International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 01, 2020
ISSN: 1475-7192

SOCCER OPINION ANALYSIS OF TWITTER MEDIA USING BACK PROPAGATION METHOD

Sunjana[1], Sri Lestari[2], Hari Supriyadi[3]

ABSTRACT---*Twitter is a social networking tool that allows users to send and read text-based messages, commonly referred to as tweets. Information and news that is done by general users right now is no longer exclusive, it can only be done by big news publishers, but it can be done by everyone. Twitter has many opinions, not only positive or neutral opinions but also negative ones. In this research, each user's tweet will be categorized into positive and negative sentiment by looking at the contents of the tweet measured using TF-IDF weighting and Backpropagation Artificial Neural Network (ANN) classification method. Before using with TF-IDF, the contents of the tweet are done in the process of preprocessing data using tokenisais, cleansing, filtering, and Steming. To facilitate the work process on the system that was created, the data used is tweet in Indonesian. This research is a sentiment towards Twitter data. After doing research on Twitter data, it can be seen the results of trials of several scenarios with the amount of data 1000 and 1500 Tweet data from the classification of Backpropagation Neural Networks with the best 1000 data scenarios in table 4.12 with an accuracy of 50.58%, a precision level of 100.00 %, Recall 50.44%, and F-measure 67.06% so when changes in learnigrate or testing parameters inputted by the user affect changes in the accuracy of the system.*

Keywords---*Twitter, tweet, Backpropagation Artificial Neural Network (ANN), and Sentiment Analysis*

I. INRTODUCTION

At this time the internet is getting more advanced, one of which is social media which has developed very rapidly. Social media needs information that can attract public attention. Because social media does present information for the public interest. Nowadays microblogging is becoming popular as a communication tool between internet users.

As one of the many social media users, Twitter, is widely used to express opinions about an event. In general opinions or messages conveyed by each user can be positive or negative.

One case in the media twitter, which received a lot of responses is the case of football. As we have understood, that soccer is a sport that is very popular with the world community, as well as the people of Indonesia. So that every time there is a problem in the management of football, of course it will be the talk of many people. Many of the soccer lovers community who express their opinions through social media, one of which uses Twitter communication media.

The large number of people who have Twitter accounts, examples of cases in Indonesia, and are actively commenting on each incident, so research on opinion analysis or sentiment analysis is a very interesting discussion. As the

[1]*Computer Science, Faculty of Engineering, Widyatama University*
Jln. Cikutra 20124 A, Bandung 40125, INDONESIA
sunjana@widyatama.ac.id

author has done. The research that the author has done is analyzing sentiments or opinions about the problems that occur in Indonesian football. The data used is opinion data on Twitter. The purpose of this study is to look for classifications of public opinion or sentiment. In this study only two classifications were sought, namely positive sentiment and negative sentiment. The method used to divide opinion into different classes is backward propagation method which is a variant of artificial neural networks. The Twitter data used is 10,000 tweets, and the results will be seen using the backward propagation method, which percentage gives a positive sentiment and what percentage gives a negative sentiment.

II. LITERATURE REVIEW

Text Mining

According to Feldman and Sanger (Feldman and Sanger, 2007), Text mining can be defined as the process of extracting information intensively to find useful value stored in documents that allows users to interact with the collection of documents from time to time using various analyzes. In a manner consistent with data mining, text mining seeks to extract useful information and data sources from the identification and expansion of patterns

Sentiment Analysis

Sentiment analysis is a field of study that seeks to understand responses, evaluation results, assessment results, attitudes and emotions of people towards an entity such as products, services, organizations, individuals, problems, events, topics and so forth. This represents a large problem space.

Coarse-grained Sentiment Analysis

In this type of Sentiment Analysis, the Sentiment Analysis conducted is at the document level. Broadly speaking, the main focus of this type of Sentiment Analysis is to consider the entire contents of the document as a positive sentiment or negative sentiment (Fink Clayton, 2011).

Fined-grained Sentiment Analysis

Fined-grained Sentiment Analysis is Sentiment Analysis at the sentence level. The main focus of fined-greined sentiment analysis is to determine the sentiment at each time in a document, where the likelihood that occurs is there is sentiment at different sentence levels in a document (Fink Clayton, 2011).

Data Labeling

The dataset in the form of text obtained on social media such as Twitter, Facebook Instagram, and other social media is data that stores certain sentiments in each opinion sentence uploaded. Text data that can be obtained from one of the Indonesian countries, the text data taken is labeled in Indonesian. Retrieval of data in an appropriate language makes it easy to label the sentences (yuan & grace 2016)

Preprocessing Analysis

A good data structure can facilitate the process of computerization automatically. In Text Mining, the information to be extracted contains information that has an arbitrary structure. Therefore, the process of converting forms into structured data is needed according to their needs for processes in data mining, which will usually be numeric values. This process is often called Text Preprocessing (Ronen Feldman, 2007).

Classification Algorithm

Classification is the process of finding a set of models or functions that describe and differentiate classes of data from the purpose of the model. Such can be used to predict the class of an object whose class is unknown (Han and Kamber, 2012).

TF (Term Frequency)

Term frequency is one method to calculate the weight of each that is proportional to the number of occurrences of the term in the text (Mark Hall & Lloyd Smith, 1999).

DF (Document Frequency)

The number of documents that contain certain terms is called the document frequency (DF). Each term that appears will increase the number of frequency documents. Based on the number of DF values, the term will be chosen. But if the DF value of the term is below a predetermined threshold, the term will be discarded..

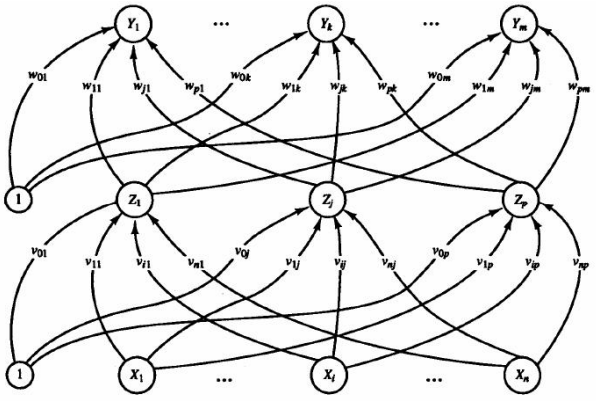
Neural Network

Artificial neural network is an information processing algorithm that mimics how the human brain thinks. The main component of this information processor is a large number of processing components that form an interconnected network called neurons. These neurons work together to find solutions to problems that must be solved.

Neural Network (backpropagation)

Neural Network (ANN) Backpropagation (BP) is a multi-layer ANN. His discovery overcomes the weakness of ANN with a single layer that resulted in the development of ANN had stalled around 1970. Backpropagation algorithm is a generalization of the delta (Widrow-Hoff) rule, namely applying the gradient descent method to minimize the error of the total square of the output calculated by the network. Many applications can be solved by backpropagation.

Backpropagation Architecture



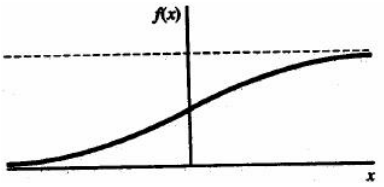
2.1 Figure of Backpropagation Architecture

BP network with one hidden layer (unit Z (Z1 ... Zp)) appears in the picture above. Output units (Y units (Y1 ..... Ym)) and hidden units have a bias. The biased weight of the output unit Y is expressed in w0k, the biased weight in the hidden

unit Zj is expressed in voj. vij is the line weight from the Xi unit to the hidden layer unit Zj. wjk is the line weight from Zj to the output unit Yk

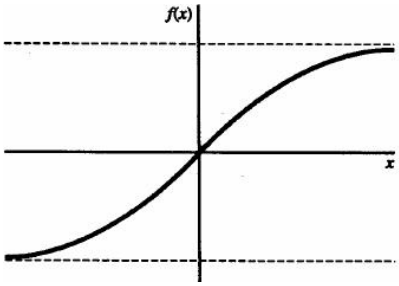
Activity function

In backpropagation, the activation function used must meet several conditions, namely: continuous, differentiable, and monotonically not decreased. One of the appropriate functions is the binary Sigmoid function with intervals (0,1).



2.2. Figure of Binary Sigmoid function

$f(x) = 1/(1+e^{-x})$ with their derivatives $f'(x) = f(x)(1-f(x))$. In addition to the sigmoid binary function, there is still a choice of other sigmoid activation functions, namely bipolar sigmoid at intervals (-1.1).



2.3. Figure of bipolar Sigmoid function

There are 3 steps that must be done in doing weight training for backpropagation neural networks. The first stage is conducting advanced training. At this stage the training is done by inputting a training pattern that is counted forward from the input screen to the output screen, using the activity function that matches the character of the data to be trained. The second stage is the backward phase. At this stage the training is carried out backwards, from the output layer continuously to the input layer. This training continues until the limits of mistakes can be tolerated. An error in artificial neural network training is the difference between the calculated value at the network output and the desired target value.

- The initial step : forward propagation

At this advanced propagation stage, the input signal pattern (= xi) is propagated to the hidden layer using predetermined activation functions. The results obtained at each hidden layer (= zj) are then propagated forward to the hidden layer above it using predetermined activity functions. And so on until the output obtained from the network (= yk).

The network output was then compared with the target that should have been achieved (= tk). If there is a difference between tk - yk that is still outside the specified tolerance limit, then the weight on each line of the network must be modified to reduce errors, so that the difference is smaller than the specified fault tolerance.

- Second step : back propagation

Because errors are still occurring, which is still a large difference between tk-yk compared to the error tolerance limits that we set, then the factors δ_k(k=1, 2, ... ,m) are calculated that we will use to distribute the errors that occur at the node the yk layer nodes to all the hidden nodes directly connected to the yk nodes. Likewise, the δ_k value will be used to modify the line weight values directly connected to the output nodes.

With the same technique, the δ_j factor at each node in the hidden layer is calculated. This δ_j value will be used as a basis for modifying the weight of each connecting line between nodes from the hidden layer below. And so on this calculation is done, so that the value of the delta of hidden layer nodes that connect directly to the input layer nodes.

- Third step : changes in weights

Modification of the weight value of artificial neural networks will be done simultaneously, after all the δ_k factor values are calculated. This δ_k value will determine the change in line weight in the lower layer.

For example, the weight of a line that goes to the output layer, the change in weight is based on the value of the δ_k at the output nodes.

III. RESULTS AND DISCUSSION

Tweet Data Analysis

At this stage the tweet data will be tested to determine the effect of the amount of data on the accuracy of data testing in several scenarios. The following explains the percentage of data in several scenarios. In each scenario, the process of testing the accuracy of documents that are appropriate and not appropriate. The test can be seen in the following table. In each scenario, the researcher uses tweet data on Indonesian soccer opinions with a total of 1000 tweet data. In this process weighting is done using the TF-IDF method and data classification using backpropagation neural network algorithms with classification determinant parameters as follows:

Table 3.1 Parameter Scenarios for data classification

scenario	learning rate	Momentum	Epoch	hiden node
1	0.1	0.0	10000	1
2	0.2	0.0	20000	2
3	0.4	0.2	40000	4
4	0.6	0.3	60000	6
5	0.1	0.0	10000	1

Based on the parameter table above, the following results are obtained:

Table 3.2 Confusion Matrix of data

sce	Amoun	Amoun	Amount	Amoun

nario	t TP (True Positive)	t FP (False Positive)	FN (False Negative)	t TN (True negative)
1	344	1	2	339
2	9	336	334	7
3	345	0	2	339
4	344	1	1	340
5	6	432	404	9

Table 3.3 Estimator Evaluation of Testing Data

sce nario	Accurati on	Precisi on	Recal l	f-measure
1	50,44%	99,71	50,37	66,93%
2	50,29%	100,00	50,29	66,93%
3	50,58%	100,00	50,44	67,06%
4	50,29%	99,71	50,29	66,86%
5	48,18%	1,37%	40,00	2,65%

In table 3.3 it can be seen that the results of the system testing process produce:

- the accuracy level for each scenario is 50.44%, 50.29%, 50.58%, 50.29%, 48.18%.
- the accuracy rate for each scenario is 99.71%, 100.00%, 100.00%, 99.71%, 1.37%.
- the recall rate for each scenario is 50.37%, 50.29%, 50.44%, 50.29%, 40.00%.
- the f-measure level for each scenario is 66.93%, 66.93%, 67.06%, 66.86%, 2.65%.

IV. CONCLUSION

Based on the five scenarios above, we can conclude that when the learning rate inputted is of minimum value, it will produce a low level of accuracy, on the contrary if the learning rate inputted is large will produce a large degree of accuracy as well.

And from the results of backpropagation neural network classification, with a total of 1000 tweet data, the best scenario is third scenario with an accuracy rate of 50.58%, a precision level of 100.00%, a recall of 50.44%, and an f-measure of 67.06 %.

BIBLIOGRAPHY

[1] Andrianto, Brian. Indriati. Adinugroho, Sigit.”Analisis Sentimen Konten Radikal Melalui Dokumen Twitter Menggunakan Metode Backpropagation”, Jurnal Pengembangan Teknologi Informasi
[2] dan Ilmu Komputer Vol. 2, No. 12, Desember 2018,
[3] Feldman , Ronen dan Sanger, James. 2007. The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, New York. Indonesia. Bandung: Universitas Telkom.
[4] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,”
[5] Found Trends Inf Retr, vol. 2, no. 1–2, pp. 1–135, Jan. 2008.
[6] Pang, Bo., dan Lee, Lillian. 2008. Opinion Mining and Sentiment Analysis. Computer Science Department, Cornell University, New York, USA.
[7] Pang, Bo., et. al. 2002 . Thumbs up? Sentiment classi?cation using machine learning techniques.Computer Science Department, Cornell University, New York, USA : Cambridge University Press.

[8] Han, J., and Kamber, M. 2006. Data Mining Concepts and Techniques. Second Edition. California: Morgan Kaufman.

[9] Hussain, H.I., Kamarudin, F., Thaker, H.M.T. & Salem, M.A. (2019) Artificial Neural Network to Model Managerial Timing Decision: Non-Linear Evidence of Deviation from Target Leverage, *International Journal of Computational Intelligence Systems*, 12 (2), 1282-1294.

[10] Clayton R. Fink, et. al. 2011 . Coarse- and Fine-Grained Sentiment Analysis of Social Media Text.Johns hopkins apl technical digest, volume 30,number 1.

[11] Solimun (2002), *Structural Equation Modeling* LISREL dan Amos, Fakultas MIPA Universitas Brawijaya, Malang.

[12] Mark A. Hall., & Lloyd A. Smith. 1999. Feature Selection for achine Learning: Comparing a Correlation-based Filter Approach to the Wrapper. In FLAIRS Conference

[13] Yiming Yang,. & Jan. O. Pedersen. 1997. A Comparative Study on Feature Selection in Text Categorization. Proceedings of ICML-97, 14th International Cenference on Machine Learning.

[14] T.Sutojo, Edy Mulyanto, Vincen Suhartono. 2010, Kecerdasan Buatan. Jakarta: Andi Offset