

Aplikasi *Opinion Mining* dengan Algoritma Naïve Bayes untuk Menilai Berita Online

Daniel Pakpahan¹⁾, Hilda Widyastuti²⁾

Politeknik Negeri Batam
Program Studi Teknik Informatika
Parkway Street, Batam Centre, Batam 29461, Indonesia

¹⁾ E-mail: kobe.daniel2@gmail.com

²⁾ E-mail: hilda@polibatam.ac.id

Abstrak

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 adalah 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 hasilnya 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 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 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 is text processing. Text processing includes tokenizing, filtering, and stemming. The next stage, producing probabilistic models whose value will be used in classification process. The core process of the classification is determining the highest probability of each category. If the results indicate the probability Bayes comments for positive category is larger then the comment, then the comments is categorized as a positive opinion and vice versa.

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

1 Pendahuluan

Opini orang lain dapat menjadi informasi sangat penting pada saat membuat keputusan atau 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 [1] 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 di mana 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 dan 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 yang tidak terstruktur. *Text mining* 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). Eksperimen untuk data *review* film dan menghasilkan tingkat akurasi 80% menggunakan metode NBC telah dilakukan oleh [1]. Penelitian [6] menunjukkan NBC merupakan metode yang konvensional dan lebih sederhana dari metode lain dalam proses teks *mining*. NBC memberikan hasil yang tepat dalam mengklasifikasikan opini dalam

bentuk paragraf yang terdiri dari beberapa kalimat menjadi kesimpulan opini positif, negatif dari opini-opini yang terdapat pada berita *online*.

Rumusan masalah penelitian ini adalah : (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*?

Adapun batasan masalah yang terdapat dalam penelitian ini adalah : (1) menggunakan berita bahasa Indonesia, (2) menggunakan komentar dengan format bahasa Indonesia yang baik dan benar, (3) *link* berita *online* yang di-*scrape* berasal dari situs *news.kompas.com*.

Adapun tujuan penelitian ini adalah : (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*.

2 Landasan Teori

2.1 Penelitian Terkait

Penelitian [6] membahas *text mining* dengan metode *Naive Bayes Classifier*(NBC) dan *Support Vector Machine*(SVM) dalam mengklasifikasikan opini berbahasa Inggris dan berbahasa Indonesia. Penelitian tersebut ingin mengetahui metode mana yang memiliki performansi lebih baik untuk diimplementasikan dalam analisis sentimen opini berbahasa Inggris dan berbahasa Indonesia.

TABEL 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

Penelitian tersebut menunjukkan bahwa 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.

Dibandingkan dengan penelitian sebelumnya yang membandingkan metode NBC dan SVM, penelitian ini menambahkan satu kategori lagi, yaitu kategori netral, sehingga terdapat 3 kategori data opini, yaitu opini positif, opini negatif dan opini netral. Penulis juga merancang sebuah sistem berbasis web yang dapat mengklasifikasikan opini secara otomatis sehingga pengujian klasifikasi opini dapat lebih mudah dilakukan.

2.2 *Opinion Mining*

Menurut [2] *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 yang dikemukakan 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 memfokuskan pada layanan analisis sentimen.

Opinion mining bisa dianggap sebagai kombinasi antara *text mining* dan *natural language processing*. Salah satu metode *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.

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 analisis keterhubungan antar dokumen. *Text mining* merupakan proses penemuan akan informasi atau trend baru yang sebelumnya tidak terungkap dengan memproses dan menganalisis data dalam jumlah besar.

Menurut [3] [4] [6] *text mining* bisa dianggap subjek riset yang tergolong baru. *Text mining* dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian atau 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*, statistika dan matematika, *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 aturan-aturan tertentu. Hasil yang diharapkan adalah informasi baru yang tidak terungkap jelas sebelumnya. Menurut [5] saat ini *text mining* telah mendapat perhatian dalam berbagai bidang antara lain:

1. Aplikasi keamanan

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

2. Aplikasi biomedis

Berbagai aplikasi *text mining* dalam literatur 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 penambangan dan analisis. Perangkat lunak *text mining* juga sedang diteliti oleh perusahaan yang berbeda yang bekerja di bidang pencarian dan pengindeks-an secara umum sebagai cara untuk meningkatkan performansinya.

4. Aplikasi media *online*

Text mining digunakan oleh perusahaan media besar seperti perusahaan Tribune, untuk menghilangkan ambiguitas informasi dan untuk memberikan pembaca pengalaman pencarian yang lebih baik, yang meningkatkan loyalitas pada *site* dan pendapatan. Selain itu, editor diuntungkan dengan mampu berbagi, mengasosiasikan dan properti paket berita yang 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 pengindeks-an untuk pencarian. Hal ini terutama berlaku dalam ilmu sains, di mana informasi yang sangat spesifik sering terkandung dalam teks tertulis.

2.4 Algoritma Naïve 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 " $x_1, x_2, x_3, \dots, x_n$ " dimana x_1 adalah kata pertama, x_2 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 persamaannya adalah sebagai berikut:

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

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis sebagai persamaan dapat ditulis sebagai berikut:

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

Persamaan diatas dapat disederhanakan menjadi sebagai berikut:

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

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 persamaannya 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

3 Analisis dan Perancangan

3.1 Deskripsi Umum Sistem



Gambar 1 Deskripsi Umum Sistem

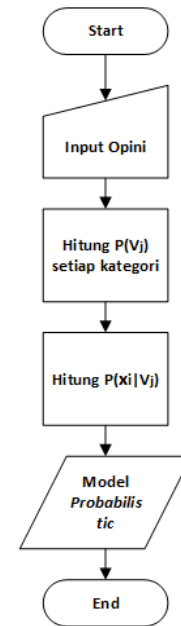
Pada gambar 1, data opini berita *online* dalam teks berbahasa Indonesia merupakan sumber data yang akan diolah ke dalam proses klasifikasi. Tahap pengolahan awal adalah proses *text preprocessing* terhadap baris-baris kalimat opini. Tahap ini terdiri dari beberapa fase yaitu tokenisasi, *filtering*, dan *stemming*.

Tokenisasi adalah memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata. Tokenisasi menghilangkan 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. Sedangkan *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.

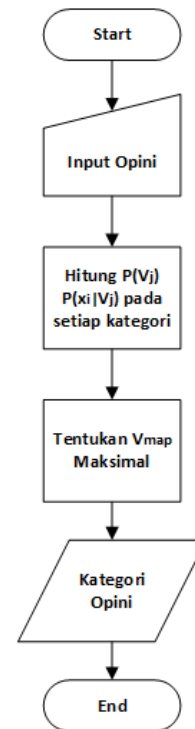
Setelah melewati tahapan *preprocessing*, proses selanjutnya adalah membuat *model probabilistic*. Tahapan ini digunakan untuk menghasilkan model yang nilainya akan digunakan pada tahapan selanjutnya yaitu proses klasifikasi. Proses ini digunakan untuk mendapatkan nilai probabilistic dari $P(V_j)$ dan $P(x_i|V_j)$. Berikut *flowchart* dari proses membuat model probabilistic pada gambar 2.

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 3.



Gambar 2 Flowchart Proses Membuat Model



Gambar 3 Flowchart Proses Klasifikasi

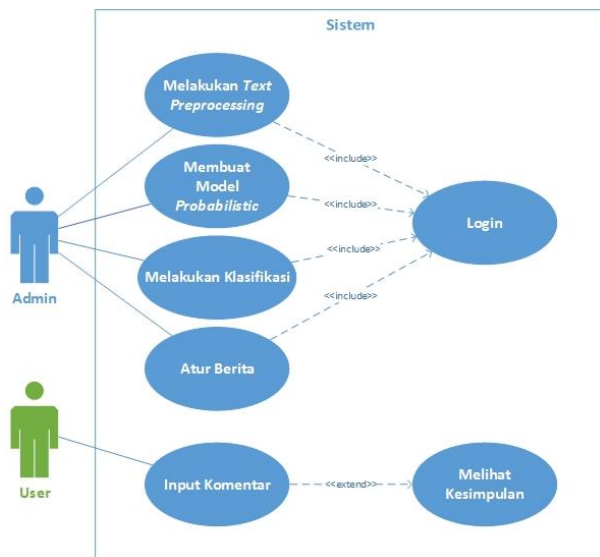
3.2 Kebutuhan Fungsional dan Non Fungsional

Kebutuhan fungsional adalah deskripsi dari layanan-layanan yang disediakan oleh sistem, sedangkan kebutuhan non fungsional adalah karakteristik dan batasan-batasan yang lain yang mendefinisikan sistem. Kebutuhan fungsional meliputi :

- Sistem dapat melakukan *text preprocessing* untuk data opini mentah.
- Sistem dapat membuat model *probabilistic* pada tahap pelatihan untuk digunakan pada proses klasifikasi.
- Sistem dapat melakukan klasifikasi untuk menghitung probabilitas tertinggi, sehingga dapat menentukan kategori opini.
- Sistem mempunyai fitur untuk mengisi opini atau komentar.
- Sistem mempunyai fitur menampilkan hasil kesimpulan berupa informasi dari opini-opini yang diolah.

Kebutuhan non fungsionalnya adalah data opini akan lebih mudah diolah, jika menggunakan bahasa Indonesia yang baik dan benar.

3.3 Use Case Diagram



Gambar 4 Diagram Use Case

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 4.

3.4 Skenario Use Case

Skenario use case ada di table 2 sampai dengan tabel 8.

TABEL 2 SKENARIO 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	
<ol style="list-style-type: none"> Admin melakukan klik terhadap <i>link login</i>. Sistem menampilkan form login untuk meminta memasukkan data <i>username</i> dan <i>password</i>. Admin mengisi form login dengan <i>username</i> dan <i>password</i>. Sistem menampilkan status login dan mengarahkan admin ke halaman yang dituju. 	

TABEL 3 SKENARIO MELAKUKAN *TEXT PREPROCESSING*

Nama use case	Melakukan <i>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	
<ol style="list-style-type: none"> Admin memilih menu <i>text preprocessing</i>. Admin memilih submenu <i>text preprocessing</i> yang diinginkan antara lain: <i>tokenizing</i>, <i>filtering</i> dan <i>stemming</i>. Admin mengisi teks bahasa Indonesia di <i>input text</i>. Admin mengklik <i>button tokenize</i>, <i>filtering</i> atau <i>stemming</i>. Sistem melakukan proses <i>text preprocessing</i> yang dipilih admin. Sistem menampilkan <i>output text</i> dari proses <i>text preprocessing</i> yang dipilih admin. 	

TABEL 4 SKENARIO MEMBAT MODEL *PROBABILISTIC*

Nama use case	Membuat <i>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	
<ol style="list-style-type: none"> Admin memilih menu model <i>probabilistic</i>. Admin melakukan proses model <i>probabilistic</i> (menambah kategori baru, menghapus kategori, menambah dokumen data latih dan menghapus dokumen data latih) Admin mengklik <i>button</i>. Sistem melakukan proses model <i>probabilistic</i> yang dipilih admin dan menambah data ke <i>database</i> atau menghapus data dari <i>database</i>. Sistem menampilkan <i>notice</i> atas aksi yang dilakukan admin. 	

TABEL 5 SKENARIO MELAKUKAN KLASIFIKASI

Nama use case	Melakukan Klasifikasi
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	
<ol style="list-style-type: none"> Admin memilih menu klasifikasi Admin mengisi teks bahasa Indonesia di bagian <i>input text</i>. Admin mengklik <i>button</i> klasifikasi Sistem melakukan proses <i>text preprocessing</i> terlebih dahulu kemudian diikuti proses klasifikasi. Sistem menampilkan kategori opini beserta skor masing-masing opini. 	

TABEL 6 SKENARIO MELIHAT KESIMPULAN

Nama use case	Melihat Kesimpulan
Deskripsi	User dapat melihat kesimpulan dari proses klasifikasi yang dilakukan. Kesimpulan berupa hasil kategori opini, jumlah opini positif, negatif, dan netral, dan sebagainya

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 <i>View Details</i> .
Skenario	
<ol style="list-style-type: none"> User telah menginput komentar. Sistem menampilkan hasil kesimpulan dari proses klasifikasi. Hasil kesimpulan berupa <i>pie chart</i>, jumlah komentar, opini negatif, positif, dan netral. 	

TABEL 7 SKENARIO INPUT KOMENTAR

Nama use case	Input Komentar
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	
<ol style="list-style-type: none"> User berada di homepage bagian kolom komentar. User mengisi komentar teks bahasa Indonesia. User mengklik <i>button post</i> Sistem melakukan proses klasifikasi komentar yang diinput. Sistem menampilkan post komentar dan sentimen opini dari komentar. 	

TABEL 8 SKENARIO ATUR BERITA

Nama use case	Atur Berita
Deskripsi	Admin dapat mengatur berita yang akan tampil di homepage dengan menginput link berita online dari situs <i>news.kompas.com</i> .
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	
<ol style="list-style-type: none"> Admin memilih menu atur berita. Admin mengisi link berita. Admin melakukan klik pada <i>button scrape</i>. Sistem melakukan proses <i>scrape</i> terhadap link berita. Sistem menampilkan gambar, judul, waktu dan headline berita. 	

3.5 Diagram Class

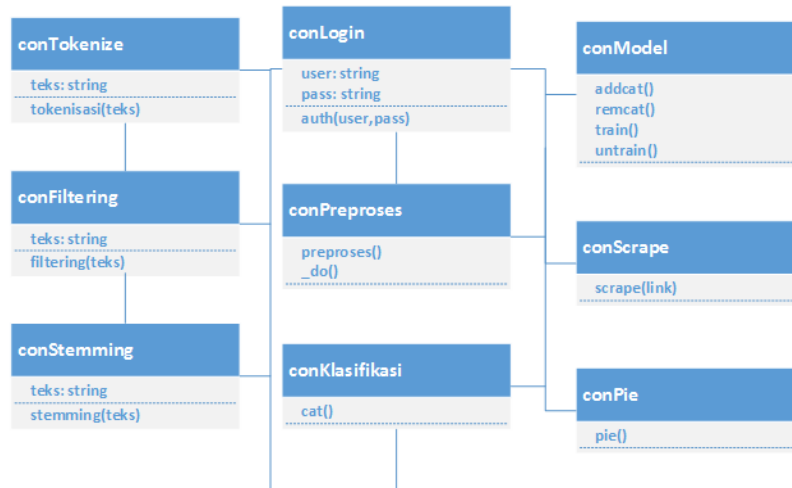
Diagram *class* perancangan menggambarkan bagaimana keterhubungan antar kelas dan berbagai metode di dalamnya yang akan diimplementasikan. Diagram *class* perancangan dapat dilihat di gambar 5. Ada sembilan *class* yang dibuat, yaitu *class* conTokenize, *class* conFiltering, *class* conStemming, *class* conLogin, *class* conPreproses, *class* conKlasifikasi, *class* conModel, *class* conScrape, dan *class* conPie.

3.6 Perancangan Basisdata

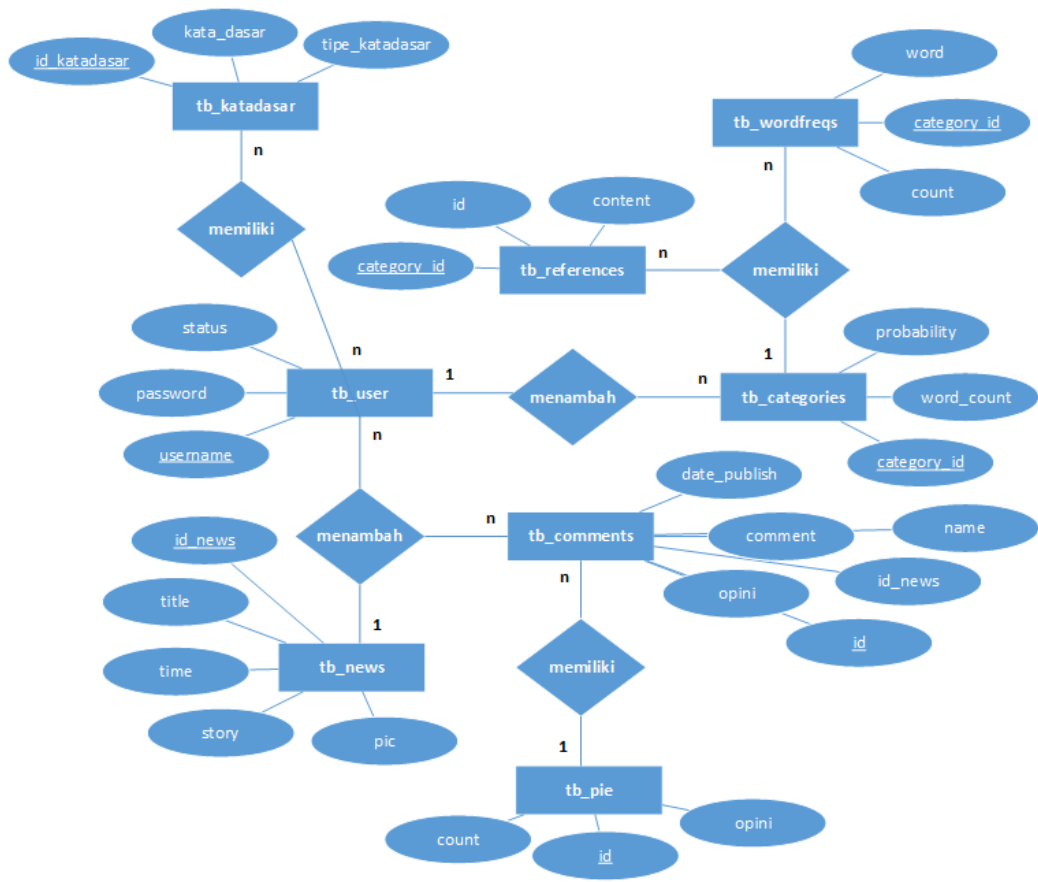
Perancangan diagram entity relationship (ER) dapat dilihat pada gambar 6. Sedangkan perancangan skema desain basisdata yang akan digunakan dalam sistem ini

dapat dilihat pada gambar 7. Penjelasan singkat mengenai gambar diagram ER adalah :

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*.



Gambar 5 Class Diagram



Gambar 6 ER Diagram

tb_user PK username password nama status	tb_categories PK category_id probability word_count	tb_news PK id_news title time pic story
tb_comments PK id id_news name comments date_publish opini	tb_katadasar PK id_katadasar katadasar tipe_katadasar	tb_references PK id category_id content
	tb_pie PK id opini count	tb_wordfreqs PK id category_id count

Gambar 7 Desain Basisdata

4 Implementasi dan Pengujian

4.1 Implementasi Antarmuka Halaman Model Probabilistic

Halaman model probabilistic di gambar 8 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 screenshot shows four panels in a 2x2 grid. Top-left: 'Add Category' with a text input for 'Name of the category', an 'Add' button, and a dropdown for 'Category to delete' (set to 'Netral'). Top-right: 'Delete a Category' with a dropdown for 'Category to delete' (set to 'Netral'). Bottom-left: 'Train Filter' with a text input for 'Document Identifier', a dropdown for 'Category for the document' (set to 'Netral'), a large text area for 'Copy/paste here the document', and a 'Train' button. Bottom-right: 'Delete a Document' with a dropdown for 'Document to be deleted' (set to 'post1 - Positif') and a 'Delete' button.

Gambar 8 Halaman Model Probabilistic

4.2 Implementasi Antarmuka Halaman Atur Berita

Halaman atur berita di gambar 9 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.

The screenshot shows the 'Atur Berita' interface. At the top, there's a 'Dashboard' and 'Atur Berita' link. Below is a description: 'Scrape berita online yang akan dijadikan berita di front page.' There's a 'KOMPAS.com' logo and a 'Link Berita:' input field with a 'Scrape' button. Below is a table with columns: Gambar, Judul, Waktu, and Headline. The first row shows a news item with a thumbnail, title 'Tim SAR Temukan 7 Obyek yang Diduga Bagian AirAsia QZ8501', time 'Kamis, 8 Januari 2015 | 20:30 WIB', and a headline starting with 'JAKARTA, KOMPAS.com — Tim SAR gabungan mendeteksi tujuh benda di bawah permukaan laut yang diduga bagian dari pesawat AirAsia...'

Gambar 9 Halaman Atur Berita

4.3 Implementasi Antarmuka Komentar

Antarmuka komentar di gambar 10 memiliki fungsi sebagai berikut:

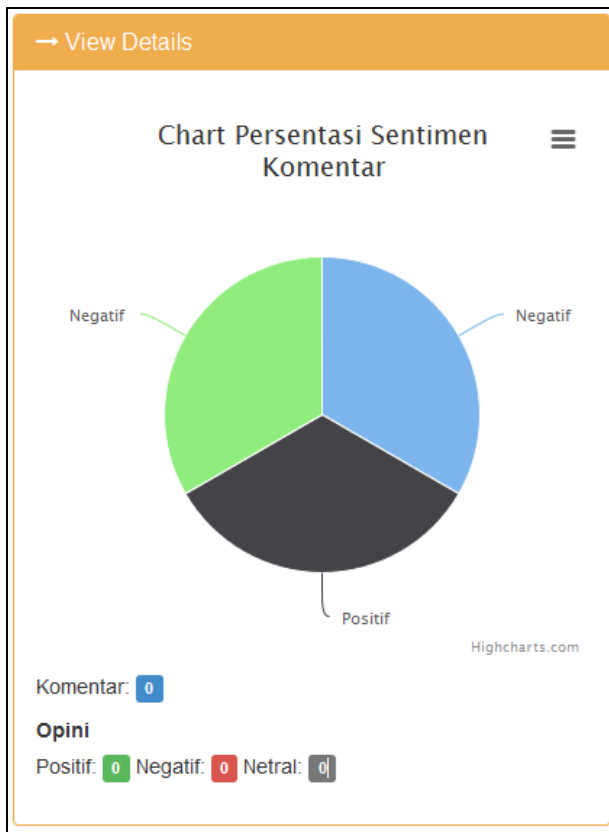
- Memasukkan nama dan komentar bahasa Indonesia yang baik dan benar.
- Setelah tombol *post* ditekan, maka *input* komentar akan diproses dan menampilkan komentar ditambah sentimen opini dari komentar tersebut.

The screenshot shows a comment form. It has a 'Komentar:' label, a text input for the name (filled with 'Leo'), a text area for the comment (filled with 'hebat'), and a 'Post' button. Below the form, the submitted comment is shown: 'Leo' with a 'Positif' sentiment label, the timestamp '2015-01-09 02:20:27', and the comment text 'hebat'.

Gambar 10 Antarmuka Komentar

4.4 Implementasi Antarmuka Melihat Kesimpulan

Antarmuka *view details* di gambar 11 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 4 Antarmuka View Details

5 Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan penelitian ini 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] I. Barber, "Bayesian Opinion Mining", Tersedia di <http://phpir.com/bayesian-opinion-mining> [diunduh 10 September 2014]
- [2] I. F. Rozi, S.H. Pramono, E. A. Dachlan, "Implementasi Opinion Mining untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi", *Jurnal EECCIS Vol. 6, No. 1*, p37-43, 2012, 2012.
- [3] B. Kurniawan, S. Effendi, S. S. Sitompul, "Klasifikasi Konten Berita dengan Metode Text Mining", *Jurnal Dunia Teknologi Informasi*, Vol. 1, No. 1, p14-19, 2012.
- [4] N. Indrawati, "Natural Language Processing (NLP) Bahasa Indonesia sebagai Preprocessing pada Text Mining", tersedia di <http://dmcittelkom.files.wordpress.com/2009/01/tulisan-penelitian-1130500862.doc> [diunduh 15 September 2014]
- [5] N. W. S. Saraswati, "Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis", *Tesis*, Program Pascasarjana Universitas Udayana, Denpasar, Indonesia, 2011.
- [6] R. Feldman, J. Sanger, "The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data", *Book*, Cambridge University Press, 2007.