JATI (Jurnal Jaringan dan Teknologi Informasi)
Vol. 4, No. 2, Juni 2026, hlm. 49-56
DOI: e-ISSN: 3063-9751

IMPLEMENTASI MULTINOMIAL NAIVE BAYES DALAM DETEKSI EMOSI PADA TEKS BAHASA MELAYU TERNATE

Nuraida Ibrahim[1], Abdul Mubarak[2*], Muhammad Fhadli[3]

[1,3]Program Studi Informatika, Fakultas Teknik, Universitas Khairun, Jl. Jati Metro, Kota Ternate
[2]Program Studi Pendidikan Profesi Insinyur, Fakultas Teknik, Universitas Khairun, Jl. Yusuf Abdurrahman, Kota Ternate

Email: [1]nuraidaibrahim21@gmail.com, [2]amuba029@unkhair.ac.id, [3]mfha@unkhair.ac.id

Abstrak

Bahasa Melayu Ternate merupakan salah satu bahasa yang digunakan dalam komunikasi sehari-hari masyarakat Maluku Utara dan memiliki karakteristik kosakata serta ekspresi emosional yang khas. Penelitian deteksi emosi berbasis teks telah berkembang pada Bahasa Indonesia, sedangkan kajian pada Bahasa Melayu Ternate masih terbatas. Penelitian ini bertujuan mengimplementasikan Multinomial Naive Bayes untuk mengklasifikasikan teks Bahasa Melayu Ternate ke dalam empat kategori emosi, yaitu Bahagia, Marah, Sedih, dan Takut. Dataset awal terdiri atas 499 teks dengan distribusi 127 data Bahagia, 129 Marah, 130 Sedih, dan 113 Takut. Tahapan penelitian meliputi pelabelan, preprocessing, augmentasi data, representasi Term Frequency-Inverse Document Frequency (TF-IDF), pembagian dataset, pelatihan Multinomial Naive Bayes, dan evaluasi. Enam teknik augmentasi meningkatkan jumlah dataset menjadi 1.040 data dengan distribusi yang sama pada setiap kelas. TF-IDF dikonfigurasi dengan maksimal 2.000 fitur dan n-gram hingga trigram, sedangkan Multinomial Naive Bayes menggunakan parameter alpha 0,5. Berdasarkan skenario eksperimen yang digunakan, pengujian terhadap 208 data memperoleh *accuracy* 88,46%, macro *precision* 89,19%, macro *recall* 88,46%, dan macro F1-score 88,59%. Hasil menunjukkan bahwa kombinasi TF-IDF dan Multinomial Naive Bayes dapat diterapkan untuk klasifikasi empat kategori emosi pada dataset Bahasa Melayu Ternate yang digunakan dalam penelitian ini.

Kata kunci: *bahasa Melayu Ternate, deteksi emosi, Multinomial Naive Bayes, TF-IDF, klasifikasi teks*

MULTINOMIAL NAIVE BAYES IMPLEMENTATION FOR EMOTION DETECTION IN TERNATE MALAY TEXT

Abstract

Ternate Malay is widely used in daily communication in North Maluku and contains distinctive vocabulary and emotional expressions. Text-based emotion detection research has increasingly been conducted for Indonesian, while studies focusing on Ternate Malay remain limited. This study aims to implement Multinomial Naive Bayes to classify Ternate Malay texts into four emotion categories: happiness, anger, sadness, and fear. The initial dataset consisted of 499 texts comprising 127 happiness, 129 anger, 130 sadness, and 113 fear samples. The research stages included labeling, preprocessing, data augmentation, Term Frequency-Inverse Document Frequency (TF-IDF) representation, dataset splitting, Multinomial Naive Bayes training, and evaluation. Six augmentation techniques increased the dataset to 1,040 samples with an equal distribution across all classes. TF-IDF was configured with a maximum of 2,000 features and n-grams up to trigrams, while Multinomial Naive Bayes used an alpha value of 0.5. Under the experimental scenario employed in this study, evaluation on 208 test samples yielded an accuracy of 88.46%, macro precision of 89.19%, macro recall of 88.46%, and macro F1-score of 88.59%. The results indicate that the combination of TF-IDF and Multinomial Naive Bayes can be applied to four-class emotion classification on the Ternate Malay dataset used in this study.

Keywords: *Ternate Malay, emotion detection, Multinomial Naive Bayes, TF-IDF, text classification*

1. PENDAHULUAN

Bahasa merupakan sarana komunikasi sekaligus bagian dari identitas sosial dan budaya masyarakat. Di Maluku Utara terdapat berbagai bahasa daerah dan ragam Melayu lokal yang digunakan dalam komunikasi sehari-hari. Penelitian mengenai pemertahanan Bahasa Tidore menunjukkan pentingnya penggunaan bahasa daerah dalam komunikasi antargenerasi untuk menjaga keberlangsungan bahasa lokal [1]. Meskipun objek bahasa tersebut berbeda dengan Bahasa Melayu Ternate, kondisi tersebut menunjukkan pentingnya perhatian terhadap bahasa lokal Maluku Utara, termasuk dalam pengembangan teknologi berbasis bahasa.

Bahasa Melayu Ternate merupakan ragam Melayu yang digunakan sebagai salah satu bahasa pengantar dalam komunikasi masyarakat Maluku Utara. Bahasa ini berkembang melalui kontak bahasa yang panjang serta memiliki kosakata dan bentuk ekspresi yang berbeda dengan Bahasa Indonesia standar. Berdasarkan penelitian terdahulu yang ditinjau, kajian deteksi emosi dan sumber daya NLP yang secara khusus mendukung Bahasa Melayu Ternate, seperti korpus dan daftar stopword, masih relatif terbatas.

Selain menyampaikan informasi, teks dapat merepresentasikan keadaan emosional seseorang. Dalam bidang *Natural Language Processing* (NLP), identifikasi emosi berdasarkan teks dikenal sebagai *emotion detection* atau *emotion classification*. Kajian deteksi emosi mencakup berbagai pendekatan, mulai dari metode berbasis aturan hingga machine learning dan deep learning [2]. Emosi dalam teks juga tidak selalu dinyatakan secara eksplisit melalui kata tertentu, tetapi dapat dipengaruhi oleh peristiwa, situasi, dan konteks yang terdapat dalam teks [3].

Deteksi emosi memiliki fokus yang berbeda dengan analisis sentimen. Analisis sentimen umumnya mengidentifikasi polaritas seperti positif, negatif, dan netral, sedangkan deteksi emosi mengidentifikasi keadaan emosional yang lebih spesifik. Penelitian ini menggunakan empat kategori, yaitu Bahagia, Marah, Sedih, dan Takut, sesuai dengan proses anotasi dataset penelitian.

Penelitian deteksi emosi pada Bahasa Indonesia telah dilakukan menggunakan berbagai pendekatan. Bata dkk. mengembangkan leksikon untuk mendukung deteksi emosi dari teks Bahasa Indonesia [4]. Ardiada dkk. menerapkan *Support Vector Machine* dan *K-Nearest Neighbour* untuk mendeteksi emosi pengguna berdasarkan teks media sosial [5]. Fadjeri dkk. juga menerapkan Naive Bayes pada deteksi emosi berbasis teks [6].

Wulandari dkk. menggunakan Naive Bayes untuk mendeteksi emosi pada teks terkait kuliah daring dan melaporkan *accuracy* sebesar 80% pada tiga kelas, 73,20% pada empat kelas, dan 60,87% pada lima kelas emosi [7]. Penelitian yang lebih baru oleh Pane dkk. menggunakan pendekatan ensemble yang menggabungkan SVM, KNN, dan XGBoost serta melaporkan *accuracy* sebesar 87,14% pada teks Bahasa Indonesia [8].

Ketersediaan sumber daya klasifikasi emosi Bahasa Indonesia juga mulai berkembang. *Emotion Dataset from Indonesian Public Opinion* menyediakan dataset Bahasa Indonesia dengan enam kategori emosi [9]. Selain itu, EmoTweetID menyediakan sumber daya tweet Bahasa Indonesia untuk klasifikasi emosi dan pengembangan representasi kata [10]. Perkembangan tersebut menunjukkan bahwa penelitian dan sumber daya deteksi emosi Bahasa Indonesia mulai meningkat, sedangkan sumber daya khusus untuk Bahasa Melayu Ternate masih relatif terbatas.

Dengan demikian, *research gap* penelitian ini tidak terletak pada kebaruan algoritma Naive Bayes. Algoritma tersebut telah banyak digunakan dalam klasifikasi teks. Gap penelitian terletak pada penerapan dan evaluasi klasifikasi emosi terhadap teks Bahasa Melayu Ternate sebagai bahasa lokal dengan keterbatasan sumber daya NLP, menggunakan dataset yang dikumpulkan dan diberi label ke dalam empat kategori emosi.

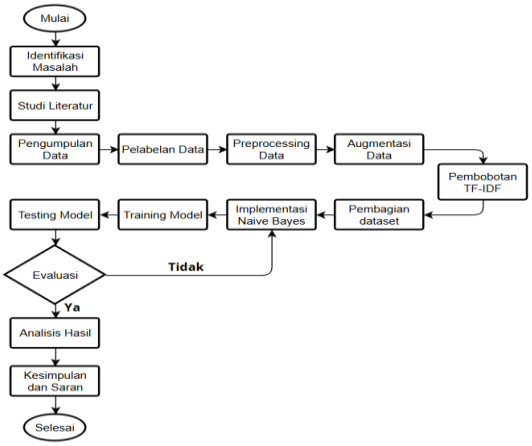
Pada proses klasifikasi, teks perlu direpresentasikan dalam bentuk numerik. Penelitian ini menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), sedangkan klasifikasi dilakukan menggunakan Multinomial Naive Bayes. TF-IDF memberikan bobot terhadap term berdasarkan kemunculannya pada dokumen dan keseluruhan koleksi, sedangkan Multinomial Naive Bayes menggunakan distribusi fitur untuk memperkirakan probabilitas suatu teks termasuk dalam kelas tertentu [11].

Dataset awal penelitian relatif terbatas sehingga digunakan augmentasi teks untuk menambah jumlah dan variasi data. Teknik augmentasi sederhana seperti *random insertion*, *random swap*, dan *random deletion* telah digunakan dalam *Easy Data Augmentation* untuk mendukung klasifikasi teks pada kondisi data terbatas [12].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengimplementasikan Multinomial Naive Bayes dalam mengklasifikasikan empat kategori emosi pada teks Bahasa Melayu Ternate menggunakan representasi TF-IDF dan mengevaluasi hasilnya menggunakan *accuracy*, *precision*, *recall*, F1-score, serta confusion matrix.

1. METODE PENELITIAN

Tahapan penelitian terdiri atas pengumpulan data, pelabelan, *preprocessing* teks, augmentasi data, representasi TF-IDF, pembagian dataset, pelatihan Multinomial Naive Bayes, dan evaluasi model.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan dan Pelabelan Dataset

Dataset terdiri atas teks percakapan Bahasa Melayu Ternate. Data dikumpulkan melalui survei langsung kepada masyarakat, wawancara untuk memperoleh contoh kalimat emosional, dan observasi terhadap percakapan masyarakat baik secara langsung maupun melalui media sosial. Struktur dataset memuat teks Bahasa Melayu Ternate, terjemahan Bahasa Indonesia sebagai referensi, dan kategori emosi.

Penelitian berhasil mengumpulkan 499 data yang terdiri atas empat kelas emosi. Distribusinya ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Dataset Awal

No.	Kategori Emosi	Jumlah Data	Persentase
1	Bahagia	127	25,45%
2	Marah	129	25,85%
3	Sedih	130	26,05%
4	Takut	113	22,65%
	Total	499	100%

Distribusi awal relatif berimbang. Kelas terbesar adalah Sedih sebanyak 130 data, sedangkan kelas terkecil adalah Takut sebanyak 113 data. Rasio kelas terbesar dan terkecil sekitar 1,15:1. Oleh karena itu, augmentasi lebih tepat dipahami sebagai upaya meningkatkan jumlah dan variasi data sekaligus menyeragamkan distribusi kelas, bukan sebagai penanganan ketidakseimbangan yang ekstrem.

Pelabelan dilakukan secara manual oleh dua anotator yang memiliki latar belakang bidang bahasa. Kategori Marah mencakup kemarahan, kejengkelan, dan nada mengancam; Bahagia mencakup kegembiraan, kepuasan, dan antusiasme; Sedih mencakup kesedihan, kekecewaan, dan keadaan melankolis; sedangkan Takut mencakup ketakutan, kecemasan, dan rasa terancam. Pada teks yang dianggap ambigu dilakukan diskusi untuk menentukan label yang paling sesuai.

Pelibatan lebih dari satu anotator penting dalam anotasi emosi karena interpretasi terhadap teks dapat bersifat subjektif. Dataset emotion classification yang lebih baru juga mendokumentasikan kualitas anotasi menggunakan pengukuran kesepakatan anotator [10]. Penelitian ini belum menghitung nilai *inter-annotator agreement*, sehingga aspek tersebut menjadi salah satu keterbatasan yang perlu diperbaiki pada penelitian berikutnya.

2.2. *Preprocessing* Teks

Preprocessing dilakukan terhadap seluruh data untuk membersihkan dan menyeragamkan bentuk teks sebelum representasi fitur. Tahapan yang digunakan terdiri atas *cleansing*, *case folding*, *tokenizing*, dan *stopword removal*.

Cleansing digunakan untuk menghapus tanda baca dan karakter yang tidak diperlukan. *Case folding* mengubah seluruh karakter menjadi huruf kecil. *Tokenizing* memecah teks menjadi token kata, sedangkan *stopword removal* menghapus kata yang dianggap kurang informatif.

Tabel 2. Contoh Hasil *Preprocessing*

Teks Asli	**Hasil *Preprocessing***	**Emosi**
Jang bagitu tara itu kalo kita tiba-tiba mati di jalan kong, ngana mo kase hidop?	jang bagitu tara kalo kita tibatiba mati jalan kong ngana mo kase hidop	Marah
Ngana tau kita dapa kase kado deng dia di dekat kampus de dong pe angkatan	ngana tau kita dapa kase kado deng dia dekat kampus de dong pe angkatan	Bahagia
dia hidop so tara ada orang tua e kasiang skali	dia hidop tara orang tua e kasiang skali	Sedih
tolong kitaa ada papancuri seng dia nae di atas pohon tu tangka sadiki ka	tolong kitaa papancuri seng dia nae atas pohon tangka sadiki ka	Takut

Contoh pada Tabel 2 diambil dari hasil *preprocessing* penelitian.

Proses *stopword removal* pada Bahasa Melayu Ternate belum sepenuhnya optimal karena belum tersedia korpus digital dan daftar stopword baku yang terdokumentasi secara luas. Penggunaan stopword list juga perlu dilakukan secara hati-hati karena daftar pada perangkat lunak NLP tidak selalu sesuai dengan karakteristik bahasa atau tokenizer yang digunakan [13].

2.3. Augmentasi Data

Augmentasi diterapkan setelah *preprocessing*. Penelitian menggunakan strategi target sebanyak dua kali jumlah kelas terbesar. Karena kelas terbesar terdiri atas 130 data, setiap kategori ditargetkan menjadi 260 data.

Enam teknik augmentasi digunakan sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Teknik Augmentasi Data

No.	Teknik	Probabilitas
1	Random Word Swap	20%
2	Random Word Deletion	15%
3	Word Duplication	20%
4	Random Insertion	15%
5	Word Replacement	15%
6	Multiple Operations	15%

Random Word Swap menukar posisi kata, *Random Word Deletion* menghapus kata secara acak, *Word Duplication* menduplikasi kata, *Random Insertion* menyisipkan kata pada posisi tertentu, *Word Replacement* mengganti kata dengan kata lain yang terdapat dalam kalimat, sedangkan *Multiple Operations* menggabungkan beberapa teknik tersebut.

Sebagian teknik tersebut memiliki prinsip yang sejalan dengan EDA yang menggunakan operasi sederhana seperti insertion, swap, dan deletion untuk meningkatkan variasi data klasifikasi teks [12].

Setelah proses augmentasi, jumlah dataset meningkat dari 499 menjadi 1.040 data.

Tabel 4. Distribusi Data Sebelum dan Setelah Augmentasi

Kategori	Sebelum	Setelah	Penambahan
Bahagia	127	260	133
Marah	129	260	131
Sedih	130	260	130
Takut	113	260	147
Total	499	1.040	541

Setelah augmentasi, masing-masing kategori memiliki 260 data atau 25% dari keseluruhan dataset.

2.4. Representasi TF-IDF

TF-IDF digunakan untuk mengubah teks menjadi representasi numerik. Secara umum bobot TF-IDF dirumuskan sebagai:

$$\mathrm{TFIDF(t,d)} = \mathrm{TF(t,d)} \times \mathrm{IDF(t)} \quad (1)$$

dengan $TF(t, d)$merupakan frekuensi term t pada dokumen d, sedangkan $IDF(t)$menunjukkan tingkat kelangkaan term pada keseluruhan dokumen [11].

Tabel 5. Konfigurasi Eksperimen

Parameter	Nilai
Dataset awal	499
Dataset setelah augmentasi	1.040
Train-test split	80:20
Data *training*	832
Data *testing*	208
max_features	2.000
min_df	1
max_df	0,9
ngram_range	(1,3)
sublinear_tf	True
Klasifikator	Multinomial Naive Bayes
Alpha	0,5

Tabel 5 menunjukkan konfigurasi TF-IDF menghasilkan matriks berukuran 1.040 × 2.000 dengan sparsitas sekitar 99,16%. Penggunaan ngram_range = (1,3) memungkinkan fitur berupa unigram, bigram, dan trigram.

2.5. Pembagian Dataset dan Multinomial Naive Bayes

Dataset yang telah melalui augmentasi dan representasi fitur dibagi menjadi 80% data *training* dan 20% data *testing* menggunakan stratified split. *Training* set terdiri atas 832 data atau 208 data pada setiap kategori, sedangkan *testing* set terdiri atas 208 data atau 52 data pada setiap kategori.

Naive Bayes merupakan metode klasifikasi probabilistik berdasarkan Teorema Bayes. Probabilitas posterior suatu kelas dapat dituliskan sebagai:

$$P(C \mid X) = \frac{P(X|C)P(C)}{P(X)} \quad (2)$$

dengan C merupakan kelas emosi dan Xmerupakan fitur dokumen [11].

Prediksi ditentukan berdasarkan kelas dengan nilai probabilitas posterior tertinggi:

$$C* = argCmaxP(C \mid X) \quad (3)$$

Penelitian menggunakan Multinomial Naive Bayes dengan parameter smoothing $\alpha = 0{,}5$.

2.6. Evaluasi Model

Kinerja model dievaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan F1-score.
Accuracy dirumuskan sebagai:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (7)$$

Pada klasifikasi multikelas, *precision*, *recall*, dan F1-score dihitung untuk setiap kategori emosi dengan pendekatan *one-vs-rest*, kemudian nilai keseluruhan dirangkum menggunakan *macro-average*.

Penggunaan beberapa metrik diperlukan agar performa model tidak hanya dievaluasi berdasarkan proporsi prediksi benar secara keseluruhan.

2. HASIL DAN PEMBAHASAN

3.1. Hasil *Preprocessing* dan Augmentasi

Preprocessing menghasilkan bentuk teks yang lebih seragam melalui pembersihan karakter, normalisasi huruf, tokenisasi, dan penghapusan sebagian stopword. Namun, hasil penelitian menunjukkan bahwa *stopword removal* belum sepenuhnya optimal karena keterbatasan daftar stopword yang tersedia untuk Bahasa Melayu Ternate.

Kondisi tersebut menunjukkan salah satu tantangan NLP untuk bahasa dengan sumber daya terbatas. Meskipun beberapa stopword masih tersisa, TF-IDF pada tahap berikutnya memberikan bobot lebih rendah terhadap term yang terlalu sering muncul sehingga dampak sejumlah kata umum dapat dikurangi.

Augmentasi meningkatkan dataset dari 499 menjadi 1.040 data dengan jumlah masing-masing kelas 260 data. Kelas Takut memperoleh jumlah data sintetis paling banyak karena merupakan kelas dengan jumlah data asli paling sedikit.

3.2. Hasil TF-IDF dan *Training* Model

TF-IDF menghasilkan 2.000 fitur untuk 1.040 dokumen dengan matriks yang bersifat sangat sparse. Model Multinomial Naive Bayes kemudian dilatih menggunakan 832 data *training*. Informasi *training* pada penelitian ini memiliki empat kelas target, yaitu Bahagia, Marah, Sedih, dan Takut, dengan 2.000 fitur masukan.

Model mempelajari prior probability setiap kelas dan likelihood fitur pada masing-masing kelas. Setelah proses *training*, model diuji menggunakan 208 data *testing*.

3.3. Hasil Prediksi Data *Testing*

Pada sepuluh contoh hasil prediksi yang telah dilakukan, sembilan data diprediksi dengan benar dan satu data mengalami kesalahan. Salah satu kesalahan terjadi ketika teks berlabel "Takut" diprediksi sebagai "Marah".

Kesalahan tersebut menunjukkan adanya potensi tumpang tindih pola term pada kategori emosi negatif. Suatu teks yang mengekspresikan ketakutan dapat mengandung kata atau konteks yang juga banyak ditemukan pada data kemarahan.

3.4. Hasil Evaluasi Model

Dari 208 data *testing*, sebanyak 184 prediksi benar dan 24 prediksi salah, sehingga Model memperoleh *macro precision* sebesar 89,19%, *macro recall* 88,46%, dan *macro F1-score* 88,59%.

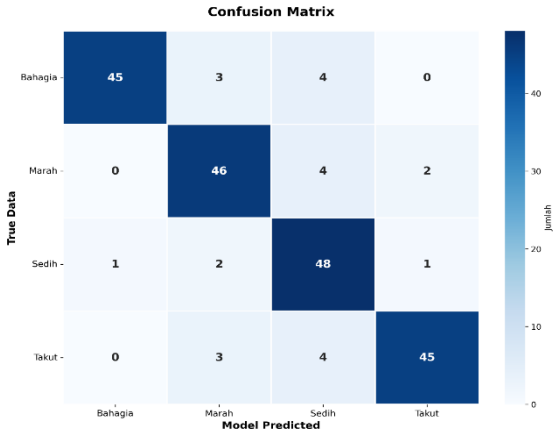
Tabel 6. Hasil Evaluasi Multinomial Naive Bayes

Metrik	Nilai
Accuracy	88,46%
Macro Precision	89,19%
Macro Recall	88,46%
Macro F1-score	88,59%
Prediksi benar	184
Prediksi salah	24
Error rate	11,54%

Accuracy dapat dihitung sebagai:

$$Accuracy = \frac{184}{208} \times 100\% = 88{,}46\% \quad (8)$$

Hasil tersebut menunjukkan bahwa pada skenario pengujian yang digunakan, sebagian besar data *testing* berhasil diklasifikasikan sesuai label.



Gambar 2. Confusion Matrix Multinomial Naive Bayes

3.5. Evaluasi Per Kelas

Evaluasi per kelas menunjukkan karakteristik performa yang berbeda. Kelas Bahagia memiliki *precision* tertinggi sebesar 97,83% dengan *recall* 86,54%. Kelas Sedih memiliki *recall* tertinggi sebesar 92,31%, tetapi *precision* sebesar 80,00%. Kelas Marah memiliki *precision* 85,19% dan *recall* 88,46%, sedangkan kelas Takut memiliki *precision* 93,75% dan *recall* 86,54%.

Precision tinggi pada Bahagia menunjukkan bahwa prediksi model terhadap kelas tersebut relatif tepat. Sebaliknya, *recall* Sedih yang tinggi dengan *precision* lebih rendah menunjukkan bahwa model berhasil mengenali sebagian besar data Sedih tetapi juga memasukkan sejumlah teks dari kelas lain ke dalam kategori Sedih.

Hasil evaluasi mencatat bahwa seluruh kelas memperoleh F1-score di atas 85%, sedangkan kesalahan terutama terjadi pada kategori emosi negatif yang memiliki kemiripan semantik.

3.6. Pengujian pada Data Baru

Selain pengujian terhadap 208 data *testing*, penelitian melakukan evaluasi tambahan menggunakan 20 teks baru yang tidak digunakan pada proses *training* maupun *testing*. Model berhasil memprediksi 14 dari 20 data dengan benar, sehingga diperoleh *accuracy* sebesar 70%.

Hasil ini lebih rendah dibandingkan *accuracy testing* utama sebesar 88,46%. Selisih tersebut menunjukkan bahwa performa model pada teks independen dapat lebih rendah daripada performa pada *testing* set penelitian.

Namun, jumlah data baru hanya 20 sampel sehingga *accuracy* 70% belum cukup untuk menjadi estimasi kemampuan generalisasi yang stabil. Pengujian eksternal dengan jumlah data yang lebih besar diperlukan untuk memperoleh kesimpulan yang lebih kuat.

3.7. Analisis Kesalahan

Sebanyak 24 dari 208 data *testing* mengalami kesalahan klasifikasi. Kesalahan tersebut terkait dengan adanya kemiripan makna antar emosi negatif, konteks kalimat yang ambigu, dan ketergantungan model pada kata kunci tertentu.

Kemiripan antara Marah, Sedih, dan Takut dapat menyebabkan pola fitur yang tumpang tindih. Multinomial Naive Bayes dan TF-IDF bekerja terutama berdasarkan distribusi term dan n-gram sehingga belum secara eksplisit memodelkan konteks semantik yang lebih luas.

Meskipun pendekatan *contextual embedding* dapat menjadi alternatif untuk penelitian lanjutan, penggunaan model yang lebih kompleks tidak otomatis menghasilkan performa lebih baik. Perbandingan harus dilakukan menggunakan dataset dan protokol evaluasi yang sama.

3.8. Perbandingan dengan Penelitian Terdahulu

Hasil *accuracy* sebesar 88,46 % pada penelitian ini tidak digunakan sebagai dasar untuk menyatakan bahwa Multinomial Naive Bayes lebih unggul dibandingkan penelitian sebelumnya karena karakteristik dataset dan protokol evaluasinya berbeda.

Wulandari dkk. melaporkan accuracy 73,20% pada skenario empat kelas menggunakan Naive Bayes [7], sedangkan Pane dkk. memperoleh accuracy 87,14% menggunakan pendekatan ensemble pada teks Bahasa Indonesia [8]. Nilai tersebut digunakan sebagai konteks bahwa performa deteksi emosi dapat berbeda bergantung pada dataset, jumlah kelas, *preprocessing*, fitur, dan skenario evaluasi.

Kontribusi penelitian ini lebih tepat diposisikan pada penerapan klasifikasi empat kategori emosi terhadap teks Bahasa Melayu Ternate serta penyediaan eksperimen awal untuk bahasa lokal dengan keterbatasan sumber daya NLP, bukan pada kebaruan algoritma Multinomial Naive Bayes.

3.9. Keterbatasan Penelitian

Penelitian mempunyai beberapa keterbatasan yang perlu diperhatikan dalam menafsirkan hasil.

Pertama, augmentasi dilakukan sebelum proses pembagian *training* dan *testing*. Kondisi ini berpotensi membuat *training* set dan *testing* set mempunyai kemiripan yang lebih tinggi daripada jika *testing* hanya menggunakan data asli.

Kedua, urutan implementasi penelitian menunjukkan transformasi TF-IDF dilakukan setelah augmentasi terhadap dataset sebelum proses pembagian data. Dalam eksperimen yang lebih ketat, vocabulary dan statistik IDF sebaiknya dibentuk menggunakan *training* set saja.

Ketiga, meskipun pelabelan melibatkan dua anotator dengan latar belakang bidang bahasa, penelitian belum melaporkan nilai *inter-annotator agreement*.

Keempat, belum tersedia daftar stopword khusus Bahasa Melayu Ternate sehingga *preprocessing* masih menggunakan sumber daya

yang belum sepenuhnya sesuai dengan karakteristik bahasa penelitian.

Kelima, pengujian data independen hanya terdiri atas 20 teks. Walaupun menghasilkan *accuracy* 70%, jumlah tersebut masih relatif kecil untuk menilai kemampuan generalisasi secara kuat.

3. KESIMPULAN

Penelitian ini mengimplementasikan Multinomial Naive Bayes untuk klasifikasi empat kategori emosi pada teks Bahasa Melayu Ternate, yaitu Bahagia, Marah, Sedih, dan Takut. Dataset awal terdiri atas 499 teks dengan distribusi 127 data Bahagia, 129 Marah, 130 Sedih, dan 113 Takut. Data diproses melalui tahapan *cleansing*, *case folding*, *tokenizing*, dan *stopword removal*, kemudian ditingkatkan menggunakan enam teknik augmentasi menjadi 1.040 data atau 260 data pada masing-masing kelas.

Representasi teks menggunakan TF-IDF dengan maksimal 2.000 fitur dan n-gram hingga trigram. Model Multinomial Naive Bayes menggunakan parameter alpha 0,5 dan dilatih menggunakan 832 data, sedangkan 208 data digunakan sebagai *testing*.

Pada skenario eksperimen yang diterapkan, model memperoleh *accuracy* sebesar 88,46%, *macro precision* 89,19%, *macro recall* 88,46%, *dan macro F1-score* 88,59%. Sebanyak 184 dari 208 data *testing* berhasil diklasifikasikan dengan benar.

Evaluasi per kelas menunjukkan bahwa Bahagia memiliki *precision* tertinggi sebesar 97,83%, sedangkan Sedih memiliki *recall* tertinggi sebesar 92,31%. Kesalahan klasifikasi terutama ditemukan pada kategori emosi negatif yang memiliki kemiripan makna dan konteks.

Pengujian tambahan menggunakan 20 data baru menghasilkan 14 prediksi benar atau *accuracy* 70%. Hasil tersebut menunjukkan perlunya evaluasi independen yang lebih besar sebelum kemampuan generalisasi model dapat disimpulkan secara kuat.

Penelitian selanjutnya disarankan melakukan pembagian data asli sebelum augmentasi, menerapkan augmentasi hanya pada *training* set, membentuk representasi TF-IDF berdasarkan *training* set, mengembangkan korpus dan daftar stopword khusus Bahasa Melayu Ternate, mengukur kesepakatan anotator, serta membandingkan Multinomial Naive Bayes dengan baseline lain seperti Logistic Regression dan Linear SVM.

DAFTAR PUSTAKA

[1] H. Altaran, M. Sondakh, and S. H. Harilama, "Peranan Komunikasi Orang Tua Dengan Remaja Dalam Melestarikan Bahasa Tidore Di Kelurahan Goto Kota Tidore Kepulauan Maluku Utara," *Acta Diurna Komunikasi*, vol. 5, no. 1, 2023.

[2] N. Alswaidan and M. E. B. Menai, "A Survey of State-of-the-Art Approaches for Emotion Recognition in Text," *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2937–2987, 2020, doi: 10.1007/s10115-020-01449-0.

[3] J. Hofmann, E. Troiano, K. Sassenberg, and R. Klinger, "Appraisal Theories for Emotion Classification in Text," in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 125–138, 2020, doi: 10.18653/v1/2020.coling-main.11.

[4] J. Bata, Suyoto, and Pranowo, "Leksikon Untuk Deteksi Emosi Dari Teks Bahasa Indonesia," in *Seminar Nasional Informatika 2015 (SEMNASIF)*, pp. 195–202, 2015.

[5] I. M. D. Ardiada, M. Sudarma, and D. Giriantari, "Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, pp. 55–60, 2019, doi: 10.24843/MITE.2019.v18i01.P08.

[6] A. Fadjeri, K. Hidayat, and D. R. Handayani, "Deteksi Emosi Pada Teks Menggunakan Algoritma Naïve Bayes," *Jurnal Riset Teknologi Informasi dan Komputer*, vol. 1, no. 2, pp. 1–4, 2021, doi: 10.53863/juristik.v1i02.365.

[7] D. R. Wulandari, C. Setianingsih, and F. M. Dirgantara, "Deteksi Emosi Berbasis Teks Untuk Menganalisis Kuliah Daring Selama Masa Pandemi Menggunakan Algoritme Naive Bayes," *eProceedings of Engineering*, vol. 9, no. 4, 2022, doi: 10.34818/eoe.v9i4.18254.

[8] S. F. Pane, F. Abdullah, and R. Habibi, "Deteksi Emosi Pada Teks Berbahasa Indonesia Menggunakan Pendekatan Ensemble," *Jurnal Teknologi Terapan*, vol. 10, no. 2, 2024, doi: 10.31884/jtt.v10i2.551.

[9] Riccosan, K. E. Saputra, G. D. Pratama, and A. Chowanda, "Emotion Dataset from Indonesian Public Opinion," *Data in Brief*, vol. 43, Art. no. 108465, 2022, doi: 10.1016/j.dib.2022.108465.

[10] K. S. Nugroho, F. A. Bachtiar, W. F. Mahmudy, M. M. Henry, M. Isnan, G. Pangestu, and B. Pardamean, "EmoTweetID: A Dataset of Indonesian Tweets for Emotion Classification and Word Embedding Construction," *Data in Brief*, vol. 68, Art. no. 113119, 2026, doi: 10.1016/j.dib.2026.113119.

[11] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.

[12] J. Wei and K. Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language*

Processing and the 9th International Joint Conference on Natural Language Processing, pp. 6382–6388, 2019, doi: 10.18653/v1/D19-1670.

[13] J. Nothman, H. Qin, and R. Yurchak, "Stop Word Lists in Free Open-source Software Packages," in *Proceedings of the Workshop for NLP Open Source Software*, pp. 7–12, 2018, doi: 10.18653/v1/W18-2502.