

FIFA 19 - PREDICTING THE POSITIONS OF PLAYERS BASED ON THEIR ATTRIBUTE VALUES

Uroš Ogrizović
Faculty of Technical Sciences

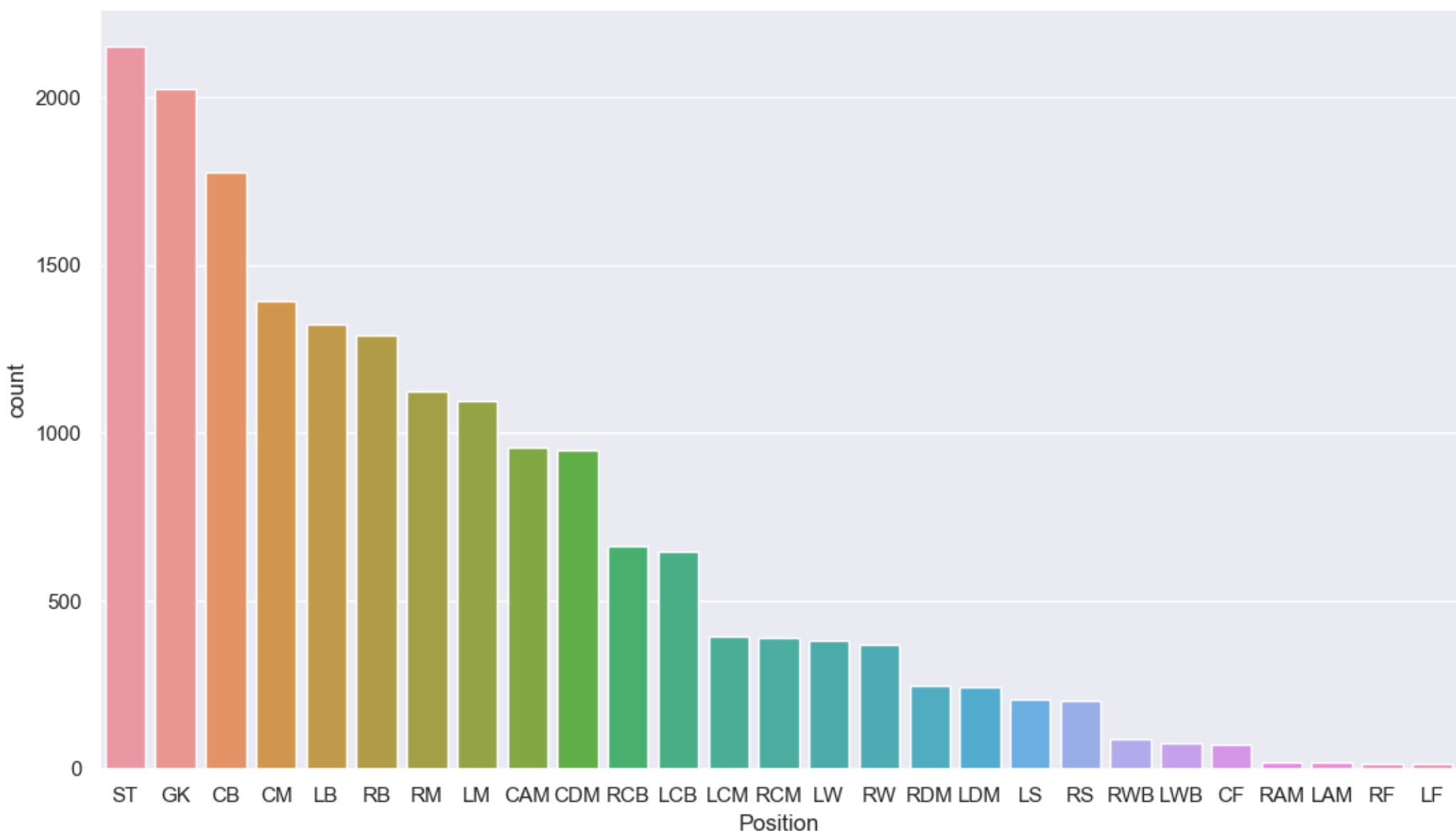


Introduction

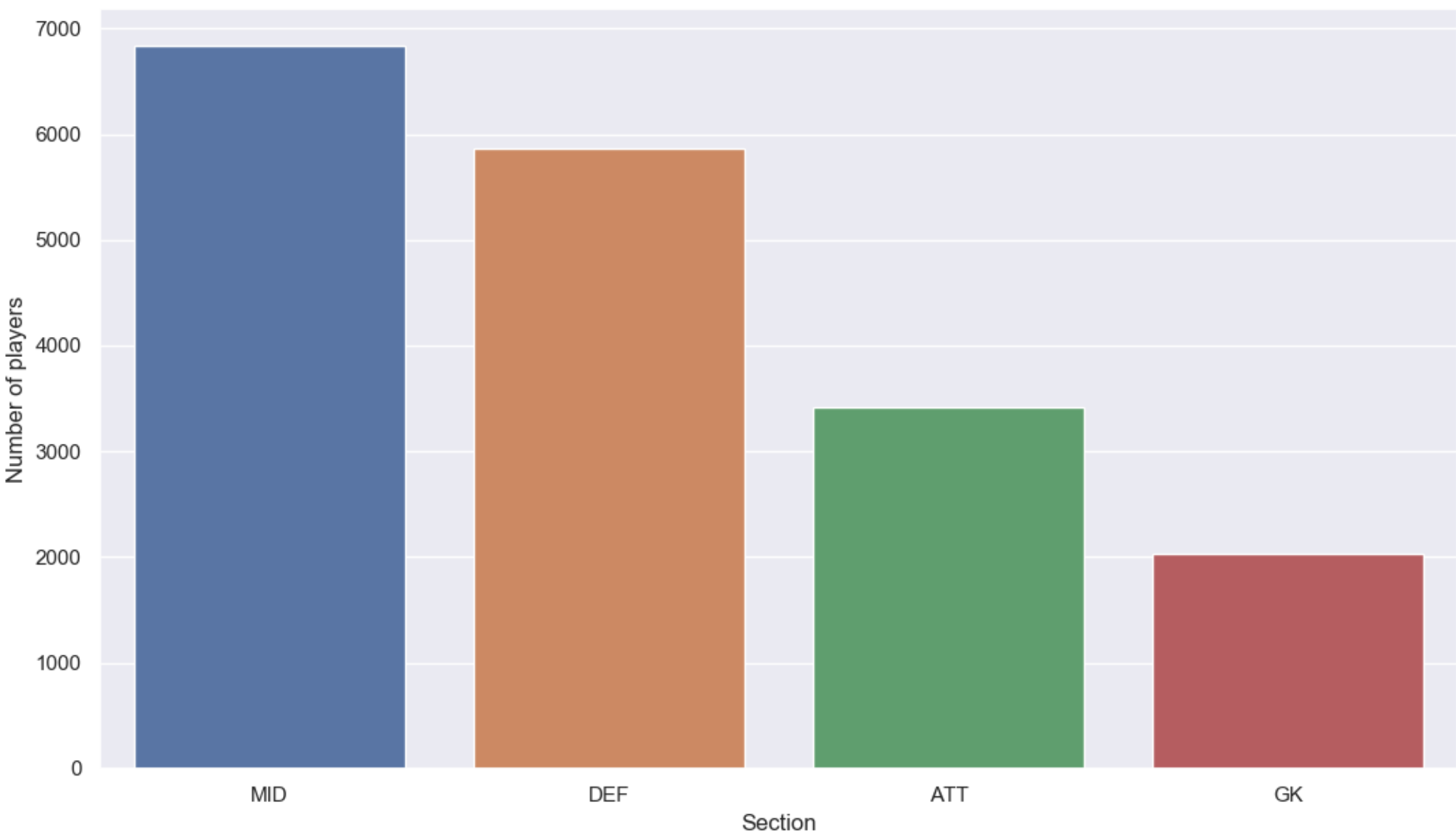
FIFA is the world's most popular football video game franchise. Over 20 million copies of FIFA 19 were sold, making it Europe's best-selling console video game in 2018 [1].

The problem being solved is a classification one. More specifically, the task is to predict the position of a footballer in FIFA 19 based on the attributes of that player.

Sections instead of positions



Number of players per position. This illustrates how unbalanced the dataset is.



Number of players per section. Although this dataset is also unbalanced, it is less so unbalanced than the aforementioned one.

Due to the dataset being unbalanced, an attempt was made at classifying the players by section. Four distinct sections emerged:

1. attack (ATT)
2. midfield (MID)
3. defense (DEF)
4. goalkeeper (GK)

The Algorithm

- „Cleaning“ the dataset:
 - Dimensionality reduction - by plotting the correlation matrix, superfluous features were detected and removed.
 - Removing null values from the dataset - there are three values that can be inserted instead of a null value:
 1. 1
 2. the column mean
 3. the column median
 - Encoding categorical features - one-hot encoding was used.
- Two models were tested:
 - A two-layer Artificial Neural Network (ANN) - it accepts a set of the footballer's attributes as input, whereas it outputs a list of probabilities, where each probability corresponds to its respective position/section. The argmax function is applied to that list, so that the most likely position/section is returned. In the case of position prediction, the ANN has 27 neurons in its output layer. In the case of section prediction, the ANN has 4 neurons in its output layer.
 - Random forest (RF) [2]

Results

A batch size of 100 was used. Other batch sizes were used, but to no increase in accuracy. All three ways of replacing null values gave the same results. 100 trees were used for the RF model.

Model	Position train acc.	Position test acc.	Section train acc.	Section test acc.	Number of epochs
ANN	51%	49%	87%	88%	10
ANN	54%	54%	88%	90%	100
RF	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
```

Several section predictions

Conclusion

Due to the dataset being unbalanced, it is difficult to predict the footballer's position. Instead, predicting the footballer's section gave a much higher accuracy. ANN and RF showed to be equally good in solving this problem.

References

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