

PENGGUNAAN SELEKSI FITUR *CHI-SQUARE* DAN ALGORITMA MULTINOMIAL *NAÏVE BAYES* UNTUK ANALISIS SENTIMEN PELANGGAN TOKOPEDIA

Tri Ernayanti^{1*}, Mustafid², Agus Rusgiyono³, Arief Rachman Hakim³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail: triernayanti99@gmail.com

DOI: 10.14710/j.gauss.11.4.562-571

Article Info:

Received: 2022-09-12

Accepted: 2022-11-28

Available Online: 2023-02-25

Keywords:

E-commerce; Sentiment Analysis; Multinomial Naïve Bayes; Feature Selection Chi-Square.

Abstract: E-commerce is a medium for online shopping that is popular among the public. Ease of access for all internet users and the completeness of products offered by e-commerce are new alternatives in meeting the needs of the community. This causes stiff competition in the e-commerce, so e-commerce need to carry out the right marketing strategy in order to compete in obtaining, retaining, and partnering with customers, one of which is by reviewing aspects of customer satisfaction. Tokopedia is an e-commerce buying and selling online that connects sellers and buyers throughout Indonesia for free. In this study, an analysis of Tokopedia's customer sentiment was carried out with the Multinomial Naïve Bayes classification. Algorithm Multinomial Nave Bayes is a model development of the Nave Bayes. The difference lies in the selection of data, if Naïve Bayes uses a Gaussian that is suitable for continue, while Multinomial Naïve Bayes is suitable for discrete data such as the number of words in a document. Multinomial Naïve Bayes is the simplest method of probability classification but is sensitive to feature selection, so the amount of data is determined by the results of Chi-Square. Multinomial Naïve Bayes is used to classify customer opinions that are positive and negative so that they can form customer satisfaction factors Tokopedia, while the Chi-Square used to measure the level of feature dependence with class (positive and negative) so as to eliminate disturbing features in the classification process. Classification performance results using Multinomial Naïve Bayes without Chi-Square obtained accuracy and kappa statistics of 88% and 75.95%, while using Chi-Square obtained accuracy and kappa statistics of 95% and 89.99%, respectively. This means that Multinomial Naïve Bayes has quite effective performance and results in analyzing Tokopedia customer satisfaction sentiment and the use of Chi-Square for feature selection can improve the accuracy of the classification process.

1. PENDAHULUAN

Peran teknologi mulai membawa peradaban memasuki bidang perdagangan, berupa penjualan dan pembelian barang secara *online* yang biasa disebut sebagai *e-commerce*. Pertumbuhan pasar *e-commerce* tidak terlepas dari perilaku masyarakat yang menginginkan kecepatan dalam berbelanja *online* serta sebagian besarnya sudah mengerti cara menggunakan internet dan *smartphone*. Indonesia telah menjadi salah satu pasar yang menarik perhatian para pengusaha *e-commerce* karena menurut survei yang dilakukan pada tahun 2020, tercatat bahwa pengguna internet di Indonesia telah mencapai 196,7 juta pengguna, angka ini mengalami pertumbuhan sebesar 8,9% dari tahun sebelumnya (APJII, 2020). Lonjakan dari pertumbuhan pengguna internet ini tentunya berdampak terhadap pertumbuhan jumlah pelanggan pada setiap perusahaan *e-commerce*.

Perusahaan *e-commerce* perlu melakukan strategi pemasaran yang tepat agar dapat bersaing dalam memperoleh, mempertahankan, dan bermitra dengan pelanggan. Salah satu strategi pemasaran yang dapat dilakukan adalah pemahaman pelanggan, salah satunya dengan mengkaji aspek kepuasan pelanggan. Secara umum kepuasan pelanggan memberikan dua manfaat utama bagi perusahaan, yaitu berupa loyalitas pelanggan dan penyebaran dari mulut ke mulut atau yang biasa disebut dengan istilah *word of mouth* (Sung, 2014). Terlebih dalam bisnis *e-commerce*, pembeli tidak melihat barang secara langsung sehingga ulasan dari pelanggan lain yang sudah menggunakan produk tersebut berfungsi sebagai acuan pada proses pertimbangan sebelum melakukan pembelian.

Tokopedia merupakan sebuah *e-commerce* jual beli *online* yang menghubungkan antara penjual dan pembeli diseluruh Indonesia dengan biaya gratis. Menurut survey yang dilakukan oleh Similar Web Tokopedia telah menjadi *marketplace* No 1 yang banyak dikunjungi dan diminati oleh pelanggan *e-commerce* dengan total kunjungan mencapai 135,1 juta pada kuartal 1 tahun 2021. Jumlah tersebut mengalami kenaikan sebesar 17,8% dari kuartal sebelumnya.

Respon pelanggan Tokopedia seringkali disampaikan kepada perusahaan melalui media sosial, salah satunya ialah *Google Play*. Penggunaan *Google Play* sebagai sumber data memungkinkan untuk memperoleh data dalam jumlah besar dan akan lebih tepat sasaran, karena data diambil dari orang yang sudah mempunyai pengalaman menggunakan layanan tersebut.

Namun permasalahan yang masih dihadapi saat ini oleh banyak perusahaan *e-commerce* di Indonesia adalah sulitnya memperoleh informasi secara keseluruhan dari semua ulasan pelanggan, karena akan membutuhkan waktu yang lama untuk membaca setiap ulasan. Oleh karenanya, diperlukan suatu metode yang mampu menggambarkan opini pelanggan yakni dengan menggunakan analisis sentimen. Menurut Kamus Besar Bahasa Indonesia (KBBI) pengertian sentimen itu sendiri adalah pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu.

Pengklasifikasian dalam penelitian ini dilakukan dengan Algoritma *Naïve Bayes*. Menurut (Wati, 2016) *Naïve Bayes* merupakan metode paling sederhana dari pengklasifikasian probabilitas, memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada *database* dengan *big data*. Selain itu, model *Naïve Bayes* juga memiliki tingkat kesalahan yang sangat minimum dibandingkan dengan algoritma klasifikasi lainnya (Liu, *et al.*, 2016). Namun, biasanya *Naïve Bayes* diinput dalam bentuk binomial hanya dengan 1 dan 0. Masalah tersebut dapat diatasi dengan menggunakan *Multinomial Naïve Bayes* yang menginput angka desimal dengan tujuan untuk memaksimalkan penggunaan metode *Naïve Bayes*. Algoritma *Multinomial Naïve Bayes* merupakan pengembangan model dari algoritma *Naïve Bayes* yang biasa digunakan dalam klasifikasi teks. Perbedaannya terletak dalam pemilihan data, jika *Naïve Bayes* menggunakan model *gaussian* yang cocok untuk data *continue*, sedangkan *Multinomial Naïve Bayes* cocok digunakan untuk data diskrit seperti jumlah kata dalam dokumen.

Data opini yang digunakan tidak terstruktur dan banyaknya jumlah fitur dapat menimbulkan masalah tingginya dimensionalitas data yang menyebabkan data menyimpang dan dapat mempengaruhi proses klasifikasi sehingga perlu dilakukan seleksi fitur. *Chi-Square* merupakan salah satu metode penyeleksian fitur yang menggunakan teori statistika untuk menguji independensi sebuah *term* dengan kategorinya. Menurut Sun *et al.* (2009) *Chi-Square* adalah salah satu seleksi fitur pada *supervised learning* yang mampu menghilangkan banyak fitur tanpa mengurangi tingkat akurasi.

Penelitian ini bertujuan untuk untuk mengimplementasikan seleksi fitur *Chi-Square* dan algoritma *Multinomial Naïve Bayes* dalam membuat analisis sentimen kepuasan pelanggan *e-commerce* Tokopedia. Algoritma *Multinomial Naïve Bayes* digunakan untuk mengklasifikasikan opini pelanggan yang bersifat positif dan negatif sehingga dapat membentuk faktor-faktor kepuasan pelanggan *e-commerce* Tokopedia, sedangkan seleksi fitur *Chi-Square* digunakan untuk mengukur tingkat ketergantungan *term* dengan kategori sehingga dapat menghilangkan fitur pengganggu pada proses klasifikasi. Penelitian ini menggunakan data ulasan pelanggan *e-commerce* Tokopedia pada situs *Google Play* dalam kurun waktu bulan Januari 2021 – Maret 2021.

2. TINJAUAN PUSTAKA

Kepuasan pelanggan merupakan perasaan senang atau kecewa yang muncul setelah membandingkan kinerja (hasil) produk yang dipikirkan terhadap kinerja yang diharapkan (Kotler, 2014). Menurut Vera Pujani (2011) ada lima faktor yang mempengaruhi kepuasan pelanggan *e-commerce* yaitu *system quality*, *information quality*, *service quality*, *features*, dan *usefulness*.

Analisis sentimen atau *opinion mining* adalah bidang studi yang menganalisis opini yakni sentimen, evaluasi, penilaian, sikap, dan perasaan terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Liu, 2012). Pengolahan data pada analisis sentimen menggunakan *Text mining*. Proses *text mining* dibagi menjadi *text pre-processing*, pembobotan kata dan seleksi fitur. *Text pre-processing* berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur. Tahap *text pre-processing* pada penelitian ini dilakukan beberapa tahapan yaitu:

1. *Case Folding* merupakan proses yang bertujuan untuk mengubah keseluruhan teks menjadi *lower case* agar seragam.
2. *Removing* merupakan proses menghilangkan beberapa kata yang tidak diperlukan untuk mengurangi *noise*
3. Menstandarkan kalimat merupakan proses mengubah kalimat yang tidak baku menjadi baku sehingga dapat dikenali sebagai bahasa yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).
4. Pelabelan data dengan menghitung skor yang diperoleh dari jumlah kata positif dikurangi jumlah kata negatif dalam setiap dokumen. Kata positif dan kata negatif dikumpulkan dalam

kamus berbentuk txt. Kalimat yang memiliki skor > 0 akan diklasifikasikan ke dalam kelas positif dan kalimat yang memiliki skor < 0 diklasifikasikan ke dalam kelas negatif (Hartanto, 2017)

5. *Stopwords removal* merupakan proses menghilangkan kata umum (*common words*) yang muncul dalam jumlah besar dan dianggap tidak memiliki makna.
6. *Stemming* merupakan proses mengubah berbagai kata imbuhan menjadi kata dasar.
7. *Tokenizing* merupakan proses membagi rangkaian kata dalam kalimat menjadi potongan kata tunggal.

Pembobotan kata atau *Term Weighting* merupakan tahapan yang bertujuan untuk memberi bobot pada setiap *term* atau kata. Perhitungan bobot ini memerlukan dua hal, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Menurut Rahman *et al.* (2017), TF merupakan banyaknya jumlah term tertentu yang ada dalam suatu dokumen. Sementara IDF adalah frekuensi kemunculan *term* pada seluruh dokumen. Pada algoritma TF-IDF digunakan rumus untuk menghitung bobot (W) dengan persamaan sebagai berikut (Feldman dan Sanger, 2007):

$$W_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \cdot \log_2 \frac{D}{d_i} \quad (1)$$

$W_{i,j}$ adalah pembobotan TF-IDF untuk *term* ke i pada dokumen ke j , $n_{i,j}$ adalah jumlah kemunculan *term* ke i pada dokumen ke j , $\sum_k n_{k,j}$ adalah jumlah kemunculan seluruh *term* pada dokumen ke j , D adalah banyaknya dokumen yang digunakan, d_i adalah banyaknya dokumen yang mengandung *term* ke i .

Setiap kolom pada kumpulan data disebut dengan fitur. Fitur yang digunakan pada penelitian ini yaitu kata yang terdapat pada sebuah ulasan. Banyaknya jumlah fitur menyebabkan data *overload*, menyimpang dan dapat mempengaruhi proses klasifikasi, sehingga perlu dilakukan seleksi fitur. Seleksi fitur merupakan proses pemilihan fitur yang relevan, dengan kriteria untuk mendapatkan fitur yang optimal, karena pada data dengan dimensi tinggi menemukan subset fitur yang optimal adalah tugas yang cukup sulit (Ling, *et al.*, 2014). *Chi-Square* merupakan salah satu metode untuk penyeleksian fitur yang menggunakan teori statistika untuk menguji independensi sebuah *term* dengan kategorinya.

Hipotesis

H_0 : *Term*/fitur independen terhadap kelas (positif dan negatif)

H_1 : *Term*/fitur dependen terhadap kelas (positif dan negatif)

Taraf signifikansi: α

Statistik uji

Uji *Chi-Square* dapat dirumuskan sebagai berikut (Manning, *et al.*, 2009):

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

O_{ij} adalah nilai sebenarnya pada baris ke- i dan kolom ke- j sedangkan E_{ij} adalah nilai harapan pada baris ke- i dan kolom ke- j .

Tabel kontingensi yang digunakan pada uji *Chi-Square*, dengan ukuran $b \times k$ dapat dilihat pada Tabel 1.

Tabel 1. Tabel Kontingensi

	Kolom ₁	Kolom ₂	...	Kolom _j	Jumlah
Baris ₁	O ₁₁	O ₁₂	...	O _{1j}	b ₁
Baris ₂	O ₂₁	O ₂₂	...	O _{2j}	b ₂
...
Baris _i	O _{i1}	O _{i2}	...	O _{ij}	b _i
Jumlah	k ₁	k ₂	...	k _j	N

O_{ij} merupakan *actual count* atau nilai sebenarnya dari dua variabel yang diamati, pada kasus ini yaitu fitur sebagai baris, dan kelas (negatif dan positif) sebagai kolom. Proses selanjutnya dalam uji *Chi-Square* adalah mencari *expected count*. Tabel *expected count* ditunjukkan oleh tabel berikut:

Tabel 2. Tabel Kontingensi *Expected Count*

	Kolom ₁	Kolom ₂	...	Kolom _j
Baris ₁	E ₁₁	E ₁₂	...	E _{1j}
...
Baris _i	E _{i1}	E _{i2}	...	E _{ij}

Mencari *expected count* dari O_{ij} atau mengisi sel E_{ij} dilakukan dengan perhitungan berikut:

$$E_{ij} = \frac{b_i k_j}{N} \quad (3)$$

dengan:

b_i : Hasil penjumlahan baris ke i

k_j : Hasil penjumlahan kolom ke j

N : Total seluruh nilai pengamatan

Kaidah pengambilan keputusan

Tolak H_0 apabila nilai $\chi^2 \geq \chi^2_{\alpha, (n-1)(m-1)}$ atau nilai signifikansi $\leq \alpha$

Algoritma *Naïve Bayes Classifier* merupakan salah satu teknik dalam *data mining*, dimana prinsip probabilitas (kemungkinan) digunakan untuk pemecahan masalah. Algoritma *Naïve Bayes* yang cocok digunakan dalam pengklasifikasian teks atau dokumen adalah *Multinomial Naïve Bayes*. Kategori dokumen pada *Multinomial Naïve Bayes* tidak hanya ditentukan berdasarkan kata yang muncul saja tetapi juga ditentukan berdasarkan jumlah kemunculannya (Witten, *et al.*, 2011). Model dari multinomial memperhitungkan frekuensi dari setiap kata yang muncul pada dokumen di mana perhitungan hasil frekuensi ini dapat membantu klasifikasi (McCallum dan Nigam, 1998). Klasifikasi dokumen bertujuan untuk menentukan kelas terbaik untuk suatu dokumen. Kelas terbaik dalam klasifikasi *Naïve Bayes* ditentukan dengan mencari *Maximum a Posteriori* (MAP) kelas C_{map} melalui persamaan:

$$C_{\text{map}} = \arg \max_{c \in C} \hat{P}(c) \prod_{k=1}^n \hat{P}(t_k | c) \quad (4)$$

$\hat{P}(c)$ dan $\hat{P}(t|c)$ dihitung pada saat *training*. Persamaannya dapat dituliskan sebagai berikut:

$$\hat{P}(c) = \frac{N_c}{N'} \quad (5)$$

$\hat{P}(c)$ adalah probabilitas prior suatu dokumen berada di kelas c , N_c adalah jumlah dokumen dengan kelas c , N adalah jumlah seluruh dokumen

$$\hat{P}(t|c) = \frac{w_{ct} + 1}{(\sum_{w' \in V} w_{ct'}) + B'} \quad (6)$$

$\hat{P}(t|c)$ adalah probabilitas frekuensi relatif *term* t dalam dokumen berada di kelas c , B' adalah jumlah IDF seluruh *term* pada *vocabulary*, w_{ct} adalah bobot TF-IDF *term* t pada dokumen dengan kelas c , $\sum_{w' \in V} w_{ct'}$ adalah jumlah bobot TF-IDF seluruh *term* pada kelas c .

Sistem dalam klasifikasi diharapkan mampu melakukan klasifikasi semua set data dengan benar, namun tidak dapat dipungkiri bahwa kesalahan akan terjadi dalam proses pengklasifikasian sehingga perlu dilakukan pengukuran kinerja dari sistem klasifikasi tersebut. Pengukuran kinerja klasifikasi, umumnya dilakukan dengan matriks konfusi (*confusion matrix*). Tabel *confusion matrix* dapat dilihat pada Tabel 3.

Tabel 3. *Confusion Matrix*

		Kelas Prediksi	
		Positif	Negatif
Kelas Aktual	Positif	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
	Negatif	<i>False Positive</i> (TN)	<i>True Negative</i> (TN)

Ukuran yang dapat digunakan untuk menilai atau mengevaluasi model klasifikasi yaitu akurasi dan *kappa statistic*. Akurasi merupakan nilai proporsi dari jumlah tiap kelas yang diklasifikasikan dengan benar terhadap total semua data (Han, Kamber, & Pei, 2012). Perhitungan *kappa statistic* didasarkan pada perbedaan antara berapa banyak kesepakatan sebenarnya dibandingkan dengan seberapa banyak kesepakatan yang diharapkan. Nilai *kappa statistic* yang semakin tinggi dianggap baik. Keduanya dapat dihitung dengan persamaan:

$$Accuracy = Po = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$P_c = \left[\left(\frac{TP + FP}{Total} \right) \left(\frac{TP + FN}{Total} \right) \right] + \left[\left(\frac{FN + TN}{Total} \right) \left(\frac{FP + TN}{Total} \right) \right] \quad (8)$$

$$K = \frac{P_o - P_c}{(1 - P_c)} \quad (9)$$

K adalah koefisien *kappa*, P_o adalah proporsi kesepakatan teramati, P_c adalah proporsi kesepakatan harapan dan 1 adalah konstanta.

Word cloud merupakan salah satu metode untuk menampilkan data teks secara visual. Dengan menggunakan *word cloud*, gambaran frekuensi kata-kata dapat ditampilkan dalam bentuk yang menarik namun tetap informatif. Ukuran huruf dalam *word cloud* menentukan frekuensi kemunculan sebuah kata, semakin besar ukuran huruf semakin besar pula kemunculan kata tersebut. Sebaliknya, semakin kecil ukuran huruf maka semakin kecil frekuensi kemunculan kata tersebut (Miley dan Read, 2011).

3. METODE PENELITIAN

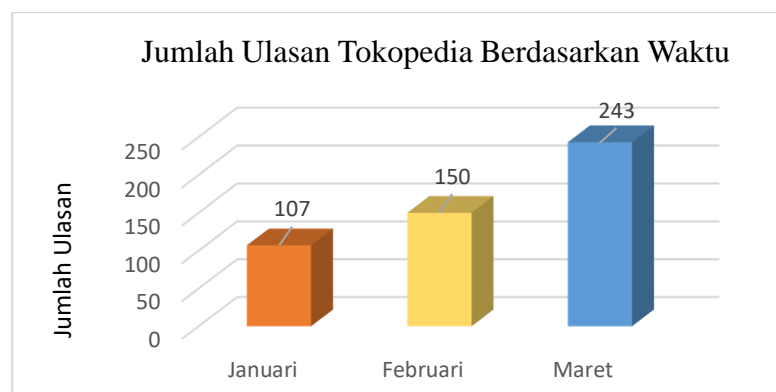
Penelitian ini menggunakan data ulasan pelanggan *e-commerce* Tokopedia. Data ini berupa data sekunder yang diperoleh langsung dari ulasan aplikasi Tokopedia pada *Google Play* sebanyak 500 ulasan. Data yang digunakan diambil dari bulan Januari 2021 sampai Maret 2021. Variabel data yang digunakan pada penelitian ini terdiri dari, tanggal yang digunakan dalam membuat ulasan, nama akun *Google*, dan ulasan pelanggan.

Penelitian ini menggunakan *software Rstudio 1.3.1073*, *Data Miner* dan *Microsoft Excel 2013*. Tahapan analisis yang dilakukan dalam penelitian ini diuraikan sebagai berikut:

1. Mengambil 500 data ulasan pelanggan Tokopedia pada *Google Play* dengan teknik *web scraping* menggunakan *software Data Miner*.
2. Melakukan *Pre-Processing* data dengan tahapan sebagai berikut: *case folding*, *remove number*, *remove punctuation*, *remove emoticon*, *strip white space*, dan menstandarkan kalimat.
3. Pelabelan data dengan *sentiment scoring*.
4. *Stopwords removal* yakni menghilangkan kata-kata yang tidak berhubungan dengan subjek utama.
5. *Stemming* yakni mengubah kata imbuhan menjadi kata dasar.
6. *Tokenizing* yakni membagi rangkaian kata dalam kalimat menjadi potongan kata tunggal.
7. Pembobotan kata dengan TF-IDF
8. Menyeleksi fitur dengan uji *Chi-Square*.
9. Membangun model klasifikasi *Multinomial Naïve Bayes* dengan perbandingan data *training* dan data *testing* sebesar 80%:20%.
10. Interpretasi hasil ke dalam bentuk visual menggunakan *Word Cloud*.

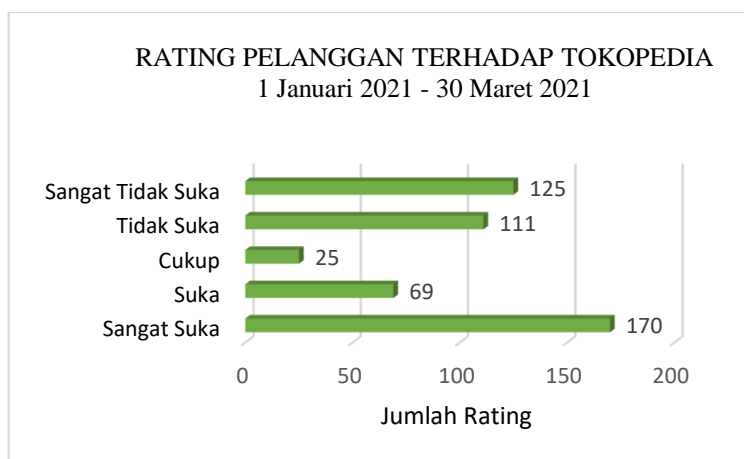
4. HASIL DAN PEMBAHASAN

Data ulasan yang akan dianalisis dilihat dari beberapa aspek diantaranya jumlah ulasan yang masuk berdasarkan urutan waktu dan *rating* aplikasi yang diberikan oleh pelanggan.



Gambar 1. Histogram Ulasan Tokopedia Berdasarkan Waktu.

Histogram pada Gambar 1 memperlihatkan bahwa setiap bulannya aplikasi Tokopedia mengalami penambahan jumlah ulasan dari pelanggan. Bulan Januari sebanyak 107 ulasan, bulan Februari sebanyak 150 ulasan dan bulan Maret sebanyak 243 ulasan. Selain ulasan, pelanggan memberikan *rating* yang direpresentasikan oleh gambar bintang untuk menggambarkan kepuasan mereka terhadap aplikasi Tokopedia.



Gambar 2. Bar Chart Rating Pelanggan Aplikasi Tokopedia

Bar Chart pada Gambar 2 memperlihatkan bahwa dari 500 ulasan terdapat sebanyak 170 pelanggan memberikan penilaian “Sangat Suka”, 69 pelanggan memberikan penilaian “Suka” dan 25 pelanggan memberikan penilaian “Cukup”, sedangkan untuk penilaian “Tidak Suka” berjumlah 111 ulasan dan 125 ulasan dengan kategori “Sangat Tidak Suka”. *Rating* yang diberikan oleh pelanggan Tokopedia terkadang tidak sesuai dengan isi ulasan yang mereka berikan, sehingga perlu dilakukan analisis terhadap isi ulasan

Data ulasan dalam penelitian ini belum sepenuhnya siap digunakan untuk proses klasifikasi. Hal ini disebabkan oleh data yang tidak terstruktur seperti terdapat angka, banyak simbol, bahasa yang tidak baku, serta singkatan yang bermacam-macam, sehingga perlu dilakukan *text pre-processing*. Tahapan *text pre-processing* yang dilakukan diantaranya sebagai berikut:

1. Tahap *case folding* dilakukan untuk menyamakan arti suatu kata yang ditulis dengan huruf kapital atau huruf kecil agar tidak terdeteksi berbeda saat tahap pembobotan kata.
2. Tahap *removing* dilakukan untuk menghapus karakter-karakter yang tidak dibutuhkan seperti angka, tanda baca, *emoticon*, dan spasi yang berlebih pada dokumen.
3. Menstandarkan kalimat perlu dilakukan untuk mengubah kata-kata yang tidak sesuai dengan KBBI seperti kesalahan ketik, singkatan maupun bahasa lokal. Proses tersebut memerlukan kamus yang dibuat secara manual sebanyak 600 kata.
4. Pelabelan data menggunakan perhitungan skor sentimen, mendapatkan hasil bahwa kelas sentimen positif lebih dominan dibandingkan kelas sentimen negatif. Jumlah ulasan yang masuk ke dalam kelas positif sebanyak 265 ulasan dan ulasan yang masuk ke dalam kelas negatif sebanyak 235 ulasan. Pelabelan menggunakan perhitungan skor sentimen masih menimbulkan kesalahan sebanyak 10 ulasan atau 2% dari 500 ulasan. Kesalahan tersebut disebabkan karena skor akhir dari masing-masing ulasan yang dihitung oleh program komputasi berbeda dengan penilaian yang diputuskan oleh manusia. Hal ini memang seringkali terjadi karena sentimen merupakan hal yang bersifat subjektif. Kesalahan dalam pelabelan tersebut diganti dengan pelabelan manual untuk mendapatkan label yang benar pada seluruh ulasan sehingga dapat memberikan performa yang optimal. Setelah dilakukan perbaikan, jumlah ulasan positif menjadi sebanyak 255 dan ulasan negatif sebanyak 245.
5. Tahap *stopwords removal* dapat mengurangi dimensi data sehingga tersisa dimensi data yang diperlukan. *Stopwords removal* memerlukan daftar kata yang akan dihapus, yaitu *stoplist* yang memuat 430 kata.
6. Tahap *stemming* menggunakan *package* *katadasaR* pada Rstudio untuk mengubah kata imbuhan menjadi kata dasar. Tahap *tokenizing* memisahkan dokumen menjadi potongan kata dengan acuan pemisah berupa spasi.

Data yang telah melalui proses *text pre-processing*, kemudian diberi bobot. Pembobotan kata pada penelitian ini menggunakan perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF). Dalam proses kali ini, setiap kata yang terdapat pada dokumen tersebut memiliki bobot tersendiri. Hasil pembobotan kata menggunakan TF-IDF ditunjukkan pada tabel 4.

Tabel 4. Hasil Pembobotan TF-IDF

Ulasan ke-	Akun	Aman	Belanja	Kecewa	Tokopedia	...	Voucher
28	0.48917	0	0	0.13379	0.02136	...	0
76	0	0.85307	0.13658	0	0	...	0
151	0.39858	0	0	0.10902	0.01740	...	0.20903
166	0	0	0	0.16352	0	...	0.31355

330	0.26904	0	0.11609	0	0.02350	...	0
393	0	0.27588	0.12899	0	0.05221	...	0
473	0	0	0	0	0.04699	...	0.72032
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
500	0	0	0	0.21024	0	...	0

Perhitungan diatas menghasilkan bobot fitur “aman” pada ulasan ke-76 sebesar 0.85307 nilai bobot kata tersebut akan digunakan dalam proses klasifikasi.

Proses selanjutnya yaitu seleksi fitur *Chi-Square* dengan tujuan untuk memilih fitur yang relevan dengan menghilangkan fitur *noise*. Perhitungan nilai *Chi-Square* dilakukan dengan program R. Nilai *Chi-Square* dan signifikansi setiap fitur terdapat pada Tabel 5.

Tabel 5. Keputusan Uji *Independensi Chi-Square*

No	Fitur	<i>Pearson Chi-Square</i>	Signifikansi	Keputusan
1	Akses	1,396	0,319	Gagal tolak H0 karena $0,319 > 0,05$
2	Akun	40,701	9×10^{-4}	H0 ditolak karena $9 \times 10^{-4} < 0,05$
3	Aman	2,301	0,210	Gagal tolak H0 karena $0,210 > 0,05$
4	Belanja	5,938	0,018	H0 ditolak karena $0,018 < 0,05$
5	Buruk	18,477	0,001	H0 ditolak karena $0,001 < 0,05$
⋮	⋮	⋮	⋮	⋮
769	Voucher	1,144	0,342	Gagal tolak H0 karena $0,342 > 0,05$

Tabel 5 menunjukkan bahwa fitur “Akses” menghasilkan nilai *Chi-Square* sebesar 1,396 dan signifikansi sebesar $0,319 \geq 0,05$ sehingga gagal tolak H0 yang berarti bahwa fitur “Akses” tidak memiliki ketergantungan kuat dengan kelas (positif dan negatif). Sedangkan fitur “Akun” menghasilkan nilai *Chi-Square* sebesar 40,701 dan signifikansi sebesar $9 \times 10^{-4} \leq 0,05$ sehingga H0 ditolak yang berarti bahwa fitur “Akun” memiliki ketergantungan kuat dengan kelas (positif dan negatif). Seleksi fitur dengan menggunakan *Chi-Square* mengakibatkan pengurangan jumlah fitur yang diperoleh. Pada taraf signifikansi 0.05, jumlah fitur yang diperoleh yaitu 160 dari jumlah fitur awal yaitu 769, berkurang sebanyak 79%.

Data yang telah melalui proses pembobotan kata, kemudian dibagi menjadi data *training* dan data *testing*. Perbandingan antara data *training* dan data *testing* adalah 80% : 20% sehingga dapat diartikan bahwa dari 500 ulasan, sebanyak 400 ulasan menjadi data *training* dan sebanyak 100 ulasan menjadi data *testing*. Pada penelitian ini digunakan data dari hasil pelabelan secara manual.

Tabel 6. Proporsi Data *Training* dan Data *Testing*

Klasifikasi	Data <i>Training</i>	Data <i>Testing</i>	Total
Positif	204	51	255
Negatif	196	49	245
Total	400	100	500

Tabel 6 menunjukkan bahwa data yang bersentimen positif sebanyak 255 data yang terdiri dari 204 data *training* dan 51 data *testing*. Sedangkan data yang bersentimen negatif sebanyak 245 data yang terdiri dari 196 data *training* dan 49 data *testing*.

Proses klasifikasi pada penelitian ini menggunakan algoritma *Multinomial Naïve Bayes*. Proses ini akan mengklasifikasikan data *testing* setelah data *training* sudah berhasil diberikan label sentimen. Pengukuran kinerja sistem klasifikasi dilakukan dengan memperhatikan matriks konfusi (*confusion matrix*). Berikut merupakan hasil dari klasifikasi *Multinomial Naïve Bayes* dengan proses seleksi fitur dan tanpa seleksi fitur.

1. Hasil Klasifikasi tanpa Seleksi Fitur

Seluruh fitur yang terdapat pada data *training* digunakan pada proses klasifikasi pertama. Jumlah fitur yang digunakan sebanyak 769 fitur. Berikut evaluasi performansi yang digambarkan melalui *confusion matrix* yang disajikan pada Tabel 7.

Tabel 7. *Confusion Matrix* tanpa Seleksi Fitur

Kelas Prediksi	Kelas Aktual	
	Negatif	Positif
Negatif	47	8
Positif	4	41

Berikut perhitungan nilai akurasi dan *kappa statistic* secara manual dari Tabel 7 berdasarkan persamaan 7, 8 dan 9.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$= \frac{41 + 47}{41 + 4 + 47 + 8} = \frac{88}{100} = 0,88 = 88\%$$

$$Po = Accuracy = 0,88$$

$$Pc = \left[\left(\frac{TP + FP}{Total} \right) \left(\frac{TP + FN}{Total} \right) \right] + \left[\left(\frac{FN + TN}{Total} \right) \left(\frac{FP + TN}{Total} \right) \right]$$

$$= \left[\left(\frac{41 + 4}{100} \right) \left(\frac{41 + 8}{100} \right) \right] + \left[\left(\frac{8 + 47}{100} \right) \left(\frac{4 + 47}{100} \right) \right]$$

$$= 0,2205 + 0,2805 = 0,501$$

$$K = \frac{Po - Pc}{(1 - Pc)} = \frac{0,88 - 0,501}{(1 - 0,501)} = 0,7595 = 75,95\%$$

2. Hasil Klasifikasi dengan Seleksi Fitur *Chi-Square*

Proses klasifikasi kedua dilakukan dengan pemilihan fitur menggunakan metode *Chi-Square*. Jumlah fitur yang digunakan sebanyak 160 fitur yang memiliki ketergantungan kuat dengan kategorinya. Berikut evaluasi performansi yang digambarkan melalui *confusion matrix* pada Tabel 8.

Tabel 8. *Confusion Matrix* dengan Seleksi Fitur *Chi-Square*

Kelas Prediksi	Kelas Aktual	
	Negatif	Positif
Negatif	49	3
Positif	2	46

Berikut perhitungan nilai akurasi dan *kappa statistic* secara manual dari Tabel 21 berdasarkan persamaan 7, 8 dan 9.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$= \frac{46 + 49}{46 + 2 + 49 + 3} = \frac{95}{100} = 0,95 = 95\%$$

$$Po = Accuracy = 0,95$$

$$Pc = \left[\left(\frac{TP + FP}{Total} \right) \left(\frac{TP + FN}{Total} \right) \right] + \left[\left(\frac{FN + TN}{Total} \right) \left(\frac{FP + TN}{Total} \right) \right]$$

$$= \left[\left(\frac{46 + 2}{100} \right) \left(\frac{46 + 3}{100} \right) \right] + \left[\left(\frac{3 + 49}{100} \right) \left(\frac{2 + 49}{100} \right) \right]$$

$$= 0,2352 + 0,2652 = 0,5004$$

memiliki memiliki performa dan hasil yang cukup efektif dalam melakukan analisis sentimen kepuasan pelanggan dengan menggunakan data ulasan aplikasi Tokopedia dan penggunaan *Chi-Square* untuk pemilihan fitur dapat meningkatkan hasil akurasi pada proses klasifikasi.

DAFTAR PUSTAKA

- APJII. (2020). *Laporan Survei Penetrasi dan Profil Perilaku Pengguna Internet Indonesia Tahun 2020*. Jakarta: Asosiasi Penyelenggara Jasa Internet Indonesia
- Feldman, R., dan Sanger, J. (2007). *The Text Mining Handbook: Advanced approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Han, J., dan Kamber, M. (2006). *Data Mining: Concept and Techniques* (2nd ed.). San Fransisco: Morgan Kaufmann Publisher.
- Hartanto. (2017). Text Mining dan Sentimen Analisis Twitter pada Gerakan LGBT. *Jurnal Psikologi Ilmiah*, IX(1), 18-25.
- Kotler dan Keller. (2014). *Manajemen Pemasaran Edisi 13 Jilid 1*. Jakarta: Erlangga.
- Ling, J., Kencana, I. P., dan Oka, T. B. (2014). *Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square*. E-Jurnal Matematika, Vol. 03.
- Liu, B. (2012). *Sentiment Analysis and Opini Mining*. San Rafael: Morgan dan Claypool Publishers.
- Liu, J., Tian, Z., Liu, P., Jiang, J. & Li, Z. (2016). *An Approach of Semantic Web Service Classification Based on Naive Bayes*. IEEE International Conference on Services Computing, Hal: 356-362.
- McCallum, A., dan Nigam, K. (1998). *A Comparison of Event Models for Naive Bayes Text Classification*. Pittsburgh: Proceedings in Workshop on Learning for Text Categorization.
- Sun, C, X. Wang, and J. Xu. (2009) "*Study on Feature Selection in Finance Text Categorization*". Science And Technology, pp. 5077-5082.
- Sung, J., dan Sung, E. (2014). *Exploring the Usefulness of a Decision Tree in Predicting People Locations*. Procedia - Social and Behavioral Sciences, 140, 447–451
- Taheri, S., dan Mammadov, M. (2013). Learning the Naive Bayes Classifier with Optimization Models. *International Journal of Applied Mathematics and Computer Science*, 787-795.
- Wati, R. (2016). *Penerapan Algoritma Genetika untuk Seleksi Fitur pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes*. Evolusi Vol. 4, No. 1, Hal: 25-31.
- Witten, L., Frank, E., dan Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). USA: Elsevier.