Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 7/1 (2014), 18-23
DOI: http://dx.doi.org/10.21609/jiki.v7i1.252

USER EMOTION IDENTIFICATION IN TWITTER USING SPECIFIC FEATURES: HASHTAG, EMOJI, EMOTICON, AND ADJECTIVE TERM

Yuita Arum Sari, Evy Kamilah Ratnasari, Siti Mutrofin, and Agus Zainal Arifin

Dept. of Informatics, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember, Jalan Teknik Kimia, Gedung Teknik Informatika Kampus ITS Sukolilo, Surabaya, 60111, Indonesia

Email: yuita.sari12@mhs.if.its.ac.id

Abstract

Twitter is a social media application, which can give a sign for identifying user emotion. Identification of user emotion can be utilized in commercial domain, health, politic, and security problems. The problem of emotion identification in twit is the unstructured short text messages which lead the difficulty to figure out main features. In this paper, we propose a new framework for identifying the tendency of user emotions using specific features, i.e. hashtag, emoji, emoticon, and adjective term. Preprocessing is applied in the first phase, and then user emotions are identified by means of classification method using kNN. The proposed method can achieve good results, near ground truth, with accuracy of 92%.

Keywords : *emotion identification, tendency, tweet, classification*

Abstrak

Sebuah *tweet* dapat mengandung dan menggambarkan kecenderungan emosi seseorang. Penelitian mengenai identifikasi emosi dapat diterapkan pada domain komersial, kesehatan, politik, dan keamanan. Teks pendek yang tidak terstruktur dalam data *tweet* menyebabkan sulit menemukan fitur-fitur penting. Pada pe-nelitian ini diusulkan sebuah model baru untuk mengidentifikasi kecenderungan emosi pengguna Twitter menggunakan fitur khusus yaitu *hashtag*, *emoji*, *emoticon*, dan kata sifat. Tahap awal dilakukan *prepro-cessing*, kemudian identifikasi emosi pengguna dengan metode klasifikasi. Hasil penelitian ini mempunyai kecenderungan emosi yang mendekati *ground truth* dengan akurasi 92% menggunakan *kNN*.

Kata Kunci : *identifikasi emosi, kecenderungan, tweet, klasifikasi*

1. Introduction

Automatic identification and extraction of emotion in text is currently an active study in the Natural Language Processing (NLP) research area. Emotion is an important element of human nature that has been widely studied in psychology and behavioral sciences. The general goal of textual emotion identification is to get information about type of emotion automatically. It is useful for analysts track attitudes and feelings in commercial, health, politics, or security domains through online forum [1]. Twitter is an application mostly used in emotion identification because user may update the hundreds of millions tweets of time in a day, to express their feeling through the broadcast of brief text post. The short text is necessary part that represents deeper understanding of user's behavior and actions.

Traditionally, due to the statistical classification nature, the most common practices adopted by researchers are mainly statistics-based models [2]. Multinomial Naïve Bayes (MNB) and LIBLINE-AR classifier is used to identify emotion by harnessing emotion related hashtag available in tweets [3]. Most of hashtag purpose is to indicate the topic and to determine the tone of message or their internal emotions. The highest accuracy of 65.57% is achieved by applied the two different machine learning algorithms to study the effectiveness of various feature combinations of unigrams, bigrams, sentiment/emotion-bearing words, and parts-of-speech information.

Classification techniques for textual emotion identification with content and hashtag, may be noisy because of not having a direct correspondence with the desired classification when each tweet is labeled manually. Therefore, the similarity between hashtag and emoji can be experimented using classification techniques[4]. The performance of labeling emotion with emoji and hashtag in using SVM via LIBSVM achieved good performance to predict some emotions such as happiness, sadness and anger than the others (fear, surprise and disgust).

In addition, user emotions identification also uses emoticon feature. Emoticon represents character of user's facial expression as their emotions

[5]. Better accuracy was achieved using character-based features about 85.9% of happiness emotion and 80% accuracy for "happy" and "fear" using distant supervision with various author-supplied conventional labels (emoji and emoticon).

Both emoji and emoticon were used in labeling data for emotion identification cause ambiguity. It is happened because not all emoji and emoticon have high relevance of emotion's label. The usage of some emoji and emoticon were not appropriated to real status [5], in which each user has different meaning.

Mostly the tweet content of informal messages written in abbreviated terms or expressive manner. The expressive manner is an effective of textual affect sensing based on adjective word features in textual messages[6]. Adjective terms in tweet can be represented using WordNet-Affect. So that WordNet-Affect was extracted automatically from Twitter corpus. This additional feature produce promising result significantly regarding its capability to recognize affective information in text from an existing corpus of informal online communication medium about 79.4% by the developed affect analysis model.

WordNet-Affect and WPARD datasets is also used in text classification of Indonesian language [7] into six basic emotion expression classes which consist of '*jijik*' (disgusted), '*malu*' (ashamed), '*marah*' (angry), '*sedih*' (sad), '*senang*' (happy), and '*takut*' (afraid) whose documents were obtaned from article. Accurateness percentage of 71.26% was obtained using K-Nearest Neighbor (kNN) classifier at k=40 as the optimum value.

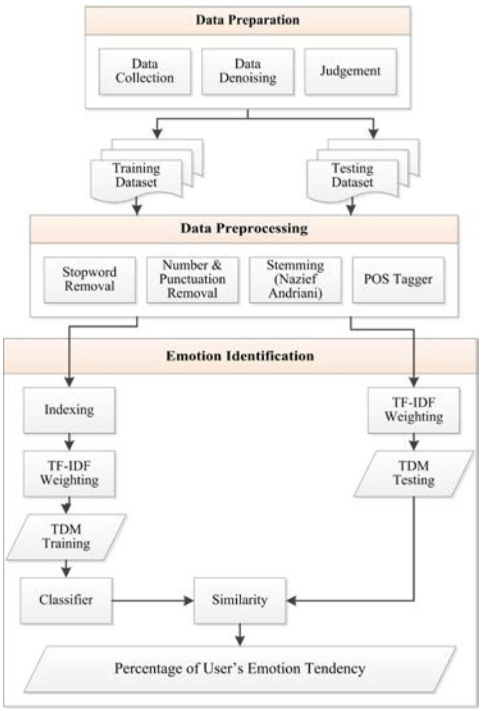

Figure 1. Proposed method.

The first important requirement in text mining is getting indexed terms, which are needed to extract. Sometimes the number of indexed term cannot always give better classification result, and whole extracted terms are not correlated well.

Therefore, we propose a new framework for identifying user emotion by its tendency based on specific features, which consists of hashtag, emoji, emoticon, and adjective term. We apply classification techniques of proving the performance in user emotions identification system. We compare some classification algorithms i.e. kNN, fuzzy kNN, and SVM to know the validity of proposed framework as emotion identification.

2. Methods

In our work, we identify user's emotion tendency using classification technique based on specific features. The first phase is collecting user's tweet, hashtag, emoji, emoticon, and adjective terms, which are compared to WordNet-Affect dataset. Then, de-noising is applied to recheck whether a tweet contains emotion or not. The rest, judgment given to label a tweet manually based on expert's opinion. Expert is a person who gives label or annotation of emotion in each tweet.

Data preprocessing is utilized in the second phase to normalize the tweets by removing stop word, eliminating numbers and punctuation marks, stemming and POS tagger. Then identify user emotions and obtain the percentage user's textual emotion by transforming each tweet document into a vector terms and using some classification techniques to compare the proposed model performance. Figure 1 shows several phases of proposed method. This experimental model is conducted in Java for extracting and cultivating tweet into term weighting matrix of training and testing data, and using MATLAB to compute classification and evaluation measurement.

Data Preparation

Preparing the data manually is the first step of cleaning data collection in tweets and giving a judge the relevance to its emotion type manually. Human expert is the one who involved to give annotation in each tweet. The experts determine emotion type of each tweet based on their own experience when express what they feel in textual messages. In each tweet must be having one emotion for representing one's feeling. Each tweet is annotated based on tweet content related to specific features word bank.

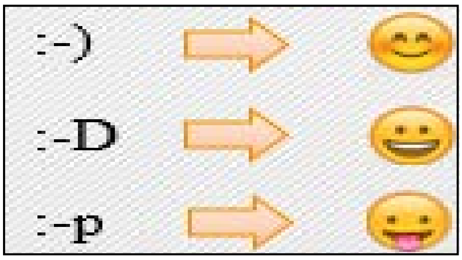

Figure 2. Example of pre-defined image.

TABLE 1
PERCENTAGE IN EACH EMOTION CLASS AMONG 764 TWEETS

Emotion Class	Percentage of Tweets
Marah	25 %
Sedih	24 %
Senang	21 %
Takut	12 %
Terkejut	17 %

In this part of data preparation, we explain the detail of collecting dataset and we used five of six emotion classes based on several previous works that unanimously basic emotion of human, i.e. fear (*takut*), anger (*marah*), surprised (*terkejut*), sadness (*sedih*), and joy (*senang*) [7]. Tweets that contain only one word, link, retweet, and informal words are deleted. We collected the tweets that used as training dataset about 764 tweets with detailed percentage of each emotion class as shown in Table 1. The testing dataset contains of 6 until 15 tweets for each of five personal accounts. In spite of the quantity of training dataset is low, it proves to identify user's emotion in a tweet consistently [8]. Besides collecting tweets, the features such hashtag, emoji, emoticon, and adjective were also collected for different each class. Table 2 shows how hastag, emoji, emoticon, and adjective were grouped manually for each emotion class.

Hashtag is a topic which is any keyword preceded by a hash sign "#" to create groupings on Twitter. Hashtag indicate the subject of user's tweets, collate tweets from different users on a shared subject, and regularly tracked specific events in real time. A predefine image shown in Figure 2 is converted codes of emoji.

Emoticon contain characters combination made by user to express their face appropriate to their emotion through the text. Other features that can describe user emotion in text content are called adjectives. Adjectives are usually considered as effective features since they can be good indicators of emotion. Some research [6][7], shows that using adjectives alone produce competitive result with those obtained by using WordNet-Affect in textual affect sensing.

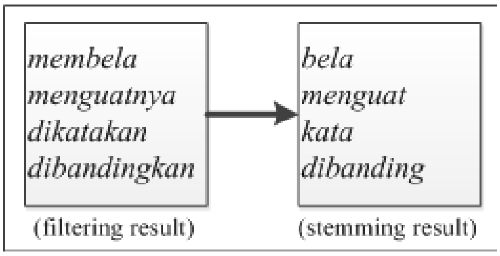

Figure 3. Stemming process.

Data Pre-Processing

Collected tweets dataset of training and testing data then be pre-processed by removing stopword, eliminating numbers and punctuation marks in tweets, stemming then POS tagger. This step is necessary because of informal form when using tweet language and may have noise in the data as characteristics of tweet. Removing stopword is performed to take off any words that can be described and not appropriate to the certain tweet topic. Nazief Andriani stemmer [9][10][11], has about 93% truth is used to transform the words contained in the tweets to the root words based on certain rules. Stemming for Indonesian text that shown in Figure 3. consists of removing the suffix (inflection and derivation),derivational prefix, infix and confix [9]. The POS tagger method represents sentence elements such as nouns, adjectives, adverbs, and others.

Emotion Detection

Classification technique is used to identify emotion in brief text require a form of words group to be process. So, each document tweet will be transformed into a vector term-space. Then the vector term-space of training and testing data are weighted using integration of Term Frequency (TF) and Inverse Document Frequency (IDF) by using equation(1) [7][12].

$$w_{t,d} = tf_{t,d} \times (log_{10} N/df_t) \qquad (1)$$

where, $tf_{t,d}$ is amount of term *t* frequency that occurred to the document *d*, and $df_{t,d}$ is amount of term *t*, while *N* is the number of document in corpus.

Weighted vector term of testing data is transformed by generate a matrix frequency called Term Document Matrix (TDM), which consist of TF-IDF weighting frequency. TDM show that rows as number of words and columns as the number of document contained in tweet corpus. This step is also applied in training data which is

TABLE 2

INSTANCE OF HASHTAG, EMOJI, EMOTICON, AND ADJECTIVE BASED ON EMOTION CLASS

<table>
<tr><th>Emotion</th><th>Adjectives</th><th>Hashtag</th><th>Emoji</th><th>Emoticon</th></tr>
<tr><td>Senang</td><td>Puas, cinta, baik</td><td>#indah</td><td>:-) | :-D</td><td>^_^ | (≧∇≦)/</td></tr>
<tr><td>Sedih</td><td>Tertekan, sengsara</td><td>#murung</td><td>:-(| :-[</td><td>('_') | (Ｔ▽Ｔ)</td></tr>
<tr><td>Marah</td><td>Benci, cemburu</td><td>#geram</td><td>:-|| | :/</td><td>(°　◇°　)</td></tr>
<tr><td>Terkejut</td><td>Heran, pesona</td><td>#takjub</td><td>°o° | :-O</td><td>＼(◎o◎)／！</td></tr>
<tr><td>Takut</td><td>Tegang, panik</td><td>#gigil</td><td>😦 | 😨</td><td>(・へ・)</td></tr>
</table>

TABLE 3

USER EMOTIONS IN EACH CLASSIFICATION TECHNIQUES USING KNN WITH K=9, FUZZY KNN WITH K=35, AND SVM

User		**Emotion**	***Marah***	***Sedih***	***Senang***	***Takut***	***Terkejut***
User 1		**Ground Truth**	**0.38**	**0.08**	**0.54**	**0.00**	**0.00**
	Output	kNN	0.31	0.08	0.54	0.00	0.08
		fuzzy kNN	0.08	0.38	0.54	0.00	0.00
		SVM	0.23	0.08	0.62	0.00	0.08
User 2		**Ground Truth**	**0.22**	**0.22**	**0.33**	**0.22**	**0.00**
	Output	kNN	0.33	0.11	0.33	0.22	0.00
		fuzzy kNN	0.00	0.56	0.44	0.00	0.00
		SVM	0.33	0.00	0.33	0.22	0.11
User 3		**Ground Truth**	**0.00**	**0.29**	**0.71**	**0.00**	**0.00**
	Output	kNN	0.29	0.14	0.43	0.14	0.00
		fuzzy kNN	0.00	0.29	0.71	0.00	0.00
		SVM	0.00	0.00	0.57	0.00	0.43
User 4		**Ground Truth**	**0.25**	**0.42**	**0.33**	**0.00**	**0.00**
	Output	kNN	0.67	0.08	0.17	0.00	0.08
		fuzzy kNN	0.00	0.75	0.25	0.00	0.00
		SVM	0.08	0.08	0.33	0.00	0.50
User 5		**Ground Truth**	**0.20**	**0.10**	**0.50**	**0.10**	**0.10**
	Output	kNN	0.50	0.00	0.50	0.00	0.00
		fuzzy kNN	0.00	0.50	0.50	0.00	0.00
		SVM	0.10	0.10	0.70	0.00	0.10

indexed first to entirely searching for obtaining the frequency of features that have been collected in a collection.

The next step is a text document *d* can be classified into a certain class. First, TDM of training and testing dataset are labeled manually by a defined class label attributes. Then, classifier model is built by analyzing TDM which describes a concept of data class. Finally, the model is tested on TDM of testing dataset to measure model's accuracy in classifying test data into certain emotion class.

In this paper, we selected kNN [13], fuzzy kNN [12], and SVM [14] as a classifier method, since they proved to outcome the classification problems.

3. Result and Discussion

The dataset were retrieved from Twitter web application manually, and separated them into two parts of training and testing. There were 764 tweets as training which have 5 emotions, then hashtag, emoji, emoticon and adjective terms are extracted in each tweet. The testing dataset is acquired by collecting 5 users' tweet. Each user may have differ-rent number of tweet in a day. It is challenging phase when we got the collection of the testing dataset, because looking for data in a few different range of times in a day is quite difficult. Therefore, the extracted specific feature is needed as explained before. Most of retrieved tweet were unstructured text, so the normalization step is applied in advance for obtaining better performance of classification. In this paper, *jijik* (disgust) is not used, because the ambiguity of clustered adjective term which is retrieved. Most of terms in *jijik* already collected into *marah* emotion.

Each user produced the percentage of emotions tendency and the accuracy in each classify-cation techniques. Figure 4 shows the tweet query that conducted by the first user.

The better classification impact the percent-tage of user emotions, in order to get the result near ground truth. Ground truth is benchmark for giving precision into the emotions prediction which is obtained as output in the system and developed manually by expert based on their experience. Experts give the percentage as tendency in each emotion label. The validity of method is given by comparing expert's observation and result of identification by system. For instance, in Figure 4 a user who has an account @widyadarma has 15 tweets, and the human expert give the emotion label manually. We take the ground truth as the tendency of @widyadarma by adding each emotion and dividing it with the number of tweets

TABLE 3
THE ACCURACY FOR EACH CLASSIFICATION TECHNIQUES IN EACH USER

Accuracy	**User 1**	**User 2**	**User 3**	**User 4**	**User 5**
kNN	**0.92**	**0.89**	0.43	**0.50**	**0.60**
fuzzy kNN	0.69	0.22	**0.71**	0.42	0.30
SVM	0.85	0.78	0.29	0.25	**0.60**
MLP	38.46	33.33	42.86	33.33	30.00

in account @widyadarma. So that, we can gain the tendency of *senang* is57.85%, *marah* is 38.46 %, *sedih* is 7.69%.

The good result is when the percentage given in ground truth is not far with the percentage in the emotions output. Table 3 shows the percentage of tendency of user emotions in each classification techniques and Table 4 shows the accuracy of user emotions in each user.

Choosing of optimal *k* parameter which is conducted by kNN and fuzzy kNN gave 9 and 35, respectively. The optimal *k* is obtained by repeated experiment in all users. In the first user the percentage of ground truth give 54%, 38%, and 8% for *senang, marah,* dan *sedih* respectively. The percentage value in that ground truth, is nearby to the emotion output when using kNN under *k*=9 with 54% of *senang*, 31% of *marah*, 8 % of *sedih*, and the rest 8% of *terkejut*. The *terkejut* result comes from "*Bayangin kalau pampersnya dibuang ke sungai .Penyakitnya juga terbawa arus kemana2 Kasihan masyarakat yang biasa langsung mengakses air sungai :(* ". From that tweet, extracted feature will only get ":(" as emoji which is represent *sedih* (sad) in the emoji's dictionary, meanwhile the ground truth detect that that the tweet is classified into *marah* (angry).

Therefore, kNN failed to detect into correct classification in that tweet because of the semantic among sentences without specific feature is not our focus. It was also happened when using fuzzy kNN and SVM classification. The best of giving emotions predictions is related to the accuracy given. The kNN classification retrieved the best accuracy in the first user's emotion identification. SVM is the second best of predicting users emotion with gives 62% of *senang*, 23% of *marah, sedih* by 8%, and *terkejut* by 8%. The *terkejut* comes same as kNN given, beside *terkejut* SVM gives the misclassified emotions, i.e. *senang*, so *senang* returned the percentage in the first user higher than the ground truth. The accuracy of SVM classification gives 85%. The poor result is obtained by fuzzy kNN classification which achieved four tweets are misclassified, and the accuracy of fuzzy KNN is 69%. The percentage of user emotions using fuzzy kNN shows 54% of *senang*, 38% of *sedih*, and 8% of *marah*. Most of errors come when the ground truth says *marah*, meanwhile the emotions detect *sedih*. It causes the extracted feature from Indonesian POS tagger is not sufficient enough to extract adjective term well, for instance *artis* (artist), *bayang* (shadow), *baru* (new), is detect as adjective that's not represent to the emotion.

The second user shows the percentage of users' emotion 33% of *senang*, 22% of *marah*, *sedih*, dan *takut.* The accuracy achieve 89%, 78%, and 22% of kNN under *k*=9, fuzzy kNN under *k*=35%, and SVM respectively. The kNN reach the best accuracy and achieve the percentage of user emotions nearly to the ground truth. The SVM classification is the second best as the first user result, and fuzzy kNN is the poor result of them.

Most of the testing dataset in the third user classified by 50% when using kNN with *k*=9 and SVM, however the fuzzy KNN classified well. It's proved with the percentage of the emotion same as ground truth. The percentage in ground truth is obtained 71% in *senang* and 29% in *sedih*. The misclassification in fuzzy kNN happened, because of the extracted feature of testing third user dataset is not exist in indexed terms. The accuracy of fuzzy-KNN gives 71%, kNN with 43 %, and the last one is SVM. The poor result is also happened, because there were several terms error when extracted using Indonesian POS Tagger. User Emotions in Each Classification Techniques Using kNN with k=9, fuzzy kNN with k=35, and SVM. The fourth user's tweet can't detect well as the other users. The accuracy gives 50% in kNN, 42% in fuzzy kNN, and 25% in SVM. It's fail when detecting the percentage of user emotions. It shows that there is relation between accuracy with the percentage of user emotions identification. The ground truth achieve the highest percentage in *sedih* by 42%, senang 33%, and the rest is *marah* 25%. The result of kNN classification reach the best accuracy but can't achieve to get the value of tendency user emotions nearly ground truth. It contains 67%, 17%, 8%, 8%, in *marah*, *senang*, *sedih*, and *terkejut*, respectively. It is shown that the highest percentage in output not fitted as ground truth value. It is also happened in fuzzy kNN which has the second best accuracy, contains 75% and 2% of *sedih* and *senang*, respectively. The rest, output emotion achieve 50% in *terkejut*, 33% in *senang*, and 8% in *marah* and *sedih* using SVM classification.

In the fifth user, kNN and SVM yield the same accuracy, i.e. 60%, however that's not same for giving percentage user emotions identification. Both of them produces the highest percentage in *senang* as a gound truth. The classification using kNN and SVM produces the same accuracy, however the most relevant percentage is given by SVM classification. The SVM yield 70% of *senang* and 10% of *marah, sedih,* and *terkejut*. The ground truth contains senang 50%, 20% of *marah*, and the rest 10% in each *sedih*, *takut*, and *terkejut.* The condition is happened because of the weighting value of the extracted features given, have different result in classification.

4. Conclusion

The proposed method is sufficient to overcome the users' emotion identification. The accuracy achieved, is related to the percentage of user emotions prediction. The kNN classification is outperform than SVM and fuzzy kNN. The second best given when using SVM classification. The best result shown in the first user emotions identification is nearly ground truth with accuracy of 92%. The misclassified result comes when the extracted feature in index terms are not related to the dictionary. The imperfect result of Indonesian POS Tagger lead the problem of getting adjective terms that are not appropriate with emotions meaning.

Further research repair the Indonesian POS Tagger which is give impact when extract the specific feature. The specific feature is the important item in the preprocessing for achieving better performance of classification, especially for user emotions identification.

References

[1] J. Wiebe, T. Wilson, & C. Cardie, "Annotating Expressions of Opinions and Emotions in Language," *Lang. Resour. Eval*, vol. 39, no. 2–3, pp. 165–210. 2005.

[2] W. Li & H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl*, vol. 41, no. 4, Part 2, pp. 1742–1749. 2014.

[3] W. Wang, L. Chen, K. Thirunarayan, & A.P. Sheth, "Harnessing Twitter 'Big Data'for Au-tomatic Emotion Identification" *In Privacy, Security, Risk and Trust (PASSAT), 2012 In-ternational Conference on and 2012 Interna-tional Conference on Social Computing (So-cialCom)*, pp. 587–592, 2012.

[4] M. Purver & S. Battersby, "Experimenting with Distant Supervision for Emotion Classification" *In Proceedings of the 13th Conference of the European Chapter of the Associ-ation for Computational Linguistics,* pp. 482–491, 2012.

[5] Z. Yuan& M. Purver, "Predicting Emotion Labels for Chinese Microblog Texts." *In pro-ceeding of: Proceedings of the 1st Interna-tional Workshop on Sentiment Discovery fr-om Affective Data (SDAD 2012)*, pp. 40–47, 2012.

[6] A. Neviarouskaya, H. Prendinger, & M. Ishizuka, "Textual Affect Sensing for Sociable and Expressive Online Communication" *In Proceedings of the 2Nd International Conference on Affective Computing and Intelligent Interaction*, pp. 218–229, 2007.

[7] Arifin, "Classification of Emotions in Indone-sian TextsUsing K-NN Method," *Int. J. Inf. Electron. Eng*, vol. 2, pp.899-903.2012.

[8] R. Kirk., et al, "EmpaTweet: Annotating and Detection Emotions on Tweeer" *in Proceedings of the Language Resources and Evalua-tion (LREC)*, pp.3806-3813, 2012

[9] J. Asian, H.E. Williams, & S.M.M. Tahaghoghi, "Stemming Indonesian". *In Proceed-ings of the Twenty-eighth Australasian confe-rence on Computer Science*, pp. 307-314, 2005.

[10] A. Z. Arifin, R. Darwanto, D.A. Navastara, & H. T. Ciptaningtyas, "Klasifikasi Online Dok-umen Berita dengan Menggunakan Algoritma Suffix Tree Clustering" *Di Seminar Sistem Informasi Indonesia (SESINDO2008)*, 2008.

[11] A. Z. Arifin, I. P. A. K. Mahendra, and H. T. Ciptaningtyas, "Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language". *In Proceedings of the 5th International Confe-rence on Information & Communication Te-chnology and Systems*, pp. 149-158, 2009.

[12] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst. Man Cybern*, vol. SMC-15, no. 4, pp. 580–585. 1985.

[13] T. Cover and P. Hart, "Nearest neighbor pattern classification", *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27. 1967.

[14] HSU, Chih-Wei &LIN, Chih-Jen. "A compa-rison of methods for multiclass support vector machines." *Neural Networks, IEEE Transac-tions on, 2002*.