NextMatch

Yuwei Liu

- Identify variables that matter in terms of winning
- Offer actionable insights for sports teams

“Many organizations spend considerable sums on collecting data, but lack the capability to … analyze the data using … machine learning.”
– Tony Strudwick, Head of Athletic Development, Manchester United FC

“(To identify variables that matter in terms of winning), run correlations across match stats, correlate it with winning“
“**(No single stats matters for winning), the only stats that matters is winning**”
– Jo Clubb, Lead Sport Scientist, Chelsea FC

Data and Algorithm

Data

La Liga (Spanish Top Division) match stats: scraped from whoscored.com

Training: 12/13-13/14, ~76 matches / team; Testing: 14/15, ~18 matches / team

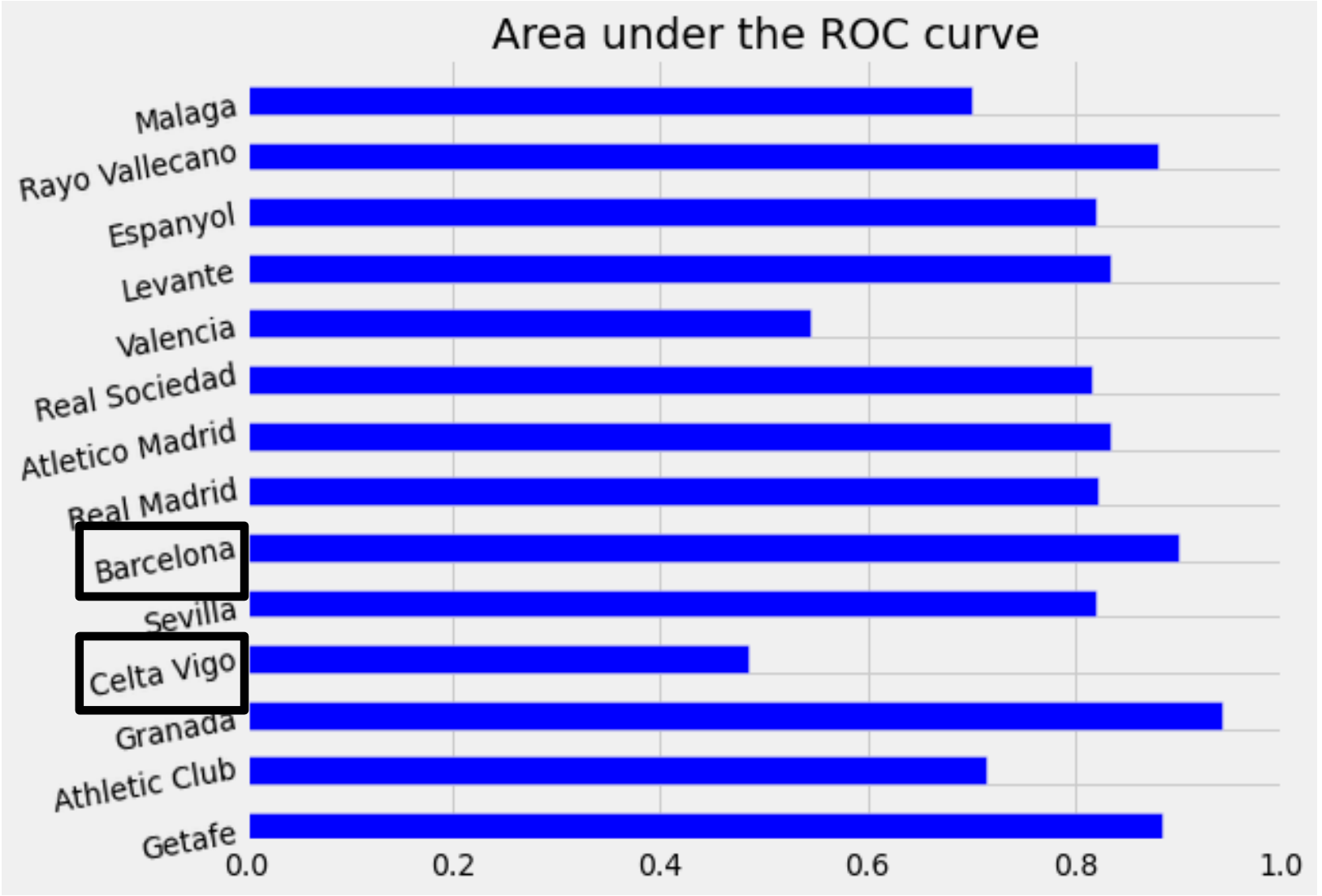
Data Sample:

Outcome of Team A's match	Features of team A	Features of team A's OP
1: win / 0: loss or tie	Pass, Shot, …	Pass, Shot, …

Algorithms

Feature Selection	Random Forest
Classification model	Logistic regression using features selected above

Model Performance (and limitation)

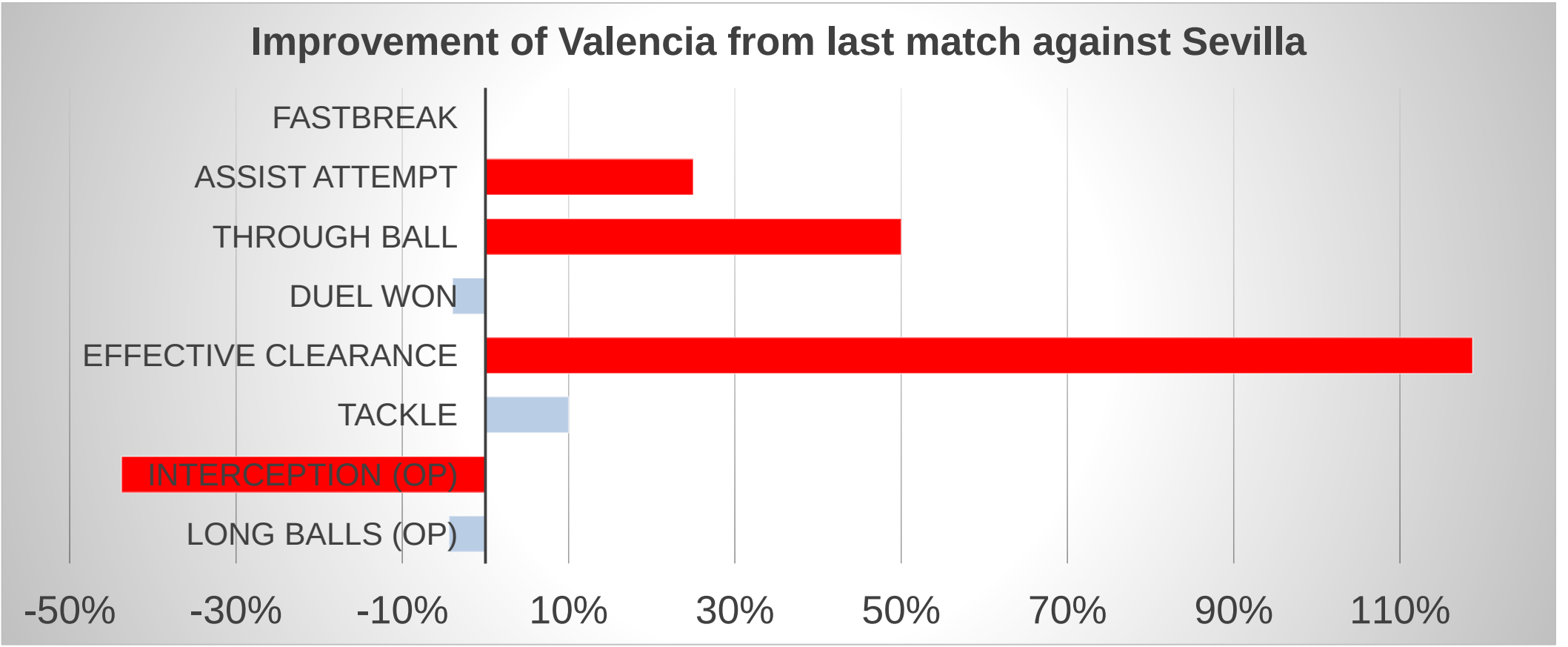


Barcelona: Changed 1 coach and 1 player in the main squad in the past 2 years

Celta Vigo: Fired 3 coaches since Feb/13; Replaced 6 out 11 players in main squad; Changed positions of all the rest

Use Case – for a coach of Valencia

Match	Valencia	Sevilla
Round 1	1	1



Round 2	3	1

Myself

Remind name; Summarize your research in 1 sentences; hobby perhaps

Data and Algorithm

Data

La Liga (Spanish Top Division) match stats: scraped from whoscored.com
Training: 12/13-13/14, ~76 matches / team; Testing: 14/15, ~18 matches / team

Outcome of Team A's match	Features of team A	Features of team A's OP
1: win / 0: loss or tie	Pass, Shot, …	Pass, Shot, …

Algorithms

Step 1: Build a classification model for each team to identify important match features

Feature Selection	Random Forest
Classification model	Logistic regression using features selected above

Step 2: Recommend actions for a particular team and their opponent

Recommend actions	a) win any opponent b) win a particular opponent
Compute probability	Apply the models to the average stats of both teams