I. INTRODUCTION

Telecommunication service providers compete to provide users with the best products. Telkomsel is one of the cellular telecommunications service providers in Indonesia. Twitter is one of the media that users widely use to express opinions by writing a tweet and then uploading it publicly [1]. Based on data from Statista [2], Indonesia is ranked fifth with the most Twitter users, with 18,45 million users. The opinions expressed on the internet products of a company can have an impact on the related company. By knowing various user reviews, the company can evaluate what needs to be improved so that user satisfaction with Telkomsel products can increase in the future [3]. Reading all tweets is considered ineffective because it will waste much time.

These problems can be solved by conducting sentiment analysis research to determine the content and value of tweets related to one or several aspects. Sentiment analysis evaluates the pattern of tweets to predict whether the tweet gives a positive, negative, or neutral rating. The aspect-based analysis uses to classify opinions on several aspects, both subjective and objective [4]. Then, sentiment analysis results can be used as evaluation material to improve service quality and user satisfaction.

One method that can be used in sentiment analysis is the Gradient Boosted Decision Tree. Aravind, et al. conducted a comparative study of product review sentiment analysis using machine learning and Lexicon-Based Approaches with Amazon user review data. Gradient Boosting produces an accuracy of 87%, precision of 88%, recall of 98%, and an F1-score of 92% [5]. The drawback in this study is that text containing emojis is not processed in pre-processing step. If emojis can be processed, it can improve accuracy prediction.

Naufal et al. (2021) research discusses the implementation of word2vec as a feature expansion in Twitter sentiment analysis. Data was obtained from Twitter using automatic data crawling with keywords and using the Twitter API. The total tweet data collected is 11,395 tweets regarding government policies. The SVM method without using feature expansion produces an accuracy of 63,95%, and the SVM method feature expansion produces an accuracy of 68,56%. This research proves that the word2vec feature expansion can improve accuracy results [6]. The drawback of this study is that the size of the dataset used is small.

Joseph et al. (2015) shows that applying Random Undersampling in research on handling imbalanced class on tweet sentiment data can improve classification performance. However, the drawback of this study is that the dataset processed is small [7]. Amin et al. (2019) research applies SMOTE and AdaBoost to overcome the imbalance class in direct marketing bank data obtained from the UCI Machine Learning Repository on marketing as much as 41,188 data. SMOTE and AdaBoost combined with the Naïve Bayes classification resulted in an accuracy value of 88,83%, the SVM classification with an accuracy value of 89,68%, and the Decision Tree classification with the highest accuracy value of 92.25% [8].

Based on the results of several studies and the author's knowledge, there has been no research related to aspect-based sentiment analysis regarding Telkomsel products that implement the Gradient Boosting Decision Tree classification method and techniques for handling imbalanced data. To produce a model that produces the best performance, researcher apply several methods such as Gradient Boosting Decision Tree with Hyperparameter Tuning, TF-IDF as a feature extraction method, Word2Vec as a feature expansion method, SMOTE and Random Undersampling as a method of handling imbalanced data. These methods have been proven to improve performance based on previous studies. The purpose of this study is to compare three scenarios that apply SMOTE and Random Undersampling as techniques for handling imbalanced data and modeling that does not apply both of those techniques.

This paper is composed of the following sections. Section II explain the data used in the experiment. Section III explain the methodology applied in the experiment. Section IV provides the result and discussion of the experiments. Section V concludes the results.