

ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA

¹Rita Apriani, ²Dudih Gustian

^{1,2}Program Studi Sistem Informasi

¹Universitas Nusa Putra, Sukabumi Indonesia

¹Jl. Raya Cibolang Kaler No. 21, Kab. Sukabumi

e-mail : ¹ritaaprianirita@gmail.com, ²dudih@nusaputra.ac.id

Korespondensi : ¹ritaaprianirita@gmail.com

ABSTRAK

Tokopedia merupakan toko online yang paling banyak dikunjungi masyarakat Indonesia yakni dengan total pengunjung mencapai 1,2 miliar yang terbagi dari 863,1 juta pengunjung dari web mobile dan 329,8 juta pengunjung dari desktop. Meskipun menduduki peringkat teratas dan rating 5 mendapatkan nilai terbanyak, tentu tidak semua memberikan komentar positif, sebagian ada yang memberikan komentar negatif, unsur kepercayaan pengguna memegang peranan penting untuk kelangsungan toko online, terkadang produk yang dijual atau dibeli melalui salah satu situs ecommerce barang (produknya) tidak sesuai dengan foto di iklan dan juga terkadang barangnya tidak sesuai dengan apa yang di inginkan. Untuk meneliti dan menganalisis hal tersebut maka diperlukan suatu metode dan analisis untuk mengklasifikasikan komentar pengguna ke dalam beberapa kategori, dimana dalam penelitian ini berupa kategori positif dan negatif. Penelitian ini menggunakan metode Naive Bayes untuk menghasilkan sentimen positif dan negatif terhadap komentar pengguna aplikasi Tokopedia di Playstore. Pengujian berdasarkan nilai class negative, class positive, recall, dan accuracy pada analisis sentimen. dengan nilai accuracy performance yang baik sebesar 97,13%, dengan nilai precision 1 Sementara pada Class Recall dihasilkan nilai 95,49% (positive class: negative). Dan nilai AUC 0,980.

Kata Kunci : Analisis Sentimen, Naive Bayes, Tokopedia

ABSTRACT

Tokopedia is the online shop most visited by Indonesians, with a total visitor number of 1.2 billion, divided from 863.1 million visitors from the mobile web and 329.8 million visitors from the desktop. Even though it is in the top rank and rating 5 gets the most value, of course not all of them give positive comments, some of them give negative comments, the element of user trust plays an important role in the continuity of an online store, sometimes products are sold or purchased through one of the goods e-commerce sites (the product is) does not match the photo in the advertisement and also sometimes the goods are not what you want. To research and analyze this, we need a method and analysis to classify user comments into several categories, which in this study are positive and negative categories. This study uses the Naive Bayes method to generate positive and negative sentiments towards comments from users of the Tokopedia application on Playstore. The test is based on the value of class negative, class positive, recall, and accuracy on the sentiment analysis. with a good accuracy performance value of 97.13%, with a precision value of 1 While the Recall class yields a value of 95.49% (positive class: negative). And the AUC value is 0.980.

Keyword : Sentiment Analysis, Naive Bayes, Tokopedia

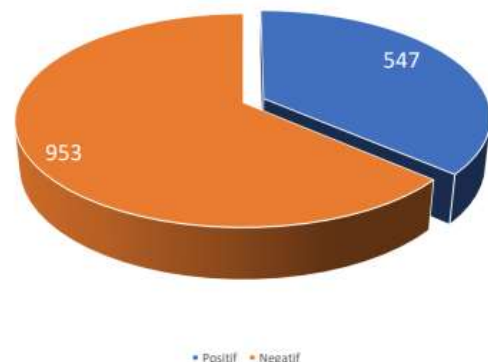
I. PENDAHULUAN

Perkembangan pemakaian smartphone di Indonesia saat ini dirasakan sangat pesat, Jumlah penduduk Indonesia yang mencapai 250 juta jiwa adalah pasar yang besar untuk pesatnya pertumbuhan pengguna smartphone. Lembaga riset digital marketing Emarketer memperkirakan pada 2018 jumlah pengguna aktif smartphone di Indonesia lebih dari 100 juta orang. Dengan jumlah sebesar itu, Indonesia akan menjadi negara dengan pengguna aktif smartphone terbesar keempat di dunia setelah Cina, India, dan Amerika [1]. Menristekdikti menyebutkan angka pengguna smartphone di Indonesia kini mencapai sekitar 25% dari total penduduk atau sekitar 65 juta orang [2], angka prosentase ini menjadi peluang pasar yang menarik sehingga beberapa kalangan di dalam maupun luar negeri kemudian mencoba untuk membuat dan memasarkan aplikasi mobile yang bermanfaat untuk kehidupan masyarakat. Ketatnya persaingan di industri e-commerce Indonesia membuat para pelaku bisnisnya harus terus menghadirkan inovasi dan langkah strategis demi bisa memenangkan pasar. Salah satu situs marketplace lokal terkemuka Tokopedia meluncurkan aplikasi mobile untuk platform Android. seluruh fitur Tokopedia bisa diakses oleh penjual dan pembeli melalui perangkat ponsel. Ini merupakan aplikasi mobile Tokopedia yang kedua setelah sebelumnya tahun lalu meluncurkan aplikasi mobile untuk perangkat BlackBerry 10. Kini Tokopedia mencoba merambah pasar yang lebih luas dengan meluncurkan aplikasi resmi Tokopedia khusus Android. Langkah ini tidak mengherankan karena platform buatan Google ini merupakan platform yang paling pesat pertumbuhan dan penggunaannya saat ini di Indonesia [3]. Menurut hasil survei situs kupon diskon CupoNation. Ditahun 2019, CupoNation merilis studi tentang situs belanja online yang paling banyak dikunjungi online shopper Indonesia sepanjang tahun 2019. Dalam studi ini, terungkap 10 situs belanja online terpopuler, baik dari platform e-commerce lokal maupun internasional. Di samping itu, CupoNation juga menganalisis sumber pengunjung yang datang dari versi desktop maupun web mobile.

Tabel 1. 10 situs belanja online lokal dan asing yang diurut berdasarkan jumlah trafik terbesar di 2019 [4]

Toko Online	Jumlah Pengunjung selama 12 bulan (desktop)	Jumlah Pengunjung selama 12 bulan (web mobile)	Jumlah pengunjung Indonesia selama 12 bulan	% pengunjung dari desktop	% pengunjung dari web mobile
Tokopedia	329,823,447	863,956,416	1,192,879,863	27.65%	72.35%
Shopee	172,275,455	664,805,215	837,080,670	20.58%	79.42%
Bukalapak	153,624,428	669,862,991	823,487,419	18.66%	81.34%
Lazada	67,129,891	378,450,756	445,580,647	15.07%	84.93%
Blibli	141,642,937	211,559,283	353,202,220	40.10%	59.90%
JD ID	34,198,152	81,219,193	105,417,345	22.95%	77.05%
Orami	2,167,378	87,660,992	89,828,371	2.45%	97.55%
Shimeka	14,817,938	48,911,817	63,729,756	22.38%	77.72%
Sociolla	5,189,789	45,984,181	51,173,970	10.14%	89.86%
Zalora	18,312,615	26,192,069	44,504,684	41.15%	58.85%

Tokopedia merupakan toko online yang paling banyak dikunjungi masyarakat Indonesia yakni dengan total pengunjung mencapai 1,2 miliar yang terbagi dari 863,1 juta pengunjung dari web mobile dan 329,8 juta pengunjung dari desktop [5]. Meskipun menduduki peringkat teratas dan rating 5 mendapatkan nilai terbanyak, tentu tidak semua memberikan komentar positif terhadap aplikasi Tokopedia, mungkin sebagian ada yang memberikan komentar negatif, sebagian lagi memberikan komentar yang netral terhadap aplikasi ini, unsur kepercayaan pengguna atau masyarakat memegang peranan penting untuk kelangsungan toko online, terkadang produk yang dijual atau dibeli melalui salah satu situs ecommerce barang (produknya) tidak sesuai dengan foto di iklan dan juga terkadang barangnya tidak sesuai atau bahkan diterima dalam keadaan rusak, tidak sesuai dengan apa yang di inginkan.



Gambar 1. Hasil Analisis Sentimen Positif dan Negatif Terhadap Komentar Di Aplikasi Tokopedia google Play

Pada gambar 1 diatas, nampak bahwa dari hasil survei yang dilakukan oleh peneliti dari total 1500 komentar menghasilkan nilai 953 komentar

atau sekitar 63% yang berasumsi negatif, sedangkan yang berasumsi positif dengan nilai 547 komentar atau sekitar 36%. Ini menandakan bahwa cukup banyak berasumsi negatif saat orang bertransaksi dengan menggunakan tokopedia misalnya. Hal ini bisa dimaklumi karena tingkat kepercayaan masyarakat untuk berbelanja secara online belum begitu banyak dan percaya.

Yane Marita Febrianti, Indriati, dan Agus Wahyu Widodo dalam penelitiannya berjudul Analisis Sentimen Pada Ulasan “Lazada” Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance. Dengan menggunakan metode K-NN Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance, pengujian berdasarkan nilai precision, recall, dan accuracy pada masing-masing analisis sentimen tanpa perbaikan kata, maupun dengan perbaikan kata. Hasil pengujian dengan nilai accuracy yang baik terdapat pada analisis sentimen dengan perbaikan kata yakni 76 %, dengan nilai precision 0,76, dengan nilai recall [6]. Santoso, Valonia & Virginia, Gloria & Lukito, Yuan. Dalam penelitiannya berjudul Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine, penelitian yang dilakukan terhadap hasil evaluasi dosen menghasilkan sentimen analisis positif, negatif dan netral dari hasil angket kuesioner yang diisi oleh mahasiswa menggunakan metode SVM dengan hasil penelitian Akurasi tertinggi didapatkan pada dataset dengan $k = 6$ dan akurasi yang dicapai adalah 67,83% [7]. Sementara Aputra, S. A., Didi Rosiyadi, Windu Gata, & Syepri Maulana Husain. Dalam penelitiannya berjudul Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization, meneliti tentang ulasan sentimen positif dan negatif pengguna terhadap aplikasi OVO pada google play/play store. Algoritma yang digunakan dalam penelitian ini adalah Naive Bayes Classifier (NB), dengan optimasi penggunaan Feature Selection (FS) Particle Swarm Optimization. Hasil dari cross validation NB tanpa FS adalah 82.30 % untuk accuracy dan 0.780 untuk AUC. Sedangkan untuk NB dengan FS adalah 83.60 % untuk accuracy dan 0.801 untuk AUC [8].

Penelitian ini memberikan solusi untuk mengetahui tingkat prosentase komentar dan respon pengguna aplikasi tersebut berdasarkan komentar yang diberikan pengguna/konsumen pada aplikasi Tokopedia, dimana dengan prosesntase dan respon tersebut akan menjadi masukan demi peningkatan mutu pelayanan, kualitas produk serta kualitas internal perusahaan.

Peneltian ini memberikan manfaat selain bisa memberikan informasi pada pihak manajemen Tokopedia agar tetap konsisten dalam pelayanan pada konsumen sehingga tingkat kepercayaan pengguna/konsumen tetap terjaga, yang akhirnya dapat meningkatkan keuntungan sesuai target yang dibuat.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Mohamad Syahrul Mubaroqa, Adiwijaya, dan Muhammad Dwi Aldhi dalam penelitiannya berjudul Aspect-based Sentiment Analysis to Review Products Using Naïve Bayes. Analisis sentimen digunakan untuk menganalisis dan ekstrak polaritas sentimen pada ulasan produk berdasarkan pada aspek spesifik produk. Penelitian ini dilakukan di PT Tiga Fase, seperti preprocessing data yang melibatkan penandaan part-of-speech (POS), pemilihan fitur menggunakan Chi Square, dan klasifikasi polaritas sentimen aspek menggunakan Naïve Bayes. Berdasarkan hasil evaluasi, diketahui bahwa sistem mampu melakukan analisis sentimen berbasis aspek dengan F1-Measure tertinggi 78,12% [9].

Dudih Gustian, Indah Suciati, dan Sudin Saepudin dalam penelitiannya Sistem Pakar Dengan Algoritma Naive Bayes Untuk Prediksi Hasil Produksi Ayam Broiler Plasma (Studi Kasus : Pt.Sekawan Sinar Surya) memberikan kontribusi kepada peternak atau pihak perusahaan dengan memberikan parameter yang berpengaruh untuk menentukan hasil produksi. Penggunaan algoritma Naive Bayes mampu menghasilkan prediksi klasifikasi keuntungan dan kerugian sebelum perhitungan realisasi pada bagian produksi. Untuk mengetahui nilai keakuratan pengujian dilakukan dengan menggunakan Software Weka. Akurasi rata-rata dari algoritma yang dihasilkan, Naive Bayes dari 96.36% mendekati angka 100% menunjukkan hasil yang maksimal. dan dengan

nilai Receiver Operating Curve (ROC) 0,9995 mendekati angka 1 maka model yang dihasilkan lebih baik [10].

Achmad Bayhaqy, Kaman Nainggolan, Sfenrianto Sfenrianto, dan Emil R. Kaburuan dalam penelitiannya Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes, bahwa data mining dapat digunakan untuk analisis data pengguna media sosial yang mengunjungi ECommerce. Penelitian ini menggunakan data teknik penambangan yang bertujuan membandingkan klasifikasi dalam analisis sentimen dari pandangan pelanggan ECommerce yang telah ditulis di Twitter. Kumpulan data berasal dari tweet tentang E-Commerce di Tokopedia dan Bukalapak Teks Preprosesing teknik penambangan, tokenize, batang, klasifikasi, dll digunakan untuk membangun klasifikasi dan sentimen analisis. Rapidminer juga digunakan untuk membantu dalam membuat analisis Sentimen untuk perbandingan dengan menggunakan tiga klasifikasi berbeda dalam dataset dengan Decision Tree, K-NN, dan Naïve Bayes Classifier mendekati akurasi terbaik. Hasil tertinggi dari penelitian ini adalah pendekatan Naïve Bayes dengan akurasi 77%, presisi 88,50% dan recall 64% [11].

Fajar Ratnawati dalam penelitiannya berjudul Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter, menjelaskan bahwa Banyaknya opini yang dituliskan di twitter membutuhkan pengklasifikasian sesuai sentimen yang dimiliki agar mudah untuk mendapatkan kecenderungan opini tersebut terhadap film apakah cenderung beropini positif atau negatif. Algoritma yang digunakan adalah Algoritma Naive Bayes. Berdasarkan hasil eksperimen yang telah dilakukan dan berdasarkan hasil analisis yang didapatkan dalam penelitian tersebut maka didapat hasil analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90 % dengan rincian nilai precision 92%, recall 90% dan f-measure 90% [12].

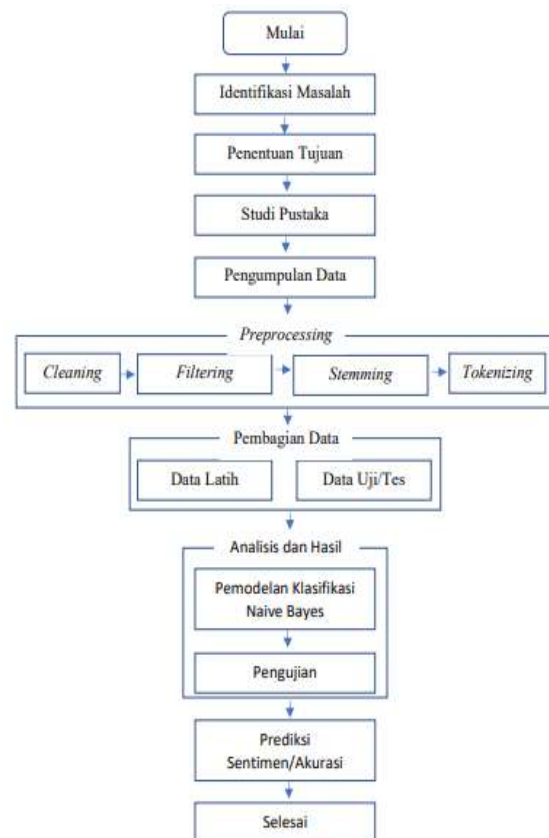
2.2 Kerangka Pemikiran



Gambar 2. Kerangka Pemikiran

III. METODOLOGI PENELITIAN

3.1 Tahapan Penelitian



Gambar 3. Tahapan penelitian

Berikut adalah penjelasan dari Alur Metode Penelitian :

1. Identifikasi Masalah

Mengamati dan menemukan permasalahan yang terjadi pada Aplikasi Tokopedia dilihat dari komentar pengguna aplikasi pada Google Playstore

2. Penentuan Tujuan

Berfungsi untuk lebih memperjelas kerangka apa saja yang menjadi sasaran penelitian ini. Seperti yang sudah ditulis di Bab I bahwa tujuan penelitian ini adalah untuk melakukan klasifikasi terhadap komentar pengguna aplikasi Tokopedia di Google Playstore dengan menggunakan algoritma Naive Bayes serta menghitung akurasi di Rapidminer.

3. Studi Pustaka

Bertujuan untuk mengetahui teori-teori apa yang akan digunakan untuk menyelesaikan permasalahan yang akan diteliti, serta mendapatkan dasar-dasar referensi yang kuat bagi peneliti

4. Pengumpulan Data

Proses pengumpulan data dan input dalam sistem ini dimulai dengan data web scraping, proses web scraping yang dilakukan adalah dengan menggunakan dataminer, salah satu chrome extension yang bisa digunakan untuk scraping data web. Keluaran dari sistem ini adalah berupa data testing dengan nilai output berupa sentimen Positif dan Sentimen Negatif yang diklasifikasikan oleh sistem berdasarkan pembelajaran data training

5. Preprocessing

Preprocessing merupakan salah satu tahapan yang penting untuk data pada proses mining. Data yang digunakan dalam proses mining tidak selamanya dalam kondisi yang ideal untuk diproses. Terkadang pada data tersebut terdapat berbagai permasalahan yang dapat mengganggu hasil dari proses mining itu sendiri seperti diantaranya adalah missing value, data redundant, outliers, ataupun format data yang tidak sesuai dengan sistem. Oleh karenanya untuk mengatasi permasalahan tersebut dibutuhkan tahap Preprocessing. Preprocessing merupakan salah satu tahapan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data, Pada

tahap ini terdapat tahapan yang di lalui diantaranya :

a. Cleaning Proses cleaning yang dilakukan yaitu merubah teks menjadi huruf kecil (case folding), menghapus karakter selain huruf, menghapus username pengguna atau mentions (@), menghapus hashtag (#), menghapus URL atau link dari setiap komentar.

b. Filtering Pembuangan kata-kata tidak penting dari hasil token. Selain juga dilakukan penghapusan tanda baca dan stopword. Stopword diproses pada sebuah kalimat jika mengandung kata-kata yang sering keluar dan di anggap tidak penting seperti waktu, penghubung, dan lain sebagainya

c. Stemming Pada tahapan stemming, yaitu pengubahan kata berimbuhan menjadi kata dasar

d. Tokenizing Pada tahapan tokenizing, setiap kata akan dipisahkan berdasarkan spasi yang ditemukan

6. Pembagian Data

Data yang telah melalui proses pelabelan, dilakukan pembagian data menjadi 2, yaitu data latih dan data uji. Untuk masing-masing pembagian data data latih dan uji ini penulis membagi kembali menjadi 2 yaitu : - Pertama, data latih berjumlah berjumlah lebih kecil dari data uji - Kedua, data latih berjumlah lebih banyak dari data uji.

7. Analisis dan Hasil Pada tahap analisis dan hasil penelitian ini klasifikasi yang digunakan adalah algoritma Naive Bayes Sebuah bentuk klasifikasi probalistik yang berdasarkan Teorema Bayes (dari statistic Bayesian dengan strong (naïve) independence assumption atau asumsi bebas. Sebuah aturan yang lebih deskriptif untuk dasar sebuah model yang akan menjadi “model yang memiliki ciri-ciri kebebasan”.

3.2 Input

Input yang dimasukkan ke sistem adalah dokumen berupa komentar dari pengguna aplikasi Tokopedia di Google Playstore. Teknik scrapping data Menggunakan dataminer salah satu chrome extension, data komentar yang di scrap

menggunakan bahasa Indonesia seperti yang dilihatkan dalam tabel berikut ini :

KOMENTAR	OPINI
 <p>Wira Rana ★★★★★ 10 Jan 2020</p> <p>Aku suka sekali Tokopedia menggunakan BTS sebagai brand ambassador. Apalagi sebentar lagi akan ada event dengan BTS, AMHM Indonesia mengucapkan terima kasih sebesar-besarnya kepada Tokopedia. Alhamdulillah, BTS dapat tanpa harus datang dan mencari ke nonton streaming. Cukup di TV atau app Tokopedia Tokopedia...</p> <p>[Harian Kompas]</p>	Positive
 <p>Kaf Syah Aziz ★★★★ 5 Agustus 2020</p> <p>Aplikasi ini bagus, gua beli pulsa yang 100.000 nggak masuk sama sekali ke nomor hp gua. Di aplikasi nya sudah berhasil tapi gua cek ngak ada dan CS K, nggak ngasih solusi sama sekali, bilang buat dan nya harus waiting cek sistem padahal mau duit nya aja...</p>	Negative
 <p>Elva Rahmatika Lestari ★★★★ 5 Agustus 2020</p> <p>Belanja kali belanja disini yg mau barangnya keserunya, penjual pasti semaksimalnya gitu. Ganti barang tanpa kembalikan, tapi barang juga sering nggk yg di kirim otomatis tapi barang di kembalikan ready. Btw anak aplikasi sebelah ah kalo belanja disini ada kemunduran dari penjual. Btw ya uninstal aja...</p>	Negative
 <p>Bintang Widya ★★★★ 10 Jan 2020</p> <p>Suka banget belanja disini, pengiriman yang cepat banget, barangnya bagus banget dan promo-promo menarik. Di kamariin kita bisa dapat cash back juga. Btw barangnya bagus banget, pengiriman cepat, dan juga barangnya bagus. Btw dan kalo perlu sudah banyak yang jadiin BTS. Btw ya mau saya akan beli barang...</p>	Positive

Gambar 4. Komentar yang di scrap

3.3 Evaluasi

Langkah selanjutnya adalah melakukan evaluasi yang bertujuan untuk mengetahui sejauhmana tingkat akurasi hasil klasifikasi data menggunakan metode naive bayes terhadap data testing/uji.

IV. HASIL DAN PEMBAHASAN

4.1 Pengumpulan data training dan data testing

Pengumpulan data dilakukan dengan dua tahap, tahap pertama adalah mengumpulkan data yang akan dijadikan data latih/training dan tahap kedua mengumpulkan data yang akan dijadikan data uji/testing. Data yang di ambil adalah komentar pengguna aplikasi Tokopedia di Google Playstore dengan menggunakan Dataminer. Langkah pertama adalah mengunjungi Aplikasi Tokopedia di situs google playstore dan kemudian melakukan scraping data komentar pengguna Aplikasi dengan mengklik menu new recipe pada extention chrome dataminer. Untuk data training dipilih sebanyak 1.000 komentar secara acak yang kemudian diberikan sentimen positif sebanyak 500 dan negatif sebanyak 500 pada setiap komentarnya secara manual. Rincian data training bisa dilihat pada tabel dibawah ini :

Tabel 2. Rincian Jumlah Data Training

Data	Sentimen Positif	Sentimen Negatif	Total Sentimen
Komentar Pengguna	500	500	1.000

Untuk data testing dipilih sebanyak 1.500 komentar terbaru dengan pengumpulan data yang sama dengan data training. Masing-masing data yang telah di scrap dari playstore tersebut diberi nama `training_tokopedia.xlsx` dan `testing_tokopedia.xlsx`, untuk selanjutnya dilakukan proses pembobotan kata dan pemberian sentimen secara otomatis melalui aplikasi yang dibuat.

4.2 Pengkategorian sentimen positif dan negatif data training

Kategori sentimen pada data training dilakukan secara manual. Kategorisasi sentimen ini dilakukan agar data training yang dimiliki mempunyai kategori sehingga bisa dilakukan proses pelatihan dan pengujian data. Dalam melakukan kategorisasi secara manual ini, data komentar yang dimiliki dibagi menjadi dua sentimen data yaitu komentar yang diisi sentimen Positif dan komentar yang diisi sentimen negatif.

Tabel 3. Contoh Data Training

Komentar	Sentimen
Saya merasa rugi berbelanja menggunakan aplikasi ini, barang yang dipesan tidak sesuai dengan dtg. Pengajuan pengembalian dana ditolak pihak tokped, bukti foto video semua lengkap. Udh sering kejadian, pihak tokped selalu memihak ke penjual	Negative
Payah... benar - benar payah. Ada atau tidaknya koneksi jaringan internet, aplikasi tidak bisa dibuka secara keseluruhan. Slalu aja ada gangguannya. Not recommended for this application. First star for this apps	Negative
Saya sangat senang belanja di sini, mulai dari harganya yg lbih murah dri aplksi lain, barangnya bagus, aman, lengkap dan pstny sangat nyaman blnja di sini	Positive
Pokoknya Cuma di sini yang pasti Gratis ongkir, mau tunai mau transfer pake apa aja. Di tempat lain banyak Rintangannya utk dapetin free ongkir. Terima kasih tokopedia, Semoga Sukses dan jadi tempat belanja online favorite kami.	Positive

4.3 Proses Analisis Data Testing

Proses klasifikasi naïve bayes membutuhkan pengukuran setiap kata yang ada dalam data training. Pengukuran ini dilakukan pada data

training dengan menghitung probabilitas positif dan probabilitas negatif. Hasil dari pengukuran ini untuk selanjutnya digunakan sebagai bahan acuan untuk penentuan sentimen positif dan sentimen negatif data testing. Penentuan sentimen dilakukan dengan cara menghitung probabilitas dokumen data testing dengan merujuk kepada probabilitas data training. Proses ini menggunakan algoritma naïve bayes. Klasifikasi sentimen ini dilakukan secara otomatis oleh aplikasi yang dibuat dengan mengimplementasikan algoritma naïve bayes. Proses ini membandingkan bobot setiap kata pada kata data testing dengan kata data training. Hasilnya setiap dokumen training ini dijumlah bobot kata probabilitas positif dan probabilitas negatifnya, selanjutnya bobot dokumen dibandingkan, apabila bobot dokumen probabilitas positif lebih besar maka hasil sentimen adalah positif. Sebaliknya apabila probabilitas negatif lebih besar maka hasil sentimen adalah negatif.

Tabel 4. Contoh Data Testing

Komentar	Sentimen
Kesel banget ,padahal order barang masih banyak eh di batalin otomatis -_- ,trs ada kendala ga bisa pake saldo ,kapan bisa lagi pake saldo tokped? -_-	?
Tokopedia tolong ya,akun saya jadi ga bisa buka apa apa kalo mau buka disuruh tunggu sebentar download lagi proses,ini sudah 2 hari loh.	?
saya suka sekali dengan aplikasi toko pedia saya alhamdulillah bisa mengikuti pelatihan prakerja di sini tapi untuk pembelannya sudah dan mengikuti tes evaluasi. tapi kenapa untuk melanjutkan ke yg lain knpa masih belum selesai tolong dong toko pedia di bantu mkasih	?
tokopedia aplikasi belanja online yg telah lama berkiprah di bidangnya dan siapa sih yang gak kenal dengan kata tokopedia.... semoga sukses selalu tokopedia	?

Hasil analisis sentimen negatif dan positif data testing pada aplikasi yang dibangun adalah sebagai berikut :



Gambar 5. Hasil Analisis Sentimen

Waktu yang dibutuhkan aplikasi untuk analisis 1.500 data testing adalah 0,4 detik dengan hasil analisis mendapatkan jumlah sentimen 547 (36,47%) positif dan 953 (63,53%) sentimen negatif.

4.4 Hasil Pengukuran Akurasi

Dari hasil pengujian yang dilakukan Rapidminer terhadap 1.500 data komentar pengguna dengan menggunakan proses TF-IDF dan Term Frequency menghasilkan 48 nilai yang sama, Pengujian berdasarkan nilai accuracy, Class Recall dan Class Precision pada analisis sentimen, dengan nilai accuracy performance sebesar 97,13%, dengan nilai precision 1 Sementara pada Class Recall dihasilkan nilai 95,49% (positive class: negative). Dan nilai AUC 0,980.

4.5 Pengolahan data testing klasifikasi sentimen negatif

Dari data testing yang dianalisis oleh aplikasi yang dibangun, ternyata data komentar yang mendapat sentimen negatif sebesar 63,53% lebih besar dari sentimen positif yang hanya mendapat 36,47%. Hal ini ternyata berbanding terbalik dengan jumlah rating yang didapat Aplikasi Tokopedia. Sesuai dengan data yang diperoleh dari Google Playstore, jumlah rating (bintang) 5 adalah sebanyak 1.941.464, bintang 4 sebanyak 262.360, bintang 3 sebanyak 131.180, bintang 2 sebanyak 78.708 dan bintang 1 sebanyak 209.888.

V. PENUTUP

5.1 Kesimpulan

Metode Naive Bayes terbukti dapat menganalisis sentimen secara otomatis. Uji coba dilakukan menggunakan data testing secara realtime menggunakan aplikasi yang dibangun dan rapidminer, setiap kata diklasifikasikan dengan sentimen positif atau negatif. Performa yang dihasilkan dari hasil pengujian yang dilakukan Rapidminer terhadap 1.500 data testing dihasilkan nilai akurasi sebesar 97,13%, dengan nilai precision 1 Sementara pada Class Recall dihasilkan nilai 95,49% (positive class: negative). Dan nilai AUC 0,980. Data komentar yang mendapat sentimen negatif sebesar 63,53% lebih besar dari

sentimen positif yang hanya mendapat 36,47%. Hal ini ternyata berbanding terbalik dengan jumlah rating yang didapat Aplikasi Tokopedia. Penulis mencoba menganalisis dari jumlah kata-kata yang sering muncul di dokumen data komentar kategori negatif dan diurutkan terbanyak terdapat dikata 'tokopedia' muncul sebanyak 580 kali di 294 komentar/dokumen, kemudian 'barang' muncul sebanyak 408 kali di 211 komentar/dokumen. Hal ini ternyata banyak komentar negatif kepada brand/nama tokopedianya itu sendiri. Disusul karena kondisi barang yang tidak sesuai saat barang di terima, barang rusak saat diterima atau barang yang dikirim tidak sesuai dengan pesanan. Berkaitan dengan komentar negatif terhadap aplikasi, katakata 'aplikasi' muncul sebanyak 254 kali di 153 dokumen.

5.2 Saran

Pada penelitian selanjutnya penulis menyarankan menambahkan beberapa metode lain misalnya dengan seleksi fitur information gain dengan tujuan untuk lebih mengurangi bias agar dapat meningkatkan akurasi algoritma naive bayes secara signifikan. Adanya penambahan data kamus bahasa gaul pada proses stopword, karena pada komentar pengguna di Google Playstore masih banyak penggunaan bahasa-bahasa yang kurang baku.

DAFTAR PUSTAKA

- [1] Indah Rahmayani, "Indonesia Raksasa Teknologi Digital Asia", 02 Oktober 2015. [Online]. Tersedia :https://kominformo.go.id/content/detail/6095/indonesia-raksasa-teknologidigital-asia/0/sorotan_media, [diakses : 5 Juni 2020]
- [2] -----, "Smartphone Rakyat Indonesia", 11 Januari 2017. [Online]. Tersedia :<https://www.ristekbrin.go.id/kabar/smartphone-rakyat-indonesia> [diakses : 5 Juni 2020]
- [3] Avi Tejo Bhaskoro, "Tokopedia semakin mudah diakses dengan aplikasi mobile untuk Android", 12 Juni 2014. [Online]. Tersedia :<https://dailysocial.id/post/tokopedia-android>. [diakses : 5 Juni 2020]
- [4] Agung Pratnyawan, "Ini situs belanja online terpopuler di Indonesia sepanjang 2019", 21 Januari 2020. [Online]. Tersedia :<https://www.suara.com/tekno/2020/01/21/160327/ini-situs-belanja-online-terpopuler-di-indonesia-sepanjang-2019?page=all> [Diakses ; 05 Juni 2020].
- [5] -----, "Survei tokopedia dan Shope", [Online]. Tersedia :<https://inet.detik.com/business/d-4872404/survei-tokopedia-dan-shopee-adu-kuat-mendominasi> [diakses : 5 Juni 2020]
- [6] Y.Febrianti., I. Indriati, I., A. Widodo, "Analisis Sentimen Pada Ulasan "Lazada" Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) dengan Perbaikan Kata Menggunakan Jaro Winkler Distance". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, no. 10, p. 3689-3698, feb. 2018.
- [7] Santoso, Valonia, Virginia, Gloria, Lukito, Yuan, "Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen dengan Metode Support Vector Machine", Jurnal Transformatika. 14. 72. 10.26623/transformatika.v14i2.439, 2017.
- [8] S.A.Aputra, D. Rosiyadi, W.Gata, S.M. Husain, "Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization", Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 3(3), 377 – 382, 2019.
- [9] Mubarak, M. Syahrul, A.Wijaya, Aldhi, M.Dwi, "Aspect-Based Sentiment Analysis to Review Products Using Naïve Bayes", In: AIP Conference Proceedings. AIP Publishing LLC, 2017. p. 020060.
- [10] D. Gustian, I. Suciati, S. Saepudin, "Sistem Pakar dengan Algoritma Naive Bayes untuk Prediksi Hasil Produksi Ayam Broiler Plasma (Studi Kasus : PT.Sekawan Sinar Surya), JURNAL RISET SISTEM INFORMASI DAN TEKNOLOGI INFORMASI (JURSISTEKNI), VOL 1 NO 3 (2019) .
- [11] Bayhaqy, Achmad, et al, "Sentiment Analysis About e-Commerce From Tweets using Decision Tree, K-Nearest Neighbor, And Naïve Bayes", 2018 international conference on orange technologies (ICOT). IEEE, 2018. p. 1-6.

- [12] F. Ratnawati, “ Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter”, JURNAL INOVTEK POLBENG - SERI INFORMATIKA, VOL. 3, NO. 1 , JUNI 2018.