Feature Extraction and Prediction Models for Context-Aware and Adaptive Soccer Analytics

Jordan B.R. Makins, Dr. S. Mistry, Prof. A. Krishna. PhD

1. Introduction

The ability to extract important features from soccer data and use computational methods to predict performance has drawn great interest and holds immense value in the soccer industry. To measure and predict performance, we propose training and deploying the first instance of the Forest Deep Neural Network (fDNN) in soccer analytics. Additional use cases of this algorithm include live data driven recommendations such as automated tactical substitutions, formation changes, automated identification of opposition tactics, playing style for exploitation, and player valuation. The algorithm is fast, adaptable and excels with a large number of features and a smaller number of samples. A variety of aggregates are collected across pitch locations. By omitting goals and assists from the feature vector, we avoided overfitting and we were able to better value non-goal scoring contributions by players in all positions.

2. Algorithm & System Design

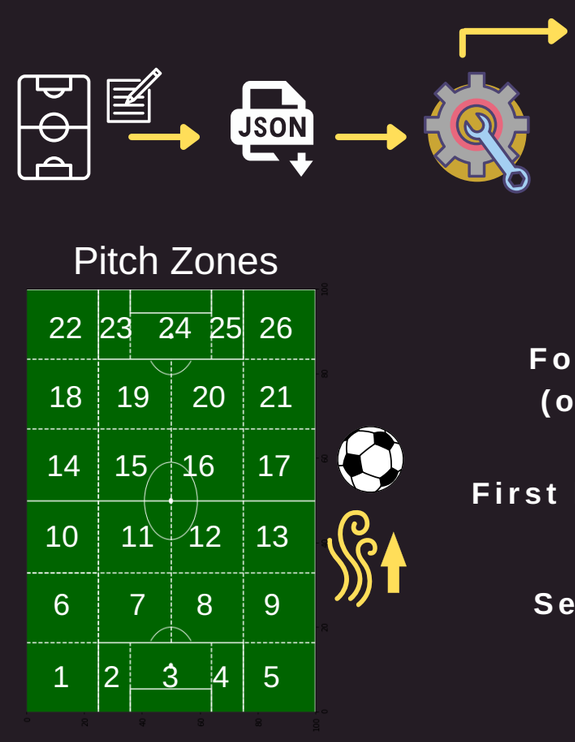
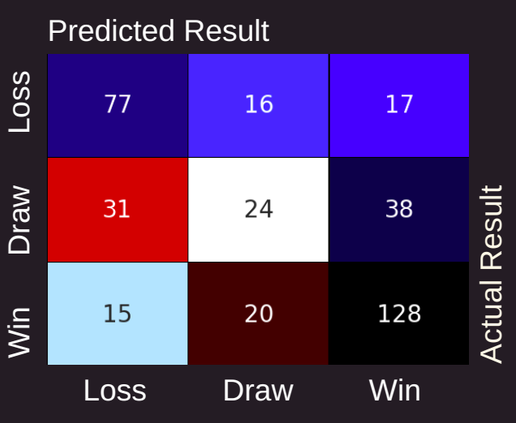


Figure 1: fDNN

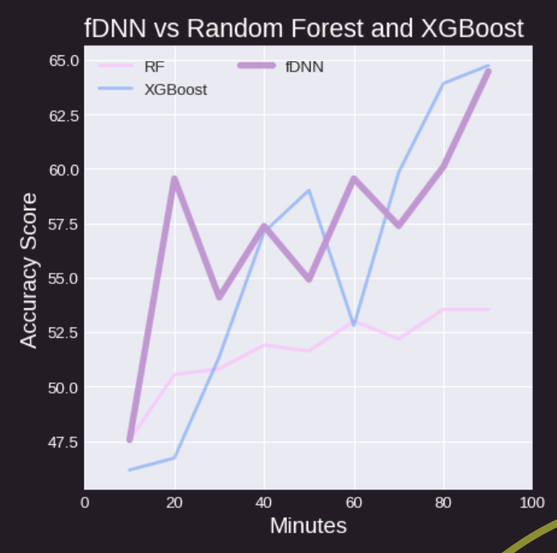
3. Results & Discussion

3.1 Feature Extraction and Machine Learning Evaluation

Predicting soccer match results live 60 minutes into match using extracted features and aggregated player ratings (ex goals and assists)

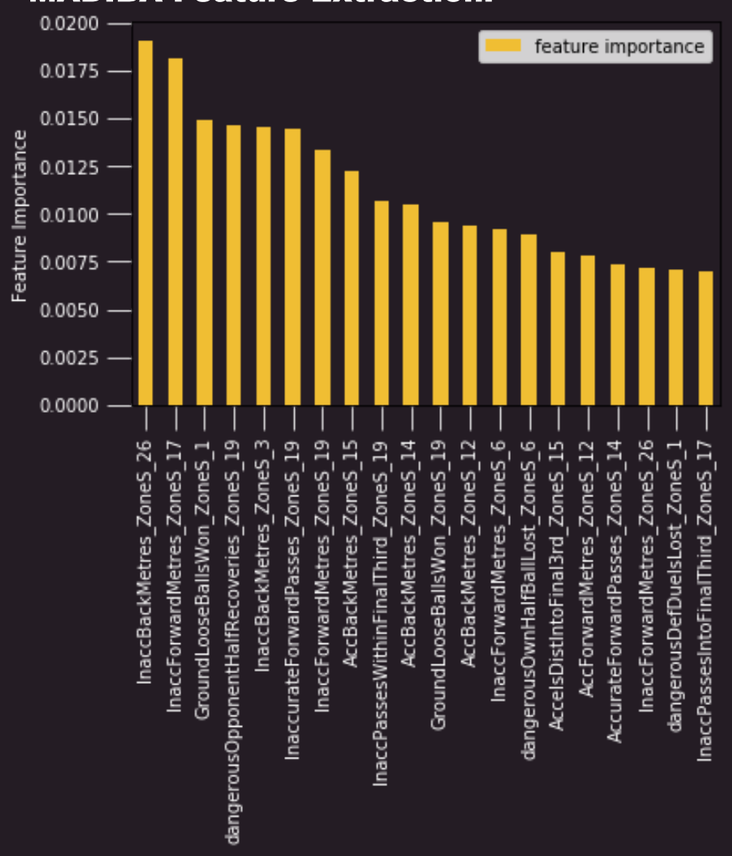


Confusion Matrix for fDNN predictions



Accuracy on live match result predictions at various time intervals based on extracted player and team features

MADIBA Feature Extraction:

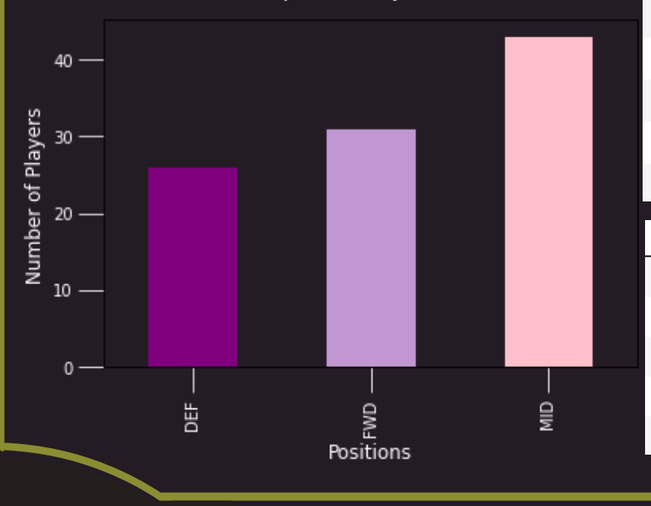


Machine Learning objectives:

- To determine and explain which relevant player and team performance attributes are important in predicting both match outcome and player value
- Explain and justify decisions for live match changes to improve performance such as automated tactical substitutions

3.2 MADIBA Player Ratings System:

Positions of Top 100 Players



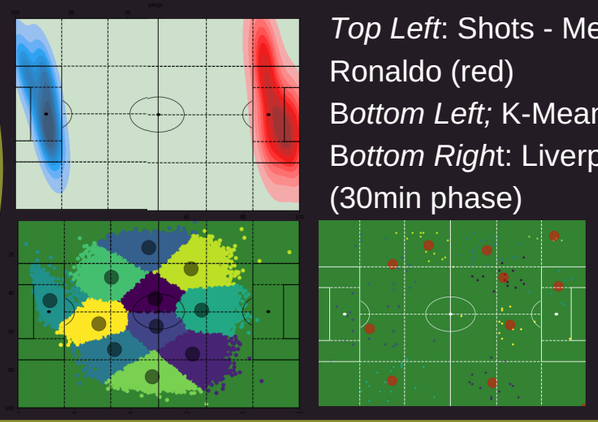
Top 5 Overall Players

Name	Position	Mins Played	playerRatings
L. Messi	FWD	2880.0	1812.6
L. Insigne	FWD	2960.0	1620.3
M.Hamsik	MID	2960.0	1604.3
K. De Bruyne	MID	2960.0	1600.5
K. Koulibaly	DEF	2800.0	1494.7

Name	Position	Mins Played	PR_perMatch
T. Kroos	MID	2160.0	94.0
L. Messi	FWD	2880.0	93.0
Neymar	FWD	1600.0	92.6
David Silva	MID	2320.0	90.7
Cristiano Ronaldo	FWD	2160.0	90.7

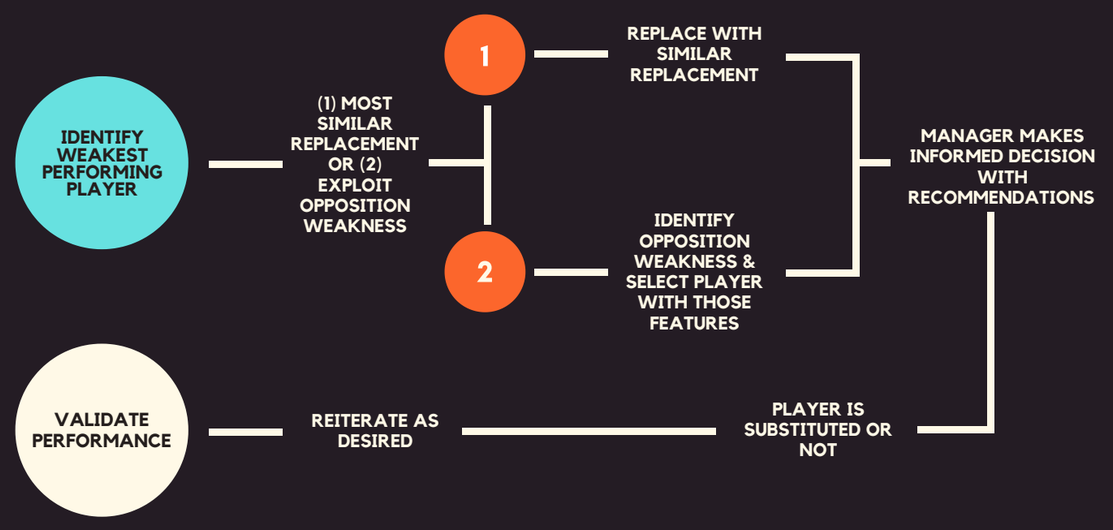
Top 5 Players per Match Played

3.3 Live Match Adaptive Analytics



Top Left: Shots - Messi (blue), Ronaldo (red)
B*ottom Left;* K-Means formation
B*ottom Right*: Liverpool formation (30min phase)

3.4 Automated Tactical Substitutions:



4.Conclusion

- **General adaptable formation, playing style, and tactical detection methods**
- **Novel processed representation of raw soccer event stream data**
- **Novel ratings system for players and teams** both live in-match and historical
- **Novel deployment of fDNN in soccer analytics** for feature extraction and performance prediction
- **Novel automated and adaptable tactical substitutions**

5. Acknowledgements

The auhors acknowledge CSIRO, A.Backhaus, T.Bonu, E.Matwiejew, and A.Subramanya for providing support, mentorship, and training throughout this and other research projects this summer.

A list of references and the original dataset available on GitHub:
hhttps://github.com/bjornmakins/soccer1/blob/master/README.md