Prediction of Results of a so Soccer Match at the World Cup Using Backpropagation Artificial Neural Network

Falano Rajib, Esmeralda C.Djamal, Fatan Kasyidi
Department of Informatics, Jenderal Achmad Yani University, West Java, Indonesia.
falano.rajib@student.unjani.ac.id, esmeralda.contessa@lecture.unjani.ac.id,
fatan.kasyidi@lecture.unjani.ac.id

Abdul Talib Bon
Department of Production and Operations, University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract
Prediction of the results of a soccer match is carried out using several variables such as historical data. Even though the score in a soccer match cannot be predicted, however, match statistics and previous match patterns can be used to predict the outcome of the next match. So that it can help the team to evaluate and prepare strategies before competing against the opposing team. In this study, a model was made to predict match results based on historical data from FIFA world cup match results, using one of the Artificial Neural Network (ANN) algorithms, Backpropagation. The results showed the prediction of soccer results in the world cup using Backpropagation Artificial Neural Networks produces an accuracy of 80.78%.

Keywords
Soccer, World Cup, Predictions, Artificial Neural Network, Backpropagation.

1. Introduction
Soccer is a sport that is very popular sport in the world. In 2017 more than 700 million people respond well to soccer throughout America, Europe, the Middle East, and Asia. There are about 108 professional soccer leagues located in 71 countries around the world. With so many soccer fans in the world that there is a lot of data used by researchers, statisticians, and data analysis to analyze data and make some conclusions (Nielsen Sports, 2018).

Predictions can be made in various fields such as predictions on the stock market (Bharne and Prabhune, 2019), sales (Omar, Hoang and Liu, 2016), foreign tourists (Vihikan, Putra and Dharmaadi, 2017). Making predictions can help planning for the future. One prediction in the field of sports is that soccer predictions can be used for team management, managers and other coaching staff members need to know about their own teams and opposing teams in detail. The soccer prediction is done by using some data depending on variables such as historical data of football match results which can be seen from match statistics and match patterns that can be used to predict match results so that each coach can identify the strengths and weaknesses of the team and prepare special strategies and preparations for players to face your opponent (Zaveri, 2018).

Previous research explained that to predict match results, there were seven match-based criteria chosen as prediction parameters. These criteria consist of the team, the condition of the team in the last few weeks, the condition of the team in the league, the quality of the opposing team in the previous match, league match, match results of the week. The parameters are considered as input from the Neural Network model and the results obtained from the league and last week are the

output from the Neural Network models (Arabzad, Araghi and Soheil, 2014). Another study Bayesian Network is used to predict a match in the Premier League, every game is considered separately because each game has a value of different factors and vary based on the condition of the game concerned. Therefore, the trial predictions were repeated for all 380 matches in three separate seasons. Then the results are compared between one season and another.

The results of the study used a data set that has several attributes including:

1. Home Team Shots
2. Away Team Shots
3. Home Team Shots on Target
4. Away Team Shots on Target
5. Home Team Corners
6. Away Team Corners
7. Home Team Fouls Committed
8. Away Team Fouls Committer

During the first season the accuracy of the prediction was 75.26% (286/380), the second season showed an accurate prediction of 79.47% (302/380) and the third season showed accuracy prediction of 70.53% (268/380)(Razali *et al.*, 2017).

In addition, other studies that predict soccer player transfer exchanges by utilizing social media Twitter. Football data analysis on Twitter is done to map popularity and predictions using Support Vector Engine (SVM) which results in an accuracy of 65%. This accuracy yields quite reasonable results given the fact that this study was based entirely on user tweets and transfer rumor that were not verified by experts. But research to produce a high enough accuracy depends on the amount of data used if the amount of data used is greater than that, prediction of transfer can be done more effectively (Murali, Shrivastava and Krishnan, 2018).

In this study, makes a computational model was made to predict the results of world cup matches based on home team goals, away team goals, half-time home goals, half-time away goals and attendance using the Backpropagation method. Prediction is done by statistics on the results of previous matches.

2. Research Methodology

The process of predicting the results of soccer matches in the world cup can be seen in Figure 1. The stages of the research began with the acquisition of FIFA World Cup match results statistics which consisted of soccer match statistics which will be used as training data and test data, then followed by pre-processing consisting of feature extraction and normalization. Furthermore, machine learning is carried out using the Backpropagation algorithm to predict soccer matches at the world cup. In this prediction system have three classes: win, draw, and lose.

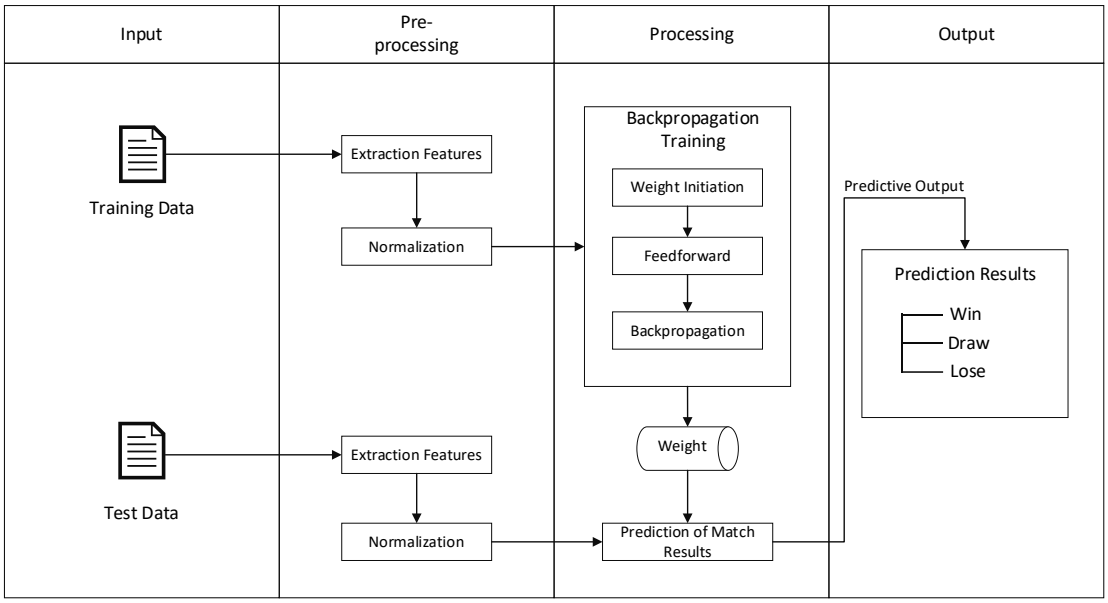

Figure 1. Football Match World Cup Results Prediction

2.1 Data Acquisition

Data obtained is historical data from FIFA World Cup matches. The data contains statistics about the results of each match between the teams in the participating world cup from various countries. From the report of each match, there is 11 statistical item: home team goal, away team goals, away, win conditions, attendance, half—time home goals, half-time away goals, referee, assistant 1, assistant 2, Round ID, and Match ID, which represents an index of the team's ability to win a match can be seen in Table 1. From these statistical data, apply an MLP neural network to predict the winning from two teams at the next stage games by means of their statistic data from previous games.

Table 1. Sample Match Result Statistics

Year	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	Attendance	Half-time Home Goals	Half-time Away Goals
2014	Brazil	3	1	Croatia	62103	1	1
2014	Mexico	1	0	Cameroon	39216	0	0
2014	Spain	1	5	Netherlands	48173	1	1
2014	Chile	3	1	Australia	40275	2	1
2014	Colombia	3	0	Greece	57174	1	0
2014	Uruguay	1	3	Costa Rica	58679	1	0
2014	England	1	2	Italy	39800	1	1
2014	Ivory Coast	2	1	Japan	40267	0	1
2014	Switzerland	2	1	Ecuador	68351	0	1
2014	France	3	0	Honduras	43012	1	0
2014	Argentina	2	1	Bosnia and Herzegovina	74738	1	0
2014	Germany	4	0	Portugal	51081	3	0

2014	IR Iran	0	0	Nigeria	39081	0	0
2014	Ghana	1	2	USA	39760	0	1

2.2 Pre-processing

This pre-processing phase is carried out to process data before it is used for training and prediction preparation. At the pre-processing stage, there are two steps carried out at the pre-processing stage, namely feature extraction and normalization.

2.2.1 Extraction Features

From the parameters in the statistics of the results of the world cup matches in the dataset. This study selects five items to represent each team's match statistics, namely HTG = Home Team Goals, ATG = Away Team Goals, A = Attendance, HTHG = Half Time Home Goals, and HTAG = Half Time Away Goals (Huang and Chen, 2011). From the selection feature, apply an MLP neural network as an input layer can be seen in Figure 2.

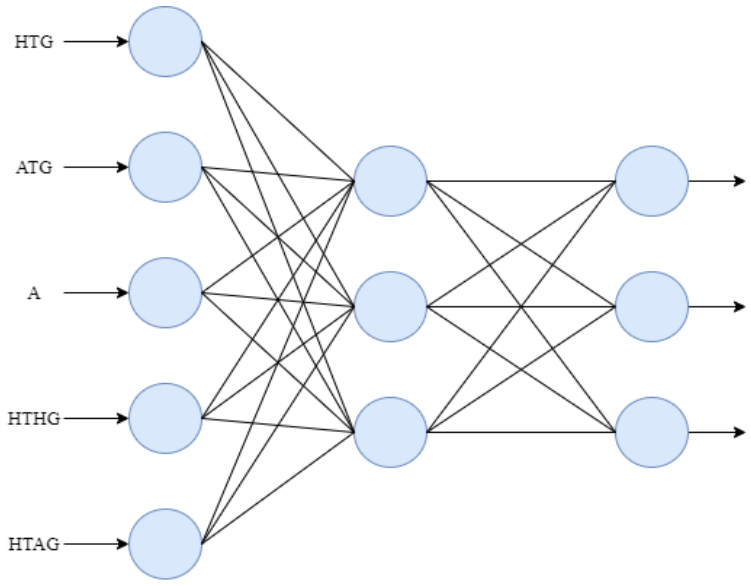

Figure 2. MLP network used in predicting

2.2.2 Normalization

The normalization stage is done to convert the data to a new range between 0 to 1. Normalization can be done using the min-max normalization method (Vamsidhar *et al.*, 2010). The selection pre-processing greatly affects accuracy based on data compatibility with the normalization method performed. In preprocessing stage is normalized using the Min-Max normalization method which can achieve convergence in a short time with a slight epoch. The Min-Max normalization process can be performed using equations 1.

$$v' = \frac{v - min_a}{max_a - min_a} \quad (1)$$

Where:

v' : Value of normalized data
v : Value data not yet normalized
min_a : Minimum value of data from attribute to-a
max_a : Maximum value of data from the to-a attribute

2.3 Backpropagation

Artificial Neural Network (ANN) is one of the artificial representations of the human brain that always tries to simulate the learning process in the human brain(Wanto *et al.*, 2017). Artificial neural networks are information processing systems designed to mimic the structure of the human brain and its functions. this can be the basis of making powerful and widely used machine learning algorithms (Zhou *et al.*, 2018). Previous research of backpropagation related to the prediction was used to forecasting analysis of organic red rice's demand using artificial neural networks using the 3-20-1 architecture with MSE value of 0.002 where forecasting results indicate that the highest demand for organic red rice is in august amounting to 22.259 kg and the lowest in April is 12.809 kg (Holik, Bachtiar and Setiadevi, 2019). As for predicting rainfall with the best tests using Alpha 0.7, the epoch 10000 and MSE values are 0.022 generating an accuracy of 80% in forecasting rainfall data both in the training and testing stage (Lesnussa *et al.*, 2018). Analysis of the backpropagation algorithm in predicting the world's largest number of Internet users reaches 92% accuracy by using the 3-50-1 architecture model and the model selection and parameters of the network architecture used (Setti and Wanto, 2019).

In the Backpropagation, there is a Multilayer Perceptron architecture which is used to alter the weight associated with the neurons found in the hidden layer and the binary sigmoid function in the layer output. The determination of the number of hidden layers to be used can use equation 2 and the binary sigmoid function can be seen in Equation 3.

$$p = \sqrt{m\ x\ n} \tag{2}$$

Where:

m : Number of input neurons
n : Number of output neurons

$$f'(x) = f(x)(1 - f(x)) \tag{3}$$

Where:

e : Exponent value
x : Activated data Input

The training with the Backpropagation method consists of three phases. The first phase is feed-forward i.e., and any neuron input will transmit the inserted signal on the hidden layer. Each neuron in the hidden layer is multiplied by a weight and added to the bias shown in equation 4.

$$z_{netj} = v_{j0} + \sum_{i=1}^{n} x_i\ v_{kj} \tag{4}$$

Where:

i : Index of Layer input units
j : Index of hidden layer units
v_{j0} : Refractive weight from layer inputs to hidden layers
n : Number of neurons in layer inputs
x_i : Input value Data to-I
v_{kj} : Weight from input layer to hidden layer

Each of the output neurons is multiplied by weights and aggregated with bias using equation 5.

$$y_{netk} = w_{k0} + \sum_{j=1}^{p} z_j\ y_{kj} \tag{5}$$

Where:

j : Neuron index hidden layers
p : Number of neurons on hidden layers
k : Neuron Index output
w_{k0} : Refractive weight from hidden layer to output layer
z_j : The activation value of the hidden layer
y_{kj} : From the hidden layer to the output layer

The second phase is Backpropagation; each neuron output receives a target pattern that has been achieved according to the input pattern when learning then calculates the change in bias using equation 6.

$$\Delta wk0 = \alpha \text{ x } \delta k \quad (6)$$

Where:

α : Value learning rate

δk : Component error on layer output

The third phase is weight change after all factors δ are calculated. This stage serves to lower the errors that occur. The error calculations are performed on each hidden layer, the bias, renewal, and the weight of each neuron output resulting in a new weight and bias. Calculations are done using Equation 7. The calculations on each hidden neuron start from the 1st neuron up to the p neuron using equation 8.

$$W_{kj}(\text{new}) = W_{kj}(\text{long}) + \triangle W_{kj} \quad (7)$$

$$V_{ji}(\text{new}) = V_{ji}(\text{long}) + \triangle V_{ji} \quad (8)$$

Where:

W_{kj} *(long)* : Neuron weight hidden to the old output
$\triangle W_{kj}$: Weight correction W
V_{ji} *(long)* : Input neuron weight to the old hidden
$\triangle V_{ji}$: Weight correction V

The training is done until the Mean Square Error (MSE) has reached a value smaller than the value of the error applied.

3. Result and Discussion

This study uses Backpropagation to predict the results of soccer matches in the world cup. The results of this study conducted several experiments on training and testing by changing the maximum iteration value, learning rate and division of the amount of training data and test data.

3.1 Maximum Iteration Testing

Maximum iteration testing is used to determine the maximum number of iterations. Testing is done by changing the iteration value. Each iteration test is used by different initial weights to find the maximum iteration. The result from maximum iteration testing can be seen in Table 2.

Table 2. Result Testing of Maximum Iteration

No	Maximum Iteration	Accuracy (%)	MSE
1.	100	63.26%	0.543
2.	200	70.23%	0.438
3.	300	69.43%	0.479

This maximum iteration test is used to determine the maximum number of iterations based on the MSE values obtained. Based on Table 2, the use of iteration 200 shows a smaller MSE value and greater accuracy so that other experiments are carried out using iteration 200.

3.2 Learning Rate Testing

The next experiment is done by changing the Learning Rate (alpha) value used for training and testing. This is done to get the optimal learning rate. However, each level of learning is carried out with different initial weights.

Table 3. Result Accuracy Learning Rate

No	Learning Rate	Training Data		Testing Data	
		Loss	Accuracy (%)	Loss	Accuracy (%)
1.	0.001	0.284	80.78%	0.484	70.38%
2.	0.005	0.256	79.32%	0.456	69.63%
3.	0.010	0.229	79.52%	0.329	69.84%
4.	0.050	0.186	80.64%	0.286	70.42%
5.	0.100	0.014	80.36%	0.114	70.68%

Testing the value of the learning rate is used to determine the best level of learning that can be used. Based on Table 3, the learning rate value of 0.001 shows a greater accuracy value of 70.38% accuracy, but the resulting loss value is 0.284. A good level of learning is shown in learning rate 0.100, where the resulting loss is smaller than other learning rates.

3.3 Amount of Training Data dan Test Data Testing

After testing the maximum iteration and learning rate test, the result is the number of iterations and the best learning rate that will be used as a parameter for testing the amount of training data.

Table 4. Result of Training Data and Test Data

No	Training Data	Testing Data	MSE
1.	245	27	0.462
2.	218	54	0.278
3.	191	81	0.354

From the results of testing the amount of training data and test data using the amount of training data, 218 and test data 27 shows the smallest MSE 0.278.

4. Conclusion

This research has produced a prediction computation model of the results of soccer matches at the world cup using Backpropagation. The number of hidden neurons used in this study affects the resulting accuracy. The results of tests conducted using learning rates.

To make predictions using backpropagation artificial neural networks requires training data to be used as a training source and then processed at the testing stage. Based on testing that has been done the best iteration produced is at number 200 with MSE 0.438, and the best learning rate test is at number 0.001 with the resulting accuracy of 80.78%, while for testing training data and test data yielding MSE of 0.278 with training data 218 and test data 54.

References

Arabzad, S. M., Araghi, M. E. T. and Soheil, S. (2014) 'Applied Research on Industrial Engineering Football Match Results Prediction Using Artificial Neural Networks ; The', (October).

Bharne, P. K. and Prabhune, S. S. (2019) 'Stock market prediction using artificial neural networks', *2019 International*

Conference on Intelligent Computing and Control Systems, ICCS 2019, 5(1), pp. 64–68.
Holik, A., Bachtiar, R. R. and Setiadevi, S. (2019) 'Forecasting Analysis of Organic Red Rice's Demand Using Artificial Neural Networks', *Jurnal Manajemen dan Agribisnis*, 16(2), pp. 123–132.
Huang, K.-Y. and Chen, K.-J. (2011) 'Multilayer Perceptron for Prediction of 2006 World Cup Football Game', *Advances in Artificial Neural Systems*, 2011(2), pp. 1–8.
Lesnussa, Y. A. *et al.* (2018) 'Application of Backpropagation Neural Networks in Predicting Rainfall Data in Ambon City', *International Journal of Artificial Intelligence Research*, 2(2).
Murali, R., Shrivastava, S. and Krishnan, S. (2018) 'Analysis of Football Data on Twitter for Popularity Mapping and Transfer Predictions', *International Journal of Engineering & Technology*, 7(3.12), p. 452.
Nielsen Sports (2018) 'World Football 2018 | Nielsen Sports', p. 35. Available at: https://nielsensports.com/reports/world-football-2018/.
Omar, H., Hoang, V. H. and Liu, D. R. (2016) 'A Hybrid Neural Network Model for Sales Forecasting Based on ARIfile:///C:/Users/USER/Downloads/Documents/5993-17071-1-PB.pdfMA and Search Popularity of Article Titles', *Computational Intelligence and Neuroscience*, 2016.
Razali, N. *et al.* (2017) 'Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL)', *IOP Conference Series: Materials Science and Engineering*, 226(1).
Setti, S. and Wanto, A. (2019) 'Analysis of Backpropagation Algorithm in Predicting the Most Number of Internet Users in the World', *Jurnal Online Informatika*, 3(2), p. 110.
Vamsidhar, E. *et al.* (2010) 'Prediction of rainfall using backpropagation neural network model', *International Journal on Computer Science and Engineering*, 2(4), pp. 1119–1121.
Vihikan, W. O., Putra, I. K. G. D. and Dharmaadi, I. P. A. (2017) 'Foreign tourist arrivals forecasting using Recurrent Neural Network Backpropagation Through Time', *Telkomnika (Telecommunication Computing Electronics and Control)*, 15(3), pp. 1257–1264.
Wanto, A. *et al.* (2017) 'Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process', *Journal of Physics: Conference Series*, 930(1).
Zaveri, N. (2018) 'Prediction of Football Match Score and Decision Making Process', *International Journal on Recent and Innovation Trends in Computing and Communication*, 6(2), pp. 162–165.
Zhou, G. *et al.* (2018) 'Artificial neural networks and the mass appraisal of real estate', *International Journal of Online Engineering*, 14(3), pp. 180–187.

Biographies

Falano Rajib is currently studying in Informatics Department at Universitas Jenderal Achmad Yani.

Esmeralda Contessa Djamal received a Bachelor's degree in Engineering Physics from Institut Teknologi Bandung in 1994, a Master's degree in Instrument and Control from Institut Teknologi Bandung in 1998. She finished a Ph.D. degree in Engineering Physics from Institut Teknologi Bandung in 2005. Her doctoral thesis focused on signal processing and pattern recognition. Since 1998 until now, she has published about 40 papers in an international journal or proceeding primarily in pattern recognition, machine learning, and signal processing. Currently, she is a lecturer of the Informatics Department, Universitas Jenderal Achmad Yani.

Fatan Kasyidi is a lecturer of Informatics in the Faculty of Science and Informatics at the University of Jenderal Achmad Yani, Indonesia since 2018. He studied Informatics which is focus on Intelligence System in Institut Teknologi Bandung for which he was awarded the Master's Degree in the year 2017. He's bachelor degree in Informatics which was obtained from the University of Jenderal Achmad Yani. He is active in research about Artificial Intelligence, Speech Recognition and Affective Computing.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He's bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.