

**ANALISIS SENTIMEN ULASAN PRODUK *SKINCARE*
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*
(STUDI KASUS: FORUM FEMALE DAILY)**



Disusun Oleh:

Nama : Willy Wildan Kamal

NIM : 17523162

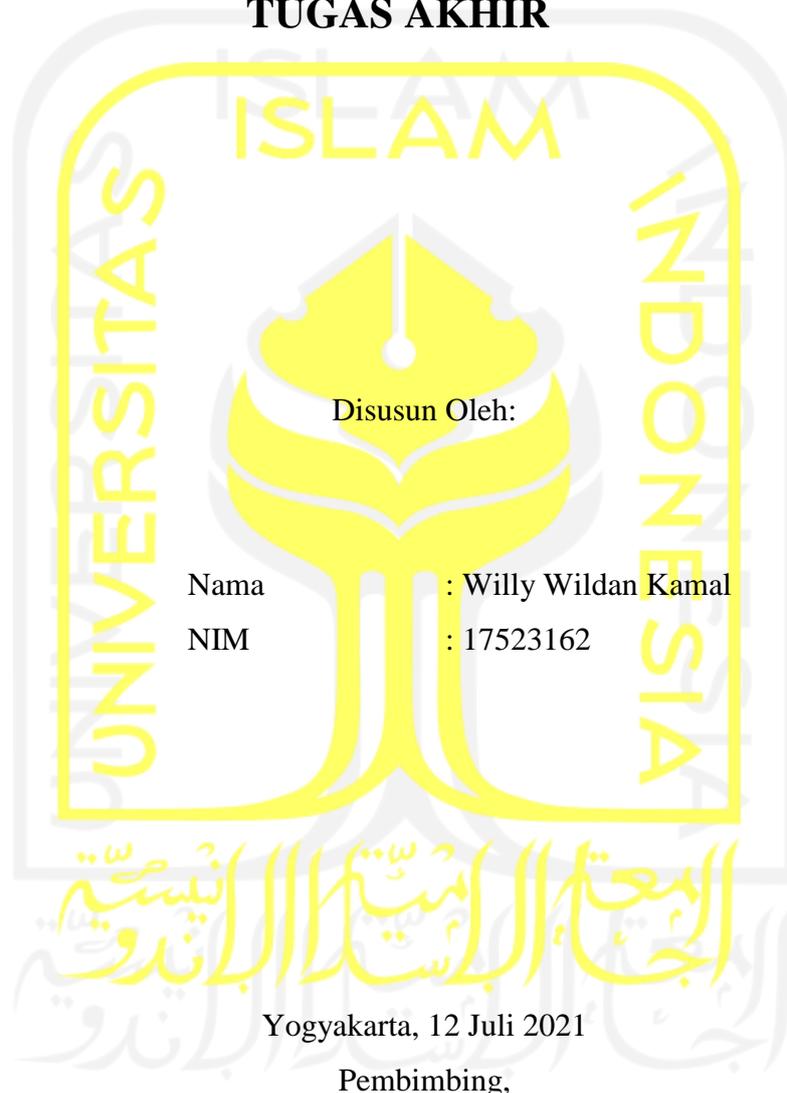
**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2021

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**ANALISIS SENTIMEN ULASAN PRODUK *SKINCARE*
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*
(STUDI KASUS: FORUM FEMALE DAILY)**

TUGAS AKHIR



(Chanifah Indah Ratnasari, S.Kom., M.Kom.)

HALAMAN PENGESAHAN DOSEN PENGUJI

**ANALISIS SENTIMEN ULASAN PRODUK *SKINCARE*
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*
(STUDI KASUS: FORUM FEMALE DAILY)**

TUGAS AKHIR

Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Teknik Informatika di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 6 Agustus 2021

Tim Penguji

Chanifah Indah Ratnasari, S.Kom.,
M.Kom.

Anggota 1

Izzati Muhimmah, S.T., M.Sc. Ph.D.

Anggota 2

Sheila Nurul Huda, S.Kom., MCs.

Mengetahui,

Ketua Program Studi Teknik Informatika – Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia



(Dr. Raden Teduh Dirgahayu, S.T., M.Sc.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Willy Wildan Kamal

NIM : 17523162

Tugas akhir dengan judul:

**ANALISIS SENTIMEN ULASAN PRODUK *SKINCARE*
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*
(STUDI KASUS: FORUM FEMALE DAILY)**

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila dikemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung resiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 12 Juli 2021



(Willy Wildan Kamal)

HALAMAN PERSEMBAHAN

Keluarga

Mamah, papah serta kakak-kakak saya yang selalu memberikan semangat, doa, dan dukungan. Dukungan baik berupa materi maupun non materi.

Sahabat

Yang saling bertukar cerita beratnya mengerjakan ini, bertukar cerita malasnya mengerjakan ini, bertukar cerita pusingnya mengerjakan ini, bertukar cerita capenya mengerjakan ini. Sehingga aku tidak merasa sendirian dan akhirnya kita saling memberi dukungan, doa, serta saling menguatkan. Selamat/Sukses yang terbaik untuk orang yang baik terimakasih.

Musisi

Yang telah menciptakan lagu-lagu untuk menemani mengerjakan tugas akhir ini.

Diri sendiri

Yang telah mau berjuang dengan segala keterbatasan fasilitas, wawasan dan pengetahuan untuk menyelesaikan tugas akhir, kamu hebat wil.

HALAMAN MOTO

“Sesungguhnya sesudah kesulitan itu ada kemudahan” Q.S Al insyirah

Tak ada yang tahu kapan kau mencapai tuju dan percayalah bukan urusanmu untuk menjawab itu katakan pada dirimu besok mungkin kita sampai, besok mungkin tercapai - hindia.



KATA PENGANTAR

Assalamu alaikum wa rahmatullahi wa barakaatuh,

Alhamdulillah, puji serta syukur kehadiran Allah SWT yang telah melimpahkan nikmat, rahmat, dan hidayah-Nya, sehingga Tugas Akhir dengan judul “Analisis Sentimen Ulasan Produk *Skincare* Menggunakan Metode *Support Vector Machine* (Studi Kasus: Forum Female Daily)” ini dapat diselesaikan. Selawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW beserta keluarga, sahabat serta para pengikut beliau hingga akhir zaman.

Tugas Akhir merupakan salah satu syarat yang harus dipenuhi dalam menyelesaikan jenjang strata satu di Jurusan Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia. Selama penyelesaian tugas akhir ini, penulis menyadari banyak pihak yang telah memberikan dukungan, arahan, bimbingan, dan bantuan baik materi maupun non materi, oleh karena itu penulis ingin mengucapkan ucapan terima kasih kepada:

1. Hari Purnomo, Prof., Dr., Ir., M.T. Selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia, beserta seluruh jajarannya.
2. Hendrik, ST., M.Eng. Selaku ketua Jurusan Informatika beserta seluruh jajarannya.
3. Dr. Raden Teduh Dirgahayu, S.T., M.Sc. Selaku ketua Program Studi Informatika beserta seluruh jajarannya.
4. Chanifah Indah Ratnasari, S.Kom., M.Kom. selaku dosen pembimbing yang selalu sabar dan sangat berjasa dalam penyelesaian tugas akhir ini hingga terselesaikan.
5. Seluruh dosen pengajar dan staf prodi Informatika yang telah memberikan bekal ilmu dan bantuannya dalam proses belajar, semoga menjadi amal kebaikan Bapak/Ibu.
6. Semua pihak yang terlibat dari awal hingga akhir dalam pembuatan tugas akhir.

Penulis menyadari bahwa dengan keterbatasan wawasan serta pengalaman penulis, membuat Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun. Semoga Tugas Akhir ini dapat memberikan manfaat. Akhir kata, semoga kita semua mendapatkan rahmat dan selalu ada dalam lindungan Allah SWT.

Wassalamu alaikum wa rahmatullahi wa barakaatuh.

Yogyakarta, 12 Juli 2021



(Willy Wildan Kamal)

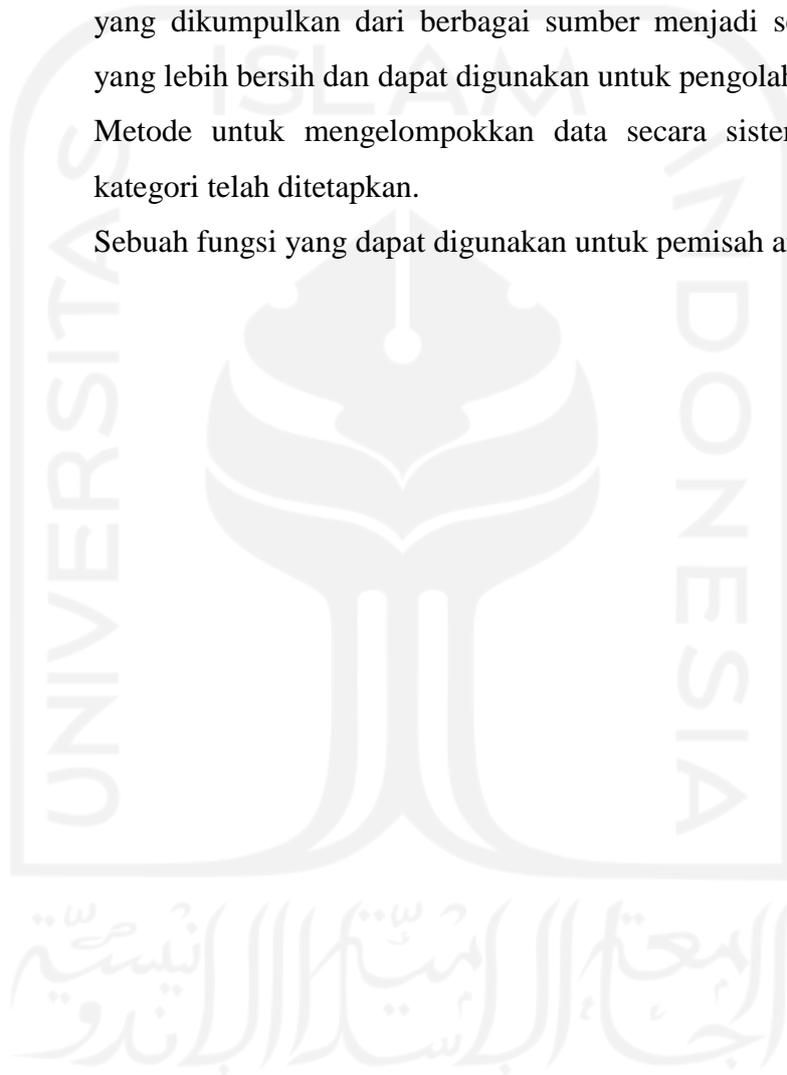
SARI

Seiring pertumbuhan *e-commerce*, peranan media sosial saat ini telah bertransformasi menjadi *social commerce*. Pertumbuhan ini secara tidak langsung mengubah cara berinteraksi *word of mouth marketing* (WOMM) dari tradisional menjadi modern. Media sosial dapat memberikan ruang bagi konsumen untuk saling memberikan ulasan dan rekomendasi. Female Daily merupakan salah satu forum *online* kecantikan terbesar di Indonesia yang menyediakan interaksi sosial untuk saling memberikan ulasan dan rekomendasi. Mengevaluasi ulasan produk dengan jumlah yang banyak membutuhkan analisis sentimen untuk mengelompokkan ulasan konsumen menjadi opini positif, negatif, atau netral. Data ulasan akan dilabeli dan dianalisis dengan menggunakan metode *Support Vector Machine* (SVM) untuk mengklasifikasikan data ulasan. Sentimen tersebut dapat membantu baik perusahaan maupun individu untuk mengetahui kualitas produk secara detail. Dataset terdiri dari 1260 ulasan terbagi menjadi 628 ulasan positif, 174 ulasan netral, dan 458 ulasan negatif. Klasifikasi dengan metode SVM dengan kernel *linear* diperoleh tingkat akurasi sebesar 86.9 %.

Kata kunci: *Support Vector Machine* (SVM), klasifikasi, analisis sentimen, Female Daily.

GLOSARIUM

<i>Web Scraping</i>	Metode pengambilan data tertentu secara semi-terstruktur pada sebuah halaman web.
<i>Parsing HTML</i>	Sebuah alat yang digunakan untuk mengubah kode-kode HTML tertentu menjadi suatu kode dengan identitas unik.
<i>Preprocessing</i>	Teknik awal sebelum pemrosesan data untuk mengubah data mentah yang dikumpulkan dari berbagai sumber menjadi sebuah informasi yang lebih bersih dan dapat digunakan untuk pengolahan selanjutnya.
Klasifikasi	Metode untuk mengelompokkan data secara sistematis ke dalam kategori telah ditetapkan.
<i>Hyperplane</i>	Sebuah fungsi yang dapat digunakan untuk pemisah antar kelas.



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI	viii
GLOSARIUM	ix
DAFTAR ISI	x
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Lingkup Penelitian	2
1.4 Tujuan Penelitian	2
BAB II TINJAUAN PUSTAKA	3
2.1 Female Daily	3
2.2 <i>Support Vector Machine</i>	3
2.2.1 SVM Pada Data Tidak Terpisah Secara Linear.....	6
2.3 Penelitian Terdahulu	9
BAB III METODOLOGI PENELITIAN	11
3.1 Tahapan Penelitian	11
3.2 Pengumpulan Data	11
3.3 <i>Preprocessing</i>	19
3.4 Pembobotan TF-IDF	22
3.5 Pengujian Klasifikasi	24
3.6 Penyimpanan Model	26
3.7 Perancangan Perangkat Lunak	26
3.7.1 Diagram alir.....	26
3.7.2 Perancangan <i>Interface</i>	27
BAB IV IMPLEMENTASI DAN PENGUJIAN	29
4.1 Pengumpulan Data	29
4.2 Implementasi <i>Preprocessing</i>	30
4.3 Data Latih dan Data Uji	39
4.4 Klasifikasi <i>Support Vector Machine</i>	40
4.5 Penyimpanan Model	42
4.6 Pengujian Analisis Sentimen Pada Antarmuka Sistem.....	42
4.7 Kekurangan Sistem	43
BAB V KESIMPULAN DAN SARAN	44
5.1 Kesimpulan	44
5.2 Saran.....	44
DAFTAR PUSTAKA	45
LAMPIRAN	47

DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian terdahulu.....	10
Tabel 3.1 Contoh perhitungan TF-IDF	24
Tabel 4.1 Contoh hasil ulasan setelah dilakukan <i>regular expression</i>	31
Tabel 4.2 Contoh hasil ulasan setelah dilakukan <i>case folding</i>	32
Tabel 4.3 Contoh hasil ulasan setelah dilakukan <i>tokenizing</i> pertama.....	32
Tabel 4.4 Contoh kata <i>slang word</i> dan perbaikannya	34
Tabel 4.5 Contoh singkatan dan perbaikannya	34
Tabel 4.6 Contoh kata yang salah dalam penulisan (<i>typo</i>) dan perbaikannya	34
Tabel 4.7 Contoh kata asing dan perbaikannya	34
Tabel 4.8 Contoh kata dengan karakter berulang	35
Tabel 4.9 Contoh hasil ulasan setelah dilakukan <i>formalization</i>	35
Tabel 4.10 Contoh hasil ulasan setelah dilakukan <i>tokenizing</i> kedua.....	36
Tabel 4.11 Kata yang dihilangkan dari daftar <i>stopword</i>	37
Tabel 4.12 Contoh kata tambahan untuk daftar <i>stopword</i>	37
Tabel 4.14 Contoh hasil ulasan setelah dilakukan <i>filtering</i>	38
Tabel 4.15 Contoh hasil ulasan setelah dilakukan <i>stemming</i>	39
Tabel 4.16 Perbandingan data latih dan data uji	40
Tabel 4.17 Perbandingan penggunaan metode kernel pada klasifikasi SVM.....	40
Tabel 4.18 <i>confusion matrix</i>	41

DAFTAR GAMBAR

Gambar 2.1 Dua kelas yang dipisahkan oleh <i>hyperlane</i>	3
Gambar 2.2 Transformasi data pada ruang fitur	7
Gambar 3.1 Tahapan penelitian	11
Gambar 3.2 grafik perbandingan ulasan pada produk aztec mask.....	12
Gambar 3.3 grafik perbandingan ulasan pada produk estee lauder	13
Gambar 3.4 grafik perbandingan ulasan pada produk pond's flawless white	13
Gambar 3.5 grafik perbandingan ulasan pada produk garnier sakura white	14
Gambar 3.6 grafik perbandingan ulasan pada produk innisfree	14
Gambar 3.7 grafik perbandingan ulasan pada produk kyusoku bihaku.....	15
Gambar 3.8 grafik perbandingan ulasan pada produk l'oreal micellar water	15
Gambar 3.9 grafik perbandingan ulasan pada produk nivea creme.....	16
Gambar 3.10 grafik perbandingan ulasan pada produk oriflame.....	16
Gambar 3.11 grafik perbandingan ulasan pada produk sari ayu by marta tilaar	17
Gambar 3.12 grafik perbandingan ulasan pada produk bedak saripohatji.....	17
Gambar 3.13 grafik perbandingan ulasan pada produk serum atau essence.....	18
Gambar 3.14 grafik perbandingan ulasan pada produk the body shop.....	18
Gambar 3.15 grafik perbandingan ulasan pada produk viva <i>skincare</i>	19
Gambar 3.16 grafik perbandingan ulasan pada produk ponds white beauty	19
Gambar 3.17 Tahapan penelitian untuk <i>preprocessing</i>	20
Gambar 3.18 Diagram alir analisis sentimen.....	27
Gambar 3.19 Rancangan tampilan <i>interface</i>	28
Gambar 4.1 Perintah instalasi <i>library beautifulsoup4</i>	29
Gambar 4.2 Inisiasi <i>scraping</i> pada halaman yang dituju.....	29
Gambar 4.3 Inisiasi <i>library</i> Python <i>beautifulsoup4</i> dan penyimpanan <i>file</i>	30
Gambar 4.4 Contoh teks hasil implementasi <i>scraping</i>	30
Gambar 4.5 Contoh kode program <i>regular expression</i> dan <i>case folding</i>	31
Gambar 4.6 Kode program <i>formalization</i> dan <i>tokenizing</i> pertama.....	33
Gambar 4.11 Perintah instalasi <i>library sastrawi</i>	37
Gambar 4.12 Contoh kode program <i>filtering</i>	38
Gambar 4.13 Kode program <i>stemming</i>	39
Gambar 4.14 Kode program penyimpanan model.....	42
Gambar 4.15 Uji coba sampel negatif teks pada tampilan sistem	42

Gambar 4.16 Uji coba sampel netral teks pada tampilan sistem43

Gambar 4.17 Uji coba sampel positif teks pada tampilan sistem43



BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini, peranan *e-commerce* mengubah peran media sosial menjadi *social commerce* (Hajli, 2014). Interaksi sosial antar konsumen melalui media sosial seperti forum *online* memberikan ruang bagi konsumen untuk *sharing of user-generated review* dan memberikan rekomendasi (Casaló, Flavián, & Guinalú, 2011). Female Daily merupakan forum *online* kecantikan terbesar di Indonesia dengan kapasitas *member* lebih dari 250 ribu, *page view* lebih dari 7,8 juta, dan pengunjung per bulan berjumlah 1,5 juta (Wardani, 2017). Forum ini merupakan wadah bertukar informasi seputar produk kecantikan termasuk *skincare*. Konsumen dapat menuliskan ulasan produk tanpa adanya batasan karakter dengan tujuan untuk mengetahui kualitas dari produk secara detail.

Semakin banyaknya produk *skincare* di pasaran seharusnya semakin meningkatkan selektivitas konsumen dalam memilih suatu produk, Hal ini dikarenakan tidak semua produk memiliki kualitas yang bagus dan setiap orang memiliki tipe kulit yang berbeda. Bahkan tipe kulit yang sama pun belum tentu cocok dengan produk *skincare* yang sama. Banyak cara untuk mendapatkan produk *skincare* dengan kualitas yang bagus, salah satunya dengan cara melihat ulasan produk. Hampir 50% dari pengguna internet bergantung pada rekomendasi *word-of-mouth* (opini) sebelum membeli atau menggunakan suatu produk, termasuk opini-opini yang ada dalam sebuah forum (Fanani, 2017). Berdasarkan hasil survei Desember 2019 yang dilakukan terhadap 1.005 pengguna internet AS, sebanyak 76% orang mempercayai *online review* sebagaimana rekomendasi dari keluarga dan teman (Murphy, 2019).

Membaca ulasan produk secara keseluruhan dapat memakan waktu, namun jika hanya sedikit ulasan yang dibaca evaluasi akan menjadi bias (Utami, 2015). Untuk mengatasi masalah tersebut, maka analisis sentimen penting dilakukan untuk mengklasifikasikan ulasan menjadi opini positif, netral, dan negatif. Pengklasifikasian ulasan membantu konsumen untuk mengetahui kualitas produk dan mencari produk yang sesuai dengan tipe kulitnya. Ada banyak metode yang dapat digunakan untuk analisis sentimen, salah satunya *Support Vector Machine*. Metode ini dipilih karena memiliki nilai akurasi yang tinggi di antara metode yang lainnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dalam penelitian ini permasalahan yang akan dirumuskan adalah:

1. Berapa persentase ketepatan analisis sentimen menggunakan metode *Support Vector Machine* dalam mengklasifikasi domain perawatan kulit berdasarkan forum Female Daily?
2. Bagaimana cara melakukan analisis sentimen pada ulasan produk *skincare* forum Female Daily?

1.3 Batasan Lingkup Penelitian

Batasan lingkup penelitian yang ditentukan untuk menghindari perluasan pembahasan sehingga tujuan dari penelitian ini tercapai. Dalam analisis sentimen ini hanya dibatasi pada:

1. Data yang akan diklasifikasi hanya data ulasan pada forum online Female Daily.
2. Metode yang digunakan untuk analisis sentimen adalah metode *Support Vector Machine*.
3. Ulasan yang dianalisis dan diklasifikasi merupakan produk kecantikan jenis *skincare*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Dapat digunakan untuk membantu calon konsumen mengetahui kualitas suatu produk *skincare*.
2. Dapat digunakan oleh pemilik produk kecantikan untuk mengetahui pandangan pelanggan terhadap kualitas produk kecantikan mereka.
3. Mengetahui seberapa tepat analisis sentimen dengan menggunakan metode *Support Vector Machine* dari data ulasan *skincare*.
4. Mengetahui cara melakukan analisis sentimen pada ulasan produk *skincare* forum Female Daily.

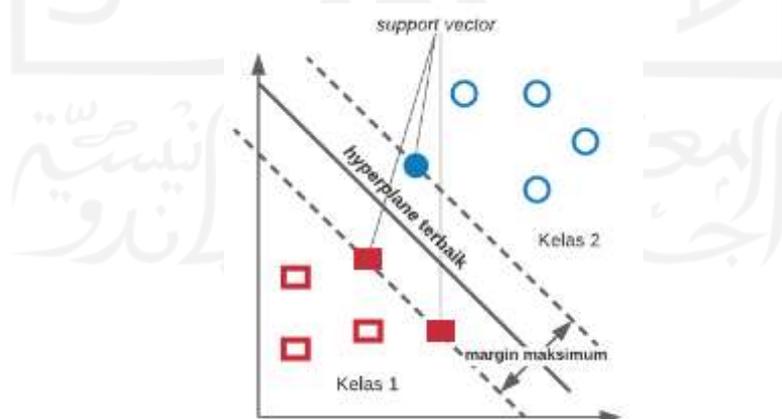
BAB II TINJAUAN PUSTAKA

2.1 Female Daily

Female Daily merupakan *startup* kecantikan yang berangkat dari sebuah blog pribadi dengan konten *fashion* dan kecantikan. Female Daily kini berada di bawah badan hukum PT Daily Dinamika Kreasi dengan memiliki 7,8 juta *page view* per bulan dan 18 ribu forum *thread* (Wardani, 2017). Layanan Female Daily dapat diakses melalui aplikasi android, *ios*, dan web. Female Daily memiliki fitur ulasan produk, *workshop* kecantikan, tutorial kecantikan, opini pakar, opini konsumen, dan forum.

2.2 Support Vector Machine

SVM merupakan salah satu metode *supervised learning* yang memiliki kemampuan untuk generalisasi dalam mengklasifikasi suatu pola dan memiliki kelebihan untuk mengidentifikasi *hyperlane* terpisah yang memaksimalkan margin antar kelas (Kristiyanti, 2015). SVM bekerja pada prinsip *Structural Risk Minimization* (SRM) dan memiliki konsep landasan teori yang lebih matang, sehingga dapat dianalisis dengan jelas secara matematis dibandingkan dengan klasifikasi lainnya. Namun metode ini memiliki kekurangan, yaitu sensitif terhadap pemilihan fitur. Selain itu, penggunaan parameter akan mempengaruhi hasil akurasi klasifikasi. Inti dari proses pelatihan *Support Vector Machine* (SVM) adalah usaha untuk mencari lokasi *hyperlane* yang optimal.



Gambar 2.1 Dua kelas yang dipisahkan oleh *hyperlane*

Hyperlane memiliki fungsi untuk memisahkan antara kedua kelas pada *input space*. Divisualisasikan garis solid seperti pada Gambar 3.3 merupakan *hyperlane* terbaik. Posisinya

tepat berada di antara kedua kelas. *Hyperlane* terbaik adalah bidang yang memisahkan data dan memiliki margin yang besar. Margin merupakan jarak antara *hyperlane* dengan data dari masing-masing kelas. *Support vector* atau *subset data training set* merupakan data terluar yang berada paling dekat dengan bidang pembatas (garis putus-putus). Data dinotasikan sebagai $\{x_i, \dots, x_n\}$, sedangkan label masing-masing kelas dinotasikan $y_i \in \{+1, -1\}$ untuk $i = 1, 2, 3, \dots, n$. Pada visualisasi bidang pembatas pertama membatasi kelas pertama (persamaan (1)) sedangkan pembatas kedua membatasi kelas kedua (persamaan (2)) sehingga diperoleh pertidaksamaan:

$$x_i \cdot w + b \geq +1, \text{ untuk } y_i = +1 \quad (1)$$

$$x_i \cdot w + b \leq -1, \text{ untuk } y_i = -1 \quad (2)$$

Diketahui w merupakan bidang normal dan b merupakan posisi bidang relatif terhadap pusat koordinat. Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara jarak dan titik terdekatnya, yaitu $\frac{1}{|w|}$ sama dengan meminimalkan $|w|^2$ dan jika kedua bidang pembatas direpresentasikan dalam pertidaksamaan (3).

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad (3)$$

Sehingga dapat dirumuskan sebagai *Quadratic Programming (QP) problem*, yaitu mencari titik minimum persamaan (4).

$$\min \frac{1}{2} |w|^2 \quad (4)$$

$$\text{s.t. } y_i(x_i \cdot w + b) - 1 \geq 0$$

Masalah ini dapat diselesaikan lebih mudah dengan teknik komputasi, di antaranya dengan formula *lagrangian multiplier* sehingga dinyatakan menjadi:

$$\min_{w,b} L_p(w, b, \alpha) = \frac{1}{2} |w|^2 - \sum_{i=1}^n \alpha_i y_i ((x_i \cdot w + b) - 1) + \sum_{i=1}^n \alpha_i \quad (5)$$

$\alpha_i \geq 0$ merupakan nilai *lagrange multiplier*. Kemudian dengan meminimalkan L_p terhadap w dan b , maka $\frac{\partial}{\partial b} L_p(w, b, \alpha) = 0$ diperoleh (6) dan dari $\frac{\partial}{\partial w} L_p(w, b, \alpha) = 0$ diperoleh (7)

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (6)$$

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (7)$$

Vektor w bernilai besar hingga tak terhingga, tetapi nilai α_i terhingga, sehingga formula *lagrangian* L_p (*primal problem*) diubah kedalam *dual problem* L_D . Dengan substitusi persamaan (7) ke L_p diperoleh L_D dengan konstanta yang berbeda.

$$L_D(\alpha) \equiv \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \quad (8)$$

Jadi persoalan untuk pencarian pemisahan bidang terbaik dirumuskan sebagai berikut:

$$\max_{\alpha} L_D \equiv \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \quad (9)$$

$$s. t \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0 \quad (10)$$

Nilai α_i digunakan untuk menemukan w , setiap data latih memiliki nilai α_i . Data latih yang mempunyai nilai $\alpha_i > 0$ merupakan *support vector* sedangkan sisanya mempunyai nilai $\alpha_i = 0$, sehingga fungsi keputusan yang dihasilkan hanya dipengaruhi oleh *support vector*.

Rumus untuk pencarian *hyperlane* terbaik adalah permasalahan *quadratic programming* sehingga nilai maksimum dari α_i selalu bisa ditemukan, maka kelas pengujian x dapat ditentukan berdasarkan nilai fungsi keputusan:

$$f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i x_i x_d + b \quad (11)$$

Diketahui:

α_i = Koefisien *lagrange*

x_i = *Support vector*

y_i = Kelas data

ns = Jumlah *support vector*

x_d = Data yang diklasifikasikan

2.2.1 SVM Pada Data Tidak Terpisah Secara Linear

Rumus SVM harus dimodifikasi agar dapat ditemukan solusi pada kasus klasifikasi data yang tidak dapat dipisahkan secara linier. Sehingga kedua bidang pembatas tersebut harus diubah agar lebih fleksibel dengan penambahan variabel ξ_i ($\xi_i \geq 0, \forall_i: \xi_i = 0$ jika x_i diklasifikasikan dengan benar) menjadi pertidaksamaan (12).

$$\begin{aligned} x_i \cdot w + b &\geq 1 - \xi_i, \text{ untuk kelas 1} \\ x_i \cdot w + b &\leq -1 + \xi_i, \text{ untuk kelas 2} \end{aligned} \quad (12)$$

Hyperlane terbaik dicari dengan menambahkan variable ξ_i disebut dengan *softmargin hyperlane*. Sehingga berubah menjadi:

$$\begin{aligned} \min \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } y_i(x_i \cdot w + b) \geq 1 - \xi_i, \xi_i \geq 0 \end{aligned} \quad (13)$$

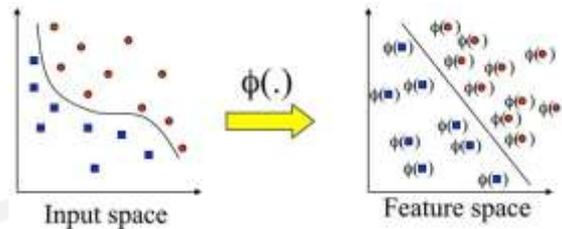
C adalah parameter yang menentukan besarnya hukuman dalam klasifikasi data. Persamaan (13) memenuhi prinsip SRM yaitu meminimalkan $\frac{1}{2} |w|^2$ ekuivalen dengan meminimalkan dimensi VC dan meminimalkan $C(\sum_{i=1}^n \xi_i)$ berarti meminimalkan *error* pada data latih.

Bentuk *primal problem* sebelumnya berubah menjadi:

$$\min L_p(w, b, \alpha) = \frac{1}{2} ||w||^2 - C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} - \sum_{i=1}^n \mu_i \xi_i \quad (14)$$

Perubahan L_p pada *dual problem* menghasilkan rumus yang sama dengan persamaan (7) sehingga pencarian pembatas bidang terbaik nonlinier caranya hampir sama dengan kasus linier tetapi rentang nilai α_i adalah $0 \geq \alpha_i \geq C$. Contoh yang bernilai $\alpha_i = C$ disebut *bounded support vector*. Metode lain untuk mengklasifikasikan masalah data nonlinier adalah dengan cara transformasi data ke dalam dimensi ruang fitur sehingga dapat dipisahkan secara linier pada ruang fitur.

Transformasi data pada ruang fitur dilakukan dengan cara memetakan data menggunakan fungsi pemetaan ke dalam ruang fitur sehingga data dapat terpisahkan oleh *hyperlane* sesuai dengan kelasnya Gambar 2.2.



Gambar 2.2 Transformasi data pada ruang fitur

Dengan fungsi transformasi $x_k \rightarrow \phi(x_k)$, maka diperoleh nilai $w = \sum_{i=1}^{ns} \alpha_i y_i \phi(x_i)$ dan fungsi hasil latih adalah:

$$f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i \phi(x_i) \phi(x_d) + b \quad (15)$$

Ruang fitur berdimensi lebih tinggi dari pada vektor input. Akibatnya komputasi pada ruang fitur sangat besar sehingga memungkinkan ruang fitur memiliki jumlah fitur tak terhingga. Sehingga untuk mengatasi masalah tersebut pada SVM digunakan “*kernel trick*”. Dari persamaan (13) terdapat *dot product* $\phi(x_i) \phi(x_d)$. Jika terdapat fungsi *kernel* K sehingga $K(x_i, x_d) = \phi(x_i) \phi(x_d)$, maka fungsi $\phi(x_k)$ tidak perlu diketahui. Sehingga didapatkan fungsi dari pelatihan:

$$f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i K(x_i, x_d) + b \quad (x_i = \text{support vector}) \quad (16)$$

Diketahui:

α_i = Koefisien *lagrange*

x_i = *Support vector*

y_i = Kelas data

ns = Jumlah *support vector*

x_d = Data yang diklasifikasikan

Beberapa *kernel* yang umum dipakai pada SVM adalah:

a. *Linear Kernel*

Kernel *linear* merupakan fungsi kernel yang paling sederhana. Kernel ini cocok digunakan ketika data yang dianalisis sudah terpisah secara *linear*. Kernel ini sering digunakan untuk menyelesaikan masalah klasifikasi, karena memiliki akurasi pelatihan dan akurasi prediksi paling baik. Kernel *linear* dapat dinyatakan dalam persamaan (17).

$$K(x_i, x) = x_i^T x \quad (17)$$

b. *Polynomial Kernel*

Kernel *polynomial* merupakan fungsi kernel yang digunakan ketika data tidak terpisah secara linear. *Polynomial* cocok untuk permasalahan ketika dataset *training* sudah dinormalisasi. Adapun kernel ini dapat dinyatakan dalam persamaan (18).

$$K(x_i, x) = (\gamma \cdot x_i^T x + r)^p, \gamma > 0 \quad (18)$$

c. *Radial Basis Function (RBF) Kernel*

Kernel *radial basis function* atau *gaussian* merupakan fungsi kernel yang digunakan ketika data tidak terpisah secara linear. Kernel ini memiliki dua parameter yaitu *Cost* dan *Gamma*. Parameter *Cost* disebut sebagai C berfungsi memaksimalkan SVM untuk menghindari kesalahan klasifikasi di setiap sampel dalam dataset *training*. Parameter *Gamma* berfungsi menentukan seberapa jauh pengaruh dari satu sampel dataset *training* dengan nilai jauh dan nilai dekat. Adapun kernel RBF dapat dinyatakan dalam persamaan (19).

$$K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0 \quad (19)$$

d. *Sigmoid Kernel*

Kernel *sigmoid* merupakan fungsi kernel yang digunakan ketika data tidak terpisah secara linear. *Sigmoid* merupakan pengembangan dari jaringan saraf tiruan dan memiliki karakteristik kurva berbentuk “S”. Kernel ini dapat dinyatakan dalam persamaan (20).

$$K(x_i, x) = \tanh(\gamma x_i^T x + r) \quad (20)$$

2.3 Penelitian Terdahulu

Beberapa penelitian tentang analisis sentimen terdahulu sudah dilakukan. Penelitian tersebut sangatlah penting dalam penelitian ini sebagai kajian untuk mengetahui keterkaitan antara penelitian terdahulu dengan penelitian yang akan dilakukan, untuk menghindari terjadinya tindakan duplikasi yang dilakukan oleh penulis. Tujuan dari tinjauan pustaka ini adalah untuk mengkaji metode yang akan dipakai dalam penelitian ini. Dalam penelitian analisis sentimen terdapat 2 metode yang paling populer digunakan yaitu *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM). Masing-masing metode memiliki nilai akurasi yang berbeda. Pada penelitian terdahulu yang penulis kaji, penelitian dengan metode *Naive Bayes Classifier* memiliki nilai akurasi yang terbilang tinggi. Contohnya pada penelitian tentang analisis sentimen pada *review* kosmetik yang dilakukan oleh Indrayuni (2019) memiliki nilai akurasi 90.50%, sedangkan penelitian yang dilakukan oleh Utami (2018) memiliki nilai akurasi 80%. Sementara itu, penelitian lain yang dilakukan oleh Rahayu, Kusri, dan Sismoro (2018) tentang *review* pengguna *marketplace online* memiliki nilai akurasi sebesar 87%.

Selain *Naive Bayes Classifier*, beberapa penelitian juga membahas metode *Support Vector Machine*. Metode ini memiliki nilai akurasi yang tinggi, seperti pada penelitian Yulietha, Faraby, dan Adiwijaya (2017) tentang klasifikasi sentimen *review* film dengan pembobotan TF-IDF memiliki nilai akurasi 84.9%. Penelitian lain yaitu yang dilakukan oleh Fikria (2018) membahas tentang analisis sentimen *review* aplikasi *e-ticketing* pada aplikasi KAI *access* dengan pembobotan TF-IDF yang memiliki nilai 89% dan pada aplikasi *tiket.com* memiliki nilai akurasi 84.68%. Pemilihan fitur dapat digunakan untuk meningkatkan nilai akurasi, seperti pada penelitian Fauzi (2018) tentang analisis sentimen *review* barang berbahasa Indonesia dengan pembobotan *term-binary*, penelitian menggunakan metode *Support Vector Machine* menghasilkan nilai akurasi 94,75%. Nilai akurasi tersebut meningkat menjadi 96,25% ketika dilakukan penggabungan dengan fitur *Query Expansion*. Begitu pula dengan penelitian Kristiyanti (2015) yang membahas tentang *review* produk kosmetik dengan pembobotan TF-IDF, memiliki akurasi 89%. Penggabungan metode pemilihan fitur dengan *Swarm Optimization* mampu meningkatkan nilai akurasi menjadi 97%. Untuk meningkatkan penelitian tersebut, penelitiannya menyarankan untuk menggunakan data *review* dari domain yang berbeda.

Pada penelitian kali ini akan menggunakan metode *Support Vector Machine* dengan menggunakan data *review* dari domain perawatan kulit. Hasil kajian penelitian ini diharapkan dapat menjadi rujukan metode pada analisis sentimen untuk penelitian terkait ulasan produk

skincare. Gambaran mengenai beberapa penelitian analisis sentimen dapat dilihat pada Tabel 2.1 yang membandingkan metode dan nilai akurasi penelitian terdahulu.

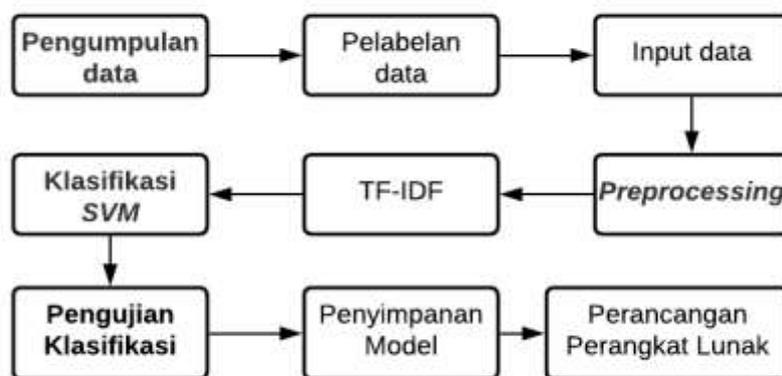
Tabel 2.1 Perbandingan penelitian terdahulu

Peneliti, Tahun	Bahasan	Metode	Nilai Akurasi
Muthia (2017)	Analisis sentimen pada <i>review</i> restoran	NBC	70.43% - 88.03%
Fauzi (2018)	Analisis sentimen <i>review</i> barang pada toko <i>online</i> menggunakan pembobotan <i>term binary</i>	SVM	94,75%
		SVM dengan <i>Query</i>	96,25%
Wati (2016)	Analisis sentimen <i>review</i> jasa maskapai penerbangan	NBC	60.00%
		NBC dengan Algoritma Genetika	89.50%
Gunawan, Fauzi, dan Adikara (2017)	Analisis sentimen pada ulasan aplikasi <i>BCA mobile</i>	NBC	94,4%
		<i>Naive Bayes Classifier</i> dengan <i>Levenshtein Distance</i>	96,9%
Utami (2018)	Analisis sentimen <i>review</i> kosmetik bahasa Indonesia	NBC	80%
Kristiyanti (2015)	Analisis sentimen <i>review</i> produk kosmetik menggunakan pembobotan TF-IDF	SVM	89.00 %
		SVM berbasis <i>Particle Swarm Optimization</i>	97.00 %
Yulietha, Faraby, dan Adiwijaya (2017)	Klasifikasi sentimen <i>review</i> film menggunakan pembobotan TF-IDF	SVM	84.9%
Indrayuni (2019)	Klasifikasi <i>text mining review</i> produk kosmetik menggunakan pemodelan bahasa <i>N-gram</i>	NBC	90.50%
Rahayu, Kusriani, dan Sismoro (2018)	Analisis sentimen <i>review</i> pengguna <i>marketplace</i>	NBC	87%
Fikria (2018)	Analisis sentimen <i>review</i> aplikasi <i>e-ticketing</i> KAI Access dan <i>tiket.com</i> menggunakan pembobotan TF-IDF	SVM (<i>KAI Access</i>)	89,60%
		SVM (<i>tiket.com</i>)	84,68%

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Sistem yang akan dibangun dalam penelitian ini memiliki empat tahapan yaitu tahap pengumpulan data, tahap *preprocessing*, tahap klasifikasi, dan pengujian. Tahapan penelitian yang dibuat dapat dilihat pada Gambar 3.1.



Gambar 3.1 Tahapan penelitian

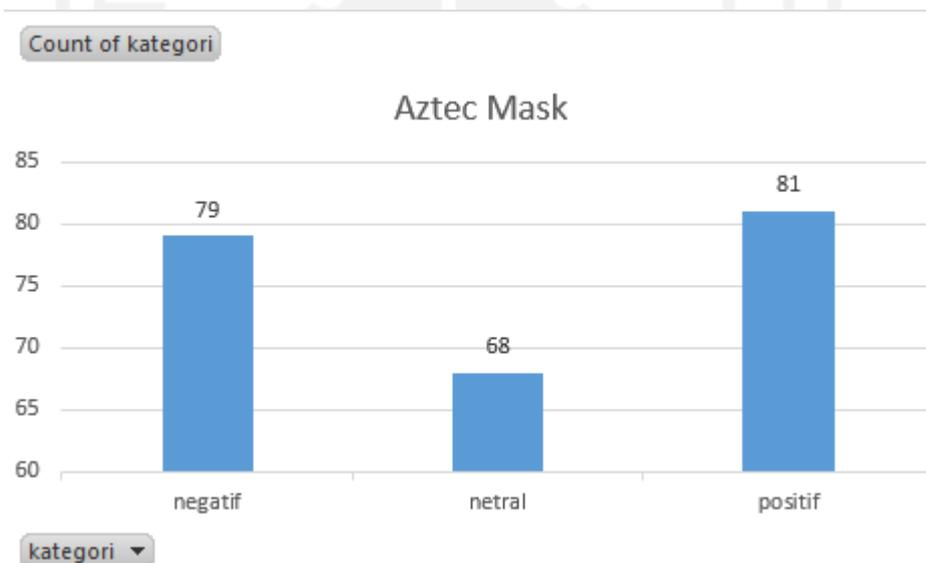
Tahapan penelitian yang pertama dimulai dengan pengumpulan data ulasan dari forum Female Daily yang digunakan sebagai data. Setelah data ulasan terkumpul akan dilakukan pelabelan data yaitu data ulasan akan dikategorikan menjadi positif, netral, dan negatif. Selanjutnya data ulasan yang telah diberikan label akan melalui tahapan *preprocessing* yaitu meliputi *regular expression*, *case folding*, *tokenizing*, *formalization*, *filtering*, dan *stemming*. Setelah data dibersihkan pada tahap *preprocessing*, maka data akan melalui tahap pembobotan TF-IDF yang bertujuan untuk mendapatkan nilai bobot setiap kata pada data latih. Kemudian data akan melalui tahap klasifikasi menggunakan metode *Support Vector Machine*. Model dari hasil klasifikasi tersebut kemudian disimpan untuk dimuat dalam tampilan web. Selanjutnya hasil klasifikasi tahap pengujian untuk melihat performa dari sistem yang dibuat.

3.2 Pengumpulan Data

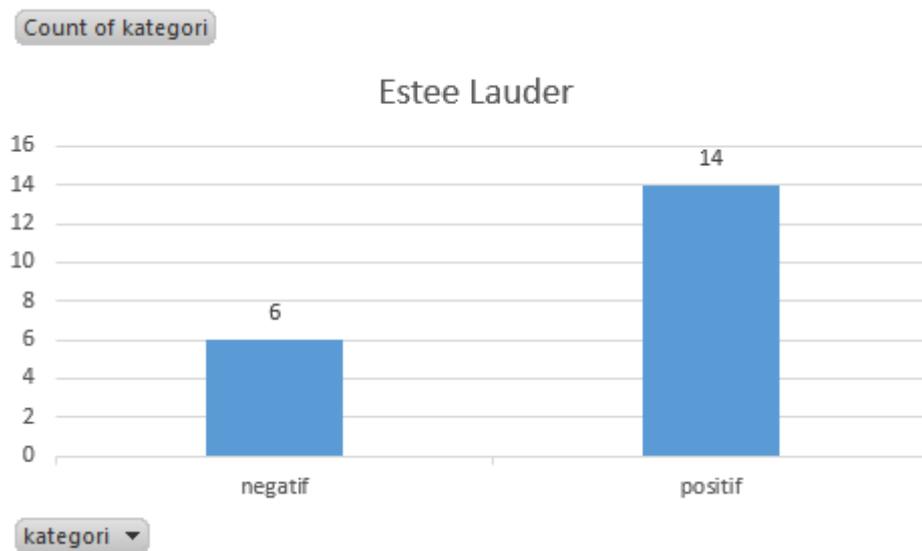
Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari salah satu forum *online* yaitu Female Daily. Pengumpulan data dengan cara mengambil data ulasan dari

forum dengan produk *skincare*. Produk *skincare* merupakan produk perawatan kulit untuk menjaga kecantikan dan kesehatan kulit. Forum Female Daily dipilih dikarenakan mengandung banyak kalimat tidak beraturan dan mudah didapatkan. Data ulasan tersebut diperoleh dengan memanfaatkan ekstraksi data dari forum Female Daily menggunakan metode web *scraping*. Web *scraping* merupakan teknik pengambilan data tertentu secara semi-terstruktur pada sebuah halaman web. Ulasan yang akan dilakukan *scraping* adalah ulasan produk *skincare* yang diambil 15 Desember 2020 sampai dengan 13 Juli 2021.

Data ulasan yang digunakan terdiri dari 15 produk *skincare* yang ada pada forum Female Daily. Berikut merupakan grafik perbandingan ulasan positif, ulasan negatif dan ulasan netral pada setiap produk *skincare*. Pada Gambar 3.2 menunjukkan bahwa pada produk aztec mask terdapat 79 ulasan negatif, 68 ulasan netral, dan 81 ulasan positif. Selanjutnya pada Gambar 3.3 menunjukkan bahwa produk estee lauder terdapat 6 ulasan negatif dan 14 ulasan positif.

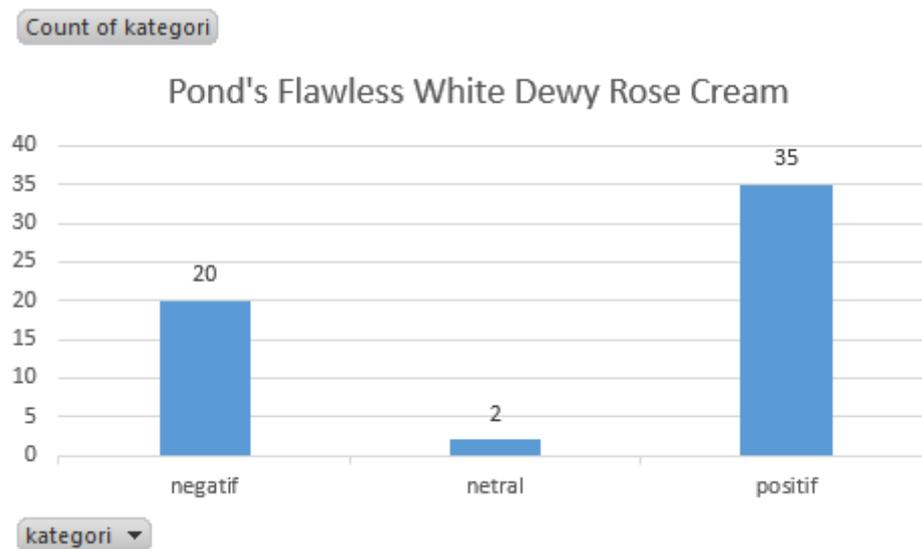


Gambar 3.2 grafik perbandingan ulasan pada produk aztec mask

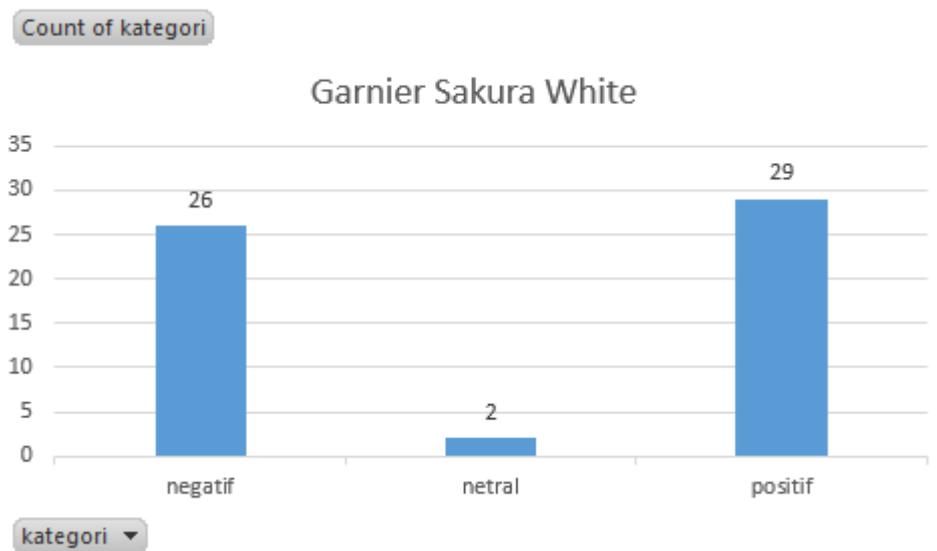


Gambar 3.3 grafik perbandingan ulasan pada produk estee lauder

Gambar 3.4 menunjukkan bahwa pada produk pond's flawless white dewy rose cream terdapat 20 ulasan negatif, 2 ulasan netral, dan 35 ulasan positif. Selanjutnya pada Gambar 3.5 menunjukkan produk garnier sakura white terdapat 26 ulasan negatif, 2 ulasan netral, dan 29 ulasan positif.

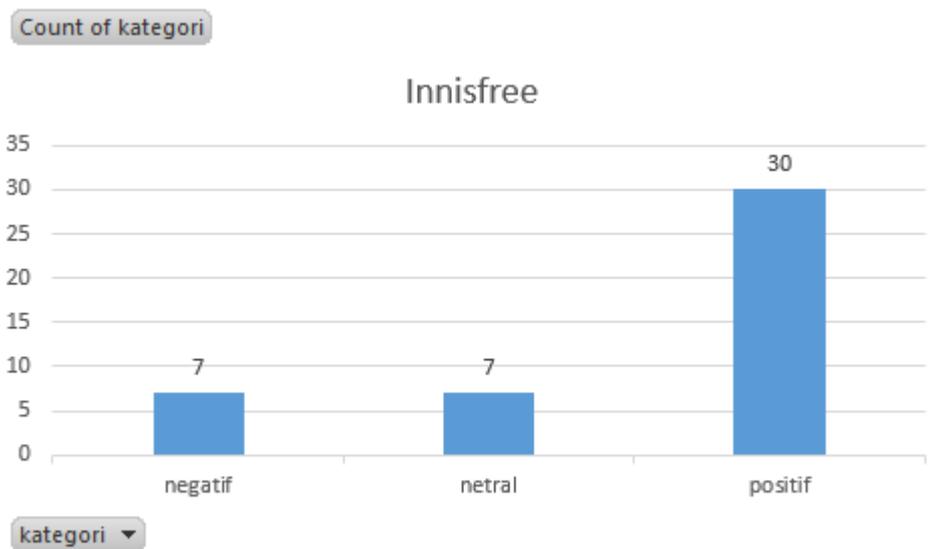


Gambar 3.4 grafik perbandingan ulasan pada produk pond's flawless white

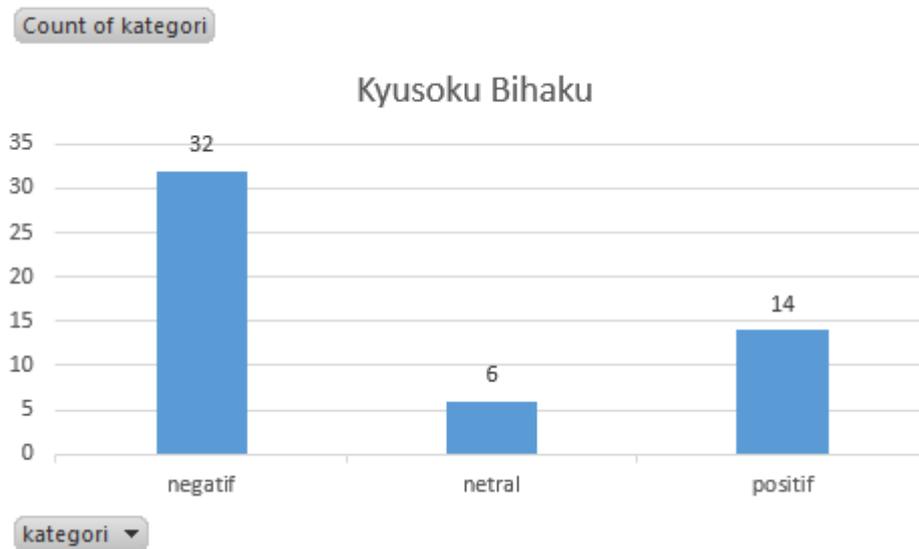


Gambar 3.5 grafik perbandingan ulasan pada produk garnier sakura white

Gambar 3.6 memperlihatkan bahwa pada produk innisfree terdapat 7 ulasan negatif, 7 ulasan netral, dan 30 ulasan positif. Selanjutnya pada Gambar 3.7 produk kyusoku bihaku memiliki 32 ulasan negatif, 6 ulasan netral, dan 14 ulasan positif.

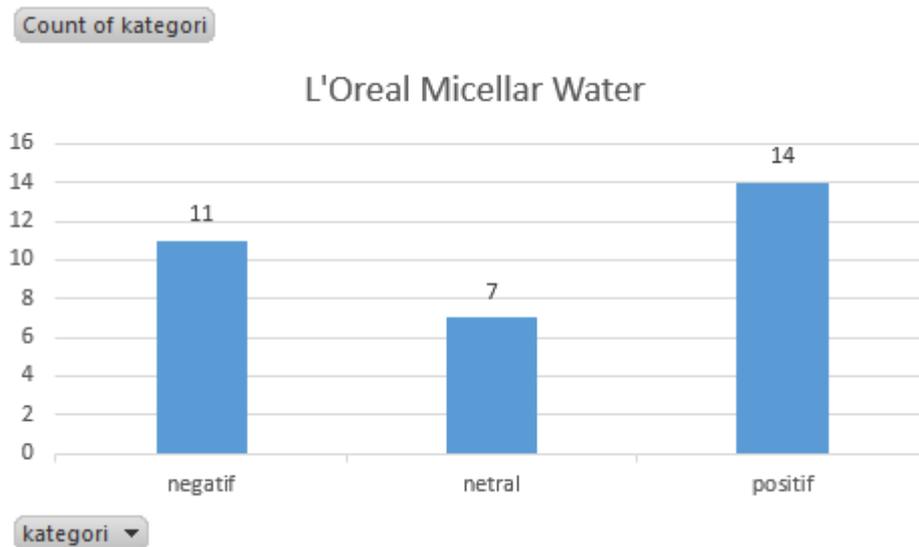


Gambar 3.6 grafik perbandingan ulasan pada produk innisfree

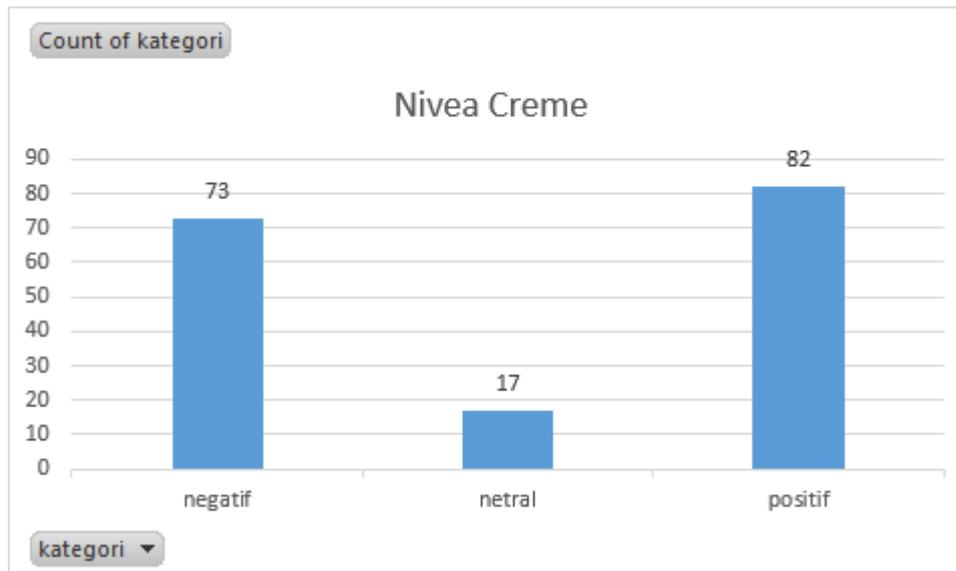


Gambar 3.7 grafik perbandingan ulasan pada produk kjusoku bihaku

Gambar 3.8 memperlihatkan bahwa pada produk l'oreal micellar water memiliki 11 ulasan negatif, 7 ulasan netral, dan 14 ulasan positif. Sedangkan pada Gambar 3.9 memperlihatkan bahwa pada produk nivea creme memiliki 73 ulasan negatif, 17 ulasan netral, dan 82 ulasan positif.

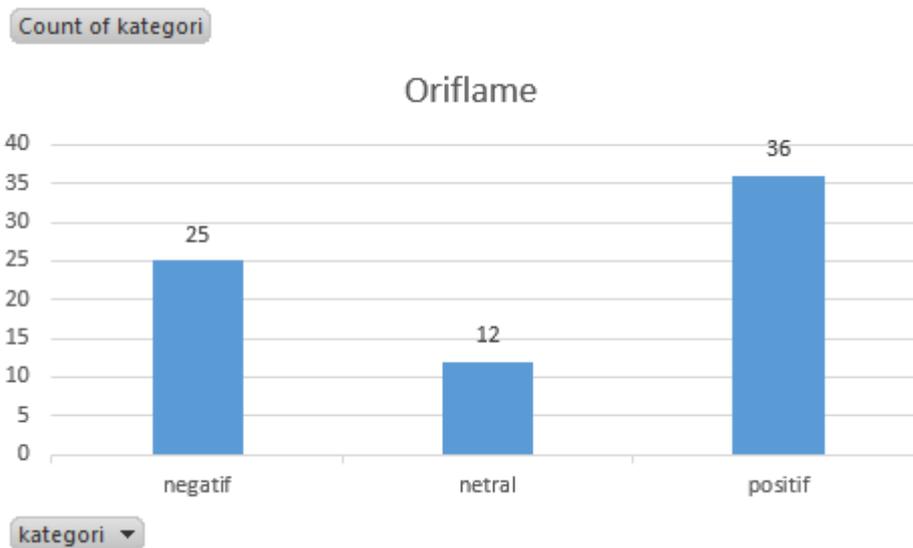


Gambar 3.8 grafik perbandingan ulasan pada produk l'oreal micellar water

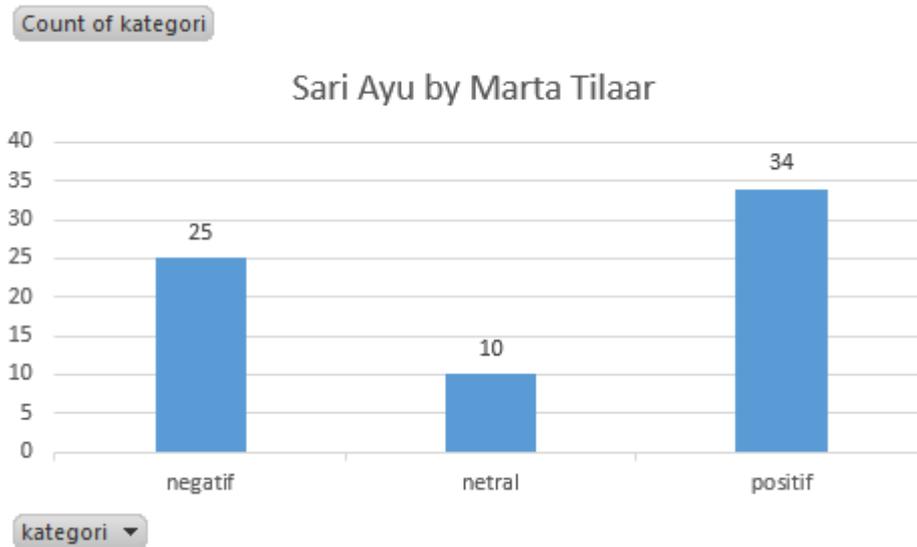


Gambar 3.9 grafik perbandingan ulasan pada produk nivea creme

Gambar 3.10 memperlihatkan bahwa pada produk oriflame memiliki 25 ulasan negatif, 12 ulasan netral, dan 36 ulasan positif. Sedangkan Gambar 3.11 memperlihatkan bahwa pada produk sari ayu by marta tilaar memiliki 25 ulasan negatif, 10 ulasan netral, dan 34 ulasan positif.

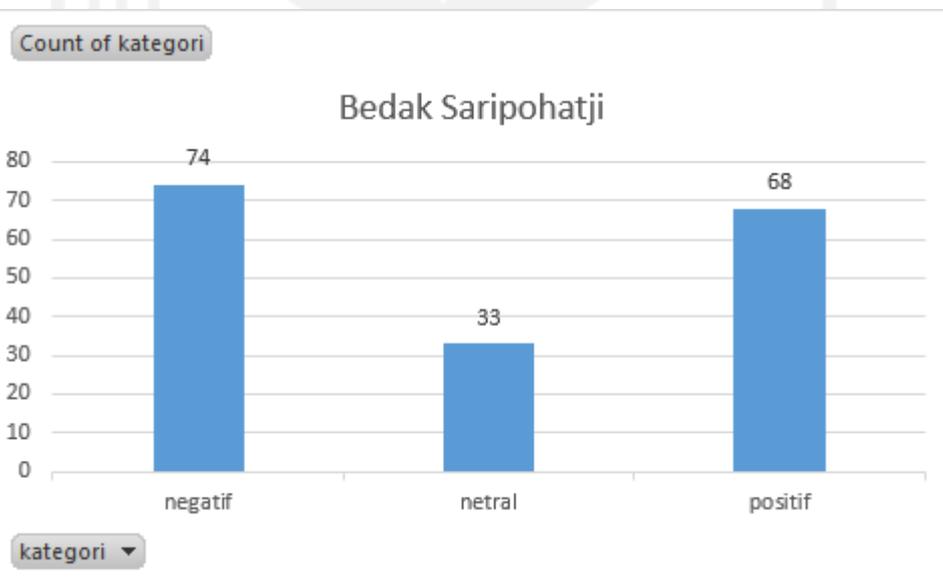


Gambar 3.10 grafik perbandingan ulasan pada produk oriflame

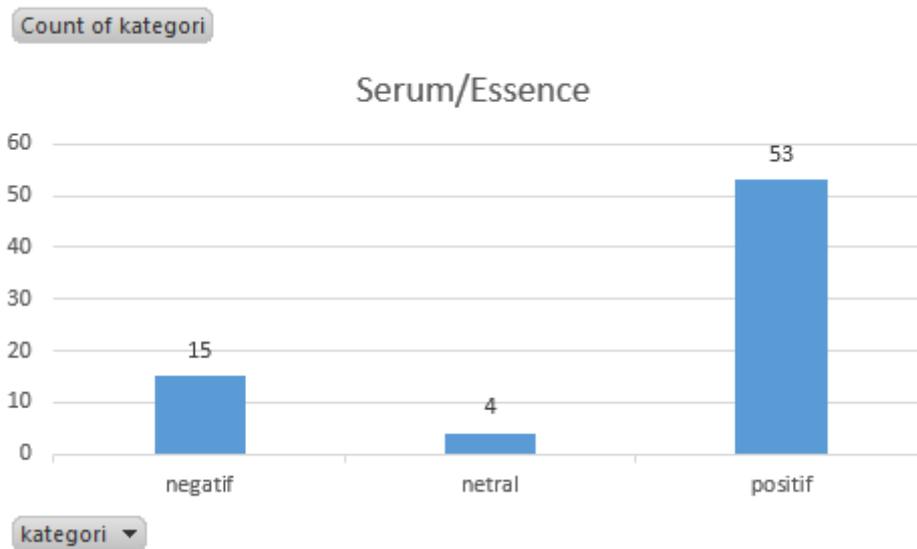


Gambar 3.11 grafik perbandingan ulasan pada produk sari ayu by marta tilaar

Gambar 3.12 memperlihatkan bahwa pada produk bedak saripohatji memiliki 74 ulasan negatif, 33 ulasan netral, dan 68 ulasan positif. Sedangkan Gambar 3.13 memperlihatkan bahwa pada produk serum atau essence memiliki 15 ulasan positif, 4 ulasan netral, dan 53 ulasan positif.

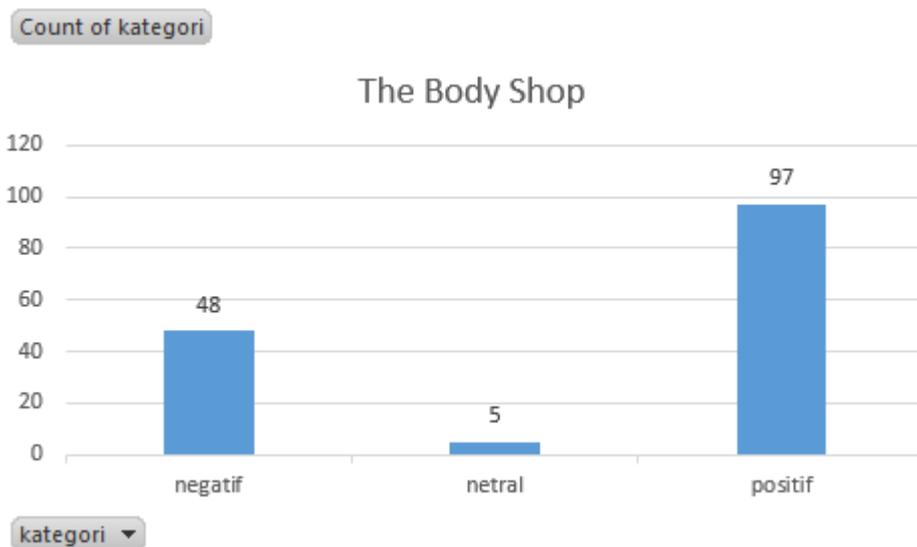


Gambar 3.12 grafik perbandingan ulasan pada produk bedak saripohatji

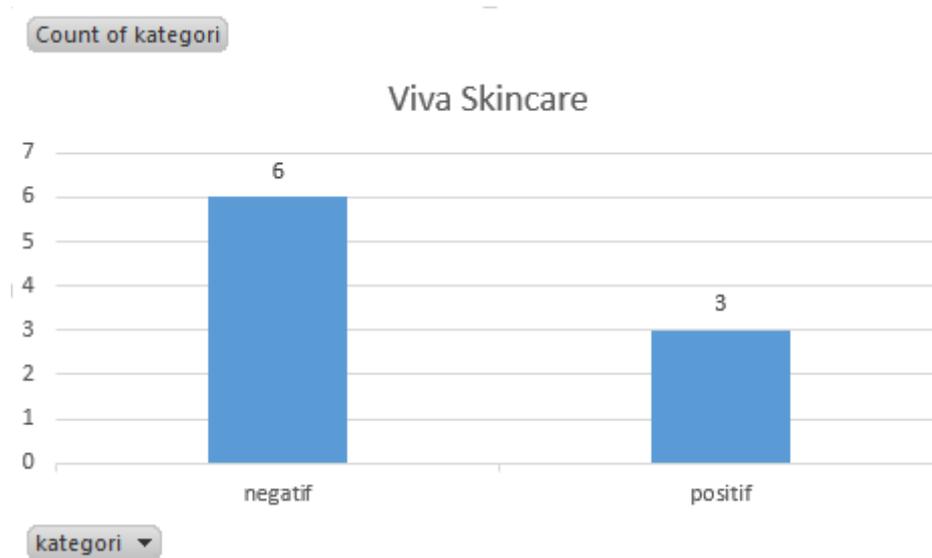


Gambar 3.13 grafik perbandingan ulasan pada produk serum atau essence

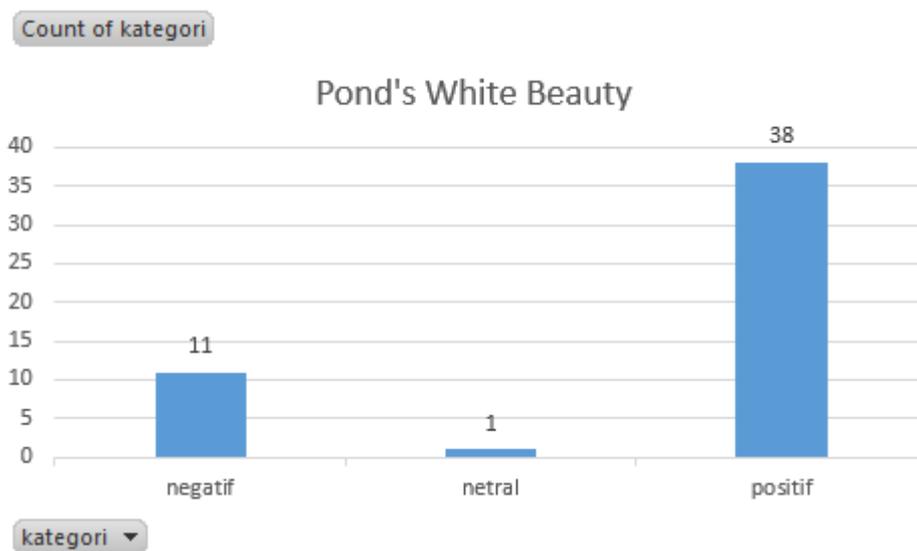
Gambar 3.14 memperlihatkan bahwa pada produk the body shop memiliki 48 ulasan negatif, 5 ulasan netral, dan 97 ulasan positif. Selain itu, Gambar 3.15 memperlihatkan bahwa pada produk viva *skincare* memiliki 6 ulasan negatif dan 3 ulasan positif. Sedangkan Gambar 3.16 memperlihatkan bahwa pada produk ponds white beauty memiliki 11 ulasan negatif, 1 ulasan netral, dan 38 ulasan positif.



Gambar 3.14 grafik perbandingan ulasan pada produk the body shop



Gambar 3.15 grafik perbandingan ulasan pada produk viva *skincare*



Gambar 3.16 grafik perbandingan ulasan pada produk ponds white beauty

3.3 Preprocessing

Preprocessing merupakan tahapan membersihkan dan mempersiapkan teks terlebih dahulu sebelum teks dianalisis oleh sistem. Pada Tahapan ini terdapat beberapa runtutan tahap yang dilakukan yang dilihat pada Gambar 3.17.



Gambar 3.17 Tahapan penelitian untuk *preprocessing*

Tahap *preprocessing* dilakukan untuk menghindari nilai data yang hilang, data yang berlebihan, dan data yang tidak konsisten. Adapun tahapan *preprocessing* yang akan dilakukan pada penelitian ini adalah:

a. *Regular Expression*

Regular expression bertujuan untuk mengambil karakter huruf abjad ‘a’ sampai dengan ‘z’ pada data ulasan. Selain karakter huruf abjad akan dihilangkan termasuk karakter tanda baca, angka, *hashtag*, dan *username*.

b. *Case Folding*

Case folding merupakan proses mengubah huruf ke dalam bentuk yang sama, yaitu kapital semua atau kecil semua. Dalam tahap ini, semua huruf yang ada dalam kalimat diseragamkan menjadi huruf kecil.

c. *Tokenizing*

Tokenizing adalah proses pemisahan kata pada kalimat menjadi satuan kata atau beberapa kata. Penelitian ini menggunakan dua kali tahap *tokenizing*. *Tokenizing* pertama sesuai dengan tujuan awal *tokenizing* yaitu untuk memisahkan kata pada kalimat menjadi satuan kata atau biasa disebut “token”, agar memudahkan proses *formalization*. *Tokenizing* kedua dilakukan setelah proses *formalization*, bertujuan untuk penggabungan dua kata atau lebih yang memiliki satu makna menjadi satu token contohnya “jerawat_menghilang”. Selain itu, dataset yang berperan sebagai token akan memiliki nilai dalam matriks dokumen yang akan dianalisis.

d. *Formalization*

Formalization adalah proses mengubah bentuk kata yang tidak standar ke bentuk standar sesuai struktur KBBI (Kamus Besar Bahasa Indonesia). Dalam tahap ini, penulis membuat kamus lokal yang disimpan di dalam *file* dengan format *csv*. Kamus tersebut bertujuan untuk mengubah struktur kata yang tidak sesuai standar KBBI seperti kata gaul (*slang word*), singkatan, dan salah penulisan kata (*typo*).

e. *Filtering*

Filtering merupakan proses pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan proses untuk menghilangkan kata yang dianggap tidak memiliki makna. Kata tersebut meliputi kata penghubung, kata ganti orang, kata seruan dan kata lainnya yang tidak memiliki arti dalam penentuan kelas suatu dokumen.

f. *Stemming*

Stemming adalah proses pengambilan kata yang memiliki makna dan mengubah kata ke bentuk dasarnya sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Pada tahap ini digunakan *library* Python *Sastrawi*. *Stemming* bertujuan untuk mentransformasikan kata menjadi kata dasarnya dengan menghilangkan kata imbuhan awalan, sisipan, akhiran serta gabungan awalan dan akhiran.

3.4 Pembobotan TF-IDF

Data yang telah melewati tahap *preprocessing* harus berbentuk angka, agar data yang masih berupa kata-kata dapat diolah dan dihitung. Data tersebut diubah menjadi vektor kemudian diberi nilai dan bobot untuk setiap kata kemudian diolah menggunakan algoritma prediksi.

Pada tahap pembobotan kata, menggunakan algoritma *Term-Frequency Times Inverse Document Frequency* (TF-IDF). Metode ini dipilih karena efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Invers Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen. Secara sederhana, metode ini digunakan untuk menghitung sering munculnya suatu kata atau *term* dalam dokumen.

Term Frequency (TF) merupakan frekuensi kemunculan sebuah kata atau *term* dalam dokumen. Semakin besar jumlah term yang muncul (TF tinggi) maka semakin besar bobot dokumen atau memberikan nilai kemiripan yang semakin besar. Ada beberapa jenis formula *Term Frequency* (TF):

- a. TF biner (*binary TF*)

Hanya mengamati ada tidaknya kata atau *term* pada sebuah dokumen, jika ada maka diberi nilai satu (1) jika tidak ada maka diberi nilai nol (0).

- b. TF murni (*raw TF*)

Nilai TF berlandaskan pada jumlah munculnya kata atau *term* dalam dokumen. Contohnya jika muncul empat (4) kali maka kata tersebut bernilai empat (4).

- c. TF logaritmik

Perhitungan nilai TF ini untuk mengatasi dominasi dokumen yang memiliki sedikit *term* dalam *query* tetapi memiliki frekuensi yang tinggi.

$$tf = 1 + \log(tf) \quad (21)$$

- d. TF normalisasi

Perhitungan ini membuktikan perbandingan antara frekuensi kata atau *term* dengan jumlah semua kata atau *term* pada sebuah dokumen.

$$tf = 0.5 + 0.5x \left(\frac{tf}{maxtf} \right) \quad (21)$$

Inverse Document Frequency (IDF) merupakan metode *statistik numeric* yang menghitung berapa diperlukannya kata dalam sebuah dokumen. Metode ini digunakan sebagai

bobot dalam penelusuran informasi dalam *text mining*. Rumus IDF dapat dituliskan sebagai berikut:

$$idf_j = \log\left(\frac{D}{df_j}\right) \quad (21)$$

Diketahui:

D : Jumlah keseluruhan dokumen dalam koleksi

df_j : Jumlah dokumen yang memiliki *term* t

Diperoleh rumus umum TF-IDF dengan menggabungkan perhitungan TF murni dan rumus IDF dengan mengalikan keduanya:

$$w_{ij} = tf_{ij} \times idf_j \quad (21)$$

$$w_{ij} = tf_{ij} \times \log\left(\frac{D}{df_j}\right) \quad (21)$$

Diketahui:

w_j : Bobot term t_j terhadap dokumen d_j

tf_{ij} : Jumlah munculnya term t_j dalam dokumen d_j

D : Jumlah keseluruhan dokumen dalam koleksi

df_j : Jumlah dokumen yang memiliki term t_j

Perhitungan (TF-IDF) akan dijelaskan berdasarkan rumus yang telah dijelaskan sebelumnya, pembobotan dilakukan dengan menggunakan tiga (3) dokumen berdasarkan kata ‘cerah’, ‘lumayan’, dan ‘belang’ sebagai berikut:

D_1 : Tapi udah agak lumayan sih, minimal kakinya cerahan.

D_2 : aku dicampur air mawar sejauh ini agak cerahan lumayan tapi dikit.

D_3 : Gw juga lumayan agak belang dikaki dan wajah gw.

Ketiga dokumen tersebut diolah melalui tahapan *preprocessing* sehingga akan mengalami perubahan dalam dokumen baru sebagai berikut:

D_1 : agak lumayan minimal cerah.

D_2 : agak cerah lumayan sedikit.

D_3 : lumayan agak belang.

Adapun *term* yang terbentuk dari hasil *preprocessing* ketiga dokumen diatas yang dijadikan sebagai contoh adalah agak, lumayan, minimal, cerah, sedikit, dan belang. Tahap selanjutnya adalah mengubah dokumen dalam bentuk vektor dengan elemen sebanyak *term*

query yang ada pada setiap dokumen. Vektor tersebut merupakan bobot dari setiap *query* yang dihitung berdasarkan metode TF-IDF. Contoh perhitungan TF-IDF dari *term* ‘mudah’ pada dokumen satu (D_1) adalah sebagai berikut:

$$tf_{cerah} = 1 \text{ Jumlah kemunculan } term \text{ ‘cerah’ dalam dokumen } d_1$$

$$D = 3 \text{ Jumlah semua dokumen dalam koleksi}$$

$$df_{cerah} = 2 \text{ Jumlah dokumen yang mengandung } term \text{ ‘cerah’}$$

Perhitungan dimulai dengan mencari nilai IDF dari *term* ‘cerah’ dengan persamaan:

$$idf_j = \log\left(\frac{D}{df_{cerah}}\right) = \log\left(\frac{3}{2}\right) = 0.176$$

Sehingga dihasilkan nilai IDF pada *term* ‘cerah’ adalah sebesar 0.176, selanjutnya akan dihitung nilai TF-IDF dari *term* ‘cerah’ dengan persamaan:

$$tfidf_j = tf_{cerah} \times idf_{cerah}$$

$$tfidf_{cerah} = 1 \times 0.176 = 0.176$$

Perhitungan TF-IDF dihitung pada tiap kata pada masing-masing dokumen kemudian dijumlahkan pada setiap dokumen. Sehingga bobot pada dokumen satu (D_1) adalah sebagai berikut:

$$w_{d1} = tfidf_{cerah} + tfidf_{lumayan} + tfidf_{belang}$$

$$w_{d1} = 0.176 + 0 + 0 = 0.176$$

Kemudian dapat dihitung nilai bobot untuk setiap *term* pada *query* dalam masing-masing dokumen adalah seperti tabel 3.1:

Tabel 3.1 Contoh perhitungan TF-IDF

Term				df	IDF	TF-IDF		
	D ₁	D ₂	D ₃			D ₁	D ₂	D ₃
cerah	1	1	0	2	0.176	0.176	0.176	0
lumayan	1	1	1	3	0	0	0	0
belang	0	0	1	1	0.477	0	0	0.477
Nilai bobot masing-masing dokumen						0.176	0.176	0.477

3.5 Pengujian Klasifikasi

Pengujian suatu klasifikasi dilakukan menggunakan sebuah himpunan data uji, yang tidak digunakan dalam pelatihan klasifikasi tersebut dengan ukuran yang telah ditentukan.

Terdapat beberapa ukuran yang dapat digunakan untuk mengevaluasi model klasifikasi, diantaranya adalah *accuracy*, *precision* dan *recall*. Model klasifikasi yang dibuat merupakan pemetaan dari baris data dengan sebuah hasil prediksi kelas dari data tersebut. Klasifikasi yang memiliki dua output label disebut dengan klasifikasi biner yang direpresentasikan dalam $\{0,1\}$, $\{\text{ya, tidak}\}$, $\{+1,-1\}$ atau $\{\text{positive; negative}\}$ untuk setiap data input yang diberikan.

Pada proses evaluasi klasifikasi ada empat kemungkinan yang terjadi dari hasil klasifikasi suatu data. Jika data positif dan diprediksi positif maka akan dihitung sebagai *true positive* dan jika data positif diprediksi negatif maka akan dihitung sebagai *false negative*. Pada data negatif jika diprediksi negatif maka akan dihitung *true negative* dan jika diprediksi positif maka akan dihitung *false positive* (Fawcett, 2006).

		Aktual	
		Positive	Negative
Prediksi	Positive	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	Negative	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

a. *Accuracy*

Accuracy adalah tingkat ketepatan prediksi yang benar. Rumus *accuracy* dapat dilihat pada persamaan (22).

$$Accuracy = \frac{TP+FN}{TP + FP + TN + FN} \quad (22)$$

b. *Precision*

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban sistem. Rumus *precision* dapat dihitung melalui persamaan (23).

$$Precision = \frac{TP}{TP + FP} \quad (23)$$

c. *Recall*

Recall adalah perhitungan ketepatan prediksi yang digunakan sebagai ukuran tingkat keberhasilan sistem dalam menemukan nekembali sebuah informasi. Rumus *recall* dapat dihitung melalui persamaan (24).

$$Precision = \frac{TP}{TP + FN} \quad (24)$$

3.6 Penyimpanan Model

Setelah melatih model diperlukan untuk menyimpan model tersebut dengan tujuan agar tidak perlu melatih ulang model yang sudah dilatih. Dalam penyimpanan model ini menggunakan model bawaan *python* yaitu *pickle*.

3.7 Perancangan Perangkat Lunak

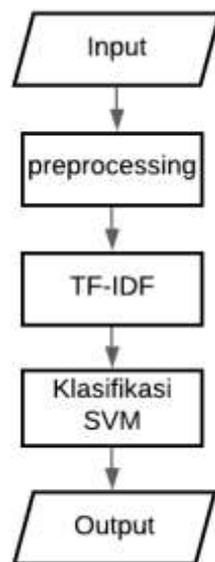
Tahapan ini adalah proses untuk merancang perangkat lunak berdasarkan hasil analisis yang telah dilakukan. Perancangan bertujuan untuk mempermudah dalam membangun perangkat lunak. Perangkat lunak bertujuan untuk menjabatani tampilan agar lebih *user-friendly*. Perangkat lunak yang digunakan dalam penelitian ini berbasis web. Perangkat lunak diimplementasikan menggunakan bahasa pemrograman Python dan *framework* Django.

3.7.1 Diagram alir

Berdasarkan analisis perancangan sistem yang telah dilakukan, pembuatan diagram alir sistem terbagi menjadi beberapa bagian yaitu diagram alir analisis sentimen dan diagram alir *preprocessing*.

Diagram Alir Analisis Sentimen

Gambar 3.18 merupakan visualisasi tentang perencanaan umum tahapan ulasan yang akan dianalisis. Pada gambar masukan teks ulasan akan melalui tahap *preprocessing*, kemudian tahap TF-IDF, kemudian tahap klasifikasi menggunakan algoritma *Support Vector Machine*.



Gambar 3.18 Diagram alir analisis sentimen

3.7.2 Perancangan *Interface*

Perancangan *interface* bertujuan untuk menggambarkan tampilan antarmuka yang akan dibuat. *Interface* merupakan perantara pengguna sistem untuk dapat berkomunikasi dengan sistem. Rancangan *interface* yang dibangun hanya terdapat satu halaman saja menampilkan hasil prediksi sentimen dari ulasan produk.

Rancangan *interface* yang akan dibuat terlihat pada Gambar 3.19. *Interface* ini berisi form *input*, *output* dan tombol untuk mengetahui hasil prediksi dari ulasan yang dianalisis. *Form input* berupa *textfield* untuk ulasan produk yang akan dianalisis sedangkan *output alert* berupa hasil dari prediksi dari ulasan yang telah dianalisis.

Analisis Sentimen

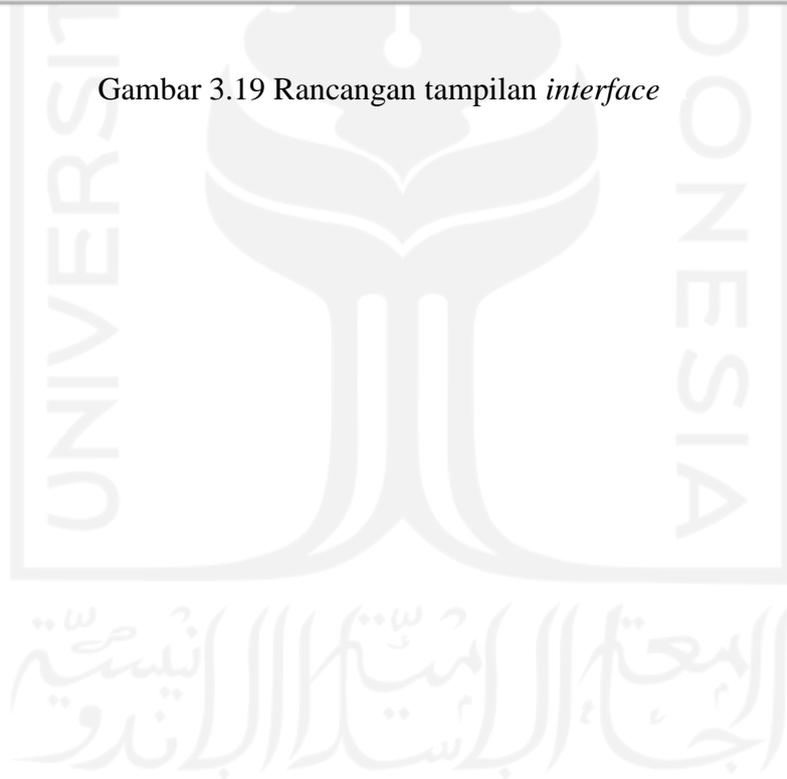
Review Produk *Skincare*:

Bagus banget produknya, wajah engga jerawat lagi tambah mulus dan cerah

submit

POSITIF! Bagus banget produknya, wajah engga jerawat lagi tambah mulus dan cerahx

Gambar 3.19 Rancangan tampilan *interface*



BAB IV

IMPLEMENTASI DAN PENGUJIAN

4.1 Pengumpulan Data

Proses pengumpulan dataset yang digunakan pada penelitian ini melalui forum Female Daily. Data yang dikumpulkan merupakan pengambilan data dari 15 Desember 2020 sampai dengan 13 Juli 2021. Dari 3600 data yang diperoleh hanya 1260 data yang dipakai, data tidak bisa dipakai karena data tersebut bukan termasuk data ulasan produk. Tahapan ini melibatkan *library* Python yaitu *beautifulsoup4*. *Library* tersebut berfungsi untuk melakukan proses parsing XML dan HTML dokumen yang sudah terunduh. Sebelum menggunakan *library* tersebut, peneliti melakukan proses instalasi dengan cara mengetikkan perintah *pip* pada *terminal visual studio code*. Gambar 4.1 merupakan visual dari perintah instalasi *library* Python *beautifulsoup4*.

```
pip install beautifulsoup4
```

Gambar 4.1 Perintah instalasi *library beautifulsoup4*

Penelitian ini menggunakan *library* Python *beautifulsoup4* karena mudah untuk digunakan. Dataset yang diambil merupakan ulasan produk *skincare* berdasarkan pada halaman web forum Female Daily. Proses pengambilan dataset berdasarkan halaman yang dituju diimplementasikan ke dalam konstruktor yang terdapat pada Gambar 4.2. Adapun keseluruhan halaman website forum Female Daily yang dilakukan *scraping* terdapat pada Lampiran A.

```
pages = []
for i in range (1,30):
    res = 'https://forum.femaledaily.com/showthread.php?21723-NIVEA-Make-Up-
        clear'+ '/page'+str(i)
    pages.append(res)
```

Gambar 4.2 Inisiasi *scraping* pada halaman yang dituju

Teknik parsing HTML digunakan untuk mengambil dataset. Pengambilan dataset dilakukan dengan cara mencari *id* dan *class_* yang dituju. Dataset yang akan diambil selanjutnya disimpan ke dalam *file* yang memiliki format *csv* dan diberi nama *review*. Di dalam *file csv* tersebut terbagi menjadi dua kolom yaitu kolom “*title*” untuk nama produk dan kolom

“review” untuk ulasan produk. Berikut merupakan implementasi *library* Python *beautifulsoup4* dan penyimpanan dataset dapat dilihat pada Gambar 4.3.

```
#scrapping halaman yang dituju
...
f = csv.writer(open('review.csv', 'w+'))
f.writerow('title','review')

for i in pages:
    page = request.get(it)
    soup = BeautifulSoup(page.content,'html.parser')
    week = soup.find(id='postlist')
    items = week.find_all(class_='postbitlegacy postbitim postcontainer old')
    for item in items:
        titles = item.find('h2').text.strip()
        reviews= item. Find(class_='postcontent restore').text.strip()
        fh = open("review.csv")
        for line in fh:
            print(re.sub(r"b'",' ', line), end= '')
        f.writerow([titles, reviews])
        print(type(titles))
```

Gambar 4.3 Inisiasi *library* Python *beautifulsoup4* dan penyimpanan *file*

Contoh hasil dari implementasi tahap pengambilan dataset dapat dilihat pada Gambar 4.4. Dapat dilihat bahwa teks mengandung beberapa yaitu teks *username*, isi ulasan, dan beberapa *noise*, yaitu terdapat kata “Hidden Content”. Hasil ini masih membutuhkan tahap *preprocessing* agar mendapatkan teks ulasan yang diinginkan sebelum ke tahap selanjutnya. Ulasan yang dikumpulkan dari hasil *scraping* dapat dilihat pada Gambar 4.4 berikut:

```
Aku sendiri juga sedang minum pil KB ini,udh jalan 2 mingguan.. berasa kulit lbh rata wrnanya dan lbh alusan
Seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit @H Hihhi
@ladygatha: aku uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya. Hidden Content
coba pakek aja dan sekarang SENENG BANGET hidden Content ..nie bedak bisa bikin muka GAK jerawatan lagi hidden Content
@kuecoklat,, aku udah nyobain ni masker n sumpe bikin ketagihan,,sampe swamiku jg ikutan maskeran,,hahahah...
pake Aztec kulitku makin berminyak dan jerawatan .... hidden Content
i love it ! perlahan bruntusan ilang, jidat mulai mulus lagi...
@hidden Content hiks sayang juga yah ternyata ga cocok di muka kamu..
Hidden Content gw pake si kb ya gitulah..... bagus en putihan dikit.
```

Gambar 4.4 Contoh teks hasil implementasi *scraping*

4.2 Implementasi *Preprocessing*

Preprocessing diimplementasikan melalui beberapa tahapan yang memiliki fungsi masing-masing.

a. *Regular Expression*

Regular expression bertujuan untuk membersihkan data dengan cara menghapus *username* (@), *hashtag* (#), tanda baca, angka, *single character*, dan mengganti *multiple space* menjadi *single space*. Data tersebut dibersihkan bertujuan untuk

mengurangi beban pemrosesan dan dianggap tidak relevan untuk dianalisa. Contoh hasil implementasinya dapat dilihat pada Tabel 4.1. Kolom sebelah kiri merupakan ulasan sebelum dilakukan *regular expression* dan kolom sebelah kanan merupakan ulasan yang sudah melalui tahap *regular expression*. Sedangkan implementasi *regular expression* dapat dilihat dari fungsi *regex* yang ada pada Gambar 4.5.

```
def regex(text):
    username = re.sub(r"@(\w+)", " ", small, flags = re.I)
    hashtag = re.sub(r"#(\w+)", " ", username, flags = re.I)
    tanda = re.sub(r"\W", " ", hashtag, flags=re.I)
    number = re.sub(r"\d", " ", tanda, flags = re.I)
    single = re.sub(r"\b[a-zA-Z]\b", " ", number, flags = re.I)
    spaces = re.sub(r"\s+", " ", single, flags = re.I)
    text = spaces.lower()
    return text
```

Gambar 4.5 Contoh kode program *regular expression* dan *case folding*

Tabel 4.1 Contoh hasil ulasan setelah dilakukan *regular expression*

Ulasan	Ulasan setelah dilakukan <i>regular expression</i>
@kuecoklat,, aku udah nyobain 2 bulan ni masker sumpe bikin ketagihan,,sampe swamiku jg ikutan maskeran,,hahahah...	aku udah nyobain bulan ni masker sumpe bikin ketagihan sampe swamiku jg ikutan maskeran hahahah
Seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit 😊👍👍👍 Hihi	Seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit Hihi
@ladygatha: gw uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya. Hidden Content	gw uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya Hidden Content
coba pakek aja dan sekarang SENENG BANGET hidden Content ..nie bedak bisa bikin muka GAK jerawat lagi	coba pakek aja dan sekarang SENENG BANGET hidden Content nie bedak bisa bikin muka GAK jerawat lagi

b. *Case folding*

Setelah proses pembersihan data, kemudian dilakukan proses *case folding*. *Case folding* dalam penelitian ini bertujuan untuk menyeragamkan semua huruf menjadi huruf kecil. Implementasi proses *case folding* menggunakan *method* Python yaitu *lower()*, dapat dilihat Gambar 4.5 baris ke-8 pada proses *regular expression*.

Sedangkan untuk contoh hasil implementasinya dapat dilihat pada Tabel 4.2. Kolom sebelah kiri merupakan ulasan sebelum dilakukan *case folding* dan kolom sebelah kanan merupakan ulasan yang sudah melalui tahap *case folding*.

Tabel 4.2 Contoh hasil ulasan setelah dilakukan *case folding*

Ulasan hasil tahap sebelumnya	Ulasan setelah dilakukan <i>case folding</i>
aku udah nyobain bulan ni masker sumpe bikin ketagihan sampe swamiku jg ikutan maskeran hahahah	aku udah nyobain bulan ni masker sumpe bikin ketagihan sampe swamiku jg ikutan maskeran hahahah
Seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit Hihi	seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit hihi
gw uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya Hidden Content	gw uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya hidden content
coba pakek aja dan sekarang SENENG BANGET hidden Content nie bedak bisa bikin muka GAK jerawatan lagi	coba pakek aja dan sekarang seneng banget hidden content nie bedak bisa bikin muka gak jerawatan lagi

c. *Tokenizing* pertama

Tokenizing pertama bertujuan untuk memisahkan kata pada kalimat menjadi satuan kata atau biasa disebut “token”. Implementasi *tokenizing* menggunakan modul *word_tokenize* dari *library* Python yang bernama *nltk* dapat dilihat Gambar 4.6 baris ke-8 pada tahapan *formalization*. Hasil dari proses *tokenizing* dapat dilihat pada Tabel 4.3.

Tabel 4.3 Contoh hasil ulasan setelah dilakukan *tokenizing* pertama

Ulasan hasil tahap sebelumnya	Ulasan setelah dilakukan proses <i>tokenizing</i>
aku udah nyobain bulan ni masker sumpe bikin ketagihan sampe swamiku jg ikutan maskeran hahahah	'aku', 'udah', 'nyobain', 'ni', 'masker', 'sumpe', 'bikin', 'ketagihan', 'sampe', 'swamiku', 'jg', 'ikutan', 'maskeran', 'hahahah'

seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit hihi	'seneng', 'bgt', 'yaaa', 'rasanya', 'udah', 'nemu', 'jodoh', 'yg', 'terbaik', 'untuk', 'kulit', 'hihi'
gw uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya hidden content	'gw', 'uda', 'bbrp', 'minggu', 'ini', 'ga', 'pake', 'nivea', 'ini', 'cz', 'kulitku', 'kok', 'jadi', 'menghitam', 'ya', 'hidden', 'content'
coba pakek aja dan sekarang seneng banget hidden content nie bedak bisa bikin muka gak jerawat lagi	'coba', 'pakek', 'aja', 'dan', 'sekarang', 'seneng', 'banget', 'hidden', 'content', 'nie', 'bedak', 'bisa', 'bikin', 'muka', 'gak', 'jerawatan', 'lagi', 'hidden', 'content'

d. Formalization

Formalization bertujuan untuk mengubah bentuk kata yang tidak standar ke bentuk standar sesuai struktur KBBI (Kamus Besar Bahasa Indonesia). Tahapan ini memiliki dua fungsi, pertama fungsi *get_lookup()* untuk memanggil kamus lokal yang disimpan pada *file* yang memiliki format *csv* bernama *kamus.csv* dan fungsi *formalize()* untuk mengubah bentuk kata sesuai dengan kamus lokal. Implementasi dari *formalization* dapat dilihat pada Gambar 4.6.

```
def get_lookup():
    df = pd.read_csv("/content/drive/My Drive/Skripsi/kamus.csv")
    df_record = df.to_dict('records')
    return { entry['slang']: entry['formal'] for entry in df_record}

lookup = get_lookup()
def formalize(text):
    text = nltk.tokenize.word_tokenize(text)
    formalized = [lookup[token] if token in lookup
                  else token for token in text]
    if formalized:
        re_tokens = formalized
    else:
        re_tokens = text
    freq_tokens = nltk.FreqDist(formalized)
    return " ".join(freq_tokens)
```

Gambar 4.6 Kode program *formalization* dan *tokenizing* pertama

Bentuk kata tidak standar yang ditemukan pada data ulasan ada berbagai macam bentuk. Bentuk kata yang mengalami perbaikan adalah *slang word*, singkatan, salah penulisan kata (*typo*), bahasa asing, karakter berulang dan lainnya. Tabel 4.3, Tabel

4.4, Tabel 4.5, Tabel 4.6, dan Tabel 4.7 merupakan contoh dari masing-masing bentuk kata dan perbaikannya. Adapun daftar kata *slang*, singkatan, salah penulisan kata (*typo*), bahasa asing, dan kata karakter berulang beserta perbaikannya terdapat pada Lampiran B.

Tabel 4.4 Contoh kata *slang word* dan perbaikannya

<i>Slang word</i>	Perbaikan <i>slang word</i>
murce	murah
sabi	bisa
hahahah	tawa
geday	gede
plesong	palsu

Tabel 4.5 Contoh singkatan dan perbaikannya

Singkatan	Perbaikan singkatan
sdh	sudah
rb	ribu
pst	pasti
prnh	pernah
tkt	takut

Tabel 4.6 Contoh kata yang salah dalam penulisan (*typo*) dan perbaikannya

<i>Typo</i>	Perbaikan <i>typo</i>
uda	sudah
sdikit	sedikit
gateell	gatal
cucok	cocok
tremasuk	termasuk

Tabel 4.7 Contoh kata asing dan perbaikannya

Bahasa asing	Perbaikan bahasa asing
answer	jawab

another	lain
applied	terapkan
atleast	paling sedikit
pore	pori

Tabel 4.8 Contoh kata dengan karakter berulang

Bahasa asing	Perbaiki karakter berulang
yaaa	ya
aluuuus	halus
ajaaa	saja
ahhhh	ah
adaaa	ada

Adapun contoh ulasan hasil implementasi *formalization* dapat dilihat pada Tabel 4.9. Bentuk kata yang diperbaiki ditulis dengan huruf tebal. Kolom sebelah kiri merupakan ulasan sebelum dilakukan *formalization* dan kolom sebelah kanan merupakan hasil ulasan yang sudah melalui tahap *formalization*.

Tabel 4.9 Contoh hasil ulasan setelah dilakukan *formalization*

Ulasan hasil tahap sebelumnya	Ulasan setelah dilakukan <i>formalization</i>
aku udah nyobain bulan ni masker sumpe bikin ketagihan sampe swamiku jg ikutan maskeran hahahah	aku sudah mencoba bulan ini masker sumpah bikin ketagihan sampai suamiku juga ikutan maskeran tawa
seneng bgt yaaa rasanya udah nemu jodoh yg terbaik untuk kulit hihi	senang banget ya rasanya sudah menemukan jodoh yang terbaik untuk kulit tawa
gw uda bbrp minggu ini ga pake nivea ini cz kulitku kok jadi menghitam ya hidden content	aku sudah beberapa minggu ini enggak pakai nivea karena kulitku kok jadi menghitam ya hidden content

coba pakek aja dan sekarang seneng banget hidden content nie bedak bisa bikin muka gak jerawat lagi	coba pakai saja dan sekarang senang banget hidden content nih bedak bisa bikin wajah enggak jerawat lagi
---	--

e. *Tokenizing* kedua

Tokenizing kedua bertujuan untuk penggabungan dua kata atau lebih yang memiliki satu makna menjadi satu token. Hasil dari proses tersebut diberi tanda “_” untuk penggabungan kata contohnya “jerawat_menghilang”. Proses ini menggunakan *library* yang sama dengan sebelumnya, akan tetapi kali ini menggunakan modul *MWETokenizer*. Hasil proses *tokenizing* kedua ini ditulis dengan huruf tebal dapat dilihat pada Tabel 4.10.

Tabel 4.10 Contoh hasil ulasan setelah dilakukan *tokenizing* kedua

Ulasan hasil tahap sebelumnya	Ulasan setelah dilakukan proses <i>tokenizing</i>
aku sudah mencoba bulan ini masker sumpah bikin ketagihan sampai suamiku juga ikutan maskeran tawa	'aku', 'sudah', 'mencoba', 'ini', 'masker', 'sumpah', 'bikin', 'ketagihan', 'sampai', 'suamiku', 'juga', 'ikutan', 'maskeran', 'tawa'
senang banget ya rasanya sudah menemukan jodoh yang terbaik untuk kulit tawa	'senang', 'banget', 'ya', 'rasanya', 'sudah', 'menemukan', 'jodoh', 'yang', 'terbaik', 'untuk', 'kulit', 'tawa'
aku sudah beberapa minggu ini enggak pakai nivea karena kulitku kok jadi menghitam ya hidden content	'aku', 'sudah', 'beberapa', 'minggu', 'ini', ' enggak_pakai ', 'nivea', 'karena', 'kulitku', 'kok', 'jadi', 'menghitam', 'ya', 'hidden', 'content'
coba pakai saja dan sekarang senang banget hidden content nih bedak bisa bikin wajah enggak jerawat lagi	'coba', 'pakai', 'saja', 'dan', 'sekarang', 'senang', 'banget', 'hidden', 'content', 'nih', 'bedak', 'bisa', 'bikin', 'wajah', ' enggak_jerawat ', 'lagi'

f. *Filtering*

Setelah berhasil mengubah kata ke bentuk formal dan tokenisasi kembali, tahapan selanjutnya adalah *filtering* atau *stopword*. Tahapan *filtering* bertujuan untuk

menghilangkan kata yang dianggap tidak bermakna. Tahap ini dilakukan dengan cara menggunakan *library* Python yaitu *sastrawi*. Sebelum menggunakan *sastrawi*, peneliti melakukan instalasi *library* menggunakan perintah *pip* sebagaimana ditunjukkan oleh Gambar 4.11.

```
pip install Sastrawi
```

Gambar 4.11 Perintah instalasi *library sastrawi*

Pada *library* Python *sastrawi* memiliki *stopword* tersendiri, namun pada penelitian ini ada beberapa kata yang dihilangkan dari *stopword sastrawi* karena dianggap penting dalam proses analisis. Tahap ini dilakukan dengan cara menggunakan perintah *remove words*. Kata yang dihilangkan dari daftar *stopword sastrawi* dapat dilihat pada Tabel 4.11.

Tabel 4.11 Kata yang dihilangkan dari daftar *stopword*

Kata yang dihilangkan
ingin
enggak
agak
tidak

Selain itu, ada beberapa kata yang ditambahkan pada *stopword sastrawi* karena dianggap tidak bermakna dalam proses analisis. Tahap ini dapat dilakukan dengan cara membuat perintah *extend words*. Contoh kata tambahan untuk daftar *stopword* dapat dilihat pada Tabel 4.12. Baik kata tambahan maupun yang dihilangkan dari daftar *stopword* merupakan hasil dari analisis data ulasan. Adapun daftar *stopword* yang digunakan terdapat pada lampiran C.

Tabel 4.12 Contoh kata tambahan untuk daftar *stopword*

banyak	semakin	pakai	halo	ah
tawa	terlihat	si	yah	mak
hasilnya	super	pas	sih	kemarin

Implementasi proses *filtering* menggunakan *library sastrawi* dapat dilihat pada Gambar 4.12 dan hasil dari implementasi proses *filtering* dapat dilihat pada Tabel 4.14. Bentuk kata yang termasuk *stopword* pada ulasan ditulis dengan huruf tebal. Kolom sebelah kiri merupakan ulasan sebelum dilakukan *filtering* dan kolom sebelah kanan merupakan hasil ulasan yang sudah melalui tahap *filtering*.

```
factory = StopWordRemoverFactory()
stopword = factory.create_stop_word_remover()
newlist = ['banyak', 'semakin', 'tawa', 'terlihat', 'hasilnya', 'ah']
stopword.dictionary.words.remove('tidak')
stopword.dictionary.words.remove('agak')
stopword.dictionary.words.remove('enggak')
stopword.dictionary.words.remove('ingin')
stopword.dictionary.words.extend(newlist)
def stopwords(text):
    cleaned = stopword.remove(text)
    return cleaned
```

Gambar 4.12 Contoh kode program *filtering*.

Tabel 4.14 Contoh hasil ulasan setelah dilakukan *filtering*

Ulasan hasil tahap sebelumnya	Ulasan setelah dilakukan <i>filtering</i>
aku sudah mencoba bulan ini masker sumpah bikin ketagihan sampai suamiku juga ikutan maskeran tawa	sumpah ketagihan suamiku ikutan
senang banget ya rasanya sudah menemukan jodoh yang terbaik untuk kulit tawa	senang menemukan jodoh terbaik
aku sudah beberapa minggu ini enggak_pakai nivea karena kulitku kok jadi menghitam ya hidden content	enggak_pakai menghitam
coba pakai saja dan sekarang senang banget hidden content nih bedak bisa bikin wajah enggak_jerawat lagi	senang enggak_jerawat

g. Stemming

Setelah menghilangkan *stopword*, tahapan selanjutnya adalah *stemming*. *Stemming* merupakan proses mengubah kata berimbuhan ke bentuk dasarnya sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Tahap ini dilakukan dengan cara mengecek data ulasan per token menggunakan fungsi *StemmerFactory()* yang telah diimpor. Implementasi *stemming* menggunakan *library sastrawi* dapat dilihat pada Gambar 4.13.

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()
def stemming(text):
    list_tokens = text.split(' ')
    text = [(token + ":" + stemmer.stem(token)) for token in list_tokens]
    text = [stemmer.stem(token) for token in list_tokens]
    return " ".join(text)
```

Gambar 4.13 Kode program *stemming*.

Hasil implementasi *stemming* yang dilakukan pada ulasan dapat dilihat pada Tabel 4.15 berikut:

Tabel 4.15 Contoh hasil ulasan setelah dilakukan *stemming*

Ulasan hasil tahap sebelumnya	Ulasan setelah dilakukan <i>stemming</i>
sumpah ketagihan ikutan	sumpah tagih ikut
senang menemukan jodoh terbaik	senang temu jodoh baik
enggak_pakai menghitam	enggak pakai hitam
senang enggak jerawatan	senang enggak jerawat

4.3 Data Latih dan Data Uji

Data latih digunakan oleh algoritma klasifikasi bertujuan membentuk sebuah model *classifier*. Model tersebut merupakan hasil representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada sebelumnya. Semakin banyak data latih yang digunakan, maka akan semakin baik mesin dalam memahami pola data. Data uji digunakan untuk mengukur keakuratan *classifier* melakukan klasifikasi dengan benar. Data yang digunakan memiliki label kelas, dengan perbandingan data latih dan data uji 79.3% : 20.7%. Perbandingan jumlah data latih dan data uji dapat dilihat pada Tabel 4.16.

Tabel 4.16 Perbandingan data latih dan data uji

Klasifikasi	Jumlah	Data Latih (79.3%)	Data Uji (20.7%)
Positif	628	498	129.9 \approx 130
Netral	174	147	27
Negatif	458	363,1 \approx 363	94,8 \approx 95
Total	1260	1008	252

4.4 Klasifikasi *Support Vector Machine*

Data latih yang telah dibentuk akan dipelajari melalui proses klasifikasi. Pada data latih terdapat ulasan negatif, ulasan netral, dan ulasan positif, kemudian pola data tersebut akan dipelajari menggunakan algoritma SVM berdasarkan ciri-ciri data pada masing-masing label kelas yang telah dibentuk. Selanjutnya akan dilakukan proses *machine learning* yaitu hasil pelatihan pada algoritma SVM akan diuji untuk mengetahui nilai akurasi dalam memprediksi data baru.

Penelitian ini melakukan percobaan beberapa kernel untuk memperoleh klasifikasi dengan akurasi terbaik. Metode kernel yang digunakan yaitu kernel *linear*, *Polynomial*, *Radial Basis Function* (RBF), dan *Sigmoid*. Berikut merupakan hasil perbandingan setiap kernel yang telah diuji coba dapat dilihat pada Tabel 4.17.

Tabel 4.17 Perbandingan penggunaan metode kernel pada klasifikasi SVM

Kernel	Akurasi
<i>Linear</i>	86.9 %
<i>Polynomial</i>	62.3 %
<i>RBF</i>	83.73 %
<i>Sigmoid</i>	84.92 %

Pada Tabel 4.15 terlihat bahwa dari perbandingan keempat *kernel* yang telah diuji, kernel *linear* memiliki nilai akurasi yang paling tinggi. Dengan demikian kernel *linear* akan dipilih untuk proses klasifikasi.

4.4.1 Evaluasi Sistem Klasifikasi

Proses evaluasi klasifikasi dilakukan dengan cara membuat *machine learning* menggunakan data latih dan data uji secara acak. Dalam melakukan proses evaluasi penelitian ini menggunakan metode *confusion matrix*. *Confusion matrix* digunakan untuk memudahkan dalam proses perhitungan akurasi dengan mengetahui jumlah data uji yang terklasifikasi dengan benar dan jumlah data uji yang salah pengklasifikasiannya. Tabel 4.18 merupakan *confusion matrix*.

Tabel 4.19 *confusion matrix*

		Aktual		
Prediksi		Positif	Netral	Negatif
	Positif	117	8	8
	Netral	4	15	0
	Negatif	9	4	87
		Akurasi		
		86.9%		

Tabel 4.18 menjelaskan besarnya persentase data yang diprediksi menggunakan *confusion matrix*. Analisis sentimen ulasan produk *skincare* menggunakan metode *Support Vector Machine* diperoleh hasil prediksi bahwa pada kelas positif, dari 130 ulasan positif yang diuji, terdapat 117 ulasan terklasifikasi secara tepat dan terjadi kesalahan prediksi sebesar 4 ulasan untuk kelas netral dan 9 ulasan untuk kelas negatif. Selanjutnya pada ulasan netral yang diuji, dari total 27 ulasan terdapat 15 ulasan yang terklasifikasi dengan tepat, namun terjadi kesalahan prediksi sebanyak 8 ulasan untuk ulasan positif dan 4 untuk ulasan negatif. Sedangkan pada ulasan negatif yang diuji, dari total 95 ulasan terdapat 87 ulasan yang terklasifikasi dengan tepat sebagai ulasan negatif, namun terjadi kesalahan prediksi terdapat 8 ulasan yang masuk ke dalam ulasan positif. Kemudian dari nilai *confusion matrix* tersebut diperoleh nilai akurasi sebesar 86.9%, artinya dari 252 data ulasan yang diuji, terdapat 219 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine* (SVM). Adapun rumus perhitungan nilai akurasi dilakukan dengan menggunakan persamaan (25).

$$Accuracy = \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah semua data yang di uji}} \times 100\% \quad (25)$$

Sehingga, nilai akurasi dari metode SVM diperoleh dari perhitungan berikut:

$$Accuracy = \frac{117 + 15 + 87}{117 + 4 + 9 + 8 + 15 + 4 + 8 + 0 + 87} \times 100\%$$

$$Accuracy = \frac{219}{252} \times 100\%$$

$$Accuracy = 86.9\%$$

4.5 Penyimpanan Model

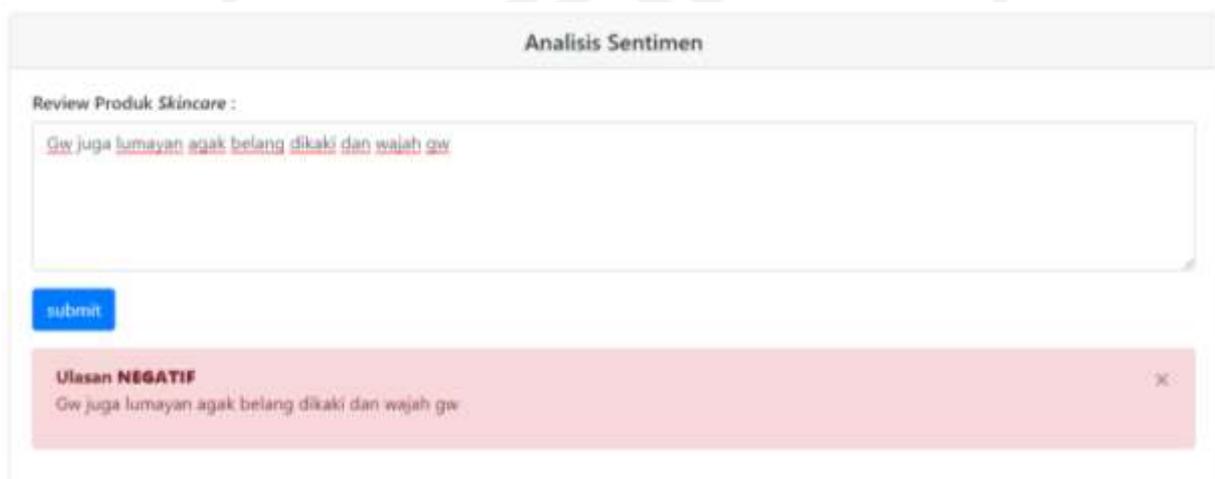
Setelah melatih model diperlukan untuk menyimpan model tersebut dengan tujuan agar tidak perlu melatih ulang model yang sudah dilatih. Dalam penyimpanan model ini menggunakan model bawaan *python* yaitu *pickle*. Model tersebut disimpan dalam *file* berbentuk *.pkl* yang bernama *model.pkl*. Implementasi menyimpan model menggunakan *pickle* dapat dilihat pada Gambar 4.14.

```
With open("/content/drive/My Drive/Skripsi/test/model.pkl", 'wb') as file:
    Pickle.dump(model, file)
```

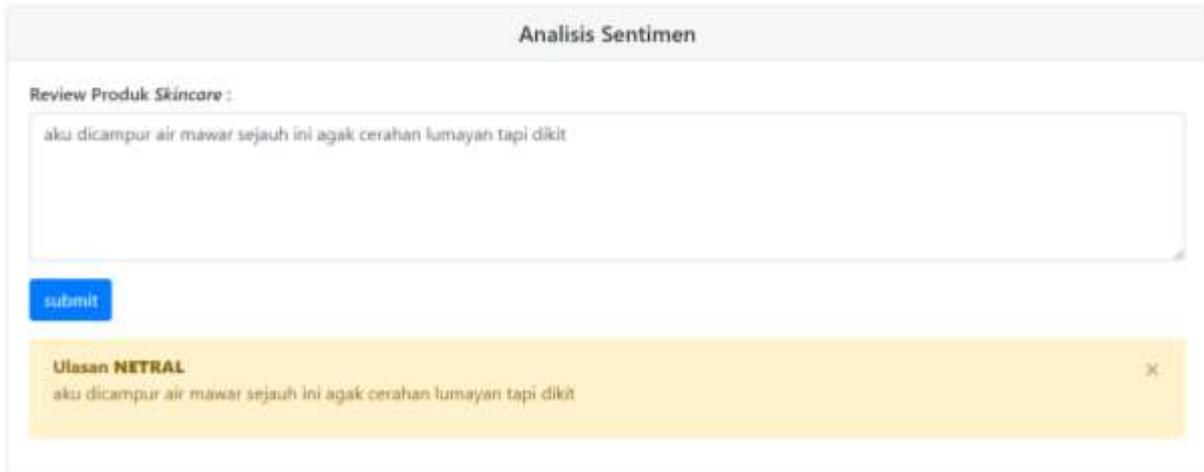
Gambar 4.14 Kode program penyimpanan model.

4.6 Pengujian Analisis Sentimen Pada Antarmuka Sistem

Dalam membangun sistem, peneliti menggunakan *framework* Django. *Framework* tersebut digunakan karena mendukung pengembangan sistem dalam bahasa pemrograman Python. Sistem yang dibangun adalah bagian halaman utama yang berfungsi untuk menganalisis ulasan yang dimasukkan oleh *user*. Adapun pengujian antarmuka akan mengambil data sampel dari salah satu ulasan dari data mentah tanpa proses *preprocessing*, karena proses tersebut akan dilakukan di dalam sistem.



Gambar 4.15 Uji coba sampel negatif teks pada tampilan sistem



Gambar 4.16 Uji coba sampel netral teks pada tampilan sistem



Gambar 4.17 Uji coba sampel positif teks pada tampilan sistem

Sistem menunjukkan dapat mengklasifikasi sampel data negatif pada Gambar 4.15, sampel data netral pada Gambar 4.16, dan sampel data positif pada Gambar 4.17. Sistem dapat mengklasifikasi data dengan baik dan dapat menampilkan hasilnya.

4.7 Kekurangan Sistem

Dalam membangun sistem, peneliti menyadari bahwa ada kekurangan dalam penelitian ini sebagai berikut:

1. Dataset yang digunakan baik jumlah data positif, jumlah data netral, maupun jumlah data negatif tidak memiliki perbandingan yang sama.
2. Sistem belum mampu membedakan kata-kata antara kondisi user saat ini dengan hasil setelah pemakaian produk, sehingga masih ada kesalahan dalam prediksi.

BAB V

KESIMPULAN DAN SARAN

Bab ini akan mengulas tentang kesimpulan yang diperoleh dari penelitian ini berdasarkan hasil yang diperoleh dari analisis proses, algoritma, dan implementasi perancangan sistem yang telah dibangun. Saran akan memberikan catatan dan perbaikan yang diperlukan untuk penelitian ini untuk kedepannya.

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka diperoleh beberapa kesimpulan sebagai berikut:

- a. Penelitian ini menggunakan algoritma *Support Vector Machine* dengan kernel *linear* memiliki tingkat akurasi 86.9 %.
- b. Analisis sentimen dilakukan dengan cara melalui beberapa tahapan seperti tahap pengumpulan data, tahap pelabelan data, tahap *preprocessing*, tahap klasifikasi, dan pengujian.

5.2 Saran

Berdasarkan hasil analisis dan kesimpulan, berikut beberapa saran dari penelitian ini untuk pengembangan sistem lebih lanjut:

- a. Diharapkan pada penelitian selanjutnya, pada dataset yang digunakan baik jumlah data positif, jumlah data netral maupun jumlah data negatif memiliki perbandingan kelas yang seimbang.
- b. Diharapkan pada penelitian selanjutnya, Sistem mampu membedakan kata-kata antara kondisi user saat ini dengan hasil setelah pemakaian produk.
- c. Bagi peneliti selanjutnya, dapat menggunakan pendekatan *machine learning* lain sebagai pembanding performa algoritma *Support Vector Machine*.

DAFTAR PUSTAKA

- Casaló, L. V., Flavián, C., & Guinalú, M. (2011). Understanding the intention to follow the advice obtained in an online travel community. *Computers in Human Behavior*, 27(2), 622–633. <https://doi.org/10.1016/j.chb.2010.04.013>
- Fauzi, M. A. (2018). *Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion*. (September).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fikria, N. (2018). *ANALISIS KLASIFIKASI SENTIMEN REVIEW APLIKASI E-TICKETING MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN ASOSIASI*.
- Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, 56(3), 387–404. <https://doi.org/10.2501/IJMR-2014-025>
- Indrayuni, E. (2019). *Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes*. VII(1), 29–36.
- Kristiyanti, D. A. (2015). *ANALISIS SENTIMEN REVIEW PRODUK KOSMETIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN PARTICLE SWARM OPTIMIZATION SEBAGAI*.
- Murphy, R. (2019). Local Consumer Review Survey. Retrieved from <https://www.brightlocal.com/research/local-consumer-review-survey/>
- Rahayu, S., Kusriani, & Sismoro, H. (2018). *JURNAL INFORMATIKA DAN TEKNOLOGI INFORMASI PROGRAM STUDI TEKNIK INFORMATIKA – FAKULTAS TEKNIK - UNIVERSITAS JANABADRA*. 3(3).
- Utami, L. D. (2015). *Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naive Bayes*. 1(2), 120–126.
- Utami, P. D. (2018). *Analisis Sentimen Review Kosmetik Bahasa Indonesia Menggunakan Algoritma Naive Bayes*.
- Wardani, A. K. (2017). *PENGARUH ELECTRONIC WORD OF MOUTH PADA FORUM ONLINE*. 4(2), 1–15.
- Yulietha, I. M., Faraby, S. Al, & Adiwijaya. (2017). *KLASIFIKASI SENTIMEN REVIEW FILM*

MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE SENTIMENT CLASSIFICATION OF MOVIE REVIEWS. 4(3), 4740–4750.



LAMPIRAN

A. Halaman Website yang digunakan

<https://forum.femaledaily.com/showthread.php?6175-Aztec-Mask>

<https://forum.femaledaily.com/showthread.php?14034-Bedak-Saripohatji>

<https://forum.femaledaily.com/showthread.php?24107-L-Oreal-Micellar-Water>

<https://forum.femaledaily.com/showthread.php?23067-POND-S-Flawless-White->

[Dewy-Rose-Cream](https://forum.femaledaily.com/showthread.php?23067-POND-S-Flawless-White-Dewy-Rose-Cream)

<https://forum.femaledaily.com/showthread.php?3517-Kyusoku-Bihaku>

<https://forum.femaledaily.com/showthread.php?24488-Atasi-Kulit-Kusam->

[Membandel-dengan-Garnier-Sakura-White](https://forum.femaledaily.com/showthread.php?24488-Atasi-Kulit-Kusam-Membandel-dengan-Garnier-Sakura-White)

<https://forum.femaledaily.com/showthread.php?3208-Sariayu-Skincare>

<https://forum.femaledaily.com/showthread.php?23649-NIVEA-Creme>

<https://forum.femaledaily.com/showthread.php?21723-NIVEA-Make-Up-Clear>

<https://forum.femaledaily.com/showthread.php?23816-POND-S-White-Beauty>

<https://forum.femaledaily.com/showthread.php?6548-Oriflame-Skincare>

<https://forum.femaledaily.com/showthread.php?16527-The-Body-Shop-Moisture-white%99-shiso-2-in-1-brightening-eye-cream>

<https://forum.femaledaily.com/showthread.php?2383-Viva-Skincare>

[https://forum.femaledaily.com/showthread.php?20943-THE-BODY-SHOP-\(Face\)-Skincare-Part-2](https://forum.femaledaily.com/showthread.php?20943-THE-BODY-SHOP-(Face)-Skincare-Part-2)

B. Daftar kata *slang*, singkatan, salah penulisan kata (*typo*), bahasa asing, dan kata karakter berulang beserta perbaikannya.

abiss	habis
abissss	habis
abiz	habis
able	sanggup
about	tentang
ademm	adem
aer	air
ahhh	ah
ahahaa	tawa
agree	setuju
aje	saja
akhirny	akhirnya
alhamdulillah	alhamdulillah
answer	menjawab
apalgi	apalagi
bagusss	bagus
bahayaaa	bahaya
bajus	bagus
bajuss	bagus
bangett	banget
bangeeettt	banget
banged	banget
baruuuu	baru
basahin	membasahi
bbrp	beberapa
bbrpa	beberapa
becanda	bercanda
because	karena
become	menjadi
coz	karena
bete	kesal
betewe	ngomong-ngomong
bget	banget

bntr	bentar
bntu	bantu
bnyak	banyak
bnyaknya	banyaknya
bole	boleh
brangkat	berangkat
brangnya	barangnya
brati	berarti
brigheting	cerah
bright	cerah
brp	berapa
brsih	bersih
brtambah	bertambah
bruntus	beruntus
bruntusan	beruntusan
btw	ngomong-ngomong
bulann	bulan
buleh	boleh
bulmat	bulu mata
bumil	ibu hamil
bwa	bawa
bwat	buat
bwh	bawah
came	datang
cancel	membatalkan
cespleng	langsung
cewe	perempuan
cidder	cuka apel
cider	cuka apel
cma	cuma
cman	cuma
cmn	cuma
cmpur	campur

cocook	cocok
collagen	kolagen
color	warna
combo	kombinasi
come	datang
comedo	komedo
comes	datang
control	kontrol
controlnya	kontrolnya
cracking	retak
cream	krim
cpd	cepat
cpt	cepat
creamny	krimnya
creamnya	krimnya
critanya	cerita
crumble	hancur
cucok	cocok
cucoknya	cocoknya
cucook	cocok
cus	langsung
cuss	langsung
cussss	langsung
cz	karena
daaannn	dan
daan	dan
dapetny	dapatny
dapetnya	dapatnya
dapettttt	dapat
detil	detail
detoksnya	detoksnya
detox	detoks
diajakin	diajak
diakalin	diakali

dbahas	dibahas
dbales	dibalas
dbawa	dibawa
dbhas	dibahas
dbikin	dibuat
dbilas	dibilas
dng	dengan
dngan	dengan
dipijet	dipijat
disc	diskon
dkt	dekat
dl	dahulu
dlam	dalam
dlm	dalam
dlu	dahulu
doang	saja
doank	saja
donk	dong
doong	dong
doongg	dong
dpn	depan
dpt	dapat
dptnya	dapatnya
eike	aku
emgnya	memangnya
enaakk	enak
gaaaa	enggak
gaakk	enggak
gaje	enggak jelas
gabole	enggak boleh
gada	enggak ada
gadpt	enggak dapat
gituu	begitu
glowing	bersinar

glow	bersinar
glowy	bersinar
gmana	bagaimana
gmn	bagaimana
gmna	bagaimana
gmpng	mudah
gone	hilang
good	baik
gpp	enggak apa-apa
grasy	berminyak
greasy	berminyak
gtu	begitu
gtuh	begitu
gurls	perempuan
gw	aku
haaiii	hai
hadirrr	hadir
haha	tawa
halusssss	halus
haluus	halus
hayuk	ayo
hayuu	ayo
hayuuk	ayo
hehe	senyum
hiks	sedih
hits	terkenal
hitz	terkenal
hmpir	hampir
hmpr	hampir
hydrate	hidrasi
hydrating	hidrasi
infoin	informasikan
inpo	informasi
inpoh	informasi
now	sekarang

ingre	bahan
inih	ini
inii	ini
iy	iya
iyaa	iya
iyaaa	iya
iye	iya
iyess	iya
jerawian	jerawatan
jerewi	jerawat
jerewian	jerawatan
jerewinya	jerawatnya
jg	juga
jga	juga
jgn	jangan
jrawat	jerawat
jrawatan	jerawatan
jrawatnya	jerawatnya
jwb	jawab
jwban	jawaban
nice	bagus
nich	nih
kcuali	kecuali
kdang	kadang
kenyel	kenyal
keringg	kering
keringgg	kering
kite	kita
kl	kalau
krn	karena
okai	oke
ky	seperti
kya	seperti
love	suka
oilnya	minyaknya

C. Daftar *slangword* yang digunakan

semenjak	penuh	bekas_jerawat	campuran	lutut
komedo_putih	komedo_hitam	tercampur	siku	dipalsukan
jenisnya	sejenis	kadaluwarsanya	pekan	akhir
pelembap	berapa	berubah	terjadi	entah
jenisnya	tipe	paling	waduh	aduh
pemakaian	dibawa	kalinya	perasaan	pemakaian
bahkan	menurutku	apapun	bakal	kata
flek_hitam	sinar	makan	masing	kondisi
cuka_apel	kerasa	rasa	ditambah	teksturnya
segala	minyaknya	besokannya	pengguna	termasuk
pakainya	mineral	semacam	dikulitku	ponds
mengoles	macam	kuas	awalnya	dipakainya
sedang	kadang	warnanya	seluruh	besoknya
maskernya	maksudnya	kemasannya	produknya	serbuknya
walaupun	makanya	maka	bertahun	tahun
berbahan	bahannya	mencampur	baunya	kontrolnya
merasa	alhasil	orang	pipi	dagu
sekali	padahal	besok	namanya	bilang
sewajah	kakinya	badanku	kelihatan	oke
kulitnya	loreal	sangat	hampir	hari
minggu	hidden	content	efeknya	kok
kyusoku	bihaku	dong	sih	tawa
garnier	aztec	saripohatji	aku	lah
oriflame	nge	thread	kamu	banget
shiso	toner	lihat	nih	lebih
innisfree	serum	nya	terlalu	diwajah
sapaan	losion	sepertinya	sari_ayu	ter
rasanya	sa	jadi	serbuk	wajahnya
an	warna	ibu	krim	micellar
kan	loh	si	hai	nivea
jadinya	ternyata	deh	berasanya	menurutku
yah	veet	rose	iya	ah

