

BAB I

INTRODUCTION

In recent years of information advancement, the rapid advancement of information technology makes it easier to access and collect information to be presented to the public, specifically via the internet. The ultimate goal of using the internet is the use of social media. Indirectly, social media facilitates information and news for the Indonesian people to consume [1]. From this, it has advantages that can be used, for example, in branding a business. The branding technique was prevalent before the internet and is very profitable, allowing businesses or influencers to easily reach their audience with their brand to the broader community [2].

The brand is also the identity of the business itself because the brand is the primary reflection that will be seen first by the public. Due to the development of information technology, the business brand can reach out quickly to the broader community, especially on online news portals that are easily accessible by the public to get the latest information instantly [2]. Not all brands have a favorable view in society; along with the ease of dissemination of information, public opinion can quickly reach a broader audience. These opinions are usually undetectable with their harmful content, especially opinions that point to a particular brand [2]. Negative opinions or opinions indirectly have an impact on the image and product branding of the business. Therefore, negative opinions must be mitigated to maintain a good picture of the business brand to the public; this is where the role of technology becomes crucial to facilitate the mitigation by using sentiment analysis [3].

Sentiment analysis uses analytical texts to obtain various data sources from the internet and other information platforms by retrieving information about a consumer's perception of a product, service, or brand [4]. Sentiment Analysis applies Natural Language Processing (NLP), Neural Processing Language is part of computer science and linguistics that studies the interaction between computers and human (natural) language to extract, transform, and interpret text information and classify them into different semantic groups [5].

In this study, we analyze the indications of positive and negative news from news titles on an online news portal source in Indonesia using the BERT algorithm to help mitigate and analyze alarming news. BERT algorithm is an open source technology based on a neural network for pre-training NLP [6]. BERT algorithm will be used to classify information with the Sentiment Analysis technique to be grouped and processed into two types of polarity; positive polarity, categorized as news that does not contain harmful elements. The negative

polarity is categorized as news that contains harmful elements with the scheme system design, system testing by using multilingual pre-trained provided from BERT base model then comparing the results against the Indonesian-only pre-trained, and analyzing the work of the best model accuracy from the results of testing.

In this study, we use the following references as a theoretical basis and the following references for research, journals, or international papers on sentiment analysis with similar BERT designs. The sentiment analysis with BERT in the research "BERT: Pre-training of deep bidirectional transformers for language understanding" uses pre-trained models to improve system accuracy and performance to provide excellent and accurate analysis results, especially in understanding different types of languages [6]. In research conducted by Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, the BERT algorithm and other algorithms were trained to analyze datasets from various sources such as General Language Understanding Evolution (GLUE), Stanford Question Answering Dataset (SquAD), and Situations With Adversarial Generations (SWAG). BERT algorithm has the highest accuracy percentage compared to other algorithms, with F1-Score results of 91.0%, %, 93.2%, and 86.3% from each dataset source [6]. Specifically, we used Indonesia Language only in our research. We also used two polarity detection based on the research "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets" this sentiment analysis is carried out similarly to the polarity target for reactions from cultural differences to the global crisis [7]. Sentiment analysis is also popular with famous problems such as noisy text using several pre-processing techniques to overcome noisy text in the research "Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work". Noisy text can interfere with and complicate classification in a sentiment analysis because of the ambiguity and meaning of different words [8]. For our research metrics, we adopt the same metrics on BERT model; in the research, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model" [9]. Different metrics can be implemented on different BERT model, but we specifically used confusion matrix to visualize the predictive analytics to give us direct comparison of different values.

This study used the same Bidirectional Encoder Representation from Transformers (BERT) algorithm to perform sentiment analysis. Because this algorithm will make it easier for the system to understand the context of the search intended by the user by examining the correlation of each keyword entered. BERT algorithm has a multi-layer bidirectional transformer encoder architectural model [6]. The sentiment analysis system in this study was

built to improve analyzing aspect of news identification specifically on Indonesian Language so it can be suitable for different model testing on the similar topic.