BAB I

With the rapid growth of the Internet, the number of marketplace users in Indonesia has greatly increased. In 2018, about 64.8% of people in Indonesia used the internet [1]. This caused a significant impact on changes in consumer behavior, especially before making a purchase decision. Online consumers can only rely on the product description provided by the seller because they cannot interact with the product. How to make sure if they will buy the right product? As stated in [2], 70% of consumers around the world trust online reviews written by previous consumers. This makes online reviews play an important role in consumer decision-making.

Most marketplaces already provide features for writing online reviews and giving an overall rating. Along with that, online forums for sharing product reviews are also rapidly growing, which makes finding product reviews much easier now than before. Increasing media coverage of beauty products in Indonesia has led consumers to pay more attention to cosmetics and personal care [3]. Online forums are filled with beauty products review, making it the most discussed topic among Indonesian women. This caused an increasing number of reviews with all sorts of comments and opinions. However, it will be very time-consuming for a consumer to read all the reviews. Other than that, focusing on the overall rating of a product alone will not be enough for the consumer to decide. Thus, mining subjective information from each aspect would be useful to help the consumer make decisions. Extracting information from the review can be done with an automated machine. Natural Language Processing (NLP) allows the machine to understand human language from review text [4]. One of NLP application that can be used to extract information is opinion mining. Most of opinion mining works studies reviews on specific domain, such as: restaurant, movie, hotel. Not many studies discussed opinion mining on beauty product domain. A study on Indonesian PRONER is one of information extraction related studies on product review domain [5].

Opinion mining is the automation method of finding out people's opinions from written language or text and analyze the opinionated data [6]. Opinion mining itself has many names including sentiment analysis, opinion extraction, and others. Opinion mining can be done on a document level or sentence level. The needs to get detailed information about specific aspect on the review lead to an idea for aspect-based opinion mining method.

Aspect-based opinion mining is a method that learns consumers' sentiment polarities on each aspect according to the words they used in reviews [7]. One of the main tasks of aspect-based opinion mining is to determine the sentiment tendency of an aspect-opinion (positive and negative). This research focused on mining information of sentiment tendency or polarity from aspect-based opinion.

TABLE I.   EXAMPLE OF LIPSTICK REVIEW WITH ASPECT-OPINION POLARITY

Review Text	Aspect - Opinion	Polarity of Aspect-Opinion
"*Lipstick ini warnanya oke dan cukup pigmented, also less oily. Recomended banget buat pemakaian sehari-hari karena tekstur yang ringan nggak berasa berat. Cuma agak kurang suka dengan packagingnya yang menurutku kurang anak muda.*" (This lipstick color is okay and quite pigmented, also less oily. Really recommended for daily use because the light texture does not feel heavy. However, I do not like the packaging, which in my opinion is lacking in youthful style.)	Color – "*warnanya oke*" (color is okay)	Positive
	Pigment – "*cukup pigmented*" (quite pigmented)	Positive
	Packaging – "*kurang suka dengan packagingnya*" (I do not like the packaging)	Negative

Table I shows an example of a review with the polarity of aspect-opinion for lipstick product from Female Daily. The reviewer gives a positive review of the *color* and *pigment* aspects of the lipstick product, while for *packaging* aspect is the opposite. Such reviews can add more value based on the commented aspect and at the same time knowing what is lacking from the product. For this reason, an aspect-based opinion mining is needed to discover the strengths or weaknesses of the products and to provide users with useful information to improve their satisfaction.

In this study, we performed aspect-based opinion extraction using Naïve Bayes Classifier. Naïve Bayes is a classification technique that uses probability and statistical methods based on Bayes' theorem. This method strongly assumes the independence of each class [8]. A study compared the performance of Naïve Bayes with other classifiers for aspect-based sentiment classification and the results show that Naïve Bayes performance surpassed other classifiers for accuracy, precision, recall, and F1-Score [9]. Another study also performed aspect-based sentiment analysis on the restaurant domain using Naïve Bayes [10]. The research in [10] was conducted by using Bag of N-gram features and resulted in high F1-Measure for the sentiment classification task. Different from [10] which only combined random undersampling and random oversampling techniques, our approach adopts Synthetic Minority Oversampling Technique (SMOTE) to add and ensuring a variety of the data. The contributions of our study are as follows:

- Since there is no labeled dataset available that is ready to be used for this research, we annotated the dataset as a training dataset in the Indonesian language for the beauty product review domain.

- We performed aspect-based opinion mining on Indonesian product review text on the beauty product domain. Not many aspect-based opinion mining studies discussed the beauty product domain, despite its unique and challenging text characteristics.

- We performed the classification experiment under two settings: with sampling and without sampling since the dataset class distribution is imbalanced.

- We performed aspect-based opinion extraction under several preprocessing settings and analyze the effect of different preprocessing settings. We conducted this because preprocessing might affect classification accuracy [11].