

ANALISIS SENTIMEN DAMPAK ENDORSEMENT PRODUK MAKEUP PADA AKUN INSTAGRAM BEAUTY INFLUENCER TASYA FARASYA

Fikeu Destri Nursari Kusumah¹⁾, Dika Jatnika²⁾

Bachelor of Digital Business Study Program, Universitas Padjadjaran / Management and Business, Universitas Padjadjaran
fikeu20001@mail.unpad.ac.id

Abstract

The increasing use of social media, particularly Instagram, has transformed it from a mere entertainment platform into a significant marketing venue for businesses. One of the businesses leveraging this opportunity is makeup brands. These brands often employ beauty influencers like Tasya Farasya to endorse their products. This study aims to analyze the impact of makeup product endorsements by Tasya Farasya through audience comments on Instagram. The method used is sentiment analysis with Naïve Bayes and Support Vector Machine (SVM) algorithms to classify 608 comments into three sentiment categories: positive, neutral, and negative. The results show that the SVM algorithm performs better with an F1-Score of 91%, compared to Naïve Bayes which only reaches 80%. The makeup product endorsements by Tasya Farasya on Instagram have a positive impact on the brand, as evidenced by the comments being predominantly positive compared to neutral and negative sentiments.

Article History

Submitted: 18 Juli 2024

Accepted: 24 Juli 2024

Published: 25 Juli 2024

Key Words

Sentiment analysis, Instagram, Endorsement, Naïve Bayes, Support Vector Machine.

Abstrak

Peningkatan penggunaan media sosial Instagram membuat Instagram tidak lagi hanya menjadi tempat untuk mencari hiburan, tetapi juga menjadi salah satu tempat pemasaran bagi bisnis. Salah satu bisnis yang memanfaatkan hal tersebut adalah makeup brand. Makeup brand ini sering menggunakan jasa beauty influencer seperti Tasya Farasya untuk endorse produk mereka. Penelitian ini bertujuan untuk menganalisis dampak endorsement produk makeup oleh Tasya Farasya melalui komentar audiens di Instagram. Metode yang digunakan ialah analisis sentimen dengan algoritma Naïve Bayes dan Support Vector Machine (SVM) untuk mengklasifikasikan 608 komentar ke dalam tiga kelas sentimen: positif, netral, dan negatif. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa lebih baik dengan F1-Score sebesar 91%, dibandingkan Naïve Bayes yang hanya mencapai 80%. Endorsement produk makeup yang dilakukan Tasya Farasya di Instagram memberikan dampak positif bagi brand, terbukti dari komentar yang didominasi oleh sentimen positif dibandingkan dengan sentimen netral dan negatif.

Sejarah Artikel

Submitted: 18 Juli 2024

Accepted: 24 Juli 2024

Published: 25 Juli 2024

Kata Kunci

Analisis sentimen, Instagram, Endorsement, Naïve Bayes, Support Vector Machine.

Pendahuluan

Tren peningkatan pengguna internet di Indonesia terus berlanjut. Internet kini sudah menjadi sumber referensi informasi dan berita yang paling utama dibandingkan media konvensional (Kementerian Komunikasi & Informatika). Berdasarkan Menurut (KataData, 2023) jumlah pengguna internet di Indonesia per Januari 2023 mencapai 213 juta orang, meningkat sebanyak 5,44% dari tahun sebelumnya. Pengguna internet di Indonesia paling banyak mengakses media sosial, salah satunya Instagram yang mana merupakan media sosial yang paling banyak

diakses ketiga sebesar 30,51%, setelah YouTube sebesar 65,41%, dan Facebook sebesar 60,24% (Asosiasi Penyelenggara Jasa Internet Indonesia, 2023). Media sosial tidak lagi hanya sekedar menjadi tempat untuk mencari hiburan. Namun, kini media sosial juga menjadi salah satu tempat pemasaran bagi bisnis. Ada berbagai macam cara yang dapat dilakukan untuk memasarkan produk melalui media sosial, mulai dari memasang iklan hingga menggunakan jasa influencer (Rahayu Pancaningsih, et al., 2022).

Seorang influencer sering kali bekerja sama dengan brand untuk melakukan endorsement produk. Setiap influencer memiliki fokus ketertarikan yang berbeda-beda, ada yang berfokus di bidang F&B, lifestyle, kecantikan, dll. Meningkatnya jumlah orang yang berprofesi sebagai influencer tidak terlepas dari perkembangan industrinya itu sendiri. Salah satu industri yang kini berkembang pesat ialah industri kecantikan kosmetik. Menurut (Wisnubrata, 2023), pada tahun 2021 berdasarkan data dari Badan Pusat Statistik (BPS), industri kosmetik di Indonesia mengalami pertumbuhan sebanyak 9,61%, dan pada tahun 2022 terdapat kenaikan jumlah perusahaan kosmetik sebanyak 20,6%. Industri kosmetik ini meliputi banyak kategori seperti makeup, skincare, bodycare, haircare, dan masih banyak lagi. Makeup merupakan salah satu yang paling banyak digunakan karena telah menjadi kebutuhan utama bagi para wanita (Afifah, 2022). Seorang influencer yang berfokus pada bidang kecantikan seperti makeup, skincare, bodycare disebut sebagai beauty influencer.

Ketika sebuah brand membutuhkan jasa seorang influencer untuk memasarkan produknya, brand ini harus mengetahui terlebih dahulu efektifitas yang dihasilkan dari kegiatan pemasaran yang telah dilakukan oleh influencer tersebut. Beauty influencer yang telah terkenal efektifitasnya sejak lama ialah Tasya Farasya. Ketenaran Tasya Farasya sebagai seorang beauty influencer tidak diragukan lagi dengan pengikut sebanyak 6,5 juta (per tanggal 22 Januari 2024) di akun Instagram milik pribadinya @tasyafarasya. Dengan jumlah pengikut yang begitu banyak dan keterampilannya dalam me-review suatu produk membuat banyak brand makeup ingin endorse produknya ke Tasya Farasya. Melalui honest review produk yang dibagikannya di akun Instagram pribadinya @tasyafarasya ketika mendapatkan endorsement dari suatu brand, sehingga membuat audiens yang melihatnya percaya dan ingin membeli produk tersebut. Tidak hanya itu, karena honest review yang dibagikannya dan terbukti terpercaya oleh audiens yang membeli produk yang dipromosikan, berujung memunculkan sebuah tren dengan tagline/ hashtag #tasyafarasyaapproved. Hashtag #tasyafarasyaapproved merupakan tanda untuk produk-produk yang telah di-review oleh Tasya Farasya kemudian terbukti bagus dan sesuai dengan klaim yang dituliskan oleh produk tersebut. Dengan boomingnya tren tersebut banyak brand yang ketika menjual produknya menyertakan tagline/ hashtag #tasyafarasyaapproved untuk meningkatkan ketertarikan calon pelanggan produknya.

Oleh karena itu, perlu dilakukannya analisis sentimen untuk mengetahui dampak endorsement produk makeup yang dilakukan oleh beauty influencer Tasya Farasya pada akun Instagram pribadinya. Dengan analisis sentimen nantinya komentar pada postingan akun Instagram pribadi Tasya Farasya akan diklasifikasikan ke dalam sentimen negatif, netral, dan

positif menggunakan algoritma Naïve Bayes dan Support Vector Machine untuk mengetahui algoritma mana yang lebih baik. Pemilihan algoritma Naïve Bayes didasari penelitian yang dilakukan oleh (Bayhaqy et al., 2018), metode klasifikasi Naïve Bayes menjadi pilihan terbaik untuk diterapkan pada dataset media sosial karena mampu memberikan hasil yang lebih akurat dan prediksi yang tepat. Sedangkan pemilihan algoritma Support Vector Machine, berdasarkan penelitian yang dilakukan sebelumnya oleh (Asshiddiqi & Lhaksmana, 2020) dengan membandingkan kedua algoritma Decision Tree dan Support Vector Machine, terbukti bahwa algoritma Support Vector Machine mendapatkan akurasi yang lebih baik sebesar 94.36% sedangkan Decision Tree sebesar 87.45%.

Tujuan dari penelitian ini adalah mengetahui bagaimana hasil evaluasi kinerja klasifikasi analisis sentimen dari algoritma Naïve bayes, mengetahui bagaimana hasil evaluasi kinerja klasifikasi analisis sentimen dari algoritma Support Vector Machine, mengetahui perbandingan algoritma mana yang lebih baik untuk klasifikasi analisis sentimen dampak endorsement produk makeup di akun Instagram beauty influencer Tasya Farasya., dan yang terakhir mengetahui dampak yang dihasilkan dari endorsement produk makeup yang dilakukan oleh beauty influencer Tasya Farasya berdasarkan komentar audiens atau followers pada setiap postingan endorsement di akun Instagram miliknya

Metode Penelitian

1. Lokasi & Objek

Penelitian ini dilakukan secara online dengan mengumpulkan data dari Instagram. Objek penelitian adalah seorang *beauty influencer*, dengan fokus pada analisis sentimen di kolom komentar feeds Instagram terkait dampak *endorsement* produk *makeup*.

2. Metode Penelitian

Dalam penelitian ini, peneliti akan menggunakan dua algoritma, yaitu Naïve Bayes classifier dan Support Vector Machine, untuk mengkategorikan data menjadi sentimen positif, netral, dan negatif.

3. Jenis & Sumber Data

Data yang digunakan dalam penelitian ini adalah data kualitatif, khususnya data tekstual berupa opini *audiens* atau *followers* terkait endorsement produk makeup. Sumber data dalam penelitian ini adalah akun media sosial Instagram @tasyafarasya, sehingga opini *audiens* atau *followers* ini dikategorikan sebagai data sekunder.

4. Pengambilan Data

Data untuk penelitian ini akan diambil dari postingan Instagram @tasyafarasya yang terkait dengan *endorsement* produk *makeup*, yang dipublikasikan dari tanggal 1 Januari 2024 hingga 29 Februari 2024. Hanya komentar berbahasa Indonesia dan bersifat eksplisit yang akan digunakan. Data dikumpulkan dengan *scraping* menggunakan tools APIFY Instagram Comments Scraper serta secara manual pada postingan *feeds endorsement* produk makeup di akun Instagram pribadi @tasyafarasya.

5. Metode Analisis Data

5.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah proses investigasi awal terhadap data untuk memahami fitur-fitur utama, kecenderungan, dan pola-pola dalam dataset.

5.2 Labeling

Dari data yang telah didapatkan dari proses *scraping* dan manual, selanjutnya akan masuk kedalam proses *labeling* data secara manual yang dibagi ke dalam 3 kategori, yaitu positif, netral, negatif.

5.3 Data Preprocessing

5.3.1 Data Cleansing

Pada tahap data cleansing ini akan dilakukan *remove mention* (@username), untuk memastikan fokus tetap pada konten utama komentar tanpa terganggu oleh informasi yang tidak relevan terkait pengguna yang disebutkan. Selanjutnya akan dilakukan juga *remove punctuation* dan *case folding*.

5.3.2 Normalisasi

Normalisasi adalah proses mengubah data dengan variasi dalam penulisan kata menjadi bentuk yang seragam, meskipun maknanya tetap sama (Manning, et al., 2009).

5.3.3 Stopword Removal

Stopword removal adalah proses menghapus kata-kata umum yang dianggap tidak memberikan nilai tambah dalam pemrosesan atau analisis teks.

5.3.4 Tokenizing

Menurut (Shah et al., 2021) *tokenizing* adalah cara untuk memisahkan kata-kata dalam teks/kalimat menjadi bagian-bagian yang lebih kecil yang disebut token.

5.3.5 TF-IDF

Pembobotan kata menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF) adalah teknik dalam analisis teks yang digunakan untuk menilai pentingnya sebuah kata dalam suatu dokumen atau kumpulan dokumen.

6. Evaluasi

Evaluasi dalam penelitian ini akan dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* membantu mengukur kinerja model atau algoritma dalam memprediksi atau mengklasifikasikan data baru. Matriks yang akan digunakan dalam penelitian ini meliputi nilai *precision*, *recall*, dan *F1-score*. Sementara itu, nilai *accuracy* tidak akan digunakan untuk mengevaluasi model.

Hasil dan Pembahasan

1. Pengumpulan Data

Penelitian ini dimulai dengan mengumpulkan komentar mengenai endorsement produk makeup di akun Instagram @tasyafarasya dari 1 Januari 2024 hingga 29 Februari 2024. Dalam periode tersebut, terdapat 7 postingan terkait endorsement produk makeup. Data dikumpulkan dengan dua metode: menggunakan tools APIFY Instagram Comments

Scraped and manually. Results from both methods were successful in collecting a total of 1,703 Instagram comments.

Tabel 4. 1 Rincian jumlah data pada 7 postingan

| Postingan | Jumlah Komentar |
|-------------|-----------------|
| Postingan 1 | 143 |
| Postingan 2 | 385 |
| Postingan 3 | 528 |
| Postingan 4 | 104 |
| Postingan 5 | 251 |
| Postingan 6 | 126 |
| Postingan 7 | 166 |

2. Data Selection & Labeling

2.1 Data Selection

All data that has been collected was selected manually based on criteria that has been determined by the researcher. Selection criteria include two main aspects, namely comments written in Indonesian and comments that are explicit. After the selection process is carried out, there are a total of 612 comments that are saved in Excel format, which are then ready to be given labels.

Tabel 4. 2 Rincian jumlah data pada 7 postingan sesudah seleksi

| No. | Jumlah komentar yang sudah terseleksi |
|-------------|---------------------------------------|
| Postingan 1 | 103 |
| Postingan 2 | 91 |
| Postingan 3 | 84 |
| Postingan 4 | 95 |
| Postingan 5 | 83 |
| Postingan 6 | 84 |
| Postingan 7 | 72 |

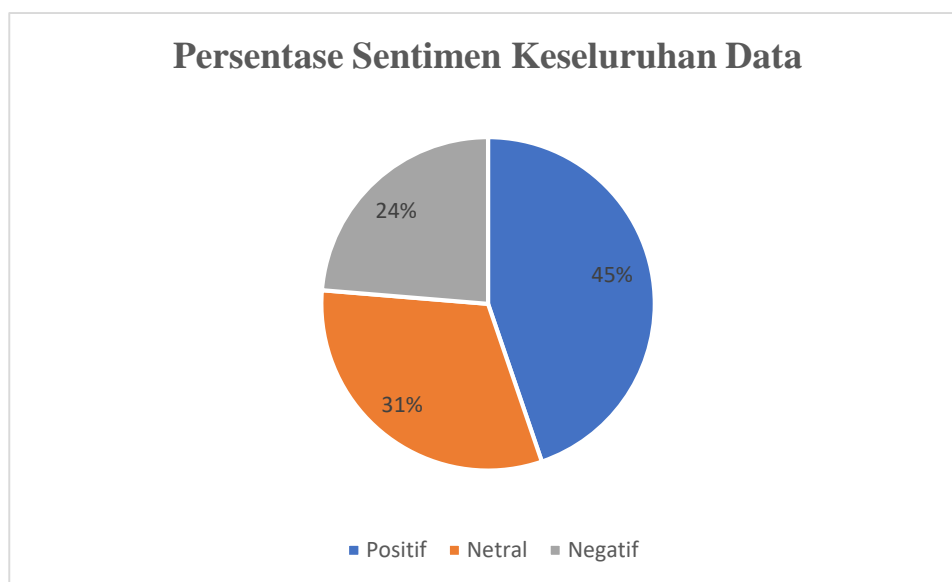
2.2 Labeling

After the data selection process is complete, the labeling process is carried out. Data is classified into three categories: positive, negative, and neutral. Labeling is carried out manually by the researcher by considering context and keywords in

setiap komentar. Misalnya, komentar yang memuji produk atau memberikan respon positif diberi label positif, komentar yang menunjukkan ketidakpuasan atau kritik diberi label negatif, dan komentar yang tidak menunjukkan sentimen jelas atau bersifat informatif tanpa emosi diberi label netral.

Tabel 4. 3 Contoh pemberian label sentimen secara manual

| Text | Sentimen |
|--|----------|
| Eyeshadow dan blushon nya emg sebgus itu woy mana warna nya bgus ² dan shof bngt | Positif |
| Lipglossnya pake apaa | Netral |
| setelah beberapa jam malah crack parah dan luntur kecewa berat | Negatif |



Gambar 4. 1 Persentase sentimen data

Persentase sentimen dari keseluruhan data menunjukkan bahwa kelas sentimen positif mendominasi dengan 45% atau berjumlah 274 data. Sentimen terbanyak kedua adalah kelas sentimen netral memiliki persentase paling sedikit, yaitu 31% atau sebanyak 193 data. Terakhir, kelas sentimen negatif yang mencakup 24% atau sebanyak 145 data.

3. *Exploratory Data Analysis (EDA)*

Exploratory Data Analysis (EDA) adalah proses awal untuk memahami karakteristik dataset. Setelah melakukan EDA, ditemukan beberapa masalah yang memerlukan pembersihan data lebih lanjut, seperti keberadaan emoji, tanda baca yang tidak diperlukan, duplikasi data, dan kata-kata yang perlu dinormalisasi. Selain itu, beberapa kata yang tidak diperlukan harus dihapus melalui proses *stopword removal*.

4. Data Pre-processing

4.4.1 Data Cleansing

Semua proses pada *data cleansing* dilakukan menggunakan software KNIME. Tahap awal dalam *data cleansing* adalah menghapus komentar duplikat dengan menggunakan node Duplicate Row Filter, sehingga tidak ada dua atau lebih komentar yang sama dalam dataset. Proses ini sangat penting untuk menghindari bias pada data. Setelah proses ini dilakukan, jumlah data yang semula sebanyak 612 berubah menjadi 607 data.

Langkah berikutnya adalah menghapus mention, yaitu referensi kepada pengguna yang ditandai dengan simbol '@'. Langkah ini penting untuk memastikan fokus analisis tetap pada konten utama komentar tanpa gangguan informasi pengguna yang tidak relevan. Proses ini dilakukan menggunakan node String Manipulation. Tabel 4.4 di bawah ini menunjukkan contoh hasil dari proses remove mention.

Tabel 4. 4 Contoh sebelum dan setelah *remove mention*

| Sebelum <i>remove mention</i> | Setelah <i>remove mention</i> |
|--|--|
| Pecinta skintintnyaaa YOu always repeat order!!! @youbeauty_idn | Pecinta skintintnyaaa YOU always repeat order!!! |
| @luciamonicaa samaaa diaku cakeep pake tipis aja lumayan cover dan emng seawwweddddddh itu | samaaa diaku cakeep pake tipis aja lumayan cover dan emng seawwweddddddh itu |

Setelah *remove mention*, dilakukan *remove punctuation* dan *case folding* secara bersamaan menggunakan node Text Preprocessing. Proses ini menghapus tanda baca dan mengubah semua huruf menjadi kecil. Selain itu, emoji seperti "👉", "❤️", dan lainnya juga dihapus. Tanda baca dan emoji sering kali tidak memberikan nilai informasi yang signifikan dalam analisis.

Tabel 4. 5 Contoh sebelum dan setelah *remove punctuation* dan *case folding*

| Sebelum <i>remove punctuation & case folding</i> | Setelah <i>remove punctuation & case folding</i> |
|--|--|
| Kalo mama ayang udh review tandanya apa???? Beli!!!! | kalo mama ayang udh review tandanya apa beli!!!! |
| Pngen banget cobaaa | pngen banget cobaaa |

4.4.2 Normalisasi

Tahap normalisasi penting untuk menyatukan variasi kata, termasuk kata gaul dan singkatan, ke dalam bentuk standar. Proses ini dilakukan menggunakan Google Colab, dimulai dengan pembuatan kamus normalisasi secara manual.

Kamus ini dibuat berdasarkan dataset yang telah dibersihkan untuk memastikan konsistensi dan akurasi data. Hasil analisis menunjukkan ada 935 kata yang memerlukan normalisasi untuk meningkatkan kualitas data.

Tabel 4. 6 Contoh sebelum dan setelah normalisasi kata

| Sebelum normalisasi | Setelah normalisasi |
|---------------------|---------------------|
| foundie | foundation |
| syuka | suka |

4.4.3 Stopword Removal

Stopword removal menghapus kata-kata umum yang sering muncul dalam teks tetapi tidak memberikan nilai informasi yang signifikan bagi analisis. Dalam penelitian ini, *stopword removal* dilakukan menggunakan Google Colab dan kamus stopwords Bahasa Indonesia dari akun GitHub bernama genediazjr. Kata-kata dalam kamus tersebut diubah menjadi string kosong dan dihapus dari teks.

Tabel 4. 7 Contoh sebelum dan setelah stopwords removal

| Sebelum <i>stopword removal</i> | Setelah <i>stopword removal</i> |
|--|---|
| mukanya engga rusak apa ya ganti ganti makeup terus | mukanya engga rusak ganti ganti makeup |
| oksidasi banget kecewa | oksidasi kecewa |

4.4.4 Tokenizing

Tokenizing adalah proses memecah sebuah teks menjadi kata-kata individu yang disebut token. Untuk melakukan proses *tokenizing* digunakanlah software Google Colab, dengan library scikit-learn. Scikit-learn adalah library Python yang menyediakan alat-alat sederhana dan efisien untuk data mining dan analisis data.

```
'memakai' 'membantu' 'membeli' 'membersihkan' 'memburuk' 'menang'
```

Gambar 4. 2 Hasil tokenizing

4.4.5 TF-IDF

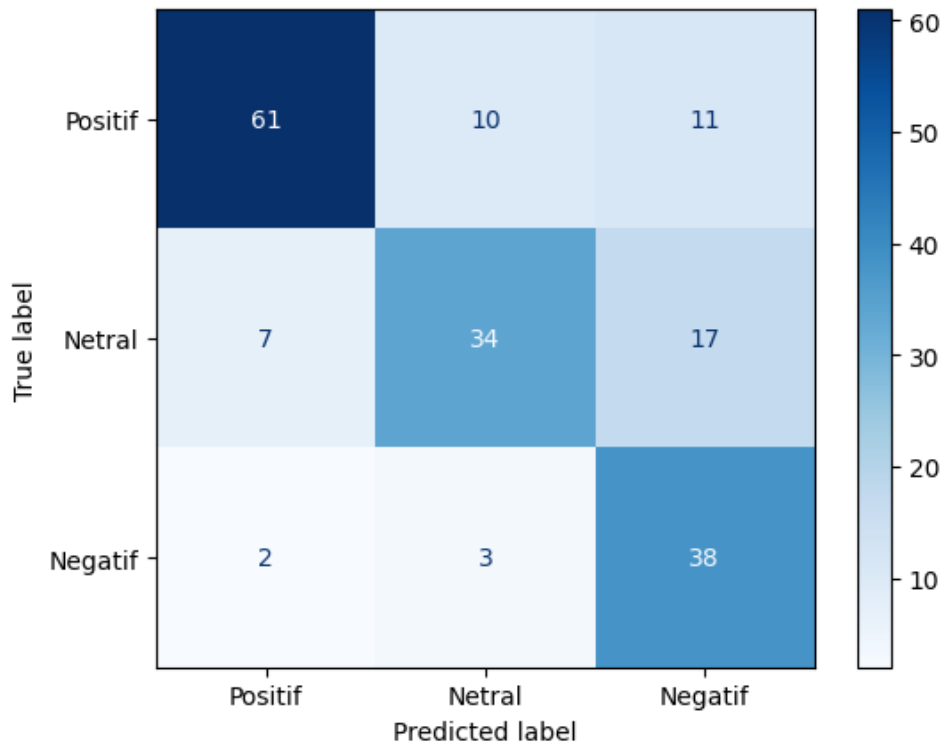
Pembobotan TF-IDF digunakan untuk mengidentifikasi kata-kata kunci paling penting dalam sebuah dokumen atau koleksi dokumen. Kata-kata dengan pembobotan TF-IDF tinggi cenderung memiliki makna lebih penting atau spesifik dalam konteks teks yang dianalisis. Proses pembobotan ini dilakukan menggunakan Google Colab dengan modul TF-IDF Vectorizer dari library scikit-learn.

5. Klasifikasi Algoritma Naive Bayes Classifier

Setelah data pre-processing, tahap berikutnya adalah data processing dengan menerapkan model klasifikasi menggunakan algoritma Naive Bayes di Google Colab.

Langkah pertama adalah membagi dataset menjadi dua bagian: data latih dan data uji, dengan perbandingan 70:30. Artinya, 70% data digunakan untuk pelatihan dan 30% sisanya untuk pengujian, menggunakan metode stratified sampling.

Setelah membagi data, langkah berikutnya data latih digunakan untuk melatih model dengan menghitung probabilitas setiap fitur untuk setiap kelas. Model yang dihasilkan berisi informasi probabilistik untuk mengklasifikasikan data baru. Selanjutnya, performa model diuji dengan data uji, yang merupakan 30% dari total data. Evaluasi dilakukan dengan menghitung metrik *precision*, *recall*, dan *F1-score*.



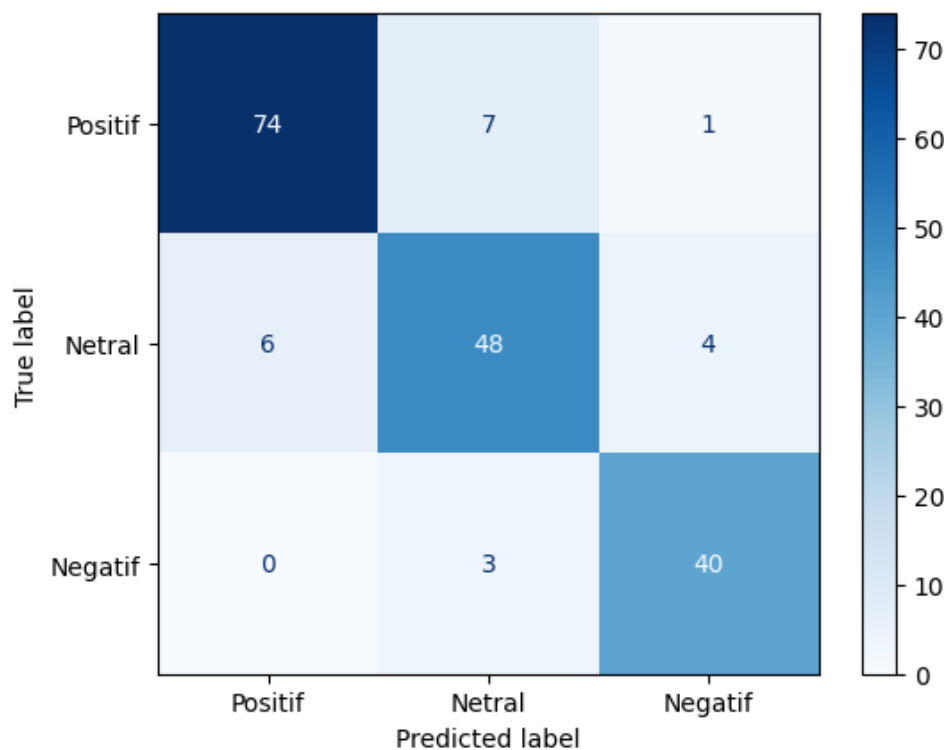
Gambar 4. 3 Confusion Matrix 3x3 Klasifikasi Algoritma Naïve Bayes

6. Klasifikasi Algoritma Support Vector Machine

Untuk mengimplementasikan algoritma Support Vector Machine (SVM), penulis menggunakan Google Colab, yang menawarkan sumber daya komputasi kuat dan akses langsung ke berbagai library di cloud, sehingga meningkatkan efisiensi implementasi dan evaluasi model. Langkah pertama adalah membagi dataset menjadi data latih dan data uji dengan perbandingan 70:30 menggunakan stratified sampling. Dengan cara ini, 70% data digunakan untuk pelatihan dan 30% untuk pengujian. Pembagian ini penting untuk memastikan model diuji secara independen dengan data yang belum pernah dilihat sebelumnya.

Setelah membagi data, langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear. Model SVM dilatih dengan data latih, di mana parameter model disesuaikan untuk memaksimalkan margin antar kelas. Kernel linear dipilih karena efektif untuk data yang dapat dipisahkan secara linear dan lebih cepat dalam komputasi dibandingkan kernel non-

linear. Selanjutnya, model dievaluasi menggunakan data uji, yang memberikan gambaran akurat tentang kinerja model pada data yang belum pernah dilihat sebelumnya.



Gambar 4. 4 Confusion Matrix 3x3 Klasifikasi Algoritma Suppor Vektor *Machine*

7. Evaluasi

7.1 Evaluasi Kinerja Klasifikasi Algoritma Naive Bayes

Tabel 4. 8 Evaluasi kinerja klasifikasi Naive Bayes

| Naive Bayes | Precision | Recall |
|-------------|-----------|--------|
| Positif | 87% | 74% |
| Netral | 72% | 59% |
| Negatif | 58% | 88% |

Untuk kelas sentimen positif, nilai *precision* adalah 87%, yang berarti 87% dari semua prediksi positif yang dibuat oleh model benar-benar positif. Nilai *recall* sebesar 74% menunjukkan bahwa model mampu mengidentifikasi 74% dari data yang sebenarnya positif. Hal ini terjadi karena dalam kelas sentimen positif terdapat kata-kata dengan makna negatif atau netral, yang menyebabkan model memprediksi komentar sebagai sentimen negatif atau netral. Contohnya, komentar seperti "studio tropic bagus" dan "fokus lipstick cantik" seharusnya positif karena kata "bagus" dan "cantik", namun model memprediksi sebagai sentimen netral

karena kata "studio tropic" (nama *brand*) dan "lipstik" (produk *makeup*) dianggap netral.

Untuk kelas sentimen netral, nilai *precision* adalah 72%, berarti 72% dari semua prediksi netral oleh model benar-benar netral. Nilai *recall* sebesar 59% menunjukkan bahwa model berhasil mengidentifikasi 59% dari data yang sebenarnya netral. Hal ini disebabkan oleh adanya kata-kata dengan makna positif atau negatif dalam kelas netral, yang membuat model salah memprediksi sentimen sebagai positif atau negatif. Contohnya, komentar "udah pakai biar engga cepet habis" seharusnya netral karena bersifat informatif tanpa emosi, namun model memprediksi sebagai sentimen negatif karena kata "engga" dianggap negatif.

Untuk kelas sentimen negatif, nilai *precision* adalah 58%, berarti hanya 58% dari prediksi negatif model yang benar-benar negatif. Namun, nilai *recall* sebesar 88% menunjukkan bahwa model berhasil mengidentifikasi 88% dari data yang sebenarnya negatif. Kesalahan ini terjadi karena komentar negatif sering kali mengandung kata-kata dengan makna netral atau positif, yang menyebabkan model salah memprediksi sentimen sebagai netral atau positif. Contohnya, komentar seperti "engga awet dipakai makan" dan "engga aman kulit berjerawat kecewa" seharusnya negatif, tetapi diprediksi sebagai netral karena kata "awet" dan "aman" dianggap positif oleh model, sehingga model mengalami kebingungan.

7.2 Evaluasi Kinerja Klasifikasi Algoritma Support Vector Machine

Tabel 4. 9 Evaluasi kinerja klasifikasi Support Vector Machine

| SVM | Precision | Recall |
|---------|-----------|--------|
| Positif | 93% | 90% |
| Netral | 83% | 83% |
| Negatif | 89% | 93% |

Untuk kelas sentimen positif, nilai *precision* mencapai 93%, berarti 93% dari prediksi positif model benar-benar positif. Nilai *recall* sebesar 90% menunjukkan model berhasil mengidentifikasi 90% data positif dengan benar. Kesalahan terjadi pada komentar seperti "studio tropic bagus," di mana kata "studio tropic" dianggap netral oleh model, padahal itu adalah nama *brand* kecantikan.

Selanjutnya, untuk kelas sentimen netral, nilai *precision* adalah 83%, berarti 83% dari prediksi netral model benar-benar netral. Nilai *recall* sebesar 83% menunjukkan model mampu mengidentifikasi 83% data netral dengan benar. Kesalahan klasifikasi pada sentimen netral disebabkan oleh kata "engga" yang dianggap negatif oleh model, padahal komentar "pertimbangan oke engga dikulit" sebenarnya netral karena tidak menunjukkan sentimen jelas.

Terakhir, untuk kelas sentimen negatif, nilai *precision* sebesar 89% berarti 89% dari prediksi negatif model benar-benar negatif. Nilai *recall* sebesar 93% menunjukkan model berhasil mengidentifikasi 93% data negatif dengan benar. Kesalahan terjadi karena stopword removal menghapus kata "biasa" dari komentar

seperti “lipcreamnya biasa saja,” sehingga komentar hanya tersisa sebagai “lipcreamnya,” yang dianggap netral karena tidak menunjukkan sentimen jelas.

7.3 Perbandingan Kinerja Klasifikasi Algoritma Naive Bayes dan Support Vector Machine

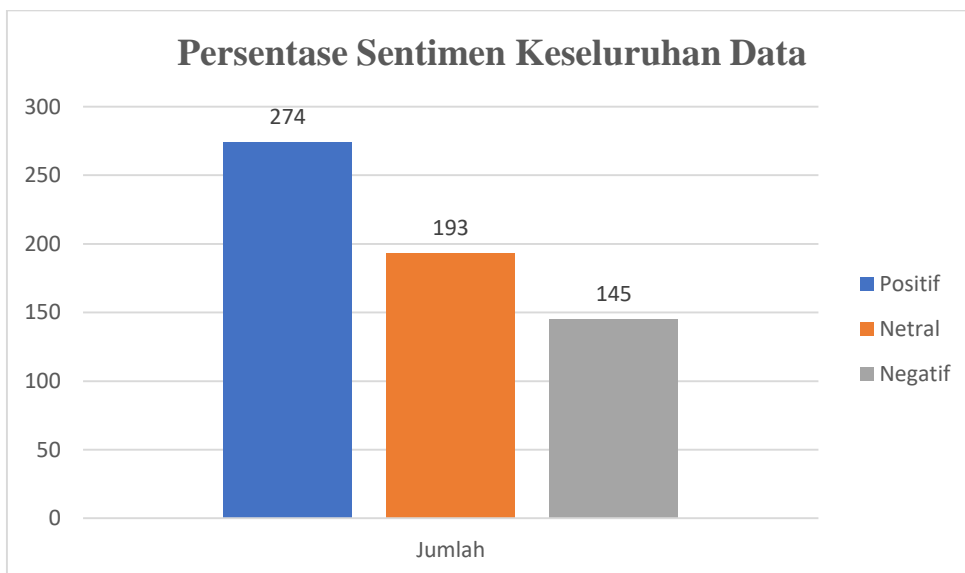
Tabel 4. 10 Perbandingan Nilai F1-Score Naive Bayes & Support Vector Machine

| Kelas Sentimen | Naive Bayes | SVM |
|----------------|-------------|-----|
| Positif | 80% | 91% |
| Netral | 65% | 83% |
| Negatif | 70% | 91% |

Tabel 4.10 menunjukkan perbandingan F1-Score antara algoritma Naive Bayes dan Support Vector Machine (SVM) untuk tiga kelas sentimen: positif, netral, dan negatif. Untuk sentimen positif, SVM mencapai F1-Score 91%, sedangkan Naive Bayes hanya 80%. Untuk sentimen netral, SVM memperoleh 83%, sementara Naive Bayes mendapatkan 65%. Pada sentimen negatif, SVM memperoleh 91%, sedangkan Naive Bayes hanya 70%.

Berdasarkan nilai F1-Score, algoritma SVM menunjukkan kinerja yang lebih baik dibandingkan Naive Bayes. Hal ini sejalan dengan penelitian yang dilakukan (Yunita & Kamayani, 2023) menunjukkan bahwa kinerja SVM lebih unggul dibandingkan Naive Bayes.

7.4 Dampak Endorsement yang dilakukan Tasya Farasya



Gambar 4. 5 Persentase sentimen keseluruhan data

(SVM) lebih efektif dalam mengurangi prediksi yang salah. Oleh karena itu, Support Vector Machine (SVM) terbukti lebih unggul.

Endorsement produk makeup oleh Tasya Farasya di akun Instagram pribadinya berdampak sangat positif bagi brand, terbukti dari dominasi komentar dengan sentimen positif dibandingkan netral dan negatif. Dampak positif ini mencakup peningkatan citra produk dan peningkatan minat beli di kalangan audiens atau followers.

Referensi

- Afifah, F. R. (2022). *Pengaruh Promosi, Desain dan Kualitas Produk terhadap Keputusan Pembelian pada Maskara Maybelline*.
- Annur, C. M. (n.d.). *Pengguna Internet di Indonesia Tembus 213 Juta Orang hingga Awal 2023*. 2023. Retrieved October 27, 2023, from <https://databoks.katadata.co.id/datapublish/2023/09/20/pengguna-internet-di-indonesia-tembus-213-juta-orang-hingga-awal-2023>
- Asosiasi Penyelenggara Jasa Internet Indonesia. (2023). *Survei Penetrasi dan Perilaku Internet 2023*. <https://apjii.or.id/>
- Asshiddiqi, M. F., & Lhaksmana, K. M. (2020). *Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI*. 7(3).
- Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies, ICOT 2018*, 1–6. <https://doi.org/10.1109/ICOT.2018.8705796>
- Kementerian Komunikasi & Informatika. (n.d.). *Kementerian Komunikasi dan Informatika*. Retrieved October 27, 2023, from https://www.kominfo.go.id/index.php/content/detail/5421/Kemkominfo%3A+Internet+Jad+Referensi+Utama+Mengakses+Berita+dan+Informasi/0/berita_satker
- Pancaningsih, R., Habiibika Baaq, S., Nurul Latifah, R., & Nur Wahit, A. (2022). *Pengaruh Media Sosial Terhadap Penjualan Produk Kecantikan Pada Online Shop Kumara Store Di Sukoharjo*. 1(4), 134–148.
- Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N., & Shah, N. K. (2021). Sentiments Detection for Amazon Product Review. *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*. <https://doi.org/10.1109/ICCCI50826.2021.9402414>
- Wisnubrata. (2023). *Seperti Apa Tren Kecantikan di Indonesia pada Tahun 2023?* Kompas. <https://lifestyle.kompas.com/read/2023/02/07/111401620/seperti-apa-tren-kecantikan-di-indonesia-pada-tahun-2023>
- Yunita, R., & Kamayani, M. (2023). Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi. *Indonesian Journal of Computer Science*, 12(5), 2879–2890. <https://doi.org/10.33022/ijcs.v12i5.3415>