

Analisis Sentimen Ulasan Produk Pada Marketplace Shopee Menggunakan Metode Naive Bayes

Abelta Cesharilo^{*1}, Untung Subagyo²

¹ Informatika, STMIK EL Rahma Yogyakarta

² Sistem Informasi, STMIK EL Rahma Yogyakarta

email: ¹abelcsharilo@gmail.com, ²untungsubagyo@stmikelrahma.ac.id

^{*}Correspondence author

Abstrak

Shopee merupakan salah satu marketplace yang banyak digunakan masyarakat untuk melakukan transaksi secara online. Fitur ulasan produk di Shopee memungkinkan para pengguna menyampaikan pendapat tentang barang yang sudah dibeli. Banyak ulasan produk di Shopee membuat analisis secara manual sulit dilakukan, sedangkan penelitian sebelumnya sering mengabaikan normalisasi bahasa yang informal, sehingga menjadi celah dalam penelitian ini. Penelitian ini bertujuan melakukan analisis sentimen ulasan produk Shopee menggunakan metode Multinomial Naïve Bayes (MNB) berbasis pembobotan TF-IDF dan normalisasi teks kustom, yang dievaluasi secara komparatif terhadap algoritma Support Vector Machine (SVM) serta diuji keandalannya menggunakan 5-Fold Cross Validation. Dataset yang digunakan terdiri dari 500 ulasan seimbang (250 positif dan 250 negatif) dari Kaggle. Tahap preprocessing meliputi case folding, cleansing, normalisasi, tokenizing, stopword removal, dan stemming. Hasil evaluasi Hold-Out (80:20) menunjukkan SVM memperoleh akurasi sebesar 96,00%, presisi 92,59%, recall 100,00%, dan F1-score 96,15%, sedangkan MNB meraih akurasi 91,00%, presisi 84,75%, recall 100,00%, dan F1-score 91,74%. Pengujian 5-Fold Cross Validation mengonfirmasi kestabilan kedua model dengan rata-rata akurasi MNB sebesar 90,80% ($\pm 3,25\%$) dan SVM sebesar 89,60% ($\pm 2,50\%$). Sebagai kesimpulan, penggabungan normalisasi kustom dan TF-IDF terbukti menghasilkan model klasifikasi yang stabil dan akurat untuk ulasan e-commerce.

Kata kunci—Sentiment Analysis, Shopee Marketplace, Text Mining, Naïve Bayes, TF-IDF.

1. PENDAHULUAN

Perkembangan teknologi informasi dan internet memiliki pengaruh yang sangat besar terhadap kegiatan perdagangan secara digital [1]. Salah satu perkembangan tersebut adalah *marketplace* yang memudahkan masyarakat dalam melakukan transaksi jual beli secara online [2]. Shopee merupakan salah satu tempat belanja online yang populer di Indonesia karena menawarkan banyak fitur yang membuat membeli barang menjadi mudah dan cepat [3]. Selain itu, Shopee juga menyediakan fitur untuk memberikan ulasan produk yang memungkinkan pengguna memberikan rating, foto, video, dan menceritakan pengalaman mereka secara langsung setelah menggunakan suatu barang [4].

Ulasan produk pada *marketplace* memiliki peran penting karena dapat menjadi sumber informasi bagi calon konsumen sebelum melakukan pembelian [5]. Ulasan ini bisa berupa pendapat baik atau buruk tentang kualitas barang, kecepatan pengiriman, respon penjual, atau kesesuaian deskripsi produk [6]. Banyaknya ulasan yang ada membuat analisis secara manual menjadi sulit dan membutuhkan waktu lama [7]. Oleh karena itu, diperlukan suatu metode yang mampu melakukan klasifikasi sentimen secara otomatis berbasis *text mining* secara cepat terhadap ulasan pengguna [8].

Analisis sentimen merupakan salah satu cabang dari text mining yang digunakan untuk mengidentifikasi opini, emosi, atau penilaian seseorang berdasarkan data teks terstruktur maupun tidak terstruktur [9]. Dalam proses analisis sentimen, diperlukan metode klasifikasi yang mampu mengelompokkan data teks secara efektif. Algoritma Naïve Bayes sering menjadi pilihan favorit dalam klasifikasi teks karena memiliki beberapa keunggulan, seperti proses perhitungan probabilitas yang mudah, persyaratan data latih yang tidak terlalu banyak, serta kecepatan dalam pemrosesannya yang singkat [10],[11]. Kendati demikian, performa Naïve

Bayes sangat bergantung pada kualitas representasi teks masukan [6]. Karakteristik ulasan pada marketplace Indonesia umumnya sarat akan penggunaan bahasa informal, singkatan, kata slang, *typo*, dan karakter acak yang menjadi *noise* bagi model komputasi [4].

Research gap pada penelitian-penelitian sentimen e-commerce terdahulu terletak pada sering diabaikannya tahap normalisasi bahasa informal kustom yang spesifik serta belum diujinya stabilitas model probabilistik saat berhadapan dengan fitur teks yang terdistribusi tidak merata [4], [6]. Selain itu, sebagian besar penelitian terdahulu mengenai ulasan Shopee atau Tokopedia hanya melakukan pengujian dengan satu skema pemisahan data (*hold-out*) tanpa menyertakan analisis pengujian kestabilan model melalui *cross-validation* serta tanpa adanya perbandingan langsung terhadap algoritma klasifikasi *supervised learning* lainnya pada dataset yang sama [12]. Hal ini menyebabkan keunggulan dan tingkat efektivitas model probabilistik belum dapat dievaluasi secara objektif.

Untuk mengatasi kekurangan dalam penelitian sebelumnya, inovasi dan kontribusi utama penelitian ini berupa penggabungan preprocessing terstruktur meliputi pembentukan kamus normalisasi kata slang kustom dan penguncian kata negasi pada stopword removal dengan pembobotan TF-IDF dan algoritma Multinomial Naïve Bayes [13], [14]. Selanjutnya, untuk memberikan penilaian yang objektif dan tidak memihak, penelitian ini melibatkan algoritma lain yaitu *Support Vector Machine* (SVM) sebagai pembanding serta menguji konsistensi hasil model dengan metode *K-Fold Cross Validation* [15]. Hasil penelitian ini diharapkan dapat memberikan baseline model klasifikasi sentimen otomatis yang andal, teruji secara komparatif, dan transparan bagi para pelaku bisnis e-commerce [5].

2. METODE PENELITIAN

Analisis Permasalahan

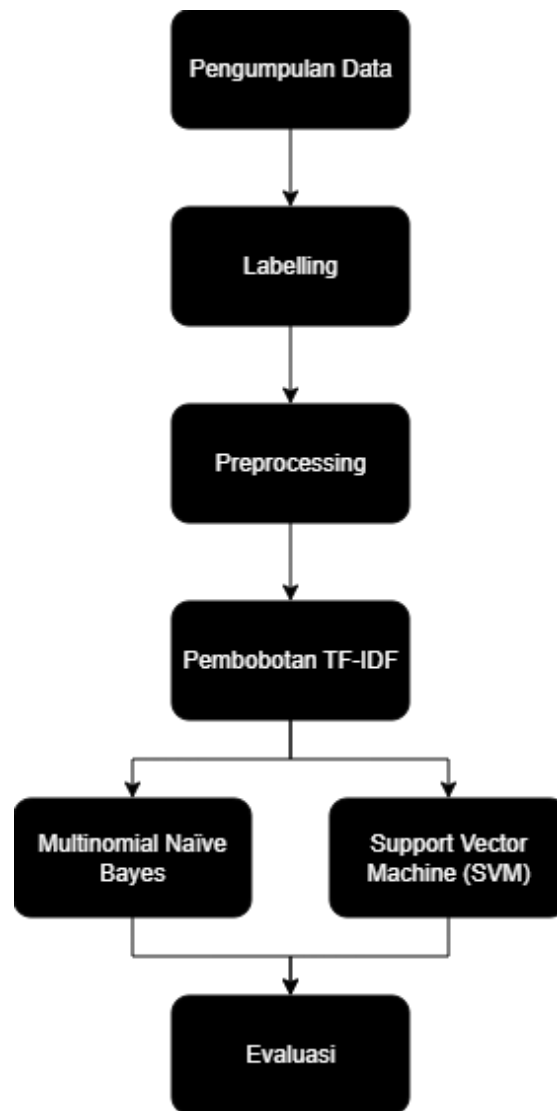
Marketplace Shopee menyediakan fitur untuk menulis ulasan produk yang bisa dipakai oleh pengguna untuk memberikan pendapat tentang barang yang telah mereka beli. Ulasan ini bisa berisi komentar positif atau negatif yang menunjukkan seberapa puas pengguna terhadap produk dan layanan dari penjual [4]. Namun, analisis manual terhadap volume ulasan yang melimpah menjadi tidak efisien [5]. Selain itu, karakteristik teks ulasan yang sarat akan kata singkatan, *typo*, dan bahasa informal khas media sosial menjadi *noise* yang dapat menurunkan akurasi model komputasi [6]. Masalah ini diselesaikan dengan membangun sistem analisis sentimen otomatis berbasis text mining menggunakan algoritma Multinomial Naïve Bayes dan pembobotan fitur TF-IDF [4]. Untuk menguji efektivitas dan keunggulan model secara objektif, penelitian ini menambahkan algoritma *Support Vector Machine* (SVM) sebagai pembanding serta menerapkan teknik *K-Fold Cross Validation* untuk menguji stabilitas performa klasifikasi [5].

Lingkungan Eksperimen dan Library Python

Pengembangan sistem dan pengujian komputasi dilakukan pada lingkungan Google Colab menggunakan bahasa pemrograman Python 3 [4]. Proses pengolahan dan manipulasi data dilakukan menggunakan *library* Pandas dan NumPy, terutama untuk pengelolaan struktur data berbentuk *dataframe* dan pemrosesan matriks numerik. Pada tahap *text preprocessing*, NLTK (*Natural Language Toolkit*) dan Sastrawi digunakan untuk melakukan *tokenizing*, penyaringan *stopword*, dan *stemming* teks berbahasa Indonesia. Selanjutnya, Scikit-Learn digunakan untuk ekstraksi fitur menggunakan *TfidfVectorizer*, implementasi model klasifikasi Multinomial Naïve Bayes dan *Support Vector Machine* (SVM), pembagian data menggunakan *train_test_split*, validasi model menggunakan *cross_val_score*, serta evaluasi hasil klasifikasi melalui *confusion matrix*. Hasil pengujian kemudian divisualisasikan menggunakan Matplotlib dan Seaborn, terutama dalam bentuk *heatmap confusion matrix* untuk mempermudah interpretasi kinerja model klasifikasi. Penggunaan berbagai *library* tersebut mendukung seluruh tahapan eksperimen secara terintegrasi, mulai dari praproses data, ekstraksi fitur, pelatihan dan validasi model, hingga evaluasi serta visualisasi hasil klasifikasi.

Kerangka Penelitian

Tahapan penelitian dilaksanakan secara sistematis meliputi pengumpulan data, pelabelan, *preprocessing* teks kustom, pembobotan TF-IDF, pemodelan klasifikasi (Multinomial Naïve Bayes dan SVM), serta pengujian stabilitas model dengan *K-Fold Cross Validation* dan *Confusion Matrix*. Tahapan Penelitian yang akan dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

Pengumpulan Data dan Pelabelan Data

Dataset ulasan produk Shopee berbahasa Indonesia diperoleh dari repositori publik Kaggle format CSV. Pelabelan sentimen dilakukan secara otomatis berdasarkan nilai rating bintang. Ulasan dengan rating 4 dan 5 dikategorikan sebagai sentimen positif, sedangkan rating 1 dan 2 dikategorikan sebagai sentimen negatif. Ulasan ber-rating 3 dieliminasi untuk menghindari ambiguitas sentimen netral. Berdasarkan proses tersebut, diperoleh sampel dataset seimbang (*balanced dataset*) sebanyak 500 ulasan (250 ulasan positif dan 250 ulasan negatif) [4].

Preprocessing

Tahap *preprocessing* dilakukan untuk membersihkan dan menyeragamkan data teks sebelum digunakan dalam proses ekstraksi fitur dan klasifikasi. Tahapan ini bertujuan

mengurangi *noise* pada teks serta menghasilkan representasi data yang lebih konsisten sehingga dapat diproses oleh model klasifikasi. *Preprocessing* dalam penelitian ini meliputi *case folding*, *cleansing*, normalisasi kata, *tokenizing*, *stopword removal*, dan *stemming*.

Tahap pertama adalah *case folding*, yaitu mengubah seluruh karakter alfabet menjadi huruf kecil untuk menyeragamkan penulisan kata. Sebagai contoh, kata “Amanah” diubah menjadi “amanah” sehingga kata yang sama tidak dianggap sebagai fitur yang berbeda hanya karena perbedaan penggunaan huruf kapital. Contoh hasil proses *case folding* ditunjukkan pada Tabel 1. Selanjutnya dilakukan *cleansing* untuk menghapus tanda baca, angka, simbol, dan karakter lain yang tidak diperlukan dalam proses klasifikasi. Tahap ini menghasilkan teks yang lebih bersih sebagaimana ditunjukkan pada Tabel 2.

Tabel 1. Case Folding

Sebelum	Sesudah
Barang sudah sampai,kemasan ga rapih, isinya hancur.trus yg rasa strawberry masa kemasan udh sobek udh dibuka.ga Amanah.	barang sudah sampai,kemasan ga rapih, isinya hancur.trus yg rasa strawberry masa kemasan udh sobek udh dibuka.ga amanah.

Tabel 2. Cleansing

Sebelum	Sesudah
barang sudah sampai,kemasan ga rapih, isinya hancur.trus yg rasa strawberry masa kemasan udh sobek udh dibuka.ga amanah.	barang sudah sampai kemasan ga rapih isinya hancur trus yg rasa strawberry masa kemasan udh sobek udh dibuka ga amanah

Setelah proses *cleansing*, dilakukan normalisasi kata untuk mengubah kata tidak baku, singkatan, maupun variasi penulisan menjadi bentuk yang lebih baku dan konsisten. Sebagai contoh, kata “ga” dinormalisasi menjadi “tidak”, “yg” menjadi “yang”, “trus” menjadi “terus”, dan “udh” menjadi “sudah”. Contoh hasil normalisasi ditunjukkan pada Tabel 3. Teks yang telah dinormalisasi kemudian melalui proses *tokenizing*, yaitu memecah kalimat menjadi unit-unit kata atau *token*. Hasil proses *tokenizing* ditunjukkan pada Tabel 4.

Tabel 3. Normalisasi Kata

Sebelum	Sesudah
barang sudah sampai kemasan ga rapih isinya hancur trus yg rasa strawberry masa kemasan udh sobek udh dibuka ga amanah	barang yang sudah sampai kemasan tidak rapih isinya hancur terus yang rasa strawberry masa kemasan sudah sobek sudah dibuka tidak amanah

Tabel 4. Tokenizing

Sebelum	Sesudah
barang yang sudah sampai kemasan tidak rapih isinya hancur terus yang rasa strawberry masa kemasan sudah sobek sudah dibuka tidak amanah	'barang', 'yang', 'sudah', 'sampai', 'kemasan', 'tidak', 'rapih', 'isinya', 'hancur', 'terus', 'yang', 'rasa', 'strawberry', 'masa', 'kemasan', 'sudah', 'sobek', 'sudah', 'dibuka', 'tidak', 'amanah'

Tahap berikutnya adalah *stopword removal* untuk menghilangkan kata-kata umum yang dianggap tidak memberikan informasi penting dalam proses klasifikasi. Meskipun demikian, kata negasi seperti “tidak” dipertahankan karena memiliki pengaruh terhadap makna dan polaritas sentimen suatu ulasan. Hasil proses *stopword removal* ditunjukkan pada Tabel 5. Tahap terakhir adalah *stemming*, yaitu mengubah kata berimbuhan menjadi bentuk kata dasarnya. Sebagai contoh, kata “kemasan” direduksi menjadi “kemas”, “isinya” menjadi “isi”, dan “dibuka” menjadi

“buka”. Hasil proses *stemming* ditunjukkan pada Tabel 6. Setelah seluruh tahapan tersebut selesai, diperoleh *clean text* yang digunakan sebagai masukan pada tahap ekstraksi fitur.

Tabel 5. Stopword Removal

Sebelum	Sesudah
'barang', 'yang', 'sudah', 'sampai', 'kemasan', 'tidak', 'rapih', 'isinya', 'hancur', 'terus', 'yang', 'hancur', 'rasa', 'strawberry', 'kemasan', 'rasa', 'strawberry', 'masa', 'kemasan', 'sudah', 'sobek', 'dibuka', 'tidak', 'amanah'	'barang', 'kemasan', 'tidak', 'rapih', 'isinya', 'hancur', 'rasa', 'strawberry', 'kemasan', 'sobek', 'dibuka', 'tidak', 'amanah'

Tabel 6. Stemming

Sebelum	Sesudah
'barang', 'kemasan', 'tidak', 'rapih', 'isinya', 'hancur', 'rasa', 'strawberry', 'kemasan', 'sobek', 'dibuka', 'tidak', 'amanah'	'barang', 'kemas', 'tidak', 'rapih', 'isi', 'hancur', 'rasa', 'strawberry', 'kemas', 'sobek', 'buka', 'tidak', 'amanah'

Pembobotan TF-IDF

Data teks yang telah melalui tahap *preprocessing* selanjutnya ditransformasikan menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) [4]. Transformasi ini diperlukan karena algoritma klasifikasi tidak dapat memproses data teks secara langsung. TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat kelangkaannya pada keseluruhan korpus. Dengan demikian, kata yang sering muncul pada suatu dokumen tetapi relatif jarang ditemukan pada dokumen lain akan memperoleh bobot yang lebih tinggi dan dianggap lebih informatif dalam membedakan karakteristik antarulasan [4]. Matriks TF-IDF yang dihasilkan selanjutnya digunakan sebagai fitur masukan pada proses pemodelan klasifikasi.

Pemodelan Klasifikasi dan Parameter Model

Pemodelan klasifikasi dilakukan menggunakan dua algoritma supervised learning, yaitu Multinomial Naïve Bayes (MNB) dan Support Vector Machine (SVM). MNB merupakan algoritma klasifikasi probabilistik berbasis Teorema Bayes yang banyak digunakan pada klasifikasi teks dan sesuai untuk data yang direpresentasikan dalam bentuk fitur berbobot seperti TF-IDF. Pada pengujian awal, model MNB menggunakan parameter smoothing $\alpha = 1,0$. Sementara itu, SVM digunakan untuk membentuk batas keputusan (hyperplane) yang memisahkan data antar kelas secara optimal. Model SVM menggunakan kernel linear dengan parameter regularisasi $C = 1,0$, yang sesuai untuk representasi teks berdimensi tinggi.

Secara umum, dasar probabilistik yang digunakan pada Naïve Bayes mengikuti Teorema Bayes sebagaimana ditunjukkan pada Persamaan (1). Pada Persamaan (1), C merupakan kelas sentimen ulasan dan X merupakan data atau fitur dari ulasan. $P(C|X)$ menunjukkan probabilitas suatu ulasan termasuk dalam kelas C berdasarkan fitur X , sedangkan $P(X|C)$ menunjukkan probabilitas kemunculan fitur X pada kelas C . Selanjutnya, $P(C)$ merupakan probabilitas awal (prior probability) dari kelas C dan $P(X)$ merupakan probabilitas kemunculan data atau evidence. Berdasarkan nilai probabilitas tersebut, Naïve Bayes menentukan kelas yang memiliki probabilitas posterior terbesar sebagai hasil prediksi. Pendekatan ini memungkinkan model mengklasifikasikan setiap ulasan berdasarkan pola kemunculan fitur kata yang merepresentasikan karakteristik masing-masing kelas sentimen.

$$P(C|X) = \frac{(P(X|C) \times P(C))}{P(X)} \quad (1)$$

Evaluasi Model dan K-Fold Cross Validation

Evaluasi kinerja model dilakukan menggunakan dua skema, yaitu *hold-out validation* dan *K-Fold Cross Validation*. Pada skema *hold-out validation*, dataset yang terdiri atas 500 ulasan dibagi menjadi 80% data latih dan 20% data uji. Dengan pembagian tersebut, sebanyak 400 ulasan digunakan untuk melatih model dan 100 ulasan digunakan untuk menguji kinerja model terhadap data yang tidak digunakan selama proses pelatihan. Hasil prediksi kemudian dianalisis menggunakan *confusion matrix* untuk memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Penggunaan beberapa metrik tersebut memberikan gambaran kinerja model yang lebih komprehensif dibandingkan hanya menggunakan nilai akurasi.

Selain *hold-out validation*, penelitian ini menerapkan *5-fold cross validation* untuk mengevaluasi kestabilan kinerja model. Dataset dibagi menjadi lima bagian (*fold*), kemudian proses pelatihan dan pengujian dilakukan sebanyak lima iterasi. Pada setiap iterasi, empat *fold* digunakan sebagai data latih dan satu *fold* lainnya sebagai data validasi, dengan bagian validasi bergantian pada setiap iterasi. Nilai akurasi dari kelima iterasi selanjutnya dihitung nilai rata-rata (*mean*) dan standar deviasinya. Nilai rata-rata digunakan untuk menggambarkan performa model secara keseluruhan, sedangkan standar deviasi digunakan untuk menunjukkan tingkat variasi kinerja antar-*fold*. Dengan pendekatan ini, evaluasi tidak hanya bergantung pada satu pembagian data, tetapi juga memberikan gambaran mengenai konsistensi dan kemampuan generalisasi model.

3. HASIL DAN PEMBAHASAN

Hasil Pengujian Evaluasi Hold-Out

Pengujian eksperimen untuk klasifikasi sentimen dari ulasan produk pada *marketplace* Shopee dengan menggunakan algoritma Multinomial Naïve Bayes dilakukan melalui simulasi di Google Colab. Berdasarkan parameter eksperimen yang telah dirancang, total dataset bersih terdiri dari 500 ulasan yang dibagi menjadi 80% untuk data latih (400 ulasan) yang digunakan untuk melatih model dan 20% untuk data uji (100 ulasan), kinerja model probabilistik Multinomial Naïve Bayes (MNB) dievaluasi secara komparatif terhadap algoritma *Support Vector Machine* (SVM) berbasis *kernel linear* [5]. Berdasarkan hasil pengujian yang telah dilakukan, diperoleh hasil evaluasi model seperti yang ditunjukkan pada Tabel 7.

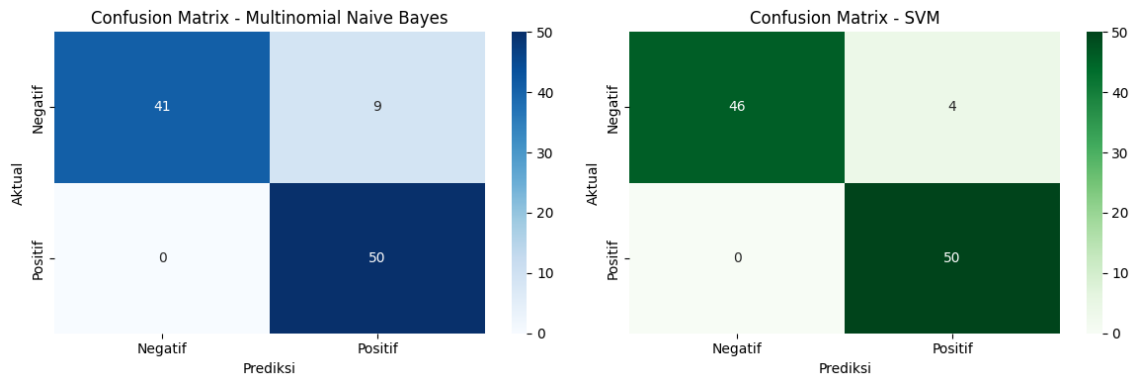
Tabel 7. Perbandingan Evaluasi Model

Parameter	MNB	SVM
Accuracy	91,00%	96,00 %
Precision	84,75%	92,59 %
Recall	100,00%	100,00 %
F1-Score	91,74%	96,15 %

Berdasarkan data pada Tabel 7 diatas, kedua model klasifikasi menunjukkan performa yang sangat baik dalam mengklasifikasikan sentimen ulasan produk Shopee [4]. Pada skema *Hold-Out*, algoritma SVM mencatatkan tingkat akurasi sebesar 96,00%, lebih tinggi dibandingkan Multinomial Naïve Bayes yang memperoleh akurasi 91,00%. Nilai *Precision* SVM juga mencapai 92,59% (dibandingkan MNB sebesar 84,75%), yang mengindikasikan bahwa SVM lebih presisi dalam meminimalisir kesalahan prediksi pada ulasan bermakna negatif [5].

Confusion Matrix dan Visualisasi Heatmap

Rincian sebaran prediksi dari 100 sampel data uji dianalisis menggunakan matriks *Confusion Matrix* yang dipetakan dalam bentuk *Heatmap* berdampingan yang bisa ditunjukkan pada Gambar 2.



Gambar 2. Visualisasi Heatmap Confusion Matrix Model MNB dan SVM

Berdasarkan hasil *confusion matrix* terhadap 100 data uji yang terdiri atas 50 ulasan positif dan 50 ulasan negatif, model Multinomial Naïve Bayes (MNB) berhasil mengidentifikasi seluruh 50 ulasan positif secara tepat (*True Positive* = 50) dan 41 ulasan negatif secara tepat (*True Negative* = 41). Model MNB tidak menghasilkan kesalahan *False Negative* ($FN = 0$), tetapi masih terdapat 9 ulasan negatif yang salah diklasifikasikan sebagai positif (*False Positive* = 9). Sementara itu, Support Vector Machine (SVM) menunjukkan hasil yang lebih baik dalam membedakan ulasan negatif dengan mengidentifikasi 46 dari 50 ulasan negatif secara tepat (*True Negative* = 46). Seluruh 50 ulasan positif juga berhasil diklasifikasikan dengan benar (*True Positive* = 50), sehingga tidak terdapat kesalahan *False Negative* ($FN = 0$), sedangkan jumlah *False Positive* berkurang menjadi 4 data.

Tidak ditemukannya *False Negative* pada kedua model menghasilkan nilai *recall* kelas positif sebesar 100,00%, yang menunjukkan bahwa seluruh ulasan positif pada data uji berhasil teridentifikasi dengan benar. Kondisi tersebut diduga berkaitan dengan karakteristik ulasan positif pada dataset yang banyak menggunakan kosakata pujian yang relatif seragam dan berulang, seperti “bagus”, “enak”, “cepat”, dan “murah” [4]. Representasi TF-IDF memberikan bobot pada kata-kata yang memiliki daya pembeda dalam korpus sehingga pola fitur pada ulasan positif dapat dikenali oleh model. Meskipun kedua model memiliki *recall* yang sama, jumlah *False Positive* yang lebih rendah pada SVM menunjukkan adanya perbedaan kemampuan kedua model dalam membedakan ulasan negatif dari ulasan positif pada data uji.

Pengujian Stabilitas Model dengan K-Fold Cross Validation

Untuk menghindari bias akibat pemisahan data secara acak pada skema *Hold-Out* serta menguji kestabilan performa model secara transparan, pengujian dilanjutkan menggunakan skema *5-Fold Cross Validation* [5]. Dataset 500 ulasan dibagi secara proporsional menjadi 5 *fold*, di mana secara bergantian setiap *fold* diuji terhadap model yang dilatih pada 4 *fold* sisanya [5]. Hasil pengujian *5-Fold Cross Validation* disajikan pada Tabel 8.

Tabel 8. Hasil Pengujian Stabilitas 5-Fold Cross Validation

K-Fold	MNB	SVM
Fold-1	87,00 %	87,00 %
Fold-2	96,00 %	92,00 %
Fold-3	89,00 %	93,00 %
Fold-4	89,00 %	89,00 %
Fold-5	93,00 %	87,00 %
Rata-rata akurasi	90,80 % ($\pm 3,25$ %)	89,60 % ($\pm 2,50$ %)

Hasil *5-Fold Cross Validation* pada Tabel 8 membuktikan bahwa kedua model memiliki konsistensi dan stabilitas yang sangat tinggi, dengan rata-rata akurasi MNB sebesar 90,80% ($\pm 3,25\%$) dan SVM sebesar 89,60% ($\pm 2,50\%$). Tingginya rata-rata akurasi *cross-validation* ini

mengonfirmasi bahwa efektivitas klasifikasi terjadi bukan karena faktor kebetulan pemisahan data (*overfitting*), melainkan karena integrasi *preprocessing* kustom dan TF-IDF yang optimal [5].

Pembahasan dan Analisis Kesalahan

Meskipun kedua model tersebut memberikan hasil yang baik, masih ada kesalahan False Positive, yaitu 9 data pada model MNB dan 4 data pada model SVM [4], [5]. Berdasarkan analisis kesalahan terhadap data tekstual, kesalahan ini dipicu oleh karakteristik ulasan negatif konsumen yang menggunakan bahasa sarkasme atau kalimat deskripsi panjang [6]. Sebagai contoh, ulasan negatif: "*Kualitas produk sih lumayan bagus, tapi sayang pengirimannya lambat banget, respon penjual juga cuek, jadi kecewa.*" Setelah melewati tahap *preprocessing*, kata hubung seperti 'sih', 'tapi', dan 'jadi' terhapus oleh *custom stopwords*. Token yang tersisa dan dihitung oleh TF-IDF didominasi oleh kata bermakna positif seperti 'kualitas', 'produk', 'lumayan', dan 'bagus'. Pada MNB yang bekerja berdasarkan asumsi independensi bersyarat tanpa memperhatikan urutan kata, tingginya frekuensi token positif tersebut mengelabui perhitungan probabilitas posterior sehingga ulasan keliru diklasifikasikan sebagai sentimen positif [4]. Sebaliknya, SVM mampu menekan False Positive lebih baik karena bekerja dengan mencari *hyperplane* pemisah berjarak maksimum (*maximum margin*) antar batas vektor.

Dalam konteks perbandingan dengan literatur terdahulu, pencapaian akurasi penelitian ini memberikan gambaran positif mengenai efektivitas kamus kustom dan penguncian kata negasi [11]. Sebagai contoh, penelitian analisis sentimen di Tokopedia oleh Rahmawati dan Santoso [12] berhasil mencapai akurasi sebesar 83,90%, sedangkan penelitian tentang ulasan makanan oleh Hajaroh dan tim mencapai akurasi 75,90%. Meski demikian, perbandingan tersebut ditangani secara bijak dan tidak dianggap sebagai bukti superioritas mutlak, karena setiap penelitian menggunakan dataset dan skala domain yang berbeda.

Keterbatasan Penelitian

Penelitian ini memiliki keterbatasan pada kuantitas dataset yang terbatas pada sampel seimbang 500 ulasan. Ukuran dataset ini menyebabkan kamus normalisasi slang belum mampu mencakup seluruh variasi istilah tidak baku baru yang berkembang cepat di media sosial. Selain itu, model yang diuji belum dilengkapi pemrosesan *ngram* konteks untuk menangani kalimat sarkasme yang kompleks [5].

4. KESIMPULAN

Berdasarkan hasil pengujian eksperimen yang telah dilakukan, dapat disimpulkan bahwa penerapan metode Multinomial Naïve Bayes (MNB) dan Support Vector Machine (SVM) yang dikombinasikan dengan pembobotan fitur TF-IDF serta *text preprocessing* kustom terbukti efektif dalam mengklasifikasikan sentimen ulasan produk pada marketplace Shopee Indonesia. Pada pengujian *Hold-Out Validation* (80:20) terhadap 100 data uji, algoritma SVM mencatatkan kinerja terbaik dengan akurasi sebesar 96,00%, presisi 92,59%, recall 100,00%, dan F1-score 96,15%, sedangkan Multinomial Naïve Bayes memperoleh akurasi sebesar 91,00%, presisi 84,75%, recall 100,00%, dan F1-score 91,74% [4], [5]. Pengujian stabilitas menggunakan *5-Fold Cross Validation* membuktikan bahwa kedua model memiliki konsistensi performa yang sangat tinggi dan teruji bebas dari *overfitting*, dengan rata-rata akurasi MNB sebesar 90,80% ($\pm 3,25\%$) dan SVM sebesar 89,60% ($\pm 2,50\%$) [5]. Penelitian ini memberikan bukti ilmiah bahwa dengan mengintegrasikan kamus kustom dalam proses normalisasi kata dan mengunci kata negasi saat menghapus stopwords, konteks sentimen negatif dari ulasan pendek dapat dipertahankan dengan baik serta meningkatkan performa klasifikasi secara signifikan. Kendati demikian, penelitian ini memiliki keterbatasan pada ukuran sampel dataset yang terbatas pada 500 ulasan seimbang, sehingga model masih menghadapi kesalahan *False Positive* pada ulasan negatif yang menggunakan bahasa sarkasme atau kalimat deskripsi panjang akibat mengabaikan urutan konteks antar kata [4], [5]. Sebagai saran untuk pengembangan ke depan, disarankan untuk memperbesar volume dataset secara signifikan, menerapkan pemodelan fitur *N-Gram* (seperti

Bigram/Trigram), serta menguji metode seleksi fitur tambahan seperti Chi-Square untuk mereduksi token ambigu untuk meningkatkan generalisasi model klasifikasi secara keseluruhan [5].

DAFTAR PUSTAKA

- [1] T. Ernayanti, Mustafid, A. Rusgiyono, and A. R. Hakim, "Penggunaan Seleksi Fitur Chi-Square dan Algoritma Multinomial Naïve Bayes Untuk Analisis Sentimen Pelanggan Tokopedia," *Jurnal Gaussian*, vol. 11, no. 1, pp. 562–571, 2023, doi: 10.14710/j.gauss.11.4.562-571.
- [2] R. L. Atimi and E. E. Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *Jurnal Sains dan Informatika*, vol. 8, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.
- [3] T. P. R. Sanjaya, A. Fauzi, and A. F. N. Masruriyah, "Analisis Sentimen Ulasan Pada E-commerce Shopee Menggunakan Algoritma Naive Bayes dan Support Vector Machine," *Jurnal Informatika Teknologi*, vol. 4, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.
- [4] V. G. Shintarani, R. Mayasari, and M. Jajuli, "Analisis Sentimen Ulasan Konsumen Pada Produk Ponsel Pintar Menggunakan Metode Naïve Bayes," *Jurnal Cahaya Mandalika*, vol. 4, no. 1, pp. 771–781, 2023, doi: 10.36312/jcm.v4i1.2101.
- [5] A. Nurian, M. S. Ma'arif, I. N. Amalia, and C. Rozikin, "Analisis Sentimen Pengguna Aplikasi Shopee Pada Situs Google Play Menggunakan Naïve Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, pp. 97–105, 2024, doi: 10.23960/jitet.v12i1.3631.
- [6] R. R. Rismansyah, A. Sudiarjo, and T. Mufizar, "Analisis Sentimen Ulasan Shopee Menggunakan Algoritma Naïve Bayes," *Jurnal Elektro dan Informatika Swadharma (JEIS)*, vol. 5, no. 1, pp. 32–35, 2025, doi: 10.56486/jeis.vol5no1.661.
- [7] R. L. Simanjuntak, T. R. Siagian, V. Anggriani, and Arnita, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 3, no. 3, pp. 23–39, 2023, doi: 10.55606/teknik.v3i3.2411.
- [8] A. D. M. Putri, N. Sulistianingsih, and R. Rismayati, "Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF Terhadap Akurasi Klasifikasi Sentimen Teks Multi-Domain," *Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 4, pp. 675–688, 2025, doi: 10.35746/jtim.v7i4.756.
- [9] M. Yunus, M. Husni, and M. M. Mufadhdhal, "Klasifikasi sentimen terhadap Badan Penyelenggara Jaminan Sosial (BPJS) pada media sosial Twitter menggunakan Naive Bayes," *STIKI Informatika Jurnal*, vol. 11, no. 2, pp. 81–91, 2021, doi: 10.32664/smatika.v11i02.577.
- [10] K. A. Rokhman, Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *Journal of Information System Management*, vol. 2, no. 2, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
- [11] Hajaroh, T. Suprpti, and R. Narasati, "Implementasi Algoritma Naive Bayes Untuk Analisis Sentimen Ulasan Produk Makanan dan Minuman di Tokopedia," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, pp. 111–118, 2024, doi: 10.36040/jati.v8i1.8237.
- [12] L. Rahmawati and D. B. Santoso, "Implementasi Metode Naive Bayes Untuk Klasifikasi Ulasan Aplikasi E-Commerce Tokopedia," *Journal of Information Technology and Computer Science*, vol. 6, no. 1, pp. 116–124, 2023, doi: 10.52985/insyst.vlil.36.

- [13] A. Muzaki, V. Febriana, and W. N. Cholifah, "Analisis Sentimen Pada Ulasan Produk di E-Commerce Dengan Metode Naive Bayes," *Jurnal Riset dan Aplikasi Mahasiswa Informatika*, vol. 5, no. 4, pp. 126–129, 2024, doi: 10.24176/simet.v10i2.3.
- [14] A. C. Wibisono, T. S. Nadira, and T. Sutabri, "Analisis Sentimen Pelanggan pada Platform Shopee Menggunakan Metode Naive Bayes," *Nusantara Journal of Multidisciplinary Science*, vol. 2, no. 6, pp. 10101018, 2025, doi: <https://jurnal.intekom.id/index.php/njms/article/view/1010>.
- [15] D. W. Putri and M. A. Soeleman, "Penerapan Algoritma Naïve Bayes Terhadap Sentimen Ulasan Produk Skincare Pada E-Commerce Shopee," *Building of Informatics, Technology and Science*, vol. 7, no. 4, pp. 2218–2228, 2026, doi: 10.47065/bits.v7i4.9209.