

ANALISIS SENTIMEN MASYARAKAT TERHADAP TOKO *ONLINE* APLIKASI SHOPEE MENGGUNAKAN METODE *MULTINOMIAL NAÏVE BAYES*

Sanrilla¹⁾, Natalis Ransi¹⁾, La Surimi¹⁾ Andi Tenriawaru¹⁾ La Ode Saidi¹⁾

¹⁾Program Studi Ilmu Komputer, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo, Kendari, Indonesia
E-mail: rylla1208@gmail.com

ABSTRAK

Tujuan penelitian ini adalah untuk mengetahui tingkat sentimen masyarakat dari ulasan produk yang ada pada toko *online* 3second dalam aplikasi shopee dengan menggunakan metode *Multinomial Naïve Bayes*. Analisis ulasan secara mudah dapat dilakukan dengan melihat jumlah bintang yang diberikan oleh pembeli, tetapi jumlah bintang tidak dapat mewakili isi dari keseluruhan ulasan. Perlu untuk melihat seluruh isi komentar ulasan untuk dapat mengetahui keseluruhan maksud ulasan. Sangat mungkin untuk menganalisis ulasan secara manual dengan melihat satu persatu, namun apabila ulasan yang dimiliki banyak akan lebih cepat menggunakan sistem analisis sentimen dengan cara melakukan teknik kalsifikasi. Algoritma *Multinomial Naïve Bayes* merupakan pengembangan dari *Naïve Bayes* yang digunakan dalam pembuatan analisis sentiment, karena algoritma ini bertujuan untuk mode klasifikasi dalam kategori positif dan negatif. Berdasarkan hasil pengujian yang dilakukan, metode *Multinomial Naïve Bayes* terbukti berfungsi baik pada penelitian ini. Hal ini dibuktikan dengan hasil *confusion matrix* dari pengujian klasifikasi sistem dan klasifikasi dari ahli bahasa mendapat nilai akurasi 91% dan memperoleh tingkat sentiment yang dibukti hasil *recall* dan *precision* yang bernilai 65.93% dan 60% untuk kelas positif, sedangkan untuk kelas negatif bernilai 34.06% dan 31%.

Kata kunci: Analisis sentimen, *Multinomial Naïve Bayes*, *Black box testing*, *Confusion matrix*, *Precision* dan *Recall*

ABSTRACT

The purpose of this study was to determine the level of public sentiment from product reviews on the 3second online store in the shopee application using the Multinomial Naïve Bayes method. Review analysis can easily be done by looking at the number of stars given by buyers, but the number of stars cannot represent the content of the entire review. It is necessary to look at the entire content of the review comments to be able to understand the overall intent of the review. It is possible to analyze reviews manually by looking at one by one, but if you have many reviews, it will be faster to use a sentiment analysis system by using a classification technique. Multinomial Nave Bayes Algorithm is the development of Nave Bayes which is used in making sentiment analysis, because this algorithm aims to classify modes in positive and negative categories. Based on the results of the tests carried out the Multinomial Naïve Bayes method proved to work well in this study. This is evidenced by the results of the confusion matrix from testing the classification system and classification from linguists who got an accuracy value of 91% and obtained a sentiment level as evidenced by the recall and precision results which were worth 65.93% and 60% for the positive class, while for the negative class it was worth 34.06% and 31%.

Keywords: *Sentiment analysis, Multinomial Naïve Bayes, Black box testing, Confusion matrix, Precision and Recall.*

1. Pendahuluan

Marketplace merupakan perkembangan dari *e-commerce* sebagai media daring berbasis internet untuk tempat melakukan kegiatan bisnis dan transaksi antara pembeli dan penjual. Pada marketplace pembeli dapat mencari barang yang ingin dibeli dari banyak penjual atau toko. Berdasarkan data Peta *e-commerce* Indonesia yang dikeluarkan oleh iPrice, Shopee adalah marketplace

dengan jumlah pengunjung terbanyak di Indonesia pada kuartal pertama tahun 2020 dengan jumlah pengunjung perbulan mencapai 71 juta pengunjung (Muktafin et al., 2020).

Shopee merupakan sebuah aplikasi yang bergerak di bidang jual beli secara *online* yang dapat diakses secara mudah dengan menggunakan *handphone*, dalam aplikasi Shopee ada banyak terdapat toko *online* yang berbeda-beda,

sehingga memudahkan penggunaanya untuk berbelanja diberbagai toko yang berbeda dalam 1 produk pencarian. Pada kegiatan pembelian produk disebuah toko *online* dalam aplikasi Shopee, pembeli dapat memberikan ulasan setelah menerima barang yang dibeli. Ulasan pembelian produk terdiri dari bintang dan isi komentar ulasan yang berisi tanggapan apresiasi maupun kritik dan masukan pada produk yang telah dibeli tersebut. Semakin banyak jumlah bintang yang dimiliki, maka semakin baik pula reputasi yang dimiliki toko tersebut. Ulasan pembelian produk memiliki pengaruh yang signifikan terhadap minat beli dari pembeli lain. Ulasan pembelian produk dapat digunakan oleh penjual untuk mendapat informasi sebagai bahan perbaikan pada produk dan layanan sehingga tercipta kepuasan pelanggan terhadap tokonya.

Analisis ulasan secara mudah dapat dilakukan dengan melihat jumlah bintang yang diberikan oleh pembeli, tetapi jumlah bintang tidak dapat mewakili isi dari keseluruhan ulasan. Diperlukan melihat seluruh isi komentar ulasan untuk dapat mengetahui keseluruhan maksud ulasan. Sangat mungkin untuk menganalisis ulasan secara manual dengan melihat satu persatu, namun apabila ulasan yang dimiliki banyak akan lebih cepat menggunakan sistem analisis sentimen dengan cara melakukan teknik kalsifikasi.

Text Mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dokumen dimana *Text Mining* merupakan variasi dari *Data Mining* yang berusaha menemukan pola menarik dari sekumpulan data tekstual yang berjumlah besar (Kurniawan et al., 2012). Naïve bayes biasa digunakan sebagai klasifikasi opini atau pun untuk pengklasifikasian. Sehingga untuk mempercepat proses perhitungan digunakan algoritma *Multinomial Naïve Bayes*. Algoritma *Multinomial Naïve Bayes* adalah pengembangan dari Naïve Bayes yang *Text Mining* adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dokumen dimana *Text Mining* merupakan variasi dari *Data Mining* yang berusaha menemukan pola menarik dari sekumpulan data tekstual yang berjumlah besar (Kurniawan et al., 2012). Naïve bayes biasa digunakan sebagai klasifikasi opini atau pun untuk pengklasifikasian. Sehingga untuk mempercepat proses perhitungan digunakan algoritma *Multinomial Naïve Bayes*. Algoritma *Multinomial Naïve Bayes* adalah pengembangan dari Naïve Bayes yang digunakan dalam pembuatan analisis sentiment, karena algoritma ini bertujuan

untuk mode klasifikasi dalam kategori positif dan negatif.

Maka dari itu pada penelitian ini peneliti melakukan analisis sentimen masyarakat pada ulasan produk dari salah satu toko dalam aplikasi Shopee, sehingga kita bisa tahu sentimen dari masyarakat terhadap suatu produk dari toko tersebut, dan dapat juga membantu masyarakat lain supaya lebih berhati-hati dalam membeli sebuah produk di toko *online*.

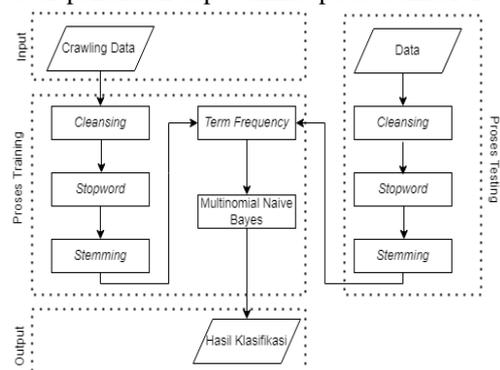
Adapun toko online yang menjadi objek penelitian ini yaitu toko *local brand* 3Second. Di lansir dari artikel yang ditulis oleh media online *shopee.co.id*, toko *online* tersebut termasuk lima dari *local brand* yang kualitasnya tidak jauh beda dengan kualitas dari produk-produk *brand* luar. Data tersebut diperoleh dari sumber *novellmag.com*, dengan peringkat pertama 3Second.

Menjadi urutan pertama toko *local brand* dengan kualitas terbaik, akan membuat toko 3second semakin dikenal oleh masyarakat dan membuat tingkat penjualan semakin banyak. Oleh karea itu Kepuasan pelanggan merupakan hal penting yang menjadi tujuan dari semua toko *online*. Selain kepuasan pelanggan yang menjadi hal penting, tak lupa pula dalam penjualan *online* tentu tidak jauh dari sentimen masyarakat atau konsumen oleh karea itu penulis berdasarkan ulasan di atas mengangkat judul yaitu “Analisis Sentimen Masyarakat Terhadap Toko *Online* Pada Aplikasi Shopee Menggunakan Metode *Multinomial Naïve Bayes*”

2. Kajian Pustaka

2.1 Prosedur Penelitian

Pada penelitian ini sistem mempunyai dua tahapan proses yaitu tahap data training adalah proses pembelajaran dan tahap data testing yang nantinya merupakan output dari sistem ini. Nilai *output* berupa sentimen ulasan yang bernilai positif atau negatif yang diklasifikasikan oleh sistem berdasarkan pembelajaran data *training*. Berikut alur prosedur penelitian dapat dilihat pada Gambar 1.

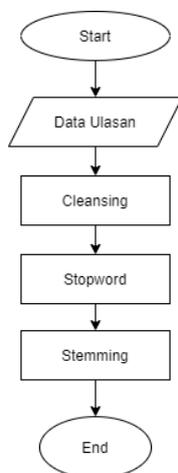


Gambar 1. Alur Prosedur Penelitian

Berdasarkan Gambar 1 alur prosedur penelitian adalah sebagai berikut.

1. Pertama melakukan input data dengan proses *crawling* data
2. Setelah *crawling* data lanjut ketahap proses training yang terdiri dari tahap *preprocessing*, menghitung *term frequency*, dan menghitung nilai *conditionalprobabilitas* dengan metode Multinomial Naive Bayes
3. Setelah tahap perhitungan dengan metode Multinomial Naive Bayes selanjutnya ketahap proses testing yang pertama input data testing lalu proses *preprocessing*, menghitung *term frequency*, dan menghitung nilai *probabilitas* dengan metode Multinomial Naive Bayes
4. Terakhir yaitu *output* hasil akhir klasifikasi yang akan di klasifikasikan ke dalam masing-masing kelasnya.

Flowchart berikut ini adalah menjelaskan sistem sedang melakukan proses *Preprocessing*. Flowchart desain yang akan digunakan dalam penelitian ini dapat dilihat pada Gambar 2.



Gambar2. Alur Pre-Processing

Text preprocessing adalah tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Inputan awal pada proses ini adalah berupa dokumen. Text preprocessing pada penelitian ini terdiri dari beberapa tahapan, yaitu (Mustaqhfi et al., 2012). Proses *preprocessing* merupakan hal yang penting untuk tahap selanjutnya, yaitu mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi.

1. *Cleansing*, tahapan *cleansing* merupakan tahap pembersihan kata yang tidak berpengaruh sama sekali terhadap hasil klasifikasi sentimen.

2. *Stopword*, tahap *stopword* merupakan tahan menghilangkan kata yang tidak sesuai dengan topik dokumen, jika ada kata tersebut tidak mempengaruhi akurasi dalam klasifikasi sentiment dokumen.

3. *Stemming*, teknik *stemming* diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

2.2 Jembatan Brown

2.2.1 Proses Training

Sebuah dokumen *training* sudah diklasifikasi-

No.	Ulasan	Kategori
1	harga agak mahal padahal kain biasa saja	N
2	alhamdulillah pesanan sampai selamat kualitas bagus pengirim cepat	P
3	barang bagus banget	P
4	warna tidak sesuai bahan tipis sablon jelek harga mahal kualitas bisa sesuai harga padahal parah	N
5	sampe produk baik suka sama barang	P
6	kualitas barang baik kecepatan pengirim baik respon penjual baik sering kasih diskon	P

kan secara *manual* dan sudah dilakukan proses *preprocessing* seperti pada Tabel 1.

Dari data Tabel 1 dibuat sebuah model *prior probabilitas* dengan mengacu pada Persamaan (1).

$$P(c) = \frac{N_c}{N} \quad (1)$$

$$P(\text{positif}) = \frac{4}{6}$$

Kelas positif muncul 4 kali dalam dokumen *training*. Total dokumen adalah 6 dokumen. Begitu pula untuk kelas negatif, setiap dokumen yang muncul dibagi dengan total seluruh dokumen. Setiap term yang ditemukan akan dihitung probabilitas kemunculan untuk setiap kelas kemudian dibagi dengan total *termfrequency* yang ditemukan. Contoh *termfrequency* yang ditemukan terdapat pada Tabel 2.

Tabel 2. Contoh *Term Frequency* Pada Dokumen

No.	Term	Positif	Negatif
1	alhamdulillah	1/30	0/23
2	bahan	0/30	1/23
3	pesanan	1/30	0/23
4	sesuai	0/30	2/23
5	sampai	2/30	0/23
6	kain	0/30	1/23
7	dengan	2/30	0/23
8	selamat	1/30	0/23
9	biasa	0/30	1/23
10	kualitas	2/30	1/23

Dalam proses perhitungan kemungkinan term tidak terdapat dalam suatu kelas. Apabila term tersebut tidak terdapat pada kelas maka akan bernilai 0. Hal ini dapat menyebabkan seluruh perhitungan menjadi tidak sesuai. Untuk itu digunakan teknik *smoothing*. Teknik *smoothing* yang digunakan adalah *Laplace smoothing*.

Berikut adalah contoh perhitungan *conditional probabilitas* untuk term “kualitas” pada kelas positif dan negatif mengacu pada Persamaan (2).

$$P(t|c) = \frac{T_{ct}+1}{\sum_{t \in V} T_{ct}+B} \quad (2)$$

$$P(\text{kualitas}|\text{positif}) = \frac{2+1}{30+37} = \frac{3}{67} = 0.0447$$

$$P(\text{kualitas}|\text{negatif}) = \frac{1+1}{23+37} = \frac{2}{60} = 0.0333$$

Pada persamaan diatas angka 2 mewakili term kualitas yang ditemukan. Angka 30 mewakili total term pada kelas positif sedangkan 37 adalah total term pada seluruh dokumen. Probabilitas tersebut nantinya dicari nilai yang paling besar dan dijadikan sebagai kelas untuk term tersebut. Misalnya saja kualitas kelas positif menunjukkan 0.0447 sedangkan kelas negatif 0.0333 Dengan hasil tersebut didapatkan term telkomsel pada kelas positif karena nilai positif lebih besar.

2.2.2 Proses Testing

Alur proses testing kurang lebih sama dengan alur training, yang dimana proses testing adalah proses uji coba dari data training hanya saja proses testing disini dilakukan dengan nilai probabilitas akhir yang diambil dari nilai *conditional probabilitas* yang dipangkatkan dengan *term frequency*-nya, kemudian dikalikan dengan nilai prior probabilitasnya. Berikut contoh data ulasan yang akan dihitung nilai probabilitas akhirnya seperti pada Tabel 3.

Tabel 3. Data Testing

No.	Ulasan	Kategori
7	Maskernyakebesaran padahal motifnyasudah bagus	?
8	Suka banget sama jaketnya ukuran juga pas warnya soft lucu banget	? ? ?
9	Pengiriman cepat tapi warna tidak sesuai dengan pesanan	? ?

Perhitungan probabilitas

berikut mengacu pada Persamaan 2.1 yang dimana nilai *conditional probabilitas* dipangkat dengan nilai dari *term frequency* masing-masing katanya kemudian hasilnya dijumlahkan sehingga memperoleh nilai akhir.

- $$P(\text{Ulasan7}|\text{Positif}) = P(\text{masker}|\text{Positif}) + P(\text{kebesaran}|\text{Positif}) + P(\text{padahal}|\text{Positif}) + P(\text{motif}|\text{Positif}) + P(\text{sudah}|\text{Positif}) + P(\text{bagus}|\text{Positif})$$

$$= 0,02985075 + 0,02985075 + 0,02985075 + 0,02985075 + 0,02985075 + 0,04477612^2$$

$$= 0,02985075 + 0,02985075 + 0,02985075 + 0,02985075 + 0,02985075 + 0,0020049$$

$$= 0,15125865$$
- $$P(\text{Ulasan7}|\text{Negatif}) = N(\text{masker}|\text{Negatif}) + N(\text{kebesaran}|\text{Negatif}) + N(\text{padahal}|\text{Negatif}) + N(\text{motif}|\text{Negatif}) + N(\text{sudah}|\text{Negatif}) + N(\text{bagus}|\text{Negatif})$$

$$= 0,033333333 + 0,033333333 + 0,05^2 + 0,033333333 + 0,033333333 + 0,033333333$$

$$= 0,033333333 + 0,033333333 + 0,0025 + 0,033333333 + 0,033333333 + 0,033333333$$

$$= 0,169166667$$

Setelah nilai akhir didapat maka selanjutnya akan dikalikan dengan nilai prior probabilitas yang dimana nilai prior probabilitas berasal dari jumlah total ulasan kelas positif/negatif dibagi dengan total keseluruhan ulasan. Berikut cara perhitungan algoritmanya.

- $$P(\text{Ulasan7}|\text{Positif}) = 4/6 \times 0,15125865 = 0,1008391$$

2. $P(\text{Ulasan7|Negatif}) = 2/6 \times 0,169166667 = 0,056388889$

Dari hasil perhitungan probabilitas diatas diketahui bahwa probabilitasulasan 7 terlihat kelas positif memiliki nilai yang paling besar dibandingkan dengan kelas negatif, sehingga ulasan 7 diklasifikasikan kedalam kelas positif.

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Data yang dikumpulkan diambil dari web *scraper*. Web *scraper* adalah cara untuk mendapatkan banyak data dari website dengan lebih cepat dan dalam waktu yang singkat. Langkah awal yaitu mengunjungi situs web *screper.io* kemudian menginstall web *scraper* yang *free* (tidak berbayar) pada web chrome, untuk menjalankan web scrapernya pertama membuat sitemap sebagai tempat untuk memudahkan mencari data yang akan di *scraping*, setelah itu masuk ketahap *scarping* dengan menentukan bagian mana data yang akan dilakukan *scraping* tersebut. Setelah data berhasil di *scraping* selanjutnya data akan disimpan kedalam file excel sekaligus melakukan klasifikasi dan pembagian kategori secara manual. Data yang dikumpulkan masing-masingnya terdiri dari 300 data positif dan 300 data negatif, jadi totalnya sebanyak 600 data ulasan.

3.2 Pengujian

Pengujian terhadap kinerja dari sistem analisis sentimen untuk mengetahui sentimen ulasan masyarakat dari suatu toko. Jadi pada tahap ini kita bisa mengetahui kekurangan dari sistem ini, agar dilakukan perbaikan sehingga kesalahan-kesalahan tersebut dapat diminimalisasi. Pada pengujian ini akan dilakukan dengan *black box* testing, pengujian kebenaran hasil klasifikasi menurut ahli bahasa melalui wawancara, pengujian *confusion matrix* dan pengujian *recall* dan *precision*.

3.2.1 Black Box Testing

Pengujian dengan menggunakan metode *Black Box* guna untuk berfokus pada kebutuhan fungsional perangkat lunak yang dibuat. Sehingga, pengujian dengan metode *Black Box* memungkinkan untuk membuat *input*-an yang melatih seluruh syarat-syarat fungsional suatu sistem. Berikut rencana pengujian sistemnya.

Tabel 4. Pengujian Halaman Login

Data Masukkan	Hasil Yang Diharapkan	Hasil Pengujian	Kesimpulan Pengujian
Username admin : Password admin : Klik tombol masuk	Data yang salah maka sistem akan menampilkan pesan kesalahan	menampilkan pesan kesalahan	Sesuai

Data Masukkan	Hasil Yang Diharapkan	Hasil Pengujian	Kesimpulan
Username admin : Password admin : Klik tombol masuk	Data yang salah maka sistem akan menampilkan pesan kesalahan	menampilkan pesan kesalahan	Sesuai

Hasil pengujian halaman *login* menggunakan pengujian dengan memasukan data *login* valid yang sudah terdaftar dalam basis data maka hasil pengujian berhasil dan dapat mengakses sistem. Tetapi apabila data yang dimasukkan salah, maka sistem akan menampilkan pesan kesalahan yaitu “Salah kombinasi username dan password” dan tidak dapat mengakses sistem.

Tabel 5. Pengujian Halaman Training

Data Masukkan	Hasil Yang Diharapkan	Hasil Pengujian	Kesimpulan
Klik menu training	Menampilkan sub menu data, <i>cleansing</i> , <i>stopword stemming</i> , <i>term frequency</i> , dan <i>conditional probabilities</i> masing-masing menu sudah ada datanya	Menampilkan sub menu data, <i>cleansing</i> , <i>stopword stemming</i> , <i>term frequency</i> , dan <i>conditional probabilities</i> masing-masing menu sudah ada datanya	Sesuai

Hasil pengujian halaman training, admin akan di tampilkan beberapa sub menu yang masing-mising menunya berisi data-data yang sudah diolah melalui proses training.

Table 6. Pengujian Halaman Testing

Data Masukkan	Hasil Yang Diharapkan	Hasil Pengujian	Kesimpulan Pengamatan
Klik menu testing	Menampilkan form pengisian data testing	Menampilkan form pengisian data testing	Sesuai

Pilih data testing Masukan data testing: Klik tombol proses perhitungan	Menampilkan sub menu data, <i>cleansing, stopword stemming</i> , dan hasil masing-masing menu sudah ada datanya	Menampilkan sub menu data, <i>cleansing, stopword stemming</i> , dan hasil masing-masing menu sudah ada datanya	Sesuai
---	---	---	--------

Hasil pengujian halaman training, admin akan memasukkan data ulasan pada form masukkan data testing sistem akan mengolah ulasan tersebut dan dicocokkan dengan data yang ada pada database. Jika data cocok maka sistem akan menampilkan hasil.

3.2.2 Hasil Pengujian Menurut Ahli Bahasa

Hasil yang diperoleh menurut ahli bahasa menyatakan untuk membedakan sebuah komentar atau ulasan suatu barang, baik secara lisan ataupun tulisan tergantung dari hasil yang diharapkan dari barang tersebut apakah sudah memenuhi harapannya. Kalimat yang diungkapkan bisa diketahui melalui penggunaan kata yang lebih dominan kedalam kata sifat. Data *testing* yang di berikan kepada ahli bahasa sebanyak 200 data yang telah diuji pada sistem sebelumnya, 182 data tersebut dikatakan telah sesuai menurut ahli bahasa bapak Harmin, S.Pd.,M.Pd.

3.2.3 Pengujian Confusion Matrix

Pengujian *Confusion Matrix* juga sering disebut *error matrix*. *Confusion Matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya. Tingkat akurasi algoritma analisis sentimen dilakukan dengan cara membandingkan hasil klasifikasi yang diperoleh dari ahli bahasa untuk mendapatkan hasil klasifikasi yang tepat pada sistem. Berikut hasil *confusion matrix* dapat dilihat pada Tabel 7.

Tabel 7. Hasil Confusion Matrix

		True Class	
		Positif	Negatif
Predicted Class	Positif	120	7
	Negatif	11	62

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Accuracy = \frac{120 + 62}{120+62+11+7} \times 100\%$$

$$= \frac{182}{200} \times 100\%$$

$$= 91\%$$

Recall adalah perhitungan dokumen yang relevan dari seluruh dokumen relevan yang berada didalam data. *Precision* adalah jumlah kelompok dokumen relevan dari total jumlah dokumen yang ditemukan.

$$Recall = \frac{\text{jumlah dokumen relevan yang ditemukan}}{\text{total dokumen relevan yang ditemukan dalam koleksi}}$$

$$Precision = \frac{\text{jumlah dokumen relevan yang ditemukan}}{\text{jumlah total dokumen yang ditemukan}}$$

Dari total 200 data ulasan yang uji oleh sistem ditemukan 127 data bernilai positif dan 73 data bernilai negatif. Berikut ini nilai *recall* dan *precision* untuk kelas positif dapat dilihat pada Tabel 8.

Tabel 8. Recall Dan Precision Kelas Positif

Kelas Positif	Relevan	Tidak Relevan
True	120	7
False	0	0

Pada tabel diatas pengukuran *recall* dan *precision* yaitu:

$$Recall = \frac{120}{120 + 62} \times 100\%$$

$$= 65.93\%$$

$$Precision = \frac{120}{127 + 73} \times 100\%$$

$$= 60\%$$

Berikut ini nilai *recall* dan *precision* untuk kelas negatif dapat dilihat pada Tabel 9.

Tabel 9. Recall dan Precision Kelas Negatif

Kelas Negatif	Relevan	Tidak Relevan
True	62	11
False	0	0

Pada tabel diatas pengukuran *recall* dan *precision* yaitu:

$$Recall = \frac{62}{62 + 120} \times 100\%$$

$$= 34.06\%$$

$$Precision = \frac{62}{127 + 73} \times 100\%$$

$$= 31\%$$

Akurasi yang didapatkan dari adalah *recall* 65.93% dan *precision* 60% untuk kelas positif. Untuk kelas negatif sebesar *recall* 34.06% dan *precision* 31%. Nilai tersebut didapatkan dari *recall* yang membagi antara jumlah relevan yang ditemukan dengan jumlah semua data relevan yang ditemukan dan *precision* yang membagi antara jumlah relevan yang ditemukan dengan jumlah semua data yang ditemukan.

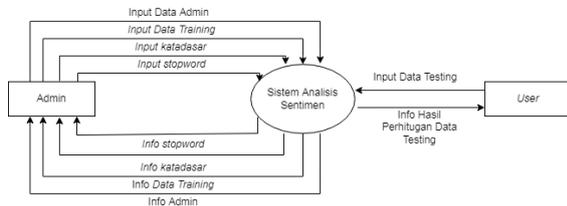
3.3 Perancangan Sistem

3.3.1 Data Flow Diagram

Data Flow Diagram (DFD) adalah suatu diagram yang menggambarkan aliran data dari sebuah proses yang sering disebut dengan sistem informasi.

1) Konteks Diagram

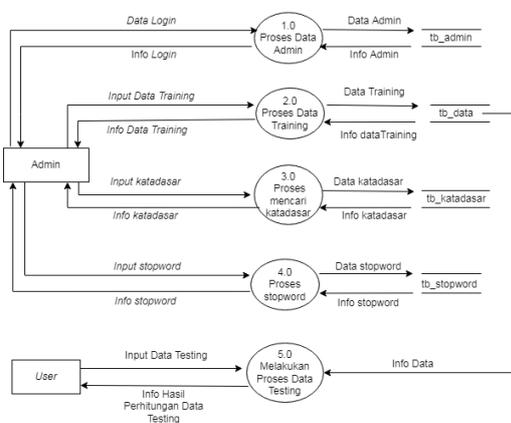
Diagram konteks sering disebut juga dengan Level-0 dan menjadi penentu utama padasebuah sistem yang dimodelkan dalam DFD.



Gambar 3. Diagram Konteks

2) Data Flow Diagram Level 1

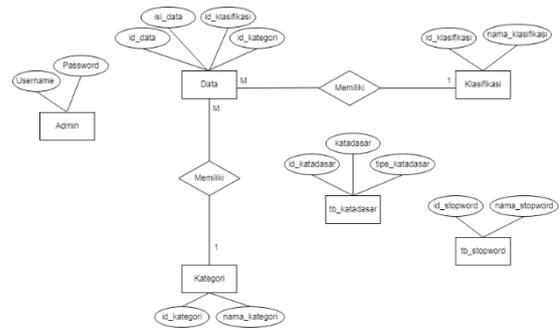
Data Flow Diagram Level 1 adalah kelanjutan dari diagram konteks, menggambarkan interaksi yang menjelaskan secara terperinci sistem analisis sentimen masyarakat yang terdapat pada Gambar 4.



Gambar 4. Data Flow Diagram Level 1

3.3.2 Entity Relationship Diagram

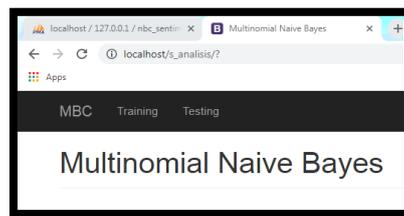
Entity Relationship Diagram (ERD) merupakan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. Rancangannya dapat dilihat pada Gambar 5.



Gambar 5. Entity Relationship Diagram

3.4 Tampilan Antarmuka

Tampilan antarmuka adalah untuk merancang tata letak sistem sesuai dengan kebutuhan sistem. Pada awal tampilan antarmuka sistem ini, menampilkan menu training yaitu tempat data latih dan menu testing untuk proses data uji. Tampilan awal antarmuka dapat dilihat pada Gambar 6.



Gambar 6. Tampilan Awal Antarmuka

3.4.1 Proses Training

Data training merupakan kumpulan data yang sudah ada untuk dijadikan sebagai data latih atau pembelajaran. Sebanyak 400 data ulasan yang dijadikan sebagai data latih yang akan diolah, beberapa tahapan proses training yaitu proses preprocessing, menghitung *term frequency*, dan menghitung nilai *conditional* probabilitas. Berikut data training dapat dilihat pada Gambar 7.

Cleansing	
No	Cleansing
1	terimakasih produk baik suka sama produknya
2	bagus aku suka modelnya pengiriman juga cepat
3	bagus nggak salah merek ini sesuai ukuran makasih
4	bajunya bagus bahannya lumayan ternyata pas banget sama badan aku
5	cepat banget sampenya mantap bagus jaketnya juga

Gambar 7. Proses Cleansing

1) Proses Stopword

Proses *Stopword* merupakan proses menghilangkan kata penghubung dan kata yang tidak sesuai dengan dokumen. Tampilan proses *stopword* dapat dilihat pada Gambar 8.

Stopword	
No	Stopword
1	terimakasih produk baik suka sama produknya
2	bagus suka modelnya pengiriman cepat
3	bagus nggak salah merek sesuai ukuran makasih
4	bajunya bagus bahannya lumayan ternyata pas banget sama badan
5	cepat banget sampainya mantap bagus jaketnya

Gambar 8. Proses Stopword

2) Proses Stemming

Proses *Stemming* adalah proses mengubah kata ke dalam bentuk kata dasar. Tampilan proses *stemming* dapat dilihat pada Gambar 9.

Stemming	
No	Stemming
1	terimakasih produk baik suka sama produk
2	bagus suka model pengiriman cepat
3	bagus nggak salah merek sesuai ukuran makasih
4	baju bagus bahan lumayan nyata pas banget sama badan

Gambar 9. Proses Stemming

3) Term Frequency

Menghitung *term frequency* (frekuensi kata) yang diperoleh dari tahap preprocessing. Semua frekuensi kata yang muncul akan di hitung jumlah kemunculannya pada tiap kelas. Tampilan *term frequency* dapat dilihat pada Gambar 10.

Term Frequency			
No	Term	Positif	Negatif
1	terimakasih	39	2
2	produk	38	31
3	baik	23	14
4	suka	31	6
5	sama	12	20

Gambar 10. Term Frequency

4) Conditional Probabilitas

Menghitung nilai *conditional probabilitas* dan memasukkan teknik *laplace smoothing* sebagai proses penambahan nilai 1 untuk menghindari nilai nol. Tampilan proses *conditional probabilitas* dapat dilihat pada Gambar 11.

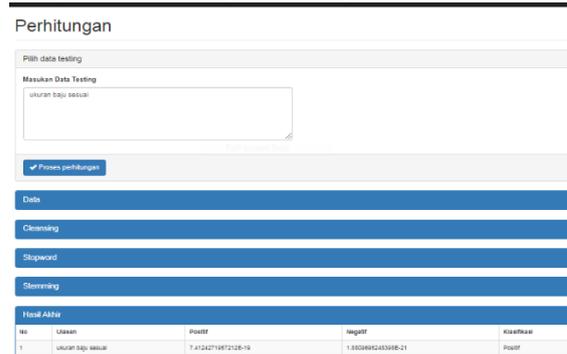
Conditional Probabilitas				
No	Term	Positif	Negatif	Klasifikasi
1	terimakasih	0.010913043470261	0.000207929898944	Positif
2	produk	0.01286985217391	0.009671207430341	Positif
3	baik	0.0035479200898665	0.0048426828482972	Positif
4	suka	0.011130434702609	0.002167162662307	Positif
5	sama	0.0048217391304348	0.008793281733746	Negatif

Gambar 11. Perhitungan Conditional Probabilitas

3.4.2 Proses Testing

Data testing ini merupakan data uji coba dari data training yang akan dilakukan sebanyak 200 data

ulasan masyarakat. Berikut ini tampilan proses testing data ulasan dapat dilihat pada Gambar 12.



Gambar 12. Proses Testing

4 Penutup

4.4 Kesimpulan

1. Metode *Multinomial Naive Bayes* terbukti berfungsi baik pada penelitian ini. Hal ini dibuktikan dengan hasil *confusion matrix* dari pengujian klasifikasi sistem dan klasifikasi dari ahli bahasa mendapat nilai akurasi 91% dari 200 data ulasan produk.
2. memperoleh tingkat sentiment masyarakat terhadap toko *online 3Second* pada aplikasi *shopee* dengan sentimen positif yang cukup banyak. Hal ini terbukti hasil *recall* dan *precision* yang bernilai 65.93% dan 60% untuk kelas positif, sedangkan untuk kelas negatif bernilai 34.06% dan 31%.

Ucapan Terimakasih. Penulis menyampaikan rasa terimakasih terhadap para pembimbing dan semua dosen penguji pada program studi ilmu komputer FMIPA.

Daftar Pustaka

- [1] Kurniawan, B., Effendi, S., & Sitompul, O. S. (2012). Klasifikasi Konten Berita Dengan Metode *Text Mining*. *Jurnal Dunia Teknologi Informasi*. Vol 5 (1), 14–19.
- [2] Muktafin, E. H., Kusriani., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language. *Jurnal Eksplora Informatika*. 32-42.
- [3] Mustaqhfiri, M., Abidin, Z., & Kusumawati, R. (2012). Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode *Maximum Marginal Relevance*. 134-147.

Diterima pada tanggal 23 Mei 2022.
Terbit online pada tanggal 28 Juli 2022