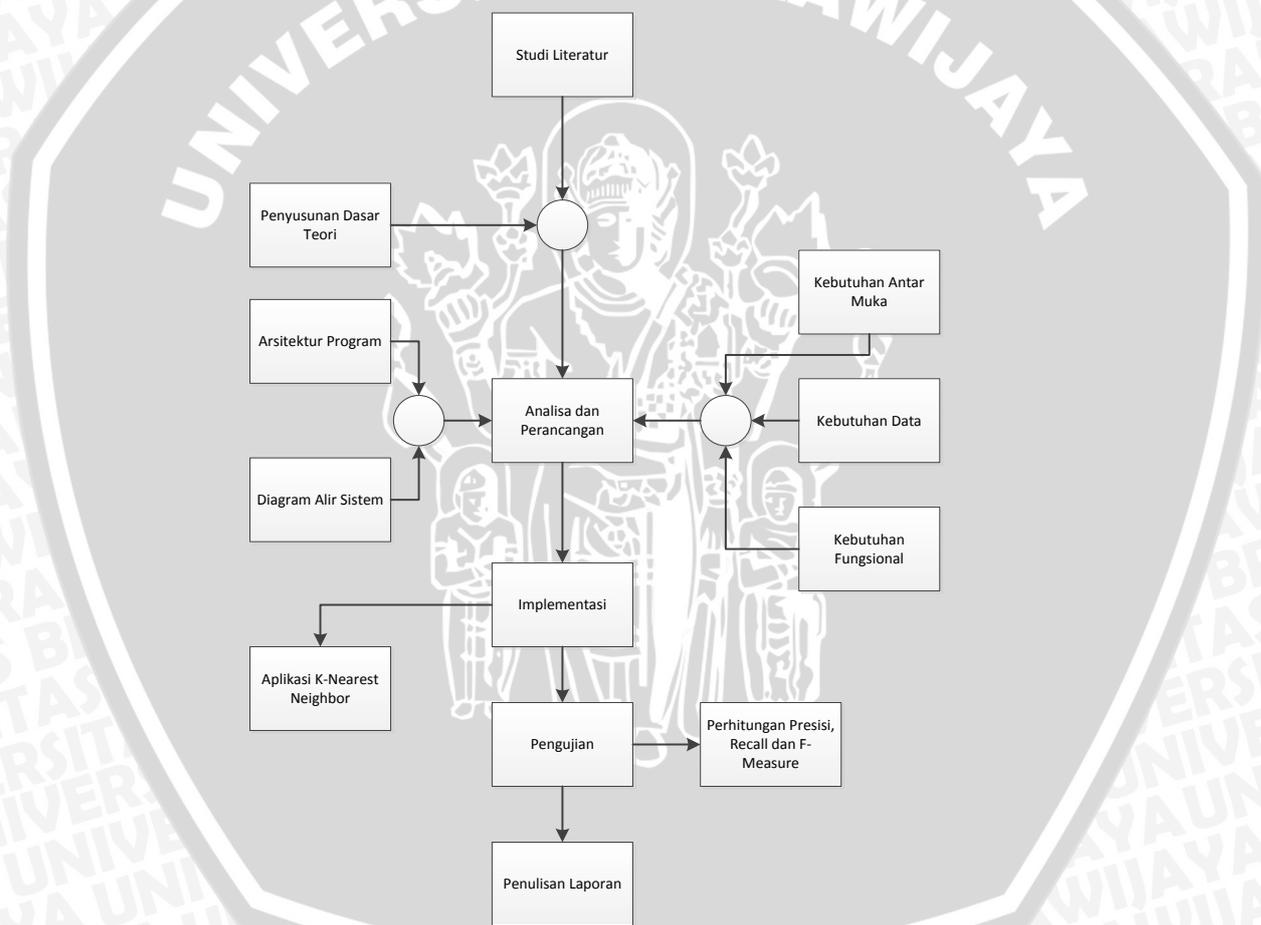


BAB III

METODE PENELITIAN DAN PERANCANGAN

Pada bab ini akan dijelaskan langkah-langkah yang dilakukan dalam pengerjaan tugas akhir, yaitu studi literature, penyusunan dasar teori, analisa dan perancangan, implementasi, analisis dan pengujian dari aplikasi perangkat lunak yang akan dibuat, hingga penulisan laporan. Kesimpulan dan saran disertakan sebagai catatan atas aplikasi dan kemungkinan arah pengembangan aplikasi selanjutnya. Gambar 3.1 menunjukkan desain skripsi secara umum.



Gambar 3. 1 Desain Skripsi

Sumber: Perancangan

3.1 Studi Literatur

Studi literatur mempelajari mengenai penjelasan dasar teori yang digunakan untuk menunjang penulisan skripsi. Teori-teori pendukung tersebut diperoleh dari buku, jurnal, e-book, skripsi sebelumnya, dan dokumentasi project.

3.2 Penyusunan Dasar Teori

Penyusunan dasar teori dilakukan setelah mendapatkan referensi yang tepat untuk mendukung penulisan skripsi ini. Teori-teori pendukung tersebut meliputi:

1. *Text Mining*

Meliputi tahap-tahap *Cleaning, Casefolding, Parsing, Filtering/stopword removal, stemming* untuk menghasilkan term yang nantinya akan di proses oleh program.

2. *K-Nearest Neighbor*

Meliputi tahap-tahap pengklasifikasian untuk menghasilkan ringkasan sesuai dengan metode yang telah di pilih oleh peneliti.

3.3 Analisa dan Perancangan

Analisis kebutuhan bertujuan untuk mendapatkan semua kebutuhan yang diperlukan oleh sistem yang akan dibangun. Analisis kebutuhan dilakukan dengan mengidentifikasi kebutuhan sistem dan siapa saja yang terlibat di dalamnya. Berikut analisis kebutuhan dalam aplikasi *sentiment analysis* pada *review* barang ini adalah:

3.3.1 Kebutuhan Antar Muka

Kebutuhan-kebutuhan untuk pengembangan perangkat lunak ini sebagai berikut :

1. Program yang akan dibangun harus mempunyai tampilan yang familiar bagi pemakai.
2. Program yang di bangun harus mempunyai tampilan yang memungkinkan user untuk menginputkan teks halaman yang akan diproses.

3. Program yang di bangun harus mampu menampilkan ringkasan yang telah di proses sebelumnya.

3.3.2 Kebutuhan Data

Data yang diolah oleh perangkat lunak ini adalah:

- 1) Data kata dasar bahasa indonesia yang berfungsi pada saat proses *stemming* dan pemisahan kata kerja/kata benda.
- 2) Data berupa *review* atau komentar yang mempunyai tatanan bahasa yang terstruktur.
- 3) Data hasil ringkasan yang akan di simpan di dalam database maupun aplikasi tambahan lainnya.

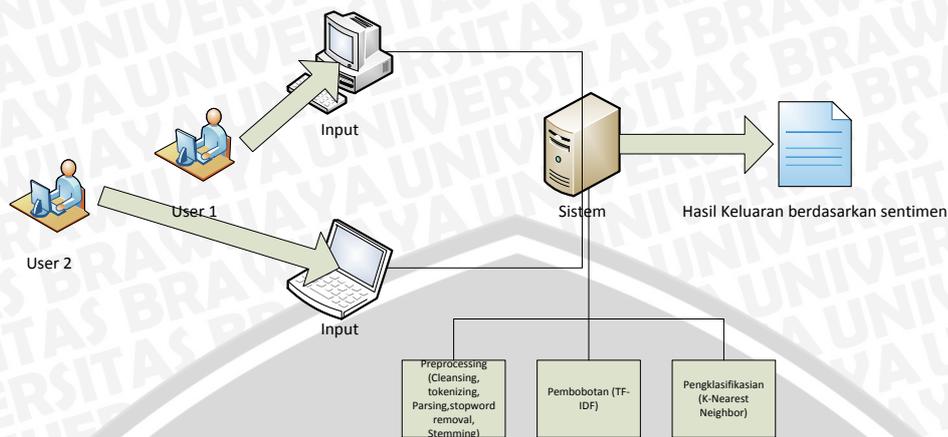
3.3.3 Kebutuhan Fungsional

Fungsi-fungsi yang dimiliki oleh perangkat lunak ini adalah:

- 1) Program harus mampu melakukan proses *preprocessing*, yakni berupa *Cleaning, parsing, Casefolding, stopword removal dan stemming*.
- 2) Program harus mampu membedakan kata yang mempunyai sentimen positif atau negatif dari setiap kalimat yang telah melalui *preprocessing*.
- 3) Perangkat lunak harus mampu melakukan proses pembobotan pada setiap kalimat berdasarkan metode yang akan digunakan.
- 4) Perangkat lunak harus mampu menghasilkan ringkasan berdasarkan peringkat dari bobot yang telah di hitung sebelumnya.

3.3.4 Arsitektur Program

Perancangan aplikasi *sentiment analysis* untuk teks *review online* dapat dilihat lebih jelas pada gambar arsitektur di bawah :



Gambar 3. 2 Arsitektur Perancangan Aplikasi

Sumber: Perancangan

Pada gambar diatas dijelaskan bahwa user menginputkan link yang berisikan *review* atau komentar seseorang atau beberapa orang untuk sebuah produk yang terdapat pada web Bhinneka.com.

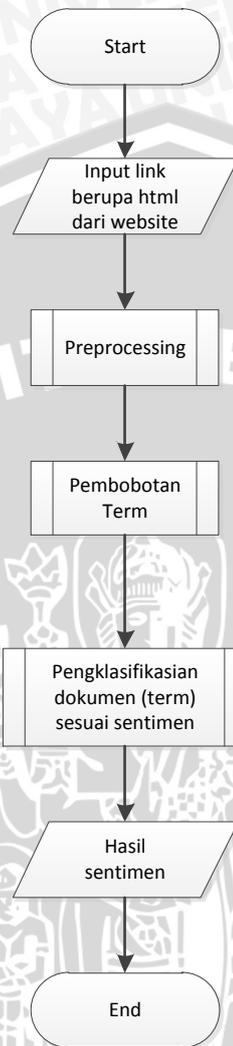
Sistem yang menerima teks masukan akan memproses dengan dimulai dari proses *preprocessing* yang terdiri dari *Cleaning* untuk menghapus inputan selain huruf, setelah *Cleaning* proses selanjutnya adalah *Casefolding* yang mengolah isi *term/kata* untuk mengecilkan semua huruf besar agar hasil menjadi lebih baik, selanjutnya dari proses *Casefolding* adalah proses *parsing* untuk memecah dokumen menjadi kata-kata. Proses selanjutnya adalah *stopword removal* adalah proses untuk menghilangkan kata-kata kurang penting atau kurang berguna untuk sistem. Proses terakhir pada *preprocessing* adalah *stemming* yang akan mengubah seluruh kata berimbuhan menjadi kata dasar.

Untuk mengubah kata/*term* agar bisa dibaca oleh komputer, data diubah menjadi numerik dengan cara pembobotan. Setelah diberikan bobot pada masing-masing *term*, masuk ke proses pengklasifikasian untuk didapatkan hasilnya apakah termasuk dalam kelas sentimen positif, negatif atau netral.

3.3.5 Diagram Alir Sistem

Diagram alir menggunakan notasi-notasi untuk menggambarkan arus data yang membantu dalam proses memahami jalannya aplikasi. Gambar 3.3

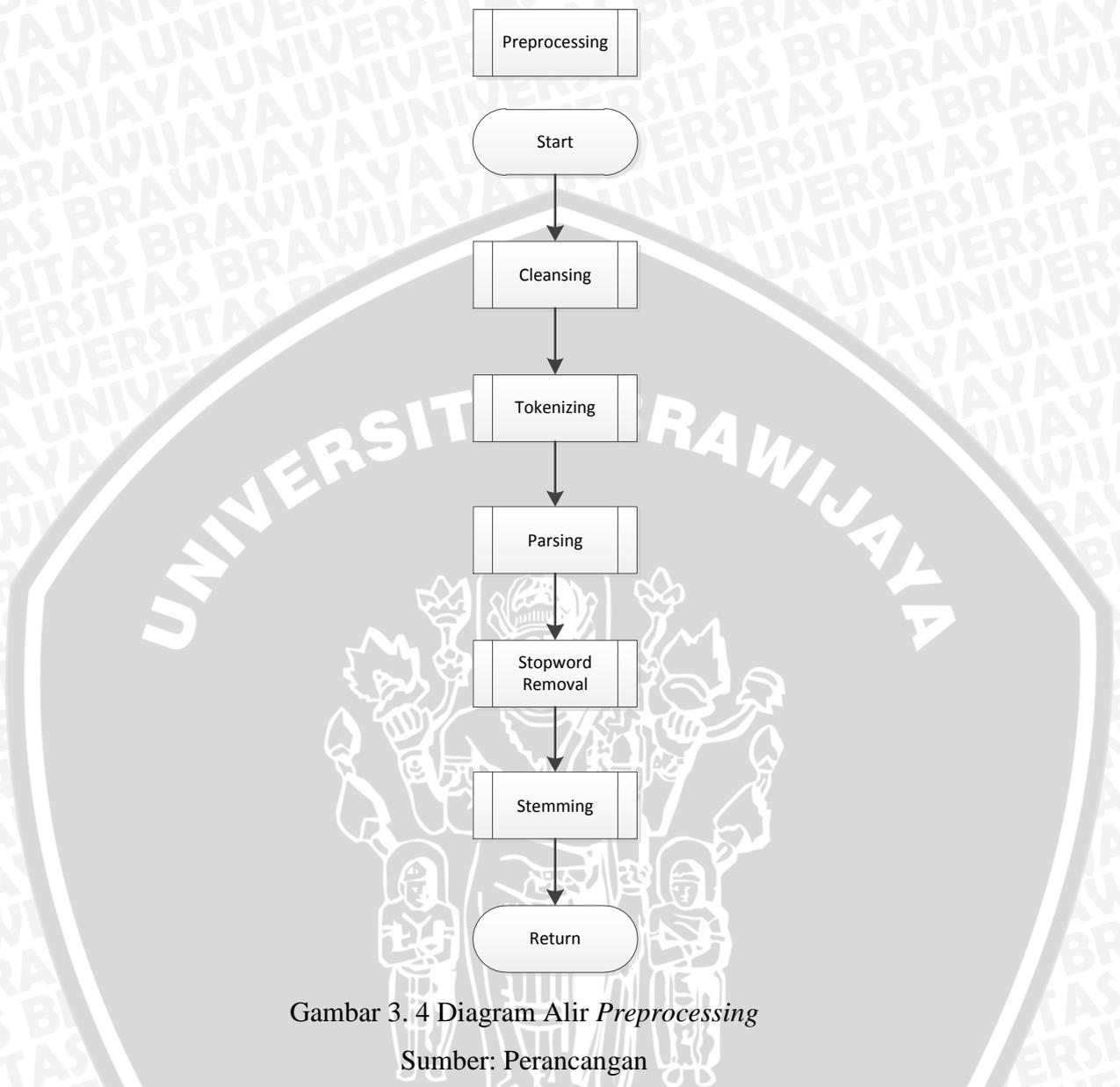
menunjukkan diagram alir aplikasi *Sentiment analysis* pada *review* barang berbahasa Indonesia dengan metode K-Nearest Neighbor.



Gambar 3. 3 Diagram Alir Sistem Keseluruhan

Sumber: Perancangan

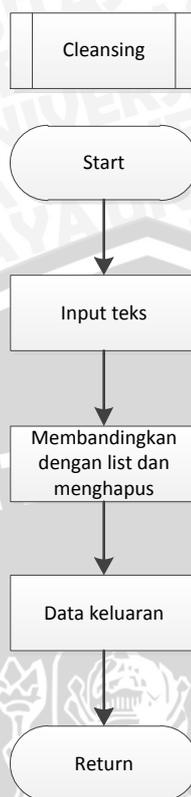
Pada proses sistem secara keseluruhan, user mulai dengan memilih teks hasil dari pemilihan dokumen pada *website* *bhinneka.com*. Teks yang dipilih user tadi akan diolah oleh sistem didalam proses *preprocessing* dan dilanjutkan proses pembobotan *term*. Setelah data sudah diberi bobot, sistem mengolah data untuk selanjutnya diklasifikasikan sesuai sentimennya. Hasil sentimen dari proses yang dilakukan sistem akan ditampilkan oleh *user interface* agar mudah untuk dianalisa dan dilihat. Selanjutnya gambar 3.4 akan menjelaskan bagaimana proses *preprocessing* berjalan dan bagaimana *step per step* prosesnya.



Gambar 3. 4 Diagram Alir *Preprocessing*

Sumber: Perancangan

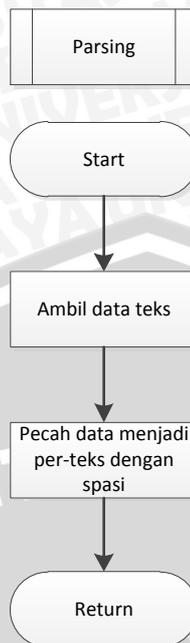
Pada proses *preprocessing* yang terdiri dari *Cleaning* untuk menghapus inputan selain huruf, setelah *Cleaning* proses selanjutnya adalah *Casefolding* yang mengolah isi *term/kata* untuk mengecilkan semua huruf besar agar hasil menjadi lebih baik, selanjutnya dari proses *Casefolding* adalah proses *parsing* untuk memecah dokumen menjadi kata-kata. Proses selanjutnya adalah *stopword removal* adalah proses untuk menghilangkan kata-kata kurang penting atau kurang berguna untuk sistem. Proses terakhir pada *preprocessing* adalah *stemming* yang akan mengubah seluruh kata berimbuhan menjadi kata dasar



Gambar 3. 5 Proses *Cleaning*

Sumber: Perancangan

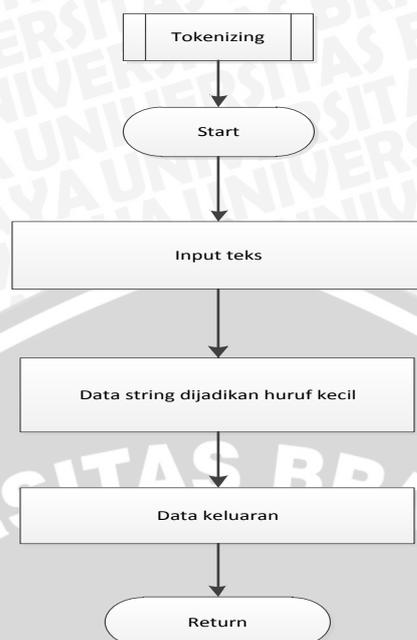
Cleaning adalah proses membersihkan dokumen dari karakter-karakter yang tidak diperlukan untuk mengurangi *noise* seperti emotikon dan simbol-simbol.



Gambar 3. 6 Proses Parsing

Sumber: Perancangan

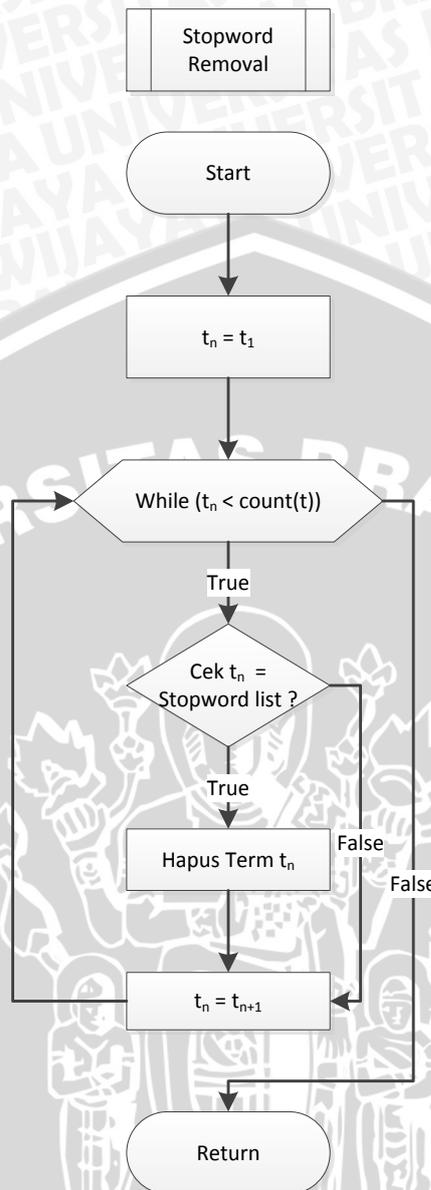
Proses *parsing* adalah proses untuk memecah teks bebas yang besar menjadi bagian-bagian yang disebut kalimat [CDF-08]. Kalimat-kalimat yang dihasilkan kemudian dipecah lagi menjadi kata-kata melalui proses Casefolding.



Gambar 3. 7 Proses Casefolding

Sumber: Perancangan

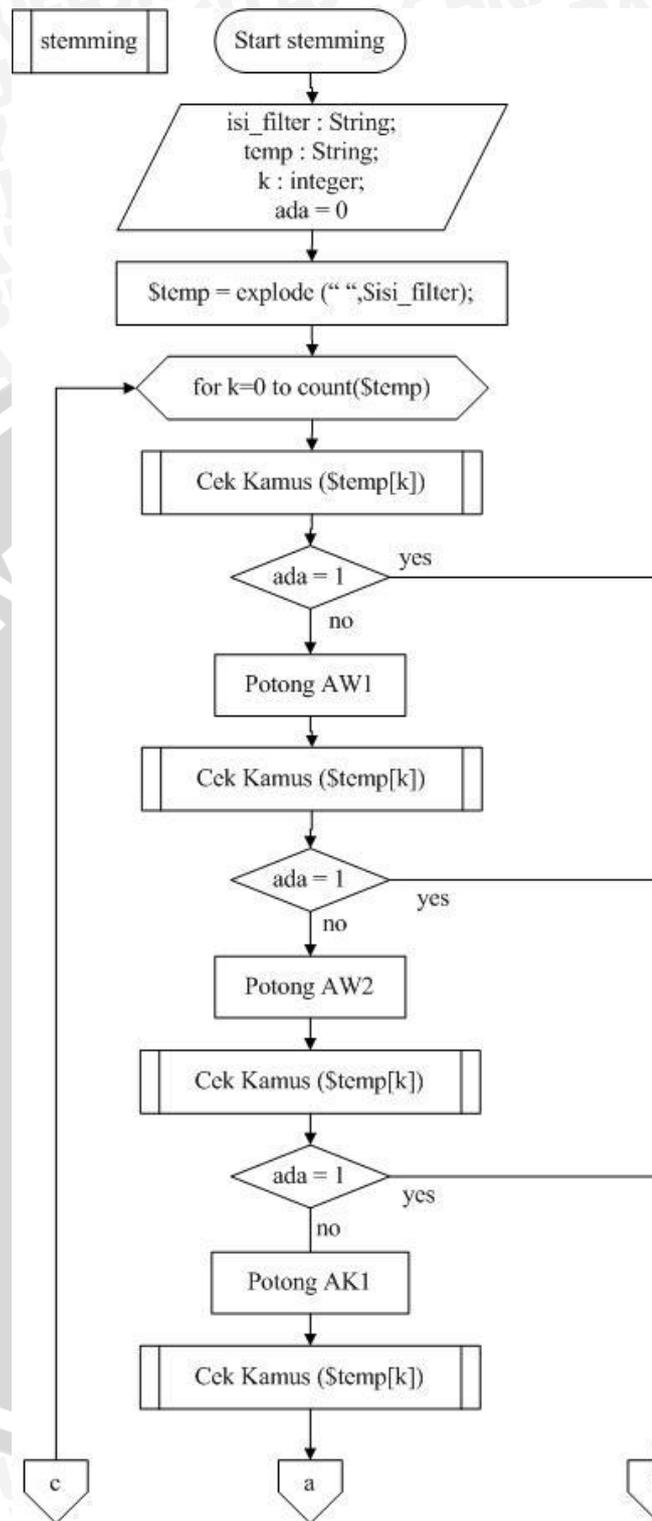
Proses *Casefolding* dilakukan untuk mengolah isi *term/kata* dan mengecilkan semua huruf besar agar hasil menjadi lebih baik. Setelah dokumen di dapatkan kemudian siap diproses untuk menghilangkan kata-kata *noise* pada proses *stopword removal*. Gambar 3.7 menunjukkan diagram alir *Casefolding*.

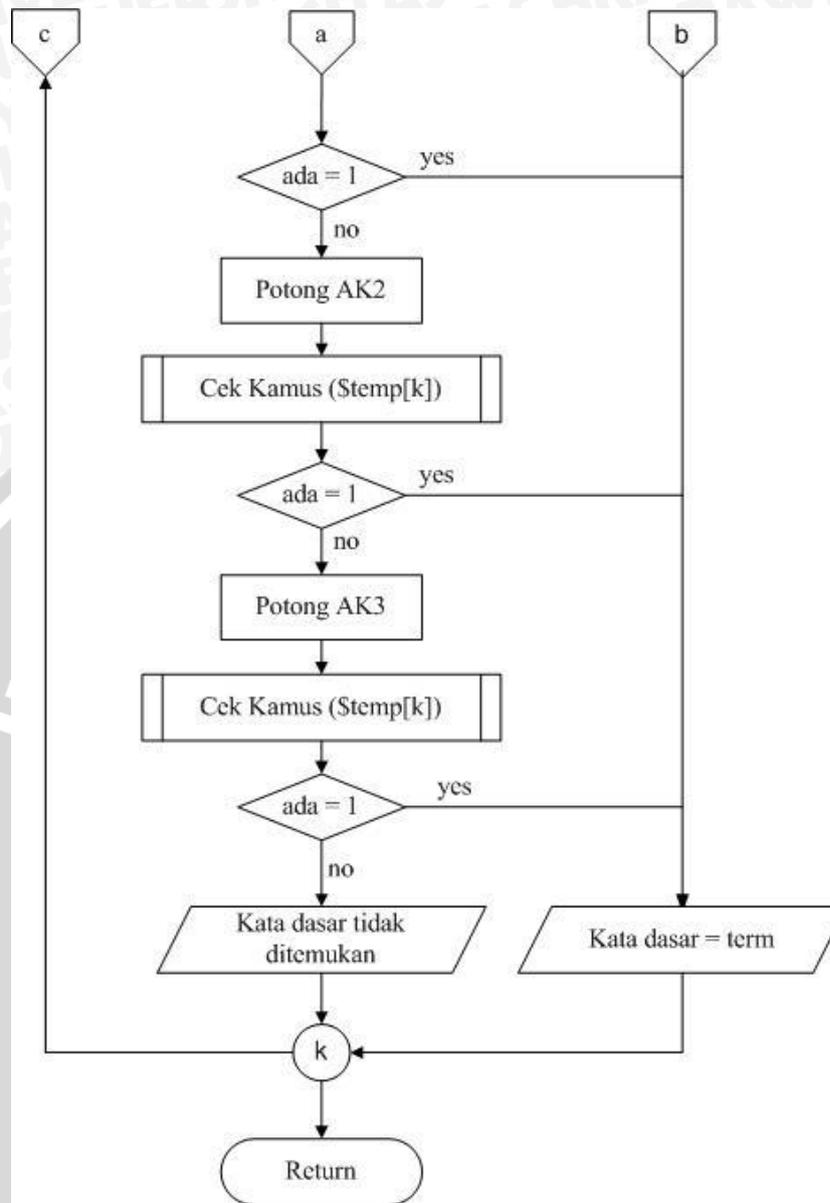


Gambar 3. 8 Proses Stopword Removal

Sumber: Perancangan

Stopword Removal adalah sebuah proses penyaringan untuk menghilangkan kata yang 'tidak relevan' pada hasil Casefolding sebuah dokumen teks dengan cara membandingkannya dengan *Stoplist (Stopword list)* yang ada [AHA-10]. Contoh dari *Stopword* misalnya, kata sambung, artikel dan preposisi.





Gambar 3. 9 Diagram Alir Proses Stemming Arifin-Setiono

Sumber:Perancangan

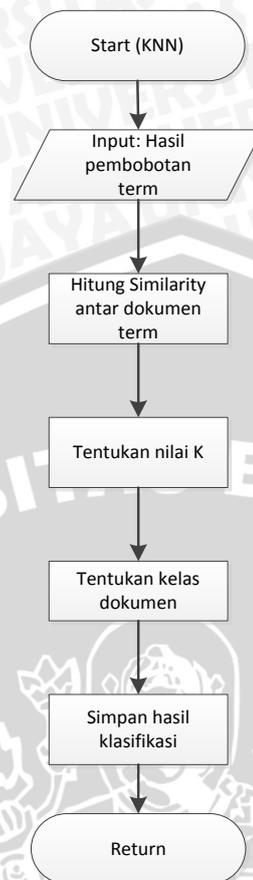
Stemming arifin-setiono dilakukan untuk mendapatkan kata dasar dari sebuah *term*. Proses ini digunakan untuk mencari kata dasar dari suatu *term*. Gambar 3.9 menunjukkan diagram alir *stemming* arifin-setiono.



Gambar 3. 10 Proses Pembobotan

Sumber: Perancangan

Term yang sudah diproses melalui *stemming* kemudian dihitung bobotnya dengan menggunakan TF-IDF. Pembobotan digunakan untuk menjadikan setiap kata dari dokumen untuk menjadi numeric agar dapat di proses oleh computer. Gambar 3.10 menjelaskan alur dari proses pembobotan.



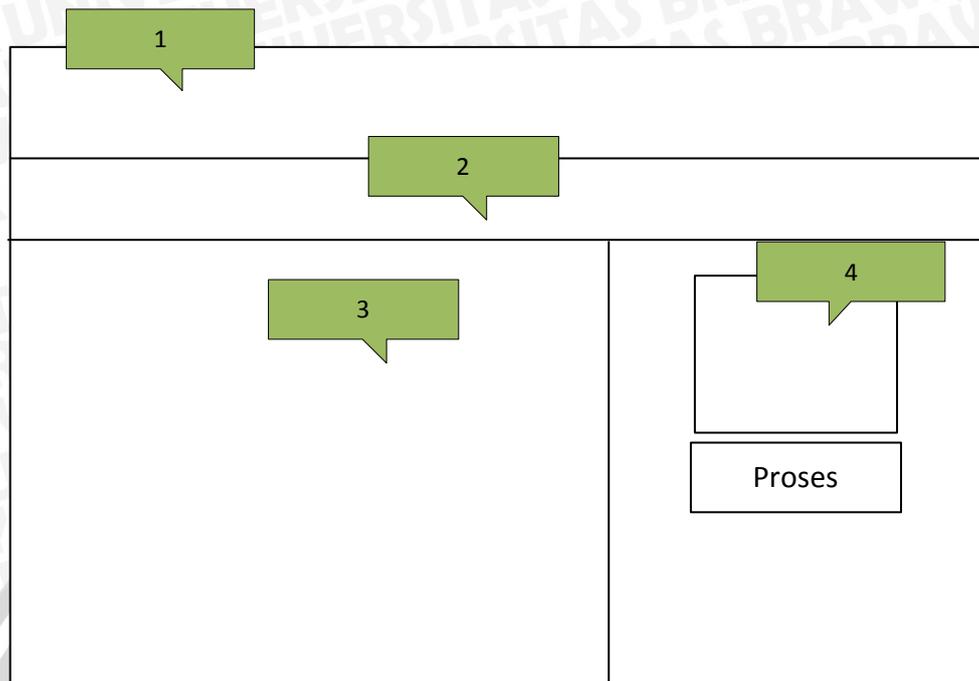
Gambar 3.11 Proses K-Nearest Neighbor (KNN)

Sumber: Perancangan

Term yang sudah diproses melalui *stemming* kemudian dihitung bobotnya dengan menggunakan TF-IDF. Hasil dari perhitungan bobot kemudian akan disimpan untuk proses klasifikasi dengan menggunakan algoritma *K-Nearest Neighbor* (KNN). Gambar 3.11 menunjukkan diagram alir KNN.

3.4 Desain Antar Muka

Perancangan desain antar muka sistem analisa sentimen ini dapat dilihat pada gambar 3.12. Perancangan desain ini bertujuan agar pengguna dapat nyaman menggunakan aplikasi yang akan dirancang.



Gambar 3. 12 Perancangan Desain Antar Muka Halaman Awal Pengujian

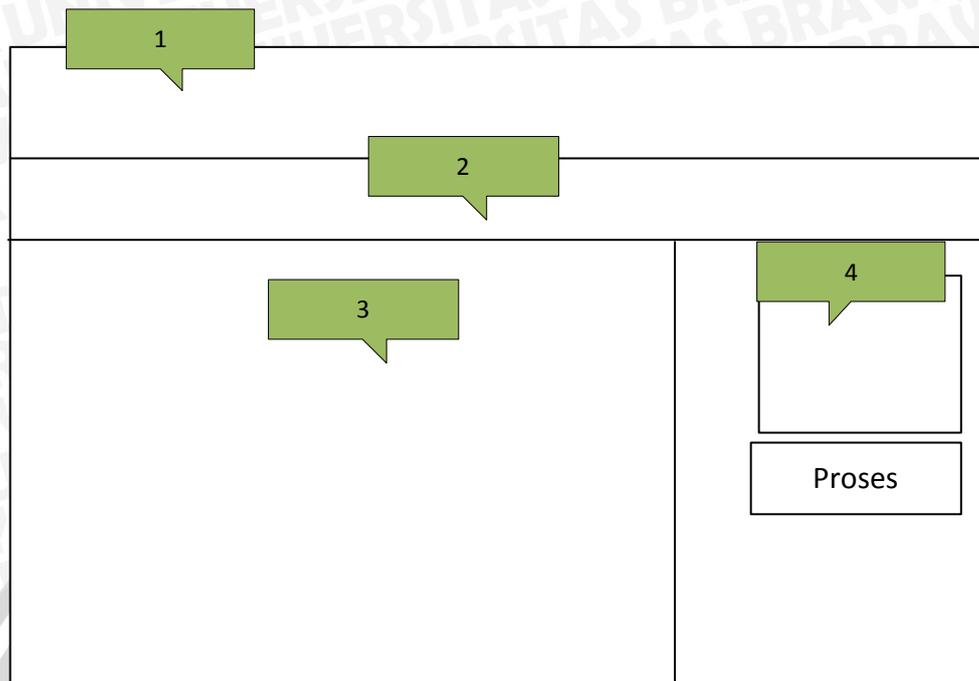
Sumber: Perancangan

Keterangan gambar 3.12:

1. *Header* dan judul aplikasi
2. Tab menu
3. Kumpulan data
4. Tombol Proses

Desain antar muka sistem ini terdiri dari empat komponen utama. Isi pada antar muka sistem adalah field untuk menerima input pengguna berupa tombol *check box* untuk memilih data yang akan diproses dan satu tombol proses untuk mengirim masukan dari pengguna ke sistem.

Pada halaman hasil, akan ditampilkan sampel dokumen yang diuji dan grafik hasil sentiment analysis. Perancangan desain antarmuka untuk halaman hasil dapat dilihat pada gambar 3.13.



Gambar 3. 13 Perancangan Desain Antar Muka Halaman Hasil

Sumber: Perancangan

Keterangan gambar 3.13:

1. Header dan judul
2. Tab menu
3. Hasil keluaran aplikasi
4. Tombol proses

3.5 Manualisasi Analisis Sentimen

Penghitungan manual berfungsi untuk memberikan gambaran umum perancangan sistem yang akan dibangun. Contoh manualisasi pelatihan sistem analisis sentimen pada *review* barang dengan metode *K-Nearest Neighbor* adalah sebagai berikut:

1. Input dokumen *review*Tabel 3. 1 Tabel Manualisasi Dokumen *Review*

ID	Review	Sentimen
1	Kamera itu bagus, saya suka sekali	positif
2	D7000 pertama memegang...wah.....bingung...karna selama ini makenya d5000...setelah seharian utak atik....layaknya seorang profesional....apalagi fokusnya yg 39 itu loh...sangat mantap...salut buat d7000	Positif
3	Saya terus terang tidak suka d7000 karena kurang mantap rasanya	negatif
4	Jelek banget bro, buat diluar ga jelas	negatif
5	Kamera ini berharga 5.400.000	positif
x	Ga nyangka bagus banget, ga sabar beli barangnya	?

Sumber: Perancangan

2. Preprocessing

a. *Cleaning*Tabel 3. 2 Tabel Manualisasi Dokumen *Cleaning*

ID	Review	Sentimen
1	Kamera itu bagus, saya suka sekali	positif
2	D pertama memegang wah bingung karna selama ini makenya d setelah seharian utak atik layaknya seorang profesional apalagi fokusnya yg itu loh sangat mantap salut buat d	Positif
3	Saya terus terang tidak suka d karena kurang mantap rasanya	negatif
4	Jelek banget bro, buat diluar ga jelas	negatif
5	Kamera ini berharga	positif
x	Ga nyangka bagus banget, ga sabar beli barangnya	?

Sumber: Perancangan

b. *Casefolding*Tabel 3. 3 Tabel Manualisasi Dokumen *Casefolding*

ID	Review	Sentimen
1	kamera itu bagus, saya suka sekali	positif
2	d pertama memegang wah bingung karna selama ini makenya d setelah seharian utak atik layaknya seorang profesional apalagi fokusnya yg itu loh sangat mantap salut buat d	Positif
3	saya terus terang tidak suka d karena kurang mantap rasanya	negatif
4	jelek banget bro, buat diluar ga jelas	negatif
5	kamera ini berharga	positif
x	ga nyangka bagus banget, ga sabar beli barangnya	?

Sumber: Perancangan

c. *Parsing*Tabel 3. 4 Tabel Manualisasi Dokumen *Parsing*

ID	Review	Sentimen
1	kamera // itu // bagus // saya // suka // sekali	positif
2	d // pertama // memegang // wah // bingung // karna // selama // ini // makenya // d // setelah // seharian // utak // atik // layaknya // seorang // profesional // apalagi // fokusnya // yg // itu // loh // sangat // mantap // salut // buat // d	positif
3	saya // terus // terang // tidak // suka // d // karena // kurang // mantap // rasanya	negatif
4	jelek // banget // bro // buat // diluar // ga // jelas	negatif
5	kamera // ini // berharga //	positif
x	ga // nyangka // bagus // banget // ga // sabar // beli // barangnya	?

Sumber: Perancangan

d. *Stopword Removal*

Tabel 3. 5 Tabel Manualisasi Dokumen *Stopword Removal*

ID	Review	Sentimen
1	kamera bagus suka	positif
2	pertama bingung selama seharian layaknya seorang fokusnya sangat mantap salut	Positif
3	terang tidak suka kurang rasanya	negatif
4	jelek banget buat diluar jelas	negatif
5	kamera berharga	positif
x	bagus sabar beli barangnya	?

Sumber: Perancangan

e. *Stemming*

Tabel 3. 6 Tabel Manualisasi Dokumen *Stemming*

ID	Review	Sentimen
1	kamera bagus suka	positif
2	pertama bingung lama hari layak orang fokus sangat mantap salut	Positif
3	terang tidak suka kurang rasa	negatif
4	jelek buat luar jelas	negatif
5	kamera harga	positif
x	bagus sabar beli barang	?

Sumber: Perancangan

3. Pembobotan

Setelah dilakukan proses *preprocessing*, selanjutnya dibuat tabel informasi dokumen yang berisi frekuensi *term* dan frekuensi dokumen.

Tabel 3. 7 Tabel Informasi Dokumen

informasi dokumen										TF-IDF					
no	term	D1	D2	D3	D4	D5	Dx	DF	IDF	D1	D2	D3	D4	D5	Dx
1	kamera	1	0	0	0	1	0	2	0,398	0,398	0,000	0,000	0,000	0,398	0,000
2	bagus	1	0	0	0	0	1	1	0,699	0,699	0,000	0,000	0,000	0,000	0,699
3	suka	1	0	0	0	0	0	1	0,699	0,699	0,000	0,000	0,000	0,000	0,000
4	pertama	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
5	bingung	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
6	lama	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
7	hari	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
8	layak	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
9	orang	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
10	fokus	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
11	sangat	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
12	mantap	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
13	salut	0	1	0	0	0	0	1	0,699	0,000	0,699	0,000	0,000	0,000	0,000
14	terang	0	0	1	0	0	0	1	0,699	0,000	0,000	0,699	0,000	0,000	0,000
15	tidak	0	0	1	0	0	0	1	0,699	0,000	0,000	0,699	0,000	0,000	0,000
16	kurang	0	0	1	0	0	0	1	0,699	0,000	0,000	0,699	0,000	0,000	0,000
17	rasa	0	0	1	0	0	0	1	0,699	0,000	0,000	0,699	0,000	0,000	0,000
18	jelek	0	0	0	1	0	0	1	0,699	0,000	0,000	0,000	0,699	0,000	0,000
19	buat	0	0	0	1	0	0	1	0,699	0,000	0,000	0,000	0,699	0,000	0,000
20	luar	0	0	0	1	0	0	1	0,699	0,000	0,000	0,000	0,699	0,000	0,000
21	jelas	0	0	0	1	0	0	1	0,699	0,000	0,000	0,000	0,699	0,000	0,000
22	sabar	0	0	0	0	0	1	1	0,699	0,000	0,000	0,000	0,000	0,000	0,699
23	beli	0	0	0	0	0	1	1	0,699	0,000	0,000	0,000	0,000	0,000	0,699
24	barang	0	0	0	0	0	1	1	0,699	0,000	0,000	0,000	0,000	0,000	0,699
25	harga	0	0	0	0	1	0	1	0,699	0,000	0,000	0,000	0,000	0,699	0,000

Sumber: Perancangan

Proses perhitungan dimulai dari pembobotan TF yaitu menghitung pembobotan kata (*term*) yang didasarkan pada perhitungan jumlah kata yang muncul pada suatu dokumen, misal kata **kamera** muncul satu kali pada dokumen 1 dan muncul satu kali pada dokumen 5.

Tabel 3. 8 Tabel Informasi TF

Informasi Term Frequency							
no	term	D1	D2	D3	D4	D5	Dx
1	kamera	1	0	0	0	1	0

Sumber: Perancangan

Nilai TF akan digunakan untuk mencari nilai DF yaitu menghitung pembobotan kata (*term*) yang didasarkan pada perhitungan jumlah kata muncul pada semua dokumen. Kata **kamera** muncul pada dokumen 1 dan 5 dan didapatkan nilai pembobotan untuk DF adalah 2.

Tabel 3. 9 Tabel Informasi DF

Informasi Term Frequency								
no	term	D1	D2	D3	D4	D5	Dx	DF
1	kamera	1	0	0	0	1	0	2

Sumber: Perancangan

Selanjutnya dihitung nilai IDF (*Inverse Document Frequency*) untuk *term* **kamera** dengan cara:

$$\text{Log} \frac{(\text{jumlah dokumen})}{(\text{nilai DF } n)}$$

$$\text{Log} \frac{5}{2} = 0,398$$

Tabel 3. 10 Tabel Informasi IDF

informasi Term Frequency									
no	term	D1	D2	D3	D4	D5	Dx	DF	IDF
1	kamera	1	0	0	0	1	0	2	0,398

Sumber: Perancangan

Setelah didapatkan nilai dari TF-IDF, dilakukan proses perhitungan *Cosine Similarity* dengan menggunakan rumus pada persamaan 2.4 maka didapatkan hasil:

Tabel 3. 11 Tabel Hasil Perhitungan *Cossine Similarity*

	D1	D2	D3	D4	D5
x	0,328	0,000	0,000	0,000	0,000

Sumber: Perancangan

Lalu hasil dari perhitungan menggunakan *Cossim* akan diurutkan dengan hasil:

Tabel 3. 12 Tabel Hasil Pengurutan *Cossine Similarity*

	sim(x,di)	kategori
D1	0,328	positif
D2	0,000	positif
D3	0,000	negatif
D4	0,000	negatif
D5	0,000	Positif

Sumber: Perancangan

3.6 Klasifikasi

Nilai *K-Values* permisalan yang dimasukkan *user* ke sistem adalah 2 ($K=2$), maka dokumen uji x masuk pada sentimen yang sama dengan dokumen 1 (D1) dan dokumen 2 (D2) sesuai dengan tabel perhitungan *cossine similarity* yang mempunyai nilai tertinggi. Untuk perhitungannya dengan menggunakan rumus dari persamaan 2.2.



Maka hasil akhir dari sistem setelah dari proses klasifikasi adalah:

Tabel 3. 13 Tabel Hasil Akhir Sistem

ID	Review	Sentimen
1	Kamera itu bagus, saya suka sekali	positif
2	D7000 pertama memegang...wah.....bingung...karna selama ini makenya d5000...setelah seharian utak atik....layaknya seorang profesional....apalagi fokusnya yg 39 itu loh...sangat mantap...salut buat d7000	Positif
3	Saya terus terang tidak suka d7000 karena kurang mantap rasanya	negatif
4	Jelek banget bro, buat diluar ga jelas	negatif
5	Kamera ini berharga 5.400.000	positif
x	Ga nyangka bagus banget, ga sabar beli barangnya	positif

Sumber: Perancangan

3.7 Implementasi

Implementasi aplikasi *sentiment analysis* dilakukan dengan mengacu pada perancangan sistem. Implementasi perangkat lunak dilakukan menggunakan bahasa pemrograman PHP dan database MySQL. Implementasi aplikasi ini meliputi:

1. Pembuatan *user interface* berupa halaman-halaman web yang menerima inputan user.
2. Melakukan proses *preprocessing* pada teks yang di masukkan oleh pengguna.
3. Melakukan proses pembobotan setiap kalimat dengan algoritma *TF-IDF*.
4. Penghitungan nilai *Similarity* dengan menggunakan algoritma *Cosine Similarity*.
5. Penghitungan pengklasifikasian Sentimen menggunakan algoritma *K-Nearest Neighbor*.

6. Menghasilkan kelompok sentimen dari teks yang telah diinputkan sebelumnya.

3.8 Pengujian

Pengujian perangkat lunak pada skripsi ini dilakukan agar dapat menunjukkan bahwa perangkat lunak telah mampu bekerja sesuai dengan spesifikasi dari kebutuhan yang melandasinya. Pengujian yang dilakukan meliputi:

1. Pengukuran tingkat akurasi program menggunakan jumlah dokumen data benar dibagi jumlah dokumen data keseluruhan.
2. Pengujian hasil program dengan cara membandingkan output program dengan output manual berdasarkan tingkat proporsi ringkasan yang di hasilkan.
3. Pengujian verifikasi metode *K-Nearest Neighbor* dengan melihat kesesuaian antara hasil implementasi dan perancangan.

Tabel 3. 14 Contoh Tabel Pengujian

No	Nilai K	Akurasi

Sumber: perancangan

3.8.1. Skenario Pengujian

Skenario pengujian dibuat untuk memenuhi perancangan pengujian dan hal-hal apa saja yang perlu direncanakan untuk pengujian. Skenario pengujian yang dilakukan meliputi:

1. Pengukuran jumlah data latih (*training*) dan data uji (*testing*).
2. Pengukuran nilai K yang tepat untuk meningkatkan nilai akurasi.

Tabel 3. 15 Contoh Tabel Skenario Pengujian

Nama Percobaan	Jumlah Data Latih	Jumlah Data Uji

Sumber: perancangan

3.9 Penulisan Laporan

Laporan skripsi ditulis setelah semua proses pengerjaan tugas akhir dilalui. Laporan berisi dokumentasi perancangan aplikasi yang akan berguna untuk pengembangan aplikasi selanjutnya.

