

Sentiment Analysis Mobile JKN Reviews Using SMOTE Based LSTM

Abstrak

Aplikasi Mobile JKN berperan penting dalam memberikan akses yang mudah dan cepat terhadap layanan kesehatan bagi pengguna JKN-KIS. Namun, ulasan pengguna menunjukkan ketidakpuasan terhadap beberapa aspek aplikasi, seperti masalah login dan kode OTP, yang dapat memengaruhi pengalaman pengguna secara keseluruhan. Tantangan lain yang dihadapi adalah ketidakseimbangan kelas pada dataset ulasan, yang dapat memengaruhi kinerja analisis sentimen. Penelitian ini menggunakan Long Short-Term Memory (LSTM) yang dikombinasikan dengan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas. Data ulasan dikumpulkan dari Google Play Store dan Kaggle, kemudian dilakukan preprocessing mencakup lemmatization, tokenization, dan padding. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa LSTM dengan SMOTE mencapai akurasi 88%, presisi 90%, recall 88%, dan F1-score 89%. SMOTE berhasil meningkatkan kinerja pada kelas minoritas meskipun terdapat sedikit penurunan pada akurasi dibandingkan model tanpa SMOTE. Visualisasi word cloud mengungkapkan sentimen positif terkait kemudahan penggunaan aplikasi, sementara sentimen negatif menunjukkan area yang memerlukan perbaikan. Penelitian ini menegaskan pentingnya penanganan dataset tidak seimbang untuk menghasilkan analisis sentimen yang lebih akurat.

Kata kunci— Analisis Sentimen, Data Tidak Seimbang, LSTM, Mobile JKN, SMOTE

Abstract

The JKN Mobile application plays an important role in providing JKN-KIS users with easy and fast access to health services. However, user reviews indicate dissatisfaction with several aspects of the application, such as login issues and OTP codes, which can affect the overall user experience. Another challenge faced is class imbalance in the review dataset, which can affect the performance of sentiment analysis. This research is using Long Short-Term Memory (LSTM) combined with Synthetic Minority Oversampling Technique (SMOTE) to manage the class imbalance. Review data is collected from the Google Play Store and Kaggle platform, then preprocessed including lemmatization, tokenization, and padding. Model performance was evaluated using the metrics accuracy, precision, recall, and f1-score. The research results show that LSTM with SMOTE achieves 88% accuracy, 90% precision, 88% recall, and 89% F1-score. SMOTE successfully improved performance in the minority class although there was a slight decrease in accuracy compared to the model without SMOTE. Word cloud visualization reveals positive sentiments regarding the ease of use of the application, while negative sentiments indicate areas that need improvement. This study emphasizes the importance of handling imbalanced datasets to produce more accurate sentiment analysis.

Keywords— Sentiment Analysis, Imbalanced Data, LSTM, Mobile JKN, SMOTE

1. INTRODUCTION

BPJS Kesehatan plays an important role in enhancing the quality of health services for Indonesian citizens [1], such as developing a mobile application services for healthcare services. Mobile JKN is one of the most applications that gave a mobile access to health facilities and services. This app have been released by the BPJS Kesehatan for health facilities and services in July 2017, the aim of this product is enhancing services for JKN-KIS users [2]. Through Mobile JKN, users can easily access a variety of health services, such as registering for health facilities, viewing membership status, and making direct payments for healthcare services [3]. This mobile application innovation is expected to improve users efficiency and comfort in accessing healthcare.

However, over time, it's very important for service providers to continue improving the application quality based on user feedback and needs. Feedback is crucial for evaluating the current application quality. Feedback can containing insights into user satisfaction and suggestions for improvement or even complaints about performance of the application [4]. According to Presidential Regulation Number 49 of 2024, consumer empowerment who are able to make optimal decisions and understand their preferences [5]. Based on this information, application developers need to gain knowledge on app shortcomings or suggestions, which can be analyzed using sentiment analysis.

Sentiment analysis is the study of text mining and natural language processing with the aim of automatically extracting insights from text data, such as in review text [6]. This analysis often used for find out how users respond to a product on the market [7]. In addition, sentiment analysis is an automatic extraction process that extracting attitudes, opinions and emotions packaged in textual data [8]. By using sentiment analysis to analyze user review data will provide important information for application development [9]. Advanced analytics with Long Short-Term Memory (LSTM) can improve this process by effectively handling sequential data and can capture long-term dependencies in text, which can provide deeper insight into user sentiment [10].

User review data can contain information that is very important for assessing whether a product is well received or not by users. By using machine learning, this review data can be processed into valuable insights that regarding customer satisfaction with an application service. Review data is usually grouped into a satisfaction index ranging from 1 to 5, but it is often imbalanced. Because users usually give a rating of more than 4 or 5, even though the contents of the review show complaints or dissatisfaction with the app service. This issue is known as imbalanced dataset. For example, in the review data Mobile JKN on play store, there are many more ratings 4 and 5 than other ratings. This imbalance can have a negative impact on sentiment analysis results.

This Imbalance dataset can causes a problem known as data bias [11], where the majority of the data has high values, while data with the low values become minority class. This condition has the potential to affect the performance of machine learning model, this is because the model will only prioritize majority data and ignore important information from minority data. This can reduce the effectiveness of the model in objectively assessing customer satisfaction.

Based on this issue, an oversampling method is needed during the data preprocessing stage of sentiment analysis to handle imbalanced datasets and ensure that the model can work optimally with the data. One possible solution is using Synthetic Minority Oversampling Technique (SMOTE). SMOTE is one of the oversampling methods that adds

synthetic data to the minority class, so that the number is the same as the majority class. Several previous research have shown that SMOTE can effectively manage class imbalance problems and improve model performance.

There are several previous studies related to sentiment analysis using SMOTE for imbalance data that have been carried out in several studies, such as research [12][13][14], which applies Naive Bayes and SMOTE to handle the imbalance data and shows significantly improved accuracy. On the other research using SMOTE to manage imbalance data with other models like random forest [15] and K-Nearest Neighbors [16]. This research has shown that SMOTE kindly improves recall and provides table performance across the models. The consistent improvement across variant studies demonstrates the reliability of SMOTE in addressing class imbalance, making it a better choice for managing the imbalance dataset in sentiment analysis and process them with LSTM model.

This research will consist of several stages, including data collection from Google Play Store and Kaggle, data preprocessing consisting of labeling, lemmatization tokenization and padding, implementation of the SMOTE method to handle class imbalance, model training using the LSTM model, and the final step is evaluating model performance using performance metrics (accuracy, precision, recall, and f1 score). The dataset consists of 100,000 user reviews of the Mobile JKN application collected from Google Play Store and Kaggle.

2. METHODS

This research method provides a detailed description of the data collection used in this research. There are two main processes in this research, namely: data preprocessing and building sentiment analysis using the SMOTE method based on the LSTM model. This research will use Python as a programming language and Jupyter Notebook as an IDE. The research method will be shown in Figure 1 below.

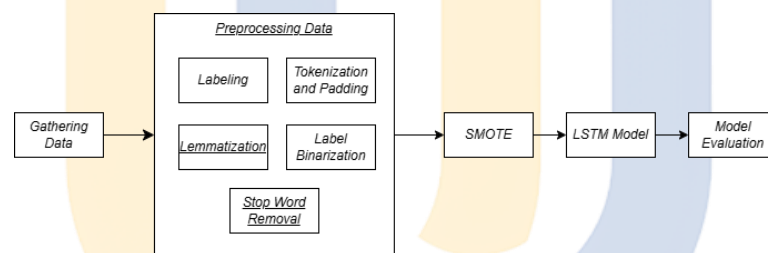


Figure 1 Research Method

Based on Figure 1 above, this research method will contain several important stages of sentiment analysis, starting from data collection, preprocessing data, and evaluation of results. These several stages aim to ensure that the data obtained can be accepted as input for the model and provide the best results. The research method will be explained in detail below.

2.1 Gathering Data

This research begins with collecting a dataset. The data collected is sourced from the Kaggle platform, using the relevant keyword “Ulasan Mobile JKN” to gather the dataset from Mobile JKN Google Play Store reviews. The dataset can be accessed from Kaggle platform at this link (<https://www.kaggle.com/datasets/nuricahyono/mobile-jkn>). Below is an example of a dataset that will be displayed in Table 1.