



## FOOTBALL MATCH WINNING TEAM PREDICTION USING MACHINE LEARNING

Mr. Nishant. G. Hulwan  
Computer Department  
MGM's College of Engineering and Technology  
Navi Mumbai, India

Prof. Sachin chavan  
Computer Department  
MGM's College of Engineering and Technology  
Navi Mumbai, India

**Abstract:** Football is the first well-known tournament within the world. The Federation Internationale de Football Affiliation (FIFA) is the institution that controls and directs this tournament and facilitates all the associations around the world, at both the proficient association and nation levels. The World Cup is the most significant sports competition in the world, which is held once every four years. The national groups play regional qualifiers organized by confederations. Those qualifiers decide the 32 groups that will advance to the final competition. FIFA makes the final draw for the game by vote, and the 32 groups apportioned into eight bunches of four. The beat two groups from several gathers move into the knockout stages (i.e., circular 16, quarter-finals, semi-finals and the ultimate)

**Keywords:** Prediction, Artificial Intelligence, Learning.

Cup, two competing models for the expectation of the FIFA World Cup fitted and examined.

### I. INTRODUCTION

1.1 In sports, football is the game, which is the most exciting & favorite game in the world. We have a massive amount of data related to football available on the internet. In any sports league, there is a keen interest to know which team will be the champion at the end of the championship. This data can be used to predict the outcome of the future matches as proposed by **Sujit Nair [1]**. Many efforts have been made to predict football matches result and selecting significant variables in football. Prediction is very useful in helping managers and clubs make the right decision to win leagues and tournaments as proposed by **Darwin Prasetyo [2]**. Analysis of the results of sports matches has always been a hot topic for both data analysts and the general public. Modern methods for processing data, in combination with the strong computational power of computers, allow us to predict the results of the next matches by applying machine learning algorithms on collected data as proposed by **Norbert Danisik [3]**.

1.2 In this article an approach for the examination and forecast of soccer coordinate comes about is proposed. This proposed system defines on a regularized Poisson relapse demonstrates that incorporates different possibly influential covariates depicting the national teams' victory in past FIFA World Glasses. Also, comparative to Bradley-Terry-Luce models, the contrast of team-specific impacts of the competing groups available in these models can be identified, and the result can predict out of it. It talked about that inside the generalized linear demonstrate (GLM) system the team-specific implications can either joined within the shape of settled or random effects. In arrange to realize variable determination and shrinkage, we utilize custom-made Rope approaches. Based on the three going before the FIFA World

### II. OBJECTIVES

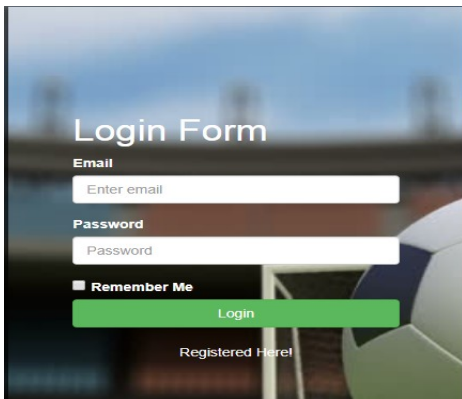
The purpose of this paper is to derive conclusion different from the previous method where the winning probabilities for each group were gotten basically by accumulating winning chances from a few online bookmakers. Based on these winning probabilities, by reverse competition reenactment, team-specific capacities can be computed by matched comparison models. Utilizing this procedure the impacts of the tournament draw can strip. Next, pairwise probabilities for each conceivable amusement at the comparing competition can anticipate and, at long last, the entire race can simulate. It ought to be famous that this strategy will continuously predict the group that has the most reduced (regular) bookmaker chances as the competition victor and, thus, is undoubtedly accepting that the bookmaker's ability covers all accessible information. This tournament is often not unreasonable, as one can undoubtedly anticipate bookmakers to utilize advanced models when setting up their chances, as they have substantial financial motivations to rate the group qualities of soccer groups correctly.

**(a) Registration**



The registration form is displayed on a background image of a football stadium. It includes the following fields and elements:

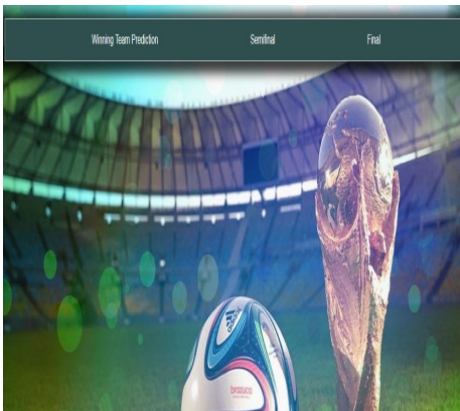
- Registration Form** (Title)
- Username**: Input field with placeholder "Enter User Name"
- Email**: Input field with placeholder "Enter email"
- Password**: Input field with placeholder "Password"
- Confirm Password**: Input field with placeholder "confirm Password"
- Registration** (Submit button)
- [Already have an account? Sign in.](#)



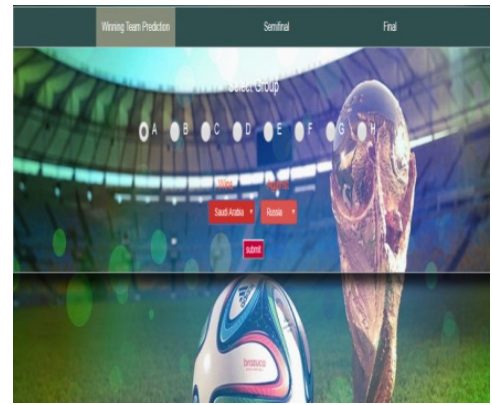
The login form is displayed on a background image of a football stadium. It includes the following fields and elements:

- Login Form** (Title)
- Email**: Input field with placeholder "Enter email"
- Password**: Input field with placeholder "Password"
- Remember Me**
- Login** (Submit button)
- [Registered Here!](#)

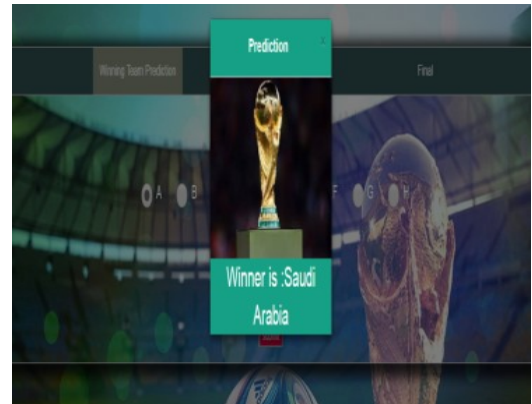
**(b) Login**



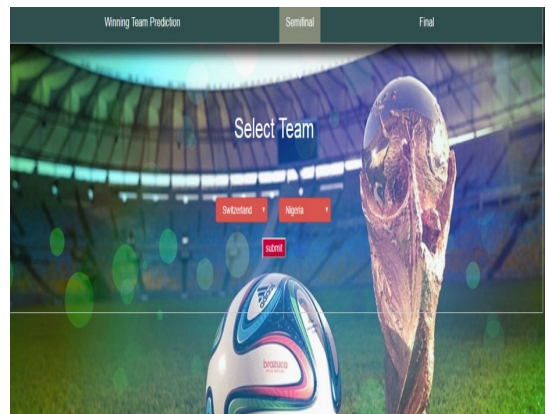
**(c) Home Page**



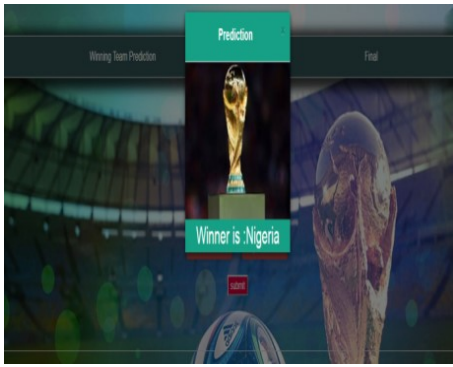
**(d) Winning team prediction Group wise**



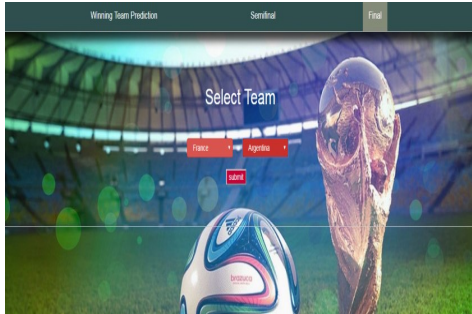
**(e) Prediction Result**



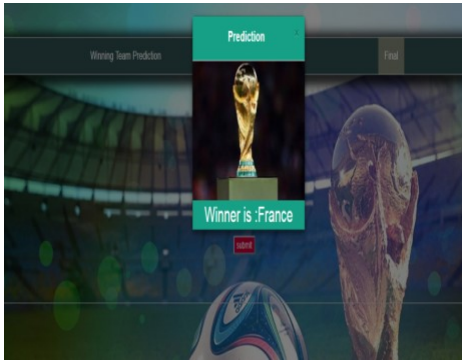
**(f) Semi-final Prediction**



(g) Prediction Result



(h) Final Prediction



(i) Prediction Result

### III. SCOPE

World Cup Prediction Meter is hugely important as the previous method doesn't have a methodical way of analyzing which leads to the moment category of approaches that based on regression models. A straightforward standard straight relapse approach was used by Stoy *et al.* (2010) to analyze the victory of national groups at the FIFA World Cup. The triumph of a group at a World Cup measure by a defined point scale that's gathered to be ordinarily conveyed. Adjacent to some sport-specific covariates too political-economic, socio-geographic as well as some devout and mental impact factors considered. Based on this demonstrate, a winning prediction for the FIFA World Cup given

### IV. OUTLINE

While highlighting this issue, a game-changing procedure thought where the teams and players are studied by their performance and while analyzing the games played. Based on which we derive into an inference of which driven to the creation of this thought. It also appears exciting to use Data Mining techniques to estimate future insights. We focus primarily on the deriving conclusion based on the series of victories, which takes after considering the performance statistics of players instrumental in winning the world cup. This player would avoid the churning of players of the season since they would be pulled in to getting more focused and way better players through such expectations.

### V. LITERATURE SURVEY

#### INTRODUCTION

We studied a few kinds of literature which included a few of the expectation calculations to foresee football comes about for diverse countries and played by little kids on alleys and streets it is a favorite sport in the world.

We utilized the discretion based on Bradley Terry Model a probability model which predict the outcome based on paired comparison to predict the outcome of the game which has been used by various football clubs namely Tottenham Hotspur football club.

The winning probabilities for each team playing in a league obtained by aggregating winning chances from a few online bookmakers that predict the outcome of matches based on players performance and team results. Based on these winning probabilities, by converse competition simulation team-specific capacities can be computed by matched comparison models. Using this method the impacts of the competition draw stripped.

It should be eminent that this methodology will predict the team that has the lowest (average) bookmaker chances as the tournament winner and, hence, is implicitly assuming that the bookmaker's expertise covers all the information. This approach is realistic, as one can undoubtedly anticipate bookmakers to use advanced models when setting up their odds, as they have substantial economic grounds to rate the team strengths of soccer teams correctly Considering all things, when a group suddenly wins a significant competition, as, for the case, Greece at the EURO 2004. Even though any measurable show will have genuine issues to anticipate accurately beneath such an unforeseen occasion, it would be alluring to at slightest draw outcomes from such an incident concerning future competitions.

Further, we can discuss the second approach that based on regression models. A standard linear regression approach was used by Stoy *et al.* (2010) to analyze the winner of World Cups. The success of a team in a World Championship is evaluated by an explained point scale that is supposed to generally distribute

among the groups. Beside some sport-specific covariates also political-economic and socio-geographic as well as some religious and cerebral influence variables are considered. Forecast for the FIFA World Cup 2010 was acquired based on this model. In contrast to Stoy et al, most of the regression approaches directly replica the digit of goals scored in a single soccer matches, assuming that the number of goals scored by each team follows a Poisson Distribution model, for example, Maher (1982), Lee (1997), Dyte and Clarke (2000), Rue and Salvesen (2000) and Goldman-Sachs Global Investment Research 2014.

Based on a survey of International matches Goldman-Sachs Global Investment Research has set up a regression model which depends on the statistics of the number of goals scored by each side in the game. They are dependent on the Poisson distribution. They have utilized six explanatory covariates: Difference between two teams in Elo rankings, the average number of goals scored in the tournament.

The name of the game won by the competing teams over the last ten years, out of which five mandatory international competitions, a dummy variable used to indicate whether the regarding match was a World Cup match.

A dummy variable used to indicate whether the team played is in its home country, a team-specific dummy variable was used to show whether the considered team played on its home continent.

And, finally based on the estimated regression parameters, the probability distribution for the outcome of each game in the championship must be obtained, and a Monte Carlo simulation with 100,000 draws used to generate or identify the probabilities of teams which reaching particular stages of the tournament, up to winning the championship.

### 1.3 Literary Review

We want to mention a completely different prediction approach, which cannot compare with one of the two major categories of statistical methods for modeling soccer data. It called as Soccer Match Index (SMI). SMI is a rating system, which uses significant historical data for both international and club level to predict the winning predicament of a match.

This algorithmic system uses a handful of years of data, taking into account goals scored and allowed and quality of the lineup fielded, and the location of the match. Also, the index weights recent events more heavily, and also takes into account the importance of the game.

### Summary of Literature Survey

Finally, we can compare the results of both approaches to regulate a final model, which uses for predicting the current FIFA World Cup 2014. It can recognize that is contradictory to other team sports, such as basketball, ice-hockey or handball, in pure soccer chance plays an important role. An important reason for this is that, compared to other sports, in soccer, fewer points (goals) are scored, and thus unique game situations can have a tremendous effect on the outcome of the match. One consequence is that now and then alleged (and

unpredictable) underdog win tournaments<sup>4</sup>. Nevertheless, it can be fascinating to scrutinize the connection and dominion structure between different potentially dominant covariates and the success of different soccer teams (in our case regarding the number of goals they score). Besides, we hope to get more awareness into which covariates are already covered by bookmakers' odds and which covariates may give some additional and useful information.

## VI. OVERVIEW

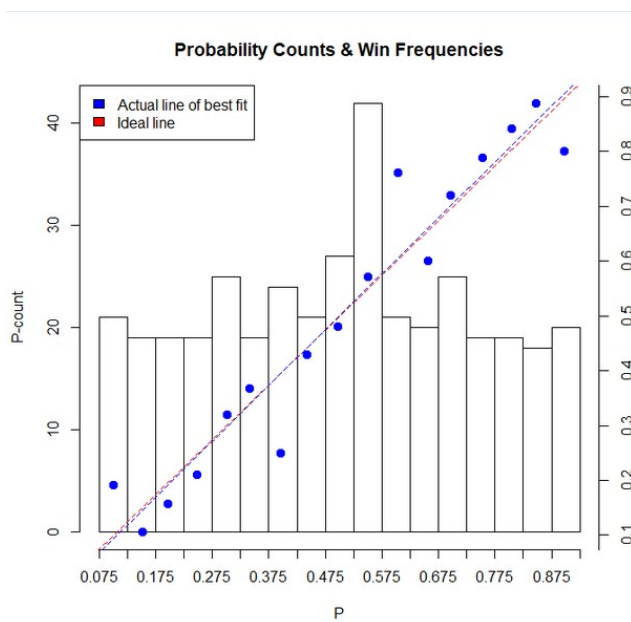
Our Model defines our analysis which concentrates on the number of goals a team scores against a specific opponent. For each group, special attack and defense parameters considered. Furthermore, the covariates of both sides, which might influence the number of scored goals, are found in the form of differences. Based on the data obtained, we derive a conclusion and predict the winning team.

The following variables are used to describe the structure of the teams. Each variable observed with the final squad of 23 players nominated for the respective World Cup.

The inclusion of a maximum number of teammates in a Football team can be hazardous because each player would have different attributes and according to player's performance, we need to predict the outcome. If maximum players from one club play together in a national team, this automatically leads to an improved presentation of the team as the teammates know each other better. Therefore, covariates are counted and included by maximum and second maximum number of members in the club. The median age of all 23 players is perceived to bring possible differences between old and young teams.

Many Champions players – For Example, The European club leagues are evaluated to be the best federation in the world; therefore, the competitions from teams between the best European teams, namely the UEFA Champions League and the Union of Football European Association Europa League can see as the most prestigious and valuable competitions on a club association level. As a computation of the success of the players on a club level, the number of players in the semifinals (taking place only weeks before the respective World Cup) of these competitions counted.

The number of players playing abroad from each team considered as the player from overseas could have better skills than the player from the unit itself. Finally, we can conclude that the national organizations strongly differ in the names of different players playing in a particular league of the respective country and players from leagues of other countries. For each group, the number of players playing in clubs abroad counted.



## VII. EXISTING SYSTEM ARCHITECTURE

Existing architecture uses two methodologies a] Poisson regression and b] Bradley-Terry-Luce models A] Poisson Regression - Poisson variables depends on two influence variables of each team and the match place. Poisson regression is used to approximate parameters for the model and based on these parameters the matches played during FIFA World Cup can be simulated. B] Bradley Terry Luce – is a probability model that predicts the outcome of the paired comparison as proposed by **Erlandson. F. Saraiva [4]**

Earlier various statistical approaches had proposed the most popular had been and they have divided into two major categories. I] the first ones are based on the prospective information contained in bookmaker's chances. II] Next approach is the probabilities of winning teams are paired and derived into a conclusion.

## VIII. ALGORITHM TECHNIQUES.

We need a considerable amount of data to store all the information about team and players so which algorithm would play an essential role? Big Data. Big data is a Collection of team data, match data, and player data. Logistic Regression also plays a critical role in Football winning prediction. In machine learning Back, propagation used for prediction, for example, **Purucker [5]** conducted initial studies on predicting results in NFL using the ANN model. His results from eight rounds were used consisting of yards gained, rushing yards gained, turnover margin, betting line odds. His findings were excellent he achieved 61% compared with 72% accuracy of experts.

## IX. LOGISTIC REGRESSION

How to use logistic regression? A classifier algorithm. How does this algorithm work? It measures the bonding between

the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function — specifically the continuous logistic distribution. In other words, logistic regression attempts to forecast a consequence (a Win or a Loss) given a set of data points (statistics) that likely influence that consequence as proposed by **Nilay Zaveri [6]**

In a football match, Sports prediction treats as a Classification problem with each class to win, draw or lose to be predicted as proposed by **D. Prasitio [7]**.

For example features in sports result, data can divide into different subsets and split the functions into match related and standing features. **Tax and Joustra [8]** considered how a hybrid model of betting odds and public data features compared with a feature set of betting odds alone. One can also investigate which classifier and feature selection algorithm together produces the best accuracy.

## X. ADVANTAGES

General win/loss predictions were projecting reliably the winners of the sporting events who made who made specific score lead. This advantage was noticeable regardless of whether reward occasions was relative (i.e., only the participant with the highest overall presentation received cash) or individual (participants received money for every correct prediction).

Predicting winners by experts reliably when making general bets when making specific bets. It shows that even in cases where greater knowledge may offer an advantage that focusing on learning can disrupt decision-making. Thus, lifelong baseball fans are more likely to pick the winning team who has never watched a game, for either person a quick prediction about the winner is more accurate than one that follows deep reflection.

## XI. CONCLUSION

The model we devised based on statistical analysis of past football games. We will be able to make reasonably accurate predictions. Although the accuracy of this model is pretty good, it's not guaranteed to be always right, and there is a lot of width for future work in this regard. We could bring in sentiment analysis, features such as individual player and team performance metrics, studying the trending hash-tags on Twitter on match day, the posts from fans on social media, etc. to further intensify the accuracy of the model.

## XII. REFERENCES

- [1] Sujit Nair, Varun Salian, Milind Dave, Piyush Kadve, PREDICTION OF RESULTS OF FOOTBALL MATCHES, International Journal of Research In Science & Engineering,2017.
- [2] Darwin Prasetio, Dra. Halili, Predicting football match results with logistic regression, 2016 International Conference on Advanced Informatics: Concepts, Theory, And Application (ICAICTA), 2016.
- [3] Norbert Danisik; Peter Lacko; Michal Farkas, Football Match Prediction Using Players Attributes, 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA),2018.

- [4] Erlandson F. Saraiva, Adriano K. Suzuki, Ciro A. O. Filh, Francisco Louzada Predicting football scores via Poisson regression model, *Commun. Stat. Appl. Methods* 2016; 23:297-319, 2016
- [5] M.C. Purucker Neural network quarterbacking *IEEE Potentials*, 15 (1996), pp. 9-15
- [6] Nilay Zaveri, Utkarsh Shah, Prediction of Football Match Score and Decision Making Process, *International Journal on Recent and Innovation Trends in Computing and Communication*.
- [7] D. Prasitio, D. Harlili, Predicting football match results with logistic regression, in: *Proceedings of the 2016 International Conference On Advanced Informatics: Concepts, Theory, And Application (ICAICTA)*, 16–19 Aug. 2016, Penang, Malaysia, 2016.
- [8] N. Tax, Y.P. Joustra Predicting the Dutch football competition using public data: A machine learning approach *Trans. Knowl. Data Eng.*, 10 (10) (2015), pp. 1-13