

Part II Project Progress Presentation

“Predicting the Outcomes of English Premier League Matches”

Joseph Marchant

Introduction

- English Premier League (EPL) has 20 teams
- Season is 38 matches
- Data covers 26 seasons
- Goal is to predict the outcome of matches from 3 classes: Home, Draw, Away.
- Can only use data that is known before game begins.
- Goal is 40% accuracy.
- Focus is on finding unique/ effective features.

Data Collection - 11v11.com

- A website for international football statistics
- Web-scraped over 10000 pages
- Stored into csv

Data Link	Referee	Stadium
arsenal-v-norwich-city-15-august	Alan Gunn	Highbury
arsenal-v-oldham-athletic-15-august	Jim Borrett	Stamford Bridge
arsenal-v-city-v-middlesbrough-15-august	Howard King	Highfield Road

HomePlayers	AwayPlayers	HomeScorers	AwayScorers
'4377(90)', '7800(90)', '37097(90)', '98(58)', '38680(90)', '846(90)'	'196', '484'	'511', '2266', '2266', '38680'	
'90)', '38546(90)', '31113(76)', '210(90)', '3620(90)', '1427(90)', '488'	'37518'	'1427'	
'4796(90)', '34414(75)', '24329(90)', '356(54)', '33173(90)', '280'	'1062', '5386'	'11296'	

Arsenal v Norwich City, 15 August 1992

John Jensen made his debut after moving from Brondy (Denmark) in July.

Score	4-2 to Norwich City
Referee	Alan Gunn
Competition	Premier League
Venue	Highbury
Attendance	24,030

[View Arsenal v Norwich City head to head](#)

Arsenal	2	Norwich City	4
Goals:		Goals:	
Steve Bould	28 	Mark Robins	69 
Kevin Campbell	39 	David Phillips	72 
		Ruel Fox	84 
		Mark Robins	84 
Starting lineup:		Starting lineup:	
+ David Seaman	Goalkeeper	+ Bryan Gunn	Goalkeeper
+ Lee Dixon	Defender/Right back	+ Ian Culverhouse	Defender
+ Tony Adams	Defender	+ Ian Butterworth	Defender/Centre back
+ Steve Bould	Defender	+ John Polston	Defender/Centre back
+ Nigel Winterburn	Defender/Left back	+ Rob Newman	Defender/Forward
+ David Hillier	Midfielder	+ Chris Sutton	Defender/Forward
+ John Jensen	Midfielder	+ David Phillips	Defender/Midfielder
+ Anders Limpar	Midfielder	+ Mark Bowen	Defender/Left back
+ Paul Merson	Midfielder	+ Ruel Fox	Midfielder
+ Kevin Campbell	Forward	+ Jeremy Goss	Midfielder
+ Alan Smith	Forward	+ Gary Megson	Midfielder

Data Collection - Other data sources

- Web-scraping
 - premierleague.com
 - Wikipedia.org
 - transfermarkt.co.uk
- Downloads
 - Football-data.co.uk
- Manual Collection
 - Physioroom.com

Data Cleaning/ Preparing for Models

- Missing/ incorrect values
- League table generator
 - Points Per Game, Goals Per Game, Clean Sheets Per Game, Position, Goals Conceded Per Game
- Prepare for models
 - Remove string columns
 - Add target column: {H = Home Win, D = Draw, A = Away Win}

A	B	C	D	E	F	G	H	I	J	K
hPos	hGSPG	hGCPG	hCSPG	hPtsPG	aPos	aGSPG	aGCPG	aCSPG	aPtsPG	target
6	2	1.5	0.5	2	17	0.5	1.5	0	0.5	3
22	0.5	2.5	0	0	8	1	0.5	0.5	2	1
9	1.5	1	0.5	1.5	7	1.5	1	0	2	3
1	3	1.5	0	3	3	2	0.5	0.5	2	1

Binary Models

- 80-20 train-test splits used
- Binary Classification
 - 1 = Home Team Win, 0 otherwise
- 5 “difference” features instead of 10
- Logistic Regression
 - 62% Accuracy
 - 51% Precision
 - 61% Recall
- Tensorflow NN
 - 61% accuracy best

3-way Classification Models

- 3-class Classification
 - Points: 3 = Home Team Win, 1 = Draw, 0 = Home Team Loss
- Tensorflow NN
- 30 hidden layers
- Learning rate = 0.1
- 100 epochs
- `Tf.nn.sparse_softmax_cross_entropy_with_logits`
- 49% accuracy on test set

Next Steps

- Keras NN
- More features
- Other ML Techniques: kNN, SVM, RandomForest
- Pipeline for predicting new matches
- In-depth evaluation

Thank you for listening.

Joseph Marchant