



PROJECT REPORT
EVALUATION OF RATINGS AND REVIEW
SENTIMENTS USING SVM AND RANDOM FOREST FOR
DETECTING PRODUCT AUTHENTICITY ON
TOKOPEDIA

RAFAEL HARYO ABHISENA
21.K1.0053

Faculty of Computer Science
Soegijapranata Catholic University
2025

ABSTRACT

In the digital age, online shopping faces significant challenges in ensuring product authenticity, especially for highly sought-after branded goods that are prone to counterfeiting. This study examines the inconsistency between product ratings and customer reviews—where high ratings may contain negative reviews, and vice versa—which makes authenticity assessment difficult to perform accurately. A total of 2,180 Adidas product-related data were collected using Python-based web scraping techniques. The data underwent preprocessing including normalization, tokenization, and TF-IDF feature extraction. Two machine learning models, Support Vector Machine (SVM) and Random Forest, were used with an 80:20 training-testing data split. The evaluation results showed that the SVM model achieved an accuracy of 93.1% and an F1 score of 92.5%, while the Random Forest model achieved an accuracy of 83.7% and an F1 score of 82.7%. This study provides a strong foundation for the development of an authenticity detection system based on sentiment analysis on e-commerce platforms.

Keywords: product authenticity, sentiment analysis, Support Vector Machine, Random Forest, e-commerce