

Perbandingan Analisis Sentimen Ulasan Produk pada Platform E-Commerce Menggunakan Algoritma *Naïve Bayes* dan *Random Forest*

Afif Budi Andy B¹, Kusnaeni², Irwan Usman³, Muhammad Hidayatullah⁴, Muh. Rifandi⁵

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Sulawesi Barat^{1,3,4},

Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Sulawesi Barat⁵,

Program Studi Sains Data, Institut Teknologi Bacharuddin Jusuf Habibie²,

afifbudiandy.b@unsulbar.ac.id¹, kusnaeni25@ith.ac.id², irwan.usman@unsulbar.ac.id³,

muh.hidayatullah@unsulbar.ac.id⁴, muhrifandi@unsulbar.ac.id⁵

Corresponding Author: Afif Budi Andy B, Email: afifbudiandy.b@unsulbar.ac.id

ABSTRACT

Sentiment analysis has become increasingly important in e-commerce because product reviews influence consumer purchasing decisions and provide feedback for sellers to evaluate product quality and improve services. The large number of online reviews on e-commerce platforms makes manual analysis inefficient and time-consuming, thereby requiring automated sentiment classification methods that are accurate and computationally efficient. This study aims to compare the performance of the Multinomial *Naïve Bayes* and *Random Forest* algorithms in classifying sentiment in Tokopedia product reviews using the PRDECT-ID dataset, which consists of 5,400 Indonesian-language reviews. The research methodology involved several preprocessing stages, including case folding, cleaning, normalization, tokenization, stopword removal, and stemming using the Sastrawi library, followed by feature extraction using the TF-IDF method. The dataset was divided using a stratified random split approach with 80% training data and 20% testing data, and the models were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The results indicate that Multinomial *Naïve Bayes* outperformed *Random Forest*, achieving an accuracy of 93.59%, precision of 91.82%, recall of 94.65%, F1-score of 93.21%, and ROC-AUC of 0.9813. In comparison, *Random Forest* achieved an accuracy of 90.35%, precision of 85.63%, recall of 93.67%, F1-score of 89.47%, and ROC-AUC of 0.9635. In addition to its superior classification performance, Multinomial *Naïve Bayes* also demonstrated greater computational efficiency with significantly faster training time. These findings suggest that Multinomial *Naïve Bayes* is a more effective approach for sentiment classification of Indonesian-language e-commerce product reviews.

KEYWORDS

Sentiment Analysis, Multinomial Naïve Bayes, Random Forest, TF-IDF, E-Commerce

ARTICLE DOI:

<https://doi.org/10.53696/venn.v5i3.458>

A. Pendahuluan

Kemajuan teknologi digital mendorong pertumbuhan pesat perdagangan di Indonesia melalui hadirnya berbagai platform e-commerce. Peningkatan penggunaan e-commerce tidak terlepas dari tingginya akses internet oleh masyarakat. Berdasarkan hasil survei dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) mencatat bahwa pada tahun 2024 jumlah pengguna internet di Indonesia mencapai sekitar 221 juta jiwa dari total populasi 278 juta jiwa ([Aisyah et al., 2025](#)). Peningkatan jumlah pengguna internet berkontribusi langsung terhadap bertambahnya basis pelanggan pada berbagai platform e-commerce di Indonesia. Meskipun pertumbuhan e-commerce di Indonesia menunjukkan perkembangan pesat, proses pembelian daring masih menghadapi tantangan, terutama terkait ketidakpastian kualitas produk yang diterima konsumen ([Kembau et al., 2025](#); [Sutarso et al., 2024](#)).

Pesatnya perkembangan platform e-commerce telah mendorong lonjakan jumlah ulasan produk yang dihasilkan oleh pengguna dalam bentuk teks. Ulasan tersebut merepresentasikan pengalaman, opini, serta penilaian subjektif konsumen yang krusial bagi ekosistem digital saat ini. Analisis sentimen berperan penting untuk memahami opini konsumen yang terkandung dalam ulasan produk maupun interaksi di media sosial. Ulasan konsumen sering kali memuat opini positif, negatif, maupun netral yang memberikan gambaran mengenai kepuasan, kualitas produk, maupun pelayanan dari penjual. Metode analisis sentimen memungkinkan transformasi informasi tidak terstruktur, seperti ulasan di forum, blog, maupun media sosial, menjadi data terstruktur yang lebih mudah dianalisis dan diproses secara otomatis ([Darwis et al., 2021](#)). Dengan demikian, analisis sentimen membantu pengguna e-commerce dan pelaku bisnis untuk memperoleh wawasan yang lebih jelas terkait persepsi konsumen, mendukung pengambilan keputusan strategis, serta memfasilitasi perbaikan kualitas layanan dan produk ([Atimi & Pratama, 2022](#); [Hasugian et al., 2023](#)).

PRDECT-ID (Product Reviews Dataset for Emotions Classification Tasks), sebuah dataset ulasan produk berbahasa Indonesia yang menyediakan kumpulan ulasan dengan label emosi dan sentimen biner (positif dan negatif) ([Sutoyo et al., 2022](#)). Data tersebut dikumpulkan dari salah satu raksasa e-commerce di Indonesia bernama Tokopedia. Dataset ini berisi ulasan produk dari 29 kategori produk di Tokopedia yang menggunakan bahasa Indonesia. Informasi yang terkandung di dalamnya sangat berpotensi dimanfaatkan untuk memperoleh gambaran umum mengenai persepsi konsumen, khususnya dalam mengidentifikasi kecenderungan sentimen positif maupun negatif terhadap suatu produk atau layanan.

Penerapan analisis sentimen tidak hanya bergantung pada pemrosesan bahasa alami, tetapi juga memerlukan metode klasifikasi yang mampu membedakan opini positif dan negatif secara akurat. Berbagai pendekatan telah digunakan, termasuk algoritma machine learning klasik seperti Random Forest dan Multinomial Naïve Bayes. Random Forest dikenal sebagai metode *ensemble learning* yang menggabungkan banyak *Decision Tree* untuk meningkatkan akurasi klasifikasi serta mengurangi risiko overfitting ([Pandu W et al., 2025](#)). Sementara itu, Multinomial Naïve Bayes memiliki kompleksitas perhitungan yang relatif rendah karena menggunakan pendekatan probabilistik berdasarkan *Bayes' Theorem* ([Azis, 2024](#); [Syah et al., 2024](#)). Dengan karakteristik tersebut, Random Forest dan Multinomial Naïve Bayes mewakili dua pendekatan yang berbeda dalam pemrosesan data teks. Multinomial Naïve Bayes unggul dalam hal kecepatan dan efisiensi komputasi, terutama pada klasifikasi teks berdimensi tinggi ([Enjelia et al., 2025](#)), sedangkan Random Forest lebih baik dalam menangkap hubungan kompleks antar fitur dan memiliki ketahanan yang baik terhadap overfitting ([Pandu W et al., 2025](#)).

Walaupun Random Forest dan Multinomial Naïve Bayes telah banyak digunakan dalam analisis sentimen, sebagian besar penelitian terdahulu masih terbatas pada penerapan salah satu metode saja. Beberapa penelitian berikut menggambarkan hasil penerapan algoritma tersebut dalam berbagai konteks analisis sentimen. Beberapa studi menyoroti penggunaan Random Forest pada ulasan Tokopedia di Play Store dengan akurasi 93% dan F1-Score 96%, serta mengidentifikasi kualitas produk dan kinerja aplikasi sebagai masalah utama ([Rayhan et al., 2018](#)). Studi lain menganalisis sentimen Twitter tentang Tokopedia menggunakan Naïve Bayes, dengan penerapan SMOTE meningkatkan akurasi menjadi 86,93% pada Naïve Bayes ([Andreyestha & Azizah, 2022](#)). Selain itu, analisis sentimen publik Pilkada 2024 di Twitter/X menunjukkan Random Forest 92,78% dan mengungguli Naive Bayes 89,43%, dengan sentimen negatif dominan terkait kebijakan dan kinerja kandidat ([Panggabean & Junika, 2024](#)).

Meskipun berbagai penelitian sebelumnya menerapkan Random Forest maupun Naïve Bayes dalam analisis sentimen, sebagian besar studi tersebut masih berfokus pada satu pendekatan atau

konteks dataset tertentu. Penelitian yang secara langsung membandingkan metode machine learning Random Forest dan Naïve Bayes pada ulasan e-commerce Indonesia, khususnya pada data PRDECT-ID, masih relatif terbatas. Berbeda dengan penelitian terdahulu, penelitian ini secara khusus melakukan perbandingan langsung antara Random Forest dan Naïve Bayes pada dataset ulasan Tokopedia dengan skema evaluasi yang sama, sehingga perbedaan kinerja kedua pendekatan dapat dianalisis secara lebih objektif.

Berdasarkan celah penelitian tersebut, penelitian ini berfokus pada masalah bagaimana performa algoritma Random Forest dan Naïve Bayes dalam mengklasifikasikan ulasan produk e-commerce Tokopedia serta metode mana yang lebih efektif dalam menangkap opini konsumen. Penelitian ini bertujuan untuk membandingkan kinerja kedua algoritma tersebut melalui penerapan ekstraksi fitur TF-IDF pada Random Forest dan Naïve Bayes, dengan evaluasi performa menggunakan metrik akurasi, precision, recall, F1-score dan ROC-AUC. Manfaat penelitian ini diharapkan dapat membantu pelaku e-commerce dalam memahami opini konsumen secara lebih cepat dan akurat, mendukung pengambilan keputusan strategis, serta menjadi dasar pengembangan sistem analisis sentimen yang lebih efektif.

B. Metodologi

1. Sumber Data

Dataset dalam penelitian ini berupa data pada platform Kaggle yang berisi 5.400 baris data ulasan konsumen terhadap berbagai produk di Tokopedia dengan nama dataset PRDECT-ID singkatan dari *Product Reviews Dataset for Emotion Classification Task*. Dataset merupakan ulasan produk dalam bahasa Indonesia yang telah diberi anotasi sentimen oleh psikolog klinis. Sentimen dalam dataset terbagi menjadi dua kategori, yaitu Positif dan Negatif. Penelitian ini hanya berfokus pada *Customer Review*. Kolom *Customer Review* berisi teks ulasan yang menggambarkan pengalaman pengguna dalam bertransaksi.

2. Preprocessing Data

Preprocessing merupakan prosedur untuk mengubah dokumen menjadi data yang terstruktur yang melibatkan serangkaian membersihkan, mengatur, dan menyesuaikan data agar dapat digunakan dengan efisien dalam proses analisis atau pemodelan. Prosedur ini perlu dipersiapkan sebelum dilakukan ekstraksi fitur menggunakan TF-IDF. *Preprocessing* terdiri dari, yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, serta *stemming* dengan memanfaatkan pustaka Sastrawi. Beberapa prosedur yang dilakukan pada tahap ini yaitu:

- a. *Case folding* adalah proses mengubah huruf kapital menjadi huruf kecil
- b. *Basic Cleaning* adalah proses menghapus elemen yang tidak relevan seperti URL, tanda baca, angka, simbol, dan karakter khusus lainnya.
- c. *Normalization* adalah proses perubahan kata-kata yang tidak sesuai dengan keseragaman makna seperti kata singkatan, kata tidak baku, kesalahan penulisan, dan sebagainya menjadi bentuk formal baik itu dalam bahasa Indonesia maupun bahasa Inggris,
- d. *Tokenization* adalah proses memisahkan kata dalam suatu dokumen
- e. *Stopword Removal* adalah proses menghapus kata umum yang tidak memiliki pengaruh terhadap sentiment.
- f. *Stemming* adalah proses untuk mengubah kata berimbuhan menjadi kata dasar menggunakan pustaka Sastrawi.

3. Ekstraksi fitur TF-IDF

Tahap ekstraksi fitur TF-IDF merupakan transformasi data teks yang telah melalui *preprocessing* dan dilabeli menjadi representasi vektor numerik, agar dapat digunakan sebagai input dalam pelatihan model klasifikasi. Metode yang digunakan adalah memberikan bobot pada kata berdasarkan dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). TF mengukur frekuensi kemunculan suatu kata dalam dokumen, di mana semakin sering kata tersebut muncul maka semakin besar nilai TF-nya. Sementara itu, IDF berfungsi untuk mengukur tingkat keunikan kata dalam seluruh korpus dokumen, di mana kata yang jarang muncul akan memiliki nilai IDF yang lebih tinggi. Bobot akhir TF-IDF diperoleh dari hasil perkalian nilai TF dan IDF. TF-IDF memiliki rumus perhitungan sebagai berikut ([Panggabean & Junika, 2024](#)).

$$W_{t,d} = W_{tf,t,d} \times idf_t \quad (1)$$

$$W_{tf,t,d} = f_{t,d} \quad (2)$$

$$idf_t = \log \left(\frac{1 + N}{1 + df_t} \right) + 1 \quad (3)$$

Keterangan:

df_t : jumlah dokumen yang mengandung term (kata) t

N : jumlah total dokumen dalam data

$f_{t,d}$: jumlah kemunculan term (kata) t pada dokumen d

$W_{t,d}$: Bobot TF-IDF untuk term (kata) ke- t pada dokumen ke- d

$W_{tf,t,d}$: Bobot Term Frequency TF untuk term (kata) ke- t pada dokumen ke- d

idf_t : Inverse Document Frequency dari term (kata) ke- t

4. Naive Bayes

Naïve Bayes adalah metode klasifikasi yang didasarkan pada prinsip dasar teorema Bayes yang digunakan untuk menangani missing value, perhitungan probabilitas, serta memprediksi peluang pada dataset yang besar. Naïve Bayes digunakan untuk mengklasifikasikan sentiment. Secara umum, pengklasifikasian Naive Bayes sederhana dilakukan dengan menghitung probabilitas untuk menetapkan dokumen ke kategori sentimen. Setiap kategori sentimen dianggap termasuk ke kategori sentimen sehingga formula probabilitas yang digunakan adalah

$$P(C_i|d) = \frac{P(C_i) \times P(d|C_i)}{P(d)} \quad (4)$$

Untuk menentukan dokumen ke kategori sentimen yang sesuai yaitu dengan membandingkan nilai posterior dari setiap dokumen dan memilih dokumen dengan nilai posterior tertinggi. Persamaan yang digunakan untuk menghitung Maximum a Posteriori (MAP) dapat dilihat pada persamaan berikut:

$$C_{map} = \underset{j}{\operatorname{argmax}} P(C_i) \prod_j P(d|C_i) \quad (5)$$

Dalam hal ini, diasumsikan bahwa dataset yang terdiri dari dokumen bersifat independen satu sama lain, sehingga kategori sentimen yang sesuai dengan nilai probabilitas maksimum yang diperoleh oleh teks masukan 'd' adalah kategori teks 'd'.

Multinomial Naïve Bayes merupakan varian dari Naïve Bayes yang umum digunakan dalam klasifikasi teks, seperti analisis sentimen. Algoritma ini dirancang untuk menangani fitur berbasis frekuensi kata, seperti yang diperoleh melalui TF-IDF. Model Multinomial Naïve Bayes mengasumsikan bahwa setiap fitur bersifat independen secara kondisional terhadap fitur lainnya dalam satu kelas, serta mengikuti distribusi multinomial. Probabilitas suatu dokumen dikalkulasi

berdasarkan frekuensi kemunculan kata dalam dokumen, dan kelas dengan nilai kemungkinan terbesar dipilih sebagai hasil klasifikasi (Dewi et al., 2023). Dalam penelitian ini, Multinomial Naïve Bayes digunakan dengan parameter $\alpha=0.1$ sebagai teknik *smoothing* untuk menghindari probabilitas nol. Proses tuning hyperparameter dilakukan menggunakan GridSearchCV dengan 5-fold cross-validation, dengan rentang α mulai dari 0.1 hingga 10.0 untuk menemukan konfigurasi terbaik berdasarkan akurasi.

5. Random Forest

Random Forest merupakan salah satu metode yang dapat meningkatkan hasil akurasi dalam membangkitkan atribut untuk setiap *node* yang dilakukan secara acak. Random Forest terdiri dari sekumpulan *decision tree*, dimana kumpulan pohon keputusan ini digunakan untuk mengklasifikasi data ke suatu kelas. Metode ini bertujuan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan. Random Forest menggunakan *information gain* untuk perhitungan dalam membangun pohon (Iqbal et al., 2025). *Information gain* adalah perubahan *entropy* setelah dilakukan pembagian berdasarkan sebuah atribut ke dalam subset yang lebih kecil. Secara matematis untuk menghitung nilai *information gain* menggunakan persamaan di bawah ini

$$Gain(A) = Entropy(S) - Entropy_A(S) \quad (6)$$

Kemudian untuk menghitung nilai $Entropy(S)$ dan $nEntropy_A(S)$ dapat menggunakan persamaan sebagai berikut:

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (7)$$

$$Entropy(S_i) = \sum_{i=1}^n \frac{A_i}{c} \quad (8)$$

$$Entropy_A(S) = \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (9)$$

6. Confusion Matrix

Confusion Matrix adalah metode yang digunakan dalam analisis performa model klasifikasi karena memberikan informasi yang rinci mengenai hasil prediksi model terhadap data uji. Matriks ini membagi hasil prediksi ke dalam empat kategori: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dengan demikian, *Confusion Matrix* memungkinkan pengukuran kesalahan serta keberhasilan model dalam mengklasifikasikan setiap kelas. *True Positive* (TP) adalah jumlah prediksi positif yang benar, berarti model berhasil memprediksi bahwa sampel termasuk dalam kelas positif. *True Negative* (TN) menunjukkan jumlah prediksi negatif yang benar, di mana model mampu mengidentifikasi sampel yang tidak termasuk dalam kelas positif. *False Positive* (FP) adalah model salah mengklasifikasikan sampel negatif sebagai positif, sementara *False Negative* (FN) adalah model salah memprediksi sampel positif sebagai negatif. Klasifikasi ini membantu dalam mengidentifikasi kekuatan dan kelemahan model dalam melakukan prediksi pada data uji (Saputra & Budiman, 2026).

7. Evaluasi Model

Tahap evaluasi model bertujuan mengukur kinerja model setelah digunakan. Evaluasi ini dilakukan menggunakan data uji yang tidak digunakan dalam proses pelatihan, untuk memberikan gambaran tentang kemampuan model dalam memprediksi data baru (Aji et al., 2025). Beberapa metrik digunakan dalam evaluasi ini, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Setiap metrik memberikan perspektif berbeda mengenai kinerja model, mulai dari ketepatan prediksi hingga kemampuan model dalam menangani ketidakseimbangan kelas (Saadah et al., 2026). Secara matematis, dirumuskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

Selain melihat nilai akurasi, keakuratan hasil klasifikasi dapat dilihat dengan menggunakan nilai AUC/ROC. AUC/ROC yaitu area di bawah kurva karakteristik operasi penerima, yang menggambarkan kinerja model dalam mengklasifikasikan data dalam kategori positif atau negatif (Nawawi et al., 2024). Aturan pengambilan keputusan dalam menentukan nilai AUC/ROC disajikan pada tabel 1.

TABEL 1. Nilai AUC/ROC

Nilai	Klasifikasi
0,90-1,00	Sangat Baik
0,80-0,90	Baik
0,70-0,80	Sedang
0,60-0,70	Buruk
0,50-0,60	Salah

C. Hasil Penelitian dan Pembahasan

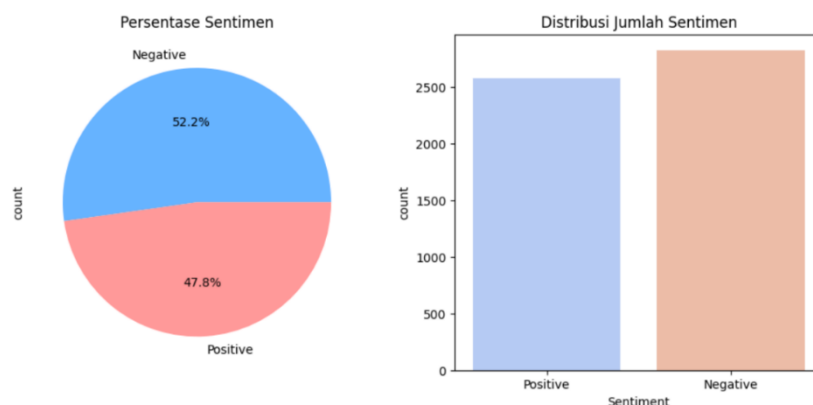
1. Deskripsi Data Penelitian

Dataset yang digunakan dalam penelitian ini adalah PRDECT-ID Dataset, yang berisi 5.400 ulasan dari 29 produk yang telah diberi label oleh psikolog klinis. Penelitian ini melakukan berfokus pada dua kolom utama, yaitu *Customer Review* dan *Sentimen*, untuk *Customer Review* berisi teks ulasan yang menjadi input utama dalam proses *preprocessing* dan ekstraksi fitur, sedangkan label sentimen dalam dua kategori, yaitu positif dan negatif. Pemilihan kedua kolom ini didasarkan pada hubungan langsung antara opini pelanggan dan sentimen yang diberikan. Tabel 2 menyajikan dataset PRDECT-ID yang digunakan.

TABEL 2. Beberapa List pada dataset PRDECT-ID

Customer Review	Sentimen
Alhamdulillah berfungsi dengan baik. Packaging aman. Respon cepat dan ramah. Seller dan kurir amanah	Positif
barang bagus dan respon cepat; harga bersaing dengan yg lain.	Positif
barang bagus; berfungsi dengan baik; seler ramah; pengiriman cepat	Positif
Sangat kecewa. Baru 4 bulan scroll sudah rusak.	Negatif
kecewa belanja di tempat kamu; lain di pesan lain dikirim ?	Negatif
gak bisa di pakai??	Negatif

Distribusi data kategori sentimen disajikan pada Gambar 1 dari 5400 ulasan, terdapat 47.8% (2579 baris) ulasan positif dan 52.2% (2821 baris) ulasan negatif. Distribusi data yang relatif seimbang antara kedua kategori sentimen ini memberikan peluang yang memadai untuk melatih model *machine learning* secara optimal pada kedua jenis sentimen.



Gambar 1. Distribusi Sentimen pada data PRDECT-ID

2. Preprocessing Data

Preprocessing data pada ulasan pelanggan dengan beberapa tahapan sistematis untuk mengubah teks mentah menjadi bentuk yang lebih terstruktur dan siap dianalisis. Proses diawali dengan *case folding*, yaitu mengubah ulasan menjadi huruf kecil agar konsisten. Selanjutnya dilakukan *cleaning* untuk menghapus tanda baca sehingga teks menjadi lebih bersih. Tahap *normalization* kemudian memperbaiki kata-kata yang tidak baku, seperti “respons” menjadi “respon”. Setelah itu, dilakukan *tokenization* dengan memecah teks menjadi unit kata. Pada tahap *stopword removal*, kata-kata umum yang tidak memiliki makna signifikan seperti “dengan” dan “dan” dihilangkan. Tahap berikutnya adalah *stemming*, yaitu mengubah kata ke bentuk dasar, misalnya “berfungsi” menjadi “fungsi” dan “packaging” menjadi “package”. Hasil akhir dari rangkaian proses ini adalah teks yang lebih ringkas dan informatif, yaitu “alhamdulillah fungsi baik package aman respon cepat ramah seller kurir amanah”, yang siap digunakan untuk analisis lebih lanjut seperti klasifikasi atau pemodelan machine learning. Berikut tahapan pre-pemrosesan data beserta hasil tahapan pre-pemrosesan ditampilkan pada Tabel 3.

TABEL 3. Tahapan pra-pemrosesan data beserta hasilnya

<i>Preprocessing</i>	Customer Reviewer
Teks Awal	Alhamdulillah berfungsi dengan baik. Packaging aman. Respons cepat dan ramah. Seller dan kurir Amanah
<i>Case Folding</i>	alhamdulillah berfungsi dengan baik. packaging aman. respons cepat dan ramah. seller dan kurir Amanah
<i>Cleaning</i>	alhamdulillah berfungsi dengan baik packaging aman respons cepat dan ramah seller dan kurir Amanah
<i>Normalization</i>	alhamdulillah berfungsi dengan baik packaging aman respon cepat dan ramah seller dan kurir amanah
<i>Tokenization</i>	['alhamdulillah', 'berfungsi', 'dengan', 'baik', 'packaging', 'aman', 'respon', 'cepat', 'dan', 'ramah', 'seller', 'dan', 'kurir', 'amanah']
<i>Stopword Removal</i>	['alhamdulillah', 'berfungsi', 'baik', 'packaging', 'aman', 'respon', 'cepat', 'ramah', 'seller', 'kurir', 'amanah']

Stemming	['alhamdulillah', 'fungsi', 'baik', 'package', 'aman', 'respon', 'cepat', 'ramah', 'seller', 'kurir', 'amanah']
Hasil Akhir	alhamdulillah fungsi baik package aman respon cepat ramah seller kurir amanah

3. Ekstraksi Fitur

Pada tahap ekstraksi fitur, teks ulasan yang telah melalui *preprocessing data* direpresentasikan ke dalam vektor numerik menggunakan metode TF-IDF dengan bantuan *TfidfVectorizer*. Hasil ekstraksi fitur menunjukkan bahwa kata “tidak” memiliki skor TF-IDF kumulatif tertinggi, diikuti oleh “barang”, “kirin”, “sesuai”, dan “bagus”, yang mencerminkan aspek utama dalam ulasan pelanggan dan berperan penting dalam membedakan sentimen positif dan negatif. Visualisasi kata dengan skor TF-IDF tertinggi dapat dilihat pada Gambar 2, yang menunjukkan kata-kata dominan dengan bobot tertinggi dan berkontribusi signifikan dalam proses klasifikasi sentimen.



Gambar 2. Distribusi Kata Data Training dengan Skor TF-IDF

4. Modelling

Langkah pertama dalam pemodelan adalah membagi dataset menggunakan metode *stratified random split* menjadi data pelatihan dan data pengujian dengan perbandingan 80% data latih dan 20% data uji, sehingga proporsi kelas sentimen tetap terjaga pada kedua subset data. Selanjutnya, penelitian ini menerapkan dua model klasifikasi, yaitu Random Forest dan Naïve Bayes, untuk mengklasifikasikan sentimen ulasan ke dalam kategori Negatif dan Positif guna membandingkan kinerja kedua model.

a. Random Forest:

Model Random Forest diterapkan menggunakan fitur TF-IDF yang telah dihasilkan pada tahap sebelumnya. Model diinisialisasi dengan parameter bawaan dari *library scikit-learn* dan menggunakan `random_state = 42` untuk menjaga konsistensi hasil pelatihan.

b. Multinomial Naïve Bayes:

Model Multinomial Naive Bayes yang digunakan dalam penelitian ini diimplementasikan menggunakan kelas *MultinomialNB* dengan parameter `alpha=0.1`, di mana nilai tersebut berfungsi untuk mengatasi kemungkinan probabilitas nol pada fitur yang jarang muncul. Pendekatan ini sangat cocok untuk data teks berbasis frekuensi seperti hasil TF-IDF, karena mengasumsikan bahwa distribusi fitur mengikuti distribusi multinomial. Dengan nilai `alpha` yang relatif kecil, model tetap

memberikan keteraturan namun tidak terlalu mengaburkan distribusi asli data, sehingga dapat meningkatkan performa klasifikasi, khususnya dalam tugas analisis sentimen.

5. Evaluasi Model

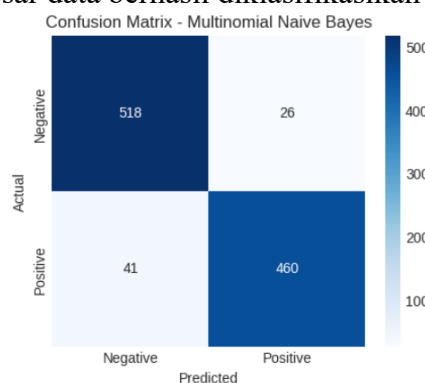
Pada tahap evaluasi model Random Forest dan Multinomial Naïve Bayes diuji menggunakan data uji yang terpisah dari data pelatihan untuk mengukur kemampuan klasifikasi sentimen dataset PRDECT-ID ke dalam kelas Positif dan Negatif. Secara rinci Tabel 4 menjelaskan hasil evaluasi metrik *accuracy*, *precision*, *recall*, dan *F1-score* dari kemampuan masing masing model

TABEL 4. Hasil performa pada dataset *testing*

Model	Akurasi	Recall	Precision	F1-Score	ROC-AUC
Naïve Bayes	0.935885	0.946502	0.918164	0.932118	0.981313
Random Forest	0.903349	0.936681	0.856287	0.894682	0.963488

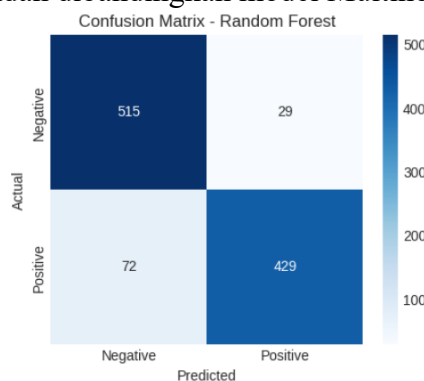
Berdasarkan hasil evaluasi, Naïve Bayes menunjukkan hasil yang baik dalam mengklasifikasikan sentimen ulasan dengan *accuracy* 93% pada data uji. Model Naïve Bayes mencapai *precision* 0,92, *recall* 0,95, dan *F1-score* 0,93. Adapun model Random Forest menunjukkan kinerja baik dalam mengklasifikasikan sentimen ulasan dengan *accuracy* 90% pada data uji. Model mencapai *precision* 0,85, *recall* 0,93, dan *F1-score* 0,89. Berdasarkan hasil evaluasi metrik ROC-AUC bahwa model Multinomial Naive Bayes dan Random Forest mengklasifikasikan dengan sangat baik untuk membedakan sentimen positif dan negatif. Model Multinomial Naive Bayes memperoleh nilai ROC-AUC sebesar 0,9813, sedangkan Random Forest memperoleh nilai 0,9635. Nilai ROC-AUC yang mendekati 1 menunjukkan kedua model memiliki kemampuan diskriminasi yang tinggi dalam klasifikasi data. Namun, jika dibandingkan langsung, Multinomial Naive Bayes menunjukkan performa yang lebih baik dibandingkan Random Forest, karena ROC-AUC yang lebih tinggi, sehingga model tersebut optimal dalam memisahkan kedua kelas sentimen pada dataset yang digunakan dalam penelitian ini. Sehingga Model Naïve Bayes memperoleh akurasi sedikit lebih unggul dibandingkan Random Forest.

Selanjutnya kedua model dievaluasi menggunakan *confusion Matrix*. Gambar 3 merupakan *confusion matrix* pada model Multinomial Naive Bayes. Pada model Multinomial Naive Bayes, terdapat 518 data sentimen negatif yang diprediksi dengan benar (*True Negative*) dan 460 data sentimen positif yang diprediksi dengan benar (*True Positive*). Sementara itu, terdapat 26 data negatif yang salah diprediksi sebagai positif (*False Positive*) serta 41 data positif yang salah diprediksi sebagai negatif (*False Negative*). Berdasarkan nilai tersebut diperoleh *accuracy* sebesar 93,59%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh model.



Gambar 3. *Confusion Matrix* – Multinomial Naïve Bayes

Sementara itu, Gambar 4 merupakan *confusion matrix* pada model Random Forest, terdapat 515 data sentimen negatif yang diprediksi dengan benar (*True Negative*) dan 429 data sentimen positif yang diprediksi dengan benar (*True Positive*). Namun model ini menghasilkan 29 kesalahan ketika data negatif diprediksi sebagai positif (*False Positive*) serta 72 kesalahan ketika data positif diprediksi sebagai negatif (*False Negative*). Berdasarkan hasil tersebut diperoleh *accuracy* sebesar 90,35%, yang masih baik namun lebih rendah dibandingkan model Multinomial Naive Bayes.



Gambar 4. Confusion Matrix – Random Forest

Berdasarkan *confusion matrix* yang ditampilkan, dilakukan analisis kinerja model dengan membandingkan Multinomial Naive Bayes dan Random Forest menggunakan metrik *precision*, *recall*, dan *F1-score* pada masing-masing sentiment yang ditunjukkan pada table 5.

TABEL 5. Hasil performa pada dataset pada masing-masing sentimen

Model	Sentimen	Presisi	Recall	F1-Score
Random Forest	Positif	94,65%	91,82%	93,21%
	Negatif	92,67%	95,22%	93,93%
Naïve Bayes	Positif	93,67%	85,63%	89,47%
	Negatif	87,73%	94,67%	91,06%

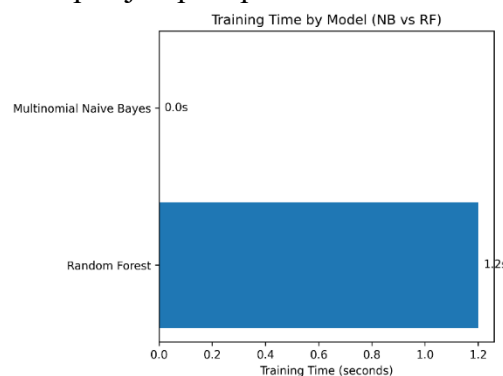
Jika dianalisis berdasarkan masing-masing kelas sentimen, Model Random Forest menunjukkan pada sentimen positif diperoleh *precision* sebesar 94,65%, yang menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan model benar merupakan sentimen positif. Nilai *recall* sebesar menunjukkan bahwa model mampu mendeteksi sebagian besar data yang benar-benar memiliki sentimen positif. Kombinasi kedua metrik tersebut menghasilkan *F1-score* sebesar 93,21%, yang menandakan keseimbangan yang baik antara ketepatan dan kemampuan deteksi model pada kelas positif. Untuk sentimen negatif, diperoleh *precision* sebesar 92,67%, yang berarti sebagian besar prediksi negatif adalah benar. Nilai *recall* sebesar 95,22% menunjukkan bahwa model sangat baik dalam mengenali data yang benar-benar negatif. Dari kedua nilai tersebut diperoleh *F1-score* sebesar 93,93%, yang mengindikasikan performa klasifikasi yang sangat baik pada kelas negatif.

Model Naïve Bayes menunjukkan sentimen positif diperoleh *precision* sebesar 93,67%, yang menunjukkan bahwa sebagian besar prediksi positif cukup akurat. Akan tetapi *recall* sebesar 85,63% menunjukkan bahwa masih terdapat cukup banyak data positif yang tidak berhasil terdeteksi oleh model. Nilai tersebut menghasilkan *F1-score* sebesar 89,47%, yang menunjukkan keseimbangan performa yang cukup baik tetapi masih lebih rendah dibandingkan model Multinomial Naive Bayes. Untuk sentimen negatif, diperoleh *precision* sebesar 87,73%, yang berarti masih terdapat beberapa prediksi negatif yang sebenarnya merupakan data positif. Sementara itu *recall* sebesar 94,67%

menunjukkan bahwa model cukup baik dalam mendeteksi data yang benar-benar negatif. Kombinasi kedua nilai tersebut menghasilkan *F1-score* sebesar 91,06%.

Secara keseluruhan, hasil perbandingan menunjukkan bahwa Multinomial Naive Bayes memiliki performa yang lebih unggul dibandingkan Random Forest. Hal ini terlihat dari nilai *accuracy* yang lebih tinggi, serta nilai *recall* dan *F1-score* pada sentimen positif yang lebih baik, yang menunjukkan bahwa model tersebut lebih efektif dalam mendeteksi ulasan dengan sentimen positif. Dengan demikian, dalam penelitian ini Multinomial Naive Bayes lebih optimal digunakan untuk tugas klasifikasi sentimen dibandingkan Random Forest.

Selanjutnya pengujian hasil pengukuran waktu pelatihan model, terlihat bahwa Multinomial Naive Bayes memiliki waktu pelatihan, mendekati 0 detik, sedangkan Random Forest memerlukan waktu pelatihan sekitar 1,2 detik. Perbedaan ini menunjukkan variasi dalam tingkat kompleksitas algoritma yang digunakan oleh kedua model. Multinomial Naive Bayes memiliki mekanisme perhitungan yang relatif sederhana karena menghitung probabilitas kemunculan kata pada setiap kelas berdasarkan prinsip teorema Bayes, sehingga proses pelatihannya dilakukan dengan cepat dan efisien. Sebaliknya, Random Forest memiliki struktur model yang kompleks karena kumpulan banyak decision tree yang dibangun melalui proses *bagging* serta pemilihan fitur secara acak, sehingga membutuhkan waktu komputasi yang lebih lama. Oleh karena itu, pemilihan model sebaiknya disesuaikan dengan kebutuhan sistem, tujuan analisis, serta keseimbangan antara efisiensi komputasi dan kemampuan model dalam mempelajari pola pada data.



Gambar 5. *Training Time* oleh model

Secara praktis, hasil analisis sentimen dalam penelitian ini dapat dimanfaatkan oleh platform *e-commerce*, khususnya Tokopedia, untuk memantau opini dan tanggapan konsumen secara lebih otomatis dan sistematis. Informasi mengenai sentimen positif dan negatif yang terdapat pada ulasan produk dapat membantu dalam mengidentifikasi produk atau penjual yang memiliki tingkat keluhan yang tinggi. Selain itu, hasil analisis tersebut juga dapat digunakan sebagai bahan evaluasi terhadap kualitas layanan serta sebagai dasar dalam meningkatkan sistem rekomendasi dan pelayanan kepada pelanggan. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam aspek akademis, tetapi juga memiliki manfaat praktis dalam mendukung peningkatan pengelolaan serta kualitas layanan pada platform *e-commerce*.

Hasil penelitian ini menunjukkan bahwa Multinomial Naive Bayes memiliki performa yang lebih optimal dibandingkan Random Forest dalam klasifikasi sentimen ulasan produk Tokopedia. Hal tersebut ditunjukkan oleh nilai *accuracy* sebesar 93,59%, *precision* 91,82%, *recall* 94,65%, *F1-score* 93,21%, dan ROC-AUC 0,9813 pada model Multinomial Naive Bayes, sedangkan Random Forest

memperoleh *accuracy* 90,35%, *precision* 85,63%, *recall* 93,67%, *F1-score* 89,47%, dan ROC-AUC 0,9635. Keunggulan tersebut dipengaruhi oleh karakteristik data teks hasil ekstraksi TF-IDF yang bersifat sparse dan berdimensi tinggi, sehingga lebih sesuai dengan pendekatan probabilistik pada Multinomial Naïve Bayes. Selain itu, kompleksitas perhitungan yang relatif sederhana menyebabkan waktu pelatihan model menjadi lebih cepat dibandingkan Random Forest. Hasil penelitian ini juga sejalan dengan beberapa penelitian terdahulu yang menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang baik dalam klasifikasi sentimen pada data teks dan ulasan e-commerce. Meskipun demikian, perbedaan performa antar penelitian dapat dipengaruhi oleh karakteristik dataset, distribusi data, serta metode ekstraksi fitur yang digunakan. Pada penelitian ini, penggunaan fitur TF-IDF pada dataset PRDECT-ID yang memiliki karakteristik teks terstruktur dan berdimensi tinggi lebih sesuai dengan pendekatan Multinomial Naïve Bayes, sehingga menghasilkan performa klasifikasi yang lebih optimal dibandingkan Random Forest.

D. Kesimpulan

Penelitian ini membandingkan performa model Multinomial Naive Bayes dan Random Forest dalam melakukan klasifikasi sentimen terhadap ulasan produk Tokopedia. Hasil penelitian menunjukkan bahwa Multinomial Naive Bayes memberikan kinerja yang lebih baik dibandingkan Random Forest berdasarkan beberapa metrik evaluasi. Model Multinomial Naive Bayes memperoleh *accuracy* sebesar 93,59%, dengan *precision* 94,65%, *recall* 91,82%, dan *F1-score* 93,21% pada kelas sentimen positif, serta nilai ROC-AUC sebesar 0,9813. Sementara itu, model Random Forest memperoleh *accuracy* sebesar 90,35%, dengan *precision* 93,67%, *recall* 85,63%, dan *F1-score* 89,47% pada kelas sentimen positif, serta ROC-AUC sebesar 0,9635. Hasil tersebut menunjukkan bahwa Multinomial Naive Bayes memiliki kemampuan yang lebih baik dalam mengklasifikasikan sentimen positif dan negatif pada data ulasan produk.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Dataset yang hanya terbatas pada dua kategori sentimen, yaitu positif dan negatif, sehingga belum mempertimbangkan kemungkinan adanya sentimen netral. Selain itu, penelitian ini hanya membandingkan dua model klasifikasi, yaitu Multinomial Naive Bayes dan Random Forest, sehingga belum mengeksplorasi metode lain yang berpotensi memberikan performa yang lebih baik. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas penggunaan dataset yang lebih beragam serta melakukan perbandingan dengan metode klasifikasi lainnya agar diperoleh hasil analisis sentimen yang lebih komprehensif dan representatif.

Daftar Pustaka

- Aisyah, A., Siregar, H., & Apriliani, R. R. (2025). Analisis Korelasi Statistik Antara Populasi Jumlah Penduduk dan Pengguna Internet Di Indonesia. *Journal of Artificial Intelligence and Digital Business (RIGGS)*, 4(3), 4776–4781. <https://doi.org/10.31004/riggs.v4i3.2684>
- Aji, B. P., Sri, C., & Aditya, K. (2025). Klasifikasi Sentimen Ulasan Produk pada Platform E-Commerce di Indonesia dengan Menggunakan Model Pre-Trained IndoBERT. *Building of Informatics, Technology and Science*, 6(4). <https://doi.org/10.47065/bits.v6i4.6968>
- Andreyestha, & Azizah, Q. N. (2022). Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE. *Infotek : Jurnal*

Informatika Dan Teknologi, 5(1), 108–116. <https://doi.org/10.29408/jit.v5i1.4581> e-ISSN

- Atimi, R. L., & Pratama, E. E. (2022). Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia. *Jurnal Sains Dan Informatika*, 8, 88–96. <https://doi.org/10.34128/jsi.v8i1.419>
- Azis, A. R. (2024). Analisis Komparasi Algoritma Machine Learning dalam Prediksi Performa Akademik Mahasiswa : Literature Review. *Jurnal Ilmu Komputer Dan Informatika (JIKI)*, 4(2), 143–150. <https://doi.org/10.54082/jiki.212>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal Tekno Kompak*, 15(1), 131–145. <https://doi.org/10.33365/jtk.v15i1.744>
- Dewi, C., Chen, R., Christanto, H. J., & Cauteruccio, F. (2023). Multinomial Naive Bayes Classifier for Sentiment Analysis of Internet Movie Database. *Vietnam Journal of Computer Science*, 10(4), 485–498. <https://doi.org/10.1142/S2196888823500100>
- Enjelia, L., Cahyana, Y., Rahmat, & Wahiddin, D. (2025). Comparison of K-Nearest Neighbors and Naive Bayes Classifier Algorithms in Sentiment Analysis of 2024 Election in Twitter (X). *Journal of Applied Informatics and Computing (JAIC)*, 9(3), 946–954. <https://doi.org/10.30871/jaic.v9i3.9593>
- Hasugian, A. H., Fakhriza, M., & Zukhoiriyah, D. (2023). Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 6, 98–107. <https://doi.org/10.53513/jsk.v6i1.7400>
- Iqbal, M., Abdullah, D., & Afrillia, Y. (2025). Penerapan Algoritma Random Forest untuk Menentukan Kelayakan Penerima BLT. *RABIT: Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2), 1220–1230. <https://doi.org/10.36341/rabit.v10i2.6503>
- Kembau, A. S., Putri, J. G. E., & Makarawung, R. J. N. (2025). Customer Indecisiveness pada E-Commerce Indonesia: Peran Price Sensitivity, Product Involvement, Risk, dan Social Inference. *Indonesian Journal of Digital Business Journal*, 5(2), 460–472. <https://doi.org/10.17509/ijdb.v5i2.86161>
- Nawawi, H. M., Hikmah, A. B., Mustopa, A., & Wijaya, G. (2024). Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir. *Jurnal Saintekom : Sains, Teknologi, Komputer Dan Manajemen*, 14(1), 13–25. <https://doi.org/10.33020/saintekom.v14i1.512>
- Pandu W, M. A., Saputro, R. E., & Rohmah, U. A. (2025). Analisis Tingkat Akurasi Metode Naive Bayes dan Random Forest dalam Prediksi Penjualan Emas. *Jurnal Pendidikan Dan Teknologi Indonesia (JPTI)*, 5(7), 1809–1821. <https://doi.org/10.52436/1.jpti.732>
- Panggabean, S., & Junika, A. (2024). Sentiment Analysis on Public Opinions Regarding the 2024 Regional Elections Using Long Short-Term Memory (LSTM), Random Forest , and Naive Bayes. *JOISTECH: Journal of Information System and Technology*, 01(02), 67–75.

- Rayhan, F. M., Wijoyo, S. H., Hayuhardhika, W., Putra, N., Studi, P., Informatika, T., Komputer, F. I., Brawijaya, U., Ulasan, S., Forest, R., & Analysis, R. C. (2018). ANALISIS SENTIMEN ROOT CAUSE ANALISIS KEPUASAN PENGGUNA APLIKASI TOKOPEDIA PADA ULASAN MENGGUNAKAN METODE RANDOM FOREST. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(1).
- Saadah, S., Auditama, K. M., Fattahila, A. A., & Amorokhman, F. I. (2026). Implementation of BERT , IndoBERT , and CNN-LSTM in Classifying. *JURNAL RESTI(Rekayasa Sistem Dan Teknologi Informasi)*, 5(158), 2–7. <https://doi.org/10.29207/resti.v6i4.4215>
- Saputra, F. D., & Budiman, F. (2026). Comparison of Random Forest and LSTM for Tokopedia Sentiment Analysis. *Journal of Applied Informatics and Computing (JAIC)*, 10(1), 630–639. <https://doi.org/10.30871/jaic.v10i1.12042>
- Sutarso, Y., Suminar, B., & Ilfitriah, A. M. (2024). Do shopping anxiety and data leakage risks matter to e- commerce customers ? Evidence from the largest economy in Southeast Asia. *Jurnal Manajemen Dan Pemasaran Jasa*, 17(1), 97–116. <https://doi.org/10.25105/jmpj.v17i1.18673>
- Sutoyo, R., Achmad, S., Chowanda, A., Widhi, E., & Isa, S. M. (2022). PRDECT-ID : Indonesian product reviews dataset for emotions classification tasks. *Data in Brief*, 44, 108554. <https://doi.org/10.1016/j.dib.2022.108554>
- Syah, A., Nurdiyansyah, F., & Rahman, A. Y. (2024). Analisis Sentimen Aplikasi Shopee, Tokopedia, Lazada dan Blibli Menggunakan Leksikon dan Random Forest. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 12(3). <https://doi.org/10.23960/jitet.v12i3S1.5155>