

Analisis Sentimen Pada Review Produk Kosmetik Bahasa Indonesia Dengan Metode Naive Bayes

Hendry Ardian^{*1}, Sandy Kosasi²

^{1,2}Jurusan Teknik Informatika; STMIK Pontianak. Jl. Merdeka No.372 Pontianak, 0561-735555
e-mail: ^{*1}hendryardian16@gmail.com, ²Sandykosasi@gmail.com

Abstrak

Analisis sentimen adalah sebuah bidang mempelajari opini, tanggapan, dan emosi seseorang terhadap suatu entitas yang dalam penelitian ini adalah review produk. Salah satu cara mengetahui kualitas produk adalah dengan membaca review produk konsumen lain. Produk kosmetik yang beredar di pasaran sangat beragam dari segi jenis maupun brand. Namun tidak semua kosmetik memiliki kualitas yang baik dan hal ini harus diperhatikan oleh konsumen. Pengkajian ulang tentang review produk kosmetik dengan cara pengklasifikasian review ke dalam kelas positif dan negatif adalah suatu cara terbaik untuk mengetahui tanggapan konsumen lain tentang suatu produk secara cepat dan tepat. Teknik klasifikasi yang paling sering digunakan adalah Naive Bayes. Naive Bayes adalah teknik yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, Naive Bayes memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan sering terjadi kesalahan dalam klasifikasi. Oleh karena itu, dalam penelitian ini digunakan metode pemilihan fitur tf-idf agar bisa meningkatkan kinerja dari Naive Bayes. Tujuan dari penelitian ini adalah mengembangkan sistem yang dapat mengklasifikasikan review menjadi positif atau negatif. Pengukuran akurasi naive bayes dilakukan dengan confusion matrix dan menghasilkan akurasi dari 69% - 82%.

Kata kunci : Analisis Sentimen , Naive Bayes, Klasifikasi, Review Produk Kosmetik, tf-idf

Abstract

Sentiment analysis is a computational study of the opinions, behaviors and emotions of people toward the entity which in this research is a product review. By reading the review of products based on the experiences of other consumers, it will be recognized the quality of a product. As cosmetic products on the market are very diverse, both in terms of type and brand. However, not all cosmetics have good quality and it is to be noticed by the consumer. So, reexamination of the cosmetic product review by classifying these reviews into positive and negative class is an excellent way to determine the response of other consumers about the product quickly and accurately. The classification technique that mostly used is Naive Bayes. Naive Bayes is a popular classification technique for text because very simple, efficient, and has good performance in many subject. But, Naive Bayes has weaknesses that is too sensitive when there are too many features which cause wrong classification. Because of that, in this research tf-idf is used as feature selection to improve Naive Bayes performance. This research aim is to produce a system that can classify reviews to their respective positive or negative class. The accuracy measurement of Naive Bayes is done by using confusion matrix and indicate the results of average accuracy from 69% to 82%.

Keywords : Sentiment Analysis, Naive Bayes, Classification, Cosmetics Product Review, tf-idf

1. PENDAHULUAN

Pada era yang terus berkembang ini, sentimen atau opini masyarakat semakin bertambah luas dan bebas diungkapkan di berbagai media. Sentimen dapat ditemukan dalam berbagai bentuk salah satunya dalam bentuk teks. Sentimen dalam bentuk teks mengandung informasi yang penting dan banyak didapatkan dari berbagai sumber seperti buku, surat kabar, situs web dan sebagainya [1]. Informasi berbentuk teks dalam sentimen dapat diambil dengan menggunakan *text mining*. *Text mining* merupakan proses mengambil informasi dari teks [2]. *Text mining* adalah proses awal dari analisis sentimen yang sangat penting karena digunakan sebagai pengidentifikasi emosional suatu pernyataan [3]. Setelah *text mining* dilakukan, maka yang selanjutnya dilakukan adalah analisis sentimen.

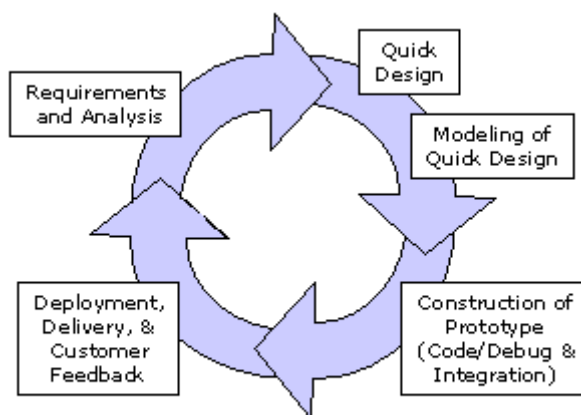
Dalam perkembangannya sentimen semakin mudah ditemukan dalam berbagai media seperti media *online*. Di era ini sentimen atau opini banyak bermunculan pada website yang memberikan review produk. Banyaknya *website* untuk *review* menyebabkan ketersediaan data *review* sangat banyak. Ketersediaan data *review* inilah yang menjadi alasan penelitian ini menggunakan sumber data berupa *review* produk untuk analisis sentimen. *Review* produk dapat memberikan manfaat besar bagi konsumen karena dengan membaca *review* konsumen mendapat informasi yaitu dapat mengetahui kualitas suatu produk dari *review* konsumen lain yang membagikan pengalamannya tentang produk terkait.

Pengguna yang sangat banyak yang memberikan *review* untuk suatu produk ditambah produk yang ada sangat banyak dan beragam menghasilkan *review* dalam jumlah yang sangat banyak untuk setiap produk [4]. Menghadapi jumlah data yang besar dan sangat banyak pada *website*, menganalisis, dan menyimpulkan informasi secara manual adalah tidak mungkin [5]. Selain jumlah data yang besar dan sangat banyak, masalah lain yang dihadapi adalah informasi yang ada dalam opini. Opini dapat berisi informasi yang tidak lengkap, informasi yang tidak konsisten, atau informasi yang berperilaku pada tren tertentu [6]. Jadi cara untuk mengekstrak informasi penting dan membangun sistem yang dapat menentukan kualitas produk secara objektif dan secara otomatis untuk menangani informasi teks yang sangat besar dan banyak adalah dengan menggunakan sistem analisis sentimen. Analisis sentimen akan mencoba menemukan polaritas sentimen dari suatu pernyataan atau kalimat dan mengkategorikan menjadi kelas positif dan negatif.

Sebelumnya telah dilakukan penelitian terdahulu yang berkaitan dengan topik ini. Penelitian tersebut diantaranya meneliti analisis sentimen *review* dataset menggunakan Naive Bayes dan K-NN *Classifier*. Hasil dari penelitian ini adalah Naive Bayes memberikan hasil yang jauh lebih baik dari K-NN dalam kasus *review* film sedangkan untuk *review* hotel Naive Bayes menunjukkan tingkat akurasi yang lebih rendah dibanding dengan *review* film, tetapi tingkat akurasi Naive Bayes dan K-NN hampir sama dalam kasus *review* hotel. Penelitian ini dilakukan untuk membandingkan tingkat akurasi dari algoritma Naive Bayes dan K-NN classifier [7]. Penelitian lain yaitu analisis sentimen dilakukan untuk menganalisis *review* produk Amazon. Analisis sentimen dilakukan dengan ekstraksi fitur dan teknik pemilihan yang berbeda-beda, sehingga langkah-langkah yang diperlukan juga berbeda [8]. Penelitian selanjutnya dilakukan dengan mempelajari seberapa berguna mengklasifikasikan *review* produk secara semantic. Dalam penelitian ini juga dijelaskan pentingnya analisis sentimen dan juga kelebihan dan kekurangan dari sistem analisis sentimen yang sudah ada. Peneliti mengajukan cara-cara yang berbeda seperti pengoreksi ejaan dalam teks *review* dan pengklasifikasian komentar. Peneliti menggunakan algoritma gabungan yang mengkombinasikan naive bayes dan decision tree. Dalam penelitian ini sumber data yang digunakan bersumber dari komentar flipkart untuk handphone MOTO X Play. Naive Bayes Classifier digunakan untuk mengklasifikasi kata-kata positif dan negatif dan decision tree digunakan untuk menghitung polaritas secara keseluruhan [9].

2. METODE PENELITIAN

Penelitian ini menggunakan bentuk penelitian yang beracuan pada studi literatur, yaitu dengan cara melakukan analisis dan klasifikasi fakta yang dikumpulkan dalam penelitian yang dilakukan. Dalam penelitian ini menggunakan metode penelitian *Research and Development* (R&D) yang merupakan metode yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut [10]. Metode pengumpulan data yang digunakan adalah data primer dan data sekunder sedangkan teknik pengumpulan data meliputi observasi dan studi dokumentasi. Metode perancangan perangkat lunak menggunakan metode prototype dengan langkah-langkah komunikasi dan pengumpulan data awal dengan melakukan komunikasi dengan user kemudian mengidentifikasi kebutuhan yang diperlukan dalam perancangan sistem meliputi kebutuhan untuk mengambil data, memproses data, membuat data latih dan data uji, input dan output sistem yang diinginkan. Tahap kedua desain cepat yaitu setelah mengetahui kebutuhan sistem, maka akan dilakukan pembentukan model desain dari sistem yang mana pemodelan dibuat dengan menggunakan UML dan flowchart. Tahap ketiga pembentukan prototype yaitu dilakukan pembentukan dari sistem yang telah didesain sebelumnya dimana pembentukan prototype ini meliputi fungsi, tampilan, dan cara kerja yang telah diidentifikasi. Tahap keempat adalah evaluasi dimana pada tahap ini sistem yang telah dibentuk akan dievaluasi apakah telah layak digunakan dan fungsinya telah sama dengan kebutuhan user. Tahap kelima adalah perbaikan yang merupakan tahap penyempurnaan kesalahan yang terjadi saat dilakukan evaluasi. Tahap terakhir adalah produksi akhir merupakan tahap dimana sistem yang dibangun telah sesuai dengan kebutuhan user dan sistem telah dapat digunakan oleh user [11].



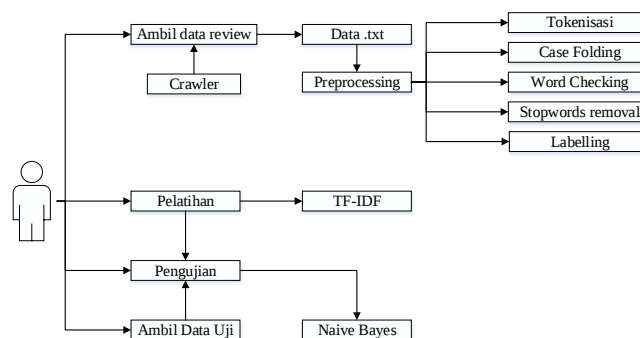
Gambar 1 Tahapan Metode Prototype

Dalam penelitian ini, perangkat analisis dan permodelan sistem yang digunakan adalah UML (Unified Modeling Language) dan flowchart. UML digunakan untuk menjelaskan langkah-langkah yang dilakukan dalam sistem yang sangat bernuansa perangkat lunak (*software intensive system*). UML yang digunakan untuk pemodelan sistem adalah use-case diagram digunakan untuk menggambarkan interaksi yang terjadi dalam sistem, sequence diagram yang digunakan untuk menjelaskan aliran perangkat lunak yang dikembangkan dan activity diagram digunakan untuk menggambarkan proses-proses ataupun kegiatan dalam sistem yang dibangun [12]. Flowchart digunakan untuk menggambarkan alur kerja sistem.

3. HASIL DAN PEMBAHASAN

Perancangan system analisis sentimen menggunakan metode *prototype* dimana tahapannya dimulai dari tahap analisis kebutuhan yang berorientasi kepada analisa sistem yang didalamnya berisikan analisa atas kebutuhan pengguna sistem. Analisa sistem dilakukan dengan melakukan observasi dan studi dokumentasi dengan mempelajari dokumen, laporan serta data yang berhubungan untuk mengetahui kebutuhan dan fitur apa saja yang dibutuhkan dalam pengembangan sistem analisis sentimen. Dari pengumpulan data tersebut dapat diambil hasil analisis kebutuhan sistem sebagai berikut : 1) Kebutuhan Fungsional: Sistem melakukan pencarian dengan proses crawling untuk mendapatkan data review produk yang akan di analisis, kemudian data review diproses dengan melakukan preprocessing dan ekstraksi fitur yang kemudian diklasifikasikan dengan naive bayes 2) Kebutuhan Antarmuka: sistem dibuat pada dengan sistem operasi windows dan memerlukan search box untuk mulai menjalankan sistem 3) Kebutuhan Unjuk Kerja : sistem dapat mengeksekusi preprocessing dan ekstraksi fitur pada review produk, sistem dapat mengklasifikasikan review produk ke dalam kelas sentimennya masing-masing dan menampilkan kelas sentimen setiap review.

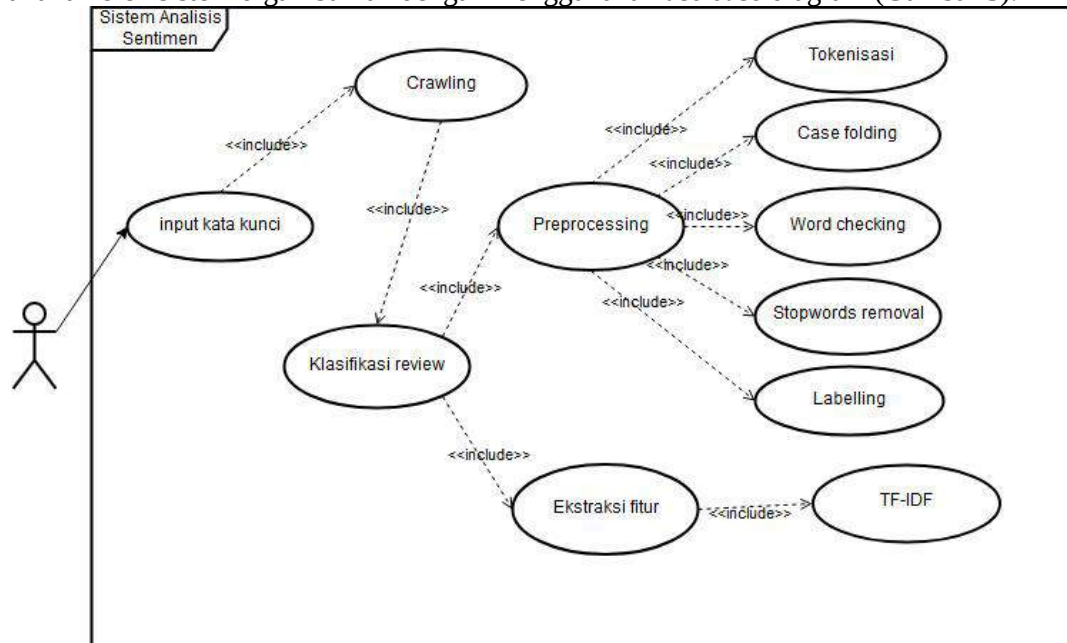
Setelah mengetahui kebutuhan sistem, tahap berikutnya adalah tahap *quick design* untuk membentuk model desain dari sistem yang ingin dibangun. Desain akan dibuat menggunakan *Unified Modelling Language (UML)* yang terdiri dari perancangan *Use Case Diagram* dan *Sequence Diagram*. Flowchart juga digunakan untuk menjelaskan setiap proses pada sistem. UML dan flowchart digunakan untuk membuat model dari sistem yang akan dibangun agar lebih jelas untuk mengetahui proses apa saja yang diperlukan oleh sistem. Tahap selanjutnya adalah pembentukan *prototype*. Pada tahap ini sistem telah dibangun sesuai dengan desain yang telah dibuat dimana sistem yang dibuat telah meliputi fungsi, tampilan, cara kerja yang telah diidentifikasi. Setelah *prototype* terbentuk selanjutnya dilakukan evaluasi pada sistem yang telah dibuat untuk mengetahui apakah sistem sudah layak digunakan dan sesuai kebutuhan user. Setelah memastikan bahwa sistem sudah layak dan sesuai kebutuhan maka sistem siap digunakan. Untuk mempermudah memahami alur kerja sistem arsitektur sistem analisis sentimen akan digambarkan untuk menjelaskan proses dan alur kerja dari sistem (Gambar 2).



Gambar 2 Arsitektur Sistem Analisis Sentimen

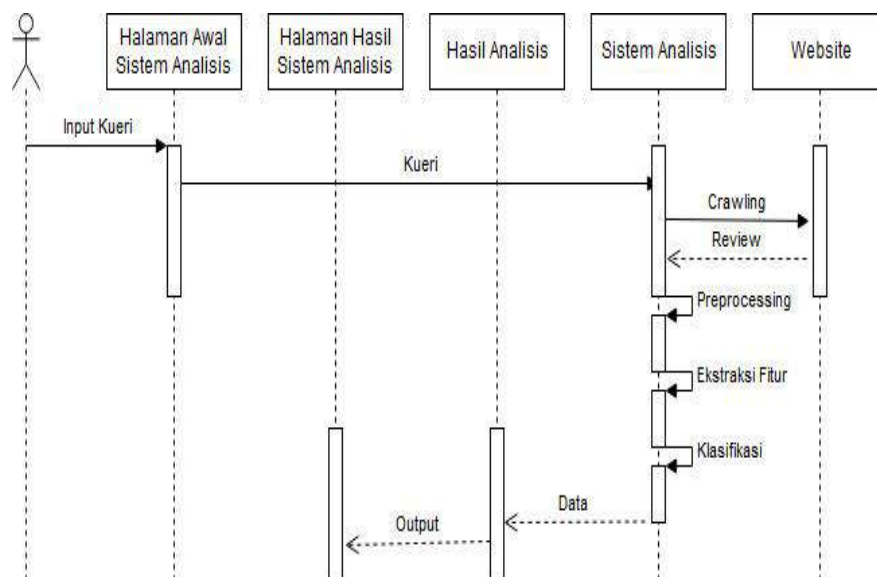
Pada penelitian ini proses penting yang dilakukan oleh sistem adalah pengambilan data dengan menggunakan crawler, preprocessing, ekstraksi fitur menggunakan tf-idf, dan klasifikasi naive bayes. Use case diagram digunakan untuk menjelaskan fungsionalitas dari sistem yang dirancang. Untuk menggunakan sistem ini user dapat melakukan input kueri berupa nama brand, kemudian kueri tersebut diproses dengan melakukan crawling untuk mencari data

review dari nama brand yang bersangkutan. Review yang telah ditemukan diklasifikasi oleh sistem dengan menggunakan metode naive bayes. Dalam proses pengklasifikasian dilakukan preprocessing yang meliputi tokenisasi, case folding, word checking, stopwords removal, labelling, dan kemudian dilakukan ekstraksi fitur dengan metode tf-idf. Setelah semua proses tersebut dilakukan, proses terakhir adalah sistem melakukan klasifikasi review. Proses yang dilakukan oleh sistem digambarkan dengan menggunakan *use case diagram* (Gambar 3).



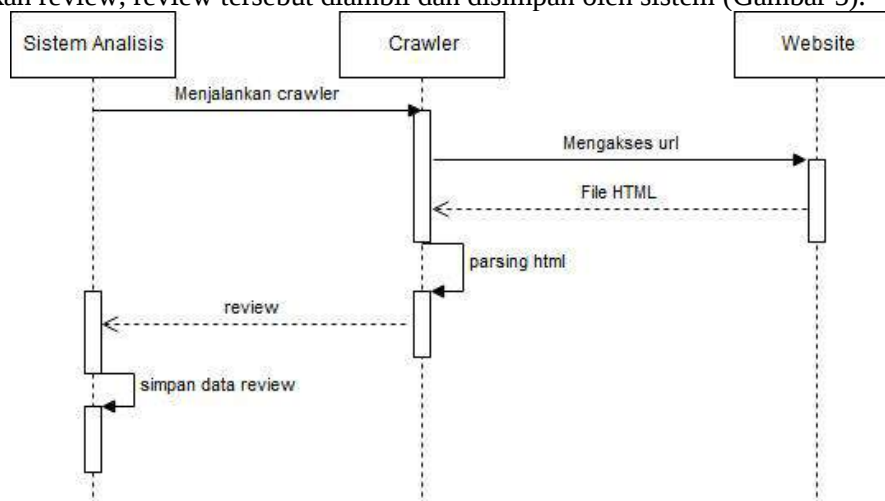
Gambar 3 Diagram Use Case

Sequence diagram digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai suatu respon dari kejadian untuk menghasilkan output tertentu. Berikut beberapa sequence diagram dalam rancangan sistem analisis sentimen ini. Sequence diagram yang menggambarkan cara kerja sistem secara umum. User menginput nama brand pada halaman awal sistem. Kueri diproses dengan melakukan crawling untuk menemukan review dari nama brand inputan. Review yang ditemukan diambil oleh sistem dan dilakukan preprocessing, ekstraksi fitur, dan klasifikasi. Data review yang telah diklasifikasi kemudian ditampilkan pada halaman hasil sistem analisis sentimen (Gambar 4).



Gambar 4 Sequence Diagram Analisis Sentimen

Sequence diagram crawling menggambarkan untuk mengambil review produk crawler mengakses url sumber review. Kemudian halaman yang diakses tersebut akan diambil dalam bentuk file HTML. File HTML yang berhasil didapatkan oleh crawler akan diproses dengan melakukan parsing. Parsing dilakukan untuk mempermudah menemukan informasi yang diperlukan dari file HTML. Informasi yang dibutuhkan adalah teks review. Setelah berhasil menemukan review, review tersebut diambil dan disimpan oleh sistem (Gambar 5).



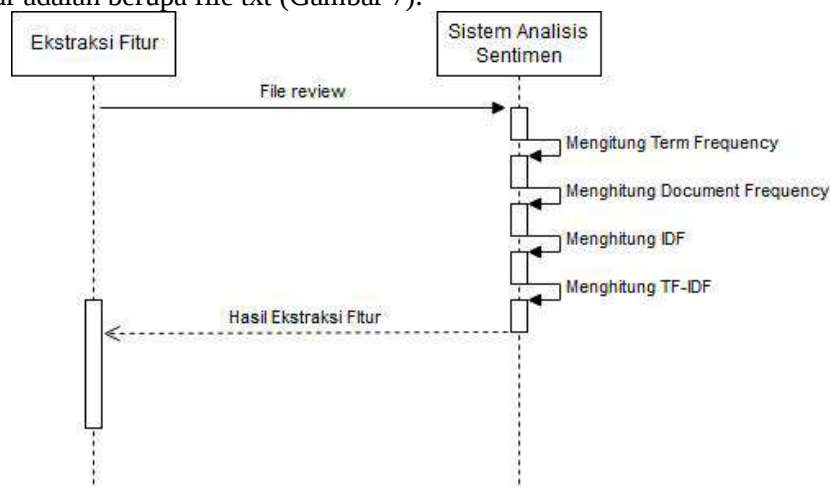
Gambar 5 Sequence Diagram Crawler

Sequence diagram preprocessing menggambarkan bahwa pada tahap preprocessing dimulai dengan mengakses file review yang telah disimpan. Setelah itu data review melalui beberapa tahap, review akan ditokenisasi yaitu dipotong menjadi per karakter, kemudian casefolding dimana semua karakter berupa huruf diubah menjadi huruf kecil, selanjutnya adalah wordchecking dilakukan untuk mengubah kata gaul menjadi kata baku, setelah itu dilakukan stopwords removal yaitu pembuangan kata yang tidak penting dengan menghapus kata tersebut dari data. Proses yang terakhir adalah labelling yaitu mencocokkan kata penguat dan kata negasi dengan kamus yang digunakan (Gambar 6).



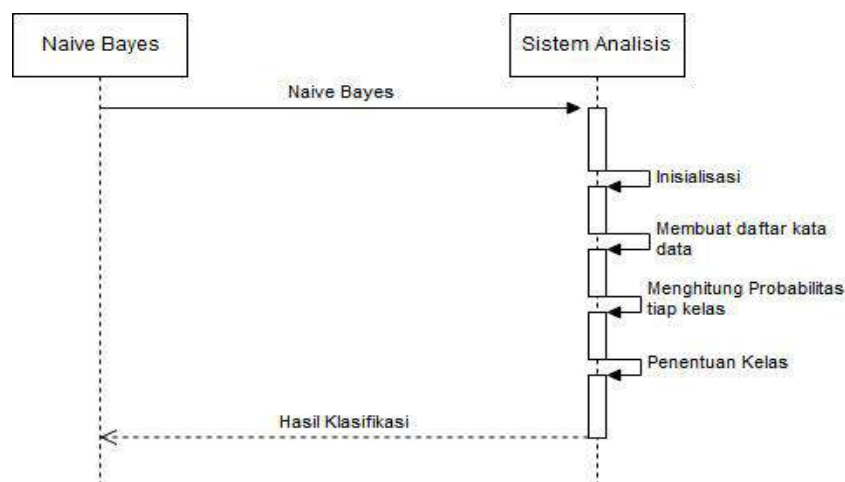
Gambar 6 Sequence Diagram Preprocessing

Sequence diagram ekstraksi fitur menggambarkan pada tahap ekstraksi fitur dilakukan perhitungan dimulai dengan menghitung term frequency (frekuensi kemunculan kata), kemudian menghitung document frequency (frekuensi kemunculan kata pada tiap dokumen), dilanjutkan dengan menghitung IDF, dan terakhir dilakukan perhitungan TF-IDF. Hasil ekstraksi fitur adalah berupa file txt (Gambar 7).



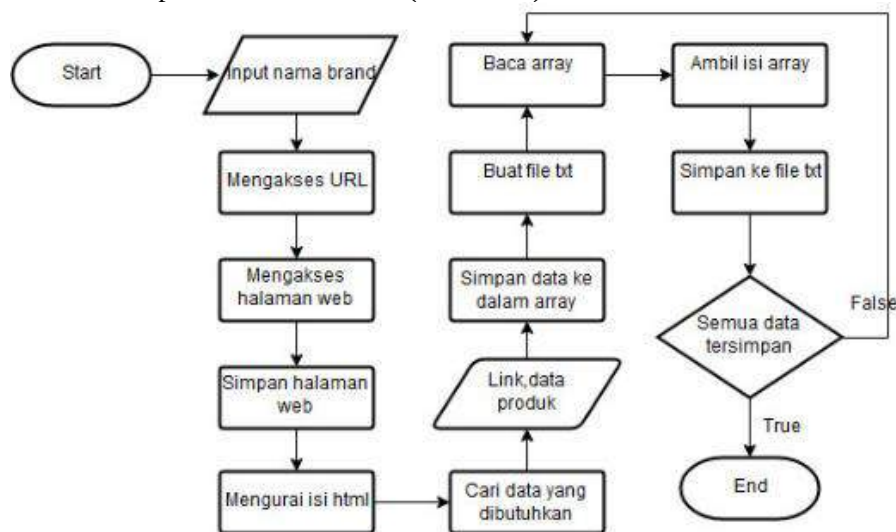
Gambar 7 Sequence Diagram Ekstraksi Fitur

Sequence diagram klasifikasi naive bayes menggambarkan pada tahap ini untuk mendapatkan hasil klasifikasi, akan dilakukan beberapa proses dimulai dari membuat daftar kata dari data review dimana proses ini dilakukan pada tahap ekstraksi fitur karena pada tahap tersebut setiap kata pada data telah dibuat daftarnya, kemudian melakukan proses perhitungan untuk mencari nilai probabilitas tiap kelas, setelah itu review baru dapat diklasifikasi oleh system ke dalam kelasnya masing-masing (Gambar 8).



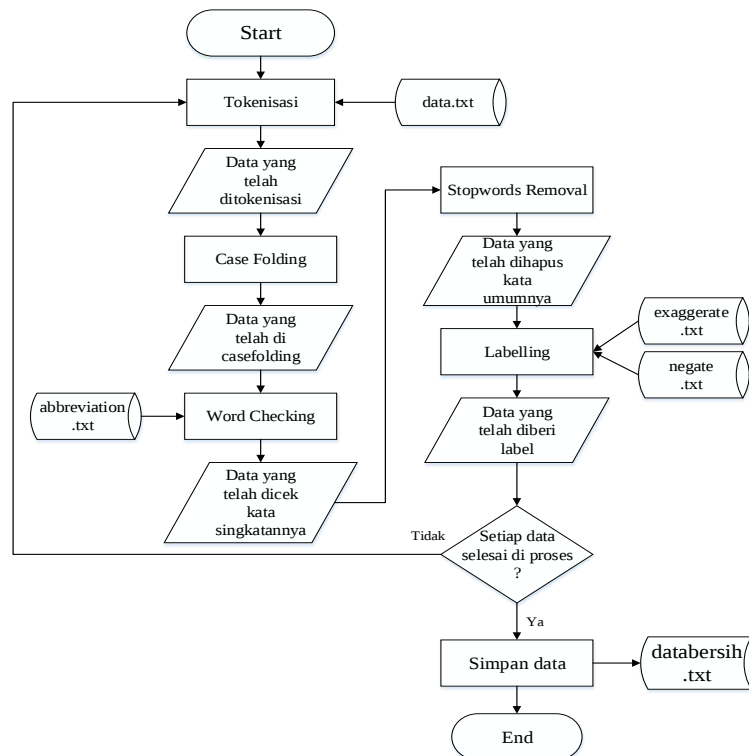
Gambar 8 Sequence Diagram Klasifikasi Naive Bayes

Flowchart digunakan untuk menggambarkan setiap proses yang dilakukan sistem lebih detail dan secara bertahap mulai dari awal berjalannya suatu proses sampai dengan akhir dari proses tersebut. Flowchart digunakan untuk menjelaskan proses apa saja yang dilakukan oleh crawler, pada saat preprocessing, melakukan ekstraksi fitur, dan melakukan klasifikasi. Dimulai dari flowchart crawler yang menggambarkan crawler mengakses url dari website yang diambil data reviewnya. Setelah mengakses url, halaman web tersebut disimpan dengan format html, kemudian dilakukan penguraian isi html untuk mencari data yang diperlukan yaitu link dan data produk serta review. Setelah ditemukan, data tersebut disimpan dalam array dan kemudian dipindahkan dan disimpan ke dalam file txt (Gambar 9).



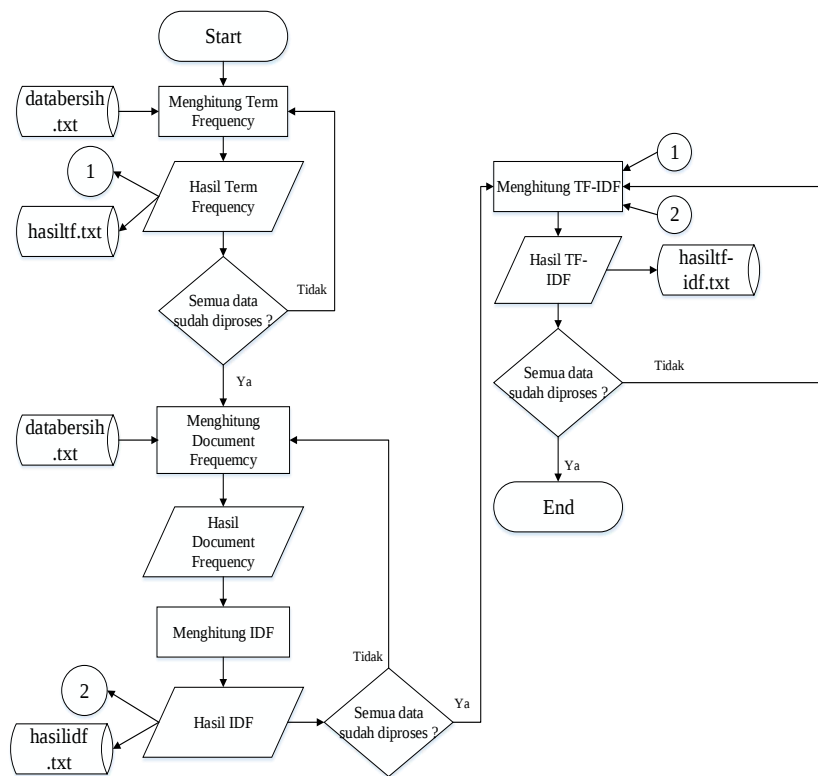
Gambar 9 Flowchart Crawler

Flowchart preprocessing menggambarkan proses yang dilakukan pada tahap ini yaitu tokenisasi proses pemotongan kata menjadi token, case folding mengubah semua huruf menjadi huruf kecil, word checking mengubah kata singkatan menjadi kata baku dengan mencocokkan kata pada review dengan kata pada kamus abbreviation, stopwords removal pembuangan kata umum yang tidak mempengaruhi makna dari suatu kalimat, labelling pemberian label terhadap kata negasi dan kata penguat sesuai dengan kamus kata yang telah disediakan. Setelah itu data disimpan untuk digunakan pada proses selanjutnya (Gambar 10).



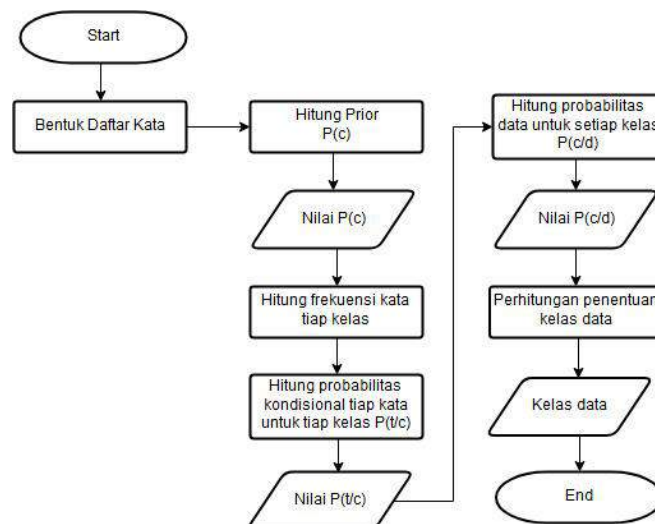
Gambar 10 Flowchart Preprocessing

Flowchart ekstraksi fitur menggambarkan tahapan-tahapan yang dilakukan untuk mendapatkan nilai bobot kata dengan menghitung term frequency, menghitung document frequency, menghitung idf, dan menghitung tf-idf. Hasil perhitungan tf-idf kemudian disimpan dalam bentuk file txt (Gambar 11).



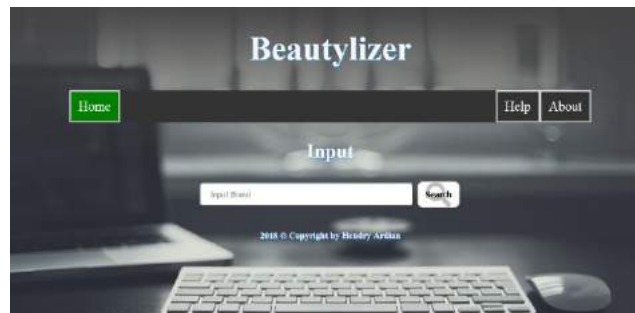
Gambar 11 Flowchart Ekstraksi Fitur

Flowchart klasifikasi naive bayes menggambarkan proses yang dilakukan agar klasifikasi dapat dilakukan dimulai dari perhitungan prior dilanjutkan perhitungan frekuensi kata, kemudian menghitung probabilitas tiap kata untuk tiap kelas, selanjutnya perhitungan probabilitas data untuk setiap kelas, dan terakhir perhitungan untuk menentukan kelas dari data (Gambar 12).



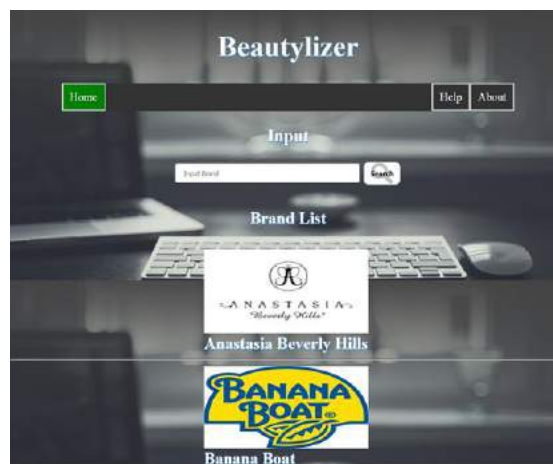
Gambar 12 Flowchart Klasifikasi Naive Bayes

Berikut ini merupakan sejumlah antarmuka dari sistem analisis sentimen yang dihasilkan. Diantaranya adalah halaman utama, halaman brand list, halaman product list, halaman ulasan produk. Rancangan antarmuka halaman utama berisi kotak pencarian untuk menginput nama brand (Gambar 13).



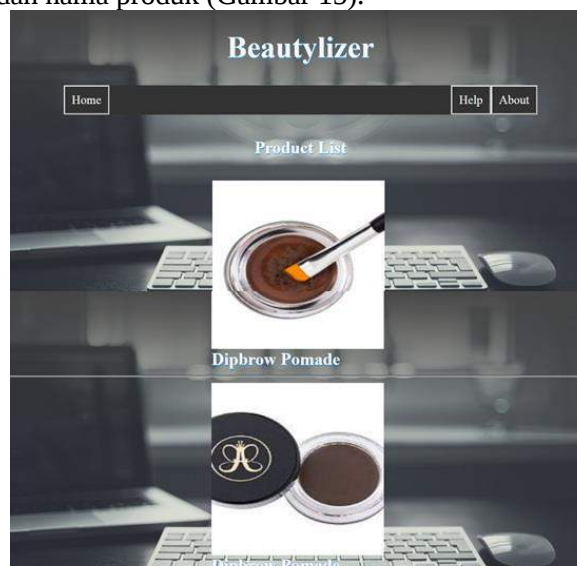
Gambar 13 Halaman Utama

Rancangan antarmuka halaman brandlist digunakan untuk menampilkan brand yang sesuai dengan kueri pencarian. Yang ditampilkan pada halaman ini adalah gambar dan nama brand (Gambar 14).



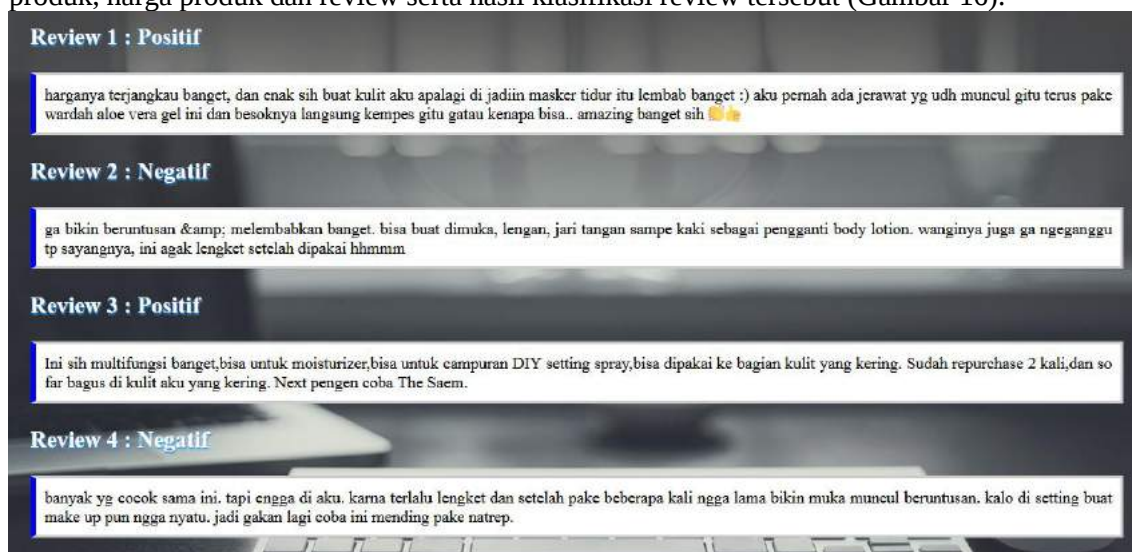
Gambar 14 Halaman Brand List

Rancangan antarmuka halaman product list digunakan untuk menampilkan produk-produk yang termasuk kelompok brand yang dicari. Produk – produk tersebut ditampilkan dalam bentuk gambar dan nama produk (Gambar 15).



Gambar 15 Halaman Product List

Rancangan antarmuka halaman ulasan produk digunakan untuk menampilkan ulasan produk dan hasil klasifikasi sentimen. Pada halaman ini juga memuat gambar produk, nama produk, harga produk dan review serta hasil klasifikasi review tersebut (Gambar 16).



Gambar 16 Halaman Ulasan Produk

Dalam penelitian ini sistem yang dihasilkan dilakukan pengujian terhadap fungsional sistem terkait dengan hal-hal teknis sistem. Tahap ini menggunakan metode pengujian *black box testing*. Berikut adalah tabel pengujian *black box* pada sistem :

Tabel pengujian *black box* pada halaman utama menunjukkan bahwa dari hasil pengujian tersebut didapat bahwa semua skenario pengujian menghasilkan output yang sesuai dengan yang diharapkan (Tabel 1).

Tabel 1 Pengujian Blackbox Halaman Utama

Skenario pengujian	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
Menampilkan halaman utama web	Muncul halaman utama web	Sesuai harapan	Sah
Tombol pada menu bar	Ketika ditekan akan membuka halaman yang dituju	Sesuai harapan	Sah
Menginput namabrand dikotak pencarian	Nama brand dapat diinputkan	Sesuai harapan	Sah
Tombol pencarian bekerja	Melakukan pencarian dengan baik	Sesuai harapan	Sah

Tabel pengujian *black box* pada halaman brandlist menunjukkan bahwa dari hasil pengujian tersebut didapat bahwa semua skenario pengujian menghasilkan output yang sesuai dengan yang diharapkan (Tabel 2).

Tabel 2 Pengujian Blackbox Halaman Brand List

Skenario pengujian	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
Menampilkan daftar brand	Muncul daftar brand	Sesuai harapan	Sah
Tombol pada menu bar	Ketika ditekan akan membuka halaman yang dituju	Sesuai harapan	Sah
Menginput namabrand	Nama brand dapat	Sesuai harapan	Sah

Analisis Sentimen Pada Review Produk Kosmetik Bahasa Indonesia Dengan Metode Naive Bayes

dikotak pencarian	diinputkan		
Tombol pencarian bekerja	Melakukan pencarian dengan baik	Sesuai harapan	Sah

Tabel pengujian *black box* pada halaman productlist menunjukkan bahwa dari hasil pengujian tersebut didapat bahwa semua skenario pengujian menghasilkan output yang sesuai dengan yang diharapkan (Tabel 3).

Tabel 3 Pengujian Blackbox Halaman Product List

Skenario pengujian	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
Menampilkan halaman product list	Product list muncul	Sesuai harapan	Sah
Tombol pada menu bar	Ketika ditekan akan membuka halaman yang dituju	Sesuai harapan	Sah
Menampilkan daftar produk	Muncul gambar dan nama produk	Sesuai harapan	Sah
Gambar produk dan nama produk	Halaman produk ditampilkan	Sesuai harapan	Sah

Tabel pengujian *black box* pada halaman ulasan produk menunjukkan bahwa dari hasil pengujian tersebut didapat bahwa semua skenario pengujian menghasilkan output yang sesuai dengan yang diharapkan (Tabel 4).

Tabel 4 Pengujian Blackbox Halaman Product List

Skenario pengujian	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
Menampilkan halaman review produk	Muncul informasi produk dan review	Sesuai harapan	Sah
Tombol pada menu bar	Ketika ditekan akan membuka halaman yang dituju	Sesuai harapan	Sah
Menampilkan hasil klasifikasi review	Menampilkan review	Sesuai harapan	Sah

Pengujian juga dilakukan pada model naive bayes untuk mengetahui tingkat akurasi sistem dalam mengklasifikasikan review. Pengujian ini terbagi berdasarkan rasio antara data latih dan data uji dengan data uji sebesar 40%, 30%, 20%, dan 10% dari 500 jumlah data. Pengujian akurasi dilakukan menggunakan confusion matrix untuk setiap rasio pembagian data latih dan data uji (Tabel 5).

Tabel 5 Perhitungan Data Uji 40%

Jumlah Data	Kelas Sebenarnya	
	Negatif	Positif
Prediksi	Negatif	68
	Positif	32

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% = \frac{70+68}{70+68+32+30} = 69\% \quad (1)$$

Pengujian perhitungan nilai precision, recall dan akurasi rasio data latih dan data uji 70%:30% (Tabel 6).

Tabel 6 Perhitungan Data Uji 30%

Jumlah Data	Kelas Sebenarnya	
	Negatif	Positif

Prediksi	Negatif	50	24
	Positif	25	51

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{51 + 50}{51 + 50 + 25 + 24} = 67,33\% \quad (2)$$

Pengujian perhitungan nilai precision, recall dan akurasi rasio data latih dan data uji 80%:20% (Tabel 7).

Tabel 7 Perhitungan Data Uji 20%

Jumlah Data	Kelas Sebenarnya	
	Negatif	Positif
Prediksi	Negatif	36
	Positif	14

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{35 + 36}{35 + 36 + 14 + 15} = 71\% \quad (3)$$

Pengujian perhitungan nilai precision, recall dan akurasi rasio data latih dan data uji 90%:10% (Tabel 8).

Tabel 8 Perhitungan Data Uji 10%

Jumlah Data	Kelas Sebenarnya	
	Negatif	Positif
Prediksi	Negatif	18
	Positif	7

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{23 + 18}{23 + 18 + 7 + 2} = 82\% \quad (4)$$

Daftar rangkuman hasil pengujian perhitungan akurasi dari setiap rasio data latih dan data uji (Tabel 9).

Tabel 9 Nilai Akurasi

Persentase Data Latih	Persentase Data Uji	Jumlah Data Latih	Jumlah Data Uji	Akurasi
60%	40%	300	200	69%
70%	30%	350	150	67,33%
80%	20%	400	100	71%
90%	10%	450	50	82%

4. KESIMPULAN

Sistem analisis sentimen yang dirancang dapat berjalan sesuai dengan fungsinya dan membantu user untuk membaca review lebih mudah dan menghemat waktu yang diperlukan untuk membaca review. Akan tetapi sistem yang telah dibangun memiliki beberapa kelemahan yaitu membutuhkan data latih dalam jumlah banyak, seringkali terjadi kesalahan pengklasifikasian sentimen saat terdapat sentimen positif dan sentimen negatif dalam satu review, dan juga sistem hanya dapat mengambil review dalam jumlah terbatas. Performa sistem analisis sentimen yang dirancang cukup baik berdasarkan hasil pengujian akurasi yang telah dilakukan. Dari perhitungan akurasi diketahui bahwa sistem memiliki performa paling baik untuk mengklasifikasi sentimen dengan menggunakan rasio data latih dan data uji 90%:10% dengan nilai akurasi sebesar 82%. Dapat disimpulkan bahwa dengan jumlah data latih yang lebih banyak dan beragam maka performa sistem analisis sentimen akan semakin baik.

5. SARAN

Diharapkan untuk penelitian selanjutnya, pengembangan sistem analisis sentimen perlu memperhatikan dan melakukan peningkatan pada bagian preprocessing agar mengurangi jumlah

variabel yang tidak penting yang lolos dari tahap preprocessing yang kemudian menyebabkan hambatan tertentu saat proses klasifikasi. Diharapkan juga membuat sistem analisis sentimen yang dapat menganalisis sentimen untuk setiap kalimat untuk menghindari kekeliruan pengklasifikasian sentimen. Diharapkan juga untuk mengembangkan crawler agar dapat mengambil review dengan jangkauan yang lebih luas dan jumlah yang banyak. Untuk penelitian selanjutnya dapat juga mencoba menggunakan metode lain seperti SVM dan ANN.

DAFTAR PUSTAKA

- [1] Ling, J. Eka,N.K., Oka,T.B., 2014, Analisis Sentimen menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur Chi Square, *E-Jurnal Matematika Vol. 3 (3)*, pp. 92-99.
- [2] Saraswati, N.W.S., 2011, Text Mining dengan Metode Naïve Bayes Classifier dan Support Vector Machine untuk Sentimen Analysis, *Tesis*, Program Studi Teknik Elektro, Program Pasca Sarjana, Universitas Udayana, Denpasar.
- [3] Zhang, L., Ghosh, R., Dekhil, M., Hsu, M. & Liu,B., 2011. Combining Lexiconbased and Learning-based Methods for Twitter Sentiment Analysis, Chicago: Hewlett-Packard Development Company, L.P.
- [4] Jinturkar ,M., Gotmare ,P., 2016, Sentiment Analysis of Customer Review Data using Big Data: A Survey, *International Journal of Computer Applications (0975 – 8887)*, 3-8.
- [5] Kaur.G., Singla.A., 2016, Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm, *International Journal of Advanced Research in Computer Engineering & Technology*, Vol.5, Issue 1, 148-153.
- [6] Mubarak.M.S, Adiwijaya, Aldhi.M.D., 2017, Aspect-based sentiment analysis to review products using Naïve Bayes, *International Conference on Mathematics: Pure, Applied and Computation*, 1-8.
- [7] Dey.L., Chakraborty.S., Biswas.A., Bose.B., Tiwari.S., 2016, Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier, *I.J. Information Engineering and Electronic Business*, no.4, 54-62.
- [8] Shaikh.T., Deshpande., D, 2016, Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews, *International Journal of Computer Trends and Technology*, vol. 36, no.4, 225-230.
- [9] Gurneet.K., and Abhinash.S., 2016, Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm, *International Journal Of Advanced Research in Computer Engineering & Technology (IJARCET)*, Vol. 5, ISSN: 2278- 1323.
- [10] S. Pressman, Roger.,2010.*Software Engineering A Practitioner's Approach*. Seventh Edition.
- [11] Pressman, R.S.Ph.D., 2012, *Rekayasa Perangkat Lunak (Pendekatan Praktisi) Edisi 7 : Buku 1* , Andi, Yogyakarta.
- [12] Haviluddin , 2011, *Memahami Penggunaan UML (Unified Modeling Language)*, Vol 6 No 1, Universitas Mulawarman, Samarinda.