Football Match Outcome Prediction

using DeepSet Player Aggregation

Group 12
Mohammad Bahrami
Sayed Masoud Mousavi

Outline

- Problem Understanding Phase
  - Problem Definition
  - Previous Work
  - Our Idea
- Data Preparation
- Exploratory Data Analysis

Problem Understanding Phase

❖ Problem Definition

- Association Football is the world’s most popular sport.
- How is football Played?
- Home Team Advantage
- What is the players’ role?
- The Conditions of aggregating the players

Outline

- Problem Understanding Phase
  - Problem Definition
  - Previous Work
  - Our Idea
- Data Preparation
- Exploratory Data Analysis

Problem Understanding Phase

❖ Previous Work

- **Elo Rating System**
- **Blade Chest**
- **Simple GNN**

Outline

- Problem Understanding Phase
  - Problem Definition
  - Previous Work
  - Our Idea
- Data Preparation
- Exploratory Data Analysis

Problem Understanding Phase

❖ Our Idea

- Our Data
- Our Solution: Deepset

$$V_T = \mathrm{MLP}_\theta\left(\sum_{p \in T} \mathrm{MLP}_\phi(p)\right)$$

Outline

- Problem Understanding Phase
- Data Preparation
  - Outlier Detection
  - Transformation and Standardization
  - Reclassifying the Categorical Features
  - Binning
- Exploratory Data Analysis

Data Preparation

❖ Outlier Detection

- Z-Score

	postGame_goals
37467	2
50688	2
50757	2
50877	2
51107	2
...	...
67259	4
47229	4
27278	4
26430	4
18678	5

[944 rows x 1 columns]

- IQR

	postGame_goals
39708	1
51438	1
51409	1
51407	1
51406	1
...	...
27278	4
47229	4
49567	4
67259	4
18678	5

[8211 rows x 1 columns]

Outline

- Problem Understanding Phase
- Data Preparation
  - Outlier Detection
  - Transformation and Standardization
  - Reclassifying the Categorical Features
  - Binning
- Exploratory Data Analysis

Data Preparation

❖ Transformation and Standardization

```
preGame_overall
Original data skewness: 0.11
After transformation if necessary data skewness: -0.053
----------------------------------------
preGame_potential
Original data skewness: 0.094
After transformation if necessary data skewness: -0.007
----------------------------------------
preGame_marketValueMilEuro
Original data skewness: 9.837
After transformation if necessary data skewness: 0.104
----------------------------------------
preGame_ageDays
Original data skewness: 0.28
After transformation if necessary data skewness: 0.067
----------------------------------------
```

Outline

- Problem Understanding Phase
- Data Preparation
  - Outlier Detection
  - Transformation and Standardization
  - Reclassifying the Categorical Features
  - Binning
- Exploratory Data Analysis

Data Preparation

❖ Reclassifying the Categorical Features

	preGame_position	preGame_rc_position
0	DC	D
1	FW	F
2	MC	M
3	MR	M
4	MC	M
...	...	...
80075	AMC	M
80076	DMC	M
80077	FW	F
80078	FW	F
80079	AMC	M

Outline

- Problem Understanding Phase
- Data Preparation
  - Outlier Detection
  - Transformation and Standardization
  - Reclassifying the Categorical Features
  - Binning
- Exploratory Data Analysis

Data Preparation

- Binning

Methods: Equal width, Equal Frequency
#bins = random between 3 and 9

```
numerical_fp = ['postGame_minPlayed',
  'preGame_overall', 'preGame_potential', 'preGame_marketValueMilEuro',
  'preGame_ageDays', 'postGame_error',
  'postGame_clearance', 'postGame_index', 'postGame_shots',
  'postGame_shots_on_target', 'postGame_shots_left_foot',
  'postGame_shots_right_foot', 'postGame_shots_head',
  'postGame_shots_other', 'postGame_goals', 'postGame_goals_left_foot',
  'postGame_goals_right_foot', 'postGame_goals_head',
  'postGame_goals_other', 'postGame_xG', 'postGame_cross',
  'postGame_cross_success', 'postGame_pass', 'postGame_pass_success',
  'postGame_pass_final_third', 'postGame_pass_final_third_success',
  'postGame_pass_forward', 'postGame_pass_forward_success',
  'postGame_dribble', 'postGame_dribble_success', 'postGame_tackle',
  'postGame_tackle_success', 'postGame_interception',
  'postGame_challenge', 'postGame_ball_recovery', 'postGame_ball_lost',
  'postGame_key_pass', 'preGame_xgpm', 'preGame_xppm',]
```

```
numerical_gk = ['postGame_minPlayed',
      'preGame_overall',
       'preGame_potential', 'preGame_marketValueMilEuro',
       'preGame_ageDays', 'preGame_xgpm',
       'preGame_xppm', 'postGame_error', 'postGame_clearance',
       'postGame_index', 'postGame_pickUp', 'postGame_punch', 'postGame_save']
```

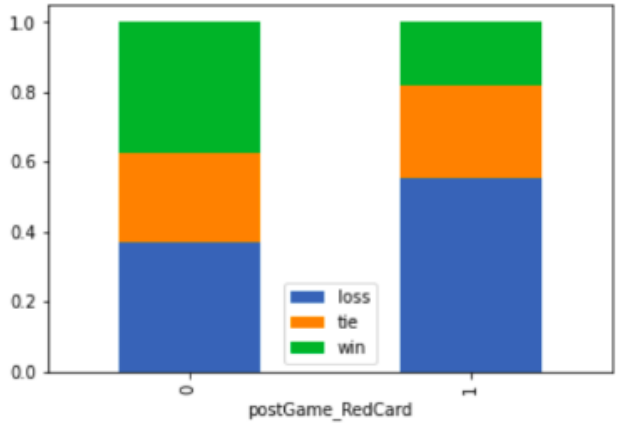
	postGame_pass	postGame_pass_binned_6
0	52	(32.667, 65.333]
1	16	(-0.196, 32.667]
2	51	(32.667, 65.333]
3	46	(32.667, 65.333]
4	98	(65.333, 98.0]
...	...	...
80075	14	(-0.196, 32.667]
80076	9	(-0.196, 32.667]
80077	22	(-0.196, 32.667]
80078	8	(-0.196, 32.667]
80079	10	(-0.196, 32.667]

Outline

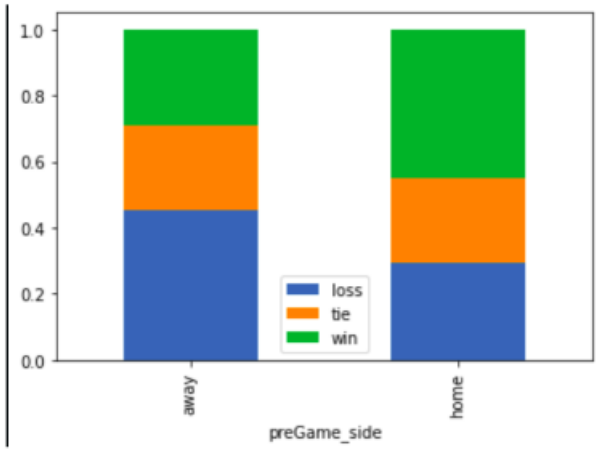
- Problem Understanding Phase
- Data Preparation
- Exploratory Data Analysis
  - Univariate Relations with the Target Value
  - Multivariate Relations
  - Binning Based on Predictive Value
  - Extracting New Features

Exploratory Data Analysis

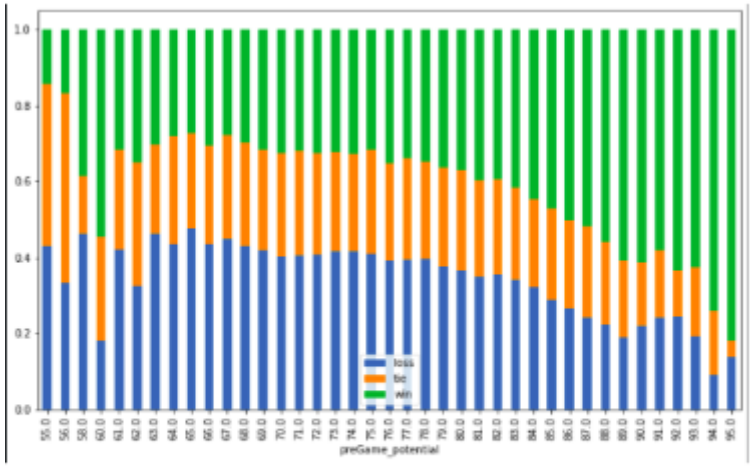
❖ Univariate Relations with the Target Value



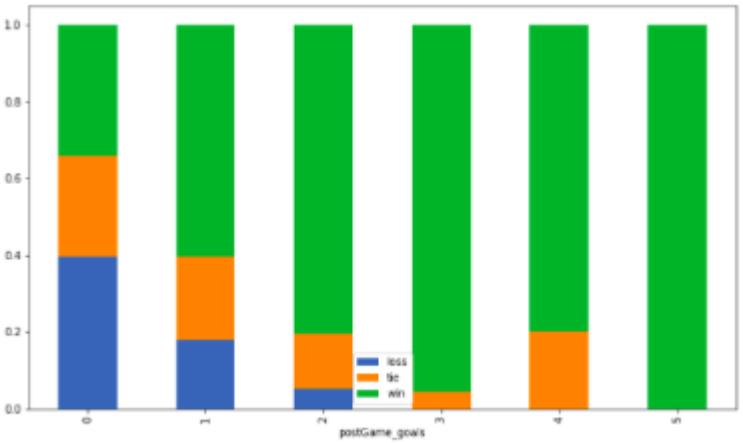
(a) weather a player received a red card or not



(b) weather a player is playing for the home team or not



(c) pre-game potential from the video game features



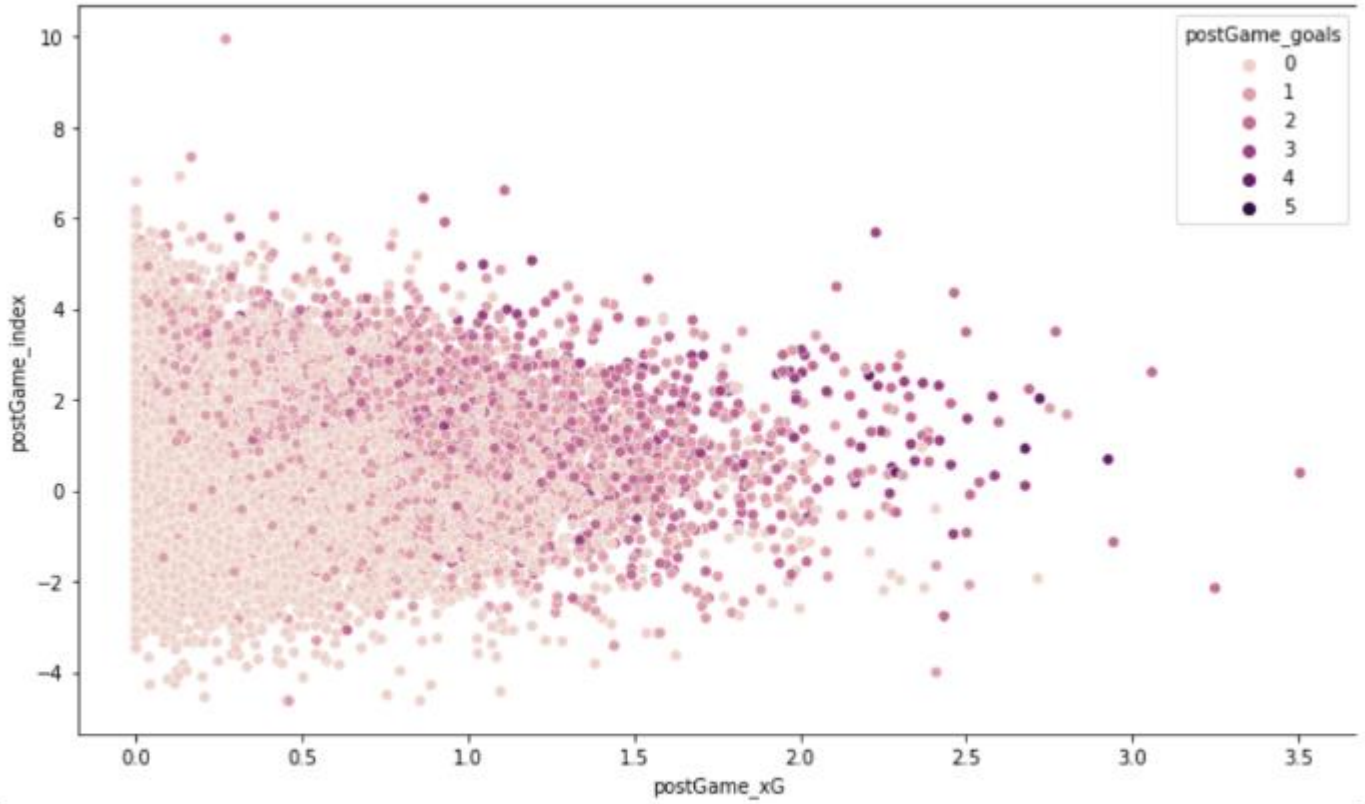
(e) the number of scored goals by a player

Outline

- ❖ Problem Understanding Phase
- ❖ Data Preparation
- ❖ Exploratory Data Analysis
  - Univariate Relations with the Target Value
  - Multivariate Relations
  - Binning Based on Predictive Value
  - Extracting New Features

Exploratory Data Analysis

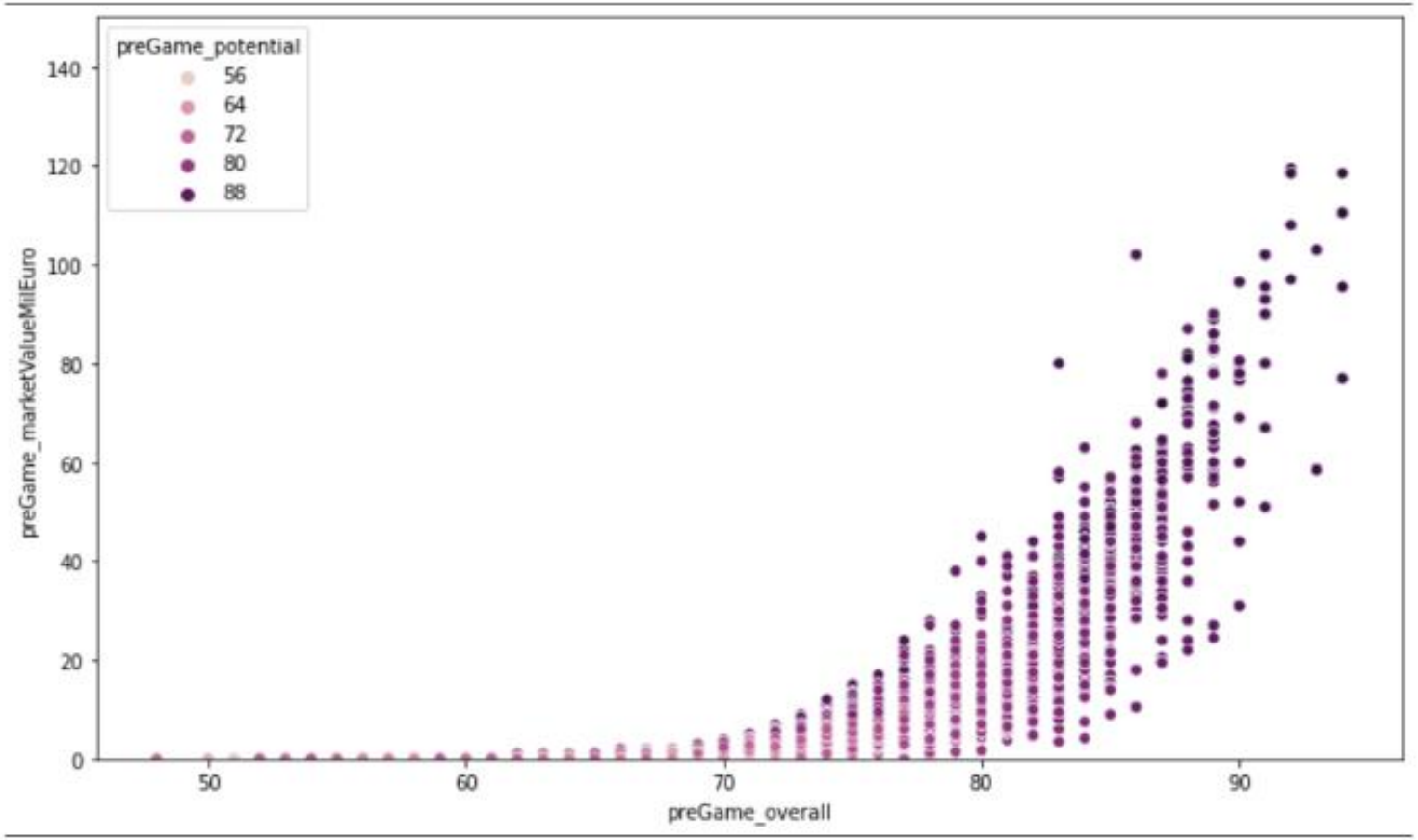
❖Multivariate Relations



(a) the relation between stephenson index, xG index and the number of goals of a player

Exploratory Data Analysis

❖Multivariate Relations



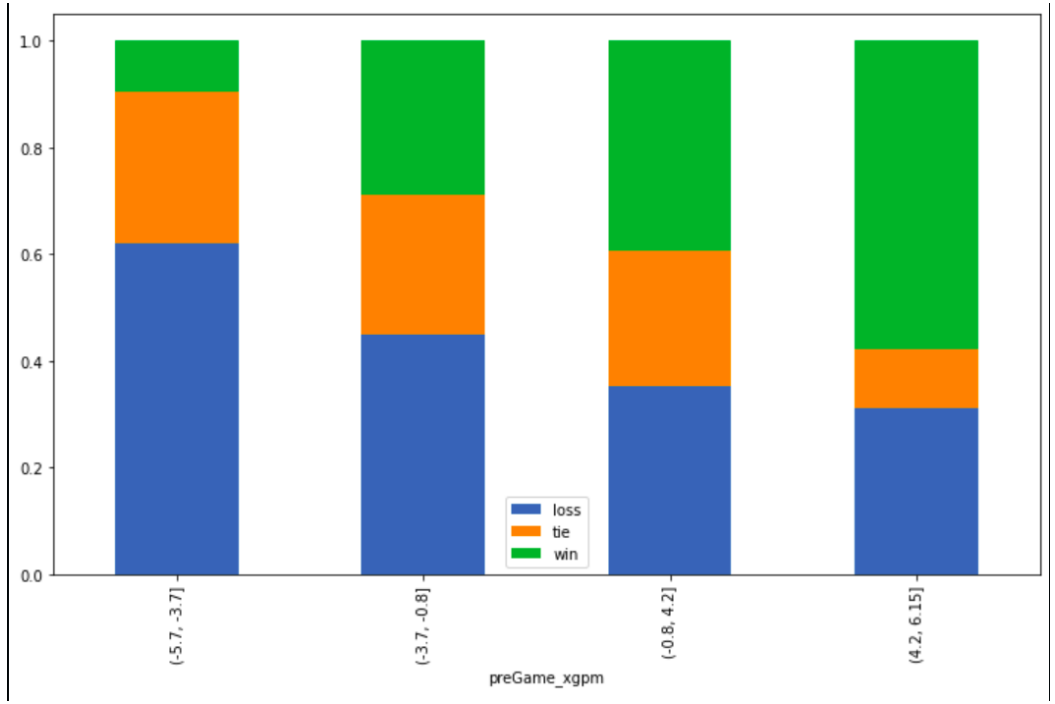
(b) the relation between overall, market value and potential video game scores

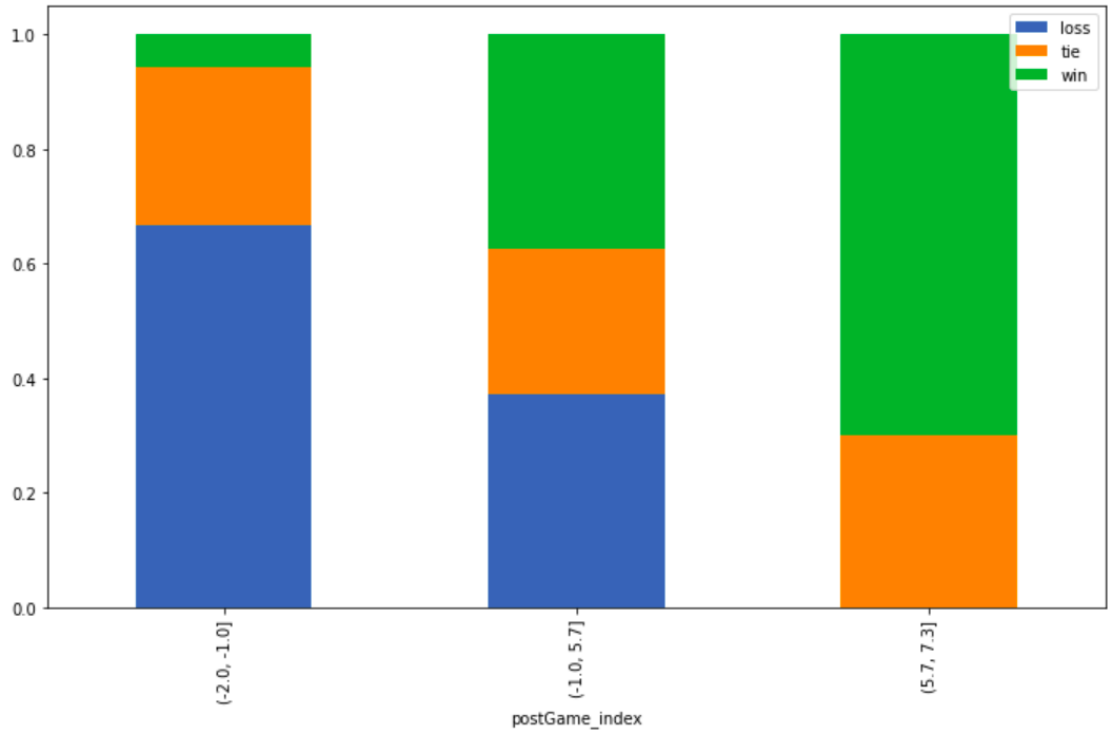
Outline

- Problem Understanding Phase
- Data Preparation
- Exploratory Data Analysis
  - Univariate Relations with the Target Value
  - Multivariate Relations
  - Binning Based on Predictive Value
  - Extracting New Features

Exploratory Data Analysis

❖ Binning Based on Predictive Value





Outline

- Problem Understanding Phase
- Data Preparation
- Exploratory Data Analysis
  - Univariate Relations with the Target Value
  - Multivariate Relations
  - Binning Based on Predictive Value
  - Extracting New Features

Exploratory Data Analysis

❖ Extracting New Features

	agg_postGame_error	agg_postGame_clearance	agg_postGame_index	agg_postGame_shots	agg_postGame_shots_on_target	agg_postGame_shots_left_foot
80075	0.0	2.25	1.167869	0.5	0.25	0.0
80076	0.0	1.25	1.737601	1.5	0.75	0.5
80077	0.0	1.75	0.717848	1.0	0.75	0.5
80078	0.0	6.75	0.391377	0.0	0.00	0.0
80079	0.0	0.00	1.234975	0.0	0.00	0.0

THANK YOU!