INTRODUCTION

Technological developments have significantly changed smartphone use in Indonesia. In 2023, the number of smartphone users in Indonesia reached over 190 million [1]. This is in line with the increasing number of applications available for smartphones. One of the widely downloaded applications on the Google Play Store is GoPay.GoPay is a digital wallet launched by GoTo Indonesia [2]. Initially, GoPay was a digital payment service in the Gojek application, but in July 2023, GoTo released the GoPay application as an independent application. The popularity of GoPay results in many reviews that can be found on Google Play Store [3]. These reviews can be negative or positive impression, general or specific to certain aspects of the GoPay application. Through the sentiment analysis process, the review data can be utilized as an essential source of information.

Sentiment analysis is a subfield of Natural Language Processing (NLP) that focuses on analyzing emotions and sentiments from text [4]. Sentiment analysis aims to classify a text into certain categories, such as positive, negative, or neutral. There are several levels in sentiment analysis, namely document, sentence, and aspect levels [5]. In contrast to sentiment analysis at the sentence and document levels, sentiment analysis at the aspect level classifies sentiment based on certain components or aspects relevant to the text [6]. Through aspect-level sentiment analysis, the classification of GoPay application reviews can be done in more detail because the classification is carried out based on several aspects that are relevant to the GoPay application, such as application performance, application security, or application appearance.

One of the classification methods that can be used in the sentiment analysis process is Multilayer Perceptron (MLP). In previous research, MLP was used for various sentiment analysis cases [5], [7], [8], [9]. In previous research, MLP was compared with several other algorithms, such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Support Vector Machine (SVM), and Naive Bayes (NB) in sentiment analysis cases. MLP demonstrated prominent performance compared to other machine learning algorithms and can outperform other deep learning algorithms in certain situations [5], [8]. In addition, MLP is significantly less computationally intensive compared to more advance neural network models like RNN and CNN. It has the shortest execution time for model training in most datasets, as proven in research [5].

Other research shows that the use of word2vec and fastText as word embedding can improve the performance of various classification models, including MLP. The use of word2vec in MLP and CNN can produce accuracy of 95% and 92%, respectively [7]. Furthermore, in the sentiment analysis of the new normal in Indonesia, the utilization of fastText on MLP, NB, and SVM resulted in an average F1 value of 91% for MLP, 92% for SVM, and 72% for NB [9]. In addition, the effectiveness of word2vec and fastText as word embeddings was also proven in [10]. This study compared the use of fastText, word2vec, and GloVe on several classification models in sentiment analysis. The best results were obtained by fastText and word2vec, which outperformed GloVe on almost every classification model. While GloVe captures global co-occurrence statistics

useful for understanding overall semantic relationships, fastText and word2vec are better suited for sentiment analysis especially in the aspect level, as they capture local context and subword information more effectively.

Further research compares classic word embeddings such as fastText, word2vec, and Glove with contextual word embeddings such as Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from Language Models (ELMo) where the comparison is done on various datasets and classification models. Although the accuracy of contextual embeddings is better than classic embeddings, the difference in accuracy is not very significant. On most datasets, the difference varies around 1-5%, with contextual embeddings showing an improvement over classic embeddings [11], [12]. However, it's important to note that while the accuracy difference is not high, contextual embeddings require significantly more computing time and memory resources [13].

Previous research has proven the effectiveness of MLP in various sentiment analysis cases but it stills in sentence-level or document-level sentiment analysis [5], [7], [8], [9]. Hence, exploration and research on MLP in aspect level sentiment analysis, also known as Aspect-Based Sentiment Analysis (ABSA), still needs improvement. Previously, some studies have conducted ABSA on five types of Twitter datasets. Various classification methods have been applied, including MLP, NB, Random Forest (RF), and Support Vector Classifier (SVC). MLP showed significant potential by achieving higher accuracies than other classification methods for every dataset, with average accuracies of 78.99%, 84.09%, 80.38%, 82.37%, and 84.72%. However, the use of word embeddings as feature extraction has not been applied. In addition, further research is still needed to validate the effectiveness of MLP in other ABSA cases [14].

Another significant research contribution is a more focused study on aspect-level sentiment analysis in the context of smartphone application reviews [15], [16]. One of the first to address aspect-level sentiment analysis, specifically on smartphone application reviews.This research builds two baseline models using MLP and SVM for aspect category classification and aspect sentiment classification, achieving F1 scores of 32%, 31%, and 29% and accuracies of 66%, 67%, and 64% across different aspects. This research also introduces the AWARE dataset, featuring 11,323 reviews across three aspects: Productivity, Social Networking, and Games, and encourages further exploration in this field [15]. Continuing this exploration, subsequent research implemented CNN and several word embeddings on the same dataset, resulting in significant performance improvements, with accuracy gains of 87.88%, 93.75%, and 31.25% in aspect category classification and improvements of 16.43%, 23.35%, and 3.72% in aspect sentiment classification, demonstrating the effectiveness of using word embeddings to enhance model performance in ABSA task [16].

Based on previous research, MLP provides a balance of strong performance and computational efficiency, outperforming several classic machine learning models while being less computationally intensive than more advanced deep learning models [5], [8]. Additionally,

fastText and word2vec have proven to improve the performance of classification models, including MLP. Moreover, fastText and word2vec still offer great performance when compared to contextual embeddings methods like BERT or ELMo, while being less computationally intensive. This makes fastText and word2vec compatible for ABSA, which is a relatively straightforward classification task [7], [9], [10], [11], [12], [13].

This research aims to address this gap by exploring the relatively unexplored area of aspect-level sentiment analysis for smartphone application reviews using the GoPay review dataset. By implementing a Multilayer Perceptron (MLP) for both aspect category and sentiment classification with fastText and word2vec, which are chosen for their proven performance and computational efficiency. The aspects used are Feature and Functionality, App Interface, and User Satisfaction. This research contributes to sentiment analysis on the aspect level of the GoPay review dataset, which is still relatively new, considering that GoPay was launched independently in July 2023. In addition, this research also aims to expand the exploration of previous research, particularly in enhancing MLP performance in aspect-based sentiment analysis [15].