

Predikcija konačnog plasmana u igri PUBG

Ognjen Vladislavljević SW-38/2015, Luka Bilić SW-49/2015, Damir Hadžić SW-41/2015

1. Motivacija

Kako su Battle Royale igrice postale popularne, tako se otvorio prostor za nova istraživanja vezana i za samu igru. Cilj ove vrste igara je da od 100 igrača koji se nalaze na ostrvu, ostanu kao poslednji preživeli. Da bi se stiglo po pobjedničkoj mjestu, potrebno je da pronađu oružja, vozila i sve ostale predmete kojima se olakšava eliminisanje protivničkih igrača. Zbog same kompleksnosti igre, postoji mnoštvo parametara koji se menjaju u zavisnosti od strategije za koju se igrač opredelio. Nas motiv je bio da analizom tog mnoštva parametara pokušamo tačno da prediktuju pobjednika jedne partije. Gorenavedeni parametri su javno dostupni za preko 65000 igara. Nas cilj je da ih obradimo i obucimo naš model da radi predikciju na osnovu tih igara.

2. Research questions

- **DBNOs** - Broj osamucenih (knocked) protivnika
- **assists** - Broj protivnika koje je dati igrač pogodio pre nego što su ubijeni od strane saigrača.
- **boosts** - Broj iskoriscenih predmeta za povećanje stete
- **damageDealt** - Ukupan broj skinutih poena drugim igračima.
- **headshotKills** - Broj ubijenih protivnika pogotkom u glavu.
- **heals** - Broj iskoriscenih predmeta za povećanje životnih bodova.
- **Id** - ID igrača.
- **killPlace** - pozicija u meču po broju ubistava.
- **killPoints** - Nezavisno rangiranje igrača na osnovu broja ubistava. Ako je vrednost rankPoints-a razlicita od -1, onda bi svaka 0 u killPoints trebala da se tretira kao "None".
- **killStreaks** - Najveći niz ubistava u kratkom vremenskom periodu.
- **kills** - Broj ubistava protivničkih igrača. prevodi kenja neće ni gledati
- **longestKill** - Najveća udaljenost između igrača, i protivnika u vreme smrti. Ovo može da zavara, posto osamucivanje igrača pracoeno voznjom kola u suprotnom smeru vodi do visokog longestKill parametra.
- **matchDuration** - Duzina meca u sekundima.
- **matchId** - ID meca. Nema istih meceva u train i test skupu podataka.
- **matchType** - String koji određuje rezim igre. Standardni rezimi su "solo", "duo", "squad", "solo-fpp", "duo-fpp" i "squad-fpp"; svi ostali rezimi su iz event-ova koji se održavaju.
- **rankPoints** - Rangiranje igrača na osnovu parametara razlicitih od samih pozicija osvajanih u prethodnim mecevima.
- **revives** - Broj ozivljavanja saigrača.
- **rideDistance** - Ukupna udaljenost predjena vozilom merena u metrima.
- **roadKills** - Broj ubistava iz vozila.
- **swimDistance** - Udaljenost preplivana u metrima.
- **teamKills** - Broj ubijenih saigrača.
- **vehicleDestroys** - Broj unistenih vozila.
- **walkDistance** - Ukupna prepesacena distanca u metrima.
- **weaponsAcquired** - Ukupan broj pokupljenih oružja.
- **winPoints** - Rangiranje igrača koje se zasniva na pobjedama.

- **groupId** - ID grupe saigraca u partiji. Ako ista grupa igaca igra vise razlicitih meceva, uvek ce imati razlicit ID.
- **numGroups** - Broj grupa u partiji.
- **maxPlace** - Najgori rang u mecu.
- **winPlacePerc** - Ovu vrednost treba da prediktujemo. Predstavlja procenat pozicije igraca u partijama. 1 odgovara prvom mestu dok 0 odgovara poslednjem mestu u partiji. Racuna se na osnovu maxPlace kolone.

3. Related work

Vecina radova na ovu temu se zasnivaju na nekom tipu EDA(exploration data analysis) i regresiji.

4. Methodology

Samu predikciju smo vrsili ensamble random forest modelom. Medjum, neki od prvobitnih podataka koje smo preuzeli su uticali negativno na sam model koji obucavamo, te smo morali da vrsimo preprocesiranje podataka. Za ovaj zadatak smo vrsili 2 stvari: uklanjanje outliersa i label encoding za neke analiticke podatke. Radjeno je visestruko uklanjanje outliersa po odredenim kriterijuma.

Kriterijumi su vecinom bili vezani za igrace koji koriste *cheat*:

- Ubista bez pomeranja
- Velik broj ubistava iz kola
- Velik broj ubistava
- Ubistva sa prevelikih udaljenosti
- Igraci koji su previse setali
- Igraci koji su previse vozili
- Igraci koji su previse plivali
- Igraci koji su promenili vise od 80 oruzja
- Igraci koji su koristili previse predmeta za lecenje

Obelezja kojima je bio potreban label encoding su matchId i groupId. Za ova dva obelezja je vrsena dodela kategorije, da bi se izbeglo pojavljivanje komplikovanih stringva koji sadrze slova i brojeve, i tako dobila jednostavna numericka vrednost.

Dobili smo inspiraciju za dodavnjem novih obelezja na nekim od radova sa slicnom temom. Ova obelezja su izvodjena iz postojećih obelezja iz data set-a. Novo izvedena obelezja su:

- **playersJoined** - grupisanje svih igraca sa istim matchId-jem. Potrebno za normalizaciju svih ostalih novoizvedenih obelezja.
- **killsNorm** - broj ubistava normalizovan u odnosu na playersJoined.
- **damageDeltNorm** - nacinjena steta normalizovana u odnosu na playersJoined.
- **maxPlaceNorm** - najbolje rangiranje normalizovano u odnosu na playersJoined.
- **matchDurationNorm** - duzina meca normalizovana u odnosu na playersJoined.

5. Discussion

Obzirom da su dataset-ovi ogromni, uzimali smo uzorak nasumičnih redova(sample) iz test/trening skupa. Time smo dobili drugacije i realnije rezultate prilikom svakog pokretanja. Za validaciju smo koristili cross-validation, tako što smo uzimali određeni procenat trening skupa(12%). Za evaluaciju smo koristili MAE(Mean absolute error). Poredili smo tačne vrednosti iz test skupa sa prediktovanim vrednostima. Povećanjem `n_estimators` parametra smo dobili malo bolje rezultate, kao i promenom `max_features` parametra. Međutim rezultati su zbog prethodnog pretprocesiranja podataka već bili dovoljno dobri i bez dodatnog podešavanja parametara.

6. References

- <http://benalexkeen.com/mapping-categorical-data-in-pandas/>
- <https://www.kaggle.com/>