



Penerapan Algoritma *Random Forest* Berbasis TF – IDF untuk Meningkatkan Akurasi Deteksi Sentimen Ulasan Produk *E-Commerce*

Munaldi¹

¹Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

dosen01573@unpam.ac.id

Abstract

The rapid growth of e-commerce results in a huge and unstructured volume of customer reviews, making manual analysis inefficient. This research aims to apply the TF-IDF-based Random Forest algorithm to improve the accuracy of automatic sentiment detection of e-commerce product reviews. The dataset used is Amazon Product Reviews with a total of 20,000 stratified data, consisting of two main sentiment classes, namely positive and negative. The main problem faced is data imbalance, where positive reviews are much more dominant, thus reducing the model's ability to recognize negative reviews. To overcome this, this study applies preprocessing stages, TF-IDF feature extraction, data balancing using SMOTE, and model optimization through hyperparameter tuning and Stratified Cross Validation. The results of the experiments showed that the early model (without SMOTE and tuning) was only able to achieve a negative class recall of 0.20, despite its relatively high accuracy. After the implementation of SMOTE and parameter optimization, the model performance improved significantly with an accuracy of 86.87% and the negative class recall increased dramatically to 0.83, and the ROC-AUC value increased to 0.89. Analysis of feature importance and WordCloud also shows that the model can recognize words that are strong indicators of customer sentiment. Thus, the proposed method has proven to be effective in improving the quality of sentiment classification and can be the basis for the development of an opinion analysis system on e-commerce platform s.

Keywords: Sentiment Analysis, E-Commerce, Random Forest, Product Reviews, TF-IDF.

Abstrak

Pertumbuhan pesat e-commerce menghasilkan volume ulasan pelanggan yang sangat besar dan tidak terstruktur, sehingga analisis manual menjadi tidak efisien. Penelitian ini bertujuan menerapkan algoritma *Random Forest* berbasis TF-IDF untuk meningkatkan akurasi deteksi sentimen ulasan produk e-commerce secara otomatis. Dataset yang digunakan merupakan *Amazon Product Reviews* dengan total 20.000 data yang dipilih secara stratified, terdiri dari dua kelas sentimen utama yaitu positif dan negatif. Permasalahan utama yang dihadapi adalah ketidakseimbangan data, di mana ulasan positif jauh lebih dominan sehingga menurunkan kemampuan model dalam mengenali ulasan negatif. Untuk mengatasi hal tersebut, penelitian ini menerapkan tahapan preprocessing, ekstraksi fitur TF-IDF, penyeimbangan data menggunakan SMOTE, serta optimasi model melalui *hyperparameter tuning* dan *Stratified Cross Validation*. Hasil percobaan menunjukkan bahwa model awal (tanpa SMOTE dan *tuning*) hanya mampu mencapai *recall* kelas negatif sebesar 0,20, meskipun akurasinya relatif tinggi. Setelah penerapan SMOTE dan optimasi parameter, performa model meningkat secara signifikan dengan akurasi 86,87% dan *recall* kelas negatif naik drastis menjadi 0,83, serta nilai ROC-AUC meningkat hingga 0,89. Analisis *feature importance* dan *WordCloud* juga memperlihatkan bahwa model dapat mengenali kata-kata yang menjadi indikator kuat sentimen pelanggan. Dengan demikian, metode yang diusulkan terbukti efektif dalam meningkatkan kualitas klasifikasi sentimen dan dapat menjadi dasar pengembangan sistem analisis opini pada platform e-commerce.

Kata kunci: Analisis Sentimen, E-Commerce, Random Forest, Ulasan Produk, TF-IDF.

1. Pendahuluan

Pertumbuhan industri e-commerce dalam lima tahun terakhir mengalami peningkatan signifikan seiring dengan perubahan pola belanja masyarakat yang beralih ke platform digital. Pada setiap transaksi,

pelanggan memberikan ulasan atau komentar terhadap produk yang dibeli [1]. Ulasan tersebut mengandung opini yang dapat merepresentasikan tingkat kepuasan pelanggan sehingga menjadi sumber informasi yang sangat berharga bagi perusahaan dalam menilai kualitas produk maupun layanan. Namun, ulasan pelanggan



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

umumnya dalam bentuk teks tidak terstruktur dan jumlahnya sangat besar, sehingga diperlukan sistem analisis otomatis untuk mengolahnya secara efisien [2]. Permasalahan yang muncul dalam pemanfaatan ulasan pelanggan adalah adanya variasi bahasa, subjektivitas pendapat, serta distribusi kelas sentimen yang tidak seimbang, di mana ulasan positif jauh lebih dominan dibandingkan ulasan negatif. Ketidakseimbangan ini membuat model pembelajaran mesin cenderung bias terhadap kelas mayoritas sehingga performa klasifikasi terhadap sentimen negatif menjadi rendah [3]. Padahal, ulasan negatif sangat penting untuk mengidentifikasi kekurangan produk yang perlu diperbaiki perusahaan.

Permasalahan utama dalam pemanfaatan ulasan pelanggan sebagai sumber informasi adalah keragaman gaya bahasa, subjektivitas opini, serta ketidakseimbangan distribusi kelas sentimen [4]. Ulasan positif sering kali jauh lebih dominan dibandingkan ulasan negatif sehingga model pembelajaran mesin cenderung bias dan kesulitan mengenali keluhan pelanggan [5]. Padahal, ulasan negatif lebih kritis dalam menunjukkan kelemahan produk. Selain itu, tantangan lainnya adalah bagaimana memilih metode yang tidak hanya akurat, tetapi juga cukup efisien untuk digunakan pada data berukuran besar yang umum terdapat pada *platform e-commerce*.

Beberapa penelitian sebelumnya menunjukkan bahwa teknik pembelajaran mesin mampu melakukan analisis sentimen dengan cukup baik. Misalnya, penelitian oleh [6] menunjukkan bahwa *SVM* dapat mengklasifikasikan sentimen *e-commerce* dengan akurasi tinggi meskipun performanya menurun saat data tidak seimbang. Studi lain oleh [7] menyebutkan bahwa metode *deep learning* seperti *LSTM* mampu menghasilkan hasil lebih baik, tetapi membutuhkan komputasi tinggi. Sementara itu, penelitian oleh [8] membuktikan bahwa penggunaan *Random Forest* pada data ulasan publik memberikan kestabilan performa yang baik terhadap variasi fitur teks.

Untuk menyelesaikan permasalahan distribusi data yang tidak seimbang tersebut, penelitian ini mengusulkan penggunaan metode *Synthetic Minority Oversampling Technique (SMOTE)* agar model tetap mampu mengenali pola pada kelas sentimen negatif. Selain itu, *TF-IDF* dipilih sebagai metode ekstraksi fitur karena efektif dalam menonjolkan kata-kata yang bermakna dalam ulasan sehingga mendukung akurasi dalam proses klasifikasi [9]. Dengan pendekatan ini, kualitas data masukan menjadi lebih baik sebelum dilakukan pelatihan model. Pada penelitian ini digunakan algoritma *Random Forest* karena memiliki keunggulan dalam menangani data berukuran besar dan berdimensi tinggi seperti teks ulasan. *Random Forest* bekerja dengan menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting*. Penelitian ini Selain itu, penggunaan *hyperparameter tuning* dan *Stratified Cross Validation*

dilakukan untuk mengoptimalkan model dan memastikan generalisasi prediksi yang lebih stabil pada data baru.

Pendekatan *TF-IDF + SMOTE + Random Forest* ini memiliki beberapa kelebihan dibandingkan metode lain, yaitu tetap efisien dalam komputasi, memiliki interpretabilitas melalui *feature importance*, dan tidak memerlukan spesifikasi lingkungan komputasi yang kompleks seperti pada *deep learning* [10]. Hal ini menjadikan metode yang diusulkan ideal untuk diterapkan pada penelitian terapan di bidang analisis sentimen *e-commerce* [11]. Dengan demikian, penelitian ini diharapkan dapat memberikan solusi yang efektif dalam menganalisis sentimen ulasan produk *e-commerce* secara lebih akurat, seimbang dalam mendeteksi sentimen positif maupun negatif, serta mendukung perusahaan dalam meningkatkan kualitas layanan melalui pemanfaatan opini pelanggan sebagai dasar pengambilan keputusan yang strategis.

2. Metode Penelitian

Metode penelitian ini dirancang untuk mengembangkan dan mengevaluasi model klasifikasi sentimen ulasan produk *e-commerce* secara otomatis dengan memanfaatkan pendekatan machine learning [12]. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data, preprocessing teks, ekstraksi fitur menggunakan *TF-IDF*, penanganan ketidakseimbangan kelas dengan *SMOTE*, pelatihan model menggunakan algoritma *Random Forest*, serta peningkatan performa melalui hyperparameter tuning dan *Stratified Cross Validation* [13].



Gambar 1. Tahapan Metode Penelitian

Berikut adalah penjelasan dari tahapan metode diatas:

a. Pengumpulan Data

Dataset yang digunakan merupakan *Amazon Product Reviews* yang setelah melalui proses *cleaning* dan pembuangan data kosong berjumlah 20.000 ulasan, terdiri dari dua kelas sentimen yaitu positif dan negatif. Data kemudian dibagi menggunakan metode *stratified split* menjadi 80% data latih (16.000 data) dan 20% data uji (4.000 data) untuk menjaga proporsi label yang seimbang pada kedua subset. Tahapan *preprocessing* mencakup *lowercasing*, penghapusan *URL*, tanda baca, angka, *stopwords*, serta *stemming* menggunakan *Porter Stemmer* untuk mendapatkan teks bersih yang siap diolah.

b. *Preprocessing Teks*

Proses *preprocessing* dilakukan untuk meningkatkan kualitas teks pada input model. Tahapan meliputi *cleaning*, *casefolding*, *stopwords removal*, dan *stemming*. Hasil *preprocessing* menunjukkan bahwa data menjadi lebih bersih dan representatif sehingga model dapat fokus mempelajari kata-kata penting yang mengandung informasi sentimen [14]. Langkah ini berperan penting dalam meningkatkan akurasi dan mengurangi *noise*.

c. *TF – IDF*

TF-IDF digunakan untuk mengubah teks menjadi fitur numerik berdasarkan frekuensi dan tingkat kepentingan kata. *TF-IDF* membantu menonjolkan kata-kata yang menjadi indikator sentimen seperti “*great*”, “*love*”, “*bad*”, “*disappoint*”. Fitur ini terbukti efektif dalam representasi teks karena menghasilkan performa model yang tinggi pada evaluasi.

d. *SMOTE*

Dataset bersifat *imbalanced* dengan dominasi ulasan positif yang mengakibatkan model bias hanya pada kelas mayoritas. Oleh karena itu digunakan *SMOTE* untuk menyeimbangkan distribusi kelas melalui oversampling pada kelas negatif. Hasil *balancing* membuat model lebih sensitif dalam mendeteksi ulasan negatif, yang sebelumnya memiliki *recall* sangat rendah. Tahap ini merupakan kunci peningkatan performa model [8].

e. *Training Model*

Random Forest dipilih karena mampu menangani jumlah fitur yang besar dari *TF-IDF* dan tidak mudah *overfitting*. Model awal menunjukkan akurasi 86.87% dengan kinerja sangat baik pada kelas positif namun masih kurang pada kelas negatif. Tahap ini memberikan *baseline* sebagai

perbandingan setelah dilakukan *tuning* dan *balancing*.

f. *Hyperparameter Tuning dan Cross Validation*

Untuk meningkatkan generalisasi model, dilakukan *GridSearchCV* dengan *Stratified K-Fold Cross Validation*. Metode ini menemukan kombinasi parameter terbaik dan mengevaluasi model melalui pembagian dataset menjadi beberapa fold. Hasil *tuning* meningkatkan *f1-score* dan *recall* kelas negatif, membuktikan bahwa model menjadi lebih adil dalam mengklasifikasikan kedua kelas [15].

g. *Evaluasi dan Feature Importance*

Evaluasi menggunakan *Accuracy*, *Confusion Matrix*, *Classification Report*, dan *ROC-AUC*. Model memiliki *ROC-AUC* 0.89, menunjukkan kemampuan yang sangat baik dalam membedakan kelas sentimen. *Feature Importance* menunjukkan bahwa kata negatif seperti “*disappoint*”, “*waste*”, “*worst*” dan kata positif seperti “*great*”, “*love*” menjadi faktor utama prediksi. Analisis ini mendukung interpretabilitas model (*Explainable AI*) dan menjelaskan alasan prediksi model terhadap ulasan.

3. Hasil dan Pembahasan

Hasil pengujian dan evaluasi yang diperoleh dari penerapan metode klasifikasi sentimen ulasan produk *e-commerce* menggunakan algoritma *Random Forest* berbasis *TF-IDF* serta penyeimbangan data dengan *SMOTE*. Seluruh tahapan analisis dilakukan secara bertahap, dimulai dari *preprocessing* data, pelatihan model, hingga proses optimasi melalui *hyperparameter tuning* dan validasi silang. Hasil yang ditampilkan pada bab ini mencakup metrik evaluasi seperti akurasi, *precision*, *recall*, *f1-score*, *confusion matrix*, dan *ROC-AUC*, disertai visualisasi pendukung serta analisis interpretatif untuk memahami kinerja model dalam mengklasifikasikan sentimen positif dan negatif secara akurat. Pembahasan dilakukan secara berurutan berdasarkan output yang dihasilkan untuk menjelaskan efektivitas model dalam menyelesaikan permasalahan ketidakseimbangan data dan meningkatkan performa pada kelas minoritas.

a. *Akurasi dan Clasification Report*

Hasil evaluasi model menunjukkan bahwa algoritma *Random Forest* memiliki akurasi sebesar 0.8687 atau 86.87%, yang mengindikasikan bahwa model berhasil mengklasifikasikan mayoritas ulasan secara benar. Namun, terlihat adanya ketidakseimbangan kinerja antar kelas. Untuk kelas positif (label 1), model menunjukkan performa sangat baik dengan *recall* sebesar 0.9976 dan *f1-score* 0.9271,

menandakan hampir semua ulasan positif berhasil terdeteksi dengan benar. Sebaliknya, untuk kelas negatif (label 0), walaupun *precision* tinggi (0.9444), *recall* sangat rendah (0.2082), yang berarti banyak ulasan negatif salah diklasifikasikan sebagai positif. Hal ini mengindikasikan bahwa dataset cenderung imbalanced dan model lebih condong memprediksi kelas positif.

```

Accuracy: 0.8686666666666667

Classification Report:
      precision    recall  f1-score   support

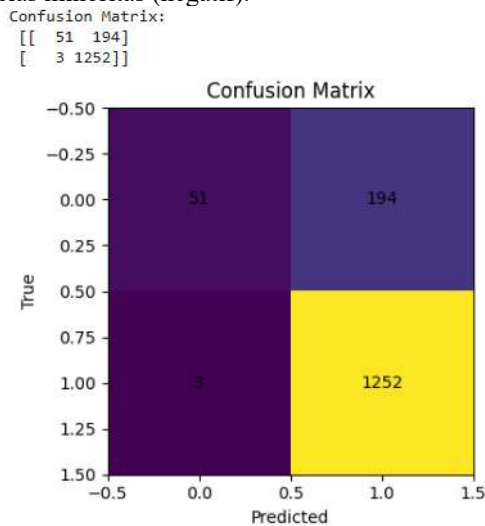
   0       0.9444     0.2082     0.3411         3
   1       0.8658     0.9976     0.9271       1252

 accuracy         0.8687         1255
 macro avg       0.9051     0.6029     0.6341         1255
 weighted avg    0.8787     0.8687     0.8314         1255
    
```

Gambar 2. Akurasi dan *Classification Report*

b. *Confusion Matrix*

Confusion matrix menunjukkan bahwa dari 245 ulasan negatif, hanya 51 yang diprediksi dengan benar, sedangkan 194 salah diklasifikasikan sebagai positif. Sebaliknya, untuk ulasan positif, model berhasil memprediksi dengan sangat baik yaitu 1.252 dari total 1.255 data positif. Distribusi prediksi ini semakin menegaskan bahwa terdapat ketimpangan data antara sentimen positif dan negatif, sehingga model lebih fokus mengidentifikasi kelas mayoritas (positif). Hal ini otomatis berdampak pada rendahnya performa di kelas minoritas (negatif).

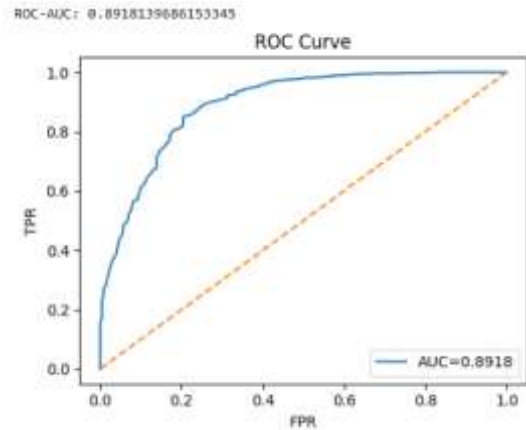


Gambar 3. *Confusion Matrix*

c. *ROC Curve* dan *ROC – AUC*

Nilai *ROC-AUC* sebesar 0.8918 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan kelas positif dan negatif pada berbagai ambang keputusan. *ROC Curve*

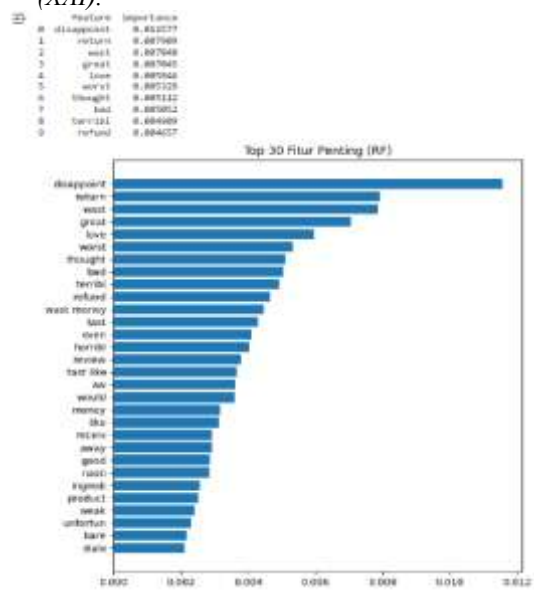
yang terlihat jauh di atas garis diagonal menunjukkan bahwa prediksi probabilitas model berada pada performa yang kuat untuk klasifikasi biner. Dengan nilai *AUC* mendekati 0.9, dapat disimpulkan bahwa model cukup robust dan efektif pada tugas analisis sentimen meskipun terdapat ketidakseimbangan label dalam dataset.



Gambar 4. *ROC - AUC*

d. *Feature Importance*

Analisis fitur penting mengungkap kata-kata yang paling berpengaruh dalam menentukan sentimen ulasan. Kata seperti "disappoint", "return", "wast", "worst", dan "bad" mendominasi kelas negatif, sedangkan kata seperti "great" dan "love" berkaitan dengan sentimen positif. Hal ini menunjukkan bahwa model cukup memahami konteks emosional dari teks ulasan. Selain itu, fitur yang berkaitan dengan kepuasan pelanggan dan pengalaman buruk menjadi indikator kuat dalam pemisahan kelas sentimen, sehingga mendukung interpretabilitas model sebagai bagian dari kajian machine learning yang *Explainable (XAI)*.



Gambar 5. *Feature Importance*

- Applicants Text Classification,” *J. Media Inform. Budidarma*, vol. 4, no. 4, 2020.
- [12] D. Pakpahan, V. Siallagan, and S. Siregar, “Classification of E-commerce Product Descriptions with The Tf-Idf and Svm Methods,” *sinkron*, vol. 8, no. 4, 2023, doi: 10.33395/sinkron.v8i4.12779.
- [13] B. Baskoro, Bayu, I. Susanto, and S. Khomsah, “Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR),” *J. Informatics, Inf. Syst. Softw. Eng. Appl. Anal.*, vol. 3, no. 2, 2021.
- [14] N. Rofiq and S. L. M. Sitio, *Pengenalan Dasar Analisis Data dengan Python di Google Colab*. Eureka Media Aksara, 2024.
- [15] S. Lina, M. Sitio, and N. Rofiq, “Classification of Creditworthy Customer Using Support Vector Machine Algorithm,” vol. 10, no. 2, pp. 339–345, 2025, doi: 10.31572/inotera.Vol10.Iss2.2025.ID502.