Evaluation of Classification Methods for Indonesian Text Emotion Detection

Muljono, Nurul Anisa Sri Winarsih, Catur Supriyanto
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
muljono@dsn.dinus.ac.id, nurulanisasriw@gmail.com, catur.supriyanto@dsn.dinus.ac.id

***Abstract*—This paper presents Indonesian text emotion detection and evaluates the performances of four different classification methods: Naïve Bayes (NB), J48, K-Nearest Neighbor (KNN) and Support Vector Machine-Sequential Minimal Optimization (SVM–SMO). The experiment uses Indonesian text corpus, containing 1000 sentences which consists of six emotion classes: anger, disgust, fear, joy, sadness, and surprise. Preprocessing step which consists of tokenization, case normalization, stopword removal, stemming and TFIDF are used to extract the features of text emotion. We conduct 10-fold cross validation and split validation for the experiment. Based on the result, we conclude that SVM-SMO classifier gives the best performance. In the 10-fold cross validation, the result shows that the accuracy of NB, J48, KNN and SVM-SMO are 80.2%, 80.8%, 68.1%, and 85.5% respectively. The same conclusion is also demonstrated by the split validation, the highest accuracy of 86% is also achieved by SVM-SMO.**

Keywords—Indonesian Text Emotion Detection, SVM-SMO, J48, Naïve Bayes, KNN

I. INTRODUCTION

Human emotion has a significant role in the daily communication between two persons or among people interactions. A successful interaction system between human being and computer can be achieved when the system is able to recognize, interpret, and proceed the expression of human being emotion accurately [1].

Most researchers have attempted to detect the computer user's emotion through some ways such as face expression, voice, and text. Identifying human emotion through text is considered as the most essential one. Because, in this globalization era, most worldwide people convey their knowledge or ideas over the internet through text [2]. In addition to text is a medium of knowledge transferring that is simpler to adapt and to proceed than other media [3].

An emotion detection can be developed and integrated into the text to speech application [4]. A storyteller that can represent the emotion of every single character in the story which is still being performed inspires us to conduct this research. The emotion of human speech can be detected by modifying the prosody, pitch, intensity, and the speech signal duration of the speaker's utterance [4]. The emotion detection system can be applied to many sectors as well, such as in business (to recognize the customers' impression based on their statement on the product offered), in education (used to detect the students' eagerness for an online-learning system), in computer gaming, in mental health, in national defense, and others [5].

There are three types of emotion approach models in psychology: categorical approach, dimensional approach, and appraisal approach [6]. This research uses categorical approach which is based on the six different emotion classes: *anger*, *disgust*, *fear*, *joy*, *sadness*, and *surprise* (based on 'Ekman') [6].

There are two techniques for analyzing emotion represented by a text, knowledge based and machine learning [7]. This research uses machine learning technique which can give better result than lexical technique, especially in sentiment analysis. In addition, machine learning can well adapt the different domains [7] [8]. A supervised machine learning technique is used in this research to arrange the classification of the emotion text by comparing some methods such as NB, J48, KNN, and SVM-SMO.

The selection of these classification methods is based on our related work. Some researchers used these methods to classify the emotion text. Chaffar and Inkpen [7] have studied the comparison of NB, J48, and SVM-SMO for English dataset. Meanwhile, KNN has been used for Indonesian text emotion [3] [9]. Our contribution is especially for the Indonesian text emotion. Therefore, we evaluate these classification methods for the comparison.

II. RELATED WORK

Several text based emotion detection were proposed. Chaffar and Inkpen [7] evaluated three classification methods NB, J48, and SVM-SMO to recognize six basic emotions (anger, disgust, fear, happiness, sadness and surprise). The experiment was carried out using several datasets: Text Affect, Alm's dataset, Aman's dataset and the Global dataset. Based on the experiments, SVM-SMO outperformed to the other classification methods.

Silva and Haddela [10] proposed the extension of Term Frequency Invers Document Frequency (TFIDF) for text emotion corpus. Their proposed term weighting is based on the concept that a term in a certain class will have little importance if it appears in large documents of other classes and vice versa.

Nivet Chirawichitchai [11] presents Thai text emotion classification by using several machine learning algorithm and various term weighting methods. There are three term weighting schemes in the experiments, such as Boolean weighting, term frequency weighting, and TFIDF weighting. Support Vector Machine (SVM) with Boolean weighting gave the best performance compared to Naïve Bayes (NB), K-Nearest Neighbor (KNN) and Decision Tree (DT). Thai emotion classification is also studied by Inrak and Sinthupinyo [12]. They proposed to use Singular Value Decomposition (SVD) method to reduce the dimension of the vector. In the experiment, SVM is the best classifier compared to NB and DT.

Li and Xu [13] proposed to use emotion cause extraction to support the emotion classification model. The cause of emotion was considered as an important factor for emotion detection. The model uses chi-square test to select the best features from the corpus. In the classification step, they use Support Vector Regression (SVR) which is variant of SVM.

Another approach is proposed by Jun Li et. al. [14]. They present Chinese text emotion classification based on emotion dictionary. The methods include WordNet for vector construction, SVM and NB for classification. In the comparison of classification methods, SVM gave the best accuracy compared to NB.

Arifin et. al. [9] present tweet emotion detection in Indonesian Language. Non-Negative Matrix Factorization (NMF) is proposed to reduce the number of features. KNN was used to classify 764 tweets from various emotions. Arifin and Ketut Eddy Purnama [3] also present emotion classification in Indonesian Language. KNN and SVM were used to classify the text corpus. As the result, SVM outperforms KNN in term of accuracy.

III. RESEARCH METHOD

Indonesia has a large number of islands, human races, ethnics and local languages. Most Indonesians communicate each other in their own local languages (such as Javanese, Sundanese, Balinese, etc.) in their daily activities. However, the Indonesians use Indonesian in the formal situation (such as at school, office, in the books, in the newspaper, on TV, on the radio, etc.). Indonesian was declared in the "Sumpah Pemuda" (The Youth Pledge) charter in 1928, and was established as the official language in 1945 during the declaration of the Indonesian Independence Day.

A. Dataset

In this paper, we collect dataset from various websites (www.dongengceritarakyat.com, www.dongeng.org, and www.pendongeng.com). The dataset consists of 1-3 sentences. We use WordNet Affect List [15] to obtain the emotional tag in the sentences. We labeled manually the sentences which doesn't contain WordNet Affect List, for example "*Para tentara Abrahah kembali dalam keadaan binasa di mana daging-daging dari tubuh mereka berceceran di jalan*" (The Abrahah warriors' bodies were brought back home and some of their flesh scattered on the ground). This sentence is labelled into class *disgust*. We collected 1000 sentences and recorded in text format file. Table I shows the detail description of the dataset.

TABLE I. DATA DESCRIPTION

Emotion	No. of Sent.	Example word from WordNet Affect List	Example of Sentence
Anger	161	Wrath, angry, anger, enviable, jealous, etc.	*"Amat besar **murka** Prabu Kertamarta ketika mengetahui cerita yang sebenarnya."* (The king Kertamarta was enormously **angry** as he recognized what had really happened.)
Disgust	48	Disgust, refuse, repulse, smelly, etc.	*"Aku sudah bosan memakan tikus tubuh mereka **bau** sehingga membuatku menjadi mual."* (I'm tired of eating rats because they are **smelly** and make me nauseous.)
Fear	136	Horror, afraid, scary, hysterical, etc.	*"Tubuhnya yang mulia bergetar denga keras dan beliau merasakan **ketakutan** dan kegelisahan."* (The king's body trembled violently and he felt **scary** and anxious.)
Joy	278	Happy, spirited, cheerful, peaceful, loving, etc.	*"Suatu hari, kabar yang sangat **menyenangkan** terdengar oleh sang Raja, istrinya kini sedang mengandung dan kemudian melahirkan."* (One day, the king heard **good** news that his wife had delivered a baby)
Sadness	194	Sad, frustrated, sorrow, guilty, depressed, etc.	*"Minah mendatangi Batu Batangkup dengan perasaan sangat **sedih**."* (Minah came to Batangkup Stone **in sorrow**.)
Surprise	183	Surprise, awe, astonishment, dumbfounded, etc.	"*Putri penggiling gandum tersebut semakin **terkejut** melihat jerami benar-benar berubah menjadi emas*." (Daughter of grain grinders are increasingly **surprised** to see that the straw really turned into gold.)

B. Preprocessing

In order to create the feature vector, we conduct some preprocessing steps: tokenization, case normalization, stopword removal, and stemming. These processes are included in the Lucene library. Therefore, the experiment did not create the code from scratch. There are some functions in the library to perform the processes.

C. Term Weighting

A TFIDF (Term Frequency Invers Document Frequency) is a popular term weighting scheme [16]. It can be computed by Eq. 1.

$$TFIDF(t) = TF * \log \frac{N}{df} \quad (1)$$

Where t is a term, TF is the total number of term t occurs in the document, N is the total document, and df is the number of document containing the term t.

D. Classification

Generally, there are two methods in machine learning, supervised learning and unsupervised learning. The difference between them is in the existence of the label in data training. Supervised learning uses label and otherwise. In this paper, four algorithms of classification (NB, J48, KNN, and SVM) are compared to find the best classification algorithm for Indonesian text emotion detection. The experiment of these classification methods are carried out using WEKA (Waikato Environment for Knowledge Analysis).

E. Evaluation

In our experiment, we use 10-fold validation and split validation. Cross-validation divide the data into training and testing data. It means that if the multiple number of the data is 10-folds, it must be divided into ten portions and runs ten times. The nine portions (90% of the total data) is used for training and the rest portion (10% of the total data) is used for testing. Otherwise, split validation is simpler than cross validation. It is only divide the data into training and testing data, when we set the split ratio 90% means that 90% of the data is the training data and 10% is the testing data.

The process of 10-cross validation and split validation produces confusion matrix: True Positive TP and True Negative TN; False Positive FP and False Negative FN. TP and TN represent the correct classification while FP and FN represent the incorrect classification. From the confusion matrix, the average value of accuracy, precision, recall, and f-measure can be obtained. Each of those formulas are shown as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$precision = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

$$f - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

IV. Experiment and Result

We have conducted the experiment in two schemes by using 10-fold validation and split validation. Table II shows the precision, recall, f-measure, and accuracy of classification methods using 10-fold cross validation. Based on the table, it can be concluded that SVM-SMO gives the best performance comparing with the other methods (NB, J48, and KNN). Before we compared KNN to the other methods, we tried several values of k for the KNN. Based on Fig. 1, the highest accuracy of KNN is shown at $k = 5$.

TABLE II. Average Results of Classification Methods using 10-Fold Validation

	Average Precision (%)	Average Recall (%)	Average F-Measure (%)	Accuracy (%)
Naïve Bayes	75.98	76.27	76.10	80.2
J48	79.82	74.68	76.63	80.8
SVM-SMO	**83.22**	**79.12**	**80.68**	**85.5**
KNN	70.40	57.95	56.85	68.1

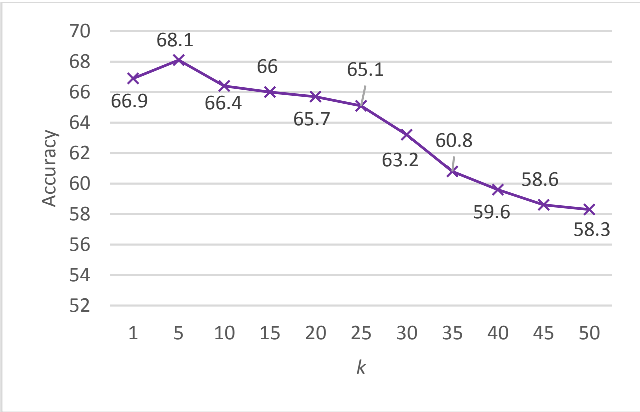

Fig. 1. Accuracy of KNN using 10-fold validation

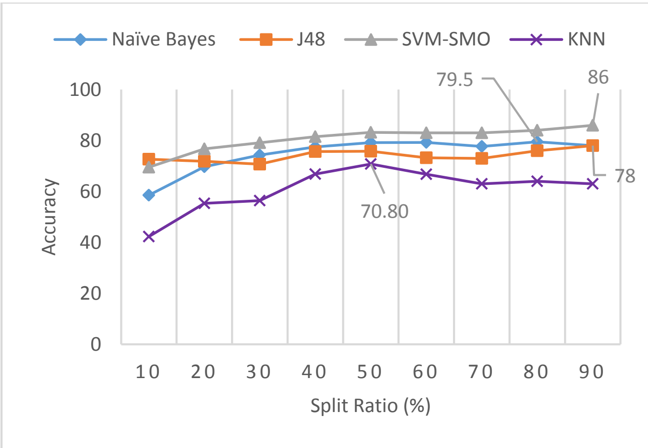

Fig. 2. Accuracy of classification methods using split validation

Fig. 2 shows the accuracy of classification methods using split validation in various split ratio. The percentage number appear in the line is the highest accuracy of each classification method. The first evaluation starts from 10% of percentage split in which the highest accuracy is achieved by J48 method with the value of 72.67%. However, for the next split ratio seemed that SVM-SMO dominate the value of accuracy. The highest accuracy (86%) is showed by SVM-SMO method in the split ratio of 90%. In the figure, the accuracy is increased when the number of split ratio is high. This applies to all classification methods. It means that large number of data training increase the accuracy of classification methods.

V. CONCLUSION

This paper presents the Indonesia emotion text classification using four machine learning techniques (NB, J48, KNN, and SVM-SMO). Our dataset is based on Ekman emotion class collected from Indonesian fairy tales. In the first evaluation, 10-fold validation is applied to gain global accuracy including precision, recall, f-measure, and accuracy of each emotion class. The highest value in global accuracy is showed by SVM-SMO which reaches 85.5%.

In the second evaluation by using split validation, SVM-SMO also gives the best performance by achieving the highest accuracy of 86%. By using 10-fold validation and split validation, SVM-SMO yield the highest accuracy and KNN yield the lowest accuracy. This research is going to be developed further by increasing the accuracy of SVM-SMO and integrating it with some optimization methods, such as genetic algorithm and particle swarm optimization.

ACKNOWLEDGMENT

We would like to thank Indonesian Directorate General of Higher Education for financial support of our research via HIBAH BERSAING research program 2016.

REFERENCES

[1] D. K. Kirange and R. R. Deshmukh, "Emotion Classification of News Headlines using SVM," *Asian Journal Of Computer Science And Information Technology,* 2012.

[2] M. Sunghwan, "Recognising Emotions and Sentiments in Text," Thesis, Dept. Elect. and Inform. Eng., University of Sydney, Sydney, Australia, 2011.

[3] Arifin and K. E. Purnama, "Classification of Emotions in Indonesia Texts using K-NN Method," *International Journal of Information and Electronics Engineering,* vol. 2, no. 6, 2012.

[4] C. O. Alm, D. Roth and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005.

[5] H. Binali and V. Potdar, "Emotion Detection State of the Art," in *Proceedings of the CUBE International Information Technology Conference*, 2012.

[6] P. Ekman, "An Argument for Basic Emotions: Cognition," *Cognition and Emotion,* vol. 6, no. 3-4, pp. 169-200, 1992.

[7] S. Chaffar and D. Inkpen, "Using a Heterogeneous Dataset for Emotion Analysis in Text," in *Advances in Artificial Intelligence*, Springer Berlin Heidelberg, 2011, pp. 62-67.

[8] P. Melville, W. Gryc and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009.

[9] A. Z. Arifin, Y. A. Sari and E. K. Ratnasari, "Emotion Detection of Tweets in Indonesian Language using Non-Negative Matrix Factorization," *International Journal of Intelligent Systems and Applications,* 2014.

[10] J. D. Silva and P. S. Haddela, "A Term Weighting Method for Identifying Emotions," in *2013 IEEE 8th International Conference on Industrial and Information Systems, ICIIS* , 2013.

[11] N. Chirawichitchai, "Emotion Classification of Thai Text based Using Term weighting and Machine Learning Techniques," in *11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2014.

[12] P. Inrak and S. Sinthupinyo, "Applying Latent Semantic Analysis To Classify Emotions In Thai Text," in *2010 2nd International Conference on Computer Engineering and Technology (ICCET)*, 2010.

[13] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Systems with Applications journal,* vol. 41, p. 1742–1749 , 2014.

[14] J. Li, Y. Xu, H. Xiong and Y. Wang, "Chinese Text Emotion Classification Based On Emotion Dictionary," in *2010 IEEE 2nd Symposium on Web Society*, 2010.

[15] C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet," in *Proceedings ofthe 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 2004.

[16] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in *Proceedings of the 10th European Conference on Machine Learning*, 1997.