

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 73

**Predviđanje rezultata kuglačkih
utakmica primjenom metoda
nadziranog učenja**

Karlo Boroš

Zagreb, lipanj 2023.

Zagreb, 10. ožujka 2023.

ZAVRŠNI ZADATAK br. 1124

Pristupnik: **Karlo Boroš (0036531517)**

Studij: Elektrotehnika i informacijska tehnologija i Računarstvo

Modul: Računarstvo

Mentor: doc. dr. sc. Luka Humski

Zadatak: **Predviđanje rezultata kuglačkih utakmica primjenom metoda nadziranoga učenja**

Opis zadatka:

Zadaća prediktivnoga modeliranja jest predvidjeti budući ishod na temelju podataka iz prošlosti. Koristi se u različitim područjima, a primjenu ima i u području sporta - primjerice, za predviđanje rezultata utakmica. Nadzirano učenje vrsta je strojnoga učenja u kojemu se na temelju poznatih ulaznih i izlaznih podataka izrađuje model koji bi onda na temelju poznatih ulaznih podataka trebao čim preciznije odrediti (predvidjeti) nepoznate izlazne podatke. U sklopu rada potrebno je proučiti i opisati primjenu metoda nadziranoga učenja na predviđanje rezultata sportskih utakmica te prikupiti i pripremiti podatkovni skup vezan za odigravanje kuglačkih utakmica, provesti eksploratornu analizu toga skupa, identificirati ključne čimbenike koji utječu na rezultat koji kuglač/ekipa ostvaruju na utakmici te razviti model za predviđanje: 1. broja čunjeva koje će pojedini kuglač srušiti u pojedinoj utakmici 2. konačnog ishoda kuglačke utakmice. Konačno rješenje potrebno je programski implementirati te provjeriti njegovu učinkovitost.

Rok za predaju rada: 9. lipnja 2023.

SADRŽAJ

1. Uvod	1
2. Povijest i razvoj kuglanja	2
2.1. Povijest kuglanja	2
2.2. Europsko kuglanje na 9 čunjeva	3
2.3. Natjecateljsko kuglanje	3
2.4. Kuglanje u Hrvatskoj	5
3. Povezani rad	6
4. Formalni opis modela	9
4.1. Model za predviđanje pojedinačnih rezultata	10
4.1.1. Primjer izračuna modela	11
4.2. Model za predviđanje ekipnog pobjednika	13
4.2.1. Primjer izračuna modela	14
5. Treniranje modela i verifikacija	17
5.1. Podatkovni skup	17
5.2. Treniranje modela	18
5.3. Verifikacija modela i rezultati	18
5.3.1. Model za predviđanje pojedinačnih rezultata	19
5.3.2. Model za predviđanje ekipnog pobjednika	20
6. Rasprava	21
7. Programska izvedba i korištene tehnologije	22
7.1. Pregled korištenih tehnologija	22
7.1.1. Programski jezik R	22
7.1.2. Google Sheets	24

7.2. Isječci koda	24
8. Zaključak	26
Literatura	27

1. Uvod

Sport je vrlo nepredvidiv, zato je i zanimljiv mnogima. No, iako nepredvidiv, uzorci postoje i mogu se pronaći ako se zna gdje tražiti. Sportska analiza područje je koje pomoću raznih tehnika analize podataka istražuje aspekte određenog sporta (nastupi igrača, posjećenost utakmica, povećanje popularnosti i slično). Takvim analizama može se doći do mnogih korisnih informacija, a čest je cilj pronaći način kako učiniti igrače boljima i uspješnijima. Danas su te analize sve prisutnije, ali s napretkom sporta i sve većim uvođenjem tehnologije u njega, to postaje neophodno. U suprotnome, ekipe koje ne koriste dostupne podatke zaostat će u odnosu na ekipe koje su u toku s novim tehnološkim dostignućima. Popularniji sportovi, kao što su košarka, nogomet i tenis, tema su tisuća znanstvenih radova. Iako je analiziran svaki detalj, svaki pokret i svaka mogućnost u utakmici ili meču, i dalje svakim danom nastaju nova istraživanja. U kuglanju nije takav slučaj. Ovaj sport ima usku bazu navijača, primarno su to ljudi koji kuglaju i njihovi bližnji, no ima potencijal postati bolji i popularniji, a istraživanja kao što je ovo i rad na promociji sigurno mogu pomoći sportu.

Ideju za rad dobio sam kao dugogodišnji kuglač, zaljubljenik u sport, ali i u statistiku i brojeve. Želja mi je bila pronaći način za predvidjeti pojedinačne rezultate igrača na utakmicama. U tome bi mi trebali pomoći rezultati iz prošlosti koje bih iskoristio za analizu i buduća predviđanja. To pripada grani prediktivnog modeliranja, a moj zadatak bio je proučiti kako ono funkcionira, kako primijeniti to na konkretan zadatak te dobiti zadovoljavajuće rezultate.

Cilj završnog rada jest osmisliti te potom programski implementirati model za predviđanje:

1. broja čunjeva koje će pojedini kuglač srušiti u pojedinoj utakmici
2. pobjednika kuglačke utakmice.

Kroz sljedeća poglavlja rada bit će opisana kratka povijest kuglanja, radovi na slične teme, opis modela uz slike i primjere, rezultati i opis korištenih tehnologija.

2. Povijest i razvoj kuglanja

2.1. Povijest kuglanja

Kuglanje je sport s bogatom poviješću, a ona seže unatrag tisuće godina. Sami početak kuglanja teško je točno odrediti jer su razne civilizacije imale njegove varijante, a Egipat i Rim su jedni od ranijih primjera. Kuglanje je stvoreno s ciljem zabave, ali i natjecanja, a takvu strukturu zadržava i do današnjeg dana. [1]

Kroz povijest, kuglanje se razvijalo, kao i sve, tako da je prošlo kroz razne promjene. Rimski legionari bacali su kamen u želji da sleti što bliže drugom kamenu. To se kasnije odmaknulo od kuglanja, a danas je to sport po imenu boćanje. No, u grobu egipatskog dječaka pronađeni su ostaci primitivnih kugli te čunjeva, a smatra se da su stari oko 5000 godina.

Razvoju sporta svakako je doprinijelo uvođenje standardiziranih pravila i modernizacija opreme, kao što je automatsko postavljanje čunjeva. Danas razlikujemo dvije glavne podgrane kuglanja - američko kuglanje (*bowling*) te europsko kuglanje. [2] Moj rad se temelji na europskom kuglanju te ću pružiti detaljniji uvid u njega u sljedećem potpoglavlju. Na slici 2.1 je prikaz kuglane koja je napravljena za europsko kuglanje.



Slika 2.1: Kuglana

2.2. Europsko kuglanje na 9 čunjeva

Glavnu razliku između europskog i američkog kuglanja nalazimo u najosnovnijim dijelovima igre, kuglama i čunjevima. Uz mnoge druge razlike, ove dvije mijenjaju princip igre u potpunosti. Čunjeva ima 9, za razliku od 10 u *bowlingu*, te su oni postavljeni u obliku romba kao što se može vidjeti na slici 2.2. Kugle su lakše te nemaju tri utora za prste kao što imaju kugle za *bowling*.

Bowling je jako popularan u svijetu, no on je više društvena igra, primarno korištena u svrhu zabave i *team-buildinga*, a europsko kuglanje je ipak tradicionalnija sportska aktivnost. Najveću popularnost ima u Njemačkoj, koja ujedno ima i najjaču ligu u Europi. Uz Njemačku, Hrvatska, Mađarska i Srbija također imaju velik broj kuglača, pa je i razumljivo zašto su sve te države kuglačke velesile.



Slika 2.2: Raspored čunjeva

2.3. Natjecateljsko kuglanje

Natjecateljsko kuglanje može se podijeliti na:

- dobne kategorije
- kuglačke discipline
- vrstu igre.

Dobne kategorije određuju veličinu i težinu kugli kojima igrači igraju. Iako postoje juniorske i kadetske kategorije, u ovome radu analizira se seniorsko kuglanje. Ono je

na najozbiljnijoj razini koja omogućuje rezultatsku stabilnost koja čini predviđanja preciznijima. Naime, kuglači sportsku zrelost u ovome sportu rijetko postižu u juniorskim kategorijama te su zbog toga podložni velikim oscilacijama koje je teško predvidjeti.

Natjecanja se osim u raznim kategorijama odigravaju i u različitim disciplinama. Discipline koje postoje u kuglanju su:

- pojedinačno
- ekipno
- parovno
- sprint
- tandem.

Iako je svako od ovih natjecanja posebno i zanimljivo za analizu, u sklopu rada promatrat će se ekipno natjecanje, točnije 1. kuglačka muška liga. Ona obuhvaća dvije kategorije, pojedinačnu i ekipnu. U ligi se natječe deset klubova što daje veliku količinu podataka za razmatranje.

Nakon što kuglači navrš 12. godinu, oni počnu kuglati na jednak način do kraja svoje karijere. Njihov nastup sastoji se od nastupa na četiri staze, od kojih na svakoj bace 30 hitaca. Tih 30 hitaca dijelimo na 15 hitaca u "pune" te 15 hitaca u "čišćenje". U prvih 15 hitaca se nakon svake odigrane kugle srušeni čunjevi postavljaju ponovo, dok se preostalih 15 hitaca igra čišćenje. Ono započinje hicem u pune nakon kojeg se spuštaju čunjevi koji nisu srušeni u prethodnom hicu. Taj se postupak ponavlja sve dok se ne sruše svi čunjevi nakon čega se ponovno spušta svih 9 i kreće čišćenje nove postave i tako 15 hitaca. To zajedno čini stazu, a gleda se ukupni rezultat u ta dva segmenta igre.

Kuglačka utakmica u 1. ligi sastoji se od šest igrača u svakoj ekipi. Tada se igra igrač protiv igrača kroz već navedene četiri staze. Igrač koji na kraju staze ima više srušenih čunjeva osvaja set, a na kraju nastupa, igrač s više setova donosi poen ekipi. U slučaju izjednačenog broja setova, pobjednik je igrač s više čunjeva. Ako su i u čunjevima izjednačeni, svaka ekipa dobiva pola poena. Tako se ekipama dodjeljuje ukupno šest poena, a preostala dva dobiva ekipa s većim ukupnim brojem srušenih čunjeva. Pobjednik utakmice je ekipa s više osvojenih poena. [3]

2.4. Kuglanje u Hrvatskoj

Hrvatska je sportska nacija i na to su mnogi ponosni. Uspješni smo u mnogo sportova, a kuglanje nije iznimka. Iako je kroz zadnjih desetak godina muška seniorska reprezentacija bila manje uspješna, ženska reprezentacija je osvojila sve što je mogla, uz rušenje više svjetskih rekorda. Seniori nisu toliko uspješni, no očekuje se kako će uskoro početi donositi medalje kući sa svjetskih prvenstava. Juniorska reprezentacija također redovito donosi medalje sa svjetskih i europskih prvenstava, a i ovdje je Hrvatska vlasnik više svjetskih rekorda.

Klupski uspjesi nisu iznimka, jer naši nas klubovi svake godine obraduju s osvojenim medaljama na svjetskim i europskim kupovima, te nastupima u završnici kuglačke lige prvaka. Starijima nisu strana imena Nikole Dragaša i Štefice Krištof, koji su obilježili svjetsko kuglanje u vrijeme Jugoslavije.

Što se statistike tiče, samo u zadnjih 10 godina, na svjetskim i europskim prvenstvima, reprezentacija Hrvatske je osvojila čak 22 medalje u seniorskoj konkurenciji, od kojih 6 zlatnih, i 62 u juniorskoj konkurenciji. [4]

3. Povezani rad

Kuglanje ne pripada najpopularnijim sportovima. Nažalost, to je i razlog nedostatku radova na tu temu. No, sport je sport, i tko zna gdje tražiti, uvijek može pronaći sličnosti u radovima vezanima za druge sportove. Postoje radovi koji se mogu smatrati povezanim s ovim radom. Iako se sportovi dosta razlikuju, određeni segmenti su slični te se mogu iskoristiti. Poglavlje će sadržavati sažetak triju radova, od kojih će dva biti vezana za tenis, a jedan za košarku. Radovi vezani za tenis usmjerili su moj način razmišljanja kada je u pitanju određivanje ishoda kuglačke utakmice. Uzimajući u obzir njihove prednosti i mane, u raspravi s mentorom smo zajednički došli do zaključka koju je metodu najbolje koristiti. Pošto se rezultat utakmice većinski temelji na pojedinačnim rezultatima igrača, što bolje bude predviđanje pojedinačnih rezultata, to će bolje predviđanje ukupnih rezultata model dati. Treći rad dao je inspiraciju za taj dio jer je vezan za košarku i govori o predikciji poena igrača iz utakmice u utakmicu u odnosu na njihov sezonski prosjek.

Rad [5] pruža pregled znanstvenih radova o predviđanju rezultata teniskih mečeva. Razmatraju se modeli koji su jednostavnije strukture, temeljeni na regresiji, pa sve do kompleksnijih modela koji su temeljeni na strojnom učenju. Prednost tenisa je svakako vrlo jasna struktura i sustav bodovanja, što olakšava razvoj statističkih modela.

U radu je jedan od najbitnijih pristupa modeliranje teniskih mečeva pomoću tzv. točkovnih modela. To su modeli koji se temelje na procjeni vjerojatnosti osvajanja poena prilikom serviranja. Ta se vjerojatnost potom koristi za izračun vjerojatnosti da igrač osvoji meč. Ti modeli uglavnom koriste Markovljeve lance za opisivanje teniskih mečeva. To su stohastički procesi, a njihova glavna karakteristika je da njihovo buduće stanje ovisi isključivo o trenutnom stanju. Tri glavne kategorije modela za predviđanje teniskih mečeva su: točkovni modeli, modeli uparenih usporedbi i modeli temeljeni na strojnom učenju. Za svaku od tih kategorija ponuđen je teorijski uvod i najvažniji koncepti na kojima se modeli zasnivaju.

Modeli navedeni u prethodnom odlomku primarno se koriste za predviđanje rezultata prije početka meča, a postoje i modeli koji se mogu koristiti za predviđanje ishoda

meča koji je trenutno u tijeku. Ovisno o aktualnom stanju meča, koriste se različite metode za izračun vjerojatnosti osvajanja seta ili meča.

U drugom radu [6] istražuju se napredne metode za predviđanje ishoda teniskih mečeva. Cilj je unaprijediti efikasnost sportskog klađenja u stvarnom vremenu. Iako tradicionalne metode za predviđanje teniskih rezultata jesu uspješne, one koriste rekurzije temeljene na Markovljevim lancima. Problematika koja dolazi s njima je velika računaska zahtjevnost što ih čini sporijima, a kada se radi u stvarnom vremenu, to je velika mana. U ovom radu, predlaže se kombinatorni pristup koji je temeljen na binomnoj distribuciji. Autori tvrde kako je jednako efikasan, ali znatno brži u izvođenju što bi ga efektivno učinilo puno boljim.

U radu je detaljno opisana navedena binomna distribucija. Opisane su matematičke formule pojedinačnih igara i setova te je objašnjeno kako se te formule kombiniraju za predviđanje konačnog rezultata meča. Distribucija nam omogućuje odrediti sve ishode igre, a time možemo precizno predvidjeti rezultate meča. U obzir se uzimaju razni faktori, kao što su forma igrača, povijest njihovih međusobnih susreta i statistika sličnih protivnika. Svjestan koliko ovakvi faktori mogu biti bitni, u moj model su implementirani neki od tih faktora koji su primjenjivi na kuglanje.

Ovom radu na važnosti daje činjenica da se validacija predikcijskih modela provodila na stvarnim podacima u teniskim mečevima. Rezultati su pokazali da je kombinatorni pristup dostigao sličnu preciznost kao i rekurzivni modeli uz veću brzinu. Pošto je cilj ovoga rada bio poboljšati efikasnost sportskog klađenja u stvarnom vremenu, brzina kombinatornog pristupa uz zadržavanje pouzdanosti predviđanja rekurzivnog modela pokazuje se kao velika prednost.

Ukratko, ovaj rad predstavlja novi pristup predviđanju ishoda teniskih mečeva koristeći kombinatorni pristup temeljen na binomnoj distribuciji. Prednosti ovog pristupa su brzina izvođenja, visoka preciznost predviđanja i praktičnost primjene u sportskom klađenju. Ova istraživanja mogu imati šire implikacije u području analize sportskih rezultata i razvoju strategija za klađenje na teniske mečeve.

Posljednji rad koji ću komentirati nije baziran na određivanju pobjednika meča, već se fokusira na pojedince koji sudjeluju u utakmici. Rad [7] je temeljen na podacima iz NBA lige, tj. američke nacionalne košarkaške lige.

Jedan od ključnih aspekata pri analizi igre u NBA ligi je identifikacija indikatora uspješnosti igrača. U radu su istaknute neke karakteristike koje imaju potencijal da budu dobri pokazatelji uspješnosti igrača, kao što su broj šuteva unutar 3 metra od koša i broj šuteva suparničke momčadi blizu koša.

Postoje i drugi faktori koji mogu biti bitni. Uočeno je da na performanse igrača

utječe i njihova podložnost ozljedama. Također, igrači kojima ugovor ističe pokušavaju pružiti dobre nastupe i igrati iznad svojih prosjeka kako bi si potencijalno osigurali novi ugovor.

Iako nije najbitnija stavka rada, analizirano je što je bitno za predviđanje pobjednika utakmice, a kao najbitniji faktori pokazali su se: obrambeni skokovi, broj poena postignutih od suparničke momčadi te broj blokada i asistencija koje suparnička momčad ostvari. To nas navodi na zaključak da dobra obrana može više značiti za ishod utakmice od dobrog napada.

Performanse igrača često se mjere pomoću ocjene efikasnosti igrača (engl. Player Efficiency Rating - PER). PER zbraja pozitivne i negativne statističke podatke igrača te daje ocjenu performansi igrača po minuti. Njegove mane su nedostatak konteksta jer se mjere temelje na razlici u poenima, nebitno je li ukupna razlika 30 poena ili 3 poena. Također, statističke informacije često u fokus stavljaju ofenzivne karakteristike igre, što smanjuje objektivnost kod igrača koji su jači u defenzivnom dijelu igre.

Ovaj rad je pokazao kako je bitno shvaćati sport koji se analizira. Nisu za svaki sport isti faktori bitni i ključni za ishod, bilo pojedinca ili ekipe. Dubinskim razumijevanjem sporta može se doći do važnih saznanja koja mogu biti bitna u određivanju uspješnosti modela. U sljedećim poglavljima tematika će se prebaciti na kuglanje, faktore i principe koji su bitni u tom sportu te metode korištene za implementaciju modela.

4. Formalni opis modela

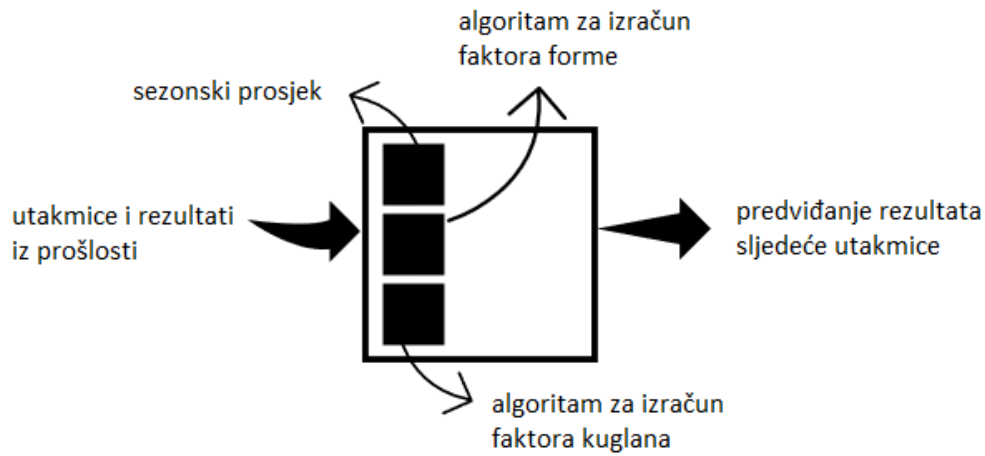
Prediktivno modeliranje [8] podgrana je podatkovne analize i strojnog učenja. Njihov cilj je izgraditi model koji može predvidjeti buduće ishode, a za to koristi dostupne podatke. Prediktivno modeliranje ima razne primjene kao što su financije, marketing, zdravstvo, sport... Ono može potencijalno predvidjeti uspjeh proizvoda na tržištu, zdravstvene ishode, detektirati prijevare, optimizirati procese. Zbog velikog potencijala postoje još brojni načini upotrebe koji nisu nabrojani.

Prediktivno modeliranje oslanja se na statističke metode, strojno učenje i duboko učenje kako bi se stvorili modeli koji mogu naučiti iz podataka i donositi predikcije. Metode nadziranog učenja korištene su u izradi modela. Nadzirano učenje je vrsta strojnoga učenja, a ono na temelju poznatih ulaznih i izlaznih podataka izrađuje model koji bi tada na temelju poznatih ulaznih podataka trebao što bolje predvidjeti nepoznate izlazne podatke.

Modeli izrađeni u sklopu ovoga rada zasnivaju se na prediktivnom modeliranju. Koriste algoritme specifične za sport i potrebe izračuna, a vrlo je bitno poznavati sport o kojemu se u istraživanju radi kako bi korištene metode imale smisla. Dva su modela, jedan koji predviđa pojedinačne rezultate igrača, a drugi koji predviđa ukupnog pobjednika utakmice.

4.1. Model za predviđanje pojedinačnih rezultata

Na slici 4.1 prikazana je arhitektura modela za predviđanje pojedinačnih rezultata:



Slika 4.1: Arhitektura modela

Formulom 4.1 se to može prikazati na sljedeći način:

$$rezultat = prosjek \cdot f_{kuglana} \cdot f_{forma} \quad (4.1)$$

Osnovni dio modela jest sezonski prosjek igrača. To je dobra baza jer, iako oscilacije kroz sezonu sigurno postoje, prosjek je pokazatelj igara kroz sezonu i logično je pretpostaviti kako je igrač s većim prosjekom uistinu bolji igrač. Prosječan sezonski rezultat dobiva se dijeljenjem svakog rezultata sa pripadajućim faktorom kuglane na kojoj je nastupio te dijeljenjem s brojem nastupa kao što se može vidjeti u formuli 4.2:

$$prosijek = \frac{1}{n} \frac{\sum_{i=1}^n X_i}{f_{kuglana, i}} \quad (4.2)$$

Problem nastaje zbog razlike u kuglanama te je zbog toga potrebno normalizirati rezultate igrača iz prošlosti. To je riješeno pridjeljivanjem faktora svakoj kuglani kojim potom množimo rezultat.

Iako laiku vjerojatno svaka kuglana izgleda isto, one imaju svoje razlike koje mogu biti ključne - od održavanja kuglane i pripreme kuglane za utakmicu do vrste čunjeva i kugli. Rezultati igrača mogu varirati na različitim kuglanama. Matematički bi se to moglo zapisati pomoću formula 4.3 i 4.4 na sljedeći način:

$$prosijek_k = \frac{\sum_{i=1}^n R_{k,i}}{n} \quad (4.3)$$

$$f_{\text{kuglana}, i} = \frac{\text{prosjeck}_i}{\max\{\text{prosjeck}_1, \text{prosjeck}_2, \dots, \text{prosjeck}_m\}} \quad (4.4)$$

gdje je n broj igrača koji su nastupili na kuglani k . Rezultat $R_{k,i}$ predstavlja rezultat igrača i na kuglani k . Kako bi se dobio faktor kuglane, njen prosjek se dijeli s najvećim od svih i time se normaliziraju nastupi igrača na različitim kuglanama.

Sljedeći faktor svakako je forma, a forma igrača mijenja se iz tjedna u tjedan. Za izračun trenutne forme analiziraju se rezultati igrača u tjednima koji prethode predmetnoj utakmici i uspoređuju se s prosjekom igrača u cijeloj sezoni. Ako su nastupi igrača u prethodnim tjednima bolji u odnosu na njegov cjelosezonski prosjek, faktor forme bit će veći od 1, što će utjecati na to da njegov predviđeni rezultat bude veći od onog predviđenog samo na temelju njegova dotadašnjeg prosjeka i faktora kuglane. Isto tako, igrač u tjednima koji prethode utakmici može postizati i lošije rezultate u odnosu na svoj cjelosezonski prosjek i tada možemo reći da je u lošoj formi te će mu faktor forme biti manji od 1.

Pri izračunu faktora forme promatraju se utakmice odigrane unutar posljednja četiri tjedna, međutim smatrao sam kako utakmica odigrana vremenski bliže promatranoj utakmici bolje odražava trenutnu formu igrača, tj., primjerice, utakmica odigrana u prethodnom tjednu značajnija je za određivanje trenutne forme igrača od utakmice odigrane prije četiri tjedna. S obzirom na to, svakom tjednu pridružuje se odgovarajući koeficijent značajnosti toga tjedna. Što je taj koeficijent veći, značajniji je utjecaj rezultata ostvarenih u tom tjednu na određivanje trenutne forme igrača. Kako bismo utjecaj faktora forme umanjili ili pojačali, uveden je koeficijent regulacije. Faktor forme izračunava se formulom 4.5:

$$f_{\text{forma}} = \frac{\sum_{i=1}^4 \frac{X_i}{\bar{X}} \cdot k_i}{\sum_{i=1}^4 k_i} \cdot k_{\text{regulacija}} \quad (4.5)$$

gdje je X_i normalizirani prosjek rezultata promatranog tjedna, \bar{X} je sezonski (normalizirani) prosjek igrača, k_i koeficijent utjecaja pojedinog prethodnog tjedna na trenutnu formu, a $k_{\text{regulacija}}$ koeficijent regulacije faktora forme.

Koeficijente utjecaja pojedinog tjedna na trenutnu formu i koeficijent regulacije faktora forme potrebno je odrediti iz skupa za učenje (treniranje) modela primjenom pristupa nadziranog učenja.

4.1.1. Primjer izračuna modela

Model koji je osmišljen i programski implementiran trebalo je i trenirati. Na primjeru 18. kola bit će prikazani međurezultati koje model koristi kako bi došao do izlaznog

podatka, predviđanja rezultata.

Iako se liga sastoji od 10 klubova, utakmice se igraju na 9 različitih kuglana, a faktori određeni neposredno prije zadnje utakmice sezone izgledali su kao na slici 4.2:

	Grad	Kuglana	Prosjek	Faktor_kuglane
1	Bjelovar	Gradska kuglana Bj	593.1783	0.968
2	Zagreb	Grmoščica	592.1674	0.967
3	Plitvička Jezera	Hotel Jezero	579.5280	0.946
4	Zadar	Mocire	605.1182	0.988
5	Osijek	Pampas	603.7328	0.986
6	Split	Poljud	595.7659	0.973
7	Đurđenovac	Sportski centar	596.9324	0.974
8	Čakovec	SRC Mladost	575.7835	0.940
9	Zaprešić	Zaprešić	612.5734	1.000

Slika 4.2: Prikaz koeficijenata kuglana

Kao što se može vidjeti, model smatra kako je Zaprešić najbolja kuglana lige, a kuglana SRC Mladost u Čakovcu najlošija. Iako 6 % razlike između kuglana na prvu ne izgleda puno, to u utakmici može napraviti razliku od 40 čunjeva za jednako dobru igru. Sa predviđanjem težine kuglane se u potpunosti slažem sa modelom.

Model je kroz testiranje kontinuirano učio i prilagođavao faktore kuglane. Inače bi bilo dovoljno analizirati rezultate iz proteklih nekoliko sezona i izračunati prosjeke, no došlo je do velikih promjena u regulativi zbog kojih rezultati iz prošlih sezona više nisu bili validni te je model učio ispočetka i svakom utakmicom ponovno računao koeficijente.

Budući da je forma prolazna, ona se također računala prije svake utakmice. Recimo da igrač ima sezonski normalizirani prosjek od 600 čunjeva. Ako u dva nastupa zaredom sruši 570 čunjeva, igrač je vidno u lošoj formi. U utakmici koja slijedi za tjedan dana, razumno bi bilo smanjiti predviđen rezultat shodno njegovoj formi. No, ako igrač s istim prosjekom u prvoj utakmici ima 570, a u sljedećoj 585, još uvijek se može reći da je u lošoj formi, ali rezultat pokazuje da se vraća u svoju standardnu formu. To je također potrebno uzeti u obzir. Faktor forme izračunava se za svakoga igrača zasebno, a svodi se na odstupanje rezultata igrača od njegovog prosjeka na utakmicama koje su se odvale u zadnjih 30 dana.

Model prije krajnjeg koraka svakom igraču dodijeli pripadajuću kuglanu i njen koeficijent. Također, dodijeli i faktor forme igrača te stvara izlaz u obliku rezultata. To

je prikazano na slici 4.3.

	Igrač	Klub	Std_uk	Kuglana	Faktor_kuglane	Faktor_forme	Rezultat
1	A. Božičević	Zadar	601.4214	Hotel Jezero	0.946	1.000	568.94
2	A. Burnać	Poštari	622.9880	Poljud	0.973	1.008	611.02
3	A. Kovač	Mertojak	655.7997	Sportski centar	0.974	1.003	640.67
4	A. Kujundžić	Zaprešić	667.8341	Grmošćica	0.967	0.995	642.57
5	A. Ratković	Poštari	613.0227	Poljud	0.973	0.981	585.14
6	B. Bakran	Zadar	617.5338	Hotel Jezero	0.946	0.997	582.43
7	B. Bogdanović	Osijek	624.0670	Pampas	0.986	1.019	627.02
8	B. Gašpert	Grmošćica	617.3058	Grmošćica	0.967	0.974	581.41
9	B. Grandić	Đurđenovac	601.3730	Sportski centar	0.974	1.006	589.25
10	B. Košak	Željezničar Čk	604.0393	Pampas	0.986	0.998	594.39
11	B. Manev	Zaprešić	659.5508	Grmošćica	0.967	0.999	637.15
12	C. Barany	Zadar	643.0954	Hotel Jezero	0.946	0.990	602.28
13	D. Domnjak	Đurđenovac	604.3751	Sportski centar	0.974	1.024	602.79
14	D. Duda	Grmošćica	640.3288	Grmošćica	0.967	1.003	621.06
15	D. Fučkar	Grmošćica	654.4275	Grmošćica	0.967	1.032	653.08

Slika 4.3: Nakon izračuna svih faktora

Preglednosti radi, igrači se grupiraju prvo prema klubovima, zatim i prema utakmicama. Slika 4.4 prikazuje kako je izgledalo predviđanje rezultata pojedinačnih igrača na primjeru utakmice Poštari-Bjelovar.

	Igrač_H	Klub_H	Rezultat_H	Igrač_G	Klub_G	Rezultat_G
1	A. Burnać	Poštari	611	Ž. Košutić	Bjelovar	608
2	M. Kolić	Poštari	600	M. Vokšan	Bjelovar	606
3	T. Odrlić	Poštari	596	S. Borčak	Bjelovar	587
4	L. Žuljević	Poštari	593	M. Teskera	Bjelovar	587
5	A. Ratković	Poštari	585	S. Šantek	Bjelovar	569
6	G. Pofuk	Poštari	575	N. Rosandić	Bjelovar	568

Slika 4.4: Pojedinačna predviđanja jedne utakmice

4.2. Model za predviđanje ekipnog pobjednika

Model za predviđanje ekipnog pobjednika koristio je model za predviđanje pojedinačnih rezultata. Razlog tome je što ekipni rezultat kroje pojedinačni rezultati. Utakmica se sastoji od 6 "mini utakmica" gdje igra igrač protiv igrača, a igrač s većim brojem čunjeva u većini slučajeva osvaja i poen. Većom preciznošću pojedinačnih rezultata postigla bi se i veća preciznost predviđanja ukupnih pobjeda. Iako model predviđa

rezultate svih igrača, kod računanja ukupnog rezultata, on u obzir uzima najboljih 6 svake ekipe. Računaju se sve kombinacije sastava dviju ekipa, a poen je pridodan igraču sa većim predviđanjem. U ovom slučaju, za to se koristi formula za permutaciju elemenata bez ponavljanja, ali gdje je redoslijed bitan. Broj permutacija izračunat je formulama 4.6 i 4.7:

$$P(n) = n! \quad (4.6)$$

$$P(6) = 6! = 720 \quad (4.7)$$

Ukupnog pobjednika određuje veći broj dodijeljenih poena. Postoje tri ishoda: pobjeda domaćina, neriješeni rezultat i pobjeda gosta. Ekipe koje u izračunatih 720 ishoda ima više pobjeda je ekipa za koju model predviđa pobjedu.

4.2.1. Primjer izračuna modela

Primjer izlaza modela na pravim podacima prikazan je slikom 4.5:

Domaćin	Gost	Pobjednik	Vjerojatnost
Poštari	Bjelovar	Poštari	0.83

Slika 4.5: Predviđanje pobjednika utakmice

Zbog jednostavnosti ćemo pretpostaviti da se utakmica sastoji od 3 "mini utakmice". Igrači ekipe A imaju sljedeće predviđene rezultate: 621, 600, 606, dok je igračima ekipe B predviđeno: 618, 613 te 604. Postupak modela prikazan je tablicama 4.1-4.6, dok konačno predviđanje nakon izračuna pokazuje tablica 4.7:

ekipa A			ekipa B		
igrač	predviđeni rezultat	poen	poen	predviđeni rezultat	igrač
1	621	1	0	618	1
2	600	0	1	613	2
3	606	1	0	604	3
ukupni poeni:		2	1		
pobjednik:		ekipa A			

Tablica 4.1: Kombinacija 1

ekipa A			ekipa B		
igrač	predviđeni rezultat	poen	poen	predviđeni rezultat	igrač
1	621	1	0	618	1
2	600	0	1	604	3
3	606	0	1	613	2
ukupni poeni:		1	2		
pobjednik:		ekipa B			

Tablica 4.2: Kombinacija 2

ekipa A			ekipa B		
igrač	predviđeni rezultat	poen	poen	predviđeni rezultat	igrač
1	621	1	0	604	3
2	600	0	1	618	1
3	606	0	1	613	2
ukupni poeni:		1	2		
pobjednik:		ekipa B			

Tablica 4.3: Kombinacija 3

ekipa A			ekipa B		
igrač	predviđeni rezultat	poen	poen	predviđeni rezultat	igrač
1	621	1	0	604	3
2	600	0	1	613	2
3	606	0	1	618	1
ukupni poeni:		1	2		
pobjednik:		ekipa B			

Tablica 4.4: Kombinacija 4

ekipa A			ekipa B		
igrač	predviđeni rezultat	poen	poen	predviđeni rezultat	igrač
1	621	1	0	613	2
2	600	0	1	604	3
3	606	0	1	618	1
ukupni poeni:		1	2		
pobjednik:		ekipa B			

Tablica 4.5: Kombinacija 5

ekipa A			ekipa B		
igrač	predviđeni rezultat	poen	poen	predviđeni rezultat	igrač
1	621	1	0	613	2
2	600	0	1	618	1
3	606	1	0	604	3
ukupni poeni:		2	1		
pobjednik:		ekipa A			

Tablica 4.6: Kombinacija 6

Pobjednik	Broj kombinacija
Ekipa A	2
Neriješeno	0
Ekipa B	4
Odluka:	Ekipa B

Tablica 4.7: Izračun krajnjeg predviđanja

5. Treniranje modela i verifikacija

5.1. Podatkovni skup

Ulazni podaci preuzeti su sa službene stranice Hrvatskog kuglačkog saveza [9]. U aplikaciju implementiranu u ovu stranicu suci upisuju pojedinačne rezultate svih utakmica. Ti podaci su potom ubačeni u Google Sheets dokument koji je izgledao kao na slikama 5.1 i 5.2. Model je iz dokumenta uzimao podatke koji su mu bili potrebni za izračun izlaznih varijabli.

ID_ut	Kolo	Br_ut	Datum	Domaćin	Gost	Kuglana
9660	1	1	10.09.2022	Đurđenovac	Grmoščica	Sportski centar
9661	1	2	10.09.2022	Osijek	Mertojak	Pampas
9662	1	3	10.09.2022	Plitvice	Željezničar Čk	Hotel Jezero
9663	1	4	09.09.2022	Poštar	Zadar	Poljud
9664	1	5	10.09.2022	Zaprešić	Bjelovar	Zaprešić

Slika 5.1: Google Sheets - popis utakmica

ID_ut	Igrač	Gost	Poz	Pr	Pu	Či	Uk	S	P
3560	K. Milinković	D	1	0	360	186	546	1,0	0,0
3560	V. Galjanić	D	2	1	389	232	621	2,0	1,0
3560	D. Fučkar	D	3	0	407	218	625	4,0	1,0
3560	D. Špiljar	D	6	1	415	236	651	4,0	1,0
3561	M. Grivičić	D	1	1	392	245	637	3,0	1,0
3561	H. Marinović	D	3	1	400	228	628	4,0	1,0

Slika 5.2: Google Sheets - rezultati igrača

Uz prikazane podatke, bilo je još i mnogih drugih: suci i njihovi rangovi, sezona u kojoj se utakmica igra, kolo i broj utakmice, rang natjecanja, nastupi igrača po stazama (pune, čišćenje, prazne kugle, ukupni rezultat, setovi), rezultati ekipa (ukupni čunjevi, razlika čunjeva, osvojeni setovi, poeni i slično). Velik raspon dostupnih podataka omogućio je raznovrsnost i dobar izbor podataka koji su mogli pomoći pri analizi.

5.2. Treniranje modela

U postupku treniranja modela potrebno je odrediti koeficijente utjecaja pojedinog tjedna na trenutnu formu i koeficijent regulacije faktora forme. Spomenuti koeficijenti određuju se tako da se traže oni uz koje će prosječna greška predviđanja na skupu za treniranje biti najmanja.

Nasumično je generirano 500 kombinacija ovih koeficijenata. Prva četiri koeficijenta bila su u rasponu 1-5 sa pomakom od 0,5, dok je peti ograničen na raspon 0,6-1,2 s pomakom 0,01.

Kombinacije koeficijenata koje su dale najbolje rezultate na skupu za treniranje prikazane su tablicom 5.1:

R b.	Koeficijent tjedan -1	Koeficijent tjedan -2	Koeficijent tjedan -3	Koeficijent tjedan -4	Koeficijent regulacije	Prosječno odstupanje
1	4,5	3,5	5,0	2,5	1,08	22,60722
2	1,0	4,5	3,0	5,0	0,98	22,64019
3	4,0	2,5	4,5	2,0	0,92	22,64235
4	5,0	3,5	4,5	2,0	0,86	22,66299
5	4,5	2,5	4,0	1,5	0,92	22,67155
6	4,0	2,5	3,0	5,0	1,06	22,72701
7	2,0	4,0	1,0	3,5	0,84	22,73494
8	4,5	4,0	3,5	5,0	1,10	22,74254
9	3,0	3,5	4,0	2,0	1,08	22,74353
10	4,5	2,5	4,0	1,0	1,18	22,76546

Tablica 5.1: Najboljih 10 generiranih kombinacija koeficijenata

Za daljnje primjene odabrana je kombinacija koeficijenata koja prema tablici daje najbolje rezultate (4,5; 5; 2; 3,5; 1,12).

5.3. Verifikacija modela i rezultati

Model je testiran na stvarnim podacima. Svi podaci su iz sezone 2022./2023. i podijeljeni su na dvije skupine, skup za treniranje i skup za testiranje. U sklopu rada izrađena su dva modela, model za predviđanje pojedinačnih rezultata i model za predviđanje ekipnog pobjednika. Skup za treniranje oba modela sastojao se od prvih devet kola ove sezone, dok je skup za učenje sadržavao preostalih devet kola sezone.

5.3.1. Model za predviđanje pojedinačnih rezultata

Model je na skupu za treniranje odredio najbolju kombinaciju koeficijenata forme. Sa tim izborom, prosječno odstupanje čunjeva cijelog skupa bilo je 23,11. U tablici 5.2 prikazano je prosječno odstupanje za svako kolo u skupu za testiranje.

Tablica 5.2: Predviđanje pobjednika utakmice

	Odstupanje (čunjevi)	Odstupanje (%)
Kolo 10	17,10	2,88 %
Kolo 11	23,25	3,85 %
Kolo 12	23,86	4,01 %
Kolo 13	22,92	3,79 %
Kolo 14	21,89	3,69 %
Kolo 15	27,17	4,48 %
Kolo 16	26,27	4,35 %
Kolo 17	24,87	4,19 %
Kolo 18	22,40	3,71 %
Ukupno	23,40	3,90 %

Po prikazanome, rezultati variraju između 17 i 28 čunjeva po igraču, što pripada intervalu između 2,8 i 5 %, a prosječno odstupanje je nešto manje od 4 %.

5.3.2. Model za predviđanje ekipnog pobjednika

Svako kolo sastoji se od pet utakmica, što daje ukupan broj od 45 utakmica u drugoj polusezoni. Tablica 5.3 prikazuje broj točno predviđenih pobjednika u svakome kolu te ukupni rezultat.

Tablica 5.3: Predviđanje pobjednika utakmice

	Broj pogodaka	Postotak pogodaka
Kolo 10	5/5	100 %
Kolo 11	4/5	80 %
Kolo 12	4/5	80 %
Kolo 13	3/5	60 %
Kolo 14	5/5	100 %
Kolo 15	2/5	20 %
Kolo 16	4/5	80 %
Kolo 17	2/5	40 %
Kolo 18	5/5	100 %
Ukupno	34/45	75,56 %

Može se vidjeti kako je postotak pogođenih rezultata otprilike 75 %. To bi značilo da model ne pogodi pobjednika svake četvrte utakmice. No postoje kola u kojima je bio precizan u svim utakmicama, dok je za par kola bio poprilično neprecizan sa manje od 50 % pogođenih pobjednika.

6. Rasprava

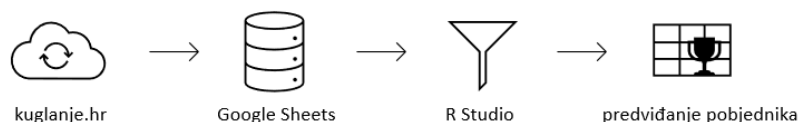
Ukupno, testirano je 45 utakmica te više od 250 pojedinačnih nastupa. Zadovoljan sam ovim rezultatima. Nisam očekivao da će postotak pobjeda biti 100%. Mislim kako to nije ni realno, sport ne bi bio zanimljiv da uvijek predviđeni pobjednik uistinu i pobijedi. U sportu sam više od 10 godina i mislim da imam određeno znanje koje ne bi trebalo biti zanemareno, a složio sam se sa skoro svim previđanjima modela što me čini sretnim. Kriva predviđanja modela bila su iznenađenje i za model, ali i za mene. Uglavnom je uzrok bio odmaranje najboljih igrača ili neočekivano loši nastupi ekipe koja je bila favorit. Takve pobjede komplicirale su situaciju na tablici i činile borbu za ostanak zanimljivijom.

Što se pojedinačnih predviđanja tiče, iako prostora za napredak ima, smatram kako je odstupanje unutar zadovoljavajućih granica. Odstupanje je manje od 5 %, a na ovako velik broj čunjeva gdje jedna devetka može promijeniti rezultat za petnaestak čunjeva, rekao bih kako je moj model dosta precizan. Kao što loš nastup ekipe može promijeniti očekivani ishod utakmice, tako i neočekivano dobar ili loš nastup igrača može jako utjecati na prosječno odstupanje. Naime, bila je situacija gdje je moj suigrač imao vjerojatno najgori nastup u posljednjih par godina. Rezultatski je bio 70 čunjeva ispod očekivanoga. U istom tom kolu je najbolji igrač na svijetu, inače igrač Zaprešića, nakon nekoliko prosječnih utakmica skoro došao do rekorda kuglane, prebacivši očekivani rezultat za 60 čunjeva. Njih dvojica sami su povećali odstupanje za više od tri čunja, a još je bilo potrebno analizirati preko 50 igrača. To govori kako stršeće vrijednosti imaju velik utjecaj na rezultate modela, obzirom na velik broj čunjeva koji je u pitanju.

7. Programska izvedba i korištene tehnologije

7.1. Pregled korištenih tehnologija

Ovaj rad zahtijevao je korištenje određenih tehnologija. Prije samoga početka izrade, bilo je potrebno razmisliti i odlučiti koji se alat činio najboljim i najprikladnijim za ovakav tip rada. Izrada je uključivala rad s podacima, njihovu analizu i prilagodbu. Kvalitetnom prilagodbom mogli su se izvući bolji rezultati. Širi izbor programskog jezika uključivao je Python, R, Matlab, Julia i slične. Odabir je na kraju pao na programski jezik R zbog predznanja koje sam imao o njemu prije početka izrade završnog rada. Na slici 7.1 može se vidjeti kratki pregled arhitekture koji će u narednim odlomcima biti objašnjen.



Slika 7.1: Arhitektura

7.1.1. Programski jezik R

Programski jezik R [10] je moćan alat, a svoju primjenu je našao primarno u statistici. Odličan je za analizu podataka i razvoj prediktivnih modela. Kao i svaki programski jezik, on ima svoje vrline i mane te bolje i lošije primjene.

R se odlikuje širokom funkcionalnošću. Relativno je jednostavno pomoću njega manipulirati podacima, raditi statističku analizu i vizualizirati podatke. Pruža velik

potencijal pri izradi prediktivnih modela i strojnom učenju. Budući da je tema bila prediktivno modeliranje u kuglanju, širok spektar paketa i biblioteka je svakako bio od pomoći.

Navedene činjenice su veliki plusevi koji olakšavaju izradu, no R nije bez mana. Krivulja učenja je poprilično strma. Sintaksa R-a nije toliko intuitivna, a početnici koji nisu upoznati sa statističkim programiranjem mogu smatrati njegove koncepte poprilično zbunjujućima. Također, brzina nije jača strana R-a. On može biti sporiji u izvođenju od nekih drugih programskih jezika, no ovaj nedostatak može se minimizirati dobrom programskom paradigmom.

R se pokazao kao dobar alat za analizu velike količine podataka. Omogućuje korisniku jednostavnu vizualizaciju i statističke mogućnosti, no sve dok brzina nije od kritične važnosti, R će biti odličan. Iako na trenutke može biti podosta kompleksan, kada se uđe u korak sa sintaksom, postaje ugodan za korištenje.

Korištene biblioteke

R u sebi sadrži velik broj paketa i biblioteka. Njihova svrha je olakšati određen dio posla. Pošto ih ima mnogo, u radu nisu sve bile potrebne, no neke ipak jesu. Sljedeći odlomci će se baviti bibliotekama koje su korištene za izradu modela.

Lubridate [11] je biblioteka u programskom jeziku R, a ona pojednostavljuje rad s datumima i vremenima. Sadrži intuitivne i lako razumljive funkcije i time olakšava manipulaciju, parsiranje i formatiranje datuma i vremena. Naročito korisna bila je pri parsiranju datuma, izračunu razlika između datuma te radu s vremenskim intervalima.

Sljedeća korištena biblioteka je *dplyr* [12] koja je bila jedna od korisnijih jer pripada najmoćnijim bibliotekama za manipulaciju podacima u R-u. U sebi ima mnogo jednostavnih i efikasnih funkcija za filtriranje, sortiranje, grupiranje, spajanje i preoblikovanje podataka. Omogućila je kvalitetnije čitanje i pisanje koda te je olakšala prilagodbu podataka kako bi zamišljeni oblik podataka bio i ostvaren.

Biblioteka *stringr* [13] ima namjenu pomoći u redu s tekstualnim podacima. Pruža velik broj funkcija za manipulaciju tekstualnim nizovima. Neke od funkcija pomažu u zamjeni, manipulaciji i pretragama unutar teksta. Pomogla je pri pripremi i čišćenju podataka za analizu.

Tidyr [14] je još jedna korisna biblioteka u programskom jeziku R koja se koristi za preoblikovanje i organizaciju podataka. Osigurava spajanje podataka, njihovu pretvorbu i razne druge mogućnosti. U kombinaciji s bibliotekom *dplyr* često se koristi za prilagodbu prije same analize.

Podaci su preuzimani iz Google Sheets dokumenta. Kako bi to bilo moguće, bila je potrebna biblioteka *googledrive* [15]. Zahvaljujući njoj, bilo je moguće autorizirati Google Drive račun čime je olakšan rad s njima, a moguće je i preuzimati dokumente.

Nakon autorizacije, potrebna je bila biblioteka koja je strogo namijenjena radu s Google Sheetsima. Tu ulogu je preuzela biblioteka *googlesheets4* [16] koja pruža funkcije za čitanje, pisanje i ažuriranje podataka izravno iz Google Sheets tablica. Korištena je za preuzimanje podataka koji su potom korišteni u analizi.

Ukratko, R je moćan alat, ali ono što ga čini još moćnijim su upravo njegove biblioteke. Bez njih bi rad bio kompliciraniji i nepregledniji, a rezultati teži za razumijevanje. Zbog navedenih razloga je prvi korak bio kvalitetno istraživanje te pronalazak korisnih biblioteka koje olakšavaju rad koji će biti potreban za izradu modela.

7.1.2. Google Sheets

Google Sheets je alat za rad s tablicama i podacima u oblaku. Korišten je kao mjesto za pohranu podataka. On pruža razne mogućnosti, među kojima je stvaranje tablica s više listova, oblikovanje podataka... Google Sheets je *web*-aplikacija i moguće je uređivati ju na bilo kojem računalu s pristupom internetu. Budući da su ovdje bili spremljeni svi potrebni podaci, mogućnost integracije Google Sheetsa i programskog jezika R je bila veliki plus.

7.2. Isječci koda

Ovo potpoglavlje pokazat će nekoliko primjera koji su omogućili ostvarenje zadanih modela.

Faktor kuglane jedan je od ključnih dijelova modela jer omogućuje usporedbu igrača ovisno o kuglani. Matematički je to bilo opisano pomoću formula 4.3 i 4.4, a programski kod je bio kako slijedi:

```
player_avg_long <- games_players %>%
  select(player, alley, total) %>%
  aggregate(total ~ player + alley, mean)
player_avg_wide <- pivot_wider(player_avg_long,
  names_from = alley, values_from = total)

bowling_alleys <- aggregate(total~alley,
  data=player_avg_long, mean)
```

```
bowling_alleys$factor <- round(
  bowling_alleys$average /
  max(bowling_alleys$average), 3)
```

Sljedeći kod prikazuje implementaciju filtriranja nastupa igrača koji su bitni za izračun forme te njihovo grupiranje po tjednima u kojima su nastupali:

```
games_last_28_days <- std_results %>%
  select(Igra , Datum, Std_uk) %>%
  mutate(Datum = ymd(Datum)) %>%
  mutate(days_since_last_game = as.numeric(
    difftime(datum, Datum, units = "days"))) %>%
  mutate(interval = cut(days_since_last_game,
    breaks = c(0, 7, 14, 21, 28),
    labels = c("Week1", "Week2",
      "Week3", "Week4"))) %>%
  group_by(Igra , interval) %>%
  summarise(prosjek = mean(Std_uk)) %>% ungroup()
```

8. Zaključak

Tema ovoga rada bila je stvaranje modela za predviđanje rezultata na kuglačkim utakmicama. Model se temelji na prediktivnom modeliranju koje uzima rezultate prošlih utakmica i stvara predviđanja. Za svakog igrača izračunava se njegov prosjek i faktor forme. Dodijeli mu se i faktor kuglane na kojoj nastupa u sljedećoj utakmici, a množenjem svih navedenih faktora dobiva se traženo predviđanje. Ekipni pobjednik računa se kombinatorikom uzimajući u obzir pojedinačne rezultate. Testiranje modela je prošlo zadovoljavajuće. Pojedinačni rezultati su odstupali za manje od 5%, dok je ekipni pobjednik ispravno pogođen u više od 70 % slučajeva.

U modelu postoje mjesta za poboljšanje. Mislim kako bi predviđanja postala preciznija ako bi se razdvojili domaći nastupi od gostujućih. Prednost domaćeg terena na određenim kuglanama je velika i to bi mogao biti razlog precjenjivanja predviđenih rezultata u gostima. Također, iako se izračun ukupnog pobjednika mogao jednakom uspješnošću odrediti i na jednostavniji način, ovaj način pruža potencijal implementiranja pogađanja točnog rezultata utakmice.

LITERATURA

- [1] Povijest kuglanja, Svibanj 2023. URL <https://www.sportilus.com/sportopedia/povijest-kuglanja/>. [Online] Pristupljeno: 8. svibnja 2023.
- [2] Podjela kuglanja, Svibanj 2023. URL [https://repozitorij.kif.unizg.hr/islandora/object/kif\\$\\%\\$3A139/datastream/PDF/view](https://repozitorij.kif.unizg.hr/islandora/object/kif$\\%$3A139/datastream/PDF/view). [Online] Pristupljeno: 8. svibnja 2023.
- [3] Kuglačka pravila, Svibanj 2023. URL <https://www.kuglanje.hr/hks/dokumenti>. [Online] Pristupljeno: 8. svibnja 2023.
- [4] Hrvatski uspjesi u kuglanju, Svibanj 2023. URL <https://www.kuglanje.hr/hks/povijest-i-uspjesi>. [Online] Pristupljeno: 8. svibnja 2023.
- [5] Vranić M. Pinter D. Krajna A. Šarčević, A. Predictive modeling of tennis matches: a review. In *MIPRO*, pages 1099–1104, Opatija, 2022. MIPRO Croatian Society.
- [6] Vranić M. Pinter D. Šarčević, A. A Combinatorial Approach in Predicting the Outcome of Tennis Matches. *Sciendo*, 31(3):525 – 538, 2021. doi: <https://doi.org/10.34768/amcs-2021-0036>.
- [7] Šarčević A. Vranić M. Pinter D. Zovak, T. Game-to-Game Prediction of NBA Players' Points in Relation to Their Season Average. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1266–1270, Opatija, 2019. MIPRO Croatian Society. doi: 10.23919/MIPRO.2019.8756733.
- [8] Prediktivno modeliranje, Svibanj 2023. URL <https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling>. [Online] Pristupljeno: 16. svibnja 2023.

- [9] Pristup ulaznim podacima, Svibanj 2023. URL <https://www.kuglanje.hr/kuglacke-lige/199>. [Online] Pristupljeno: 10. svibnja 2023.
- [10] R studio, Travanj 2023. URL <https://www.r-project.org/about.html>. [Online] Pristupljeno: 10. travnja 2023.
- [11] lubridate, Travanj 2023. URL <https://cran.r-project.org/web/packages/lubridate/lubridate.pdf>. [Online] Pristupljeno: 10. travnja 2023.
- [12] dplyr, Travanj 2023. URL <https://dplyr.tidyverse.org/>. [Online] Pristupljeno: 10. travnja 2023.
- [13] stringr, Travanj 2023. URL <https://www.rdocumentation.org/packages/stringr/versions/1.5.0>. [Online] Pristupljeno: 10. travnja 2023.
- [14] tidyr, Travanj 2023. URL <https://cran.r-project.org/web/packages/tidyr/index.html>. [Online] Pristupljeno: 11. travnja 2023.
- [15] googledrive, Travanj 2023. URL <https://googledrive.tidyverse.org/>. [Online] Pristupljeno: 11. travnja 2023.
- [16] googlesheets4, Travanj 2023. URL <https://www.rdocumentation.org/packages/googlesheets4/versions/0.1.1>. [Online] Pristupljeno: 11. travnja 2023.

Predviđanje rezultata kuglačkih utakmica primjenom metoda nadziranog učenja

Sažetak

Tema ovog rada bilo je stvoriti model za predviđanje rezultata kuglačkih utakmica, a pritom naučiti više o analizi zadanog podatkovnog skupa te baratanju istim. Cilj je svakako bio ostvariti predviđanja koja su približna stvarnima i za pojedinačne rezultate i za ekipne. Korišteno je prediktivno modeliranje kao širi skup. Preciznije, do rezultata se došlo pomoću metoda nadziranog učenja. Model je ostvaren u programskom jeziku R. Rezultati su zadovoljavajući. Predviđanje pobjednika utakmice je uspješno u prosječno 3 od 4 utakmice, a pojedinačni rezultati pripadaju prostoru unutar 5 %. Iako prostora za napredak ima, ovaj rad je pokazao kako kuglanje ipak ima konstantu koja se može iskoristiti za predikcije. Budući rad mogao bi se fokusirati na smanjenje pogreške i odstupanja, testiranje drugih metoda i slično, no metoda nadziranog učenja ispostavila se kao dobra metoda za predviđanje kuglačkih rezultata.

Ključne riječi: Kuglanje, nadzirano učenje, prediktivno modeliranje, analiza podatkovnih skupova, programski jezik R

Predicting the 9-pin bowling matches results using supervised learning methods

Abstract

The aim of this study was to create a model for predicting the outcomes of bowling matches while gaining a deeper understanding of analyzing the given dataset and handling it effectively. The ultimate goal was to achieve predictions that closely resemble the actual results for both individual and team matches. The chosen approach was predictive modeling, specifically utilizing supervised learning methods and the R programming language as the primary tool. The prediction of match winners was successful in approximately 3 out of 4 matches, and individual scores fell within a 5 % margin. Although there is room for improvement, this study demonstrated that predictive methods can be applied to a sport such as bowling. Future research could focus on reducing errors in individual predictions, exploring alternative methods, and so on. Nonetheless, supervised learning has proven to be an effective approach for predicting 9-pin bowling match results.

Keywords: 9-pin bowling, supervised learning, predictive modeling, data set analysis, R programming language