



Multiview Sentiment Analysis with Image-Text-Concept Features of Indonesian Social Media Posts

Esther Irawati Setiawan^{1*} Hans Juwiantho¹ Joan Santoso¹ Surya Sumpeno^{2,3,4}
 Kimiya Fujisawa⁵ Mauridhi Hery Purnomo^{2,3,4}

¹Department of Information Technology, Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

²Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Indonesia

³Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Indonesia

⁴The Science and Technology Center of Artificial Intelligence for Healthcare and Society (PUI AIHeS), Indonesia

⁵School of Media Science, Tokyo University of Technology, Tokyo, Japan

* Corresponding author's Email: esther@istts.ac.id

Abstract: Social media development makes it possible for everyone to express their opinions and information through text, speech, video, or images. Multiview sentiment analysis in current studies generally combines two modalities, text and image. It seeks to classify social media posts into two or more polarities, such as positive, neutral, or negative. To improve the performance of multiview Sentiment Analysis, we added another modality, which is concepts derived from text and image. Our proposed model integrates three views into a fusion with an ensemble approach by a meta-classifier. We performed text classification with Deep Convolutional Neural Networks. The input feature is Word2Vec for text representation in order to preserve semantic meaning. Additionally, we analyzed concepts from texts with SenticNet 5 as a knowledge base model and extracted concepts from images using the DeepSentiBank model. We obtained 2089 Adjective Noun Pairs and classified it with Multi-Layer Perceptron. Then we combined predicted probabilities from each classifier for Image, Text, and Concept by Ensemble Learning. A meta-classifier was implemented to predict the final sentiment from a fusion of Image-Text-Concept features. The fusion for multiview sentiment analysis works well and could achieve the best accuracy of 70% by applying the ensemble approach with Logistic Regression as the meta-classifier.

Keywords: Sentiment analysis, Multimedia computation, Fusion model, Ensemble learning, Social media analysis.

1. Introduction

There has been growing public and corporate interest of social networking sites over the last few years. Social media analysis can be developed for many applications, such as health care, finance, voting intentions, security and defense, disaster response, NLP-based modeling, NLP-based social media visualization, entertainment, and media monitoring [1]. Studies in social media have been developed on various modalities such as text [2-5], image [6], video [7], and network graph [8].

Unstructured Big Data content is rapidly growing in social media [9]. Previously text-based, nowadays, social media posts contain images and videos. Aside

from text, users may also post images to share their life experiences as well as their opinions. Since Twitter recommended users to post images for better social media engagement [10], users have actively posted images to accompany text-based posts. In 2012, there are 400 million tweets per day, which are 36% of users posted images, which is the largest type of multimedia shared [11], and image posted is more likely to be retweeted [12].

A user on social media can post photos of themselves celebrating New Year and caption it with their comment regarding their feeling about the celebration. The text, as well as facial expressions, may indicate their emotion during the celebration [13]. Therefore, image and text should be viewed together since these two domains complement each other [14].



Figure. 1 Example of Image and text tweet

Users also express their sentiment in social media posts through emoticons and videos. Sometimes the processing of text alone is not enough due to the ambiguity of the sentence. Thus the use of images can help analyze sentiment.

Fig. 1 gives an example of two posts where image and text are clarified by each other. The top image shows three ladies with various expressions in which the text in English means “the script creator did not think hence causing army name bad in elf eyes”. Therefore, we can conclude that it is a negative post. The second image also shows football activities, which is accompanied by text that expresses negativity “This tactical analysis shows how France killed Argentina’s high line structure”.

Sentiment analysis and affective computing have become one of the main areas of research in social media [15]. Sentiment analysis includes an analysis of the positive or negative tendency of a user towards a product, brand, or company, and an analysis of

social roles, popularity, or values in social networks. Sentiment analysis classifies user opinions into two binary polarities, which are positive and negative. However, Twitter posts also contain emotionless posts which needs to be classified as neutral. Other analyzes of sentiment are classified into numeric rating [16]. Furthermore, given that Twitter posts contain an average of 33 characters [17], it is challenging to acquire a finer granularity such as obtaining a numeric rating. The most common method used is supervised learning with different types of features.

Twitter is a popular social media in Indonesia, and in 2017 Indonesia is in the top five Twitter users in the world [18]. In the same year, more than 143 million Indonesians use the internet, and approximately 90 percent of these people are Twitter, Facebook, and Instagram users. Hence, Tweets are compact and reliable data for sentiment analysis in Indonesian [19]. Not only events and news, but also emotional expressions are shown in social media posts. This gives researchers an opportunity to analyze the behavior of users in Indonesia. Industry and event organizers also need to analyze public opinion. As a result, a variety of studies are carried out on predictive sentiment analysis.

Social Media Analysis in Indonesian and English languages face similar challenges, though there are differences in their grammatical structure. By machine learning, approach in English Sentiment Analysis could also be applied for Indonesian, such as the use of external knowledge [20]. The need for external knowledge is due to Indonesian as a low resource language. We also need to address concerns in Indonesian social media posts such as entity extraction [21], informal language [20], mixed language, slang words, inappropriate abbreviations, and sarcasm. Classical machine learning methods in Sentiment Analysis require the extraction of hand-crafted features. However, Deep Learning approaches perform end-to-end feature representation and classification together [22]. There are also still opportunities to develop Indonesian Image-Text Sentiment Analysis.

In addition to analyzing text-based user posts, image and social networks need to be analyzed in a comprehensive study to understand sentiments of social media users. Text and images are posted together, more than other modalities, so we use text and images in this study. Several researchers have presented sentiment analysis based on image-text multiview features [23]. [24] shown in their paper that image and text in social media have a correlation and complemented each other, and display emotional relevance. Moreover, studies in Indonesian

Sentiment Analysis that combine features from image and text are rarely seen. In conjunction with the image, text in social media posts could express the user's feelings [25].

On the other hand, Image Classification usually detects objects from an image, such as an image of the sky or a dog. However, Image Sentiment Analysis should extract adjectives correlated with images together with nouns for obtaining sentiment strength. Thus, concept which is defined as Adjective Noun Pairs could provide semantic representation for obtaining better sentiment from an image. A pre-trained model called DeepSentiBank [26] is used in our research, representing the concept features. Resources in Indonesian languages for obtaining concepts from images is limited. Therefore, we proposed an approach that could be adopted in Indonesian Sentiment Analysis with English based DeepSentiBank.

Unlike conventional text sentiment analysis that uses bag of words with a co-occurrence count, we need to analyze words based on commonsense concepts. Implicit expressions need to be analyzed in relation to concepts. Therefore, we use external knowledge to enhance text sentiment analysis with SenticNet 5 [27]. We use a localized version of SenticNet, which was developed especially for the English Sentiment Analysis. [28].

The main contributions of this work can be summarized as follows:

- A novel framework for detecting sentiment from microblog posts by fusion of text, image, and concept features. Text posts were classified with Convolutional Neural Networks. Concepts from the text are extracted with a concept ontology, which is SenticNet 5. Moreover, concept was extracted from images with DeepSentiBank pre-trained model, which is a model based on Deep Convolutional Neural Networks.
- Combining classifier results with Stacking Based Ensemble Learning to balance each model's performance with a Meta-Classifer. Ensemble Learning is believed to yield better performance since there is a diversity between models for Image, Text, and Concept Fusion [29].
- We also created an Indonesian dataset for Image-Text Sentiment Analysis from Social Media.
- To the best of our knowledge, we are the first to perform multiview sentiment analysis, which includes a fusion of image, text, and concept in Indonesian. We believe our ensemble model of Image-Text-Concept Fusion would advance the research of Sentiment Analysis, especially in Indonesian.

The subsequent section reviews the related works of Image-Text Sentiment Analysis. Then, the Ensemble Learning of image-text-concept fusion Sentiment Analysis Framework is presented in Section 3. Section 4 provides a description of the Image-Text dataset for Indonesian Sentiment Analysis. The results of experiments with analysis are reported in Section 5. Finally, the last section describes the conclusion and future works.

2. Related works on image-text sentiment analysis

As explained in the previous section, researchers have focused on the study of text-based sentiment analysis to develop frameworks for predicting customer opinion, political elections, measuring economic indicators, and so on. In addition to text, social media users express their opinions using images and videos. Sentiment analysis on large-scale textual and visual posts can help to improve on obtaining user sentiments toward events or topics.

2.1 Text-based sentiment analysis

Several studies have been reported on text sentiment analysis using classical machine learning methods such as Bayesian Network [30], Multi Layer Perceptron [31], and Logistic Regression [32]. However, there have been numerous studies that use deep learning for sentiment analysis. Particularly with Convolutional Neural Networks as in [33] and [34], which achieved promising results compared to conventional methods.

Sentiment Analysis for English has used Convolutional Neural Networks in a study by [35]. Experiments were carried out on four types of CNN. The highest accuracy is achieved on the non-static CNN model. This model uses Word2Vec as an initialization word in vector.

Further research on deep learning for sentiment analysis is done by comparing the Long Short Term Memory (LSTM) and Dynamic Convolutional Neural Network (DCNN) algorithms with classical machine learning, i.e., Gaussian Naive Bayes (NB), Support Vector Machine (SVM), and Maximum Entropy (ME) in Thai with data obtained from Twitter [36]. The results of the experiments presented in the paper show that deep learning algorithms are superior to classical machine learning with the best accuracy rate of 75.35% by the DCNN method.

Sentiment Analysis features vary from Bag-Of-Words to concepts. Research in [27] applies named entities with concepts as features, which is called SenticNet 5. Several studies have been developed

with SenticNet 5, such as the creation of Commonsense Ontology [37] and the addition of SenticNet to LSTM cells [38] for Targeted Aspect Based Sentiment Analysis.

The Indonesian language belongs to the Austronesian language family, which is a Western Malayo-Polynesian language. It is the national language of different ethnic citizens. There is a variety of Indonesian languages due to the social linguistic environment in this country. Education, mass media, government, and religion are using a more formal language. On the contrary, informal language is used in everyday language and social media posts. The existence of abbreviation, foreign words, emoji, emoticons, abbreviations in social media posts needs special treatment in Sentiment Analysis [20].

Indonesian and English use Latin-based characters. However, the grammatical structure of both languages is quite different. Sentiment analysis in English takes into account the grammatical structure that has been proposed with Tree-structured Recursive Neural Networks [14]. Although this approach is rarely seen in Indonesian, due to limited language resources.

Sentiment Analysis research in Indonesian is mostly based on classical machine learning methods, such as Naive Bayes (NB) [39], Support Vector Machine (SVM) [40], Decision Tree [41], and Maximum Entropy (ME) [42] which is based on a bag-of-words model that does not concern the order of a word in a sentence. On the other hand, deep learning algorithms take into account the order of words in a sentence. However, it is only in the past few years that deep learning for Indonesian sentiment analysis has begun to be proposed. One of the early studies is text-based sentiment analysis with Convolutional Neural Networks[20]. Deep Learning for Indonesian text needs an adequate amount of data and external knowledge to improve the performance of the models. Tools and resources are currently limited in Indonesian. Some of the external knowledge available is translated from English resources.

Studies conducted in Indonesian sentiment analysis uses methods of deep neural networks in [4, 43, 44]. It was found that using a word embedding to create a similarity matrix could increase accuracy. The results of the deep neural network outperformed classical machine learning methods. Moreover, syntactic analysis is not performed in this study. We used a neural language-based word representation approach, whereas a neural language model is formed unsupervisedly from a collection of documents.

2.2 Image-based sentiment analysis

Studies on Image Sentiment Analysis are frequently performed by extracting low-Level features to be used on mid-level attributes and by detecting facial expression [45]. Linear SVM and Logistic Regression are used to perform the Image Sentiment Analysis classification process.

Another study reported by [46] processes Image Sentiment Analysis using a YouTube videos dataset with content videos of reviews or product recommendations. Each video showed an object in the form of a person speaking, and the process of sentiment analysis is done by processing facial characteristic points on the face and results in facial features to determine sentiment.

Recent works use Deep Learning, especially Convolutional Neural Networks for image sentiment analysis, which uses mid-level features of an image together with SentiBank [47], VGG16 [48], and DeepSentiBank based predictor [26]. Our study did not detect facial expressions or obtain facial features. On the contrary, we extract the concepts contained in the image due to social media posts images that do not always contain a display of the facial image.

2.3 Fusion of image-text for sentiment analysis

Recent reports on machine learning techniques of combining image and text features for sentiment analysis have been published. The authors of the paper [49] state that the number of people combining texts and images in social media to express their opinions is increasing, and therefore multiview-based visual sentiment analysis is emerging. In this analysis, a mid-level representation called Adjective Noun Pronoun is used with the following features: word for text, ANP for images, and emoticon symbols.

A previous study by [50] combined text in conjunction with image features using Stacking Based Ensemble Learning. The ensemble classifier was Random Forest on a publicly available MVSA dataset [51] in English. Another approach is sentiment analysis on the transportation domain described in article [52]. The study in the article uses Instagram's captions, images, and comments. The framework uses three supervised learning algorithms, which are SVM, RF, and NB classifiers. NB uses the Bayes theorem and assume the independence of the features.

One of the deep learning approaches has been reported by the authors of the article [13]. They propose a cross-modality consistent regression model that could combine both the state-of-the-art techniques for visual and textual sentiment analysis.

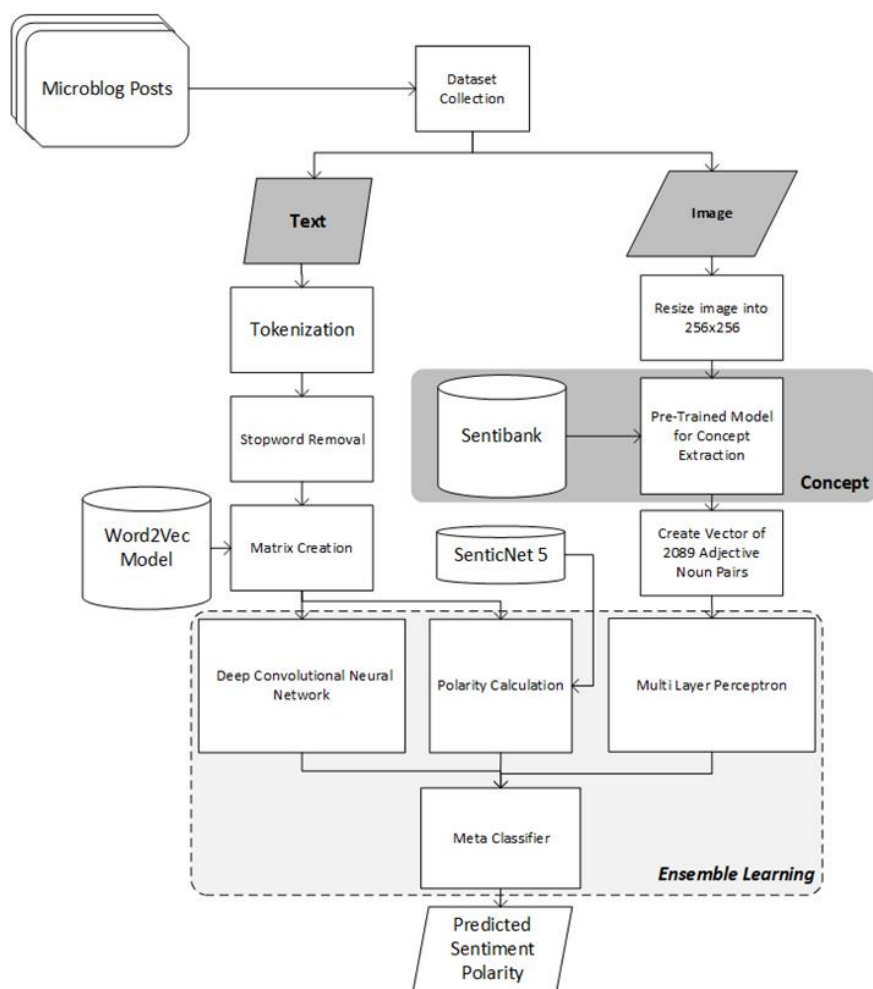


Figure. 2 Proposed architecture

First, they fine-tuned a Convolutional Neural network for image sentiment analysis. They then trained a paragraph vector model for textual sentiment analysis and later trained their multi view regression model.

Works have also been done on sentiment analysis by ensemble learning in [53] and [54]. Bayesian Learning is also implemented in [55]. An approach in [46] extracted audio, visual, and writing for multi view sentiment analysis. In this study, SenticNet 3.0 is used as a lexical source containing 30,000 concepts. The polarity value for each concept ranges from -1 to 1. SenticNet includes all of the concepts in WordNet Affect.

The feature merging process used by Poria et al. [56] is feature-level fusion and decision level fusion. Feature-level fusion combines vector features from three modalities to create one long feature vector. This method has the advantage of simplicity but still produces high accuracy. In addition to feature-level fusion, an experiment with decision level fusion was also carried out using the results of sentiment classification of each modality (sound, image, and text) as input and producing a final sentiment label as

output by using an equal-weighted scheme to represent the weight of each modality. The weight of each modality is 0.33.

Our research implements decision level fusion with the Stacking based Ensemble. We combine Logistic Regression as a meta-classifier together with Ensemble Learning, which is also implemented in [57] on text classification, and in [58] to predict bug within projects.

3. Ensemble learning for sentiment analysis

This study is expected to contribute to research on sentiment analysis in Indonesian. Sentiment analysis in Indonesian is mostly text-based. Nevertheless, despite the growth of modalities used in sentiment analysis, Deep Learning is rarely seen in Indonesian multi view sentiment analysis. Therefore, this study aims to propose a method for conducting sentiment analysis with image-text-concept features in Indonesian. Classification of sentiment polarity is usually performed by conventional machine learning. However, this work combines Deep Learning

techniques and Stacking based Ensemble for combining Image, Text, and Concept Features.

We propose an Ensemble Learning for Sentiment Analysis (ELSA) framework of sentiment analysis with textual, image, and concept features in microblog posts. The text data is processed with Deep Convolutional Neural Network and polarity calculation is performed with external knowledge. The image data is analysed with DeepSentiBank as the pre-trained model. Finally, sentiment polarity is determined by Ensemble Learning with Meta-Classifer. An overview of our proposed methodology is shown in Fig. 2.

3.1 Text sentiment analysis with DCNN

We describe several steps in text-based sentiment analysis. The first is preprocessing, which consists of two stages, i.e., tokenization and stopword removal. Tokenization is extracting words from sentences and cleaning characters that are not alphabetical, such as question marks, colons, and commas. Tweets also contain hashtags, mentions, and URLs thus we removed the \#, @, and URL prefix characters. The tokenization results were still not clean enough as they contain HTML symbols and tags. Hence, we removed the noise. Then, we delete insignificant words via stopwords removal.

Deep Convolutional Neural Network Algorithm is a network model that receives matrix input. In image processing, pixel numbers are inputted as contents of the matrix. Since text data have no numbers, it is necessary to convert the data into numbers.

Our model achieved the best result with five layers. The layers are embedding layer, followed by

convolutional 1d layer, appended by max pooling and global max pooling layer. The convolutional 1d layer learn with 50 filter size and 3, 4, 5 window size. Then, a fully-connected layer is added with a hidden layer and a softmax layer at the end. The fully-connected layer has independent weights with 50 neurons. Each neuron from the fully connected layer is fed as input to the Softmax layer. There are three neurons in the softmax layer, based on the number of sentiment targets in this study.

Kim [35] presented a study by comparing the initialization of numbers used for text data in the Convolutional Neural Network method. The accuracy when using pre-trained word vectors is much higher compared to initialization with random numbers. Word vector is used to represent text in vectors from learning results with the Neural Network algorithm. Each word has a vector that represents the meaning of the word. Words with similar meanings produce a vector with almost the same number.

Word Embedding is a language modeling technique and a feature learning technique in Natural Language Processing that maps words or phrases in the vocabulary to vectors. Embedding layer is a process to give vector values with lengths and specific values for each ID of vocabulary words. The vector length is one of the training parameters that is changed for experimental purposes - or known as embedding size. The vector values for each word ID in the vocabulary word are obtained from the embedding table. The vector for each word ID is initialized by a random value. Updates are made on the values as the training process progress. Training is carried out using a neural network. After training, words with similar meanings have vectors whose

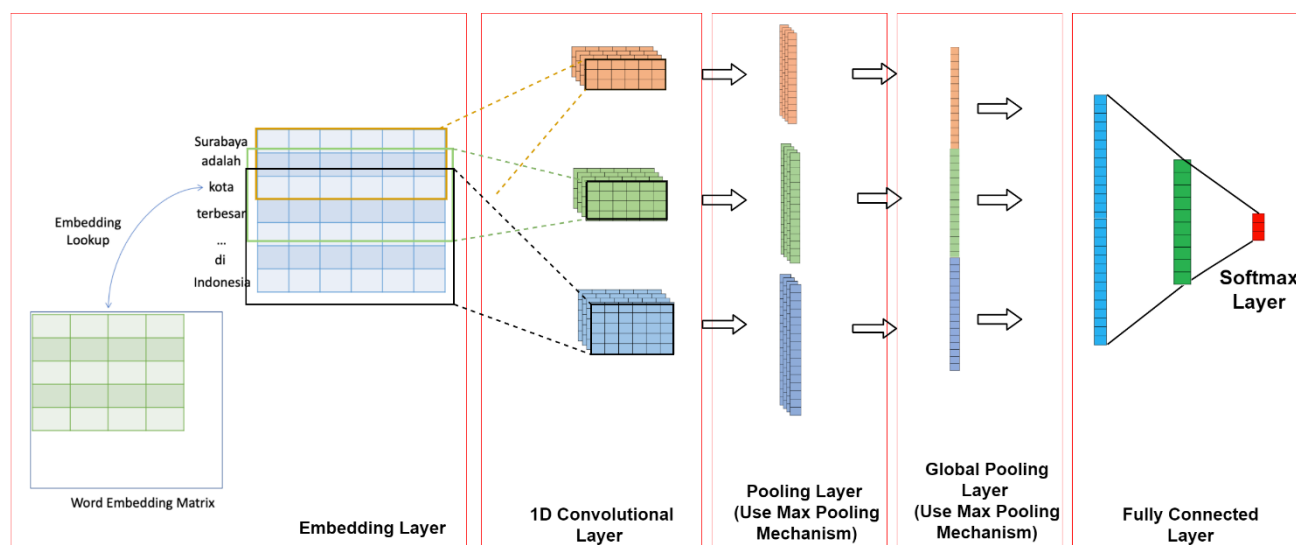


Figure. 3 CNN architecture for text sentiment analysis

values are close to one another. The Deep Convolutional Neural Network architecture used in this paper in Fig. 3 refers to the architecture derived from Kim [35] with additional modifications. If $x_i \in \mathbb{R}^k$ represents k dimensional Word2Vec based on the word i in the sentence, the sentence with length n is written as a function in Eq. (1).

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus x_n \quad (1)$$

The \oplus operator is a merging operator used to combine words that have been vectorized into matrix forms. The Convolution operation uses a filter $w \in \mathbb{R}^{hk}$, with the window size h as the number of sequence words processed. The window size used in this research is 3, 4, and 5. The c_i feature is generated from the dot product with the input $x_{i:i+h-1}$ where h is the window size followed by the rectifier linear unit activation function. The feature function is written in Eq. (2).

$$c_i = \max(0, w \cdot x_{i:i+h-1}) \quad (2)$$

Filters are applied to each word in a sentence $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$ to produce a feature map with $c \in \mathbb{R}^{n-h+1}$. The feature map function is written in Eq. (3).

$$c = [c_1, c_2, \dots, c_{n-h+1}] \quad (3)$$

In the effort to avoid overfitting, a dropout layer is required to randomly select neurons not to be used during the training phase. The probability of the dropout layer being used is 0.5. The training process continues by pooling the feature map. The pooling method used is MaxPooling, which is taking the maximum value $\hat{c} = \max\{c\}$ as a feature based on a filter. The purpose of the pooling process is to obtain the most essential features that represent other features for each feature map.

During the training phase, an optimizer is required to reduce the loss function or errors found in neurons and filters by updating neurons and filters in each iteration. In this study, Adam optimizer was used for Deep Convolutional Neural Network. The Adam optimizer algorithm calculates exponential averages of gradients and gradient ranks. The learning rate parameter of the Adam Optimizer is 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$.

3.2 Sentic based text sentiment analysis

Text Sentiment Analysis performance could be enhanced with the use of external knowledge. In Indonesian, there are several types of external knowledge, such as noun ontology [59] and SenticNet [27]. Nevertheless, to improve the quality of text Sentiment Analysis, we use SenticNet which is widely used as external knowledge in Sentiment Analysis to obtain better semantics and sentiment.

Detection of concept is done by searching concepts in sentences based on a database from SenticNet 5. Each word in a sentence is considered to be a concept composed by itself by doing the tokenization process, i.e., giving an id or token on each word. Each concept is entered or checked in the SenticNet 5 library. If the concept is found in the SenticNet 5 library, then the value of the concept is obtained, otherwise the polarity value is 0.

Each concept value obtained is used to calculate the polarity score based on the SenticNet 5 database. Fig. 4 contains examples of concepts contained in the SenticNet 5. The blue font is given to the polarity value for each concept.

Calculation of polarity is performed on all concepts contained in a sentence. The polarity value derived from each concept has a range of -1 to +1. -1 is a concept with very negative sentiment, and +1 is a concept with very positive sentiment. The sentence polarity value is calculated by summing up each polarity value of each concept in a sentence to determine the sentiment of the text. If the concept is not found in the library, then the value given is 0 for the concept. The polarity value is then compared with the sentiment label to find the error value for each sentence. All sentences that have obtained the polarity value are searched for the boundary between negative, neutral, and positive.

The boundary search process is done by attempting from the smallest to the largest number, and for each experiment, the accuracy value of all sentences predicted by SenticNet with the original sentiment polarity is analyzed. There are two boundary values sought, i.e., negative and neutral separators and neutral and positive separators. After the boundary values are obtained, all values that are below the negative and neutral boundary are made 0. Values above the neutral and positive boundary are made as a value of 2, and among them are given a value of 1. The value represents the sentiment label used. Each value is then changed to a 1x3 matrix with

```
senticnet["gurih"] = [0, 1, 0, 0, "#bunga", "#bunga", 0.333, "sangat menarik", "menarik", "bunga", ...]
senticnet["abadi"] = [0.99, 0.939, 0, 0, "#kegembiraan", "#bunga", 0.643, "membuntuti", "bertahan", ...]
```

Figure. 4 Examples of Concepts that are obtained from SenticNet

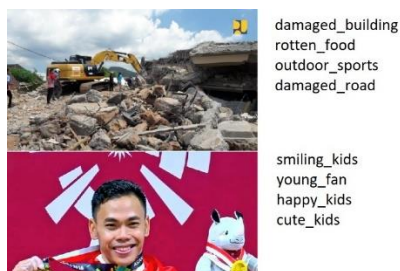


Figure. 5 The result of concept extraction

values in the column given values 1 and 0 in the other columns. This value becomes the probability of sentiment polarity. The probability result of each sentiment is used in the ensemble.

3.3 Concept extraction for image sentiment analysis

In recent works, concept-level sentiment analysis is starting to be applied in obtaining sentiment polarity [60]. Concept-level sentiment analysis does not only process text but also utilize web-based ontologies to extract implicit sentiments from concepts. Getting the concept of text-based content is usually done by extracting information in the text using parsing methods such as [61, 62].

However, in this research, we try to analyse the concept not only from the text but also from the contents of the image by using Adjective Noun Pairs. Therefore, we do not need grammar parsing. Using this pre-trained model, we can obtain some adjective noun pairs that describe sentiment of the image rather than the text itself. Therefore, concept extraction is performed on each image to get the concept contained in the image data. The library used to extract the concept is DeepSentiBank [26]. The concept extraction process is done by processing each image in Deep Convolutional Neural Network which has been trained in advance as many as 250,000 epochs. The model used is made up of 8 layers, which are five convolutional layers and three fully connected layers. Images entered on the model will be adjusted by changing the image size to 256 x 256 x 3 pixels with a range of values from pixels 0 to 255 normalized to 0 to 1. Image data is then processed on each layer up to a softmax layer with a size of 2,089, which is used to represent each concept. The resulting concept is in English, and the data is used in the classification process.

The results obtained from DeepSentiBank are concept features on images formed from adjectives and nouns that become adjective noun pairs (ANP) to represent sentiments at the mid-level such as "adorable dog" or "strange food". An example of the Adjective Noun Pair is displayed in Fig. 5.

Chen [26] introduced DeepSentiBank as a Deep Convolutional Neural Network model that had been trained to obtain ANP based on 2089 concept labels. The model has been trained as many as 250,000 iterations or 77 epochs.

The data used in the DeepSentiBank training model uses image data from the Flickr website. It was obtained by searching the appropriate Adjective Noun Pair (ANP) and obtained at most 1000 images for each ANP tag. The total numbers of obtained images are one million from 3316 ANP. Then, the image is selected on the condition that it has 120 images. The total ANP obtained is 2089 with 867,919 image data after being selected based on having at least 120 images. For each ANP, 20 data are taken for testing, and the rest is used for training. Each ANP have at least 100 image data for training and 20 data for testing.

A total of 2089 Adjective Noun Pairs is used as an image feature. The model in our framework is displayed in Fig. 6. All of these values are fed into a MultiLayer Perceptron with a size of 1000. The output of the architecture are three neurons representing negative, neutral, and positive as written in Eqs. (4) and (5).

$$h = \tanh(x \cdot w + b) \tag{4}$$

x is the concept vector from DeepSentiBank, h is the hidden layer.

$$y = \text{softmax}(h \cdot w + b) \tag{5}$$

w is the neuron weight, b is the bias.

3.4 Stacking based ensemble learning

This step combines the features of two modalities of image and text that have been previously obtained from the classification process. A common approach is by decision fusion [63]. Each sentiment probability that has been obtained is used as a classification score. One of the approaches is by using an equal-weighted scheme that provides a probability score for each classification score. A calculation is made to produce the final label using the rule-based approach in Eq. (6).

$$l' = \text{argmax}_i (q_1 s_i^{\text{image}} + q_2 s_i^{\text{text}} + q_3 s_i^{\text{sentitic}}) \tag{6}$$

Where q_1 , q_2 and q_3 are the weights for each modality. Weights $q_1 = q_2 = q_3 = 0.33$ or equal for

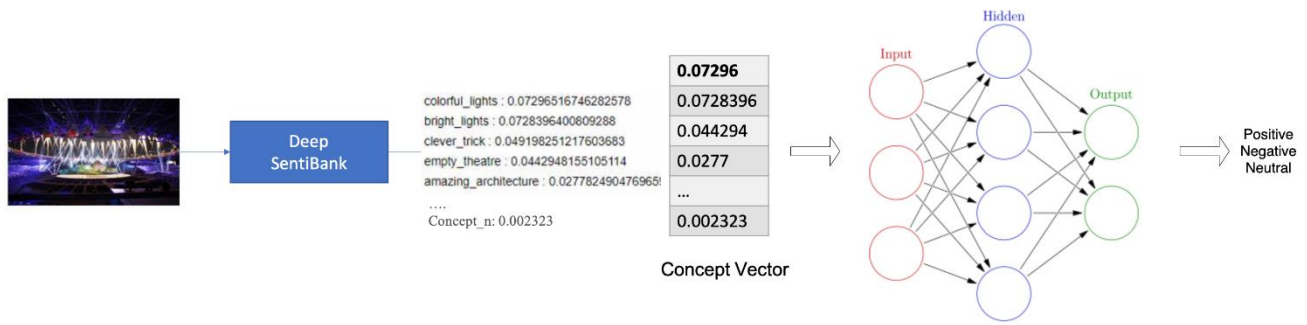


Figure. 6 Concept extraction from the deepsentibank model

each modality. s_i^{image} , s_i^{text} , s_i^{sentic} are sentiment scores of image, text, and sentic value from SenticNet 5 respectively.

We also made an effort to find the best weight for the fusion with exhaustive search. We searched for a probability value from 0% to 100% for each modality. The repetition is done as many as 5,151, from the percentage of text 0% to 100%, image 0% to 100%, and SenticNet from 0% to 100%. The total of the three values is 100%. Therefore, if both of the text and images have been given a value of 40%, then the SenticNet value is given 20% so that the total becomes 100%.

Algorithm 1 Algorithm for Ensemble Approach
Input : n : number of training data of D , x : text, I : image, y : target sentiment
Output : E : Ensemble Classifier
Learn Base Classifier
<ol style="list-style-type: none"> 1. FOR ALL $i = 1$ TO n DO 2. $h_{text}(x) \leftarrow text(x_i)$ 3. $h_{concept}(I) \leftarrow concept(I_i)$ 4. $h_{polarity}(x) \leftarrow polarity(x_i)$ 5. END FOR
Construct Dataset for Prediction
<ol style="list-style-type: none"> 1. FOR ALL $j = 1$ TO n DO 2. Create a row of predicted values which contains z_j^{prob}, y_j where $z_j^{prob} = \{h_{text}(x_j), h_{concept}(I_j), h_{polarity}(x_j)\}$ and add to D^{new} 3. END FOR
Learn Meta-Classifier
<ol style="list-style-type: none"> 1. Learn a meta-classifier E based on created D^{new} 2. RETURN E

Our proposed framework uses Stacking based Ensemble Learning to combine the predicted values from the image-text-concept [64]. However, we adopted a stacked generalization method [65] with a meta-classifier that previously only processes one type of input into image-text-concept features. This method obtains class probabilities from each image-Text-Concept, not class predictions. We

experimented with several meta-classifiers, which is reported in more detail in the experiments section.

In our model, we assume that there are n training data of $D = \{x, I, y\}$, where x is the text, I is an image, and y is the target sentiment. There are three major steps in Algorithm 1. The first is learning the base classifier from the original image-text dataset. From the dataset, we obtained $h_{text}(x), h_{concept}(I), h_{polarity}(x)$. $h_{text}(x)$ contains predicted probabilities of each target class, obtained from the DCNN text classifier $text(x_i)$ on text data x_i as input. $h_{concept}(I)$ contains predicted probabilities of each target class, obtained from the results of MultiLayer Perceptron model $concept(I_i)$ fed with 2089 Adjective Noun Pairs from image I_i as features. $h_{polarity}(x)$ contains predicted probabilities of each target class, obtained from the results of sentic model $polarity(x_i)$ from text x_i . Probability scores from step 1 are then added to a new dataset D^{new} in step two, which is building a new dataset for prediction. Probability scores from $h_{text}(x_j), h_{concept}(I_j), h_{polarity}(x_j)$ of each row from dataset D are concatenated to form z_j^{prob} . z_j^{prob} and y_j are then added to the new dataset D^{new} . The final step is learning a meta-classifier to predict the final class from the new dataset D^{new} .

4. Image and text dataset for Indonesian sentiment analysis

We collected tweets taken from Twitter based on criteria by which these tweets must have text and images. Our approach on dataset creation is based on [51]. We believe we have created a novel image-text, complemented with concept dataset for Indonesian Multi-View Sentiment Analysis. The collection of datasets via the Twitter website is done by entering a number of base words in Indonesian. The base words used have positive and negative meanings based on the Indonesian Sentiment Lexicon presented by [66], which is based on [67], and popular words based on trending topic in Twitter by the time of data collection. The positive and negative terms are

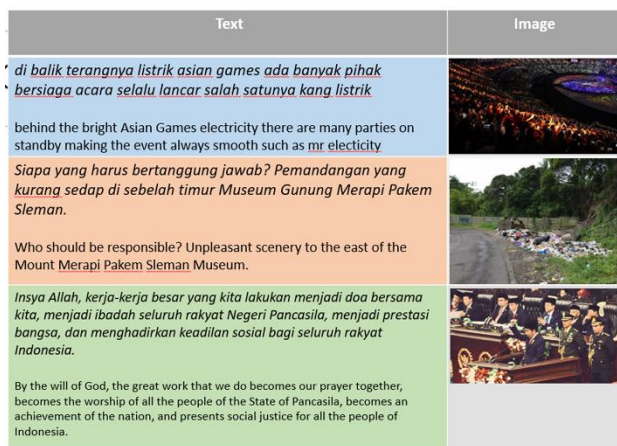


Figure. 7 Dataset sample

Table 1. Number of sentiment for each polarity

Sentiment	Annotator 1	Annotator 2	Annotator 3
Negative	727	1,201	747
Neutral	2,184	736	1,918
Positive	1,170	2,144	1,416
Total	4,081	4,081	4,081

selected based on Indonesian Sentiment Lexicon, which are manually translated in Indonesian, i.e., ‘bagus’ (good) and ‘jelek’ (bad).

In Fig. 7, the first column defines the tweet obtained from Twitter, and the second column shows the image that is posted together with the tweet. As shown in Fig. 7, the blue color defines the neutral sentiment, the green color defines the positive sentiment, and the red color defines the negative sentiment. This dataset is used in this study for the multiview sentiment analysis.

We gathered 9,962 tweets in August 2018. We selected posts that are written in Indonesian. The total data after filtering is 4,081. Three annotators manually annotated the filtered data. There are three sentiment polarities, which are positive, neutral, and negative. Based on the results of the label of three annotators, at least two annotators must agree on the same label so that the data can be used. The sentiment amount for each polarity from each annotator is in Table 1. We obtained 3,961 data on which at least two annotators have an agreement as displayed in Table 1. Out of 4,081 data labelled by three annotators, 120 data have different labels from each annotator. We achieved moderate inter-annotator agreement with

Table 2. Number of Sentiment for Each Polarity

Dataset	Positive	Negative	Neutral
InMVSA	1,398	724	470
MVSA	1,531	855	1,575

0.42 Fleiss’ kappa score. The number of each sentiment is in Table 2 with comparison from MVSA Dataset [51].

5. Experimental study

In this section, we report experiments scenarios on our dataset. Experiments were carried out with 10-fold cross-validation. The experiment has been carried out ten times. Text modality has been trained using the DCNN model. The amount of text data is 3,961 posts, and the model used the same parameters as in [35].

The accuracy of each modality when classified separately, is displayed in Table 3. Then we experimented by combining each modality using a decision level fusion weight of 0.33 or an equal-weighted scheme. The results of decision level fusion based on the research by [56] with an equal-weighted scheme are displayed in Table 4. The results obtained from a decision-level fusion with an un-weighted average scheme with the best accuracy on the data were 54.9% on the 9th fold.

We have conducted experiments to find the best weight on each fold of cross-validation with image data. We obtained the best accuracy on fold 9 with an accuracy value of 73% and an average of 70% as shown in Fig. 8. The probabilities for each modality in fold 7 are as follows: text 90%, image data 8%, and SenticNet 2%. When compared to concept data, the use of text data is 63% and concept data 34%.

Table 3. Average cross validation accuracy of each modality

Method	Text	Image	SenticNet 5
Averaging	0,680	0,519	0,457

Table 4. Accuracy of decision level fusion

Fold	Result with Weight 0.3
1	0.473
2	0.468
3	0.481
4	0.489
5	0.534
6	0.523
7	0.496
8	0.506
9	0.549
10	0.494
Averaging	0.497

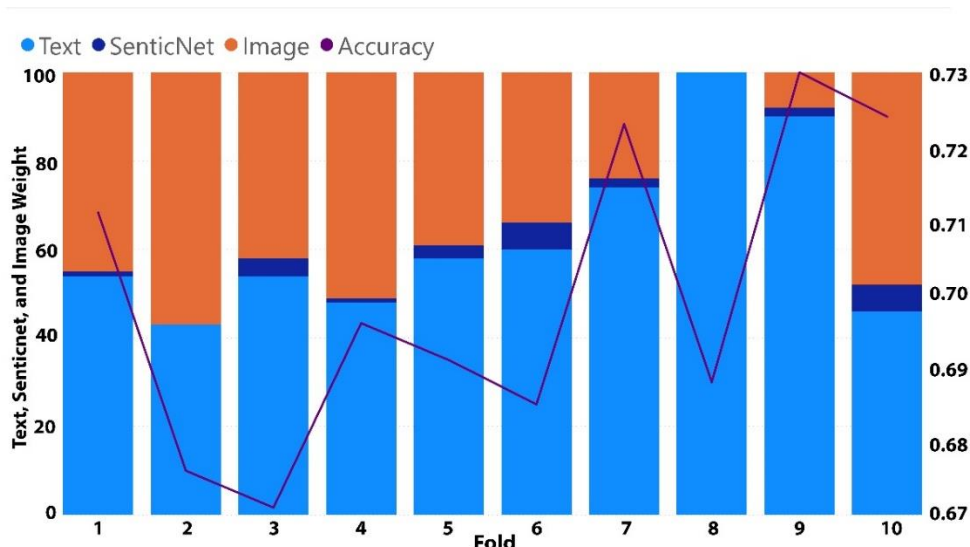


Figure. 8 Decision level with finding the best weight

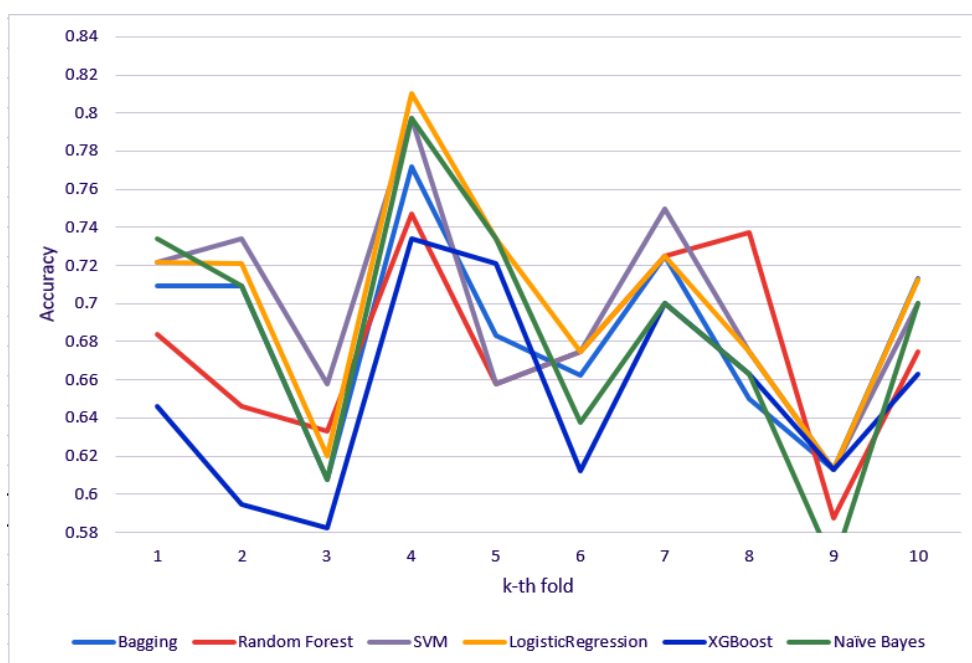


Figure. 9 Meta classifier experiments graph with image-text-concept fusion

Table 5. Accuracy of meta-classifiers

Fusion	RF	SVM	LOGIT	XGB	NB
Image-Text	0.663	0.697	0.698	0.652	0.685
Concept-Text	0.643	0.696	0.696	0.649	0.683
Image-Concept	0.478	0.486	0.527	0.493	0.513
Image-Text-Concept	0.677	0.698	0.701	0.653	0.685

Table 6. Evaluation with baseline models on image-text-concept data

Approach	Best Accuracy	Average Accuracy
Un-weighted Average	0.549	0.497
Decision Level Fusion (Weighted)	0.730	0.700
Ensemble Learning with Logistic Regression	0.810	0.701

(LOGIT), XGBoost (XGB), and Naive Bayes (NB).

We have also experimented with various Meta-Classifiers, which are Bagging, Random Forest (RF), Support Vector Machine (SVM), Logistic Regression

The experiment results are displayed in Table 5. The best-averaged accuracy was achieved with Logistic Regression at 81% on fold 4, as displayed in Fig. 9. We used a line chart to show the stability of

the performance of our proposed model.

The comparison of previous models and our approach is in Table 6. We compared with un-weighted average approach based on [46] and Decision Level Fusion by [56]. The best accuracy is obtained with Ensemble Learning with Logistic Regression with a slightly higher average accuracy.

6. Conclusion, limitation, and future works

We have developed an ELSA framework with a fusion of text-image-concept to achieve better Sentiment Analysis performance with Stacking based Ensemble Learning. Based on the results of our conducted experiments, text sentiment analysis has the best accuracy of 71% from average accuracy of 68%, image feature for sentiment analysis has the best accuracy of 55% from average accuracy of 52%, and the SenticNet results the best accuracy of 49% from average accuracy of 46%.

Multimodal sentiment analysis can increase the accuracy rate to 81% by merging the modality of text, image, and concept in the 4th fold with Logistic Regression as Meta-Classifer. The performance is 13% higher compared to only text accuracy of 68% or image accuracy of 52% on the same fold. The average accuracy of Logistic Regression as the Meta-Classifer is 70%.

Among the three features, the accuracy of polarity calculation with SenticNet5 needs to be further analyzed. SenticNet5 has an excellent performance in English. However, in Indonesian, an additional phase is necessary due to the informal words and a variety of terms used in Indonesian Social Media.

Moreover, the use of balanced data for each sentiment might improve accuracy as well as increasing the amount of data used can help improve the accuracy of prediction results. Content restrictions that are not only based on text data, but also on image data or in a particular image environment can also be done for further research. We also present a multiview dataset in Indonesian that consists of Image-Text Posts that could be analyzed in future research.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Esther I. Setiawan, Mauridhi H. Purnomo, and Suryo Sumpeno; methodology, Esther I. Setiawan and Joan Santoso; software, Esther I. Setiawan and Hans Juwiantho; validation, Joan Santoso; formal analysis, Esther I. Setiawan, Suryo

Sumpeno and Mauridhi H. Purnomo; writing—original draft preparation, Esther I. Setiawan and Joan Santoso; writing—review and editing, Mauridhi H. Purnomo, Surya Sumpeno, and Kimiya Fujisawa; visualization, Esther I. Setiawan, Hans Juwiantho and Surya Sumpeno; supervision, Surya Sumpeno, Kimiya Fujisawa and Mauridhi H. Purnomo; project administration, Esther I. Setiawan and Hans Juwiantho; funding acquisition, Esther I. Setiawan and Mauridhi H. Purnomo.

References

- [1] A. Farzindar and D. Inkpen, "Tutorial Applications of Social Media Text Analysis", *EMNLP2015 Tutor.*, 2015.
- [2] K. Liu and L. Chen, "Medical Social Media Text Classification Integrating Consumer Health Terminology", *IEEE Access*, Vol. 7, pp. 78185–78193, 2019.
- [3] M. Ling, Q. Chen, Q. Sun, and Y. Jia, "Hybrid Neural Network for Sina Weibo Sentiment Analysis", *IEEE Transactions on Computational Social Systems*, 2020.
- [4] E. I. Setiawan, F. Ferry, J. Santoso, S. Sumpeno, K. Fujisawa, and M. H. Purnomo, "Bidirectional GRU for Targeted Aspect-Based Sentiment Analysis Based on Character-Enhanced Token-Embedding and Multi-Level Attention", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 5, pp. 392–406, 2020.
- [5] F. Nurifan, R. Sarno, and K. R. Sungkono, "Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 6, pp. 47–58, 2019.
- [6] I. Amerini, C.-T. Li, and R. Caldelli, "Social network identification through image classification with CNN", *IEEE Access*, Vol. 7, pp. 35264–35273, 2019.
- [7] M. H. Alsafrjalani, "An Extensible, Modular Framework for Classifying YouTube Videos Using Web and Social Media", In: *Proc. of 2019 IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR)*, San Jose, California, pp. 459–462, 2019.
- [8] E. I. Setiawan, C. P. Susanto, J. Santoso, S. Sumpeno, and M. H. Purnomo, "Preliminary study of spam profile detection for social media using Markov Clustering: Case study on Javanese people", In: *Proc. of 2016 International Computer Science and*

- Engineering Conf. (ICSEC)*, Chiang Mai, Thailand, pp. 1–4, 2016.
- [9] J. K. P. Seng and K. L.-M. Ang, “Multimodal Emotion and Sentiment Modeling From Unstructured Big Data: Challenges, Architecture, & Techniques”, *IEEE Access*, Vol. 7, pp. 90982–90998, 2019.
- [10] S. Rogers, “What fuels a Tweet’s engagement?”, Twitter, California, United States, 2014.
- [11] S. Laird, “Twitter: A Day in the Life”, Mashable, New York, United States, 2012.
- [12] G. Roberts, “Twitter Facts & Statistics 2019”, Neighbourhood, 2019.
- [13] Q. You, J. Luo, H. Jin, and J. Yang, “Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia”, In: *Proc. of the Ninth ACM International Conf. on Web Search and Data Mining*, San Francisco, USA, pp. 13–22, 2016.
- [14] Q. You, L. Cao, H. Jin, and J. Luo, “Robust visual-textual sentiment analysis: When attention meets tree-structured recursive neural networks”, In: *Proc. of the 24th ACM International Conf. on Multimedia*, Amsterdam, Netherlands, pp. 1008–1017, 2016.
- [15] E. Cambria, “Affective computing and sentiment analysis”, *IEEE Intelligent Systems*, Vol. 31, No. 2, pp. 102–107, 2016.
- [16] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, *A practical Guide to Sentiment Analysis*, Springer, 2017.
- [17] S. Perez, “Twitter’s doubling of character count from 140 to 280 had little impact on length of tweets”, Techcrunch, San Francisco, United States, 2018.
- [18] Herman and D. A. Mononimbar, “Indonesia Fifth-Largest Country in Terms of Twitter Users”, JakartaGlobe, Jakarta, Indonesia, 2017.
- [19] W. Budiharto and M. Meiliana, “Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis”, *Journal of Big Data*, Vol. 5, No. 1, p. 51, 2018.
- [20] T. A. Le, D. Moeljadi, Y. Miura, and T. Ohkuma, “Sentiment analysis for low resource languages: A study on informal Indonesian tweets”, In: *Proc. of the 12th Workshop on Asian Language Resources (ALR12)*, Osaka, Japan, pp. 123–131, 2016.
- [21] J. Santoso, E. I. Setiawan, E. M. Yuniarno, M. Hariadi, and M. H. Purnomo, “Hybrid Conditional Random Fields and K-Means for Named Entity Recognition on Indonesian News Documents”, *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 3, pp. 233–245, 2020.
- [22] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep learning based text classification: A comprehensive review”, *arXiv Prepr. arXiv2004.03705*, 2020.
- [23] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, “A survey of multimodal sentiment analysis”, *Image and Vision Computing*, Vol. 65, pp. 3–14, 2017.
- [24] T. Chen, H. M. SalahEldeen, X. He, M.-Y. Kan, and D. Lu, “VELDA: Relating an Image Tweet’s Text and Images.”, In: *Proc. of The Twenty-Ninth AAAI Conf. on Artificial Intelligence*, Austin, United States, pp. 30–36, 2015.
- [25] A. Vempala and D. Preo\ctiuc-Pietro, “Categorizing and Inferring the Relationship between the Text and Image of Twitter Posts”, In: *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, pp. 2830–2840, 2019.
- [26] T. Chen, D. Borth, T. Darrell, and S.-F. Chang, “DeepSentibank: Visual sentiment concept classification with deep convolutional neural networks”, *arXiv Prepr. arXiv1410.8586*, 2014.
- [27] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, “SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings”, In: *Proc. of The Thirty-Second AAAI Conf. on Artificial Intelligence (AAAI-18)*, New Orleans, Louisiana, USA, 2018.
- [28] Y. Xia, X. Li, E. Cambria, and A. Hussain, “A Localization Toolkit for Sentic Net”, In: *Proc. of 2014 IEEE International Conf. on Data Mining Workshop*, Shenzhen, China, pp. 403–408, 2014.
- [29] L. Rokach, “Ensemble-based classifiers”, *Artificial Intelligence Review*, Vol. 33, No. 1–2, pp. 1–39, 2010.
- [30] H. Liang, U. Ganeshbabu, and T. Thorne, “A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution”, *IEEE Access*, Vol. 8, pp. 54164–54174, 2020.
- [31] D. A. Alboaneen, H. Tianfield, and Y. Zhang, “Sentiment analysis via multi-layer perceptron trained by meta-heuristic optimisation”, In: *Proc. of 2017 IEEE International Conf. on Big Data (Big Data)*, Boston, United States, pp. 4630–4635, 2017.
- [32] F. Ali, E.-S. Shaker, P. Khan, and K.-S. Kwak, “Feature-based Transportation Sentiment Analysis Using Fuzzy Ontology and SentiWordNet”, In: *Proc. of 2018 International Conf. on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea, pp. 1350–1355, 2018.

- [33] W. Li, Z. Li, and X. Fang, "Short text sentiment analysis based on convolutional neural network", In: *Proc. of 2018 14th International Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Limassol, Cyprus, pp. 291–295, 2018.
- [34] D. Goularas and S. Kamis, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data", In: *Proc. of 2019 International Conf. on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, Istanbul, Turkey, pp. 12–17, 2019.
- [35] Y. Kim, "Convolutional neural networks for sentence classification", In: *Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014.
- [36] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on Thai Twitter data", In: *Proc. of 2016 13th International Joint Conf. on Computer Science and Software Engineering (JCSSE)*, Pattaya, Thailand, pp. 1–6, 2016.
- [37] M. Dragoni, S. Poria, and E. Cambria, "OntoSenticNet: A commonsense ontology for sentiment analysis", *IEEE Intelligent Systems*, Vol. 33, No. 3, pp. 77–85, 2018.
- [38] W. L. K. Khine and N. T. T. Aung, "Applying Deep Learning Approach to Targeted Aspect-based Sentiment Analysis for Restaurant Domain", In: *Proc. of 2019 International Conf. on Advanced Information Technologies (ICAIT)*, Yangon, Myanmar, pp. 206–211, 2019.
- [39] C. Fiarni, H. Maharani, and R. Pratama, "Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique", In: *Proc. of 2016 4th International Conf. on Information and Communication Technology (ICoICT)*, Bandung, Indonesia, pp. 1–6, 2016.
- [40] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, "Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine", *Journal of Information System Engineering and Business Intelligence*, Vol. 4, No. 1, pp. 57–64, 2018.
- [41] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naive Bayes, Decision Tree, and Random Forest Algorithm", *Procedia Computer Science*, Vol. 161, pp. 765–772, 2019.
- [42] B. A. D. Putri, A. U. Khasanah, and A. Azzam, "Sentiment Analysis on Grab User Reviews Using Support Vector Machine and Maximum Entropy Methods", In: *Proc. of 2019 International Conf. on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, pp. 468–473, 2019.
- [43] A. Ilmania, S. Cahyawijaya, A. Purwarianti, and others, "Aspect Detection and Sentiment Classification Using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis", In: *Proc. of 2018 International Conf. on Asian Language Processing (IALP)*, Bandung, Indonesia, pp. 62–67, 2018.
- [44] H. Juwiantho, E. I. Setiawan, J. Santoso, and M. H. Purnomo, "Sentiment Analysis Twitter Bahasa Indonesia berbasis Word2Vec menggunakan Deep Convolutional Neural Network", *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 7, No. 1, pp. 181–188, 2020.
- [45] J. Yuan, Q. You, and J. Luo, "Sentiment analysis using social multimedia", *Multimedia Data Mining and Analytics*, Springer, pp. 31–59, 2015.
- [46] S. Poria, H. Peng, A. Hussain, N. Howard, and E. Cambria, "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis", *Neurocomputing*, Vol. 261, pp. 217–230, 2017.
- [47] J. Mandhyani, L. Khatri, V. Ludhrani, R. Nagdev, and S. Sahu, "Image sentiment analysis", *International Journal of Engineering Science*, Vol. 4566, 2017.
- [48] W. Fengjiao and M. Aono, "Visual Sentiment Prediction by Merging Hand-Craft and CNN Features", In: *Proc. of 2018 5th International Conf. on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, Krabi, Thailand, pp. 66–71, 2018.
- [49] R. Ji, D. Cao, and D. Lin, "Cross-modality sentiment analysis for social multimedia", In: *Proc. of 2015 IEEE International Conf. on Multimedia Big Data*, Beijing, China, pp. 28–31, 2015.
- [50] E. I. Setiawan, H. Juwiantho, J. Santoso, S. Sumpeno, K. Fujisawa, and M. H. Purnomo, "Sentiment analysis using stacking based ensemble learning of textual and image features on social media posts", In: *Proc. of The 6th IIEEJ International Conf. on Image Electronics and Visual Computing (IEVC 2019)*, Denpasar, Bali, 2019.
- [51] T. Niu, S. Zhu, L. Pang, and A. El Saddik, "Sentiment analysis on multi-view social data", In: *Proc. of International Conf. on Multimedia Modeling*, Miami, United States, pp. 15–27, 2016.

- [52] G. T. Giancristofaro and A. Panangadan, "Predicting sentiment toward transportation in social media using visual and textual features", In: *Proc. of 2016 IEEE 19th International Conf. on Intelligent Transportation Systems (ITSC)*, Rio De Janeiro, Brazil, pp. 2113–2118, 2016.
- [53] Y. Su, Y. Zhang, D. Ji, Y. Wang, and H. Wu, "Ensemble Learning for Sentiment Classification", *Workshop on Chinese Lexical Semantics*, Wuhan, Cina, pp. 84–93, 2012.
- [54] D. Kumari, "A study of meta-learning in ensemble based classifier", *Engineering Science and Technology: An International Journal*, Vol. 2, No. 1, pp. 36–41, 2012.
- [55] E. Fersini, E. Messina, and F. A. Pozzi, "Sentiment analysis: Bayesian ensemble learning", *Decision Support System*, Vol. 68, pp. 26–38, 2014.
- [56] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content", *Neurocomputing*, Vol. 174, pp. 50–59, 2016.
- [57] N. Sultana and M. M. Islam, "Meta Classifier-Based Ensemble Learning For Sentiment Classification", In: *Proc. of International Joint Conference on Computational Intelligence*, pp. 73–84, 2020.
- [58] D. Di Nucci and A. De Lucia, "The role of meta-learners in the adaptive selection of classifiers", In: *Proc. of 2018 IEEE Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE)*, Campobasso, Italy, pp. 7–12, 2018.
- [59] J. Santoso, J. N. Nugraha, E. M. Yuniarno, and M. Hariadi, "Noun ontology generation from Wikipedia article using Map Reduce with pattern based approach", In: *Proc. of 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Sanur, Bali, pp. 373–378, 2015.
- [60] E. Cambria, "An introduction to concept-level sentiment analysis", In: *Proc. of Mexican International Conf. on Artificial Intelligence*, Mexico, pp. 478–483, 2013.
- [61] A. T. Hermawan, J. Santoso, and others, "Natural language grammar induction of indonesian language corpora using genetic algorithm", In: *Proc. of 2011 International Conf. on Asian Language Processing*, Penang, Malaysia, pp. 15–18, 2011.
- [62] J. Santoso, Gunawan, H. V Gani, E. M. Yuniarno, M. Hariadi, and M. H. Purnomo, "Noun phrases extraction using shallow parsing with C4.5 decision tree algorithm for Indonesian Language ontology building", In: *Proc. of 2015 15th International Symposium on Communications and Information Technologies (ISCIT)*, Yogyakarta, Indonesia, pp. 149–152, 2015.
- [63] H.-N. Tran and E. Cambria, "Ensemble application of ELM and GPU for real-time multimodal sentiment analysis", *Memetic Computing*, Vol. 10, No. 1, pp. 3–13, 2018.
- [64] C. C. Aggarwal, *Data Classification: Algorithms and Applications*. CRC Press, Florida, United States, 2014.
- [65] K. M. Ting and I. H. Witten, "Issues in stacked generalization", *Journal of Artificial Intelligence Research*, Vol. 10, pp. 271–289, 1999.
- [66] D. H. Wahid and S. N. Azhari, "Peringkasan sentimen ekstraktif di twitter menggunakan hybrid TF-IDF dan cosine similarity", *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, Vol. 10, No. 2, pp. 207–218, 2016.
- [67] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web", In: *Proc. of the 14th International Conf. on World Wide Web*, Cardiff, United Kingdom, pp. 342–351, 2005.