

FIFA 19 - PREDVIĐANJE POZICIJE FUDBALERA NA OSNOVU VREDNOSTI ATRIBUTA

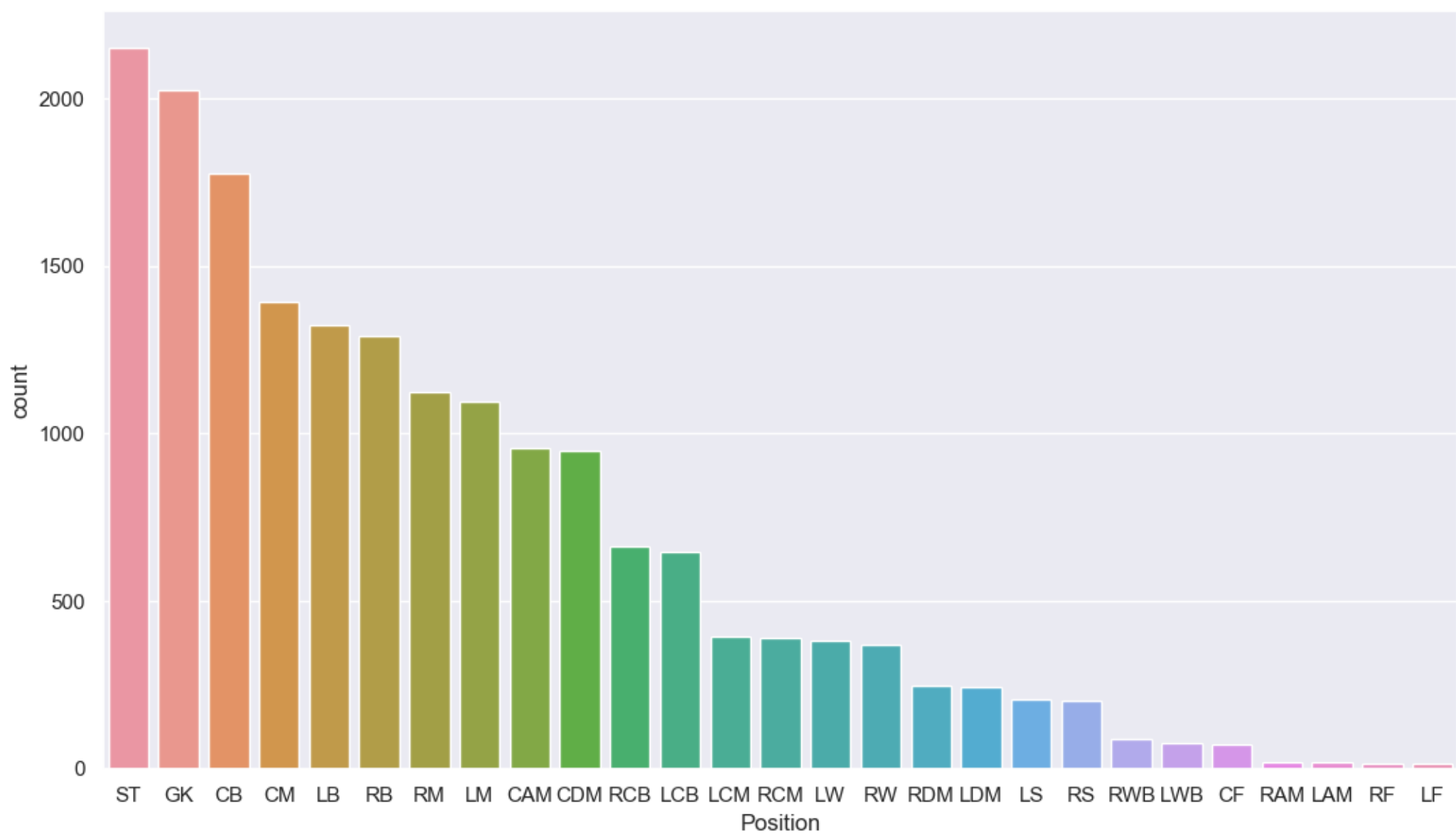
Uroš Ogrizović
Fakultet tehničkih nauka



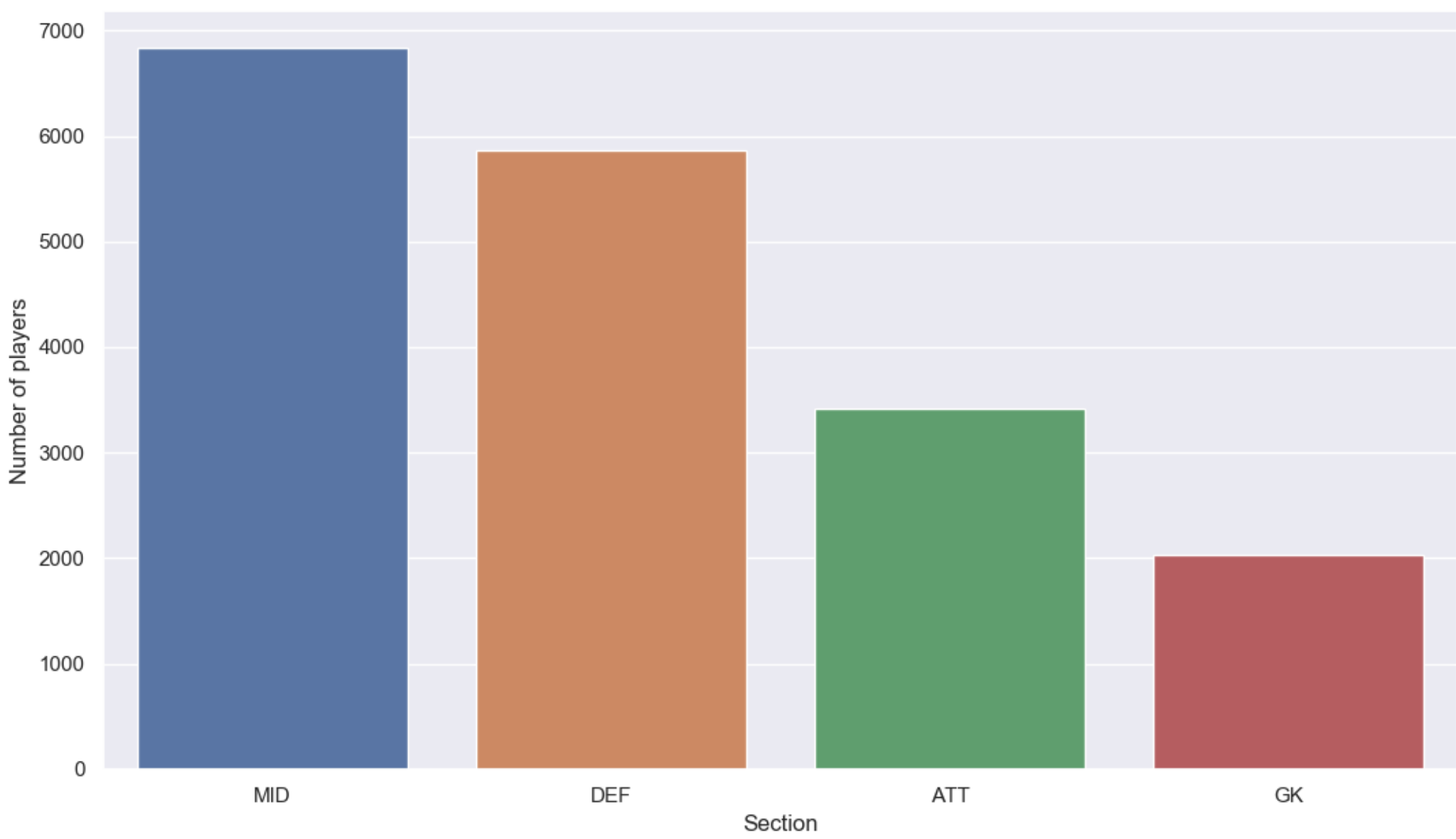
Uvod

FIFA je najpopularnija franšiza fudbalskih video igara na svetu. Prodato je preko 20 miliona primeraka FIFE 19, što je čini najprodavanijom konzolnom video igrom u Evropi u 2018. godini [1]. Problem koji se rešava je klasifikacione prirode. Konkretnije, treba predvideti poziciju fudbalera u video igri FIFA 19 na osnovu vrednosti atributa tog fudbalera.

Sekcije umesto pozicija



Broj igrača po pozicijama. Uočljiva je nebalansiranost skupa podataka.



Broj igrača po sekcijama. Iako je i ovaj skup podataka nebalansiran, on je u mnogo manjoj meri nebalansiran od skupa podataka sa slike. Usled nebalansiranosti skupa podataka, pokušano je sa klasifikacijom igrača po sekcijama. Uočene su četiri sekcije:

1. napad (ATT)
2. sredina (MID)
3. odbrana (DEF)
4. golman (GK)

Algoritam

- „Sređivanje“ *dataset*-a:
 - Redukcija dimenzionalnosti - crtanjem matrice korelacije je uočeno koja obeležja nisu u korelaciji ni sa jednim drugim obeležjem, pa su ta obeležja odstranjena.
 - Odstranjivanje *null* vrednosti iz skupa podataka. Ono se vrši tako što se *null* vrednost zameni nekom drugom vrednošću, a to može da se vrši na tri načina:
 1. ubaci se 1
 2. ubaci se srednja vrednost kolone
 3. ubaci se medijana kolone
 - Enkodovanje kategoričkih obeležja - korišćen je *one-hot encoding*.
- Testirana su dva modela:
 - Dvoslojna neuronska mreža (*Artificial Neural Network*, ANN) - ulaz u mrežu je skup atributa fudbalera, a izlaz iz mreže je lista verovatnoća. Ta lista verovatnoća ima onoliko elemenata koliko pozicija/sekcija postoji, te vrednost na *i*-tom indeksu predstavlja verovatnoću da igrač igra *i*-tu poziciju/sekciju. Na tu listu se primeni *argmax* funkcija, to jest, vrati se pozicija/sekcija koja odgovara najvećoj verovatnoći iz liste. U slučaju predviđanja pozicije igrača, mreža u izlaznom sloju ima 27 neurona, a u slučaju predviđanja pozicije igrača, mreža u izlaznom sloju ima 4 neurona.
 - *Random forest* [2]

Rezultati

Korišćen je *batch size* = 100, zbog toga što je uočeno da promena veličine *batch size*-a ne utiče na *accuracy*. Sva tri načina zamenjivanja *null* vrednosti su dala iste rezultate. Za *Random forest* model je korišćeno 100 stabala odlučivanja.

Model	Position train acc.	Position test acc.	Section train acc.	Section test acc.	Broj epoha
ANN	51%	49%	87%	88%	10
ANN	54%	54%	88%	90%	100
<i>Random forest</i>	99%	27%	100%	89%	-

```
Predicted section: DEF, Actual section: DEF
Predicted section: MID, Actual section: MID
Predicted section: MID, Actual section: MID
Predicted section: MID, Actual section: ATT
Predicted section: MID, Actual section: MID
Predicted section: GK, Actual section: GK
Predicted section: DEF, Actual section: DEF
Predicted section: MID, Actual section: MID
```

Nekoliko predviđanja sekcija

Zaključak

Usled nebalansiranosti skupa podataka je teško predvideti poziciju fudbalera. Umesto toga, mnogo bolji rezultati su postignuti u predviđanju sekcije koju fudbaler igra. ANN i *Random forest* su se pokazali kao jednako dobri u rešavanju ovog problema.

Reference

References

- [1] <https://gadgets.ndtv.com/games/news/fifa-19-sales-flat-thanks-to-fifa-18-says-ea-1989300>
- [2] https://www.youtube.com/watch?v=J4Wdy0Wc_xQ