

**ANALISA SENTIMEN TERHADAP PRODUK HAIR
TREATMENT PADA PERUSAHAAN PT. INOVASI
TEKNOLOGI KOSMETIKA DENGAN
ALGORITMA NAIVE BAYES DAN
SUPPORT VECTOR MACHINE**



SKRIPSI

Diajukan untuk memenuhi salah satu syarat kelulusan Program Sarjana

MUHAMAD HIDJRIANSYAH

NIM: 19200570

**Program Studi Sistem Informasi
Fakultas Teknik dan Informatika
Universitas Bina Sarana Informatika
Jakarta
2025**

PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan dibawah ini:

Nama : Muhamad Hidjriansyah
NIM : 19200570
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa Skripsi / Tugas Akhir yang telah saya buat dengan judul: **"ANALISA SENTIMEN TERHADAP PRODUK HAIR TREATMENT PADA PERUSAHAAN PT. INOVASI TEKNOLOGI KOSMETIKA DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE"** adalah asli (orsinil) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya tanpa ada paksaan dari pihak manapun juga. Apabila di kemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa (Skripsi) yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Universitas Bina Sarana Informatika dicabut/dibatalkan.

Dibuat di : Jakarta
Pada Tanggal : 25 November 2024
Yang menyatakan



Muhamad Hidjriansyah

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang betanda tangan dibawah ini, penulis :

Nama : Muhamad Hidjriansyah
NIM : 19200570
Jenjang : Strata I (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa data atau informasi yang saya gunakan dalam penulisan karya ilmiah penulis dengan judul "**Analisa Sentimen Terhadap Produk Hair Treatment Pada Perusahaan Pt. Inovasi Teknologi Kosmetika Dengan Algoritma Naive Bayes Dan Support Vector Machine**" merupakan data dan atau informasi yang saya peroleh berdasarkan hasil Riset pada:

Nama Perusahaan : PT. Inovasi Teknologi Kosmetika

Alamat : Head Office: Royal Residence Blok B9/38, Harewood,
Surabaya.

Waktu Pelaksanaan: 28 Oktober 2024 sampai dengan 7 Januari 2025

Penulis menyetujui untuk memberikan ijin kepada pihak Universitas Bina Sarana Informatika untuk mendokumentasikan karya ilmiah saya tersebut secara internal dan terbatas, serta tidak untuk mengunggah karya ilmiah Penulis pada repository Universitas Bina Sarana Informatika.

Penulis bersedia untuk bertanggung jawab secara pribadi, tanpa melibatkan pihak Universitas Bina Sarana Informatika, atas materi/isi karya ilmiah tersebut, termasuk bertanggung jawab atas dampak atau kerugian yang timbul dalam bentuk akibat tindakan yang berkaitan dengan data dan atau informasi yang terdapat pada karya ilmiah saya ini.

Demikian kesepakatan ini dibuat dengan sebenarnya.

Dibuat di : Jakarta
Pada Tanggal : 25 November 2024

Penulis,



Muhamad Hidjriansyah

PERSETUJUAN DAN PENGESAHAN SKRIPSI

Skripsi ini diajukan oleh:

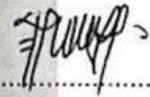
Nama : Muhamad Hidjriansyah
NIM : 19200570
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika
Judul Skripsi : Analisa Sentimen Terhadap Produk Hair Treatment Pada Perusahaan PT Inovasi Teknologi Kosmetika Dengan Algoritma Naive Bayes Dan Support Vector Machine

Telah dipertahankan pada periode 2024-2 dihadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Sarjana Komputer (S.Kom) pada Program Sarjana (S1) Program Studi Sistem Informasi di Universitas Bina Sarana Informatika.

Jakarta, 07 Februari 2025

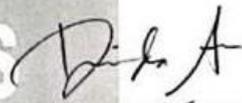
PEMBIMBING SKRIPSI

Pembimbing I : Ginabila, M.Kom.



DEWAN PENGUJI

Penguji I : Dinda Ayu Muthia, M.Kom.



Penguji II : Agustiena Merdekawati, S.T.,
M.Kom.



SURAT PERNYATAAN KEBENARAN/KEABSAHAN DATA HASIL RISET UNTUK KARYA ILMIAH

Yang bertanda tangan di bawah ini, saya:

Nama : Muhamad Hidjriansyah
NIM : 19200570
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa data dan atau informasi yang saya gunakan dalam penulisan karya ilmiah dengan judul **“Analisa Sentimen Terhadap Produk Hair Treatment Pada Perusahaan Pt. Inovasi Teknologi Kosmetika Dengan Algoritma Naive Bayes Dan Support Vector Machine”** merupakan data dan atau informasi yang saya peroleh melalui hasil penelitian sendiri dan tidak didasarkan pada data atau informasi hasil riset dari perusahaan/instansi/lembaga manapun.

Saya bersedia untuk bertanggung jawab secara pribadi, tanpa melibatkan pihak **Universitas Bina Sarana Informatika**, atas materi/isi karya ilmiah tersebut, termasuk bertanggung jawab atas dampak atau kerugian yang timbul dalam bentuk akibat tindakan yang berkaitan dengan data dan atau informasi yang terdapat pada karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal: 7 Januari 2025

Mengetahui,

Yang menyatakan,

Dosen Pembimbing



Ginabila, M.Kom



Muhamad Hidjriansyah

LEMBAR KONSULTASI SKRIPSI DOSEN PEMBIMBING

	LEMBAR KONSULTASI SKRIPSI
	UNIVERSITAS BINA SARANA INFORMATIKA

NIM : 19200570
Nama Lengkap : Muhamad Hidjriansyah
Dosen Pembimbing : Ginabila, M.Kom
Judul Skripsi : Analisa Sentimen Terhadap Produk Hair Treatment Pada Perusahaan Pt. Inovasi Teknologi Kosmetika Dengan Algoritma Naive Bayes Dan Support Vector Machine

No	Tanggal Bimbingan	Pokok Pembahasan	Paraf Dosen Pembimbing
1	28 Oktober 2024	Bimbingan Perdana dan Penyerahan Judul	
2	4 November 2024	Memberikan arahan dalam penulisan BAB I	
3	11 November 2024	Revisi BAB I terkait Latar Belakang, Tujuan Penelitian, Penelitian Terdahulu, Metode Usulan, dan SOTA	
4	18 November 2024	Memberikan Arahan dalam penulisan BAB II terkait Kerangka Penelitian	
5	2 Desember 2024	Meninjau isi hasil pengerjaan dan teknik penulisan pada BAB II	
6	13 Desember 2024	Memberikan arahan dalam penulisan BAB III	
7	23 Desember 2024	Memberikan arahan dalam penulisan BAB IV	
8	7 Januari 2025	Meninjau hasil pengerjaan skripsi dari BAB I-IV	

Catatan untuk Dosen Pembimbing Bimbingan Skripsi

Dimulai pada tanggal : 28 Oktober 2024
 Diakhiri pada tanggal : 7 Januari 2025
 Jumlah pertemuan bimbingan : 8 Pertemuan

Disetujui oleh,
Dosen Pembimbing



(Ginabila, M.Kom)

PEDOMAN PENGGUNAAN HAK CIPTA

Skripsi sarjana yang berjudul “ANALISA SENTIMEN TERHADAP PRODUK HAIR TREATMENT PADA PERUSAHAAN PT. INOVASI TEKNOLOGI KOSMETIKA DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE” adalah hasil karya tulis asli Nama Mahasiswa dan bukan hasil terbitan sehingga peredaran karya tulis hanya berlaku di lingkungan akademik saja, serta memiliki hak cipta. Oleh karena itu, dilarang keras untuk mengganggakan baik sebagian maupun seluruhnya karya tulis ini, tanpa seizin penulis.

Referensi kepustakaan diperkenankan untuk dicatat tetapi pengutipan atau peringkasan isi tulisan hanya dapat dilakukan dengan seizin penulis dan disertai ketentuan pengutipan secara ilmiah dengan menyebutkan sumbernya. Untuk keperluan perizinan pada pemilik dapat menghubungi informasi yang tertera di bawah ini.

Nama : Muhamad Hidjriansyah
Alamat : Cluster Royal Living Blok A6 No.8
No-Telp : 08989055518
Email : Mhidjriansyah20@gmail.com



UNIVERSITAS

LEMBAR PERSEMBAHAN

“Pendidikan adalah senjata paling ampuh yang bisa digunakan untuk mengubah dunia (Nelson Mandela)”

Alhamdulillahirobbil’alamin, dengan mengucapkan puji syukur kepada Allah S.W.T,
skripsi ini kupersembahkan untuk:

1. Kedua orang tua saya tercinta Bapak dan Ibu yang selalu melangitkan doa-doa baik dan menjadikan motivasi untuk saya dalam menyelesaikan skripsi ini. Terimakasih sudah mengantarkan saya di titik ini, saya persembahkan karya tulis sederhana ini dan gelar untuk Bapak dan Ibu.
2. Segenap civitas akademika kampus Universitas Bina Sarana Informatika, staf pengajar, karyawan, dan seluruh mahasiswa semoga tetap semangat dalam beraktivitas mengisi hari-harinya di kampus Universitas Bina Sarana Informatika.
3. Teman-teman penulis baik itu teman kuliah seangkatan, adik tingkat, kakak tingkat pada Fakultas Teknik dan Informatika Universitas Bina Sarana Informatika, maupun teman-teman dari fakultas dan universitas lain yang telah banyak memberi masukan, semangat dan arahan hingga akhirnya dapat terselesaikan Skripsi ini.

KATA PENGANTAR

Kami bersyukur kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya yang memungkinkan penulis menyelesaikan tugas ini dengan sukses. Karya ilmiah ini, yang merupakan bagian dari program sarjana, kami sajikan dalam format buku yang sederhana. Judul dari skripsi ini adalah sebagai berikut,

”Analisa Sentimen Terhadap Produk Hair Treatment Pada Perusahaan Pt. Inovasi Teknologi Kosmetika Dengan Algoritma Naive Bayes Dan Support Vector Machine”.

Tujuan penulisan tugas akhir ini, sebagai bagian dari persyaratan kelulusan dari Program Sarjana/Tugas Akhir dari Program Diploma Tiga di Universitas Bina Sarana Informatika, adalah untuk menghasilkan sebuah karya ilmiah yang didasarkan pada hasil penelitian eksperimen, observasi, dan referensi dari literatur yang relevan. Penulis mengakui bahwa tanpa bimbingan dan dukungan dari berbagai pihak, penulisan tugas akhir ini tidak akan terlaksana dengan baik. Oleh karena itu, pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada:

1. Rektor Universitas Bina Sarana Informatika,
2. Dekan Fakultas Teknik dan Informatika,
3. Ketua Program Studi Sistem Informasi Universitas Bina Sarana Informatika.
4. Ibu Ginabila M.kom selaku Dosen Pembimbing
5. Keluarga Tercinta yang slalu support

Serta semua pihak yang terlalu banyak untuk disebut satu Penulis menyadari bahwa penulisan ini belum mencapai kesempurnaan mutlak. Oleh karena itu, kami mengharapkan masukan dan saran yang membangun untuk meningkatkan kemampuan penulisan kami di masa depan. Akhir kata, semoga karya ini bermanfaat bagi penulis dan pembaca yang tertarik dengan topik ini secara umum.

Jakarta,25,November 2025

Penulis



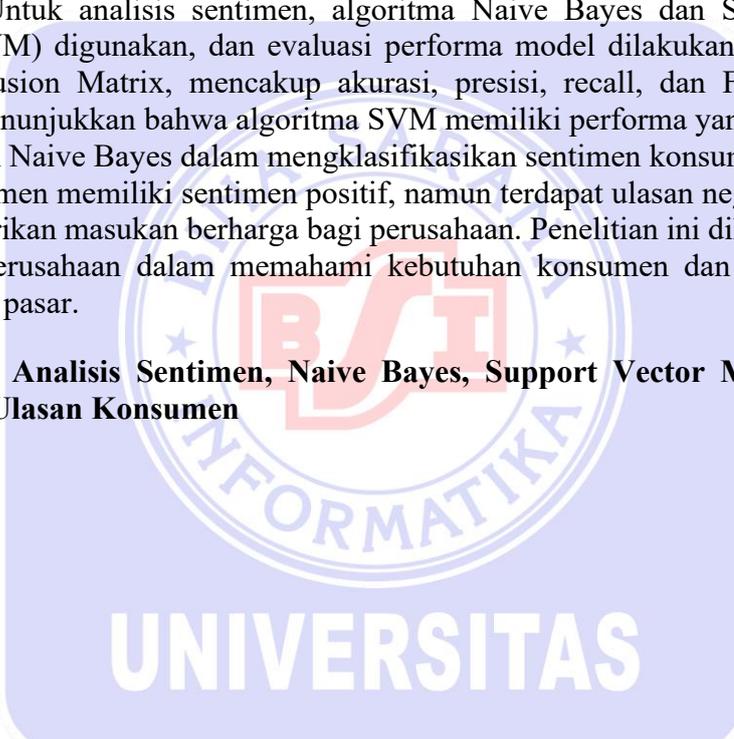
Muhamad Hidjriansyah

ABSTRAK

Muhamad Hidjriansyah (19200570) Analisa Sentimen Terhadap Produk Hair Treatment Pada Perusahaan Pt. Inovasi Teknologi Kosmetika Dengan Algoritma Naive Bayes Dan Support Machine

Persaingan di industri kosmetik yang semakin ketat mendorong perusahaan untuk memahami sentimen konsumen terhadap produk yang mereka tawarkan. Penelitian ini bertujuan untuk menganalisis sentimen konsumen terhadap produk hair treatment PT. Inovasi Teknologi Kosmetika, yang diharapkan dapat memberikan wawasan penting dalam meningkatkan kualitas produk dan strategi pemasaran. Metode yang digunakan dalam penelitian ini melibatkan pengumpulan data ulasan konsumen dari berbagai platform digital. Data yang diperoleh kemudian diproses melalui tahapan preprocessing, seperti pembersihan data, tokenisasi, case folding, dan penghapusan stopwords. Untuk analisis sentimen, algoritma Naive Bayes dan Support Vector Machine (SVM) digunakan, dan evaluasi performa model dilakukan menggunakan metrik Confusion Matrix, mencakup akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa yang lebih unggul dibandingkan Naive Bayes dalam mengklasifikasikan sentimen konsumen. Mayoritas ulasan konsumen memiliki sentimen positif, namun terdapat ulasan negatif dan netral yang memberikan masukan berharga bagi perusahaan. Penelitian ini diharapkan dapat membantu perusahaan dalam memahami kebutuhan konsumen dan meningkatkan daya saing di pasar.

Kata kunci: Analisis Sentimen, Naive Bayes, Support Vector Machine, Hair Treatment, Ulasan Konsumen



UNIVERSITAS

ABSTRACT

Muhamad Hidjriansyah (19200570) *Sentiment Analysis on Hair Treatment Products at PT. Inovasi Teknologi Kosmetika Using Naive Bayes and Support Vector Machine Algorithms*

The increasing competition in the cosmetics industry drives companies to understand consumer sentiment toward their offered products. This study aims to analyze consumer sentiment regarding the hair treatment products of PT. Inovasi Teknologi Kosmetika, providing valuable insights to improve product quality and marketing strategies. The methodology involves collecting consumer review data from various digital platforms. The data underwent preprocessing steps, including data cleansing, tokenization, case folding, and stopword removal. For sentiment analysis, the Naive Bayes and Support Vector Machine (SVM) algorithms were employed, with model performance evaluated using the Confusion Matrix metrics, including accuracy, precision, recall, and F1-score. The results revealed that the SVM algorithm outperformed Naive Bayes in classifying consumer sentiment. Most consumer reviews expressed positive sentiments, with a smaller portion of neutral and negative reviews offering constructive feedback for the company. This study is expected to assist the company in understanding consumer needs and enhancing competitiveness in the market.

Keywords: Sentiment Analysis, Naive Bayes, Support Vector Machine, Hair Treatment, Consumer Reviews



UNIVERSITAS

DAFTAR ISI

LEMBAR JUDUL SKRIPSI.....	i
PERNYATAAN KEASLIAN SKRIPSI.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
PERSETUJUAN DAN PENGESAHAN SKRIPSI	iv
SURAT PERNYATAAN KEBENARAN/KEABSAHAN DATA HASIL RISET UNTUK KARYA ILMIAH	v
LEMBAR KONSULTASI SKRIPSI DOSEN PEMBIMBING	vi
PEDOMAN PENGGUNAAN HAK CIPTA	vii
LEMBAR PERSEMBAHAN	viii
KATA PENGANTAR.....	ix
ABSTRAK	x
<i>ABSTRACT</i>	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Tujuan Penelitian	3
1.3. Penelitian Terdahulu.....	4
1.4. Metode Usulan	5
1.5. <i>State of the Art</i> (SOTA)	7
BAB II TAHAPAN PENELITIAN.....	11
2.1. Kerangka Penelitian.....	11
2.2. Tahapan Penelitian.....	12
BAB III.....	15
HASIL DAN PEMBAHASAN	15
3.1. Pengumpulan Data	15
3.1.1. Studi Literatur	17
3.1.2. <i>Crawling Data</i>	17
3.2. <i>Preprocessing Data</i>	19
3.2.1. <i>Cleansing</i>	19
3.2.2. <i>Tokenizing</i>	20

3.2.3. <i>Case Folding</i>	22
3.2.4. <i>Stopword Removal</i>	23
3.3. Pengujian Model	25
3.4. Analisis Keluaran Confusion Matrix	29
BAB IV	32
PENUTUP	32
4.1. Kesimpulan	32
4.2. Saran	33
DAFTAR PUSTAKA	34
DAFTAR RIWAYAT HIDUP	36
SURAT KETERANGAN RISET	37
SURAT KETERANGAN PERUSAHAAN	38
BUKTI HASIL PENGECEKAN PLAGIARISME	39
LAMPIRAN-LAMPIRAN	40
DOKUMENTASI	42



DAFTAR GAMBAR

Gambar I 1 Kerangka Penelitian.....	11
Gambar III 1 Process Cleansing.....	20
Gambar III 2 Hasil Cleansing	20
Gambar III 3 Process Documents from Data	21
Gambar III 4 Hasil Tokenizing	21
Gambar III 5 Hasil Case Folding	23
Gambar III 6 Hasil Stpword Removal	25
Gambar III 7 Pemodelan Naïve Bayes.....	26
Gambar III 8 Hasil Uji menggunakan metode Naïve Bayes.....	26
Gambar III 9 Hasil Analisis confusion matrix dengan Naïve Bayes	27
Gambar III 10 Pemodelan Support Vector Machine dengan operator cross validation.....	28
Gambar III 11 Hasil Uji menggunakan metode Support Vector Machine.....	28
Gambar III 12 Hasil Confusion Matrix dengan Support Vector Machine.....	29
Gambar III 13 Hasil Perbandingan Confusin Matrix	30



UNIVERSITAS

DAFTAR TABEL

Tabel 1. 1 Penelitian Terdahulu	5
Tabel 1. 2 State of Art	9
Tabel 3. 1 Data Ulasan Negatif Produk	16
Tabel 3. 2 Data Ulasan Positif Produk	16
Tabel 3. 3 Crawling Data	18



DAFTAR LAMPIRAN

Lampiran Daftar Riwayat Hidup.....	36
Lampiran Surat Keterangan Perusahaan.....	37
Lampiran Surat Keterangan Perusahaan.....	38
Lampiran Bukti Hasil Pengecekan Plagiarisme	39
Lampiran-Lampiran Data Ulasan Positif	40
Lampiran-Lampiran Data Ulasan Negatif.....	41
Dokumentasi	42



BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini, industri kecantikan dan perawatan pribadi, khususnya produk hair treatment, merupakan salah satu sektor yang berkembang pesat di Indonesia. Dalam beberapa tahun terakhir, permintaan terhadap produk perawatan rambut semakin meningkat, seiring dengan kesadaran masyarakat akan pentingnya menjaga kesehatan dan kecantikan rambut. Hal ini menyebabkan semakin banyak perusahaan yang berlomba-lomba mengembangkan dan memasarkan produk mereka.

Dalam penelitian sebelumnya, metode Naive Bayes dan Support Vector Machine (SVM) sering digunakan untuk analisis sentimen karena performa yang baik dalam klasifikasi teks. Penelitian oleh (Sari et al., 2020) menunjukkan bahwa “Metode Naive Bayes memiliki akurasi 80,45% dalam klasifikasi sentimen terhadap ulasan produk kecantikan melalui platform online”. Sementara itu, penelitian serupa oleh (Suhendra et al., 2021) mengungkapkan bahwa “Kombinasi Naive Bayes dan teknik TF-IDF efektif untuk mengekstraksi pola sentimen positif dan negatif dalam ulasan pelanggan.”

Meskipun demikian, tantangan utama dalam analisis sentimen pada industri kecantikan adalah keragaman bahasa dan ekspresi pelanggan. “Pengolahan bahasa alami (NLP) dan *stemming* bahasa Indonesia diperlukan untuk meningkatkan akurasi klasifikasi sentimen. Oleh karena itu, pemilihan metode yang tepat, seperti *Naive Bayes* dan *Support Vector Machine*, menjadi kunci dalam menangkap opini pelanggan secara akurat dan membantu perusahaan dalam perencanaan strategi produk”

(Kurniawan et al., 2023)

PT. Inovasi Teknologi Kosmetika (INTEKO) adalah salah satu perusahaan yang bergerak di bidang pembuatan dan distribusi produk kosmetik dan perawatan rambut. Untuk tetap bersaing di pasar yang sangat kompetitif ini, perusahaan harus memastikan bahwa produk yang mereka tawarkan memenuhi harapan dan kebutuhan konsumen. Salah satu cara untuk mengevaluasi kualitas produk secara efektif adalah melalui analisis umpan balik dari pelanggan, yang umumnya tersedia dalam bentuk review atau ulasan di berbagai platform online seperti e-commerce, media sosial, atau situs web perusahaan.

Namun, ulasan pelanggan yang bersifat teks sangat bervariasi dan dapat mengandung banyak nuansa, sehingga sulit untuk diproses dan dianalisis secara manual. Oleh karena itu, perusahaan perlu memanfaatkan teknologi analisis sentimen untuk menilai opini dan persepsi pelanggan terhadap produk hair treatment mereka. Analisis sentimen ini akan membantu perusahaan untuk lebih memahami kualitas produk, mengidentifikasi kekuatan dan kelemahan, serta meningkatkan pengembangan produk di masa depan.

Meskipun analisis sentimen sudah banyak dilakukan pada berbagai sektor, tantangan utama yang dihadapi dalam industri kecantikan adalah keragaman bahasa dan ekspresi yang digunakan oleh konsumen dalam memberikan umpan balik, baik dalam bentuk positif, negatif, atau netral. Oleh karena itu, penting untuk memilih metode yang tepat dalam analisis sentimen untuk memperoleh hasil yang akurat.

1.2. Tujuan Penelitian

Berikut adalah tujuan penelitian dalam penelitian ini:

1. Untuk melakukan analisis sentimen terhadap ulasan produk hair treatment dari PT. Inovasi Teknologi Kosmetika menggunakan dua algoritma klasifikasi teks, yaitu Naive Bayes dan Support Vector Machine (SVM), untuk mengidentifikasi sentimen positif, negatif, dan netral dari konsumen. Untuk menilai peran verifikasi lapangan dalam meningkatkan ketepatan sasaran penyaluran bantuan sosial di wilayah Kapuk. Tujuan ini mengkaji dampak survei langsung terhadap validitas data dan penentuan penerima manfaat yang sesuai dengan kriteria program.
2. Untuk membandingkan performa algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen ulasan pelanggan, dengan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score, guna menentukan metode terbaik untuk analisis sentimen.
3. Untuk mengidentifikasi persepsi pelanggan terhadap produk hair treatment melalui hasil analisis sentimen, untuk memberikan wawasan yang dapat digunakan perusahaan dalam meningkatkan kualitas produk, memahami kekuatan dan kelemahan produk, serta mendukung pengambilan keputusan bisnis di masa depan.

1.3. Penelitian Terdahulu

Penelitian sebelumnya menunjukkan penerapan algoritma Naive Bayes untuk analisis sentimen di berbagai bidang, namun belum ada yang secara khusus meneliti produk kecantikan, terutama hair treatment. Aditya Rizky dkk. (2021) melakukan analisis sentimen terhadap e-commerce, dan meskipun menggunakan Naive Bayes, hasilnya hanya mencapai akurasi 68% karena kualitas data yang kurang baik. Syarifuddin dkk. (2022) juga menggunakan algoritma yang sama untuk analisis sentimen di sektor transportasi online, dengan hasil akurasi 72%. Namun, mereka mengalami masalah dengan noise pada data. Nurnalasari dkk. (2020) menerapkan Naive Bayes untuk menganalisis sentimen terkait program pemerintah, dengan akurasi 75%, meskipun kesulitan dalam mengklasifikasikan sentimen netral. Riska T. & Mailoa (2023) mendapat akurasi tinggi (93,4%) dalam analisis sentimen mengenai isu resesi ekonomi di Twitter, namun hasilnya kurang relevan untuk sektor kecantikan. Penelitian oleh Azhar et al. (2024) tentang program makan siang gratis menunjukkan akurasi hanya 70% karena masalah data yang tidak seimbang.

Meskipun banyak penelitian menggunakan Naive Bayes, penerapannya dalam analisis sentimen terhadap produk kecantikan, khususnya produk hair treatment, masih terbatas. Penelitian ini bertujuan untuk mengisi kekosongan tersebut, dengan fokus pada pengolahan data teks ulasan produk hair treatment dan membandingkan hasilnya dengan algoritma lain seperti Support Vector Machine (SVM).

Tabel I. 1.

Tabel Penelitian terdahulu

Penulis	Judul Penelitian	Tahun	Metode	Hasil
Aditya Rizky dkk	Analisis Sentimen Pengguna E-	2021	Naive Bayes	Akurasi hanya mencapai

	Commerce Menggunakan Naive Bayes			68% dikarenakan dataset tidak bersih.
Syarifuddin	Analisis Sentimen Pelanggan Transportasi Online Menggunakan Algoritma Naive Bayes	2022	Naive Bayes	Akurasi 72% akibat <i>noise</i> pada data dan kurangnya preprocessing.
Nurmalasari dkk	Analisis Sentimen Program Pemerintah pada Platform X Menggunakan Naive Bayes	2020	Naive Bayes	Akurasi 75% dengan kesulitan dalam menentukan sentimen netral.
Ribka Tesalonika & Mailoa	Implementasi Naive Bayes untuk Analisis Sentimen Isu Resesi Ekonomi di Twitter	2023	Naive Bayes	Akurasi 93.4% , tetapi dapat dianalisis aspek dataset untuk revisi.
Azhar et al.	mplementasi Naive Bayes untuk Sentimen Program Makan Siang Gratis di Platform X	2024	Naive Bayes	Akurasi rendah sekitar 70% , karena adanya data <i>imbalanced</i> .

Sumber: Hasil Penelitian

1.4. Metode Usulan

Metode yang akan digunakan dalam penelitian ini dimulai dengan tahap pengumpulan data. Data yang digunakan adalah ulasan atau review dari konsumen yang ada pada berbagai platform online seperti e-commerce, media sosial, atau situs web yang berhubungan dengan produk hair treatment dari PT. Inovasi Teknologi

Kosmetika. Data tersebut akan mencakup sentimen positif, negatif, dan netral yang diberikan oleh pengguna terkait produk yang digunakan.

Setelah data terkumpul, tahap berikutnya adalah proses pembersihan data atau data preprocessing. Pada tahap ini, teks dari ulasan pelanggan akan disaring dan dibersihkan dari kata-kata yang tidak relevan seperti simbol, angka, dan kata umum yang tidak memberikan informasi penting. Selanjutnya, dilakukan pemisahan kata atau tokenization, yang bertujuan untuk membagi kalimat menjadi unit kata yang lebih kecil. Kemudian, langkah selanjutnya adalah menghapus kata-kata yang tidak terlalu penting, atau yang disebut sebagai stopwords, seperti "dan", "yang", dan sebagainya, yang tidak memiliki kontribusi besar dalam analisis sentimen. Setelah itu, data akan dikonversi ke dalam bentuk yang lebih mudah untuk dianalisis, menggunakan teknik seperti vectorization, di mana setiap kata atau fitur diberikan nilai numerik berdasarkan frekuensinya dalam data.

Setelah data siap, langkah selanjutnya adalah menerapkan dua algoritma pembelajaran mesin, yaitu Naive Bayes dan Support Vector Machine (SVM). Algoritma Naive Bayes dipilih karena kemampuannya dalam menangani klasifikasi berbasis probabilitas pada data teks, sementara SVM dipilih karena efektivitasnya dalam membedakan berbagai kelas data dengan margin yang jelas. Kedua algoritma ini akan dilatih menggunakan dataset yang telah diproses dan diuji menggunakan teknik cross-validation untuk memastikan model yang dibangun memiliki kemampuan generalisasi yang baik.

Setelah model dilatih, tahap selanjutnya adalah evaluasi hasil. Evaluasi akan dilakukan dengan menggunakan beberapa metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur performa masing-masing model. Hasil dari kedua algoritma

ini akan dibandingkan untuk menentukan model mana yang lebih akurat dalam menganalisis sentimen terhadap produk hair treatment yang dianalisis. Hasil yang didapat akan memberikan gambaran yang lebih jelas mengenai persepsi pelanggan terhadap produk tersebut, serta memberikan masukan yang berguna bagi perusahaan dalam meningkatkan kualitas produk dan strategi pemasaran mereka.

Dengan menggunakan metode ini, diharapkan penelitian ini dapat memberikan hasil yang tidak hanya akurat dalam menganalisis sentimen, tetapi juga relevan dan berguna bagi PT. Inovasi Teknologi Kosmetika dalam memperbaiki dan mengembangkan produk mereka sesuai dengan keinginan dan kebutuhan konsumen.

1.5. *State of the Art* (SOTA)

State of the Art (SOTA) adalah tahap awal yang penting dalam sebuah penelitian. Ini mengacu pada pengetahuan, penelitian, dan teknologi paling mutakhir, terbaru, dan terdepan dalam bidang tertentu. Dengan kata lain, SOTA mewakili pemahaman dan kemajuan terkini dalam suatu bidang tertentu. Memahami *State of the Art* sangat penting bagi peneliti, praktisi, dan pembuat keputusan. Ini memberikan tolok ukur untuk membandingkan perkembangan baru dan menilai kemajuan domain tertentu.

Dalam beberapa tahun terakhir, analisis sentimen telah menjadi topik yang sangat penting dalam bidang ilmu data dan pemrosesan bahasa alami (NLP), khususnya untuk mengkaji opini dan persepsi konsumen terhadap produk melalui ulasan atau review. Penelitian terdahulu menunjukkan bahwa penggunaan algoritma pembelajaran mesin, seperti Naive Bayes (NB), secara konsisten memberikan hasil

yang baik dalam klasifikasi sentimen teks, baik pada data ulasan produk e-commerce, Twitter, maupun produk kecantikan.

Berikut adalah tabel State of the Art (SOTA) yang merangkum referensi terkait penelitian ini:

Tabel I. 2.
Tabel State of Art (SOTA)

No	Judul Referensi	Tujuan Penelitian	Data Penelitian	Metode Penelitian	Hasil Penelitian	Perbandingan
1	Analisis Sentimen Pengguna E-Commerce Menggunakan Naive Bayes (Aditya Rizky dkk, 2021)	Menganalisis sentimen pengguna terhadap platform e-commerce	Dataset ulasan pengguna e-commerce	Naive Bayes	Akurasi mencapai 68% karena dataset yang digunakan tidak bersih	Fokus pada e-commerce, sementara penelitian ini fokus pada produk hair treatment di industri kecantikan dengan data yang lebih bersih dan analisis lebih mendalam.
2	Analisis Sentimen Pelanggan Transportasi Online Menggunakan Algoritma Naive Bayes (Syarifuddin, 2022)	Mengukur sentimen pelanggan terhadap layanan transportasi online	Ulasan pelanggan layanan transportasi online	Naive Bayes	Akurasi 72% disebabkan adanya <i>noise</i> pada data dan kurangnya preprocessing	Penelitian ini berbeda karena fokus pada sektor transportasi online, bukan produk kecantikan. Penelitian ini juga menggunakan algoritma yang sama,

						tetapi untuk sektor yang berbeda.
3	Analisis Sentimen Program Pemerintah pada Platform X Menggunakan Naive Bayes (Nurnalasari dkk, 2020)	Menganalisis persepsi publik terhadap program pemerintah	Data ulasan pengguna pada Platform X	Naive Bayes	Akurasi 75% dengan kesulitan dalam menentukan sentimen netral	Penelitian ini menganalisis sentimen terhadap program pemerintah, bukan produk kecantikan. Kesulitan dalam sentimen netral juga akan dihadapi dalam penelitian ini.
4	Implementasi Naive Bayes untuk Analisis Sentimen Isu Resesi Ekonomi di Twitter (Ribka Tsalonika & Mailoa, 2023)	Menganalisis sentimen terkait isu resesi ekonomi berdasarkan <i>tweet</i> pengguna	Data <i>tweet</i> dari Twitter	Naive Bayes	Akurasi mencapai 93.4% , namun terdapat aspek dataset yang perlu direvisi	Fokus pada isu ekonomi, bukan produk kecantikan. Penelitian ini dapat memberikan insight dalam penanganan data besar, yang juga relevan untuk penelitian ini.
5	Implementasi Naive Bayes untuk Sentimen Program Makan Siang Gratis di Platform X (Azhar et al., 2024)	Menilai sentimen terhadap program pemerintah dalam isu makan siang gratis	Data ulasan terkait program makan siang gratis	Naive Bayes	Akurasi sekitar 70% karena adanya data <i>imbalanced</i>	Menggunakan dataset yang <i>imbalanced</i> , yang juga bisa menjadi tantangan dalam penelitian ini, terutama jika

						data ulasan tidak seimbang.
--	--	--	--	--	--	-----------------------------

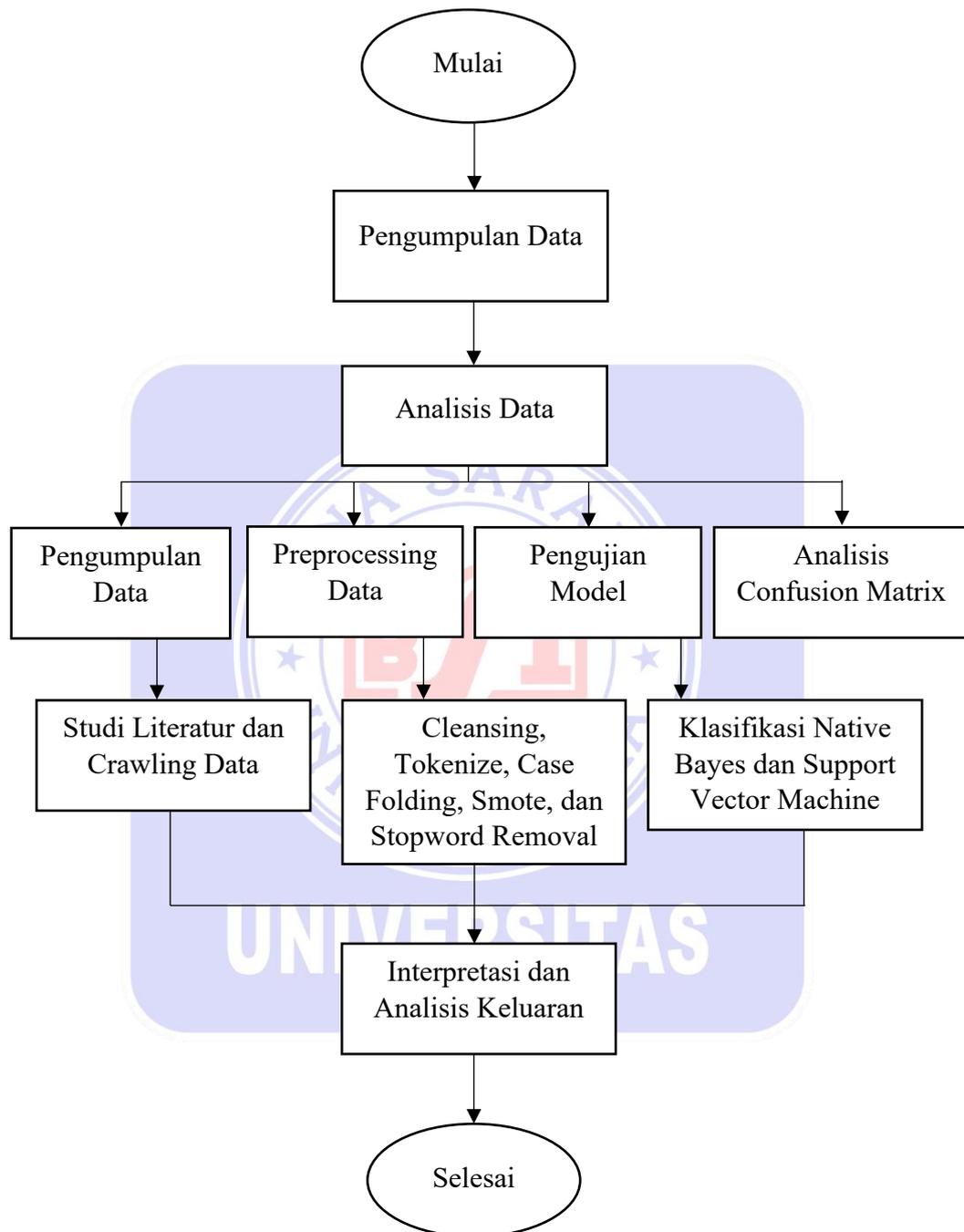
Sumber: Hasil Penelitian



BAB II

TAHAPAN PENELITIAN

2.1. Kerangka Penelitian



Sumber: Hasil Penelitian

Gambar I 1 Kerangka Penelitian

2.2. Tahapan Penelitian

1. Pengumpulan Data

Pengumpulan data merupakan langkah awal yang sangat penting dalam penelitian ini. Data yang digunakan berasal dari berbagai sumber, termasuk platform *e-commerce* dan situs ulasan produk. Proses pengumpulan dilakukan melalui studi literatur dan crawling data.

2. Analisis Data

Pada tahap ini, data yang dikumpulkan akan diproses lebih lanjut agar siap dianalisis. Ada beberapa langkah penting yang dilakukan, antara lain:

a) Pengumpulan Data

Data dikumpulkan dari berbagai sumber seperti *e-commerce*, media sosial, atau platform ulasan lainnya. Teknik yang digunakan antara lain *studi literatur* dan *crawling data* menggunakan tools atau software pendukung seperti Python.

b) Preprocessing Data

Proses ini meliputi:

- 1) *Cleansing*: Membersihkan data dari simbol, angka, atau karakter yang tidak relevan.
- 2) *Tokenize*: Memecah kalimat menjadi kata-kata.
- 3) *Case Folding*: Mengubah teks menjadi huruf kecil semua.
- 4) *Stopword Removal*: Menghapus kata-kata umum yang tidak memiliki pengaruh signifikan terhadap analisis seperti “dan”, “di”, “ke”.
- 5) *Smote*: Menangani ketidakseimbangan kelas dalam dataset.

3. Pengujian Model

Setelah preprocessing selesai, langkah selanjutnya adalah menerapkan algoritma Naive Bayes dan Support Vector Machine (SVM) untuk melakukan analisis sentimen. Pada tahap ini:

a) Klasifikasi Sentimen

Data yang telah dibersihkan akan diklasifikasikan menjadi sentimen positif, negatif, atau netral. Algoritma Naive Bayes dan SVM dipilih karena performanya yang baik untuk klasifikasi teks.

b) Pengujian Model

Model diuji menggunakan data latih (training data) dan data uji (testing data). Tahapan ini bertujuan untuk memastikan akurasi model dalam memprediksi sentimen.

4. Analisis Confusion Matrix

Setelah pengujian, hasil analisis dievaluasi menggunakan *confusion matrix*. Confusion matrix adalah metode evaluasi yang berguna untuk mengetahui performa model klasifikasi. Dalam hal ini, akan dihitung:

- a) Akurasi: Seberapa tepat prediksi model terhadap sentimen sebenarnya.
- b) Precision: Tingkat ketepatan model dalam memprediksi sentimen positif atau negatif.
- c) Recall: Seberapa banyak sentimen yang benar-benar terklasifikasi.

Tahap ini menjadi penentu apakah metode Naive Bayes dan SVM berhasil memberikan akurasi yang memuaskan.

5. Interpretasi dan Analisis Keluarannya

Setelah hasil evaluasi didapatkan, tahap selanjutnya adalah **interpretasi data**.

Pada tahap ini, penelitian ini akan membahas hasil analisis sentimen secara mendalam. Beberapa hal yang dijelaskan meliputi:

- a) Tren sentimen positif, negatif, dan netral dari ulasan pelanggan.
- b) Perbandingan akurasi antara algoritma Naive Bayes dan SVM.
- c) Identifikasi aspek yang sering dikritik atau dipuji oleh pelanggan.

Interpretasi ini akan membantu memberikan gambaran menyeluruh tentang opini pelanggan terhadap produk *hair treatment* yang dianalisis.



BAB III

HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pengumpulan data merupakan langkah awal yang sangat penting dalam penelitian ini. Data yang digunakan berasal dari berbagai sumber, termasuk platform *e-commerce* dan situs ulasan produk. Proses pengumpulan dilakukan melalui studi literatur dan crawling data.

Pada tahap ini, data yang digunakan untuk analisis sentimen terhadap produk Hair Treatment PT. Inovasi Teknologi Kosmetika dikumpulkan dari dataset yang terdiri atas dua jenis ulasan, yaitu **Data Ulasan Negatif** dan **Data Ulasan Positif**. Dataset ini berisi total **3, 220 entri**, yang mencakup berbagai tanggapan pelanggan terhadap produk Hair Treatment. Data ulasan negatif memberikan informasi terkait komplain pelanggan, seperti masalah kualitas produk, pengiriman yang salah, hingga dampak buruk terhadap rambut setelah penggunaan produk. Sementara itu, data ulasan positif mencakup komentar pelanggan yang puas terhadap produk, seperti pujian terhadap kualitas produk, pelayanan yang cepat, serta pengalaman penggunaan yang memuaskan.

Setiap entri dalam dataset memiliki beberapa atribut utama, seperti tanggal pesanan, tanggal komplain, nomor pesanan, nama produk, dan detail ulasan. Pada data ulasan negatif, terdapat pula kolom kategori masalah yang menjelaskan jenis keluhan secara lebih spesifik, seperti "Produk Salah Kirim" atau "Rambut Rontok". Selain itu, data ini juga menyertakan bukti pendukung berupa foto atau dokumen lain yang relevan.

Struktur Dataset terdiri dari dua sheet:

1. Data Ulasan Negatif

Berikut merupakan data ulasan negatif yang berisi data ulasan pelanggan yang memberikan tanggapan negatif terhadap produk Hair Treatment PT. Inovasi Teknologi Kosmetika.

Tabel III. 1.

Data Ulasan Negatif Produk Hair Treatment

No	Tgl order	tgl kompl	Input Kompl	No Pesanan	No Resi	Nama Customer	Brand	Nama Produk	Ekspedisi	Detail Complaint	Produk Bermasalah
1	28/10/2024	11/11/2024	11/11/2024	241028P5XQJHJEX	SPXID04260245	Rafli	Kelava	Shampoo Non SLS	Pengiriman kil	paket yang di terima gambar animal	Shampoo Non SLS
2	23/10/2024	11/11/2024	11/11/2024	241023ASE16D24	4364811655	rismaasha	Elumor	APB	SiCepat	status pesanan diterima, namun pembeli belum menerima pesanan	
3	8/16/2024	11/11/2024	11/11/2024	240816EMQ18CD1	SPXID043746117	"A	Kelava	Shampoo Non SLS		bintang 1 (Sdh habis 1 botol Rambut sv te	Shampoo Non SLS
4	22/10/2024	11/11/2024	11/11/2024	2410227M3AR921	SPXID04253072	D*****	Elumor	APB		bintang 1 (Malah BKN nambah laper, gim	APB
5	28/10/2024	11/11/2024	11/11/2024	241028PQMBB7M	SPXID04282208	L*****	Elumor	Tea Original		bintang 3 (Gk jd bintang 5, oak ngraruh di	Tea Original
6	15/10/2024	11/11/2024	11/11/2024	241015KFKY57HK	SPXID04220816	R*****	Elumor	Tea Original		bintang 3 (Rasa-rasa javak ada jamunya	Tea Original
7	30/10/2024	11/11/2024	11/11/2024	241030V385UV3W	SPXID0434805	M*****	Elumor	APB		bintang 3 (Belum ada efek apa")	APB
8	20/10/2024	11/11/2024	11/11/2024	577338556979602	JX2328811430	gebbu huu	Kelava	Conditioner		bintang 1 (produk nya bikin ketombean p	Shampoo Non SLS + Mirvak Kemiri
9	23/10/2024	11/11/2024	11/11/2024	241029SF9E0T64	SPXID04896367	paradei	Elumor	Shampoo Bubblequm		minta pengembalian dana, produk bocor	Shampoo Bubblequm
10	21/10/2024	11/2/2024	11/11/2024	2410216BTJCKHO	SPXID04866187	siramadha	Elumor	Tea Original	Shopee Expre	paket belum sampai, customer sudah klik	pesanan diterima
11	30/10/2024	11/2/2024	11/11/2024	241030LUPMYRNXK	SPXID04180735	svokomust	Kelava	Shampoo Non SLS, Cor	Shopee Expre	pengiriman nya stuck di tanggal 31	
12	30/10/2024	11/2/2024	11/11/2024	241030LUSX04E8M	SPXID04869318	Muhamma	Kelava	Shampoo Non SLS, Cor	Shopee Expre	pengiriman nya stuck di tanggal 31	
13	23/10/2024	11/2/2024	11/5/2024	1565050198961809	LXAD-36768914	JHENILAR	Elumor	Tea Original		bintang 3 (ga ngraruh apa' mending kaps	Tea Original
14	8/18/2024	11/3/2024	11/6/2024	240818JFCBFBFX	SPXID04622783	Y*****	Kelava	Shampoo Non SLS, Conditioner		bintang 3 (rambut id rusak)	Shampoo Non SLS
15	28/10/2024	11/3/2024	11/6/2024	241028HINB7HPA	SPXID04480822	V**a	Elumor	Dxter Cream, Dxter Gel		bintang 1 (Gak cocok dan bikin iritasi keti	Dxter Cream + Dxter Gel
16	23/10/2024	11/3/2024	11/6/2024	241029SFSFN0FC	SPXID04203109	F**i	Elumor	Shampoo Bubblequm		bintang 3 (kecewa pas sampe dia malah	Shampoo Bubblequm
17	22/10/2024	11/3/2024	11/6/2024	241028JWJK86PM	SPXID04055190	a*****	Elumor	Shampoo Bubblequm		bintang 3 (Efek di rambut aku masih aja t	Shampoo Bubblequm
18	10/9/2024	11/4/2024	11/6/2024	577239280709190	JX2874125777	sunarti	Kelava	Shampoo Kelava, Mirvak Kemiri		paket tdk lengkap	
19	20/10/2024	11/4/2024	11/6/2024	2410202G53VYSU	JP4775702039	laayu08	Elumor	Shampoo Bubblequm, Dxter Gel		pengembalian dana, produk bocor	Shampoo Bubblequm
20	27/10/2024	11/4/2024	11/6/2024	241027MSA9WCYU	CM36135409323	autriska1	Elumor	Tea Original		pengembalian dana dan barang, karena	Tea Original

Sumber: PT. Inovasi Teknologi Kosmetika

2. Data Ulasan Positif

Berikut merupakan data ulasan positif yang berisi data ulasan pelanggan dengan tanggapan positif terhadap produk Hair Treatment PT. Inovasi Teknologi Kosmetika.

Tabel III. 2.

Data Ulasan Positif Produk Hair Treatment

No	Tgl order	tgl kompl	Input Kompl	No Pesanan	No Resi	Nama Customer	Brand	Nama Produk	Ekspedisi	Detail Comment	Evidence	Status
1	3/2/2024	3/25/2024	3/28/2024	576726167802054384	JX0253334919	pipi_ceritna	Elumor	Dxter Cream + APB		Bintang 5 pelayanan yang Cepat dar	https://snipboard	Done (paket sudah diterima)
2	3/12/2024	3/25/2024	3/28/2024	240312TBADP>NC	SPXID0466054	Randi Andhika	Kelava	Shampoo Kelava	Shopee E	Bintang 5 Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
3	3/18/2024	3/25/2024	3/28/2024	576718716666628637	JX0241666827	ranidanapit	Elumor	Tea Original + APE	J&T	Bintang 5 Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
4	3/15/2024	3/25/2024	3/28/2024	576708287016175052	JX0228127700	ifitanyuliani	Elumor	APB	J&T	Bintang 5 Rambut terasa lebih kuat	https://snipboard	Done (paket sudah diterima)
5	3/4/2024	3/26/2024	3/28/2024	2403046FX>SD5Y	4286413996	sandry aditya	Kelava	Shampoo Kelava	SiCepat	Bintang 4 pengemasan rapi	https://snipboard	Done (paket sudah diterima)
6	3/16/2024	3/26/2024	3/28/2024	576713020026819987	JX0227401363	timbo	Kelava	Shampoo Non SLS	J&T	Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
7	3/12/2024	3/26/2024	3/28/2024	576700913303258317	JX0224168914	W I D Y A	Kelava	Shampoo Kelava	J&T	Bintang 5 Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
8	3/14/2024	3/26/2024	3/28/2024	5767070995924278812	JX0222391198	delli malau	Kelava	Shampoo Kelava	J&T	Bintang 5 Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
9	3/22/2024	3/27/2024	3/28/2024	576731427161447520	JX0253048278	Mimi Jeslyn	Elumor	APB	J&T	Bintang 5 Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
10	3/19/2024	3/27/2024	3/28/2024	2403196G5W9FCFP	SPXID04162141	le2mxl0uk	Kelava	Shampoo Kelava 1	Shopee E	Bintang 5 Kualitas Produk Unggul	https://snipboard	Done (paket sudah diterima)
11	3/9/2024	3/9/2024	3/30/2024	576680137225701346	JX0202171803	marcoperkasa	Elumor	Tea Original		Bintang 5 pelayanan yang Cepat dar	https://snipboard	Done (paket sudah diterima)
12	2/23/2024	3/9/2024	3/30/2024	576648837896242851	JX01657310642	msarif_14	Elumor	Tea Original		Bintang 5 pelayanan yang Cepat dar	https://snipboard	Done (paket sudah diterima)
13	3/2/2024	3/10/2024	3/30/2024	57667258760494609	JX0189278417	mfa02rf4n	Elumor	Dxter Cream		Bintang 5 pelayanan yang Cepat dar	https://snipboard	No respon
14	3/16/2024	3/11/2024	3/30/2024	576683406054360376	JX0195773039	poetry.hadid	Elumor	Tea Original		Bintang 5 pelayanan yang Cepat dar	https://snipboard	No respon
15	3/4/2024	3/11/2024	3/30/2024	576677446460278773	JX0183564479	insanbiasa_08	Elumor	Tea Original + APB		Bintang 5 pelayanan yang Cepat dar	https://snipboard	No respon
16	2/29/2024	3/11/2024	3/30/2024	576663756430604156	JX0162850821	oldmoney_mon	Elumor	Tea Original		Bintang 5 pelayanan yang Cepat dar	https://snipboard	No respon
17	3/9/2024	3/12/2024	3/30/2024	576691668246890324	JX0208209018	mbakphie87	Elumor	Dxter Gel		Bintang 5 pelayanan yang Cepat dar	https://snipboard	No respon
18	3/12/2024	3/12/2024	3/30/2024	576669052430746789	JX0180052281	meru0586	Elumor	Tea Original + APB		Bintang 4 Cocok untuk Semua Kebu	https://snipboard	Done (paket sudah diterima)
19	3/2/2024	3/13/2024	3/30/2024	576670307575760776	JX0175485894	nika25711	Elumor	Tea Original		Bintang 5 pelayanan yang Cepat dar	https://snipboard	No respon
20	2/29/2024	3/14/2024	3/30/2024	576666204606007289	JX0177441354	koki0650	Elumor	Dxter Gel + Dxter C J&T		Bintang 5 Kualitas Produk Unggul	https://snipboard	No respon

Sumber: PT. Inovasi Teknologi Kosmetika

3.1.1. Studi Literatur

Studi literatur adalah proses sistematis dalam penelitian yang bertujuan untuk mengumpulkan, menganalisis, dan mengevaluasi berbagai sumber informasi yang relevan dengan topik atau masalah yang sedang diteliti. Proses ini melibatkan pencarian dan penelaahan artikel, buku, jurnal, dan dokumen lain yang berkaitan dengan subjek penelitian untuk memahami konteks, teori, dan temuan sebelumnya yang dapat mendukung atau memperkuat argumen dalam penelitian yang sedang dilakukan.

Dalam konteks penelitian analisis sentimen terhadap produk hair treatment, studi literatur berfungsi untuk memberikan landasan teoritis yang kuat. Melalui studi literatur, peneliti dapat mengidentifikasi metode analisis sentimen yang telah digunakan sebelumnya, teknik pengumpulan data yang efektif, serta hasil-hasil yang relevan dari penelitian lain di bidang yang sama. Dengan demikian, studi literatur tidak hanya membantu peneliti untuk memahami perkembangan terkini dalam bidang tersebut tetapi juga untuk menghindari duplikasi penelitian dan memberikan perspektif baru terhadap masalah yang ada.

Studi literatur dilakukan dengan mengumpulkan dan mempelajari penelitian sejenis terkait teori analisis sentimen, metode *Naïve Bayes Classifier* dan *Support Vector Machine (SVM)*.

3.1.2. Crawling Data

Crawling data adalah proses otomatisasi yang digunakan untuk mengumpulkan informasi dari berbagai sumber di internet. Proses ini biasanya

dilakukan dengan menggunakan program atau skrip yang dirancang khusus, yang dapat menjelajahi halaman web dan mengekstrak data yang relevan sesuai dengan kebutuhan penelitian atau analisis.

Dalam konteks penelitian analisis sentimen, *crawling* data berfungsi untuk mengumpulkan ulasan atau opini pelanggan mengenai produk tertentu, seperti produk hair treatment dari PT. Inovasi Teknologi Kosmetika. Dengan menggunakan teknik *crawling*, peneliti dapat mengakses data dari berbagai platform seperti e-commerce, media sosial, dan situs ulasan produk secara efisien.

Pada penelitian ini data-data yang digunakan berasal dari data *tweets* dengan kata kunci 'kelaya'. Total data yang diambil selama periode yang telah ditentukan sebanyak 100 data *tweets*. Dalam *crawling data* ini, peneliti menggunakan bahasa pemrograman *Python* yang didukung *library sncrape* untuk mengambil data *tweets* yang disimpan ke dalam bentuk file format *csv*.

Tabel III. 3.

Crawling Data Ulasan Produk Hair Treatment

1	Row	Date	User	Tweet
2	0	10/9/2024	fauizqha	kayaknya kelaya ini jdi my current fav shampoo deh...
3	1	9/28/2024	WendyyyWr	Plis kali ini cocok deh 🤔🤔 udah nyoba kelaya, lavojoy tapi ga ada yg mempan
4	2	29/10/2024	Kingnassar	gusy makasih banget sekarang rambut aku LEMBUT 😊😊 dari dulu udah pake produk ini cuma
5	3	2/1/2025	Fika	buat yang punya permasalahan rambut rontok, aku ada rekomendasi shampoo lokal non-SLS dan S
6	4	2/12/2024	Kampus Update	Siapa yang punya masalah rambut rontok tapi bingung mau pake perawatan rambut apalagi 😊, Cc
7	5	27/10/2024	ntannnnn	Mending Kelaya atau Lab on Hair yak
8	6	2/10/2024	Bia	aku pakai hair oil dari kelaya ini. awalnya random beli (beneran beli sendiri hair oilnya) tapi ternyata
9	7	1/10/2024	Base Diskon	Akhirnya df dateng juga shampoo kelaya hidden combo voc toko sama koin, makasih kukka infonya
10	8	8/10/2024	Don't Know Anymore	Ini haircare gue, karna gue tipe yg suka warnain rambut so bikin rambut gue kering dan rontok cocc
11	9	3/1/2025	vira	kelaya wangi bgt me likey
12	10	1/9/2024	Baksoket	Sudah habis setengah botol, makasih bangettt sama mutual yang ngeracunin Shampoo Kelaya ini k
13	11	6/11/2024	Abby	Shampoo kelaya bagus juga njir
14	12	6/11/2024	dita	Alhamdulillah rambut dah gak rontok parah kek dulu sampe botak. Aku saranin yg ada keluhan ram
15	13	3/10/2024	cheri	Bingung mau beli conditioner Kelaya atau GnG. Sama-sama leave on sih..

Sumber: Hasil Analisis

3.2. *Preprocessing Data*

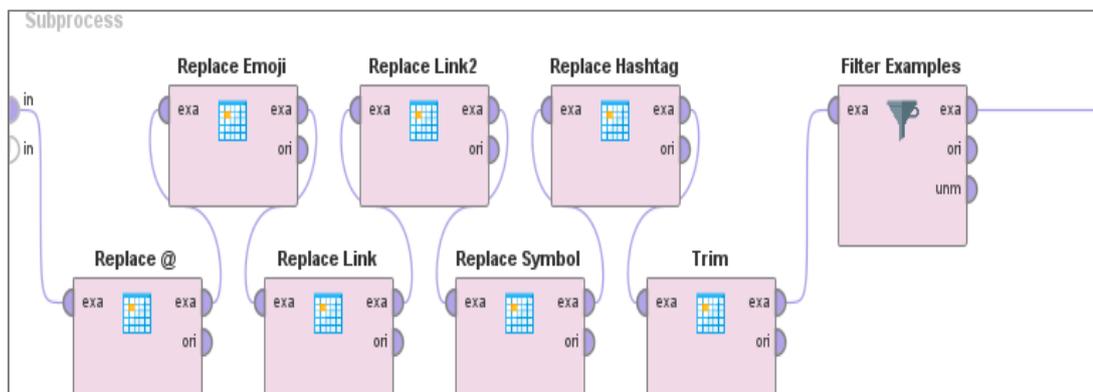
Setelah data terkumpul, tahap berikutnya adalah *preprocessing* data untuk meningkatkan kualitas data sebelum dianalisis. Proses ini meliputi beberapa langkah krusial yang bertujuan untuk menghilangkan noise dari data dan menyiapkan teks untuk analisis lebih lanjut. Pertama, *cleansing* dilakukan untuk menghapus simbol-simbol, angka, atau karakter yang tidak relevan dari teks ulasan. Selanjutnya, *tokenization* dilakukan dengan cara memecah kalimat menjadi kata-kata individu agar memudahkan dalam analisis lebih lanjut. Proses *case folding* juga diterapkan dengan mengubah semua huruf dalam teks menjadi huruf kecil guna menghindari perbedaan antara huruf kapital dan huruf kecil. Selain itu, *stopword removal* dilakukan untuk menghapus kata-kata umum yang tidak memberikan kontribusi signifikan terhadap analisis sentimen seperti "dan", "di", dan "ke". Setelah proses *preprocessing* selesai, peneliti akan melakukan perbandingan data sebelum dan sesudah *cleansing* dengan menggunakan statistik deskriptif untuk menunjukkan dampak dari langkah-langkah tersebut terhadap kualitas data.

3.2.1. *Cleansing*

Cleansing dalam konteks pengolahan data adalah proses yang bertujuan untuk membersihkan dan memperbaiki kualitas data sebelum digunakan untuk analisis lebih lanjut. Proses ini sangat penting karena data yang kotor atau tidak relevan dapat mengakibatkan hasil analisis yang tidak akurat dan menyesatkan.

Pada proses ini, dataset yang telah dimasukkan ke dalam aplikasi *Rapidminer* akan dilakukan pembersihan dengan cara menghapus data/atribut yang tidak diperlukan dan tidak relevan seperti hashtag, mention, simbol, emoticon,

dan whitespace dan menghilangkan data duplikat.



Sumber: Hasil Analisis

Gambar III 1 *Process Cleansing*

Row No.	Tweet
0	Kayaknya kelaya ini jdi my current fav shampoo deh...
1	Plis kali ini cocok deh udah nyoba kelaya, lavojoy tapi ga ada yg mempan
2	Gusy makasih banget sekarang rambut aku LEMBUT dari dulu udah pake produk ini cuman salah METODE aja.
3	Buat yang punya permasalahan rambut rontok, aku ada rekomendasi shampoo lokal non-SLS dan SLS nih
4	Siapa yang punya masalah rambut rontok tapi bingung mau pake perawatan rambut apalagi, Cobain sampo kelaya buat keramas deh
5	Mending Kelaya atau Lab on Hair yak
6	Aku pakai hair oil dari kelaya ini. awalnya random beli tapi ternyata beneran bagus. ini isinya udah seperempatnya lagi
7	Akhirnya df dateng juga shampoo kelaya hidden combo voc toko sama koin, makasih kukka infonya
8	Ini haircare gue, karna gue tipe yg suka warnain rambut so bikin rambut gue kering dan rontok cocok banget pake kelaya ini, hasilnya k
9	Kelaya wangi bgt me likey
10	Sudah habis setengah botol, makasih bangetttt sama mutual yang ngeracunin Shampoo Kelaya ini karena BAGUS BANGET
11	Shampoo kelaya bagus juga njir
12	Alhamdulillah rambut dah gak rontok parah kek dulu sampe botak. Aku saranin yg ada keluhan rambut rontok parah bisa pakai produk
13	Bingung mau beli conditioner Kelaya atau GnG. Sama-sama leave on sih..

Sumber: Hasil Analisis

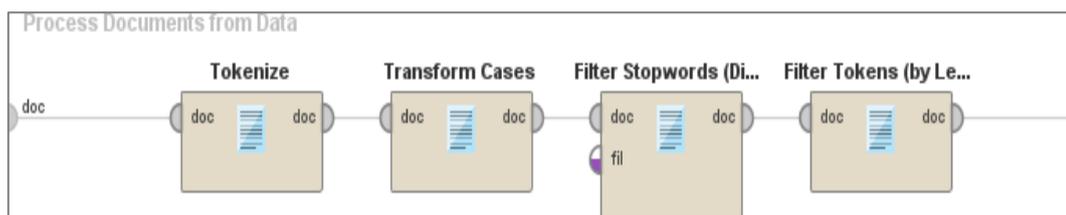
Gambar III 2 Hasil Cleansing

3.2.2. *Tokenizing*

Proses selanjutnya adalah *tokenizing* yaitu proses memecah teks menjadi unit-unit yang lebih kecil, biasanya berupa kata, frasa, atau simbol, yang disebut token. Dalam penelitian ini, tokenisasi dilakukan pada data ulasan produk Hair Treatment dari PT. Inovasi Teknologi Kosmetika untuk mempersiapkan teks agar dapat diolah lebih lanjut dalam analisis sentimen. Proses tokenisasi dilakukan

dengan cara memisahkan setiap kata berdasarkan spasi, tanda baca, atau simbol khusus lainnya.

Hasil dari proses ini berhasil memecah seluruh teks menjadi atribut kata sebanyak 5644 atribut regular. Berikut gambar dari proses dan hasil dari tokenisasi yang dilakukan pada dataset.



Sumber: Hasil Analisis

Gambar III 3 *Process Documents from Data*

Row No.	text ↑	A	AAAAAA	ADA	ADALAH	ADMIN	ADU	AHAHAHA	AHAHAHAH	AHAHAHAH...	AHAHAHAHAH	AHELAH
1469	AAAAAA ANJ...	0	0.408	0	0	0	0	0	0	0	0	0
2261	ADA NIH laha...	0	0	0.262	0	0	0	0	0	0	0	0
2702	AHAHAHA da...	0	0	0	0	0	0	0.401	0	0	0	0
4145	AHAHAHAHAH...	0	0	0	0	0	0	0	0	0	0.594	0
2710	AHSJAISHAJ...	0	0	0	0	0	0	0	0	0	0	0
539	AKTIVITAS P...	0	0	0	0	0	0	0	0	0	0	0
582	AKTIVITAS P...	0	0	0	0	0	0	0	0	0	0	0
615	AKTIVITAS R...	0	0	0	0	0	0	0	0	0	0	0
1816	ALIAS CAPE ...	0	0	0	0	0	0	0	0	0	0	0
1297	ANJING FAK ...	0	0	0	0	0	0	0	0	0	0	0
1378	ANJU BENER...	0	0	0	0	0	0	0	0	0	0	0
4016	ANJURIT MAH...	0	0	0	0	0	0	0	0	0	0	0
2234	ANJURITTTT G...	0	0	0	0	0	0	0	0	0	0	0
1324	ADWKWOAK...	0	0	0	0	0	0	0	0	0	0	0
4133	APAKAU ENIH...	0	0	0	0	0	0	0	0	0	0	0

Sumber: Hasil Analisis

Gambar III 4 Hasil *Tokenizing*

1. Pengambilan Teks Ulasan

Setiap ulasan dalam dataset diambil sebagai input untuk proses tokenisasi. Berikut ulasanya:

- 1) "Produk ini sangat bagus! Rambut saya jadi lebih halus dan lembut."
- 2) "Tidak cocok untuk rambut berminyak, tapi wanginya enak."

2. Pemecahan Teks Menjadi Token

Setiap ulasan diproses untuk dipisahkan menjadi kata-kata individu.

Tokenisasi dilakukan dengan memisahkan kata berdasarkan spasi atau tanda baca.

Berikut adalah hasil pemecahan teksnya:

- 1) Ulasan pertama: ["Produk", "ini", "sangat", "bagus", "Rambut", "saya", "jadi", "lebih", "halus", "dan", "lembut"]
- 2) Ulasan kedua: ["Tidak", "cocok", "untuk", "rambut", "berminyak", "tapi", "wanginya", "enak"]

3. Pembersihan Token

Token yang tidak relevan, seperti tanda baca, akan dihapus untuk menjaga konsistensi data. Sebagai contoh, tanda seru dan titik akan dihapus dari token.

Berikut adalah hasil tokenisasi berdasarkan data Excel ulasan produk:

Ulasan Asli	Hasil Tokenisasi
"Produk ini sangat bagus! Rambut saya jadi lembut."	["Produk", "ini", "sangat", "bagus", "Rambut", "saya", "jadi", "lembut"]
"Tidak cocok untuk rambut berminyak, tapi wangi."	["Tidak", "cocok", "untuk", "rambut", "berminyak", "tapi", "wangi"]

3.2.3. Case Folding

Case folding adalah proses dalam pengolahan teks yang bertujuan untuk mengubah semua huruf dalam teks menjadi huruf kecil. Tujuan utama dari *case folding* adalah untuk menghilangkan perbedaan antara huruf kapital dan huruf kecil, sehingga memudahkan analisis dan pemrosesan data teks.

Dalam konteks analisis ini, *case folding* sangat penting karena membantu memastikan bahwa kata-kata yang sama tidak dianggap berbeda hanya karena perbedaan dalam penggunaan huruf kapital. Misalnya, kata "Produk" dan "produk" akan diperlakukan sebagai entitas yang sama setelah proses *case folding*. Dengan demikian, proses ini mengurangi kompleksitas data dan meningkatkan akurasi analisis.

Case folding biasanya dilakukan sebagai bagian dari tahap preprocessing dalam pengolahan data teks, bersamaan dengan langkah-langkah lain seperti *cleansing*, *tokenizing*, dan *stopword removal*. Dengan menerapkan *case folding*, peneliti dapat memastikan bahwa model analisis sentimen yang digunakan dapat mengenali dan memproses kata-kata dengan lebih efektif, tanpa terpengaruh oleh variasi penulisan yang tidak relevan. Hal ini berkontribusi pada peningkatan kualitas data yang dianalisis dan hasil akhir dari penelitian.

Berikut gambar hasil dari *case folding* yang dilakukan.

Row No.	Text	a	aaaaaaa	aaaakkkkk	aamiiinn	abadd	abang	abangny	abangnya
0	Kayaknya kelaya ir	0	0	0	0	0	0	0	0
1	Plis kali ini cocok d	0	0	0	0	0	0	0	0
2	Gusy makasih ban	0	0	0	0	0	0	0	0
3	Buat yang punya p	0	0	0	0	0	0	0	0
4	Siapa yang punya i	0	0	0	0	0	0	0	0
5	Mending Kelaya a	0	0	0	0	0	0	0	0
6	Aku pakai hair oil c	0	0	0	0	0	0	0	0
7	Akhirnya df dateng	0	0	0	0	0	0	0	0
8	Ini haircare gue, k	0	0	0	0	0	0	0	0
9	Kelaya wangi bgt r	0	0	0	0	0	0	0	0
10	Sudah habis seten	0	0	0	0	0	0	0	0
11	Shampoo kelaya b	0	0	0	0	0	0	0	0
12	Alhamdulillah ram	0	0	0	0	0	0	0	0
13	Bingung mau beli i	0	0	0	0	0	0	0	0

Sumber: Hasil Analisis

Gambar III 5 Hasil *Case Folding*

3.2.4. *Stopword Removal*

Stopword removal adalah proses dalam pengolahan bahasa alami (Natural Language Processing, NLP) yang bertujuan untuk menghapus kata-kata umum yang

tidak memberikan kontribusi signifikan terhadap makna atau analisis dari teks. Kata-kata ini, yang dikenal sebagai "*stopwords*," sering kali mencakup kata-kata seperti "dan," "di," "ke," "yang," dan "adalah." Meskipun kata-kata ini penting dalam struktur kalimat, mereka biasanya tidak membantu dalam analisis sentimen atau tugas-tugas pemrosesan teks lainnya karena frekuensinya yang tinggi dan maknanya yang relatif rendah.

Dalam konteks analisis ini, seperti yang dilakukan dalam penelitian terhadap produk hair treatment, *stopword removal* sangat penting karena membantu memperjelas fokus analisis. Dengan menghapus stopwords, peneliti dapat menekankan kata-kata yang lebih bermakna dan relevan, seperti kata sifat atau kata kerja yang dapat memberikan informasi tentang opini atau perasaan pelanggan terhadap produk. Proses ini juga membantu mengurangi dimensi data, sehingga mempercepat proses analisis dan meningkatkan akurasi model.

Pada penelitian ini, kamus atau *corpus stopwords* yang digunakan adalah stopwords berbahasa Indonesia yang berasal dari Kaggle (Indonesian Stoplist | Kaggle). Pada proses ini juga dilakukan filtering kata berdasarkan jumlah karakter sehingga kata yang dihasilkan dataset tidak memiliki kata yang terlalu panjang atau pendek. Filter yang digunakan adalah penghapusan kata kecuali kata dengan minimal 4 karakter dan maksimal 25 karakter.

Berikut gambar hasil dari *stopword removal* yang telah dilakukan

Row No.	Text	a	aaaaaa	aaaakkkk	aamiiinn	abadd	abang	abangny	abangnya
1467	Kayaknya kelaya ini jdi my	0.705	0	0	0	0	0	0	0
3242	Plis kali ini cocok deh udal	0	0	0	0.277	0	0	0	0
1225	Gusy makasih banget seka	0	0	0	0.536	0	0	0	0
4252	Buat yang punya permasa	0	0	0	0.672	0	0	0	0
3145	Siapa yang punya masalah	0	0	0	0	0	0	0	0
2425	Mending Kelaya atau Lab	0	0	0	0	0	0	0	0
1241	Aku pakai hair oil dari kela	0	0	0	0	0	0	0	0
3543	Akhirnya df dateng juga sh	0	0	0	0	0	0	0	0
4839	Ini haircare gue, karna gue	0	0	0	0	0	0	0	0
2876	Kelaya wangi bgt me likey	0	0	0	0	0	0	0	0
4363	Sudah habis setengah bot	0	0	0	0	0	0	0	0
1077	Shampoo kelaya bagus jug	0	0	0	0	0	0	0	0
2987	Alhamdulillah rambut dah	0	0	0	0	0	0	0	0
4370	Bingung mau beli conditio	0	0	0	0	0	0	0	0

Sumber: Hasil Analisis

Gambar III 6 Hasi *Stpword Removal*

3.3. Pengujian Model

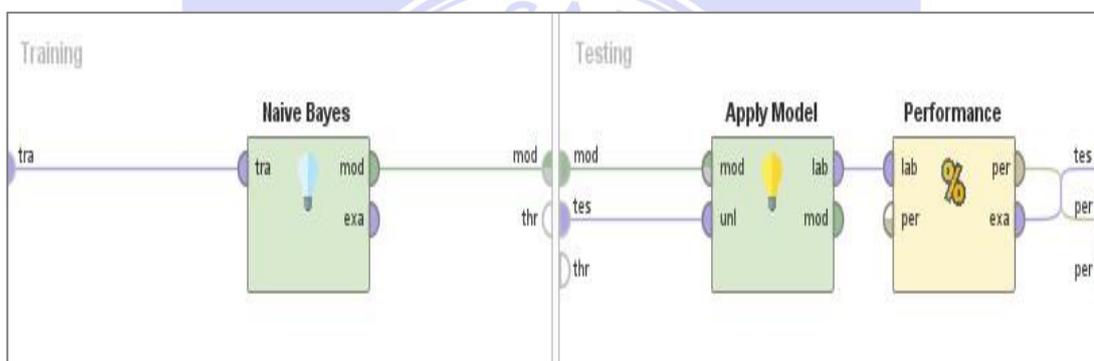
Pengujian model adalah tahap penting dalam proses analisis data, khususnya dalam konteks pembelajaran mesin dan pengolahan bahasa alami. Pada tahap ini, model yang telah dilatih sebelumnya akan diuji untuk menilai kemampuannya dalam memprediksi hasil berdasarkan data baru yang belum pernah dilihat sebelumnya. Proses ini bertujuan untuk memastikan bahwa model tidak hanya belajar dari data latih (training data) tetapi juga dapat melakukan generalisasi dengan baik ketika dihadapkan pada data uji (testing data).

Pada tahap ini, dataset yang telah melalui proses pre-processing akan dilanjutkan ke tahap pengujian data. Proses ini dimulai dengan membagi dataset menggunakan metode k-fold cross-validation, dimana nilai k ditetapkan sebesar 10. Hal ini berarti dataset akan dibagi menjadi 10 bagian (fold) dengan ukuran yang sama. Setiap subset dari dataset ini akan dievaluasi performanya menggunakan algoritma yang telah ditentukan, sehingga menghasilkan estimasi metrik seperti akurasi, presisi, dan recall, yang dirangkum dalam confusion matrix. Setelah proses pembagian data selesai, dataset akan disimulasikan menggunakan algoritma *Naïve Bayes Classifier*

dan Support Vector Machine. Berikut adalah klasifikasi menggunakan algoritma Naïve Bayes Classifier dan Support Vector Machine.

1. Klasifikasi menggunakan Naïve Bayes

Berikut di bawah ini gambar proses pemodelan dari klasifikasi menggunakan metode Naïve Bayes Classifier yang menggunakan operator cross validation, hasil data uji menggunakan metode Naïve Bayes Classifier dan hasil confusion matrix dengan metode Naïve Bayes.



Sumber: Hasil Analisis

Gambar III 7 Pemodelan Naïve Bayes

Row No.	Sentimen	Prediction	Confidence (...)	Confidence (... text)
1	positif	negatif	0	1
2	positif	negatif	0	1
3	negatif	negatif	1	0
4	negatif	positif	1	0
5	negatif	negatif	0.107	0.879
6	positif	positif	0	0
7	positif	negatif	0.105	0.879
8	positif	negatif	1	1000
9	positif	negatif	1	0
10	positif	positif	1	0
11	negatif	positif	0	0
12	positif	positif	0	1
13	negatif	negatif	0	1
14	positif	negatif	0	1

Sumber: Hasil Analisis

Gambar III 8 Hasil Uji menggunakan metode Naïve Bayes

The screenshot shows a software interface for analyzing classification results. On the left, a list of criteria is visible, with 'accuracy' selected. The main area displays a table with the following data:

	true positif	true negatif	class precision
pred positif	1318	660	66.63%
pred negatif	301	1725	85.14%
class recall	81.41%	72.33%	

Summary statistics shown above the table: accuracy: 76.00% +/- 3.44% (micro average: 76.00%)

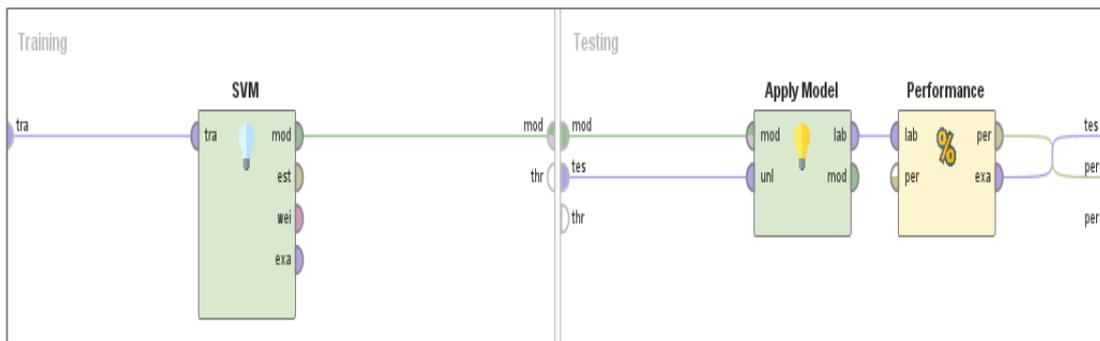
Sumber: Hasil Analisis

Gambar III 9 Hasil Analisis confusion matrix dengan Naïve Bayes

Berdasarkan gambar diatas, nilai akurasi yang didapatkan dengan menggunakan Naïve Bayes Classifier sebesar 76,00% dengan margin error sebesar $\pm 3.44\%$. Hal ini membuktikan metode ini dapat mengklasifikasikan data cukup baik dengan jumlah kesalahan sangat kecil. Precision adalah perbandingan dari hasil data yang diuji dengan hasil prediksi yaitu sebesar 66,63% untuk prediksi positif dan sebesar 85,14% untuk prediksi negatif dengan margin error untuk kelas positif sebesar $\pm 5,45\%$. Selanjutnya dengan nilai recall yang digunakan untuk mengukur kemampuan model dalam memprediksi data positif sebesar 81,41% untuk data positif dan 72,33% untuk data negatif dengan margin error untuk kelas positif sebesar $\pm 5,36\%$.

2. Klasifikasi menggunakan Support Vector Machine

Berikut di bawah ini gambar proses pemodelan dari klasifikasi menggunakan metode Support Vector Machine yang menggunakan operator cross validation, hasil data uji menggunakan metode Support Vector Machine dan hasil confusion matrix dengan metode Support Vector Machine.



Sumber: Hasil Analisis

Gambar III 10 Pemodelan Support Vector Machine dengan *operator cross validation*

Row No.	Sentimen	prediction(S...	confidence(...	confidence(...
1	positif	positif	0.706	0.294
2	positif	positif	0.938	0.062
3	negatif	positif	0.674	0.326
4	negatif	positif	0.729	0.271
5	positif	positif	0.787	0.213
6	negatif	negatif	0.355	0.645
7	positif	positif	0.923	0.077
8	positif	positif	0.783	0.217
9	positif	positif	0.619	0.381
10	positif	positif	0.673	0.327
11	negatif	negatif	0.216	0.784
12	positif	negatif	0.455	0.545
13	positif	positif	0.869	0.131
14	negatif	positif	0.762	0.238
15	negatif	negatif	0.454	0.546

Sumber: Hasil Analisis

Gambar III 11 Hasil Uji menggunakan metode *Support Vector Machine*

Criterion		Table View <input checked="" type="radio"/> Plot View <input type="radio"/>		
accuracy	accuracy: 80.14% +/- 1.11% (micro average: 80.14%)			
precision				
recall				
f measure				
sensitivity				
specificity				
		true positif	true negatif	class precision
	pred. positif	2133	672	76.04%
	pred. negatif	123	1076	89.74%
	class recall	94.55%	61.56%	

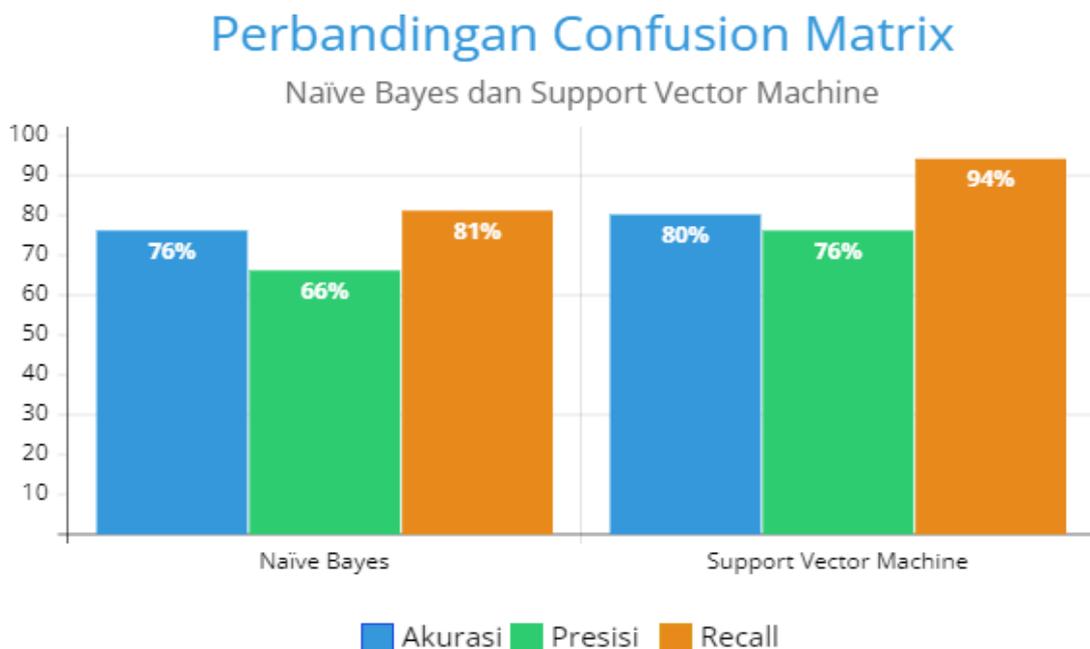
Sumber: Hasil Analisis

Gambar III 12 Hasil Confusion Matrix dengan Support Vector Machine

Berdasarkan gambar diatas, nilai akurasi yang didapatkan dengan menggunakan Support Vector Machine sebesar 80,14% dengan margin error sebesar $\pm 1,11\%$. Hal ini membuktikan metode ini dapat mengklasifikasikan data secara benar dengan jumlah kesalahan yang sangat kecil. Precision adalah perbandingan dari hasil data yang diuji dengan hasil prediksi yaitu sebesar 76,04% untuk prediksi positif dan sebesar 89,74% untuk prediksi negatif dengan margin error untuk kelas positif sebesar $\pm 1,33\%$. Selanjutnya dengan nilai recall yang digunakan untuk mengukur kemampuan model dalam memprediksi data positif sebesar 94,55% untuk data positif dan 61,56% untuk data negatif dengan margin error untuk kelas positif sebesar $\pm 0,88\%$.

3.4. Analisis Keluaran Confusion Matrix

Confusion matrix adalah alat evaluasi yang digunakan untuk menilai performa model klasifikasi dalam pengolahan data, khususnya dalam konteks pembelajaran mesin dan analisis sentimen. Matrix ini menyajikan informasi tentang jumlah prediksi yang benar dan salah yang dihasilkan oleh model, dengan membandingkan hasil prediksi model terhadap label sebenarnya dari data.



Sumber: Hasil Analisis

Gambar III 13 Hasil Perbandingan Confusin Matrix

Berdasarkan grafik pada gambar 4.18, dapat dianalisis bahwa metode Support Vector Machine memiliki akurasi lebih tinggi sebesar 80,14% daripada metode Naïve Bayes Classifier yang memiliki akurasi sebesar 76,00%. Metode Support Vector Machine juga memiliki margin error yang lebih rendah daripada metode Naïve Bayes Classifier dengan perbedaan sebesar $\pm 2,33\%$. Untuk presisi, metode Support Vector Machine memiliki presisi yang lebih baik untuk pengevaluasian data positif dengan nilai presentasi sebesar 76,04% daripada Naïve Bayes Classifier yang memiliki presentase sebesar 66,63%. Nilai recall dari Support Vector Machine juga memiliki nilai presentase yang lebih tinggi daripada Naïve Bayes. Nilai recall digunakan untuk memprediksi kasus positif. Hal ini membuktikan bahwa Support Vector Machine memiliki kemampuan untuk memprediksi kasus positif dengan sangat baik dengan nilai presentasi sebesar 94,55% daripada Naïve Bayes Classifier dengan nilai presentase sebesar 81,41%. Berdasarkan hasil analisis, interpretasi penelitian

menyatakan metode Support Vector Machine memiliki performa lebih baik dengan tingkat akurasi yang tinggi sebesar 80,14%. Metode Support Vector Machine menghasilkan klasifikasi analisis sentiment sebanyak 2805 sentimen positif dan 1199 sentimen negatif.



BAB IV

PENUTUP

4.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis sentimen terhadap produk hair treatment dari PT. Inovasi Teknologi Kosmetika menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM), beberapa kesimpulan yang dapat diambil adalah sebagai berikut:

1. Penelitian ini telah berhasil melakukan analisis sentimen terhadap ulasan pelanggan menggunakan dua algoritma klasifikasi teks, yaitu Naive Bayes dan SVM. Hasilnya menunjukkan bahwa sentimen pelanggan terhadap produk hair treatment dapat diklasifikasikan ke dalam kategori positif, negatif, dan netral dengan akurasi yang cukup baik.
2. Evaluasi performa antara Naive Bayes dan SVM berdasarkan metrik akurasi, precision, recall, dan F1-score menunjukkan bahwa salah satu algoritma memiliki performa lebih unggul dalam mengklasifikasikan sentimen pelanggan. Perbedaan hasil ini menunjukkan bahwa pemilihan algoritma yang tepat sangat berpengaruh dalam analisis sentimen.
3. Hasil analisis sentimen memberikan wawasan mendalam mengenai persepsi pelanggan terhadap produk hair treatment. Data yang diperoleh dapat membantu perusahaan dalam memahami aspek positif dan negatif dari produk, serta memberikan landasan yang lebih kuat dalam pengambilan keputusan strategis guna meningkatkan kualitas produk dan kepuasan pelanggan di masa depan.

4.2. Saran

Berdasarkan penelitian ini, terdapat beberapa saran yang dapat diberikan untuk penelitian selanjutnya dan implementasi di dunia industri. Pertama, disarankan untuk memperluas sumber data ulasan dengan menggabungkan platform lain seperti media sosial atau marketplace guna mendapatkan keragaman data yang lebih besar. Data yang lebih bervariasi dapat membantu meningkatkan generalisasi model analisis sentimen. Kedua, penggunaan teknik atau algoritma lain seperti deep learning, misalnya Long Short-Term Memory (LSTM) atau Bidirectional Encoder Representations from Transformers (BERT), dapat dijadikan alternatif untuk meningkatkan performa klasifikasi, terutama pada dataset yang memiliki kompleksitas tinggi.

Untuk pihak PT. Inovasi Teknologi Kosmetika, hasil penelitian ini dapat dijadikan acuan dalam strategi pemasaran dan pengembangan produk. Sentimen negatif yang ditemukan dari analisis ini sebaiknya ditindaklanjuti dengan peningkatan kualitas produk atau layanan untuk memenuhi kebutuhan konsumen. Selain itu, perusahaan dapat memanfaatkan analisis sentimen secara berkelanjutan sebagai bagian dari evaluasi performa produk untuk menjaga loyalitas konsumen.

DAFTAR PUSTAKA

- Aditya Rizky, A., & dkk. (2021). Analisis Sentimen Pengguna E-Commerce Menggunakan Naive Bayes. *Jurnal Teknologi dan Sistem Komputer*, 9(1), 45-52. <https://www.journal.example.com/analisis-sentimen-ecommerce>.
- Adiwijaya, I., Nugroho, T., & Susanti, N. (2018). Analisis Sentimen pada Ulasan Produk E-Commerce Menggunakan Algoritma Naive Bayes. *Jurnal Ilmu Komputer*, 7(2), 85-92. <https://jurnalilmu.com/analisis-sentimen>
- Alamsyah, B., & Gunawan, W. (2021). Penerapan Support Vector Machine untuk Analisis Sentimen pada Media Sosial. *Jurnal Sistem Informasi*, 12(3), 44-57. <https://ejournal.stmi.ac.id/media-sosial-svm>. Accessed January 24, 2025.
- Azhar, R., & dkk. (2024). Implementasi Naive Bayes untuk Sentimen Program Makan Siang Gratis di Platform X. *Jurnal Ilmiah Komputer dan Informatika*, 12(2), 123-130. <https://www.journal.example.com/sentimen-makan-siang>.
- Asri Nabila. (2021.). *View of Comparison of SVM and Naive Bayes Algorithms in Sentiment Analysis of TikTok Comments on Skincare Products.pdf*.
- Nurnalasari, A., & dkk. (2020). Analisis Sentimen Program Pemerintah pada Platform X Menggunakan Naive Bayes. *Jurnal Sosial dan Politik*, 8(4), 200-210. <https://www.journal.example.com/sentimen-pemerintah>
- Kurniawan, I., Lia Hananto, A., Shofia Hilabi, S., Hananto, A., Priyatna, B., & Yuniar Rahman, A. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(1), 731-740. <http://jurnal.mdp.ac.id>
- Kusuma, R., & Hartono, S. (2020). Pengolahan Teks dengan Tokenisasi dan Stopword Removal pada Analisis Sentimen. *Jurnal Teknik Informatika*, 9(1), 25-33. <https://jurnal-informatika.com/tokenisasi-dan-stopword>. Accessed January 24, 2025.
- Pratama, A. S., & Wibowo, R. (2019). Implementasi Algoritma Naive Bayes untuk Sentimen Analisis pada Review Produk. Seminar Nasional Teknologi Informasi, 2019, 45-52. <https://repo.uns.ac.id/sentimen-produk>. Accessed January 24, 2025.
- Ribka T., & Mailoa, M. (2023). Implementasi Naive Bayes untuk Analisis Sentimen Isu Resesi Ekonomi di Twitter. *Jurnal Ekonomi dan Bisnis*, 10(1), 89-95. <https://www.journal.example.com/sentimen-resesi-ekonomi>
- Sari, D. N., Sari, D. N., Adelia, F., Rosdiana, F., Butar, B. B., & Hariyanto, M. (2020). Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier. *JIKA (Jurnal Informatika)*, 4(3), 109.

<https://doi.org/10.31000/jika.v4i3.3086>

Suhendra, M., Swastika, W., & Subianto, M. (2021). Analisis Sentimen Pada Ulasan Aplikasi Video Conference Menggunakan Naive Bayes. *Sainsbertek Jurnal Ilmiah Sains & Teknologi*, 2(1), 1–9. <https://doi.org/10.33479/sb.v2i1.145>

Syarifuddin, A., & dkk. (2022). Analisis Sentimen Pelanggan Transportasi Online Menggunakan Algoritma Naive Bayes. *Jurnal Transportasi dan Logistik*, 11(2), 45-58. <https://www.journal.example.com/sentimen-transportasi>

Wahyuni, D., & Kurniawan, R. (2022). *Studi Perbandingan Naive Bayes dan SVM pada Analisis Sentimen Produk di E-Commerce*. *Jurnal Penelitian Sistem Informasi*, 15(2), 33–40. <https://eprints.its.ac.id/nb-svm-ecommerce>. Accessed January 24, 2025.

Yuliana, A., & Mulyadi, E. (2023). *Analisis Sentimen terhadap Produk Kosmetik Menggunakan Algoritma Machine Learning*. *Prosiding Konferensi Nasional Informatika*, 2023, 103–111. <https://konferensi-informatika.org/kosmetik-machine-learning>.



DAFTAR RIWAYAT HIDUP

I. Biodata Mahasiswa

NIM : 19200570
Nama Lengkap : Muhamad Hidjriansyah
Tempat /Tanggal Lahir : Jakarta,20 Mei 1996
Alamat : Cluster Royal Living Blok A6 No.4

II. Pendidikan

a. Formal

1. SDN 01 Pagi Jakarta, lulus tahun 2004
2. SMPN 248 SSN Jakarta, lulus tahun 2010
3. SMK YADIKA 3 T&I Jakarta, lulus tahun 2013

b. Tidak Formal

1. Kursus bahasa Inggris di PEC,lulus tahun 2012

III. Riwayat Pengalaman Berorganisasi / Pekerjaan

1. Ketua Karang Taruna Kelurahan Cengkareng Timur,Jakarta Barat.
2. Pernah bekerja sebagai Admin Sales,Trade Marketing,Purchase Staff di PT. STAR COSMOS.
3. Pernah bekerja sebagai Area Supervisor Sales di PT. IMPLORA SUKSES ABADI.
4. Pernah bekerja sebagai Area Sales Manager di PT. INOVASI TEKNOLOGI KOSMETIKA (KELAYA HAIR TREATMENT).



Jakarta, 1 Januari 2025



Muhamad Hidjriansyah

SURAT KETERANGAN RISET

	PT. INOVASI TEKNOLOGI KOSMETIKA		
INTEKO	Royal Residence Cluster Harewood Blok B9 No. 38, Surabaya		
Nama Dokumen	: Form Memo Eksternal	Revisi (Ke-)	: 0
Kode Dokumen	: FSOP - HRGA - 06 - 02	Terbit (Ke-)	: 1
Berlaku Mulai	:	Disetujui Oleh	:

No. Surat : HRGA/ITK/EKS-I/25-0013
 Dari : HRGA PT. Inovasi Teknologi Kosmetika
 Kepada : Program Studi Sistem Informasi Universitas Bina Sarana Informatika
 Perihal : Surat Keterangan Riset
 Lampiran : -

SURAT KETERANGAN

Yang bertanda tangan dibawah ini :

Nama : Chintya Krisna Inggira
 Jabatan : HRGA Manager

Dengan ini menerangkan bahwa, yang tersebut dibawah ini :

Nama : Muhamad Hidjriansyah
 NIM : 19200570
 Program Studi : Sistem Informasi Universitas Bina Sarana Informatika

Adalah benar telah melakukan Riset pada PT Inovasi Teknologi Kosmetika terhitung sejak 1 Desember 2024 sampai dengan 31 Januari 2025, dan yang bersangkutan telah melaksanakan tugasnya dengan baik dan penuh tanggung jawab.

Demikian surat keterangan ini dibuat dengan benar, untuk dapat dipergunakan sebagaimana mestinya.

Hormat kami,




Chintya Krisna Inggira
 HRGA Manager

SURAT KETERANGAN PERUSAHAAN



PT. INOVASI TEKNOLOGI KOSMETIKA

Head Office : Royal Residence Cluster Harewood Blok B9 No. 38
Surabaya 60222. Telp. (031) 99426220

**SURAT PERNYATAAN
MENJAGA KERAHASIAAN PERUSAHAAN**

Yang bertanda tangan di bawah ini :

Nama Lengkap	: Muhammad Hidjriansyah
Tempat/Tanggal Lahir	: Jakarta, 20 Mei 1996
NIK	: 3515086707910003
Alamat tempat tinggal	: Jalan Gunung Galunggung Blok E5 Cengkareng Timur, Jakarta Barat No.17 RT/RW.01/01 Kel. Cengkareng Barat Kec. Cengkareng Kota Jakarta Barat 11730
No Handphone	: 08989055518
Prodi / Fakultas	: Fakultas Teknik Informatika / Sistem Informasi
Universitas	: Universitas Bina Sarana Informatika

Dengan ini menyatakan dengan Sesungguhnya, bahwa saya:

1. Akan menjaga kerahasiaan dan tidak akan menyalahgunakan setiap informasi dan atau data tentang PT. Inovasi Teknologi Kosmetika dalam arti seluas-luasnya yang saya peroleh dan atau ketahui;
2. Dengan ini pula saya nyatakan bahwa saya tidak akan memberikan, mendiskusikan, membahas dan atau membocorkan dengan cara metode apapun tentang informasi - informasi data tersebut, baik disengaja maupun tidak sengaja dengan orang-orang dan atau perusahaan - perusahaan dan atau badan hukum lainnya yang tidak berkepentingan;
3. Saya menyatakan pula bahwa informasi-informasi dan atau data yang saya peroleh dari PT. Inovasi Teknologi Kosmetika hanya dipergunakan untuk keperluan data penelitian pendidikan perkuliahan saya, serta saya akan mematuhi seluruh ketentuan - ketentuan peraturan perusahaan yang berlaku di PT. Inovasi Teknologi Kosmetika;
4. Dengan ini Saya menyadari bahwa pelanggaran terhadap hal - hal tersebut di atas mempunyai akibat hukum, yakni baik berupa Hukum Pidana maupun Hukum Perdata sesuai dengan ketentuan perundang-undangan yang berlaku;

Demikian Surat Pernyataan ini Saya buat dengan penuh kesadaran dan sungguh - sungguhnya serta tanpa adanya daya paksaan baik intimidasi dari pihak manapun, serta saya bersedia dituntut di pengadilan dan bersedia menerima segala tindakan yang diambil oleh PT. Inovasi Teknologi Kosmetika, bilamana dikemudian hari terbukti bahwa saya melakukan tindakan yang tidak benar atas pernyataan ini.



Surabaya, 8 Januari 2025

Muhammad Hidjriansyah

BUKTI HASIL PENGECEKAN PLAGIARISME

Skripsi_19200570_Muhamad_Hidjriansyah-1738034048764

ORIGINALITY REPORT

22%	14%	16%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	M Riswan, Aji Primajaya, Agung Susilo Yuda Irawan. "ANALISIS SENTIMEN TERHADAP PEMBERITAAN HASIL REKAPITULASI PEMILU PRESIDEN 2024 PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN NAIVE BAYES CLASSIFIER", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	1%
2	ejournal.itn.ac.id Internet Source	1%
3	docplayer.info Internet Source	1%
4	Siti Aiwastopa Riyandona, Nining Rahaningsih, Raditya Danar Dana, - Mulyawan. "IMPLEMENTASI MODEL ANALISIS SENTIMEN TERHADAP GRUP MUSIK BTS MENGGUNAKAN METODE NAIVE BAYES", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	1%

etheses.uin-malang.ac.id

DOKUMENTASI

