



Analisis Sentiment Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naïve Bayes Classification

Alisia Silver Stone, Fathoni*

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹09031381823084@unsri.ac.id, ^{2,*}fathoni@unsri.ac.id

Email Penulis Korespondensi: alisiasilverstone101@gmail.com

Abstrak—Analisis sentimen atau opinion mining merupakan analisis yang bertujuan untuk melihat sentimen masyarakat atau kelompok mengenai entitas tertentu. Sentimen yang diekspresikan masyarakat bisa dalam bentuk positif, negatif dan netral. Salah satu media yang bisa diberikan opini oleh masyarakat adalah pada aplikasi e-commerce yaitu aplikasi shopee, shopee memiliki fitur komentar atau penilaian terhadap produk yang telah dibeli. Toko yang dijadikan sampel penelitian ialah toko online shop amreta, berdasarkan hasil identifikasi permasalahan didapatkan bahwa faktanya banyak komentar yang tidak sesuai dengan bintang yang diberikan sehingga dapat dikatakan bahwa rating tersebut belum bisa mewakili bahwa kinerja toko sudah baik atau belum. Oleh karena itu untuk meningkatkan profit kerja toko, pihak amreta masih perlu melakukan evaluasi toko. Dalam melakukan evaluasi, pihak toko membutuhkan pengklasifikasian terhadap komentar positif, negatif ataupun netral. Analisis sentimen pelanggan terhadap penilaian produk pada toko online shop amreta menggunakan metode naive bayes classification. Penggunaan data uji pada penelitian ini didapat dari sentiment konsumen online shop amreta sebanyak 2014 data, kemudian data diolah melalui proses cleaning data menghasilkan data bersih sebesar 1899 data. Selanjutnya proses preprocessing data yang terbagi menjadi 3 tahap yaitu Tokenize Data, Transform case dan Stopword removal. Setelah itu analisis data tahap pelabelan otomatis menggunakan Text Vectorize dari proses tersebut didapatkan pembagian data menjadi 3 kelompok data 71% atau 1343 data positif, 3% atau 52 data negatif dan 26% atau 504 data netral. selanjutnya diolah dengan menggunakan tools rapidminer sedangkan untuk operator dalam bentuk algoritma menggunakan model sentiment Naïve Bayes Classification melalui perhitungan otomatis. Hasil penelitian dapat disimpulkan bahwa data uji yang didapat memiliki tingkat Akurasi sebesar 97.16% dengan menggunakan model Naïve Bayes Classification.

Kata Kunci: Sentiment; Online Shop; Shopee; Naïve Bayes; E-Commerce

Abstract—Sentiment analysis or opinion mining is an analysis that aims to see the sentiment of people or groups regarding certain entities. The sentiments expressed by society can be in positive, negative and neutral form. One media that can be given an opinion by the public is in the e-commerce application, namely the shopee application, shopee has a comment or assessment feature on the product that has been purchased. Toko which was used as a sample of research is an amreta online shop store, based on the results of the identification of the problem, it was found that the fact was that many comments did not match the stars given so it can be said that the rating cannot represent that the store's performance is good or not. Therefore, to increase the profit of shop work, the amreta still needs to evaluate the store. In conducting an evaluation, the store needs to classify positive, negative or neutral comments. Analysis of customer sentiment towards product assessments in amreta online shop stores using the naive bayes classification method. The use of test data in this study was obtained from the sentiment of amreta online shop consumers as much as 2014 data, then the data was processed through the data cleaning process resulting in net data of 1899 data. Furthermore, the data preprocessing process is divided into 3 stages, namely Tokenize Data, Transform case and Stopword removal. After that, the analysis of data for the automatic labeling stage using Text Vectorize from the process obtained data division into 3 data groups of 71% or 1343 positive data, 3% or 52 negative data and 26% or 504 neutral data. furthermore, it is processed using rapidminer tools while for operators in the form of algorithms using the Sentiment Naïve Bayes Classification model through automatic calculations. The results of the study can be concluded that the test data obtained have an accuracy level of 97.16% using the Naïve Bayes Classification model.

Keywords: Sentiment; Online Shop; Shopee; Naïve Bayes; E-Commerce

1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat dan mempermudah masyarakat dalam memberikan pendapat dan opini dengan berbagai cara seperti menyampaikan langsung atau menggunakan media sosial, peningkatan kecepatan internet serta perluasan cakupan area atau jangkauan internet di Indonesia menjadi faktor utama masyarakat mudah dalam mengakses serta dalam menyampaikan opini di media online. Analisis sentimen atau opinion mining merupakan analisis yang bertujuan untuk melihat sentimen masyarakat atau kelompok mengenai entitas tertentu. Sentimen yang diekspresikan masyarakat bisa dalam bentuk positif atau negatif [1].

Salah satu media yang bisa diberikan opini oleh masyarakat adalah E-Commerce, sebut saja salah satunya Shopee merupakan salah satu pusat belanja online yang menawarkan berbagai macam jenis produk menjamin kenyamanan konsumen ketika konsumen browsing produk yang sedang dicari dan juga menjamin opsi pembayaran yang aman. Sebagai perusahaan yang menjadi wadah salah satu pelaku bisnis terbesar di Indonesia, Shoppe juga memiliki fitur komentar atau penilaian terhadap produk yang telah dibeli. Salah satu toko yang dijadikan sampel peneliti ialah Toko Online Shop Amreta. Amreta merupakan salah satu pelaku bisnis yang ada di online shop shopee, dengan memiliki pengikut lebih dari 7.8 ribu pengikut serta memiliki rating penjualan 4.9 dari 5.0, maka pihak amreta pasti memiliki banyak opini atau pendapat oleh masyarakat yang sudah berbelanja pada online shop tersebut, opini masyarakat yang berupa review tersebut juga disertai review menggunakan star dimana penilaian menggunakan star berperan sebagai pengukur kepuasan customer dalam



berbelanja di toko Online Shop Amreta. Dari review yang peneliti amati di toko Amreta pada platform shopee. Online Shop Amreta di aplikasi shopee sudah memiliki rating 4.9 tetapi faktanya setelah ditinjau kembali banyak komentar yang belum sesuai dengan bintang diberikan, sehingga dapat diasumsikan bahwa rating tersebut tidak dapat mewakili bahwa kinerja toko tersebut sudah baik. Ketidaksesuaian tersebut dapat dilihat secara kasat mata dikolom komentar yang menilai bahwa layanan di Amreta sudah baik tetapi memberikan nilai bintang yang sedikit ataupun sebaliknya. Oleh karena itu Online Shop Amreta masih perlu melakukan evaluasi terhadap kinerja toko agar meningkatkan profit dan menjaga citranya. Oleh karena itu penulis melakukan penelitian analisis sentimen pada toko amreta untuk membantu pihak toko dalam memberikan acuan pada saat mereka akan melakukan evaluasi. Tujuan dari penelitian ini ialah melakukan klasifikasi pada sentimen masyarakat dengan mengimplementasi menggunakan metode *Naïve Bayes Classification*, metode naïve bayes adalah untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan [2] selanjutnya ialah mengukur tingkat akurasi dengan menerapkan metode naïve bayes dalam mengelola data sentimen.

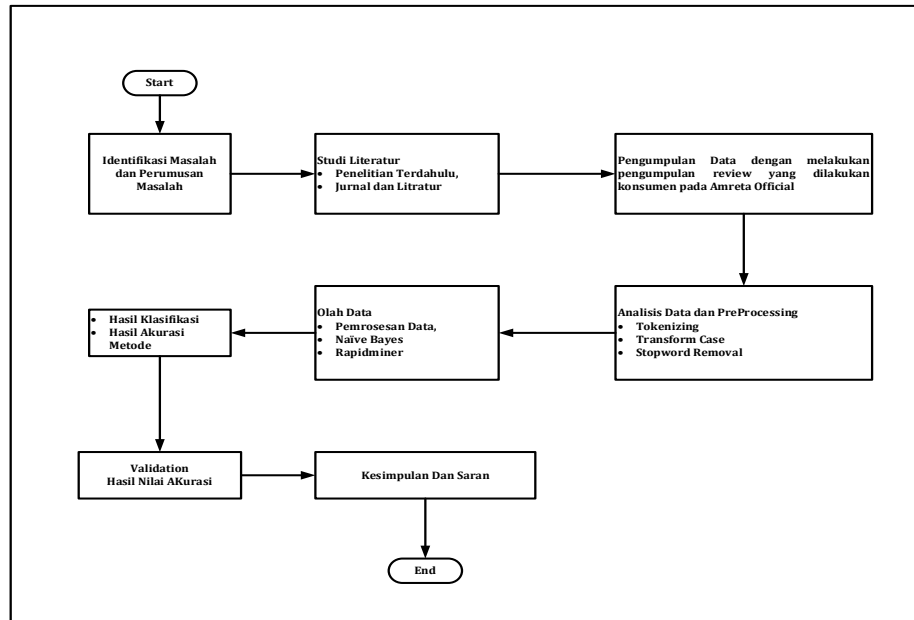
Analisis sentimen juga merupakan merupakan topik riset yang penting dan sedang marak dilakukan saat ini. Opinion mining merupakan cabang penelitian dari text mining. Fokus dari opinion mining adalah melakukan analisis opini dari suatu dokumen teks. Terdapat tiga buah subproses dari opinion mining yaitu, document subjectivity, opinion orientation dan target detection. Dalam dunia bisnis, opinion mining banyak digunakan untuk menganalisis secara otomatis opini pelanggan tentang produk dan pelayanannya [5]. Adapun penelitian terkait yang dijadikan peneliti sebagai referensi untuk penelitian yang dijalankan yaitu penelitian Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter, Berdasarkan hasil eksperimen, analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90 % dengan rincian nilai precision 92%, recall 90% dan f-measure 90% [2]. Sedangkan pada penelitian yang berjudul Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion, Hasil pengujian akhir menghasilkan rata-rata akurasi sebesar 96,25% dengan parameter nilai learning rate = 0,001, nilai lambda = 0,1, nilai complexity = 0,01 dan iterasi maksimal adalah 50. Berdasarkan hasil pengujian, diperoleh akurasi metode Support Vector Machine dan Query Expansion sebesar 96,25% dan akurasi menggunakan metode Support Vector Machine tanpa Query Expansion sebesar 94,75% [4].

Berdasarkan pernyataan tersebut peneliti ingin melakukan penelitian analisis sentimen terhadap penilaian produk oleh masyarakat pada Toko Online Shop Amreta, alasan peneliti ingin melakukan penelitian analisis sentimen pada Toko Online Shop Amreta ialah untuk menghindari ketidaksesuaian antara bintang dan komentar yang diberikan dengan adanya ketidaksesuaian dari komentar dan bintang yang diberikan pelanggan pada kolom penilaian atau ulasan membuat pihak toko amreta ragu dalam menentukan apakah penilaian tersebut bersifat positif, negatif atau netral, sehingga untuk mengetahui pengkelompokan dari pendapat konsumen terhadap penilaian suatu produk ketika sudah melakukan pembelian produk tersebut pada Toko Online Shop Amreta apakah termasuk sentimen yang positif atau termasuk sentimen negatif melalui proses pengolahan data opini atau lebih dikenal text mining. Untuk mendukung dan membantu proses penelitian, peneliti menggunakan metode naïve bayes. Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Naïve Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Untuk menerapkan metode ini peneliti menggunakan alat bantu yaitu rapidminer yang merupakan salah satu tools yang digunakan untuk pengolahan data seperti data mining [4]. Dari latar belakang diatas, maka akan dibahas lebih lanjut dalam tugas akhir dengan judul "Analisis Sentimen Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naive Bayes Classification"

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metode naïve bayes dalam proses pembobotan teks. Teks yang sudah melalui proses preprocessing akan dilakukan perhitungan untuk menentukan review terhadap sebuah produk ataupun sentimen oleh konsumen itu sendiri. Perhitungan naïve bayes digunakan untuk menentukan nilai teks tersebut dalam topik dan sentimen. Kata hasil pembobotan data yang telah dihitung fitur kemunculanannya yang digunakan sebagai input awal. Setelah didapatkan fitur kemunculan kata maka akan ditentukan nilai bobot kata terhadap topik dan bobot kata terhadap sentimen. Nilai bobot pada topik dan sentimen yang diambil adalah nilai tertinggi dari beberapa kemungkinan yang telah dihitung. Berikut merupakan proses mengukur akurasi menggunakan metode naïve bayes pada rapidminer:

**Gambar 1.** Tahapan Penelitian

Deskripsi penjelasan dari Metode Tahapan penelitian Gambar 1 diatas:

1. Proses pertama kali yang dilakukan ialah mengidentifikasi masalah serta perumusan masalah terhadap penelitian Analisis Sentimen Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naive Bayes Classification.
2. Melakukan pencarian referensi terhadap penelitian terdahulu atau studi literatur dengan mencari referensi penelitian sejenis atau penelitian terdahulu.
3. Pengumpulan Data dengan melakukan pengumpulan review yang dilakukan konsumen pada Amreta Official,
4. Melakukan analisis data terhadap data review yang sudah dikumpulkan dan melakukan proses preprocessing yang meliputi proses tokenizing, transform case, dan stopwords removal data pada Analisis Sentimen Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naive Bayes Classification”.
5. Proses selanjutnya melakukan pemrosesan data yang dimana data yang sudah dikumpulkan akan diolah terlebih dahulu sebelum dilakukan klasifikasi dengan model yang telah ditentukan yaitu algoritma Naive Bayes Classification.
6. Olah data dengan tools rapidminer dengan memasukkan operator-operator yang sudah disediakan serta melakukan pelabelan data dari analisis sentimen yang sudah didapat.
7. Hasil dari olah data pada Analisis Sentimen Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naive Bayes Classification akan didapat hasil akurasi dari analisis yang didapat.
8. Setelah olah data dijalankan maka akan didapat hasil klasifikasi dan akurasi terhadap review produk yang sudah dibeli oleh konsumen pada akun belanja online shopee Amreta Official.
9. Kesimpulan dan saran atau hasil akhir dari penelitian Analisis Sentimen Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naive Bayes Classification.

2.2 Metode Pengumpulan Data

Tahap pengumpulan data merupakan tahap yang penting dari suatu penelitian. Data yang digunakan untuk penelitian harus benar-benar akurat dan jelas sumbernya sesuai dengan kebutuhan penelitian Analisis Sentimen Pelanggan Terhadap Penilaian Produk Pada Toko Online Shop Amreta Menggunakan Metode Naive Bayes Classification. Pengumpulan data sebagai berikut.

2.3 Cleaning Data

Pada Tahap ini merupakan tahap dimana teks inputan yang berasal dari review konsumen pada Amreta Official di Shopee pertama diproses. Cleaning Data merupakan proses penyamaan huruf. Pada data review terdapat kata yang berhuruf besar dan terdapat teks yang berhuruf kecil. Perbedaan bentuk huruf pada data review tersebut akan mengganggu proses selanjutnya maka dari itu diperlukan proses cleaning data untuk merubah *review menjadi huruf kecil semua*.

2.4 Tokenizing

Tahap tokenizing merupakan tahap yang berfungsi untuk memisahkan antar kata yang ada didalam teks review. Proses ini menggunakan bantuan karakter spasi dalam pemisahan kata-nya. Pada sebuah teks review terdiri dari



beberapa kata yang saling terhubung dan dipisahkan dengan spasi. Untuk mempermudah dalam pemrosesan sebuah teks maka masing-masing kata dalam sebuah kalimat tersebut harus dipisahkan.

2.5 Stemming

Kalimat yang sudah melewati proses pemisahan kata melalui tokenizing selanjutnya akan dilakukan proses tranform case. Proses dimana akan menghapus kata-kata yang tidak penting atau tidak memiliki makna seperti kata sambung. Pada proses tranform case dibutuhkan sebuah database yang menyimpan daftar stopword yang digunakan sebagai acuan dalam penghapusan kata-kata hasil proses tokenizing digunakan sebagai inputan pada proses tranform case.

Setelah didapatkan kata tersebut maka tiap kata akan dilakukan pengecekan kata dengan data pembanding yakni kata yg ada dalam stopwords. Jika kata tersebut termasuk dalam kumpulan kata stopwords maka kata tersebut akan dihapus. Setelah proses pengecekan kata stopwords dilakukan maka dilakukan pengecekan kembali untuk jumlah kata. Jika dalam pengecekan ditemukan jumlah kata lebih dari satu maka kata tersebut akan dilakukan pengecekan stopwords kembali.

2.6 Stopword Removal

Proses setelah filtering adalah proses stemming. Proses stemming merupakan proses pencarian kata dasar terhadap sebuah kata yang telah dilakukan proses filtering. Pada proses ini kata akan diubah menjadi kata dasar sehingga kata yang memiliki imbuhan depan atau imbuhan belakang akan dihapus dan disesuaikan kata dasarnya.

3. HASIL DAN PEMBAHASAN

Pada penelitian yang berjudul “Analisis Sentiment Pelanggan Terhadap Penilaian Produk Pada Toko *Online Shop Amreta Menggunakan Metode Naive Bayes Classification*”, pada penelitian ini bertujuan untuk mengetahui klasifikasi dari sentimen atau opini penilaian produk pada toko *online shop amreta* dan yang kedua adalah mengetahui tingkat akurasi menggunakan metode *naive bayes classification* dalam pengolahan data. Penelitian tersebut di lakukan karna adanya ketidak sinkron antara penilaian menggunakan *star* yang diberikan dengan kalimat opini review yang di berikan, adapun penelitian terkait yang dijadikan referensi adalah penelitian penulis yang bernama Ratnawati pada tahun 2018 yang berjudul “Implementasi Algoriitama *Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter*” analisis tersebut menghasilkan 90% tingkat keakurasiannya.

Untuk memperoleh data penelitian mengambil data dari *review* pada akun Amreta oleh konsumen melalui aplikasi *shopee* data tersebut dikumpulkan kedalam file *excel* sampel data review yang terkumpul adalah sebesar 2014 sentimen data terdapat pada Gambar 2.

	A
1996	Pengiriman cepat, bahan sesuai harga. Sesuaiiii lah yaaaaa...
1997	Bahan Tasnya bagus... dalamnya ada kain lapisan, hrg terjangkau.. pengiriman cpt. Tp respon agak lambat...
1998	hunting outer berujung tas sekalian karena mau dipake kuliah eheheheheheheh. bahan dan jahitannya bagus, dengan harga segini keliatan premium banget! tapi ini
1999	Barang sudah diterima dengan baik, pengemasan dan pengiriman cepet banget, packaging aman, dan tasnya bagussss, keren sih ini, harganya murah tapi kualitas
2000	Pengiriman standar, bahannya bagus, tapi kenapa ga ada tulisan inisialnya di tas, padahal aku sudah kasih di note nya :-(huhu
2001	bahann nyaa tebeell, bagusss, jataaan nyaa jg rapihhhhh
2002	Tasnya cantik banget, bahannya bagus, model simple, gak begitu besar tapi muat banyak banget
2003	bagus banget bahannya, packingnya juga rapi jadi ga perlu takut barangnya rusak.
2004	Bagusss bisa dipake buat kuliah hihhihi thankyouuu 🥰🥰🥰
2005	Bagus bgt tasnya tebel kaya mahal loh wkwk makasih yaa seller
2006	Bgus sih tpi di packingnya kurang mntap ada garis lipatan di sisi tas
2007	Bahan bagus
2008	Bagusss banget, muat laptop, bahannya oke banget 🥰
2009	Bahannya bagusss banget, keliatan mahal and elegan gitu dengan harga ga bikin dompet tetep ada isinya 🥰
2010	Terimakasih paket sudah sampai
2011	Bete banget jamuran :(
2012	baguuuus tasnya muat banyak padahal keliatannya kecil 🥰
2013	Bagusssssss bangetttt sukaaaa 🥰🥰🥰🥰
2014	bahan ny bagus, barang ny bagus

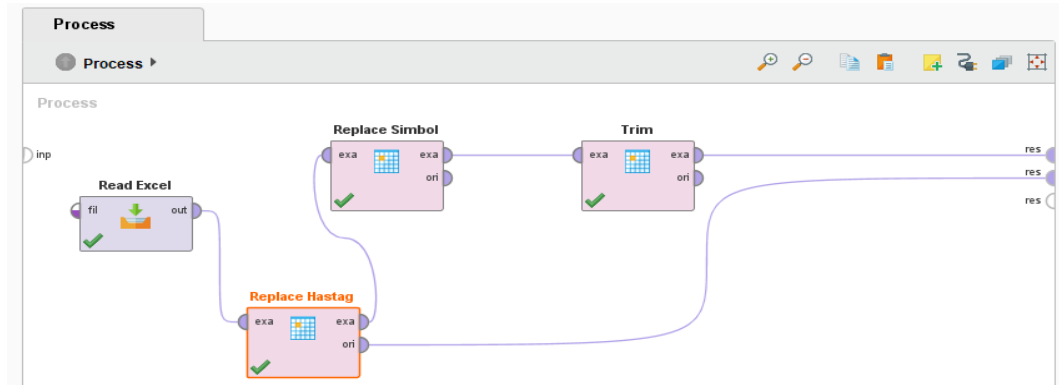
Gambar 2. Sampel Data Review

Cleaning Data tahap ini merupakan tahap dimana teks inputan yang berasal dari review konsumen pada Amreta Official, tahap ini merupakan proses penyamaan huruf Perbedaan bentuk huruf pada data *review* tersebut akan mengganggu proses selanjutnya maka dari itu diperlukan proses *cleaning data* untuk merubah *review* menjadi huruf kecil, langkah pertama yang dilakukan adalah penerapan operator semua dapat dilihat pada Gambar 3.

Row No.	Review
1	Bagusss tasnya, bahannya juga lumayan solid gitu, cuman aga kurang sreik sama dalamannya yg warna merah hehe
2	Tokonya baik bgtt, sempat ketuker tasnya yg dikirim dan bisa
3	Ga ekspetasi bakal bagus kaya gini, ternyata bagus bgt woi 🥰 Tapi emang produk amreta tuh sebagai itu aku udah punya 3 tas nya d...
4	Buy 1 get 2 🥰🥰🥰 bgttt asmara bagus bgttt gatau lg aku punya agni buat seharizilnangis bgttt bagus smua aduhhhhhhhhh
5	Udah punya ananta 1 yg warna merah masih awet juga, sekarang beli lagi yg ada emboss nya hini worth the price
6	Cakep banget astagaaa ayunda bener bener se waw itu ya, km wni hitam jd masuk pakai apa aja huhu nyssel juga cmn beli 1
7	Tasnya bagus banget, sukaa, cepet sampainya, makasih ya kak, sukses selalu yaa
8	Sabar yaa yang enggak kabegian Agni :))))))
9	Baguuuuuuuu bangeeettt pertama kali beli disini baguuuuuuuuu dahh luw banyak2
10	Luv bangetttt Semoga awet tasnyaaaaa 🥰🥰🥰🥰 sering promosi seller 🥰🥰
11	rm so grateful to have this bag 🥰🥰 detailnya parah sii. Makin lama makin cakep Pembelian kedua nih 🥰
12	JUARA ULUM III
13	Paketnya cepet banget sampe, bagus tas nya asmara bahan nya lembut gitu warna aslinya ternyata lebih cantik... Yg alexa juga bagus ba...
14	Bagusss bangetttt Sweater nya juga gak panas Asmara jugaas soft banget kullnyaas Sukses terus amreta 🥰
15	Bagus bangettttt amreta the best 🥰 seallu pesan kardigan disini. Uдах punya ayunda outer 4 dan Ayunda 2 🥰 Thank you amreta 🥰

Gambar 3. Data Mentah Sebelum Proses *Cleaning Data*

Selanjutnya data di proses untuk menghilangkan white space yang tidak perlu menggunakan operator Read Excel, Replace Hastag, Replace Simbol dan operator trim yang terdapat pada Gambar 4..



Gambar 4. Penerapan operator *Trim*

Dengan menghasilkan output dataset sebagai berikut terdapat pada Gambar 5.

Row No.	Review
1	The position of the example in the (filtered) view on the example table.
2	Baguss tasnya bahannya juga lumayan solid gitu cuman aga kurang sre sama dalemannya yg warna merah hehe
3	Tokonya baik bgttsempet ketuker tasnya yg dikirim dan bisa
4	Ga ekspetasi bakal bagus kaya gini ternyata bagus bgt woi Tapi emang produk amreta tuh sebgas itu aku udah punya 3 tas nya da...
5	Buy 1 get 2cryyy bgtttt asmara bagus bgtttt gatau lg aku punya agnii buat sehari2nangis bgtttt bagus smua aduhhhhhhhhh
6	Udah punya ananta 1 yg wrna merah masih awet juga sekarang beli lagi yg ada emboss nya hihi worth the price
7	Cakep banget astagaaa ayunda bener bener se waw itu ya krn wrn hitam jd masuk pakai apa aja huhu nyesel juga cmn beli 1
8	Tasnya bagus banget sukaa cepet sampainyaa makasih ya kak sukses selalu yaa
9	Sabar yaa yang enggak kebagian Agni
10	Baguuuuuuuu bangeeettttt pertama kali beli disiniiii baguuuuuuuu dahh luvv banyak2
11	Luv bangeet Semoga awet tasnyaaaaaa sering sering promonya seller
12	Im so grateful to have this bag detailnya parah sii Makin lama makin cakep Pembelian kedua nih
13	JUARAA UMUMM
14	Paketnya cepet banget sampe bagus tas nya asmara bahan nya lembut gitu warna aslinya ternyata lebih cantikk Yg alexa juga bagu...
15	Baguss bangetttt Sweater nyaa juga gak panass Asmara jugaaa soft banget kulitnyaa Sukses teruss amretaa
16	Bagus bangettt amreta the best☑ sealu pesan kardigan disini Uдах punya ayunda outer 4dan Ayunya 2 Thank you amreta☑

Gambar 5. Hasil Penerapan Operator *Trim*

Selanjutnya penulis menerapkan Operator *Remove Duplicate* ialah untuk menghilangkan review duplikat setelah penerapan operator remove duplicate maka menghasilkan 1899 dataset dari 2014 dataset, berikut hasil penerapan dari operator *remove duplicate* dapat dilihat pada Tabel 1. berikut ini.

Tabel 1. Hasil Sebelum dan Sesudah Penerapan *Remove Duplicate*

Sebelum	Filter (2,014 / 2,014 examples): all
Sesudah	Filter (1,899 / 1,899 examples): all

Setelah rangkaian penerapan operator tersebut maka di hasilkan hasil output dari seluruh rangkaian proses *cleaning data* yang terdapat pada Gambar 6.

Review	A
1 Review	
2 Baguss tasnya bahannya juga lumayan solid gitu cuman aga kurang sre sama dalemannya yg warna merah hehe	
3 Tokonya baik bgttsempet ketuker tasnya yg dikirim dan bisa	
4 Ga ekspetasi bakal bagus kaya gini ternyata bagus bgt woi Tapi emang produk amreta tuh sebgas itu aku udah punya 3 tas nya da...	
5 Buy 1 get 2cryyy bgtttt asmara bagus bgtttt gatau lg aku punya agnii buat sehari2nangis bgtttt bagus smua aduhhhhhhhhh	
6 Uдах punya ananta 1 yg wrna merah masih awet juga sekarang beli lagi yg ada emboss nya hihi worth the price	
7 Cakep banget astagaaa ayunda bener bener se waw itu ya krn wrn hitam jd masuk pakai apa aja huhu nyesel juga cmn beli 1	
8 Tasnya bagus banget sukaa cepet sampainyaa makasih ya kak sukses selalu yaa	
9 Sabar yaa yang enggak kebagian Agni	
10 Baguuuuuuuu bangeeettttt pertama kali beli disiniiii baguuuuuuuu dahh luvv banyak2	
11 Luv bangeet Semoga awet tasnyaaaaaa sering sering promonya seller	
12 Im so grateful to have this bag detailnya parah sii Makin lama makin cakep Pembelian kedua nih	
13 JUARAA UMUMM	
14 Paketnya cepet banget sampe bagus tas nya asmara bahan nya lembut gitu warna aslinya ternyata lebih cantikk Yg alexa juga bagu...	
15 Baguss bangetttt Sweater nyaa juga gak panass Asmara jugaaa soft banget kulitnyaa Sukses teruss amretaa	
16 Bagus bangettt amreta the best☑ sealu pesan kardigan disini Uдах punya ayunda outer 4dan Ayunya 2 Thank you amreta☑	
17 Wawwww ayuuu pink baguss banget udah punya yang cream dan beli lagi karna emang sebgas itu	
18 nagh punya tas dr amreta Pertama punya puan dan sekarang punya sling bag dr amreta Bagus banget kualitasnya Ini pun free anastasya	
19 SUKAK BGT bahannya baguss bukan main dapet free ayunda juga the best pokoknya	
20 Luv banget Sukaanaa lancar rezeki buat seller yaaaa Biar bisa adain promo promo lainnya aamiin	
21 Definisi buy 1 get 1 yang nyata terimakasih kak semoga usahanya berkah selalu dan sukses selalu	
22 Suka banget ini pembelian yang ke-2 dan enggak pernah mengecewakan Bahannya bagus apalagi ini beli waku 1 gratis 1	
23 Pembelian telah disembuyikan karena mengandung spam konten yang kurang pantas. Keis akan ditarik kembali dari penjualan ini.	

Gambar 6. Hasil *Cleaning Data*



Langkah selanjutnya yang dilakukan penulis ialah tahap *preprocessing* yang dibagi menjadi 3 proses diantaranya ialah:

1. Tokenize Data

Pada proses *tokenize* yang dimana merupakan proses memecah *review* menjadi kata-kata atau memecah kalimat kedalam beberapa *text*. Proses ini menggunakan ekstensi *text preprocessing* yang ada pada *rapidminer Transform Case*, berikut ini sampel data hasil pengolahan data *tokenize* yang diambil beberapa sampel untuk dapat dilihat terdapat pada Gambar 7. seperti berikut ini.

Word	Attribute Name	Total Occurrences ↓	Document Occurrences
bagus	bagus	661	596
banget	banget	465	399
nya	nya	445	339
dan	dan	347	295
juga	juga	333	293
harga	harga	307	291
tas	tas	268	234
sesuai	sesuai	266	250
bgt	bgt	255	213
yg	yg	248	203

Gambar 7. Hasil Proses *Tokenize*

Setelah penerapan operator *tokenize* maka menghasilkan data seperti Gambar 4.28 diatas, pada hasil data diatas merupakan 10 kata yang paling sering digunakan oleh konsumen untuk melakukan review terhadap produk dari amreta *official*, terlihat kata “bagus” menjadi kata yang sering digunakan yaitu sebanyak 661 kali.

Tabel 2. Proses Perubahan Data Pada *Tokenize*

Sentimen	
Sebelum Proses <i>Tokenize</i>	Sesudah Proses <i>Tokenize</i>
Ok lah dgn hrga stgu tp kaku bgt buat hp msk kestu jd kek sempit bgt	Ok, lah, dgn, hrga, stgu, tp, kaku, bgt, buat, hp, msk, kestu, jd, kek, sempit, bgt,
kualitas produk baik dan original kecepatan pengiriman bagus	kualitas, produk, baik, dan, original, kecepatan, pengiriman, bagus
Barang bagus Mendarat dengan selamat tmksh Seller	Barang, bagus, Mendarat, dengan, selamat, tmksh, Seller
Tasnya bagus dan bahanny jg bagus banget Suka bangettt	Tasnya, bagus, dan, bahanny, jg, bagus, banget, Suka, bangettt

Pada proses diatas terlihat setiap sentimen dipisahkan satu persatu kata-kata dari sebuah kalimat, proses ini dinamakan pemecahan kalimat menjadi kata-kata.

2. Transform case

Berikut ini merupakan data setelah proses transform case yang dilakukan terdapat pada Gambar 8. seperti berikut ini:

Word	Attribute Name	Total Occurrences ↓	Docum...
bagus	bagus	871	779
banget	banget	467	400
nya	nya	449	342
dan	dan	364	307
tas	tas	347	302
harga	harga	346	327
juga	juga	336	296
pengiriman	pengiriman	323	320
sesuai	sesuai	297	280
tasnya	tasnya	291	273

Gambar 8. Data Setelah *Transform Cases*



Terlihat pada Gambar yang sebelumnya kata-kata terbanyak adalah “bagus” memiliki 661 kali penggunaan kata, maka setelah proses *transform cases* menjadi 871, hal tersebut dikarenakan adanya kata “Bagus” yang menggunakan huruf kapital, sehingga ketika menggunakan operator *transform cases* kata-kata tersebut menjadi “bagus”. Berikut ini merupakan perbedaan tabel jumlah kata-kata ketika menggunakan operator *tokenize* terdapat pada Tabel 3.

Tabel 3. Perbedaan Jumlah Kata

<i>Tokenize</i>	4755 regural word
<i>Transform Case</i>	4113 regural word

Berikut ini merupakan contoh perubahan data pada proses transform case yang dapat dilihat pada Tabel 4. berikut ini.

Tabel 4. Proses Sesudah Dan Sebelum *Tokenize* Data

Sentimen	
<i>Sebelum Proses Tokenize</i>	<i>Sesudah Proses Tokenize</i>
Ok lah dgn hrga stgu tp kaku bgt buat hp msk kestu jd kek sempit bgt kualitas produk baik dan original kecepatan pengiriman bagus BARANG BAGUS MENDARAT DENGAN SELAMAT TMKSH SELLER Tasnya bagus dan BAHANNY jg bagus banget Suka bangettt	ok, lah, dgn, hrga, stgu, tp, kaku, bgt, buat, hp, msk, kestu, jd, kek, sempit, bgt, kualitas, produk, baik, dan, original, kecepatan, pengiriman, bagus barang, bagus, mendarat, dengan, selamat, tmksh, seller tasnya, bagus, dan, bahanny, jg, bagus, banget, suka, bangettt

Berdasarkan hasil diatas menghasilkan perubahan data dari kata-kata yang awalnya mengandung unsur huruf *capital* atau *uppercase* berubah menjadi kata-kata dengan huruf kecil atau *lowercase*.

3. Stopword Removal

Berikut ini merupakan perubahan data dari proses *stopword removal* menghilangkan kata-kata yang tidak perlu seperti Tabel 5. data berikut ini.

Tabel 5. Proses Data *Stopword Removal*

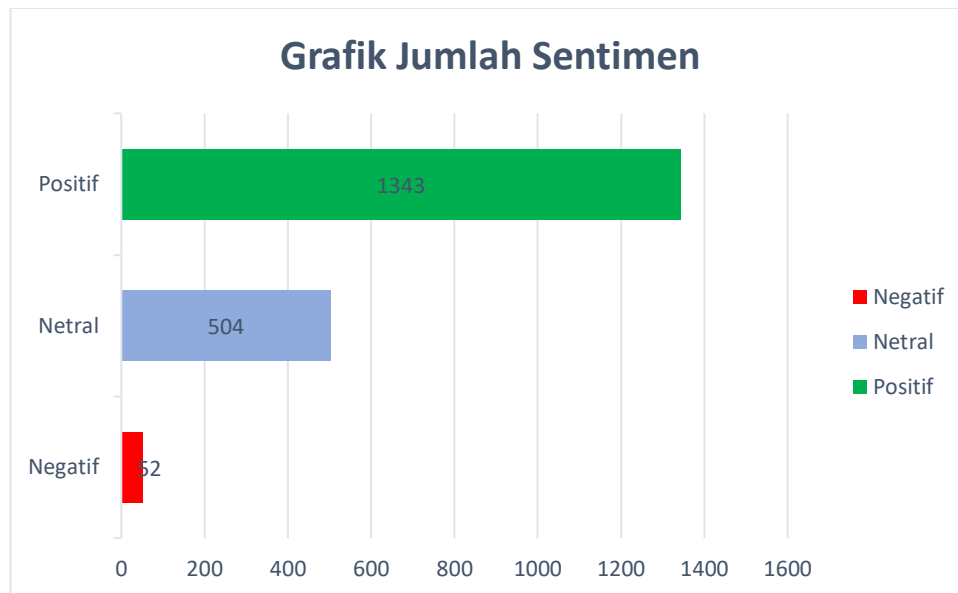
Kata yang Dihilangkan
ada
adalah
adanya
adapun
agak
agaknya
agar
akan
akankah
akhir
.....

Tahap selanjutnya adalah proses pelabelan menggunakan *Text Vectorize* dalam menentukan sentimen analisis pada dataset pada hasil text vectorization akan membaca sentimen yang ada dalam file excel secara otomatis. Dari penggunaan text vectorization di dapatkan Tabel. Merupakan hasil pembagian menjadi 3 kelompok. Dari jumlah 2014 *review* diolah lagi untuk menghilangkan duplikat dan menghasilkan 1899 *review*.

Tabel 7. Pembagian Data

No	Keterangan	Jumlah
1	Negatif	52
2	Netral	504
3	Positif	1343
Total		1899

Sentiment terhadap toko Amreta yang berisi kalimat negatif sebesar 52 atau 3%, lalu sentimen netral sebanyak 504 data atau 26%, serta sentimen positif memiliki jumlah yang paling besar yaitu 1343 atau 71%. Hasil tersebut membuktikan bahwa sebanyak 71% pelanggan toko Amreta memberikan positif terhadap produk dari toko *Online Shop* tersebut. Adapun grafik persentase data sentimen terhadap toko *Online Shop* Amreta dapat dilihat pada Gambar berikut.

**Gambar 9.** Grafik Sentimen

Selanjutnya setelah proses tahap pelabelan secara otomatis dilakukan proses uji data dengan menggunakan metode *Naïve Bayes Classification* dari proses tersebut didapatkan hasil akurasi sebesar 97.16%.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dan pembahasan yang telah diuraikan, maka dapat disimpulkan bahwa penggunaan data uji pada penelitian ini didapat dari sentiment konsumen melalui komentar review pada toko online shop amreta sebanyak 2014 data set. Kemudian data diolah melalui proses cleaning data sehingga menghasilkan data uji sebesar 1899 data set. Setelah tahapan melalui tahap preprocessing dilakukan tahap labelisasi yang menghasilkan 71% atau 1343 dataset yang bernilai positif, 3% atau 52 dataset yang bernilai negatif dan 26% atau 504 dataset yang netral. selanjutnya data tersebut diolah dengan menggunakan tools rapidminer sedangkan untuk operator dalam bentuk algoritma menggunakan model sentiment *Naïve Bayes Classification* melalui perhitungan otomatis didapatkan nilai akurasi sebesar 97.16%.

REFERENCES

- [1] Zuhri, F. N., & Alamsyah, A. (2017). Menggunakan *Naïve Bayes Classifier* Di Forum Kaskus Publik Sentiment Analysis Of Smartfren Brand Using *Naïve Bayes Classifier* On Kaskus Forum. E- Proceeding of Management, 4(1), 242–251.
- [2] Ratnawati, F. (2018). Implementasi Algoritma *Naive Bayes* Terhadap Analisis Sentimen Opini Film Pada Twitter. INOVTEK Polbeng - Seri Informatika, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>
- [3] Haryanto, D. J., Muflikhah, L., & Fauzi, M. A. (2018). Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 2(9), 2909–2916.
- [4] Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode *Naïve Bayes*. Jurnal Media Informatika Budidarma, 5, 987–994. <https://doi.org/10.30865/mib.v5i3.3030>
- [5] Rozi, I., Pramono, S., & Dahlan, E. (2012). Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi. Jurnal EECCIS, 6(1), 37–43.
- [6] Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment Analysis of YouTube Movie Trailer Comments Using *Naïve Bayes*. Bulletin of Computer Science and Electrical Engineering, 1(1), 26–32. <https://doi.org/10.25008/bcsee.v1i1.5>
- [7] Novianti, D. (2019). Implementasi Algoritma *Naïve Bayes* Pada Data Set Hepatitis Menggunakan Rapid Miner. Paradigma: Jurnal Komputer Dan Informatika Universitas Bina Sarana Informatika, 21(2), 143–148. <https://doi.org/10.31294/p.v20i2>
- [8] Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. Edik Informatika, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- [9] Lestari, A. R. T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naïve Bayes* dan Pembobotan Emoji. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 1(12), 1718–1724. <http://j-ptiik.ub.ac.id>
- [10] Aprilia, D., Aji Baskoro, D., Ambarwati, L., & Wicaksana, I. W. S. (2013). Belajar Data Mining Dengan Rapid Minner. <https://doi.org/->
- [11] Asiyah, A. (2021). Pengaruh Penilaian Produk Dan Ongkos Kirim Terhadap Keputusan Pembelian Konsumen Shopee Di Kabupaten Ponorogo. IAIN Ponorogo.



- [12] Chandani, V. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. *Journal of Intelligent Systems*, 1(1), 56–60.
- [13] Dillak, R. Y., Pangestuty, D. M., & Bintiri, M. G. (2012). Klasifikasi Jenis Musik Berdasarkan File Audio Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantization. *Seminar Nasional Informatika, 2012(semnasIF)*, 122–125.
- [14] Dupakuntla, V. B. P., Veeraboina, H., Reddy, M. V. K., Satyanarayana, M. M., & Sameer, Y. S. (2020). Learning Based Approach For Hindi Text Sentiment Analysis Using. *INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY*, 7(8).
- [15] Pratama, Y., Roberto Tampubolon, A., Diantri Sianturi, L., Diana Manalu, R., & Friez Pangaribuan, D. (2019). Implementation of Sentiment Analysis on Twitter Using Naïve Bayes Algorithm to Know the People Responses to Debate of DKI Jakarta Governor Election. *Journal of Physics: Conference Series*, 1175(1). <https://doi.org/10.1088/1742-6596/1175/1/012102>
- [16] Puspitasari, N., Tampubolon, W., & Taruk, M. (2021). Analisis Metode EUCS Dan Hot-Fit Dalam Mengevaluasi Penerapan Sistem Informasi Manajemen Kepegawaian (Simpeg). *Jurnal SITECH*, 4(1).