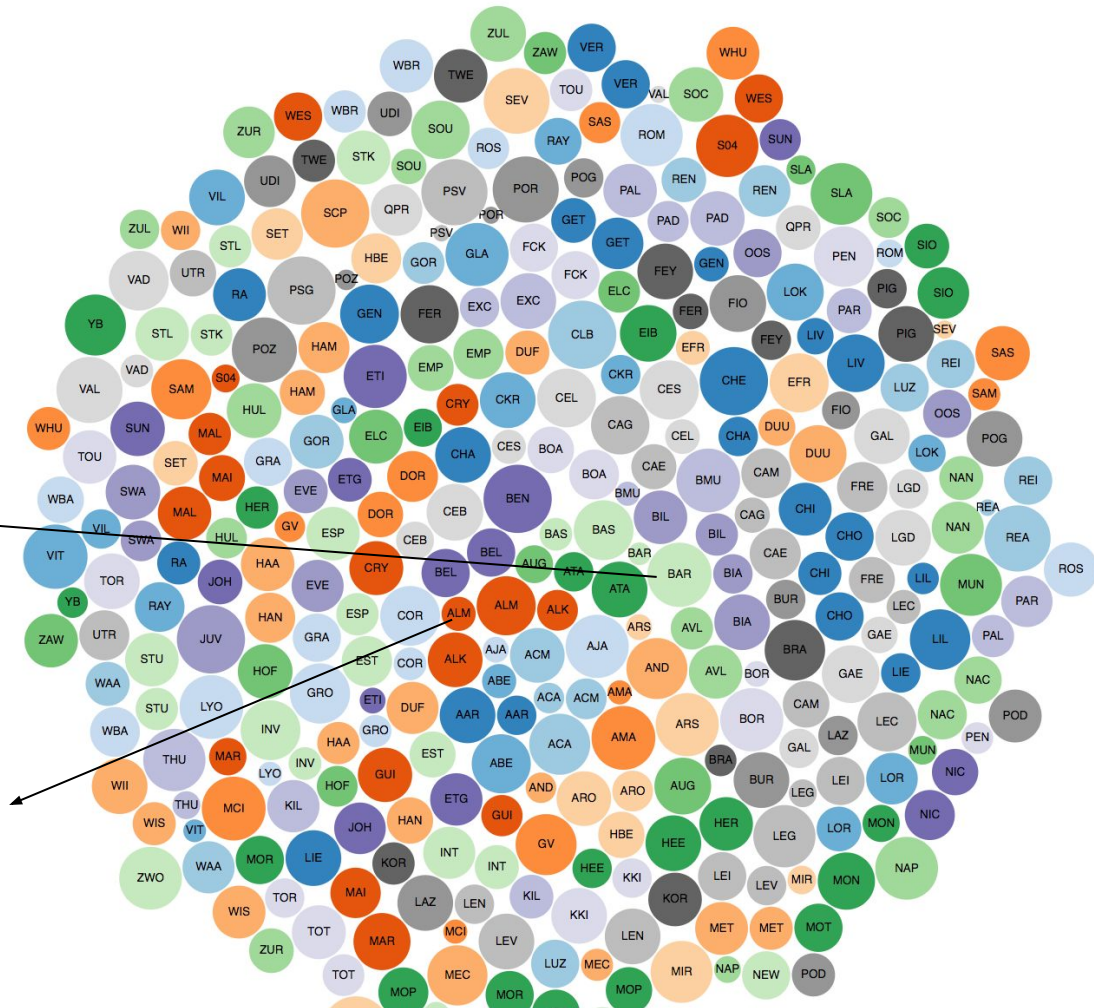

Soccer Match Analysis

Keerthi Pamulaparthi
Feb 15, 2017



EUROPEAN SOCCER DATA

SOURCE	Kaggle
Season	2014/2015
# Teams	188
# Countries	11
# Matches	3081



Home Team Win % = 100%

Home Team Win % = 20%

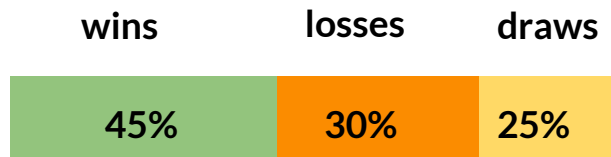
Prediction

Home Team Match
Outcome

FEATURES

& CLASSES

GOAL FEATURES



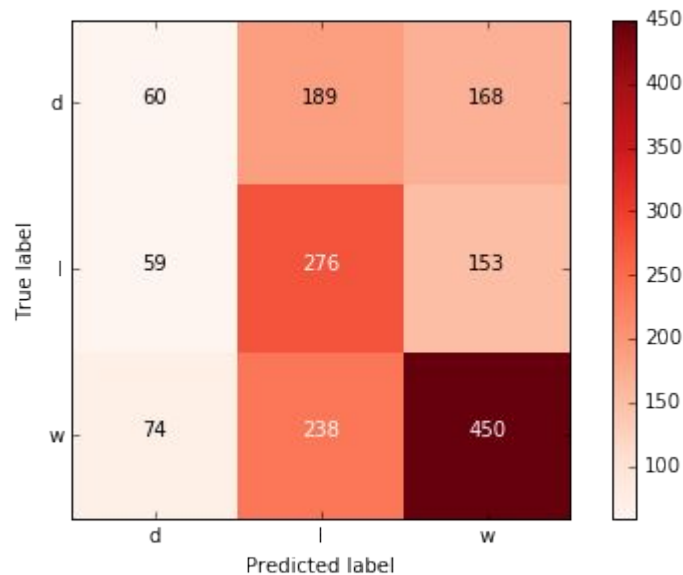
TEAM FEATURES

Model : Scenario 1

Goal Features

Best Model : Logistic Regression

Accuracy Score : 0.502

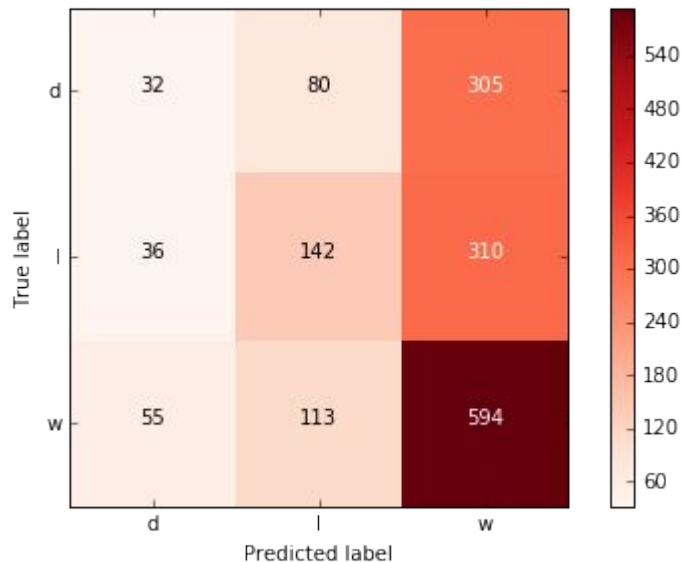


Model : Scenario 2

Team Features

Best Model : **Random Forest**

Accuracy Score : **0.47**

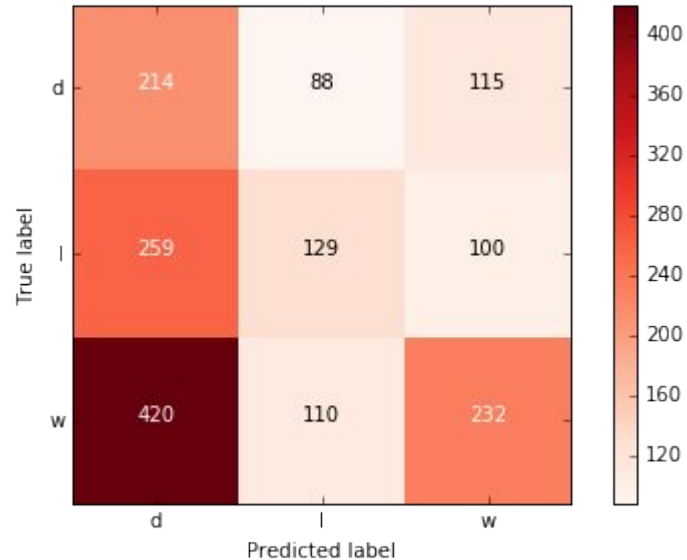


Model : Scenario 3

Goals + Team Features

Best Model : **Random Forest**

Accuracy Score : **0.49**

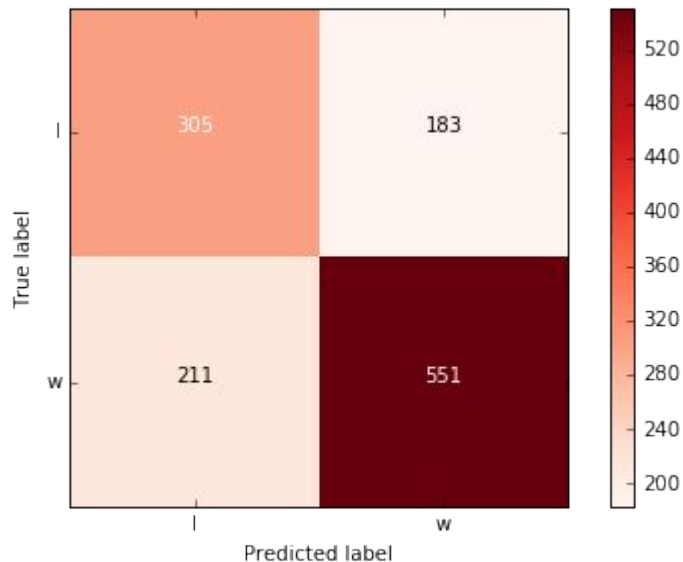


Model : Scenario 4

Goal Features - Draws removed

Best Model : **Logistic Regression**

Accuracy Score : **0.68**

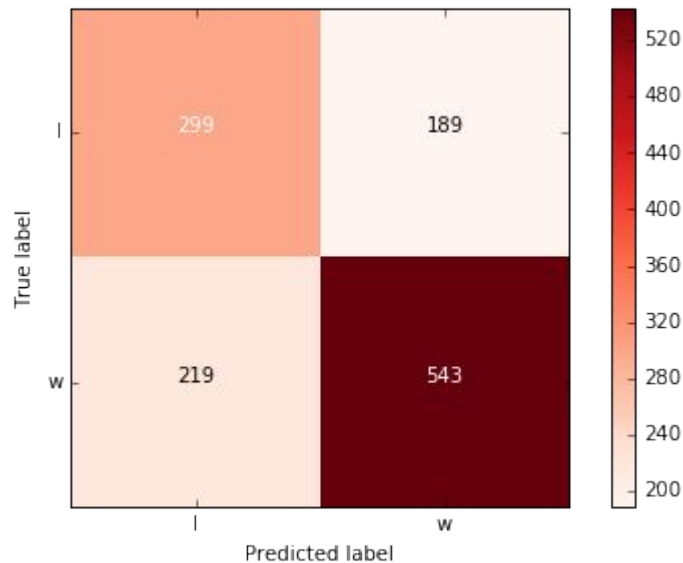


Model : Scenario 5

Goal Features - Draws removed + Team Features

Best Model : **Logistic Regression**

Accuracy Score : **0.67**

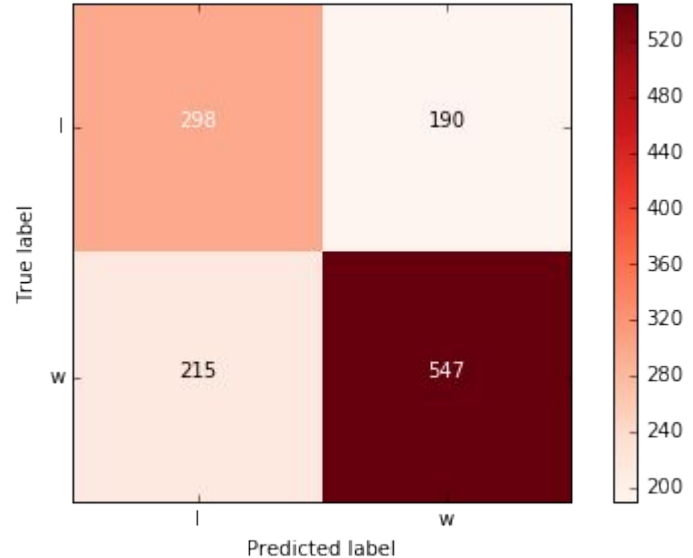


Model : Scenario 6

**Goal Features - Draws + Win Streak
+Team Features**

Best Model : **Logistic Regression**

Accuracy Score : **0.67**



ANALYSIS

1. Better Prediction For Home Team With Draws Removed
2. Having Team Features In The Model Did Not Make It Better

CONCLUSION

Home Team
Wins/Losses can be
predicted with
Greater Accuracy
than Draws

Next steps

More Features

Players data

D3 Visualization and Flask Web App

THANK YOU

Questions?

Features

Team Features:

- buildUpPlaySpeed
- buildUpPlayDribblingClass
- buildUpPlayPassing
- buildUpPlayPositioningClass
- chanceCreationPassing
- chanceCreationCrossing
- chanceCreationShooting
- chanceCreationPositioningClass
- defencePressure defenceAggression
- defenceTeamWidth defenceDefenderLineClass

GOAL FEATURES

(+/- DRAW, WIN STREAK)

- cumulative HOME TEAM goals FOR
- cumulative HOME TEAM goals AGAINST
- cumulative AWAY TEAM goals FOR
- cumulative AWAY TEAM goals AGAINST

Only Team Features : Scores

Logistic Regression	0.48	0.44
Logistic Regression (2nd degree)	0.64	0.40
Decision Tree (best parameters)	0.51	0.40
Random Forest (best parameters)	0.94	0.47