

CAMPUS SENTIMENT ANALYSIS E-COMPLAINT USING PROBABILISTIC NEURAL NETWORK ALGORITHM

^aMohammad Zoqi Sarwani, ^b Wayan Firdaus Mahmudy

^{a,b}Magister Ilmu Komputer/Informatika

Universitas Brawijaya, Jl. Veteran No.8 Malang

E-mail: zaki_swr@yahoo.com

Abstract

E-complaint is one of the technologies which is used to collect feedback from customers in the form of criticism and suggestions using electronic systems. For some companies or agencies, e-complaint is used to provide better services to its customers. This study is aimed to perform sentiment analysis of an e-complaint service, with the case of Brawijaya University. There are three main stages for the proposed system, i.e. Text Preprocessing, Text Weighting, and PNN for the classification. Tokenization, filtering, and stemming are done in the text preprocessing. Resulted text from the preprocessing stage is weighting using Term Inverse Document Frequent (TFIDF). To classify the negative or positive complaints, PNN are used in the last stage. For the experiments, 70 data are used as the training data, and 20 data are used as the testing data. The experimental results based on the combination of the number of training and testing dataset, showed that the accuracy achieved up to 90%.

Keywords: Probabilistic Neural Network, PNN, E-Complaint, TF/IDF

INTRODUCTION

Services on an agency or company must have quality standards that need to be improved if you want to make the maximum contribution to the users of services. At this time, various institutes or companies are competing to improve the quality contained in the institution. One of the driving factors of tight competition is the standard quality of service. Standard service becomes important for services are an image or the first impression is given by an institution to its clients. Standard services too often become a special assessment for the client in determining their interest. There are various ways to measure standard of service provided by an institution. One way is using survey to the clients or providing a dropbox box in a specific place to accommodate collect the feedback received clients.

However this technique is likely ineffective because it is hard to manage such paper based information. Especially when it comes to the assessment and understanding of the customers suggest about the University's services. As the development of technology, modern methods that can be used are e-complaint. E-Complaint is a web-based technology that provides online services to clients to provide such information in response to the performance of services on an institution. E-Complaint in higher education is generally used as a quality assurance to process and fulfillment of reduces rotational quality standards consistently and continuously, and gives satisfaction to the customers in making sustainable development (continuous improvement). To perform continuous development and providing customer satisfaction with the services provided, customer complaints is an expression of dissatisfaction on a consumer's behalf to a responsible party. It's very valuable and should be followed up. Customer complaints submitted via e-complaint where then used as a reference for the evaluation and early detection of systems and standards needed to achieve a predicate as World Class University [1].

Brawijaya University is one of the agencies that have implemented the E-Complaint services. There are several aspects complained of by the system offers various services to users including students, graduate, curriculum, human resources, teaching, academic atmosphere, information technology services

and other aspects [1]. However, the system used in UB is still not able to provide analysis on the feedbacks submitted by the customers automatically. This in particular suggests that the system could be extended by adding sentiment analysis feature to help synthesizes the comments collected in the system, Sentiment analysis is a field of science that studies on how to analyze sentiment, evaluations, opinions, attitudes, judgments, and emotions of an entity can be a product, service, organisms, people, issues, events, and topics [2].

Research on sentiment analysis has been done by some previous investigators that as practiced by [3] and [4] which uses fuzzy sets for classification. While the research conducted by [5] using fuzzy inference system Tsukamoto. Tsukamoto fuzzy method has also been used by [6]. Selection method performed by previous researchers provides results that have a low accuracy rate. The fuzzy logic method is not included as one of the machine learning methods. In addition, the method has a high level of complexity in the application on a classification of text-mining. Therefore, it takes a machine learning algorithm that is capable of providing a high degree of accuracy. There are several algorithms machine learning that is often used for text classification such as Naive Bayes algorithm used in previous studies by [7]. Besides Naive Bayes algorithm, there are other algorithms that Artificial Neural Network that has the capacity that was adopted from the workings of the human nervous system [2]. Simon Haykim (1994) states that the speed of a human neuron in processing an event very quickly because it has a lot of neurons are organized so well that the human neural network system faster than even the fastest computers today [8].

In Artificial Neural Network (ANN) there are different types. Therefore, it takes foresight in determining the appropriate type of ANN algorithms for cases taken. In a previous study conducted by [9] that uses one of the ANN algorithm to solve the case of text mining. In the study, the ANN method chosen is the Support Vector Machine (SVM). In this study, we chose Probabilistic Neural Network (PNN) PNN is a method for developing a theory Bayes ANN that uses statistical calculations. This method provides good performance and fast in

the classification because it is done in one step [10].

Regarding the problem, to analyze sentiment E-Complaint needed a fast and accurate method and has a high accuracy. Therefore, the researchers used machine learning methods for the analysis of sentiment e-complain that the algorithm Probabilistic Neural Network (PNN).

With the development of information technology today, more and more computer systems are offered to facilitate the work that had been done manually. Sentiment analysis is one example of a computer system that is currently a lot of research, both sentiment analysis for a product, e-complaint and so forth. There are a variety of previous studies that do this, one of which is carried out by [4]. In the study, the method used is a fuzzy set which in the study to get fairly accurate results. However, the level of accuracy obtained is still less when compared to the results of research conducted by[5]. The study uses data in the form of e-complaint contained in UB, which is then analyzed using a Fuzzy Inference System Tsukamoto.

In the study conducted by [5], the selection method used for classification is not one of the methods contained in Machine Learning. Tsukamoto Fuzzy method has low accuracy when compared to other methods of machine learning. In Tsukamoto fuzzy method, there is a stage to make the boundaries of membership and membership degree. At this stage is a stage that could determine a high degree of accuracy. In the study conducted by[5], a stage to determine the membership function searched using the weighting method TF / IDF. In addition, the Tsukamoto fuzzy method has a high level of difficulty in the case study text mining, as well as the process of calculation for the classification. Previous studies have not investigated text mining using fuzzy inference system. Because this method is one method used for cases that are gray or uncertainty.

METHODOLOGY

To perform sentiment analysis process e-complaint there are various stages must be passed. Figure 1 shows the adopted in this study which include making the text preprocessing beforehand. This process is done to remove the words and symbols that have no meaning. After that, the next step is to do the

weighting word using TF / IDF before the analysis using algorithms PNN. PNN algorithm is used to classify e-complaint's document, whether e-complaint document is negative or positive complaint.

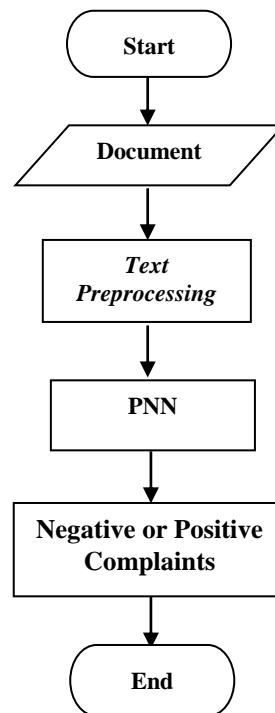


Figure 1. Research Flowchart

In the proposed e-complaint system, there are several stages i.e., text processing, word weighting, and the classification process uses algorithms PNN.

Data

The data used in this study is collected from e-complaint system of UB. Data taken from the e-complaint is the raw data that must be carried out pre-processing process in advance before could be in the next stage

Text Processing

Text processing is a process that is done to perform text processing. At this stage, there are several steps taken of text processing, i.e., tokenization, filtering, and stemming.

Prior to tokenization step, a process called 'To Lower Case' should be done in which the entire text will be converted to lowercase. Once it is done, tokenization process, is carried out to

break up the sentence based on the delimiter spaces and eliminate the numbers and symbols [11].

Filtering is a process that is done after the process of tokenization. This process aims to eliminate words that are considered stopword. And the last is stemming, the process undertaken to obtain basic words by omitting suffix to a word [12][13]. Flow performed on text processing process is presented in Figure 2.

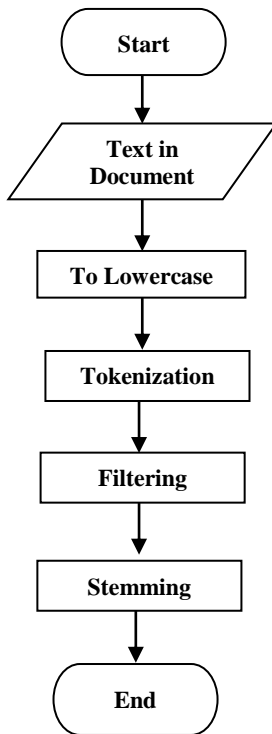


Figure 2. Text Processing Flowchart

Text Weighting

Text weighting method used in this study is the Term Frequent Inverse Document Frequent (TFIDF), where the weighting process is performed weighting text each word contained in the document to be used as input to the input layer PNN algorithm. This method was chosen because it can provide a good solution. The formula used in this method is shown in Equation (1).

$$X_{jk} = TF_{jk} \times tdf_k \quad (1)$$

where :

j is 1,2,3,4,..., n

k is 1,2,3,4,..., m

X_{jk} = the weight of word j contained in the document k

TF_{jk} = number of appear of word j in the document k .

tdf_k = equals to $\log \frac{n}{df_k}$, where n is index number of document

df_k = number of documents.

Probabilistic Neural Network Algorithm (PNN)

Probabilistic Neural Network (PNN) is the model established by the appraiser function of solid opportunities. This model offers a classification process that is good and fast because the process is only performed one stage.

PNN was used in this study which has three layers include: Input Layer, Layer Pattern, and summation Layer [14]. Words contained in each document used as input layer and in the pattern layer is frequent of words in each document. A number of neuron in pattern layer is based on the number of words in each document. PNN method in this study used only one-time training. The training process is very fast when compared to other methods such as Neural Network Feed-Forward Neural Network [10].

On the input layer is input a document to be classified. Then of the layer pattern is the result of the weighing of words in the document by using the formula in Equation (1) is used as an input to the formula In Equation (2).

$$F_{ki} (d_j) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{d_j^n W_{k,i} - 1}{\sigma^2}\right) \quad (2)$$

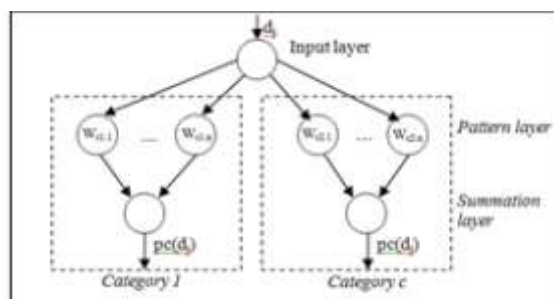


Figure 3. PNN Algorithm design [14]

For the next stage is the summation layer. In this layer adding up, all the values obtained in the previous layer for each category using the formula in equation (3).

$$p_i(d_j) = \sum_{k=1}^{N_i} F_{k,i}(d_j), i = 1,2,3,\dots,c \quad (3)$$

RESULT AND DISCUSSION

We Have Analyzed The E-Complaint Of UB Using Probabilistic Neural Network. Before The Analysis Process Using The Input Method In The Form Of E-Complaint Process Is Carried Out Pre-Processing Is The Process Of Normalizing Text Can Be Input Into The Method. The Method Used To Perform Weighting was That By Using TF / IDF.

In this study, a system built has two stages, namely the first stage of the process of training data or training data. In this experiment, the data used for training are 70 and for testing are 20. The process used to calculate the accuracy of the system by using some combination of training data and testing the data to get a combination of the amount of training data and testing data is to obtain a high degree of accuracy. Table 1 presents a combination of training data and testing the data and the level of accuracy obtained for each combination. To get the level of accuracy in any combination used a formula to calculate the number of the correct document on testing data is divided by the number of documents used in the data testing.

In Table 1 It Can Be Seen A Combination Of Training Data And Testing Data To Give A High Degree Of Accuracy Is A Combination Of Training Data With The Number 70 With The Number Of Testing Data 20 Has An Accuracy Rate Of 90%.

REFERENCES

- [1] "Beranda | E-Complaint UB." [Online]. Available: <https://e-complaint.ub.ac.id/>. [Accessed: 11-Dec-2015].
- [2] Suyanto, Artificial Intelligence : Searching, Reasoning, Planning, dan Learning (Revisi Kedua). Bandung: Informatika Bandung, 2014.
- [3] Liu, B, "Sentiment Analyze and Opinio Mining. In Sentiment Analyze and Opinion Mining (p. 7)," Chic. Morgan Clypool Publ., 2012.
- [4] Gi Fiu and X Wang, "Chinesee Sentence-Level Sentiment Classification Based On Fuzzy Set," Coling PosterColing, pp. 312–319, 2010.

Table 1. Results of testing using the PNN

No	Number of Training Data	Number of Testing Data	Accuracy of Naïve Bayes	Accuracy of PNN
1	70	20	90%	90%
2	50	20	80%	86.67%
3	35	20	50%	80%
4	20	20	50%	80%
5	10	20	15%	75%

As seen in Table 1, increasing the number of makes the accuracy of the classification increased. We also used the Naïve bayes for the classification stage as the comparison of our proposed method. The proposed system gives the same accuracy compared with the Naïve Bayes for the classification stage. However, the Naive Bayes methods cannot achieve a better accuracy compare to PNN, if the number of training was a small.

CONCLUSION

Our proposed system consists of three main stages, i.e. text preprocessing, text weighting, and classification of the negative or positive complaints. We use TF/IDF method to calculate the weight of the words in document, and PNN for the classification. There are three layers are used in PNN, first, the input layer, second, the pattern layer, and the third is the summation layer. The experiments showed that our proposed system achieved up to 90% accuracy.

For further research another weighting method can be used and combine with PNN or other classification method to get a better classification accuracy.

- [5] Yudhistira, B.N, "Sentiment Analisis E-Complaint Kampus Menggunakan Fuzzy Inference System Metode Tsukamoto," Univ. Brawijaya, 2014.
- [6] M. Sholihin, N. Fuad, and N. Khamiliyah, "Sistem Pendukung Keputusan Penentuan Warga Penerima Jamkesmas Dengan Metode Fuzzy Tsukamoto," *J. Tek. Vol.*, vol. 5, no. 2, 2013.
- [7] M. Z. Sarwani and W. F. Mahmudy, "Analisis Twitter Untuk Mengetahui Karakter Seseorang Menggunakan Algoritma Naive Bayess Classifier," *SESINDO 2015*, pp. 1–6, 2015.
- [8] Simon Haykim, "Neural Networks: A Comprehensive Foundation," Macmillan Publ. Co. N. Y., 1994.
- [9] S. T. Sinaga and M. L. Khodra, "Restricted Content Classification Based On Video Metadata And Comments (Case Study: youtube.com)," *Kursor*, vol. 7, no. 4, 2016.
- [10] E. E. Saputri, R. S. Wahono, and V. Suhartono, "Pemilihan Parameter Smoothing pada Probabilistic Neural Network dengan Menggunakan Particle Swarm Optimization untuk Pendeteksian Teks Pada Citra," *J. Intell. Syst.*, vol. 1, no. 1, pp. 22–26, 2015.
- [11] Weiss, S.M, Indurkha, N, Zhang, T, and Damerou, F, *Text Mining: Predictive Methods for Analyzing Unstructured Information*. New York: Springer, 2005.
- [12] Feldman, R and Sanger, J, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press, 2007.
- [13] M. W. Berry and J. Kogan, Eds., *Text mining: applications and theory*. Chichester, U.K: Wiley, 2010.
- [14] P. M. Ciarelli, E. Oliveira, C. Badue, and A. F. De Souza, "Multi-label text categorization using a probabilistic neural network," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 1, pp. 133–144, 2009.