

# 1. INTRODUCTION

Twitter is one of the most popular online media for people of all ages. According to data released by Statista [1], Indonesia is included in the category of ten countries with the most active Twitter users worldwide, 15.7 million users in July 2021 that can say. It can be understood the importance of Twitter's social media channel as one of the an alternative to marketing media. In addition, it can be used as a forum to accommodate user responses related to certain aspects, such as Telkomsel. PT Telkomsel, Tbk is one of the providers of mobile access service providers that has the largest users in Indonesia as many as 150 million [2]. A fairly good marketing strategy is the main factor in a large number of users of this operator. However, to become a better company and want to continue to grow, Telkomsel is able to accept criticism and suggestions from its users. Complaints from users who refer to aspects of Telkomsel's services and signals are often made in Twitter tweets with harsh or kind language. The written comments can be used as evaluation material to increase Telkomsel user satisfaction and a study is needed to analyze whether the comments are positive, negative or neutral.

Sentiment analysis discusses the behavior of a person or society based on opinions related to certain topics, where these opinions will produce positive, negative, and neutral polarity values [3]. Bing Liu mentions [4], in general, sentiment analysis has three main levels, including Document Level, Sentence Level, and Aspect Level. In this study, Aspect-based [5] is one type of sentiment analysis technique that can be a solution to determine individual views on each aspect such as Telkomsel signals and services. In addition, this aspect-based sentiment analysis is carried out to identify the tendency of user loyalty to the product. The FastText feature expansion is used to help improve accuracy [6] which is based on the skip-gram model.

This research is based on various literature sources taken from previous studies related to research methods and objects. Felia et al. [7] conducted an aspect-based sentiment analysis using the Naïve Bayes and Chi-Square classification methods as selection features. mentions that adding n-gram and TF-IDF to feature selection helps to improve the F1-Score even more. This study obtained the best accuracy results, namely 80.18%, recall 72.49%, precision 77.25% and f1-score as much as 74.73%. The drawback of this study is the distribution of the data used is not balanced and there is no handling imbalanced data in this study. Furthermore, similar research discusses sentiment analysis in the research of A.N Muhammad et al. [8], the results of the study stated that NBSVM method is claimed to have very good performance. In the test, the researcher divides the data into train data and test data with varying scales. A stronger level of accuracy and performance is obtained at a scale of 7:3 or 70% train data and 30% test data. 91% precision, 83% recall, and 87% F1-Score. Then the use of the NBSVM classification method is also found in the study of F. F. Zain et al [8] to classify film reviews. Model evaluation is done by comparing the performance of three methods, namely Nave Bayes, SVM, and NBSVM. Based on the evaluation results, the model gets accuracy as high as 88.8% when using NBSVM with unigram and bigram features, compared to using individual methods, namely NB and SVM.

Erwin *et al.*[6], implement feature expansion in Twitter sentiment analysis. In his research, the classification used is SVM, Logistic Regression and Nave Bayes. They mention that feature expansion is proven to help improve accuracy. The highest accuracy is obtained in the Logistic Regression classification. However, in this study have not implemented word embedding in feature expansion testing. The next research is done by Dimuthu Lakmal *et al.*[9] evaluate the types of word embedding, namely FastText, Word2vec, and GloVe for the Sinhalese language. The evaluation method used is intrinsic and extrinsic evaluation. From the two evaluation processes, FastText with 300 vector dimensions managed to get the best overall accuracy compared to Word2vec and GloVe.

This study aims to determine the effect of feature expansion with FastText and handling imbalanced data on aspect-based Twitter sentiment analysis. A comparison of aspect-based sentiment analysis methods in this study was carried out because each method's performance results can determine high and low aspects. In its design, this research will be built using various methods, namely, Nave Bayes – Support Vector Machine (NBSVM) will be used to classify the model. Feature extraction is carried out using the TF-IDF method based on a predetermined number of features. Then a performance comparison will be made before and after the implementation of the FastText feature expansion to find out the influential features. The number of data classes that is not balanced will affect the level of accuracy of the prediction results of a model. This problem can be handled with the SMOTE [10] technique and AdaBoost [11]. These methods are proven to get good results. Therefore, the authors decided to use all these methods to produce the best model. Based on the knowledge of the author and several studies, there has been no research similar using NBSVM as classification method, TF-IDF as feature extraction, FastText as feature expansion, SMOTE and AdaBoost as handling imbalanced data method, and Telkomsel tweets as the dataset.