

1. INTRODUCTION

The Industrial Revolution 4.0 has increased the number of internet users in Indonesia [1]. Based on data collected by the Indonesian Internet Service Providers Association (APJII), the number of internet users in Indonesia in 2024 reached 221.5 million people, with a penetration rate of 79.5% of the total population of 278.7 million people [2]. Digital services such as mobile banking depend on the internet [3]. In response to the global trend of banking digitalization, Bank Negara Indonesia (BNI) launched the Wondr by BNI application to meet customers' needs for fast and easily accessible services. The app has various innovative features that help users manage their finances easily through their mobile devices [4]. According to download data on the Google Play Store, the Wondr by BNI application has been downloaded 1 million times as of October 2024, with a rating of 3.4.

App ratings are usually accompanied by user reviews, which indicate the user's experience with the app such as convenience, security, and transaction speed. Reviews on the Google Play Store are quite numerous and unstructured, so a method is needed to find out how user reviews of the application [5]. Sentiment analysis on review data is very important because this text data processing method not only processes the text but also reveals the information contained in it. It is able to identify and divide user responses into positive and negative responses. The results of the analysis can be used to repair and improve service quality [6].

There has been a lot of research on sentiment analysis, such as research related to sentiment analysis on e-commerce application using CNN method getting an accuracy result of 86,6% [7]. Research related to cellular device user optimization using the CNN-LSTM hybrid model obtained an accuracy result of 92% [8]. In the banking sector, a comparison of Logistic Regression, Random Forest, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Naive Bayes methods was conducted. LSTM achieved the best accuracy rate of 91%, SVM was second with an accuracy rate of 89% [9]. The Long Short-Term Memory method was used to review mutual fund investment applications with 99.3% accuracy [10]. The Zoom application, which uses the Convolutional Neural Network (CNN) method, also uses review data from the Google Play Store, with an accuracy result of 91.5% [11].

Sentiment analysis in Indonesian is not as easy as in English. This is due to abbreviations, informal language usage, morphological complexity, and language mixtures often found in user reviews. Text pre-processing and pre-processing stages are essential to obtain good model performance due to such issues. Therefore, to improve the classification model input, word representation techniques such as Word2Vec and FastText are used. These word representation techniques have the ability to collect contextual meaning and sub-word related information.

Based on previous research, the author in this study will use the CNN method combined with LSTM. This combination was chosen because LSTM has a better ability to handle data with long sequences and maintain important temporal context [12], while CNN is effective in extracting important features from review text [13]. In [8], the feature extraction process has not been applied, so the potential accuracy improvement cannot be fully utilized. In addition, previous studies still have gaps because they do not use word representation techniques such as Word2Vec and FastText. These techniques can enrich text features and enhance the model's ability in sentiment classification. Therefore, to improve data representation quality, this study will employ the Word2Vec and FastText word representation methods. It is anticipated that model performance will be enhanced by using these methods. However, the final results depend on the amount of data, class balance, and language variation in the reviews

This study aims to assess the effect of feature extraction scenarios with hyperparameter optimization on the performance of the CNN-LSTM model. In addition, this study also evaluates the accuracy of the model in classifying user reviews into positive and negative sentiment categories. This research aims to provide a more complete picture of how effective the incorporation of Word2Vec and FastText is in improving the representation of text data in sentiment classification tasks. In addition, the goal of this research is to find the optimal hyperparameter configuration to improve the accuracy of the CNN-LSTM model. Therefore, the findings of this study can be used as a reference for future research in the field of natural language processing and the development of deep learning models for sentiment analysis.