

1. INTRODUCTION

In recent years, the extraordinary evolution of technology has resulted in smartphones becoming important devices in people's daily lives. Information and communication are now within a "click" distance, making life easier and more practical. As a result, today's smartphones impact many people's lives, with more and more people owning smartphones. Based on data from the Statista Research Department in 2022, by 2020, as many as 67.15% of the population in Indonesia will have at least one smartphone [1]. The use of smartphones in Indonesia is expected to continue to grow yearly.

Smartphone products have sprung up a lot in Indonesia, where one of the most popular products today is Xiaomi. The brand issues various products such as gadgets, laptops, smartwatches, televisions, etc. Every product launched is sure to get various kinds of opinions/sentiments from the public [2]. Lei Jun founded Xiaomi in 2010 with the vision of "innovation for everyone".

Based on a market research report from Canalsy, for the 1st (one) quarter of 2021 and 2022 that Xiaomi is in the third position of the 5 (five) largest vendors on the market [3]. From the table above, Xiaomi's market share is recorded at 14% in the first (first) quarter of 2021 and 13% in the first (first) quarter of 2022. Despite a 20% decline, Xiaomi still maintains its third position with 39.2 million units. Smartphone devices to various regions in the first quarter of 2022.

Along with the increase in smartphone users, the development of social media has changed the way humans communicate. Many people use social media to express their opinions, and other things that are of public concern [4]. Such things are often called sentiments. Sentiment analysis is needed in order to understand, extract, and process textual data automatically so that sentiment information is obtained in an opinion sentence [5] that are usually found on social media, especially on Twitter. Twitter is a social media network that is popular and widely used by people in Indonesia. According to a report by Statista, there were 18.45 million users of the Twitter application in Indonesia as of January 2022. This places Indonesia as the fifth country with the most Twitter users in the world [6].

A technique that aims to detect opinions about a subject (for example, a product or an organization) in a data set is called Sentiment analysis [7]. It is also a process for determining sentiment, which is made in text form and divided based on several categories of positive, neutral, or negative sentiment [8]. Sentiment analysis analyzes sentiment values from a text document, the entire document, paragraphs, sentences, or clauses [9].

Apart from extracting sentiments, people usually want to know is, when did sentiment change occur and what caused the sentiment to change. This is important because by knowing what causes sentiment to change, the parties concerned can make better decisions [10]. In the business industry, it is used to view each customer's review or opinion, which is useful for providing customer information. In making business decisions, this information will be useful [11].

One tool that performs sentiment classification is BERT. BERT is a neural network-based technique for pre-training natural languages [12]. The way BERT works is by being able to train a Language model based on a whole set of words in a sentence or query. BERT allows the language model to learn the context of words according to those around them [13]. Bidirectional Encoder Representations from Transformers is a pre-trained contextual word representation model based on MLM (Masked Language Model), using two-way Transformers [14].

The BERT model architecture is a multi-layer bidirectional transformer encoder-decoder structure [15]. Transformers follow this architecture using self-attention and point-wise stacked, fully connected encoders and decoders [16]. There are 2 steps in the performance of the BERT framework, namely pre-training and fine-tuning. BERT pre-training does not use the traditional left-to-right or right-to-left method but uses Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) for pre-training data [17]. MLM fills in the blanks, where the model uses the context word around the mask token to predict what word should be, while NSP is the prediction of the next sentence with the two models given. After pre-training data, BERT will perform fine-tuning where fine-tuning is initialized with previously trained parameters, and all fine-tuning parameters use labeled data from downstream tasks [18]. BERT is pre-trained using 800 million words of BookCorpus data and 2.5 billion words of English Wikipedia. With rich pre-training data and pre-training tasks that make the BERT model understand every word deeply, only fine-tuning is needed to use BERT in various tasks [19].

IndoBERT is the Indonesian version of the BERT model [13]. IndoBERT is contextual embedding because it is the most advanced language model for Indonesian based on the BERT model [20]. Previous studies have used the BERT and IndoBERT models to conduct product sentiment analysis. It was stated that the IndoBERT model labeling classification has the highest f1 score among the 5 other models, namely Naive Bayes, Logistic Regression, BiLSTM w/FastText, MBERT, MalayBERT, namely achieving 84.13% [21]. In addition, research produces an F1 score obtained from precision and recall, which is 98.9% and an accuracy value of 99%, concluding that the pre-trained BERT method is very effective for implementing sentiment analysis [22]. Furthermore, some studies produce a fairly high accuracy, equal to 73.7%. This accuracy has been proven good and far enough compared to using the Naive Bayes algorithm for the classification process [23]. In other studies, some studies produce the highest accuracy of 84% [24].

This topic was chosen because in 2021 research has been carried out on products with the Xiaomi brand to determine public opinion towards Xiaomi products. It uses the Naive Bayes method with data of 2,078 tweets and producing a fairly good accuracy of 71.88% with a percentage distribution of 39% positive sentiment, 51% neutral and 10% negative [2]. This study aims to conduct a sentiment analysis of public opinion on Xiaomi smartphone products on social media Twitter using IndoBERT with hyperparameters that can optimize the performance of IndoBERT to

produce excellent accuracy. The BERT and IndoBERT methods were used because they performed well in previous studies. The BERT method has advantages over the CNN, RNN, OpenAI GPT, and ELMo methods [19]. Likewise, the IndoBERT method outperforms 5 other models, namely Naive Bayes, Logistic Regression, BiLSTM w/FastText, MBERT, and MalayBERT [21].