ABSTRACT

As users increasingly turn to YouTube and TikTok for travel inspiration and information, these platforms have emerged as key hubs where travelers share their journeys and seek recommendations from fellow users. Amid this surge in user-generated content, a significant challenge arises from the vast amount of unstructured data in the form of comments. This study addresses this challenge by focusing on the Special Region of Yogyakarta, the most favored holiday destination in 2022. The primary objective of this research is to aggregate and analyze public sentiment towards tourism in this region using data from YouTube and TikTok. The results of this research include the development of a data warehouse for integrating social media and tourism data using a snowflake schema. This study also produced a classification model using the Naïve Bayes Classifier algorithm, which is stored and integrated into the ETL program using the Python programming language. The resulting dashboard provides actionable insights into public sentiment dynamics, social media effectiveness, sentiment trends over time, and influential user contributions. This research is expected to be beneficial for tourism managers to continually improve tourism quality and for potential tourists to gain the best possible travel experiences.

Keywords—Data Warehouse, Sentiment Analysis, Social Media, Tourism, Special Region of Yogyakarta