

**APLIKASI OPINION MINING DENGAN
ALGORITMA NAÏVE BAYES UNTUK MENILAI
BERITA ONLINE**

TUGAS AKHIR

Oleh :

Daniel Pakpahan 3311201020

Disusun untuk memenuhi syarat kelulusan Program Diploma III



**PROGRAM STUDI TEKNIK INFORMATIKA
POLITEKNIK NEGERI BATAM
BATAM
2015**

HALAMAN PENGESAHAN

**APLIKASI OPINION MINING DENGAN ALGORITMA NAÏVE BAYES
UNTUK MENILAI BERITA ONLINE**

Oleh :

Daniel Pakpahan 3311201020

Tugas Akhir ini telah diterima dan disahkan
sebagai persyaratan untuk memperoleh gelar

Ahli Madya

di

**PROGRAM STUDI DIPLOMA III TEKNIK INFORMATIKA
POLITEKNIK NEGERI BATAM**

Batam, 31 Januari 2015

Disetujui oleh;

Pembimbing,

Hilda Widyastuti, M.T

NIK. 197705122012122001

HALAMAN PERNYATAAN

Dengan ini, saya:

Nama : Daniel Pakpahan

NIM : 3311201020

adalah mahasiswa Teknik Informatika Politeknik Batam yang menyatakan bahwa tugas akhir dengan judul:

APLIKASI OPINION MINING DENGAN ALGORITMA NAÏVE BAYES UNTUK MENILAI BERITA ONLINE

disusun dengan:

1. Tidak melakukan plagiat terhadap naskah karya orang lain.
2. Tidak melakukan pemalsuan data
3. Tidak menggunakan karya orang lain tanpa menyebut sumber asli atau tanpa izin pemilik.

Jika kemudian terbukti terjadi pelanggaran terhadap pernyataan di atas, maka saya bersedia menerima sanksi apapun termasuk pencabutan gelar akademik.

Lembar pernyataan ini juga memberikan hak kepada Politeknik Batam untuk mempergunakan, mendistribusikan ataupun memproduksi ulang seluruh hasil Tugas Akhir ini.

Batam, 31 Januari 2015

Daniel Pakpahan

3311201020

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa, atas segala nikmat dan karunia-Nya yang tak terhingga sehingga Laporan Tugas Akhir yang berjudul “Aplikasi *Opinion Mining* dengan Algoritma Naïve Bayes untuk menilai Berita *Online*” ini dapat terselesaikan.

Laporan yang ditujukan guna melengkapi syarat kelulusan Program Studi Teknik Informatika Politeknik Negeri Batam ini dalam pengerjaannya mendapat banyak arahan dan bimbingan dari berbagai pihak, oleh karena itu dalam kesempatan ini penulis hendak mengucapkan ungkapan terima kasih yang tak terhingga kepada:

1. Tuhan Yang Maha Esa yang senantiasa memberikan limpahan anugerah dan nikmat-Nya berupa kesehatan yang selalu mengiringi penulis.
2. Kedua orangtua yang selalu menjadi inspirasi dan memberikan semangat, kasih sayang serta dukungan kepada penulis.
3. Kakak dan abang yang telah menjadi saudara-saudara yang luar biasa.
4. Ibu Hilda Widyastuti, M.T, selaku pembimbing Tugas Akhir yang telah meluangkan banyak waktu dan memberikan motivasi, bimbingan serta arahan demi kelancaran Tugas Akhir ini.
5. Bapak Dwi Ely Kurniawan, M. Kom, selaku koordinator Tugas Akhir yang selalu memberikan motivasi, arahan, dan semangat dalam pengerjaan Tugas Akhir ini.
6. Teman-teman seperjuangan TA yang tidak bisa penulis sebutkan satu-persatu.
7. Para dosen, sahabat-sahabat tercinta, dan seluruh pihak yang ikut serta memberikan motivasi dan membantu dalam menyelesaikan laporan Tugas Akhir ini.

Penulis menyadari selama pelaksanaan Tugas Akhir ini masih banyak hal-hal yang dapat dipelajari dan semuanya tidak lepas dari kekurangan dan kesalahan. Akhir

kata mohon maaf atas kesalahan dan kekurangan serta keterbatasan baik pada aplikasi maupun pada dokumentasi. Semoga aplikasi ini bermanfaat dan bisa dikembangkan pada masa yang akan datang.

Batam, Januari 2015

Penulis

ABSTRAK

APLIKASI OPINION MINING DENGAN ALGORITMA NAÏVE BAYES UNTUK MENILAI BERITA ONLINE

Opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Salah satu metode *text mining* yang bisa digunakan untuk menyelesaikan masalah *opinion mining* adalah *Naïve Bayes Classifier* (NBC).

Sumber data yang akan diolah dalam proses klasifikasi adalah data opini atau komentar pada berita *online*. Sebelum data opini atau komentar diolah ke dalam proses klasifikasi, tahap pertama yang harus dilewati ada proses *text preprocessing* berupa *tokenizing*, *filtering*, dan *stemming*. Tahap selanjutnya adalah menghasilkan model *probabilistic* yang nilainya akan digunakan pada proses klasifikasi.

Proses inti adalah proses klasifikasi untuk menentukan probabilitas tertinggi dari tiap kategori. Jika hasil menunjukkan probabilitas Bayes komentar untuk kategori positif lebih besar maka komentar tersebut masuk kategori opini positif demikian juga sebaliknya.

Kata kunci: *opinion mining*, naïve bayes, *text preprocessing*, klasifikasi, model *probabilistic*

ABSTRACT

OPINION MINING APPLICATIONS WITH NAÏVE BAYES ALGORITHM TO ASSES NEWS ONLINE

Opinion mining is the process of understanding, extracting and processing textual data automatically to get the sentiment of information contained in an opinion sentence. One of text mining methods that can be used to solve the problem of opinion mining is the Naïve Bayes Classifier (NBC).

Source of data to be processed in the process of data classification is the opinion or comment on the news online. Before the opinion or comment data is processed into the classification process, the first step that must be passed no such tokenizing the text preprocessing, filtering, and stemming. The next stage is to produce probabilistic models whose value will be used in the classification process.

The process is the core of the classification process to determine the highest probability of each category. If the results indicate the probability Bayes comments for positive category is larger then the comment is categorized as a positive opinion and vice versa.

Keywords: opinion mining, naïve bayes, text preprocessing, classification, probabilistic models

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
KATA PENGANTAR	iv
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan.....	3
1.5 Sistematika Penulisan	3
BAB II LANDASAN TEORI	4
2.1 Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis.....	4
2.2 <i>Opinion Mining</i>	5
2.3 Text Mining	5
2.4 Algoritma Naive Bayes Classification (NBC).....	7
2.5 MySQL dan PHP	10
2.6 HTML.....	10
BAB III ANALISIS DAN PERANCANGAN	12
3.1 Deskripsi Umum Sistem.....	12
3.2 Kebutuhan Fungsional dan Nonfungsional	14
3.2.1 Kebutuhan Fungsional.....	14
3.2.2 Kebutuhan Nonfungsional	15
3.3 Use Case Diagram	16

3.4	Skenario <i>Use Case</i>	17
3.4.1	Skenario <i>Use Case</i> Login	17
3.4.2	Skenario <i>Use Case</i> Melakukan <i>Text Preprocessing</i>	18
3.4.3	Skenario <i>Use Case</i> Membuat <i>Model Probabilistic</i>	19
3.4.4	Skenario <i>Use Case</i> Melakukan Klasifikasi	20
3.4.5	Skenario <i>Use Case</i> Melihat Kesimpulan	21
3.4.6	Skenario <i>Use Case</i> Input Komentar	21
3.4.7	Skenario <i>Use Case</i> Atur Berita.....	22
3.5	Diagram <i>Sequence</i>	23
3.5.1	Diagram <i>Sequence</i> Login.....	23
3.5.2	Diagram <i>Sequence</i> Melakukan <i>Text Preprocessing</i>	24
3.5.3	Diagram <i>Sequence</i> Membuat Model Probabilistic	24
3.5.4	Diagram <i>Sequence</i> Melakukan Klasifikasi.....	25
3.5.5	Diagram <i>Sequence</i> Melihat Kesimpulan	26
3.5.6	Diagram <i>Sequence</i> Atur Berita	26
3.5.7	Diagram <i>Sequence</i> Input Komentar.....	27
3.6	Diagram <i>Class</i>	28
3.7	Perancangan Basisdata	29
3.7.1	ER Diagram	29
3.7.2	Desain Basisdata.....	30
3.8	Perancangan Antarmuka Sistem.....	31
3.8.1	Rancangan Antarmuka Halaman Login	31
3.8.2	Rancangan Antarmuka Halaman Text Preprocessing	32
3.8.3	Rancangan Antarmuka Halaman Membuat Model Probabilistic	33
3.8.4	Rancangan Antarmuka Halaman Melakukan Klasifikasi.....	34
3.8.5	Rancangan Antarmuka Melihat Kesimpulan.....	35
3.8.6	Rancangan Antarmuka Input Komentar	35
3.8.7	Rancangan Antarmuka Halaman Atur Berita.....	36
BAB IV IMPLEMENTASI DAN PENGUJIAN.....		37
4.1	Implementasi Struktur Tabel Basisdata	37
4.1.1	Tabel Categories	37

4.1.2	Tabel Comments	37
4.1.3	Tabel Katadasar	37
4.1.4	Tabel News	38
4.1.5	Tabel Pie	38
4.1.6	Tabel References	38
4.1.7	Tabel User	38
4.1.8	Tabel Wordfreqs	39
4.2	Implementasi Class	39
4.3	Implementasi Antarmuka	40
4.3.1	Implementasi Antarmuka Halaman Homepage	40
4.3.2	Implementasi Antarmuka Halaman <i>Login Admin</i>	41
4.3.3	Implementasi Antarmuka Halaman Adminpage	41
4.3.4	Implementasi Antarmuka Halaman <i>Tokenizing</i>	42
4.3.5	Implementasi Antarmuka Halaman <i>Filtering</i>	43
4.3.6	Implementasi Antarmuka Halaman <i>Stemming</i>	43
4.3.7	Implementasi Antarmuka Halaman Model <i>Probabilistic</i>	44
4.3.8	Implementasi Antarmuka Halaman Klasifikasi	45
4.3.9	Implementasi Antarmuka Halaman Atur Berita	45
4.3.10	Implementasi Antarmuka Komentar	46
4.3.11	Implementasi Antarmuka Melihat Kesimpulan (<i>View Details</i>) ..	47
4.4	Pengujian	48
4.4.1	Strategi Pengujian	48
4.4.2	Deskripsi Pengujian	48
4.4.3	Hasil Pengujian	49
BAB V KESIMPULAN DAN SARAN		55
5.1	Kesimpulan	55
5.2	Saran	55
DAFTAR PUSTAKA		56

DAFTAR GAMBAR

Gambar III-1 Deskripsi Umum Sistem	12
Gambar III-2 Flowchart Proses Membuat Model	13
Gambar III-3 Flowchart Proses Klasifikasi.....	14
Gambar III-4 Diagram <i>Use Case</i>	16
Gambar III-5 Diagram Sequence Login.....	23
Gambar III-6 Diagram Sequence Melakukan <i>Text Preprocessing</i>	24
Gambar III-7 Diagram Sequence Membuat Model Probabilistic	24
Gambar III-8 Diagram Sequence Melakukan Klasifikasi.....	25
Gambar III-9 Diagram Sequence Melihat Kesimpulan	26
Gambar III-10 Diagram Sequence Atur Berita	26
Gambar III-11 Diagram Sequence Input Komentar.....	27
Gambar III-12 Class Diagram.....	28
Gambar III-13 ER Diagram	29
Gambar III-14 Desain Basisdata	30
Gambar III-15 Antarmuka Halaman Login.....	31
Gambar III-16 Rancangan Antarmuka Halaman <i>Text Preprocessing</i>	32
Gambar III-17 Rancangan Antarmuka Halaman Membuat Model <i>Probabilistic</i> . 33	
Gambar III-18 Rancangan Antarmuka Halaman Melakukan Klasifikasi	34
Gambar III-19 Rancangan Antarmuka Melihat Kesimpulan.....	35
Gambar III-20 Rancangan Antarmuka Input Komentar	35
Gambar III-21 Rancangan Antarmuka Halaman Atur Berita	36
Gambar IV-1 Halaman Homepage.....	40
Gambar IV-2 Halaman <i>Login Admin</i>	41
Gambar IV-3 Halaman <i>Adminpage</i>	42
Gambar IV-4 Halaman <i>Tokenizing</i>	42
Gambar IV-5 Halaman <i>Filtering</i>	43
Gambar IV-6 Halaman <i>Stemming</i>	43
Gambar IV-7 Halaman Model <i>Probabilistic</i>	44
Gambar IV-8 Halaman Klasifikasi.....	45

Gambar IV-9 Halaman Atur Berita.....	46
Gambar IV-10 Antarmuka Komentar	46
Gambar IV-11 Antarmuka View Details	47

DAFTAR TABEL

Tabel II.1 Perbandingan metode NBC dan SVM.....	4
Tabel III.1 Skenario <i>Use Case</i> Login.....	17
Tabel III.2 Skenario <i>Use Case</i> Melakukan Text Preprocessing	18
Tabel III.3 Skenario <i>Use Case</i> Membuat Model Probabilistic	19
Tabel III.4 Skenario <i>Use Case</i> Melakukan Klasifikasi	20
Tabel III.5 Skenario <i>Use Case</i> Melihat Kesimpulan	21
Tabel III.6 Skenario <i>Use Case</i> Input Komentar.....	21
Tabel III.7 Skenario <i>Use Case</i> Atur Berita	22
Tabel III.8 Deskripsi Halaman Login	31
Tabel III.9 Deskripsi Halaman <i>Text Preprocessing</i>	32
Tabel III.10 Deskripsi Halaman Membuat Model <i>Probabilistic</i>	33
Tabel III.11 Deskripsi Halaman Melakukan Klasifikasi.....	34
Tabel III.12 Deskripsi Antarmuka Melihat Kesimpulan.....	35
Tabel III.13 Deskripsi Antarmuka Input Komentar	35
Tabel III.14 Deskripsi Halaman Atur Berita.....	36
Tabel IV.1 Tabel Categories	37
Tabel IV.2 Tabel Comments	37
Tabel IV.3 Tabel Katadasar	37
Tabel IV.4 Tabel News	38
Tabel IV.5 Tabel Pie	38
Tabel IV.6 Tabel References.....	38
Tabel IV.7 Tabel User.....	39
Tabel IV.8 Tabel Wordfreqs	39
Tabel IV.9 Implementasi Class	39
Tabel IV.10 Hasil Pengujian	49

BAB I PENDAHULUAN

1.1 Latar Belakang

Opini orang lain dapat menjadi informasi sangat penting ketika saatnya untuk membuat keputusan atau memilih di antara beberapa pilihan. Informasi telah menjadi kebutuhan utama dalam kehidupan manusia. Informasi bisa dikatakan sebagai pengetahuan yang didapatkan dari pembelajaran, pengalaman, atau instruksi. Selain itu, pengetahuan tentang peristiwa-peristiwa tertentu atau situasi yang telah dikumpulkan atau diterima melalui proses komunikasi, pengumpulan intelegen, ataupun didapatkan dari berita juga dinamakan informasi. Berita merupakan informasi baru atau informasi mengenai sesuatu yang sedang terjadi, disajikan lewat bentuk cetak, siaran, *internet*, atau dari mulut ke mulut kepada orang ketiga atau orang banyak.

Di era perkembangan teknologi ini, media-media berita seperti koran harian semisal Kompas telah memiliki *website*. Sehingga memudahkan para pengguna media berita *online* untuk bisa mengakses dan membaca berita kapanpun diinginkan bahkan memberikan komentar. Menurut Ian Barber (2010), Web adalah tempat yang baik bagi orang-orang untuk mengekspresikan pendapat mereka, pada berbagai topik. Bahkan pemberi opini secara profesional, seperti reviewer film, memiliki blog dimana publik dapat mengomentari dan merespon apa yang mereka pikirkan. Kemampuan untuk mengekstrak pendapat tersebut dari baris-baris teks dapat menjadi sangat berguna, dan ini adalah area studi yang banyak dikaji, tidak diragukan karena kemungkinan nilai komersialnya. Baris-baris teks tersebut dapat diolah untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. *Opinion mining* dilakukan untuk melihat pendapat atau kecenderungan opini terhadap masalah atau topik berita oleh seseorang, apakah cenderung beropini negatif, positif atau netral. Sehingga diharapkan opini-opini yang terkumpul dapat menjadi informasi yang berguna.

Informasi yang terdapat pada *berita online* merupakan informasi data teks digital dan dalam bentuk tidak terstruktur. *Text mining* menjadi sangat diperlukan untuk menangani teks yang tidak terstruktur tersebut. *Text mining* mengacu pada

proses mengambil informasi berkualitas tinggi dari teks. Kegiatan penting dalam *text mining* adalah klasifikasi atau kategorisasi. Metode klasifikasi yang digunakan dalam penelitian ini adalah metode *naive bayes* yang sering disebut *Naive Bayes Classifier* (NBC). Ian Barber dalam *Bayesian Opinion Mining*, 2010 telah melakukan eksperimen untuk data *review* film dan menghasilkan tingkat akurasi 80% menggunakan metode NBC. Saraswati (2011), dalam penelitiannya menunjukkan NBC merupakan metode yang konvensional dan lebih sederhana dari metode lain dalam proses teks *mining* dan NBC memberikan hasil yang tepat dalam mengklasifikasikan opini dalam bentuk paragraf yang terdiri dari beberapa kalimat. Sehingga dengan adanya penelitian tugas akhir ini dapat dihasilkan kesimpulan opini positif, negatif atau netral dari opini-opini yang terdapat pada berita *online*.

1.2 Rumusan Masalah

Adapun pokok masalah yang terdapat dalam penelitian ini, antara lain :

1. Bagaimana merancang aplikasi *opinion mining* dengan metode *naive bayes* untuk menilai berita *online*?
2. Bagaimana membuat aplikasi *opinion mining* dengan metode *naive bayes* untuk menilai berita *online*?

1.3 Batasan Masalah

Adapun batasan masalah yang terdapat dalam penelitian ini, antara lain :

1. Berita yang digunakan dalam penelitian adalah berita bahasa Indonesia.
2. Masukan komentar menggunakan format bahasa Indonesia yang baik dan benar.
3. *Link* berita *online* yang di *scrape* berasal dari situs news.kompas.com.

1.4 Tujuan

Adapun tujuan penelitian yang terdapat dalam penelitian ini, antara lain :

1. Merancang aplikasi *opinion mining* dengan metode *naive bayes* untuk menilai berita *online*.
2. Membuat aplikasi *opinion mining* dengan metode *naive bayes* untuk menilai berita *online*.

1.5 Sistematika Penulisan

Sistematika penyusunan Proyek Akhir II ini disusun berdasarkan:

- Bab I : Pendahuluan. Berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan untuk memberikan gambaran isi laporan tugas akhir ini.
- Bab II : Tinjauan pustaka. Berisi tentang teori yang akan digunakan sebagai landasan pengerjaan tugas akhir.
- Bab III : Analisis dan perancangan. Berisi tentang deskripsi umum sistem, kebutuhan fungsional dan nonfungsional, *use case* sistem, *sequence diagram*, dan *class diagram*.
- Bab IV : Implementasi dan pembahasan. Berisi tentang implementasi dan pembahasan tampilan, *coding*, dan uji coba.
- Bab V : Kesimpulan dan saran dari pembuatan dan pengembangan aplikasi.

BAB II LANDASAN TEORI

2.1 Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis

Penelitian dilakukan oleh Saraswati (2011) yang melakukan *text mining* dengan metode *Naive Bayes Classifier* dan *Support Vector Machine* dalam mengklasifikasikan opini berbahasa Inggris dan berbahasa Indonesia. Penelitian menunjukkan bahwa kedua metode tersebut (*Naive Bayes Classifier* dan *Support Vector Machine*) banyak digunakan dalam kategorisasi teks. Penelitian tersebut ingin mengetahui metode yang mana yang memiliki performansi yang lebih baik untuk diimplementasikan dalam *sentiment analysis* opini berbahasa Inggris dan berbahasa Indonesia. Beberapa metode lain yang juga dapat digunakan untuk proses teks *mining* adalah C45, K-Nearest Neighbor, K-Means dan algoritma genetika.

Kesimpulan dari penelitian tersebut menunjukkan metode SVM memberikan unjuk kerja yang lebih baik daripada metode NBC untuk mengklasifikasikan opini berbahasa Inggris dan opini positif berbahasa Indonesia. Sedangkan, NBC memberikan unjuk kerja yang lebih baik dalam mengklasifikasikan data uji opini negatif berbahasa Indonesia.

Tabel II.1 Perbandingan metode NBC dan SVM

Metode	Opini Positif		Opini Negatif	
	Bahasa Indonesia	Bahasa Inggris	Bahasa Indonesia	Bahasa Inggris
<i>Naive Bayes Classifier</i>	74.29%	80.18%	87.14%	83.66%
<i>Support Vector Machine</i>	78.20%	80.15	78.14%	98.95

Terkait dengan penelitian sebelumnya yang membandingkan metode NBC dan SVM, penulis menambahkan data kategori opini netral ke dalam penelitian tugas akhir. Sehingga terdapat 3 kategori data opini, yaitu opini positif, opini negatif dan opini netral. Penulis juga merancang sebuah sistem berbasis web dimana sistem tersebut dapat mengklasifikasikan opini secara otomatis sehingga pengujian klasifikasi opini dapat lebih mudah dilakukan.

2.2 *Opinion Mining*

Menurut (Rozi *et al.*, 2012), *opinion mining* atau *sentiment analysis* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang menfokuskan pada layanan analisis sentimen.

Opinion mining bisa dianggap sebagai kombinasi antara *text mining* dan *natural language processing*. Salah satu metode dari *text mining* yang bisa digunakan untuk menyelesaikan masalah *opinion mining* adalah *Naive Bayes Classifier* (NBC). NBC bisa digunakan untuk mengklasifikasikan opini ke dalam opini positif atau negatif. NBC bisa berfungsi dengan baik sebagai metode pengklasifikasi teks. Penelitian tentang NBC sebagai metode pengklasifikasi teks telah dilakukan oleh SM Kamaruzzaman dan Chowdury Mofizur Rahman serta Ashraf M Kibriya *et.al* pada tahun 2004. Dari proses pengujian secara kualitatif disebutkan bahwa teks bisa diklasifikasikan dengan akurasi yang tinggi.

2.3 **Text Mining**

Text mining memiliki definisi menambang data yang berupa teks di mana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata - kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. *Text mining* juga merupakan proses penemuan akan informasi atau trend baru yang sebelumnya tidak terungkap dengan memproses dan menganalisa data dalam jumlah besar.

Menurut Adiwijaya (dalam Nur Indrawati, 2009) disebutkan bahwa *text mining* bisa dianggap subjek riset yang tergolong baru. *Text mining* dapat

memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian / pengelompokkan dan menganalisis *unstructured text* dalam jumlah besar.

Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti *Data mining*, *Information Retrieval*, Statistik dan Matematik, *Machine Learning*, *Linguistic*, *Natural Language Processing*, dan *Visualization*. Kegiatan penelitian untuk *text mining* antara lain ekstraksi dan penyimpanan teks, *preprocessing* akan konten teks, pengumpulan data statistik dan *indexing* dan analisis konten.

Dalam menganalisis sebagian atau keseluruhan *unstructured text*, *text mining* mencoba untuk mengasosiasikan satu bagian *text* dengan yang lainnya berdasarkan aturanaturan tertentu. Hasil yang di harapkan adalah informasi baru atau *insight* yang tidak terungkap jelas sebelumnya. Menurut Saraswati (2011), saat ini *text mining* telah mendapat perhatian dalam berbagai bidang, antara lain:

1. Aplikasi keamanan

Banyak paket perangkat lunak *text mining* dipasarkan terhadap aplikasi keamanan, khususnya analisis *plain text* seperti berita internet.

2. Aplikasi biomedis

Berbagai aplikasi *text mining* dalam literature biomedis telah disusun. Salah satu contohnya adalah PubGene yang mengkombinasikan *text mining* biomedis dengan visualisasi jaringan sebagai sebuah layanan internet.

3. Perangkat lunak dan aplikasi

Departemen riset dan pengembangan perusahaan besar, termasuk IBM dan Microsoft, sedang meneliti teknik *text mining* dan mengembangkan program untuk lebih mengotomatisasi proses pertambangan dan analisis. Perangkat lunak *text mining* juga sedang diteliti oleh perusahaan yang berbeda yang bekerja di bidang pencarian dan pengindeksan secara umum sebagai cara untuk meningkatkan performansinya.

4. Aplikasi media *online*

Text mining sedang digunakan oleh perusahaan media besar, seperti perusahaan Tribune, untuk menghilangkan ambigu informasi dan untuk memberikan pembaca dengan pengalaman pencarian yang lebih baik, yang

meningkatkan loyalitas pada *site* dan pendapatan. Selain itu, editor diuntungkan dengan mampu berbagi, mengasosiasi dan properti paket berita, secara signifikan meningkatkan peluang untuk menguangkan konten.

5. Aplikasi pemasaran

Text mining juga mulai digunakan dalam pemasaran, lebih spesifik dalam analisis manajemen hubungan pelanggan.

6. Aplikasi akademik

Masalah *text mining* penting bagi penerbit yang memiliki database besar untuk mendapatkan informasi yang memerlukan pengindeksan untuk pencarian. Hal ini terutama berlaku dalam ilmu sains, dimana informasi yang sangat spesifik sering terkandung dalam teks tertulis.

2.4 Algoritma Naive Bayes Classification (NBC)

Algoritma *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dalam penelitian ini yang menjadi data uji adalah dokumen opini. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Dalam algoritma *naive bayes classifier* setiap dokumen direpresentasikan dengan pasangan atribut "x₁, x₂, x₃, ...x_n" dimana x₁ adalah kata pertama, x₂ adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori opini. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{map}), dimana persamaanya adalah sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\text{arg max}} \left(\frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \right) \quad [1]$$

Untuk P(x₁, x₂, x₃, ...x_n) nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis sebagai persamaan dapat ditulis sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\text{argmax}} \left(P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \right) \quad [2]$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{arg\,max}} \prod_{i=1}^n (P(x_i|V_j)P(V_j)) \quad [3]$$

Keterangan:

V_j = Kategori opini

$P(x_i|V_j)$ = Probabilitas x_i pada kategori V_j

$P(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dan $P(x_i|V_j)$ dihitung saat pelatihan dimana persamaanya adalah sebagai berikut:

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad [4]$$

$$P(x_i|V_j) = \frac{n_k + 1}{n + |kosakata|} \quad [5]$$

Keterangan:

$|docs\ j|$ = jumlah dokumen pada kategori j

$|contoh|$ = jumlah dokumen dari semua kategori

n_k = jumlah kemunculan kata x_i pada kategori V_j

n = jumlah kata dalam setiap kategori

$|kosakata|$ = jumlah semua kata dari semua kategori

Contoh:

Tentukan sentimen dari kalimat berikut "*I enjoyed the movie*". Diketahui:

Probabilitas dari kategori positif = 0.5

Probabilitas dari kategori negatif = 0.5

Menghitung probabilitas kata pada tiap kategori menggunakan rumus:

$$P(x_i|V_j) = \frac{n_k + 1}{n + |\text{kosakata}|}$$

Probabilitas 'enjoyed' pada kategori negatif:

Jumlah kemunculan kata 'enjoyed' pada kategori negatif (n_k) = 4

Jumlah kata dalam setiap kategori (n) = 68247

Jumlah semua kata dari semua kategori ($|\text{kosakata}|$) = 136189

$$P(e | n) = \frac{4+1}{68247+1} = \frac{5}{68248} = 7.3261941537E - 5$$

Probabilitas tiap kata sesuai dengan kategori masing-masing setelah menggunakan rumus.

$$P(\text{enjoyed}|\text{negatif}) = 7.3261941537E-5$$

$$P(\text{movie}|\text{negatif}) = 0.003898530591481$$

$$P(\text{enjoyed}|\text{positif}) = 6.3684594696543E-5$$

$$P(\text{movie}|\text{positif}) = 0.0026355624574415$$

Menghitung probabilitas kalimat pada tiap kategori menggunakan rumus:

$$V_{MAP} = \arg \max_{V_j \in V} \prod_{i=1}^n (P(x_i|V_j)P(V_j))$$

$$P(\text{kalimat}|\text{negatif}) = P(V_j)(P_{x_i|V_j}) = 4.76742182331028E-08$$

$$P(\text{kalimat}|\text{positif}) = P(V_j)(P_{x_i|V_j}) = 8.39223634497933E-08$$

Dari semua kategori dokumen yang diujikan menunjuk bahwa probabilitas tertinggi adalah kategori positif. Maka, berdasarkan nilai probabilitas tertinggi kalimat "I enjoyed the movie" mengandung sentimen positif.

Catatan: kata 'I' dan 'the' diabaikan karena tidak mengandung makna.

2.5 MySQL dan PHP

MySQL adalah *multiuser database* yang menggunakan bahasa *Structured Query Language* (SQL). Sedangkan SQL adalah bahasa standar yang digunakan untuk mengakses database server. MySQL sendiri merupakan implemetasi sistem manajemen basisdata relasional (RDBMS) yang didistribusikan secara gratis. Adapun kelebihan yang dimiliki oleh MySQL adalah source-nya yang bersifat gratis, sintaks dan proses pengaksesan *database* lebih mudah dipahami, mendukung *record* yang memiliki kolom dengan panjang tetap atau panjang bervariasi. Sedangkan PHP merupakan bahasa pemrograman yang berjalan pada sebuah web-server (serverside). Oleh karenanya PHP dapat melakukan apa saja yang bisa dilakukan program CGI lain, yaitu mengolah data dengan tipe apapun, menciptakan halaman web yang dinamis, serta menerima dan menciptakan cookies(Haris Saputro, 2012).

PHP dapat digunakan pada semua sistem operasi antara lain Linux, UNIX, Microsoft Windows, Mac OS, dan RISC OS. PHP juga mendukung banyak web server seperti Apache, IIS, Personal Web Server, Netscape, Xitami dan masih banyak lainnya. PHP tidak terbatas pada hasil keluaran HTML namun juga memiliki kemampuan untuk mengolah keluaran gambar, file pdf dan flash movie serta menghasilkan text seperti XHTML dan file XML.

2.6 HTML

HTML adalah bahasa untuk memarkup atau memformat. HTML memberitahu web browser bagaimana cara menampilkan konten atau isi. HTML memisahkan konten berupa kata-kata, gambar, audio, video dan lain sebagainya dengan menggunakan beberapa elemen yang telah didefinisikan untuk mengidentifikasi tipe-tipe konten, Elemen memiliki tag yang mengekspresikan konten yang diawali dengan penutup siku dan tag penutup.

Adapun Fungsi HTML antara lain:

1. Membuat, mendesain dan mengontrol tampilan webpage dan isinya.
2. Mempublikasikan dokumen secara online sehingga bisa diakses, dilihat dari seluruh dunia.

3. Menambahkan objek-objek berupa image, audio, video dan java applet dalam dokumen html.

Dengan menggunakan HTML berkas yang dibuat dalam perangkat lunak pengolah kata disimpan dalam format ASCII normal sehingga menjadi halaman web dengan perintah-perintah HTML.

BAB III ANALISIS DAN PERANCANGAN

3.1 Deskripsi Umum Sistem



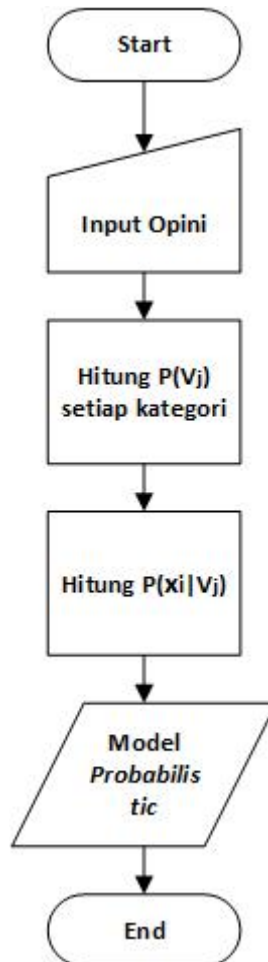
Gambar III-1 Deskripsi Umum Sistem

Data opini berita *online* dalam teks berbahasa Indonesia merupakan sumber data yang akan diolah ke dalam proses klasifikasi. Tahap pengolahan awal adalah proses *text preprocessing* terhadap baris-baris kalimat opini. Tahap ini terdiri dari beberapa fase yaitu tokenisasi, *filtering*, dan *stemming*.

Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata. Tokenizing menghilangkan delimiter-delimiter seperti tanda titik(.), koma(,), spasi dan karakter angka yang ada pada kata tersebut. Tokenisasi juga melakukan fungsi *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil. Filtering merupakan fase menghilangkan kata-kata yang tidak mengandung makna atau *stopword*. Stopword adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Misalnya "di", "oleh", "pada", "sebuah", "karena" dan lain sebagainya. Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya. Tujuan dari proses stemming adalah menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata.

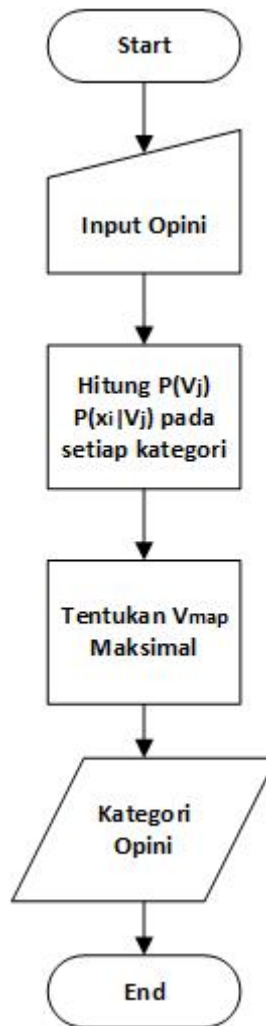
Setelah melewati tahapan preprocessing, proses selanjutnya adalah membuat *model probabilistic*. Tahapan ini digunakan untuk menghasilkan model yang nilainya akan digunakan pada tahapan selanjutnya yaitu proses klasifikasi.

Proses ini digunakan untuk mendapatkan nilai probabilistic dari $P(V_j)$ dan $P(x_i|V_j)$. Berikut *flowchart* dari proses membuat model probabilistic pada gambar III.2.



Gambar III-2 Flowchart Proses Membuat Model

Proses inti adalah proses klasifikasi untuk menentukan sebuah kalimat sebagai anggota opini positif atau sebagai anggota opini negatif berdasarkan nilai perhitungan probabilitas *Bayes* yang lebih besar. Jika hasil probabilitas *Bayes* kalimat tersebut untuk opini positif lebih besar maka kalimat tersebut masuk kategori opini positif demikian juga sebaliknya. Berikut *flowchart* dari proses klasifikasi pada gambar III.3.



Gambar III-3 Flowchart Proses Klasifikasi

3.2 Kebutuhan Fungsional dan Nonfungsional

Dalam membangun aplikasi ini mempunyai kebutuhan fungsional dan nonfungsional. Kebutuhan fungsional adalah deskripsi dari layanan-layanan yang disediakan oleh sistem, sedangkan kebutuhan nonfungsional adalah karakteristik dan batasan-batasan yang lain yang mendefinisikan sistem.

3.2.1 Kebutuhan Fungsional

- a. Sistem dapat melakukan *text preprocessing* untuk data opini mentah.
- b. Sistem dapat membuat model *probabilistic* pada tahap pelatihan untuk digunakan pada proses klasifikasi.

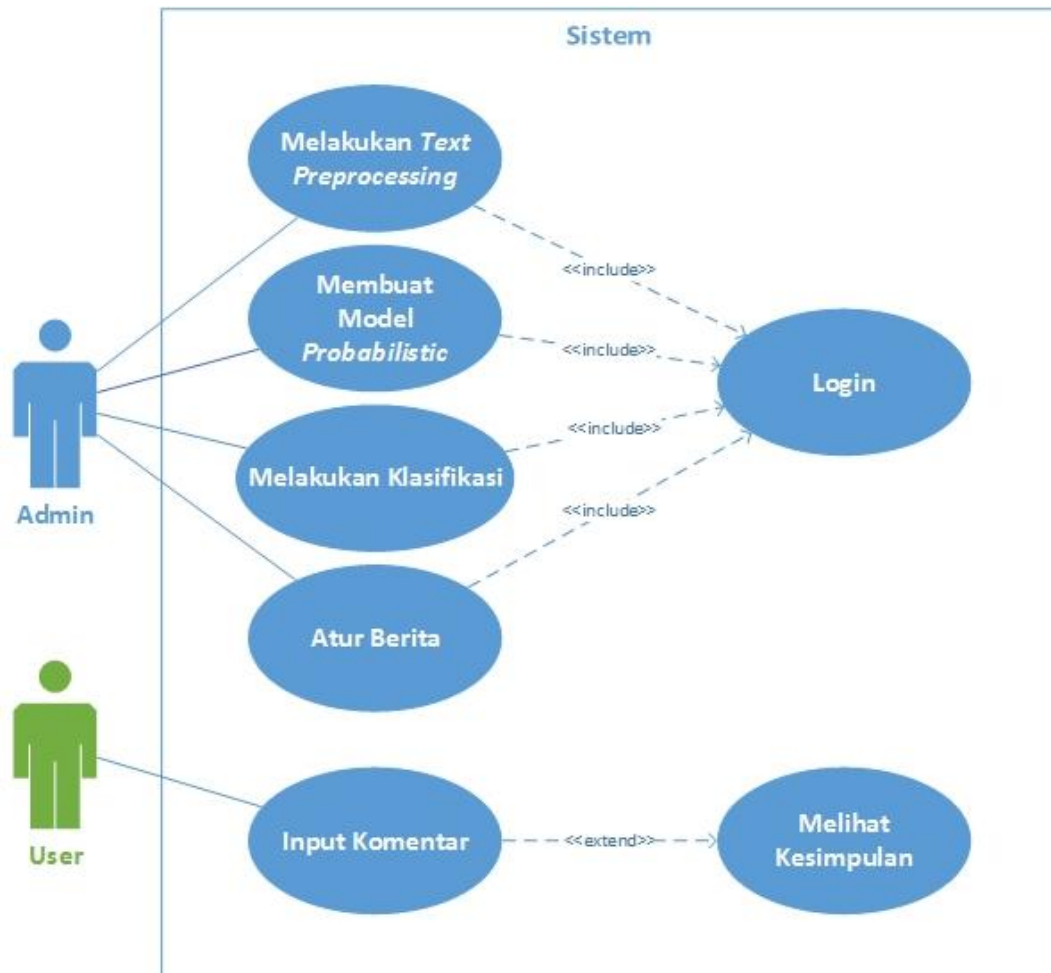
- c. Sistem dapat melakukan klasifikasi untuk menghitung probabilitas tertinggi, sehingga dapat menentukan kategori opini.
- d. Sistem mempunyai fitur untuk mengisi opini atau komentar.
- e. Sistem mempunyai fitur menampilkan hasil kesimpulan berupa informasi dari opini-opini yang diolah.

3.2.2 Kebutuhan Nonfungsional

- a. Data opini akan lebih mudah diolah, jika menggunakan bahasa Indonesia yang baik dan benar.

3.3 Use Case Diagram

Pemodelan kebutuhan system dibuat dengan menggunakan diagram *use case*. Diagram *use case* ini menggambarkan kebutuhan sistem secara keseluruhan. Diagram use case yang digunakan dapat dilihat pada gambar III.4.



Gambar III-4 Diagram Use Case

3.4 Skenario Use Case

3.4.1 Skenario Use Case Login

Tabel III.1 Skenario Use Case Login

Nama use case	<i>Login</i>
Deskripsi	Untuk mengakses sistem ini admin harus melakukan login terlebih dahulu, karena admin memiliki akses fungsi lebih yang disediakan oleh sistem daripada user biasa.
Aktor	Admin
Kondisi awal	Admin tidak bisa menggunakan fungsi yang disediakan oleh sistem.
Kondisi akhir	Admin bisa menggunakan fasilitas yang disediakan oleh sistem.
Skenario	
Aktor	Sistem
1. Admin melakukan klik terhadap <i>link login</i> .	
	2. Sistem menampilkan form login untuk meminta memasukkan data <i>username</i> dan <i>password</i> .
3. Admin mengisi form login dengan <i>username</i> dan <i>password</i> .	
	4. Sistem menampilkan status login dan mengarahkan admin ke halaman yang dituju.

3.4.2 Skenario Use Case Melakukan Text Preprocessing

Tabel III.2 Skenario Use Case Melakukan Text Preprocessing

Nama use case	<i>Melakukan Text Preprocessing</i>	
Deskripsi	Admin dapat melakukan fungsi <i>text preprocessing</i> untuk mengolah data opini.	
Aktor	Admin	
Kondisi awal	Admin bersiap mengisi opini di kolom komentar, sistem siap melakukan <i>text preprocessing</i> .	
Kondisi akhir	Sistem melakukan proses <i>text preprocessing</i> dan hasilnya disimpan ke dalam <i>database</i> .	
Skenario		
Aktor	Sistem	
1. Admin memilih menu <i>text preprocessing</i> .		
2. Admin memilih submenu text preprocessing yang diinginkan antara lain: <i>tokenizing, filtering</i> dan <i>stemming</i> .		
3. Admin mengisi teks bahasa Indonesia di <i>input text</i> .		
4. Admin melakukan klik terhadap <i>button tokenize, filtering</i> atau <i>stemming</i> .		
	5. Sistem melakukan proses <i>text preprocessing</i> yang di pilih admin.	
	6. Sistem menampilkan <i>output text</i> dari proses <i>text preprocessing</i> yang dipilih admin.	

3.4.3 Skenario Use Case Membuat Model Probabilistic

Tabel III.3 Skenario Use Case Membuat Model Probabilistic

Nama use case	<i>Membuat Model Probabilistic</i>	
Deskripsi	Admin dapat melakukan fungsi membuat model <i>probabilistic</i> . Proses ini digunakan mengestimasi probabilitas kemunculan kata sebagai sentimen positif, negatif dan netral dengan melihat kumpulan data latih sentimen positif, negatif dan netral dan menghitung seberapa sering kata tersebut muncul dalam tiap kelas.	
Aktor	Admin	
Kondisi awal	Admin sudah login ke dalam sistem dan memilih menu model <i>probabilistic</i> .	
Kondisi akhir	Sistem melakukan proses model <i>probabilistic</i> berdasarkan aksi yang dilakukan admin dan hasilnya disimpan ke dalam <i>database</i> .	
Skenario		
Aktor	Sistem	
1. Admin memilih menu model <i>probabilistic</i> .		
2. Admin melakukan proses model <i>probabilistic</i> (menambah kategori baru, menghapus kategori, menambah dokumen data latih dan menghapus dokumen data latih)		
3. Admin melakukan klik pada <i>button</i> .		
	4. Sistem melakukan proses model <i>probabilistic</i> yang dipilih admin dan	

	menambah ke atau menghapus dari database.
	5. Sistem menampilkan <i>notice</i> atas aksi yang dilakukan admin.

3.4.4 Skenario *Use Case* Melakukan Klasifikasi

Tabel III.4 Skenario *Use Case* Melakukan Klasifikasi

Nama use case	<i>Melakukan Klasifikasi</i>	
Deskripsi	Admin dapat melakukan proses klasifikasi untuk mengetahui kategori opini tertinggi yang diinput.	
Aktor	Admin.	
Kondisi awal	Admin sudah login ke sistem dan memilih menu klasifikasi.	
Kondisi akhir	Sistem melakukan proses klasifikasi dan menampilkan output tabel kategori opini beserta skor masing-masing.	
Skenario		
Aktor	Sistem	
1. Admin memilih menu klasifikasi		
2. Admin mengisi teks bahasa Indonesia di <i>input text</i> .		
3. Admin melakukan klik pada <i>button</i> klasifikasi.		
	4. Sistem melakukan proses <i>text preprocessing</i> terlebih dahulu kemudian diikuti proses klasifikasi.	
	5. Sistem menampilkan kategori opini beserta skor masing-masing opini.	

3.4.5 Skenario Use Case Melihat Kesimpulan

Tabel III.5 Skenario Use Case Melihat Kesimpulan

Nama use case	<i>Melihat Kesimpulan</i>	
Deskripsi	User dapat melihat kesimpulan dari proses klasifikasi yang dilakukan. Kesimpulan berupa hasil kategori opini, jumlah opini positif, negatif, dan netral, dsb.	
Aktor	User.	
Kondisi awal	User berada di halaman homepage dan telah menginput komentar.	
Kondisi akhir	User bisa melihat hasil kesimpulan dari proses klasifikasi pada menu View Details.	
Skenario		
Aksi Aktor	Respon Sistem	
1. User telah menginput komentar.		
	2. Sistem menampilkan hasil kesimpulan dari proses klasifikasi. Hasil kesimpulan berupa <i>pie chart</i> , jumlah komentar, opini negatif, positif dan netral.	

3.4.6 Skenario Use Case Input Komentar

Tabel III.6 Skenario Use Case Input Komentar

Nama use case	<i>Input Komentar</i>	
Deskripsi	User dapat menginput komentar di kolom komentar agar dapat melakukan uji sentimen komentar.	
Aktor	User.	
Kondisi awal	User sudah berada di homepage pada siap menginput komentar pada kolom komentar.	
Kondisi akhir	Sistem melakukan proses klasifikasi dan menampilkan output post komentar beserta kategori opini sentimen.	
Skenario		

Aktor	Sistem
1. User berada di homepage bagian kolom komentar.	
2. User mengisi komentar teks bahasa Indonesia.	
3. User melakukan klik pada <i>button post</i> .	
	4. Sistem melakukan proses klasifikasi dari komentar yang diinput.
	5. Sistem menampilkan post komentar dan sentimen opini dari komentar.

3.4.7 Skenario Use Case Atur Berita

Tabel III.7 Skenario Use Case Atur Berita

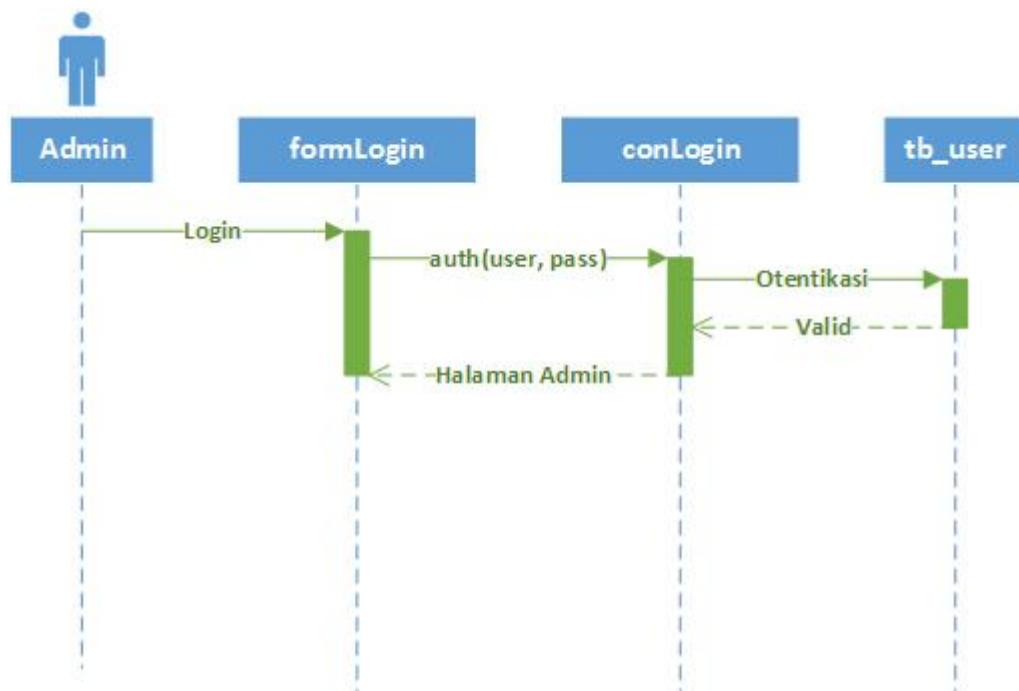
Nama use case	<i>Atur Berita</i>	
Deskripsi	Admin dapat mengatur berita yang akan tampil di homepage dengan menginput link berita online dari situs news.kompas.com.	
Aktor	Admin.	
Kondisi awal	Admin sudah login ke sistem dan memilih menu atur berita.	
Kondisi akhir	Sistem menampilkan gambar, judul, waktu dan headline berita.	
Skenario		
Aktor	Sistem	
1. Admin memilih menu atur berita.		
2. Admin mengisi link berita.		
3. Admin melakukan klik pada <i>button scrape</i> .		

	4. Sistem melakukan proses <i>scrape</i> terhadap link berita.
	5. Sistem menampilkan gambar, judul, waktu dan headline berita.

3.5 Diagram Sequence

Pada bagian ini ditampilkan diagram *sequence* yang menggambarkan interaksi elemen-elemen yang terdapat di dalam sistem. Diagram *sequence* ini dibuat sesuai masing-masing *use case*.

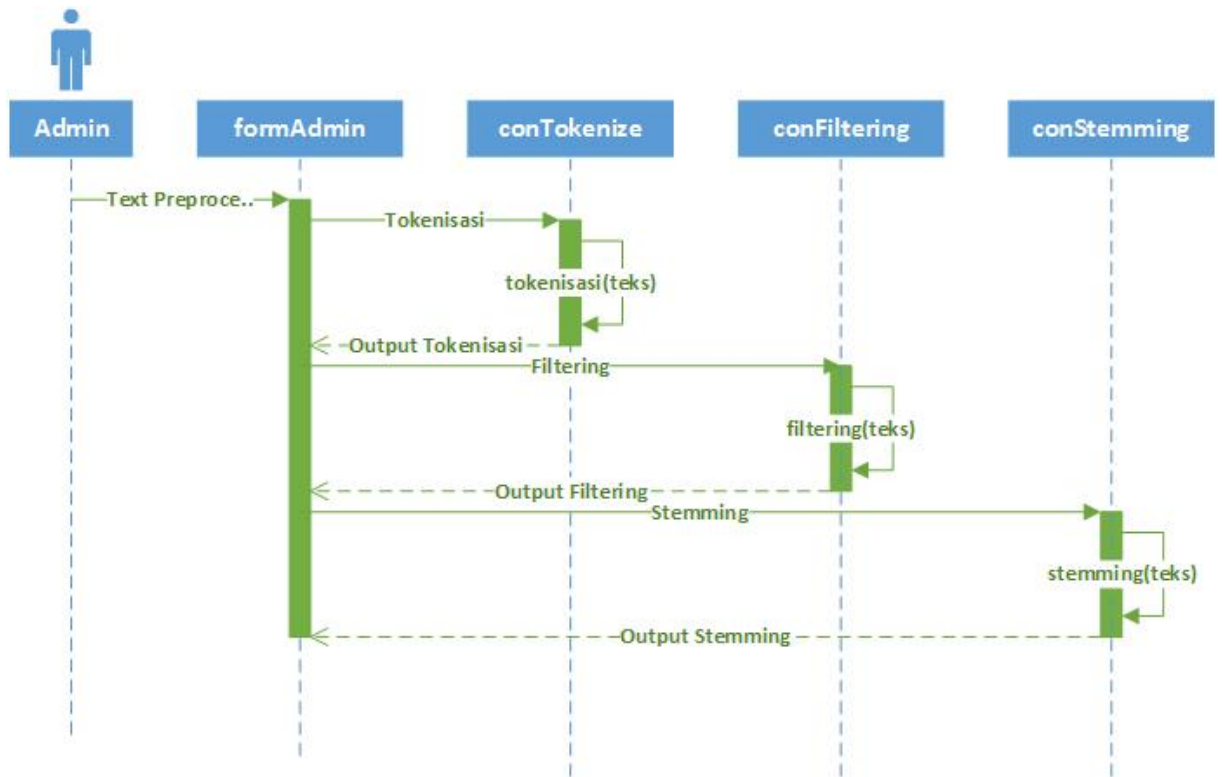
3.5.1 Diagram Sequence Login



Gambar III-5 Diagram Sequence Login

Admin harus melakukan login terlebih dahulu agar dapat menuju *Adminpage*. Admin melakukan login di halaman login dengan menginputkan *username* dan *password* yang valid. Apabila input *username* dan *password* valid, maka admin akan diarahkan menuju halaman *Adminpage*.

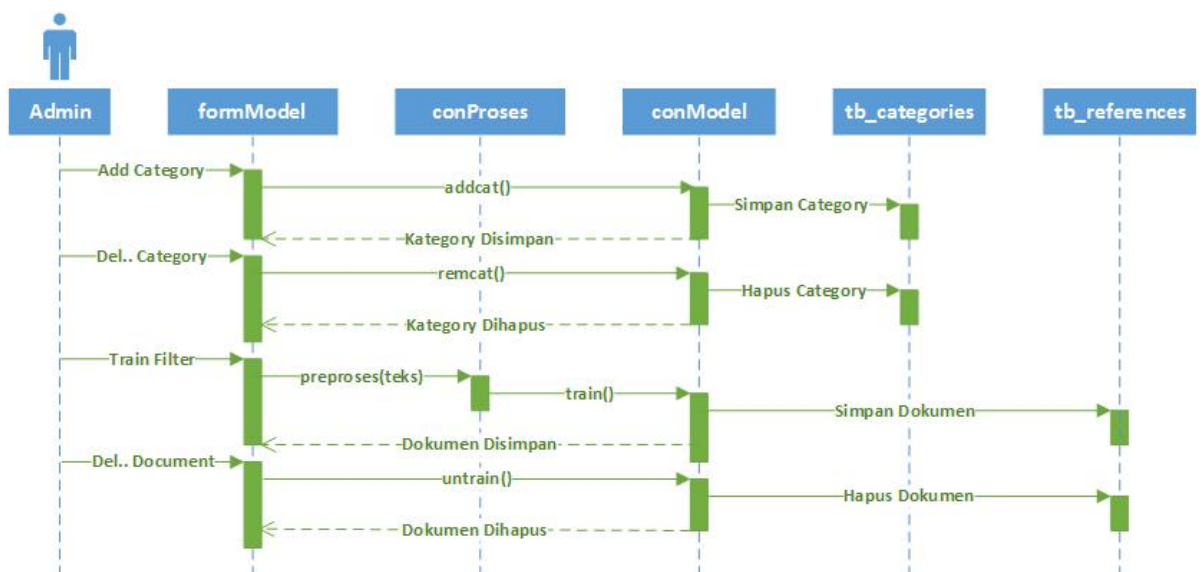
3.5.2 Diagram Sequence Melakukan *Text Preprocessing*



Gambar III-6 Diagram Sequence Melakukan *Text Preprocessing*

Admin dapat melakukan proses text preprocessing yang terdiri dari 3 fungsi yaitu *tokenizing*, *filtering* dan *stemming*.

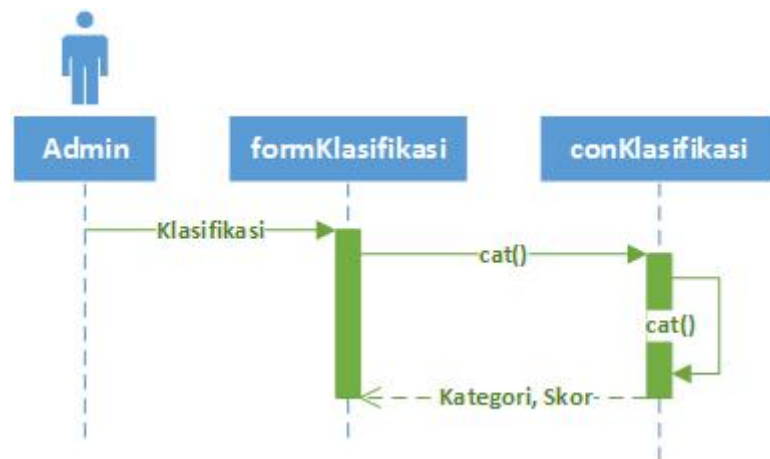
3.5.3 Diagram Sequence Membuat Model Probabilistic



Gambar III-7 Diagram Sequence Membuat Model Probabilistic

Pada *sequence* membuat model *probabilistic* memiliki fungsi yaitu `addcat()`, `remcat()`, `train()` dan `untrain()`. Admin dapat melakukan proses membuat model *probabilistic* pada halaman model *probabilistic*. Fungsi `addcat()` digunakan untuk menambah kategori baru yang kemudian disimpan ke dalam `tb_categories`. Fungsi `remcat()` digunakan untuk menghapus kategori dari `tb_categories`. Fungsi `train()` digunakan untuk menambah dokumen data latih yang kemudian disimpan ke dalam `tb_references`. Fungsi `untrain()` digunakan untuk menghapus dokumen data latih dari `tb_references`.

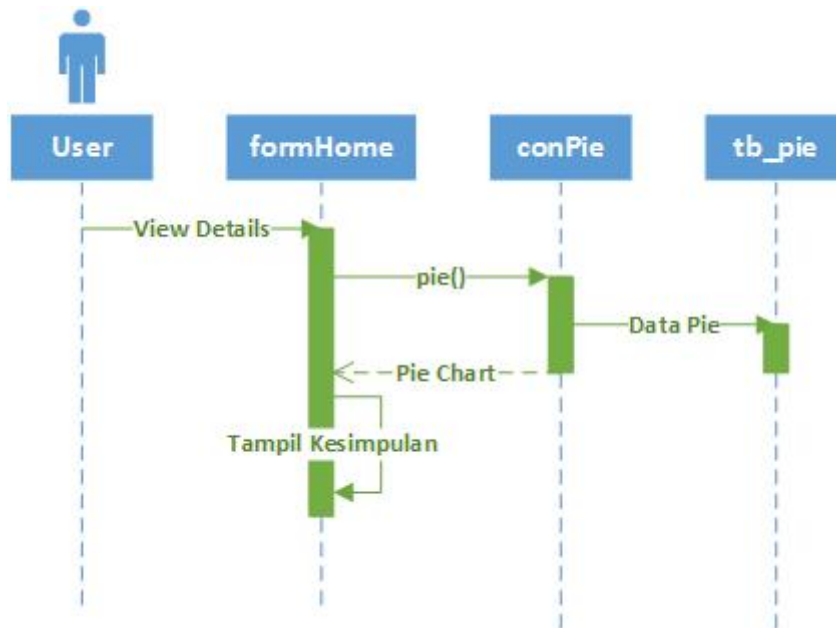
3.5.4 Diagram *Sequence* Melakukan Klasifikasi



Gambar III-8 Diagram *Sequence* Melakukan Klasifikasi

Pada *sequence* melakukan klasifikasi fungsi utama adalah `cat()`. Admin melakukan fungsi klasifikasi pada halaman klasifikasi kemudian menginput teks. Fungsi `cat()` digunakan untuk melakukan klasifikasi terhadap masukan teks dan menampilkan keluaran berupa kategori beserta skornya.

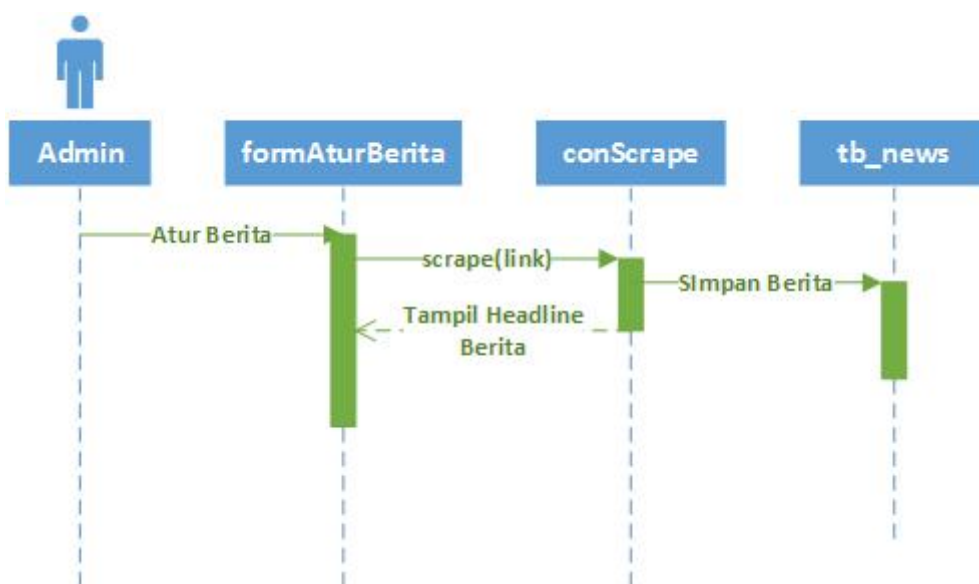
3.5.5 Diagram Sequence Melihat Kesimpulan



Gambar III-9 Diagram Sequence Melihat Kesimpulan

User dapat melakukan proses melihat kesimpulan. User melakukan proses melihat kesimpulan pada halaman *Homepage*. Keluaran dari kesimpulan berupa *pie chart* perbandingan kategori komentar. Fungsi utama adalah `pie()` yang menampilkan data kategori dan jumlah komentar dari `tb_pie`.

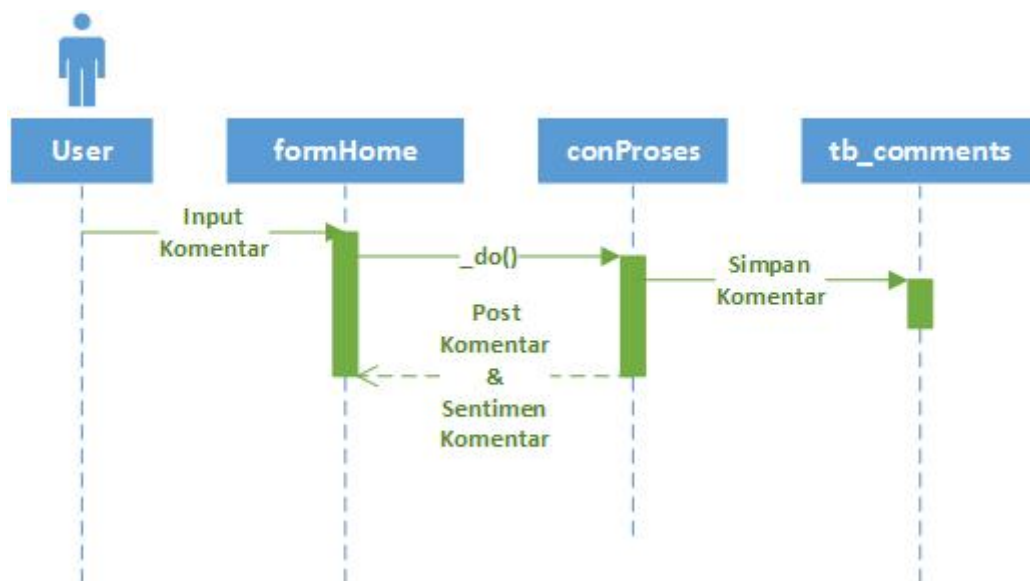
3.5.6 Diagram Sequence Atur Berita



Gambar III-10 Diagram Sequence Atur Berita

Pada *sequence* atur berita Admin dapat melakukan proses mengatur berita yang akan tampil pada halaman Homepage. Admin menginput *link* berita yang kemudian akan di proses oleh fungsi *scrape()*. Fungsi *scrape()* digunakan untuk melakukan ambil konten dari link berita online yang di masukan. Keluaran dari proses *scrape()* adalah judul, gambar, waktu dan isi berita *online*.

3.5.7 Diagram Sequence Input Komentar

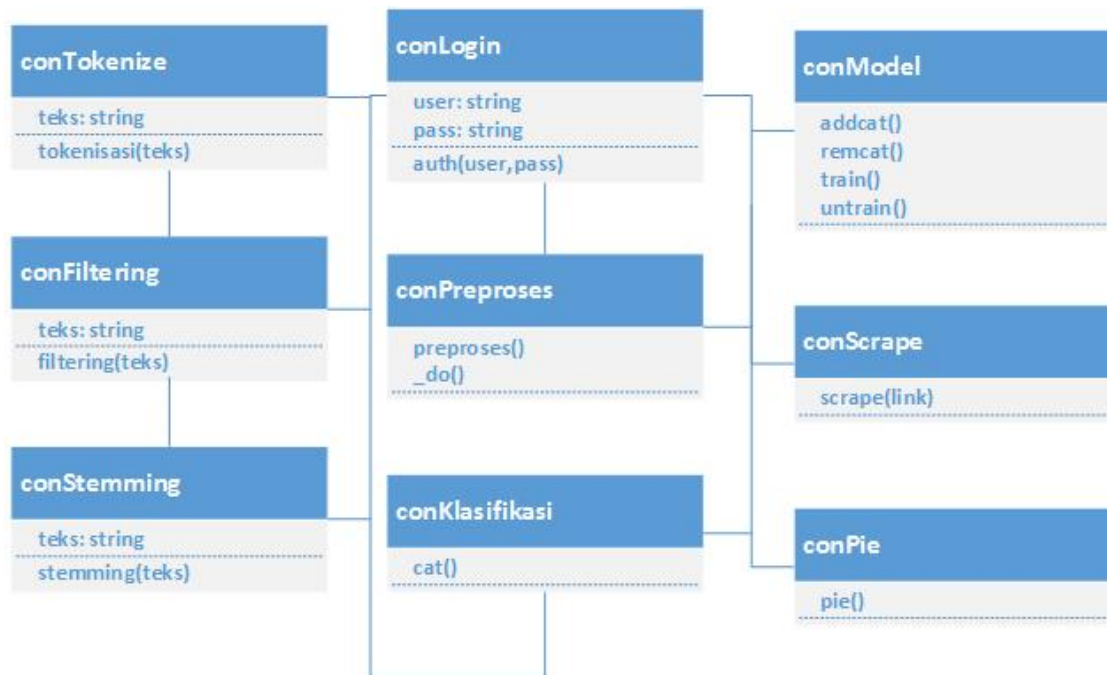


Gambar III-11 Diagram Sequence Input Komentar

Pada *sequence* input komentar user dapat melakukan proses masukan komentar terhadap berita *online*. Masukan komentar akan di proses pada fungsi *_do()*. Fungsi *_do()* merupakan gabungan proses *text preprocessing* dan klasifikasi. Hasil proses *_do()* menampilkan kategori opini dari masukan komentar. Masukan komentar juga disimpan ke dalam *tb_comments*.

3.6 Diagram Class

Diagram *class* perancangan menggambarkan bagaimana keterhubungan antar kelas dan berbagai *method* di dalamnya yang akan diimplementasikan. Diagram *class* perancangan dapat dilihat pada gambar III.12.

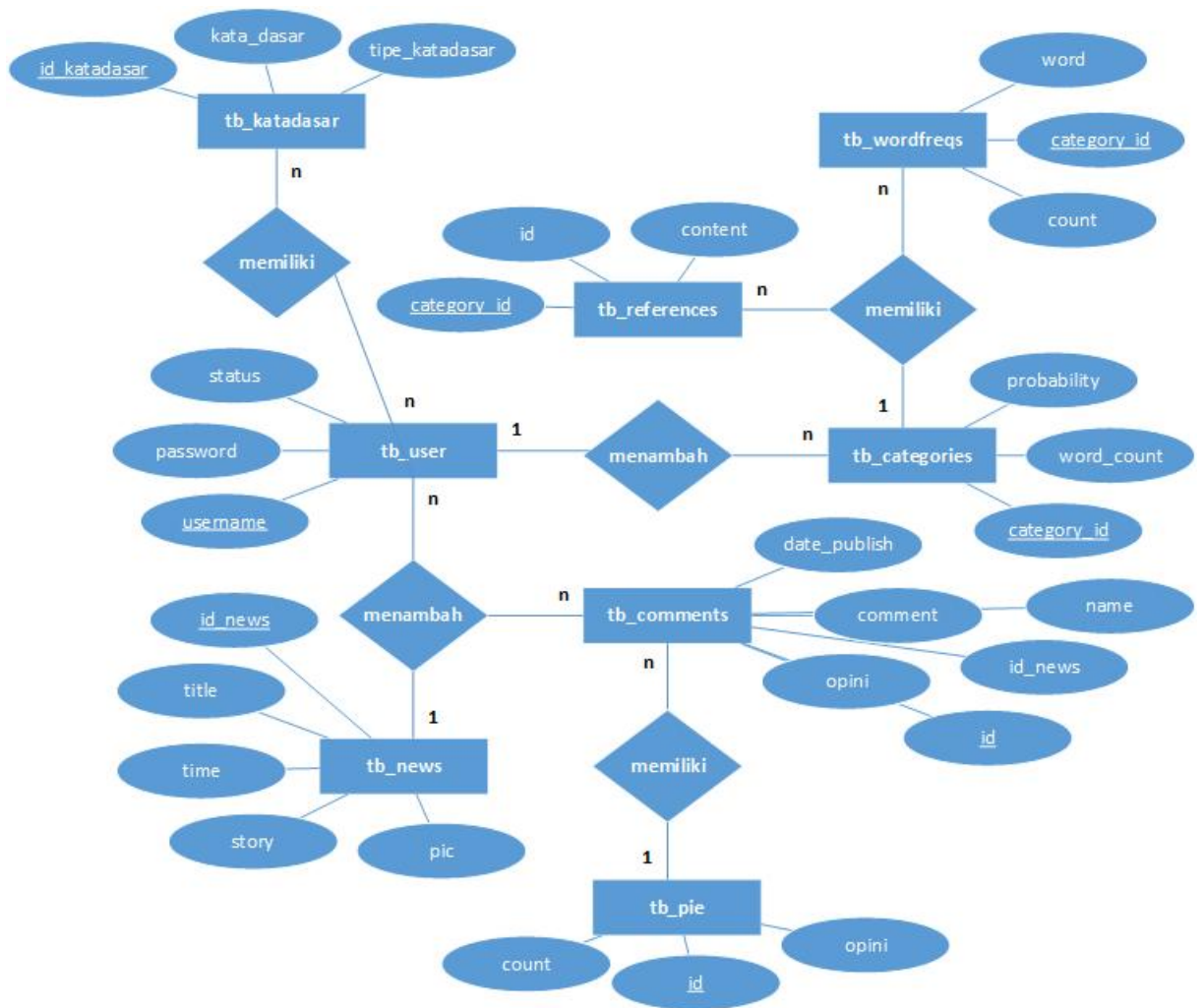


Gambar III-12 Class Diagram

3.7 Perancangan Basisdata

3.7.1 ER Diagram

Perancangan skema er diagram yang akan digunakan dalam sistem ini dapat dilihat pada gambar III.13.



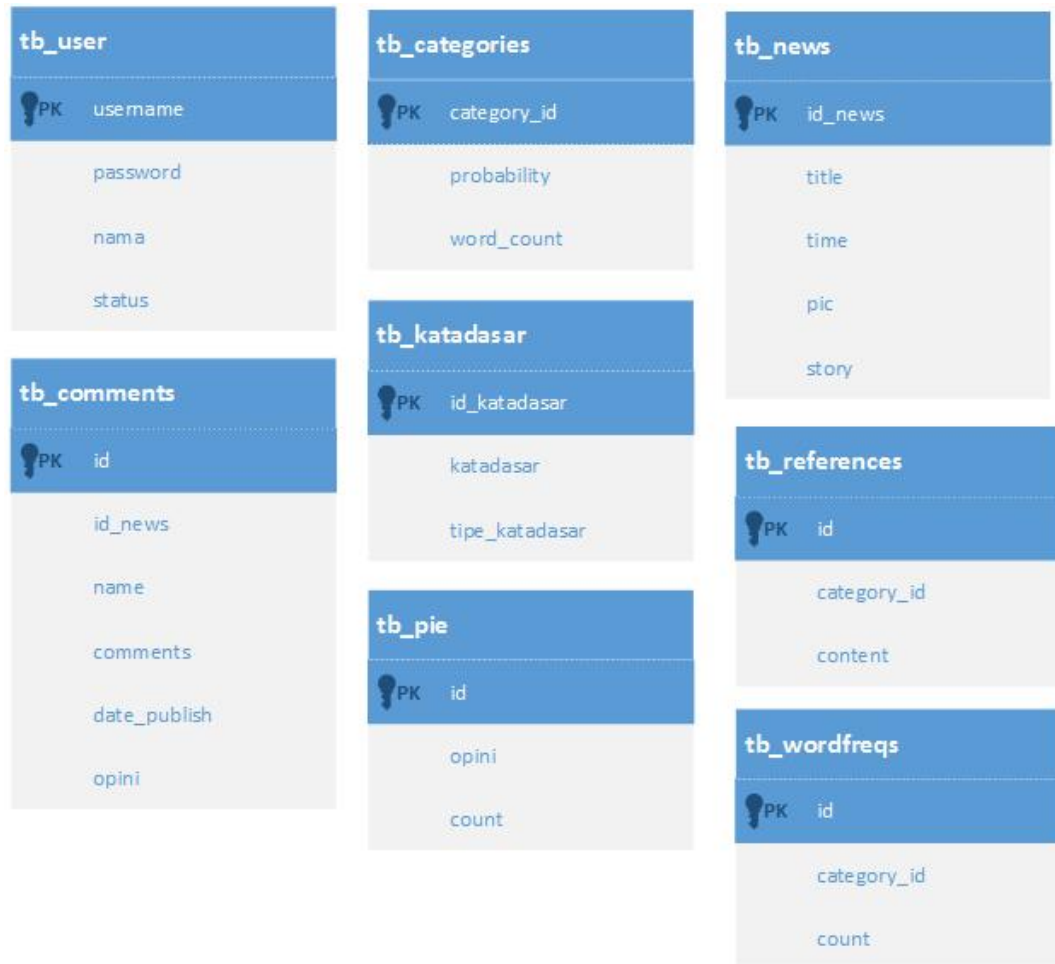
Gambar III-13 ER Diagram

Penjelasan singkat mengenai gambar ER Diagram diatas:

1. *User* dapat menambah kategori, berita dan komentar.
2. *User* memiliki kumpulan kata dasar yang digunakan untuk proses *stemming*.
3. Setelah *user* menambahkan kategori, entitas kategori memiliki kumpulan frekuensi kata dan data referensi untuk model *probabilistic*.
4. Setiap komentar yang ditambah *user*, entitas komentar memiliki persentasi komentar berupa *pie chart*.

3.7.2 Desain Basisdata

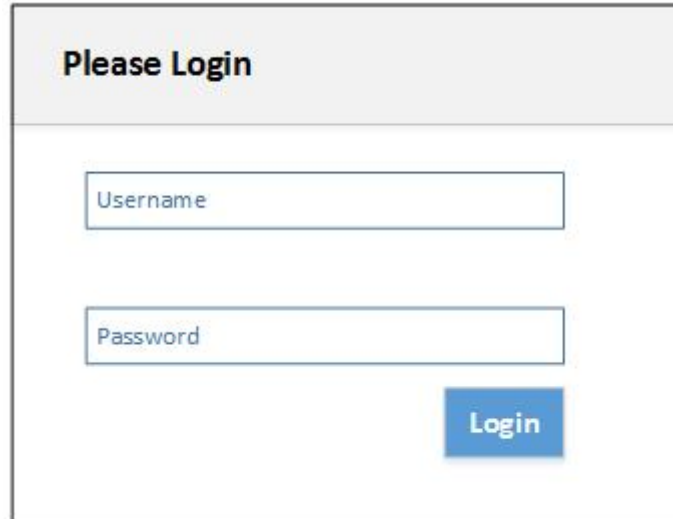
Perancangan skema desain basisdata yang akan digunakan dalam sistem ini dapat dilihat pada gambar III.14.



Gambar III-14 Desain Basisdata

3.8 Perancangan Antarmuka Sistem

3.8.1 Rancangan Antarmuka Halaman Login



The image shows a login interface with a light gray header containing the text "Please Login". Below the header, there are two white input fields with blue borders. The first field is labeled "Username" and the second is labeled "Password". To the right of the "Password" field is a blue button with the text "Login" in white.

Gambar III-15 Antarmuka Halaman Login

Tabel III.8 Deskripsi Halaman Login

Id Objek	Jenis	Nama	Keterangan
txtusername	<i>Input text</i>	<i>Username</i>	Diisi dengan username pengguna.
txtpassword	<i>Password</i>	<i>Password</i>	Diisi dengan password pengguna
btnlogin	<i>Button</i>	<i>Login</i>	Tombol untuk melakukan proses login.

3.8.2 Rancangan Antarmuka Halaman Text Preprocessing



Gambar III-16 Rancangan Antarmuka Halaman *Text Preprocessing*

Tabel III.9 Deskripsi Halaman *Text Preprocessing*

Id Objek	Jenis	Nama	Keterangan
txtcomment	Input textarea	Txtcomment	Diisi dengan teks bahasa Indonesia.
txtoutput	Blockquote	Txtoutput	Menampilkan hasil proses text preprocessing.
btnTokenize	Button	Tokenize	Tombol untuk melakukan proses tokenize.

3.8.3 Rancangan Antarmuka Halaman Membuat Model Probabilistic

Gambar III-17 Rancangan Antarmuka Halaman Membuat Model *Probabilistic*

Tabel III.10 Deskripsi Halaman Membuat Model *Probabilistic*

Id Objek	Jenis	Nama	Keterangan
txtcat	Input text	nameCategory	Diisi dengan nama kategori.
txtdocid	Input text	Docid	Melakukan proses login.
txtdoc	Input textarea	Document	Diisi dengan dokumen data latih.
category	Select category	Categories	Memilih jenis kategori dokumen untuk dilatih.
remcat	Select category	Categories	Memilih jenis kategori dokumen untuk dihapus.
untrain	Select document	Document	Memilih jenis dokumen untuk dihapus.
btnadd	Button	Add	Tombol untuk melakukan proses tambah kategori.
btntrain	Button	Train	Tombol untuk melakukan proses train filter atau dokumen data latih.
btndelcat	Button	Delete	Tombol untuk melakukan proses hapus kategori.

btnuntrain	Button	Delete	Tombol untuk melakukan proses hapus dokumen data latih.
------------	--------	--------	---

3.8.4 Rancangan Antarmuka Halaman Melakukan Klasifikasi

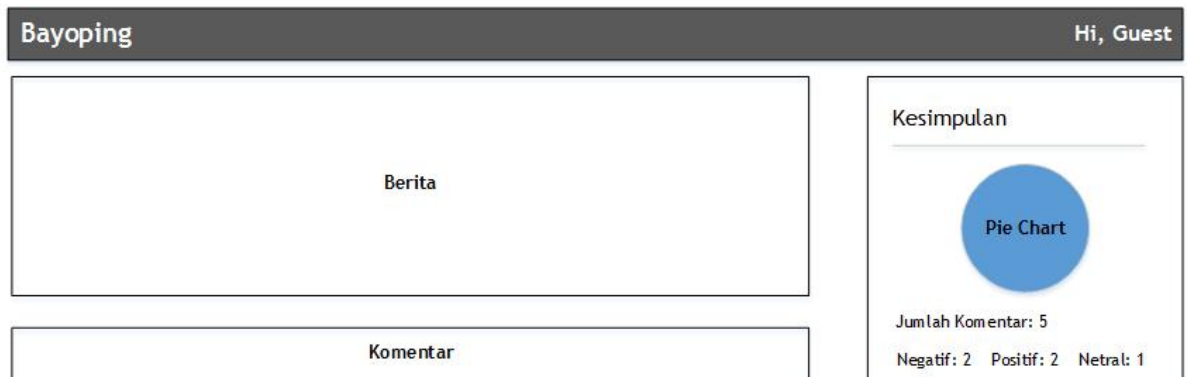


Gambar III-18 Rancangan Antarmuka Halaman Melakukan Klasifikasi

Tabel III.11 Deskripsi Halaman Melakukan Klasifikasi

Id Objek	Jenis	Nama	Keterangan
txtcomment	Input textarea	txtcomment	Diisi dengan teks bahasa Indonesia.
txtoutput	Blockquote	txtoutput	Menampilkan hasil proses klasifikasi.
btnklasifikasi	Button	Klasifikasi	Tombol untuk melakukan proses klasifikasi.

3.8.5 Rancangan Antarmuka Melihat Kesimpulan

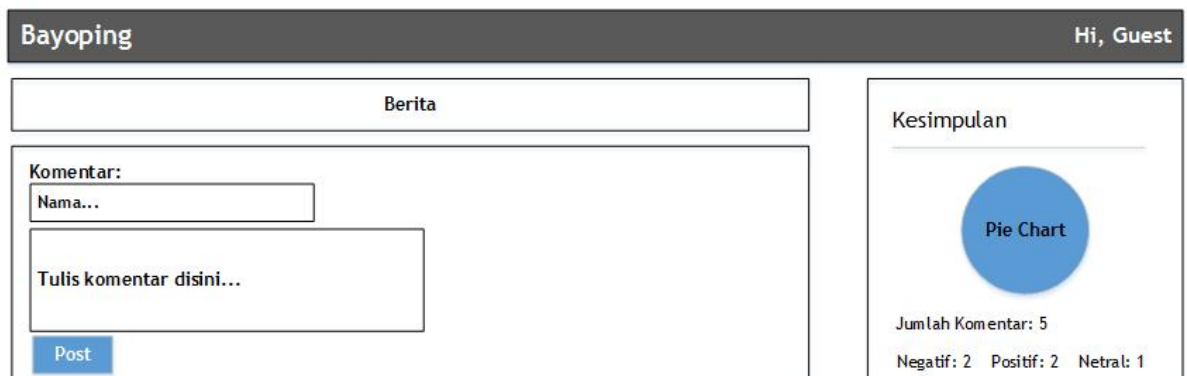


Gambar III-19 Rancangan Antarmuka Melihat Kesimpulan

Tabel III.12 Deskripsi Antarmuka Melihat Kesimpulan

Id Objek	Jenis	Nama	Keterangan
piechart	Chart	piechart	Menampilkan <i>pie chart</i> kesimpulan dari komentar yang diinput.

3.8.6 Rancangan Antarmuka Input Komentar



Gambar III-20 Rancangan Antarmuka Input Komentar

Tabel III.13 Deskripsi Antarmuka Input Komentar

Id Objek	Jenis	Nama	Keterangan
txtnama	Input text	txtnama	Diisi dengan nama.
txtkomentar	Input textarea	txtkomentar	Diisi dengan komentar.
btnPost	Button	Post	Tombol untuk proses <i>submit</i> komentar

3.8.7 Rancangan Antarmuka Halaman Atur Berita



Gambar III-21 Rancangan Antarmuka Halaman Atur Berita

Tabel III.14 Deskripsi Halaman Atur Berita

Id Objek	Jenis	Nama	Keterangan
txtlink	Input text	txtlink	Diisi dengan link berita.
tblBerita	Table	tblBerita	Menampilkan berita hasil <i>scrape</i> .
btnScrape	Button	Scrape	Tombol untuk melakukan proses <i>scrape</i> berita <i>online</i> .

BAB IV IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Struktur Tabel Basisdata

4.1.1 Tabel Categories

Tabel dibawah ini merupakan file implementasi basis data tabel categories.

Tabel IV.1 Tabel Categories

Nama Field	Jenis	Volume	Keterangan
category_id	Varchar	100	PK
probability	Double	-	
word_count	Big Integer	20	

4.1.2 Tabel Comments

Tabel dibawah ini merupakan file implementasi basis data tabel comments.

Tabel IV.2 Tabel Comments

Nama Field	Jenis	Volume	Keterangan
id	Integer	10	PK
id_news	Integer	11	
name	Varchar	40	
comments	Text	-	
date_publish	Timestamp	-	On update current_timestamp
opini	Varchar	12	

4.1.3 Tabel Katadasar

Tabel dibawah ini merupakan file implementasi basis data tabel katadasar.

Tabel IV.3 Tabel Katadasar

Nama Field	Jenis	Volume	Keterangan
id_katadasar	Integer	10	PK
katadasar	Varchar	70	
tipe_katadasar	Varchar	25	

4.1.4 Tabel News

Tabel dibawah ini merupakan file implementasi basis data tabel news.

Tabel IV.4 Tabel News

Nama Field	Jenis	Volume	Keterangan
id_news	Integer	100	PK
title	Varchar	999	
time	Varchar	999	
pic	Varchar	999	
story	Varchar	9999	

4.1.5 Tabel Pie

Tabel dibawah ini merupakan file implementasi basis data tabel pie.

Tabel IV.5 Tabel Pie

Nama Field	Jenis	Volume	Keterangan
id	Integer	11	PK
opini	Varchar	100	
count	Integer	4	

4.1.6 Tabel References

Tabel dibawah ini merupakan file implementasi basis data tabel references.

Tabel IV.6 Tabel References

Nama Field	Jenis	Volume	Keterangan
id	Varchar	250	PK
category_id	Varchar	100	
content	Text	-	

4.1.7 Tabel User

Tabel dibawah ini merupakan file implementasi basis data tabel user.

Tabel IV.7 Tabel User

Nama Field	Jenis	Volume	Keterangan
username	Varchar	12	PK
password	Varchar	12	
nama	Varchar	12	
status	Varchar	12	

4.1.8 Tabel Wordfreqs

Tabel dibawah ini merupakan file impementasi basis data tabel wordfreqs.

Tabel IV.8 Tabel Wordfreqs

Nama Field	Jenis	Volume	Keterangan
word	Varchar	100	PK
category_id	Varchar	100	
count	Big Integer	20	

4.2 Implementasi Class

Berdasarkan pada perancangan yang telah dilakukan, maka hasil implementasi kelas dan antarmuka yang dibuat secara detail dari perancangan yang telah dilakukan kemudian dilanjutkan pada implementasi yang menghasilkan kelas-kelas yang saling terhubung antara *controller*, model, dan *view*. Perbedaan antara perancangan dan implementasi dapat dilihat pada tabel berikut:

Tabel IV.9 Implementasi Class

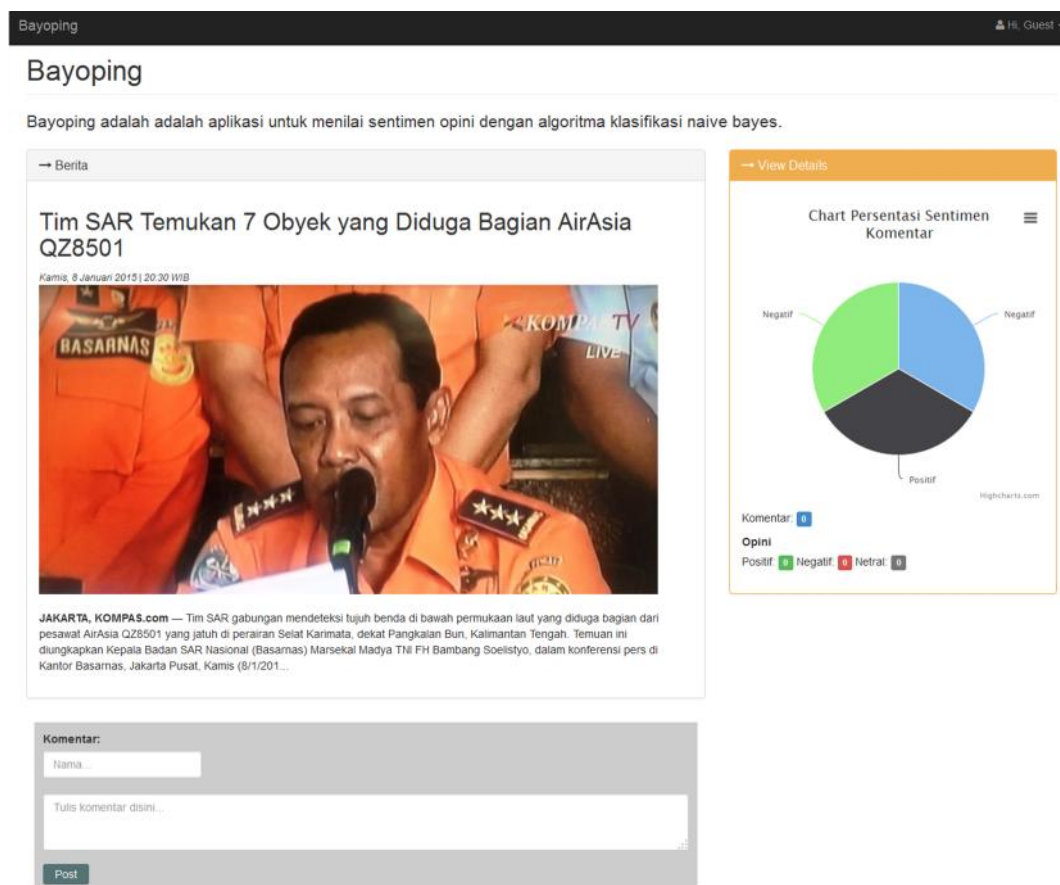
No	Nama Kelas Perancangan	Nama File Antarmuka	Nama File Implementasi
1	conTokenize	tokenizing.php	pre.tokenizing.php
2	conFiltering	filtering.php	pre.filtering.php
3	conStemming	stemming.php	pre.stemming.php
4	conKlasifikasi	klasifikasi.php	bayes.php
5	conModel	modprob.php	bayes.php
6	conScrape	newsconf.php	scrape.php

4.3 Implementasi Antarmuka

Sebelum dilakukan pengujian sistem, maka perlu dilakukan perancangan antarmuka yang menjadi penghubung antara sistem dan user. Dengan adanya antarmuka yang baik maka juga akan meningkatkan kualitas dari aplikasi. Aplikasi berkualitas tentunya sangat menguntungkan dalam hal interaksi manusia dan komputer.

4.3.1 Implementasi Antarmuka Halaman Homepage

Halaman homepage merupakan halaman utama bagi *guest* atau user. Pada halaman ini terdapat konten berita, konten komentar, dan konten kesimpulan opini. User dapat membaca konten berita, menginput komentar, dan mendapatkan kesimpulan opini.

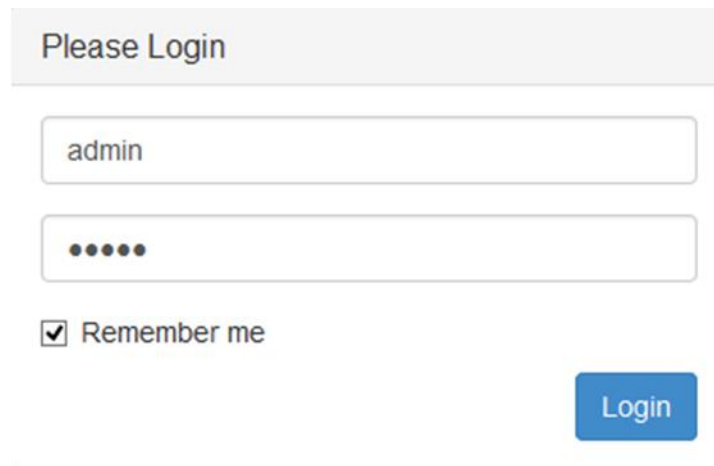


Gambar IV-1 Halaman Homepage

4.3.2 Implementasi Antarmuka Halaman *Login Admin*

Halaman login memiliki fungsi sebagai berikut:

- a. Memasukkan *username* dan *password* untuk login admin.
- b. Setelah menekan tombol *login*, sistem akan melakukan otentikasi *username* dan *password* ke database, jika data valid maka akan dialihkan ke *Adminpage*. Sebaliknya jika tidak valid, maka akan dialihkan ke *Homepage*.



Please Login

admin

.....

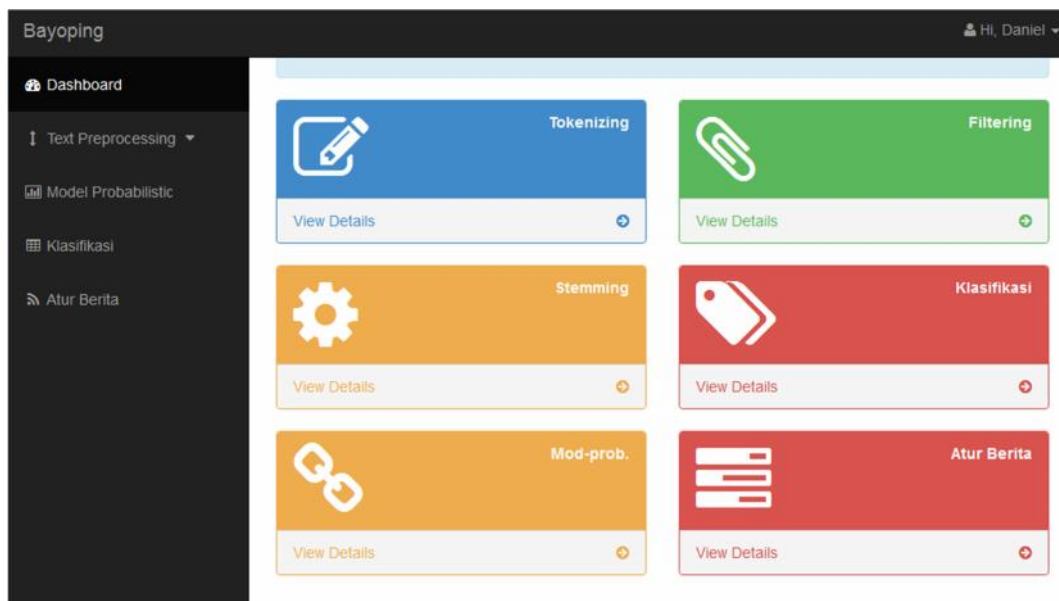
Remember me

Login

Gambar IV-2 Halaman *Login Admin*

4.3.3 Implementasi Antaramuka Halaman *Adminpage*

Halaman *Adminpage* merupakan halaman yang ditujukan untuk admin. Pada halaman ini terdapat menu-menu yang dapat dilakukan seorang admin. Menu-menu tersebut seperti, *tokenizing*, *filtering*, *stemming*, klasifikasi, modprob, dan atur berita.



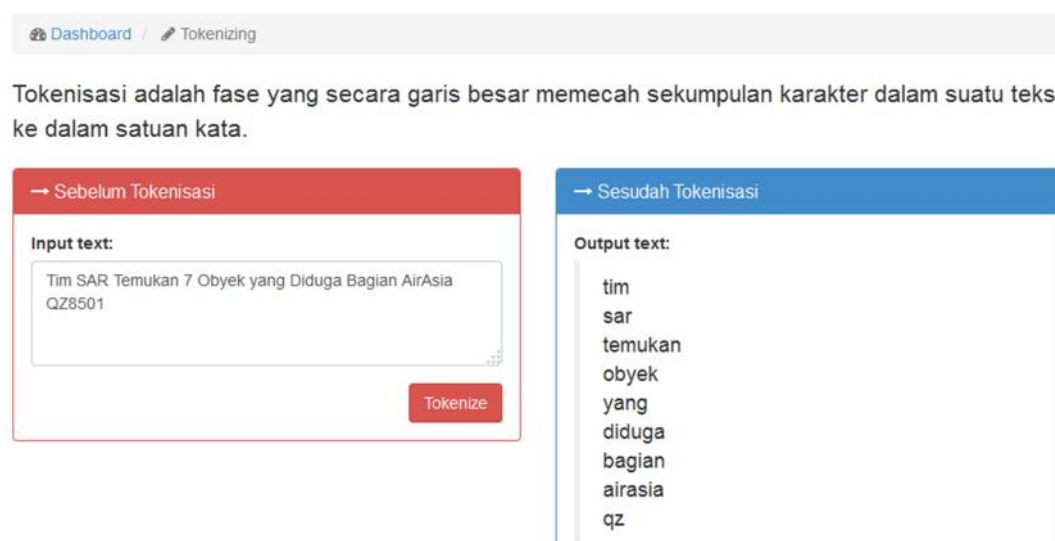
Gambar IV-3 Halaman Adminpage

4.3.4 Implementasi Antarmuka Halaman *Tokenizing*

Halaman tokenizing memiliki fungsi sebagai berikut:

- a. Memasukkan *input text* berupa teks bahasa Indonesia
- b. Setelah menekan tombol *tokenize*, maka *output text* akan menampilkan hasil proses *tokenize*.

Tokenizing



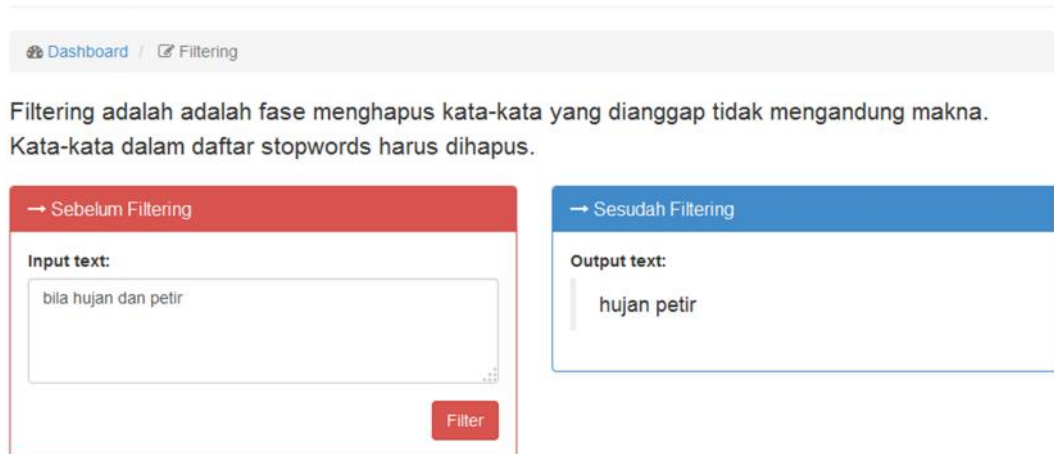
Gambar IV-4 Halaman *Tokenizing*

4.3.5 Implementasi Antarmuka Halaman *Filtering*

Halaman tokenizing memiliki fungsi sebagai berikut:

- Memasukkan input text berupa teks bahasa Indonesia.
- Setelah menekan tombol filter, maka output text akan menampilkan hasil proses filtering.

Filtering



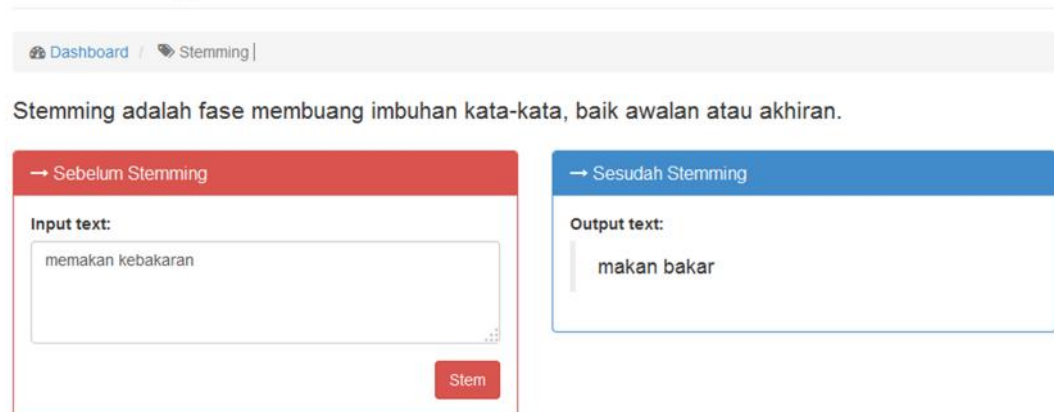
Gambar IV-5 Halaman *Filtering*

4.3.6 Implementasi Antarmuka Halaman *Stemming*

Halaman tokenizing memiliki fungsi sebagai berikut:

- Memasukkan *input text* berupa teks bahasa Indonesia.
- Setelah menekan tombol stem, maka *output text* akan menampilkan hasil proses *stemming*.

Stemming



Gambar IV-6 Halaman *Stemming*

4.3.7 Implementasi Antarmuka Halaman Model *Probabilistic*

Halaman modprob ini memiliki fungsi sebagai berikut:

- Add category* merupakan fungsi untuk menambahkan kategori baru. Kategori yang dimaksud adalah positif, negatif, dan netral. Nama kategori ditulis di text field dan diproses dengan menekan tombol *add*.
- Delete a category* merupakan fungsi untuk menghapus kategori yang telah ada atau yang telah ditambahkan. Kategori dihapus dengan memilih terlebih dahulu nama kategori dan diproses dengan menekan tombol *delete*.
- Train filter* merupakan fungsi untuk menambah dokumen data latih. Untuk melakukan *train filter* harus memasukkan id dokumen, memilih kategori dokumen, dan isi dokumen sendiri. Setelah itu dengan menekan tombol *train*, maka dokumen akan diolah menjadi data latih.
- Delete a document* merupakan fungsi untuk menghapus dokumen data latih yang telah ada atau yang telah ditambahkan. Dokumen dihapus dengan memilih nama dokumen terlebih dahulu dan diproses dengan menekan tombol *delete*.

The image displays four distinct panels from the Probabilistic Model interface:

- Add Category:** Features a text input field for the category name and a blue 'Add' button.
- Delete a Category:** Includes a dropdown menu for selecting a category (currently set to 'Netral') and a red 'Delete' button.
- Train Filter:** Contains a document identifier input field, a category dropdown menu (set to 'Netral'), a large text area for pasting the document content, and a blue 'Train' button.
- Delete a Document:** Shows a dropdown menu for selecting a document (currently 'pos1 - Positif') and a red 'Delete' button.

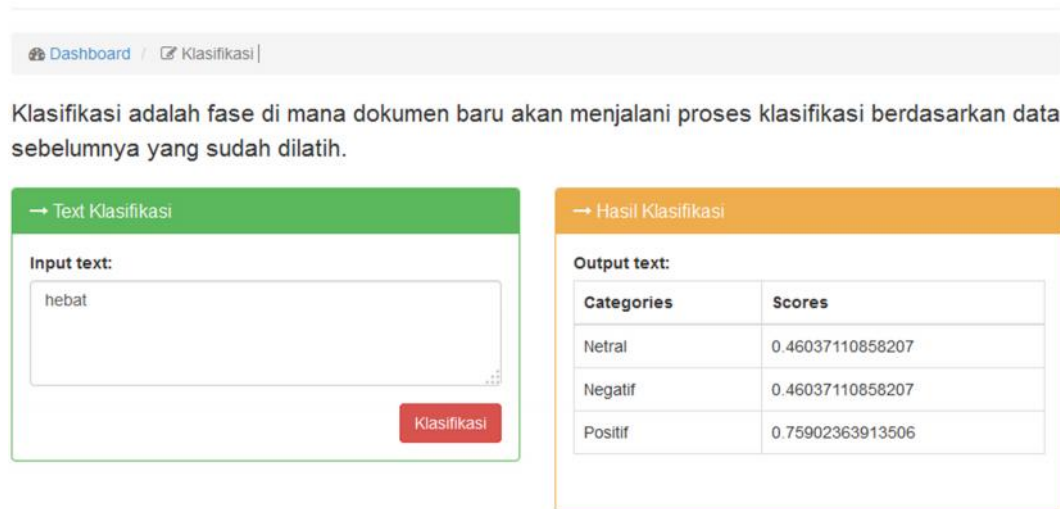
Gambar IV-7 Halaman Model *Probabilistic*

4.3.8 Implementasi Antarmuka Halaman Klasifikasi

Halaman klasifikasi memiliki fungsi sebagai berikut:

- Memasukkan *input text* berupa teks bahasa Indonesia.
- Setelah menekan tombol klasifikasi, maka *output text* akan menampilkan hasil proses klasifikasi.

Klasifikasi



Gambar IV-8 Halaman Klasifikasi

4.3.9 Implementasi Antarmuka Halaman Atur Berita


Halaman atur berita memiliki fungsi sebagai berikut:

- Memasukkan *link* berita dari situs berita *online* news.kompas.com. Setelah menekan tombol scrape, maka *input link* berita akan diproses sehingga menampilkan *output* berupa gambar, judul, waktu, dan *headline* berita.


Atur Berita

[Dashboard](#) / [Atur Berita](#)

Scrape berita online yang akan dijadikan berita di front page.



Link Berita:

Gambar	Judul	Waktu	Headline
	Tim SAR Temukan 7 Obyek yang Diduga Bagian AirAsia QZ8501	Kamis, 8 Januari 2015 20:30 WIB	JAKARTA, KOMPAS.com — Tim SAR gabungan mendeteksi tujuh benda di bawah permukaan laut yang diduga bagian dari pesawat AirAsia QZ8501 yang jatuh di perairan Selat Karimata, dekat Pangkalan Bun, Kalimantan Tengah. Temuan ini diungkapkan Kepala Badan SAR Nasional (Basarnas) Marsekal Madya TNI FH Bambang Soelistyo, dalam konferensi pers di Kantor Basarnas, Jakar...

Gambar IV-9 Halaman Atur Berita

4.3.10 Implementasi Antarmuka Komentar

Antarmuka komentar memiliki fungsi sebagai berikut:

- Memasukkan nama dan komentar bahasa Indonesia yang baik dan benar.
- Setelah tombol *post* ditekan, maka *input* komentar akan diproses dan menampilkan komentar ditambah sentimen opini dari komentar tersebut.

Komentar:

Leo

hebat

Leo

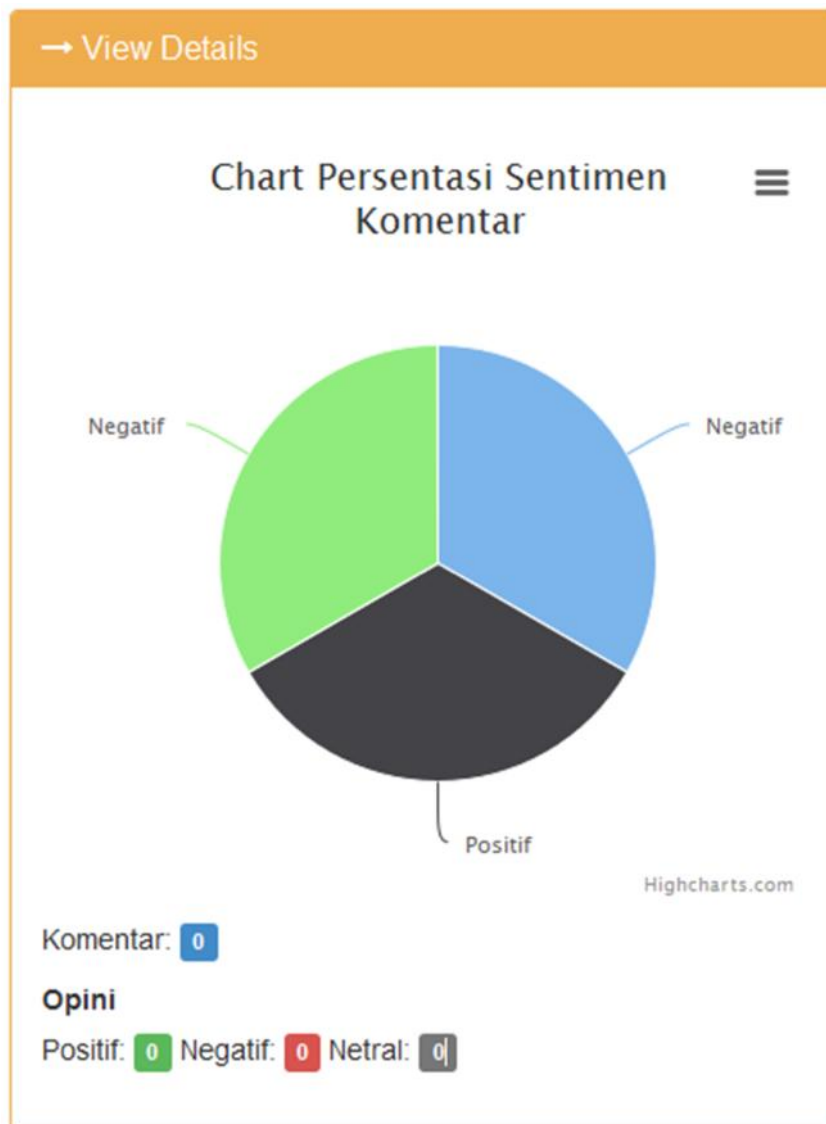
2015-01-09 02:20:27

hebat

Gambar IV-10 Antarmuka Komentar

4.3.11 Implementasi Antarmuka Melihat Kesimpulan (*View Details*)

Antarmuka *view details* merupakan antarmuka untuk menampilkan kesimpulan sentimen opini dari komentar-komentar yang telah diinput. Kesimpulannya berupa *pie chart*, jumlah komentar, dan jumlah opini positif, negatif, dan netral.



Gambar IV-11 Antarmuka View Details

4.4 Pengujian

4.4.1 Strategi Pengujian

Strategi pengujian pada aplikasi ini menggunakan metode *black box* yaitu berfokus pada kebutuhan deskripsi fungsional guna mengetahui jalannya sistem secara lengkap.

4.4.2 Deskripsi Pengujian

Deskripsi pengujian dalam aplikasi ini adalah sebagai berikut:

1. Membuka halaman homepage.
 - a. Melakukan proses input komentar.
 - b. Melihat kesimpulan sentimen komentar.
2. Melakukan proses login
 - a. Login sebagai Admin (memasukkan *username*, *password*) menuju *Adminpage*.
3. Login sebagai admin.
 - a. Melakukan proses *text preprocessing*.
 - *Tokenizing*
 - *Filtering*
 - *Stemming*
 - b. Melakukan proses membuat model *probabilistic*.
 - c. Melakukan proses klasifikasi.
 - d. Melakukan proses atur berita.

4.4.3 Hasil Pengujian

Pengujian aplikasi telah dilaksanakan dengan baik sesuai dengan metode *Black Box* yang digunakan dan hasilnya dapat dilihat di table berikut ini:

Nama Penguji: Jeffri Bahisindra

Umur: 20 Tahun

Jenis Kelamin: Laki-laki

Tabel IV.10 Hasil Pengujian

No	Use Case	Fungsi	Skenario	Data Uji	Target	Pengujian
1	Login	Otentikasi	1. Memasukkan <i>username</i> dan <i>password</i> 2. Menekan tombol login	Data Benar Username: admin Password: admin	- Berhasil masuk ke adminpage	Ok
				Data Salah Username: user Password: user	- Memberikan pesan kesalahan invalid login.	Ok
2	Melakukan <i>text preprocessing</i>	Simulasi <i>tokenizing</i>	1. Admin memilih menu <i>tokenizing</i> .	Input text: “5554 Perekonomian-INDONESIA...//”	Output text: “perekonomian indonesia”	Ok

			<ol style="list-style-type: none"> 2. Admin memasukkan teks bahasa Indonesia untuk ditokenisasi. 3. Admin menekan tombol <i>tokenize</i>. 			
		Simulasi <i>filtering</i>	<ol style="list-style-type: none"> 1. Admin memilih menu <i>filtering</i>. 2. Admin memasukkan teks bahasa Indonesia untuk difilter. 3. Admin menekan tombol <i>filter</i>. 	Input text: “Indonesia dan Malaysia sedang melakukan negosiasi.”	Output text: “indonesia malaysia negosiasi”	Ok
		Simulasi <i>stemming</i>	<ol style="list-style-type: none"> 1. Admin memilih menu <i>stemming</i>. 2. Admin memasukkan teks bahasa Indonesia untuk distem. 3. Admin menekan tombol <i>stem</i>. 	Input text: ”Perekonomian Indonesia sedang dalam pertumbuhan yang membanggakan”	Output text: “ekonomi indonesia sedang dalam tumbuh yang bangga”	Ok

3	Melakukan klasifikasi	Simulasi klasifikasi	<ol style="list-style-type: none"> 1. Admin memilih menu klasifikasi. 2. Admin memasukkan teks bahasa Indonesia untuk diklasifikasi. 3. Admin menekan tombol klasifikasi. 	Input text: “hebat”	Output text: Categories Scores Netral 0.4603 Negatif 0.4603 Positif 0.7590	Ok
4	Melakukan proses atur berita	Menginput berita online	<ol style="list-style-type: none"> 1. Admin memilih menu atur berita. 2. Admin memasukkan link berita <i>online</i> dari situs news.kompas.com. 3. Admin menekan tombol <i>scrape</i>. 	Link berita: “http://nasional.kompas.com/read/2015/01/09/10200101/3.Jenazah.Korban.AirAsia.Tiba.di.Pangkalan.Bun.Total.46.Jenazah.yang.Ditemukan”	Link berhasil proses dan menampilkan gambar, judul, waktu, dan headline berita.	Ok
5	Membuat model probabilistic	Menambah kategori (<i>Add category</i>)	<ol style="list-style-type: none"> 1. Admin memilih menu model probabilistic. 2. Admin memasukkan nama kategori yang ingin ditambah. 	Name of the category: Positif	<ul style="list-style-type: none"> - Muncul pesan “kategori telah ditambah”. - Kategori masuk ke database. 	Ok

			3. Admin menekan tombol add.			
	Menghapus kategori (Delete a category)	<ol style="list-style-type: none"> 1. Admin memilih menu model probabilistic. 2. Admin memilih nama kategori yang ingin dihapus. 3. Admin menekan tombol delete. 	Category to delete: Positif	<ul style="list-style-type: none"> - Muncul pesan “kategori telah dihapus”. - Kategori dihapus dari database. 	Ok	
	Latih dokumen (Train filter)	<ol style="list-style-type: none"> 1. Admin memilih menu model probabilistic. 2. Admin memasukkan nama id dokumen. 3. Admin memilih kategori untuk dokumen yang dilatih. 4. Admin memasukan isi dokumen. 	Document identifier: pos1 Category for the document: Positif Copy/paste here the document: menghargai kerja keras BASARNAS dan TIM GABUNGAN yang sudah bekerja keras,lama tidaknya proses pencarian relatif	<ul style="list-style-type: none"> - Muncul pesan “filter dokumen telah dilakukan”. - Dokumen latih masuk ke database. 	Ok	

			5. Admin menekan tombol train	tergantung dari sudut pandang mana kita melihat, jangan bisanya cuma mencaci pdhal belum tentu kalian di beri tanggung jawab yang sama bisa menjalankan kewajiban itu dengan baik. tapi apapun itu semoga keluarga korban yang ditinggalkan mendapat ketabahan dan para korban mendapatkan tempat terbaik...		
	Menghapus dokumen (Delete a document)	1. Admin memilih menu model probabilistic. 2. Admin memilih nama kategori yang ingin dihapus.	Document to deleted: pos1 - Positif	- Muncul pesan “filter dokumen dihapus”. - Dokumen latih yang dipilih dihapus dari database.	Ok	

			3. Admin menekan tombol delete.			
6	Input Komentar	Menentukan sentimen dari komentar yang diinput.	<ol style="list-style-type: none"> 1. User berada di halaman homepage 2. User membaca berita online kemudian menulis nama user dan komentar berdasarkan berita. 3. User menekan tombol post. 	<p>Nama: Udin</p> <p>Komentar: “semoga korban pesawat cepat ditemukan”</p>	<ul style="list-style-type: none"> - Muncul komentar yang telah di post. - Muncul sentimen berdasarkan komentar (Positif). 	Ok

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari tugas akhir “Aplikasi Opinion Mining dengan Algoritma Naïve Bayes untuk menilai Berita *Online*” adalah sebagai berikut:

1. Aplikasi dapat menentukan sentimen dari komentar atau opini berdasarkan algoritma Naïve Bayes.
2. Aplikasi dapat melakukan proses text preprocessing berupa *tokenizing*, *filtering*, dan *stemming*.
3. Keakuratan klasifikasi sentimen komentar bergantung pada dokumen data latih.

5.2 Saran

Sebagai langkah pengembangan aplikasi di masa yang akan datang, saran yang dapat diberikan adalah:

1. Menambah fungsi menghitung keakuratan model.
2. Mendeteksi simbol yang mengandung makna tertentu.

DAFTAR PUSTAKA

- [1] Indrawati, Nur. *Natural Language Processing (NLP) Bahasa Indonesia sebagai Preprocessing pada Text Mining*. [Online]. Tersedia di: <http://dmcittelkom.files.wordpress.com/2009/01/tulisan-penelitian-1130500862.doc> [diunduh 15 September 2014]
- [2] Barber, Ian. 2010. *Bayesian Opinion Mining*. [Online]. Tersedia di: <http://phpir.com/bayesian-opinion-mining> [diunduh 10 September 2014]
- [3] Saraswati, Ni Wayan Sumartini. 2011. *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis*. [Online]. Tersedia di: http://www.pps.unud.ac.id/thesis/pdf_thesis/unud-209-236721286-tesis.pdf [diunduh 10 September 2014]
- [4] Berry, M. W.; Kogan, J. 2010. *Text Mining: Application and Theory*. [Online]. Tersedia di: www.mohamedrabee.com/books/book1_1165.pdf [diunduh 15 September 2014]
- [5] Feldman, R.; Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. [Online]. Tersedia di: www.roelsbeestenboel.nl/text.pdf [diunduh 15 September 2014]
- [6] Rozi, Iman F.; Sholeh H. P. dan Erfan A. D. 2012. *Implementasi Opinion Mining untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi*. [Online]. Tersedia di: <http://jurnaleccis.ub.ac.id/index.php/eccis/article/viewFile/164/142> [diunduh 12 September 2014]
- [7] Kurniawan, B.; Syahril E. dan Opim S. S. 2012. *Klasifikasi Konten Berita dengan Metode Text Mining*. [Online]. Tersedia di: <http://jurnal.usu.ac.id/index.php/duniait/article/view/409/212> [diunduh 08 Oktober 2014]