

Measuring happiness in large population

Annabelle Wenas¹, Smita Sjahputri^{1,2}, Bagus Takwin², Alfindra Primaldhi², and Roby Muhamad^{1,2}

¹ProveticLab, Jl. Kerinci I No. 2, Jakarta, 12120 Indonesia

²Faculty of Psychology, University of Indonesia, Depok, 16424 Indonesia

Corresponding Author's E-mail: annabellewenas@provetic.com

Abstract. The ability to know emotional states for large number of people is important, for example, to ensure the effectiveness of public policies. In this study, we propose a measure of happiness that can be used in large scale population that is based on the analysis of Indonesian language lexicons. Here, we incorporate human assessment of Indonesian words, then quantify happiness on large-scale of texts gathered from twitter conversations. We used two psychological constructs to measure happiness: valence and arousal. We found that Indonesian words have tendency towards positive emotions. We also identified several happiness patterns during days of the week, hours of the day, and selected conversation topics.

1. Introduction

Governing complex modern societies requires some basic measurements in the societal level. These measurements will ensure that public policies are effective and meet the ever changing demand. However, currently, the most common aggregate measures of societies are economic measures such as economic growth. Yet, as modern societies grow more complex, there is a need to develop other measurements beyond economic measures especially for psychological measures that can capture subjective well-being [7]. It is reasonable to think that a combination of economic and psychological measures can provide more comprehensive view of a society which, in turn, will be useful for formulating better public policies and their evaluations. In this paper we propose an approach to measure psychological characteristics for large population that based on text data.

Our focus is on the measurement of emotional states and we follow Russell [9] who asserted that emotion, mood, and other emotionally charged events are states that are simply combination between feeling good or bad and energized or worn out. Russell addressed these emotional states as core affect, and mapped its structure into circumplex model [4-6, 8-10]. Horizontal axis of circumplex model is valence, which is a measure of emotion ranging from negative to positive emotions. Whereas its vertical axis is arousal, a measure of emotional intensity. Thus, for example, anger is a negative emotion with high intensity and lethargic is a negative emotion with low intensity. On the other end of the spectrum, excited and calm are positive emotions with high and low intensity respectively. Note that happiness is a positive emotion with moderate intensity.

Core affect is the fundamental of happiness measurement in present study. Moreover we propose a measure for happiness that can be used in large scale population that is based on the analysis of



Indonesian language lexicons. Similar works for the measurement of happiness for English and other languages have been done extensively [1-3], but their measure of happiness based solely on the valence dimension.

Here, in addition to measuring word valence, we also included the measurement of arousal dimension. The reason to include arousal is because positive valence is necessary but not sufficient component of happiness – Simply because there are other states that have positive valence like excitement and calm. Thus, level of arousal is the key to differentiate excitement, happiness, and calm; Three of them indeed have positive valence, yet their arousal level are variable from high, moderate, to low respectively. Hence, we propose a more comprehensive way to measure happiness, using combination of valence and arousal measurement.

2. Measures and Methods

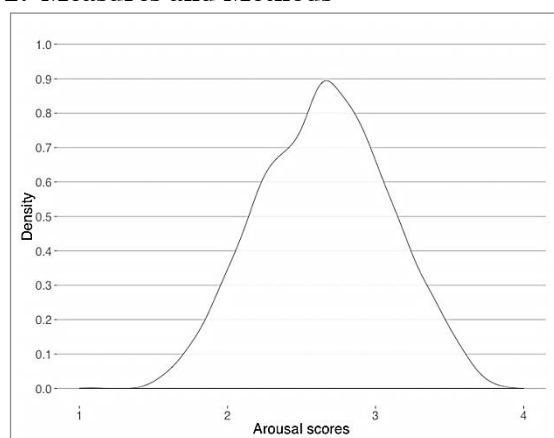


Figure 2. Density of arousal scores

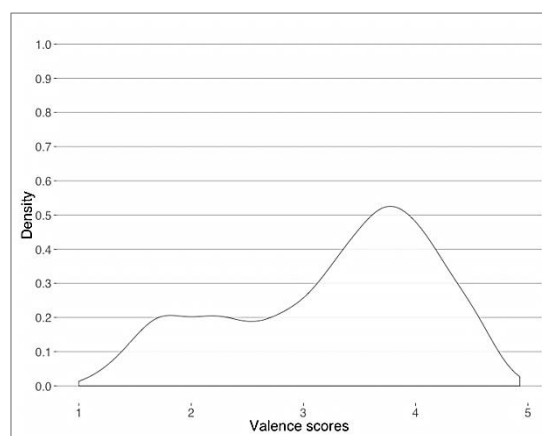


Figure 3. Density of valence scores

In order to explore arousal and valence contents in Indonesian language, we constructed a corpus consisting 3,000 Indonesian words. The source of corpus is the most frequently used words in twitter conversation [1–3] from 2010-2012. Then we measured how those words perceived by individuals in terms of arousal and valence. Therefore 83 students were recruited to rate each word – Rating process was administered using online panel. Arousal rating scale ranging from 1 (calm) to 4 (excited), whereas valence ranging from 1 (negative) to 5 (positive), both with 1 incremental point. Average rating scores then calculated for each word, see figure 2 and 3 for the results.

In terms of arousal, rating average is near symmetrically distributed. Whereas for valence, rating average is negatively skewed. This valence result is consistent with study which reveals that human language has positivity bias [3]. Further, we found that there is no significant correlation between valence and arousal ($R^2 = 0.028$, $F(1, 2981) = 86.89$, $p < .001$). This result is consistent with Russell's core affect theory [7-8], in which valence and arousal dimensions are orthogonal to each other.

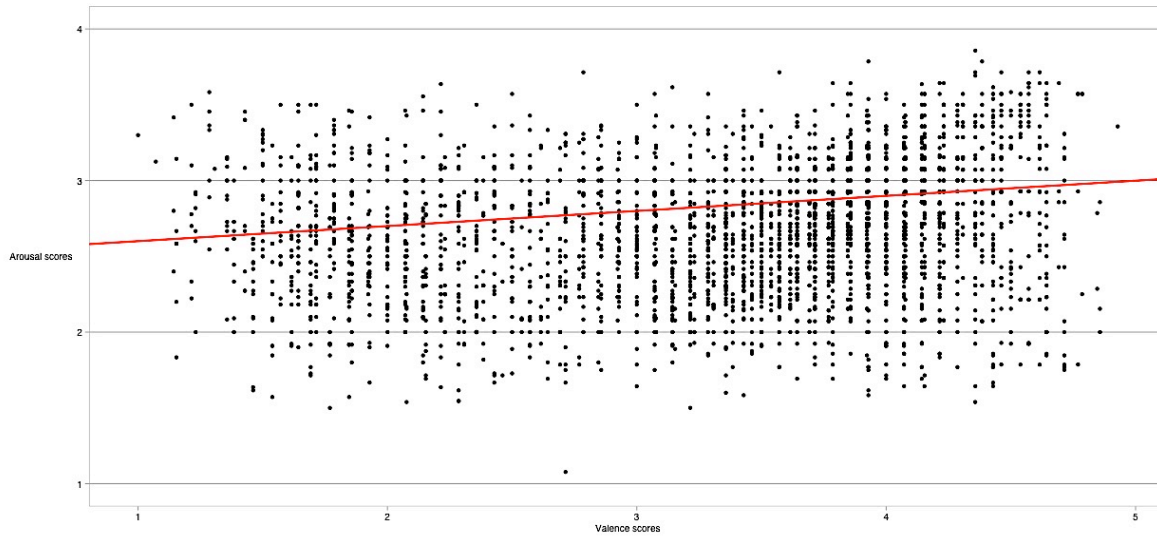


Figure 4. Valence and arousal scores distribution

From figure 5, notice that word with highest arousal is “*spektakuler*” (spectacular), and with lowest arousal is “*sunyi*” (silence). Whereas word with most positive valence is “*optimistis*” (optimistic), and with most negative valence is “*membunuh*” (murder). Thus in a sense, structural consistency with circumplex model is evident in this result.

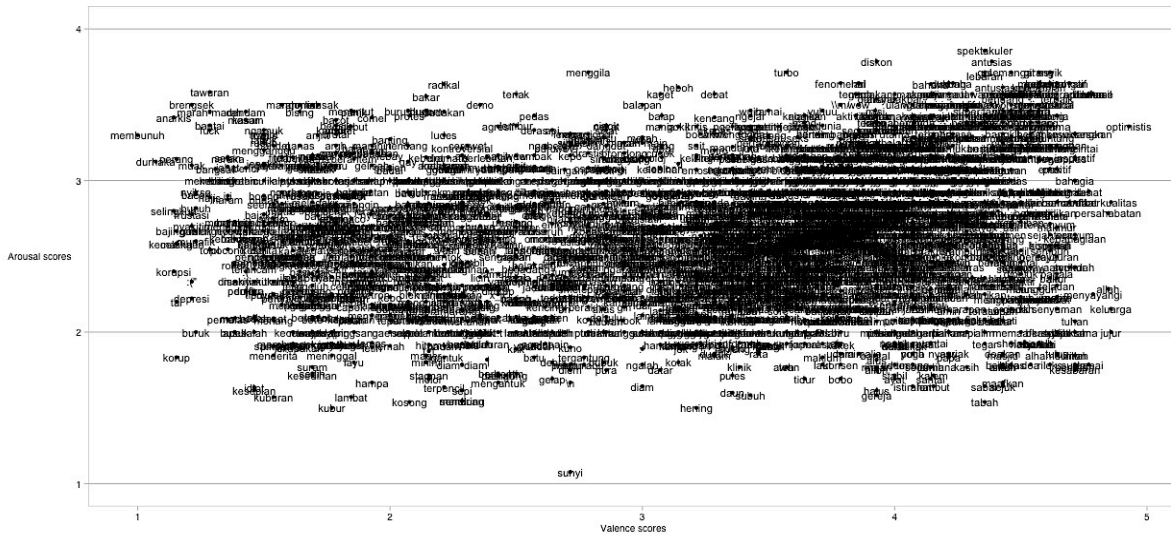


Figure 5. Distribution of words based on valence and arousal scores

Valence text score estimation [1] used to estimate both arousal (a_{text}) and valence (v_{text}) score for a text, which in this case tweets. We identify frequency of word(s) from corpus that mentioned in the text, and compute a weighted average of the arousal and valence of words as

$$a_{text} = \frac{\sum_{i=1}^n a_i f_i}{\sum_{i=1}^n f_i} \quad (1)$$

$$v_{text} = \frac{\sum_{i=1}^n v_i f_i}{\sum_{i=1}^n f_i} \quad (2)$$

Note that a_i and v_i are recorded average arousal and valence, respectively, from every word of our corpus (i). Moreover f_i is lexicon i frequency in text. As illustration let's take this tweet “@MerryRiana: Mungkinkah di balik tembok keputusan yang sedang menghadang Anda terletak sebuah jalan besar menuju kesuksesan? #MR” The seven underlined words recorded in corpus with average valences of 2.50 (under), 2.07 (wall), 2.07 (something), 2.58 (road), 2.93 (big), 2.66 (toward), and 3.30 (success) respectively. Overall valence score for the tweet therefore would be

$$v_{text} = \frac{1}{7} (1 \times 2.50 + 1 \times 2.07 + 1 \times 2.07 + 1 \times 2.58 + 1 \times 2.93 + 1 \times 2.66 + 1 \times 3.30) \approx 2.6$$

From calculation above we found that tweet's valence text score falls in 2.6 from the scale of 1 to 5, making it moderately positive.

3. Results

Using this method, we calculated arousal and valence scores for various topics in public debates and also how it vary within temporal and topical dimensions. We randomly collected tweets and analyze it within three contexts, which are: day of the week, hour of the day, and topic (see Table 1).

Table 1. Explanation of data contexts

Context	Number of Variables	Number of Tweets
Day of the week	7	1,513,085
Hour of the day	24	1,564,505
Topic	8	738,967

3.1. Day of The Week

We collected data from Indonesia's local twitter conversations in 2014. In order to overcome limitations of the server, hence randomly chosen 1,200 timestamps points for every day of the week (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday). Suppose all Mondays in 2014 had a total of 74,880 hours, of all randomly chosen 1,200 hours; We repeat this procedure for the rest of the days. Then from every point of these days, selected 200 tweets. In total, 1,513,085 tweets have successfully collected.

As we can see in figure 6, conversation on Thursday has the highest arousal, and immediately followed with lowest arousal on Friday. Whereas for valence, Monday has the highest valence compared to the rest of days. This is contrary to popular belief “Monday blues”, which says back to work day is identical to sadness. Although one thing to keep in mind is Eid al-Fitr was on Monday (July 28th, 2014); The day when social media crowded with positive words of greetings, like “*selamat*” or congratulation ($v_{selamat} = 4.4$). This fact might explain why Monday has the very high positive valence compared to the other days' measurement. Monday's moderate arousal and high valence indicate happiness is evident.

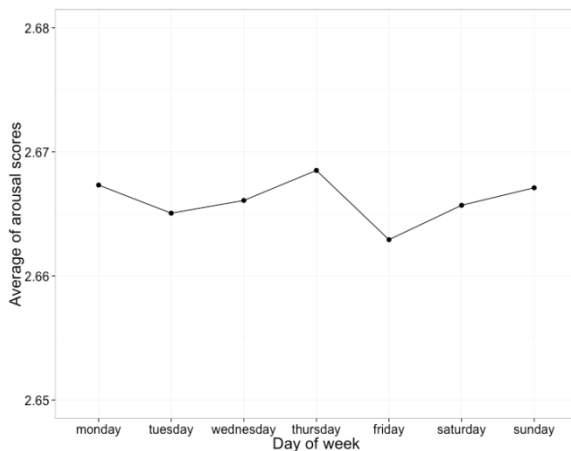


Figure 6. Arousal scores distribution per day of the week

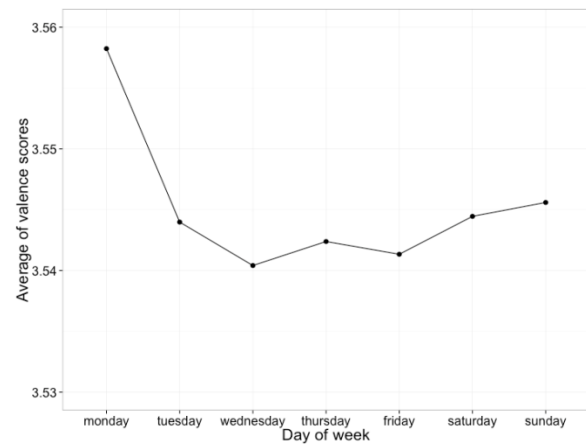


Figure 7. Valence scores distribution per day of the week

3.2. Hour of The Day

For the hour of the day, we use the same method as for the days, yet this time we chose 365 time stamps points for each hour. Therefore total points from all hours are 8,760. Further, from each point of these hours selected 200 tweets. Total tweets that we have collected are 1,564,505.

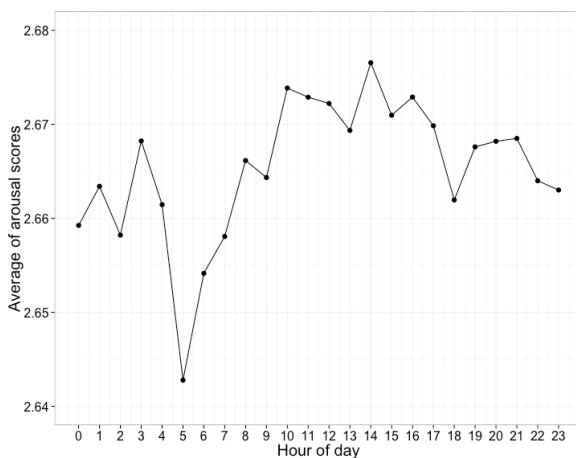


Figure 8. Arousal scores distribution per hour the of day

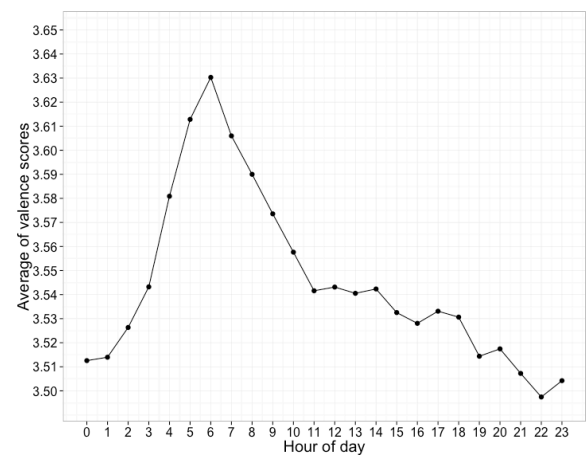


Figure 9. Valence scores distribution per hour of the day

Variations are more pronounced for both arousal and valence in hour of the day compared to previous analysis for day of the week. Lowest arousal hour observed on 5 am, whereas highest peak of arousal hour on 14 pm. Moreover 6 am is the hour with highest valence, the peak is constantly decreased until reaches its lowest on 10 pm. Apparently on early hours of the day, 6 am to 8 am, moderate arousal and positive words are frequently used in social media. This finding indicates that upon those hours happiness is most strongly present. Further, there is a unique reversed pattern of data points of arousal and valence. Arousal significantly decreased from 4 am and elevated back on 6 am, whereas valence significantly increased from 2 am and shifts to decline on 7 am.

3.3. Topic

Next, we selected various twitter conversations and analyze their valence and arousal scores. We chose eight of which to compare arousal and valence of each topic. Selected topics along with further details explained in Table 2 below. From each topic we randomly chose 2,000 timestamps points, and from each of these points selected 50 tweets. Total tweets that have been successfully collected are 738,967.

Table 2. Explanation of topical dimensions

Topic	Category	Number of Tweets
Ahok (Basuki T. Purnama)	Local Leader	93,409
Ridwan Kamil	Local Leader	87,536
Banjir (<i>Flood</i>)	City Issue	92,032
Macet (<i>Traffic jam</i>)	City Issue	95,811
Golkar	Political Party	87,394
PDIP	Political Party	90,625
McDonald's	Fast Food Brand	96,524
KFC	Fast Food Brand	95,636

The order of topic in figure 10 and figure 11 below is based on the average valence score from the highest to the lowest. Thus, we see that Ridwan Kamil (Mayor of Bandung) has the highest valence, whilst both flood and traffic jam are identically have the lowest valence. Moreover both political parties have the highest arousal, with Golkar has the highest arousal level, and presumably moderate level of valence.

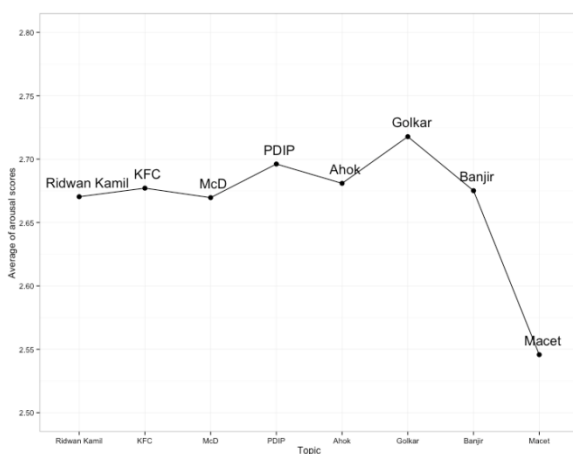


Figure 10. Arousal scores distribution per topic

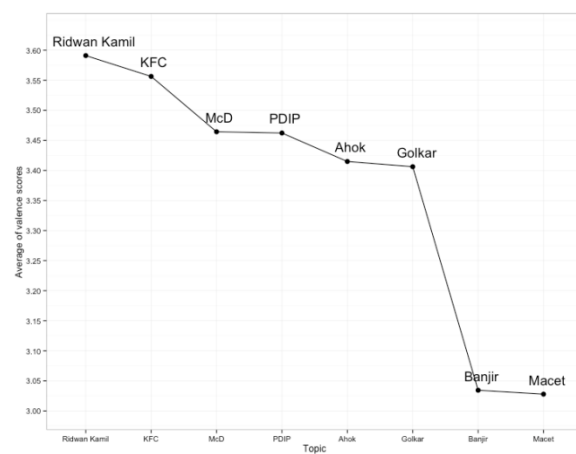


Figure 11. Valence scores distribution per topic

Same lowest valence observed for both city issues, although there is large gap between their arousal average scores. Flood has the third highest arousal, whereas traffic jam is the eight or lowest among all. KFC and McDonald's are the two most popular fast food brands in Indonesia, have almost similar level of arousal, yet KFC has higher valence compared to McDonald's. Ahok (Governor of Jakarta) has slightly higher arousal compared to Ridwan Kamil (Mayor of Bandung). However, Ridwan Kamil clearly dominates valence chart, not only upon Ahok but also

among any other topic. This finding indicates that Ridwan Kamil and KFC are the happiest topics compared to others.

4. Discussion

Core affect mapped in circumplex model has opened a new direction to the study of emotions [4-6, 8-10]. For instance to research about happiness as presented in this paper. Whereas most study measure happiness only through valence component [1-3] here we also measure arousal; Because positive valence isn't always equal to happiness, and vice versa on negative valence. Arousal measurement serves as indicator of emotional intensity, providing more comprehensive insight. As explained in introduction, positive valence with high arousal is excitement, moderate arousal is happiness, and low arousal is calm.

Inspired by the study initiated by Dodds [1-3], we constructed corpus comprising 3,000 Indonesian words, each has rating scores of arousal (on the scale of 1-4), and valence (on the scale of 1-5). We found that there is no significant correlation between valence and arousal ($R^2 = 0.028$, $F(1, 2981) = 86.89$, $p < .001$). As seen in figure 5, emotional states are present across the combination of arousal and valence level [8-9]. Our particular finding might be useful as new method in the future to label unknown emotional states. This proposed method isn't limited to social media's corpora, but any text source like speech and song lyric. More integrated corpora, which combine various text sources would be ideal.

Moreover, valence scores are negatively skewed toward positive range. Whereas, arousal is symmetrically distributed with dominance on moderate range of the score 2 and 3. Therefore in accordance to finding from study by Dodds et al. and Pollyanna hypothesis [3], tendency toward positivity bias is also present in corpus used in this study. Note that, our result shows that words used in social media conversation (in this case twitter) are indeed biased toward positive emotion and contradict the assumption that conversations in social media are dominated by negative expression.

We have calculated arousal and valence scores for various public conversations in Indonesian language and from this nation's spatial location, within temporal and topical dimensions. We randomly collected tweets and analyze it within three contexts, which are: day of the week, hour of the day, and conversation topics. Using combination of arousal and valence average scores we indicate that several data points as the happiest among others. First, from day of the week context, Monday is presumably has the highest level of happiness. Second, from hour of the week context, 6 am to 8 am are indicated of having highest level of happiness. Lastly, Ridwan Kamil (Mayor of Bandung) and fast food brand KFC, are the first and second most topics with highest level of happiness. Notice that we reduce score range up close to enhance differences between points.

We conclude by noting that our approach has the potential to be used as a measure of emotions for large population in multitude domains. Further development of this approach will include tests for sensitivity and robustness, and also the inclusion of other psychological measures such as moral judgments, values, and personality.

Keywords: arousal, computational psychology, emotion, happiness, hedonometer, psychology, valence

Acknowledgment

We thank Judotens Budiarto, Muhammad Robee, and Heri Purnomo who have developed the technology for data collection and analysis. We are also grateful to Jehan Amanda, Steffi Hartanto, Prasakti Tenri, and Karina Suhenda who have helped collecting the data and performing initial data analysis.

References

- [1] Dodds, P. S., & Danforth, C. M. (2009). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4), 441-456.
- [2] Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one*, 6(12), e26752.
- [3] Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K.D., Kloumann, I. M., Bagrow, J. P., Megerdoomian, K., McMahon, M. T., Tivnan, B. F., & Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, 112(8), 2389-2394.
- [4] Feldman, L. A. (1995a). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology*, 69(1), 153.
- [5] Feldman, L. A. (1995b). Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, 21(8), 806-817.
- [6] Gregg, V. H., & Shepherd, A. J. (2008). Factor structure of scores on the state version of the four dimension mood scale. *Educational and Psychological Measurement*.
- [7] Helliwell, J. F., Layard, R., & Sachs, J. (2015). World happiness report 2015. New York: *Sustainable Development Solutions Network*.
- [8] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161.
- [9] Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1), 145.
- [10] Widhiarso, W., & Prawitasari, J. E. (2010). Struktur Semantik Kata Emosi dalam Bahasa Indonesia. *Jurnal Psikologi*, 37(2), 153-164.