

Pengaruh Jenis *Stopwords* terhadap Akurasi Model Multinomial Naïve Bayes dalam Proses Sentimen Analisis

Jimmy Tjen

Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Widya Dharma Pontianak
Jalan H.O.S Cokroaminoto No. 445, Pontianak, Kalimantan Barat, Indonesia
Email: jimmy.tjen@mathmods.eu

Abstract. *The Impact of Stopwords on the Accuracy of the Multinomial Naïve Bayes for Sentiment Analysis Process.* Implementing machine learning in business has enabled producers and sellers to assess product quality by analyzing customer reviews through Sentiment Analysis (SA). This study investigates the impact of different stopwords categories on the accuracy of the Multinomial Naïve Bayes (MNB) model for SA. This research considered ten stopwords categories: general, conjunctions, slang, temporal terms, nouns, pronouns, interjections, adverbs, and single-letter words. A Friedman test conducted on commentary from three shoe products revealed that removing conjunction stopwords (MNB-conjunction) could potentially improve the predictive accuracy of the MNB model for SA by approximately 1%. A T-test further validated this result, showing that two out of three datasets provided evidence that MNB-conjunction outperformed the MNB model without removing stopwords.

Keywords: customer reviews, Multinomial Naïve Bayes (MNB), Sentiment Analysis (SA), *stopword*

Abstrak. Penerapan dari machine learning dalam bisnis telah memungkinkan produsen atau penjual untuk mengetahui kualitas produk dagangan mereka berdasarkan pada analisis ulasan pelanggan menggunakan Sentiment Analysis (SA). Penelitian ini bertujuan untuk mengetahui pengaruh dari jenis stopwords terhadap akurasi dari metode Multinomial Naïve Bayes (MNB) dalam proses SA. Terdapat 10 jenis stopwords yang digunakan dalam penelitian ini: umum, konjungsi, bahasa gaul, keterangan waktu, kata benda, kata ganti orang, kata seruan, kata kerja, dan kata dengan satu huruf. Berdasarkan pada uji Friedman pada tiga ulasan dari tiga produk sepatu, diketahui bahwa menghilangkan stopwords konjungsi (MNB-konjungsi) dapat meningkatkan akurasi model MNB dalam proses SA sebesar 1%. Hasil uji T pada dua dari tiga himpunan data menunjukkan bahwa MNB-konjungsi memiliki akurasi yang lebih baik ketimbang MNB tanpa menghilangkan stopwords.

Kata Kunci: ulasan pelanggan, Multinomial Naïve Bayes (MNB), Sentiment Analysis (SA), *stopword*

1. Pendahuluan

Revolusi industri 4.0 telah menjadikan teknologi memiliki peranan penting dalam kehidupan manusia, salah satunya di bidang *marketing*, dimana berbagai algoritma *machine learning* telah digunakan untuk menunjang kepentingan bisnis [1], [2], [3]. Sebagai contoh, algoritma *machine learning* dapat digunakan untuk proses *sentiment analysis* (SA) yang merupakan suatu proses untuk menganalisis komentar dari konsumen terhadap produk tertentu [4], [5]. Secara umum, SA dapat digunakan untuk membantu para penjual dalam memahami perspektif dari pembeli terhadap produk yang mereka jual [6]. Dengan mengetahui dengan jelas pandangan dari konsumen maka produk penerus dari produk yang sekarang ini beredar dapat disusun sedemikian rupa sehingga memuaskan konsumen, yang pada akhirnya akan meningkatkan profit margin dari penjual [7], [8].

Salah satu proses yang memengaruhi hasil dari SA adalah proses menghilangkan *stopword* [9]. *Stopword* merupakan kata-kata yang dianggap tidak signifikan atau netral memengaruhi kelas dari komentar, seperti kata dan, atau, di, ke, dan seterusnya [10], [11]. Dengan

menghilangkan kata-kata tersebut, maka akan dihasilkan *bag of words* atau kumpulan kata yang lebih signifikan dalam mengklasifikasikan kelas dari komentar [12]. Beberapa penelitian terkait dengan menghilangkan *stopword* dalam proses SA telah dipublikasikan pada penelitian berikut: [13], [14], [15], [16], [17].

Peneliti pada [13] mengidentifikasi pengaruh dari *stopword* dalam memengaruhi hasil klasifikasi dari komentar Youtube dengan menggunakan metode *Bidirectional Long Short-Term Memory* (Bi-LSTM). Berdasarkan pada penelitian yang telah dilakukan, disimpulkan bahwa penggunaan *stopword* dapat mengurangi akurasi dari model karena terlalu banyak kata yang dibuang pada proses *stemming* atau pengembalian kata pada bentuk kata dasarnya. Penelitian menunjukkan bahwa proses membuang *stopwords* meningkatkan akurasi dari proses SA dalam mengidentifikasi 50.000 komentar yang diambil dari jejaring sosial X (dahulu Twitter) dengan akurasi 88% [14]. Proses menghapus *stopwords* dibutuhkan untuk menganalisis komentar dalam bahasa Urdu dengan menggunakan metode *Support Vector Machine* (SVM) dan *Multinomial Naive-Bayes* (MNB) [15]. Proses penghapusan *stopword* dibutuhkan untuk meringkas model SA *hybrid* [16]. Berdasarkan pada penelitian tersebut disimpulkan bahwa metode SA *hybrid* dapat mengklasifikasikan komentar dari situs iMBD dan X dengan akurasi hingga 94%. Terakhir, penelitian menunjukkan bahwa metode SVM yang digabungkan dengan *deep learning* menghasilkan akurasi sebesar 92% dalam mengklasifikasikan komentar pada jejaring sosial X [17].

Berdasarkan pada hasil penelitian yang telah dilakukan sebelumnya, terlihat bahwa untuk ukuran data yang besar [14], maka menghapus *stopword* dibutuhkan untuk menghasilkan model prediktif yang akurat. Namun, untuk data yang memiliki jumlah sampel kecil, penghapusan *stopword* dapat mengakibatkan penurunan dari akurasi model prediktif [13]. Hal ini menunjukkan bahwa penggunaan *stopword* dalam proses SA, tidak dapat digunakan secara sembarangan. Melihat permasalahan di atas, maka penentuan *stopword* yang perlu dihapus menjadi topik yang menarik untuk dibahas. Dalam kasus ini, perlu untuk dipelajari apakah menghapus seluruh *stopword* dari komentar dapat mengurangi akurasi model ataukah ada jenis *stopword* tertentu yang sebaiknya dihilangkan untuk meningkatkan performa klasifikasi dari model. Penelitian ini dilakukan untuk mengetahui bagaimana jenis *stopword* memengaruhi hasil dari proses SA. Secara khusus, penelitian ini memberikan dua kontribusi penelitian sebagai berikut. (1) Memvalidasi secara statistika (melalui uji Friedman) bahwa terdapat jenis *stopword* yang sebaiknya dihilangkan untuk meningkatkan akurasi model prediktif dari SA berbasis metode MNB. (2) Proses validasi dilakukan pada himpunan data yang diperoleh dari komentar produk sepatu yang diambil dari situs Shopee dan Tokopedia, yang merupakan situs *e-commerce* Indonesia.

Untuk memvalidasi metodologi pada penelitian ini digunakan data yang berasal dari penelitian [18]. Data yang digunakan merupakan data yang berasal dari tiga merek sepatu yang dijual di Indonesia. Pada data tersebut terdapat dua kelas data, yaitu positif dan negatif. Terkait dengan penelitian tersebut, penelitian menggunakan metode NB klasik untuk mengklasifikasikan komentar dari produk, sedangkan pada penelitian ini, digunakan metode MNB.

Penelitian ini dilakukan dengan mengikuti metodologi penelitian yang dilakukan yang menggunakan metode MNB [15]. Metode ini digunakan karena metode MNB memberikan hasil klasifikasi yang baik dengan nilai F1 sebesar 0,91 sesuai dengan penelitian yang telah dilakukan. Lebih lanjut, penelitian terkait metode MNB seperti pada penelitian [19], [20], [21] menunjukkan bahwa metode MNB memiliki performa yang baik dalam mengklasifikasikan data yang berbentuk teks. Terlebih, untuk n buah sampel, m buah fitur, dan k buah kelas, kompleksitas waktu dari metode MNB hanya berada pada $O(mn \cdot \log(k))$ [22]. Sehingga, meskipun bukan yang terbaik, namun metode ini masih dapat digunakan untuk jumlah data yang tidak terlalu besar dengan waktu perhitungan yang relatif singkat. Berbeda dari penelitian yang dilakukan sebelumnya, penelitian ini tidak bertujuan untuk mengimprovisasi akurasi model prediktif dari metode MNB. Melainkan, ingin mempelajari bagaimana setiap jenis *stopword* memengaruhi akurasi dari model MNB.

2. Metode Multinomial Naïve Bayes

Pada bagian ini akan dibahas bagaimana metode MNB digunakan untuk mengidentifikasi jenis *stopword* yang sebaiknya digunakan dalam proses SA. Silakan mengacu pada penelitian [23], [24] terkait dengan dasar dari metode MNB dan terkait dengan dasar dari proses SA [25].

Tinjau sebuah himpunan data teks T yang direpresentasikan ke dalam bentuk vektor kata, dimana $T = \{t_1, t_2 \dots t_n\}$; $T \in \mathbb{Z}^{m \times n}$ dengan $t_i = [t_i(1) t_i(2) \dots t_i(m)]^T$; $t_i \in \mathbb{Z}^m$; $i = 1, 2, \dots n$ menyatakan kata ke- i dan $t_i(k)$; $k = 1, 2, \dots m$ menyatakan ada berapa kali kata t_i muncul pada sampel teks ke- k . Dimisalkan bahwa $C = \{c_1, c_2, \dots, c_p\}$; $C \in \mathbb{Z}^p$ menyatakan kelas yang ada di dalam data T , dimana $c = [c(1) c(2) \dots c(n)]^T$; $c \in C^n$ menyatakan vektor kelas teks yang berkorespondensi dengan himpunan data T sehingga $\{\tilde{t}_i, c(i)\}$; $\tilde{t} = [t_1(i) t_2(2) \dots t_n(i)]$; $t_i \in T$ menyatakan pasangan teks dan kelas untuk sampel ke- i . Berdasarkan pada definisi di atas, maka MLB akan menentukan sebuah fungsi matematis yang menghubungkan setiap nilai dari T dengan kelas yang sesuai berdasarkan pada perhitungan probabilitas. Untuk mencapai target ini, maka terdapat dua tahapan utama, yaitu hapus *stopword* dan hitung prediksi kelas.

Tahap pertama. Tinjau definisi data teks seperti pada paragraf sebelumnya. Misalkan bahwa $T_s = [t_{a(1)} t_{a(2)} \dots t_{a(b)}]$; $T_s \subset T$ merupakan submatriks atau matriks bagian dari T yang menampung informasi *stopword* yang ingin dihilangkan dari T , dimana $a = [a(1) a(2) \dots a(b)]$; $a \in \mathbb{Z}^b$ merupakan sebuah vektor yang berisikan indeks yang menandakan kata mana saja yang termasuk *stopword* dari himpunan data T . Dalam kasus ini dapat didefinisikan sebuah matriks baru $T_{s'}$, dimana $T_{s'} \in T$ merupakan matriks yang berisikan kata bukan *stopword* atau $T_{s'}$ menyatakan komplemen dari T_s . Dengan sedikit penyalahgunaan notasi (*abuse of notation*), maka $T_{s'}$ dapat dinotasikan sesuai dengan Persamaan 1 atau $T_{s'}$ merupakan himpunan data teks selain dari yang dikategorikan sebagai *stopword*. Dengan menggunakan $T_{s'}$ dan C sesuai dengan pembahasan sebelumnya, maka dapat diturunkan model untuk memprediksi kelas dari setiap sampel W dari data yang telah dihilangkan *stopword*-nya.

$$T_{s'} = [t_i \in T: t_i \notin T_s; i = 1, 2, \dots n] \quad (1)$$

Tahap kedua. Tinjau $T_{s'}$ dan C . Untuk sebuah kelas k , misalkan $c'_k = \frac{c}{|c_k|}$ sebagai sebuah himpunan kelas komplemen dari c_k . Lebih lanjut, misalkan $\delta_{i,j}$ sebagai sebuah fungsi indikator (seperti pada *delta Dirac*) yang dinyatakan dalam Persamaan 2. Probabilitas untuk kelas c_k dapat dinyatakan dalam Persamaan 3. Dengan $|C|$ menyatakan kardinalitas atau jumlah elemen dari himpunan kelas. Lebih lanjut *prior probability* untuk kata $t_i \in T_s$ diklasifikasikan sebagai kelas c_k dapat dinyatakan dalam Persamaan 4. Persamaan 4 dapat dipahami sebagai rasio antara penjumlahan dari berapa banyak kata t_i yang tergolong dalam kelas c_k dibagi dengan jumlah semua kata yang tergolong ke dalam kelas c_k .

$$\delta_{i,j} = \begin{cases} 1 & \text{jika } i=j \\ 0 & \text{jika sebaliknya} \end{cases} \quad (2)$$

$$\mathbb{P}(c_k) = \frac{\sum_{i=1}^n \delta(c(i), c_k) + 1}{|C| + m} \quad (3)$$

$$\mathbb{P}(t_i | c_k) = \frac{\sum_{l=1}^n (\delta(c(l), c_k) \cdot \sum_{j=1}^m t_i(j)) + 1}{\sum_{l=1}^n (\delta(c(l), c_k) \cdot \sum_{j=1}^m t_i(j)) + n} \quad (4)$$

Dengan menggunakan definisi dari probabilitas kelas dan *prior probability* sesuai dengan Persamaan 3 dan Persamaan 4, maka prediksi kelas dari metode MNB dapat ditentukan. Secara khusus, prediksi dengan metode MNB untuk sebuah sampel ke- W dinyatakan dalam Persamaan

5, dimana $\hat{c}(W)$ menyatakan tebakan kelas untuk sampel ke- W . Persamaan 5 merupakan estimasi kelas untuk sampel ke- W dari himpunan data T_s . Secara umum T_s dapat menjadi lebih dari satu himpunan saja. Dalam kasus ini, tanpa mengubah keumuman (*without loss of generality*), Persamaan 3 sampai Persamaan 5 dapat disesuaikan dengan himpunan data *stopword* selain dari T_s .

$$\hat{c}(w) = \underset{c_k \in C}{\arg \max} \left(\mathbb{P}(c_k) \prod_{i=1}^n P(t_i | c_k) \right) \quad (5)$$

3. Metodologi Penelitian

Pada bagian ini akan dijelaskan data yang digunakan untuk percobaan. Lebih lanjut akan dibahas alur dan *experiment setup* dari penelitian. Terakhir, akan dibahas besaran pengukuran yang digunakan untuk menilai performa dari model untuk setiap jenis *stopword* yang dihilangkan.

3.1. Data Penelitian

Data yang digunakan merupakan data kumpulan komentar dari pembeli sepatu yang ada di *e-commerce* yang ada di Indonesia. Terdapat tiga merek sepatu yang ditinjau dalam penelitian ini. Namun, untuk menjaga netralitas dari penelitian ini, ketiga merek sepatu yang digunakan akan disamarkan namanya menjadi sepatu merek A, B, dan C. Untuk setiap merek sepatu terdapat tiga kelas, yaitu negatif, positif, dan netral. Namun, jumlah sampel dengan kelas netral sangat sedikit, sehingga pada penelitian ini, kelas netral akan diasosiasikan sebagai kelas negatif. Hal ini berdasarkan pada pertimbangan bahwa sampel yang terindikasi netral sebenarnya bisa atau cenderung negatif. Sehingga dalam kasus ini, sampel tersebut akan dikategorikan ke dalam kelas negatif.

Terdapat 1.260 sampel untuk ketiga merek sepatu. Namun, jumlah perbandingan sampel positif dan negatifnya sedikit berbeda untuk tiap merek. Untuk sepatu merek A, dari 1.260 sampel komentar, terdapat sampel positif sejumlah 1.103 dan 157 sampel memiliki kelas negatif. Untuk merek B, terdapat 1.105 sampel positif dan 155 sampel negatif. Sedangkan untuk merek C, terdapat 1.106 sampel positif dan 154 sampel negatif.

Pada penelitian ini, terlihat bahwa jumlah perbandingan antara kelas positif dan negatif sangat tidak berimbang. Oleh karena itu, diperlukan penyesuaian terhadap jumlah data. Dalam kasus ini, seluruh sampel negatif akan diambil untuk tiap merek sepatu. Kemudian akan diambil sampel positif sebanyak dua kali lipat dari sampel negatif secara acak, sehingga rasio dari sampel positif dan negatif menjadi 2:1. Sebagai contoh untuk merek B, maka akan digunakan data yang terdiri dari 155 sampel negatif dan 310 sampel positif, sehingga total terdapat 465 sampel untuk merek B, dan seterusnya. Proporsi ini dipilih tanpa ada pertimbangan khusus (secara sembarang) dan dapat disesuaikan dengan keinginan dari pengguna. Setelah itu, akan dipilih 80% data untuk melatih model MNB dan 20% yang tersisa dari tiap merek akan digunakan untuk memvalidasi akurasi dari model prediktif. Detail terkait dengan pengaturan simulasi akan dijelaskan pada sub bagian 3.2.

3.2. Konfigurasi Penelitian (*Experiment Setup*)

Penelitian ini bertujuan untuk menguji apakah ada jenis *stopword* tertentu yang meningkatkan akurasi model prediktif dari proses SA. *Stopword* yang digunakan diperoleh dari berbagai sumber, salah satunya dengan menggunakan Kamus Besar Bahasa Indonesia (KBBI), terutama untuk menentukan jenis yang tepat. Dalam penelitian ini, tidak dimungkinkan untuk menuliskan seluruh kata yang termasuk ke dalam jenis tersebut akibat kompleksitas dari bahasa Indonesia sehingga, kata yang dituliskan merupakan kata yang dapat ditemukan di dalam kamus atau sering digunakan dalam komunikasi tertulis yang diambil dengan proses *crawling*. Oleh karena itu, terdapat 10 jenis *stopword* yang akan diujikan dalam penelitian ini, seperti yang dapat dilihat pada Tabel 1.

Tabel 1. 10 Jenis Stopwords yang Akan Diujikan

No	Jenis Stopword	Keterangan
1	Umum	Stopword yang lazim dihilangkan ketika melakukan SA. Stopword ini berlaku secara umum dan dapat berisikan kata dari sembilan jenis <i>stopword</i> yang lain.
2	Konjungsi	Kata yang digunakan sebagai kata sambung. Contoh kata dari jenis ini antara lain: yang, oleh, supaya, maka, dan seterusnya.
3	Gaul	Jenis kata era modern yang merupakan hasil pelesetan dari bahasa Indonesia atau singkatan. Contoh kata jenis ini adalah gue, elo, doang, y, ngak, dan seterusnya.
4	Keterangan Waktu	Kata yang menyatakan keterangan waktu dalam sebuah kalimat. Contoh kata dari jenis ini adalah kemarin, hari ini, Senin, Rabu, Januari, dan seterusnya.
5	Keterangan Benda	Kata benda yang bersifat netral dan tidak menentukan kelas dari sebuah sampel. Contoh kata dari jenis ini adalah manusia, hal, benda, dunia, negara, dan seterusnya.
6	Kata Kerja	Kumpulan dari kata kerja yang bersifat netral dan tidak mengklasifikasikan kelas dari sebuah sampel. Contoh kata dari jenis ini adalah ada, memiliki, mempunyai, menonton, menilai, dan seterusnya.
7	Kata Ganti Orang (Pronomina)	Kata yang digunakan untuk mengganti nama dari subjek. Contoh kata dari jenis ini adalah saya, dia, mereka, kami, dan seterusnya.
8	Kata Seruan	Kata yang mewakili ekspresi tertentu. Contoh kata dari jenis ini adalah wah, ya, ayo, cuma, aja, dan seterusnya.
9	Keterangan Umum (Adverbia)	Kata keterangan yang lazim digunakan dalam sebuah kalimat, namun tidak dapat digunakan untuk klasifikasi kelas. Contoh kata dari jenis ini adalah sangat, pasti, terlalu, hanya, saja, dan seterusnya.
10	Kata Satu Huruf	Jenis satu kata yang hanya dituliskan dalam satu huruf, biasanya karena penyingkatan. Contoh kata dari jenis ini adalah y (dari kata ya), g (ngak), x (kali), dan seterusnya.

Seluruh *stopword* akan diterapkan pada himpunan data untuk mengetahui bagaimana pengaruh dari setiap jenis *stopword* terhadap akurasi model prediktif dari SA. Terdapat 12 simulasi yang dilakukan, yaitu: 10 simulasi untuk setiap jenis *stopword*, satu simulasi untuk proses SA tanpa menggunakan *stopword*, dan satu simulasi dengan menggunakan seluruh *stopword*. Untuk 10 simulasi tunggal hanya ada satu jenis *stopword* yang akan dihilangkan per simulasi. Sebagai contoh, untuk simulasi *stopword* konjungsi, maka hanya *stopword* konjungsi yang dihilangkan dari data, sementara kata yang merupakan bagian dari jenis *stopword* lain tidak akan dihilangkan. Semua proses simulasi memiliki alur proses SA yang sama, dimulai dari menghilangkan tanda baca dan *emoticon*, mengubah semua huruf besar ke huruf kecil, menghasilkan token dari sampel, dan melatih model NB [26], [27].

Untuk menentukan kualitas prediksi dari setiap model yang dibangun, maka digunakan parameter akurasi. Akurasi mengacu pada seberapa banyak tebakan yang dihasilkan model relatif terhadap jumlah sampel yang ada. Sebagai contoh, T_P atau *true positive* menyatakan jumlah tebakan kelas positif yang benar, T_N atau *true negative* menyatakan jumlah tebakan kelas negatif yang benar, F_P atau *false positive* menyatakan jumlah tebakan kelas positif yang salah (ditebak sebagai positif, namun sebenarnya negatif), dan F_N atau *false negative* menyatakan seberapa banyak tebakan negatif yang salah (ditebak negatif, namun sebenarnya positif). Berdasarkan definisi tersebut, maka akurasi dapat dinyatakan dengan Persamaan 6.

$$A\% = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \times 100\% \quad (6)$$

Untuk memastikan bahwa hasil pengukuran akurasi yang diperoleh valid, semua simulasi akan diulang sebanyak 30 kali dengan pengacakan data. Dalam kasus ini, untuk setiap percobaan, semua sampel negatif yang digunakan akan tetap sama. Untuk sampel positif, setiap pengulangan akan dipilih sampel yang berbeda dari pengulangan sebelumnya, sehingga sampel positif dan negatif yang digunakan untuk melatih model akan berbeda di setiap pengulangan.

Tahap selanjutnya adalah untuk menentukan jenis *stopword* yang sebaiknya dihilangkan untuk meningkatkan performa model MNB. Dalam kasus ini, hasil akurasi yang dihasilkan oleh 11 jenis *stopword* (simulasi tanpa *stopword* tidak dimasukkan) akan diuji dengan menggunakan uji Friedman [28]. Uji Friedman merupakan pengujian untuk menentukan apakah terdapat

perbedaan dari sekumpulan data. Dalam kasus ini, uji Friedman digunakan untuk memvalidasi apakah terdapat sekurang-kurangnya satu jenis *stopword* yang jika dihilangkan akan meningkatkan akurasi dari model MNB. Terkait permasalahan ini, maka hipotesis yang diujikan pada penelitian ini adalah sebagai berikut.

Untuk sepatu merek A:

H_0 : Tidak terdapat perbedaan dari akurasi model prediktif untuk jenis *stopword* yang berbeda dalam memprediksi kelas komentar sepatu merek A.

H_a : Terdapat sekurang-kurangnya satu jenis *stopword* yang memiliki akurasi model prediktif yang berbeda dari jenis yang lain dalam memprediksi kelas komentar sepatu merek A.

Untuk sepatu merek B:

H_0 : Tidak terdapat perbedaan dari akurasi model prediktif untuk jenis *stopword* yang berbeda dalam memprediksi kelas komentar sepatu merek B.

H_b : Terdapat sekurang-kurangnya satu jenis *stopword* yang memiliki akurasi model prediktif yang berbeda dari jenis yang lain dalam memprediksi kelas komentar sepatu merek B.

Untuk sepatu merek C:

H_0 : Tidak terdapat perbedaan dari akurasi model prediktif untuk jenis *stopword* yang berbeda dalam memprediksi kelas komentar sepatu merek C.

H_c : Terdapat sekurang-kurangnya satu jenis *stopword* yang memiliki akurasi model prediktif yang berbeda dari jenis yang lain dalam memprediksi kelas komentar sepatu merek C.

Jika hipotesis nol ditolak, maka tahap berikutnya adalah menguji signifikansi model dengan akurasi terbaik terhadap model MNB yang tidak melalui proses pembuangan *stopword*. Tujuannya untuk mengetahui apakah penggunaan *stopword* secara signifikan memengaruhi kualitas dari model prediktif. Lebih lanjut, akan ditunjukkan perbandingan dari kedua metode berdasarkan parameter lain seperti *precision*, *recall*, dan *F1-score*.

4. Hasil dan Pembahasan

Pada bagian ini akan ditampilkan hasil simulasi untuk setiap jenis *stopword* yang digunakan. Untuk mempermudah pembahasan, maka analisis akan dilakukan berdasarkan pada tiap merek sepatu.

4.1. Akurasi Model Prediktif: Sepatu Merk A

Tabel 2 menunjukkan akurasi model prediktif dari 12 simulasi yang dilakukan dengan menggunakan data dari sepatu merek A. Nomor simulasi mengacu pada definisi yang telah dijelaskan pada sub bagian 3.2. Sebagai contoh, simulasi pertama mengacu pada metode MNB yang disimulasikan dengan menghapus jenis *stopword* umum dari komentar dan seterusnya. Simulasi ke-11 mengacu pada MNB dengan menghapus semua jenis *stopword* dari nomor 1 hingga 10, sedangkan simulasi ke-12 mengacu pada model MNB yang disimulasikan tanpa menghapus *stopword* apapun.

Berdasarkan eksperimen yang dilakukan diketahui bahwa metode MNB dengan menghapus jenis *stopword* konjungsi (MNB-konjungsi) merupakan metode yang menghasilkan akurasi tertinggi, dengan nilai akurasi sebesar 95,1%. Berdasarkan Tabel 2 terlihat bahwa model MNB yang hanya menghapus *stopword* umum (MNB-umum) dan MNB yang menghapus seluruh *stopword* (MNB-semua) memiliki akurasi yang lebih rendah jika dibandingkan dengan model MNB yang dilatih tanpa menghapus *stopword* apapun (MNB-netral). Hal ini menunjukkan bahwa menghapus seluruh *stopword* dapat menurunkan akurasi model prediktif dari metode MNB.

Tabel 2. Perbandingan Rerata Akurasi Model Prediktif Metode MNB untuk Mengklasifikasikan Komentar terhadap Sepatu Merek A, B, dan C (Dalam Satuan %)

Jenis Data	Simulasi ke-											
	1	2	3	4	5	6	7	8	9	10	11	12
Merek A	93,2	95,1*	93,7	94,8	94,2	94,8	94,5	93,9	93,9	94,8	92,9	94,9
Merek B	93,5	95,4*	93,2	94,8	94,2	94,8	94,8	94,8	94,5	94,8	93,1	95,0
Merek C	94,2	96,5*	95,4	96,2	96,2	96,3	96,2	96,0	95,5	96,1	94,0	95,7

* metode dengan rerata akurasi tertinggi

Tahap berikutnya adalah uji Friedman untuk menentukan apakah ada jenis *stopword* yang memiliki akurasi yang lebih baik daripada yang lain. Hasil dari pengujian Friedman menunjukkan nilai $Q = 53,6$ dengan derajat kebebasan 10 dan rentang kepercayaan ($\alpha = 0,05$). Hasil ini setara dengan p -value sebesar $5,6 \times 10^{-8}$ yang menunjukkan bahwa terdapat satu model yang lebih baik dibandingkan dengan model yang lain, yaitu model MNB-konjungsi. Oleh karena itu, dapat disimpulkan bahwa H_0 ditolak dan H_a diterima. Berdasarkan hasil ini, tahap selanjutnya adalah mengujikan signifikansi dari model MNB-konjungsi dengan model MNB-netral.

Tabel 3. Nilai Rerata Akurasi, Precision, Recall, F1-score, dan Uji T untuk Sepatu Merek A

Parameter	MNB-Konjungsi	MNB-Netral
A%	95,10%	94,48%
P%	95,65%	95,35%
R%	97,89%	98,21%
F1	96,69%	96,68%
St. Deviasi	0,035	0,038
p -value (sepihak)		0,163
p -value (dua pihak)		0,326

Tabel 3 menunjukkan rerata dari akurasi, *precision*, *recall*, *F1-score*, dan hasil uji T untuk metode MNB-konjungsi dan MNB-netral. Hipotesis yang diujikan adalah sebagai berikut.

H_0 : Tidak terdapat perbedaan dari akurasi model prediktif antara MNB-konjungsi dan MNB-netral dalam mengklasifikasikan komentar sepatu merek A.

H_a : Akurasi model prediktif metode MNB-konjungsi lebih tinggi daripada metode MNB-netral dalam mengklasifikasikan komentar sepatu merek A.

4.2. Akurasi Model Prediktif: Sepatu Merek B

Berdasarkan hasil pada Tabel 2, untuk sepatu merek B terlihat bahwa MNB-konjungsi menghasilkan akurasi tertinggi jika dibandingkan dengan model MNB lain. Terkait dengan uji Friedman, diperoleh hasil $Q = 42,71$ dengan derajat kebebasan 10 dan rentang kepercayaan 5%, dimana nilai ini setara dengan p -value sebesar $5,59 \times 10^{-6}$. Hasil ini menunjukkan bahwa H_0 ditolak untuk sepatu merek B dan disimpulkan bahwa terdapat MNB-konjungsi lebih baik dibanding metode yang lain.

Tabel 4 menunjukkan hasil uji T antara metode MNB-konjungsi dan MNB-netral. Seperti pada proses sebelumnya, hipotesis yang diajukan untuk sepatu merek B adalah sebagai berikut.

H_0 : Tidak terdapat perbedaan dari akurasi model prediktif antara MNB-konjungsi dan MNB-netral dalam mengklasifikasikan komentar sepatu merek B.

H_b : Akurasi model prediktif metode MNB-konjungsi lebih tinggi daripada metode MNB-netral dalam mengklasifikasikan komentar sepatu merek B.

Berdasarkan hasil dari Tabel 4, terlihat bahwa nilai p -value kurang dari 0,05. Hal ini menandakan bahwa hanya ada peluang lebih kecil dari 5% untuk menyatakan bahwa hasil ini terjadi secara kebetulan, sehingga terdapat bukti secara statistik untuk menolak H_0 dan menerima H_b . Hasil ini menunjukkan bahwa dengan membuang *stopword* dari kategori konjungsi berpotensi untuk meningkatkan performa model MNB dan tidak diperlukan untuk membuang seluruh *stopword* karena berpotensi untuk mengurangi akurasi model prediktif.

Tabel 4. Nilai Rerata Akurasi, Precision, Recall, F1-score, dan Uji T untuk Sepatu Merek B

Parameter	MNB-Konjungsi	MNB-Netral
A%	95,44%	93,2%
P%	95,38%	93,0%
R%	97,95%	98,1%
F1	96,59%	95,4%
St. Deviasi	0,030	0,030
p -value (sepihak)		0,025
p -value (dua pihak)		0,051

4.3. Akurasi Model Prediktif: Sepatu Merek C

Konsisten dengan kedua hasil sebelumnya, terlihat dari Tabel 2 bahwa metode MNB-konjungsi memiliki akurasi prediksi tertinggi dengan nilai sebesar 96,5%. Hasil dari uji Friedman juga menunjukkan bahwa terdapat model MNB yang lebih baik dari yang lain (dalam kasus ini model MNB-konjungsi) dengan p -value sebesar $6,97 \times 10^{-9}$ dan $Q = 58,49$.

Tabel 5 menunjukkan hasil uji T dari metode MNB-konjungsi dengan MNB-netral. Adapun hipotesis yang diajukan untuk sepatu merek C adalah sebagai berikut.

H_0 : Tidak terdapat perbedaan dari akurasi model prediktif antara MNB-konjungsi dan MNB-netral dalam mengklasifikasikan komentar sepatu merek C.

H_c : Akurasi model prediktif metode MNB-konjungsi lebih tinggi daripada metode MNB-netral dalam mengklasifikasikan komentar sepatu merek C.

Tabel 5. Nilai Rerata Akurasi, Precision, Recall, F1-score, dan Uji T untuk Sepatu Merek C

Parameter	MNB-Konjungsi	MNB-Netral
A%	96,53%	95,69%
P%	96,24%	95,74%
R%	98,66%	98,01%
F1	97,37%	96,79%
St. Deviasi	0,028	0,027
p -value (sepihak)		0,0014
p -value (dua pihak)		0,0028

5. Kesimpulan

Penelitian ini bertujuan untuk mengidentifikasi bagaimana jenis *stopword* memengaruhi akurasi model prediktif dari metode MNB dalam proses SA. Berdasarkan pada penelitian yang telah dilakukan disimpulkan bahwa dari 10 jenis *stopword* yang diajukan, terdapat bukti secara statistik bahwa metode MNB yang dilakukan dengan menghilangkan *stopword* berjenis konjungsi memiliki akurasi model prediktif yang lebih baik daripada MNB yang menghilangkan semua *stopword*. Untuk ketiga himpunan data yang digunakan terlihat bahwa metode MNB-konjungsi memiliki akurasi model prediksi tertinggi jika dibandingkan dengan metode MNB yang dibangun dengan *stopword* umum dan semua *stopword*. Hal ini menunjukkan bahwa membuang *stopword* berjenis konjungsi dapat meningkatkan performa dari metode MNB dalam proses SA.

Hasil uji T terhadap MNB-netral menunjukkan bahwa metode MNB-konjungsi memiliki akurasi model prediktif yang lebih baik daripada MNB-netral. Hal ini mengisyaratkan bahwa pemilihan dari *stopword* hendaknya tidak dilakukan secara sembarang tanpa memperhatikan jenis *stopword* yang digunakan. Arah penelitian lanjutan dari penelitian ini adalah dengan menggunakan kosakata yang lebih besar untuk setiap jenis *stopword*. Saat ini, penelitian ini hanya memanfaatkan kamus KBBI dan pengamatan dalam menyusun kamus *stopword*. Penelitian lanjutan dapat meninjau bagaimana efisiensi waktu dari setiap metode untuk menghasilkan model yang efisien namun tetap handal.

Referensi

- [1] M. K. Gourisaria, R. Agrawal, G. Harshvardhan, M. Pandey, and S. S. Rautaray, "Application of Machine Learning in Industry 4.0," *Studies in Big Data*, pp. 57–87, 2021.
- [2] T. V. N. Rao, A. Gaddam, M. Kurni, and K. Saritha, "Reliance on Artificial Intelligence, Machine Learning and Deep Learning in the Era of Industry 4.0," *Smart Healthcare System Design: Security and Privacy Aspects*, 2021, pp. 281–299. doi: 10.1002/9781119792253.ch12.
- [3] T. H. Gan, J. Kanfoud, H. Nedunuri, A. Amini, and G. Feng, "Industry 4.0: Why Machine Learning Matters?," *Lecture Notes in Mechanical Engineering (LNME)*, pp. 397–404, 2021, doi: 10.1007/978-981-15-9199-0_37.

- [4] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 55, pp. 5731–5780, 2022, doi: 10.1007/s10462-022-10144-1.
- [5] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, 2021, doi: 10.1007/s13278-021-00776-6.
- [6] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, no. 1, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.
- [7] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, "Sentiment analysis for customer review: Case study of Traveloka," *Procedia Computer Science*, vol. 216, pp. 682–690, 2023, doi: 10.1016/j.procs.2022.12.184.
- [8] H. Huang, A. A. Zavareh, and M. B. Mustafa, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions," *IEEE Access*, vol. 11, pp. 90367–90382, 2023, doi: 10.1109/ACCESS.2023.3307308.
- [9] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences*, vol. 12, no. 17, p. 8765, 2022, doi: 10.3390/app12178765.
- [10] A. A. Wazrah and S. Alhumoud, "Sentiment Analysis Using Stacked Gated Recurrent Unit for Arabic Tweets," *IEEE Access*, vol. 9, pp. 137176–137187, 2021, doi: 10.1109/ACCESS.2021.3114313.
- [11] T. H. Jaya Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," *Procedia Computer Science*, vol. 197, pp. 660–667, 2022, doi: 10.1016/j.procs.2021.12.187.
- [12] Y. S. Mehanna and M. Bin Mahmuddin, "A Semantic Conceptualization Using Tagged Bag-of-Concepts for Sentiment Analysis," *IEEE Access*, vol. 9, pp. 118736–118756, 2021.
- [13] A. Santosa, I. Purnamasari, and R. Mayasari, "Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 6, no. 1, pp. 81–93, 2022.
- [14] U. D. Gandhi, P. M. Kumar, G. C. Babu, and G. Karthick, "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wireless Personal Communications*, 2021, doi: 10.1007/s11277-021-08580-3.
- [15] M. Z. Ali, Ehsan-Ul-Haq, S. Rauf, K. Javed, and S. Hussain, "Improving Hate Speech Detection of Urdu Tweets Using Sentiment Analysis," *IEEE Access*, vol. 9, pp. 84296–84305, 2021, doi: 10.1109/ACCESS.2021.3087827.
- [16] K. L. Tan, C. P. Lee, K. M. Lim, and K. S. M. Anbananthen, "Sentiment Analysis With Ensemble Hybrid Deep Learning Model," *IEEE Access*, vol. 10, pp. 103694–103704, 2022, doi: 10.1109/ACCESS.2022.3210182.
- [17] C. P. D. Cyril, J. R. Beulah, N. Subramani, P. Mohan, A. Harshavardhan, and D. Sivabalaselvamani, "An automated learning model for sentiment analysis and data classification of Twitter data using balanced CA-SVM," *Concurrent Engineering*, vol. 29, no. 4, pp. 386–395, 2021, doi: 10.1177/1063293X211031485.

- [18] R. Setiawan and J. Tjen, "Identifying Opinions of Footwear Products in Indonesia via Sentiment Analysis." *International Conference on Digital Business Innovation and Technology Management (ICONBIT)*, vol. 1, no. 1, pp. 561-565, 2024.
- [19] K. Kaushik *et al.*, "Multinomial Naive Bayesian Classifier Framework for Systematic Analysis of Smart IoT Devices," *Sensors*, vol. 22, no. 19, p. 7318, 2022, doi: 10.3390/s22197318.
- [20] T.-T. Wong and H.-C. Tsai, "Multinomial naïve Bayesian classifier with generalized Dirichlet priors for high-dimensional imbalanced data," *Knowledge-Based Systems*, vol. 228, p. 107288, 2021, doi: 10.1016/j.knosys.2021.107288.
- [21] N. E. Allali, M. Fariss, H. Asaidi, and M. Bellouki, "Multinomial Naive Bayes Categorization for Semantic Web Services," in *2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)*, IEEE, 2021, pp. 74–79. doi: 10.1109/ICDATA52997.2021.00023.
- [22] A. Askari, A. D', A. Laurent, and E. Ghaoui, "Naive Feature Selection: Sparsity in Naive Bayes," 2020. [Online]. Available: <https://github.com/aspremon/NaiveFeatureSelection>.
- [23] S. Xu, Y. Li, and Z. Wang, "Bayesian Multinomial Naïve Bayes Classifier to Text Classification," *Lecture notes in mechanical engineering*, 2017, pp. 347–352. doi: 10.1007/978-981-10-5041-1_57.
- [24] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," in *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, IEEE, 2019, pp. 593–596. doi: 10.1109/ICACTM.2019.8776800.
- [25] J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8469–8510, 2023, doi: 10.1007/s10462-022-10386-z.
- [26] C. Therence and J. Tjen, "Analysis of Cosmetic Product Opinions on E-Commerce Based on Naïve Bayes Classifier.", *International Conference on Digital Business Innovation and Technology Management (ICONBIT)*, vol. 1, no. 1, pp. 328-334, 2024.
- [27] I. Salim and J. Tjen, "A Comparison of Online Investment Application Opinion Based on Sentiment Analysis." *International Conference on Digital Business Innovation and Technology Management (ICONBIT)*, vol. 1, no. 1, pp. 541-547, 2024.
- [28] J. Ma *et al.*, "Metaheuristic-based support vector regression for landslide displacement prediction: a comparative study," *Landslides*, vol. 19, no. 10, pp. 2489–2511, 2022, doi: 10.1007/s10346-022-01923-6.