

An Application of Linear Regression & Artificial Neural Network Model in the NFL Result Prediction

Anyama, Oscar Uzoma
Department of Computer Science
University of Port Harcourt
Nigeria

Igiri, Chinwe Peace
Department of Computer Science
University of Port Harcourt
Nigeria

Abstract—Football results prediction in has gained popularity in recent years. Non hybrid approaches have shown complex and low prediction results. Data mining tools with insufficient features, however, have also yielded low predictions. In our research, machine learning has been used to develop a hybrid football match result predictive model for NFL. We constructed a more comprehensive system with improved prediction accuracy by using a hybridized approach. Our prediction system for football match results was implemented using a hybrid of artificial neural network (ANN) and linear regression (LR) techniques with Rapid Miner as a data mining tool. The technique yielded 90.32% prediction accuracy. With this output, it is observed that the prediction accuracy is higher than those of existing systems.

Keywords—ANN; Hybrid; Machine learning; Models; Prediction

I. INTRODUCTION

Sports prediction is gradually becoming a huge business venture, sports entertainment is not just about the competitors (teams, officials) of the game and the fans showing their support week in week out to their various sports outfit. Betting markets have gradually churned out huge gaming sentiments into a multi-billion dollar venture. You can barely watch games these days without being reminded that you can bet on the results and make reasonable financial gains. Therefore, predicting game results has become an area of interest for different sports organizations [5].

In this research however, the interest of using machine learning in Artificial Intelligence as an approach in the prediction of the outcome of games is thoroughly investigated. Machine Learning approach provides an advantage of having an unbiased/objective analysis with respect to games statistics using select techniques and methods [7]. In turn, this provides and ensures that games results are predicted in a much effective manner using appropriate models.

The overall benefits of developing such a system are:

- To build a system that can help bettors beat the bookies [6]
- To help managers with team strategies and decision making
- To contribute to knowledge and learning
- To study statistics obtained from games' data.

II. LITERATURE REVIEW

Today, various mathematical games prediction models exist, ranging from result prediction models to number of goals prediction models to injury prediction model and even to half time prediction models. These models and computer programs for games predictions have long been developed and still in development. Most of them employ stochastic methods to describe uncertainty.

One of the most recent challenges especially to researchers is the deployment of an efficient Hybrid Prediction System, which must be a system that takes into account the performances of players and ranking of teams. The system must have very high prediction accuracy.

Participation in games activity is perhaps important to all of us; therefore it is not surprising that there has been a substantial amount of research work done on prediction of games. Some of the related works are discussed.

A large number of literatures have been dedicated to the development of goal modeling, result modeling, ratings and rankings for games prediction. The methods proposed in these papers and articles can be evaluated by their ability to predict the outcomes of future games. Many papers have considered methods based on various forms of mathematical models where predictions and forecasting are made for games outcome [8]

a) Related Works

A Quantitative Stock Prediction System based on Financial News was done by [2]. In their work the discrete stock price prediction using a synthesis of linguistic, financial and statistical techniques to create the Arizona Financial Text System (AZFinText) was done. The major objective of the project was to provide predictions for stock market using statistical data gathered from financial news. The lines of research approach used were Mean Squared Error (MSE), visualization tools and Machine Learning Techniques. Prediction accuracy of 71.2% was obtained with a Simulated Trading return of 8.50%

[1] developed an Ocean Model, Analysis and Prediction System using the root mean square error for sea surface height anomaly. The approach introduced a daily forecast with a new four-cycle design was introduced where four independent forecast cycles in each time-lag. The system is composed of a real-time ocean observing system, a quality control system, the Ocean Forecast Australia Model (OFAM), the BLUElink Ocean Data Assimilation System (BODAS), an adaptive

initialisation scheme and air-sea fluxes from the Australian Community Climate and Earth System Simulator (ACCESS).

An artificial neural network which is a non-linear data modeling tool that can be used to find hidden patterns and relationships within data was developed by [4]. The findings by the author show good solution to the prediction problem. The major pitfalls include: high complexity, lowly parameterized poor training of datasets. The week 15 prediction rate was 75% using the season average and only 37.5% of the games using the three week average.

[3] examined nine college football ranking systems, including several used by the BCS and considered them in addition to an indicator of home field advantage and betting spreads as predictors in regression models predicting the outcomes (point spreads) of 1,582 games from 1998 to 2001. The approach was a very robust one with lots of parameters captured. The major pitfall is its linear dependencies. A prediction accuracy of 74.7% was obtained.

III. ANALYSIS OF EXISTING SYSTEM

The present system is the research work done by [4]. In [4], an artificial neural network which is a non-linear statistical model that can be used to find hidden patterns, approximate functions and find relationships within data using the concept of neurons in the human brain. The tools used are matlab and perl script.

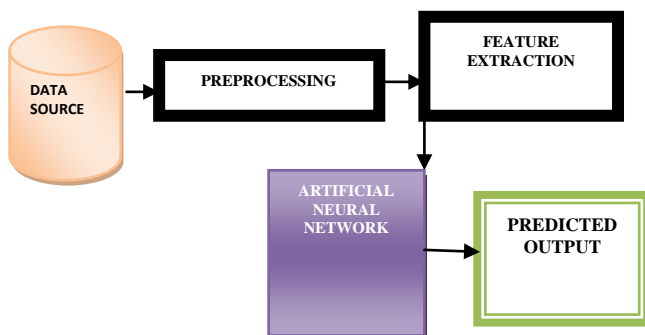


Fig 1: Joshua K., (2003) (Existing System)

In [4], the Neural Network Prediction of NFL Football Games was divided into the following stages:

- Data Collection
- Data Extraction and
- Data mining

Features used are:

- Total yardage differential
- Rushing yardage differential
- Time of possession differential (in seconds)
- Turnover differential and
- Home or away scores.

A Artificial Neural Network

The existing system develops a model using a Back Propagation approach which implies that more than one predictor variable is available.

The algorithm below represents the multi linear regression approach.

1) Algorithm (ANN)

Input

Attributes X_1, X_2, \dots, X_n

Main process algorithm

initialize network weights (often small random values)

do

 forEach training example

 prediction = neural-net-output(network, ex)

 actual = teacher-output(example)

 compute error (prediction - actual)

 compute all weights from hidden layer to output layer

 compute all weights from input layer to hidden layer

 update network weights // input layer not modified by error estimate

 until all ex classified correctly or another terminating criterion satisfied

 return the network

2) Output

$Y =$ Predicted Result used for rating

Pitfalls

- High complexity
- Low parameterized
- Data cannot be retrained

3) Prediction Accuracy

Predictions were made using both prediction sets and were tested for weeks 14 and 15 of the 2003 NFL season. In both cases, the season average prediction set was more effective in predicting the outcome of the games. For week 14, the season average prediction set generated 75% correct outcomes, whereas the three week average set correctly predicted 62.5% of the games. The week 15 prediction rate was 75% using the season average and only 37.5%

IV. PROPOSED HYBRIDIZED PREDICTION SYSTEM

In the proposed system, the use of machine learning was developed to out-perform the existing system. RapidMiner modeling tool is used to investigate the hybridized methods. The proposed model framework is a hybrid of Linear Regression Technique and Artificial Neural Network, which employs a supervised learning.

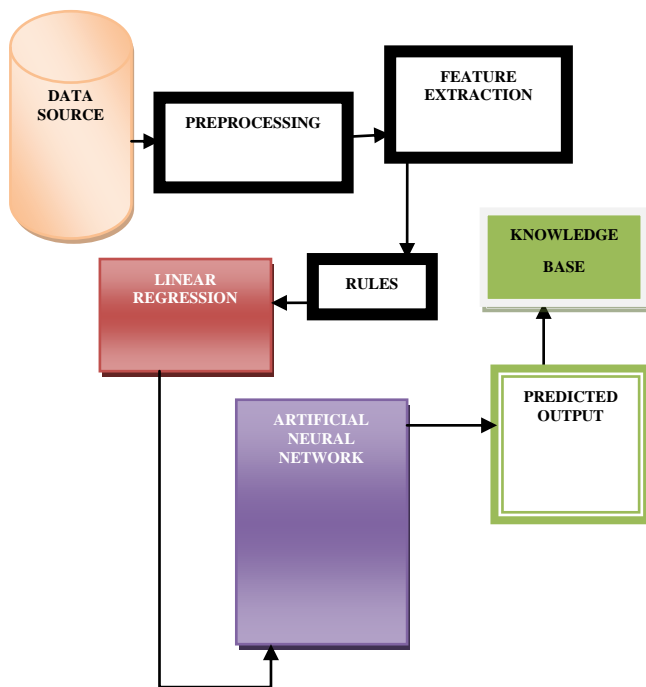


Fig. 2: proposed hybrid prediction system

The hybrid system could be used to perform non-linear mapping based on a variety of relevant statistics, hence for this project the dataset to be considered will be available from the standard games portal. The importance and weight of each statistic must be determined prior to making a prediction using linear regression as this will provide appropriate statistical weights.

The ANN will provide classification of already weighed features. Results from trained prediction set will be applied to unseen games. The resulting hybrid model provides an optimized model which in turn will yield good results with good prediction accuracy with big implication on the various dependents of the results

A) Analysis of proposed system

In designing the hybrid system, the following steps will be employed:

Step 1: Problem definition

Here the understanding will be broken into the project requirements.

Step 2: Data collection and pre-processing

The dataset location on the server will be downloaded automatically from NFL games server using Google's URL crawling tools.

Step 3: Modeling

This phase is the core of the hybrid system and will be divided into two sub steps:

Build

In this phase, the linear regression technique and ANN techniques will be developed using the features sets.

Execute

The resulting value of range -1 to +1 from the Linear Regression Technique will provide attributes that affects or contributes to the prediction results. The results will be used as input for the ANN model.

B. Linear Regression (Attribute Weighing)

The relationship between the dependent and independent variables is seen. Multiple regression models contain one measurement variable in multiple forms. The response variable is influenced by more than one predictor label. Unlike linear regression, where the response is a straight line, the response may be a curvilinear or multi-dimensional. Multiple implies more than one predictor class while linear means linear in the regression coefficients being additive.

1) Algorithm (Linear Regression)

Input

Attributes X_1, X_2, \dots, X_n

Linear Regression process

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \sigma$$

β_0 = intercept

$\beta_1, \beta_2, \dots, \beta_n$ = regression coefficients

σ = σ_{res} = residual standard deviation

2) Output

Y = Dependent variable representing resulting weights of the attributes.

C. Artificial Neural Network (ANN)

This technique learns a model by means of a feed-forward neural network trained by a back propagation algorithm (multi-layer perceptron). This operator cannot handle polynomial attributes. The results from will the linear regression will be converted using a nominal to numeric operator

1) Description

An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation (the central connectionist principle is that mental phenomena can be described by interconnected networks of simple and often uniform units) [9].

A feed-forward neural network is an artificial neural network where connections between the units do not form a directed cycle.

Back propagation algorithm is a supervised learning method which can be divided into two phases: propagation and weight update. The two phases are repeated until the performance of the network is good enough. In back propagation algorithms, the output values are compared with the correct answer to compute the value of some predefined error-function.

A multilayer perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of

nodes in a directed graph, with each layer fully connected to the next one.

2) Algorithm (ANN)

Input

Attributes X1, X2... Xn

Main process algorithm

initialize network weights (often small random values)

do

forEach training example

prediction = neural-net-output(network, ex)

actual = teacher-output(example)

compute error (prediction - actual)

compute all weights from hidden layer to output

layer

compute all weights from input layer to hidden layer

update network weights // input layer not modified by error estimate

until all ex classified correctly or another terminating criterion satisfied

return the network

3) Output

Y= Predicted Result used for rating

C) Unique features

A point of emphasis in this study is to expand the dataset beyond ordinary NFL statistics to include a couple of unique features and to explore their impact.

The following unique features will be used:

- i Players Performance index
- ii Bookmakers betting spread
- iii Moving Averages

Row No.	Winner	confidence	confidence	prediction	YdsHme	HTOw	H1sID	HAttack fac	HRating
1	Away	0	1	Away	370	2	26	1.006	3.944
2	Away	0	1	Away	287	3	18	0.922	-38.844
3	Home	1	0	Home	422	0	25	0.848	34.067
4	Away	0	1	Away	240	2	14	0.793	-17.156
5	Away	0	1	Away	297	1	17	0.882	19.933
6	Home	0.999	0.001	Home	390	1	18	0.952	-7.451
7	Away	0	1	Away	355	3	24	1.120	15.356
8	Away	0	1	Away	192	2	10	1.167	-42.689
9	Home	1	0	Home	277	2	18	0.982	34
10	Away	0	1	Away	358	4	19	0.783	-35.600
11	Home	1	0	Home	426	1	24	1.365	19.327
12	Home	1	0	Home	514	1	28	0.952	-8.244
13	Home	1	0	Home	344	3	24	0.963	29.756
14	Home	1	0	Home	379	0	20	0.952	16.222
15	Away	0	1	Away	296	3	18	0.850	-0.812
16	Home	1	0	Home	379	0	23	1.058	22.083
17	Away	0	1	Away	307	1	20	0.973	-16.729

Fig. 3.2: A screen shot of the result from Rapid Miner

ImprovedNeuralNet

Hidden 1

=====

Node 1 (Sigmoid)

- Home Team = Denver Broncos: 1.027
- Home Team = New York Jets: 0.777
- Home Team = Buffalo Bills: 0.742
- Home Team = St. Louis Rams: 0.583
- Home Team = Carolina Panthers: 0.292
- Home Team = Detroit Lions: 0.917
- Home Team = Chicago Bears: 1.645
- Home Team = Jacksonville Jaguars: 0.985
- Home Team = Dallas Cowboys: 1.023
- Home Team = Pittsburgh Steelers: -0.879
- Home Team = Cleveland Browns: 1.849
- Home Team = San Francisco 49ers: -0.031
- Home Team = Indianapolis Colts: -0.109
- Home Team = New Orleans Saints: -1.277
- Home Team = Washington Redskins: 0.230
- Home Team = San Diego Chargers: 0.081
- Home Team = New England Patriots: -0.511
- Home Team = Kansas City Chiefs: 0.515
- Home Team = Philadelphia Eagles: 0.598
- Home Team = Seattle Seahawks: -0.001
- Home Team = Baltimore Ravens: 0.043
- Home Team = Oakland Raiders: 0.110
- Home Team = Atlanta Falcons: -0.434
- Home Team = Arizona Cardinals: 0.314
- Home Team = Houston Texans: 1.067
- Home Team = Green Bay Packers: 2.223
- Home Team = Tampa Bay Buccaneers: -0.814
- Home Team = New York Giants: -0.065
- Home Team = Cincinnati Bengals: -0.940
- Home Team = Tennessee Titans: 1.591
- Home Team = Minnesota Vikings: 0.301
- Home Team = Miami Dolphins: 0.369
- Away Team = Baltimore Ravens: 0.361
- Away Team = Tampa Bay Buccaneers: 0.965
- Away Team = New England Patriots: -0.736

Away Team = Arizona Cardinals: 0.931
 Away Team = Seattle Seahawks: 0.279
 Away Team = Minnesota Vikings: 0.063
 Away Team = Cincinnati Bengals: -0.881
 Away Team = Kansas City Chiefs: 0.558
 Away Team = New York Giants: 1.997
 Away Team = Tennessee Titans: 1.293
 Away Team = Miami Dolphins: 0.531
 Away Team = Green Bay Packers: 0.800
 Away Team = Oakland Raiders: -0.468
 Away Team = Atlanta Falcons: 0.395
 Away Team = Philadelphia Eagles: 1.715
 Away Team = Houston Texans: -0.710
 Away Team = New York Jets: 0.211
 Away Team = Dallas Cowboys: -1.318
 Away Team = San Diego Chargers: 0.462
 Away Team = San Francisco 49ers: 0.428
 Away Team = Cleveland Browns: 0.671
 Away Team = Jacksonville Jaguars: 1.623
 Away Team = St. Louis Rams: -0.178
 Away Team = Detroit Lions: 0.160
 Away Team = Washington Redskins: -0.234
 Away Team = New Orleans Saints: 0.174
 Away Team = Denver Broncos: 0.282
 Away Team = Carolina Panthers: 0.921
 Away Team = Pittsburgh Steelers: 0.230
 Away Team = Indianapolis Colts: 0.391
 Away Team = Chicago Bears: 0.379
 Away Team = Buffalo Bills: 0.799

V. RESULT DISCUSSION

Looking at the thirty one games that were predicted, the three games that were incorrectly predicted by the hybrid model over the progression of weeks 16 and 17, one game could be considered “upset.” One of the remaining games was games was “too close to call.” The prediction set provided a prediction accuracy of 90.32 using a training rate of 200. This indicates that improved prediction data, using the best fit attributes would lead to more accurate prediction.

VI. CONCLUSION

This study has shown the development of a hybrid model using Linear Regression and Artificial Neural Network techniques in the prediction of the results of NFL games with an improved accuracy (90.32%) which is higher than that of the existing system (74%).

A) Research Highlights

The research highlights of this paper are:

- This paper proposes a different approach for NFL results prediction.
- The approach uses hybridized methods for implementation.
- The hybridized techniques used are Linear Regression and Neural Network.
- The results show improved prediction accuracy of 90.32%

REFERENCES

1. Brassington G.B., Freeman J., Huang X., Pugh T., Oke P.R., Sandery P.A., Taylor A., Andreu-Burillo I., Schiller A., Griffin D.A., Fiedler R., Mansbridge J., Beggs H. & Spillman C.M. Ocean Model, Analysis and Prediction System: version 2 CAWCR Technical Report. The Centre for Australian Weather and Climate Research. No. 052. 2012, Unpublished
2. Robert P.S. and Hsinchun C. A Quantitative Stock Prediction System based on Financial News, in press
3. Fair R. & Oster J. College Football Rankings and Market Efficiency, Journal of Sports Economics, vol. 8, pp 3-18, 2007.
4. Joshua K. Neural Network Prediction of NFL Football Games. ECE539, Unpublished.
5. Haghghat, M., Rastegari, H. & Nourafza, N. A review of data mining techniques for result prediction in sports. ACSIJ Advances in Computer Science: An International Journal, 2(5), 7-12, 2013.
6. Buursma, D. Predicting sports event from past result: Towards effective betting on football matches. Preceding 14th Twente Student Conference on IT. University of Twente, Faculty Electrical Engineering, Mathematics and Computer Science, Netherlands. Conference paper 7226, 2011.
7. Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R. & Honrao, V. (2013). Predicting students' performance using ID3 and C4.5 classification algorithms. International Journal of Data Mining & Knowledge Management Process, 3(5), 39-52, 2013.
8. Nivard, W. & Mei, R. D. Soccer analytics: Predicting the of soccer matches. (Master thesis: UV University of Amsterdam), 2012.
9. Sengupta, S. Neural networks and application. Lecture. Department of Electrical and Electronics Engineering. Indian Institute of Technology, Kharagpur, Youtube, 2003.