

ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR

Josua Josen A. Limbong¹, Irwan Sembiring^{*2}, Kristoko Dwi Hartomo³

^{1,2,3}Universitas Kristen Satya Wacana, Salatiga

Email: ¹josualimbong21@gmail.com, ²irwan@uksw.edu, ³kristoko@uksw.edu

^{*}Penulis Korespondensi

(Naskah masuk : 23 April 2021, diterima untuk diterbitkan: 14 Oktober 2021)

Abstrak

Saat ini internet memungkinkan pengguna untuk membuat ulasan secara *online* diberbagai jenis platform. Salah satunya aplikasi *e-commerce Shopee* pada *website google play store* dimana kelas sentimen positif dan negatif yang terdapat pada ulasan *online* jelas mencerminkan persepsi pengguna tentang berbagai jenis layanan dan produk yang ada. Selain itu, pelanggan berpotensi yang membaca ulasan *online* dapat secara signifikan terpengaruh oleh sentimen dari ulasan yang tertera pada kolom ulasan. Hal ini menandakan ulasan yang bersentimen positif ataupun negatif yang ditinggalkan oleh pengguna sangat mempengaruhi pengguna lainnya dalam memilih layanan maupun produk yang dicari. Oleh karena itu perlunya analisis sentimen untuk mengklasifikasi dataset yang begitu banyak sehingga dapat dengan mudah mengetahui apa saja sentimen pelanggan. penelitian ini menggunakan data ulasan sebanyak 500 ulasan. Kemudian ulasan tersebut diklasifikasi menggunakan aplikasi orange dengan metode *Naive Bayes* dan *K-Nearest Neighbor* (KNN). Kemudian selanjutnya menggunakan metode *word cloud* untuk mengetahui topik-topik yang sering diulas oleh pelanggan. Hasilnya setelah menggunakan metode *Naive Bayes* memperoleh hasil nilai *accuracy* 0,914, *precision* 0,915, *recall* 0,914 dan *F1 score* 0,916. Sedangkan metode *KNN* memperoleh nilai *accuracy* 0,928, *precision* 0,929, *recall* 0,928, dan *F1 score* 0,926. Hal ini membuktikan bahwa dalam penelitian ini kinerja metode *KNN* lebih baik. Kemudian berdasarkan hasil *word cloud* yang diperoleh didapatkan informasi kata dengan sentimen positif yang paling sering diulas oleh pelanggan diantaranya terkait kata: gratis, bagus, suka, murah, mudah, dan cepat. Sedangkan informasi sentimen negatif yang diperoleh seperti kata : kecewa, jelek, mahal, bohong, ribet, dan perbaiki.

Kata kunci: *analisis sentimen, e-commerce, word cloud, naive bayes, k-nearest neighbor*

ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS

Abstract

Today the internet allows users to create online reviews on various types of platforms. One of them is the *Shopee e-commerce application on the google play store website*, where the positive and negative sentiment classes contained in online reviews reflect user perceptions about the various types of services and products available. Also besides, potential customers who read online reviews can be significantly affected by the sentiment of the reviews listed in the review column. This indicates that positive or negative reviews left by users greatly influence other users in choosing the services or products they are looking for. Therefore the need for sentiment analysis to classify such a large dataset so that you can easily find out what customer sentiments are. This study uses a dataset of 500 reviews. Then the reviews are classified using the orange application with the *Naive Bayes* and *K-Nearest Neighbor* (KNN) methods. Then use the *word cloud* method to find out topics that are frequently reviewed by customers. The results, after using the *Naive Bayes* method, get the *accuracy* value of 0.914, *precision* 0.915, *recall* 0.914, and *F1 score* 0.916. Meanwhile, the *KNN* method obtained an *accuracy* value of 0.928, *precision* 0.929, *recall* 0.928, and *F1 score* 0.926. This proves that in this study the performance of the *KNN* method is better. Then based on the *word cloud* results obtained word information with positive sentiments that are most

often shared by customers related to words: free, good, like, cheap, easy, and fast. Meanwhile, the negative sentiment information obtained includes the words: disappointed, ugly, expensive, lying, complicated, and fix.

Keywords: *sentiment analysis, e-commerce, word cloud, naive bayes, k-nearest neighbor.*

1. PENDAHULUAN

Di era modern saat ini dengan persaingan bisnis yang semakin ketat pelaku bisnis harus memanfaatkan teknologi dalam memasarkan barang dagangannya, agar terus mengikuti perkembangan teknologi untuk kemajuan usahanya (Masripah and Utami 2020). Saat ini internet memungkinkan pengguna untuk membuat ulasan secara online di berbagai jenis *platform*. Selain berbagi pengalaman pribadi, pengguna juga menunjukkan emosi atau perasaan mereka dalam *user-generated content* (UGC), seperti ulasan pengguna (Wu and Li 2020). Kelas sentimen positif dan negatif yang terdapat pada ulasan *online* jelas mencerminkan persepsi pengguna tentang berbagai jenis layanan yang ada. Selain itu, pelanggan berpotensi yang membaca ulasan *online* dapat secara signifikan terpengaruh oleh polaritas sentimen dari ulasan yang tertera pada kolom ulasan. Hal ini menandakan ulasan yang bersentimen positif ataupun negatif yang ditinggalkan oleh pengguna, sangat mempengaruhi pengguna lainnya dalam memilih layanan ataupun produk yang dicari (Artanti et al. 2018).

Internet kini telah menjadi media andalan dalam berkomunikasi hingga berbisnis. Internet juga sangat membantu dan memudahkan dalam proses bisnis, sehingga saat ini bisnis *online* lebih disukai oleh masyarakat. Salah satu *marketplace* terkenal di Indonesia saat ini, yaitu Shopee. Hal ini didukung oleh data pada situs *iprice* yang menyatakan Shopee menjadi *e-commerce* dengan pengunjung situs bulanan terbesar di Indonesia. Data *iprice* menunjukkan, Shopee mendapatkan sebanyak 71,5 juta kunjungan selama kuartal I tahun 2020. Sedangkan Tokopedia dan Bukalapak menempati urutan selanjutnya, masing-masing sebanyak 69,8 juta dan 37,6 juta kunjungan (Dwi 2020). Berdasarkan data pendukung inilah peneliti tertarik menggunakan *marketplace* Shopee pada *google play store* yang dijadikan sebagai studi kasus penelitian.

Saat ini *marketplace* menjadi pilihan yang paling digemari masyarakat untuk berbelanja suatu produk karena memiliki banyak kelebihan atau kemudahan, antara lain : pembeli tidak perlu datang langsung ke tempat penjual, toko dapat diakses 24 jam, banyak diskon, promo, dan bahkan bebas ongkos kirim. Namun karena kelebihan itulah sehingga sering muncul beberapa permasalahan, seperti : sering terjadi penipuan pada saat melakukan transaksi, penjual tidak mengirimkan barang yang sudah dipesan, barang yang ada digambar tidak sesuai dengan aslinya, kemasan rusak serta juga terjadi kesalahan warna. Hal inilah yang kemudian mendorong banyak pelanggan

menggunakan media sosial untuk melampiasakan perasaannya terhadap kualitas layanan yang diberikan oleh suatu perusahaan *online marketplace*, dengan menuliskan komentarnya pada situs online tersebut. Hal ini dapat mempengaruhi kepercayaan masyarakat atau pelanggan untuk berbelanja secara *online* (Muktafin, Kusriani, and Luthfi 2020).

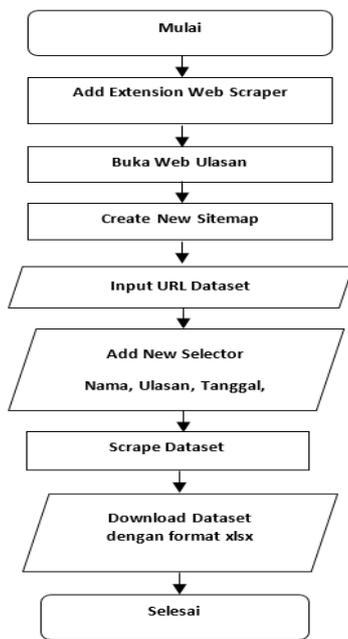
Berdasarkan ulasan-ulasan yang diberikan oleh pengguna aplikasi Shopee diharapkan adanya umpan balik bagi pihak developer untuk meningkatkan kualitas layanan. Akan tetapi, dengan begitu banyaknya data ulasan yang masuk, sehingga sistem aplikasi *e-commerce* belum mampu mengetahui hal-hal apa saja yang perlu segera dilakukan perbaikan. Belum ada metode yang sistematis dan akurat yang dapat mengklasifikasikan sebuah ulasan menjadi kelompok positif dan negatif. Dengan adanya klasifikasi data ulasan pengguna berdasarkan jenis sentimen, akan mempermudah pihak *developer* dalam mendapatkan informasi sentimen dari para pengguna aplikasi *e-commerce*.

Penelitian ini bertujuan untuk mengetahui sentimen pengguna aplikasi *e-commerce Shopee*, yang akan menghasilkan rekomendasi bagi pihak developer agar kedepannya dapat memperbaiki kualitas fitur, pelayanan, produk, serta berbagai faktor yang dapat mempengaruhi kepuasan pelanggan. Proses pengklasifikasian dalam penelitian ini menggunakan *machine learning* dengan metode algoritma *Naive Bayes* dan *K-Nearest Neighbor* untuk mengetahui hasil klasifikasi sentimen, dan juga untuk mengetahui perbandingan *accuracy*, *precision*, *recall*, dan *F1 score*. Selanjutnya untuk mengetahui faktor-faktor apa saja yang mempengaruhi kepuasan pelanggan digunakan metode *word cloud*, yaitu berdasarkan jumlah kata-kata yang sering ditulis pelanggan *e-commerce* Shopee. Hasil penelitian pada aplikasi *e-commerce* Shopee ini diharapkan dapat memberikan hasil klasifikasi teks yang baik dan akurat agar dapat menjadi sistem rekomendasi dan informasi yang bermanfaat bagi pihak yang membutuhkan.

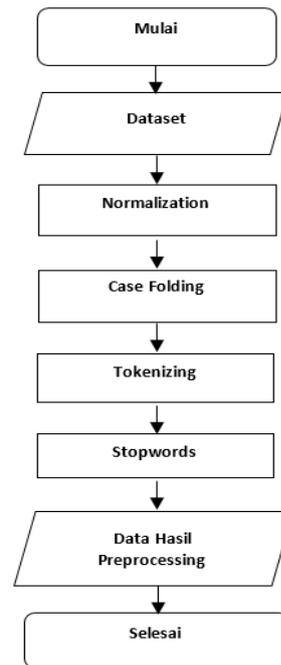
2. METODE PENELITIAN

2.1 Pengumpulan Data

Penelitian ini menggunakan dataset sebanyak 500 dataset yang didapatkan dari aplikasi *e-commerce* Shopee pada situs *website google playstore*. Pengambilan dataset dilakukan dengan menggunakan teknik *web scraping* dari *reviewer* pengguna aplikasi *e-commerce* Shopee. Adapun tahapan *scraping* dataset menggunakan *tools web scraping* sebagaimana ditunjukkan pada Gambar 1 berikut.



Gambar 1. Tahapan Scraping Dataset



Gambar 2. Tahapan Preprocessing Data

2.2 Preprocessing Data

Pada proses ini data yang telah didapatkan akan melewati proses *preprocessing* terlebih dahulu agar mendapatkan bentuk data yang siap diproses oleh data *mining*.

2.3 Klasifikasi

Proses klasifikasi dalam penelitian ini menggunakan aplikasi orange dan metode *Naive Bayes* dan *K-Nearest Neighbor (KNN)*. *Naive Bayes* merupakan salah satu algoritma pembelajaran mesin probabilistik berdasarkan teorema bayes, yang biasa digunakan dalam berbagai macam tugas klasifikasi (Ratnawati 2018).

Bentuk umum teorema bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)}$$

Keterangan:

X = Data dengan kelas yang belum diketahui.

H = Hipotesa data X merupakan suatu kelas spesifik.

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posterior probability)

P(H) = Probabilitas hipotesis H (prior probability).

Sedangkan metode *K-Nearest Neighbor (KNN)* adalah metode yang digunakan untuk klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Septian, Fahrudin, and Nugroho 2019). Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan *training sample*.

Untuk dapat menghitung jarak antara dua titik x dan y bisa digunakan jarak Euclidean, sebagai berikut:

$$d(X_1, Y_1) = \sum \frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2}$$

yang mana X_1 , $1 = 1, 2$, adalah atribut kategori, dan n_{1j} , n_1 mewakili frekuensi yang sesuai.

2.4 Word Cloud

Word cloud adalah metode *text mining* yang menampilkan grafis dari frekuensi kata yang lebih menonjolkan kata-kata yang lebih sering muncul dalam teks sumber. Semakin besar ukuran kata dalam visual, maka semakin umum kata tersebut dalam dokumen (Alamsyah and Zuhri 2017).

2.5 Validasi dan Evaluasi

Proses validasi pada penelitian ini menggunakan teknik *cross validation* sebanyak 10 kali lipat untuk memproses data *training* dan data *testing*. Kemudian dalam proses evaluasi menggunakan *confusion matrix* untuk mengukur dan melihat nilai *accuracy*, *precision*, *recall*, dan *calibration plot* untuk mengetahui hasil *F1 Score*

3. LITERATURE REVIEW

Beberapa hasil penelitian terdahulu yang penulis jadikan sebagai acuan dalam pemilihan metode penelitian ini adalah sebagai berikut.

Dalam penelitian yang dilakukan oleh (Muktafin, Kusri, and Luthfi 2020). Penelitian ini dilakukan untuk mengklasifikasikan ulasan produk “hijab instan” ke dalam kelas positif dan negatif menggunakan algoritma KNN dan TF-IDF dengan pendekatan NLP. Hasilnya, klasifikasi menggunakan

pendekatan NLP mendapat akurasi sebesar 76,92%, presisi 80,00%, dan recall 74,07%, sedangkan tanpa NLP hanya mendapat akurasi sebesar 69,23%, presisi 80,00%, dan recall 64,52%.

(Vijayaragavan, Ponnusamy, and Aramudhan 2020), melakukan analisis sentimen pada ulasan produk menggunakan metode *Support Vector Machine* (SVM), dengan dataset berupa ulasan mengenai iPod sebanyak 1.811 ulasan. Pada penelitian ini metode *Support Vector Machine* (SVM) mendapatkan akurasi sebesar 96.57% dibandingkan dengan metode yang lain seperti RBF Network yang hanya mendapatkan 88.84%, RF mendapatkan 88.34%, DT mendapatkan 86,01%, dan MLP mendapatkan 85.31%.

Penelitian dengan judul “Algoritma Klasifikasi Naive Bayes untuk Analisa Sentimen Aplikasi Shopee” diteliti oleh (Masripah and Utami 2020). Hasil penelitian ini mendapatkan hasil akurasi sebesar 71.50% dan nilai AUC (Area Under Curve) sebesar 0,500, dan algoritma Naive Bayes dapat digunakan untuk melakukan analisa sentiment pada aplikasi Shopee.

Penelitian lainnya yang dilakukan oleh (Pulakiang 2019) melakukan Analisis Kualitas Layanan E-Commerce Menggunakan Twitter Api pada Tokopedia, Lazada, dan Bukalapak. Dari ketiga *e-commerce* yang diteliti Hasil penelitian ini menunjukkan bahwa *e-commerce* yang paling banyak mendapat komentar negatif atau evaluasi dari pelanggan ada pada Lazada, sedangkan pada Tokopedia dan Bukalapak hasil analisis menunjukan kesamaan presentase. Artinya Tokopedia dan Bukalapak mempunyai tingkat evaluasi dan komentar negatif yang sama.

(Azzahra and Wibowo 2020) melakukan penelitian mengenai analisis sentimen ulasan wisata kuliner pada web tripadvisor. Hasilnya mendapatkan nilai akurasi sebesar 98,67%. Hal ini menunjukkan kombinasi yang baik menggunakan metode Naive Bayes, pelabelan multi aspek, dan konversi ikon emosi.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini menggunakan teknik *web scraping* dari *reviewer* aplikasi *e-commerce* Shopee. Data yang diambil adalah data ulasan pada tahun 2020. Hasil *scraping dataset* penelitian sebagaimana disajikan pada Gambar 3 berikut.

Gambar 3. Dataset

4.2 Text Preprocessing

Data ulasan pengguna aplikasi *Shopee* pada *website google play store* merupakan data teks yang tidak terstruktur, sehingga teks perlu diolah agar mendapatkan informasi yang berguna. Tahapan-tahapan yang dilakukan pada proses ini, yaitu proses *normalization*, *case folding*, *tokenizing*, dan *stopwords*. Tahapan pertama atau *normalization* adalah proses teks yang dilakukan untuk memperbaiki kata-kata yang salah eja atau kata yang disingkat. Misalnya, kata “tidak” dapat ditulis dengan “tdk, ngga, nggak, engga”, dan banyak lainnya. Dalam proses *spelling normalization* juga mengartikan bahasa asing khususnya *english* menjadi bahasa Indonesia. Proses *normalization* dapat dilihat pada Tabel 1. berikut.

Tabel 1. Proses Normalization

Data Input	Data Output
Suka banget sama aplikasi ini,memang sih sejauh ini gapernah ngalamin yg namanya pengembalian barang, segala macam, tapi kalo belanja di Shopee aman,cepat banget pengirimannya,banyak event " menarik, kumpulin koin bisa dituker potongan harga, dapet voucher gratis ongkirrr . GREAT JOB SHOPEE!!	Suka sekali sama aplikasi ini, memang sejauh ini tidak pernah mengalami yang namanya pengembalian barang, segala macam, tapi kalo belanja di shopee aman, cepat banget pengirimannya, banyak event-event menarik, kumpulin koin bisa ditukar potongan harga, dapat voucher gratis ongkir, Kerja Bagus shopee

Tahapan kedua yaitu *case folding*. *case folding* adalah proses untuk mengkonversi semua karakter dalam dokumen menjadi huruf besar atau huruf kecil semua, agar mempercepat perbandingan selama proses pengindeksan. Proses *case folding* disajikan pada Tabel 2 berikut. (Salam, Zeniarja, and Khasanah 2018).

Tabel 2. Proses Case Folding

Data Input	Data Output
Suka banget sama aplikasi ini,memang sih sejauh ini gapernah ngalamin yg namanya pengembalian barang, segala macam, tapi kalo belanja di Shopee	suka sekali sama aplikasi ini, memang sejauh ini tidak pernah mengalami yang namanya pengembalian barang, segala macam, tapi kalo belanja di shopee aman,

Data Input	Data Output
aman,cepat pengirimannya,banyak event!" menarik, kumpulin koin bisa dituker potongan harga, dapet voucher gratis ongkirrr, GREAT JOB SHOPEE!!	cepat banget pengirimannya, banyak event-event menarik, kumpulin koin bisa ditukar potongan harga, dapat voucher gratis ongkir, kerja bagus shopee

Kemudian lanjut ke tahapan ketiga, yaitu *tokenizing*. *Tokenizing* merupakan proses pemisahan teks menjadi potongan kata yang disebut token. Tujuan dari proses ini adalah mendapatkan potongan kata yang akan menjadi entitas yang memiliki nilai dalam matriks dokumen teks yang akan dianalisis.

Selanjutnya masuk ke tahap yang terakhir, yaitu *stopword*. Pada tahapan ini kata-kata atau simbol yang tidak memberikan informasi yang berguna akan dieliminasi (Utami 2017).

4.3 Klasifikasi Sentimen

Pada proses klasifikasi sentimen data akan dibagi menjadi data latih dan data uji yang kemudian akan diklasifikasi menggunakan metode *Naive Bayes* dan *KNN*. Data latih dan data uji yang digunakan adalah data yang telah memiliki label kelas, dengan perbandingan data uji dan data latih adalah 80% : 20%.

Tabel 3. Proses Tokenizing

Data Input			
suka sekali sama aplikasi ini memang sejauh ini tidak pernah mengalami yang namanya pengembalian barang, segala macam, tapi kalo belanja dishopee aman, cepat banget pengirimannya, banyak event-event menarik, kumpulin koin bisa ditukar potongan harga, dapat voucher gratis ongkir, kerja bagus shopee			
Data Output			
suka sekali sama aplikasi ini memang sejauh ini tidak pernah	mengalami yang namanya pengembalian barang segala tapi kalo belanja	dishopee aman cepat banget pengiriman banyak event-event menarik kumpulin koin	bisa ditukar potongan harga dapat voucher gratis ongkir kerja bagus shopee

Klasifikasi Naive Bayes

Klasifikasi menggunakan metode *Naive Bayes* mendapatkan hasil nilai *accuracy* 0,914, *precision* 0,915, dan *recall* 0,914. Pada gambar 4 menunjukkan proses klasifikasi menggunakan orange.

Tabel 4. Proses Stopword

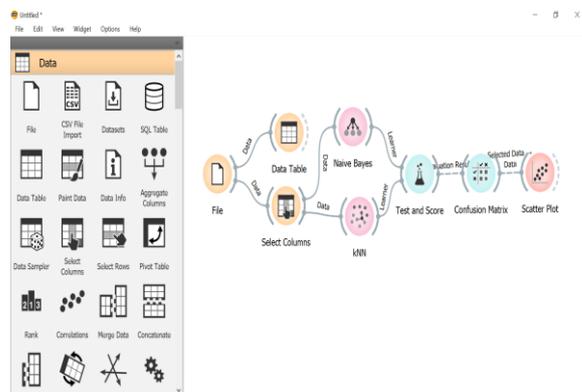
Data Input	Data Output
suka sekali sama aplikasi ini memang sejauh ini tidak pernah mengalami yang namanya pengembalian barang, segala macam, tapi kalo belanja dishopee aman, cepat banget pengirimannya, banyak event-event menarik, kumpulin koin bisa ditukar potongan harga, dapat voucher gratis ongkir, kerja bagus shopee	Suka aplikasi pengembalian barang belanja potongan harga voucher gratis ongkir aman cepat banyak

Tabel 5. Data Uji dan Data Latih

Aplikasi	Klasifikasi	Data latih (80%)	Data Uji (20%)	Jumlah
Shopee	Positif	200	50	250
	Negatif	200	50	250
Total		400	100	500

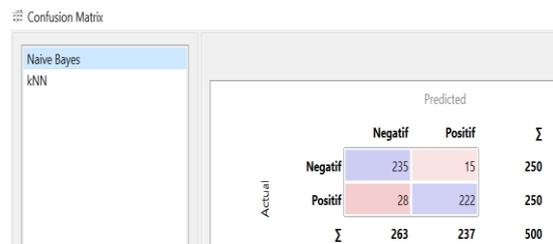
Tabel 6. Hasil Evaluasi Model Naive Bayes

Classifier	Accuracy	Precision	Recall
Naive Bayes	0,914	0,915	0,914



Gambar 4. Proses Klasifikasi dengan orange

Kemudian hasil *confusion matrix* menggunakan metode *Naive Bayes* dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix Model Naive Bayes

Gambar 5 di atas menunjukkan hasil *confusion matrix* yang didapatkan dengan menggunakan metode *Naive Bayes*, yaitu sentimen bernilai negatif yang diprediksi dengan benar sebanyak 235 dan yang salah sebanyak 15 ulasan. Kemudian sentimen positif yang diprediksi dengan benar sebanyak 222 dan yang tidak tepat sebanyak 28 ulasan.

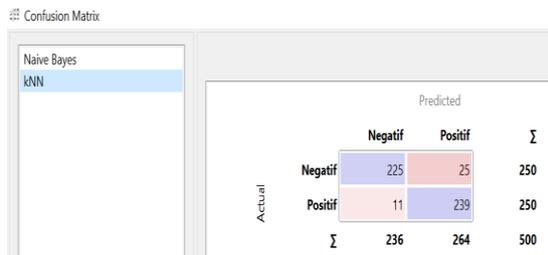
Klasifikasi K-Nearest Neighbor

Klasifikasi menggunakan metode *KNN* mendapatkan hasil nilai *accuracy* 0,928, *precision* 0,929, dan *recall* 0,928.

Tabel 7. Hasil Evaluasi Model KNN

Classifier	Accuracy	Precision	Recall
KNN	0,928	0,929	0,928

Kemudian hasil *confusion matrix* menggunakan metode *KNN* dapat dilihat pada Gambar 6 berikut.



Gambar 6. Confusion Matrix Model KNN

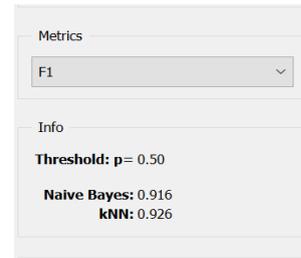
Gambar 6 di atas menunjukkan hasil *confusion matrix* menggunakan metode *KNN*, yaitu sentimen bernilai negatif yang diprediksi dengan benar sebanyak 225 dan yang salah sebanyak 25 ulasan. Kemudian sentimen positif yang diprediksi dengan benar sebanyak 239 dan yang tidak tepat sebanyak 11 ulasan.

Berdasarkan hasil klasifikasi menggunakan metode *Naive Bayes* dan *KNN*, dapat disimpulkan bahwa kinerja metode *KNN* lebih unggul dengan nilai *accuracy*, *precision* dan *recall* lebih tinggi dari pada metode *Naive Bayes*. Perbandingan evaluasi kedua metode dapat dilihat pada Tabel 8 berikut.

Selanjutnya dilakukan pencarian nilai *F1 score*, dimana *F1 score* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan. Rumus $F1\ Score = 2 * (recall * precision) / (recall + precision)$. Berikut hasil *score F1* menggunakan *widget calibration plot* dengan metode *Naive Bayes* dan *KNN*.

Tabel 8. Perbandingan Evaluasi Metode *Naive Bayes* dan *KNN*

Classifier	Accuracy	Precision	Recall
Naive Bayes	0,914	0,915	0,914
KNN	0,928	0,929	0,928



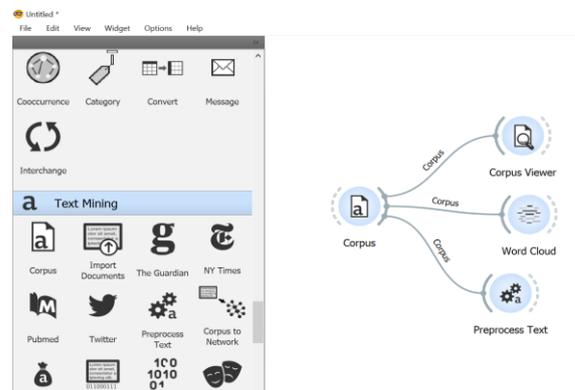
Gambar 7. Hasil F1 Score.

Gambar 7 di atas menunjukkan hasil pengujian *F1 score* dengan metode *Naive Bayes* yang memperoleh nilai sebesar 0.916 dan metode *KNN* sebesar 0.926. Hasil ini mengindikasikan bahwa model klasifikasi menggunakan metode *Naive Bayes* dan *KNN* mempunyai *precision* dan *recall* yang baik.

4.4. Visualisasi Data Review

Word Cloud

Hasil dari *word cloud* ini memvisualisasikan data review berupa nilai dari kata-kata yang sering diberikan oleh pelanggan yang memberi ulasan pada *e-commerce Shopee*. Semakin besar ukuran *font* pada *word cloud* maka topik tersebut yang sering dibicarakan oleh pelanggan yang memberi ulasan. Hasil dari *word cloud* dapat dilihat pada Gambar 9 berikut.



Gambar 8. Proses Word Cloud

Berdasarkan hasil *word cloud* ini kita dapat mengetahui kata-kata yang bersentimen positif dan negatif beserta nilai dari kata-kata tersebut. Hal ini berguna untuk mengetahui faktor-faktor yang mempengaruhi kepuasan pelanggan.

Frekuensi Kata Positif

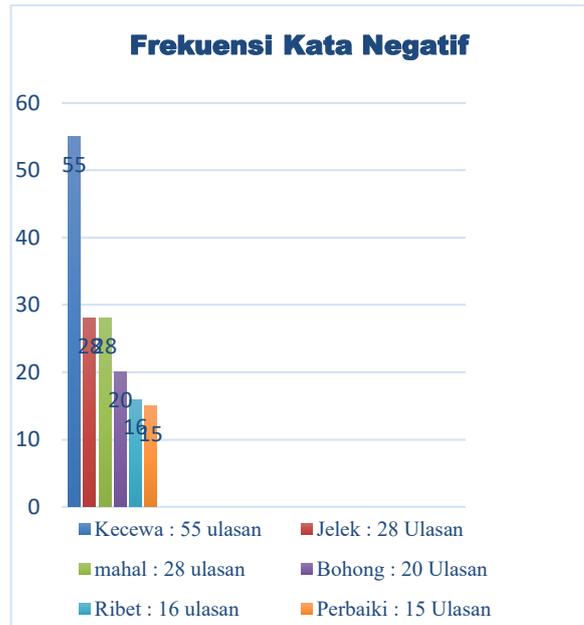
Berdasarkan hasil klasifikasi ulasan sentimen dan penggunaan metode *word cloud* dibuatlah grafik frekuensi kata positif dari 500 data ulasan dan diperoleh 6 kata positif yang paling sering keluar, diantaranya adalah : “gratis” dengan frekuensi 100 kali, “bagus” 99 kali, “suka” 47 kali, “murah” 35 kali, “mudah” 33 kali, dan “cepat” 28 kali, yang hasilnya dapat dilihat pada Gambar 10 berikut.



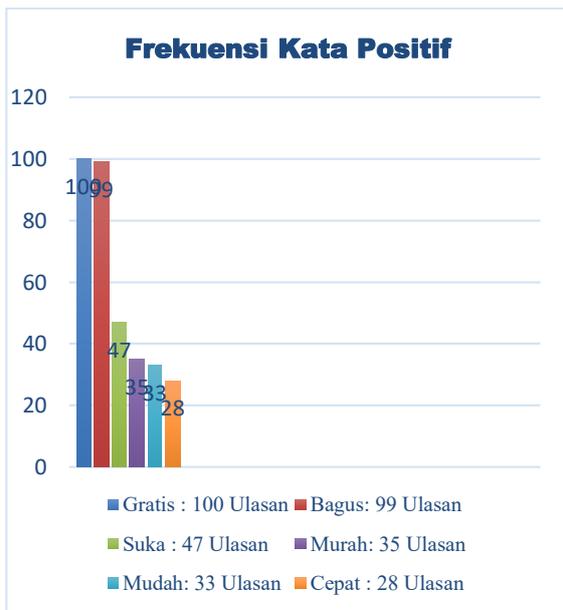
Gambar 9. Hasil Word Cloud

Frekuensi Kata Negatif

Kemudian pada hasil klasifikasi ulasan bersentimen negatif diperoleh kata yang dianggap relevan sebagai sentimen negatif, adalah: “kecewa” dengan frekuensi 55 kali, “jelek” 28 kali, “mahal” 28 kali, “bohong” 20 kali, “ribet” 16 kali, dan “perbaiki” 15 kali, yang hasilnya dapat dilihat pada Gambar 11 berikut.



Gambar 11. Frekuensi Kata Negatif



Gambar 10. Frekuensi Kata Positif

4.5 Asosiasi Kata Positif dan Negatif

Asosiasi kata positif dan negatif dilakukan untuk mendapatkan informasi terkait kata-kata yang saling berhubungan dengan topik yang diulas. Tabel 9 menampilkan asosiasi kata dengan kelas sentimen positif. Proses ini dilakukan dengan cara menyaring kata-kata yang memiliki hubungan dengan topik yang diulas. Berikut didapatkan topik-topik yang berasosiasi dengan kelas sentimen positif.

1. Kata “produk” berasosiasi dengan kata bagus, sesuai, dan suka. Yang artinya pengguna menyukai produk dari aplikasi Shopee serta barang yang diterima juga sesuai harapan.

Tabel 9. Asosiasi Kata Positif

	Produk	Pengiriman
Bagus	99	Mudah 33
Sesuai	35	Cepat 28
Suka	47	Aman 12
	Aplikasi	Pelayanan
Suka	47	Bagus 99
Mudah	33	Suka 47
Update	30	Cepat 28
	Ongkir	Voucher
Gratis	100	Puas 24
Murah	35	Membantu 19
		Mantap 18

2. Kata “aplikasi” menunjukkan informasi bahwa aplikasi yang digunakan mempermudah pengguna dan aplikasi selalu *update* sehingga fitur-fiturnya dapat memberi kemudahan.
3. Kata “ongkir” didapatkan informasi bahwa pelanggan puas dengan fitur gratis ongkir yang diberikan oleh pihak Shopee.
4. Kata “pengiriman” didapatkan informasi bahwa proses pengiriman barang yang dilakukan mudah, cepat, dan aman.
5. Kata “pelayanan” didapatkan informasi bahwa pelayanan yang didapatkan oleh pengguna aplikasi Shopee bagus dan cepat.
6. Kata “voucher” didapatkan informasi bahwa pelanggan puas dan sangat membantu pelanggan dalam berbelanja.

Selanjutnya berdasarkan Tabel 10 didapatkan asosiasi kata yang berhubungan dengan ulasan dengan kelas sentimen negatif. Berikut informasi yang didapatkan.

1. Kata “produk” mendapatkan informasi produk yang diterima memiliki kualitas yang jelek atau tidak sesuai dengan ekspektasi pengguna.
2. Kata “aplikasi” menunjukkan informasi aplikasi yang digunakan ribet serta diharapkan oleh pengguna untuk diperbaiki.

3. Kata “ongkir” berasosiasi dengan kata mahal yang menunjukkan ongkir terasa mahal oleh pengguna.
4. Kata “pengiriman” mendapatkan informasi pengiriman yang dilakukan mendapatkan masalah seperti pengiriman yang dilakukan lama dan terjadinya penipuan terhadap pengguna aplikasi Shopee.
5. Kata “pelayanan” mendapatkan informasi pelayanan yang didapatkan oleh pengguna jelek dan mengecewakan.
6. Kata “voucher” mendapatkan informasi kecewa dan bohong.

Tabel 10. Asosiasi Kata Negatif

Produk	Pengiriman		
Kecewa	55	Susah	15
Jelek	28	Nunggu	15
Mahal	28	Penipuan	10
Aplikasi	Pelayanan		
Tidak bisa	26	Jelek	28
Ribet	16	Mengecewakan	14
Perbaiki	15		
Ongkir	Voucher		
Mahal	28	Kecewa	55
		Bohong	20

5. KESIMPULAN

Berdasarkan hasil pengujian dari 500 data ulasan menggunakan aplikasi orange dengan metode *Naive Bayes* dan *K-Nearest Neighbor* dapat disimpulkan bahwa metode *KNN* mendapatkan nilai yang lebih unggul dari segi *accuracy*, *precision*, *recall* dan *F1 score*. Kemudian pada hasil *confusion matrix* metode *Naive Bayes* lebih unggul dari pada metode *KNN* dalam memprediksi ulasan bersentimen negatif dengan hasil klasifikasi benar sebanyak 235 dan yang salah sebanyak 15 ulasan. Namun dalam memprediksi ulasan bersentimen positif metode *KNN* lebih unggul dengan hasil klasifikasi sentimen positif yang benar sebanyak 239 ulasan dan yang salah sebanyak 11 ulasan.

Selanjutnya, dari hasil pengujian menggunakan metode *word cloud* didapatkan informasi topik sentimen positif dan negatif yang sering *direview* oleh pelanggan, diantaranya terkait topik: produk, aplikasi, ongkir, pengiriman, pelayanan, dan voucher.

Setelah mengetahui topik yang sering diulas oleh pelanggan kemudian dilakukannya asosiasi kata sehingga ditemukan kata yang berasosiasi dengan sentimen positif yang diulas oleh pengguna, seperti terkait kata :gratis, bagus, suka, murah, mudah, dan cepat. Sedangkan hasil asosiasi kata bersentimen negatif yang ditemukan diantaranya terkait kata : kecewa, jelek, mahal, bohong, ribet, dan perbaiki.

Hasil dari penelitian ini mendapatkan informasi mengenai sentimen positif dan negatif, dan juga faktor-faktor yang mempengaruhi kepuasan pelanggan berdasarkan hasil *word*

cloud dan asosiasi kata sehingga kedepannya pihak *developer* dapat melakukan perbaikan kearah yang lebih baik lagi.

DAFTAR PUSTAKA

- ALAMSYAH, ANDRY, & FAISHAL NURUZ ZUHRI. 2017. “Measuring Public Sentiment Towards Services Level in Online Forum Using Naive Bayes Classifier and Word Cloud.” *CRS-ForMIND International Conference and Workshop 2017* (October).
- ARTANTI, DIAN PUTRI, ABDUL SYUKUR, ADI PRIHANDONO, & DE ROSAL I MOSES SETIADI. 2018. “Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naive Bayes.” : 8–9.
- AZZAHRA, SITTI ALIYAH, & ARIEF WIBOWO. 2020. “Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi Dengan Algoritme Naive Bayes Untuk Ulasan Wisata Kuliner Pada Web Tripadvisor.” *Jurnal Teknologi Informasi dan Ilmu Komputer* 7(4): 737.
- DWI, HADYA JAYANI. 2020. “Pengunjung Situs Shopee Terbesar Di Indonesia.” 6 Juli (April): 1.<https://databoks.katadata.co.id/datapublish/2020/07/06/pengunjung-situs-shopee-terbesar-di-indonesia#>.
- MASRIPAH, SITI, & LILA DINI UTAMI. 2020. “Algoritma Klasifikasi Naive Bayes Untuk Analisa Sentimen Aplikasi Shopee.” *Swabumi* 8(2): 114–17.
- MUKTAFIN, ELIK HARI, KUSRINI KUSRINI, & EMHA TAUFIQ LUTHFI. 2020. “Analisis Sentimen Pada Ulasan Pembelian Produk Di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing.” *Jurnal Eksplorasi Informatika* 10(1): 32–42.
- PULAKIANG, ANTHONI RIVAI ;I MADE NOMO WIRATHA; JOHN REIMON BATMETAN. 2019. “Analisis Kualitas Layanan E-Commerce Menggunakan Twitter.” *Indonesian Jpurnal on Information System* 4(April): 25–31.
- RATNAWATI, FAJAR. 2018. “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter.” *INOVTEK Polbeng - Seri Informatika* 3(1): 50.
- SALAM, ABU, JUNTA ZENIARJA, & RIMA SEPTIYAN USWATUN KHASANAH. 2018. “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia).” *Prosiding SINTAK*: 480–86.
- SEPTIAN, JEREMY ANDRE, TRESNA MAULANA FAHRUDIN, & ARYO NUGROHO. 2019. “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF Dan K - Nearest Neighbor.” *Journal of Intelligent Systems and*

- Computation* (September): 43–49.
- UTAMI, LILYANI ASRI. 2017. “Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization.” 13(1): 103–12.
- VIJAYARAGAVAN, P., R. PONNUSAMY, & M. ARAMUDHAN. 2020. “An Optimal Support Vector Machine Based Classification Model for Sentimental Analysis of Online Product Reviews.” *Future Generation Computer Systems* 111: 234–40. <https://doi.org/10.1016/j.future.2020.04.046>.
- WU, BING, & PENG LI. 2020. “Influence of MOOCs EWOM on the Number of Registrations and Completions.” *IEEE Access* 8: 158826–38.

Halaman ini sengaja dikosongkan