Predicting good strategies for picking players in Fantasy premier league

Praneet Nandan and Adarsha Mohit Sahu

Recap

Till midway, we had implemented an SVM model on the data, and then improved it's output using insights from one of the reference papers to include better feature vectors

1. Based on the fact that high ranking managers have a good player choice, we included the selected by percentage of the top 8000 manager teams (which was for a lot of players much different than the total selected by percentage)
2. We removed the inbuilt team difficulty index of FPL and replaced it with betting odd, something that improved the performance.

Recap

We also identified some problems,

1. The game-week data is sequential, so maybe the time series nature of it is getting lost in SVM
2. We plan to do separate analysis for different positions, but for a single game week data individual position entries are low, for different game-weeks the number of feature vectors change

To combat these two problems, two birds one stone, we planned to replace the gameweek score data with one feature, predicted using time-series modelling

Literature we talked about in midway

- Gupta A (2019) Time Series Modeling for Dream Team in Fantasy Premier League. International Conference on Sports Engineering (ICSE'17) CoRR abs/1909.12938

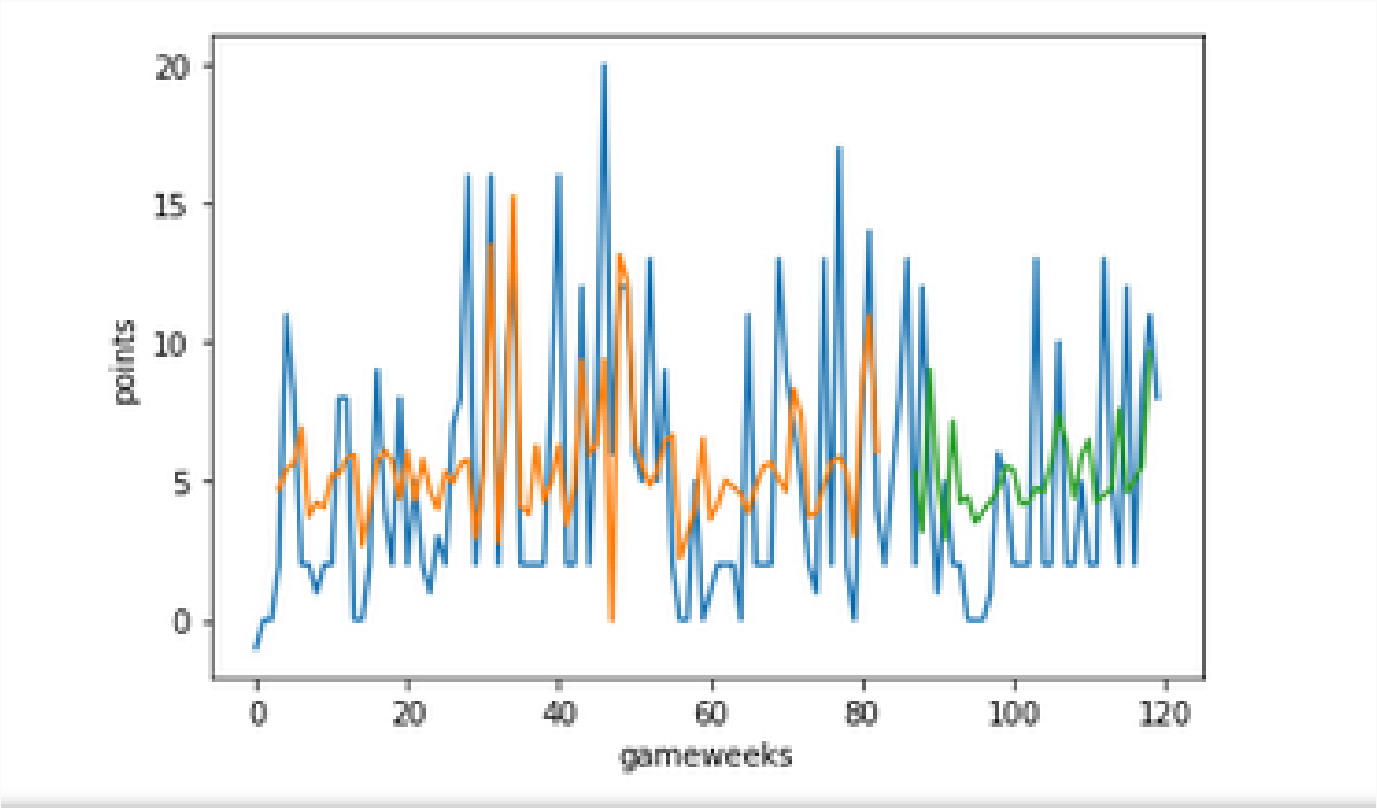
They have used LSTM neural networks and ARIMA to study performance over different seasons, and to predict for one season. We want to do that for different game-weeks
They mix both in some percentage, p1,p2 to get a hybrid predictor for the entire season
Our plan- use LSTM RNN and ARIMA to put two features as input to the SVM

LSTM

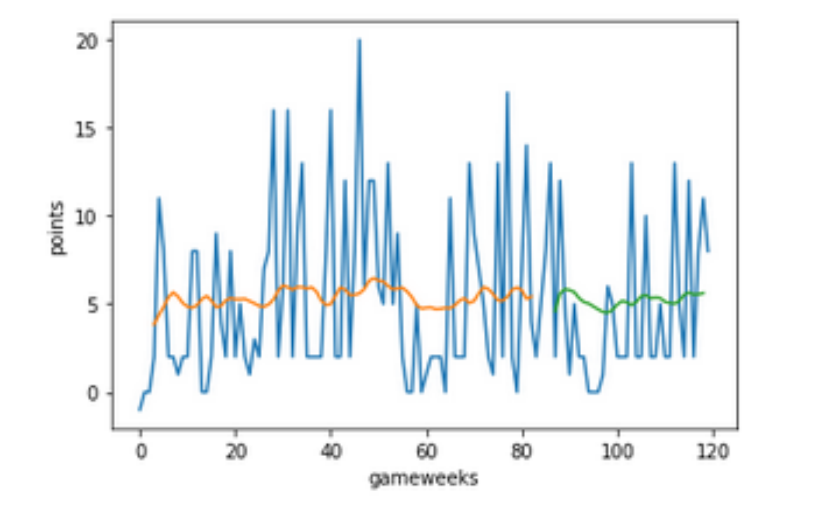
LSTMs are a form of neural networks that are good at handling sequence dependence.

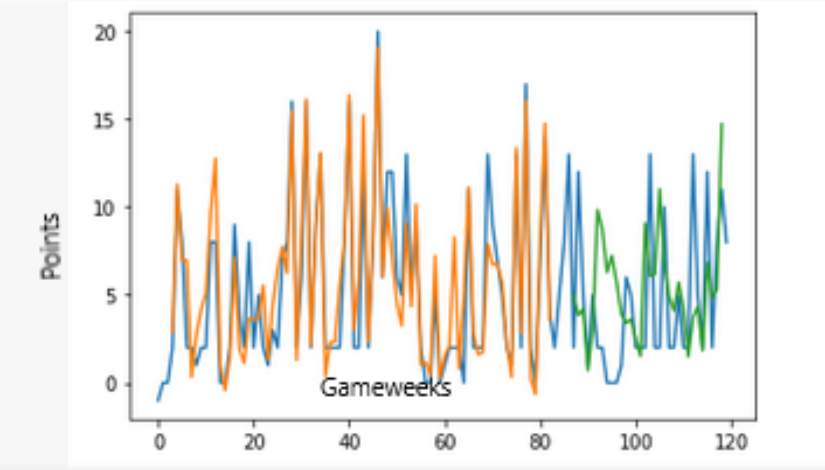
We scrapped previous season's individual game-week data and then tested out LSTM-RNN on it

Data for Jamie Vardy

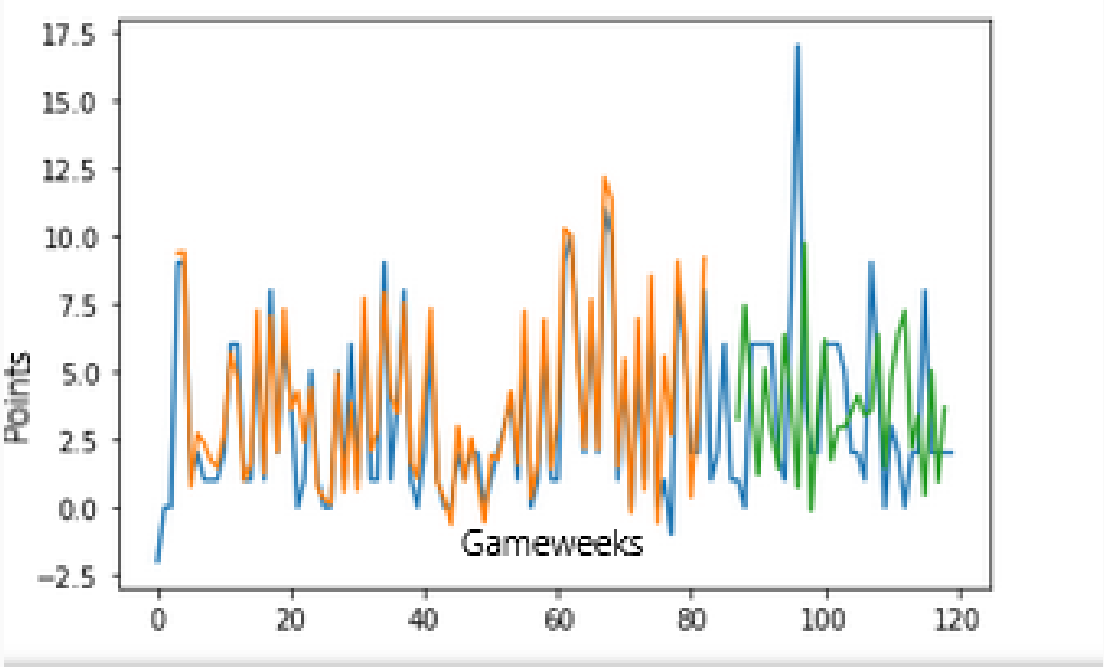


LSTM


points
gameweeks
0
5
10
15
20
0
20
40
60
80
100
120


Points
Gameweeks
0
5
10
15
20
0
20
40
60
80
100
120

LSTM



Results do not seem to be very good

Still went ahead to try it with SVM, worse output than with game-week features

ARIMA, Maybe?

ARIMA, or Autoregressive Integrated Moving Average Model is also a popular choice for time series analysis

It has three aspects, Autoregression, differencing and Moving averages

Hence it has three parameters, p,d and q

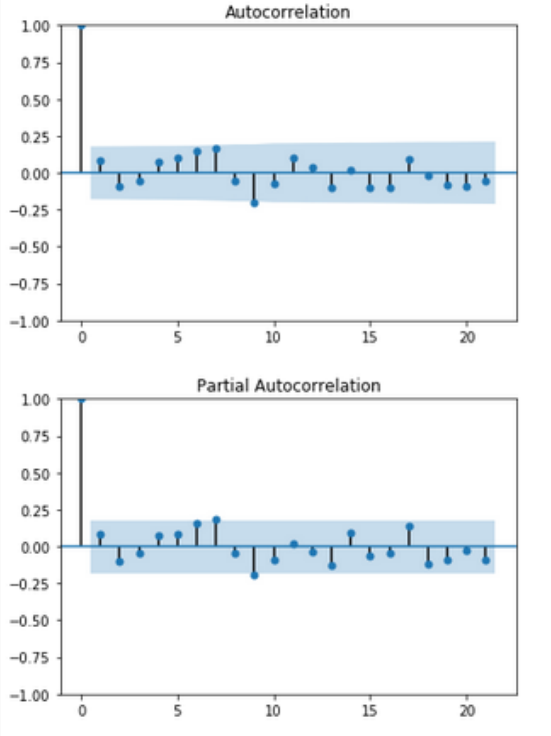
p is the order of the Auto Regressive (AR) term. It refers to the number of lags of Y to be used as predictors.

The value of d is the minimum number of differencing needed to make the series stationary. If the time series is already stationary, then d $= 0$.

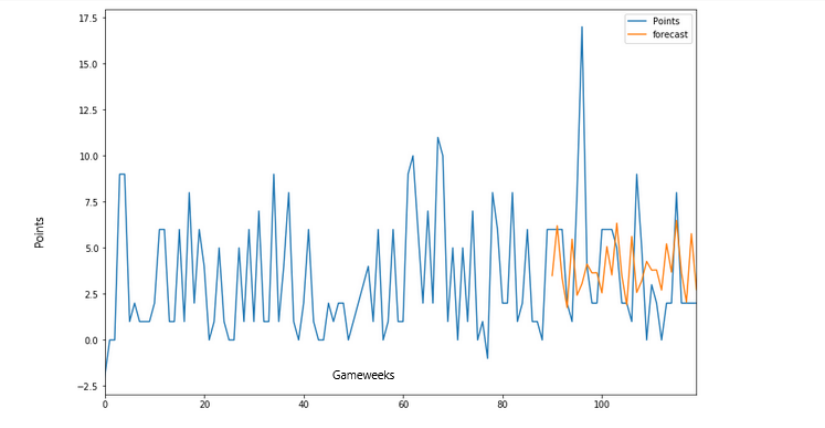
q is the order of the Moving Average (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

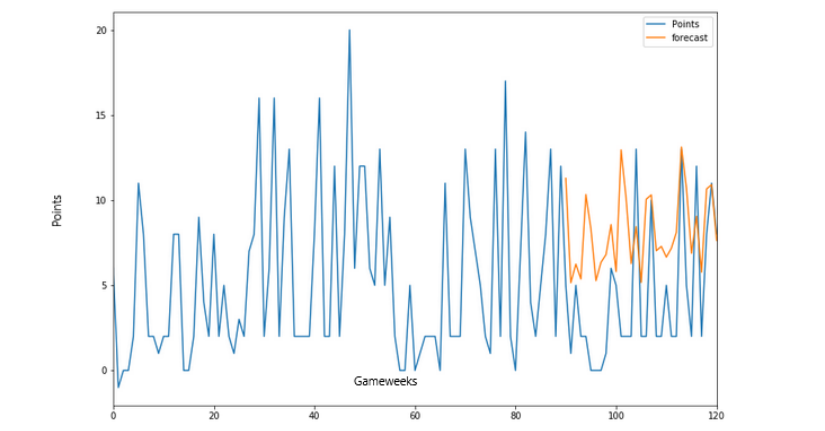
ARIMA

For finding the best ARIMA parameters, we used the Akaike's Information Criterion (AIC), and first checked that the series is stationary using Augmented Dicky-Fuller test



ARIMA


Points
forecast
Points
Gameweeks


Points
forecast
Points
Gameweeks

ARIMA

Seems that even though ARIMA is not that good either, it is able to predict which game weeks the player is going to do well in
To capture this, we ran ARIMA on all players that we had the big data set for and also took the average score of them throughout. Then we made a feature that is 1 if ARIMA predicts more than average and 0 if less. Using this and average score till that gameweek as features instead of gameweek data, we ran our SVM once again, this time separately for different positions and with more gameweeks

Results

For the Goalkeeper position,

Threshold=3

Accuracy: 0.926829268292683

Precision: 0.7142857142857143

Recall: 0.8333333333333334

Threshold=5

Accuracy: 0.926829268292683

Precision: 0.5

Recall: 0.6666666666666666

Results

For the Midfielder position,

Threshold=3

Accuracy: 0.9144736842105263

Precision: 0.8333333333333334

Recall: 0.4761904761904761 6

Threshold=5

Accuracy: 0.9539473684210527

Precision: 1.0

Recall: 0.4166666666666666

Results

For the Defender position,

Threshold=3

Accuracy: 0.872

Precision: 0.6

Recall: 0.3333333333333333

Threshold=5

Accuracy: 0.968

Precision: 1.0

Recall: 0.5

Results

For the Forward position,

Threshold=3

Accuracy: 0.8431372549019608

Precision: 0.5384615384615384

Recall: 0.7777777777777778

Threshold=5

Accuracy: 0.9019607843137255

Precision: 0.4

Recall: 0.5

References

- Bonello N, Beel J, Lawless S, Debattista J. (2019). Multi-stream Data Analytics for Enhanced Performance Prediction in Fantasy Football.
- Gupta A (2019) Time Series Modeling for Dream Team in Fantasy Premier League. International Conference on Sports Engineering (ICSE'17) CoRR abs/1909.12938
- O'Brien JD, Gleeson JP, O'Sullivan DJP (2021) Identification of skill in an online game: The case of Fantasy Premier League. PLoS ONE 16(3): e0246698. https://doi.org/10.1371/journal.pone.0246698