FIFA 2018 CHALLENGE

UpGrad PGDDS, June 2018 Cohort

ABSTRACT

Project Report: UpGrad FIFA 2018 world cup football winner prediction challenge.

Bangalore Analysts Team

Roopa Narayan

Sreevani Nagendran

Shekhar Shastry

Madan Padaki

CONTENTS

June 30, 2018

PROJECT REPORT: FIFA ’18 CHALLENGE

OBJECTIVE

The objective of the FIFA 2018 challenge is to predict the quarter-finalists, semi-finalists, finalists and the ultimate winner of the FIFA 2018 world cup being held in Russia. On the verge of Round-of-16, we have 16 teams that have qualified to fight for the world cup. The key is to identify which teams will make it to the next levels and that one team that will emerge at the top!

APPROACH

As a team, we research a host of historical data and statistics available for the top 16 teams and looked at various possible approaches to predict the winner.

Initially we considered the approach of collating historical information pertaining to each team’s performance in terms of current ranking, goals scored, top players, past world cup record and running some statistical algorithms to predict the winner. However, as we researched this, we found a lot of data available and numerous predictions already available based on complex mathematical models.

We decided to do a ***Poll-of-Polls*** and apply additional algorithms on top of the available prediction data to analyze and predict the winning team.

DATA PREPARATION & DATA CLEANSING

We collated the following data for our analysis:

1. World cup prediction data across various sources, which have been based on statistical analysis of the teams’ performance. We collated a total of 13 predictions from different sources.
2. Current team standings in term of goals scored, ranking in their respective groups, goal difference

We cleansed the data to have consistent naming, de-normalized to apply excel pivots/formulae as needed. This final data can be seen in the “Predictions Data” sheet of the excel workbook submitted.

DATA ANALYSIS

We analyzed the polls data with two approaches:

1. Team predictions' count:
   From the predictions data, we looked at the number of times each of the qualified teams (2018 Round of 16 qualified list) got predicted as a winner at each stage from quarter final to winning stage. We then looked at top 8-4-2-1 teams emerging across each stage. This analysis can be viewed in the "Poll_On_Polls_Approach A" sheet of the excel workbook. A snapshot of the final tally looked as below:

Round of 16 (Already Qualified)	Quarter Finals	Semi Finals	Finals	Winner
Brazil	12	10	8	5
Belgium	12	4	3	2
Spain	12	5	3	1
France	10	7	0	
Argentina	10	5	2	0
England	9	0	0	
Portugal	7	3	0	
Uruguay	3	2	1	1
Columbia	3	0	0	
Croatia	3	1	0	
Mexico	1	0	0	
Russia	1	0	0	
Denmark	1	1	0	
Sweden	0	0	0	
Switzerland	0	0	0	
Japan	0	0	0	

2. Weighted score for prediction accuracy:
   Since the above approach was a vertical cut, we also looked at allotting a weightage for prediction accuracy in each subsequent stage and hence the best prediction with a highest weighted score was selected. This analysis can be viewed in the "Poll_On_Poll_Approach B" sheet of the excel workbook.

   However, the biggest concern with the above two approaches was existence (or not!) of teams outside of the final qualified round of 16 list. Hence, we had to take a combination of the above analysis and the current world cup performance data to predict the winner.

FINAL PREDICTIONS

We applied the following algorithm to arrive at our final predictions as presented below:

Work with the draws for Round of 16. For each draw, apply the below logic to arrive at the winner of that game:	
A	Get the Team Predictions Count data for each team and pick the team with higher number of predictions or predicted as a better team in the best prediction as the winner of that match.
B	In case of a tie in step A above or if the team is not found in A data, look for the current match standings of the teams. The team which has *higher rank/standing* within its group gets priority over the other.
C	In case of a tie in step B above, look for the team which had a tougher or higher challenging group to play with. (Challenge here to be determined based group teams' rankings in poll predictions data count) and pick the team which has secured that position in a *higher challenging group*.
D	In case of a tie in step C above, do by goal difference of the two teams in the current standings and pick the team with a higher goal difference as the winner.

With the above logic, the predictions for the quarter finals, semi finals and finals looks as below:

Few logic examples:

1. Sweden vs. Switzerland:

   Sweden and Switzerland both not found in predictions data. Sweden topped its group. Had a similar challenging group in comparison to Switzerland and also has a higher goal difference. Hence selected.

2. Sweden vs. England:
   Sweden not found in predictions data. However, England ranked high in predictions. Hence selected.

PROJECT SUBMISSION

The following files have been submitted on the Github link provided below:

https://github.com/getnsv/FIFA2018ChallengeUpgrad

1. fifaprojectreport.pdf
2. Final_worldcup2018.xlsx

REFERENCES

https://www.fifa.com/worldcup/groups/

https://www.kaggle.com/agostontorok/soccer-world-cup-2018-winner

https://www.ladbrokes.com.au/info/soccer-world-cup/world-cup-teams/

https://www.telegraph.co.uk/football/world-cup-2018-wallchart-predictor/

https://www.printyourbrackets.com/images/fillable-world-cup.pdf

https://www.cbssports.com/soccer/world-cup/news/2018-world-cup-predictions-russia-2018-brackets-favorites-expert-picks-and-upsets-by-cbs-sports-experts/

https://www.telegraph.co.uk/world-cup/2018/06/28/world-cup-2018-groups-latest-teams-reach-last-16-standings-tables/

http://www.gracenote.com/sports/fifa-world-cup-predictions-2018/

https://www.gqindia.com/content/fifa-2018-world-cup-predictions-winners-goldman-sachs-ai/

https://www.technologyreview.com/s/611397/machine-learning-predicts-world-cup-winner/

https://www.kaggle.com/agostontorok/soccer-world-cup-2018-winner

https://arxiv.org/pdf/1806.03208.pdf