ABSTRACT

Public complaints are one of the key indicators in understanding social issues, particularly in the field of transportation and traffic. The large volume of complaints submitted through social media platforms such as X (Formerly) demands an automatic classification system capable of managing information quickly, accurately, and efficiently. This study aims to develop a text classification model for public complaints addressed to the X account of Suara Surabaya (@e100ss), utilizing the Extreme Gradient Boosting (XGBoost) model combined with various feature extraction methods, maely Term Frequency-Inverse Document Frequency (TF-IDF), CountVectorizer, and Word2Vect. The system development process follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, beginning with data collection via web scraping, text preprocessing, and model evaluation using accuracy, precision, recall, f1-score, confusion metrics, and ROC Curve. The research findings show that the combination of XGBoost and CountVectorizer achieves the best performance with 93% accuracy, 90% precision, 86% recall, and f1-score of 88%. This model is considered the most balanced in identifying complaint texts without sacrificing classification accuracy. The system is implemented as an interactive web application usingthe Streamlit framework and is equipped with a phrase management feature to accommodate newly emerging complaint phrases not previously recognized by the model. In addition, the implementation of a 0,4prediction threshold enhances the system's ability to detect rare or underrepresented complaints. Through this approach, the developed system is expected to contribute to improving the qualityof public services by leveraging machine learning technology for automated text analysis

Keywords: *text classification, xgboost, public complaints, counvectorizer, public services*