

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/309211649>

Emotion Detection of Tweets in Indonesian Language using Non-Negative Matrix Factorization

Article in *International Journal of Intelligent Systems and Applications* · August 2014

DOI: 10.5815/ijisa.2014.09.07

CITATIONS

6

READS

87

6 authors, including:



Yuita Arum Sari

Brawijaya University

60 PUBLICATIONS 50 CITATIONS

[SEE PROFILE](#)



Siti Mutrofin

Universitas Pesantren Tinggi Darul 'Ulum

12 PUBLICATIONS 15 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



IT for Agriculture [View project](#)



Tomato Ripeness Identification [View project](#)

Emotion Detection of Tweets in Indonesian Language using Non-Negative Matrix Factorization

Agus Zainal Arifin, Yuita Arum Sari, Evy Kamilah Ratnasari

Department of Informatics, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: agusza@cs.its.ac.id; yuita.sari12@mhs.if.its.ac.id; kamilah.evy12@mhs.if.its.ac.id

Siti Mutrofin

Department of Information Systems, Faculty of Engineering, University of Pesantren Tinggi Darul 'Ulum, Jombang, Indonesia

E-mail: mutrofins@gmail.com

Abstract— Emotion detection is an application that is widely used in social media for industrial environment, health, and security problems. Twitter is a short text message known as a tweet. Based on content and purposes, the tweet can describe as information about a user's emotion. Emotion detection by means of a tweet, is a challenging problem because only a few features can be extracted. Getting features related to emotion is important at the first phase of extraction, so the appropriate features such as a hashtag, emoji, emoticon, and adjective terms are needed. We propose a new method for analyzing the linkages among features and reduced semantically using Non-Negative Matrix Factorization (NMF). The dataset is taken from a Twitter application using Indonesian language with normalization of informal terms in advance. There are 764 tweets in corpus which have five emotions, i.e. happy (*senang*), angry (*marah*), fear (*takut*), sad (*sedih*), and surprise (*terkejut*). Then, the percentage of user's emotion is computed by k-Nearest Neighbor (kNN) approach. Our proposed model achieves the problem of emotion detection which is proved by the result near ground truth.

Index Terms— Emotion Detection, Tweet, Indonesian Language, Emoji, Emoticon, Hashtag, Wordnet-Affect, NMF

I. INTRODUCTION

Emotion detection is a natural language processing and text mining that automatically discover people's emotion from textual data such as a tweet. A tweet is information updated by user in Twitter personal microblog, no longer than 140 characters for networking among followers. Although the status update of users are a brief text, a tweet is updated hundreds of millions of times a day by people all over the world [1], and its content varies tremendously based on user interests and behaviors. Users express their mood as a response of an event and status through the broadcast post.

Twitter has gained attention among users, organizations, and research scholars in different disciplines because of its portability, immediacy, and ease of use, although it is a relatively new communication medium [1]. There are several studies for emotion

detection using tweet data, because of short text and mostly have informal form then regarded as a challenging problem. Detecting emotion has applications for industrial environment, health, and security [2]. Classification technique is applied to decide what emotion the tweet that can be classified based on the content and hashtag [3]. Hashtag serve many purposes, but most notably they are used to indicate the topic, for instance, hashtags indicate the tone of the message or their internal emotions.

So far emotion detection of text, classifies into six basic emotion classes i.e. joy, sadness, anger, fear, thankfulness, and surprise. Each tweet is labeled manually appropriate to the emotion class. The labels may be noisy because of ambiguous, vague or not having a direct correspondence with the desired classification [4]. Therefore, the research of finding similarity based on semantic is needed. Hashtag and emoji are applied to examine the appropriate label between test and train dataset using a classifier. Emoji is a converted code by computer as a Pre-Define image. In addition, emotion detection also using emoticon [5], which made by user from character combination to draw their face expression appropriate to their emotion.

In general, user mostly used emoji to express their emotion, but it is not sufficient to cover emotion detection problems, therefore emoticon feature is also used. Ambiguity can be happened when using emoji and emoticon for labeling data because not all emoji and emoticon have closed relevance in class emotions. The usage of some emoji and emoticon was not appropriated to real status, in which each user has different meaning [5]. Tweet content is user opinions that mostly used in informal language, so relevance between the features and corpus content can't be discovered, then emotion content in a words mostly depend on adjective features [6]. An adjective in tweet can be represented by using wordnet-affect and extracted automatically from Twitter corpus, because it produces better result significantly.

Describing user emotion through a short text of tweet

is a difficult way. Based on the previously works, we use four kinds of features, i.e. hashtag, emoji, emoticon, and adjective terms, that are extracted from data corpus. Not all features produced well during determining the emotion that related with, and it will be effective if using features reduction with semantic concept.

In this paper, we proposed emotion detection method in Indonesian language tweets using NMF to determine relevance among features precisely. Finding correlation among features is important phase to determine user's emotion, and NMF will cover this issues. NMF is used to reduce dimension in multidimensional data and can be revealed to judge the user's emotion by the meaning of extracted features, then it can be used to figure out relationship among them. The emotion in each user will be presented by percentage value using kNN approach by retrieving similarity between query and corpus under certain of k . Emotion was classified into five class i.e. *senang*, *sedih*, *marah*, *takut*, and *terkejut*. Our proposed method result will give benefit for getting user's emotion information in a day.

This paper organized as follow: In section II the research method will be explained. The experimental design and discussion will be presented in section III, and finally this paper is concluded in section IV.

II. METHOD OF RESEARCH

In our work, we detect Twitter user's emotion based on their tweet using NMF, which is semantically found by the linkages among features. First, collecting and denoising user's tweets, hashtag, emoji, emoticon, and adjectives for each class which is used in emotion relevance judgment. Second phase is data pre-processing which is performed to normalize the tweets by removing stopword, eliminating numbers and punctuation marks, stemming then POS tagger. Eventually, by transforming each document into a vector terms, we are able to detect emotion and obtain the percentage emotion by classifying each terms. Then, kNN is used to assess the percentage of emotion in each user. Fig. 1 shows several phases. This experimental model is conducted in Java for extracting and cultivating tweet into term weighting matrix of corpus and query, and using MATLAB to build NMF and compute evaluation measurement.

A. Data preparation

These steps provide three processes i.e. data collection, alleviation data, and emotion relevance judgment. Data collection is a process to judge emotion class and collect tweet manually. Then deleting tweet is needed, which contains only one word, link, retweet, and informal words. Deciding relevance of tweet and class emotion is done manually by expert based on topic such hashtag, emoji, emoticon, and adjective.

In this paper, some facial expressions that describe basic human's emotions can be seen in [7] and we used five of six emotion classes, i.e. fear (*takut*), anger (*marah*), surprised (*terkejut*), sadness (*sedih*), and joy (*senang*). The disgust (*jijik*) emotion class is not included,

because it's difficult to retrieve tweet that related to this emotion class. Since crawl the dataset, most of adjective term's feature in "disgust" already included in anger emotion and the number of retrieved disgust's tweet is far different from other emotions.

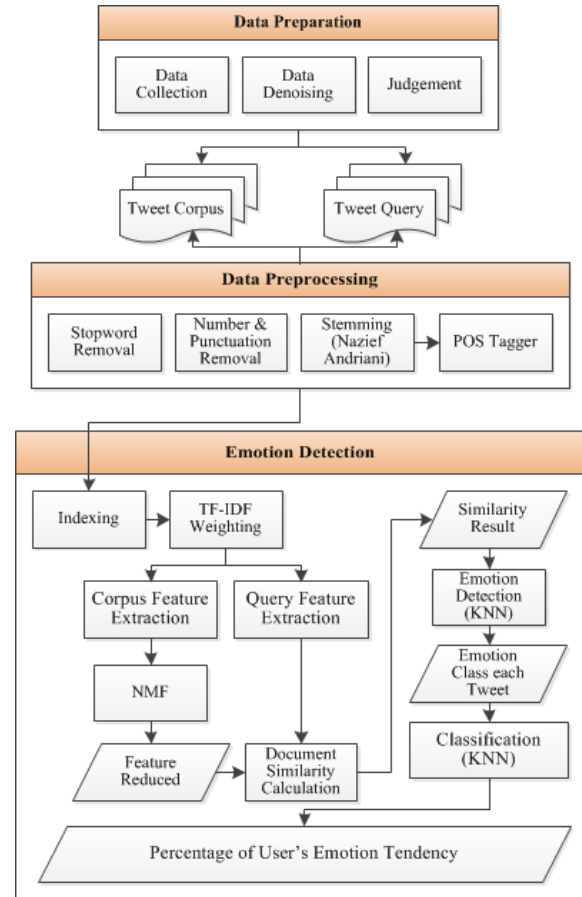


Fig. 1. Proposed Method

Table 1. Example Features in Each Emotion Class

Emotion	Adjectives	Hashtag	Emoji	Emoticon
<i>Senang</i>	<i>Puas, cinta, baik</i>	<i>#indah</i>	<i>:-D</i>	<i>(≧▽≦)/</i>
<i>Sedih</i>	<i>Sengsara</i>	<i>#munung</i>	<i>:-(</i>	<i>(T▽T)</i>
<i>Takut</i>	<i>Tegang, panik</i>	<i>#gigil</i>	<i>☐</i>	<i>(・へ・)</i>
<i>Marah</i>	<i>Benci, cemburu</i>	<i>#geram</i>	<i>:- </i>	<i>(-_-#)</i>
<i>Terkejut</i>	<i>Heran, pesona</i>	<i>#takjub</i>	<i>:-O</i>	<i>(ㄟㄟㄟ)ㄟ</i>

The data collection produce user's tweet, hashtag, emoji, emoticon, and adjectives collection for each class emotion. Table 1 shows how hashtag, emoji, emoticon, and adjective were grouped manually for each emotion class. Hashtag is a topic which is any keyword preceded by a hash sign "#". Hashtag is used to create group on Twitter. Twitter users can use hashtag to indicate the subject of their messages, to collate tweets from different users on a shared subject. In addition, twitter regularly tracked the specific events in real time. Emoji is a pre-define image from converted code. Emoticon contains characters combination made by user to express their face appropriate to their emotion through the text [8]. Other features that can describe emotion in text content are

called adjectives. It can draw emotion meaning that happening on user and the user's feeling information.

B. Data pre-processing

Data pre-processing is necessary phase because of informal tweet form and may have noise in the data as characteristics of tweet. This step consists of stopword removal, eliminating numbers and punctuation marks in tweets, stemming, and tagging with POS tagger. Stopword removal is performed for the removal of words that can't be described and not appropriate to the tweet certain themes. Stemming method using Nazief Andriani [9,10,11] has about 93% of the truth to transform the words contained in the document to the root words using certain rules. Stemming for Indonesian text consists of suffix which is removed (inflection and derivation), derivational prefix, infix, and confix [9]. POS tagger is a representing method of sentence elements, such as nouns, adjectives, adverbs, and others. In this study, Stanford tagger is applied to get the adjective terms of sentence.

C. Emotion detection

Corpus of tweet are updated by user has a brief text in each, and its text should be transformed into a vector terms in the first time to reach the goal of classification process. Indexing method is used to entirely searching for obtaining the frequency of features that have been collected in a tweets collection phase and to generate a matrix frequency called Term Document Matrix (TDM). Indexing 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 query corpus, because we use some tweet document for a user account to identify user's emotion.

The next step is prepare corpus and query documents that have been weighted using integration of Term Frequency (TF) and Inverse Document Frequency (IDF) by using (1) [7,12].

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

where $tf_{t,d}$ is amount of term t frequency that occurred to the document d , and df_t is amount of term t , while N is the number of document. The result of term weighting is a TDM, which consist of TF-IDF weighting frequency in corpus and query.

Matrix reduction of weighted TDM is applied in the corpus using NMF. The features that are extracted using TDM matrix, is then used for giving relevance emotions of query tweet which is given semantically using NMF. NMF can detect the topic based on relevance between features and the predictions of the user precisely [13]. NMF is a technique that can be used to classify documents by decomposing the matrix and have overcome to dimension problem. This technique use partition structure to the parallel cluster of mutually disjoint with the existing pattern and classifying document in the collection according to semantic features in it. NMF technique defines as follows [14]. Given non-negative matrix A , find a non-negative matrix W and H such that,

$$A_{mn} \approx W_{mk} H_{kn}, \quad (2)$$

where A_{mn} is an $m \times n$ column matrix which contained an m non-negative value (number of words) for each documents, while W_{mk} matrix is column k in W are called featured vector, and H_{kn} where each column of H is the weighting column.

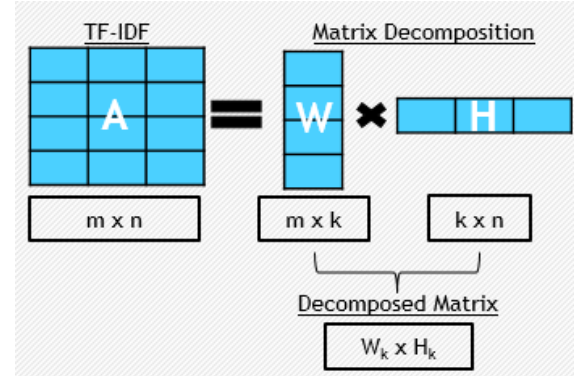


Fig. 2. NMF Feature Reduction Illustration

Fig. 2 describes a set dimensional vector m and the collected vector in $m \times n$ size where m is the number of features and n is the number of vectors. This matrix will be factorized into matrix W in $m \times k$ size and matrix H in $k \times n$ size. The k value must be smaller than n and m value ($k \ll m \times n$) so that W and H matrix will be smaller than initial A matrix. TDM is used by NMF process to produce decomposing matrix and to get smaller dimensions.

After factoring process in TDM, the next step is calculating the value of similarity or relevance between tweet query and tweet corpus. Similarity calculation between query and corpus can be solved by (3).

$$q = q^T W_k H_k, \quad (3)$$

where q is the result of a query vector similarity of a tweet from a user account with the result of training corpus, q^T is the result of a transposed vector of query of the tweet. W_k and H_k is the W and H matrix that has been reduced by k -rank, respectively. Results of vector similarity q are sorted from the largest to smallest and determined the class of a query with a tweet using the KNN approach.

Tweet documents were classified into certain class emotion. The classification process consists of two steps in deciding emotion class. First, detect emotion in each tweet using KNN approach to determine top class emotion for one tweet. Second, determining results of one tweet is classified using KNN to calculate the percentage of emotion class tendency of a user based on some their tweets. We use KNN classifier which uses supervised classification, in which the query result classification is based on class majority [7,15].

III. EXPERIMENTAL RESULT AND DISCUSSION

A. Experimental result

In this paper, we used tweets as text document for training and testing dataset. Those tweets are retrieved

from Twitter web application manually. Then, tweets are classified and labeled manually into 5 emotion classes, and save it in plain text file format as a ground truth of training dataset. The tweet document consists of tweet corpus and tweet query. It's about 764 tweets corpus as training dataset were collected with detailed, 193 tweets for *marah*, 186 tweets for *sedih*, 161 tweets for *senang*, 95 tweets for *takut*, and 129 tweets for *terkejut*. The testing dataset called as query, which are used in this paper, have 5 personal accounts that will be detected their percentage of emotion, in which each account contain 6 until 15 tweets. In each user account may have different number of tweet. It based on finding related query into term indexing in training dataset.

To get query from a user in a few different times is difficult way. There are many tweets that even don't have one of hashtag, emoji, emoticon or adjective feature. So that, we just focus on the dataset that containing those features for getting the emotion concept as proposed. The normalization process for each tweet that has informal language is also applied for achieving the performance of emotion retrieval.

The performance of emotion retrieval is done by evaluating the document similarity based on recall, precision, and F-measure. Recall is the success level to recognize a class that supposes to be recognized and

precision is the precision level of classification result of all document [7,16]. Therefore, recall entails the number of relevant and recognized entity value divided by the number of entities that is supposed to be recognized by the system, while precision is got from the number of relevant and recognized entity value divided by the number of the whole recognitions done by the system. F-measure is representing all system performance and combined recall and precision value.

B. Discussion

In each query produced the evaluation measure, such as precision, recall, and F-measure. The percentage of emotions, are happened in each query of tweet. A tweet has a target and output, in which the target is obtained by giving label manually that accordance to expert's judgment, then comparing it with the output, which is obtained from extracting of the system. Fig. 3 shows the query tweet that conducted by the first user. The 13 tweets, have emotion in each as ground truth, then we call it as emotion target. The value of emotion target and emotion output are analyzed. When the total of percentage in user's emotion has the similar value between emotion target and emotion output, it can be inferred that the system is sufficient to overcome the user's emotion prediction.

```
senang-@widydarma: I miss ʘ ̜̜ and our talks at campus :)
sedih-@widydarma: iya ujee maaf ya tadi salah panggil ^~^
marah-@widydarma: Sekali lagi ya mom jangan buang pampers sembarangan dan dibersihkan dulu sebelum dibuang #marah
marah-@widydarma: Bayangin kalau pampersnya dibuang ke sungai Penyakitnya juga terbawa arus kemana2 Kasihan masyarakat yang biasa langsung mengakses air sungai :(
marah-@widydarma: Pipis tinja yang ada di pampers ternyata bisa jadi media berkembangnya virus polio #murka
marah-@widydarma: Di pampers itu ada pipis tinja dan itu banyak sekali kumannya Biasanya ibu2 buang pampers langsung buang aja tanpa dibersihkan #panas
marah-@widydarma: Moms dan calon moms seantero timeline jangan suka buang pampers sembarangan ya Ternyata bahaya #marah
senang-@widydarma: sekarang jamannya kalem gitu senam :)
senang-@widydarma: aku suka suara jangkrik malam ini lebih suka lagi suara kamu main gitar sambil nyanyi :)
senang-@widydarma: lagian hujan hujan galau mending masak mi rebus kenyang :)
senang-@widydarma: move move :)
senang-@widydarma: tidak perlu mimpi saat jatuh cinta kenyataan lebih indah daripada mimpi :)
senang-@widydarma: aku terapung di linimasa tapi tetap tenggelam dalam samudera cintamu #nyobaromantis :)
```

Fig. 3. The Queries of the First User's Tweet

Table 2 shows the evaluation measure in each query tweet of the first user account in variety of k -rank and k . Then the number of retrieved value is represented by k as nearest neighbor determination approach. For instance, the first query tweet in Table 2 give precision 0.6, recall 0.059, and F-Measure 0.023, it means that the retrieval system can correctly retrieve by 60%, and also produce the relevant information to the query by 5%. In the first tweet of query has an emoji “:)” that can represent the emotion of happy in its tweet. The recall of the first tweet become low, because at the first tweet, there is only an emoji feature extracted from query. These conditions also happened in almost query tweet. Mostly of query tweets measurement, have higher value in precision than recall.

The average of precision, recall, and F-Measure also compute in each emotion class to get correctly retrieved emotion. The *senang* emotion has 62.86% of precision, 5.86% of recall, and 10.71% of f-measure. In *marah* emotion, it has 56%, 4.35%, and 8.08% of precision, recall, and f-measure respectively. Then, in *sedih* emotion, it produces 26.67% of precision, 2.15% of recall, and 3.98% of f-measure. The emotion target shows that *senang* is the highest percentage, the second one is *marah*, and the rest is *sedih*. This condition is appropriate with the performance using average of precision, recall, F-Measure, which also has the highest average in *senang* emotion.

Beside precision, recall, and F-Measure, each tweet has the percentage of emotions obtained by kNN approach. The highest score of percentage will be the decision of emotion detection of user's tweet. For instance in the first query tweet in Table 2, there are *marah* by 6.67%, *sedih*

by 13.3%, *senang* by 60%, *takut* by 13.33%, and *terkejut* by 6.67%. So, we can see that the first tweet query is having the highest score with *senang* by 60%. Then, we can detect that the first user has happy emotion in a day.

Table 2. Sample of Evaluation Measure of Emotions Retrieval Using kNN Approach

Query	P	R	F-Measure	Emotion Target	The Percentage of Emotions	Emotion Output
1	0.6000	0.0559	0.1023	<i>Senang</i>	<i>Marah</i> (6.67%); <i>Sedih</i> (13.3%); <i>Senang</i> (60%); <i>Fear</i> (13.3%); <i>Terkejut</i> (6.67%)	<i>Senang</i>
2	0.2667	0.0215	0.0398	<i>Sedih</i>	<i>Marah</i> (26.67%); <i>Sedih</i> (26.67%); <i>Senang</i> (20%); <i>Takut</i> (20%); <i>Terkejut</i> (6.67%)	<i>Marah, Sedih</i>
3	1.0000	0.0777	0.1442	<i>Marah</i>	<i>Marah</i> (100%)	<i>Marah</i>
4	0.2000	0.0155	0.0288	<i>Marah</i>	<i>Marah</i> (20%); <i>Sedih</i> (26.67%); <i>Takut</i> (20%); <i>Terkejut</i> (33.3%)	<i>Terkejut</i>
5	0.6000	0.0466	0.0865	<i>Marah</i>	<i>Marah</i> (60%); <i>Sedih</i> (20%); <i>Takut</i> (20%)	<i>Marah</i>
6	0.2667	0.0207	0.0385	<i>Marah</i>	<i>Marah</i> (26.67%); <i>Sedih</i> (33.3%); <i>Senang</i> (20%); <i>Takut</i> (13.3%); <i>Terkejut</i> (6.67%)	<i>Sedih</i>
7	0.7333	0.0570	0.1058	<i>Marah</i>	<i>Marah</i> (73.3%); <i>Senang</i> (20%); <i>Terkejut</i> (6.67%)	<i>Marah</i>
8	0.6000	0.0559	0.1023	<i>Senang</i>	<i>Marah</i> (6.67%); <i>Sedih</i> (13.3%); <i>Senang</i> (60%); <i>Takut</i> (13.3%); <i>Terkejut</i> (6.67%)	<i>Senang</i>
9	0.6667	0.0621	0.1136	<i>Senang</i>	<i>Marah</i> (13.3%); <i>Sedih</i> (6.67%); <i>Senang</i> (66.67%); <i>Terkejut</i> (13.3%)	<i>Senang</i>
10	0.5333	0.0497	0.0909	<i>Senang</i>	<i>Marah</i> (6.67%); <i>Sedih</i> (33.3%); <i>Senang</i> (53.3%); <i>Terkejut</i> (6.67%)	<i>Senang</i>
11	0.6000	0.0559	0.1023	<i>Senang</i>	<i>Marah</i> (6.67%); <i>Sedih</i> (13.3%); <i>Senang</i> (60%); <i>Takut</i> (13.3%); <i>Terkejut</i> (6.67%)	<i>Senang</i>
12	0.6000	0.0559	0.1023	<i>Senang</i>	<i>Marah</i> (13.3%); <i>Sedih</i> (6.67%); <i>Senang</i> (60%); <i>Takut</i> (13.3%); <i>Terkejut</i> (6.67%)	<i>Senang</i>
13	0.8000	0.0745	0.1364	<i>Senang</i>	<i>Sedih</i> (13.3%); <i>Senang</i> (80%); <i>Takut</i> (6.67%)	<i>Senang</i>
				<i>Senang</i> (53.85%); <i>Marah</i> (38.46%); <i>Sedih</i> (7.69%)		<i>Senang</i> (53.85%); <i>Marah</i> (30.77%); <i>Sedih</i> (13.85%); <i>Terkejut</i> (7.69%)

The emotion detection is undertaken same in all tweet queries, so it can give a conclusion after those. We would like to compare the result between emotion target and emotion output as proposed method. In the first user account experiment, the result show that mostly of the users has *senang* 53.85%, *marah* 30.77%, *sedih* 13.85%, and *terkejut* 7.69%. That evidence has a bit difference with target emotion class, which is manually labeled before. The percentage in emotion target is *senang* 53.85%, *marah* 38.46%, and *sedih* 7.69%. So in the case of using $k\text{-rank}=250$ and $k=15$, the system is good enough for recognizing first user's emotion.

The same way of giving performance measures and taking percentage of emotions, are done in all user's account. In this case, we have 5 users account, which is having several tweet queries in each. The evaluation method computed in $k\text{-rank}$ 250, 300, 350, and 400. The experiment shows that the critical k in nearest neighbor approach is chosen by k under 5, 15, 25, and 35. Table 3 shows the percentage emotions in all users.

In the first user using $k\text{-rank}=300$ with $k=15$ give same result as the ground truth. The average precision, recall, and F-Measure of *senang* is 56.19%, 5.24%, and 9.58% respectively, for *marah* emotion gives 49.33%, 3.83%, and 7.12% respectively, the rest is *sedih* emotion has 28%, 3.76%, and 6.663% respectively. The result will fail if there is too big value of $k\text{-rank}$ and k , for instance from $k\text{-rank}=350$ and $k=25$, the drawback was produced because of the average of measure value are not appropriate with the ground truth, that is *senang* in 30.77% and *marah* 84.6%. The ground truth of percentage *senang* emotion in first user gives 53.8%, *marah* 38.46%, and *sedih* 7.69%. It should be same value for average precision, recall, and F-Measure, it should has the highest score in happy emotion, the second one for angry emotion, and the last for sad emotion. However, when the condition may be different when $k\text{-rank}$ and k too small or too big the highest percentage of emotion is not happy, but may be other emotions. So, it can conclude that the first user is having most happy emotion.

The same condition is happened in the second user, the percentage of emotion has the same value as ground truth given, when k -rank = 300, and $k=15$. The percentage of *senang* emotion in second user is 33%, and for *sedih*, *marah*, and *takut* has the same emotion by 22.22%. In that k -rank and k , the average of precision, recall, and f-measure of happy emotion is 64.44%, 6%, and 10.98,

respectively. For *marah* emotion gives 40%, 3.11%, and 5%, respectively, for *sedih* emotion gives 56.67%, 4.57%, and 8.46% respectively, and *takut* emotion gives 63.33%, 10%, and 17.28%, respectively. The percentage given by system will be more valid when the measure of the average, precision, and recall are compatible.

Table 3. The Result of Percentage of Emotions in Each User Account, with k -rank 250, 300, 350 and 400 under $k=5, 15, 25, 35$

Number of Tweet	User 1	User 2	User 3	User 4	User 5
	13	9	7	12	10
Ground Truth (Emotion Target)	<i>Senang</i> (53.85%); <i>Marah</i> (38.46%); <i>Sedih</i> (7.69%)	<i>Senang</i> (33.33%); <i>Sedih</i> (22.22%); <i>Marah</i> (22.22%); <i>Takut</i> (22.22%)	<i>Senang</i> (42.86%); <i>Sedih</i> (28.75%); <i>Terkejut</i> (28.57%)	<i>Sedih</i> (41.67%); <i>Senang</i> (33.33%); <i>Marah</i> (25%)	<i>Senang</i> (50%); <i>Marah</i> (20%); <i>Sedih</i> (10%); <i>Takut</i> (10%); <i>Terkejut</i> (10%)
k-rank=250 & k=5	<i>Senang</i> (61.54%); <i>Marah</i> (30.77%); <i>Sedih</i> (15.38%)	<i>Marah</i> (22.22%); <i>Sedih</i> (33.33%); <i>Senang</i> (33.33%); <i>Takut</i> (11.11%); <i>Terkejut</i> (11.11%)	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (57.14%); <i>Takut</i> (28.57%)	<i>Marah</i> (33.33%); <i>Sedih</i> (16.66%); <i>Senang</i> (33.33%); <i>Takut</i> (33.33%); <i>Terkejut</i> (16.66%)	<i>Marah</i> (30%); <i>Sedih</i> (10%); <i>Senang</i> (50%); <i>Terkejut</i> (10%)
k-rank=250 & k=15	<i>Senang</i> (53.85%); <i>Marah</i> (30.77%); <i>Sedih</i> (13.85%)	<i>Marah</i> (22.22%); <i>Sedih</i> (11.11%); <i>Senang</i> (44.44%); <i>Takut</i> (11.11%); <i>Terkejut</i> (11.11%)	<i>Marah</i> (28.57%); <i>Sedih</i> (14.29%); <i>Senang</i> (57.14%); <i>Takut</i> (28.57%)	<i>Marah</i> (50%); <i>Sedih</i> (33%); <i>Senang</i> (33%)	<i>Marah</i> (40%); <i>Sedih</i> (10%); <i>Senang</i> (60%)
k-rank=250 & k=25	<i>Senang</i> (53.85%); <i>Marah</i> (38.46%); <i>Sedih</i> (23.08%)	<i>Marah</i> (44.44%); <i>Sedih</i> (22.22%); <i>Senang</i> (33.33%); <i>Takut</i> (22.22%)	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (57.14%); <i>Takut</i> (28.57%)	<i>Marah</i> (25%); <i>Sedih</i> (17 %); <i>Senang</i> (8%); <i>Takut</i> (16%); <i>Terkejut</i> (42%)	<i>Marah</i> (50%); <i>Sedih</i> (10%); <i>Senang</i> (40%)
k-rank=250 & k=35	<i>Senang</i> (53.85%); <i>Marah</i> (38.46%); <i>Sedih</i> (7.69%)	<i>Marah</i> (33.33%); <i>Sedih</i> (33.33%); <i>Senang</i> (22.22%); <i>Takut</i> (11.11%)	<i>Marah</i> (28.57%); <i>Sedih</i> (14.29%); <i>Senang</i> (28.57%); <i>Takut</i> (28.57%)	<i>Marah</i> (58%); <i>Sedih</i> (25%); <i>Senang</i> (25%); <i>Takut</i> (16%)	<i>Marah</i> (50%); <i>Senang</i> (50%)
k-rank=300 & k=5	<i>Senang</i> (53.85%); <i>Marah</i> (30.77%); <i>Sedih</i> (15.38%)	<i>Marah</i> (33.33%); <i>Sedih</i> (22.22%); <i>Senang</i> (33.33%); <i>Takut</i> (11.11%); <i>Terkejut</i> (22.22%)	<i>Marah</i> (28.57%); <i>Sedih</i> (14.29%); <i>Senang</i> (42.85%); <i>Takut</i> (42.85%)	<i>Marah</i> (42%); <i>Senang</i> (42%); <i>Takut</i> (16%)	<i>Marah</i> (40%); <i>Senang</i> (70%); <i>Terkejut</i> (10%)
k-rank=300 & k=15	<i>Senang</i> (53.85%); <i>Marah</i> (38.46%); <i>Sedih</i> (7.69%)	<i>Marah</i> (22.22%); <i>Sedih</i> (22.22%); <i>Senang</i> (33.33%); <i>Takut</i> (22.22%);	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (42.85%); <i>Takut</i> (28.57%)	<i>Marah</i> (50%); <i>Sedih</i> (25%); <i>Senang</i> (33%); <i>Takut</i> (8%); <i>Terkejut</i> (8%)	<i>Marah</i> (60%); <i>Sedih</i> (10%); <i>Senang</i> (40%)
k-rank=300 & k=25	<i>Senang</i> (53.85%); <i>Marah</i> (30.77%); <i>Sedih</i> (7.69%)	<i>Marah</i> (22.22%); <i>Sedih</i> (33.33%); <i>Senang</i> (33.33%); <i>Takut</i> (11.11%)	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (42.85%); <i>Takut</i> (28.57%)	<i>Marah</i> (41%); <i>Sedih</i> (25%); <i>Senang</i> (8%); <i>Takut</i> (25%)	<i>Marah</i> (50%); <i>Senang</i> (40%); <i>Terkejut</i> (30%)
k-rank=300 & k=35	<i>Senang</i> (46.15%); <i>Marah</i> (46.15%); <i>Sedih</i> (15.38%)	<i>Marah</i> (11.11%); <i>Sedih</i> (11.11%); <i>Senang</i> (33.33%); <i>Takut</i> (44.44%)	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (42.85%); <i>Takut</i> (28.57%); <i>Terkejut</i> (14.29%)	<i>Marah</i> (42%); <i>Sedih</i> (33%); <i>Takut</i> (16%); <i>Terkejut</i> (8%)	<i>Marah</i> (40%); <i>Sedih</i> (10%); <i>Senang</i> (50%)
k-rank=350 & k=5	<i>Senang</i> (53.85%); <i>Marah</i> (23.08%); <i>Sedih</i> (15.38%); <i>Takut</i> (15.38%)	<i>Marah</i> (22.22%); <i>Sedih</i> (33.33%); <i>Senang</i> (33.33%); <i>Takut</i> (55.55%)	<i>Marah</i> (14.29%); <i>Sedih</i> (42.86%); <i>Senang</i> (42.86%); <i>Takut</i> (42.86%)	<i>Marah</i> (25%); <i>Sedih</i> (33%); <i>Senang</i> (42%); <i>Takut</i> (33%)	<i>Marah</i> (40%); <i>Senang</i> (40%); <i>Takut</i> (10%); <i>Terkejut</i> (10%)
k-rank=350 & k=15	<i>Senang</i> (53.85%); <i>Takut</i> (30.77%); <i>Marah</i> (23.08%)	<i>Sedih</i> (44.44%); <i>Senang</i> (33.33%); <i>Takut</i> (22.22%);	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (57.71%); <i>Takut</i> (28.58%)	<i>Marah</i> (58%); <i>Sedih</i> (16%); <i>Senang</i> (25%); <i>Takut</i> (16%)	<i>Marah</i> (60%); <i>Sedih</i> (20%); <i>Senang</i> (50%)

k-rank=350 & k=25	<i>Marah</i> (69.23%); <i>Senang</i> (46.15%)	<i>Marah</i> (44.44%); <i>Sedih</i> (11.11%); <i>Senang</i> (33.33%); <i>Takut</i> (11.11%)	<i>Marah</i> (42.86%); <i>Sedih</i> (42.86%); <i>Takut</i> (14.29%)	<i>Marah</i> (67%); <i>Sedih</i> (25%); <i>Senang</i> (17%)	<i>Marah</i> (70%); <i>Sedih</i> (10%); <i>Senang</i> (30%); <i>Takut</i> (20%)
k-rank=350 & k=35	<i>Senang</i> (30.77%); <i>Marah</i> (84.6%)	<i>Marah</i> (55.55%); <i>Sedih</i> (11.11%); <i>Senang</i> (33.33%); <i>Takut</i> (11.11%)	<i>Marah</i> (14.29%); <i>Sedih</i> (28.58%); <i>Senang</i> (42.86%); <i>Takut</i> (14.29%)	<i>Marah</i> (50%); <i>Sedih</i> (17%); <i>Senang</i> (17%); <i>Takut</i> (25%)	<i>Marah</i> (50%); <i>Sedih</i> (10%); <i>Senang</i> (40%); <i>Terkejut</i> (10%)
k-rank=400 & k=5	<i>Senang</i> (84.62%); <i>Takut</i> (23.07%); <i>Marah</i> (15.38%)	<i>Sedih</i> (22.22%); <i>Senang</i> (33.33%); <i>Takut</i> (11.11); <i>Terkejut</i> (66.66%)	<i>Marah</i> (28.57%); <i>Sedih</i> (14.29%); <i>Senang</i> (57.14%); <i>Takut</i> (28.57%)	<i>Marah</i> (33%); <i>Sedih</i> (8%); <i>Senang</i> (33%); <i>Takut</i> (25%); <i>Terkejut</i> (25%)	<i>Marah</i> (40%); <i>Senang</i> (70%);
k-rank=400 & k=15	<i>Senang</i> (53.85%); <i>Takut</i> (30.77%); <i>Marah</i> (23.08%)	<i>Marah</i> (22.22%); <i>Sedih</i> (33.33%); <i>Senang</i> (33.33%); <i>Takut</i> (22.22%)	<i>Takut</i> (100%)	<i>Marah</i> (50%); <i>Sedih</i> (33%); <i>Senang</i> (25%); <i>Takut</i> (8%)	<i>Sedih</i> (60%); <i>Senang</i> (10%); <i>Takut</i> (40%); <i>Terkejut</i> (20%)
k-rank=400 & k=25	<i>Marah</i> (53.85%); <i>Senang</i> (38.46%); <i>Takut</i> (38.46%); <i>Sedih</i> (15.38%)	<i>Marah</i> (44.44%); <i>Sedih</i> (44.44%); <i>Senang</i> (22.22%); <i>Takut</i> (11.11%)	<i>Marah</i> (14.29%); <i>Sedih</i> (14.29%); <i>Senang</i> (42.57%); <i>Takut</i> (28.57%)	<i>Marah</i> (50%); <i>Sedih</i> (8%); <i>Senang</i> (25%); <i>Takut</i> (8%); <i>Terkejut</i> (8%)	<i>Marah</i> (40%); <i>Sedih</i> (40%); <i>Senang</i> (10%); <i>Takut</i> (80%); <i>Terkejut</i> (40%)
k-rank=400 & k=35	<i>Takut</i> (61.54%); <i>Sedih</i> (30.77%); (23.08%)	<i>Marah</i> (77.77%); <i>Sedih</i> (11.11%); <i>Senang</i> (11.11%); <i>Takut</i> (11.11%)	<i>Marah</i> (42.57%); <i>Sedih</i> (42.57%); <i>Takut</i> (28.57%); <i>Terkejut</i> (14.29%)	<i>Marah</i> (42%); <i>Sedih</i> (33%); <i>Senang</i> (17%); <i>Takut</i> (8%)	<i>Marah</i> (10%); <i>Sedih</i> (70%); <i>Senang</i> (10%); <i>Takut</i> (10%)

In the third user, k -rank 300 and $k=15$ can't retrieve the *terkejut* emotion as in ground truth, however it has the same percentage value in *senang* emotion, i.e. 42.85%. It is also happened when use k -rank=300 and $k=25$, which has the same value in each. The *terkejut* emotion comes when k -rank=300 and $k=35$, and k -rank=400 and $k=35$, however it can't represent the percentage of emotion near ground truth, when *senang* 42.86%, *sedih* and *terkejut* is 28.75%. The average precision, recall, and f-measure in each emotion is 40%, 3%, and 6.8% respectively for *senang*, for *sedih* emotion gives 30%, 2%, and 4.48% respectively, and the rest emotion is *terkejut* gives 13%, 1.56%, and 2.78%.

The fourth user can't retrieve well, mostly missed of retrieved information. The *marah* emotion has detected as the highest score of percentage, whereas the ground truth, the highest score is happened in *sedih* emotion. In the last user using k -rank 300 and $k=35$, the result of retrieved value near ground truth, but fear and angry can't be retrieved. The value of average precision, recall, and f-measure in each emotion is not near ground truth. Therefore, it influences the performance of giving percentage in emotion's user that also not near ground truth.

IV. CONCLUSIONS

The proposed method is sufficient to overcome the problem of user's emotion detection, which is proved by the result of the percentage emotions and the average precision, recall, and f-measure in each, near ground truth. The optimal values of k -rank and k is different in each user, it is based on the validity averagevalue of precision, recall, and f-measure in each emotion. Mostly, the best

result of emotion detection performance when k -rank=250 and 300 under $k=15$. The invalid value comes when using k -rank and using k too big or too small, because it will produce noise information retrieved that not related to the user's tweets query.

Further research will include hashtag with one or more words, as known as phrase. In ordinary phrase, in formal language still produce term, which is easily understood. Nevertheless, using hashtag in Twitter is not allowed to use space character among terms. So, it open problem, which is possible to determine a phrase correctly.

REFERENCES

- [1] F. Atefeh, W. Khreich. A Survey of Techniques for Event Detection in Twitter. Computational Intelligence, September, 2013.
- [2] S. M. Mohammad. #Emotional Tweets. Proceedings of the SemEval12, 2012, 1.
- [3] W. Wang, L. Chen, K. Thirunarayan et al. Harnessing Twitter 'Big Data' for Automatic Emotion Identification. Proceedings of SocialCom12, 2012, pp. 587-592.
- [4] M. Purver, S. Battersby. Experimenting with Distant Supervision for Emotion Classification. Proceedings of EACL '12, 2012, pp. 482-491.
- [5] Z. Yuan, M. Purver. Predicting Emotion Label for Chinese Microblog Texts. Proceedings of SDAD 2012, 2012, pp. 40.
- [6] A. Neviarouskaya, H. Predinger, M. Ishizuka. Textual Affect Sensing for Sociable and Expressive Online Communication. Proceedings of ACII'07, 2007, pp. 218-229.
- [7] Arifin, K. E. Purnama. Classification of Emotions in Indonesian Texts Using K-NN Method. International Journal of Information and Electronics Engineering, v2, n6, 2012, pp. 899-903.

- [8] J. Suttles, N. Ide. Distant supervision for emotion classification with discrete binary values. *Computational Linguistics and Intelligent Text Processing*, 2013, pp.121-136.
- [9] J. Asian, H. E. Williams, S. M. M. Tahaghoghi. Stemming Indonesian. *Proceedings of ACSC'05*, v38, 2005, 307-314.
- [10] A. Z. Arifin, R. Darwanto, D. A. Navastara, H. T. Ciptaningtyas. Klasifikasi Online Dokumen Berita Dengan Menggunakan Algoritma Suffix Tree Clustering. *Proceeding of SESINDO2008*. 2008. 17.
- [11] A. Z. Arifin, I. P. A. K. Mahendra, H. T. Ciptaningtyas. Enhanced confix stripping stemmer and ants algorithm for classifying news document in indonesian language. *Proceeding of ICTS*, 2009.
- [12] X. Yan, J. Guo, S. Liu et al. Learning topics in short texts by non-negative matrix factorization on term correlation matrix. *Proceedings of SDM13*, 2013.
- [13] A. Z. Arifin. Penggunaan Digital Tree Hibrida pada Aplikasi Information Retrieval untuk Dokumen Berita. *Prosiding Seminar Nasional Sains dan Teknologi*, 2002.
- [14] P. C. Barman, N. Iqbal, S. Y. Lee. Non-negative Matrix Factorization Based Text Mining: Feature Extraction and Classification. *Neural Information Processing*, 2006, pp. 703-712.
- [15] M. M. Rahman. Mining Social Data to Extract Intellectual Knowledge. *IJISA*, v4, n10, 2012, pp.15-24.
- [16] A. Z. Arifin, A. Rachmania, I. Lukmana. Topic Identification of Indonesian Language News Article Based on Theme Query. *IPTEK*, v1, n1, 2013.

(Unipdu), Jombang, East Java. Her research interest include computer vision, image processing, and data mining.

How to cite this paper: Agus Zainal Arifin, Yuita Arum Sari, Evy Kamilah Ratnasari, Siti Mutfrofin, "Emotion Detection of Tweets in Indonesian Language using Non-Negative Matrix Factorization", *International Journal of Intelligent Systems and Applications (IJISA)*, vol.6, no.9, pp.54-61, 2014. DOI: 10.5815/ijisa.2014.09.07

Authors' Profiles

Agus Zainal Arifin, male, received S.Kom degree from Informatics Engineering, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia received his M.Kom degree from University of Indonesia (UI), Jakarta, Indonesia and received Dr. degree in Hiroshima University, Japan. He is currently a dean of Informatics Engineering Department in ITS Surabaya. His research interest included computer vision, image processing, information retrieval for Indonesian or Arabic document, and computer aided diagnosis.

Yuita Arum Sari, female, received S.Kom degree from Computer Science at University of Brawijaya, in November 2011. She is currently a master student at Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Her research interest included data mining, text mining, computer vision, pattern recognition, and machine learning.

Evy Kamilah Ratnasari, female, received the S.Pd degree of Informatics Education from Universitas Negeri Malang, Indonesia, in 2011. She is currently a master student at Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Her research interest include data mining, computer vision, image processing and pattern recognition.

Siti Mutfrofin, female, received the S.Kom degree from Informatics Engineering at Trunojoyo University, Madura, Indonesia, in 2009. She is currently a master student at Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia and a lecturer of Informatics Engineering in Universitas Pesantren Tinggi Darul 'Ulum