

Predikcija rejtinga igrača u video igri DoTA2

Tomislav Dobrički
Fakultet tehničkih nauka
Univerzitet u Novom Sadu
Trg Dositeja Obradovića 6
21000 Novi Sad
e-mail: dtoma95@gmail.com

Stefan Colić
Fakultet tehničkih nauka
Univerzitet u Novom Sadu
Trg Dositeja Obradovića 6
21000 Novi Sad
e-mail: stefan.colic4@gmail.com

Ognjen Vlajić
Fakultet tehničkih nauka
Univerzitet u Novom Sadu
Trg Dositeja Obradovića 6
21000 Novi Sad
e-mail: ogi.vlajic@gmail.com

Abstract—Dota 2 je jedna od najpopularnijih online video igara u svetu. Poznata je po svojoj kompleksnosti i težini. Dota 2 je poznata i po razvijenoj profesionalnoj sceni, gde se neki od najbolji igrača međusobno takmiče za velike sume novca. Ovaj rad je fokusiran na Matchmaking Rating (MMR) koji je indikator veštine igrača unutar Dota 2 sistema. MMR se koristi za uparivanje igrača koji su sličnih veština kako bi mečevi bili što bolje izjednačeni. Cij ovog rada je da se razvije model, uz pomoć mašinskog učenja, koji može precizno da predvidi MMR jednog igrača na osnovu njegove istorije mečeva. Podaci za treniranje modela su prikupljeni pomoću javno dostupnih API-a (Application programming interface). U radu je ukupno opisano pet načina predstavljanja prikupljenih podataka. U svrhu pronalazjenja najboljeg modela, nad podacima je trenirano više oblika linearne regresije, Ada Boosta i konvolutivnih (CNN) i rekurentnih (RNN) neuronskih mrežaza. Najbolji rezultati su dobijeni pomoću skupa podataka, gde su karakteri iz Dota 2 predstavljeni preko svojih uloga, i konvolutivne neuronske mreže. Iz rezultata se ispostavilo da je predviđanje MMR-a igrača sa prikupljenim podacima moguć do neke mere, ali i dalje ne dovoljno precizno.

Ključne reči—Dota 2; MMR; regresija; elastic net; neuronska mreža; konvolutivna neuronska mreža; rekurentna neuronska mreža; Ada Boost;

I. UVOD

Dota 2 [1] je MOBA (*Multiplayer online battle arena*) igra razvijena od strane Američke kompanije Valve. Igra je prvobitno nastala kao custom mapa u igri *Warcraft III* 2009. godine i jedna je od prvih igara iz MOBA žanra. Dota 2 spada u elektrosnski sport (e-sport) i trenutno ima najveći nagradni fond na nekom turniru koji iznosi 25 miliona američkih dolara [4].

U jednom meču igre Dota 2, deset igrača su podeljena u dva tima od po pet igrača. Cilj igre jeste uništiti protivničku bazu odnosno glavnu kulu koja se zove *Ancient*. Sama igra je veoma kompleksna iz razloga što ima 117 heroja, sa velikom kombinaciju mogućnosti. Mečevi mogu da traju više od sat vremena, i u tom periodu postoji veliki broj odluka prilikom kojih igrači moraju što ispravnije da postupe da bi pobedili. Osim strateškog znanja, igra zahteva i određeni nivo mehaničkih sposobnosti od svojih igrača. Glavni razlog za popularnost Dote jeste u njenoj težini i u takmičarskoj atmosferi igrača što mesečno privlači oko 12 miliona jedinstvenih igrača [1].

Glavna mera za veštinu nekog igrača je MMR (*matchmaking rating*), koji je baziran na ELO sistemu koji se koristi u šahu [10]. MMR funkcioniše relativno jednostavno, za svaku pobedu igrač je nagrađen za oko 25 poena, a prilikom gubitka mu se oduzima oko 25 poena. Tačan broj nagrađenih i oduzetih poena zavisi od MMR protivničkih i savezničkih igrača u meču. Iako je MMR relativno jednostavan sistem, on se pokazao kao dovoljno dobra mera da razdvaja igrače različite veštine u igri.

Zbog kompleksnosti i višedimenzionalnosti igre, teško je odrediti koje su tačno osobine igrača na osnovu kojih je jasno razdvojiti bolje igrače od lošijih. Glavni cilj ovog rada jeste da se napravi model za predikciju MMR-a igrača na osnovu istorije mečeva. Ovim bi se pokazalo do koje mere se može proceniti veština igrača isključivo na osnovu statističkih podataka. Dodatno, sistem koji precizno prediktuje MMR igrača na osnovu njegove preformanse u mečevima može da posluži kao koristan alat za otkrivanje igrača koji zlorabljavaju MMR sistem. To su često igrači koji kupe nalog sa višim MMR-om (*boosting*) ili igrači koji namerno spuštaju svoj MMR da bi lakše pobeđivali protiv slabijih igrača (*smurfing*).

U ovom radu će se opisati primena različitih modela mašinskog učenja, uključujući više oblika linearne regresije, Ada Boosta i konvolutivnih (CNN) i rekurentnih (RNN) neuronskih mreža. Takođe je opisan proces prikupljanja podataka putem javno dostupnih servisa, zajedno sa preprekama sa kojima smo se suočili u tom precesu. Osim prikupljanja podataka, dodatani izazov čini način predstavljanja istorije mečeva nekog igrača. U ovom radu je predstavljeno pet načina izdvajanja atributa u obliku pet različitih seta podataka.

Preformanse modela su izmerene uz pomoć posebnog validacionog skupa podataka koji nije korišćen prilikom treniranja modela. Za meru performansi se priljublju RMSE (*Root mean squared error*) i MAE (*mean absolute error*).

Ovaj rad će biti organizovan na sledeći način. U sekciji II će biti napravljen osvrt na srodna istraživanja kako za igru Dota 2 tako i za druge MOBA igre. U III sekciji će biti navedeno kako su formirani setovi podataka koji su korišćeni za obučavanje modela. Takođe, u ovoj sekciji će biti opisana obeležja setova i analiza nad podacima. U sekciji IV će biti

reči o metodologiji koja je korišćena za rešavanje problema predikcije rejtinga. U V sekciji će biti prikazani rezultati modela. Na kraju u sekciji VI će biti napravljena sumariizacija rada i izneće se zaključci o obrađenoj temi.

II. SRODNA ISTRAŽIVANJA

Broj radova koji proučavaju igru Dota 2 je veoma mali i uglavnom se bavi predikcijom koji tim će pobediti u zavisnosti od igrača u svakom timu ili izbora heroja u partiji. Među njima se izdvajaju dva rada koja su slična tematici ovog rada.

Prvi rad koji se izdvaja jeste [5]. On se bavi proučavanjem prostorno-vremenskog aspekta u igri, odnosno u kakvoj korelaciji su veština igrača i njegovo kretanje po mapi. Za rad su izabrani igrači različitih veština koji su klasifikovani u 4 grupe. Sami podaci o kretanju svakog igrača su prikupljeni pomoću *replay* fajlova koji se generišu nakon svake partije i koji sadrže veliku količinu informacija. Utvrđeno je da igrači sa većom veštinom više menjaju zone na mapi i da prostorno-vremenski aspekt može biti veoma značajan u određivanju MMR-a, ali da pri tome treba uzeti u obzir i izbor heroja u meču. Zbog API-ja koji je bio dostupan nije postojala mogućnost da se iskoriste ove informacije kako bi se unapredio model iz ovog rada.

Drugi rad [6] proučava u kakvoj korelaciji su veštine igrača i tip heroja koji je izabran u partiji (*intelligence, agility i strength*). U ovom radu se pokazalo da veština igrača ne mora da se poklapa sa brojem partija odnosno iskustvom igrača, kao i da su uspešniji oni igrači koji su agresivniji i koji gledaju da imaju više ubistava nego asistencija i da što brže završe partije.

U nedostaku radova koji se bave Dotom 2 traženi su radovi koji se bave nekom drugom MOBA igrom. Pronađen je rad [7] koji se bavi igrom *Heroes of Newerth* [8] (igrom nastalom po ugledu na Dotu). Cilj ovog rada je bio da se nađe korelacija između MMR-a igrača i njegovog stila igre kroz skup obeležja putem statičke analize. Rad je pokazao da je broj odigranih mečeva u slaboj sprezi sa MMR-om što je zaključak i iz prethodno navedenog rada, dok su neki parametri kao što su K-D ratio (kill-death), K-A ratio (kill-assists), GPM (gold per minute), XPM (experience per minute), *creep denying* u relativnoj jakoj sprezi.

Kako Dota 2 spada u e-sport tražena je i sličnost sa radovima koji proučavaju klasične sportove. Rad [9] prediktuje preformanse *Quarterback*-ova iz NFL-a na osnovu njihovih prethodnih igrara. Isprobavani su SVM (*Support Vector Machine*) i neuronska mreža, ali su obe metodologije pokazale slabije rezultate usled nedostatka obeležja kojih je bilo samo dvanaest.

Pošto je tema istraživanja ovog rada MMR proučavano je i kako se on koristi u matchmaking system-u priliko nalaženja igrača za partije u MOBA igrama. Najpoznatiji ovakav sistem je Elo sistem [10], čija modifikacija se koristi i u Doti, a pored njega se često koristi i TrueSkill [11].

III. SKUP PODATAKA

U poglavljima ove sekcije su opisani preduzeti procesi za prikupljanje i pripremu podataka koji su korišćeni u ovom radu. Na kraju je data kratka analiza čitavog skupa podataka.

A. Prikupljanje podataka

Skup podataka koji je korišćen u sklopu ovog rada je samostalno sastavljen. Za prikupljanje podataka korišćena su dva servisa koji nude mogućnost dobavljanja podataka o igračima i njihovim mečevima. Prvi servis je zvanični Dota 2 API (u daljem tekstu samo Dota 2 API), koji je održan od strane proizvođača igre. Dota 2 API je pogodan za prikupljanje velikog broja podataka zato što nema ograničen broj zahteva, zbog toga je on korišćen kada god je to moguće. S druge strane, OpenDota API nudi mogućnost dobavljanja MMR-a igrača, što nije moguće kroz Dota 2 API, ali ima ograničen broj zahteva na nivou meseca.

Sam proces sastavljanja skupa podataka je podeljen na četiri faze: prikupljanje identifikatora igrača, prikupljanje MMR-a igrača, prikupljanje mečeva i dobavljanje detalja mečeva. U narednim podsekcijama je opisan način izvršavanja i rezultat svake faze.

1) Prikupljanje identifikatora igrača

Svaki igrač ima nalog sa jedinstvenim identifikatorom u obliku celog broja. Identifikatori su dodeljeni na nivou *Steam* platforme, gde je Dota 2 samo jedan od mnogih igara u ponudi. Što znači da mnogi nalozi nemaju ni jedan odigrani meč u Doti 2. Cilj prve faze jeste pronalaženje naloga koji su igrali Dotu 2 u periodu gde je MMR igrača bio javno dostupan (do kraja 2017. godine).

Identifikatori igrača se prikupljaju tako što se iz početnog, slučajno odabranog meča, preuzmu identifikatori svih igrača u njemu. Ovo se vrši pomoću *get_match_details* funkcionalnosti Dota 2 API-a, kojoj je potrebno proslediti identifikator meča koji je u pitanju. Identifikatori mečeva su celobrojna vrednost koja se inkrementira (povaćava za jedan) za svaki meč. Zbog toga je, inkrementiranjem identifikatora početnog meča, moguće pristupiti meču koji je počeo odmah nakon početka prethodnog meča. Daljim inkrementiranjem se preuzimaju mečevi, a zatim i identifikatori igrača. Krajnji rezultat ove faze je datoteka sa nizom identifikatora Dota 2 igrača.

Prilikom preuzimanja identifikatora igrača ignorisani su identifikatori koji imaju vrednost 4294967295, zato što predstavljaju igrače koji nisu dozvolili da njihovi podaci budu javni. Dodatno, uzeti su u obzir samo igrači koji su igrali u mečevima tipa *Ranked Match*. Vođeno je računa i o tome da se u rezultujućem skupu ne pojave duplikati identifikatora.

2) Prikupljanje MMR-a igrača

Za igrače koji su prikupljeni u prvoj fazi dobavljen je MMR pomoću *get_players_account_id_ratings* funkcionalnosti OpenDota API-a. OpenDota API vraća niz JSON objekata u kojima su zabeleženi *Solo* i *Party* MMR zajedno sa identifikatorom i vremenom meča na čijem kraju su zabeležene vrednosti. Primer ovakvih objekata je vidljiv na slici 1. Ovi podaci su često nepotpuni ili skroz nepostojeći za

neke igrače. Zbog toga je moguće izvući MMR od otprilike trećine igrača sakupljenih u prvoj fazi. Kod igrača gde je dostupan više od jednog podatka uzet je najnoviji.

```
{
  "account_id":83573797,
  "match_id":null,
  "solo_competitive_rank":null,
  "competitive_rank":4011,
  "time":"2015-11-18T15:57:17.058Z"
},
{
  "account_id":83573797,
  "match_id":2268559035,
  "solo_competitive_rank":3681,
  "competitive_rank":null,
  "time":"2016-04-03T03:01:08.833Z"
},
}
```

Slika 1 - Primer JSON objekta sa informacijama o MMR-u

Pošto se u ovom radu bavimo estimacijom Solo MMR-a igrača, onda samo i njega uzimamo u obzir (polje *solo_competitive_rank* na slici 1). U isto vreme, MMR jednog igrača se često menja, zbog toga nam je potreban i podatak o meču u kojem je zabeležen MMR. Krajnji rezultat ove faze je datoteka sa podacima o identifikatoru igrača, MMR-u i identifikatoru meča u kojem je taj MMR zabeležen.

3) Prikupljanje mečeva

Da bi upotpunili skup podataka, osim MMR igrača, potrebno je prikupiti i njihovu istoriju odigranih mečeva. Dota 2 API nudi funkcionalnost kojoj je potrebno proslediti identifikator igrača, zajedno sa opcionim parametrima kao što su identifikator meča od kojeg počinje prikupljanje istorije, tip mečeva, ukupan broj mečeva, identifikator heroja kojeg posmatrani igrač mora da igra itd. Mečevi za svakog igrača se dobavljaju prosleđivanjem podataka sakupljenih iz predhodnih faza, zajedno sa parametrom za ukupan broj mečeva koji se dobavlja postavljanim na 100.

Zbog nedostatka Dota 2 API-a, za većinu igrača nije moguće dobiti mečeve na opisani način. Greška nastaje ako je igrač ima više od 500 odigranih mečeva posle meča koji je prosleđen API-u (meč dobavljeni iz druge faze), u kojem se vrati prazan niz mečeva. Igrom slučaja, ova greška ne nastaje ako se specificira i identifikator heroja za kojeg se prikupljaju mečevi. Koristeći ovo rešenje, moguće je prikupiti istoriju mečeva igrača tako što se pošalje po jedan zahtev za svakog heroja u igri, a zatim se dobijeni nizovi spoje, sortiraju po vremenu početka meča i onda iz tog niza uzme prvih 100 mečeva.

Dobavljanje mečeva pojedinačno za svakog heroja je značajno sporije, pre svega zato što je broj zahteva po igraču povećan sa 1 na 117. Zbog toga su u ovoj fazi korišćena oba pristupa kombinovano. Za razliku od predhodnih faza, prikupljanje mečeva je vremenski zahtevan proces. Procenjeno je da je za 3000 igrača potrebno od 2 do 3 dana da bi se ova faza izvršila uspešno. Krajnji rezultat ove faze je JSON datoteka koja zajedno sa identifikatorom i MMR-om svakog

igrača sadrži i njihovu istoriju mečeva koja je predstavljena nizom objekata. Izgled jednog od objekta koji predstavlja meč je vidljiv na slici 2.

```
"match_id":3556849965,
"match_seq_num":3094040961,
"start_time":1510443621,
"lobby_type":7,
"radiant_team_id":0,
"dire_team_id":0,
"players":[{"account_id":4294967295, "player_slot":0, "hero_id":9},
{"account_id":313600850, "player_slot":1, "hero_id":44},
{"account_id":246685252, "player_slot":2, "hero_id":42},
{"account_id":337316935, "player_slot":3, "hero_id":47},
{"account_id":85349036, "player_slot":4, "hero_id":68},
```

Slika 2 - Primer JSON objekta sa informacijama o meču

4) Dobavljanje detalja mečeva

Mečevi dobijeni iz prehodne faze nemaju dovoljno podataka da bi bili korisni za rešavanje problema koji je pretstavljen u ovom radu. Zbog toga je potrebno iskoristiti *get_match_details* funkcionalnosti Dota 2 API-a koja za prosleđeni identifikator meča vraća detaljniji opis istog. Slično kao kod prethodne faze, i ovde je potrebno izvršiti veliki broj zahteva po svakom igraču, s toga je i dobavljanje detalja svakoig meča vremenski zahtevan proces.

Rezultat ove faze je skup JSON datoteka gde svaka sadrži niz igrača zajedno sa detaljnom istorijom mečeva. Podaci su ovaj put podeljeni u više datoteka zato što je u ovoj fazi memorijsko zauzeće značajno poraslo. Skup datoteka kojima odgovara 300 igrača zauzima oko 885 MB memorije. Detaljni prikaz meča sadrži veliki broj dodatnih atributa za svakog igrača, neki od značajnijih su:

- *hero_id* - identifikator heroja kojeg je igrač igrao
- *kills* - broj „ubistava“ koji je igrač ostvario nad drugim igračima
- *deaths* - broj „smrti“ igrača
- *assists* - broj asistencija koje je igrač ostvario, odnosno broj „ubistava“ u kojima je igrač pomogao.
- *last_hits* - broj „ubistava“ koji je igrač ostvario nad ne-igračkim karakterima protivničkog tima.
- *denies* - broj „ubistava“ koji je igrač ostvario nad ne-igračkim karakterima sopstvenog tima.
- *gold_per_min* - prosečan broj prikupljenog „zlata“ za svaki minut meča. Često skraćeno kao *gpm*.
- *xp_per_min* - prosečan broj prikupljenog „iskustva“ za svaki minut meča. Često skraćeno kao *xpm*.
- *level* - poslednji dostignuti nivo heroja na kraju meča.
- *item_1*, *item_2* ... *item_6* - svaki od ovih atributi sadrži identifikator *item*-a (predmeta) koji je bio u igračevom posedu na kraju meča.

B. Priprema podataka

Od inicijalno prikupljenog seta podataka bilo je neophodno napraviti ciljni set koji će biti korišćen u daljim analizama. S obzirom na prirodu i obim podataka, napravljeno je pet različitih setova koji se razlikuju po tome koje elemente igre pokrivaju, pri čemu se svi setovi baziraju na standardnim obeležjima (*account_id*, *mmr*, *wins*, *loses*, *gpm*, *xpm*, *kills*, *deaths*, *assists*, *last_hits*, *denies*, *level*):

- *Simple set* – U sebi sadrži gore pomenuta obeležja, pri čemu su *gpm*, *xpm*, *kills*, *deaths*, *assists*, *last_hits*, *denies*, *level* proseci kroz ukupan broj mečeva (od 50 do 100) za svakog zabeleženog igrača. Kroz iteraciju mečeva akumulirane su vrednosti za svako obeležje gde je nakon iteracija izračunat prosek tako što su vrednosti podeljene sa ukupnim brojem zabeleženih mečeva (*recorded_games*).
- *All heroes set* – Predstavlja nadogradnju *Simple seta* u takvom obliku da su posmatrani svi heroji koji figuriraju u samoj igri. Broj heroja koji postoji je 117 i za svakog su beleženi konkretni proseci (npr. *axe_wins*, *axe_loses*, *axe_gpm*, *axe_xpm* itd.). Ovakav set podataka ima za nedostatak čestu pojavu nula jer je ukupan broj heroja koji postoji u igri veći od broja zabeleženih mečeva. Kod manje raznovrsnih igrača (onih koji mnogo često igraju jednog heroja – najčešće samo onih koje znaju), broj kolona koji sadrži isključivo nule je preko 90%.
- *All roles set* – je proširena verzija *simple seta* na način koji je drugačiji od prethodno opisanog, odnosno koji zaobilazi probleme sa velikom količinom nula. Umesto što se posmatra svaki heroj, posmatraju se grupe heroja, koje su određene prema rolama koje sama igra definiše [1]:
 - 1) *Carry* – Tip heroja koji postane efektivan u završnoj fazi partije, kome je neophodno nekoliko *item-a* (puno *gold-a*) da bi bio efektivan pri čemu je najčešći izvor *damage-a* običan napad (*right click*).
 - 2) *Nuker* – Heroji koji imaju magije (sposobnosti) koje nanose velike količine *damage-a* u trenutku.
 - 3) *Initiator* – Ovakav tip heroja najčešće započinje timske borbe jer ga karakteriše izdržljivost ili posedovanje neke magije koja zahvata neku površinu (*Aoe*) pri čemu je efekat magije *damage* ili *stun*.
 - 4) *Disabler* – Osobine su im posedovanje magija koje za cilj imaju da onesposobe (*stun-uju*) protivničkog heroja.
 - 5) *Durable* – Glavna osobina im je izdržljivost, odnosno dugotrajnost u timskoj borbi.
 - 6) *Escape* – Tip heroja koji poseduje magije koje imaju efekat da u trenutku pređu određenu distancu, pri čemu je najčešća upotreba izvlačenje iz neugodnih situacija.

- 7) *Support* – Suprotno od onoga što predstavlja *Carry*. Najveća efektivnost heroja se ispoljava u ranoj fazi partije pri čemu nisu neophodni *item-i* da bi vrline tog heroja došle do izražaja. Primarna uloga je da ostalim herojima bude podrška tokom igre.
- 8) *Pusher* – Posедуje sposobnosti da brzo nanosi *damage* zgradama, koje je neophodno uništiti.
- 9) *Jungler* – Ovi heroji ranu fazu igre najviše provode u delu mape koji predstavlja džunglu. Karakteriše ih mogućnost da ubijaju neutralne *creep-ove* od samog početka partije.

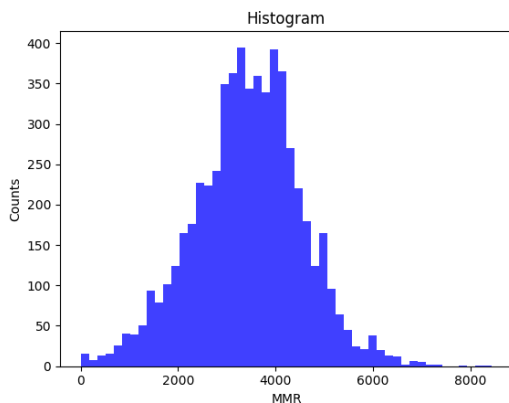
Pored navedenih uloga, u razmatranje je uzeta kategorija *Complexity* kojom igra labelira one heroje koji su teški za igranje. Obeležja su formirana po sličnom principu (npr. *carry_wins*, *carry_loses*, *carry_gpm*, *carry_xpm* itd.), s tim što se na ovaj način zaobilazi problem postojanja velikog broja nula, jer postoji 10 (9+1) kategorija (rola) naspram 117 heroja. Retka je pojava da neko u velikoj količini mečeva u potpunosti izostavi igranje heroja nekog tipa, pogotovo ako se uzme u obzir da konkretan heroj može imati više uloga istovremeno. Sama igra [12] definiše još nivoa unutar uloga (zvezde) kojima se govori u kojoj meri neki heroj ima osobine koji karakterišu određenu rolu. Najviši nivo predstavljaju tri zvezdice, dok oni sa nižim nivoem (jedna ili dve zvezdice) neke osobine ne poseduju ili se ispoljavaju samo u određenim situacijama. Pri pravljenju ciljnog seta, uzeti su heroji sa dve i tri zvezde.

- *All roles and items set* – Predstavlja proširenje gore pomenutog ciljnog seta, jer su u razmatranje ubačeni i *item-i* (predmeti koji poboljšavaju heroja) koji su beleženi za svaku ulogu. Pošto za *item-e* nema poente računati prosek, za njih je beležen broj kupovina kroz posmatrani skup mečeva (broj *item-a* koje heroj može u jednom trenutku nositi sa sobom je devet).
- *All matches one by one set* - Ovaj ciljni set je formiran na drugačiji način, gde se beleženje podataka ne zasniva na računanju proseka, već se za svaki meč beleže podaci (*gpm*, *xpm*, *kills*, *deaths*, *assists*, *last_hits*, *denies*, *level*) o igraču. Ali osim posmatranog igrača, na ovaj način je moguće zabeležiti podatke o svim igračima u meču. Ovo može da daje bolji uvid o MMR posmatranog igrača delom i na osnovu nivoa veštine njegovih protivnika i saveznika. Dodatno, posmatrani su i igrani heroji kao i kupljeni itemi, ova obeležja su predstavljena pomoću *one-hot encoding-a*. Osim podataka o pojedinačnim igračima, zabeleženi su podaci o pobedničkom timu i trajanju meča.

C. Analiza podataka

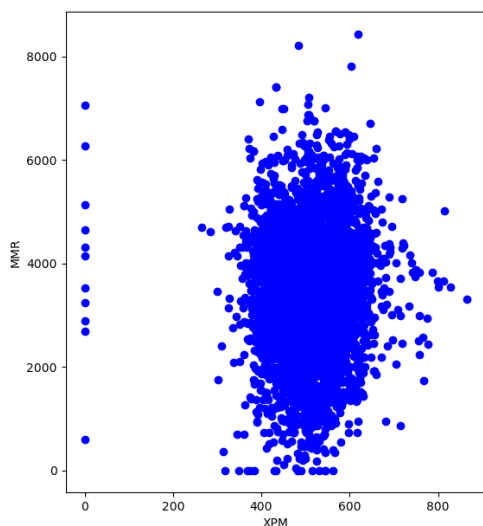
Ciljno obeležje je MMR i ukoliko se pogleda histogram sa slike 3 može da se uoči da set ima najviše igrača sa MMR od 3000 do 4500 što odgovara i stvarnom stanju u igrici koje je

posledica korišćenja modifikovanog elo siststema [10] za rangiranje igrača.



Slika 3 - Histogram broja igrača po MMR-u

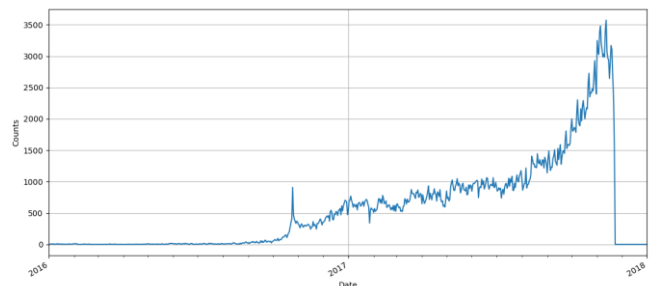
Sa iste slike, kao i sa slike 4, koja pokazuje odnos MMR-a i XMP-a, možemo videti da imamo oko 10 igrača čiji je MMR jednak 0. Ovi igrači predstavljaju grešku u podacima pošto matchmaking sistem ne dozvoljava da igrači imaju MMR ispod 1. Iz ovog razloga je odlučeno da se ovi igrači uklone prilikom obučavanja modela. Takođe, na slici 4 se može videti i da postoji par igrača čiji je XPM jednak 0. I na ove igrače se gleda kao na grešku u podacima, iako za njih postoji potencijalno šansa da su tokom prikupljenih partiji bili neaktivni sve vreme. Bez obzira na razlog i ovi igrači su izbačeni tokom obučavanja modela.



Slika 4 - XPM igrača različitog MMR-a

Da bi podaci bili što relevantniji gledalo se da oni budu iz istog patch-a. Svaki patch sa sobom nosi promene kojima se određeni heroji *buffuju* odnosno pojačavaju, dok se drugi *nerfuju* odnosno oslabljuju. Ovo je posledica toga što *developer* pokušavaju da uravnoteže igricu, kao i da spreče dominaciju jednog tima na e-sport sceni. Broj mečeva po datumu iz seta se može videti na slici 5. Može se primetiti da

je najveći broj prikupljenih mečeva odigrano u oktobru i novembru 2017. godine.



Slika 5 - Broj mečeva po datumu

IV. METODOLOGIJA

U ovoj sekciji su opisani modeli koji su korišćeni za estimaciju MMR-a Dota 2 igrača. Da bi se dostigao što bolji rezultat, isproban je veliki broj modela koji rade nezavisno jedan od drugog. Sekcija je organizovana u tri poglavlja. Prvo poglavlje nabraja regresione modele koji su korišćeni, zajedno sa opisom načina na koji su primenjeni na naš problem. U drugom poglavlju su nabrojene korišćene neuronske mreže zajedno sa opisom njihovih arhitektura i procesa treniranja. U trećem poglavlju je opisan sistem neuronskih mreža koje međusobno sarađuju da bi vršili predikciju.

A. Predviđanje MMR-a putem regresionih modela

Za predikciju *matchmaking rating*-a je korišćeno više regresionih modela sa ciljem postizanja boljih performansi. Upotrebljeni su:

- Linearna regresija,
- Linearna regresija sa *lasso* regularizacijom,
- Linearna regresija sa *ridge* regularizacijom,
- Linearna regresija sa *elastic net* regularizacijom,
- *AdaBoost* ansambl regresionih stabala.

Nakon učitavanja podataka vršena je normalizacija upotrebom *StandardScaler*-a iz *sklearn* [13] biblioteke gde je potom odrađena unakrsna validacija modela sa podelom trening i test skupa u odnosu 4:1 sa pet iteracija. Kao mere greške su korišćene mere *MAE* (*mean absolute error*) i *RMSE* (*Root mean square error*) kao najčešće korišćene metrike za meru preciznosti [14].

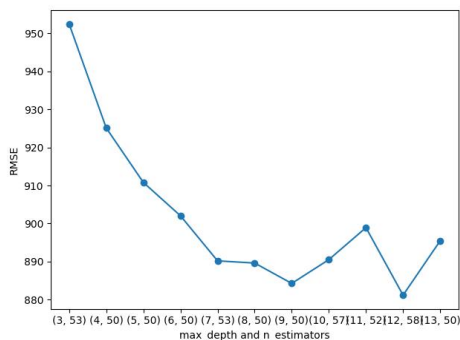
Navedeni modeli su upotrebljeni na četiri od pet skupova podataka, odnosno konkretno na *simple*, *all heroes*, *all roles* i *all roles and items* skupovima (koji su opisani u delu za pripremu podataka), dok je skup *all matches one by one* upotrebljavan za obučavanje neuronske mreže.

Optimizacija parametara je rađena za *elastic net* linearnu regresiju, dok za *lasso* i *ridge* nije, pošto isprobavanjem različitih vrednosti za ove modele nismo došli do zaključka da bi optimizacija parametara dovela do boljih rešenja. Optimizacija je vršena prema *alpha* i *l1_ratio* parametrima. Najbolje dobijene vrednosti za nevedene parametre za *simple*

set podataka su redom 0.14 i 0.95, dok za *all heroes* i *all roles* vrednosti iznose 0.88 i 0.95, odnosno 0.62 i 1.00. Za *all roles and items* set podataka je dobijena najbolja vrednost kod unakrsne validacije, pri čemu dobijene optimalne vrednosti za navedene parametre iznose 0.50 i 0.52.

Takođe je vršena optimizacija parametara i za model regresionih stabala u *AdaBoost* ansamblu, gde je optimizacija vršena prema broju estimatora (*n_estimators*) i dubine regresionog stabla (*max_depth*). Za *simple* skup podataka, dobijene vrednosti su 86 estimatora i dubina 11. Optimalan broj estimatora i dubina za *all heroes* i *all roles* setove iznose 97 i 10, odnosno 58 estimatora i dubina 12, dok za *all roles and items* te vrednosti iznose 66 i 12.

Na slici 6 prikazan je primer toka optimizacije parametara za model regresionih stabala u *AdaBoost* ansamblu za *all roles* skup podataka.



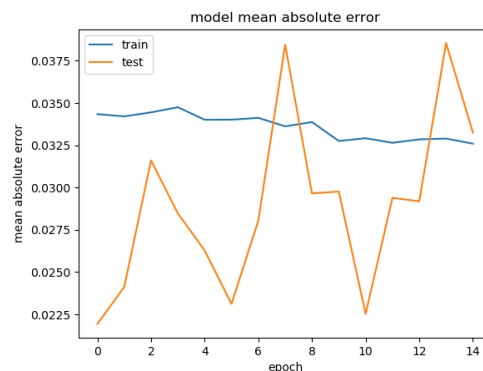
Slika 6 – AdaBoost optimizacija parametara

B. Neuronske mreže

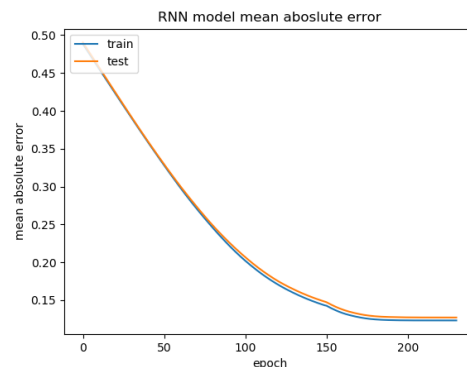
U ovom poglavlju su opisane tri neuronske mreže koje su trenirane za predikciju MMR-a igrača. Sve Neuronske mreže su implementirane uz pomoć *keras* [2] i *tensorflow* [3] biblioteka. Mreže su trenirane nad skupom podataka za treniranje, a rezultati su provereni nad testnim skupom. Treniranje i test skup su podeljeni u odnosu 4:1.

Prva mreža je trenirana na *all roles and items* setu podataka. Ova mreža se sastoji iz jednog konvolutivnog sloja (*Conv1D*), jednog *max pooling* sloja i tri potpuno-povezana međusloja na kraju. Mreža je tokom treniranja pokazala veoma dobre rezultate već i u prvoj epohi treniranja. Na slici 7 su prikazane performance mreže tokom treniranja. Mreža nije trenirana kroz veliki broj etapa zato što veoma brzo konvergira. Za treniranje je korišćen *Adam* [15] optimizator, dok je *MAE* korišćen kao funkcija gubitka. Ovako jaka oscilacija greške nad testnim skupom podataka se može pripisati i maloj količini podataka u njemu (a i generalno).

Istorija mečeva igrača može da se posmatra kao sekvenca podataka. Druga neuronska mreža iskoriscava ovu osobinu pomoću svoje arhitekture i trenirana je na *all matches one by one* setu. Ona se sastoji od dva LSTM (*Long short-term memory*) rekurentna sloja. Treniranje rekurentne mreže je trajalo znatno duže od konvolutivne, što se pre svega može pripisati i samoj količini podataka, koja je u ovom slučaju mnogo veća. Na slici 8 su prikazani podaci o istoriji treniranja mreže.

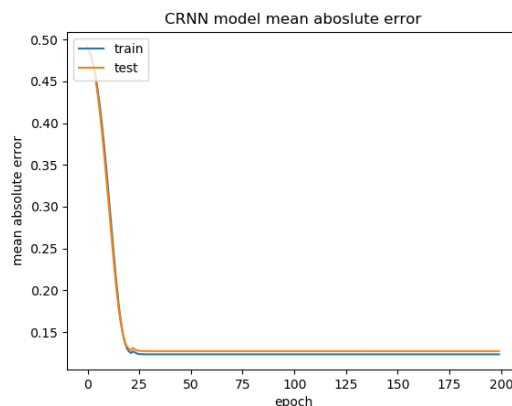


Slika 7 – Podaci o istoriji treniranja konvolutivne neuronske mreže



Slika 8 – Podaci o istoriji treniranja rekurentne neuronske mreže

Poslednja implementirana mreža je takođe rekurentna mreža koja je trenirana na *all matches one by one* setu podataka. Mreža se sastoji iz dva para konvolutivnog i *max pooling* sloja na početku, jednog rekurentnog LSTM sloja i na kraju jednog potpuno povezanog sloja. Na slici 9 se vidi da je ova mreža mnogo brže konvergirala od predhodne, ali krajnji rezultat je veoma sličan u oba slučaja.



Slika 9 – Podaci o istoriji treniranja konvolutivne, rekurentne neuronske mreže

C. Sistem neuronskih mreža

U ovom poglavlju je opisan sistem koji se sastoji iz tri neuronske mreže. Prve dve su konvolutivne mreže namenjene za izvlačenje značajnih atributa iz podataka i za redukciju dimenzionalnosti. Treća, glavna, je rekurentna mreža koja kao ulaz prima podatke koji su kombinacija izlaza konvolutivnih mreža i manjeg podskupa originalnih podataka. Na kraju se iz rekurentne mreže dobija predikcija o MMR-u igrača kao povratna vrednost.

Obe konvolutivne mreže su trenirane nezavisno od ostatka sistema, u sklopu autoenkoder arhitekture [18]. Prva mreža je obučavana kao enkoder za skup obeležja koji predstavlja kupljene *item*-e jednog igrača unutar nekog meča. Mreža je sačinjena od tri jednodimenzionalna konvolutivna sloja (*Conv1D*) i tri *max pooling* sloja. Druga konvolutivna mreža je trenirana kao enkoder za podatke o svih pet igrača unutar jednog tima nekog meča. Ona se sastoji iz tri dvodimenzionalna konvolutivna sloja (podaci o timu su predstavljeni u obliku matrice), i tri *max pooling* sloja.

Rekurentna mreža je trenirana na *all matches one by one* setu, s tim da su podaci svakog meča prvo provedeni kroz konvolutivne neuronske mreže. Ovim je značajno smanjena dimenzionalnost ulaza neuronske mreže, čime je i omogućeno brže treniranje. Mrežu čine dva LSTM sloja.

V. REZULTATI

Pre početka treniranja modela, napravljena je podela na skup podataka koji će biti korišćen za obučavanje istih, kao i za optimizaciju parametara i na podatke koji će služiti za evaluaciju istreniranih modela, pri čemu su od oba skupa nastali ciljni skupovi na način kako je to opisano u poglavlju za pripremu podataka.

Kao mere greške korišćeni su već pomenuti *MAE* i *RMSE*. Obe mere predstavljaju vrednost greške u istoj jedinici kao i ciljno obeležje. Vrednosti greške koje su dobijene su dosta velike u odnosu na prirodu MMR-a, gde svaki odigrani meč povećava, odnosno smanjuje, MMR igrača za 25, shodno ishodu meča. Ono što se može zaključiti je da prosečna greška u predikciji rejtinga oko 680, što je manje od 10% od posmatranog opsega (1-7000+). U tabeli 1 su prikazani rezultati greške validacije regresionih modela nad validacionim skupom po pitanju MAE metrike.

Konkretno vrednosti su dobijene za optimalne parametre nad trening skupom kada su u pitanju *AdaBoost* i *Elastic net*, dok su za druge regresione modele (*Ridge* i *Lasso*) korišćeni vrlo slični parametri koji daju za neke skupove podataka identične rezultate (nisu dobijeni bolji od ovih). Najbolji rezultat među, regresionim modelima, je postignut linearnom regresijom sa *elastic net* regularizacijom za ciljni skup

TABELA I – VREDNOSTI GREŠKE (MAE) NAD VALIDACIONIM SKUPOM

Skupovi podataka	Regresioni modeli (MAE)				
	<i>Linear</i>	<i>Lasso</i>	<i>Ridge</i>	<i>Elastic net</i>	<i>AdaBoost</i>
<i>Simple</i>	727.3048	727.2561	727.2991	726.1295	738.3382
<i>All heroes</i>	695.5997	668.3724	694.2278	651.9068	695.9210
<i>All roles</i>	659.0073	658.3226	658.9900	658.3226	693.4146
<i>All roles items</i>	698.8810	654.9423	698.6068	614.3647	673.9113

TABELA 2 – VREDNOSTI GREŠKE (RMSE) NAD VALIDACIONIM SKUPOM

Skupovi podataka	Regresioni modeli (RMSE)				
	<i>Linear</i>	<i>Lasso</i>	<i>Ridge</i>	<i>Elastic net</i>	<i>AdaBoost</i>
<i>Simple</i>	946.0183	945.8562	946.0039	942.6634	900.6323
<i>All heroes</i>	1014.488	962.2716	1011.655	874.3338	951.7846
<i>All roles</i>	905.2223	907.2657	905.1446	907.2657	903.6686
<i>All roles items</i>	937.7040	873.0970	937.0870	791.6570	880.0053

podataka koji posmatra heroje po ulogama, uključujući i podatke o *item*-ima. U tabeli 2 su prikazane vrednosti greške za drugu metriku, odnosno *RMSE*.

Iz ovih rezultata se vidi da skup podataka igra značajnu ulogu prilikom predikcije. Skupovi u kojima su predstavljeni heroji i itemi su pokazali bolje rezultate u odnosu na *simple* skup u svim slučajevima. Na osnovu razlike u vrednostima MAE i RMSE mere, može se zaključiti da modeli često prave jako velike greške na nekim podacima validacionog skupa.

Neuronske mreže nisu pokazale značajno bolje rezultate. Od svih ispitanih modela, najbolji rezultat je dobijen pomoću konvolutivne neuronske mreže trenirane na *all roles items* setu podataka. S druge strane, rekurentne mreže trenirane nad *all matches one by one* setu podataka su pokazale najgore rezultate od svih ispitanih modela. Iz ovih rezultata je ponovo uočeno da arhitektura mreže ne utiče značajno na poboljšanje u preciznosti i da mnogo veći značaj nosi korišćeni set podataka. Na tabeli 3 su prikazani rezultati neuronskih mreža u metrikama MAE, RMSE i R^2

TABELA 3 – VREDNOSTI GREŠKE ZA RAZLIČITE METRIKE NAD VALIDACIONIM SKUPOM

Mera greške	Tip neuronske mreže / skup podataka			
	CNN / <i>All roles items</i>	RNN / <i>One by one</i>	CRNN / <i>One by one</i>	CNN-RNN stack / <i>One by one</i>
MAE	582.0014	897.3104	897.4528	897.3810
RMSE	769.0652	1152.6550	1152.9523	1152.8049
R^2	0.5426	-0.0016	-0.0021	-0.0018

VI. ZAKLJUČAK

U ovom radu je rešavan problem predikcije MMR-a igrača u igri Dota 2 na osnovu istorije odigranih mečeva. Opisan je proces prikupljanja i pripreme podataka za potrebu treniranja modela mašinskog učenja. U sklopu pripreme podataka su predstavljena 4 tipa izvlačenja atributa u vidu računanja proseka odigranih mečeva. Takođe je predstavljen i skup podataka u kojem je svaki meč posmatran pojedinačno. Nad pomenutim skupovima podataka je trenirano više regresionih modela i neuronskih mreža da bi se pronašao model koji daje najbolje rezultate.

Iz rezultata u radu se može uočiti da način na koji su podaci predstavljeni igra najznačajni uticaj na performance. Najbolje rezultate su pokazali *all roles* i *all roles items* setovi

podataka radu, koji uz dovoljno dobar model mogu da dostignu MAE od ispod 600. Ali čak i u najboljem slučaju, greška dobijena prilikom validacije je prevelika da bi se moglo reći da sistem može uspešno da predvidi MMR nekog igrača. Iz ovoga se može pretpostaviti da kompleksnost igre kao što je Dota 2 nije moguće predstaviti samo uz pomoć statističkih podataka koji su korišćeni u ovom radu.

Dalje unapređenje bi se moglo dostići pronalaženjem boljih načina da se prikupljeni skup podataka predstav. Primer bilo prikupljanje podataka o odigranim mečevima u vidu parsiranja snimaka unutar Dota 2 klijenta. Ovim bi se dobio bolji uvid o samom ponašanju igrača u toku meča. Takođe bi se dobili podaci iz svih trenutaka meča, a ne samo na kraju kao što je u podacima koji su korišćeni u radu.

VII. REFERENCE

- [1] <http://blog.dota2.com>, [Poslednji pristup 4. maj 2019.].
- [2] <https://keras.io/>, [Poslednji pristup 6. maj 2019.].
- [3] <https://www.tensorflow.org/>, [Poslednji pristup 6. maj 2019.].
- [4] <http://dota2.prizetrac.kr/international2018>, [Poslednji pristup 4. maj 2019.].
- [5] A. Drachen, M. Yancey, J. Maguire, D. Chu, I. Y. Wnag, T. Mahlmann, M. Schubert, D. Klabjan, "Skill-Based Differences in Spatio-Temporal Team Behavior in Defence of The Ancients 2 (DotA 2)"
- [6] A. Sapienza, H. Peng, E. Ferrara, "Performance Dynamics and Success in Online Games", arXiv:1801.09783v1 [cs.SI] 29 Jan 2018
- [7] N. Caplar, M. Suznjovic, M. Matijasevic, "Analysis of player's in-game performance vs rating: Case study of Heroes of Newerth", arXiv:1305.5189v1 [physics.soc-ph] 22 May 2013.
- [8] <http://www.heroesofnewerth.com/>, [Poslednji pristup 4. maj 2019.].
- [9] R. Lutz, "Fantasy Football Prediction", arXiv:1505.06918v1 [cs.LG] 26 May 2015
- [10] A. Elo, "The Rating Of Chess Players, Past and Present", New York: Arco Publishing, 1975.
- [11] R. Herbrich, T. Minka, i T. Graepel. "TrueSkill TM: A Bayesian Skill Rating System", Technical Report MSR-TR-2006-80, Microsoft Research, June 2006.
- [12] https://liquipedia.net/dota2/Hero_Roles, [Poslednji pristup 5. maj 2019.].
- [13] <https://scikit-learn.org/stable/>, [Poslednji pristup 5. maj 2019.].
- [14] A. Botchkarev, "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology"
- [15] DP. Kingma, J. Ba Adam, "A method for stochastic optimization", arXiv preprint.
- [16] <https://github.com/joshuauffv/dota2api> [Poslednji pristup 4. maj 2019.].
- [17] <https://docs.opendota.com/> [Poslednji pristup 4. maj 2019.].
- [18] G.E. Hinton i R.R. Salakhutdinov. "Reducing the dimensionality of data with neural networks". Science, 2006 Jul, pp.504-507.