

I. INTRODUCTION

Transportation plays a vital role in daily life across Indonesia, with the total number of motor vehicles reaching 161,787,250 units in April 2024. This surge has contributed to severe traffic congestion, especially in urban centers [1]. To alleviate these issues, mobile applications offering transportation services have become an effective solution, making it easier for people to access transportation options digitally [2]. Despite its growing popularity, the Mitra Darat application has faced criticism from users regarding technical problems and the overall user experience. Out of 1,907 reviews on the Google Play Store, 1,204 were positive, while 703 were negative, pointing out an imbalance in the data that does not reflect the full spectrum of user opinions.

The Mitra Darat application is an example of an application from the transportation sector that provides various land transportation services. Initially known as “*Teman Bus*,” this app has evolved into a multi-service platform such as BRT Nusantara, KSPN, and Perintis [3]. The app aims to reduce dependency on private vehicles and decrease congestion in Indonesia.

Sentiment analysis is a text processing technique used to detect opinions and describe users sentiments toward a particular topic [4]. The multi-aspect sentiment analysis approach allows for more focused analysis and includes the identification of specific aspects in reviews and the sentiment orientation (positive, negative, or neutral) toward each aspect [5].

One of the challenges in sentiment analysis is imbalanced data, where one class dominates, reducing accuracy for the minority class [6]. This is critical in transportation applications, as user reviews are vital for assessing service quality and decision-making. Ignoring imbalanced data risks overlooking less dominant reviews, which may reveal specific issues or negative experiences. Techniques like undersampling and oversampling are used to address this and improve model performance [7].

In recent years, advancements in sentiment analysis, particularly using Support Vector Machine (SVM) and techniques like SMOTE, have significantly improved classification accuracy in handling imbalanced data. The state of the art in this area indicates that while methods like SVM have achieved high accuracy in classifying sentiment, challenges such as data imbalance and overfitting persist. This research aims to bridge these gaps by employing a combination of SVM and SMOTE to better analyze sentiment in user reviews of transportation apps, ensuring a more balanced and accurate representation of user experiences.

A study conducted by [8] discussed sentiment analysis on the Traveloka and Tiket.com applications. Of the total 4,821 data analyzed, Traveloka recorded 97.18% positive data, while Tiket.com reached only 73.38% of 204 data. Conversely, Tiket.com had a higher percentage of negative data, 26.62% from 741 data, compared to Traveloka, which recorded only 2.82% from 140 data. This study confirms the superiority of the Support Vector Machine (SVM) method in classifying user opinions with a high accuracy rate.

Additionally, another study by [9] analyzed sentiment on Gojek and Grab reviews from the Google Play Store using Naïve Bayes and Support Vector Machine (SVM) algorithms combined with Synthetic Minority Over-Sampling Technique (SMOTE) to address data imbalance. Naïve Bayes achieved 74.41% accuracy without SMOTE, which decreased to 64.93% with SMOTE. In contrast, SVM showed better performance, improving from 73.20% accuracy without SMOTE to 81.09% with SMOTE. The multi-aspect sentiment analysis approach using SVM is highly relevant in this study due to its proven effectiveness in classifying user sentiment, as shown in studies of Traveloka and Tiket.com. Research on Gojek and Grab also demonstrates that combining SVM with the SMOTE technique improves accuracy, especially in handling class imbalance. SVM achieved 97.18% accuracy in Traveloka, compared to Tiket.com’s 73.38%. Other studies report that SVM with SMOTE achieved 81.09% accuracy, surpassing Naïve Bayes. This study will use SVM to analyze transportation app reviews, applying undersampling and oversampling to address data imbalance, aiming to enhance public transportation services and reduce traffic congestion.

This study focuses on identifying user sentiment regarding the quality of service and user experience of the Mitra Darat application. The research aims to provide strategic input to improve public transportation services in Indonesia and encourage a shift from private vehicles to public transportation to reduce traffic congestion.