale metodološke razlike među radovima. Prvi doprinos ovoga rada stoga je izravna usporedba navedenih pristupa u kontroliranom eksperimentalnom okruženju, u kojem je arhitektura jedini važniji element koji se mijenja (poglavlje 5.).

Drugi doprinos odnosi se na način vrednovanja modela. Osim standardnih mjera pogreške, koje sažimaju prosječno odstupanje predviđenih bodova od stvarno ostvarenih, u usporedbu su uključene i metrike koje ispituju praktično važnije aspekte predviđanja: rangiranje igrača unutar kola, izbor kapetana, prepoznavanje osobito visokih i osobito niskih učinaka te kalibraciju vjerojatnosti nastupa. Dio tih metrika preuzet je iz postojeće literature, a dio je prilagođen specifičnostima promatrane domene (potpoglavlje 5.3.4.). Na taj se način model ne svodi na jednu ukupnu ocjenu, nego se može preciznije vidjeti koliko pogrešku čini i u kojim se okolnostima ta pogreška najčešće pojavljuje.

Treći doprinos proizlazi iz vrednovanja predviđanja u stvarnom postupku odlučivanja. Umjesto da se modeli uspoređuju samo prema metrikama izračunanim na razini pojedinačnih susreta, u radu se povijesnom simulacijom cijele sezone ispituje koliko se razlike među predviđanjima zaista prenose na konačan broj osvojenih bodova. Pritom se za upravljanje momčadi koristi isti optimizacijski mehanizam, dok se mijenja samo izvor predviđanja na kojem se odluke temelje (poglavlje 6.). Svi navedeni elementi na kraju su objedinjeni u web aplikaciju, putem koje su predviđanja i preporuke za upravljanje momčadi dostupni krajnjem korisniku (poglavlje 7.).

4. Prikupljanje, integracija i eksploratorna analiza podataka

Prediktivni je model dobar onoliko koliko su dobri podatci iz kojih uči, pa je ovo poglavlje posvećeno upravo opisu procesa izgradnje korištenog podatkovnog skupa i njegovoj analizi. Poglavlje je podijeljeno na ukupno pet potpoglavlja u kojima se redom predstavljaju i obrađuju korišteni izvori podataka (4.1.), struktura i vremenska pokrivenost dobivenog skupa (4.2.), njegova eksploratorna analiza (4.3.), kronološka podjela i problem hladnog starta (4.4.) te sažeti pregled veza između tih svojstava i odluka donesenih pri modeliranju (4.5.).

4.1. Izvori podataka

Predviđanje broja bodova koje će pojedini igrač ostvariti u nadolazećem kolu zahtijeva podatke koje nijedan izvor ne nudi u cijelosti, pa je podatkovni skup izgrađen iz triju izvora: službenih podataka FPL-a, platforme za nogometnu analitiku Understat te platforme za praćenje sportskih susreta Sofascore. U nastavku će biti više riječi o svakom od navedenih izvora.

4.1.1. Službeni podatci Fantasy Premier Leaguea

Budući da je sa službenog FPL API-ja moguće dohvatiti podatke samo za tekuću sezonu, povijesni se podatci u ovom radu preuzimaju iz repozitorija *vaastav/Fantasy-Premier-League* [16]. Riječ je o repozitoriju koji održava šira zajednica korisnika i u kojemu se podatci od sezone 2016./17. nadalje arhiviraju kontinuirano za svako odigrano kolo. On se u istraživanjima vezanima uz FPL nametnuo kao *de facto* standardni izvor podataka jer sadrži povijesne zapise na razini igrač–kolo. Ti zapisi čine okosnicu objedinjenog podatkovnog skupa u ovom radu, a često služe i kao polazište u srodnim istraživanjima.

Uz ciljnu varijablu, odnosno ukupan broj bodova koje je igrač ostvario u susretu (`total_points`), spomenuti zapisi sadrže i kompletan popis događaja susreta koji ulaze u bodovanje, odnosno tzv. *box-score*. To uključuje odigrane minute, golove, asistencije, sačuvane mreže (engl. *clean sheet*), primljene golove, vratarske obrane, obranjene i promašene jedanaesterce, autogolove, kartone te bonus bodove. Dostupan je i indeks BPS (engl. *Bonus Points System*), na temelju kojega se dodjeljuju bonus bodovi.

Zapisi dodatno sadržavaju FPL-ove kompozitne pokazatelje učinka *influence*, *creativity* i *threat*, koji redom mjere ukupan utjecaj igrača na tijek susreta, kreativnost pri stvaranju prigoda te izravnu prijetnju protivničkim vratima. Uključeno je i stanje tržišta uoči svakog kola, poput cijene igrača, broja korisnika koji ga imaju u svojoj momčadi te broja ulaznih i izlaznih transfera.

4.1.2. Podatci platforme Understat

Understat [17] mrežna je platforma za nogometnu analitiku koja objavljuje agregirane statistike igrača i momčadi te zapise o pojedinačnim udarcima. Njezina je glavna značajka pokazatelj očekivanih golova (engl. *expected goals*, xG), odnosno modelna procjena vjerojatnosti da će pojedinačni udarac završiti pogotkom. U ovom se radu koriste podatci za natjecanje *Premier League*.

Podatci s platforme Understat uključeni su na trima razinama granularnosti. Na razini igrača i susreta dostupni su ostvareni golovi i asistencije, broj udaraca i ključnih dodavanja te očekivani golovi (xG), očekivane asistencije (xA) i očekivano ostvareni golovi bez jedanaesteraca (npxG). Dostupne su i dvije mjere sudjelovanja igrača u napadačkim posjedima: xGChain i xGBuildup. Pokazatelj xGChain igraču pripisuje xG svih udaraca iz posjeda u kojima je sudjelovao, pa obuhvaća doprinosa cijelom slijedu napada, a ne samo završnom udarcu ili asistenciji. Pokazatelj xGBuildup primjenjuje istu logiku, ali izostavlja igrača koji je izveo udarac i igrača koji je uputio ključno dodavanje koje mu je prethodilo, pa je prikladniji za opis doprinosa izgradnji napada prije njegove završne faze.

Na razini momčadi i susreta dostupni su očekivani golovi (xG), primljeni očekivani golovi (xGA), PPDA te broj tzv. dubokih dovršenih dodavanja (engl. *deep completions*). PPDA (engl. *passes allowed per defensive action*) omjer je broja dodavanja koja momčad

dopušta protivniku i broja vlastitih obrambenih akcija na protivničkoj polovici. Niža vrijednost PPDA upućuje na intenzivniji pressing, odnosno na češće ometanje protivničke izgradnje napada. Pokazatelj *deep completions* broji uspješna dodavanja, bez ubacivanja, koja završavaju približno 18 metara od protivničkih vrata.

Zapisi o pojedinačnim udarcima sadržavaju informacije o položaju i obilježjima udarca, poput dijela tijela kojim je upućen, situacije iz koje je nastao i dodijeljene vrijednosti xG. Iz njih se izvode dodatni pokazatelji strukture napada, primjerice udio udaraca iz kaznenog prostora, udio udaraca glavom te udio udaraca nastalih nakon prekida.

Glavna je vrijednost ovog izvora njegova vremenska pokrivenost. Dok su FPL-ove očekivane statistike u korištenoj arhivi dostupne tek od sezone 2022./23., Understat za cijelo promatrano razdoblje pruža istovrsne pokazatelje na razini igrača, momčadi i pojedinačnih udaraca, čime je moguće konstruirati vremenski usporedive varijable za sve promatrane sezone.

4.1.3. Podatci platforme Sofascore

Sofascore [18] mrežna je platforma za praćenje sportskih susreta koja, uz rezultate u stvarnom vremenu, pruža opsežan skup statistika igrača i momčadi na razini pojedinačnog susreta. U odnosu na FPL i Understat, njezina je glavna vrijednost znatno šira pokrivenost pokazatelja individualnog učinka, osobito onih koji nisu izravno povezani s udarcima ili fantasy bodovanjem.

Na razini igrača dostupne su, među ostalim, statistike dodavanja, duela, obrambenih akcija, driblinga i akcija specifičnih za vratare, poput obrana, hvatanja ubačaja i istrčavanja. Dostupna je i Sofascoreova kompozitna ocjena nastupa na ljestvici od 3 do 10. Budući da se ocjena izračunava iz većeg broja statističkih pokazatelja prema metodologiji same platforme, u ovom se radu tumači kao sažeti, izvorno specifičan pokazatelj nastupa, a ne kao zasebno opažena statistika.

Takvi su podatci posebno vrijedni za braniče i vratare, čiji je pojedinačni učinak u službenim FPL podacima opisan ograničenim brojem pokazatelja. Za obje momčadi dostupne su i osnovne statistike susreta, poput posjeda lopte, udaraca, udaraca iz kuta i ulazaka u završnu trećinu.

4.2. Objedinjeni podatkovni skup

4.2.1. Vremenska pokrivenost i struktura podatkovnog skupa

Korišteni podatkovni skup sadrži povijesne zapise iz sedam završenih sezona, 2018./19.–2024./25. Starije su sezone pritom isključene iz dvaju razloga. Prvo, što se ide dublje u prošlost, to je Sofascoreova statistička pokrivenost oskudnija. Drugo, što je sezona starija, to slabije opisuje današnju igru: pravila su se s vremenom mijenjala (primjerice, u sezoni 2019./20. je uvedena upotreba VAR-a, a od sezone 2022./23. pa nadalje dopušteno je napraviti pet zamjena umjesto tri), a izmjenjuju se i igrači, klubovi i stilovi igre, pa se obrasci naučeni na vrlo starim sezonama slabije prenose na predviđanje današnje igre. Sezona 2025./26. nije obuhvaćena jer se od nje mijenjaju FPL-ova pravila bodovanja, ponajprije uvođenjem bodova za obrambeni doprinos (engl. *defensive contribution*), pa ciljna varijabla više nije definirana jednako kao u promatranom razdoblju, unutar kojega su pravila bodovanja igrača posve stabilna.

Rezultat postupka integracije prethodno opisanih izvora podataka je jedna tablica u CSV formatu sa 177 997 redaka i 286 stupaca (približno 155 MB na disku). Svaki redak opisuje jedan par (igrač, susret), tj. jednog registriranog igrača jedne momčadi u jednom susretu natjecanja *Premier League*, neovisno o tome je li igrač u tom susretu nastupio ili ne (pritom u kolu u kojemu momčad igra dva susreta igrač stoga ima dva retka). Stupci se prema izvoru i razini granularnosti na kojoj su zabilježeni dijele u pet različitih blokova (tablica 4.1.), pri čemu se preostalih sedam stupaca odnosi na pogodnost pomoćnog menadžera (engl. *assistant manager*), uvedenu u sezoni 2024./25., i pri učitavanju se odbacuje.

Tablica 4.1. Raspodjela stupaca objedinjenog podatkovnog skupa po blokovima.

Blok stupaca	Broj stupaca
FPL, osnovni blok: identifikatori, kontekst susreta, fantasy <i>box-score</i> , kompozitni pokazatelji učinka, stanje tržišta	54
Understat, razina igrača: xG, xA, npxG, xGChain, xGBuildup, udarci, ključna dodavanja, struktura udaraca	23
Understat, razina momčadi: momčadski xG i xGA, PPDA, <i>deep completions</i>	10
Sofascore, razina igrača: dodavanja, dueli, obrambene akcije, vratarske akcije, ocjena nastupa	98
Sofascore, razina momčadi i protivnika: statistika susreta iz perspektive igrača	94
Stupci pomoćnog menadžera (odbacuju se pri učitavanju)	7

4.2.2. Čišćenje i standardizacija pri učitavanju

Pri svakom se učitavanju skupa — kako u analizi podataka, tako i pri izgradnji modela — primjenjuje isti, fiksni korak čišćenja. Čišćenje obuhvaća četiri zahvata: (i) uklanjaju se potpuno identični retci (u trenutačnoj ih izgradnji skupa nema, no korak postoji isključivo iz opreznosti); (ii) uklanjaju se 322 retka za koje je zabilježena pozicija “AM”, tj. zapisi stvarnih klupskih menadžera koji u skup ulaze preko prethodno spomenute pogodnosti pomoćnog menadžera (ostvaruju bodove i bez odigranih minuta); (iii) 101 redak s oznakom pozicije “GKP”, anomalijom API-ja iz jednog kola sezone 2021./22., preimenuje se u “GK”; te (iv) odbacuje se devet stupaca: sedam stupaca vezanih uz pomoćnog menadžera, jedna interna zastavica izvorne arhive i jedan potpuno prazan stupac. Očišćeni skup ima 177 675 redaka, s distribucijom po pozicijama: MID 74 884, DEF 60 493, FWD 22 702 i GK 19 596.

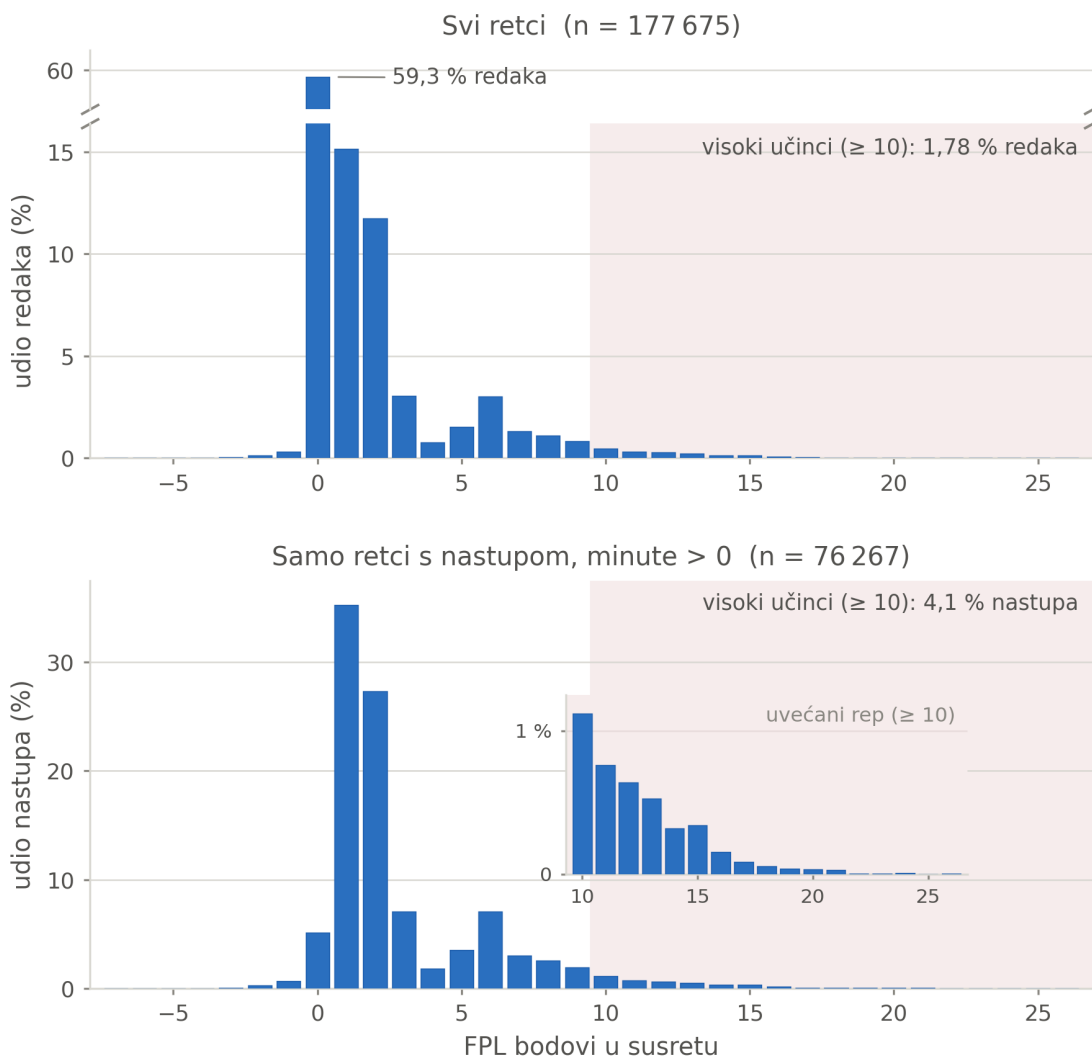
4.3. Eksploratorna analiza

Prije početka izgradnje modela, potrebno je analizirati korišteni skup podataka te utvrditi njegova strukturna svojstva kako bismo pri njegovoj izgradnji mogli donijeti adekvatne odluke. U ovom su potpoglavlju stoga predstavljeni rezultati kratke eksploratorne analize provedene na očišćenom skupu iz potpoglavlja 4.2. te su navedene odluke koje su u skladu s njima donesene u oblikovanju prediktivnih modela.

4.3.1. Ciljna varijabla

Ciljna varijabla u problemu predviđanja je broj bodova jednog igrača u jednom susretu (`total_points`). Na očišćenom skupu njezin prosjek iznosi 1,23, a standardna devijacija 2,42. Vrijednosti se kreću od -7 do 26, a distribucija, koja je vidljiva na slici 4.1., izrazito odstupa od normalne: 59,3 % svih vrijednosti jednako je nuli, dok je samo 1,78 % vrijednosti dvoznamenkasto. Uz to, desno je asimetrična (koeficijent asimetrije 2,97) te ima znatno teže repove od normalne distribucije (koeficijent zaobljenosti 13,8, naspram 3 za normalnu distribuciju).

Gomila nula gotovo je u cijelosti posljedica (ne)nastupanja, a ne loših izvedbi. Podjela redaka prema odigranim minutama (tablica 4.2.) pokazuje da je 57,1 % redaka bez



Slika 4.1. Distribucija ciljne varijable na svim retcima (gore) i samo na retcima s nastupom (dolje).

nastupa (oni po definiciji nose nula bodova i čine 96,3 % svih redaka s nulom), dok među igračima koji su nastupili samo njih 5,2 % ostaje na nuli.

Tablica 4.2. Podjela redaka prema odigranim minutama i podrijetlo nula u ciljnoj varijabli.

Odigrane minute	Broj redaka	Udio	Prosječni bodovi	Udio s 0 bodova
Bez nastupa (0 min)	101 408	57,1 %	0,00	100 %
Kratki nastup (1–59 min)	20 877	11,8 %	1,25	7,7 %
Dugi nastup (≥ 60 min)	55 390	31,1 %	3,49	4,2 %

Takva distribucija ciljne varijable opravdava ključnu arhitekturnu odluku: dvostupanjsku strukturu modela. Predviđanje se pritom razdvaja na procjenu vjerojatnosti nastupa i procjenu očekivanih bodova uz uvjet da igrač nastupi. Jedinstveni regresor

učen na svim retcima ponajprije bi učio razlikovati nastupe od izostanaka, što je u osnovi klasifikacijski problem, umjesto da razlikuje učinke među igračima koji su doista nastupili.

Distribucija s dugim desnim repom otvara i drugi problem: visoki učinci (engl. *hauls*) čine tek oko 4 % redaka s nastupom. Model koji minimizira neponderiranu kvadratnu pogrešku zato će predviđanja povlačiti prema najčešćim vrijednostima od 1–2 boda te sustavno podcjenjivati rijetke visoke učinke. Postupak učenja mora se toj pojavi izričito prilagoditi.

4.3.2. Razlike po pozicijama

Pozicija igrača jedan je od glavnih izvora heterogenosti u problemu predviđanja, jer određuje i FPL-ovo vrednovanje događaja i obrasce nastupanja u stvarnim susretima. Isti događaj ne donosi jednak broj bodova svim pozicijama: gol vrataru i braniču vrijedi šest bodova, veznom igraču pet, a napadaču četiri, dok sačuvana mreža vrataru i braniču donosi četiri boda, veznom igraču jedan, a napadaču nijedan. Pozicije se razlikuju i prema očekivanom vremenu provedenom u igri: vratar (ako je u početnom sastavu) najčešće brani cijeli susret, dok su igrači ostalih pozicija (osobito napadači) češće izloženi rotaciji i izmjenama tijekom susreta. Zajedničko promatranje svih pozicija prikriva bitne razlike u raspodjeli nastupa i bodova, što je možda i najbolje vidljivo u tablici 4.3.

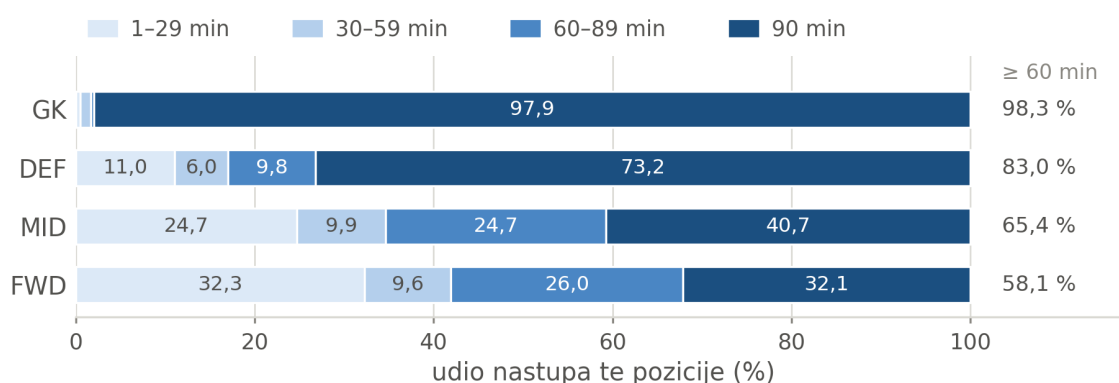
Tablica 4.3. Struktura nastupa i distribucija ciljne varijable po pozicijama.

Pozicija	Svi retci		Samo retci s nastupom				
	Broj redaka	S nastupom	≥ 60 min	Prosječni bodovi	St. dev.	Maks.	Visoki učinci
GK	19 596	27,4 %	98,3 %	3,52	2,83	16	4,57 %
DEF	60 493	42,3 %	83,0 %	2,73	2,99	24	2,95 %
MID	74 884	46,9 %	65,4 %	2,81	2,88	26	4,60 %
FWD	22 702	44,9 %	58,1 %	3,12	3,29	24	5,27 %

Nastupi vrataru su, kao što je već istaknuto, praktički binarni: vratar ili brani cijeli susret ili uopće ne nastupa, gotovo bez ičega između, što je sada vidljivo i u brojkama. Naime, vratari nastupaju u samo 27,4 % susreta za koje su registrirani, a čak 98,3 % tih nastupa traje barem 60 minuta, što znači da od tri do četiri vratara registrirana po momčadi tijekom susreta uglavnom brani samo jedan. Na ostalim je pozicijama udio nastupa sličan, između 42 i 47 %, ali se trajanje nastupa razlikuje: udio nastupa od najmanje

60 minuta iznosi 83,0 % kod braniča, 65,4 % kod veznih igrača i 58,1 % kod napadača (slika 4.2.), a prema napadačkim pozicijama raste udio kraćih ulazaka i ranijih izmjena — kod napadača je 41,9 % nastupa kraće od jednog sata. Predviđanje nastupa stoga nije jednak problem za sve pozicije: kod vratara ponajprije ovisi o statusu prvoga vratara i njegovoj raspoloživosti, a kod igrača ostalih pozicija o rotaciji momčadi i odluci trenera. To opravdava izgradnju zasebnih modela nastupa za svaku poziciju.

Unutar samih nastupa podjela na kratke i duge nastupe otvara još jedan problem: broj bodova raste s odigranim minutama (tablica 4.2.), a koliko će minuta igrač dobiti u nadolazećem susretu unaprijed nije poznato — odigrane minute su ishod susreta, pa bi njihova upotreba kao ulazne varijable značila curenje informacija. Model očekivanih bodova uz uvjet nastupa stoga mora očekivani broj minuta procijeniti posredno, iz povijesnih pokazatelja poput kliznog prosjeka odigranih minuta i udjela susreta započetih u početnoj postavi tijekom prethodnih kola.



Slika 4.2. Struktura odigranih minuta uz uvjet nastupa, po pozicijama.

Promatraju li se stupci tablice 4.3. koji daju detaljniji uvid u svojstva redaka s nastupom, može se primijetiti da se pozicije vidljivo razlikuju po prosječnom učinku, raspršenosti učinaka oko tog prosjeka te po učestalosti i dometu ekstremnih učinaka. Vratari imaju najviši prosječni učinak (3,52 boda po nastupu) i najmanju raspršenost (standardna devijacija 2,83), ali i najniži domet: nijedan vratarski učinak u svih sedam sezona ne prelazi 16 bodova. Braniči imaju najniži prosjek (2,73 boda) te najrjeđe visoke učinke (2,95 % nastupa), dok vezni igrači i napadači visoke učinke postižu znatno češće (4,60 % odnosno 5,27 % nastupa), a pojedinačni im se učinci penju do 26 odnosno 24 boda. Napadači pritom imaju i najveću raspršenost učinka (standardna devijacija 3,29).

Sve navedene razlike — različiti prosjeci, različite raspršenosti i, kako pobliže opisuje potpoglavlje 4.3.3., bitno različiti mehanizmi nastanka visokih učinaka — empirijski su argument za učenje zasebnog modela za svaku poziciju umjesto jednoga zajedničkog modela u kojem bi pozicija bila tek jedna od ulaznih varijabli.

4.3.3. Raščlamba visokih učinaka

S obzirom na to da se bodovni učinci igrača računaju pomoću FPL-ove formule, svaki se visoki učinak može pouzdano rastaviti na komponente bodovanja od kojih je nastao. Udjeli pojedinih komponenti u visokim učincima, koje možemo vidjeti u tablici 4.4., pokazuju da se mehanizam njihova nastanka od pozicije do pozicije poprilično razlikuje. Vezni igrači i napadači do visokog učinka u pravilu dolaze samo izravnim napadačkim doprinosom (gol ili asistenciju). Gotovo svi njihovi visoki učinci sadrže gol (91,5 % odnosno 95,7 %), bonus bodovi rijetko izostaju, a sačuvana mreža (koja veznom igraču donosi samo jedan bod, a napadaču nijedan) ima zanemarivu ulogu.

Nasuprot tomu, visoki učinci vratara i braniča najčešće zahtijevaju istodobno ostvarenje više rijetkih događaja. Naime, kod braniča 75,7 % visokih učinaka istodobno sadrži sačuvanu mrežu i napadački doprinos, dakle dva pojedinačno rijetka događaja koja se moraju poklopiti u istom susretu. Ta istodobnost nije slučajna: nastup, sačuvana mreža i najviša tri bonus boda braniču zajedno donose najviše devet bodova, pa je za dvozna-
menkasti učinak nužan i napadački doprinos. Vratari pak do visokih učinaka dolaze spojem sačuvane mreže (prisutna u 85,0 % slučajeva) i velikog broja obrana ili obranjenog jedanaesterca, uz gotovo neizostavne bonus bodove, dok je napadački doprinos u njima rijetkost.

Tablica 4.4. Udio visokih učinaka (nastupa s najmanje 10 bodova) koji sadrže pojedinu komponentu bodovanja, po pozicijama.

Pozicija	Gol	Asistencija	Sačuvana mreža	Bonus	Mreža i napadački doprinos
GK	0,4 %	4,5 %	85,0 %	99,2 %	3,7 %
DEF	54,6 %	53,1 %	75,7 %	96,8 %	75,7 %
MID	91,5 %	44,9 %	44,3 %	98,1 %	44,3 %
FWD	95,7 %	54,9 %	33,3 %	99,3 %	33,3 %

Taj obrazac objašnjava zašto su visoki učinci braniča najrjeđi od svih pozicija i ujedno najavljuje trajnu poteškoću pri modeliranju: predviđanje visokih učinaka braniča i vratara ostaje najtežim potproblemom, neovisno o kvaliteti ulaznih varijabli.

4.4. Kronološka podjela i problem hladnog starta

Razvoj modela u ovome radu počiva na strogo kronološkoj podjeli svih sedam promatranih sezona: skup za učenje čine sezone 2018./19.–2022./23. (120 667 redaka), skup za provjeru sezona 2023./24. (29 725 redaka), a ispitni skup sezona 2024./25. (27 283 retka). Ovakva podjela nije proizvoljna, već nužna: naime, retci su unutar iste sezone međusobno snažno ovisni (isti se igrači ponavljaju iz kola u kolo, a klizni prozori susjednih kola gotovo u cijelosti dijele istu povijest), pa bi nasumična podjela gotovo istovjetne retke rasporedila s obje strane granice. Zbog toga bi konačan model bio ispitan na onome što je pri učenju već vidio, a njegova bi uspješnost time bila precijenjena. Iz istog se razloga pri ugađanju hiperparametara u poglavlju 5. primjenjuje kronološka unakrsna provjera s rastućim skupom za učenje: svaki se model provjerava na vremenski kasnijem dijelu podataka od onoga na kojem je učen. Naposljetku, kronološki poredak odgovara i stvarnim uvjetima primjene koje ovaj rad oponaša — model naučen na prošlosti predviđa sezonu koju nikada nije vidio.

Kronološka podjela otvara i jedan strukturni problem čiji razmjer vrijedi izraziti brojkama: hladni start (engl. *cold start*). Svake sezone u ligu ulaze tri nova kluba, klubovi dovode pojačanja, a iz njihovih akademija stižu debitanti, pa u svakoj sezoni nakon prve između 19 i 30 % redaka pripada igračima kojih nema ni u jednoj ranijoj sezoni skupa — bilo zato što u ligi nastupaju prvi put, bilo zato što se u nju vraćaju nakon izbjavanja. U ispitnoj sezoni 2024./25. riječ je o 192 igrača i 5 349 redaka, što je 19,6 % cijele sezone. Hladni start ne podrazumijeva potpuni izostanak informacija. Značajke poznate uoči susreta, poput domaćinstva, cijene, udjela vlasništva i transfera, dostupne su od trenutka kada je igrač evidentiran u FPL-u. Nasuprot tomu, igračeve povijesne značajke prije prvog nastupa nisu definirane, a u nekoliko početnih susreta temelje se na vrlo malom broju opažanja. Nedostajuće se vrijednosti zato popunjavaju prosjekom odgovarajuće pozicije, izračunatim samo na pripadajućem skupu za učenje. Tako se ne uvode informacije iz budućnosti, a igrač bez povijesti dobiva neutralnu početnu procjenu koja se postupno zamjenjuje njegovim vlastitim povijesnim pokazateljima.

4.5. Od svojstava podataka do odluka pri modeliranju

Tablica 4.5. sažima cijelo poglavlje: svaki njezin redak navodi jedno utvrđeno svojstvo podataka i odluku pri oblikovanju modela koju to svojstvo povlači u poglavlju 5.

Tablica 4.5. Sažeti pregled veza između svojstava podataka i odluka pri modeliranju.

Svojstvo podataka	Posljedica pri modeliranju
57 % redaka otpada na izostanke nastupa; iz njih potječe 96 % svih nula; samo 5 % nastupa ostaje bez ijednog boda	Dvostupanjska struktura: vjerojatnost nastupa $\times$ očekivani bodovi uz uvjet nastupa
Nastup vratara gotovo je binaran; minute igrača iz polja uz uvjet nastupa bimodalno su raspodijeljene	Zasebni klasifikatori nastupa po pozicijama; očekivano vrijeme na terenu ulazi u model učinka preko kliznih minuta i udjela susreta u početnoj postavi
Desno asimetrična ciljna varijabla; visoki učinci čine tek oko 4 % redaka s nastupom	Ponderiranje primjera obrnuto proporcionalno učestalosti, kao brana stezanju predviđanja prema modu na 1–2 boda
Visoki učinci veznih igrača i napadača poživaju na jednom događaju — izravnom napadačkom doprinosu; kod vratara i braniča traže konjunktiju više rijetkih događaja	Zasebni modeli po pozicijama; dekompozicija po komponentama; prepoznavanje visokih učinaka vrednuje se odzivom među najboljih K predviđanja
FPL-ove očekivane statistike dostupne su tek od sezone 2022./23., dok Understat istovrsne pokazatelje pruža za cijelo promatrano razdoblje	Understat kao izvor povijesnih očekivanih golova, čime značajke ostaju vremenski usporedive u svim sezonama
Oko 20 % redaka ispitne sezone pripada igračima hladnog starta, bez ikakve povijesti u skupu	Nedostajuće povijesne značajke popunjavaju se prosjekom odgovarajuće pozicije, izračunatim isključivo na sezonama za učenje — nikada nulama

5. Modeli za predviđanje bodova igrača

U ovom je poglavlju opisan razvoj komponente inteligentnog asistenta čija je namjena predviđanje bodovnog učinka pojedinog igrača za nadolazeće kolo. Uspoređuju se dva pristupa modeliranju: pristup izravnog predviđanja (u nastavku: A1), u kojem očekivani broj bodova uz uvjet nastupa predviđa jedan regresijski model po poziciji, i pristup dekompozicije bodova po komponentama (u nastavku: A2), u kojem svaku bodovnu komponentu (golove, asistencije, sačuvane mreže, bonus itd.) predviđa zaseban model, a konačno se predviđanje za pojedinog igrača dobiva njihovim zbrajanjem pomoću službene FPL-ove formule bodovanja. Osim arhitekture modela, svi ostali elementi — podatci, ulazne značajke, podjela skupa podataka, referentni modeli i metrike — identični su za oba pristupa, pa se izmjerene razlike mogu prvenstveno pripisati samoj arhitekturi. Poglavlje najprije definira problem predviđanja (potpoglavlje 5.1.) te obrazlaže odabir i učenje modela (potpoglavlje 5.2.), zatim opisuje zajednički metodološki okvir (potpoglavlje 5.3.) i oba pristupa (potpoglavlja 5.4. i 5.5.) te njihovo vrednovanje (potpoglavlje 5.6.).

Dva pristupa nisu razvijena do iste razine detalja. Prvi predstavlja glavni tijek rada, pa je njegova arhitektura oblikovana u više inačica, uz ugađanje hiperparametara zasebno za svaku poziciju i ponderiranje primjera obrnuto proporcionalno njihovoj učestalosti (potpoglavlje 5.4.). Nasuprot tomu, drugi je pristup svjesno zadržan u jednostavnom obliku: uveden je kao eksperiment nad arhitekturom drugoga stupnja, a ne s namjerom da se postigne njegov najveći mogući učinak. Stoga ne uključuje ni ugađanje hiperparametara ni ponderiranje primjera (potpoglavlje 5.5.). Takva asimetrija ne umanjuje važnost usporedbe, već jasno određuje njezino tumačenje. Naime, sve razlike koje nisu izjednačene pogoduju prvom pristupu, pa se svaka prednost drugoga mora promatrati

kao rezultat ostvaren unatoč toj neravnopravnosti, a ne zbog nje.

5.1. Formulacija problema

Neka $y_{p,g}$ označava ukupan broj bodova koje igrač p ostvari u kolu g . Zadatak je uoči roka za predaju momčadi za kolo g izračunati predviđanje $\hat{y}_{p,g}$ za svakog igrača, pri čemu se smiju koristiti isključivo informacije dostupne u tom trenutku. Budući da su izvorni zapisi granularnosti igrač–susret, vrijednost $y_{p,g}$ u dvostrukom kolu predstavlja zbroj bodova ostvarenih u oba susreta igračeve momčadi.

Dobro predviđanje ne znači samo malu razliku između predviđenih i ostvarenih bodova. Asistent na temelju predviđanja predlaže početnu postavu, kapetana i transfere, pa je važno i da model igrača s većim potencijalom pravilno rangira ispred ostalih. Vrednovanje zato, uz mjere pogreške numeričkih predviđanja, obuhvaća metrike poretka igrača, odabira kapetana i kalibracije (potpoglavlje 5.3.4.).

5.2. Odabir i učenje modela

Svi su modeli u ovome poglavlju ansambli stabala odluke izgrađeni gradijentnim pojačavanjem (engl. *gradient boosting decision trees*, GBDT), ostvareni u programskom okviru LightGBM [19]. Takvi se modeli grade postupno, stablo po stablo, pri čemu se svako novo stablo prilagođava negativnome gradijentu funkcije gubitka po predviđanjima dotadašnjega ansambla, pa svaki korak ispravlja pogreške prethodnih.

Osim što su gradijentno pojačana stabla čest izbor za tablične podatke (na kojima u sustavnim usporedbama i dalje nadmašuju modele dubokoga učenja [20]), dodatan argument za njihov izbor u ovom radu daju i dva konkretna svojstva korištenih podataka: heterogenost i kolinearnost. Skup značajki opisan u potpoglavlju 5.3.1. izrazito je heterogen — kako po tipovima pojedinih značajki, tako i po rasponima vrijednosti koje one mogu poprimiti — a uz to su značajke po samoj konstrukciji izrazito kolinearne. Za stabla to nije problem: ona podatke dijele prema poretku vrijednosti, pa su neosjetljiva na monotone transformacije i ne zahtijevaju nikakvo skaliranje ni predobradu značajki. Na njihova predviđanja kolinearnost gotovo da i ne utječe, no nije bezazlena za tumačenje modela: važnost se raspršuje među gotovo istovjetnim stupcima, zbog čega mjere

važnosti značajki treba čitati s oprezom.

Još jedan praktičan razlog za upotrebu LightGBM-a jest njegova brzina. Pragove za podjelu čvora pri gradnji stabla traži nad histogramima: vrijednosti svake značajke unaprijed razvrstava u malen broj razreda (engl. *bins*), pa umjesto svih vrijednosti u čvoru ispituje tek nekoliko stotina kandidata. Stabla usto ne gradi razinu po razinu (engl. *level-wise*), nego uvijek širi onaj list koji obećava najveće smanjenje gubitka (engl. *leaf-wise growth*), čime uz jednak broj listova doseže niži gubitak. Ta je strategija ujedno i sklonija prenaučivosti na malim skupovima, pa su upravo broj listova, dubina stabla i najmanji dopušteni broj primjera u listu hiperparametri koji se ugađaju u potpoglavlju 5.4. Zajedno, ta dva mehanizma čine učenje modela dovoljno brzim da opsežan postupak vrednovanja opisan u nastavku uopće bude izvediv.

5.3. Zajednički metodološki okvir

Kako bi usporedba dvaju pristupa bila valjana, svi elementi osim same arhitekture modela moraju biti jednaki. Ovo potpoglavlje stoga definira njihov zajednički eksperimentalni okvir: postupak izgradnje značajki bez curenja informacija nakon roka za predaju momčadi (potpoglavlje 5.3.1.), dvostupanjsku strukturu predviđanja (potpoglavlje 5.3.2.), kronološku podjelu podataka (potpoglavlje 5.3.3.), metrike vrednovanja (potpoglavlje 5.3.4.) i heurističke referentne modele (potpoglavlje 5.3.5.). Uz tako postavljen okvir pristupi se razlikuju samo u modeliranju očekivanog broja bodova uz uvjet nastupa. Razlike u rezultatima stoga se mogu pripisati upravo toj komponenti modela.

5.3.1. Izgradnja značajki bez curenja informacija

Pri izgradnji značajki slijedi se jedno temeljno pravilo: predviđanje za kolo *g* smije koristiti samo podatke nastale prije kola *g*. Da bi se to pravilo dosljedno poštovalo, sve se povijesne značajke grade istim postupkom u nekoliko koraka. Najprije se retci pojedinačnih susreta agregiraju na razinu (igrač, kolo): statistike se usrednjuju uz zabilježen broj susreta, čime su dvostruka kola ispravno obuhvaćena. Zatim se klizne statistike (engl. *rolling statistics*) računaju isključivo nad odigranim kolima, uključujući i tekuće. Dobiveni se rezultati potom pridružuju natrag svim retcima i unutar svakog igrača popunjavaju posljednjom poznatom vrijednošću, pa retci bez nastupa nasljeđuju vrijednosti

iz posljednjega odigranog kola. Naposljetku se sve vrijednosti pomiču za jedno kolo, tako da svaki redak vidi strogo ranije informacije. Time stupci ishoda (sve zabilježeno tijekom susreta, od golova i minuta do naprednih statistika) u modele ulaze isključivo kao agregirane povijesne vrijednosti, dok stupci koji su istinski poznati već uoči susreta (domaćinstvo, cijena, udio vlasništva, transferi) ulaze izravno.

Tim postupkom nastaje 214 značajki, koje tablica 5.1. sažima po skupinama. Za igrače bez ikakve povijesti u skupu podataka (igrače hladnog starta iz potpoglavlja 4.4.) nedostajuće se povijesne značajke popunjavaju prosjekom za odgovarajuću poziciju, izračunatim isključivo na sezonama za učenje. Iz istoga se skupa izvode dvije liste značajki. Prvu, od 26 značajki, koriste modeli za predviđanje nastupa: čine je cijele skupine minutaže i uloge, konteksta susreta i tržišta te pet značajki o kartonima. Drugu, od 199 značajki, koriste modeli za predviđanje učinka: nju čine sve značajke osim minutaže i uloge, rednoga broja kola, razmaka između susreta momčadi i indikatorskih varijabli pozicije. Petnaest značajki (kartoni, domaćinstvo, cijena, udio vlasništva i transferi) ulazi u obje liste, pri čemu su obje liste jednake za oba pristupa.

5.3.2. Dvostupanjska struktura predviđanja

Velik udio nula u ciljnoj varijabli, koji je ponajprije posljedica nenastupanja igrača (potpoglavlje 4.3.1.), opravdava zajedničku dvostupanjsku strukturu obaju pristupa. Neka $m_{p,g}$ označava broj minuta koje je igrač p odigrao u kolu g . Budući da igrač bez nastupa osvaja nula bodova, očekivani se broj bodova može rastaviti na vjerojatnost nastupa i očekivani broj bodova uz uvjet nastupa. Predviđanje se stoga računa kao

$$\hat{y}_{p,g} = \hat{P}(m_{p,g} > 0) \cdot \hat{E}[y_{p,g} | m_{p,g} > 0]. \quad (5.1)$$

Prvi stupanj, *model nastupa*, procjenjuje prvi činitelj iz izraza (5.1). U oba je pristupa riječ o binarnome klasifikatoru ostvarenom u LightGBM-u, učenom minimiziranjem binarne unakrsne entropije (engl. *binary log-loss*). Ciljna varijabla označava je li igrač odigrao barem jednu minutu, a model se uči na svim retcima pojedine pozicije, uključujući i retke s nastupom i retke bez nastupa, uz 26 značajki za predviđanje nastupa. Drugi stupanj, *model učinka*, procjenjuje drugi činitelj iz izraza (5.1) te se uči isključivo na retcima s

Tablica 5.1. Skupine izvedenih značajki (ukupno 214), razvrstane prema tome što opisuju.

Skupina	Broj	Sadržaj
Bodovna forma	24	klizni prosjeci bodova i njihova varijabilnost; bodovi i bonus iz posljednjega odigranog kola; bodovi i BPS po susretu u sezoni; klizni bonus i <i>influence</i> ; ICT i BPS po 90 minuta; ocjena nastupa (klizno i sezonski)
Minutaža i uloga	9	klizni prosjeci minuta; minute posljednjega nastupa i broj dana od njega; odigrani susreti i prosječne minute u sezoni; udio susreta u početnoj postavi; uzastopni susreti u postavi
Sklonost visokim učincima	6	dvoznamenkasti učinci u posljednjih 5 i 10 kola; nedavni maksimumi bodova; učestalost bonusa; sezonski udio visokih učinaka
Napadački učinak	69	golovi, asistencije, <i>threat</i> i <i>creativity</i> ; stope po 90 minuta za xG, xA, npxG, xGChain, xGBuildup, udarce, ključna dodavanja, udarce u okvir, velike prilike i udarce iz kaznenog prostora; sezonske stope golova, asistencija, xG-a i xA-a te odstupanje ostvarenih od očekivanih statistika; ubačaji, duge lopte i dodiri lopte
Obrambeni učinak	30	sačuvane mreže, primljeni golovi i obrane; sezonske sačuvane mreže po susretu; oduzimanja lopte, presijecanja, izbijanja, blokade i vraćanja posjeda; dobiveni i zračni dueli (klizno, po 90 minuta i sezonski)
Vratarske akcije	4	obrane i obrane udaraca iz kaznenog prostora, uhvaćeni visoki ubačaji, boksanje lopte
Disciplina i rijetki događaji	13	klizni i sezonski žuti kartoni; crveni kartoni, autogolovi i obranjeni jedanaesterci (sezona i karijera); promašeni jedanaesterci; prekršaji, gubitci lopte, pogreške
Forma momčadi	21	postignuti i primljeni golovi, sačuvane mreže, rezultati, prepušteni fantasy bodovi; momčadski xG, xGA, npxGD, PPDA, očekivani bodovi i <i>deep completions</i> ; posjed lopte, udarci, velike prilike, ulasci u završnu trećinu
Forma protivnika	20	isto, za nadolazećega protivnika, bez <i>deep completions</i>
Povijest protiv protivnika	2	broj susreta protiv nadolazećega protivnika; karijerni bodovi po 90 minuta protiv njega
Kontekst susreta	3	domaćinstvo, redni broj kola, broj dana od prethodnoga susreta momčadi
Tržište	9	cijena i promjena cijene, udio vlasništva, ulazni i izlazni transferi, njihove stope i neto saldo
Indikatori pozicije	4	<i>one-hot</i> kodiranje pozicije

nastupom, uz 199 značajki učinka, i jedini je dio arhitekture u kojemu se dva pristupa razlikuju. Oba se stupnja uče zasebno za svaku poziciju (GK, DEF, MID i FWD), u skladu s razlikama utvrđenima u potpoglavlju 4.3.2.

5.3.3. Podjela podataka

Kronološka podjela podatkovnog skupa iz potpoglavlja 4.4. preuzima se neizmijenjena. Sezona za provjeru ima tri uloge: rano zaustavljanje (engl. *early stopping*), odabir hiperparametara te određivanje svih pragova i pravila odlučivanja (primjerice praga iznad kojega se predviđanje proglašava visokim učinkom). Ispitna sezona ni u jednu od tih svrha ne ulazi — njoj se pristupa jedanput, pri završnom vrednovanju.

5.3.4. Korišteni skup metrika

Cjelovit popis korištenih metrika donosi tablica 5.2. Dvije metodološke odluke pritom zaslužuju posebno obrazloženje.

Tablica 5.2. Skup metrika za vrednovanje modela.

Skupina	Metrike
Pogreška predviđanja	MAE, RMSE, medijan apsolutne pogreške, pristranost, R^2
Rangiranje	Spearmanov i Kendallov koeficijent korelacije ranga, NDCG@K (engl. <i>normalised discounted cumulative gain</i>) [21], preklapanje skupova najboljih K igrača
Odabir kapetana	točnost prvoga izbora, srednja recipročna vrijednost ranga (engl. <i>mean reciprocal rank</i> , MRR) koji je model dodijelio stvarno najboljem igraču kola, ostvareni udio bodova idealnoga kapetana i idealne početne postavke
Prepoznavanje događaja	preciznost, odziv i F1-mjera za visoke učinke (≥ 10 bodova) i niske učinke (engl. <i>blank</i> , ≤ 1 bod), uz pragove određene na sezoni za provjeru; odziv visokih učinaka među najboljih K predviđanja
Nastup (prvi stupanj)	AUC, binarna unakrsna entropija, Brierova ocjena (engl. <i>Brier score</i>), očekivana pogreška kalibracije (engl. <i>expected calibration error</i>)

Prva se odluka odnosi na skup redaka nad kojim se metrike računaju. Metrike rangiranja izračunate nad svim retcima te prikazane samostalno daju pogrešan uvid u kvalitetu modela, tj. njegova drugoga stupnja: velik dio njihove vrijednosti potječe od razdvajanja igrača koji nastupaju od onih koji ne nastupaju, a to je zadatak isključivo prvog stupnja modela. Model tako može dobro poredati igrače unutar kola već i samim time što pogodi tko će uopće izaći na teren. Iste metrike, ponovno izračunate isključivo na retcima s nastupom, omogućuju mjerenje sposobnosti drugoga stupnja — koliko dobro model razlikuje učinke među igračima koji jesu nastupili — pa se usporedno navode u cijelome poglavlju kao mjerodavan pokazatelj.

Druga se odluka tiče ponovljivosti rezultata. Vrijednosti metrika kojima se modeli iz ovoga poglavlja međusobno uspoređuju uglavnom se razlikuju vrlo malo, pa je prije donošenja svakoga zaključka potrebno razdvojiti stvarne razlike u kvaliteti od varijabilnosti uzrokovane slučajnošću postupka učenja. Zbog toga se cijeli postupak — učenje modela i njihovo vrednovanje — ponavlja tri puta, svaki put s drugim sjemenom (engl. *seed*) generatora slučajnih brojeva, uz nepromijenjene ostale uvjete. Ta tri pokretanja ne predstavljaju statistički test te ne govore ništa o tome kako bi se metrike ponašale na drugoj ispitnoj sezoni, već samo omogućuju dobivanje stabilnijih rezultata.

5.3.5. Referentni modeli

Na istom se skupu metrika vrednuju i tri heuristička referentna modela korištena u ovom radu: prosjek bodova po susretu od početka sezone, klizni prosjek posljednjih triju nastupa te prosjek po poziciji s proširujućim prozorom (engl. *expanding position average*). U cijelome se poglavlju modeli po svakoj metrici uspoređuju s najboljim od triju referentnih modela, a njihove glavne rezultate na ispitnoj sezoni donosi tablica 5.3.

Tablica 5.3. Heuristički referentni modeli na ispitnoj sezoni (2024./25.).

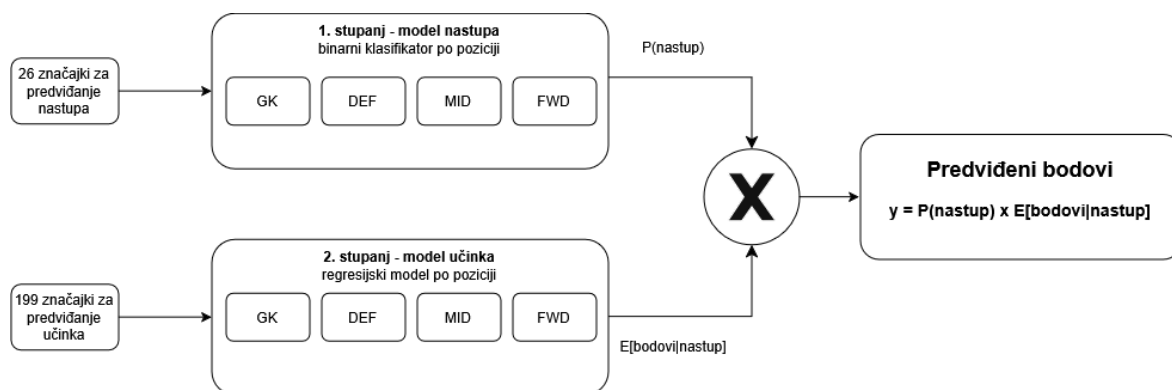
Referentni model	MAE	RMSE	Pristranost	Spearman	NDCG@10	Top-10	MRR kapet.
Bodovi po susretu u sezoni	2,141	2,630	+1,278	0,158	0,368	0,134	0,217
Klizni prosjek	2,001	2,655	+1,101	0,251	0,340	0,108	0,197
Prosijek pozicije	1,419	2,307	−0,047	0,105	0,114	0,037	0,015

5.4. Pristup 1: izravna regresija po pozicijama

5.4.1. Arhitektura i učenje

U prvome pristupu drugi stupanj čine četiri regresijska modela, po jedan za svaku poziciju, a svaki se uči isključivo na retcima s nastupom igrača te pozicije. Za funkciju gubitka odabrana je kvadratna pogreška (L2), a budući da bi samostalna uporaba tog gubitka rezultirala povlačenjem predviđanja prema najčešćim vrijednostima od 1–2 boda (potpoglavlje 4.3.1.), primjenjuje se ponderiranje primjera obrnuto proporcionalno učestalosti (engl. *inverse-frequency sample weighting*), po uzoru na metodu OpenFPL [13]. Ciljna se varijabla razvrstava u pet razreda s proizvoljno odabranim granicama {1, 4, 7, 10}, a svaki razred dobiva težinu obrnuto proporcionalnu broju svojih redaka, normaliziranu tako da prosječna težina po retku iznosi jedan. Time u funkciji gubitka svih pet razreda nosi jednaku ukupnu težinu, pa najrjeđi razred, koji čine dvoznamenkasti učinci, na učenje utječe jednako kao i mnogo brojniji razred najslabijih učinaka, čime pojedinačni redak iz njega nosi višestruko veću težinu od retka iz najbrojnijega razreda.

Konačni se modeli obaju stupnjeva uče na cijelome skupu za učenje, s hiperparametrima odabranima postupkom opisanim u nastavku i uz rano zaustavljanje na sezoni za provjeru, a cjelovita je arhitektura prikazana na slici 5.1.



Slika 5.1. Arhitektura prvoga pristupa.

5.4.2. Ugađanje hiperparametara

Hiperparametri se ugađaju zasebno za svaku poziciju i svaki stupanj, pri čemu se koristi alat Optuna [22] uz procjenitelj TPE (engl. *tree-structured Parzen estimator*). TPE nad dotad ispitanim konfiguracijama gradi dvije gustoće (jednu za konfiguracije koje su dale dobre rezultate, drugu za sve ostale), pa nove konfiguracije uzorkuje ondje gdje je njihov omjer najpovoljniji. Kriterij je pretrage prosječni gubitak dobiven trostrukom kronološkom unakrsnom provjerom s postupno rastućim skupom za učenje. Model se najprije uči na sezonama 2018./19.–2020./21., a vrednuje na sezoni 2021./22. Skup za učenje zatim se proširuje sezonom 2021./22., uz vrednovanje na sezoni 2022./23., te naposljetku i sezonom 2022./23., uz vrednovanje na sezoni 2023./24. Nasumična podjela podataka ovdje nije primjerena jer bi dovela do curenja informacija: retci iz iste sezone međusobno su snažno ovisni (potpoglavlje 4.4.).

U prvome se stupnju za svaku poziciju ispituje 30 konfiguracija, pri čemu je kriterij pretrage unakrsna entropija. U drugome se stupnju ispituje 40 konfiguracija po poziciji, osim za vezne igrače — najbrojniju i najheterogeniju skupinu — za koje se ispituje 80 konfiguracija. Kriterij je RMSE, a težine primjera ponovno se izračunavaju unutar svake podjele, koristeći isključivo njezin skup za učenje. Prostor pretrage uključuje parametre složenosti stabla: broj listova, najveću dubinu i najmanji broj primjera u listu, s rasponima prilagođenima pojedinoj poziciji. Obuhvaća i stopu učenja te L1- i L2-regularizaciju, čije se vrijednosti pretražuju na logaritamskoj skali.

5.5. Pristup 2: dekompozicija bodova po komponentama

5.5.1. Arhitektura i učenje

Kako je prethodno navedeno, ukupan broj bodova igrača nije proizvoljno određen, već proizlazi iz zbroja unaprijed definiranih bodovnih doprinosa. Svakom se relevantnom događaju — nastupu, голу, asistenciji, sačuvanoj mreži, primljenim golovima, obranama, bonus bodovima i kaznenim odbitcima — pridružuje fiksni broj bodova, pri čemu taj broj ovisi isključivo o poziciji igrača (tablica 2.1.). Ukupan učinak stoga predstavlja linearnu kombinaciju prebrojivih događaja, a zbog linearnosti očekivanja uvjetno očekivanje ukupnog broja bodova može se izraziti istom linearnom kombinacijom uvjetnih očekivanja njegovih pojedinačnih sastavnica. Slijedom toga, drugi stupanj ne zahtijeva jedinstveni regresijski model: svaka se komponenta može procijeniti zasebno, nakon čega se predviđeni ukupni učinak dobiva zbrajanjem tih procjena. Na toj je ideji utemeljen drugi pristup.

Takva je dekompozicija opravdana različitom prirodom pojedinih komponenti. Neke od njih imaju binarni oblik, primjerice podatak o tome je li momčad sačuvala mrežu ili je li igrač nastupio najmanje 60 minuta, dok druge predstavljaju broj ostvarenih događaja, kao što su golovi, asistencije, bonus bodovi i kartoni. Njihov se mehanizam nastanka također razlikuje: sačuvana mreža i broj primljenih golova određeni su učinkom momčadi, dok golovi i asistencije prvenstveno ovise o individualnom doprinosu igrača. Rastavljanjem ukupnog učinka svaka komponenta dobiva vlastitu ciljnu varijablu i funkciju gubitka prilagođenu svojoj vrsti, umjesto da se sve promatraju kao jedan skalarni ishod.

Na razini svakog retka, odnosno svakog para igrač–susret, definiraju se komponentne ciljne varijable: indikator nastupa u trajanju od najmanje 60 minuta, broj golova, asistencija, bonus bodova i žutih kartona, indikator sačuvane mreže te dvije varijable koje uključuju zaokruživanje nadolje i koje su već prije učenja izražene na bodovnoj ljestvici — bodovi za primljene golove $\lfloor \text{primljeni}/2 \rfloor$ i bodovi za obrane $\lfloor \text{obrane}/3 \rfloor$. Budući da su posljednje dvije veličine unaprijed pretvorene u bodove, njihovi modeli izravno predviđaju pripadajuće bodovne vrijednosti. Predviđanje samog broja primljenih golova ili obrana ne bi omogućilo da se dobivena procjena naknadno samo podijeli i zaokruži, jer očekivanje zaokružene veličine nije jednako zaokruženom očekivanju: takav bi postu-

pak zato zahtijevao dodatnu korekciju pri sastavljanju ukupnog predviđanja. Korištene bodovne vrijednosti odgovaraju pravilima koja su bila na snazi tijekom cijeloga promatranog razdoblja: gol donosi 6/6/5/4 boda za GK/DEF/MID/FWD, asistencija 3 boda, sačuvana mreža 4/4/1/0 bodova, primljeni golovi oduzimaju −1 bod za svaka dva gola kod GK i DEF, a obrane donose +1 bod za svake tri obrane kod GK. Osim toga, obranjeni jedanaesterac vratarima donosi +5 bodova, bonus iznosi od 1 do 3 boda, nastup nosi 1 ili 2 boda, uz standardne odbitke za kartone, autogolove i promašene jedanaesterce¹. Ispravnost ovako definiranih ciljnih varijabli provjerena je rekonstrukcijom ukupnog broja bodova za svaki redak iz njihovih vrijednosti i navedenih pravila bodovanja: jednakost vrijedi točno za svih 177 675 redaka, i to u svakoj sezoni. Time se istodobno potvrđuje da se sustav bodovanja tijekom promatranog razdoblja nije mijenjao.

Za svaki par komponente i pozicije uči se zaseban model — *glava* (engl. *head*) — ali samo ako ta komponenta za promatranu poziciju donosi bodove te se pojavljuje dovoljno često da ju je moguće pouzdano učiti iz dostupnih podataka (tablica 5.4.). Na taj se način dobivaju 23 komponentne glave koje, zajedno s četirima klasifikatorima nastupa, čine ukupno 27 modela. Događaji koji se ne modeliraju izravno (crveni kartoni, autogolovi, promašeni i obranjeni jedanaesterci te, isključivo za vratare, golovi, asistencije i žuti kartoni) pri sastavljanju predviđanja predstavljeni su konstantom specifičnom za poziciju. Njezina je vrijednost jednaka prosječnom bodovnom doprinosu tih događaja među retcima s nastupom u sezonama korištenima za učenje.

Tablica 5.4. Komponentne glave drugoga pristupa. Stope navedene u posljednjem stupcu prosjeci su na retcima s nastupom u sezonama za učenje.

Komponenta	Vrsta	Pozicije	Razlog isključenja
Nastup ≥ 60 min	vjerojatnost	GK, DEF, MID, FWD	—
Golovi	broj	DEF, MID, FWD	GK: stopa 0,0003 po susretu, ulazi kao konstanta
Asistencije	broj	DEF, MID, FWD	GK: ulazi kao konstanta
Sačuvana mreža	vjerojatnost	GK, DEF, MID	FWD: ne nosi bodove
Bodovi za primljene golove	broj	GK, DEF	MID, FWD: ne kažnjavaju se
Bodovi za obrane	broj	GK	pravilo se odnosi samo na vratare
Bonus	broj	GK, DEF, MID, FWD	—
Žuti kartoni	broj	DEF, MID, FWD	GK: stopa 0,044 po susretu, ulazi kao konstanta

Svaka se glava uči samo na retcima koji odgovaraju nastupima igrača promatrane

¹Ista su pravila prikazana u tablici 2.1. u poglavlju 2. Od sezone 2025./26. FPL ih mijenja u dvama detaljima: bodovi za vratarev gol povećani su na 10 te su uvedeni bodovi za obrambeni doprinos.

pozicije, pri čemu koristi istih 199 značajki učinka dostupnih i drugome stupnju prvog pristupa. Odabir funkcije gubitka prati vrstu ciljne varijable: glave koje procjenjuju vjerojatnosti koriste binarnu unakrsnu entropiju, dok se za glave koje predviđaju brojeve događaja primjenjuje Poissonova devijanca. U potonjem slučaju model, primjenom kanonske logaritamske funkcije veze (engl. *canonical logarithmic link function*), procjenjuje logaritam očekivane stope — sama je procijenjena stopa zato nužno nenegativna, što optimizacija kvadratne pogreške ne može osigurati.

Konačni se modeli obaju stupnjeva uče na cijelome skupu za učenje, uz unaprijed zadane hiperparametre opisane u nastavku i rano zaustavljanje na sezoni za provjeru.

5.5.2. Hiperparametri

Budući da se hiperparametri ne ugađaju ni u jednome stupnju, sve glave, kao i klasifikator nastupa prvoga stupnja, koriste jednu od dviju unaprijed zadanih konfiguracija, odabranih prema vrsti ciljne varijable. Prvu koriste vjerojatnosne glave, a kako je i ciljna varijabla prvoga stupnja binarna, njome je zadan i model nastupa: 31 list, dubina 8, stopa učenja 0,05, najmanje 50 primjera u listu te po 0,8 uzorkovanja redaka i stupaca po stablu. Drugu koriste glave za brojeve događaja; ona je nešto većega kapaciteta i uz blagu regularizaciju: 63 lista, dubina 8, stopa učenja 0,05, najmanje 50 primjera u listu, $L_1 = 0,1$ i $L_2 = 1$. Jedini prilagodljivi element ostaje broj stabala: on se za svaki model zasebno određuje ranim zaustavljanjem na sezoni za provjeru, pa se kapacitet svake glave prilagođava učestalosti događaja koji modelira.

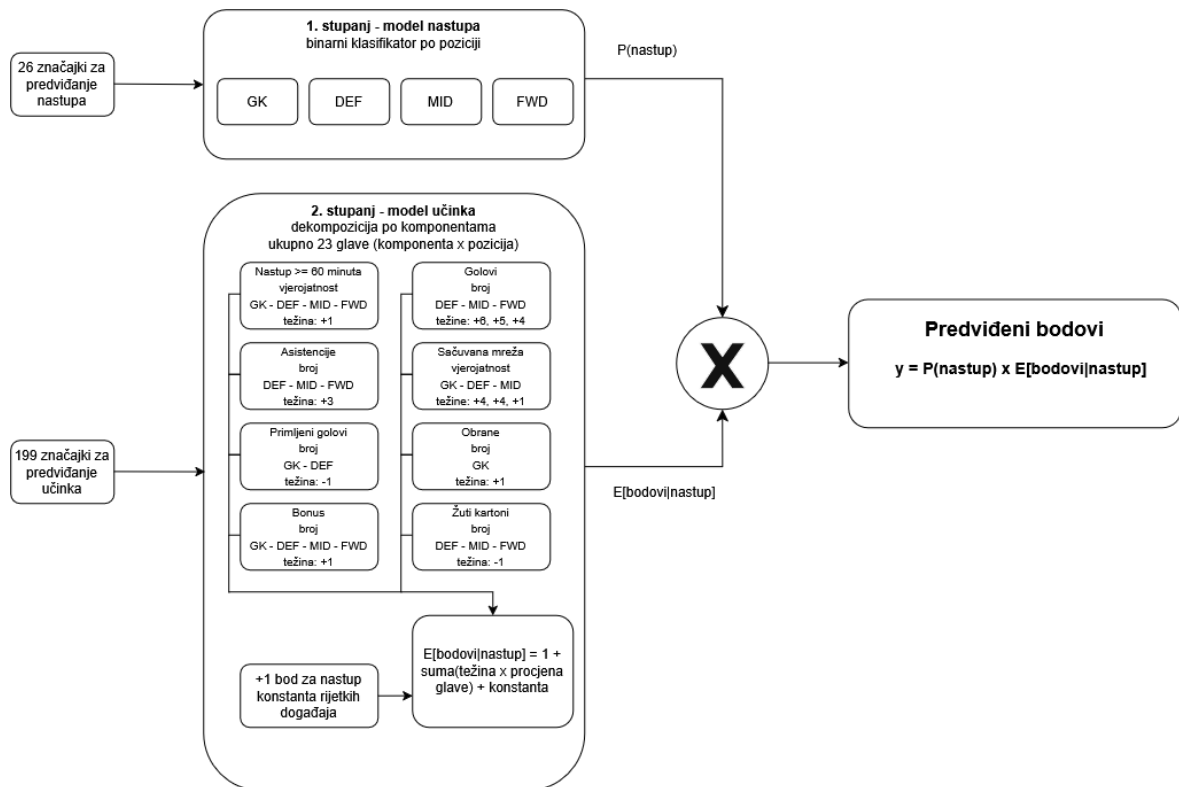
5.5.3. Sastavljanje predviđanja

Sastavljeno uvjetno očekivanje može se zapisati na sljedeći način:

$$\begin{aligned} \hat{E} [y_{p,g} \mid m_{p,g} > 0] = & 1 + \hat{P}(\geq 60 \text{ min}) + v_{\text{poz}}^{\text{gol}} \hat{E}[\text{golovi}] + 3 \hat{E}[\text{asistencije}] \\ & + v_{\text{poz}}^{\text{SM}} \hat{P}(\text{sačuvana mreža}) - \hat{E}[\text{bodovi za primljene}] + \hat{E}[\text{bodovi za obrane}] \\ & + \hat{E}[\text{bonus}] - \hat{E}[\text{žuti kartoni}] + c_{\text{poz}}. \end{aligned} \quad (5.2)$$

Indeksi igrača i kola na desnoj su strani radi čitljivosti izostavljeni, dok su sva očekivanja uvjetovana istim događajem, odnosno nastupom promatranog igrača. Prva dva člana

proizlaze iz pravila bodovanja prema kojem nastup donosi jedan bod, a nastup u trajanju od najmanje 60 minuta dva boda, pa očekivani doprinos nastupa iznosi $1 + \hat{P}(\geq 60 \text{ min})$. Koeficijenti $v_{\text{poz}}^{\text{gol}}$ i $v_{\text{poz}}^{\text{SM}}$ označavaju bodovnu vrijednost gola, odnosno sačuvane mreže, za odgovarajuću poziciju, dok se članovi koji se odnose na primljene golove i obrane uključuju samo za pozicije na koje se ta pravila odnose. Veličina c_{poz} predstavlja konstantni doprinos rijetkih događaja. Konačno predviđanje ponovno slijedi izraz (5.1). Cjelovita je arhitektura prikazana na slici 5.2.



Slika 5.2. Arhitektura drugoga pristupa.

5.6. Rezultati

5.6.1. Ukupna usporedba s referentnim modelima

Tablica 5.5. prikazuje vrijednosti metrika obaju pristupa na ispitnoj sezoni te, za svaku metriku, rezultat najboljega od triju heurističkih referentnih modela. Rezultat svakoga modela predstavlja prosjek triju pokretanja opisanih u potpoglavlju 5.3.4., pri čemu se pokretanja razlikuju isključivo prema sjemenu generatora slučajnih brojeva. Budući da sezonski prosjek može prikriti pojedine razlike, prednost se smatra potvrđenom samo ako ostaje prisutna i nakon detaljnije raščlambe: na svakoj od četiriju pozicija te u sva-

kom od 38 kola promatranom zasebno, a ne samo u ukupnom rezultatu.

Tablica 5.5. Glavni rezultati na ispitnoj sezoni (2024./25.).

Metrika	Najbolja heur.	A1	A2
<i>Svi retci (27 283 retka)</i>			
MAE	1,419	1,043	0,967
RMSE	2,307	1,950	1,882
Pristranost	−0,047	+0,132	+0,006
R^2	0,004	0,288	0,338
Spearman	0,251	0,709	0,729
Kendall	0,198	0,558	0,579
NDCG@10	0,368	0,441	0,453
NDCG@20	0,362	0,445	0,467
Preklapanje najboljih 10	0,134	0,165	0,175
Točnost odabira kapetana	0,158	0,175	0,149
MRR kapetana	0,217	0,255	0,236
Bodovi kapetana u sezoni (od 679)	305	331	332
Ostvareni udio bodova idealne postave	0,362	0,448	0,441
Odziv visokih učinaka među 30 najboljih	0,242	0,330	0,333
Odziv visokih učinaka među 50 najboljih	0,326	0,441	0,441
F1-mjera za niske učinke	0,557	0,832	0,829
Kola u kojima nadmašuje sve tri heuristike	—	38/38	38/38
<i>Samo retci s nastupom (11 566 redaka)</i>			
MAE	1,869	1,983	1,827
RMSE	2,876	2,844	2,771
R^2	0,013	0,034	0,084
Spearman	0,306	0,350	0,395
NDCG@10	0,422	0,458	0,464

Prema mjerama pogreške i korelaciji rangova oba pristupa ostvaruju nedvojbenu prednost. Bodovni učinak igrača u pojedinom kolu predviđaju uz prosječno odstupanje od približno jednoga boda, dok najbolja heuristika postiže MAE od 1,419 boda. Istodobno objašnjavaju približno trećinu varijance ciljne varijable, za razliku od heuristika koje ne objašnjavaju gotovo ništa (R^2 iznosi 0,288 i 0,338 naspram 0,004), a znatno su uspješniji i u rangiranju igrača, što potvrđuju Spearmanovi koeficijenti od 0,709 i 0,729 u odnosu na 0,251. Takvi rezultati nisu posljedica samog usrednjavanja: svaki je pristup bolji od svih triju heurističkih modela prema MAE-u, RMSE-u te Spearmanovu i Kendallovu koeficijentu u svih 38 kola i na svakoj poziciji, pri čemu prednost u MAE-u ni u jednom slučaju nije manja od 0,29 boda. Jednako je izražena razlika pri prepoznavanju niskih učinaka, gdje F1-mjera iznosi 0,832 i 0,829 naspram 0,557.

Nijedna metrika koja se odnosi na sam vrh poretka ne ispunjava navedeno mjerilo. Sezonski rezultati za NDCG@10 pogoduju modelima, s vrijednostima 0,441 i 0,453 nas-

pram 0,368, no pri analizi svakog kola zasebno svaki model nadmašuje sve tri heuristike u tek približno polovici slučajeva — u 20, odnosno 21 od ukupno 38 kola. Sezonska je prednost još manja pri usporedbi preklapanja predviđene i ostvarene najbolje desetorke kola, gdje vrijednosti iznose 0,165 i 0,175 naspram 0,134, dok broj kola u kojima modeli vode pada ispod polovice, na 11 odnosno 8 od 38. To ipak ne predstavlja poraz: u svim ostalim kolima rezultat je izjednačen ili podijeljen, a prvi je pristup slabiji od svih triju heuristika samo u jednom kolu, dok se to drugom pristupu ne događa ni jednom. Ni odabir kapetana ne razlikuje pristupe, jer strategija sezonskoga prosjeka najboljega igrača kola pogađa jednako često kao i bilo koji od dvaju modela, s vrijednostima 0,158 naspram 0,175 i 0,149, a modeli su bolji od svih triju heuristika u samo jednom, odnosno trima kolima. Metrike recipročne vrijednosti ranga i kapetanovih bodova pokazuju isti obrazac te su uključene isključivo radi potpunosti.

Kada se navedene vrijednosti prevedu u stvarne odluke koje menadžer donosi, mijenja se njihov praktični poredak: metrika s najvećom brojčanom prednošću nije i ona koja najviše utječe na ishod, dok je prednost modela ondje gdje je odluka najvažnija gotovo zanemariva. Najveću praktičnu korist pritom donosi rezultat koji kao naslovni podatak djeluje najmanje impresivno — prepoznavanje učinaka od najviše jednoga boda, jer je zadržavanje igrača koji uopće ne nastupi u momčadi česta i skupa pogreška. Sljedeći je po važnosti odabir početne postave, pri čemu jedanaest najbolje ocijenjenih igrača osvaja 44–45 % bodova koje bi ostvarila savršena postava, naspram 36 % kod najbolje heuristike. Nasuprot tomu, metrike koje opisuju sam vrh poretka najizravnije odgovaraju menadžerovim odlukama, budući da on u sastav bira jedanaest igrača, a jednoga određuje za kapetana, no upravo je na njima doprinos modela najmanji. Deset najbolje plasiranih igrača pojedinoga kola u prosjeku osvaja 12,7 boda, pri čemu je 94 % njihovih učinaka visoko, što predstavlja najslabije predvidljiv događaj ove ciljne varijable (potpoglavlje 4.3.3.). Od tih deset igrača modeli u prosjeku prepoznaju 1,7, a heuristike 1,3, pa se njihova razlika svodi na 0,4 igrača po kolu unutar skupine u kojoj i uspješniji pristup promašuje više od četiri petine slučajeva. Najskuplje je pritom nepostojanje razlike pri izboru kapetana, jer kapetanstvo predstavlja odluku s najvećom polugom u pojedinom kolu: oba pristupa ostvaruju 331 i 332 boda, a strategija sezonskoga prosjeka 305, od mogućih 679 bodova idealnoga kapetana, dok je ta razlika manja od raspona koji proizlazi već iz promjene sjemena. Naposljetku, mjerama pogreške treba pridati manju težinu nego što sugerira

njihova raširenost, jer tipična pogreška od približno dva boda po susretu u retcima s nastupom i dalje premašuje razliku na kojoj počiva većina transfernih odluka. Stoga se argumentacija ovoga poglavlja temelji na rangiranju, a ne na pogrešci.

5.6.2. Usporedba dvaju pristupa

Prema mjerama pogreške i korelaciji rangova dekompozicija dosljedno ostvaruje malu, ali jasnu prednost. Bodovni učinak u pojedinom kolu predviđa uz odstupanje od 0,967 boda, naspram 1,043 boda, dok RMSE iznosi 1,882 u odnosu na 1,950. Igrače također rangira neznatno uspješnije, sa Spearmanovim koeficijentom od 0,729 naspram 0,709 te Kendallovim koeficijentom od 0,579 naspram 0,558. Raščlamba po kolima i pozicijama potvrđuje taj obrazac gotovo bez odstupanja: dekompozicija ostvaruje manju pogrešku u svih 38 kola, pri čemu njezina najmanja prednost u pojedinom kolu iznosi 0,03 boda prema MAE-u, kao i na sve četiri pozicije. Kendallov koeficijent veći je u svih 38 kola, dok je Spearmanov veći u 37 kola, a u jedinom preostalom kolu zaostatak iznosi samo 0,002, uz istu prednost na svim četirima pozicijama. Pristranost pristupe razlikuje još jasnije, jer dekompozicija postiže vrijednost +0,006 naspram +0,132 te je bliža nuli na svakoj poziciji. Razlika je pritom sustavna, a ne slučajna: ponderiranje obrnuto proporcionalno učestalosti iz potpoglavlja 5.4. već svojom konstrukcijom pomiče model učinka izravne regresije prema višim vrijednostima. Zato taj model na retcima s nastupom precjenjuje učinak za 0,38 boda, dok sastavljeno predviđanje, koje ne primjenjuje ponderiranje, ostvaruje precjenjivanje od 0,09 boda.

Promatranje vrha poretka ne omogućuje jasno razlikovanje pristupa. Iako sezonske vrijednosti NDCG@10 i preklapanja najboljih deset igrača pogoduju dekompoziciji, s rezultatima 0,453 naspram 0,441, odnosno 0,175 naspram 0,165, i to na svakoj poziciji, analiza pojedinačnih kola ne pokazuje jednako dosljedan obrazac. Dekompozicija prema NDCG@10 vodi u 25 od 38 kola, dok prema preklapanju najboljih deset vodi u njih 13, uz 17 neriješenih kola i osam kola u kojima prednost ima izravna regresija. Smjer sezonskih rezultata stoga jest stabilan, ali nije potvrđen raščlambom po kolima, zbog čega se nijedna od tih metrika ne može smatrati prednošću. Pokazatelji izbora kapetana pritom nisu međusobno usklađeni: točnost i recipročna vrijednost ranga idu u prilog izravnoj regresiji, s vrijednostima 0,175 naspram 0,149, odnosno 0,255 naspram 0,236, no kapetani obaju pristupa tijekom sezone osvajaju gotovo jednak broj bodova, 331 i 332 od

679 bodova koje bi ostvario idealan kapetan. Prednost pri izboru koja se ne odražava na osvojenim bodovima ne predstavlja stvarnu prednost, pa ovaj skup metrika ne podupire zaključak ni u korist jednoga pristupa. Konačno, prepoznavanje niskih učinaka gotovo je jednako, 0,832 naspram 0,829, što je posljedica zajedničkoga modela nastupa, a ne razlike između dviju arhitektura.

Dva dodatna nalaza pokazuju da se razlika pojavljuje upravo ondje gdje se arhitekture stvarno razlikuju. Prednost dekompozicije povećava se na retcima s nastupom, odnosno u populaciji za koju je odgovoran isključivo njezin drugi stupanj, dok se na repu ciljne varijable obrazac obrće: ponderiranje izravne regresije smanjuje pogrešku za učinke iznad pet bodova, ali pritom ne omogućuje prepoznavanje nijednoga dodatnog visokog učinka (potpoglavlje 5.6.4.). Kada se ta prednost vrednuje zajedno s njezinom cijenom, zaključak je jasan — jedina prednost izravne regresije ne donosi nijedan dodatno prepoznat visoki učinak, dok dekompozicija vodi prema svim ostalim metrikama koje razlikuju dva pristupa. Zbog toga se dekompozicija usvaja kao konačni prediktivni model ovoga rada, a izravna regresija ostaje referentna izvedba s kojom je uspoređena.

5.6.3. Model nastupa

Oba pristupa koriste zajednički model nastupa iz potpoglavlja 5.3.2., pa velik dio rezultata koje tablica 5.5. prikazuje na svim retcima proizlazi upravo iz toga modela. Njegove su vjerojatnosti dobro kalibrirane: Brierova ocjena iznosi 0,043 za vratare, 0,096 za braniče, 0,090 za vezne igrače i 0,073 za napadače, dok očekivana pogreška kalibracije za iste pozicije iznosi 0,008, 0,017, 0,019 i 0,032. Najjednostavniji su slučaj vratari, što je očekivano zbog gotovo binarnog obrasca nastupanja opisanog u potpoglavlju 4.3.2., dok su napadači najslabije kalibrirani. Model je za njih sustavno manje siguran nego što bi trebao biti u čitavom rasponu vjerojatnosti, jer su procjene previše povučene prema 0,5, a to je najizraženije između 0,3 i 0,8: napadač s predviđenom vjerojatnošću nastupa od 0,66 doista nastupa u 76 % slučajeva, dok onaj s vjerojatnošću od 0,35 nastupa u 28 % slučajeva. Neugodena inačica istoga stupnja, koju koristi dekompozicija, postiže Brierove ocjene od 0,044, 0,096, 0,090 i 0,073, što se ni na jednoj poziciji ne razlikuje za više od 0,0003 od odgovarajućih ugođenih vrijednosti. Stoga su te dvije inačice na ispitnoj sezoni praktički jednak model, iako se njihove pogreške kalibracije nešto više razlikuju kod vratara, gdje neugodena inačica ostvaruje 0,012 naspram 0,008, premda su obje vrijednosti

i dalje male.

5.6.4. Visoki učinci i rep predviđanja

Raščlamba pogreške prema ostvarenom broju bodova pokazuje ograničenje koje oba pristupa dijele sa svim modelima očekivane vrijednosti (tablica 5.6.). Na glavnom dijelu distribucije oba su pristupa precizna, dok na njezinu repu podjednako podcjenjuju visoke učinke, gotovo za njihov puni iznos. Kada ostvareni učinak premaši pet bodova, prosječno predviđanje bilo kojega pristupa tek se neznatno povećava, s približno tri boda na najviše 3,9 boda, iako prosječan stvarni učinak raste sve do 13,3 boda. Posebno je informativna točka u kojoj se dvije krivulje sijeku: dekompozicija je uspješnija u svakom razredu do četiri boda, a slabija u svakom razredu iznad te granice. Njezina ukupna prednost prema MAE-u zato potječe iz velikoga udjela niskih učinaka, koji čine gotovo čitav skup redaka.

Tablica 5.6. Srednja apsolutna pogreška i prosječno predviđanje po ostvarenom broju bodova, na svim retcima ispitne sezone i na razini pojedinačnoga susreta.

Ostvareni bodovi	Retci	MAE		Prosječno predviđanje	
		A1	A2	A1	A2
0	16 405	0,42	0,40	0,42	0,40
1	4 513	1,25	1,01	1,97	1,80
2–4	3 919	1,29	0,93	2,86	2,55
5–6	1 077	2,80	2,98	2,97	2,71
7–10	950	4,87	5,31	3,30	2,85
više od 10	304	9,36	9,93	3,93	3,36

I raspodjele predviđanja izravno potvrđuju isti obrazac. U stvarnim se učincima 95. percentil kreće oko šest bodova za vratare, braniče i vezne igrače, odnosno osam bodova za napadače, dok sezonski maksimumi dosežu od 15 do 25 bodova. Nasuprot tomu, izravna regresija daje vrijednosti 95. percentila od 3,7 do 5,2 boda, ovisno o poziciji, pri čemu nijedno predviđanje ne premašuje 10,7 boda, dok dekompozicijski pristup procjenjuje 95. percentil između 3,2 i 4,0 boda te ne prelazi 8,3 boda. Objedinjeno je predviđanje stoga komprimiranije od predviđanja izravne regresije. Taj se rezultat ne može pripisati nedostatnoj prilagodbi modela, nego proizlazi iz prirode predviđane veličine: svaka glava procjenjuje prosječnu stopu, dok iznimno visok učinak nastaje tek istodobnim ostvarenjem više rijetkih događaja. Najveće objedinjeno predviđanje tijekom cijele

ispitne sezone, 8,25 boda, temelji se na uvjetnom očekivanju od 8,33 boda, od čega je čak 4,30 boda povezano s golovima, što odgovara 0,86 očekivanih golova, dok je za stvarno ostvarenih 15 bodova bilo potrebno da se u istome susretu poklopi nekoliko takvih rijetkih ishoda. Očekivana vrijednost ne može dosegnuti rezultat koji postoji samo kao posljedica njihove konjunkcije. To ograničenje dijele obje arhitekture, a razlikuju se po tome što izravna regresija namjerno proširuje rep raspodjele ponderiranjem obrnuto proporcionalnim učestalosti iz potpoglavlja 5.4., koje dekompozicijski pristup ne primjenjuje.

Ta dodatna kompresija ne narušava sposobnost rangiranja predviđanja. Među 50 najviših predviđanja po kolu odziv visokih učinaka iznosi 0,441 za oba pristupa, u usporedbi s 0,326 za najbolju heuristiku, dok se među 30 najviših predviđanja vrijednosti kreću od 0,330 za izravnu regresiju i 0,333 za dekompoziciju naspram 0,242 za heuristiku. Drugim riječima, kompresija sužava raspon dodijeljenih vrijednosti, ali uglavnom zadržava njihov međusobni poredak, zbog čega obje arhitekture pronalaze jednak udio visokih učinaka iako im pridružuju različite brojčane procjene. Stoga je za oba pristupa prepoznavanje visokih učinaka ograničeno istom niskom gornjom granicom, koju sama promjena arhitekture ne uspijeva povisiti. Kada se visokim učinkom smatra svako predviđanje iznad praga utvrđenoga zasebno za svaku poziciju na sezoni za provjeru, izravna regresija postiže preciznost od 0,072 uz odziv od 0,404, a dekompozicija preciznost od 0,082 uz odziv od 0,340. Pripadajuće F1-mjere, 0,122 i 0,133, u apsolutnom su smislu niske neovisno o načinu usporedbe. Među nevratarskim pozicijama najzahtjevniji slučaj i dalje predstavljaju braniči, s vrijednostima od 0,060 i 0,091, što je u skladu sa strukturom konjunkcije opisanom u potpoglavlju 4.3.3.

5.6.5. Važnost značajki

U prvom se stupnju za svaku poziciju kao daleko najvažnija izdvaja jedna značajka: broj dana od posljednjega igračeva nastupa. Što je vrijednost te značajke veća, to je izlaz prvog stupnja bliži nuli (pa je samim time i ukupan broj predviđenih bodova bliži nuli zbog umnoška), a što je ta vrijednost manja, to je izlaz bliži jedinici, pa u tim slučajevima drugi stupanj može nesmetano obavljati svoju funkciju. Njezin je dobitak dva do tri i pol puta veći od dobitka sljedeće značajke, a nakon nje slijede nedavni udio utakmica započetih u početnoj postavi te niz uzastopnih nastupa od prve minute. Posljednja od tih značajki kod veznih igrača i napadača zauzima tek šesto, odnosno sedmo mjesto, pri čemu je pre-

mašuju ulazni transferi. Procjena raspoloživosti stoga se ponajprije oslanja na svježinu i pravilnost igračevih nastupa, dok transferno tržište nadopunjuje tu procjenu informacijom koja proizlazi iz zajedničkoga vrednovanja menadžera. U izravnoj regresiji važnost značajki modela učinka pritom se mijenja među pozicijama na način koji je moguće smisleno protumačiti. Naime, za vratare prevladavaju obilježja momčadi i protivnika povezana s karakterom susreta, kod braniča najveću ulogu imaju cijena, udio vlasništva i kontekst protivnika, a kod veznih igrača i napadača to su cijena, klizni *threat*, kretanje transfera i udio vlasništva. Dva se zaključka nameću iz takva rasporeda: blok tržišnih značajki, koji čine cijena, udio vlasništva i transferi, na svim pozicijama osim vratarske djeluje kao snažna apriorna procjena nastala kolektivnom prosudbom menadžera, dok je za vratare najslabiji među svim blokovima, pri čemu je cijena tek na 89. mjestu od ukupno 199 značajki. Istodobno, Sofascoreove varijable koje opisuju momčad i protivnika osiguravaju glavninu obrambenog signala koji FPL podatci sami ne sadržavaju.

Analiza važnosti značajki po pojedinim glavama dekompozicijskog modela pokazuje da se one uglavnom oslanjaju na očekivane skupine varijabli. Tako se na vrhu glava za sačuvanu mrežu nalazi protivnikov posjed lopte, dok glavama za primljene golove najviše doprinosi protivnikov broj dodavanja. U svim trima glavama za kartone igračeva vlastita stopa prekršaja zauzima drugo mjesto, a glavu koja predviđa nastup od barem 60 minuta kod nevratarskih igrača najviše oblikuju dodiri s loptom, *influence* i nedavno osvojeni bodovi. U napadačkim glavama važnost tržišnih varijabli postupno raste: kod braniča prednost imaju stope očekivanih golova te *threat* i *creativity*, zatim se udio tržišnih varijabli povećava kod veznih igrača, a kod napadača cijena, udio vlasništva i transferi predvode obje glave. Iz toga se može zaključiti da tržište učinak napadača procjenjuje preciznije od bilo koje pojedinačne klizne statistike. Mogućnost takve provjere svake komponente zasebno sama je po sebi prednost interpretabilnosti koju jedinstveni regresijski model ne može pružiti.

6. Optimizacija upravljanja momčadi i povijesna simulacija

Modeli predstavljeni u poglavlju 5. svakomu igraču u pojedinome susretu pridružuju numeričku vrijednost. Pretvaranje tih vrijednosti u konkretno upravljanje FPL momčadi — uključujući transfere, sastavljanje početne postave i izbor kapetana — zaseban je kombinatorni problem ograničene optimizacije, koji rješava mehanizam razvijen u ovome poglavlju. Taj mehanizam ima dvojaku namjenu: u aplikaciji opisanoj u poglavlju 7. generira tjedne preporuke inteligentnog asistenta, dok se u ovome radu koristi za vrednovanje cjelokupnog sustava. U tu se svrhu povijesna sezona ponovno simulira prema službenim pravilima igre, pri čemu se isti postupak odlučivanja primjenjuje na različite izvore predviđanja. Potpoglavlje 6.1. definira problem odlučivanja, potpoglavlje 6.2. opisuje optimizacijske algoritme i obradu kola nakon roka za predaju, potpoglavlje 6.3. objašnjava simulacijski protokol i uspoređivane strategije, a potpoglavlje 6.4. iznosi dobivene rezultate.

6.1. Problem odlučivanja

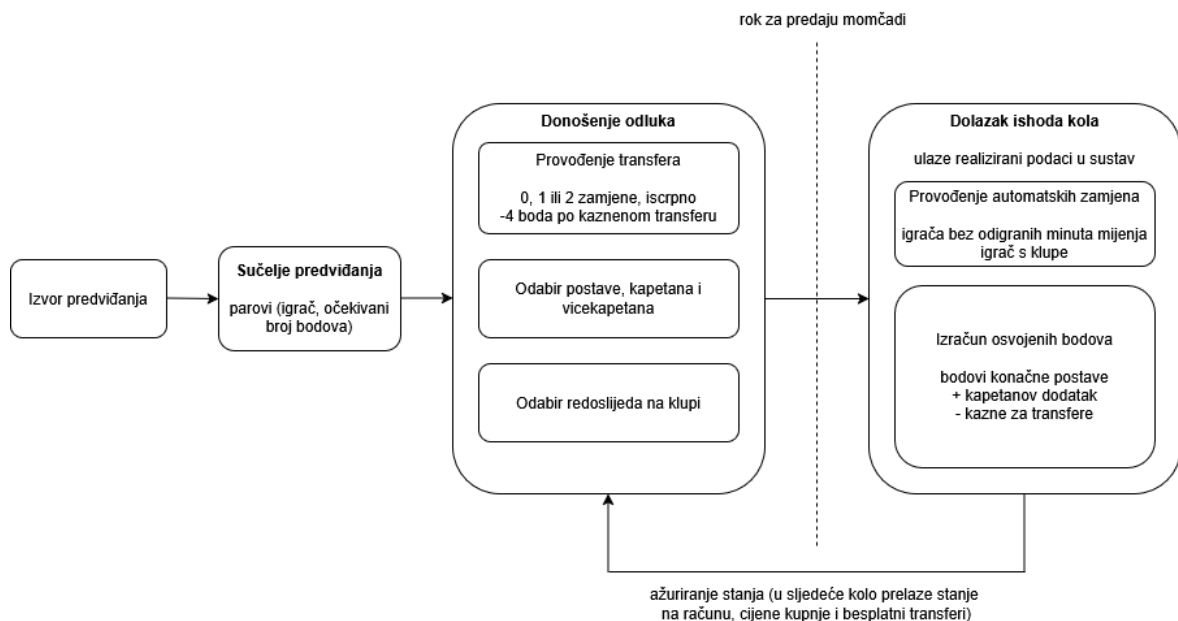
Prije svakoga roka za predaju konačne momčadi menadžer mora donijeti četiri odluke: koje transfere provesti (koga prodati, a koga kupiti), koja jedanaestorica od petnaest igrača u momčadi ulazi u početnu postavu, koga imenovati kapetanom, a koga vicekapetanom te kojim redom četvorica igrača s klupe ulaze u sastav ne nastupi li koji od igrača iz početne postave. Sve se četiri donose prema pravilima opisanima u poglavlju 2. Peta odluka igre, upotreba čipa, u ovome se vrednovanju ne razmatra.

Optimizacijski mehanizam izračunata predviđanja prima putem jedinstvenoga sučelja, i to u obliku preslikavanja koje svakomu igraču pridružuje očekivani broj bodova u susretima kola za koje se odluka donosi. Predviđanja po susretima zbrajaju se po igraču

i kolu, pa se igraču čija momčad u tom kolu igra dva susreta priznaju oba, dok se igraču bez zakazanoga susreta dodjeljuje vrijednost nula. Takav pristup odvaja postupak odlučivanja od izvora predviđanja, zbog čega isti mehanizam mogu pokretati naučeni model, jednostavna heuristika, pokazatelj kolektivne prosudbe menadžera ili čak stvarno ostvareni broj bodova. Usporedba strategija u potpoglavlju 6.4. time ostaje kontrolirana, jer razlike u konačnom sezonskom rezultatu proizlaze isključivo iz kvalitete upotrijebljenih predviđanja.

6.2. Pregled mehanizma

Slika 6.1. daje pregled mehanizma u cjelini za jedno kolo, a ostatak potpoglavlja opisuje njegove dijelove pojedinačno.



Slika 6.1. Pregled optimizacijskog mehanizma za jedno kolo.

Na početku sezone momčad, početna postava i kapetan određuju se istodobno rješavanjem cjelobrojnoga linearnog programa. Binarne varijable x_i , y_i i c_i redom označavaju pripada li igrač i momčadi, nalazi li se u početnoj postavi i je li odabran za kapetana. Cijena igrača v_i iskazana je u jedinicama od £0,1 milijuna, zbog čega početni proračun od £100 milijuna odgovara ograničenju od 1 000, dok $\hat{p}_i$ predstavlja predviđeni broj bodova. Model dodatno uključuje skupove igrača po pozicijama P_k , pripadne kvote

$q_k = (2, 5, 5, 3)$ te skupove igrača pojedinih klubova T_i , a njegova je formulacija sljedeća:

$$\max_{x,y,c} \sum_i \hat{p}_i (y_i + c_i) \quad (6.1)$$

$$\begin{aligned} \text{uz ograničenja} \quad & \sum_i x_i = 15, \quad \sum_i y_i = 11, \quad \sum_i c_i = 1, \quad y_i \leq x_i, \quad c_i \leq y_i, \\ & \sum_{i \in P_k} x_i = q_k \quad \forall k, \quad \sum_{i \in T_i} x_i \leq 3 \quad \forall t, \quad \sum_i v_i x_i \leq 1\,000, \\ & \sum_{i \in P_{GK}} y_i = 1, \quad \sum_{i \in P_{DEF}} y_i \geq 3, \quad \sum_{i \in P_{FWD}} y_i \geq 1, \quad x_i, y_i, c_i \in \{0, 1\}. \end{aligned}$$

Funkcija cilja svakomu članu početne postave pridružuje predviđene bodove jednom, pri čemu se kapetanovi bodovi računaju dvaput. U promatranoj sezoni broj kandidata po kolu kretao se između 616 i 784, a rješavač CBC optimalno rješenje je pronalazio za približno pola sekunde.

U slučaju kada je nova sezona već uzela maha te je početna momčad već odavno odabrana, menadžeru od početnih već spomenutih obveza ostaju samo odabir početne postave od igrača koje trenutačno posjeduje te odabir kapetana, a za to se koristi isti, već postojeći program, pri čemu se primjenjuje na samo petnaest igrača koje menadžer već posjeduje. Jednako kao i prije, vicekapetan postaje igrač s najvećim predviđenim brojem bodova među članovima postave bez kapetana, a redoslijed klupe i dalje slijedi pravila automatskih zamjena: pričuveni vratar zauzima prvo mjesto jer može zamijeniti isključivo vratara iz početne postave, a preostala tri mjesta pripadaju igračima nevratarskih pozicija raspoređenima silazno prema predviđenim bodovima.

Uz to, mehanizam u svakome kolu također razmatra odabir jednog od triju mogućih poteza: zadržavanje postojeće momčadi, provođenje jednoga transfera ili provođenje dvaju transfera. Skup potencijalnih kupnji za svaku poziciju čini trideset igrača s najvećim predviđanjem za trenutno kolo koji već nisu članovi momčadi, pri čemu u raspoloživi proračun ulazi i potencijalni profit koji se može steći prodajom odgovarajućeg igrača. Kako su kvote za svaku poziciju fiksne, prilikom svakog transfera mehanizam zamjenjuje igrača drugim igračem iste pozicije, pa se sve kombinacije jedne i dviju takvih zamjena nad razmatranim kandidatima pretražuju iscrpno. Svaki se plan pritom vrednuje očekivanim bodovima najbolje postave i kapetana koji iz njega proizlaze, uz odbitak od četiri boda za svaki transfer koji premašuje broj raspoloživih besplatnih transfera. Za odabir postave kod konkretnog plana pritom se ne rješava program (6.1), nego se razma-

tra svih osam dopuštenih formacija te se za svaku formaciju na svakoj poziciji odabire potreban broj igrača s najvećim predviđenim brojem bodova. Takvo je rješenje znatno brže, što je važno jer se u svakome kolu procjenjuje nekoliko desetaka tisuća kandidatskih momčadi, a pozivanje rješavača za svaku od njih onemogućilo bi pravodoban odgovor korisniku. Transfer se prihvaća samo ako daje strogo bolji rezultat od zadržavanja trenutačne momčadi.

Po završetku svakoga kola obrađuju se ostvareni ishodi, pri čemu se automatske zamjene izvode prema službenim pravilima igre: igrača iz početne postave koji nije upisao nijednu minutu zamjenjuje pričuvni igrač prema unaprijed određenome redoslijedu prioriteta. Zatim se za sljedeće kolo pohranjuje ažurirano trajno stanje, koje obuhvaća stanje na računu, nabavne cijene igrača i broj dostupnih besplatnih transfera. Vrijednosti igrača u momčadi pritom se usklađuju s cijenama evidentiranim u promatranome kolu, pa se vrijednost momčadi i prodajna cijena igrača tijekom sezone razvijaju jednako kao u igri. Nakon svakoga kola provjeravaju se ispravnost sastava i nenegativnost proračuna, a ni u jednome prikazanom pokretanju nisu utvrđene nepravilnosti.

6.3. Protokol simulacije

Za potrebe vrednovanja razvijenog optimizacijskog mehanizma pokreće se povijesna simulacija na ispitnoj sezoni 2024./25., pri čemu se provodi za svih 38 kola. U prvome kolu mehanizam iz predviđanja dobivenih za to kolo i proračuna od £100 milijuna sastavlja početnu momčad, a u svakome se sljedećem kolu izvodi već opisana petlja odlučivanja, pri čemu je za sve strategije odabrana ista konfiguracija: u svakome se kolu razmatraju najviše dva transfera, a čipovi se ne koriste (mehanizam ih predlaže zasebnim heurističkim slojem koji nije predmet ovoga vrednovanja). Korištene se strategije razlikuju isključivo u izvoru predviđanja:

- **A1 i A2** — predviđanja dobivena izravnom regresijom (potpoglavlje 5.4.) odnosno dekompozicijom po komponentama (potpoglavlje 5.5.);
- **naivne heuristike** — bodovi igrača iz prethodnoga kola, prosjek bodova po susretu od početka sezone te klizni prosjek posljednjih triju nastupa (posljednje se dvije pojavljuju i kao heuristički referentni modeli u potpoglavlju 5.3.5.);

- **kolektivna prosudba** — kao predviđanje uzima se broj menadžera koji igrača imaju u momčadi, čime nastaje najzastupljenija momčad koja zadovoljava pravila igre;
- **savršeno predviđanje** (engl. *oracle*) — kao predviđanje uzimaju se sami realizirani bodovi, čime se dobiva gornja granica koja pokazuje koliko prostora sloj odlučivanja uopće ostavlja boljim predviđanjima.

S obzirom na to da na samom početku sezone tri korištene povijesne strategije nemaju dovoljno podataka za svoj izračun, svaka od njih vlastitu veličinu preuzima iz prethodne sezone: heuristika prethodnoga kola uzima bodove iz posljednjega kola prošle sezone u kojemu je igrač imao susret, heuristika sezonskoga prosjeka prosjek bodova po susretu tijekom cijele prošle sezone, a klizni prosjek posljednja tri susreta te sezone. Treći heuristički referentni model iz potpoglavlja 5.3.5., prosjek po poziciji s proširujućim prozorom, među ovim strategijama nema svoga para. Naime, on svakom igraču jedne pozicije pripisuje jednu te istu vrijednost, pa ne razdvaja nijedan par kandidata za isto mjesto u momčadi: program odabira postaje ravnodušan prema vrlo velikom skupu momčadi koje zadovoljavaju pravila igre, a njegov odabir tada odlučuje rješavač, a ne predviđanje.

6.4. Rezultati

Tablica 6.1. donosi rezultate povijesne simulacije, a tablica 6.2. kvalitetu pojedinačnih odluka iz kojih su ti rezultati nastali. Slijedi pet čitanja tih rezultata.

Tablica 6.1. Povijesna simulacija sezone 2024./25.

Strategija	Ukupno	Bodovi/kolo	Transferi	Kažnjeni transferi
Savršeno predviđanje (ostvareni bodovi)	3 693	97,2	74	37
Bodovi po susretu u sezoni	2 193	57,7	39	3
A2 — dekompozicija po komponentama	2 192	57,7	41	4
Kolektivna prosudba	2 129	56,0	38	1
A1 — izravna regresija	1 971	51,9	46	9
Bodovi iz prethodnoga kola	1 884	49,6	74	37
Klizni prosjek	1 822	48,0	72	35

Ako se upustimo u usporedbu dvaju pristupa, lako možemo zaključiti da A2 postiže puno bolji rezultat od A1: naime, na ispitnoj sezoni je osvojio čak 221 bod više (2 192 naspram 1 971), pri čemu ga nadmašuje u 23 od 38 kola, a u jednome od njih su međusobno

Tablica 6.2. Kvaliteta pojedinačnih odluka u istim pokretanjima.

Strategija	Kapetan bio najbolji u		Kola sa zamjenom	Bodovi/transfer
	vlastitoj postavi	cijelom skupu		
Savršeno predviđanje	100,0 %	100,0 %	2,6 %	+14,3
Bodovi po susretu u sezoni	36,8 %	21,1 %	44,7 %	+2,1
A2 — dekompozicija po komponentama	44,7 %	21,1 %	28,9 %	+4,2
Kolektivna prosudba	26,3 %	13,2 %	36,8 %	+3,0
A1 — izravna regresija	31,6 %	7,9 %	36,8 %	+4,3
Bodovi iz prethodnoga kola	15,8 %	5,3 %	31,6 %	+2,2
Klizni prosjek	31,6 %	13,2 %	39,5 %	+0,6

izjednačeni. A2 pritom u transfere ulazi štedljivije: u usporedbi s A1, koji je proveo 46 transfera i dobio 9 kaznenih bodova, A2 ih je proveo 41 te pritom dobio 4 kaznena boda, što je vjerojatno rezultat njegove komprimirane ljestvice predviđanja (potpoglavlje 5.6.4.) koja rijetko dopušta razliku veću od četiri boda, koliko kažnjeni transfer mora nadmašiti. Mjereno udjelom u zbroju savršenoga predviđanja, A2 osvaja 59,4 %, a A1 53,4 %. Poredak na razini cjelovitoga sustava time potvrđuje presudu do koje je potpoglavlje 5.6.2. došlo na razini pojedinačnih susreta.

Iako A2 uspijeva nadmašiti A1 na ispitnoj sezoni, to nije slučaj i s najboljom heuristikom: strategija prosjeka bodova po susretu u sezoni završava povijesnu simulaciju s 2 193 boda, a A2 s 2 192, što je razlika od samo jednog boda za cijelu sezonu. Pri tome je omjer po kolima na strani heuristike, koja osvaja više bodova u čak 21 kolu, uz jedno izjednačeno kolo. Na ovoj su sezoni dvije strategije nerazlučive i iz tablice se ne smije čitati nikakav poredak među njima. Ostale referentne strategije A2 uspijeva nadmašiti: najzastupljeniju momčad pobjeđuje za 63 boda (21–17 po kolima), strategiju bodova iz prethodnoga kola za 308 bodova (27–10–1), a klizni prosjek triju nastupa za 370 bodova (25–13). A1 nadmašuje samo posljednje dvije, dok za sezonskim prosjekom zaostaje 222 boda (12–25–1), a za kolektivnom prosudbom 158 bodova (16–21–1).

Činjenica da oba pristupa kod problema predviđanja u svakome od 38 kola nadmašuju sva tri heuristička referentna modela (i po pogrešci i po korelaciji ranga), dok ovdje tek jedan od njih završava izjednačen s najboljom od tih heuristika, a drugi čak ispod nje, možda se na prvi pogled čini neskladnom, ali ovaj rezultat još dodatno podupire

saznanja predstavljena u potpoglavlju 5.6.1. Naime, odluka o vođenju momčadi zapravo se oslanja samo na najuži vrh poretka: jedanaest igrača koji ulaze u postavu i jednoga kapetana među približno 700 kandidata, upravo u dijelu u kojem se prednost modela pokazala najmanjom: po NDCG@10 vode u približno polovici kola, po preklapanju predviđene i ostvarene najbolje deseterke u manje od trećine kola, a po odabiru kapetana nimalo. Model može biti bolji od heuristike na cijelome skupu igrača, a da u uskome pojasu iz kojega se momčad doista sastavlja ne bude nimalo bolji.

Uz usporedbu rezultata koje različite strategije postižu na cijeloj sezoni, zanimljivo je pogledati i kvalitetu pojedinačnih važnih odluka koje pritom donose. Naime, ako pogledamo tablicu 6.2. te pogledamo retke vezane uz pristup A2 te najbolju referentnu heuristiku, možemo uočiti da iako postižu praktički isti rezultat u ispitnoj simulaciji, A2 većinom donosi bolje pojedinačne odluke: primjerice, početna postava koju sastavlja zahtijeva automatsku zamjenu u 28,9 % kola, dok je kod najbolje heuristike taj postotak nešto viši te iznosi 44,7 %. Tu razliku omogućuje kvalitetan model nastupa iz potpoglavlja 5.6.3., čija F1-mjera za niske učinke (0,83 naspram 0,56 kod heuristika) upravo i jest sposobnost da se u postavu ne uvrsti igrač koji neće izaći na teren. Slična je situacija i kod odabira kvalitetnih transfera: igrači koje A2 dovodi u kolu transfera ostvaruju 4,2 boda više od onih koje prodaje, naspram 2,1 boda kod sezonskog prosjeka te 0,6 do 3,0 boda kod ostalih referentnih strategija.

Iako je prethodno spomenuto da A2 donosi bolje pojedinačne odluke, još jedan pogled u tablicu 6.2. nam govori da to nije uvijek slučaj. Iznimka je odabir kapetana: odluka koju jednako (loše) donose obje najbolje strategije. Najboljega igrača cijeloga skupa A2 kapetanom imenuje u 21,1 % kola, jednako često kao i heuristika sezonskog prosjeka, a A1 u njih 7,9 %. Ograniči li se mjera na menadžerovu vlastitu postavu, modeli izgledaju bolje, pa A2 vodi s 44,7 naspram 36,8 %, no ta mjera ovisi o sastavu momčadi u kojoj se bira: ona kaže samo da je kapetan bio najbolji među jedanaest igrača koje strategija posjeduje, a ne i da je najbolji kandidat kola uopće bio među njima. Upravo na to pitanje odgovara stupac cijeloga skupa, a ondje prednosti nema, što je isti zaključak do kojeg se došlo na razini pojedinačnih susreta u potpoglavlju 5.6.1., gdje točnost odabira kapetana nije razdvojila nijedan od dvaju pristupa od te heuristike. Odluka s najvećom polugom u pojedinom kolu time ostaje upravo ona koju modeli ostavljaju gotovo nedirnutom.

7. Web aplikacija za potporu pri donošenju odluka

U dosadašnjim su poglavljima predstavljani glavni dijelovi sustava za potporu odlučivanja: u poglavlju 4. predstavljen je objedinjeni podatkovni skup izgrađen iz triju izvora, u poglavlju 5. su razvijena dva modela predviđanja bodovnog učinka igrača u nadolazećem kolu, a u poglavlju 6. povijesnom je simulacijom optimizacijskog mehanizma ispitano koliku praktičnu korist ta predviđanja donose. Ovo poglavlje sve te dijelove objedinjuje u web aplikaciju koja ih putem jednostavnog sučelja stavlja na raspolaganje krajnjem korisniku. Aplikacija nosi naziv *Gaffer*, prema izrazu iz engleskoga nogometnog žargona kojim se označava menadžer momčadi, a počiva na dvjema središnjim komponentama: inteligentnom asistentu, koji predviđanja modela pretvara u personalizirane preporuke za upravljanje momčadi, i statističkom pregledu, koji korisniku omogućuje samostalnu analizu podataka.

Poglavlje je podijeljeno prema glavnim dijelovima ostvarenog sustava. Potpoglavlje 7.1. predstavlja funkcionalnosti razvijene aplikacije, a potpoglavlje 7.2. opisuje arhitekturu sustava. Potpoglavlja 7.3.–7.5. zatim opisuju tri modula namijenjena korisniku, dok se u potpoglavlju 7.6. razmatra sigurnosni model konačnog programskog rješenja.

7.1. Pregled funkcionalnosti

Aplikacija je organizirana oko triju modula namijenjenih korisniku. Statistički modul omogućuje korisnicima samostalno istraživanje igrača i klubova te uvid u njihove zbirne statistike po sezonama, grafikone bodovne forme i minutaže, usporedbu očekivanoga i ostvarenoga napadačkog učinka, usporednu analizu dvaju igrača te pregled težine rasporeda (engl. *fixture difficulty*). Inteligentni asistent nadograđuje taj analitički sloj interaktivnim alatom za slaganje momčadi, koji pri svakoj izmjeni provodi službena pravila

o sastavu iz poglavlja 2., te uslugom preporučivanja, koja odabranu momčad prosljeđuje optimizacijskom mehanizmu i vraća prijedlog transfera, odabir kapetana i početnu postavu. Simulacijski modul pretvara povijesnu simulaciju iz poglavlja 6. u postupak koji korisnik sam vodi: odabire sezonu, raspon kola i strategiju, pokreće simulaciju, pregledava odluke donesene u svakome kolu te uspoređuje više pokretanja tako da se njihove krivulje prikažu na istome grafikonu.

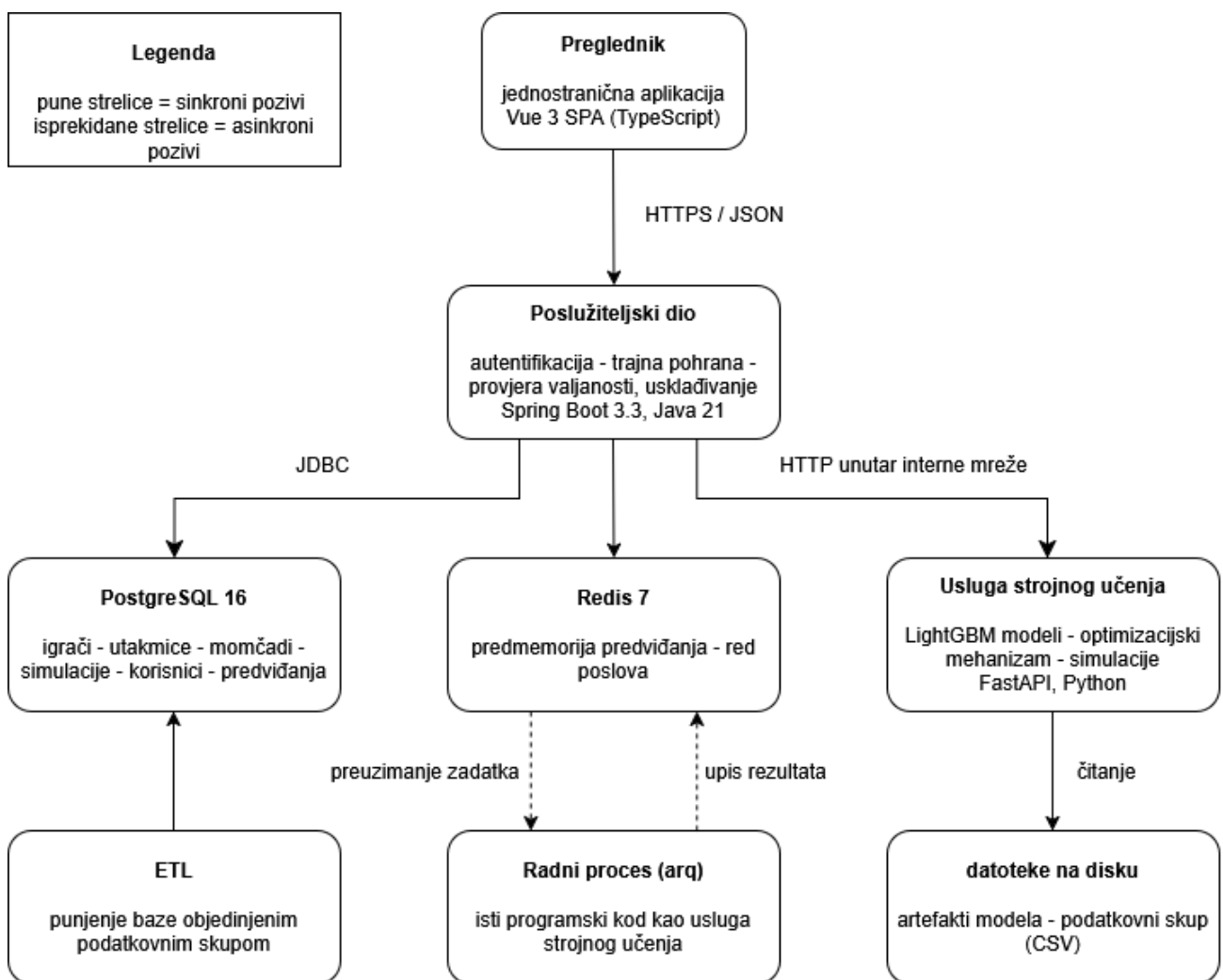
Cjelokupno je sučelje prevedeno na hrvatski i engleski jezik. Statistički modul i pokretanje simulacija dostupni su svim korisnicima, dok slaganje momčadi, preporuke te mogućnosti vezane uz korisnički račun (spremljene momčadi, povijest vlastitih simulacija i usporedba više pokretanja) zahtijevaju registraciju.

7.2. Arhitektura sustava

Gaffer je implementiran kao troslojni sustav čije se usluge (engl. *services*) mogu neovisno postavljati u pogon, kao što je i prikazano na slici 7.1. U pregledniku se izvodi jednos-tranična web aplikacija (engl. *single-page application*) ostvarena programskim okvirom Vue 3, koja s poslužiteljskim dijelom (engl. *backend*) komunicira isključivo putem protokola HTTPS. Poslužiteljski je dio napisan u programskom jeziku Java uz programski okvir Spring Boot, a zadužen je za autentifikaciju, trajnu pohranu podataka, provjeru valjanosti i usklađivanje rada ostalih usluga. Stanje aplikacije pohranjuje u bazi PostgreSQL, dok se Redis koristi za predmemoriranje (engl. *caching*) često traženih odgovora. Sva je funkcionalnost strojnog učenja izdvojena u zasebnu uslugu napisanu u Pythonu uz programski okvir FastAPI, u kojoj se učitavaju naučeni modeli, izvodi optimizacijski mehanizam i provode simulacije. Dugotrajne simulacije pritom preuzima poseban radni proces koji ih dohvaća kao zadatke iz reda poslova zasnovanog na Redisu. Baza podataka puni se unaprijed provedenim ETL postupkom, i to iz podatkovnog skupa predstavljenog u poglavlju 4. U proizvodnom okruženju sve se sastavnice izvode u kontejnerima, i to iza reverznog proxy poslužitelja (engl. *reverse proxy*) na kojemu završava TLS veza.

Takva podjela slijedi uzorak *model-as-a-service*, odnosno posluživanje naučenih modela kao zasebne usluge, a za nju postoje dva glavna razloga. Prvi je usklađenost s programskim rješenjem iz poglavlja 5. i 6., koje je razvijeno u Pythonu. Ponovno pisanje toga koda u drugom programskom jeziku značilo bi ponovno obavljanje prethodno obavlj-

nog posla, a uz to bi otvorilo mogućnost da se ponašanje sustava u proizvodnom okruženju razlikuje od onoga koje je prethodno vrednovano. Usluga u Pythonu zato je ostvarena kao tanak omotač oko istraživačkog koda, bez vlastitoga stanja, što je usput čini pogodnom za horizontalno skaliranje. Drugi je razlog sigurnost i jednostavnost održavanja: sve čemu preglednik može pristupiti okupljeno je u poslužiteljskom dijelu, pa se sigurnosni mehanizmi, ograničavanje broja zahtjeva, predmemoriranje i vođenje revizijskog traga ostvaruju na jednome mjestu, dok usluga strojnog učenja ostaje dostupna isključivo unutar interne mreže. Tablica 7.1. sažeto prikazuje glavne tehnologije upotrijebljene u svakom sloju.



Slika 7.1. Arhitektura sustava.

Tablica 7.1. Glavne tehnologije aplikacije po slojevima.

Sloj	Tehnologije	Zaduženje
Klijentski dio	Vue 3, TypeScript, Vite, PrimeVue, ECharts, Pinia, vue-i18n	Jednostranično korisničko sučelje, interaktivni grafikoni, dvojezičnost
Poslužiteljski dio	Spring Boot 3.3 (Java 21), Spring Security, JPA/Hibernate, Flyway, Bucket4j	Javni REST API: autentifikacija, trajna pohrana, provjera valjanosti, usklađivanje rada usluga, predmemoriranje, ograničavanje broja zahtjeva
Usluga strojnog učenja	Python 3.12, FastAPI, LightGBM, PuLP, pandas, arq	Posluživanje modela, optimizacijski mehanizam, provođenje simulacija
Pohrana	PostgreSQL 16, Redis 7, MinIO	Relacijski podatci, predmemorija i red poslova, objektna pohrana za sigurnosne kopije baze
Pogon i nadzor	Docker Compose, Caddy, Prometheus, Grafana, Loki, Tempo	Kontejnerizacija i postavljanje u pogon, mjerni podatci, dnevnički zapisi i tragovi izvođenja

7.3. Inteligentni asistent

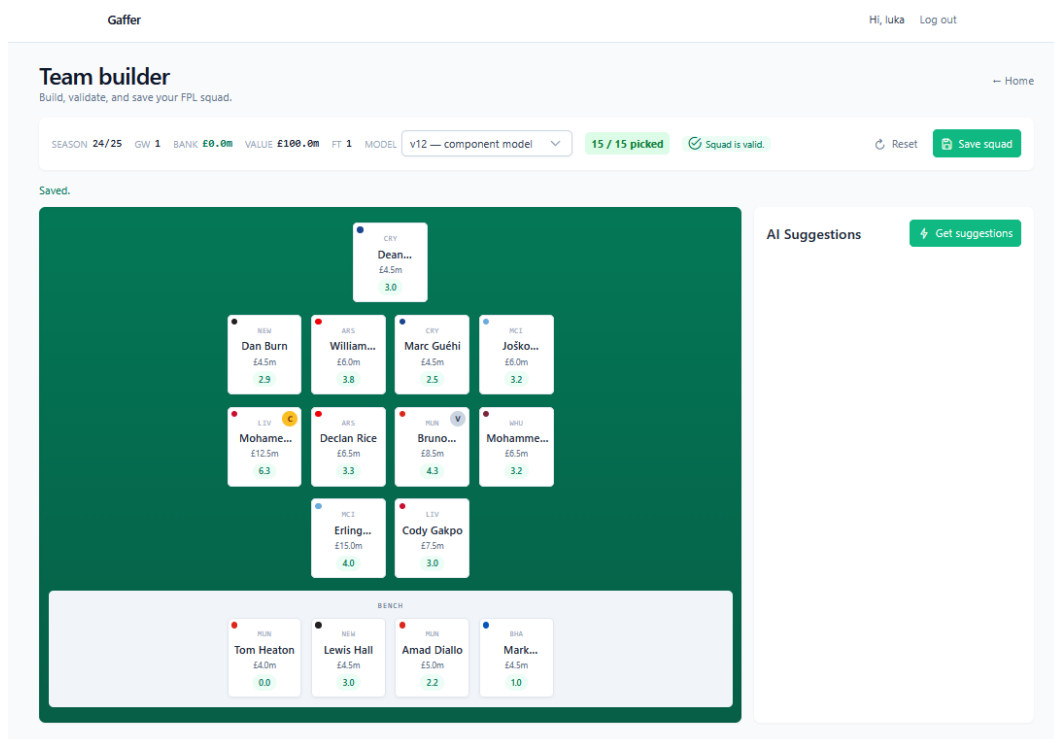
Asistent počiva na dvjema funkcionalnostima dostupnima samo registriranim korisnicima: interaktivnom alatu za slaganje momčadi i usluzi preporučivanja koja kao ulaz prima momčad složenu u tome alatu.

7.3.1. Alat za slaganje momčadi

Alat za slaganje momčadi (slika 7.2.) prikazuje cijelu momčad od petnaest igrača: početnu postavu na shematskom prikazu nogometnog terena, pri čemu se igrači određene pozicije nalaze u odgovarajućem redu, a potom i klupu s preostalim četvoricom igrača. Svaki je igrač prikazan karticom s bojama njegova kluba, cijenom i predviđenim brojem bodova koji za njega daje trenutačno odabrana inačica modela. Momčad se slaže tako da korisnik odabere prazno mjesto, čime se otvori popis igrača u kojemu se nude samo valjani odabiri za to mjesto: igrač mora igrati na traženoj poziciji i uklapati se u preostali proračun, a njegovim odabirom momčad ne smije prijeći granicu od tri igrača iz istoga kluba. Odabrani se igrači zatim mogu premješati između početne postave i klupe povlačenjem i ispuštanjem (engl. *drag-and-drop*) ili odgovarajućim naredbama s tipkovnice,

dok se kapetan bira izravno na kartici željenoga igrača.

Valjanost momčadi provjerava se na dvjema razinama. U klijentskom se dijelu provjera pokreće nakon svake izmjene i korisniku odmah prijavljuje svako kršenje pravila opisanih u poglavlju 2., među kojima su sastav momčadi 2–5–5–3, ograničenje broja igrača iz istoga kluba, proračun od £100 milijuna, dopuštena formacija te zahtjev da i kapetan i vicekapetan budu u početnoj postavi. Iste se provjere ponavljaju i na poslužiteljskom dijelu pri svakom spremanju momčadi, tako da se poslužitelj ne oslanja na stanje koje mu šalje klijentski dio. Spremljene se momčadi pohranjuju zasebno za svaki korisnički račun i ponovno učitavaju kada se korisnik vrati u aplikaciju.



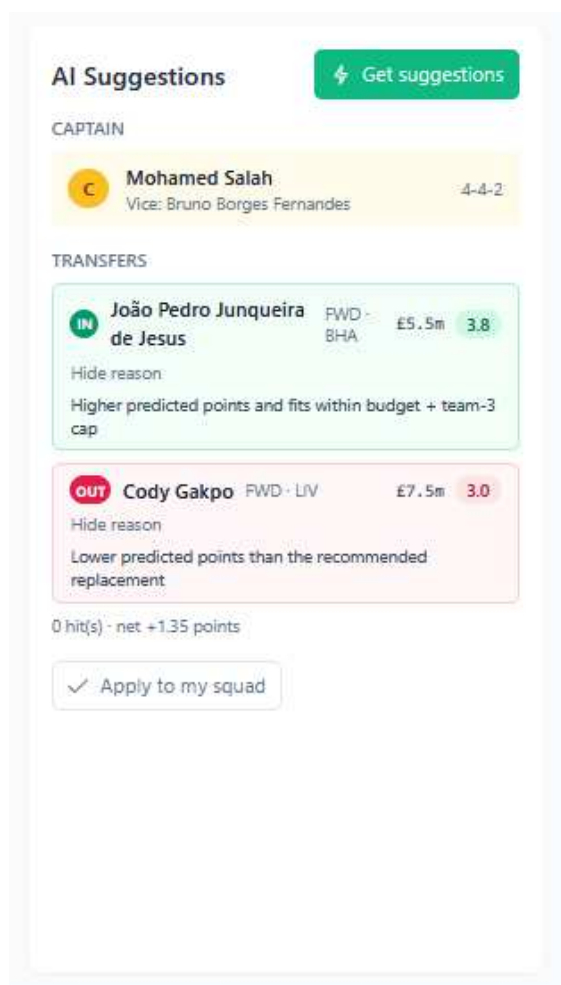
Slika 7.2. Alat za slaganje momčadi.

7.3.2. Preporuke

Usluga preporučivanja povezuje alat za slaganje momčadi s optimizacijskim mehanizmom predstavljenim u poglavlju 6., pri čemu se koristi ista njegova implementacija koju poslužuje usluga strojnog učenja. Kada korisnik zatraži savjet, poslužiteljski dio toj usluzi šalje trenutačnu momčad zajedno s odabranom sezonom, kolom i inačicom modela. Optimizacijski mehanizam zatim vrednuje moguće planove transfera prema pravilima o transferima iz potpoglavlja 6.2., odabire najbolju početnu postavu i određuje kapetana.

Vraćena preporuka sadržava konkretne transfere: za svaki od njih navodi se igrač koji izlazi iz momčadi, igrač koji u nju ulazi i predviđeni broj bodova koji model daje za obojicu, a za plan u cjelini i ukupan očekivani dobitak nakon odbitka eventualnih kaznenih bodova. Uz to su navedeni i preporučeni kapetan, vicekapetan i početna postava, a uz svaki je predloženi transfer priloženo kratko obrazloženje koje korisnik po potrebi može otvoriti (slika 7.3.).

Preporuke ne mijenjaju momčad automatski. Korisnik predloženi plan može jednim potezom prenijeti na teren, nakon toga ga i dalje uređivati, a rezultat se trajno pohranjuje tek spremanjem momčadi. Budući da se optimizacija provodi nad korisnikovom vlastitom momčadi i njegovim preostalim proračunom, savjet je personaliziran u smislu koji se od njega i traži: dva korisnika koji odaberu isti model, a imaju različite momčadi, mogu dobiti različite prijedloge transfera.

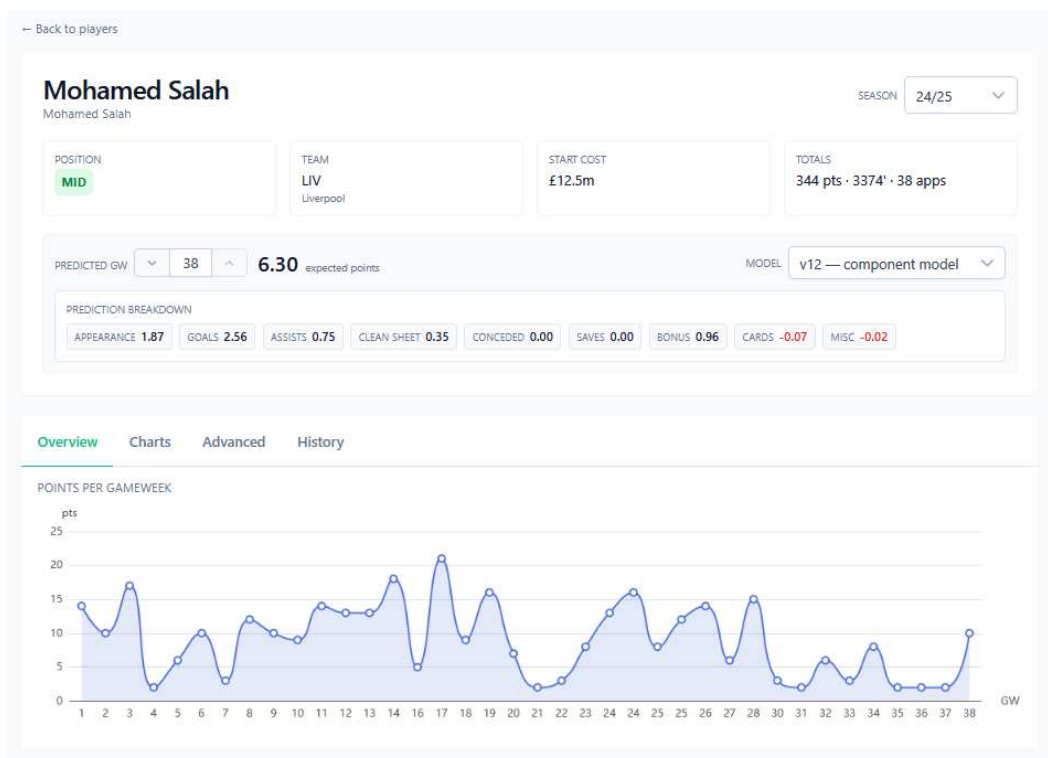


Slika 7.3. Preporuka dobivena za spremljenu momčad.

7.4. Statistički modul

Uz uslugu inteligentnog asistenta, koja korisnicima izravno pomaže pri donošenju odluka za trenutno kolo, Gaffer također pruža i potporu pri samostalnoj analizi podataka u razvijenom statističkom modulu. Odabirom ili pretraživanjem pojedinog igrača ili kluba korisnik dobiva detaljan uvid u razne statističke učinke ostvarene u svim učitanim sezonama. Primjerice, otvaranjem detaljne stranice odabranog igrača brzo se i lako mogu vidjeti njegovi bodovi po kolima, očekivani golovi u odnosu na očekivane asistencije, histogram odigranih minuta te težina igračevih susreta tijekom sezone. Zasebna kartica donosi statistike izvedene iz podataka platforme Sofascore, poput kretanja ocjene nastupa po susretima, usporedbe kumulativnih očekivanih i ostvarenih golova te obrambenih akcija po 90 minuta. Igračeva je stranica ujedno mjesto na kojemu dekompozicija po komponentama dobiva praktičnu ulogu. Naime, kada je odabran pristup A2, prikaz predviđanja razlaže očekivani učinak na devet bodovnih komponenti, među kojima su bodovi za nastup, golovi, asistencije, sačuvana mreža i preostale sastavnice. Korisnik tako ne vidi samo koliko bodova model predviđa, nego i iz čega ih očekuje (slika 7.4.). Prednost interpretabilnosti, o kojoj je za drugi pristup bilo riječi u potpoglavlju 5.5., time dolazi izravno do krajnjega korisnika.

Statistički modul dopunjuju još dva prikaza. Prikaz za usporedbu postavlja dva igrača jedan uz drugoga prema istim metrikama, u svakome retku označuje povoljniju vrijednost i jasno pokazuje kada određeni izvor nema dostupne podatke. Prikaz kluba usmjerava analizu na razinu momčadi, pa obuhvaća sezonske sažetke, raspored utakmica, kliznu napadačku i obrambenu formu izraženu golovima i očekivanim golovima te težinu nadolazećih susreta. Na kraju, matrica težine rasporeda (slika 7.5.) prikazuje sve klubove u sljedećim kolima kao toplinsku kartu, što korisniku daje poznat alat za planiranje transfera oko povoljnih i zahtjevnih nizova utakmica.



Slika 7.4. Stranica igrača uz uključen dekompozicijski model.

Fixture difficulty

All teams × upcoming gameweeks. Coloured by opponent strength quintile.

← Back to teams

SEASON 24/25 FROM GW 1 WINDOW 8

1 2 3 4 5

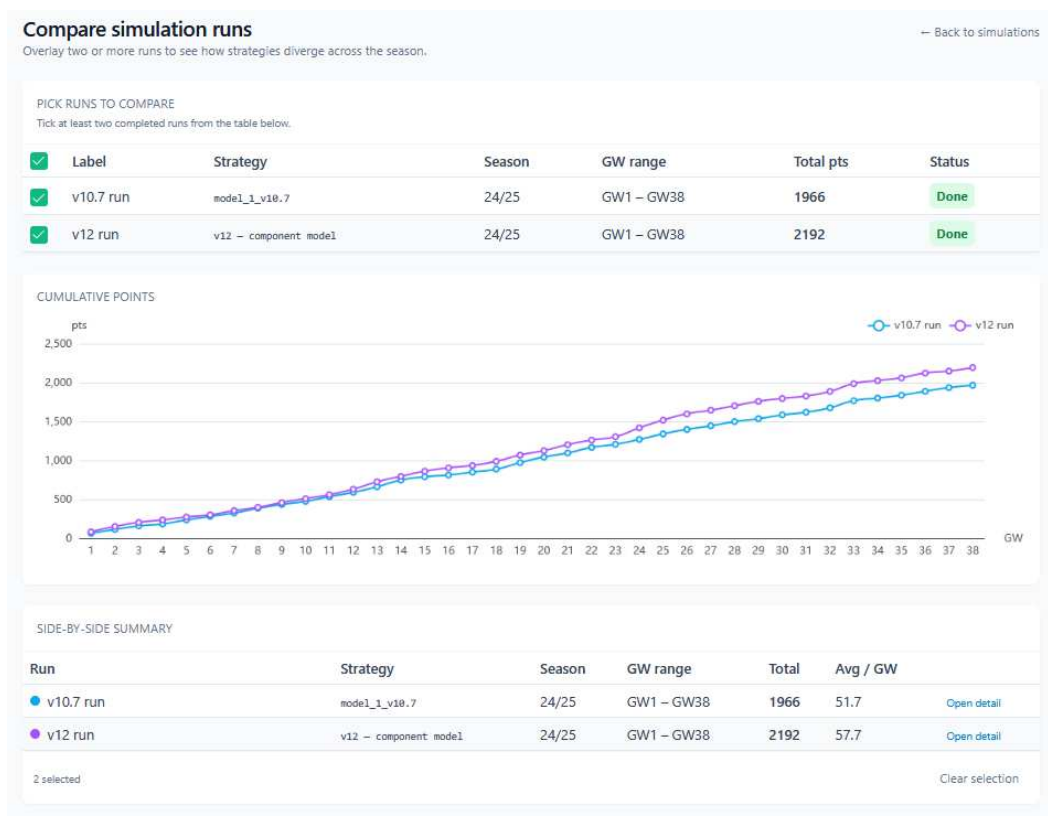
TEAM	GW1	GW2	GW3	GW4	GW5	GW6	GW7	GW8
ARS Arsenal	WOL (H)	AVL (A)	BHA (H)	TOT (A)	MCI (A)	LEI (H)	SOU (H)	BOU (A)
AVL Aston Villa	WHU (A)	ARS (H)	LEI (A)	EVE (H)	WOL (H)	IPS (A)	MUN (H)	FUL (A)
BOU Bournemouth	NFO (A)	NEW (H)	EVE (A)	CHE (H)	LIV (A)	SOU (H)	LEI (A)	ARS (H)
BRE Brentford	CRY (H)	LIV (A)	SOU (H)	MCI (A)	TOT (A)	WHU (H)	WOL (H)	MUN (A)
BHA Brighton	EVE (A)	MUN (H)	ARS (A)	IPS (H)	NFO (H)	CHE (A)	TOT (H)	NEW (A)
CHE Chelsea	MCI (H)	WOL (A)	CRY (H)	BOU (A)	WHU (A)	BHA (H)	NFO (H)	LIV (A)
CRY Crystal Palace	BRE (A)	WHU (H)	CHE (A)	LEI (H)	MUN (H)	EVE (A)	LIV (H)	NFO (A)
EVE Everton	BHA (H)	TOT (A)	BOU (H)	AVL (A)	LEI (A)	CRY (H)	NEW (H)	IPS (A)
FUL Fulham	MUN (A)	LEI (H)	IPS (A)	WHU (H)	NEW (H)	NFO (A)	MCI (A)	AVL (H)
IPS Ipswich	LIV (H)	MCI (A)	FUL (H)	BHA (A)	SOU (A)	AVL (H)	WHU (A)	EVE (H)
LEI Leicester	TOT (H)	FUL (A)	AVL (H)	CRY (A)	EVE (H)	ARS (A)	BOU (H)	SOU (A)
LIV Liverpool	IPS (A)	BRE (H)	MUN (A)	NFO (H)	BOU (H)	WOL (A)	CRY (A)	CHE (H)
MCI Man City	CHE (A)	IPS (H)	WHU (A)	BRE (H)	ARS (H)	NEW (A)	FUL (H)	WOL (A)
MUN Man Utd	FUL (H)	BHA (A)	LIV (H)	SOU (A)	CRY (A)	TOT (H)	AVL (A)	BRE (H)
NEW Newcastle	SOU (H)	BOU (A)	TOT (H)	WOL (A)	FUL (A)	MCI (H)	EVE (A)	BHA (H)
NFO Nott'm Forest	BOU (H)	SOU (A)	WOL (H)	LIV (A)	BHA (A)	FUL (H)	CHE (A)	CRY (H)
SOU Southampton	NEW (A)	NFO (H)	BRE (A)	MUN (H)	IPS (H)	BOU (A)	ARS (A)	LEI (H)
TOT Spurs	LEI (A)	EVE (H)	NEW (A)	ARS (H)	BRE (H)	MUN (A)	BHA (A)	WHU (H)
WHU West Ham	AVL (H)	CRY (A)	MCI (H)	FUL (A)	CHE (H)	BRE (A)	IPS (H)	TOT (A)
WOL Wolves	ARS (A)	CHE (H)	NFO (A)	NEW (H)	AVL (A)	LIV (H)	BRE (A)	MCI (H)

Slika 7.5. Matrica težine rasporeda.

7.5. Simulacijski modul

Simulacijski modul prenosi postupak vrednovanja iz poglavlja 6. u interaktivni dio aplikacije. Pokretanje simulacije definira se odabirom sezone, raspona kola i strategije, a dodatne postavke određuju najveći dopušteni broj transfera u pojedinom kolu te mogućnost upotrebe čipova. Korisniku je ukupno dostupno šest strategija, od kojih se dvije temelje na razvijenim modelima, dok preostale četiri odgovaraju referentnim strategijama opisanim u potpoglavlju 6.3.

Budući da simulacija cijele sezone za svako kolo ponovno izvodi petlju odlučivanja optimizacijskog mehanizma i najčešće traje između jedne i dvije minute, njezina se obrada provodi asinkrono. Nakon što korisnik preda zahtjev, poslužiteljski dio pokretanje simulacije dodaje u red poslova, zaseban radni proces izvodi izračun, a poslužiteljski dio zatim provjerava dovršetak posla i tek nakon toga pohranjuje dobiveni slijed rezultata po kolima. Stranica s rezultatima prikazuje kumulativne bodove na grafikonu te za svako kolo daje tablični prikaz bodova prije odbitka, bodova nakon odbitka kaznenih bodova za transfere, odabranog kapetana, provedenih transfera, eventualno iskorištenog čipa i preostalog iznosa na računu. Stranica za usporedbu omogućuje prikaz više takvih krivulja na istome grafikonu, kao na slici 7.6., čime se usporedbe strategija iz potpoglavlja 6.4. pretvaraju u interaktivan prikaz. Kako bi se ograničilo opterećenje sustava, jedan korisnički račun može pokrenuti najviše deset simulacija dnevno, dok se isto dnevno ograničenje primjenjuje i na IP adresu kada se simulacija pokreće bez prijave u korisnički račun.



Slika 7.6. Usporedba dvaju pokretanja simulacije tijekom sezone 2024./25.

7.6. Sigurnost

Autentifikacijski se sloj oslanja na digitalno potpisane tokene JSON Web Token (JWT), koji se prenose isključivo kolačićima (engl. *cookies*) s postavljenom oznakom `HttpOnly` i strogom postavkom atributa `SameSite`, pa ostaju nedostupni skriptama koje se izvode u pregledniku. Sjednica se temelji na kombinaciji pristupnog tokena (engl. *access token*) s trajanjem od petnaest minuta i tokena za osvježavanje (engl. *refresh token*) s trajanjem od trideset dana, pri čemu se token za osvježavanje nakon svake upotrebe rotira i u sustavu čuva isključivo u obliku sažetka. Lozinke se također ne spremaju kao izvorni tekst, nego se prije pohrane obrađuju algoritmom `bcrypt`. Ponovljeni neuspješni pokušaji prijave dovode do zaključavanja računa, a neuspjela se autentifikacija uvijek javlja namjerno istovjetnom porukom o pogrešci, kako se ne bi moglo doznati koje su adrese e-pošte već registrirane. Istek pristupnog tokena ne prekida korisnički tok rada jer klijentski dio u pozadini dohvaća novi token i zatim još jednom šalje zahtjev koji je prethodno završio neuspjehom.

Dodatne zaštite sloj po sloj sužavaju izloženu površinu napada. Broj se zahtjeva ogra-

ničava pri prijavi, registraciji računa i predaji simulacija, dok se promet s drugih izvora (engl. *cross-origin*) prihvaća isključivo s unaprijed određenog izvora klijentskog dijela. Odgovori poslužitelja nose politiku sigurnosti sadržaja (engl. *Content Security Policy*) i ostala zaštitna zaglavlja, stroga se sigurnost prijenosa (engl. *HTTP Strict Transport Security*) nameće na reverznom proxy poslužitelju, a pogreške aplikacije vraćaju se u obliku propisanom dokumentom RFC 7807, uz postojane strojno čitljive kodove, pa se klijentski dio može ravnati prema kodu umjesto da tumači tekst namijenjen čovjeku.

8. Zaključak

Cilj rada bio je izraditi sustav koji korisnicima igre *Fantasy Premier League* pruža podatkovno utemeljenu podršku pri donošenju odluka u svakom kolu. Za ostvarenje toga cilja podatci iz triju javno dostupnih izvora objedinjeni su u skup koji obuhvaća sedam sezona, pri čemu svako opažanje odgovara paru igrač–susret (poglavlje 4.). Na tako pripremljenim podacima razvijena su dva pristupa predviđanju bodovnoga učinka i međusobno uspoređena u kontroliranim uvjetima (poglavlje 5.). Dobivena se predviđanja zatim primjenjuju u optimizacijskom mehanizmu koji, uz poštovanje pravila igre, određuje transfere, početnu postavu i izbor kapetana (poglavlje 6.), dok je čitav sustav objedinjen u web aplikaciji *Gaffer* (poglavlje 7.).

Vrednovanje na razini pojedinačnih susreta pokazalo je da oba prediktivna pristupa uvjerljivo nadmašuju sva tri heuristička referentna modela. Ta se prednost pojavljuje prema mjerama pogreške i korelaciji rangova, u svakom od 38 kola i za sve igračke pozicije. U međusobnoj usporedbi dvaju pristupa dekompozicija bodova po komponentama nadmašuje izravnu regresiju prema gotovo svim metrikama koje ih razlikuju, a pritom pokazuje i osjetno manju pristranost, zbog čega je odabrana kao konačni prediktivni model. Takav je rezultat ostvaren bez ugađanja hiperparametara i bez ponderiranja primjera, odnosno uz manje prilagodbi nego što je primijenjeno u usporednom pristupu.

Najvažniji nalaz rada vidljiv je pri usporedbi dviju razina vrednovanja, jer se prednost ostvarena na cijelome skupu igrača nije prenijela na sezonski ishod. Dekompozicija je u povijesnoj simulaciji osvojila 2 192 boda, strategija zasnovana na sezonskom prosjeku bodova po susretu 2 193 boda, dok je izravna regresija zaostala za objema. Razlog takvomu raskoraku jest činjenica da se pri upravljanju momčadi od približno 700 kandidata bira samo jedanaest igrača i kapetan, odnosno odlučuje se unutar samoga vrha poretka, gdje je prednost modela najmanja. Prema NDCG@10 prediktivni pristupi vode u približno

polovici kola, prema preklapanju predviđene i ostvarene najbolje deseterke u manje od trećine kola, a pri izboru kapetana prednost ne postoji. Iz toga proizlazi općenit zaključak da model namijenjen odlučivanju treba ocjenjivati upravo u dijelu poretka iz kojega se odluka stvarno donosi, budući da prosječna točnost na cijelome skupu kandidata govori malo o njegovoj praktičnoj korisnosti.

Doprinos dekompozicijskog modela time nije umanjen, već se očituje u kvaliteti pojedinačnih odluka. Uz gotovo jednak sezonski rezultat kao najbolja heuristika, njegova početna postava zahtijeva automatsku zamjenu u 28,9 % kola, nasuprot 44,7 % kola kod heuristike, dok igrači koje model dovodi ostvaruju 4,2 boda više od igrača koje prodaje, u odnosu na 2,1 bod kod heuristike. Stoga njegova najveća praktična vrijednost nije u predviđanju izrazito visokih učinaka, nego u pouzdanijem prepoznavanju igrača koji uopće neće nastupiti.

Rezultate ipak treba tumačiti uz nekoliko ograničenja. Vrednovanje se temelji na jednoj ispitnoj sezoni i jednom slijedu odluka, pa razlika od jednoga boda između dviju vodećih strategija ne uspostavlja njihov poredak, već pokazuje da se na promatranoj sezoni ne mogu razlučiti. Obje arhitekture dijele i ograničenje svojstveno modelima očekivane vrijednosti: visok bodovni učinak nastaje istodobnim ostvarenjem više rijetkih događaja, zbog čega ga takva procjena ne može dosegnuti. Posljedice su komprimirane raspodjele predviđanja i niske F1-mjere za visoke učinke.

Navedena ograničenja ujedno upućuju na najizravnije mogućnosti daljnjega rada. Budući da dekompozicijski pristup zasebno procjenjuje pojedine komponente bodova, njihovim bi se uzorkovanjem mogla dobiti čitava raspodjela bodova umjesto jedne točkaste procjene, čime bi izbor kapetana postao odabir igrača s najvećom vjerojatnošću visokoga učinka, a ne samo s najvećom očekivanom vrijednošću. Optimizacijski mehanizam moguće je proširiti planiranjem kroz više kola, korištenjem prikupljenih besplatnih transfera i uključivanjem odluka o čipovima. Kako se najveći zaostatak pokazao pri izboru kapetana, gdje promjena arhitekture nije donijela poboljšanje, vjerojatnije je da prostor za napredak leži u podacima dostupnima bliže roku za predaju, poput najava sastava i kladioničarskih kvota, nego u daljnjim izmjenama modela. Konačno, vrednovanje kroz više sezona pružilo bi pouzdaniju sliku od one dobivene jednom simuliranom sezonom.

Literatura

- [1] Wikipedia contributors, “Fantasy football (association)”, *Wikipedia*, [https://en.wikipedia.org/wiki/Fantasy_football_\(association\)](https://en.wikipedia.org/wiki/Fantasy_football_(association)), 2026., pristupljeno: 26. svibnja 2026.
- [2] T. H. Lopez-Menchero, “How fantasy football, and fantasy premier league, went from niche hobby to mainstream boom”, *ESPN*, https://www.espn.com/soccer/story/_/id/37635033/how-fantasy-premier-league-went-niche-hobby-mainstream-boom, 2022., pristupljeno: 26. svibnja 2026.
- [3] D. Rose, “We tracked down the 12-year-old who won the inaugural telegraph fantasy football”, *Yahoo Sports*, <https://sports.yahoo.com/article/tracked-down-12-old-won-060000835.html>, 2025., pristupljeno: 27. svibnja 2026.
- [4] J. Reeves i D. Wardale, “The offline era: The history of fantasy football (part 1)”, *FPL Tips*, <https://fpltips.com/the-offline-era-of-fantasy-football/>, 2023., pristupljeno: 27. svibnja 2026.
- [5] S. Shanmugam, “History of fantasy football and origin of fpl”, *All About FPL*, <https://allaboutfpl.com/2020/05/fpl-history-and-origin-of-fantasy-football/>, 2020., pristupljeno: 27. svibnja 2026.
- [6] J. Reeves, “The fpl era: The history of fantasy football (part 3)”, *FPL Tips*, <https://fpltips.com/the-fpl-era-the-history-of-fantasy-football/>, 2023., pristupljeno: 20. lipnja 2026.
- [7] Premier League, “Fantasy Premier League — Official Fantasy Football Game of the Premier League”, <https://fantasy.premierleague.com/>, 2026., pristupljeno: 21. kolovoza 2026.

- [8] T. Matthews, S. D. Ramchurn, i G. Chalkiadakis, “Competing with humans at fantasy football: Team formation in large partially-observable domains”, u *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12)*, Toronto, Ontario, Canada, 2012., str. 1394–1400. <https://doi.org/10.1609/aaai.v26i1.8259>
- [9] W. Eilertsen, A. Gupta, i B. K. Kristiansen, “Developing a forecast-based optimization model for fantasy premier league”, diplomski ili magistarski rad, Norges teknisk-naturvitenskapelige universitet, 2018. [Mrežno]. Adresa: <https://api.nva.unit.no/publication/0198ef8a26ca-beb618aa-8ed2-4b01-a8b8-3f828eaa53c6>
- [10] N. Bonello, J. Beel, S. Lawless, i J. Debatista, “Multi-stream data analytics for enhanced performance prediction in fantasy football”, u *Proceedings of the 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2019)*, Galway, Ireland, 2019., str. 284–292. [Mrežno]. Adresa: <http://arxiv.org/abs/1912.07441>
- [11] D. Frees, P. Ravella, i C. Zhang, “Deep learning and transfer learning architectures for english premier league player performance forecasting”, 2024. [Mrežno]. Adresa: <https://arxiv.org/abs/2405.02412>
- [12] M. Tamimi i T. Tran, “Players’ performance prediction for fantasy premier league, using transformer-based sentiment analysis on news and statistical data”, *International Journal of Computer Science in Sport*, sv. 24, br. 1, str. 133–147, 2025. <https://doi.org/10.2478/ijcss-2025-0008>
- [13] D. Groos, “Openfpl: An open-source forecasting method rivaling state-of-the-art fantasy premier league services”, 2025. [Mrežno]. Adresa: <https://arxiv.org/abs/2508.09992>
- [14] E. Notelid i T. Östlund, “Enhancing fantasy premier league strategies through machine learning and large language models”, diplomski ili magistarski rad, Uppsala universitet, 2025., uPTEC I 25008.
- [15] A. A. Aribu, “Machine learning: Applying genetic algorithms for budget management & team composition in fantasy premier league (FPL)”, diplomski ili magistar-

ski rad, NOVA Information Management School, Universidade NOVA de Lisboa, 2024.

- [16] V. Anand, “FPL Historical Dataset”, *GitHub*, <https://github.com/vaastav/Fantasy-Premier-League>, 2022., pristupljeno: 25. veljače 2026.
- [17] Understat, “Understat – xg stats for teams and players”, <https://understat.com>, 2026., pristupljeno: 2. svibnja 2026.
- [18] Sofascore, “Sofascore – live scores and match statistics”, <https://www.sofascore.com>, 2026., pristupljeno: 2. ožujka 2026.
- [19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, i T.-Y. Liu, “LightGBM: A highly efficient gradient boosting decision tree”, u *Advances in Neural Information Processing Systems 30*, Long Beach, CA, USA, 2017., str. 3146–3154.
- [20] L. Grinsztajn, E. Oyallon, i G. Varoquaux, “Why do tree-based models still outperform deep learning on typical tabular data?” u *Advances in Neural Information Processing Systems 35 (NeurIPS 2022), Datasets and Benchmarks Track*, New Orleans, LA, USA, 2022., str. 507–520.
- [21] K. Järvelin i J. Kekäläinen, “Cumulated gain-based evaluation of IR techniques”, *ACM Transactions on Information Systems*, sv. 20, br. 4, str. 422–446, 2002.
- [22] T. Akiba, S. Sano, T. Yanase, T. Ohta, i M. Koyama, “Optuna: A next-generation hyperparameter optimization framework”, u *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019., str. 2623–2631.

Izjava o korištenju umjetne inteligencije

Prilikom izrade ovog rada korišteni su veliki jezični modeli Claude Fable 5 i Claude Opus 5. Pri pisanju rada korišteni su za ispravljanje gramatičkih pogrešaka, za pronalaženje nedosljednosti u tekstu (poput nedosljednog pisanja kratice *engl.*, koja se ponegdje pojavljivala kao *eng.*) te za generiranje pojedinih slika, primjerice onih koje prikazuju arhitekturu cijeloga sustava i pojedinih pristupa, a koje su naknadno ručno izrađene u alatu *draw.io*. Tijekom razvoja rješenja predstavljenog u ovom radu korišteni su za pisanje velikog broja testova kojima se ono usporedno provjeravalo, u skladu s razvojem vođenim testovima (engl. *test-driven development*).

Osvrt na korištenje umjetne inteligencije: Sav je generirani sadržaj pregledan i prihvaćen tek nakon provjere i uz potrebne izmjene: slike su poslužile kao predložak te su naknadno izrađene ručno, a generirani su testovi provjereni i po potrebi ispravljani. Budući da modeli povremeno navode netočne tvrdnje, svi su podatci i tvrdnje u radu provjereni u odnosu na izvorne podatke i vlastite rezultate.

Izjava o korištenju umjetne inteligencije: Potvrđujem da sam tijekom izrade ovog diplomskog rada koristio umjetnu inteligenciju u skladu s Politikom primjerenog korištenja umjetne inteligencije na Fakultetu elektrotehnike i računarstva, te da umjetna inteligencija nije zamijenila moje kritičko razmišljanje niti moj doprinos.