Razvoj sustava upravljanja nogometnim timovima korištenjem metoda strojnog učenja

Nikić, Anđelko

Sveučilište u Zagrebu

Ekonomski fakultet

Integrirani preddiplomski i diplomski studij

Poslovna ekonomija – smjer Menadžerska informatika

RAZVOJ SUSTAVA UPRAVLJANJA NOGOMETNIM TIMOVIMA KORIŠTENJEM METODA STROJNOG UČENJA

Diplomski rad

Anđelko Nikić

Zagreb, siječanj, 2024.

Sveučilište u Zagrebu

Ekonomski fakultet

Integrirani preddiplomski i diplomski studij

Poslovna ekonomija – smjer Menadžerska informatika

RAZVOJ SUSTAVA UPRAVLJANJA NOGOMETNIM TIMOVIMA KORIŠTENJEM METODA STROJNOG UČENJA

DEVELOPMENT OF FOOTBALL TEAM MANAGEMENT SYSTEM USING MACHINE LEARNING METHODS

Diplomski rad

Student: Anđelko Nikić

JMBAG studenta: 0067567226

Mentorica: Prof. dr. sc. Mirjana Pejić Bach

Zagreb, siječanj, 2024.

Sažetak i ključne riječi

Sport je najpopularnija vrsta zabave na svijetu, a nogomet najpopularniji od svih. On se kroz godine razvija u svim segmentima: taktički, strategijski, tehnički i fizički. Za taj razvoj zaslužna je i tehnologija koja dobiva sve više pažnje. Cilj ovoga rada je približiti čitatelju nogomet i tehnologiju koja koristi izvan terena i na terenu. U prvome dijelu ovoga rada stavljen je naglasak na svojevrsnu suradnju tehnologije i nogometa. Kreće se od najstarije upotrebe pa sve do složenih analiza i metoda strojnog učenja. Navedeni su i objašnjeni sustavi kojima se koriste nogometni timovi kako bi stekli konkurentsku prednost. U drugome dijelu rada govori se o značenju otkrivanja znanja u bazama podataka, u kojim se još industrijama ta znanstvena disciplina koristi te zašto je tako važna za svijet u kojem živimo. Prikazane su i razne metode otkrivanja znanja iz baza podataka, među njima i klaster analiza koja će biti iskorištena kasnije u istraživanju. Za kraj je ostavljeno autorsko istraživanje koje se temelji na bazi podataka koja sadrži podatke o nogometnim momčadima koje se natječu u najboljih 5 liga na svijetu u sezoni 2021.-2022. Radi se o podacima kao što je broj zabijenih golova, broj dobivenih žutih i crvenih kartona, posjed lopte i dr. U dva navrata korištena je metoda klasteriranja odnosno klaster analiza. Zaključak rada je taj da je engleska Premier liga najbolja i najgledanija liga s razlogom te da najbolje nogometne ekipe igraju nogomet visokog intenziteta, vrlo napadački te da nastoje čuvati loptu u svome posjedu što je duže moguće.

Ključne riječi: nogomet, tehnologija, otkrivanje znanja u bazama podataka, klaster analiza, rudarenje podataka

Summary and key words

Sports are the most popular form of entertainment in the world, and football is the most popular of all. Over the years, he has developed in all segments: tactical, strategic, technical and physical. Technology, which is receiving more and more attention, is also responsible for this development. The aim of this work is to bring football and the technology it uses off the field and on the field closer to the reader. In the first part of this paper, emphasis is placed on a kind of cooperation between technology and football. It ranges from the oldest usage all the way to complex analysis and machine learning methods. The systems used by football teams to gain a competitive advantage are listed and explained. The second part of the paper talks about the meaning of discovering knowledge in databases, in which other industries this scientific discipline is used and why it is so important for the world in which we live. Various methods of discovering knowledge from databases are also presented, among them cluster analysis, which will be used later in the research. Finally, author's research based on a database containing data on football teams competing in the top 5 leagues in the world in the 2021-2022 season is left. This is data such as the number of goals scored, the number of yellow and red cards received, possession of the ball, etc. On two occasions, the method of clustering or cluster analysis was used. The conclusion of the paper is that the English Premier League is the best and most watched league for a reason and that the best football teams play high-intensity, very attacking football and try to keep the ball in their possession as long as possible.

Keywords: football, technology, knowledge discovery in databases, cluster analysis, data mining

Izjavljujem i svojim potpisom potvrđujem da je diplomski rad isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

U Zagrebu,_____ Student:_____

(potpis)

STATEMENT ON THE ACADEMIC INTEGRITY

I hereby declare and confirm by my signature that the final thesis is the sole result of my own work based on my research and relies on the published literature, as shown in the listed notes and bibliography. I declare that no part of the thesis has been written in an unauthorized manner, i.e., it is not transcribed from the non-cited work, and that no part of the thesis infringes any of the copyrights. I also declare that no part of the thesis has been used for any other work in any other higher education, scientific or educational institution.

U Zagrebu,_____ Student:_____

(potpis)

Sadržaj

1. UVOD

1.1. Predmet i cilj rada

Kao glavni predmet rada *Razvoj sustava upravljanja nogometnim timovima korištenjem metoda strojnog učenja* postavljeno je istraživanje nogometne baze podataka *Football Teams* koja je preuzeta iz internetske baze podataka *Kaggle*. Kako bi čitatelj mogao što bolje razumijeti istraživanje, u sekundarnom dijelu rada opisuje se na koji način se tehnologija koristi u nogometu te što ona pruža. Prvenstveno je naglasak stavljen na analizu podataka prikupljenih nosivim uređajima koje nose igrači na terenu a zatim se ti podaci koriste u razne svrhe. Termini otkrivanja znanja u bazama podataka i rudarenje podataka su sami po sebi složeni tako da je i njima posvećen poveći dio rada. Rudarenje podataka se davno prije upotrebe u nogometu koristilo u drugim industrijama i poslovima koji su poslužili kao primjer. Kako metode otkrivanja znanja u bazama podataka pronalaze sve više primjene u svim sportovima pa tako i u nogometu, ovim radom pokušava se približiti ta, za nogometne obožavatelje, aktualna tema.

Glavni cilj ovoga rada je prikazati kako se metodama otkrivanja znanja u bazama podataka može doći do znanja koja mogu pomoći u shvaćanju nogometa. Nogomet je igra s puno aspekata a istraživanjem se otkriva koji su atributi bitni za uspjeh momčadi. Dobivena znanja mogu se upotrijebiti u planiranju treninga, utakmice ili samo kao argumenti u rasparvi. Treba shvatiti da se istraživanje temelji na usporedbi podataka različitih momčadi, a ne na stvaranju novog algoritma strojnog učenja.

1.2. Izvor podataka i metode prikupljanja

Metodologija korištena u ovome radu sastavljena je od teorijskog i empirijskog dijela. Teorijski dio temelji se na pročitanoj literaturi koja se uglavnom sastoji od knjiga, znanstvenih članaka i publikacija te internet izvora na hrvatskom i engleskom jeziku. Izvor litterature su pretežito knjižnica Ekonomskog fakulteta te internetske baze znanstvenih literatura kao što su Google Znalac, Hrčak i slično. Iz pročitane literature doneseni su zaključci i subjektivna mišljenja autora. Empirijski dio rada zasnovan je na istraživanju čiji

je predmet baza podataka koja sadrži podatke o najboljih 5 nogometnih liga u svijetu a preuzeta je s internetske stranice *Kaggle*. Na tom primjeru biti će prikazano kako se otkrivaju nova znanja iz baza podataka metodom klasteriranja. Za provođenje klaster analize upotrijebljen je računalni program koji se naziva *Weka*. Podaci dobiveni obradom u *Weki* biti će analizirani i objašnjeni. Biti će objašnjeno zašto su dobiveni podaci korisni te kako ih se može upotrijebiti.

1.3. Sadržaj i struktura rada

Rad se sastoji od četiri tematske cjeline. U uvodnom dijelu rada objašnjeni su ciljevi i predmet rada te metodologija i izvori koji su korišteni u radu. Nakon uvoda slijedi druga cjelina u kojoj su glavna tema nogomet i tehnologija odnosno njihov spoj. Ova cjelina je podijeljena na 3 glavne stavke a to su: tehnologija na terenu, nogometna analitika (tehnologija izvan terena) te nogometni sustavi i aplikacije. Kreće se u smjeru od primitivnijih tehnologija do složenijih a sve je zaokruženo sustavima koji obavljaju prikupljanje i obradu podataka. Nakon što smo približili čitatelju osnovnu temu teorijskog dijela, slijedi cjelina u kojoj je naglasak na otkrivanju znanja u bazama podataka. U toj cijelini je objašnjeno što označava taj pojam, kako se koristi, gdje se koristi te koje sve metode postoje. Otkrivanje znanja u bazama podataka koristi se u mnogim drugim branšama koje su pridonjele razvoju različitih metoda koje je zatim preuzeo nogomet i nogometni analitičari. Prethodne dvije cjeline ključne su za zadnju cjelinu koja nije teorijska već empirisjka a radi se o istraživanju nogometne baze podataka. Prvo su opisani podaci koji se koriste u istraživanju, zatim metodologija te na kraju dobiveni rezultati, rasprava i zaključak autora.

2. NOGOMET I TEHNOLOGIJA

2.1. Tehnologija na terenu

2.1.1. Tehnologija gol-linije

Tehnologija gol-linije, u svijetu poznata kao *Goal-Line Technology (GLT)*, sjajan je prikaz kako tehnologija može uvelike pomoći u donošenju ispravne odluke. I to ne bilo koje odluke, već najbitnije odluke u nogometu, a to je odluka o postignutom pogotku. Nema nikakvog prostora za manipulaciju, samo čista pravednost.

U povijesti se mnogo puta dogodilo da je donešena kriva odluka. Neki od primjera datiraju iz novijeg vremena, malo prije uvođenja ovoga sistema. Mnogi gledatelji se sjećaju kada je poništen čisti gol Ukrajine protiv Engleske na Europskom nogometnom prvenstvu 2012. godine ili kada je dvije godine ranije poništen regularan gol Engleskoj protiv Njemačke na Svjetskom prventvu u Južnoafričkoj Republici.

U takvim slučajevima, na odluku suca utječu sljedeći čimbenici:

- Položaj suca na terenu: nije poravnat s gol-linijom, nema dobar kut pogleda
- Lopta može dostići brzinu do 120 km/h, to je nemoguće ispratiti ljudskom oku
- Značajna udaljenost između (otprilike 35-40 m) između linijskog suca i gol-linije

Jedini način da se takve kontroverze potpuno izbjegnu je uvođenje GLT-a, automatiziranih sustava koji pomažu sucima u donošenju relevantnih odluka za ovakve situacije (Spagnolo et al., 2013).

Dva sustava koja su se istakla kao najbolja rješenja ovog problema su „*Hawk-Eye*“ i „*Goal Ref*“ sustavi. Oba sustava koriste velik broj brzih kamera postavljenih na raznim lokacijama po stadionu koje su usmjerene prema jednom od dva gola kako bi otkrile da li je lopta prešla gol-liniju ili ne.

2.1.1.1. Hawk-Eye

Hawkeyeov utjecaj u svijetu profesionalnog tenisa je ogroman. Od 2004. *Hawk-Eye* sustav se koristi za određivanje točne lokacije gdje lopta udari u tlo. Danas se koristi na gotovo svim razinama tenisa i postao je sveprisutan dio igre. Tenisači mogu pozvati tzv. *„Challenge"* ako nisu sigurni u odluku suca.

Hawk-Eye je predstavljen nogometnom svijetu 2012. godine kada je (zajedno s *Goal Refom*) odobren i ušao u drugu fazu testiranja. Tehnologija po prvi put testirana na međunarodnoj prijateljskoj utakmici na Wembleyju kada su igrali Belgija i Engleska, iako podaci za tu utakmicu nisu bili dostupni tijekom utakmice, samo FIFA ima pristup očitanjima sustava. U travnju 2013. *Hawk-Eye* tehnologiju službeno je odobrila Premier liga za korištenje u sezoni 2013. – 14. To je bio znak da je tehnologiju prepoznala najviša razina profesionalnog nogometa. *Hawk-eye* je prvi put korišten 17. kolovoza 2013. na stadionu Anfield u Liverpoolu (Mani, 2023).

2.1.1.2. Goal-Ref

Goal Ref je još jedna tehnologija gol linije koju preferira FIFA. Temelji se na magnetskim poljima. Dva su suprotna magnetska polja s obje strane gol linije. Lopta ima elektroniku ugrađenu unutar lopte, kao i tri magnetne trake smještene na vanjskoj površini lopte. Osim toga, senzori su smješteni unutar vratnica, kao i unutar prečke. Nakon što lopta pređe gol-liniju, antena iza gola detektira promjenu magnetskog polja i šalje signal sučevom satu unutar jedne sekunde. Ova tehnologija zahtijeva promjene na terenu i vratnicama za instalaciju sustava. Iako se ovo smatra učinkovitim i točnim, te ima mnoge iste prednosti kao i *Hawk-eye*, pojavila se zabrinutost oko lansera unutar lopte i može li se kretati unutar lopte što dovodi do pogrešnih odluka, s marginom pogreške od samo milimetara. Iako je *Goal-Ref* jeftiniji od *Hawk-Eyea*, postoji je zabrinutost oko njegove implementacije (Surujlal i Jordaan, 2013).

2.1.2. Video Assistant Referee (VAR)

Jedan od najnovijih i najkontroverznijih sustava u svijetu nogometa je sustav koji se naziva *Video Assistant Referee*, poznatiji kao VAR. VAR je uistinu podijelio nogometne

obožavatelje, jedni misle kako je potpuno uništio uzbuđenje nogometne igre dok drugi smatraju kako je donio novu dozu pravednosti koja je važnija od uzbuđenja.

Godinama su sudačke pogreške smatrane "dijelom igre", a VAR je prvi put uključen u nogometna pravila 2018. godine kako bi to promijenio. VAR je prvi puta upotrijebljen na Svjetskom prvenstvu 2018. To je vrlo lako za zapamtiti ako ste iz Hrvatske jer je to za Hrvate najbitnije natjecanje u povijesti.

VAR je sustav tehničke podrške za suce koji sucima pruža mogućnost promjene ili utjecaj na odluke, uživo. Tehnologija omogućuje suradnju suca na terenu sa video pomoćnim sucima (engl. Video Assistant Referee) izvan terena u sobi za video operacije (engl. Video Operation Room). Oni zajedno s pomoćnim VAR-om (AVAR) i operaterom za ponavljanja videosnimka (engl. Recording Operator) pregledavaju trenutak incidenta korištenjem nekoliko TV monitora. Razni vizualni elementi omogućuju različite kutove incidenta kako bi pomogli sucu na terenu da donese odluke koje je teško donijeti u stvarnom vremenu. Nova tehnologija je razvijena kako bi se pomoglo sucima u poboljšanju kvalitete sudačkih odluka, budući da su netočne odluke često utjecale na rezultat utakmica (Van Den Berg i Surujlal, 2020).

VAR sustav značajno je unaprijedio profesionalno donošenje odluka tijekom igre i vratio najviše vrijednosti poštenja i fer natjecanja u nogometu. Tijekom godina obični gledatelji svjedočili su velikom broju sudačkih pogrešaka jednostavno zbog brzine trenutka koje ljudska reakcija ponekad ne može ispratiti, a VAR je uspješno poboljšao fer uvjete koje su zahtijevali suci i gledatelji. Iz profesionalne perspektive, VAR sustavi promoviraju pravednost i točno donošenje odluka. Iako sustav ne razjašnjava sve potencijalne nejasnoće, video repriza može smanjiti pogreške, čime se poboljšava profesionalnost sudaca i igrača (Tamir i Bar-Eli, 2021).

2.2. Nogometna analitika

2.2.1. Povijest

2.2.1.1. Razdoblje poslije 2. svjetskog rata

Iako se čini kako je prikupljanje, analiziranje i rudarenje podataka tek nedavno ušlo u svijet sporta, konkretno svijet nogometa, to nije slučaj. Još sredinom prošloga stoljeća razni statističari ili samo veliki entuzijasti i zaljubljenici u sport, počeli su prikupljati podatke sa utakmica i krenuli u njihovo analiziranje ne bi li došli do rješenja. Jedan od takvih entuzijasta bio je Charles Reep.

Britanac Charles Reep se smatra prvim nogometnim analitičarem. Njegova karijera analitičara utakmica započela je tijekom utakmice između Swindon Towna i Bristol Roversa u proljeće 1950. godine. Tijekom igre, Repp je spontano izvadio svoju bilježnicu i počeo bilježiti sve što je vidio koristeći zapise i bilješke. Od tog trenutka bio je opsjednut svojom novom zanimacijom. Do kraja godine razvio je kompletan sustav bodovanja, a godinu dana kasnije izrađivao je tjedne izvještaje utakmica za Wolvese - kao savjetnik glavnog trenera Stana Cullisa pod njegovim vodstvom, Wolvesi su početkom 1950-ih postali jedan od najvećih engleskih klubova. Tri godine kasnije preselio se u Sheffield Wednesday i nakon toga obnašao druge dužnosti (Memmert i Raabe, 2018).

Ono što je bila samo Reepova zanimacija ostalo je njegova karijera sve do njegove smrti 2002. Već 1968. koristio je svoj sustav za analizu gotovo 2500 utakmica što je uistinu zapanjujući broj i ogroman skup podataka. Između ostalog zaključio je da se 80% golova postiže nakon tri dodavanja ili manje. Izračunao je i da je polovica golova postignuta nakon što se lopta vratila u protivničku polovicu, kao i da je prosječan omjer šuteva i pogodaka 10:1. Može se zaključiti kako su baš Reepove analize dovele do toga da su Englezi igrali *„long ball“* stil nogometa, što je zapravo preskakanje igre dugim loptama prema protivničkom golu (Memmert i Raabe, 2018).

2.2.1.2. Korištenje video analize u kasnim 90-ima

Korištenje videa za analizu podataka počelo je kasnih 1990-ih. Većina najboljih europskih nogometnih klubova usvojila je sustavan pristup analitici i zapošljavaju analitičare. Derby County bio je jedan od prvih klubova koji je koristio tehnologiju 1998. Klubovi sada imaju visokokvalitetne snimke i mnoštvo relevantnih statistika dostupnih klikom miša, eliminirajući potrebu za zrnatim VHS snimkama čak i za najosnovniju analizu. Treneri su polako počeli preispitivati statistike koje pružaju podatkovne tvrtke. Godine 2001. Alex Ferguson je neočekivano prodao Jaapa Stama Laziju. Ovaj potez iznenadio je sve. Ferguson to nije službeno priznao, ali statistička analiza igračevih izvedbi odigrala je faktor u dogovoru (Sekan, 2023).

2.2.1.3. Moderno doba

U posljednjih nekoliko godina analiza podataka postala je važan dio nogometa. Glavni razlog za to je sumiranje podataka koje analitika donosi nogometu. Ne samo da nogometna analitika pomaže u prikupljanju golemih količina podataka koji se ne mogu prikupiti ručno, već kada analitičari podataka obrade te podatke, timovima može pružiti mnoštvo informacija o njihovim protivnicima i njima samima.

Drugi veliki razlog je taj što se ponekad nogometne analize zanemaruju i timovi zbog toga gube konkurentsku prednost. Kada je sportska analitika prvi put ušla u sportsku industriju, naišla je na cinizam mnogih menadžera, trenera, igrača pa čak i bivših igrača jer nisu vjerovali da će im računala pružiti informacije o igri.

Najbolji primjer bio bi onaj menadžera Southamptona Harryja Redknappa iz 2005. Redknapp koji je imao mentalitet stare škole da vjeruje ljudima i njihovom iskustvu umjesto analitici, navodi se da je rekao analitičarima podataka Southamptona u to vrijeme: "Reći ću Vam, sljedeći tjedan, zašto ne bismo donijeli vaše računalo da igramo protiv njihovog računala i vidimo tko će pobijediti?"

Redknapp nikada nije koristio analitiku za planiranje utakmica te sezone i Southampton je ispao iz lige. Od 2021. 19 od 20 momčadi Premier lige koristi softver za praćenje igrača pod nazivom Prozone za prikupljanje podataka o igračima. Ponukani prošlim iskustvima nogometnih organizacija, pronašli su ravnotežu između tehničkih podataka dobivenih iz

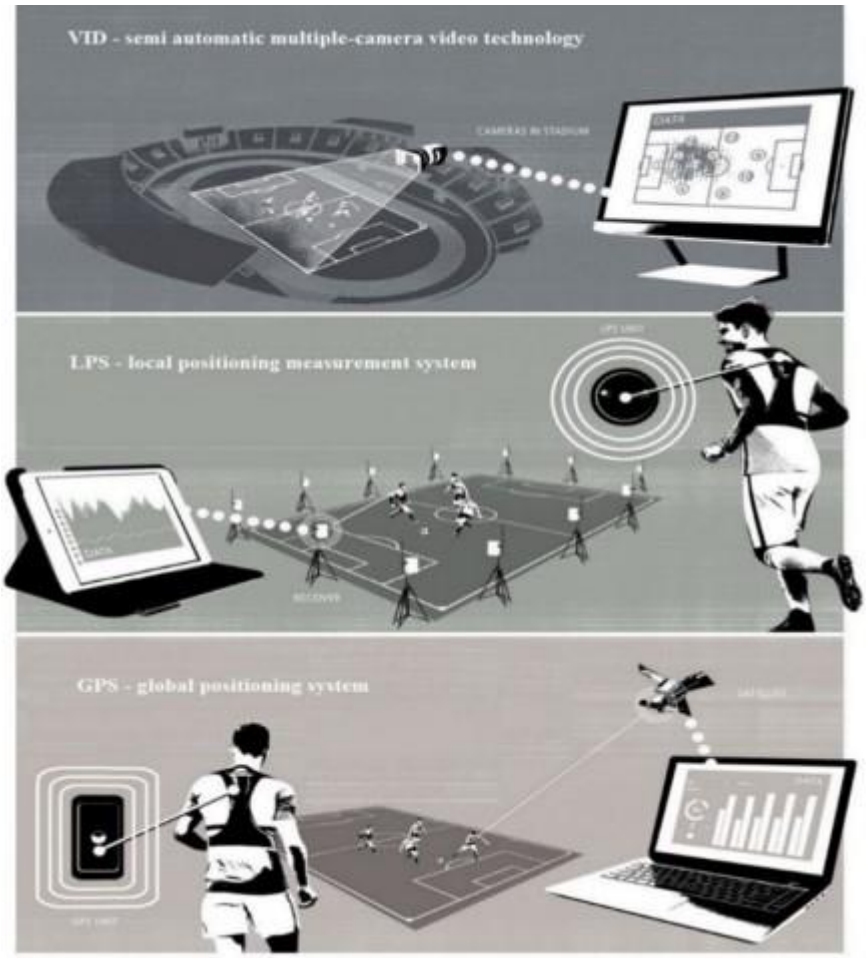
nogometne analitike i savjeta iskusnih pojedinaca koji su već godinama u nogometnoj industriji (Kumar, 2022.).

2.2.2. Prikupljanje podataka za analizu

U natjecateljskom nogometu, korištenje nosivih uređaja elektroničkog sustava za praćenje performansi (EPTS) tijekom utakmica odobreno je 2015. Nosivi EPTS uređaji uključuju: globalne (GPS) ili lokalne (LPS) sustave za pozicioniranje, akcelerometre, žiroskope i magnetometre. Često su se koristili na nogometnim treninzima za poboljšanje performansi igrača. Nakon toga je uslijedio zahtjev da igrači smiju nositi EPTS uređaje za praćenje performansi i na službenim utakmicama (Dunn, Hart i James, 2018).

Jedna od važnijih komponenti EPTS-a je i tehnologija koja se zove VID. VID je skraćeni naziv za poluautomatsku video tehnologiju s više kamera. Ova tehnologija je korištena prije nego što je dopušteno nositi ranije navedene uređaje za praćenje koji se nalaze u prsluku ispod dresa. VID je alat koji analizira ponašanje i pozicioniranje nogometaša, a za to koristi nekoliko poluautomatskih kamera visoke rezolucije koje prate igrače i postavljene su oko nogometnog igrališta po određenom uzorku. Sustav generira putanje igrača na terenu tijekom cijele utakmice i omogućuje osoblju za trening da proučava pojedinačne pokrete igrača i interakcije između njih.

Slika 1. EPTS sustav



Izvor: Bitilis, P. i Chatzipanagiotou, N. (2022). Digitalizing the Football Experience. Dostupno na: https://ceur-ws.org/Vol-3267/paper2.pdf

Podaci dobiveni iz EPTS-a mogu se podijeliti u četiri različite kategorije. Te kategorije su: fiziologija, kineziologija, neuromuskularne varijable i taktika. Fiziološke varijable povezane su s biološkim stresom koji doživljavaju nogometaši tijekom treninga ili natjecanja, poput otkucaja srca i zasićenosti kisikom. Kinematičke varijable su varijable koje se odnose na vanjska radna opterećenja, kao što su obrasci kretanja, ukupna prijeđena udaljenost i relativne udaljenosti. Ubrzanje, okrete, promjene smjera i skokove nazivamo neuromuskularnim varijablama a podatke o njima dobivamo iz već ranije spomenutih senzora kao što su troosni akcelerometri, žiroskopi i magnetometri. Taktičke varijable se

odnose na nogometnu taktiku i položaj nogometaša u odnosu na loptu ili protivnika, a mogu se dobiti iz VID-a.

Korištenjem EPTS-a, nogometaši mogu steći sveobuhvatno razumijevanje svoje nogometne izvedbe i postati svjesni svojih slabosti. Osim toga, nogometni stručni stožer može bolje procijeniti stanje svakog igrača, u skladu s tim planirati trening i poboljšati taktičko ponašanje momčadi (Bitilis and Chatzipanagiotou, 2022).

2.2.3. Upotreba prikupljenih podataka

Razvoj sportske znanosti nužan je za postizanje maksimalnih rezultata u sportu. Trenutno se gotovo sve odluke koje donose stručnjaci i menadžeri temelje na rezultatima analize i interpretacije statističkih podataka. Isto tako, popularni mediji koriste statistiku kao medij zabave, prikazujući podatke na način na koji publika može lako razumjeti. Navijači koriste statistiku za mjerenje učinka svojih omiljenih igrača ili momčadi, a zatim vode rasprave među prijateljima ili na stranicama društvenih medija.

Nogomet je sport koji se sastoji od kompleksnih događaja. Nogometna izvedba vrlo je složena i ima mnogo aspekata koje je teško definirati. Skup podataka je opsežan i sadrži informacije o događajima tijekom utakmica, uključujući vrste dodavanja, napada, udaraca, itd. Cilj statističkih metoda je učinkovito generirati relevantne informacije o procesima iz sportskih podataka, koji su često empirijski.

Analiza učinka trebala bi spajati kretanje igrača u vremenu i prostoru s opisom obrazaca igre, a komplementarnost između ta dva tipa mjerenja nudi obećavajući pristup za trenere. Bilježenje statistike koja je vezana uz izvedbu na terenu temelji se na aktivnosti koje provode igrači i momčadi tijekom igre, kao što su dodavanja, udarci, prilike, područja kretanja, obrambene akcije, dueli i obrane vratatra tijekom utakmice. Nogometna statistika može pružiti podatke vezane uz ponašanje igrača, koji se mogu koristiti kao informacije o učinku momčadi.

Nogometna analitika se koristi u različite svrhe:

- analiza utakmica
- prilagodba treninga i prevencija ozlijeda
- izviđanje novih talenata

- planiranje taktike i strategije
- predviđanje rezultata

Može se reći kako sve kreće od analiziranja utakmica. Utakmice su glavni izvor podataka koje crpe treneri te zatim u suranji sa stručnim timom donose odluke o strategiji, taktici, treninzima i ostalom. U nastavku rada govorit će se primarno o analizi utakmica te o taktici i strategiji.

2.2.3.1. Analiza utakmice

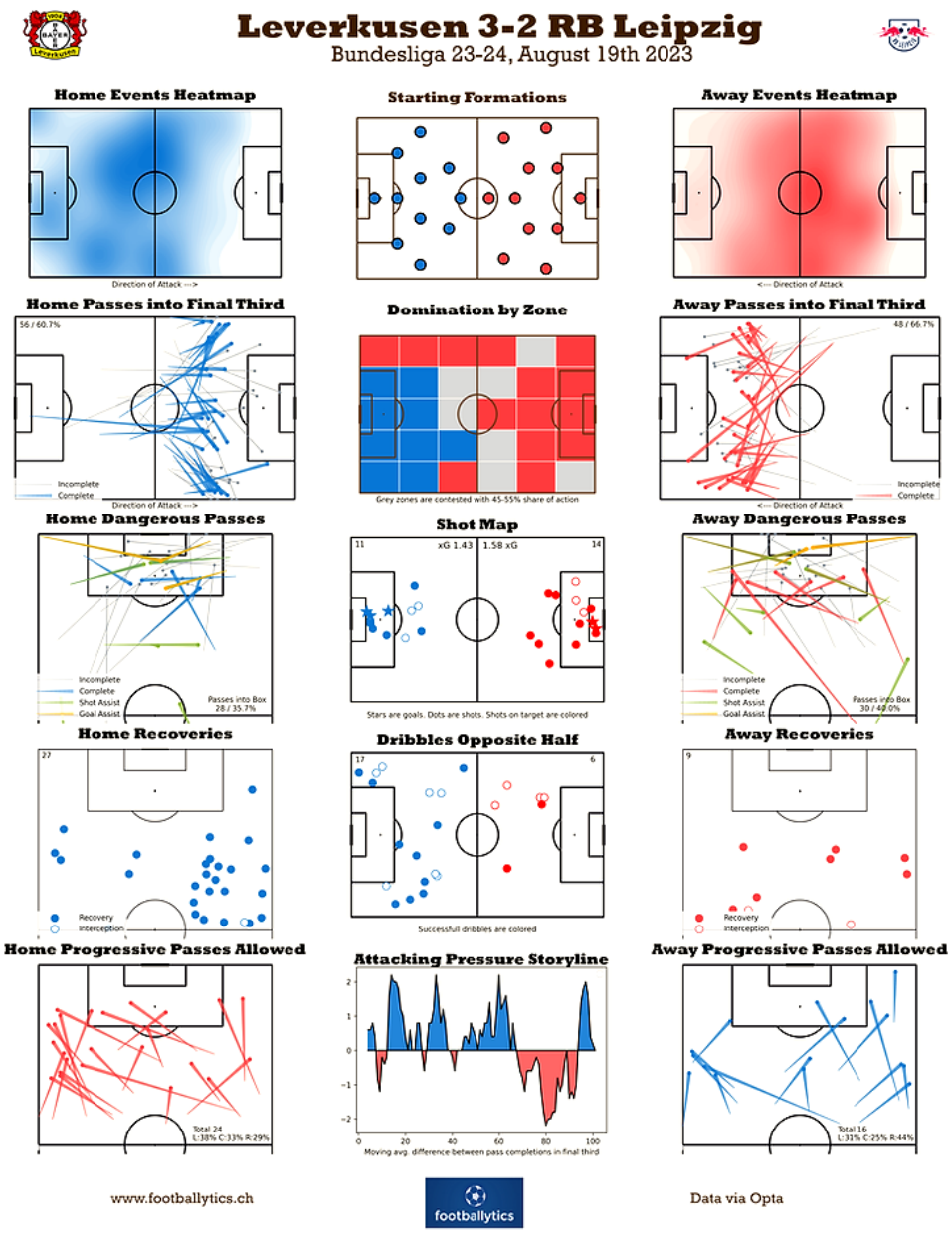
Analiza utakmice u nogometu razvila se iz jednostavnog brojanja, tj. koliko se puta određena radnja dogodila pa preko kvalitativne procjene igre od strane stručnjaka, koja također ima tendenciju biti prilično subjektivna i sve do nešto sofisticiranijeg kvantitativno vrednovanje broja driblinga, kretanja i pretrčanih udaljenosti. Konačno, došlo se do dinamičke taktičke analize gdje se napredna analitika koristi na velikim skupovima podataka za otkrivanje obrazaca, interakcija i izračunavanje složenijih KPI-jeva od učestalosti radnji (Krockel, 2019).

Ocjena igre se vrši nakon utakmice. Što se tiče ocjene igrača, sebe i igre, trener se može vratiti i planirati trening. Procjene provode treneri kako bi pružili povratne informacije o učinku igrača. Treneri mogu identificirati učinak igrača na temelju informacija prikupljenih iz prošlih nastupa (prethodnih utakmica) kako bi promijenili buduće ponašanje. (Prakoso i Lumintuarso, 2021).

Osim toga, mnoge napredne statistike vrijedne pažnje pojavile su se tijekom proteklog desetljeća, poput očekivanih golova (engl. *Expected Goals*), obrambenog pokrivanja, sekvenci i više. *Expected goals* (xG) je statistička mjera kvalitete stvorenih i primljenih prilika (očekivanih golova). xG dodjeljuje ocjenu od 0 do 1 svakoj vjerojatnosti prilike na temelju više varijabli. xG se računa za pojedinačne igrače, ali se računa i kumulativno za cijelu momčad. Model uklanja dio slučajnosti iz stvarnih postignutih golova i pruža bolji uvid u izvedbu tima. xG nije ostao imun na kritike, ali postoje slučajevi u kojima je pristup implementiran s velikim uspjehom (Apostolou i Tjortis, 2019). Nakon toga su uslijedila i mnoga druga mjerila. To uključuje očekivane asistencije (engl. *Expected Assists*), očekivane prijetnje (engl. *Expected Threats*), očekivan broj primljenih golova (engl. *Expected Goals Conceded*), vrednovanje radnji procjenom vjerojatnosti (engl. *Valuing*

Actions by Estimating Probabilities) i dopuštena dodavanja po obrambenoj akciji (engl. *Passes Allowed Per Defensive Action)* i još nekoliko (Footballytics, 2021).

Slika 2. Napredna analitika



Izvor: Footballytics (2021). Data Analytics in Football. Dostupno na: https://www.footballytics.ch/post/data-analytics-in-football

2.2.3.2. Taktika i strategija

Taktika je ključna komponenta uspjeha u modernom elitnom nogometu. Sve do nedavno, međutim, bilo je malo detaljnih znanstvenih istraživanja timskih taktika. Jedan od razloga za to je nedostatak dostupnih relevantnih podataka. To se nedavno promijenilo s razvojem napredne tehnologije praćenja. Naprotiv, sada je sve teže upravljati količinom dostupnih podataka (Rein i Memmert, 2016).

Mnogi nogometni stručnjaci još se uvijek nisu usuglasili po pitanju što je zapravo taktika a što strategija. Sama definicija taktike kaže da je to radnja ili strategija pažljivo planirana za postizanje određenog cilja. Znači riječ strategija se spominje u definiciji taktike. Opće je pririhvaćeno kako je strategija odluka donesena prije utakmice bez vremenskog ograničenja, a taktika djeluje pod jakim vremenskim ograničenjem, odnosno to su odluke donesen za vrijeme utakmice. Nadalje, strategija se odnosi na aspekte kao što su sastav momčadi i dodijeljene pozicije, dok se taktika odnosi na pozicioniranje igrača tijekom utakmice. Igrač mora biti fleksibilan i biti spreman prilagoditi se na protivnikovu igru. (Krockel, 2019). Kada je u pitanju natjecateljski nogomet, naravno, cilj je pobijediti na utakmici. Stoga je odabir prave taktike ključan za svaku pripremu prije utakmice (Krockel, 2019).

Taktička analiza se razvijala kroz godine. Tradicionalno se taktička analiza oslanjala na metode notacijske analize temeljene na prosječnim statistikama i ukupnim iznosima. Pokazatelji uključuju, na primjer, varijable dodavanja, posjed, vraćanje lopte ili stil igre. Problem kod takvih tradicionalnih metoda je što se gubi širi kontekst i slabo objašnjavaju poveznice između događaja na utakmici pa se zbog toga treneri ne oslanjaju samo na statistiku (Rein i Memmert, 2016). Korištenje statistike za pripremu taktike ili donošenje odluka tijekom utakmica pokazalo je da je 39,1% odluka bilo temeljeno na statistici, a 60,9% nije bilo temeljeno na statistici. Treneri tvrde da koriste statistiku za donošenje odluka jer su točnije i detaljnije na temelju aktivnosti koje se odvijaju na terenu. Istodobno, treneri ne koriste statistiku za donošenje odluka jer više vjeruju svojim instinktima, a uvjeti igre se brzo mijenjaju pa se odluke moraju donositi brzo (Prakoso i Lumintuarso, 2021).

Osim analiziranja samih statističkih podataka postoji i pristup usredotočen na kontrolu prostora. Na primjer, ova metoda izračunava koliki postotak terena zauzima jedna momčad u odnosu na sve igrače na terenu. Nalazi u ovom području pokazuju da napadački tim pokriva veće područje u odnosu na obrambeni tim. Isto tako, iskusni igrači pokrivaju veće

područje od manje iskusnih igrača. Kasnije se razvila metoda u kojoj se igrače proučava kao čvorove, a dodavanja između njih kao ponderirane vrhove, gdje broj dodavanja između dva igrača određuje težinu. Ovaj prikaz ponašanja timskog dodavanja omogućuje jednostavnu identifikaciju ključnih igrača u timu budući da oni pokazuju veću povezanost s drugim vrhovima kao i veće težine vrhova (Rein i Memmert, 2016).

Najsuvremenija metoda u upotrebi su algoritmi strojnog učenja. Algoritmi strojnog učenja (ML) temeljeni na podacima o pozicioniranju igrača tijekom utakmice sve se više koriste za proučavanje taktičkog odlučivanja u elitnom nogometu. Algoritmi strojnog učenja identificiraju specifične obrasce podataka u velikim skupovima podataka izgradnjom *a priori* nepoznatih modela iz podataka. Druga skupina ML metoda koja se ističe u nogometnoj literaturi koristi modeliranje neuronske mreže. Iako se o ovom pristupu u sportskim istraživanjima raspravlja već neko vrijeme, uspješne primjene tek su nedavno postale češće. Fang, Wei i Xu su u u svojem radu uspjeli uspostaviti model neuronske mreže temeljen na dugotrajnom i kratkoročnom pamćenju i eksperimentima su dokazali da algoritam može predvidjeti stopu uspješnosti dodavanja i uspješnih klizećih startova. Ono što je zajedničko ovim metodama je to što se fokusiraju na proučavanje određenog aspekta timske taktike, posebno formiranja tima, međutim, trenutno postoji nedostatak informacija o tome kako kombinirati dobivene informacije u različitim taktičkim domenama (Rein i Memmert, 2016).

2.2.3.3. Predviđanje rezultata

Prvi adekvatan model za predviđanje ishoda utakmice stvorili su 1997. Dixon i Coles. Ovaj model se smatra klasikom i mogao je izdvojiti vjerojatnost postizanja gola u igri, slijedeći Poissonovu distribuciju. Posljednjih godina istraživači su se usredotočili na izravno predviđanje pobjeda, neodlučenih rezultata i poraza, umjesto da pokušavaju predvidjeti golove ili postignute bodove. Implementirani su različiti algoritmi strojnog učenja (ML) kako bi se otkrili najdiskriminirajući čimbenici koji razlikuju pobjednike od gubitnika. Različiti istraživači su se koncentrirali na različite faktore.

Largo-Penas i njegovi suradnici su u svoj sustav rangiranja uključili udarace protivničke momčadi, ubačaje, mjesto odigravanja utakmice, posjed i sposobnosti protivničke momčadi. Harrop i Neville vjeruju da je najbolji faktor predviđanja preciznost dodavanja,

zatim broj udaraca, dodavanja i driblinga (manje to bolje) i mjesto na kojem se utakmica igra. Mao i njegovi suradnici tvrdili su da su karakteristike s najpozitivnijim učinkom udarci u okvir vrata, preciznost udaraca, dueli na zemlji te broj dobivenih duela u zraku (Apostolou i Tjortjis, 2018).

U svome radu iz 2015., Cintia, Rinzivillo i Pappalardo su uspjeli pokazati da se promatranjem određenih indikatora performansi može opisati koje nogometne taktike i strategije koriste momčadi te u kakvoj su one korelaciji s uspjehom momčadi na Svjetskom prvenstvu u Brazilu 2014.

2.3. Nogometni sustavi i aplikacije

Mnogo je nogometnih aplikacija koje se koriste u elitnom nogometu i bez kojih je ta razina nogometa gotovo pa i nezamisliva. Danas je sportski analitičar važno i traženo zanimanje jer svaka momčad koja se natječe treba stručnjake za nogometnu analizu. Mnogo je onih koji su to na vrijeme shvatili i počeli raditi na svojim sistemima i softwearima kako bi se izdignuli iz konkurencije i zauzeli svoje mjesto na tržištu. Postoji nekoliko jako poznatih sustava koji su postali standard i koji su gotovo uvijek u upotrebi u najelitnijim natjecanjima i ligama. Na slici koja slijedi biti će prikazani jedni od najvažnijih sustava koji se koriste u sferi nogometa (Cacho-Elizondo i Alvarez, 2020).

Tablica 1. Nogometni sustavi i aplikacije

Tvrtka (Zemlja podrijetla)	Opis Proizvoda/Usluge
SAP (Njemačka)	Cloud rješenje koje pokreće platforma SAP HANA, fokusirano na sportske prakse (upravljanje timom, planiranje treninga, analiza performansi).
Mediacoach (Španjolska)	Napredni profesionalni alat za video analizu pokreta, integriran s fizičkim i taktičkim podacima.
Match Analysis (SAD)	program koji analizira nogometne videe i statistiku
Wyscout (Italija)	Platforma koja omogućuje ljudima da vide utakmice koje se odvijaju diljem svijeta s računala, tableta ili mobilnog telefona.
GolStats (Meksiko)	Razvoj tehnologije te opskrba sadržajem i informacijama povezana s video zapisima za profesionalne nogometne momčadi i medije (20 milijuna podataka po utakmici s tehnologijom virtualne stvarnosti).
Opta Sports (UK)	Tvrtka za sportsku analitiku sa sjedištem u Engleskoj koja pruža podatke za 30 sportova u 70 zemalja.
Stats (SAD)	Pružatelj rješenja za uključivanje navijača i timsku izvedbu koristeći umjetnu inteligenciju.
Catapult Sports (Australija)	Pružatelj analitike sportskih performansi za podršku igračima, timovima i trenerima.
Real Track Systems (Španjolska)	Sustav za praćenje tjelesne aktivnosti koji koristi Wimu, WiFi uređaj za prikupljanje podataka i SPRO, jednostavnu i fleksibilnu aplikaciju.
Second Spectrum (SAD)	Tvrtka usmjerena na razvoj tehnoloških rješenja strojnog učenja, računalnog vida i proširene stvarnosti za sportsku industriju.

Izvor: Cacho-Elizondo i Álvarez (2020). Big Data in the Decision-Making Processes of Football Teams Integrating a Theoretical Framework, Applications and Reach.

2.3.1. ProZone

2.3.1.1. Razvoj ProZone-a

ProZone je nastao kao ideja Rama Mylvaganama, inženjera koji je bio direktor marketinga za Mars. Mylvaganam je prvi put dobio ideju za ProZone 1996. godine kada je radio za konzultantsku tvrtku za upravljanje i imao ugovor s Derby Countyjem, a kontakt je dobio preko Neila Ramsaya, bivšeg nogometnog agenta. Prva verzija ProZonea bila je prijenosna kabina s 22 masažne stolice, koje je razvio Mylvaganam s lokalnim proizvođačem, a te stolice su emitirale električne impulse i navodno opuštale mišiće igrača i povećavale njihovu fleksibilnost. ProZone je bila skraćenica od *"Professional Zone"*.

Nakon teških početaka došao je trenutak da se sustav probije na velika vrata. Tvrtka je uspjela dogovoriti suradnju sa Alexom Fergusonom, tadašnjim menadžerom Manchester Uniteda. Manchester United pristao je platiti ProZoneu 50.000 funti ako osvoji trofej te godine. Te je sezone United osvojio trostruku titulu: Ligu prvaka, Premier ligu i FA kup, a Prozone je zaradio svoj prvi ček. U svibnju 1999. Prozone je imao dva kupca i nije imao prihoda. Do kolovoza 2000. šest prvoligaških klubova kupilo je njihove usluge. Među tim klubovima bio je i Bolton koji je tada igrao nogomet na nižem rangu. Čelnici ProZonea su to shvatili kao odličnu priliku za dokazivanje svojeg sustava. U tome su i uspjeli. Bolton je pobijedio Preston u doigravanju Championshipa 3-0 i plasirali su se u Premier ligu. Između 2003. i 2007., Bolton je zabilježio uzastopne plasmane među prvih osam u Premiershipu, rekord u dosljednosti bolji samo od prva četiri. Kvalificirali su se za kup UEFA prvi put 2005. i ponovno 2006. Kada je menadžer Sam Allardyce (koji je prvi shvatio vrijednost nogometne analitike i samog sustava) otišao 2007. imali su impresivnih 39 bodova nakon 21 utakmice.

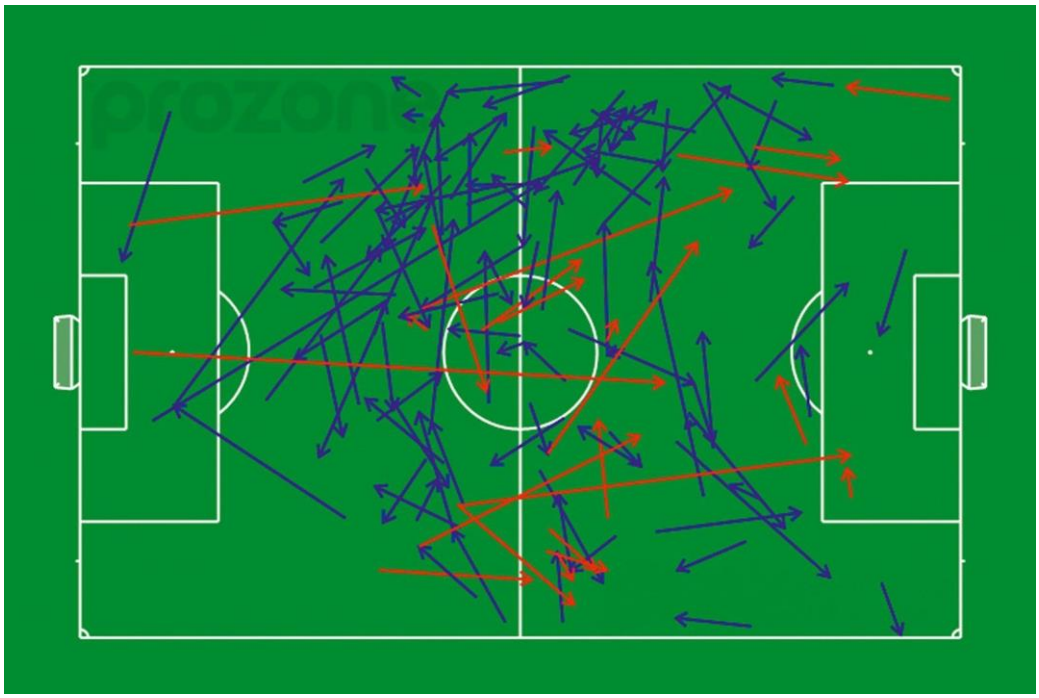
Danas 19 od 20 momčadi Premier lige koristi ProZone. Svaka momčad ima vlastiti tim analitičara performansi i znanstvenika podataka koji traže pokazatelje koji kvantificiraju učinak igrača, događaje koji određuju utakmice i trendove koji karakteriziraju sezone. Oni su znanstvenici koji seciraju najpopularniju igru na svijetu, gledajući podatke iz ProZonea i drugih izvora kako bi razumjeli što određuje pobjedu i poraz. U okruženju Premier lige vrijednom više milijuna dolara, klubovi ne samo da žele konkurentsku prednost, oni je trebaju (Medeiros, 2014).

2.3.1.2. Upotreba ProZone-a

Prozone počiva na mjerenju kretanja. Mjerenje kretanja je koncept koji se koristi u biomehanici za opisivanje kompilacije i analize bilo koje vrste 2D i 3D kretanja. Ti se podaci dobivaju iz video kamera i obrađuju kako bi se kvantificirali kinematički obrasci kretanja. Kretnje se prate pomoću kamera postavljenih oko stadiona. Osam kamera je pozicionirano tako da pokrivaju čitav teren. Nadalje, svako područje terena pokriveno je s najmanje 2 kamere za točnost, okluziju, razlučivost i otpornost. Sve kamere su kablovima spojene natrag na središnju točku i spojene unutar kutije za video distribuciju. Nakon dovršetka automatskog praćenja, izlazni podaci svih 8 kamera automatski se kombiniraju za generiranje jednog skupa podataka.

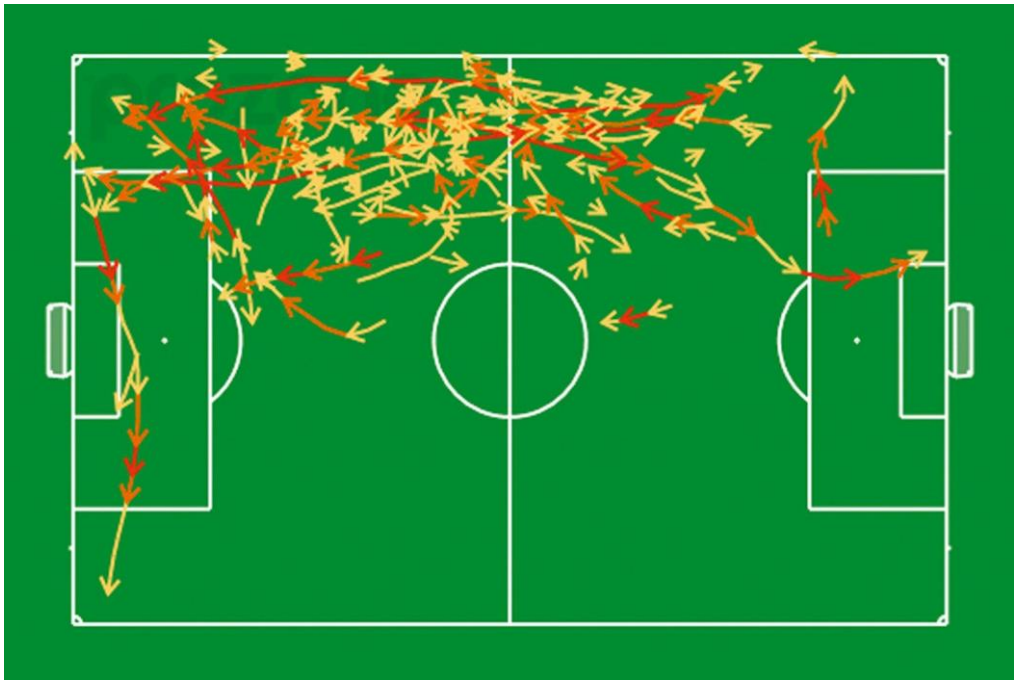
Koordinate slike videa pretvaraju se u koordinate nogometnog terena putem procesa kalibracije (homografija računalnog vida). Prozone koristi linearnu kalibraciju transformacije u 4 točke za mapiranje slike videa u koordinate na terenu, a zatim pročišćava ovu kalibraciju s vlastitim algoritmom od 50 točaka koji eliminira izobličenje s obzirom na optičke pogreške (zakrivljenost leće) i nagib terena.

Slika 3. ProZone - uspješna i neuspješna dodavanja



Izvor: Medeiros, J. (2017). How data analytics killed the Premier League's long ball game. Dostupno na: https://www.wired.co.uk/article/premier-league-stats-football-analytics-prozone-gegenpressing-tiki-taka

Slika 4. ProZone - individualne kretnje igrača na utakmici



Izvor: Medeiros, J. (2017). How data analytics killed the Premier League's long ball game. Dostupno na: https://www.wired.co.uk/article/premier-league-stats-football-analytics-prozone-gegenpressing-tiki-taka

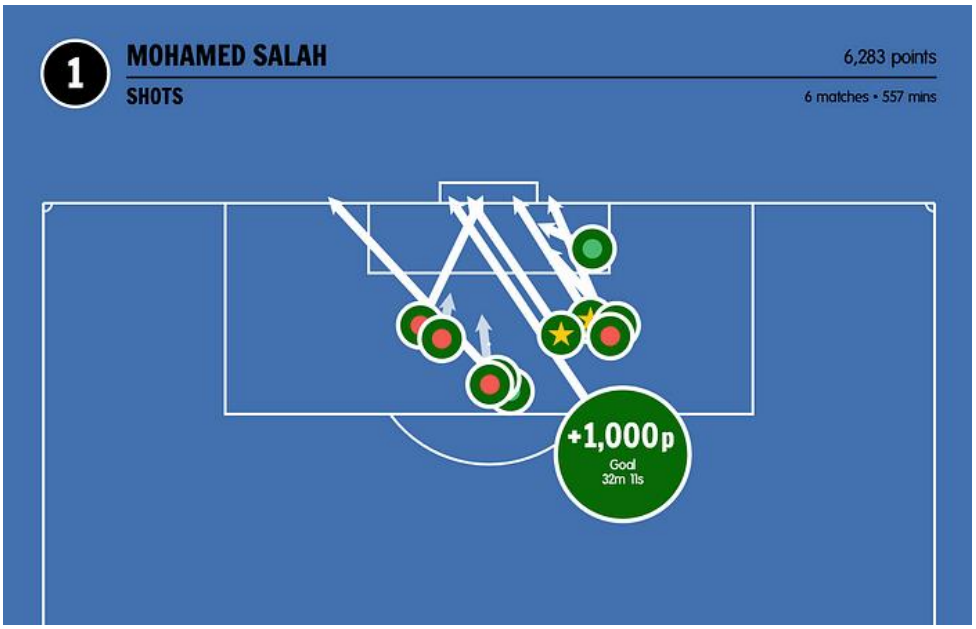
Prednost ove nove tehnologije je mogućnost praćenja svih uključenih u nogometnu utakmicu (igrača i sudaca) i kvantificiranja njihovih obrazaca kretanja. Međutim, glavni nedostaci su visoki troškovi i potreba za instaliranjem više kamera i računalnih mreža, kao i posvećenih operatera za prikupljanje i analizu podataka. Posljednih godina, ovaj sustav se koristi u kombinaciji sa nosivim sustavima za praćenje kako bi se smanjili troškovi a i povećala preciznost sustava. (Valter et al., 2006)

2.3.2. Twelve Football App

Aplikacija Twelve football ocjenjuje igrače uživo tijekom utakmice. Ocjene se temelje na rigoroznom statističkom modelu o tome kako igrači povećavaju (i smanjuju) šanse svoje momčadi za postizanje pogotka. Taj model se temelji na sustavu dvanaest točaka koji je algoritam strojnog učenja koji je razvio profesor iz Uppsale i nogometni matematičar David Sumpter.

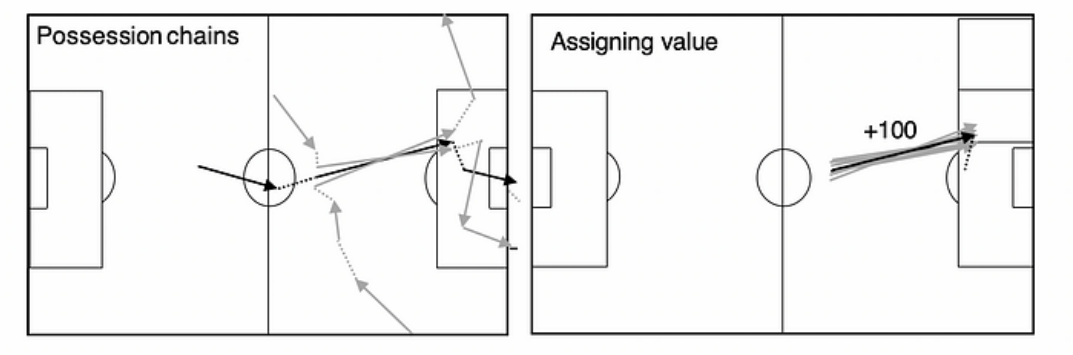
Iako su statističke metode napredne, objedinjujući koncept koji stoji iza modela iznimno je jednostavan i počinje s ciljevima. Najbolja stvar koju možete učiniti u nogometu je postići gol. Postizanje pogotka daje maksimalnih 1000 bodova. Svi ostali bodovi dodijeljuju se u odnosu na postizanje pogotka. Twelve dodjeljuje bodove na temelju toga kako određene napadačke radnje povećavaju šanse momčadi za stvaranje prilike za postizanje zgoditka (dodavanja, driblinzi, slobodni udarci...). Kako bi došli do zaključka, prvo je potrebno prikupiti i usporediti podatke iz prošlih sezona.

Slika 5. Twelve football app



Izvor: Sumpter, S. (2021). Evaluating actions in football using machine learning. Dostupno na: https://soccermatics.medium.com/evaluating-actions-in-football-using-machine-learning-69517e376e0c

Slika 6. Twelve football app

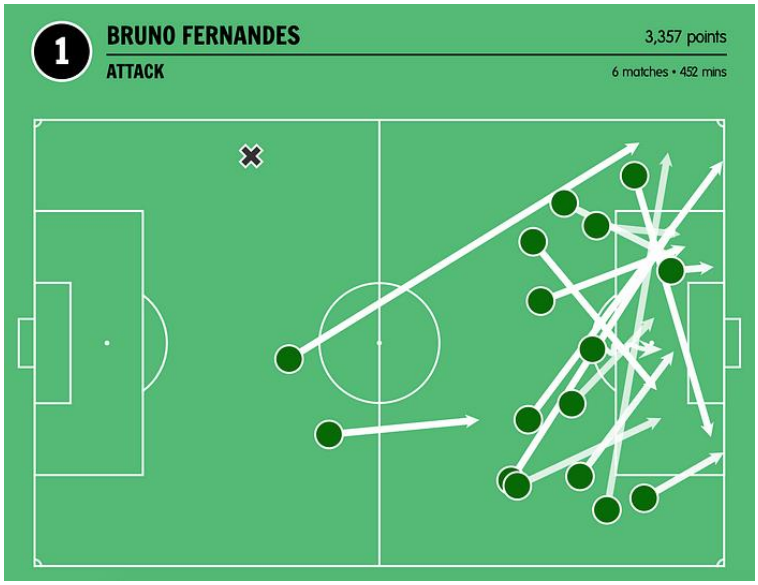


Izvor: Sumpter, S. (2021). Evaluating actions in football using machine learning. Dostupno na: https://soccermatics.medium.com/evaluating-actions-in-football-using-machine-learning-69517e376e0c

Gornje prikazano dodavanje u suparnički šesnaesterac vrijedi +100 bodova jer 10% puta kada se odigra taj pas vodi do pogotka. Znači ako pogodak vrijedi 1000 bodova, dodavanje koje daje 10% gola vrijedi 100 bodova. Važno je da se ova procjena dodavanja ne temelji na 'intuiciji' ili subjektivnom osjećaju o dodavanju već se temelji na statističkom modelu koji uzima u obzir desetke tisuća dodavanja u skupu podataka. Ovdje na scenu stupa strojno učenje. Algoritam aplikacije uči što je opasno a što manje opasno dodavanje.

Slijedi primjer ocjene igrača Manchester United-a Brune Fernandesa na uzorku od šest odigranih utakmica. Dodavanja naprijed-nazad između braniča, koja rijetko dovode do udaraca, obično vrijede samo +2 ili +3 boda. Dodavanja prema naprijed na sredini terena vrijede +20 ili +30 bodova (Sumpter, 2021).

Slika 7. Twelve football app



Izvor: Sumpter, S. (2021). Evaluating actions in football using machine learning. Dostupno na: https://soccermatics.medium.com/evaluating-actions-in-football-using-machine-learning-69517e376e0c

3. OTKRIVANJE ZNANJA U BAZAMA PODATAKA

3.1. Uvod u otkrivanje znanja u bazama podataka

Početkom 1970-ih pohranjivanje podataka ili informacija bilo je vrlo skupo. Ali zahvaljujući napretku u alatima za prikupljanje informacija i internetu proteklih dvadeset i pet godina, dostupna je golema količina podataka u elektroničkom obliku. Zbog pohranjivanja tako velike količine podataka, baze podataka rastu vrlo brzo. Ti podaci mogu biti vrlo korisni u procesu donošenja odluka u bilo kojem području. Potrebno je otkriti samo koji podaci sadrže korisne informacije. To je moguće uz pomoć rudarenja podataka tj. otkrivanja znanja iz baza podataka. Rudarenje podataka je proces izvlačenja korisnih informacija iz velikih količina prethodno nepovezanih podataka (Ahmad, Qamar i Qasim Afser Rizvi, 2015).

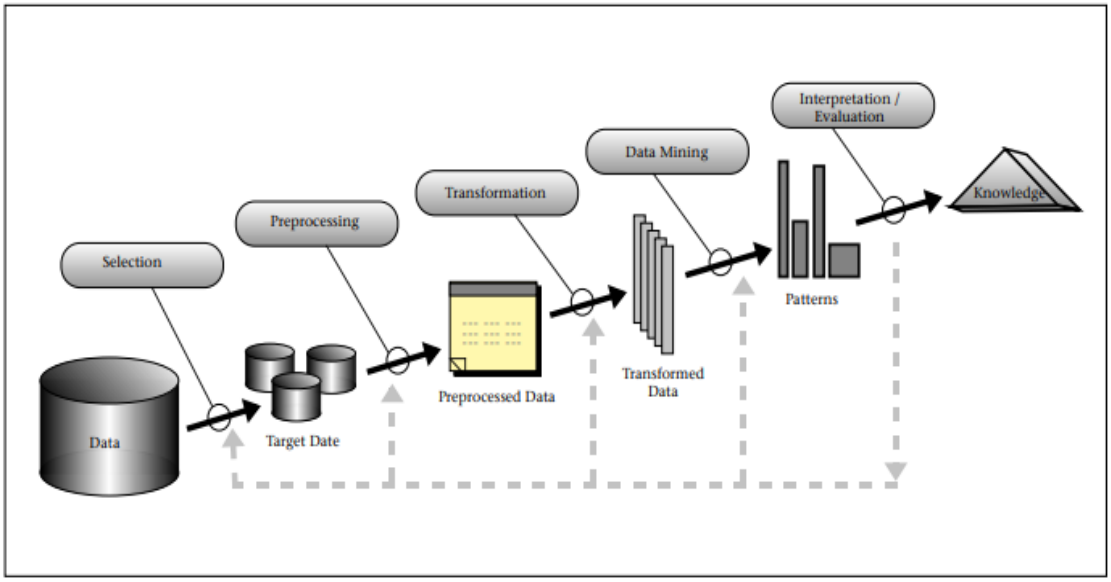
Povijesno gledano, koncept pronalaženja korisnih uzoraka u podacima imao je mnogo naziva, uključujući rudarenje podataka, ekstrakciju znanja, otkrivanje informacija, rudarenje informacija, arheologiju podataka i obradu uzoraka podataka. Izraz rudarenje podataka primarno koriste statističari, analitičari podataka i zajednica upravljačkih informacijskih sustava. Također je popularan u svijetu baza podataka. Izraz otkrivanje znanja u bazama podataka (engl. Knowledge Discovery in Databases) skovan je na prvoj radionici KDD-a 1989. kako bi se naglasilo da je znanje krajnji proizvod otkrića vođenog podacima. Stekao je popularnost u područjima umjetne inteligencije i strojnog učenja (Fayyad, Piatetsky-Shapiro i Smyth, 1996).

Rudarenje podataka je nova metodologija kojom se otkrivaju vrijedni podaci u bazama podataka poduzeća. Metoda se naziva rudarenje podataka, jer se u velikim količinama podataka traže informacije koje „vrijede zlata“. Mnogo je definicija rudarenja podataka, a istaknut ćemo sljedeće: Rudarenje podacima je traženje vrijednih informacija u velikim količinama podataka. Rudarenje podacima je istraživanje i analiza velikih količina podataka pomoću automatskih ili poluautomatskih metoda s ciljem otkrivanja smislenih pravilnosti. Do prije nekoliko godina metoda se prvenstveno razvijala unutar znanstvene zajednice. Tek se nedavno počela primjenjivati u poduzećima i jasno je prepoznato da je korištenje rudarenja podataka neizbježno za poduzeća kako bi stekla komparativnu prednost. Treba istaknuti da je rudarenje podataka više umjetnost nego znanost. Ne postoji recept za uspješno rudarenje podataka koji će nužno dovesti do pronalaska vrijednih informacija (Pejić Bach, 2005).

Prema Pejić Bach rudarenje podataka se provodi kroz nekoliko koraka:

1. Definiranje poslovnog problema
2. Priprema podataka - određivanje potrebnih podataka, transformacija i uzorkovanje, te vrednovanje podataka
3. Modeliranje - odabir metode rudarenja te izrada i vrednovanje modela
4. Implementacija - interpretacija i korištenje rezultata

Slika 8. Koraci u procesu otkrivanja znanja u bazama podataka



Izvor: Fayyad, Piatetsky-Shapiro i Smyth (1996). From Data Mining to Knowledge Discovery in Databases

3.2. Područja primjene otkrivanja znanja iz baza podataka

Jasno je da tvrtke, institucije i organizacije imaju puno podataka, ali najvažnije je kako ih koriste. Bilo u znanosti, marketingu, financijama, zdravstvu, maloprodaji ili bilo kojem drugom području, tradicionalne metode analize podataka temeljno se oslanjaju na jednog ili više analitičara koji su upoznati s podacima i služe kao sučelje između podataka, korisnika i proizvoda. Projekti analitike podataka dopiru do svih poslovnih jedinica stoga je važno da i Zaposlenici u tim jedinicama imaju temeljno znanje o načelima podatkovno-analitičkog razmišljanja jer inače neće stvarno razumjeti što se događa u poslu. Ovaj nedostatak razumijevanja mnogo je štetniji u projektima koji se tiču znanosti o podacima nego u drugim tehničkim projektima, jer znanost o podacima podržava poboljšano donošenje odluka (Provost i Fawcett, 2013). U poslovanju, glavna područja primjene otkrivanja znanja iz baza podataka uključuju marketing, financije (osobito ulaganja), otkrivanje prijevara, proizvodnju, telekomunikacije i internetske agencije (Fayyad, Piatetsky-Shapiro i Smyth, 1996). Međutim primjena se uvelike proširila i diverzificirala pa tako otkrivanje znanja iz baza podataka možemo pronaći u skoro svim sferama što

života što poslovanja. U nastavku poglavlja biti će obrađene i neke druge djelatnosti kako bi se što bolje prikazala diverzificiranost upotrebe.

3.2.1. Bankarstvo

Jedno od područja gdje je rudarenje podataka pronašlo svoju vrijednost je bankarstvo. Banke na dnevnoj bazi prikupljaju ogromne količine podataka o klijentima. Podaci o računima, transakcijama po svakom računu, kreditnim obvezama te demografski podaci vode se za svakog klijenta. Ovi se podaci bilježe u transakcijske baze podataka koje su nužne za tekuće poslovanje. Međutim, transakcijske baze podataka su vrlo velike. Zamislimo da menadžment banke želi utvrditi karakteristike prošlih insolventnih klijenata. Takve podatke možete zatražiti od IT stručnjaka tvrtke, koji uz svakodnevni posao provode dosta vremena pripremajući potrebna izvješća. Dok izvješće stigne na upraviteljev stol, možda će biti prekasno za donošenje odluke. Metoda koja može povećati stopu uspješnosti korištenja transakcijskih baza podataka za poboljšanje poslovanja naziva se rudarenje podataka.

Različite su primjene rudarenja podataka u bankarstvu. Glavni resurs banke su njeni klijenti te se rudarenje podataka provodi radi njihovog pridobivanja, zadovoljstva i zadržavanja. Klijenti se segmentiraju, a korištenjem rudarenja podataka mogu se pronaći segmenti koji su do sada bili zanemareni. Izuzetno je bitno i prepoznati životnu vrijednost klijenta. Rudarenje podataka stvara modele koji predviđaju dugotrajnu vrijednost klijenta kako bi se službenici banaka mogli više usredotočiti na klijente koji su trenutno neprofitabilni, ali bi mogli biti profitabilni u budućnosti. Jedan od modela koji se također koristi u bankarstvu jest model prodaje dodatnih proizvoda postojećim klijentima. Takva vrsta modela određuje vjerojatnost da će klijent banke kupiti dodatni proizvod. Cilj je povećati broj klijenata koji će odgovoriti na ponudu a i povećati kvalitetu odnosa s klijentima. Tako banka povećava profitabilnost jer je trošak prodaje proizvoda postojećim klijentima niži od privlačenja novih klijenata. Također, za banku su bitni i modeli koji predviđaju koja je vjerojatnost da klijent pređe u konkurentsku banku ili smanji potrošnju ukoliko se kamate podignu na normalnu razinu. Za banke je od izuzetne važnosti i model rizika. Uz pomoć tog modela banke određuju kojim klijentima će dati kredit a kojima ne. Bitno je da ne daju kredit osobi koja ga neće moći vratiti (Pejić Bach, 2005).

3.2.2. Obrazovanje

Neke zemlje pružaju besplatno školovanje od prvog razreda osnovne škole pa sve do završetka fakulteta. Stoga mnogi učenici srednjih škola nastavljaju svoje obrazovanje u nadi da će postati akademski građani. Zbog velikog broja studenata postalo je teško pružiti visoku kvalitetu predavanja i svima. Kao rezultat toga, mnogi studenti ne diplomiraju u roku. Rudarenje podataka može otkriti koji su problemi koji ometaju studente u ispunjavanju fakultetskih obaveza. Sve veća primjena tehnologije u obrazovanju svakodnevno generira velike količine podataka, što je postalo meta mnogih diljem svijeta.

Tri su izvora podataka u obrazovanju:

- tradicionalna predavanja – predavanja u obrazovnim institucijama, podaci se prikupljaju putem upitnika i promatranja učitelja i profesora
- e-učenje – podaci se prikupljaju pomoću raznih platformi za učenje
- pametni sustavi za učenje i prilagođeni sustavi učenja – specijalizirani su za prikupljanje podataka pojedinca (učenika ili studenta)

Mnogo je prostora za razvoj kada se spominje rudarenja podataka u obrazovanju. To su mnogi shvatili stoga se to područje ubrzano razvija i ima prednost što obuhvaća nove algoritme i tehnike razvijene u raznim polja rudarenja podataka i strojnog učenja. Rudarenje obrazovnih podataka pomaže u razvoju metoda za izdvajanje zanimljivih, razumljivih, korisnih i novih informacija za bolje razumijevanje učenika i njihovog okruženja u kojem uče.

Metode koje su razvijene omogućuju identificiranje rizičnih učenika, utvrđivanje prioriteta za potrebe učenja različitih grupa studenata, povećanje stope diplomiranja, učinkovito ocjenjivanje rada škole ili fakulteta, maksimiziranje resursa institucija i optimiziranje reforme predmetnog kurikuluma (Algarni, 2016).

3.2.3. Zdravstvo

Zdravstvena skrb uključuje detaljan proces dijagnosticiranja, liječenja i prevencije bolesti, ozljeda i drugih fizičkih i psihičkih oštećenja ljudi. Zdravstvena industrija raste velikom brzinom u većini zemalja. Može ju se smatrati industrijom bogatom podacima jer stvara velike količine podataka, uključujući elektroničke medicinske zapise,

administrativna izvješća i druge podatke za usporedbu. Međutim, ti se podaci u zdravstvu ne koriste na zadovoljavajuć način (Jothi, Rashid i Husain, 2015). Mnogo je odnosa skriveno u tako velikoj zbirci podataka, poput odnosa između podataka o pacijentima i broja dana koliko su bili hospitalizirani (Ahmad, Qamar i Qasim Afser Rizvi, 2015).

Tradicionalne metode pretvaranja podataka u znanje oslanjaju se na ručnu analizu i interpretaciju. Na primjer, u zdravstvenoj industriji uobičajeno je da stručnjaci redovito (npr. kvartalno) analiziraju trenutne trendove i promjene u zdravstvenim podacima. Stručnjaci potom podnose izvješće s detaljnom analizom zdravstvenoj organizaciji sponzoru i zatim to izvješće postaje temelj za buduće odluke i planiranje upravljanja zdravstvenom skrbi. S naprednim istraživanjima u zdravstvu dostupne su goleme količine podataka, ali glavna poteškoća je pretvaranje postojećih informacija u korisnu praksu (Fayyad, Piatetsky-Shapiro i Smyth, 1996). Rudarenje podataka ima ogroman potencijal da zdravstvenim sustavima omogući učinkovitiju i djelotvorniju upotrebu podataka te da se posljedično s tim poboljša skrb i smanje troškovi (Ahmad, Qamar i Qasim Afser Rizvi, 2015). Veliko bogatstvo se skriva i u mogućnosti predviđanje raznih bolesti, kao i za pomoć liječnicima u dijagnozi pri donošenju kliničke odluke (Jothi, Rashid i Husain, 2015).

Rudarenje medicinskih podataka pruža prilike za značajna otkrića. U nastavku se navode najčešći primjeri:

- otkriće novih činjenica – povezanost između lijeka i nuspojava ili potencijalno otkriće novog lijeka
- organiziranje velikih repozitorija medicinskih podataka - uvid u ono što bi se sljedeće moglo dogoditi (npr. pandemija)
- predviđanje budućnosti u različitim situacijama i scenarijima - predviđanju trendova (kao u slučaju epidemija) i prijetnji, kao i prilika (Kudyba, 2018)

Velik je broj pacijenata sa sličnim dijagnozama koji se mogu istraživati i te dobivene informacije pomažu u otkrivanju uzroka određenih bolesti i njenih ishoda. Nastavak toga je propisivanje preventivnih mjera pacijentima i spašavanje njihovih života. Nakon što su ranije navedeni općeniti primjeri otkrivanja znanja iz baza podatka, u ostatku teksta navedeni su konkretniji slučajevi. U jednom istraživanju korištena su asocijativna pravila, konkretnije apriori algoritam, uz pomoć kojeg su izrađena pravila za zdrave i bolesne

ljude. Na temelju tih pravila otkriveni su čimbenici koji uzrokuju srčane probleme kod muškaraca i žena. Logistička regresija koristi se za procjenjivanje relativnog rizika od raznih bolesti, kao što su dijabetes, angina, moždani udar itd. Isto tako, stablo odlučivanja se koristi za predviđanje stope mortaliteta kod pacijentica s rakom dojke. Naravno, održavanje zdravstvenog sustava košta. Koriste se klasifikacijske tehnike za predviđanje troškova liječenja za medicinske usluge, koji svake godine brzo rastu i postaju velika briga. Također vezano uz optimalizaciju troškova, hijerarhijsko klasteriranje osigurava učinkovito korištenje bolničkih resursa i poboljšava usluge skrbi za pacijente u zdravstvu (Ahmad, Qamar i Qasim Afser Rizvi, 2015).

3.2.4. Poljoprivreda

Otkrivanje znanja u bazama podataka koristi se i u poljoprivredi za prikaz statističkih informacija o stanju tla, klimatskim uvjetima, prošlim prinosima usjeva, vladinim strategijama, svim informacijama o pesticidima, gnojivima. Nekoliko je različitih primjena otkrivanja znanja iz baza podataka u poljoprivredi. Neke od njih se odnose na predviđanje vremenskih uvjeta i onečišćenje zraka u atmosferi. Neke tehnike rudarenja podataka često se kombiniraju s tehnikama temeljenim na GPS-u za proučavanje karakteristika tla i klasifikacije tla (Mankar i Burange, 2014.). Te tehnike koje se koriste za predviđanje vremenskih uvjeta i klasifikaciju tla nazivamo *K-Means* i *K-Nearest*. Jedna od tehnika je i korištenje neuronskih mreža za otkrivanje koje su jabuke zrele a koje još nisu ili su prezrele (Kaur, Gulati i Kundra, 2014).

Predviđanje cijena je također jedno vrlo važno pitanje za svakog poljoprivrednika jer bi trebao znati kolika je očekivana cijena njegovih usjeva. Proteklih godina predviđanja cijena temeljila su se na iskustvima poljoprivrednika s određenim usjevima i poljima. Pretpostavimo li da poljoprivrednici bilježe svoje predviđene cijen iz prošlosti, ti podaci bi se mogli pretvoriti u elektronički oblik i iskoristiti za klasifikaciju budućih cjenovnih predviđanja (Kaur, Gulati i Kundra, 2014).

3.2.5. Telekomunikacije

Telekomunikacijska industrija bila je jedna od prvih industrija koja je usvojila tehnologiju rudarenja podataka. To je vjerojatno zato što telekomunikacijske kompanije

često generiraju i pohranjuju velike količine visokokvalitetnih podataka, imaju vrlo velike korisničke baze i rade u okruženju koje se brzo mijenja i vrlo je konkurentno (Weiss, 2010). Spomenuti podaci uključuju podatke o detaljima poziva koji opisuju pozive preko telekomunikacijske mreže, mrežne podatke koji opisuju status hardverskih i softverskih komponenti u mreži te korisničke podatke koji opisuju telekomunikacijske korisnike (Weiss, 2005).

Potrebno je razumijeti podatke koji se obrađuju. U telekomunikacijskoj industriji oni se dijele na podatke prikupljene tokom poziva, podatke mreže i podatke o klijentima. Podaci koji su prikupljeni tijekom poziva uglavnom sadrže informacije o pozivatelju i primatelju poziva i vremenu kada je poziv obavljen i koliko je trajao. To nisu podaci podložni rudarenju jer nisu korisni podaci o svakom individualnom pozivu već je bitan sami klijent. Takve podatke treba najprije sortirati pa tek onda tražiti korisne informacije. Što se tiče mrežnih podataka, tu su najbitniji podaci o problemima s mrežom. Uz razvoj rudarenja podataka stvorene su metode i tehnike koje automatski rješavaju manje mrežne probleme teleoperatera bez pomoći ljudske ruke. I za kraj preostaju podaci o klijentu. Ti podaci će uključivati informacije o imenu i adresi i mogu uključivati druge informacije kao što su plan usluge i informacije o ugovoru, kreditni rezultat, prihod kućanstva i povijest plaćanja. Podaci o korisnicima često se koriste zajedno s drugim podacima kako bi se poboljšali rezultati. Na primjer, podaci o korisnicima obično se koriste za dopunu podataka o detaljima poziva kada se pokušava identificirati telefonska prijevara (Weiss, 2005).

Razloge primjene otkrivanja znanja iz baza podataka u telekomunikacijskoj industriji možemo podijeliti u 3 kategorije: marketing, otkrivanje prijevara te izolacija i predviđanje grešaka na mreži. Zbog zaista velike konkurencije, teleoperateri moraju znati kako zadovoljiti svoje korisnike. Koriste rudarenja podataka za identifikaciju korisnika, zadržavanje korisnika i maksimiziranje dobiti dobivene od svakog korisnika. Drugi veliki izazov s kojim se suočavaju su prevare. Postoje prevare vezane uz uzimanje određene usluge s ciljem da se ona ne plati nikada dok je druga vrsta krađa identiteta korisnika. Najčešća tehnika otkrivanja prevare je uspoređivanje korisnikovog trenutnog ponašanja sa starim pozivnim navikama. Preostaje još samo objasniti na koji način se izoliraju i predviđaju greške na mreži. Razvijen je alat za rudarenje podataka koji otkriva koja je povezanost između više različitih problema i povezuje ih na jedan izvor te pamti taj

obrazac i u mogućnosti je predvidjeti ostale greške kad se javi jedna koja je povezana s njima (Weiss, 2010).

3.3. Važnost primjene otkrivanja znanja iz baza podataka

Poduzeća koriste podatke kako bi stekla konkurentsku prednost, poboljšala učinkovitost i pružila vrijednije usluge klijentima. Podatke koje bilježimo o našem okruženju koristimo za izgradnju teorija i modela a zatim dobivene rezultate koristimo za izgradnju naših poslovnih planova. Budući da računala omogućuju ljudima da prikupe više podataka nego što ih možemo probaviti, prirodno je okrenuti se računalnim tehnikama koje nam pomažu otkriti smislene obrasce i strukture u velikim količinama podataka. Algoritmi i tehnologije za velike podatke omogućuju otkrivanje novih poslovnih uvida i donošenje informiranih odluka na temelju podataka. Iskorištavanje znanja skrivenog u velikim podacima poboljšava organizacijsku izvedbu i konkurentsku prednost (Pejić Bach et. al., 2020). Stoga tehnika otkrivanja znanja u bazama podataka pokušava riješiti stvarni problem koji nam je svima donijelo doba digitalnih informacija a to je preopterećenost podacima.

Tehnike rudarenja podataka koriste se u mnogim područjima poput matematike, kibernetike, genetike, marketinga, itd. Razne matematičke tehnike koriste se kako bi se izvukli podaci relevantni za određenu organizaciju (npr. što ljudi vole, kupuju, žele) te se pronalaze nova znanja koja se koriste u kampanjama, reklamama, raznim poslovnim jedinicama, opskrbnim lancima, itd. U suštini je to zapravo proces razvrstavanja velikih skupova podataka, identificiranja obrazaca i uspostavljanja veza te rješavanja problema analizom podataka. Iz toga proizlaze mogućnosti za otkrivanje skrivenih obrazaca i odnosa između podataka koji se zatim koriste za efikasnije poslovanje. Prateći taj proces možemo predvidjeti buduće trendove i proizvode, čime povećavamo prodaju i zadovoljstvo kupaca. Upravo se tu mogu primjetiti glavne prednosti i važnost otkrivanja znanja iz baza podataka.

Slika 9. Primjer analize tržišne košarice

Izvor: Lozić i Šimec (2020). Rudarenje podataka

Otkivanjem znanja iz baza podataka pokušava se steći prednost odnosno vrsta moći u fazi rudarenja. Postoje dvije vrste moći: prediktivna i opisna. Korištenjem metoda za proricanje budućih ili još nepoznatih vrijednosti dolazi se se prediktivne moći. U opisnu moć ubrajaju se zanimljivi podaci i interpretabilni uzorci koji opisuju podatke koje pronalazimo.

Uistinu je mnogo različitih primjera benfita rudarenja podataka. Predikcijska moć koju se stječe iz informacija o prošlim prodaja i ponašanju kupaca omogućuje kreiranje predikcijskih modela za buduće prodaje, nove usluge i proizvode. Kao što je i ranije spomenuto u poglavlju pod nazivom „Bankarstvo“, u financijskoj industriji koriste se tehnike rudarenja podataka za izradu rizičnih modela za detekciju prijevare u transkacijama, kao i kod posudbi i hipoteka. U velikoj, serijskoj proizvodnji metode se koriste radi poboljšanja sigurnosti proizvoda, procjene korištenosti proizvoda, identifikaciju problema sa kvalitetom, poboljšanje proizvodnje te za upravljanje lanca opskrbe. Također, velika važnost rudarenja podacima pridaje se i u marketingu gdje se koristi za povećanje zadovoljstva kupaca i za kreiranje targetiranih reklamnih kampanja. Jedan od primjera su i trgovine. Uz pomoć prikupljenih znanja, trgovine, uz sve ostale

aspekte prodaje, prilagođavaju izgled dućana kako bi poboljšale iskustvo kupaca i povećale prodaju određenih proizvoda. Na kraju krajeva, svi mogu profitirati od otkrivanja znanja iz baza podataka ovisno o poslovanju i potrebama, nebitno bio to pojedinac ili velike korporacije i državne strukture (Lozić i Šimec, 2020).

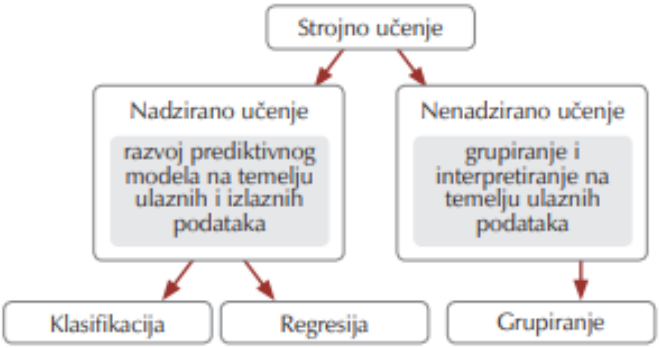
3.4. Prikaz metoda za otkrivanje znanja iz baza podataka

Iz mnogih izvora dolaze podaci. Ti podaci su integrirani i pohranjeni u neku zajedničku bazu podataka. Dio podataka se zatim uzima i oblikuju u format kako bi se pripremili i proslijedili algoritmu za rudarenje podataka koji proizvodi izlaz u obliku pravila ili neke druge vrste uzoraka. Na kraju se podaci tumače kako bi dali ono što se traži od početka a to je novo i potencijalno korisno znanje. (Bramer, 2020). Kao što spominje i Pejić Bach tako i Bramer opisuje rudarenje podataka više kao umjetnost nego egzaktnu znanost.

Prvo valja krenuti od činjenice da postoje dvije vrste podataka i više različitih vrsta baza podataka s kojima se postupa na različite načine. Prva vrsta podataka nazivaju se označeni. Imaju takav naziv jer kod njih postoji posebno određen atribut a cilj je iskoristiti ulazne podatke za predviđanje vrijednosti tog atributa za instance koje ćemo tek vidjeti. Rudarenje takve vrste podataka poznato je kao nadzirano učenje. Druga vrsta podataka su neoznačeni podaci te kod njih ne postoje posebno označeni atributi (Bramer, 2020). Izraz koji se koristi za rudarenje ovakvih tipova podataka je nenadzirano učenje. Što se tiče različitih vrsta baza podataka njih dijelimo na: relacijske, transakcijske, objektno orijentirane, prostorne i aktivne baze podataka, kao i globalne informacijske sustave (Liao, Chu i Hsiao, 2012).

Postoje dva glavna cilja rudarenja podataka: predviđanje i opis. Predviđanje se može povezati s nadziranim učenjem dok je cilj nenadziranog učenja opis. Do predviđanja se dolazi korištenjem varijabli ili polja u skupu podataka za predviđanje nepoznatih i budućih varijabli koje nas interesiraju. Kod opisa je stvar jednostavnija jer je koncentracija usmjerena na pronalaženje obrazaca koji opisuju podatke koje mogu protumačiti ljudi (Kantardžić, 2011). Ovi ciljevi se postižu različitim metodama i tehnikama rudarenja podataka koje će biti detaljnije pojašnjene u nastavku ovoga poglavlja.

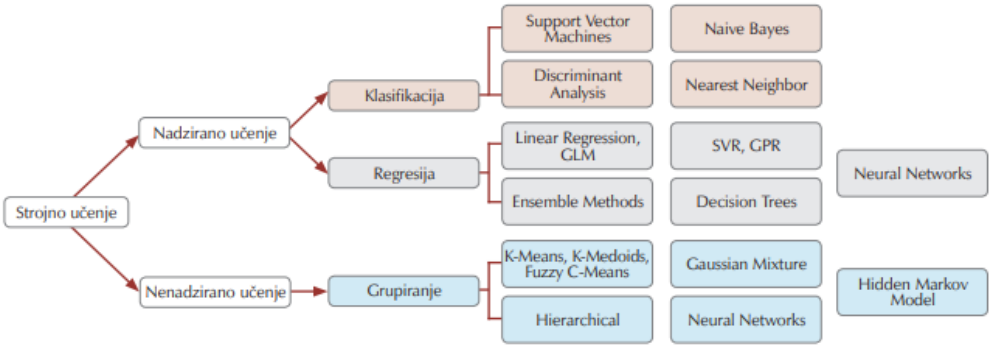
Slika 10. Osnovna podjela strojnog učenja



Izvor: Bolf; N. (2021). Osvježimo znanje: Strojno učenje. Dostupno na: https://hrcak.srce.hr/file/382926

Svaka od ovh osnovnih metoda dijeli se na još nekoliko različitih podmetoda koje će se samo spomenuti i prikazati na Slici 8 u nastavku rada jer nisu toliko bitne za samo istraživanje koje se provodi na kraju rada.

Slika 11. Podjela metoda strojnog učenja



Izvor: Bolf; N. (2021). Osvježimo znanje: Strojno učenje. Dostupno na: https://hrcak.srce.hr/file/382926

3.4.1. Nadzirano učenje

U ovom načinu učenja, algoritam koristi poznati skup ulaznih podataka (ulaz modela) i poznati skup izlaznih podataka (izlaz modela) i pokušava osposobiti model za predikciju (Bolf, 2021). Nadzirano učenje je zapravo prediktivno rudarenje podataka jer proizvodi

model sustava opisan danim skupom podataka, izražen kao izvršni kod te se koristi za izvođenje procjena, predviđanja, klasifikacije i drugih sličnih zadataka (Kantardžić, 2011).

Nadalje, iz nadziranog učenja proizlaze tehnike kojima rudarimo podatke. O tome koja tehnika će biti korištena ovisi o tipu atributa. Atributi mogu biti kvalitativni ili kvantitativni. Ako je atribut kvalitativan tehnika koja se primjenjuje je klasifikacija, ukoliko je kavntitativan koristi se regresija.

3.4.1.1. Klasifikacija

Klasifikacija je među njačešćim metodama rudarenja podataka. Klasifikacijom se ulazni podaci razvrstavaju po unaprijed određenim kategorijama ili klasama. Za razliku od nekih matematičkih metoda klasifikaciju je jednostavno objasniti i shvatiti pomoću nekoliko primjera iz svakodnevnog života. U medicini, može se koristiti za klasifikaciju tumora u benigni ili maligni (Bolf, 2021). Koristi se i u obrazovanju kada se dodijeljuju ocjene učenicima i studentima (Bramer, 2020). U poslu, kada se elektronska pošta razvrstava u željenu ili neželjenu (spam). U glasanju, kada se prikupljaju mišljenja građana o određenoj temi. Primjera je zaista mnogo.

3.4.1.2. Regresija ili predviđanje

Kao što se klasifikacijom pokušava predvidjeti kvalitativna vrijednost, tako se regresijom pokušava predvidjeti numeričku vrijednost, kao što je dobit tvrtke ili cijena dionice. Ako se ponavalja medicina kao primjer, regresija se može korisiti u slučaju kada se želi predvidjeti jerojatnost da će pacijent preživjeti (Fayyad, Piatetsky-Shapiro and Smyth, 1996). Također, regresijom se mogu predviđati sastav i kvaliteta proizvoda, kao i potražnja za istim ili promjene temperature (Bolf, 2021). Vrlo popularna tehnika modeliranja kada se govori o regresiji je neuronska mreža koja se temelji na modelu ljudskog neurona. Neuronska mreža ima više ulaza tojest podražaja koje koristi za predviđanje jednog ili više izlaza (Bramer, 2020).

3.4.2. Nenadzirano učenje

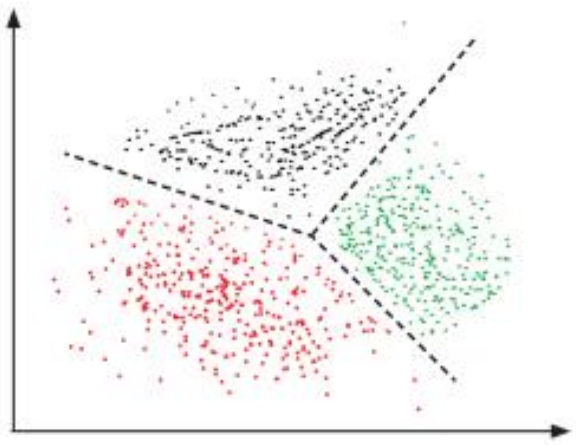
Podaci koji nemaju posebno označene atribute nazivaju se neoznačenim. Rudarenje podataka neoznačenih podataka poznato je kao učenje bez nadzora. Ovdje je cilj jednostavno izvući najviše informacija koje možemo iz dostupnih podataka.

3.4.2.1. Grupiranje (klasteriranje)

Grupiranje je tehnika koja se najčešće vezuje uz nenadzirano učenje. Grupiranjem (klasteriranjem) pronalaze se skrivene grupe (klasteri) ili obrasci. Ukoliko su bogate podacima mogu se međusobno preklapati ili biti hijerarhijski prikazane a ukoliko nisu onda su međusobno isključive i iscrpne (Fayyad, Piatetsky-Shapiro i Smyth, 1996). U procesu grupiranja algoritmi sortiraju podatke po stavkama koje su im zajedničke ili slične. Damo li za primjer, banka može grupirati svoje klijente prema dobi, prihodu, visini kredita itd. (Bramer, 2020) Klasteriranje pomaže i pri analizi tržišta i prepoznavanju objekta (Bolf, 2021).

Postoje i različite metode koje proizlaze iz grupiranja, neke od njih su medijan, klasterska analiza, *K-Means* grupiranje te grupni prosjek. Rezultati se prikazuju kao dijagram drva, tabla članstva ili *„Icicle Plot“* (Lozić i Šimec, 2020).

Slika 12. K-Means algoritam



Izvor: Bolf; N. (2021). Osvježimo znanje: Strojno učenje. Dostupno na: https://hrcak.srce.hr/file/382926

4. ISTRAŽIVANJE NOGOMETA KORIŠTENJEM METODA OTKRIVANJA ZNANJA U BAZAMA PODATAKA

4.1. Opis izvora podataka

Kao izvor podataka u ovome radu koristit će se baza podataka o nogometnim timovima prikupljena s internetske stranice *„Kaggle“*. Baza podataka preuzeta je u .xlsx formatu pod nazivom „*Football Teams“* kao što prikazuje Tablica 2. Radi se o bazi podataka koja sadrži informacije o nogometnim klubovima (Barcelona, Bayern, Manchester City itd.) iz pet najboljih nogometnih liga na svijetu a to su Premier Liga, Serie A, Bundesliga, Ligue 1 i LaLiga. Dostupni podaci su iz sezone 2021.-2022. Slijedi Tablica 1 koja prikazuje spomenutu bazu podataka. Zbog zaista velikog broja instanci u radu je prikazan samo dio tablice ali dovoljno da čitatelj može zaključiti o čemu se radi.

Tablica 2. Baza podataka "Football Teams"

Team	Tournament	Goals	Shots pg	yellow_cards	red_cards	Possession%	Pass%	AerialsWon	Rating
Manchester City	Premier League	83	15.8	46	2	60.8	89.4	12.8	7.01
Bayern Munich	Bundesliga	99	17.1	44	3	58.1	85.5	12.9	6.95
Paris Saint-Germain	Ligue 1	86	15	73	7	60.1	89.5	9.5	6.88
Barcelona	LaLiga	85	15.3	68	2	62.4	89.7	10.6	6.87
Real Madrid	LaLiga	67	14.4	57	2	57.7	87.7	11.8	6.86
Manchester United	Premier League	73	13.8	64	1	54.5	84.8	14.5	6.85
Juventus	Serie A	77	15.7	76	6	55.4	88.3	11.4	6.85
Aston Villa	Premier League	55	13.7	63	4	49.1	78.6	19.4	6.84
Borussia Dortmund	Bundesliga	75	14.6	43	1	57.5	85.5	12.8	6.84
Atletico Madrid	LaLiga	67	12.1	100	0	51.8	83.1	14.4	6.84
Atalanta	Serie A	90	16.3	66	3	53.5	83.5	16.8	6.84
Chelsea	Premier League	58	14.6	49	3	58.6	87	15.2	6.83
Liverpool	Premier League	68	16	40	0	59	85.7	14.3	6.82
AC Milan	Serie A	74	14.7	80	4	51.4	84	15.2	6.82
Lille	Ligue 1	64	12.8	67	2	52.6	83.5	15.8	6.82
Tottenham	Premier League	68	11.7	53	2	51.3	81.8	16.4	6.81
Napoli	Serie A	86	17	71	3	54.1	87	11.1	6.81
Leicester	Premier League	68	12.8	61	0	53.2	82.1	16.2	6.8
Wolfsburg	Bundesliga	61	14.1	56	3	51	78	16.9	6.8
Inter	Serie A	89	14.5	59	2	52	87	11.8	6.8
Lyon	Ligue 1	81	16.1	60	10	53.6	84.7	14.3	6.8
RB Leipzig	Bundesliga	60	16	57	0	57.3	83.2	18.6	6.78
Leeds	Premier League	62	13.7	61	1	55.1	80.8	14.5	6.77
West Ham	Premier League	62	12.3	48	3	44.5	77.8	19.9	6.77
Everton	Premier League	47	10.5	59	2	47.3	81.4	17.7	6.73
Bayer Leverkusen	Bundesliga	53	13	58	0	57.3	84.4	13.1	6.73
Eintracht Frankfurt	Bundesliga	69	13.2	80	1	52.4	79.6	17.9	6.73
Monaco	Ligue 1	76	12.8	74	7	54.2	82.7	16.5	6.73
Roma	Serie A	68	14.3	84	3	51.5	84.5	12.1	6.71
Sevilla	LaLiga	53	12.1	75	2	58.7	86.2	16.6	6.7
Borussia M.Gladbach	Bundesliga	64	13.4	61	2	51.5	82	15.3	6.7

Izvor: internetska baza podataka *Kaggle*

U tablici 3 koja slijedi biti će prikazani svi atributi, njihovi opisi, formati i modaliteti.

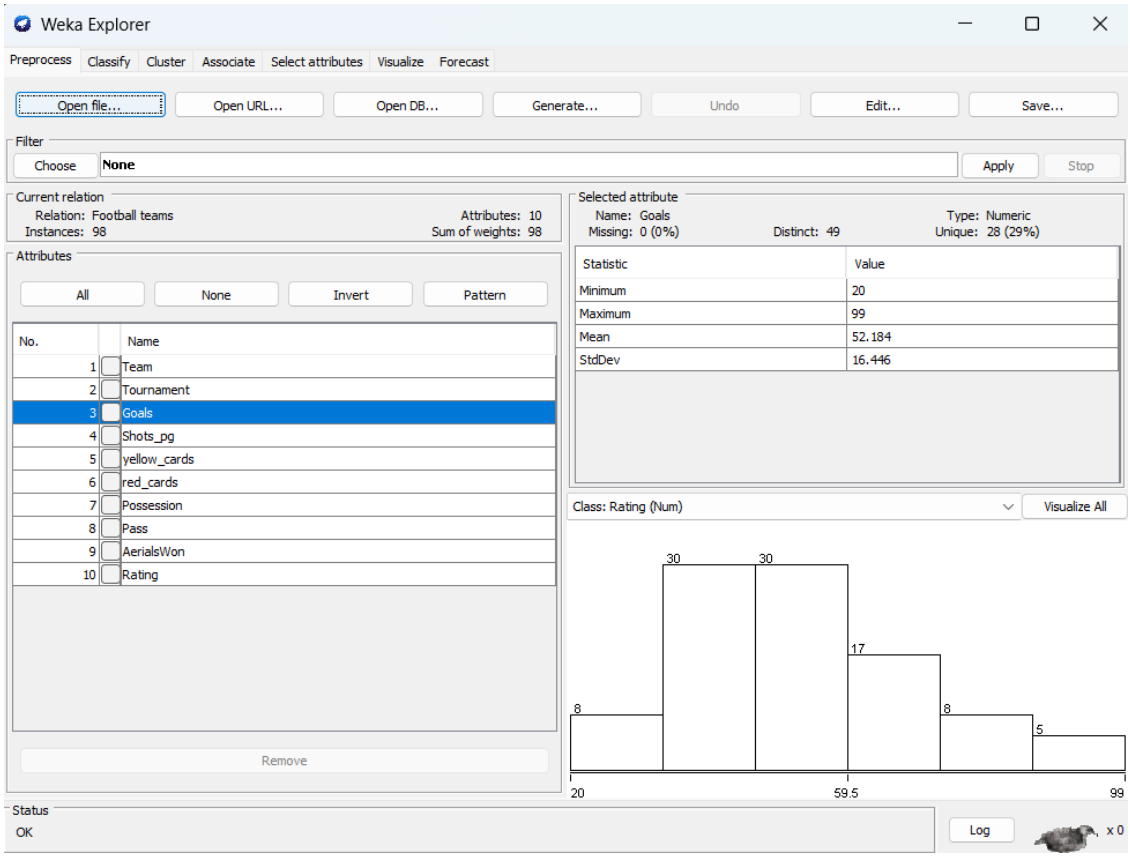
Tablica 3. Atributi

Naziv atributa	Opis atributa	Format atributa	Modalitet atributa
Team	Naziv momčadi	Nominalni	Arsenal, Chelsea, AC Milan, PSG, itd.
Tournament	Natjecanje u kojem momčad sudjeluje	Nominalni	Premier League, Bundesliga, Serie A, LaLiga, Ligue 1
Goals	Ukupan broj zabijenih golova u sezoni	Numerički	
Shots pg	Prosječan broj udaraca na gol	Numerički	
Yellow cards	Ukupan broj žutih kartona u sezoni	Numerički	
Red cards	Ukupan broj crvenih kartona u sezoni	Numerički	
Possession (%)	Posjed lopte	Numerički	
Pass (%)	Postotak točnih dodavanja	Numerički	
Aerials Won	Osvojeni zračni dueli	Numerički	
Rating	Ocjena momčadi	Numerički	

Izvor: izrada autora

Za provođenje istraživanja koristit će se program Weka. Kako bi to bilo moguće potrebno je pretvoriti bazu podataka iz .xlsx formata u .csv format. Tek tada može se pristupiti podacima u Weki. Pretvaranje se radi na način da se .xslx datoteka otvori u programu Notepad, potom se podaci koji su razdvojeni točka-zarezom razdvajaju samo zarezom, decimalni brojevi moraju biti napisani točkom a umjesto navodnih znakova se ostavlja prazno polje. Nakon tih izmjena podaci su kopirani u Microsoft Excel i spremljeni u .csv format. Zatim je .csv datoteka pretvorena u .arff format koji je standardan za Weku. Tako spremljena datoteka sada se može otvoriti u programu i njeni podaci se mogu rudariti. Otvorena datoteka prikazana je na slici ispod.

Slika 13. „Football Teams“ u Weki

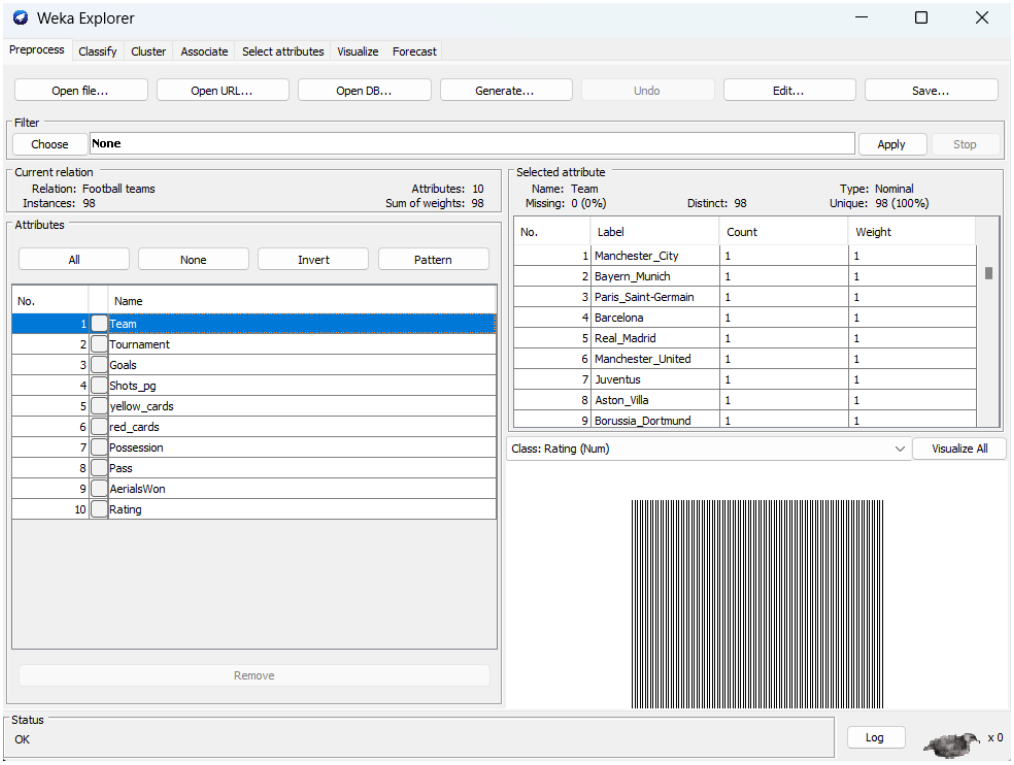


Izvor: autorsko istraživanje

Nakon što su u tablici navedeni i u Weki prikazani svi atributi koji će sudjelovati u istraživanju, slijedi detaljnije objašnjenj svakoga od njih.

Prvi atribut u ovoj bazi podataka je *Team*. Pod atribut *Team* ulaze sve nogometne momčadi koje ulaze u ovu analizu. Taj atribut je nominalan te postoji 98 različitih nogometnih ekipa.

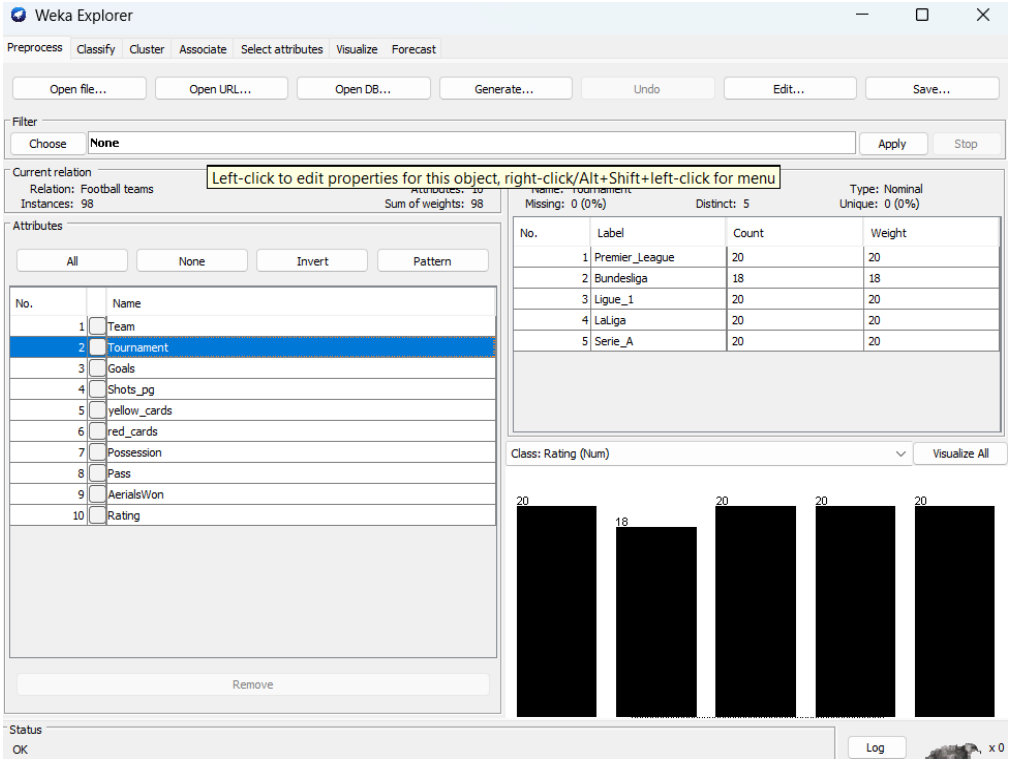
Slika 14. Atribut *Team*



Izvor: autorsko istraživanje

Drugi atribut o kojem se govori je *Tournament.* Moglo bi se reći da je bi bolji naziv za ovaj atribut bio *League* pošto se radi o ligama u kojima se ranije navedene ekipe natječu. Izraz turnir se češće upotrebljava za natjecanja koja traju kraće i gdje sudjeluje manje ekipa ali iz različitih zemalja dok u ligi uglavnom sudjeluju ekipe iz iste zemlje. Atribut *Tournament* je također nominalni i broji 5 različitih modaliteta za 5 najboljih liga na svijetu: Premier League (Engleska), LaLiga (Španjolska), Bundesliga (Njemačka), Ligue 1 (Francuska) te Serie A (Italija). Svaka liga broji 20 instanci osim Bundeslige u kojoj sudjeluje 18 momčadi.

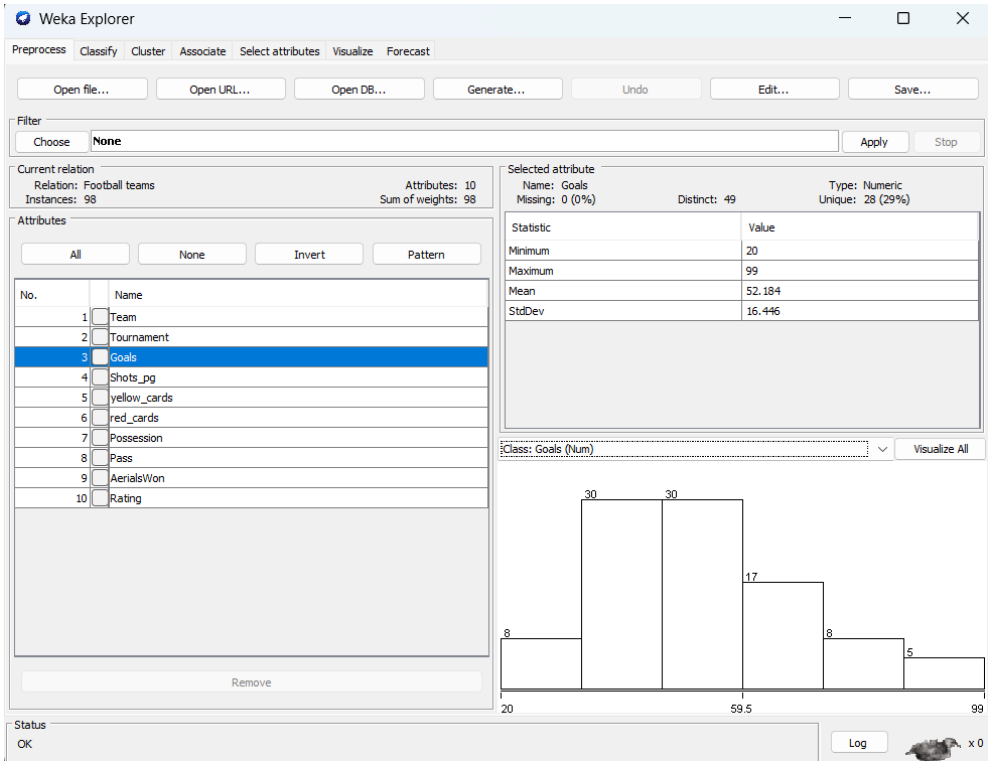
Slika 15. Atribut *Tournament*



Izvor: autorsko istraživanje

Treći aribut po redu je atribut *Goals* koji je nakon dva nominalna atributa numerički. Taj atribut nam govori koliko je golova momčad zabila u sezoni. Najmanji broj pogodaka u ovoj bazi podataka je 20 i njih je zabila ekipa Sheffield Uniteda, a najveći broj pogodaka zabio je Bayern Munchen i iznosi 99. Prosječan broj golova je 52.184 a prosječno odstupanje od prosjeka iznosi 16.446.

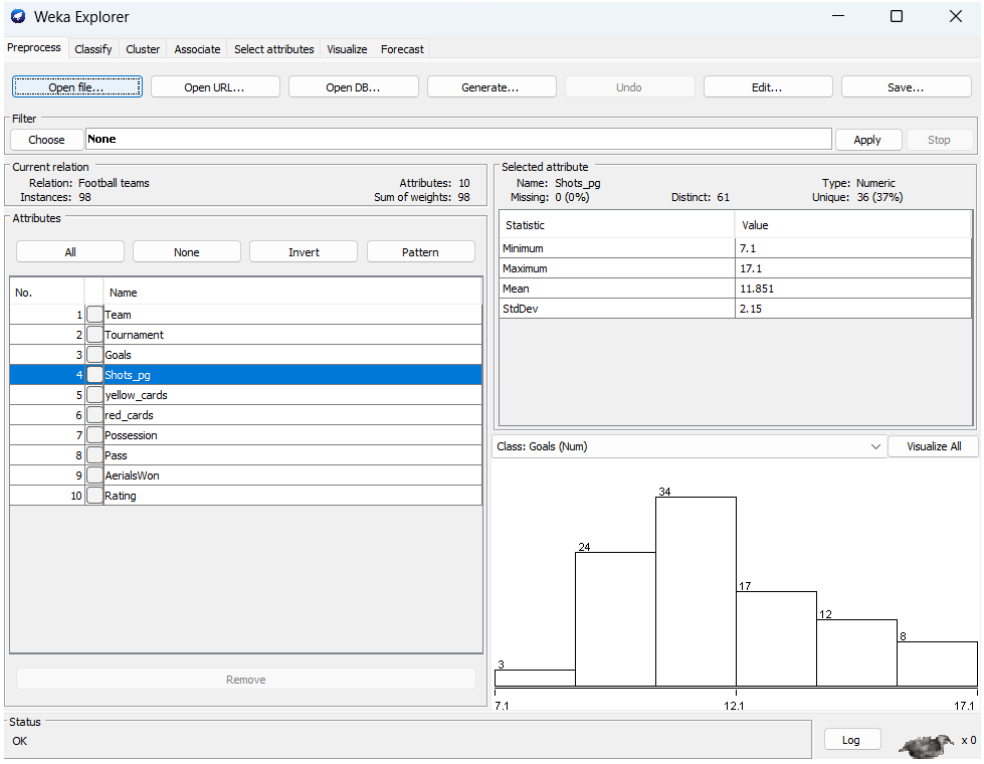
Slika 16. Atribut *Goals*



Izvor: autorsko istraživanje

Nakon atributa *Goals* slijedi atribut koji je usko vezan uz njega a to je *Shots per game*. Shots per game odnosno broj udaraca po utakmici je također numerički atribut te mu minimalna vrijednost iznosi 7.1 a maksimalna 17.1. Prosječan broj udaraca na gol u sezoni 2021.-2022. iznosio je 11.851 a standardna devijacija 2.15.

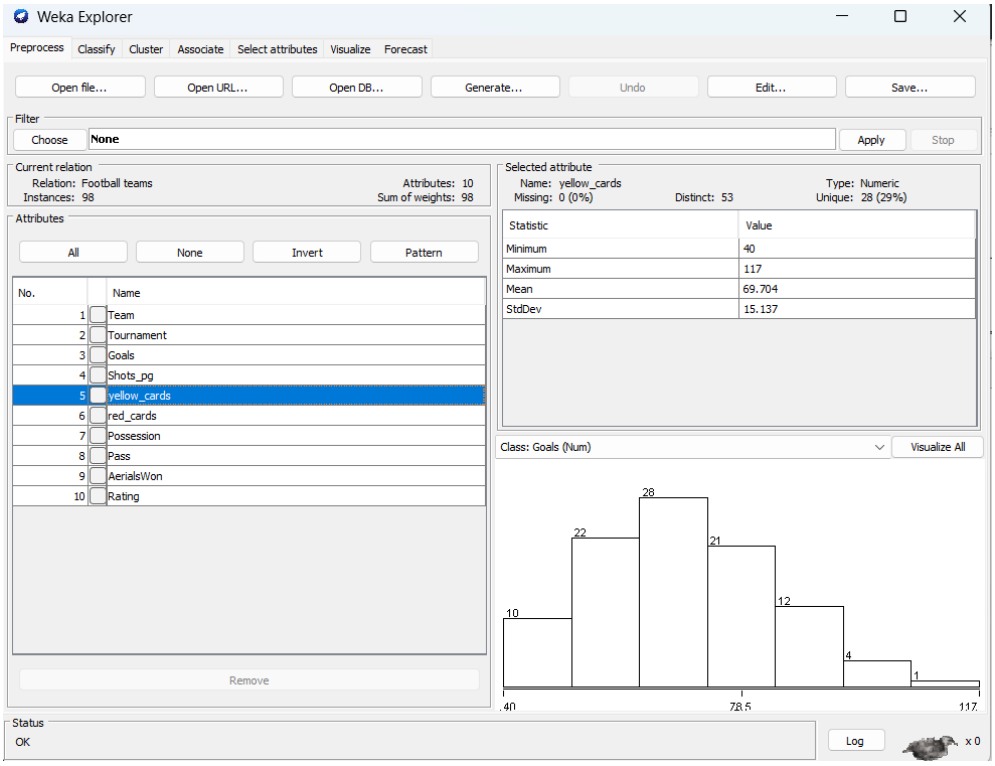
Slika 17. Atribut *Shots per game*



Izvor: autorsko istraživanje

U nastavku poglavlja slijedi atribut *Yellow cards* tj. žuti kartoni koji je po prirodi numerički atribut. Najmanji broj žutih kartona iznosi 40, taj broj je ostvarila momčad Liverpoola dok je momčad s najviše žutih kartona bila Getafe. Prosječan broj podijeljenih žutih kartona bio je 69.704 a standardna devijacija 15.137.

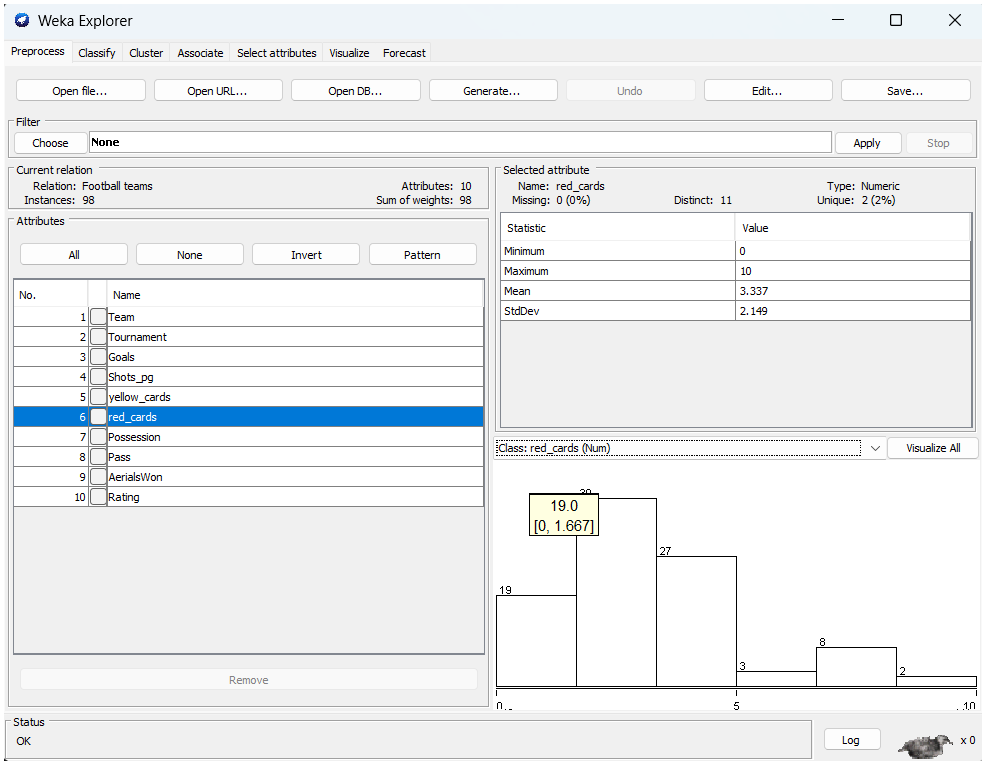
Slika 18. Atribut *Yellow cards*



Izvor: autorsko istraživanje

Nakon žutih kartona slijede po logici i crveni kartoni pod nazivom *Red cards* u bazi podataka koji imaju puno veći utjecaj na samu igru stoga ih je značajno manje od žutih. Crveni kartoni se dodijeljuju samo kad se radi o teškim prekršajima te igrač koji dobije crveni karton biva isključen iz igre. Dva kumulativna žuta kartona također rezultiraju crvenim kartonom. Najmanja vrijednost ovog numeričkog atributa je 0 a najveća 10. Prosječna vrijednost 3.337 a prosječno odstupanje od prosječne vrijednosti 2.149.

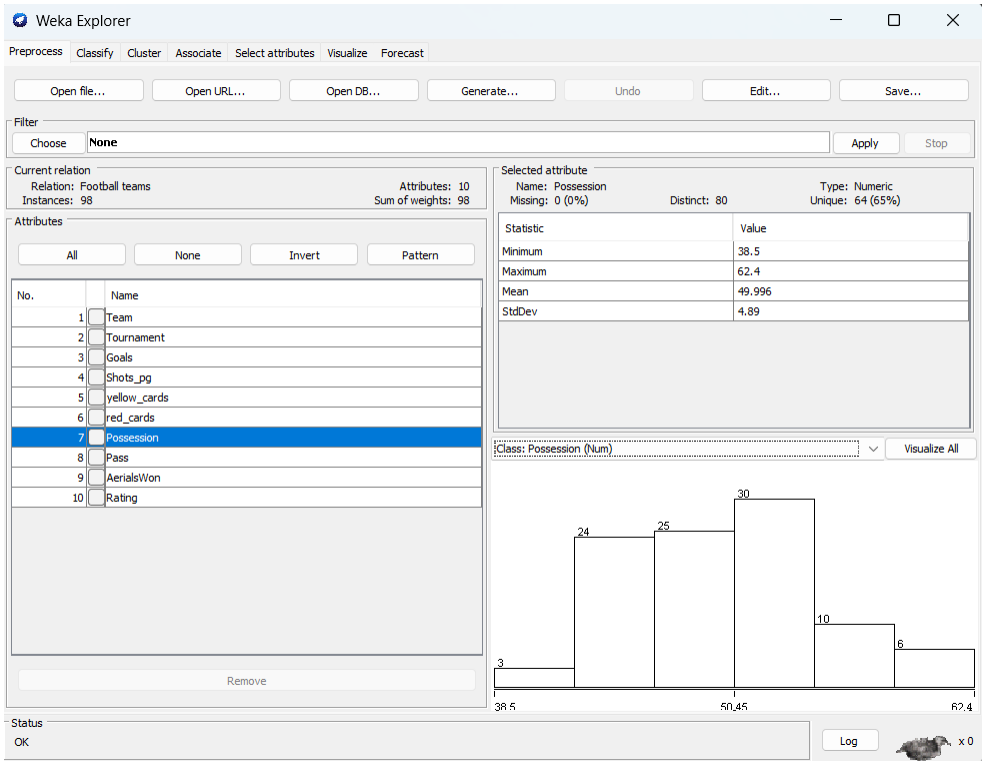
Slika 19. Atribut *Red cards*



Izvor: autorsko istraživanje

Atribut *Possession* je numerički atribut ali izražen u postotku. On predstavlja postotak vremena u kojem je momčad bila u posjedu lopte za vrijeme utakmice. U ovoj bazi podataka su uzete naravno sve utakmice odigrane u sezoni. Najmanji postotak posjeda je imala ekipa Cadiza i on je iznosio 38.5%. Najviše provedenog vremena s loptom imala je Barcelona, čak 62.3%. Središnja vrijednost iznosi 49.996% što i logika nalaže jer u utakmici sudjeluju dvije ekipe a standardna devijacija koja je uvijek izražena u apsolutnoj vrijednosti je 4.89.

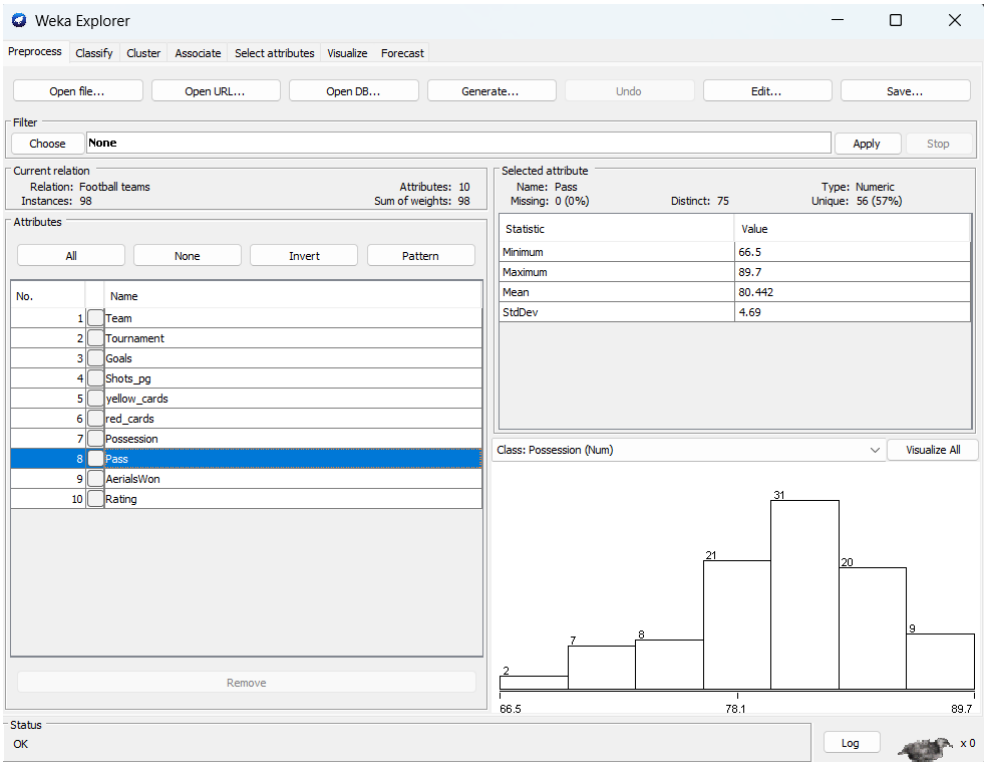
Slika 20. Atribut *Possession*



Izvor: autorsko istraživanje

Još jedan numerički atribut izražen u postotku jest *Pass*, a on odražava postotak uspješnih dodavanja od broja ukupnih dodavanja u utakmici. Prema *Football Teams* bazi podataka najveći postotak uspješnih dodavanja jest 89.7% a najmanji 66.5%. Medijan iznosi 80.442% a prosječno odstupanje od medijana u apsolutnom iznosu je 4.69.

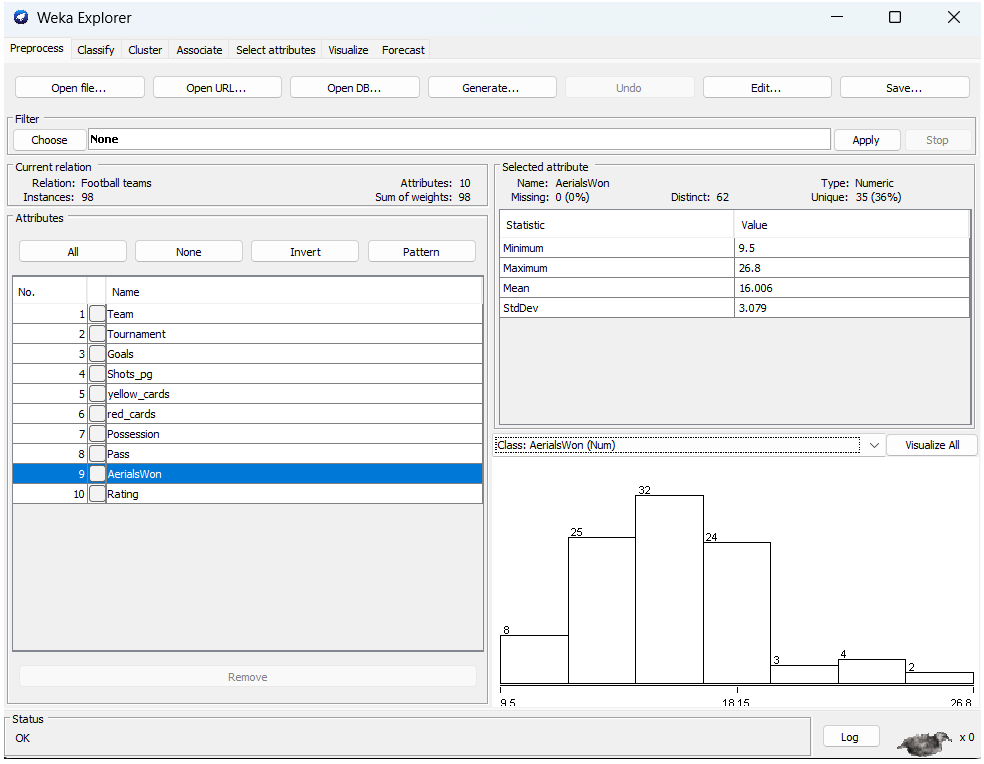
Slika 21. Atribut *Pass*



Izvor: autorsko istraživanje

Atribut *Aerials Won* označava prosječan broj osvojenih zračnih duela po utakmici. *Aerials Won* je također numerički atribut s maksimalnom vrijednosti od 26.8 dok je prikazana minimalna vrijednost 9.5. Prosječan broj osvojenih duela je 16.006. Vrijednost standardne devijacije je 3.079.

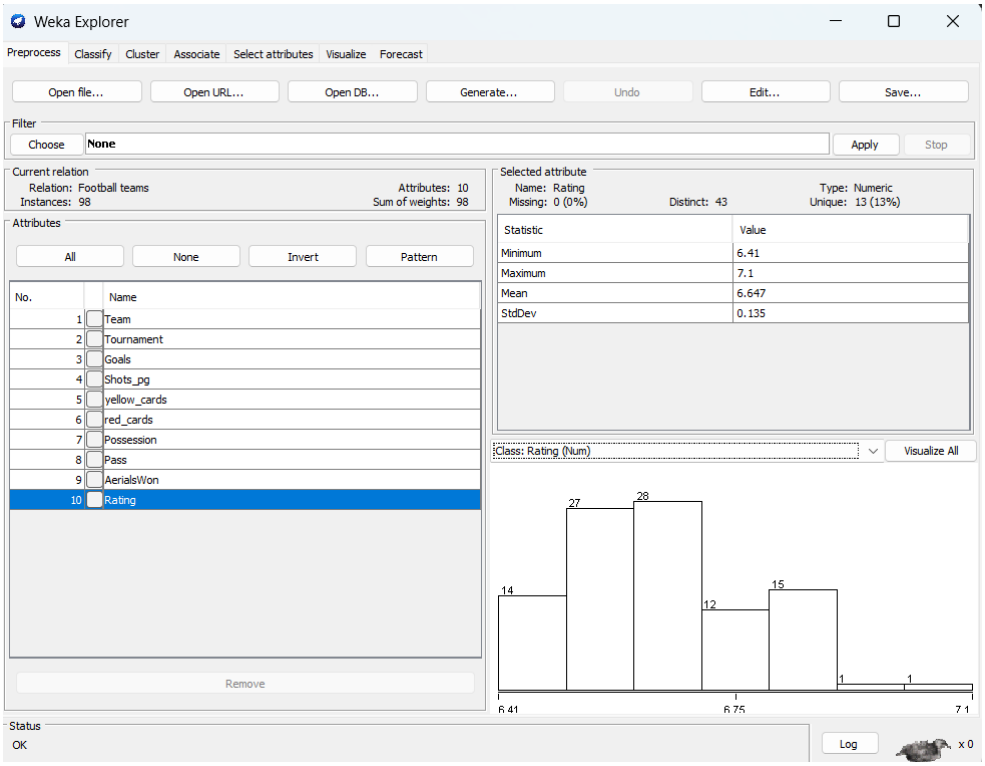
Slika 22. Atribut *Aerials Won*



Izvor: autorsko istraživanje

Posljednji odnosno 10. atribut je *Rating*. *Rating* je numerički atribut i to je ocjena dana momčadi s obzirom na njezine izvedbe u sezoni 2021-2022 te se kreće u intevalu od 1 do 10. Momčadi su ocjenjene od strane analitičkih stručnjaka te je njihova ocjena uključena u bazu podataka koja se istražuje. Najveća dana ocjena među svim ekipama je 7.1 dok je najniža tj. najlošija momčad zaradila ocjenu 6.41

Slika 23. Atribut *Rating*



Izvor: autorsko istraživanje

4.2. Metodologija istraživanja

Metoda koja je odabrana za otkrivanja znanja iz baze podataka *Football Teams* je klaster analiza. Klaster analiza je jedan od najboljih načina za razumijevanje velike količine podataka jer ih organizira u logične skupine. Klaster analiza predstavlja statističku metodu koja se koristi za grupiranje objekata na temelju njihovih sličnih karakteristika (Zoroja i Pejić Bach, 2016). Klaster analizom se pokušavaju pronaći određene pravilnosti u podacima i uzročno-posljedične veze te se na temelju toga segmentiraju u klastere. Nakon provođenja analize dobiti će se odgovori na razna pitanja koja se najčešće postavljaju a tiču se nogometne igre.

U jednom dijelu istraživanja koncentracija će biti na otkrivanju nogometnih stilova u različitim ligama tojest zemljama. U nekim ligama momčadi igraju dosta slično dok se u nekima vidi diverzificiranost. Nadalje, istražit će se koja liga je zapravo najkompetitivnija

a gdje su razlike između momčadi jako velike. Otkrivanjem te činjenice dobija se razlog zašto i koja liga je najgledanija u svijetu.

U drugom dijelu istraživanja stavit će se naglasak na otkrivanje najboljih ekipa između 98 navedenih. Te najbolje ekipe činile bi posebnu ligu o čijoj se realizaciji govori u budućnosti. Donijet će se zaključci što dovodi do toga da su baš te ekipe najbolje te koje su njihove sličnosti i razlike.

Klaster analiza počinje odabirom varijabli za analizu, nakon čega slijedi odabir postupka klasteriranja koji upravlja načinom formiranja klastera. Za potrebe ovog istraživanja odabran je postupak klasteriranja k-srednjih vrijednosti. Prema Hartiganu i Wongu, to je postupak koji dijeli "M točaka u N dimenzija u K klastera tako da je zbroj kvadrata unutar klastera minimiziran". Postupak iterativno promatra srednje vrijednosti klastera na način da se promatranja istovremeno premještaju u klaster s najbližom sredinom. Klaster analiza K-srednjih vrijednosti nastavlja ponovno izračunavati srednje vrijednosti klastera i premješta opažanja u onoliko koraka koliko je potrebno dok se nijedno opažanje ne premjesti u drugi klaster (Pejić Bach, Jaklič i Vugec, 2018).

Kao što je ranije spomenuto u poglavlju, prvo će se istražiti karakteristike različitih svjetskih nogometnih liga. Zbog toga je važno uključiti atribut *Tournament* u analizu. Svi atributi osim atributa *Team* su uključeni u klaster analizu. Kod provedbe klasteriranja potrebno je krenuti od odabira optimalnog broja klastera. Broj se određuje na način da se postepeno povećava broj klastera sve dok se *Within cluster sum of square errors* smanjuje za značajnu razliku. Jednom kada razlika postane zaista neznatna, a poslije se opet između dva sljedeća klastera naglo poveća a zatim nastavlja opet usporavati ili rasti otkrivena je prijelomna točka ili takozvana točka infleksije. Ona označava optimalan broj klastera za analizu.

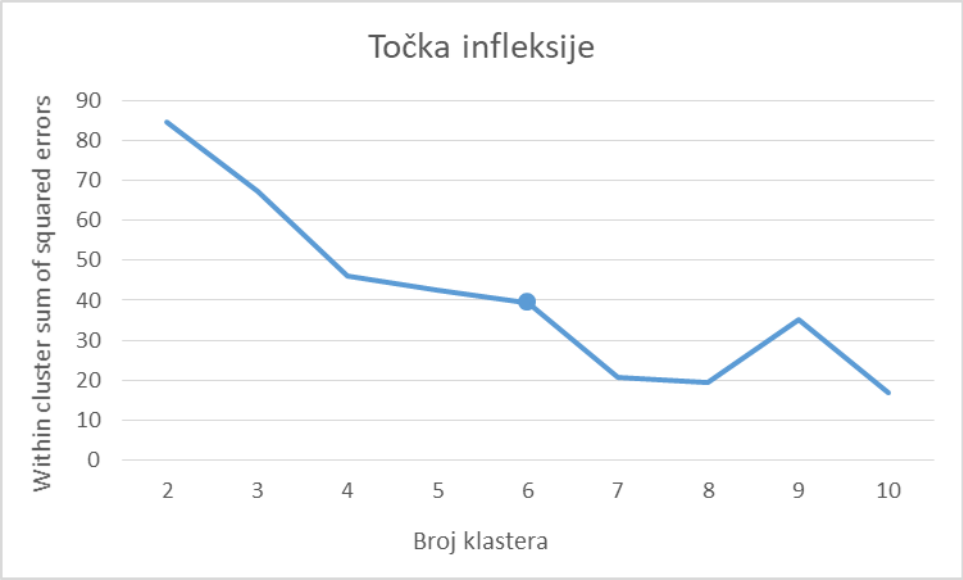
Tablica 4 u nastavku prikazuje odnos između broja klastera i njihovih *Within cluster sum of square errors.* Iz navedene tablice izvučen je graf koji prikazuje točku infleksije.

Tablica 4. Točka infleksije

Broj klastera	Within cluster sum of squared errors
2	84,78363
3	67,20334
4	46,21878
5	42,62118
6	39,39507
7	20,76567
8	19,48794
9	35,15461
10	16,82013

Izvor: izrada autora

Graf 1. Točka infleksije

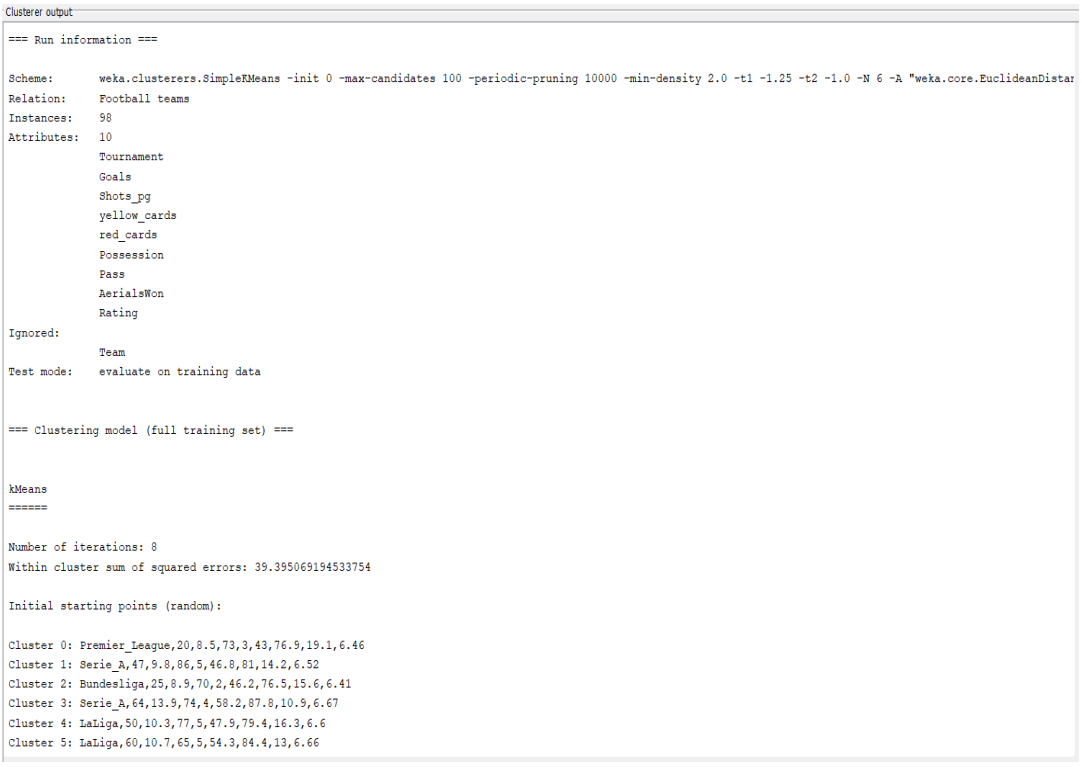


Izvor: izrada autora

Iz priložene Tablice 4 i Grafa 1 možemo vidjeti kako u slučaju klaster analize podataka iz baze podataka *Football Teams* broj *within sum of squared errors* značajno pada sve do broja klastera koji iznosi 4 te zatim pad usporava do broja 6 i poslije njega opet raste s povećanjem broja klastera na 7. To dovodi do zaključka kako je idealan broj klastera u ovom slučaju 6.

Na slici koja slijedi biti će prikazana kartica *Cluster* na kojoj su vidljive sve informacije o klaster analizi kao što su odabrani atributi, ignorirani atributi, within cluster sum of squared errors itd.

Slika 24. Informacije o prvoj klaster analizi

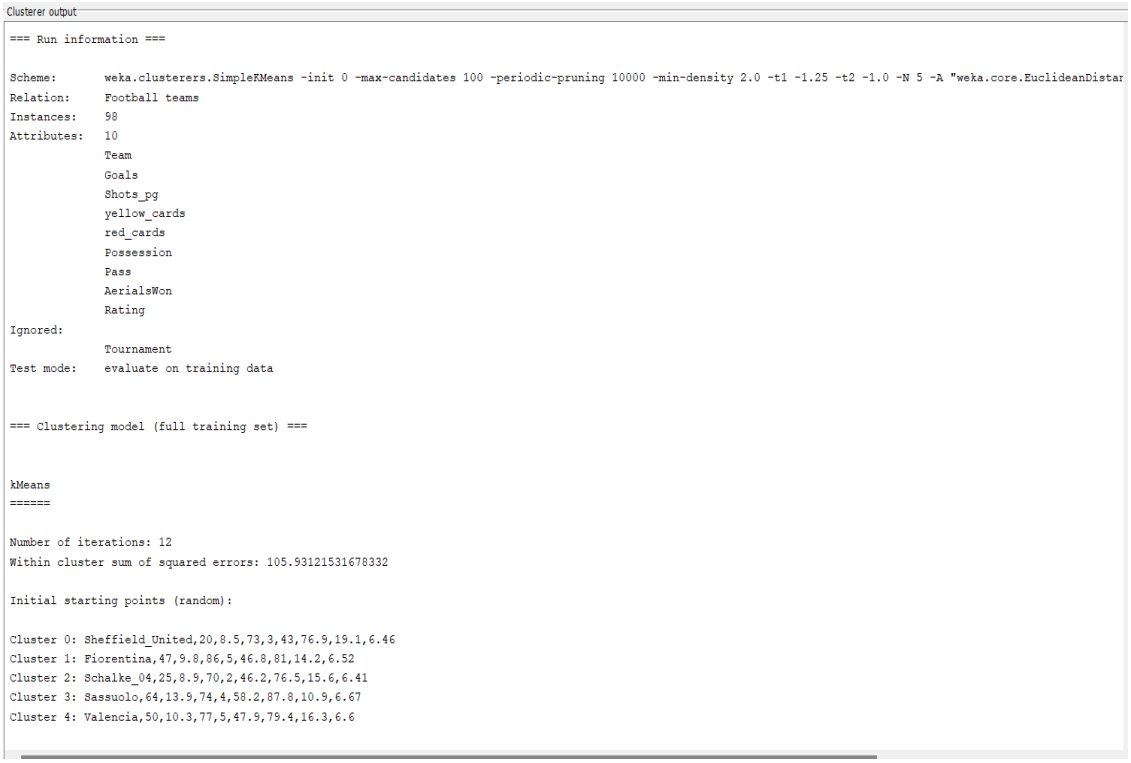


Izvor: autorsko istraživanje

U drugome dijelu istraživanja postojeće lige ćemo pretvoriti u novih 5 liga. Na taj način biti će jednostavnije uvidjeti koje momčadi su najuspješnije a koje su najmanje uspješne i zašto je to tako. Usporedba po atributima će biti pregledno prikazana te će neupućeni u nogomet moći sagledati atribute i zaključiti koja obilježja dovode do uspješnosti. Pošto

informacije o trenutnim natjecateljskim ligama nisu potrebne za ovu analizu, atribut Tournament izostavit će se u ovoj analizi a uključiti atribut *Team*.

Slika 25. Informacije o drugoj klaster analizi



Izvor: autorsko istraživanje

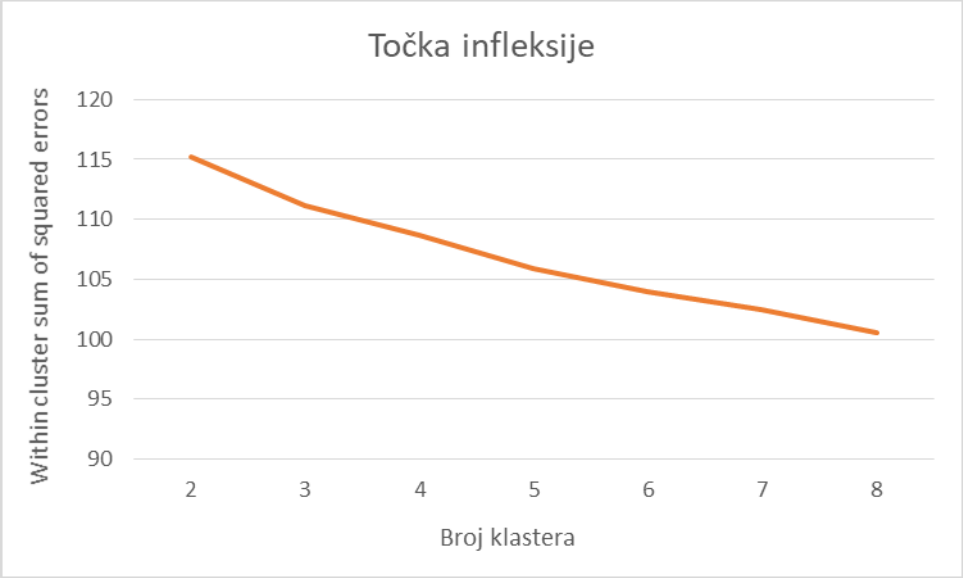
Odabrani broj klastera u ovoj analizi je proizvoljan i iznosi 5. Postoje 2 razloga zašto je odabran broj 5. Prvi razlog tome je što trenutno postoji u svijetu 5 vodećih velikih natjecateljskih liga sa sličnim brojem momčadi i želi se postići barem približan i realan broj ekipa u svakom od novih klastera koji su zamišljeni da predstavljaju novi poredak. Drugi razlog je taj što je u ovoj situaciji poprilično teško odrediti točku infleksije jer se broj *Within sum of squared errors* kontinuirano smanjuje za približno iste iznose što prikazuju tablica i graf u nastavku rada.

Tablica 5. Točka infleksije druge klaster analize

Broj klastera	Within cluster sum of squared errors
2	115,22144
3	111,19481
4	108,69376
5	105,93122
6	103,99351
7	102,43284
8	100,56852

Izvor: izrada autora

Graf 2. Točka infleksije druge klaster analize



Izvor: autorsko istraživanje

4.3. Rezultati istraživanja

U ovome poglavlju iznose se rezultati dviju klaster analiza podataka iz baze *Football Teams*. Otkriven je optimalan broj klastera te relevantni atributi kako bi se došlo do novih znanja. U nastavku će dobiveni klasteri biti prikazani i objašnjeni. Slika 27 koja slijedi prikazuje klastere iz prve analize gdje je atribut *Team* jedini ignorirani atribut.

4.3.1. Prva klaster analiza – Obilježja različitih liga

Slika 26. Klasteri prve klaster analize

```
Final cluster centroids:
                                    Cluster#
Attribute               Full Data           0           1           2           3           4           5
                          (98.0)      (21.0)      (23.0)      (21.0)      (10.0)      (13.0)      (10.0)
=============================================================================================================================
Tournament        Premier_League Premier_League      Serie_A   Bundesliga      Serie_A      LaLiga      LaLiga
Goals                    52.1837       51.1429      46.5652      50.4286         79.1     39.6923        60.3
Shots_pg                  11.851        12.081      10.9043      12.2048        15.03      9.7846       12.31
yellow_cards             69.7041            55      79.4783      60.7619         71.7     86.3846        73.2
red_cards                 3.3367        2.4762       4.3478       1.9048          4.9      4.7692         2.4
Possession               49.9959       49.9667      47.8696      49.6429         54.4     46.9154       55.29
Pass                     80.4418       80.6429      80.2174      79.1524         85.9     75.1462       84.67
AerialsWon               16.0061       16.7048      15.3913      16.5714        12.96     18.9615       13.97
Rating                     6.647        6.7171         6.56       6.6414        6.791      6.5285       6.722



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       21 ( 21%)
1       23 ( 23%)
2       21 ( 21%)
3       10 ( 10%)
4       13 ( 13%)
```

Izvor: autorsko istraživanje

U Weki je prvi klaster označen kao *Cluster#0*. Ovaj klaster sadrži 21 instancu što čini 21% od ukupnog broja podataka te su u pitanju većinom klubovi iz engleske Premier Lige. Ako se ovaj klaster uspoređuje s drugima po atributu *Rating*, dolazi se do zaključka da je *Cluster#0* treće najbolje rangirani klaster. Broj golova koji prosječno postižu momčadi iz ovog klastera u sezoni 2021.-2022. je skoro 52 dok je broj udaraca upućenih na gol malo veći od 12. Što se tiče žutih i crvenih kartona, broj žutih kartona je ravno 55 što je najmanje u usporedbi s drugim klasterima dok su se crveni kartoni također smjestili dosta nisko, na predzadnje mjesto s 2.47. Ekipe ovog klastera imaju skoro 50% posjeda i nešto više od 80% uspješnosti dodavanja u prosjeku. Što se tiče zračnih duela tu su na drugom mjestu sa više od 16 osvojenih prosječno po utakmici.

Slijedi drugi klaster koji ima naziv *Cluster#1*. *Cluster#1* sadrži 23 instance i isto toliko zauzimaju u postotku od ukupnog broja. U ovom klasteru je prosječna ocjena skoro najniža

što praktički znači da se ovdje radi o poprilično lošim momčadima i to većinski iz talijanske Serie A. Momčadi u ovome klasteru zabijale su u prosjeku ukupno svega 47 golova u sezoni uz prosječno 11 udaraca na gol po utakmici te su posjed zadržavale skoro 48% vremena. S druge strane, igrači ovih ekipa dobili su drugi najveći broj žutih kartona (77,38) a s crvenim su se smjestile u sredini s 3.33 kartona u sezoni. U zlatnoj sredini se nalaze kad se govori o postotku uspješnih dodvanja i broju osvojenih zračnih duela po utakmici.

Treći klaster se naziva *Cluster#2* te je on prema atributu Ranking na četvrtom mjestu. U ovome klasteru postoji 21 instanca te je njemačka Bundesliga predstavnik ovih momčadi. U skladu s *Ratingom* je i broj zabijenih golova koji je 50.42. Zanimljivo je da su momčadi ovog klastera najbolje u segmentu crvenih kartona gdje imaju najmanji broj zarađenih crvenih kartona koji je nešto niži od 2. Isto tako dobivaju i mali broj žutih kartona koji jedva prelazi 60. Loptu u posjedu imaju skoro 50% vremena kao i ekipe prvog analiziranog klastera ali uz nižu uspješnost točnih dodavanja (79.15%). Osvojeni broj zračnih duela je u usporedbi s drugim klasterima u sredini (16.57).

Slijedi četvrti klaster koji nosi naziv *Cluster#3*. Ovaj klaster sadrži uz *Cluster#5* najmanje instanci (10) ali zato su tu smještene najbolje rangirane ekipe s prosječnom ocjenom od 6.79. Liga predstavnik je opet Serie A kao i u ranije spomenutom drugom klasteru u kojem je prosječna ocjena druga najniža. U idućem poglavlju rada biti će objašnjeno kako i zašto je do toga došlo. Momčadi ovoga klastera čvrsto drže prvo mjesto po broju zabijenih golova koji je skoro 80 a taj ogroman broj golova je posljedica velikog broja udaraca na gol koji je viši od 15. Ove momčadi u prosjeku drže loptu najduže u svojim nogama (54.4%) što dovodi i do najvećeg broja uspješnih dodavanja (85.9%). Zanimljivo je da ove ekipe imaju najviše isključenih igrača zbog crvenih kartona, gotovo 5 u sezoni te najmanji broj osvojenih zračnih duela (12.96).

Poslije četvrtog klastera dolazi peti klaster naziva *Cluster#4*. Weka je u ovaj klaster smjestila 13% podataka odnosno 13 instanci. Ekipe ovoga klastera su prema *Ratingu* na začelju s ocjenom 6.53 a liga koja predstavlja ovaj najgori klaster je španjolska LaLiga. Ekipe ovog klastera postižu najmanje golova u sezoni (40) a to proizlazi iz naravno i najmanjeg broja upućenih udaraca na gol (10) po utakmici. Ovu podosta lošu statistiku prati i najmanji postotak posjeda (47%) te najlošiji postotak uspješnih dodavanja (75.15%). Ali su zato ove ekipe na prvom mjestu kada su u pitanju žuti kartoni (86) te na drugom

kada se spominju crveni kartoni (4.77). Interesantno je i da sudjeluju i osvajaju najviše zračnih duela po utakmici (19).

I za kraj ostaje spomenuti i *Cluster#5*. Ovaj klaster je totalno različit od prethodnog petog klastera po svim atributima osim po jednom a to je *Tournament*. Naime i ovaj klaster najviše čine ekipe iz španjolske LaLige no međutim ove ekipe krasi veliki broj zgoditaka (60), veliki broj udaraca na gol (12), najveći postotak posjeda (55.29%) te visoka uspješnost dodavanja (84.67%). Isto tako, imaju mali broj zarađenih crvenih kartona (2.4) i dobivenih zračnih duela (14). Sve to skupa daje vrlo visoku ocjenu od 6.72.

4.3.2. Druga klaster analiza – Najbolje momčadi

Nakon rezultata prve klaster analize slijede rezultati drugog klasteriranja. Ovoga puta ignoriran je atribut *Tournament* a umjesto njega je dodan atribut *Team*. Ovo je učinjeno iz razloga što dugogodišnji poznavaoci i obožavatelji nogometa znaju u koji razred pripadaju određene ekipe te se one mogu koristiti kao primjer za usporedbu s drugim ekipama (npr. Aston Villa već nekoliko sezona završava između sredine i vrha tablice Premier Lige isto kao Roma u talijanskoj Serie A). Ovoga puta će objašnjenje iznesenih atributa biti sažeto te će u fokusu biti samo najbitnije činjenice. U nastavku slijedi Slika 28 koja prikazuje nastale klastere u Weki.

Slika 27. Klasteri druge klaster analize

```
Final cluster centroids:
                                        Cluster#
Attribute                 Full Data           0           1           2           3           4
                            (98.0)      (12.0)      (16.0)      (27.0)      (17.0)      (26.0)
=============================================================================================================================
Team                  Manchester_City     Burnley        Metz     Everton Manchester_City  Aston_Villa
Goals                         52.1837     33.0833       47.25     42.5185     77.3529     57.6154
Shots_pg                      11.851        9.55     10.5062     10.7556     15.3412     12.5962
yellow_cards                  69.7041         75     85.3125     68.1481     60.4118     65.3462
red_cards                      3.3367          4      5.5625      2.3333      3.1176      2.8462
Possession                    49.9959     43.875     48.9437     46.9074     56.7176     52.2808
Pass                          80.4418     72.0583     80.825     78.5259     86.4882     82.1115
AerialsWon                    16.0061     20.925     15.0688     16.5111     13.2059     15.6192
Rating                         6.647      6.5133      6.5619      6.5693      6.8453      6.7123



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      12 ( 12%)
1      16 ( 16%)
2      27 ( 28%)
3      17 ( 17%)
4      26 ( 27%)
```

Izvor: autorsko istraživanje

Iz priložene slike može se uočiti 5 klastera. Klaster s najnižim Ratingom od 6.51 je *Cluster#0* a on broji i najmanje instanci, svega 12. Za usporedbu, *Cluster#2* koji se prema ocjeni nalazi u sredini ljestvice, sadrži duplo više instanci od prvog klastera. Najbolje ocjenjeni *Cluster#3* sa ocjenom 6.85 ima 17 instanci. Ove činjenice su nekako i logične jer je prosječnih uvijek najviše i u ostalim sferama koje se promatraju. Momčad koja predstavlja iznad prosječne je Manchester City, a Burnley je momčad koja predstavlja ispod prosječne. Momčad koja predstavlja one najbrojnije je Everton.

Manchester City postiže preko 30 golova više od Evertona a 40 više od Metza. Imaju skoro 5 udaraca više i uvijek drže loptu u posjedu duže od protivnika (57%). Sukladno podatku o posjedu imaju i puno veći postotak točnih dodavanja od najgorih ekipa poput Burnleya, razlika je 14%. Ali zato Burnley ima puno veći broj osvojenih zračnih duela (21) dok Manchester City i slične ekipe osvajaju svega 13 takvih duela. Što se tiče žutih i crvenih kartona, slabije ocjenjene momčadi dobivaju veći broj i žutih (75) i crvenih (4)

kartona. Još se ističe činjenica da prosječne ekipe poput Evertona imaju najmanji broj zarađenih crvenih kartona (2.33).

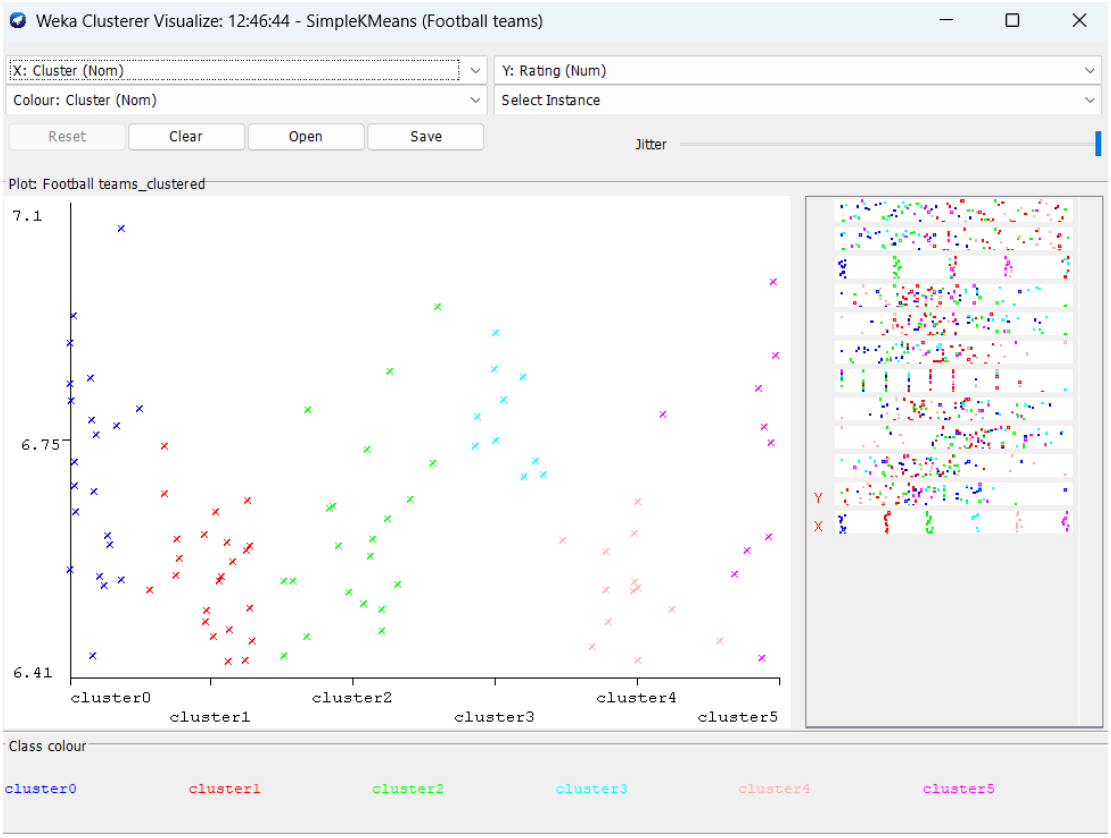
4.4. Rasprava

Nakon što su izneseni rezultati klasteriranja slijedi rasprava i subjektivno mišljenje autora rada. Klaster analizom otkrivene su određene uzročno-posljedične veze koje otkrivaju nova znanja iz baze podataka. Kako je nogomet česta tema razgovora tako su i nogometne rasprave učestale i žustre. Mnogi redovito prate nogometne utakmice i na temelju toga stvaraju svoju mišljenje i sliku o nogometu.

4.4.1. Stanje u nogometu danas

Mnogi tvrde kako je engleska Premier Liga najbolja liga na svijetu. Ostali su to često opovrgavali zbog neuspjeha engleskih ekipa u međuklupskom europskom natjecanju koje se zove Liga Prvaka. Za one koji ne znaju Liga Prvaka (engl. *Champions League*) je najjače klupsko natjecanje na svijetu. Igra se u skupinama iz kojih natjecanje nastavljaju dvije najbolje ekipe a zatim slijedi faza izbacivanja. Real Madrid je klub koji je najviše puta osvojio ovo natjecanje, a u prošlom desetljeću Real i Barcelona su u natjecanju bile najdominantnije ekipe te se zbog toga moglo diskutirati kako je ipak španjolska LaLiga najbolja liga na svijetu. Osobno mislim da je to zadnjih par godina engleska Premier Liga te ću svoj stav pokušati objasniti pomoću rezultata dobivenih klaster analizom.

Slika 28. Odnos klastera i atributa *Rating*



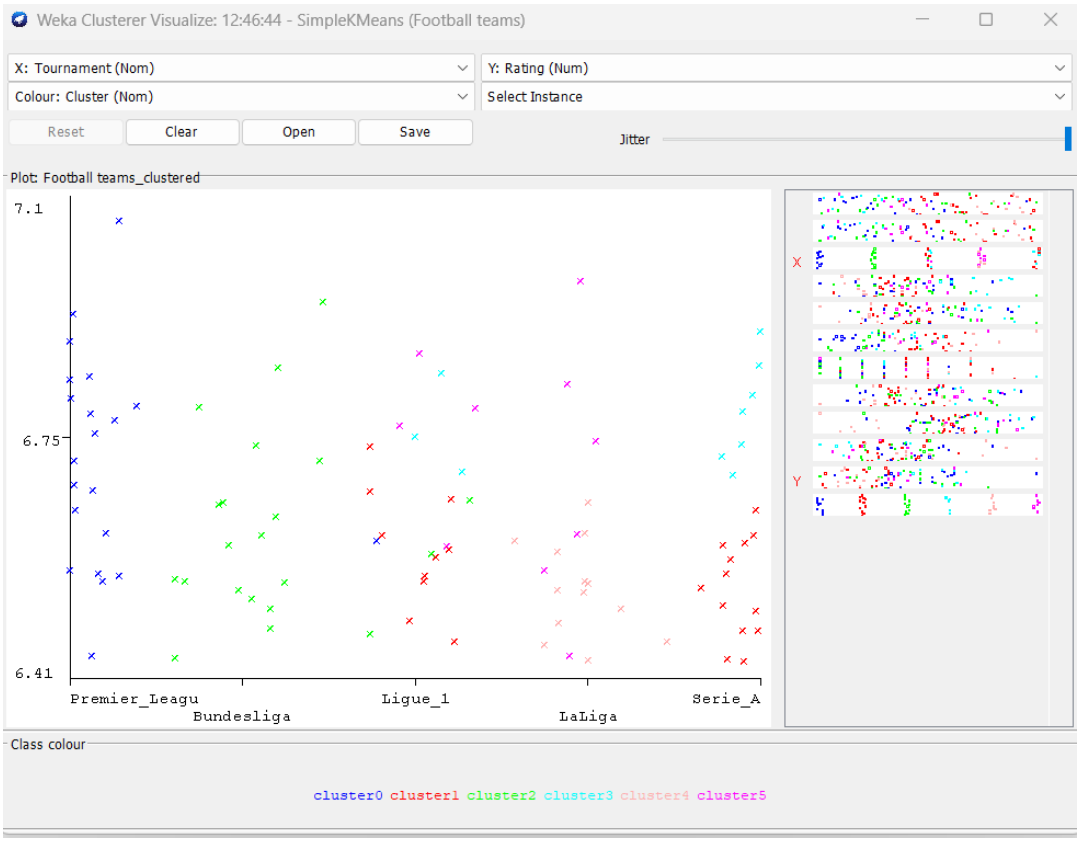
Izvor: autorsko istraživanje

Prema slici 29 iznad koja prikazuje prvu provedenu klaster analizu, talijanska Serie A i španjolska LaLiga su imale najviše momčadi u klasterima koji su imali najveći *Rating*. Na prvi mah dalo bi se zaključiti da su to dvije najkompetitivnije lige na svijetu. Zapravo, prema ovim rezultatima, talijanska Serie A je apsolutno najbolja liga na svijetu. Međutim ako se ponovno sagledaju rezultati može se primjetiti da ta dva klastera imaju jako mali broj instanci (10). Još jedan argument koji ne ide u prilog tim ligama je taj što su one istovremeno i na čelu dva klastera s najmanjim *Rating-om* odnosno najlošijim momčadima, pogotovo Serie A koja ima pozamašan broj momčadi u *Clusteru#1*. Za razliku od tih klastera, *Cluster#0* kojeg predvodi engleska Premier Liga ima čak 21 instancu a ocjenom je tek nešto ispod dviju ranije spomenutih najboljih klastera.

Sljedeća slika zorno prikazuje odnos *Rating-a*, nogometnih liga i klastera. Uočava se velika podijeljenost Serie A, LaLige i francuske Ligue 1 koja je uvelike diverzificirana. To dovodi da zaključka kako su talijanske momčadi poput Napolija, Intera ili Juventusa

postigle puno pogodaka ne samo zbog svoje iznimne kvalitete već i zato što se natječu u ligi gdje je velika razlika u kvaliteti između najboljih i najlošijih momčadi. Isto vrijedi i za španjolsku LaLigu. U usporedbi s njima Premier Liga i Bundesliga su lige sa malo outlinera što znači da su momčadi karakterno (u nogometnom smislu) vrlo slične i upravo to dovodi do visoke razine konkurentnosti a zbog jake konkurencije sve su utakmice neizvjesne i samim time zanimljivije prosječnom obožavatelju nogometa.

Slika 29. Odnos atributa *Tournament* i *Rating*



Izvor: autorsko istraživanje

Situacija je ovakva dijelomično i zbog novca. Nogomet, kao ni ostale branše, nije imun na novac. Razlika u televizijskim pravima je ogromna i daje veću financijsku moć engleskim klubovima. Klub iz Premier Lige dobiva 123 milijuna funti po sezoni, u usporedbi s LaLigom (56 milijuna funti), Serie A (52 milijuna funti), Bundesligom (52 milijuna funti) i Ligue 1 (27 milijuna funti). U prosjeku, tim iz Premier Lige ostvaruje više

nego dvostruko više prihoda od emitiranja nego njegov ekvivalent u La Ligi (Geey i Harvey, 2023.). Zbog novca, slabije ekipe u Premier Ligi imaju veću platežnu moć te mogu dovesti bolje igrače kako bi se mogli natjecati s najboljima.

U svim ligama je situacija jasna osim u francuskoj Ligue 1. Može se uočiti da je Ligue 1 kombinacija ekipa iz svih klastera što znači da ta liga nema svoj identitet, igraju se različiti stilovi nogometa i momčadi nemaju ni približno slična ulaganja. Ligue 1 u svijetu je poznata kao razvojna liga za mlade igrače u kojoj dominira brzi i iznimno fizički stil nogometa. Zbog toga što je razvojna liga kako za igrače tako i za trenere te momčadi sasvim drugačije igraju i zato ih je Weka rasporedila po različitim klasterima.

Premier Liga je liga u kojoj se istovremeno igra i tehnički i fizički izuzetno zahtjevan nogomet. Englezi su oduvijek poznati po svojoj fizičkoj igri što možemo vidjeti po velikom broju zračnih duela te broju žutih i crvenih kartona gdje su točno na sredini. Uz to još zadovoljavaju i velik broj točnih dodavanja i zabijenih golova. Zbog ovih parametara smatram da je engleski nogomet trenutno najbolji u svijetu iako su i Nijemci jako blizu sa svojom Bundesligom. U Bundesligi se igra sličan nogomet ali u prilog joj ne ide to što je Bayer Munchen gotovo svake godine prvak.

4.4.2. Što najbolje čini najboljima?

Svi su svjesni kako je cilj u nogometu postići zgoditak i samim time ekipe koje postižu najviši broj golova se mogu smatrati najboljima. Iako ponekad ima i iznimki ovog pravila jer postoji i defanzivni segment igre te ponekad ekipe koje su izrazito dobre u tome mogu ostvariti pozitivan rezultat. Klaster analizom dobiven je *Cluster#3* koji ima najbolji *Rating* i u koji su smještene momčadi koje krase ranije navedene karakteristike.

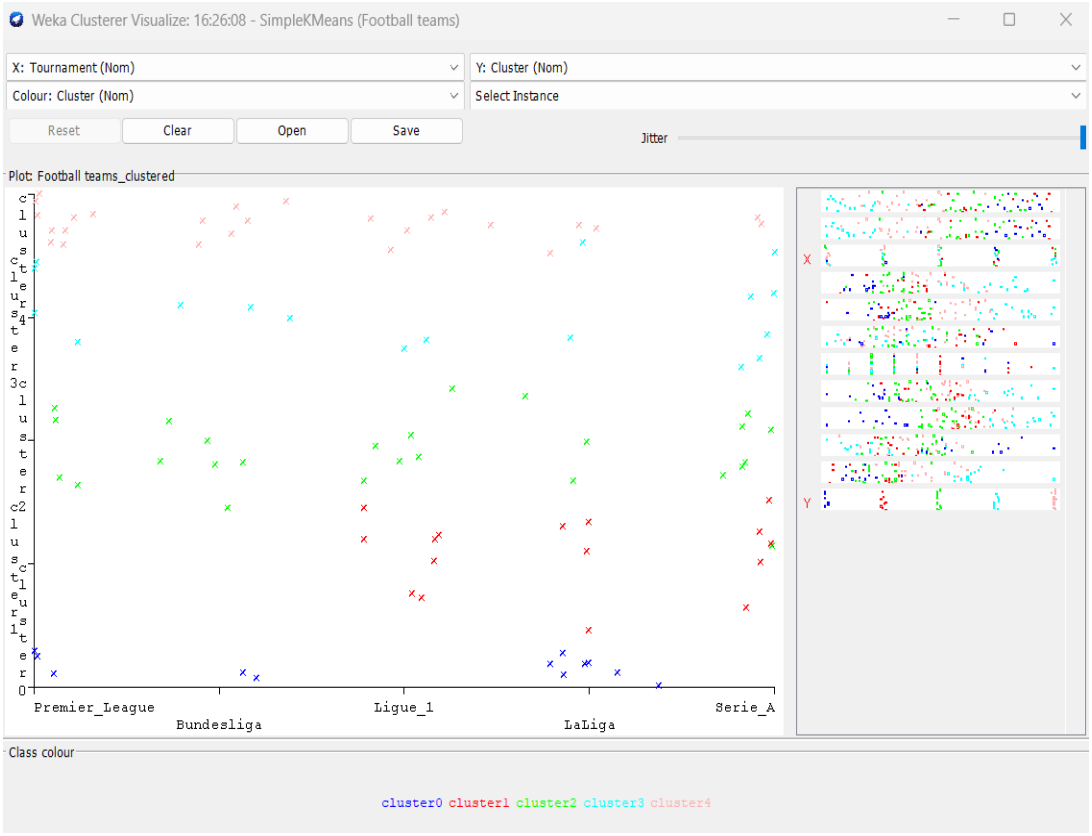
Nakon uvoda u ovo poglavlje slijedi vizualni prikaz odnosa klastera i ocjena, *Cluster#3* koji je najbitniji za ovu raspravu označen je tirkiznom bojom.

Slika 30. Ekipe razvrstane po klasterima na temelju ratinga

Izvor: autorsko istraživanje

Cluster#3 može se nazvati „Superligom“. Superliga je ideja koja je zamalo zaživjela i u praksi a pokrenuo ju je predsjednik Real Madrida Florentino Perez. Iako su određene momčadi bile spremne učestvovati u takvoj ligi, nogometna organizacija UEFA i nogometna javnost oštro su osudili ovakav pokušaj isključivanja većine nogometnih klubova iz kompetitivnog nogometa. Iako je trenutno ideja odbačena, ona i dalje živi i realno je da se jednog dana i ostvari pošto takvi primjeri postoje u drugim sportovima u Europi. U nastavku slijedi prikaz dobivene Superlige (*Cluster#3*), odnosno momčadi koje su pridružene iz najboljih 5 liga svijeta.

Slika 31. Momčadi Superlige prikazani po svojim ligama



Izvor: autorsko istraživanje

Iz gore navedene Slike 32 može se vidjeti kako na temelju statistike iz sezone 2021.-2022., Serie A ima najviše predstavnika u Superligi a svaka liga ima barem 2 predstavnika. Momčad s najboljom ocjenom te sezone bio je Manchester City stoga je taj klub i predstavnik klastera. Superliga ima 17 sudionika što je malo manje od standardnog broja učesnika u ligama petice.

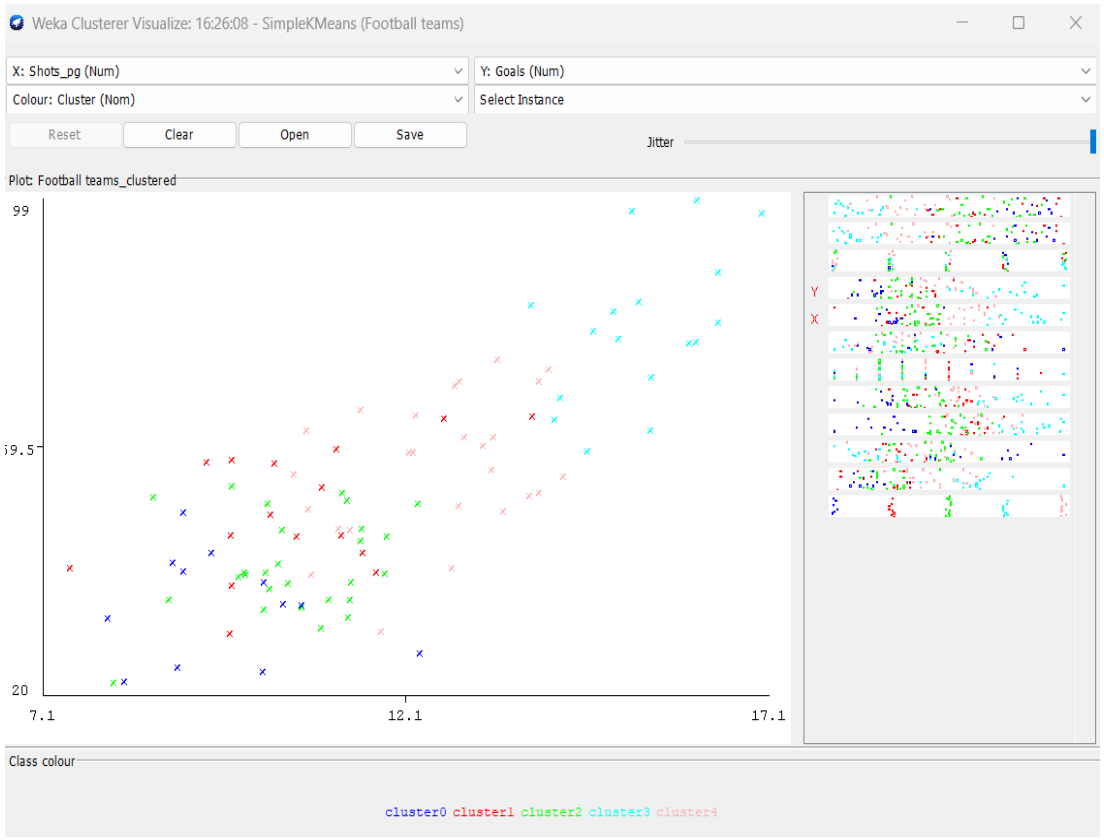
Tablica 6. Superliga

Naziv momčadi	Liga
Manchester City	Premier League
Liverpool	
Manchester United	
Chelsea	
Bayern Munchen	Bundesliga
Borussia Dortmund	
RB Leipzig	
PSG	Ligue 1
Lyon	
Real Madrid	LaLiga
Barcelona	
Napoli	Serie A
Inter	
AC Milan	
Sassuolo	
Juventus	

Izvor: izrada autora

Zašto su baš momčadi iz Tablice 6 najbolje? Netko će reći da je to zbog financijske premoći s čime se mogu složiti ali u ovome radu u prvi plan je stavljen nogometni aspekt. Iz nekoliko vizualnih pikaza dobivenih rezultata klasteriranja izvući će se informacije koje će biti sažete u odlomak kojim će se pokušati pojasniti kako ove momčadi ostvaruju pobjede.

Slika 32. Prikaz broja upućenih udaraca i postignutih golova

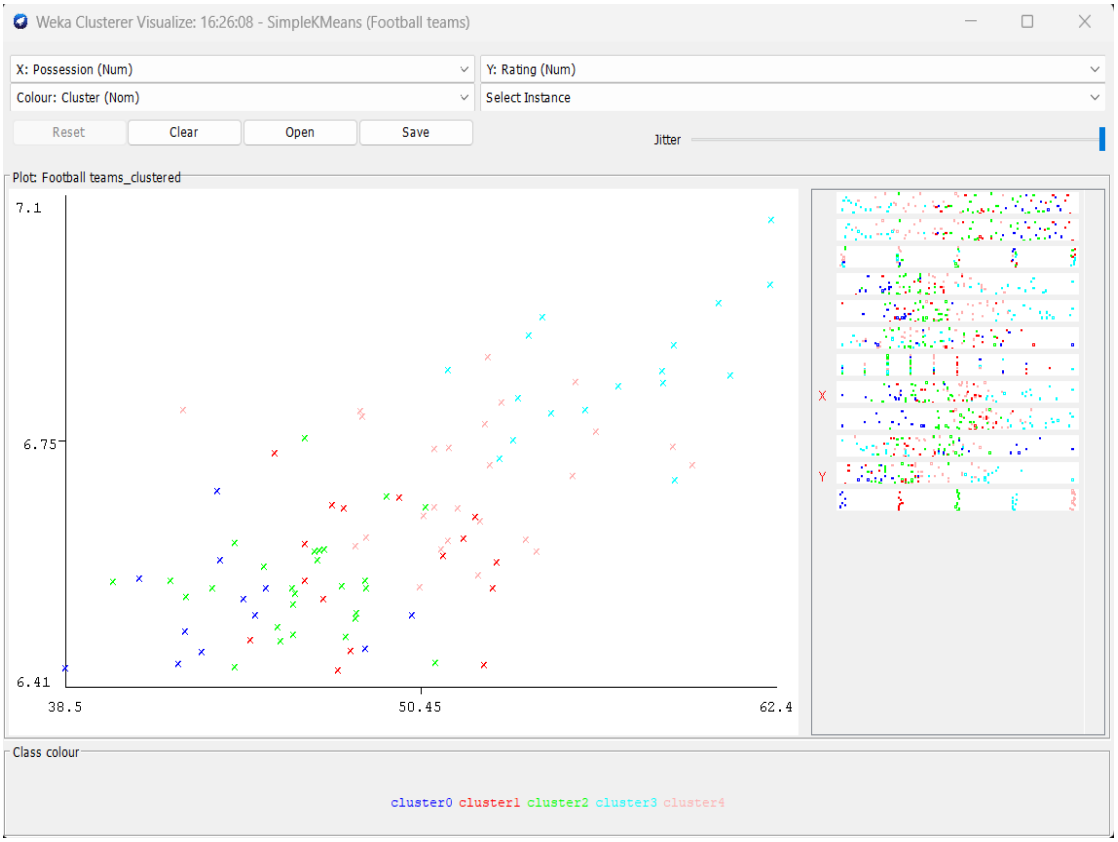


Izvor: autorsko istraživanje

Momčadi iz ovoga klastera igraju izrazito napadački nogomet a to se vidi iz velikog broja udaraca na gol a samim time postižu i velik broj golova.

Kroz godine postojanja nogometa i izmjenjivanja različitih taktika, stručnjaci i treneri su zaključili kako treba inzistirati na čuvanju posjeda i igri „od noge do noge“. Stoga moderni igrači na svim pozicijama moraju biti savršeno tehnički potkovani jer ako njihova ekipa ima posjed veća je šansa da će doći do gola a istovremeno se smanjuju šanse da prime zgoditak.

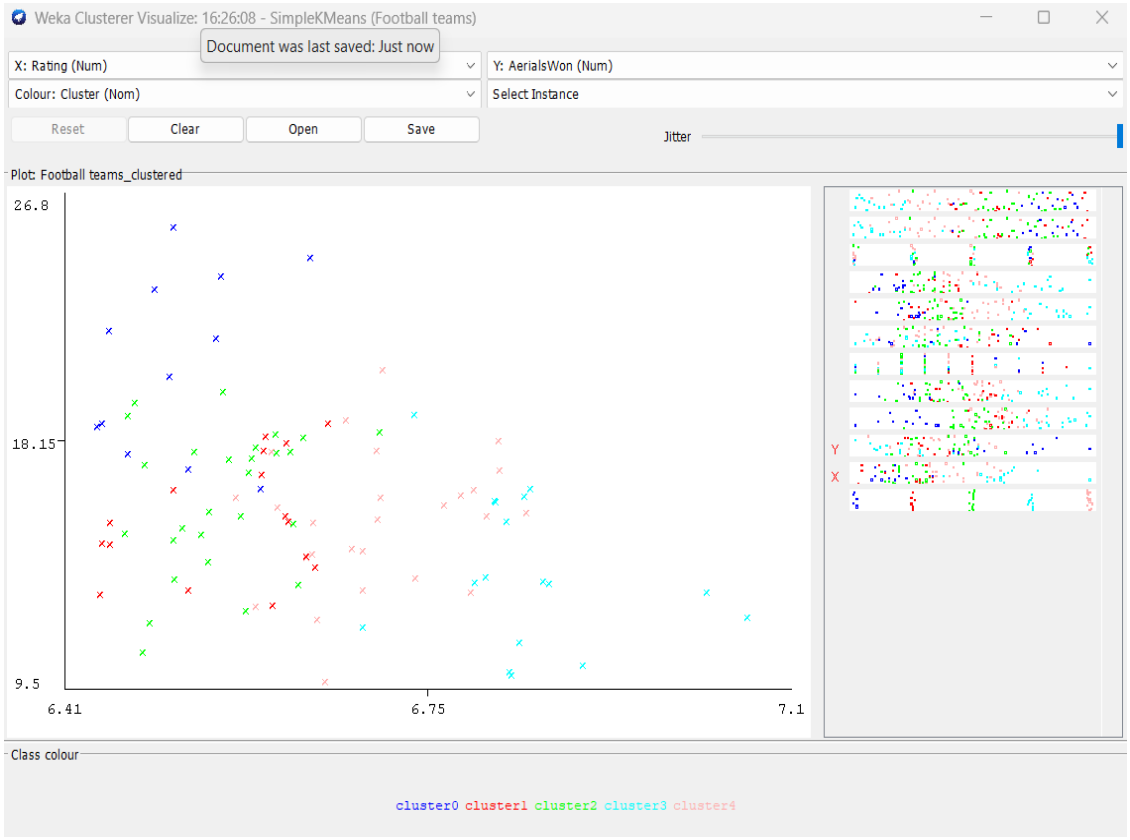
Slika 33. Odnos posjeda i zarađene ocjene



Izvor: autorsko istraživanje

Sjetimo se kako je u prošlosti prvi nogometni analitičar Charles Reep došao do zaključka kako treba dalekomentnim dodavanjima slati loptu ispred gola jer je to najefikasniji način za postići zgoditak. Godinama kasnije novi stručnjaci su dokazali kako je ta teorija neistinita jer su najbolje momčadi igrale upravo suprotno, s puno kratkih dodavanja. Baš zbog takvog stila u njihovoj igri nema puno zračnih duela.

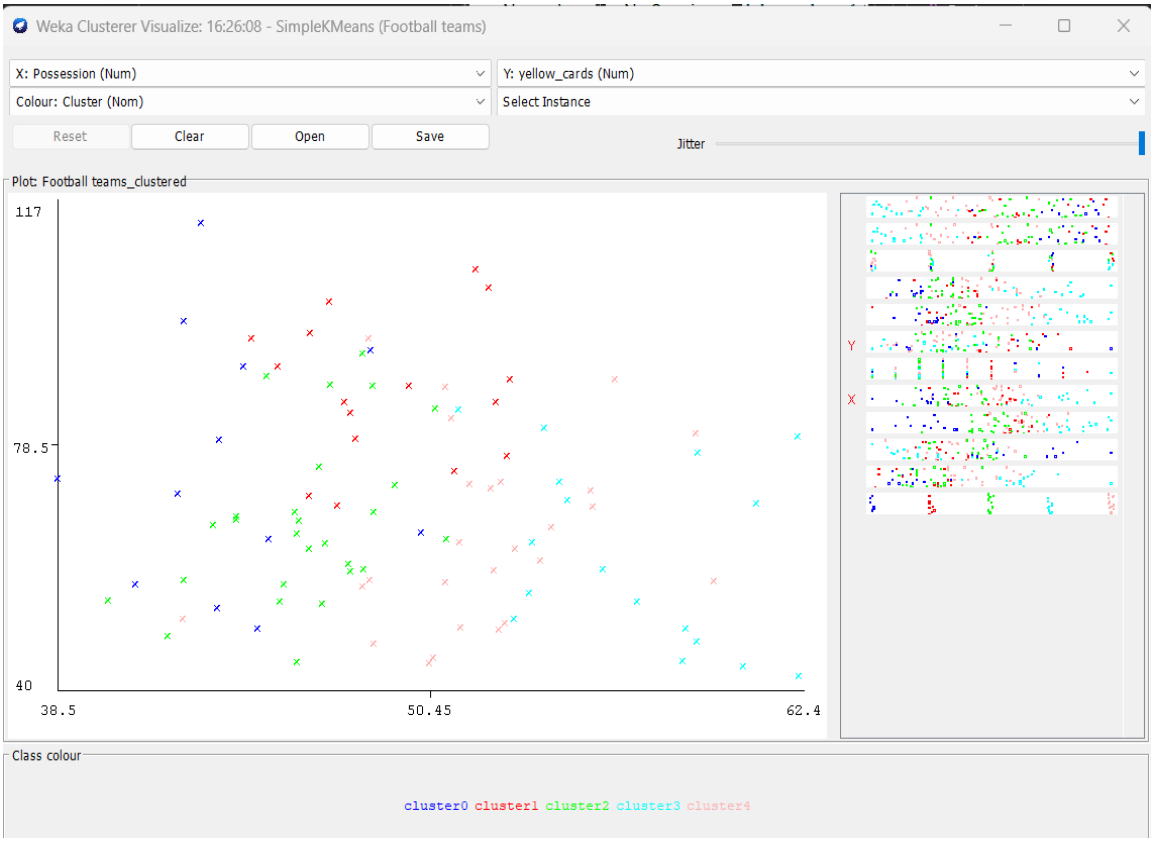
Slika 34. Odnos dobivenih zračnih duela i zarađene ocjene



Izvor: autorsko istraživanje

Za kraj su ostali prikazi o žutim i crvenim kartonima. Njihov broj proizlazi iz posjeda. Momčad uglavnom dobija kartone kada je u defanzivi i nema posjed jer pokušava zaustaviti protivnika i ponovno uzeti loptu.

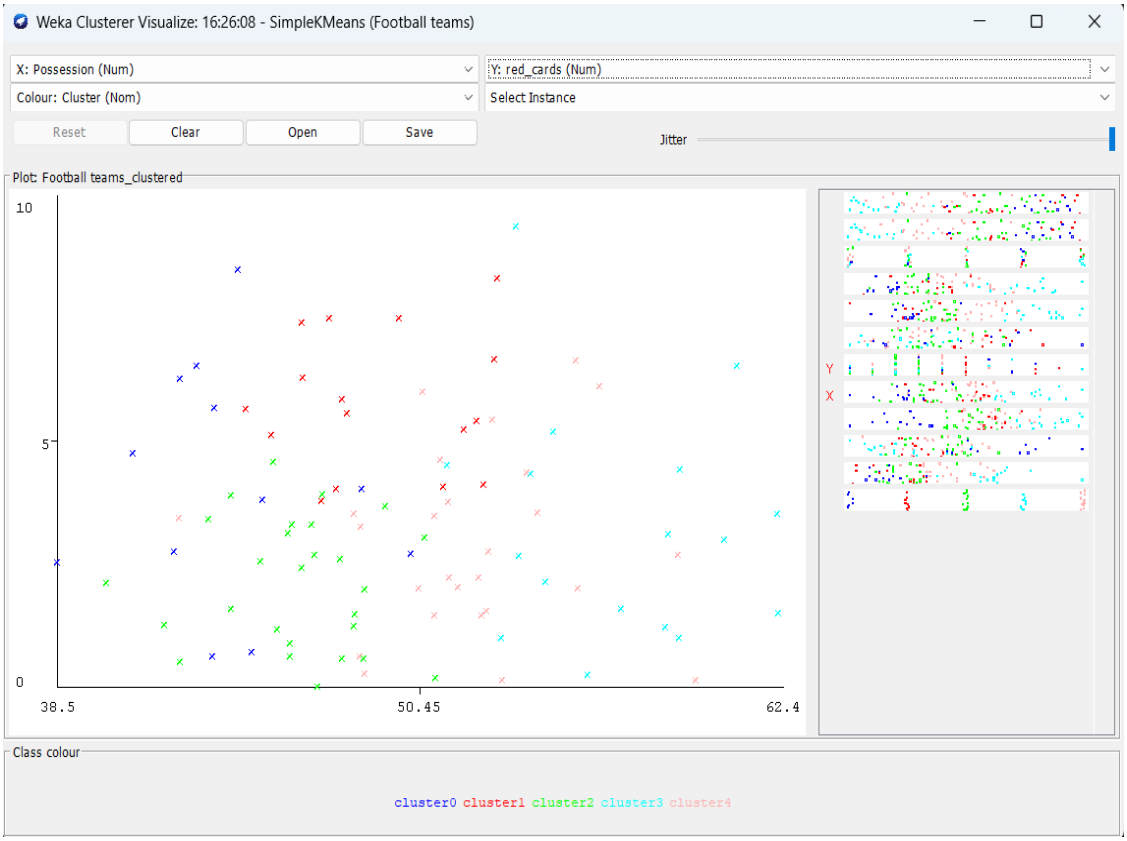
Slika 35. Odnos zarađenih žutih kartona i posjeda lopte



Izvor: autorsko istraživanje

Iz slike 36 vidimo da momčadi koje duže zadržavaju loptu u posjedu zarađuju manje žutih kartona a to opisuje najbolje momčadi.

Slika 36. Odnos zarađenih crvenih kartona i posjeda lopte



Izvor: autorsko istraživanje

Situacija s crvenim kartonima na slici 37 iznad je nešto drugačija. Zbog toga što momčadi igraju napadački i konstantno stvaraju šanse za to im je potreban posjed. Kako bi bile što opasnije po suparnički gol, igrači stoje visoko i pritišću protivničke da se povuku što bliže vlastitome golu i brane. U takvim situacijama ako dođe do veće pogreške postoji velika opasnost od kontranapada te je ponekad igrač primoran zaustaviti kontru najstrožom kaznom. To je razlog zašto broj crvenih kartona nije najmanji kao broj žutih.

5. ZAKLJUČAK

Nogomet je za mnoge najvažnija sporedna stvar na svijetu. Kroz godine se razvija i njegova popularnost samo raste. Za to je djelomično zaslužna i tehnologija, direktno i indirektno. Tehnologija se kao i nogomet kroz godine razvila, još i brže. Sve većem broju ljudi su dostupni televizori, tableti, pametni telefoni, itd. a to su sve mediji putem kojih obožavatelji mogu pratiti utakmice. U ovome radu se nismo bavili time već utjecajem tehnologije na samu igru. Nogomet je u neku ruku postao pravedniji zbog tehnologije VAR-a, gol-linije i ostalih. Prikupljanje podataka nikad nije bilo brže i jednostavnije. To je dovelo do razvoja nove discipline, nogometne analitike. Ona je u modernome nogometu postala nužnost jer dovodi do konkurentske prednosti. Ona utječe na pripremanje treninga, taktike i strategije, prevenciju ozlijeda te analizu samih utakmica i i zvedbi igrača. Mnogi treneri su ju pokušali ignorirati ali jednostavno nisu mogli. Najbolje momčadi izuzetno puno ulažu u suvremene nogometne sustave i aplikacije i zapošljavaju analitičare koji su sposobni izvući korisna znanja iz tih prikupljenih podataka. Na kraju ovoga rada analizirana je jedna tipična nogometna baza podataka. Iz nje su, uz pomoć klaster analize, uspješno izvučena i prikazana znanja o nogometnim stilovima i ligama. Naravno, nogometna mišljenja su često subjektivna i razlikuju se ali kada se sagledaju podaci dolazi se do jedne istine. Trenutna situacija u nogometu je takva da je engleska liga najdominantnija liga na svijetu. Kao i u sve drugo, u nogomet morate ulagati kako bi stvorili dobar proizvod i Englezi to rade. Nogometaši engleske Premier lige su superiorniji u odnosu na druge u skoro svim aspektima nogometne igre. Ostale lige uglavnom imaju jednu ili dvije ekipe koje dominiraju godinama i to djelomično oduzima čaroliju sporta. Stoga, ukoliko želite početi pratiti nogomet, moj savjet je da krenete od engleske lige.

Popis literature

1) Ahmad, P., Qamar, S. i Rizvi, S.Q. (2015). Techniques of Data Mining In Healthcare: A Review. International Journal of Computer Applications, 120, str. 38-50.

2) Algarni, A. (2016). Data mining in education. International Journal of Advanced Computer Science and Applications, 7(6).

3) Apostolou, K., & Tjortjis, C. (2019, July). Sports Analytics algorithms for performance prediction. In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), str. 1-4.

4) Bitilis, P. i Chatzipanagiotou, N. (2022). Digitalizing the Football Experience.

5) Bolf, N. (2021). Osvježimo znanje: Strojno učenje. Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske, 70(9-10), str. 591-593.

6) Bramer, M. (2020). Principles of data mining.

7) Cacho-Elizondo, S. i Álvarez, J. D. L. (2020). Big Data in the Decision-Making Processes of Football Teams Integrating a Theoretical Framework, Applications and Reach. Journal of Strategic Innovation and Sustainability, 15(2), str. 21-44.

8) Cintia, P., Rinzivillo, S. i Pappalardo, L. (2015). A network-based approach to evaluate the performance of football teams. In Machine learning and data mining for sports analytics workshop, Porto, Portugal.

9) Dunn, M., Hart, J. i James, D. (2018). Wearing electronic performance and tracking system devices in association football: Potential injury scenarios and associated impact energies. In Proceedings (Vol. 2, No. 6, str. 232). MDPI.

10) Fang, L., Wei, Q. i Xu, C. J. (2021). Technical and tactical command decision algorithm of football matches based on big data and neural network. Scientific Programming, str. 1-9.

11) Fayyad, U., Piatetsky-Shapiro, G. i Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), str. 37-37.

12) Footballytics (2021). Data Analytics in Football. [online] Footballytics. Dostupno na: https://www.footballytics.ch/post/data-analytics-in-football. [21.10.2023.].

13) Geey, D. i Harvey, A. (2023). Football Broadcasting Deals Across the Top 5 European Leagues. [online] DANIEL GEEY. Dostupno na: https://www.danielgeey.com/done-deal-blog/football-broadcasting-deals-across-the-top-5-european-leagues. [13.01.2024.].

14) João Medeiros (2017). How Data Analytics Killed the Premier League's Long Ball Game. [online] Wired.co.uk. Dostupno na: https://www.wired.co.uk/article/premier-league-stats-football-analytics-prozone-gegenpressing-tiki-taka. [16.11.2023.].

15) Jothi, N. I Husain, W. (2015). Data mining in healthcare–a review. Procedia computer science, 72, str. 306-313.

16) Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.

17) Kaur, M., Gulati, H. i Kundra, H. (2014). Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications. International Journal of Computer Applications, 99(12), str. 1–3.

18) Kröckel, P. (2019). Big Data Event Analytics in Football for Tactical Decision Support (Doctoral dissertation, Dissertation, Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU).

19) Kudyba, S. P. (2018). Healthcare informatics: improving efficiency through technology, analytics, and management. CRC Press.

20) Kumar, A.S. (2022). 5 Important Uses of Data Analytics in Football - Mad about Sports. [online] madaboutsports.in. Dostupno na: https://madaboutsports.in/blog/5-important-facts-data-analytics-in-football/ [17.01.2024.].

21) Liao, S. H., Chu, P. H. i Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. Expert systems with applications, 39(12), str. 11303-11311.

22) Šimec, A., & Lozić, D. (2020). Rudarenje podataka.

23) Mani, A. (n.d.).Goal-Line Technology in Soccer: Discussion and Evaluation of Systems.

24) Mankar, A. B. i Burange, M. S. (2014). Data Mining-an evolutionary view of agriculture. International journal of Application or Innovation in engineering and management, 3(3), str. 102-105.

25) Memmert, D. i Raabe, D. (2018). Data analytics in football: Positional data collection, modelling and analysis. Routledge.

26) Pejić Bach, M. (2005). Rudarenje podataka u bankarstvu. Zbornik ekonomskog fakulteta u Zagrebu, 3(1), str. 181-193.

27) Pejić Bach, M., Bertoncel, T., Meško, M., Suša Vugec, D., & Ivančić, L. (2020). Big data usage in european countries: Cluster analysis approach. Data, 5(1), 25.

28) Pejić Bach, M., Jaklič, J., & Vugec, D. S. (2018). Understanding impact of business intelligence to organizational performance using cluster analysis: does culture matter?. International Journal of Information Systems and Project Management, 6(3), 63-86.

29) Prakoso, M. L. A. i Lumintuarso, R. (2021). Analysis of the Use of Statistical Data in the Formulation of Strategies, Tactics and Evaluation of Football Matches. In 4th

International Conference on Sports Sciences and Health (ICSSH 2020), str. 33-37.

30) Provost, F. i Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.".

31) Rein, R. i Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus, 5(1), str. 1-13.

32) Sekan, F. (2023). Short History of Data Analysis in football. [online] Medium. Dostupno na: https://medium.com/@filip.sekan/short-history-of-data-analysis-in-football-ce1963e428ae. [11.12.2023.].

33) Spagnolo, P., Leo, M., Mazzeo, P., Nitti, M., Stella, E. i Distante, A. (2013). Non-invasive soccer goal line technology: a real case study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, str.1011-1018.

34) Sumpter, D. (2021). Evaluating actions in football using machine learning. [online] Medium. Dostupno na: https://soccermatics.medium.com/evaluating-actions-in-football-using-machine-learning-69517e376e0c. [18.12.2023.].

35) Surujlal, J. i Jordaan, D. B. (2013). Goal line technology in soccer: are referees ready for technology in decision making?: technology and innovation. African Journal for Physical Health Education, Recreation and Dance, 19(2), str. 245-257.

36) Tamir, I. i Bar-Eli, M. (2021). The moral gatekeeper: Soccer and technology, the case of Video Assistant Referee (VAR). Frontiers in psychology, 11, 613469.

37) Valter, D.S., Adam, C., Barry, M. i Marco, C. (2006). Validation of Prozone ®: a New video-based Performance Analysis System. International Journal of Performance Analysis in Sport, 6(1), str.108–119.

38) Van den Berg, L. i Surujlal, J. (2020). Video assistant referee: Spectator and fan perceptions and experiences. International Journal of social sciences and humanity studies, 12(2), str. 449-465.

39) Weiss, G. (2010). Data mining in the telecommunications industry. In Networking and Telecommunications: Concepts, Methodologies, Tools, and Applications, str. 194-201.

40) Weiss, G.M. (2005). Data Mining in Telecommunications. Data Mining and Knowledge Discovery Handbook, str.1189–1201.

41) Zoroja, J., & Pejić Bach, M. (2016). Impact of information and communication technology to the competitiveness of European countries-cluster analysis approach. Journal of theoretical and applied electronic commerce research, 11(1), 1-10.

Popis slika

Popis tablica

Popis grafova

Životopis studenta

Anđelko Nikić

Cirkovljanska 1

10 000 Zagreb

Mob: +385 95 8847 940

E-mail: nikic.andjelko@gmail.com

OIB: 50164770759

Obrazovanje:

- Ekonomski fakultet Sveučilišta u Zagrebu, smjer menadžerska informatika (2017. – 2024.)
- XI. Gimnazija u Zagrebu (2013. – 2017.)

Radno iskustvo:

- Pliva d.o.o. - sudent/analitičar u operacijama nabave (2022. – 2023.)
- Hotel Esplanade d.o.o. (2019. – 2021.) - konobar

Posebna znanja i vještine:

- Engleski u govoru i pisanju (B2)
- MS Office
- Ariba (SAP)
- Bizagi
- Weka
- Google Ads
- Komunikacijske vještine
- Položen vozački ispit B katgorije

Ostali interesi:

- Košarka
- Nogomet
- Pub-kvizovi
- Padel