FPA - Midterm

Prediction of football match results using virtual data

Introduction

Betting on sporting events, particularly football, has grown in recent years.

Accurately **predicting match outcomes** is of great interest to teams, coaches, and bettors, as it **can provide a competitive advantage**.

However, this task remains challenging due to the **complex nature of the game** and the many factors influencing results, with prediction **accuracies consistently below 60% in the literature**.

Match outcome prediction has the potential to enable successful betting strategies and aid players' and coaches' understanding of success factors.

Doing that is a challenge in part due to the game's low-scoring nature, high competitivity, the possibility of the game ending in a tie, and the many factors affecting it (e.g., skills, weather).

Objective

Our goal is to **explore the potential use of player ratings** from the EA FIFA videogame as a reliable source of information for **predicting the winner** of a football match.

We seek to determine whether we can utilize this data to develop relevant prediction models for the three-class problem (i.e., win, loss, draw).

Related Work

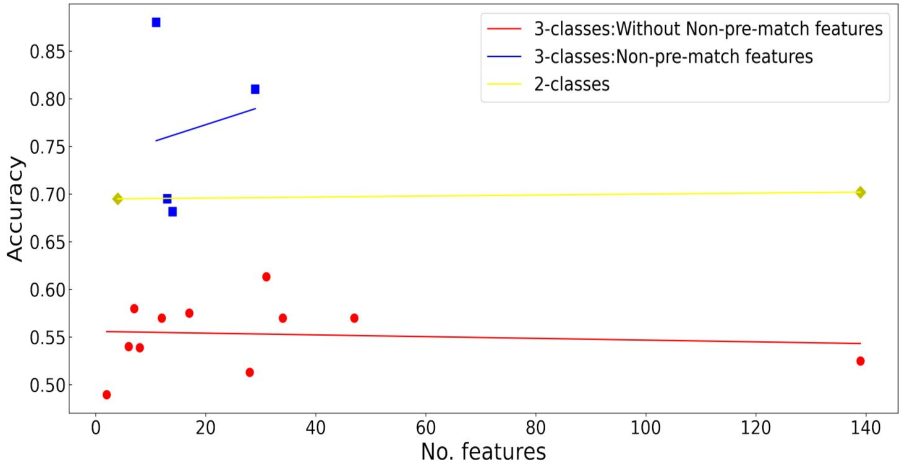
When studying match prediction in football, studies can be separated into (i) predicting the outcome of the match and (ii) predicting the final scores.

The former:

- Y. Ren and T. Susnjak, 2022, created different machine learning models in matches of the English Premier League for three seasons 2019-2021 based on the *Kelly index*.
- H. Rue and O. Salvesen, 1997, used a Bayesian dynamic generalized linear model to estimate the time-dependent skills of all teams, and to predict the outcome.
- J. gon Shin and R. Gasparya, 2014, used player ratings from the EA FIFA 2015 video game to predict match outcomes based on the information of a single tournament.

The most challenging problem is that of predicting one of three possible outcomes: *Win*, *Loss*, *Tie*, when considering only pre-match information.

The accuracy results are consistently below 60% for this problem.



Number of features against prediction accuracy for all types of literature. Taken from Y. Ren and T. Susnjak, 2022.

Data and Methods
& Exploratory Data Analysis

Databases

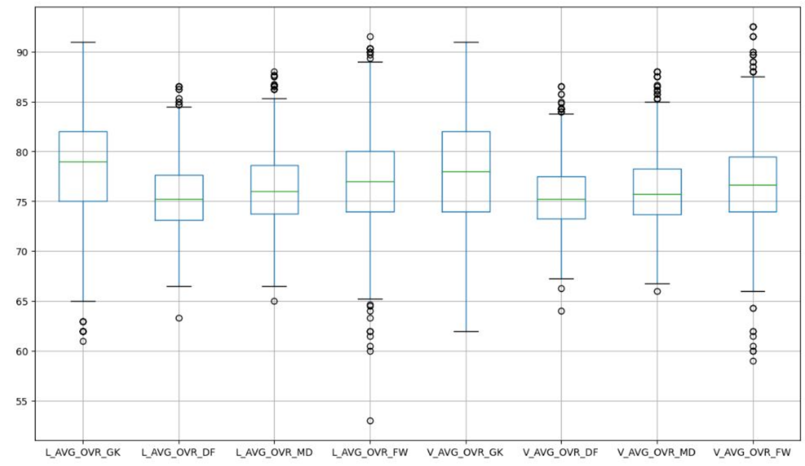
1. A public data set that contains information about soccer matches from a time window from 2016 to 2018 (we focus our analysis on matches from 2018) - **Pappalardo et al. 2019.**

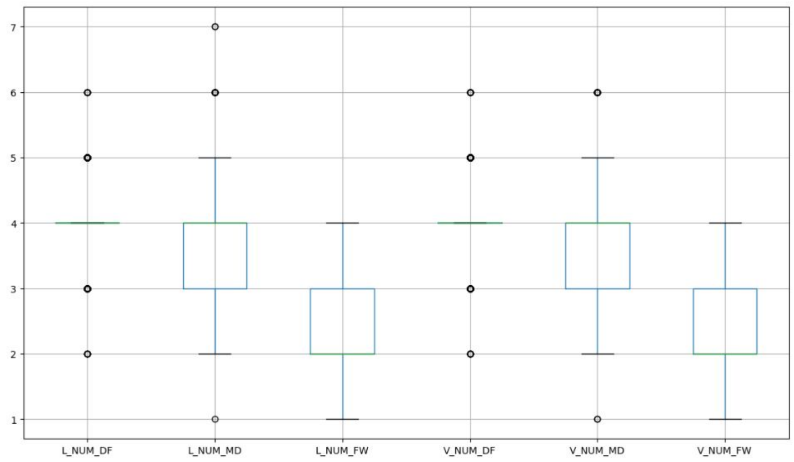
2. The player ratings in the EA FIFA video game scrapped from the SoFIFA website - **https://sofifa.com/**

Feature engineering

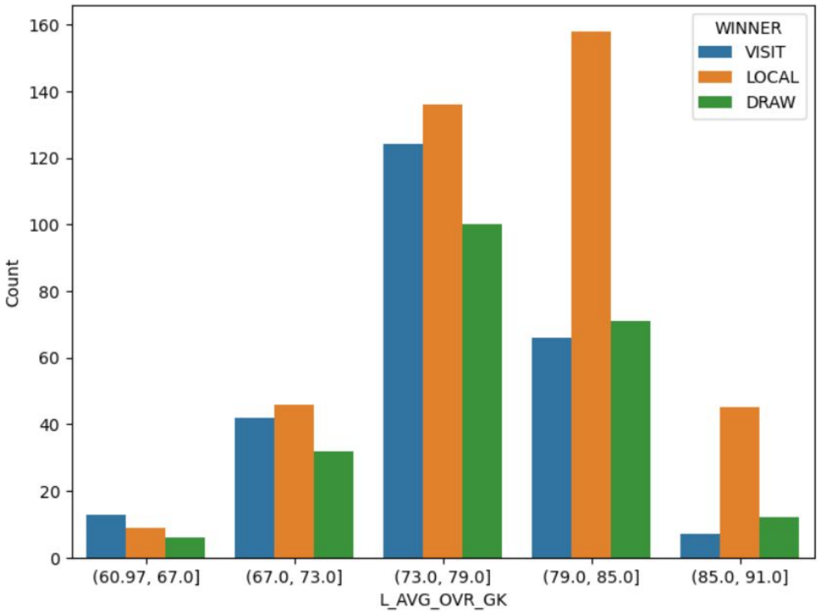
Attribute	Range	Description
OVR_GK	[0-100]	Overall rating of the goalkeeper.
AVG_OVR_DF	[0-100]	Average overall rating of defensive team players.
AVG_OVR_MD	[0-100]	Average overall rating of midfielder team players.
AVG_OVR_FW	[0-100]	Average overall rating of offensive team players.
NUM_DF	[2-6]	Number of defensive players for the team at match start.
NUM_MD	[1-7]	Number of midfielder players for the team at match start.
NUM_FW	[0-4]	Number of offensive players for the team at match start.

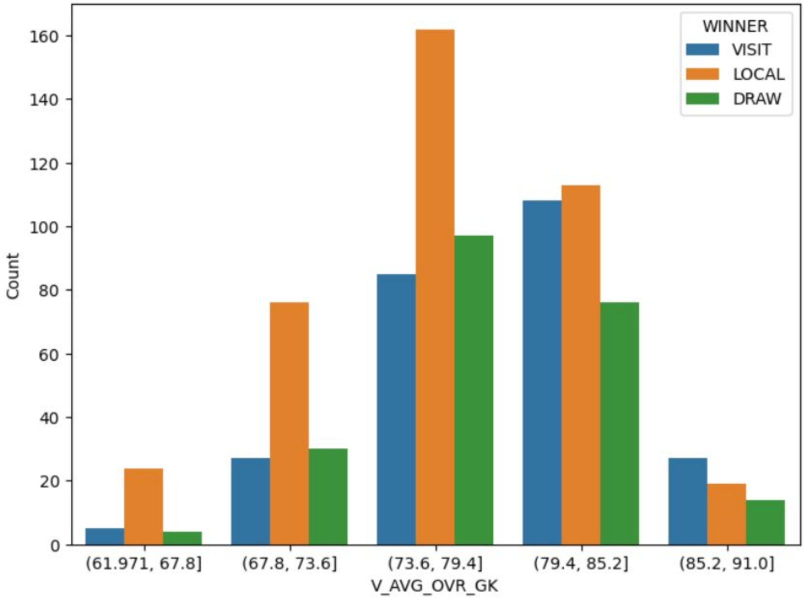
Distribution of team ratings and formations



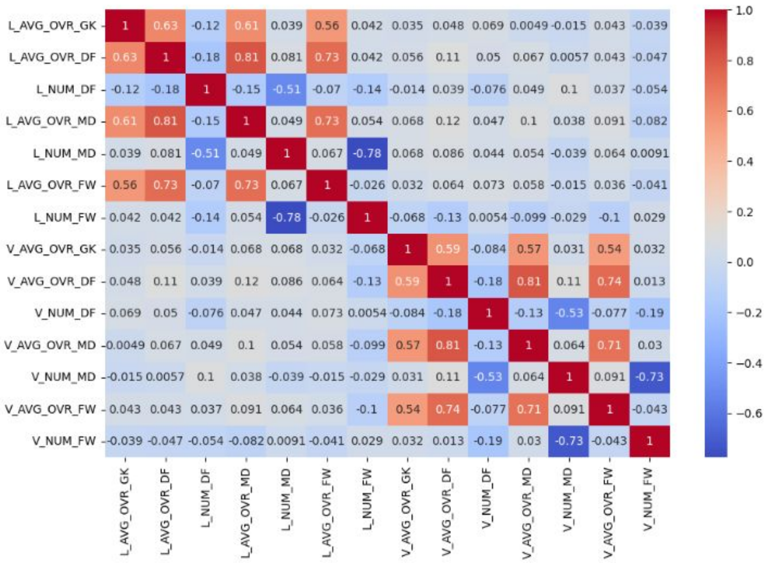


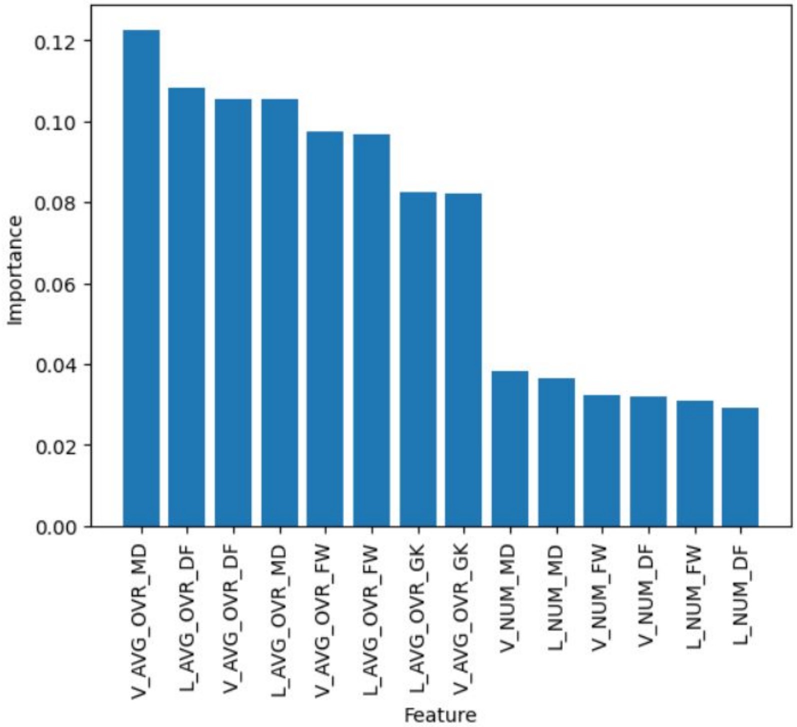
Attribute distribution by class



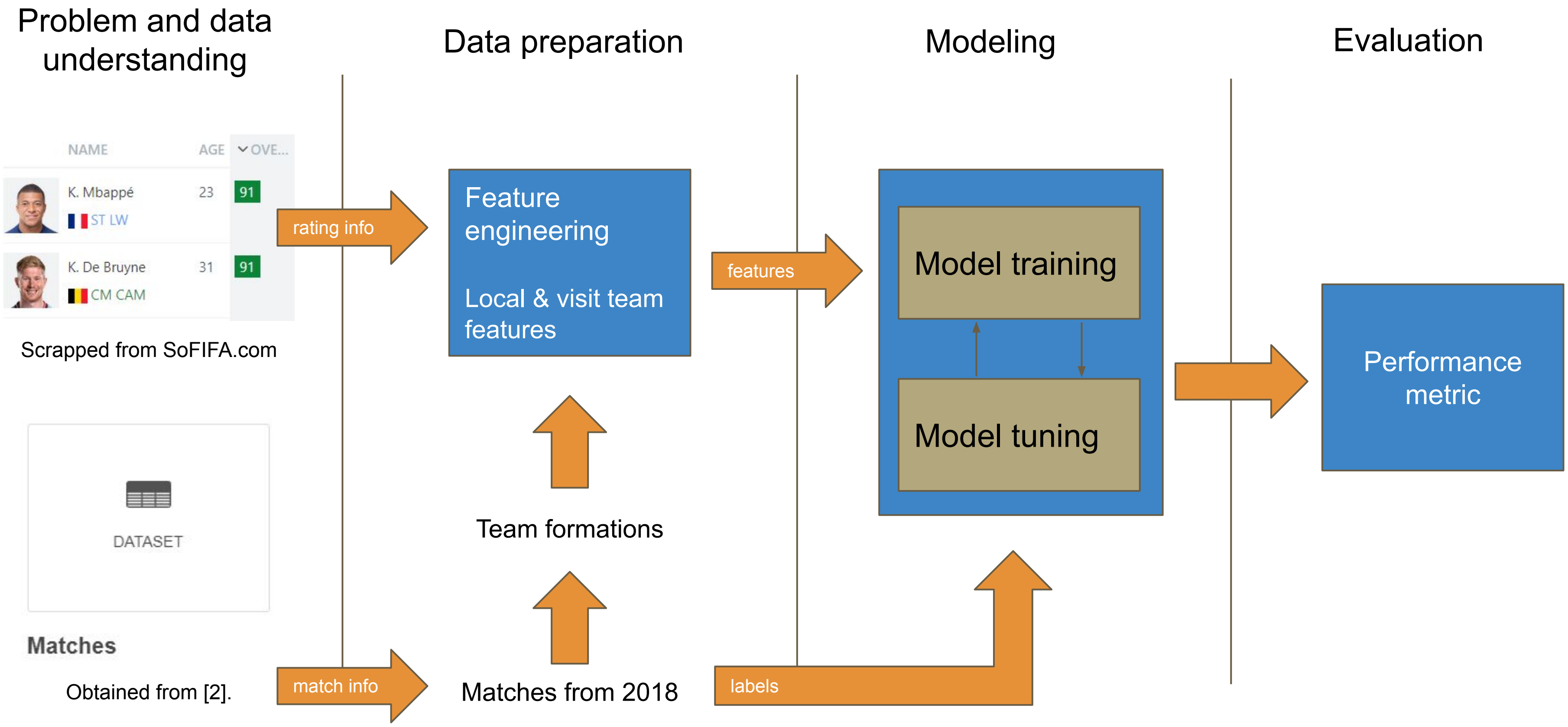


Feature correlation and importance





Approach



Problem and data understanding
Data preparation
Modeling
Evaluation
NAME
AGE
OVE...
K. Mbappé
ST LW
23
91
K. De Bruyne
CM CAM
31
91
Scrapped from SoFIFA.com
rating info
Feature engineering
Local & visit team features
features
Model training
Model tuning
Performance metric
DATASET
Matches
Obtained from [2].
Team formations
match info
Matches from 2018
labels

References

- Y. Ren and T. Susnjak k, “Predicting football match outcomes with explainable machine learning and the kelly index,” 2022.
- H. Rue and O. Salvesen, “Predicting and retrospective analysis of soccer matches in a league,” 01 1997.
- J. gon Shin and R. Gasparyan, “A novel way to soccer match prediction,” 2014.
- L. Pappalardo, P. Cintia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi, and F. Giannotti, “A public data set of spatio-temporal match events in soccer competitions,” Scientific Data, vol. 6, no. 1, p. 236, 2019. Available: https://doi.org/10.1038/s41597-019-0247-7
- “Players FIFA 23 Apr 6, 2023 SoFIFA.” [Online]. Available: https://sofifa.com/