JURNAL INFORMATIKA
PROXIES

Vol.9, No.2, January, 2026

Evaluation of Ratings and Review Sentiments Using SVM and Random Forest for Detecting Product Authenticity on Tokopediax

Rafael Haryo Abhiseno[1], Y.B. Dwi Setianto[2]

[1] Teknik Informatika, Fakultas Ilmu Komputer, Universitas Katolik Soegijapranata, 21k10053@student.unika.ac.id, erastusrafaell@gmail.com
[2] Teknik Informatika, Fakultas Ilmu Komputer, Universitas Katolik Soegijapranata, setianto@unika.ac.id

Corresponding Author Email:erastusrafaell@gmail.com

https://doi.org/10.18280/proxies.v9i2.14046

Received: 2026-06-02
Revised: 2026-07-20
Accepted: 2026-07-20
Available online: 2026-07-21

Keywords:
e-commerce, product authenticity, random Forest, sentiment analysis, support Vector Machine, tokopedia

ABSTRACT

The digital era has made online shopping easier, but it has also presented significant challenges regarding product authenticity, especially for branded goods that are often counterfeited. This study aims to detect discrepancies between product ratings and reviews for Adidas products on Tokopedia. A total of 2,180 data points were obtained using web scraping techniques with Python. The data was processed through data pre-processing and feature extraction using TF-IDF. Machine learning algorithms Support Vector Machine (SVM) and Random Forest were applied with a training and testing data split of 80:20. The final evaluation results showed that the SVM model achieved an accuracy of 93.1% and an F1 score of 92.5%, while Random Forest achieved an accuracy of 83.7% and an F1 score of 82.7%. This indicates that sentiment analysis can be effectively used to detect discrepancies between ratings and reviews, which can support the development of product authenticity detection systems on e-commerce platforms.

INTRODUCTION

The increasing use of online marketplaces such as Tokopedia has transformed consumer behaviour in the digital era. However, product authenticity remains a critical issue, especially for branded items that are often counterfeited. Customer reviews and ratings are key indicators of product quality and trustworthiness, but inconsistencies between them—such as high ratings with negative comments—make it difficult to assess authenticity.

This study focuses on analyzing such inconsistencies to support authenticity detection in Tokopedia's fashion product category. Using a dataset of Indonesian-language reviews on Adidas products, Support Vector Machine (SVM) and Random Forest algorithms are applied to classify sentiment, detect inconsistencies, and re-rate reviews as needed. Previous research has rarely addressed this direct relationship between rating-review alignment and product authenticity.

The aim of this research is to: (1) identify inconsistency patterns between ratings and reviews, (2) re-evaluate inconsistent reviews, and (3) compare the effectiveness of SVM and Random Forest in this task. The scope is limited to textual analysis using machine learning techniques without involving user metadata.

RESEARCH METHODOLOGY

Previous research on sentiment analysis in e-commerce has explored various machine learning algorithms, such as Naïve Bayes, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbours (KNN), and Recurrent Neural Networks (RNN). These studies emphasize the importance of effective preprocessing techniques including TF-IDF, SMOTE, TOMEK LINK, and NLP-based normalization. However, few have addressed the alignment between star ratings and review content in the context of product authenticity—this is the gap that this study aims to fill.

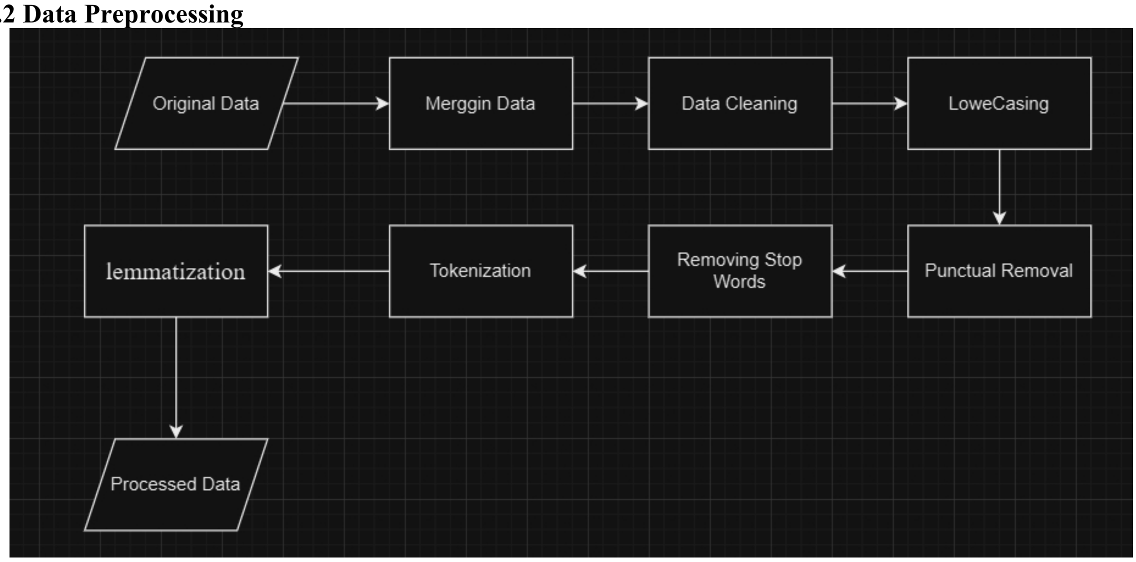
This research proposes a multi-stage sentiment analysis pipeline to detect inconsistencies between product ratings and review texts, with a focus on Adidas fashion products on Tokopedia. The stages of this research include data collection, preprocessing, labelling, feature extraction, model training, inconsistency detection, re-rating inconsistent data, and final classification. The complete research flow is shown in Figure 1.

2.1 Data Collection

```
No,Halaman,Rating,Review
1,1,5,"proses cepat, tanggap, mantab, suwun"
2,1,5,"respon dan pengiriman toko cepat, ukuran dan warna sandal sesuai orderan,, nice"
3,1,5,"pengiriman cepat seller fast respon, dan barang bagus"
4,1,5,mantap
5,1,5,"keren, bagus bgt sepatunya, nyaman, enteng dan empuk, warna cantik. asli"
6,1,5,"bagus sepatunya sesuai gambar, warnanya cantik, pengiriman aman dan cepat. tks."
7,1,5,good shoes good job
8,1,5,"joss
brg lsg kirim dan sesuai orderan
thanks seller and pak kurir"
9,1,5,"harga murah, ori sepatunya, nyaman dipakai, tdk slip"
10,1,5,ori PT MAP ðŸ'ðŸ»ðŸ'ðŸ»ðŸ'ðŸ» terimakasih seller ðŸ˜
11,2,5,"bagus banget, sepatu asli, cepet sampe, respon seller gercep bgt..pokoknya toppp"
12,2,4,"Bahan ori, bagus, kokoh. Tapi modelnya mengecewakan di bagian tali belakang itu yg jadi salah satu pertimbangan buat beli sepatu olahraga adalah, buat narik. Ini talinya..."
13,2,5,mantap
14,2,5,Produk original ðŸ'¯ Nyaman dipakai lari ðŸ' Kualitas bahan tinggi ðŸ' Produk sesuai Ekspektasi
15,2,5,"mantul, pas dan keren, makasih."
16,2,5,penjual responsif. pengiriman cepat. barang aman.
17,2,5,Modelnya keren!
18,2,5,pas mantap  Produk original ðŸ'¯
19,2,5,"Mantap brg bgus,sesuai pict,sellernya jg fast respon bgt,menjaga kualitas pokoknya tokonya amanah bgt.next bakal order lgi"
20,2,5,sesuai
21,3,5,"Fast respon meskipun tanya2 banyak,
Barang sesuai diskripsi"
22,3,5,"Sepatu sudah sampai dengan selamat, pesanan sesuai dengan deskripsi. Barang original dan pengiriman juga cepat. Sukses selalu bosq"
```

The dataset was obtained through web scraping techniques using the Python requests, BeautifulSoup, and Selenium libraries. A total of 2,180 product reviews related to Adidas fashion products (e.g., shoes, shirts, hats) were collected from Tokopedia. Each review contains the following fields: review text, product title, and user rating (1–5 stars). The data was stored in .csv format for further analysis.

2.2 Data Preprocessing



The preprocessing stage included several steps to clean and normalize the data:

- **Lowercasing:** All text was converted to lowercase.
- **Stopword Removal:** Common stopwords in Indonesian were removed using a predefined stopword list.
- **Tokenization:** Reviews were split into individual words.
- **Stemming:** Words were reduced to their root form using an Indonesian stemmer.
- **Filtering:** Only reviews with ratings of 1–2 (negative) and 4–5 (positive) were retained; neutral ratings (3) were excluded.

2.3 Labelling and Feature Extraction

Each review was labelled as positive (1) or negative (0) based on its star rating. Feature extraction was done using the TF-IDF vectorizer with unigram and bigram features, which transformed the textual data into numerical vectors suitable for machine learning models.

2.4 Sentiment Classification

Two machine learning algorithms were used to build baseline sentiment classification models:

- **Support Vector Machine (SVM):** Effective in handling high-dimensional text data.
- **Random Forest:** Able to detect complex patterns through ensemble learning.

The dataset was split into 80% training and 20% testing, and model performance was evaluated using accuracy, precision, recall, and F1-score.

2.5 Inconsistency Detection and Re-Rating

After obtaining sentiment predictions, each predicted sentiment was compared with the original rating label to identify inconsistencies. If the sentiment and the rating did not align, the review was marked as **inconsistent**.

Subsequently, a second model was trained using only **consistent data**, to learn the proper relationship between review text and rating. This model was then used to **re-rate** inconsistent reviews. The re-rated data was combined with the consistent data to create the final dataset.

2.6 Final Model and Evaluation

The final classification model was trained on the complete dataset (re-rated + consistent). SMOTE was applied to address class imbalance before training. The model's effectiveness was evaluated again using accuracy, precision, recall, and F1-score to measure its ability to classify sentiment and consistency accurately.

RESULT AND DISCUSSION

3.1 Model Evaluation Before SMOTE

Before applying class balancing using SMOTE, both Support Vector Machine (SVM) and Random Forest (RF) models showed differences in handling the imbalanced dataset. Figures 1 and 2 illustrate the evaluation metrics of both models.

```
Akurasi: 0.9380733944954128
Best Params: {'clf__C': 1000, 'clf__class_weight': None, 'clf__gamma': 0.1, 'clf__kernel': 'rbf'}
Recall: 0.9325091387961765
Precision: 0.9336809033785629
F1 Score: 0.9330889248870322

              precision    recall  f1-score   support

           0       0.92      0.91      0.91       159
           1       0.95      0.95      0.95       277

    accuracy                           0.94       436
   macro avg       0.93      0.93      0.93       436
weighted avg       0.94      0.94      0.94       436
```

Figure 1. support vector machine evaluation before smote

```
Akurasi: 0.8463302752293578
Best Params: {'class_weight': 'balanced_subsample', 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 300}
Recall: 0.8315580125396076
Precision: 0.8377747252747253
F1 Score: 0.8344056641743242

              precision    recall  f1-score   support

           0       0.81      0.77      0.79       163
           1       0.87      0.89      0.88       273

    accuracy                           0.85       436
   macro avg       0.84      0.83      0.83       436
weighted avg       0.85      0.85      0.85       436
```

Figure 2. random forest evaluation

The results of testing without SMOTE using Support Vector Machine achieved an accuracy of 94%, precision of 93%, recall of 93%, and an f1-score of 93%. Meanwhile Random Forest showed slightly different performance, with an accuracy of 85%, precision of 83%, recall of 83%, and an f1-score of 83%. These results indicate that without class balancing, both models are still capable of delivering satisfactory performance, despite the potential for bias toward the majority class due to the imbalance in label distribution in the training data.

3.2 Final Model Evaluation (After SMOTE)

To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The final evaluation results of SVM and RF are shown in figure 3, figure 4, and Table 1.

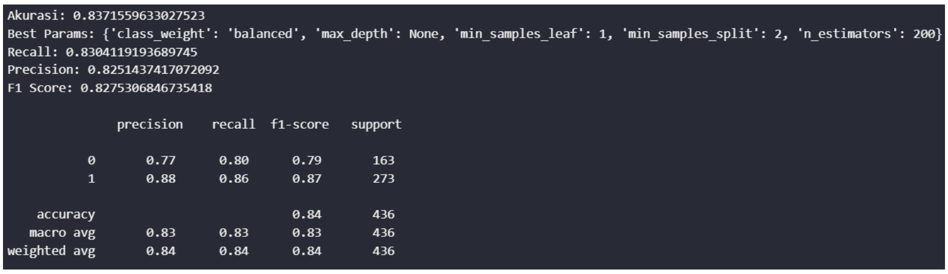
```
Akurasi: 0.9311926605504587
Best Params: {'clf__C': 10, 'clf__class_weight': None, 'clf__gamma': 'scale', 'clf__kernel': 'linear'}
Recall: 0.9230751765320255
Precision: 0.9278154058087476
F1 Score: 0.9253475789329498

              precision    recall  f1-score   support

           0       0.92      0.89      0.90       159
           1       0.94      0.95      0.95       277

    accuracy                           0.93       436
   macro avg       0.93      0.92      0.93       436
weighted avg       0.93      0.93      0.93       436
```

Figure 3. support vector machine evaluation after smote



```
Akurasi: 0.8371559633027523
Best Params: {'class_weight': 'balanced', 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
Recall: 0.8304119193689745
Precision: 0.8251437417072092
F1 Score: 0.8275306846735418

              precision    recall  f1-score   support

           0       0.77      0.80      0.79       163
           1       0.88      0.86      0.87       273

    accuracy                           0.84       436
   macro avg       0.83      0.83      0.83       436
weighted avg       0.84      0.84      0.84       436
```

Figure 4. random forest evaluation after smote

Table 1. Evaluation after smote

Model	Accuracy	Recall	Precision	F1 Score	Best Params
SVM	93.12%	92.31%	92.78%	92.53%	{'C': 10, 'kernel': 'linear', 'class_weight': None, 'gamma': 'scale'}
RF	83.72%	83.04%	82.51%	82.75%	{'n_estimators': 200, 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'class_weight': 'balanced'}

Applying SMOTE to balance the label distribution caused a consistent decline in the performance of both models, with the application of SMOTE showing a drop in accuracy, precision, recall, and F1 score compared to before SMOTE was applied.

3.3 Discussion

The SVM model consistently outperformed Random Forest in all metrics—accuracy, recall, precision, and F1-score—both before and after applying SMOTE. This result aligns with literature stating that SVM is well-suited for high-dimensional text classification, particularly when paired with TF-IDF vectorization. The use of a linear kernel with hyperparameter tuning further enhanced the margin separation between consistent and inconsistent review samples.

On the other hand, while Random Forest offered flexibility and robustness, its performance lagged behind SVM, especially in recall and precision. This is likely due to the RF's reliance on feature randomness, which may underperform in sparse textual datasets.

The application of SMOTE significantly helped improve class balance, leading to better generalization of both models. This highlights the importance of handling imbalanced data in review classification tasks, especially when detecting subtle inconsistencies between textual sentiment and numerical ratings.

CONCLUSION

The purpose of this study is to build a text-based sentiment classification system for product evaluations that can tell the difference between the rating given and the genuine sentiment indicated in the review. The many steps of the experiment, such as data preparation, initial classification, inconsistency detection, rerating of inconsistent reviews, and final classification, reveal that the proposed pipeline strategy works well to make the model more accurate and dependable.

The first classification model, which used Support Vector Machine (SVM), discovered that 207 of the 2,180 reviews (9.5%) had ratings that didn't match what they were supposed to say. By separating the reviews that didn't match and training a rerating model just on the consistent data, the system was able to modify the rating labels to better reflect the genuine thoughts expressed in the text.

The SMOTE approach made the rerating findings more realistic by changing the way the review content was spread out. The final models trained on this combined dataset did the best, with SVM attaining an accuracy of 93.1% and an F1-score (macro) of 92.5%. The Random Forest model, on the other hand, was 83.7% accurate and had an F1 score (macro) of 82.7%. The results reveal that the method of discovering and resolving label problems makes the data much better and the classification much more accurate.

But there are certain drawbacks that SMOTE might have. SMOTE produces bogus samples based on the k-nearest neighbors of the minority class. This means that if you apply it on high-dimensional and sparse representations like TF-IDF vectors, you can end up with fake data that doesn't actually indicate what text reviews are about. This could make models less accurate or recall in some cases because they learn from artificial patterns that aren't in real-world data.

So, even though this strategy works well to improve the accuracy of sentiment classification, you should be very careful when utilizing oversampling methods like SMOTE for text-based classification tasks.

ACKNOWLEDGMENT

Praise and thanks be to Jesus Christ, who has given me strength, patience, perseverance, wisdom, and guidance to complete this thesis. Without His grace, I would not have had the perseverance to complete this journey to the end.

I would like to express my deepest gratitude to my thesis advisor, Mr. Y.B. Dwi Setianto, S.T., M.Cs., for his invaluable guidance, unwavering support, and constructive feedback throughout this process. Their guidance has been a crucial factor in the successful completion of this research. I am also deeply grateful to Mr. Yonathan for his profound insights and academic encouragement, which were essential in shaping this thesis.

In addition, I would like to thank the Department of Computer Science at Soegijapranata Catholic University for the resources and facilities provided to me. Special thanks also go to the library and administrative staff, who provided tremendous support during the research process and facilitated access to important materials.

To my beloved family, especially my parents, I am indebted for their love, prayers, and unwavering encouragement. Their belief in me kept me motivated and strong during the challenging moments of this journey. I am also grateful to my friends, who have always been by my side, providing unwavering encouragement and friendship during difficult times.

Lastly, to everyone who has supported me, both directly and indirectly, I extend my heartfelt thanks. May the love of God the Father, the great gift in Jesus Christ, and the intimate fellowship in the Holy Spirit dwell in all of you.

REFERENCES

[1] W. F. Mustafa, S. Hidayat, and D. H. Fudholi, "Prediksi Retensi Pengguna Baru Shopee Menggunakan Machine Learning," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, p. 612, Jan. 2024, doi: 10.30865/mib.v8i1.7074.

[2] H. Utami, "Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network," *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 31, May 2022, doi: 10.13057/ijas.v5i1.56825.

[3] A. W. T. Hadiwibowo, F. P. Nabilla, and A. Y. P. Yusuf, "Analisis Tingkat Kepuasan Pengguna Shopee Bedasarkan Rating Dan Ulasan Google Play Store Menggunakan Naïve Bayes," *J. Ris. Inform. Dan Teknol. Inf.*, vol. 1, no. 2, pp. 43–47, Mar. 2024, doi: 10.58776/jriti.v1i2.122.

[4] S.-C. Necula, "Exploring the Impact of Time Spent Reading Product Information on E-Commerce Websites: A Machine Learning Approach to Analyze Consumer Behavior," *Behav. Sci.*, vol. 13, no. 6, p. 439, May 2023, doi: 10.3390/bs13060439.

[5] Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, and Arnita Arnita, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *J. Tek. Mesin Elektro Dan Ilmu Komput.*, vol. 3, no. 3, pp. 23–39, Nov. 2023, doi: 10.55606/teknik.v3i3.2411.

[6] S. Watmah, S. Suryanto, and M. Martias, "Komparasi Metode K-NN, Support Vector Machine Dan Random Forest Pada E-Commerce Shopee," *INSANtek*, vol. 2, no. 1, pp. 15–21, Jun. 2021, doi: 10.31294/instk.v2i1.419.

[7] N. Al Maeni, N. Suarna, and W. Prihartono, "ANALISIS SENTIMEN PENGGUNA SHOPEE BERDASARKAN DATA TWEET DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1834–1840, Apr. 2024, doi: 10.36040/jati.v8i2.8298.

[8] F. Adel Ramadhan, Rd. R. Permana Ruslan, and A. Zahra, "Sentiment Analysis Of E-Commerce Product Reviews For Content Interaction Using Machine Learning," *Cakrawala Repos. IMWI*, vol. 6, no. 1, pp. 207–220, Feb. 2023, doi: 10.52851/cakrawala.v6i1.219.

[9] I. Saputra *et al.*, "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning," *Fakt. Exacta*, vol. 13, no. 4, p. 200, Feb. 2021, doi: 10.30998/faktorexacta.v13i4.7074.

[10] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplora.v10i1.390.