HOCHSCHULE
SRH HEIDELBERG

Shotifier: A Machine Learning Pipeline for Classifying Goal Scorers

Mentored by:
Prof. Dr.Ajinkya Prabhune
Prof. Dr. Sreeganesh Thottempudi
Ashish Chouhan

Presented by:
Ankit Raj
Darshan Chandra
Sindhu Subramanya
Mahaveer Asangi
Tanmay Gaikwad
Tauseef Ahmad Hashmi

Contents

- Objective
- Introduction
- Wyscout Dataset
- Data exploration
- Architecture
- Feature Selection
- Spatial Distribution
- Zone Interaction
- Modeling
- Evaluation
- Conclusion and Future Work

Objective

In this project, we present Shotifier, a binary classification pipeline that is based on the principle of hybrid parallelism. Shotifier focuses on forwards or strikers and uses match statistics for classifying whether the shot from the striker at the opponent's goal will be converted to a goal or not.

Introduction

- Football, is one of the most popular sports, is also considered unpredictable in nature and a low scoring game.
- A multi-billion industry where an estimated 4 billion fans follow and initiate a global sphere of influence.
- In English Premier League clubs alone on an average of 1.8 billion euro was spent on the acquisition of players in the 2017 summer transfer window.

Exploratory Data Analysis

- Goals made by forwards and other players in percent.
- Which match period is more productive as per the Goal shots / Expected Goal Shots.
- Which kind of passes are being converted to goals by forwards and Non-Forwards.

Players	Number of Events	Goals	Percentage
Forwards	52784	242	0.46 %
Non-Forwards	235311	703	0.29 %

Forwards and Non-Forward

Positions
Goalkeeper
Right Centre Back
Left Centre Back
Right Winger
Right Centre Midfielder
Left Centre Midfielder
Left Winger
Attacking Midfielder
Right Centre Midfielder
Defensive Midfielder
Left Centre Midfielder

Positions
Right Defensive Midfielder
Left Defensive Midfielder
Right Attacking Midfielder
Left Attacking Midfielder
Right Centre Back (3 at the back)
Centre Back
Left Centre Back (3 at the back)
Right Wingback
Left Wingback
Right Back (5 at the back)
Left Back (5 at the back)

Positions
Right Winger
Left Winger
Second Striker
Striker
Right Wing Forward
Left Wing Forward

Non Forwards Events

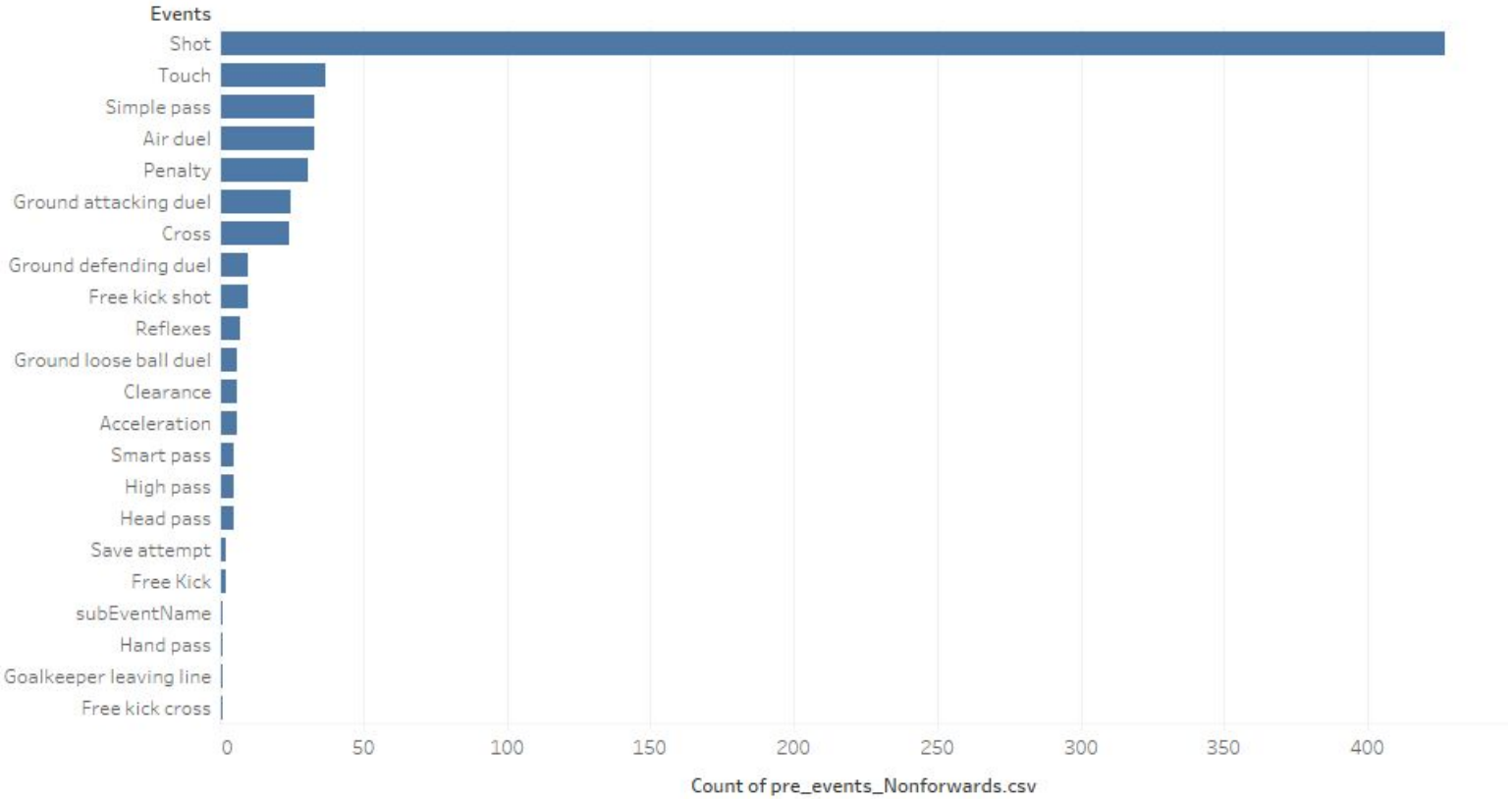


Fig. Non Forward Events

Forwards Events

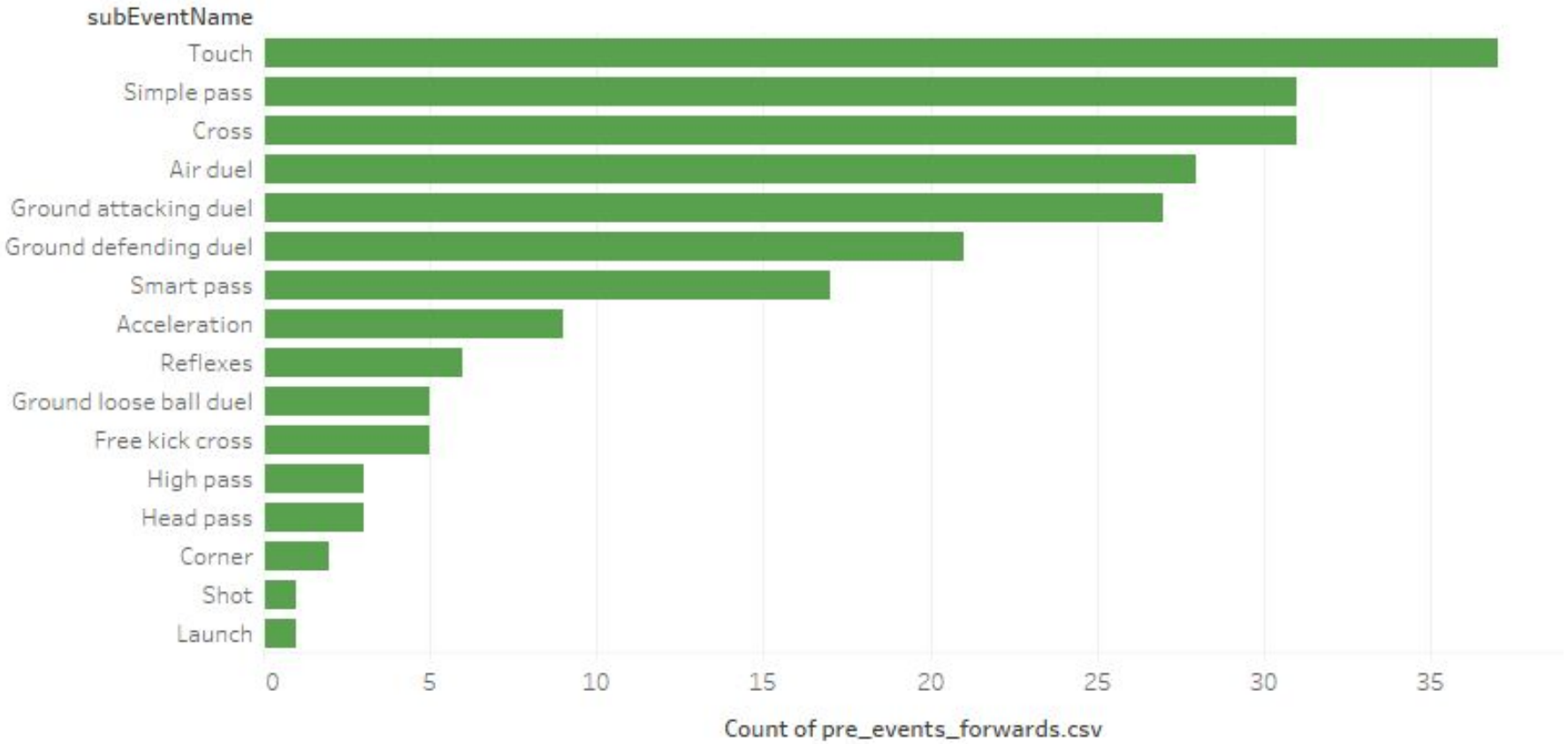
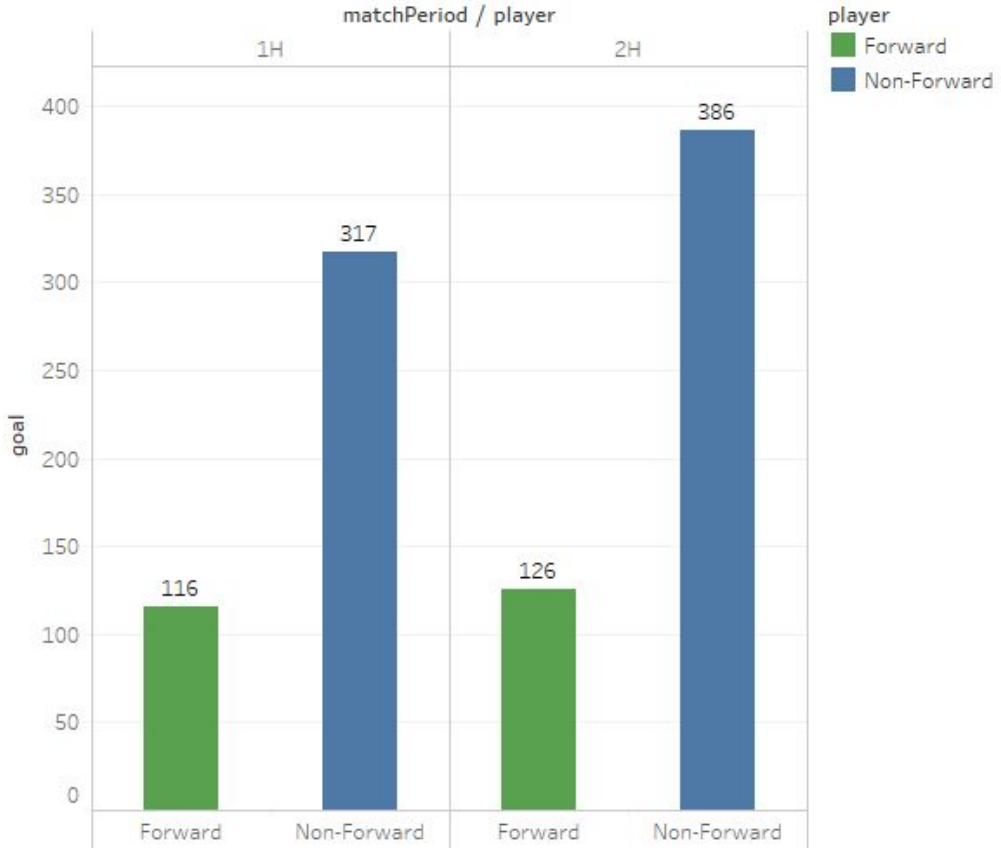
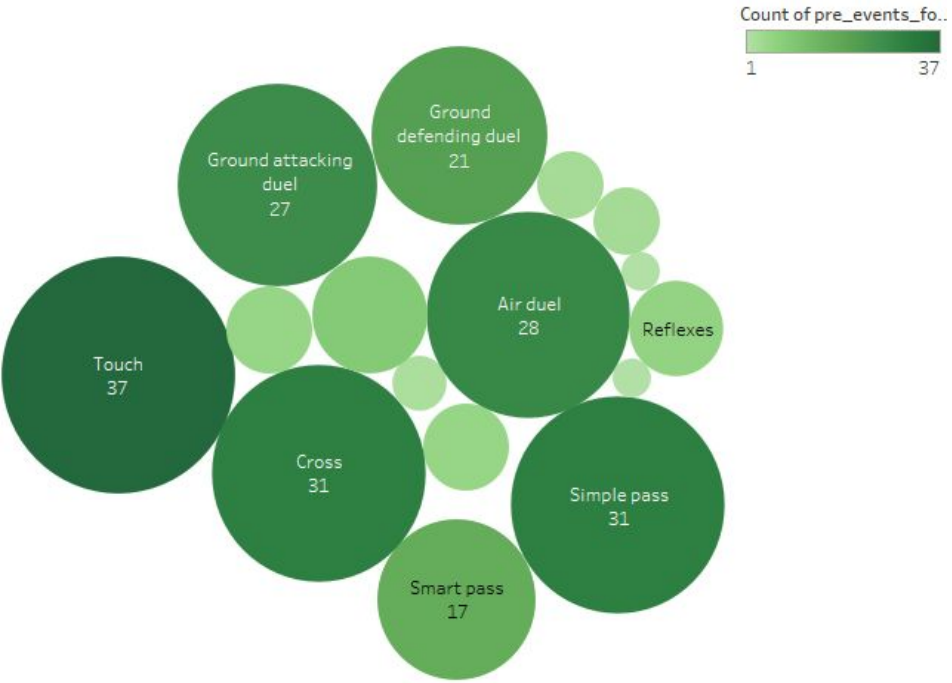


Fig. Non Forward Events

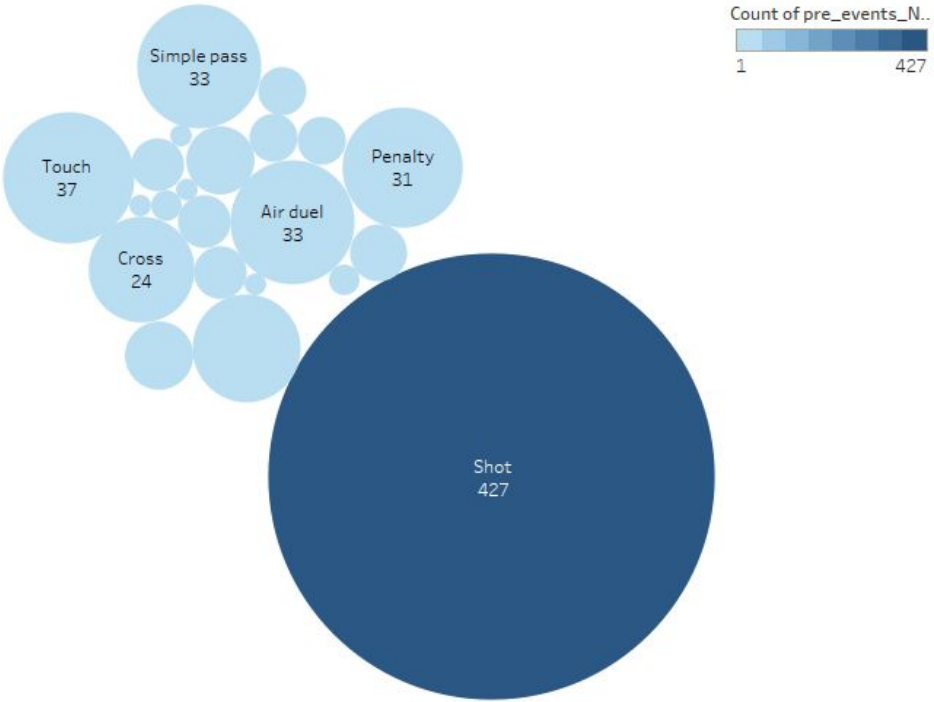
Match Period Exploration



Passes Before Goal Comparison



Types of Passes to Forwards before Goal



Types of Passes to Non-Forwards before Goal

Wyscout Dataset

- APIDocs. https://apidocs.wyscout.com
- Season 19/20 Data.
- English Premier League.
- Total of 380 matches were played with teams playing 38 matches each.

Datasets have the following stats:

- Events Data: Players, Formations, events, EventName, Event ID, SubEventName, SubEventID, positions, X-Y Coordinates,Tags, ID. Etc.

- Player Data: Player Details, list of positions, playtime percentage, yellow cards, red cards, shots, goals, assists, penalties, Duels, Duels Won, Defensive Duel, Aerial Duels, Successful Passes, Crosses, Free Kicks etc.

Shotifier Pipeline

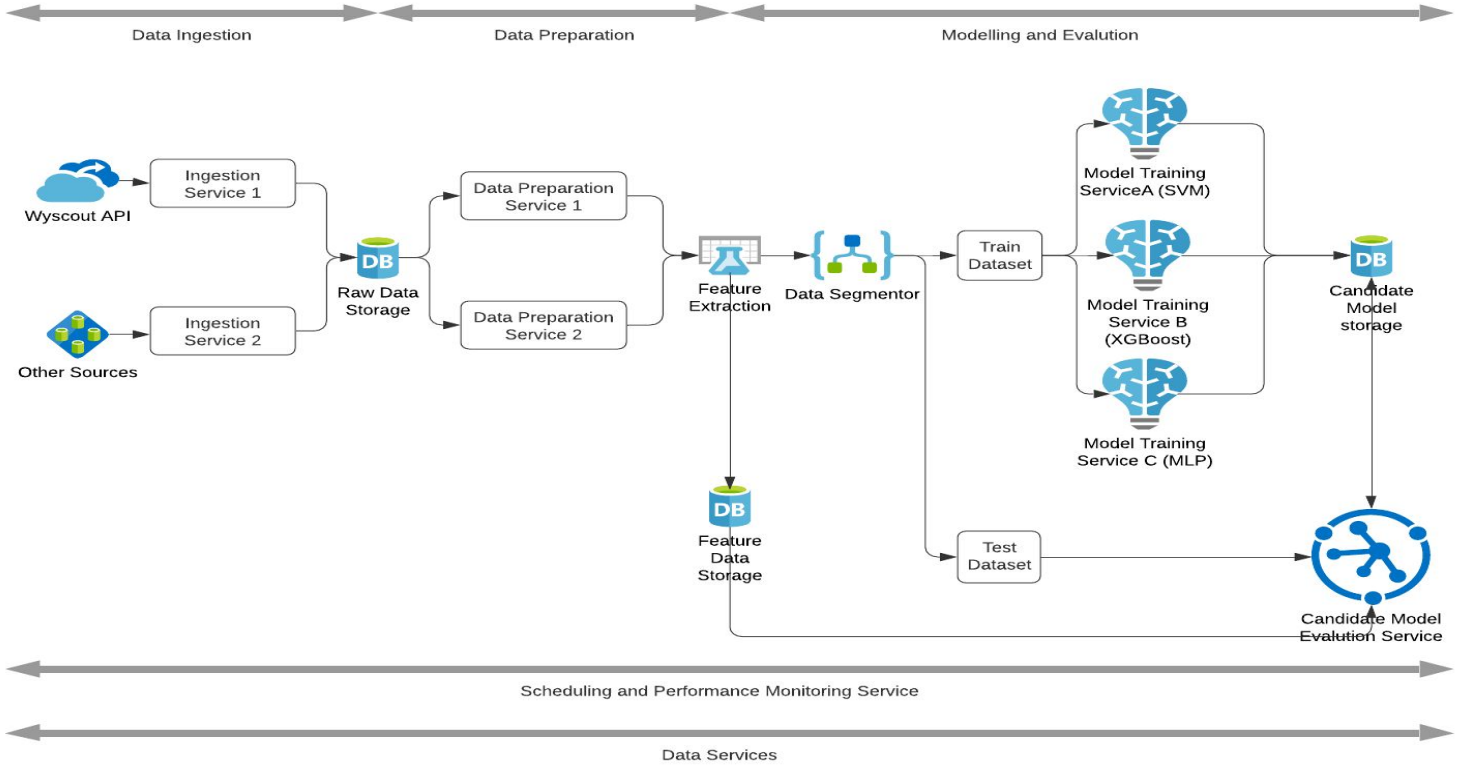


Fig. Architecture of Shotifier Framework

Feature Selection

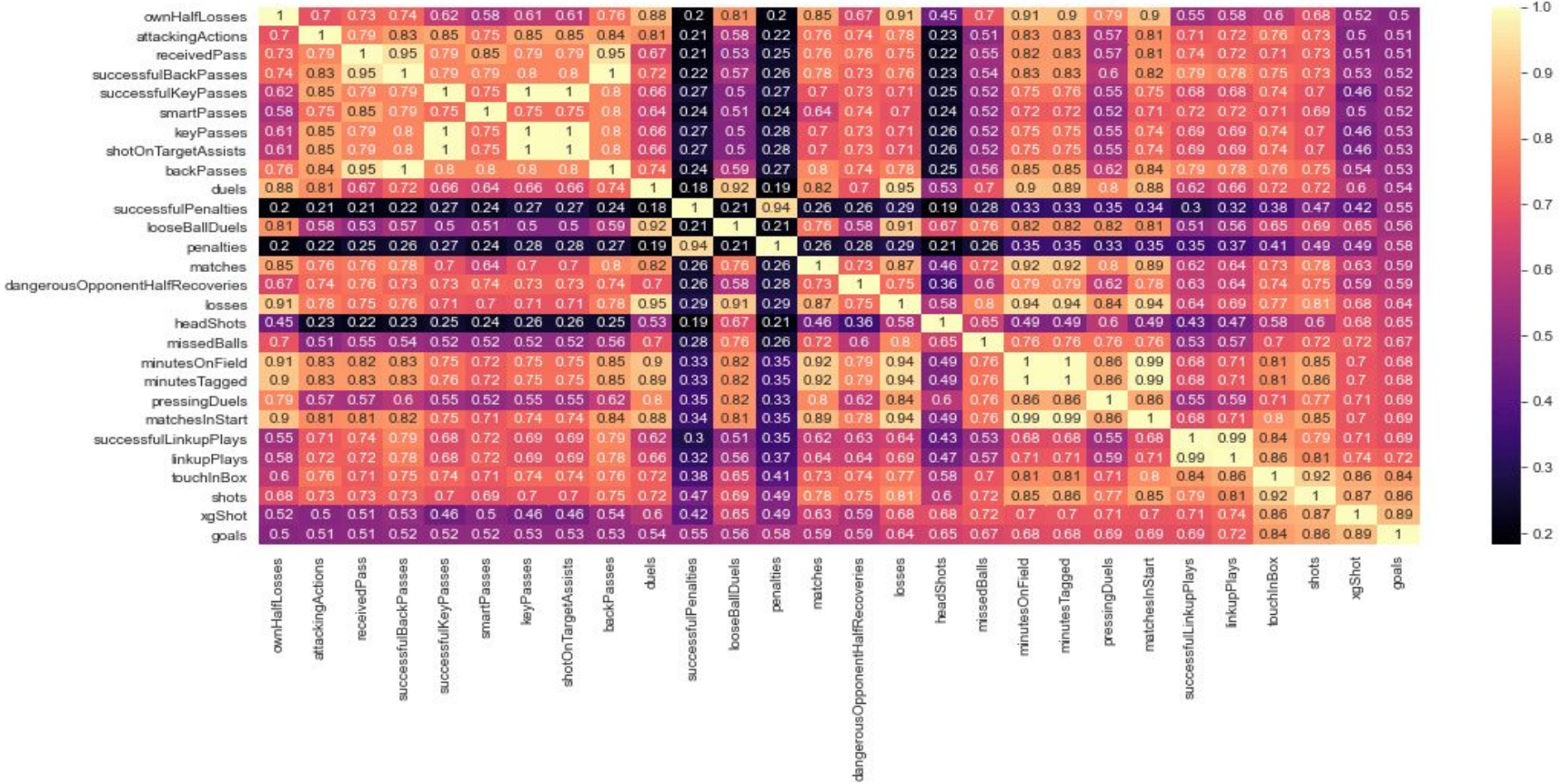


Fig. Heat Map of factors that have high correlation coefficient

Spatial Distribution

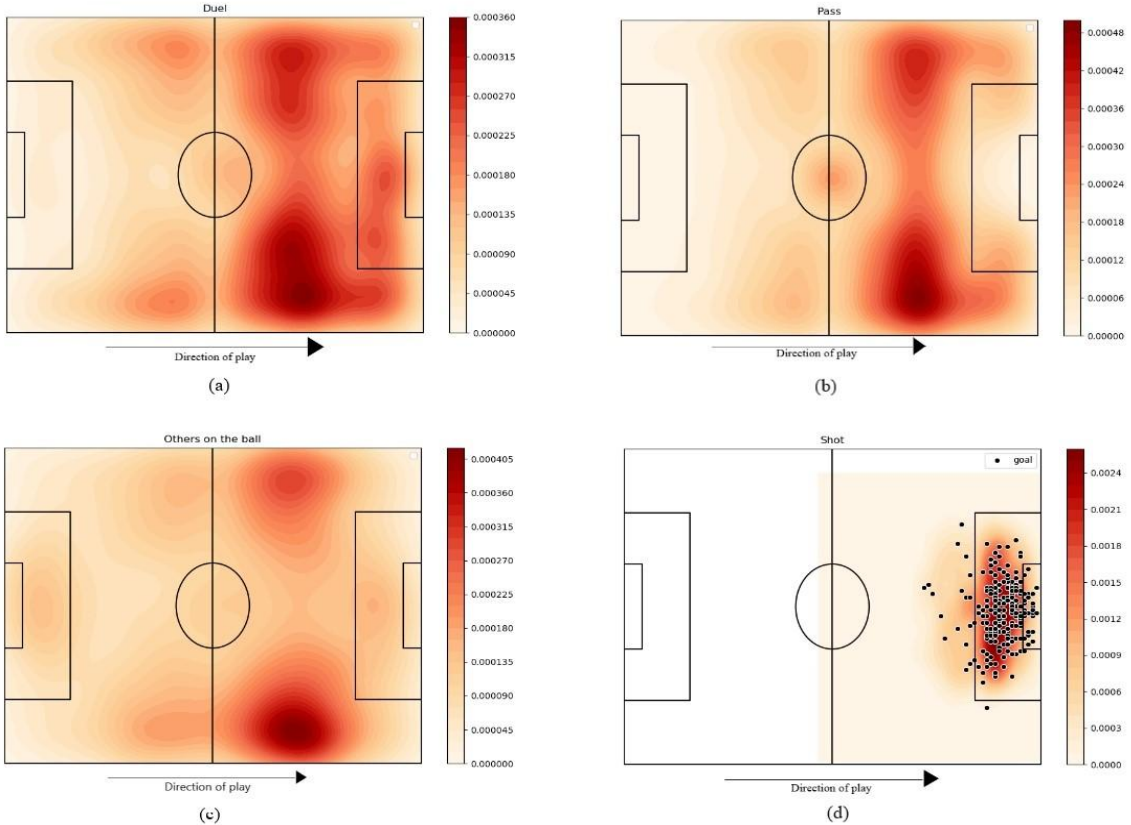
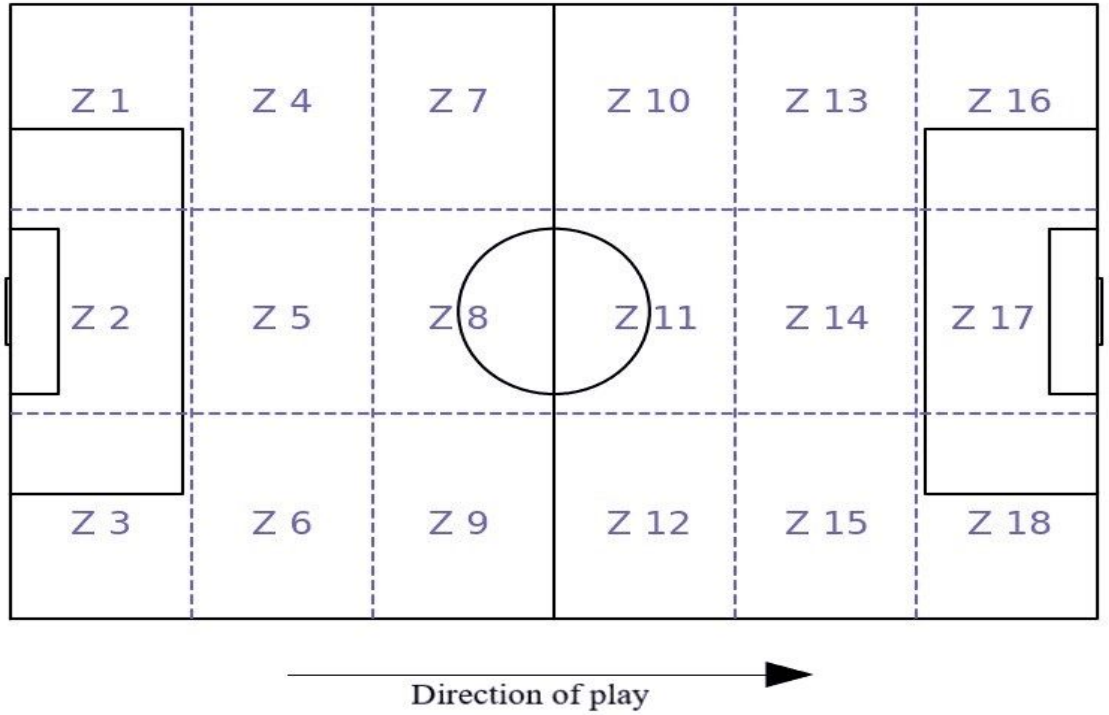


Fig.. Spatial distribution of active players with respect to events

Zone Interaction



Event Start	Event counts	Goal counts
Z 15	6345	2
Z 13	5499	2
Z 14	4210	14
Z 12	3886	0
Z 18	3673	15
Z 10	3420	0
Z 17	2875	211
Z 9	2663	0
Z 7	2488	0
Z 11	2437	0
Z 16	2356	8

Fig. Division of ground into zones

Zone 17 Interaction

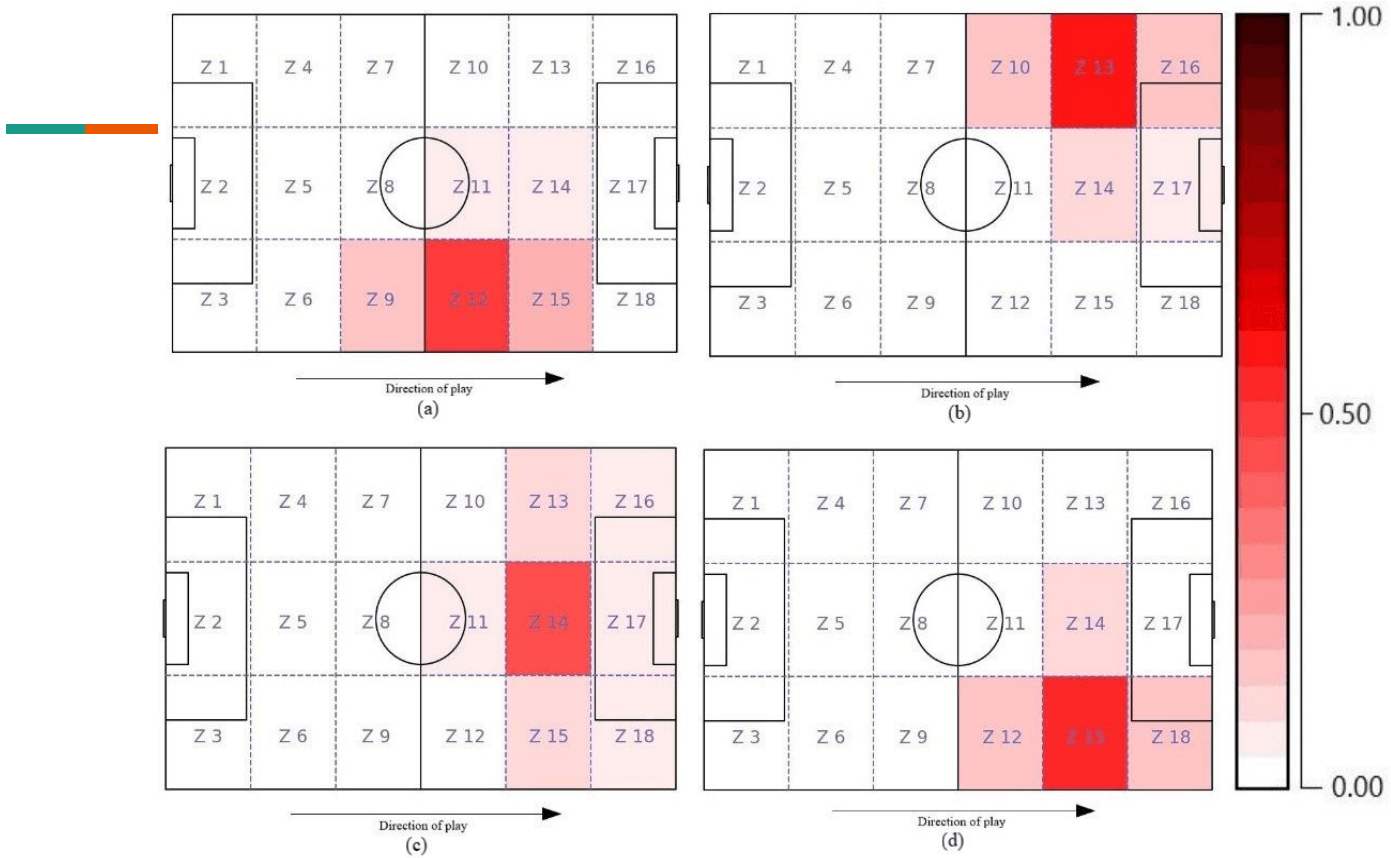


Fig. Zone interactions with other zones

Modeling and Evaluation

- Shotifier is based on hybrid parallelism which incorporates several Machine learning (ML) pipelines to train a model. In this pipeline, the training data is converted as a 2D matrix with each row as a data instance.

- To concurrently run all the machine learning models such as SVM, XGBoost, random forest and MLP a pipeline for each model is implemented.
- K-fold cross-validation is implemented to evaluate the performance of each model that uses an unseen data.

- Model monitoring is a continuous process, a shift in the prediction might lead to restructuring the model. Grafana, a tool to get monitoring dashboards is used for performance monitoring.

Model	Evaluation Score
Random Forest	84.02
XGBoost	86.92
SVM	86.92
ML	87.03

Conclusion

Above result reflects that we have been able to narrow down the players' goal-scoring abilities and positioning senses perfectly. In matches, it has enormous potential to identify most suitable players in different schemes.

The Shotifier framework can be used by coaches and managers during or after a match to better strategize their upcoming game play.

Our method has obtained surprisingly good results on both exploratory analysis and classification by capturing critical features of event tracking data.

Future Scope

As future work, we could easily improve the classification estimates by using more detailed information such as shot power, trajectory of the ball, and location of the players in the vicinity of the event. Furthermore, we plan to investigate whether the proposed method is applicable to other sports like, for instance, basketball or ice hockey.

Thank
you!