Fantasy

Premier League

Winning FPL with Machine Learning

Patrick Timmermans, Martin Vestergaard og Ruben Engmose

The game

No room for racism
Martínez
5.0
Mendy
5.2
Digne
6.1
Dallas
5.4
Alexander-Arno...
7.6
Shaw
5.3
Dias
5.9
Salah
12.9
Lingard
6.3
Ward-Prowse
5.8
Jota
6.7
Raphinha
5.5
Wood
6.6
Calvert-Lewin
7.5
Kane
11.4

Premier League
McCarthy
NEW (H)
Zouma
SHU (H)
Chilwell
SHU (H)
Cresswell
FUL (H)
Ziyech
SHU (H)
Zaha
LEE (H)
Son
WBA (A)
De Bruyne
LIV (H)
Bamford
CRY (A)
Calvert-...
MUN (H)
Kane
WBA (A)
Macey
AVL (H)
Riedewald
LEE (H)
Vesterg...
NEW (H)
Kilman
LEI (A)
GK
1
2
3

Our goal

Get as many point as possible!

And win the game

Data

name	team	assists	goals_scored	minutes	ict_index	team_a_score	team_h_score	was_home	opponent_team	total_points
Callum Wilson	Newcastle	1	1	90	9.2	2	0	False	6	12
Angelo Ogbonna	West Ham	0	1	90	8.7	1	2	True	2	9
Daniel Castelo Podence	Wolves	0	1	69	8.3	2	1	False	1	9
Tammy Abraham	Chelsea	0	0	78	2.2	0	0	True	17	2
James Justin	Leicester	0	0	90	2.4	2	1	True	8	1

Player_name	team_name	form	selected	lr_ict	was_home	games_this_gw	opponent_team_code	strength_overall_home	opp_strength_overall_away
Jayden_Bogle	Sheffield Utd	2.500000	7139	1.8	True	1	35	1000.0	1040.0
Jan_Vertonghen	Spurs	2.000000	238023	1.6	True	1	4	NaN	NaN
Sead_Kolasinac	Arsenal	0.000000	16068	0.0	True	1	14	NaN	NaN
Ollie_Watkins	Aston Villa	7.666667	384478	4.6	True	1	2	1100.0	1170.0
Roberto_Firmino	Liverpool	5.000000	1353215	6.4	False	1	57	1340.0	1140.0

The road to a complete team

- Model
    - Multiple models for different predictions
    - Criteria: Lightning fast (and accurate)
- Team selection
    - Running through approx. 10e+21 combination in seconds
    - Inspired by the idea of recursive elimination used in feature selection
- Team scoring & line-up
- Model tuning

The road to a complete team

Players round 1 - rated:

	Player_name	team_name	pos	value	point_pred_fourR	point_pred_oneR	relative_value
107578	Jordan_Pickford	Everton	1	50	9.336030	2.472648	0.186721
105260	Kepa_Arrizabalaga	Chelsea	1	50	8.000716	4.553492	0.160014
107920	Richarlison_de Andrade	Everton	4	80	12.475241	4.469689	0.155941
112214	Illan_Meslier	Leeds	1	45	6.753121	2.677089	0.150069
119890	Alex_McCarthy	Southampton	1	45	6.621116	3.254625	0.147136
105374	Andreas_Christensen	Chelsea	2	50	7.251445	3.624840	0.145029
107844	Dominic_Calvert-Lewin	Everton	4	70	10.103397	3.883943	0.144334
113278	Alisson_Ramses Becker	Liverpool	1	60	8.604009	4.442588	0.143400
124032	Lukasz_Fabianski	West Ham	1	50	7.132742	4.789048	0.142655
105032	Marcos_Alonso	Chelsea	2	60	8.477947	4.577460	0.141299
107502	Lucas_Digne	Everton	2	60	8.472981	2.075184	0.141216
109288	Marek_Rodák	Fulham	1	45	6.347299	3.607050	0.141051
105640	Reece_James	Chelsea	2	50	7.030134	3.281668	0.140603
105184	Kurt_Zouma	Chelsea	2	50	6.990171	3.823769	0.139803

Inital team round 1:

	Player_name	team_name	pos	value	point_pred_fourR	point_pred_oneR	relative_value
107578	Jordan_Pickford	Everton	1	50	9.336030	2.472648	0.186721
105260	Kepa_Arrizabalaga	Chelsea	1	50	8.000716	4.553492	0.160014
105374	Andreas_Christensen	Chelsea	2	50	7.251445	3.624840	0.145029
105032	Marcos_Alonso	Chelsea	2	60	8.477947	4.577460	0.141299
107502	Lucas_Digne	Everton	2	60	8.472981	2.075184	0.141216
105640	Reece_James	Chelsea	2	50	7.030134	3.281668	0.140603
105184	Kurt_Zouma	Chelsea	2	50	6.990171	3.823769	0.139803
108148	James_Rodríguez	Everton	3	75	10.391694	4.179337	0.138556
118864	John_Lundstram	Sheffield Utd	3	55	7.499907	2.340174	0.136362
105298	N'Golo_Kanté	Chelsea	3	50	6.627982	3.124460	0.132560
100624	Mohamed Naser_El Sayed Elneny	Arsenal	3	45	5.818248	2.418159	0.129294
105070	Jorge Luiz_Frello Filho	Chelsea	3	50	5.815241	3.264845	0.116305
107920	Richarlison_de Andrade	Everton	4	80	12.475241	4.469689	0.155941
107844	Dominic_Calvert-Lewin	Everton	4	70	10.103397	3.883943	0.144334
125362	Raúl_Jiménez	Wolves	4	85	11.880279	5.731295	0.139768

Inital team round 1 - with all restraints (team and value-cap):

	Player_name	team_name	pos	value	point_pred_fourR	point_pred_oneR	relative_value
107578	Jordan_Pickford	Everton	1	50	9.336030	2.472648	0.186721
113278	Alisson_Ramses Becker	Liverpool	1	60	8.604009	4.442588	0.143400
105374	Andreas_Christensen	Chelsea	2	50	7.251445	3.624840	0.145029
105640	Reece_James	Chelsea	2	50	7.030134	3.281668	0.140603
105184	Kurt_Zouma	Chelsea	2	50	6.990171	3.823769	0.139803
99940	Héctor_Bellerín	Arsenal	2	50	6.010505	2.442702	0.120210
113202	Virgil_van Dijk	Liverpool	2	65	7.344967	2.571936	0.112999
108148	James_Rodríguez	Everton	3	75	10.391694	4.179337	0.138556
118864	John_Lundstram	Sheffield Utd	3	55	7.499907	2.340174	0.136362
100624	Mohamed Naser_El Sayed Elneny	Arsenal	3	45	5.818248	2.418159	0.129294
100510	Willian_Borges Da Silva	Arsenal	3	80	8.804472	6.179244	0.110056
113240	Sadio_Mané	Liverpool	3	120	12.281109	9.287485	0.102343
107920	Richarlison_de Andrade	Everton	4	80	12.475241	4.469689	0.155941
125362	Raúl_Jiménez	Wolves	4	85	11.880279	5.731295	0.139768
120004	Danny_Ings	Southampton	4	85	10.881871	3.449545	0.128022

Inital team round 1 - optimized by four round point prediction and team set by one round point prediction

	Player_name	team_name	pos	value	point_pred_fourR	point_pred_oneR	relative_value
124032	Lukasz_Fabianski	West Ham	1	50	7.132742	4.789048	0.142655
107578	Jordan_Pickford	Everton	1	50	9.336030	2.472648	0.186721
105032	Marcos_Alonso	Chelsea	2	60	8.477947	4.577460	0.141299
113392	Andrew_Robertson	Liverpool	2	70	9.209394	4.079745	0.131563
105374	Andreas_Christensen	Chelsea	2	50	7.251445	3.624840	0.145029
124184	Aaron_Cresswell	West Ham	2	50	5.947127	3.358110	0.118943
122018	Matt_Doherty	Spurs	2	60	6.850802	3.026378	0.114180
100510	Willian_Borges Da Silva	Arsenal	3	80	8.804472	6.179244	0.110056
105526	Mason_Mount	Chelsea	3	70	7.706046	5.885123	0.110086
99674	Pierre-Emerick_Aubameyang	Arsenal	3	120	11.404673	2.862152	0.095039
100624	Mohamed Naser_El Sayed Elneny	Arsenal	3	45	5.818248	2.418159	0.129294
118864	John_Lundstram	Sheffield Utd	3	55	7.499907	2.340174	0.136362
125362	Raúl_Jiménez	Wolves	4	85	11.880279	5.731295	0.139768
107920	Richarlison_de Andrade	Everton	4	80	12.475241	4.469689	0.155941
107844	Dominic_Calvert-Lewin	Everton	4	70	10.103397	3.883943	0.144334

Model tuning

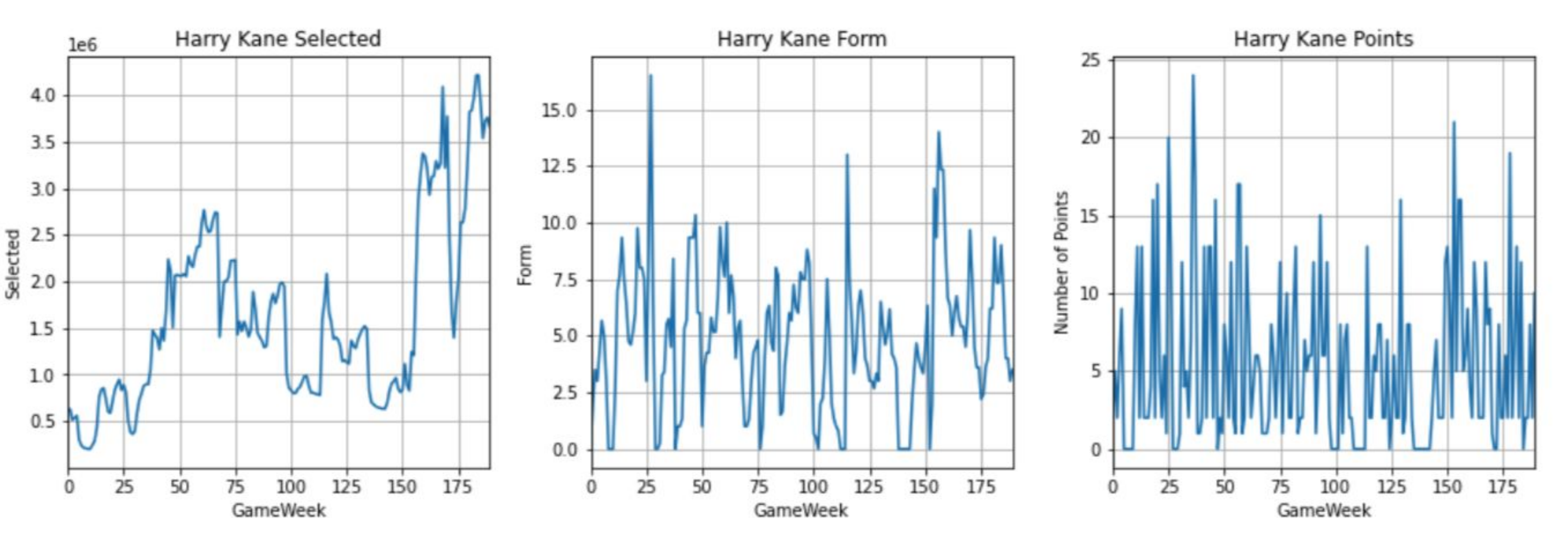
- One output but a ton of parameters
  - And training times of more than 30 minutes
- Focus on players that are likely to make the team
  - Only looking for ‘value’ players
  - Weights
- Essentially this made tuning any parameter very hard
  - Every attempt of hyper-optimisation was
  - Lack of model tuning?

(Dis)advantage opposed to expert knowledge

- It understood:
    - Match weeks
    - Team strengths
    - General player values

- It did NOT understand:
    - Basic strategies regarding positions, rotation and substitutions
        - It would buy Nick Pope, bench him for four rounds and sell him
    - Which players were definitely not playing
        - Other publicly available information
    - The more advanced rules

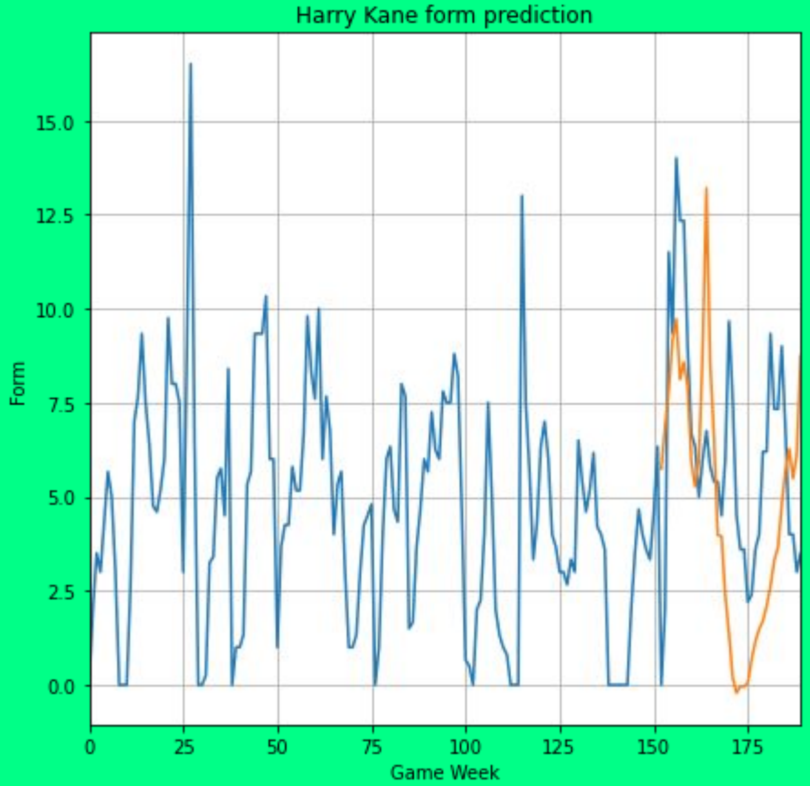
Player Example - 3 features over 5 seasons (190 gameweeks)

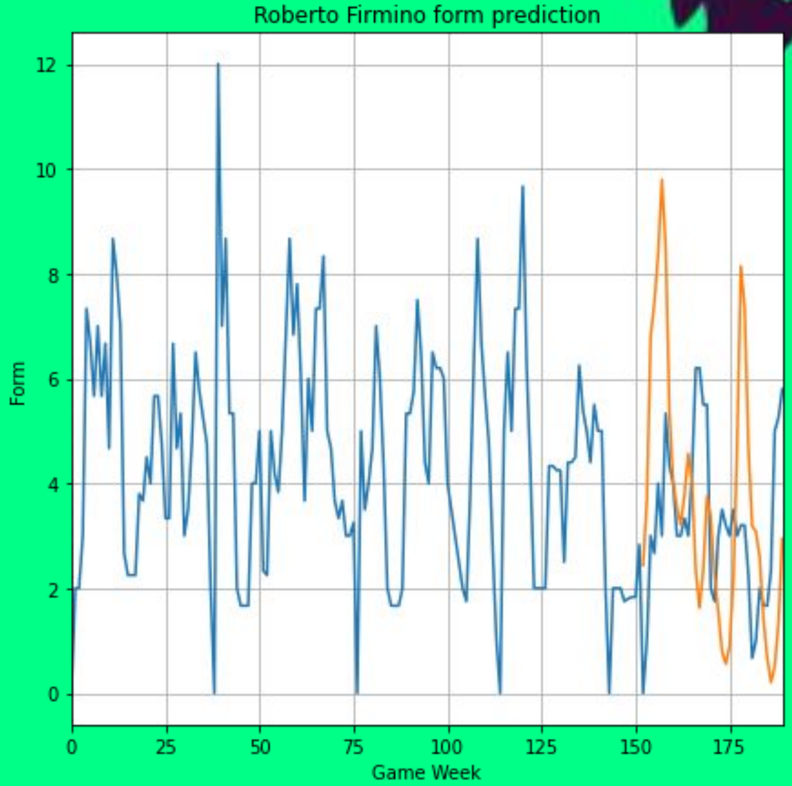


SLTM - Setup

- Goal: Predict the form of a player and use prediction as a feature
- Train-test split: Train 4 seasons - Predict 1 season (Time homogeneous)
- First idea: Train on all players and predict for each individual player
- How to create “train groups”?
- Separate model and train for each position (Hand plucked 5 players in each)
  - Keeper
  - Defence
  - Midfield
  - Striker
- Predict for individual players based on position
- Is the model and training data representative for the individual player?

Player example - SLTM prediction result





Prediction SLTM result and discussion

Possible Improvements

- Predict 1-5 matches each round with updated information (Instead a whole season)
- Choose representative players with least correlation
- How do we split our models? (Position, volatility, team, ...)
- How to reduce overfitting? (Train on more players)
- How to validate model? (Predicting on different types of players)

Fundamental constraints

- Few players with enough data (Conditional on players)
- Overall: Volatile and nearly impossible to predict (At Least not more than 5+ GW)
- “Enough data”?

Results

Rank	Team & Manager		Points
1•	Teddy Bears Utd Michael Coone	WINNER 2020-21	2680
2•	Brandvarme Drenge Jake Houe		2334
3•	Robbers tropper Ruben Larsen		2291
4•	1M		2187
5•	ML DreamTeam Robot	Top 13%	2173
6•	YellowBlueArmy Emil Hofman		2086
7•	2M		2066
8•	FC Kongpatty Patrick Timmermans		1923

Appendix

- FPL_data.ipynb
- Score_team.py
- Select_team.py
- FPL_Hacked.py
- Hyper_param_optim.py
- PL_SLTM.ipynb

References

SLTM Setup:
https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

Data:
https://github.com/vaastav/Fantasy-Premier-League

FPL_data.ipynb

Gets the available data from: https://github.com/vaastav/Fantasy-Premier-League combine the player data with the team data, to construct the dataset we use for training.
Making new variables such as Form, Last rounds ICT-index, Points scored until given gw, a team-code which is consistent over different seasons and informations about the next 5 gws.
Lastly combining rows concerning gws where a player plays more than one game, and constructing missing rows, where players have a blank gw.

gw = Gameweek/one round of the game

FPL_hacked.py

The main script keeping track of game week at iterating throughout the season. From here is made all the references to the remaining scripts. It takes FPL_DATA.csv as input generated from FPL_data.ipynb.

For each game week, it trains two LightGBM models used respectively to predict the points in the following round and another to predict multiple rounds ahead.

It calls the two files score_team.py and select_team.py, which, as their title indicates, handles the team composition and the team scoring.

FPL - Form_prediction_SLTM.ipynb

The script illustrations the variation of different features over gameweeks for different players

We set up the test size/train window and the LSTM model. Here we also set the forward, loss and optimizer functions.

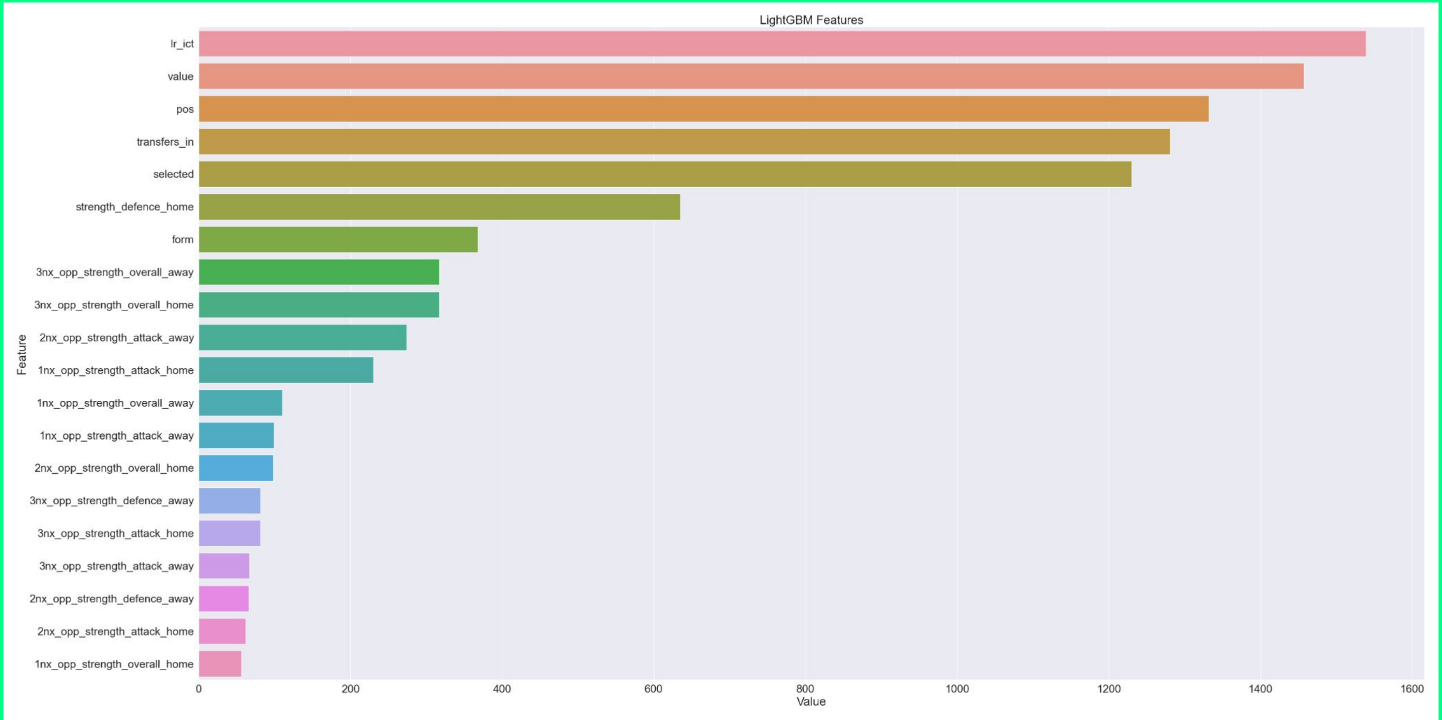
We train the model on specific groups of players by position.

The model is trained over 100 epocs for each “train player”.

At last, we plot the predicted form over the actual form for the last season and export all form predictions as a new feature in a csv file

Feature importance

Model predicting 4 rounds ahead



Feature importance

Model predicting 1 rounds ahead