



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 1 No. 1 (2022) pp: 3918-3927

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Analisis Sentimen Komentar Instagram terhadap Isu PMS menggunakan SVM

Nurul Humaera B.<sup>1</sup>, Farida<sup>2</sup>, Musliadi KH<sup>3</sup>

<sup>1,2</sup> Sistem Informasi, Institut Teknologi dan Bisnis Bina Adinata

<sup>3</sup> Teknik Informatika, Fakultas Komputer, Universitas Universal

<sup>1</sup>[humairah777@gmail.com](mailto:humairah777@gmail.com), <sup>2</sup>[faridahvarvd4@gmail.com](mailto:faridahvarvd4@gmail.com), <sup>3</sup>[musliadiKH@gmail.com](mailto:musliadiKH@gmail.com)

### Abstrak

Perkembangan media sosial sebagai ruang publik membentuk persepsi masyarakat terhadap berbagai isu sosial dan Kesehatan. Isu perubahan emosi menjelang menstruasi atau PMS (Premenstrual Syndrome) menjadi salah satu isu yang mendapat banyak perhatian pada media sosial Instagram. Tingkat interaksi pengguna pada konten viral bertema PMS menunjukkan beragam respon emosional, namun disisi lain topik ini masih sering disalahpahami dan distigmatisasi. Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna Instagram terhadap meme viral yang membahas perubahan emosi menjelang menstruasi menggunakan algoritma Support Vector Machine dengan Stemming Sastrawi. Sebanyak 2760 komentar yang terkumpul setelah melalui tahapan pre-processing yang meliputi case folding, pembersihan teks, normalisasi slang, penanganan emoji, tokenisasi, stopword removal, stemming Sastrawi dan penerjemahan bahasa, diperoleh 1201 komentar yang layak dianalisis. Proses pelabelan sentimen dilakukan secara semi-otomatis dengan bantuan LLM (Large Language Model) yang kemudian divalidasi dengan pelabelan manual oleh peneliti. Tingkat realibilitas pada pelabelan semi-otomatis dengan Uji Cohen's Kappa menunjukkan nilai kesepakatan yang kuat sebesar 0,8513, sehingga data dinilai layak untuk tahap klasifikasi. Pembagian data latih sebesar 80% dan data uji 20%. Hasil penelitian menunjukkan bahwa akurasi model SVM (Support Vector Machine) mencapai nilai 85% dalam mengklasifikasikan komentar. Penelitian membuktikan bahwa pendekatan SVM dengan stemming Sastrawi efektif dalam menganalisis sentimen pada media sosial yang bersifat Informal dan mendukung pengembangan strategi edukasi dan komunikasi kesehatan berbasis data untuk mengurangi stigma terhadap isu PMS diruang digital.

**Kata kunci:** Analisis Sentimen, Premenstrual Syndrome, SVM, Instagram

### 1. Latar Belakang

Media sosial menjadi gaya hidup manusia modern sejak dua dekade terakhir. Perubahan cara interaksi dan berbagi informasi menyebabkan masyarakat modern secara tidak langsung mengubah cara berkomunikasi. Komunikasi instan dengan beragam orang diseluruh dunia, memutuskan jarak geografis. Kemudahan akses melalui fitur berbagi teks, video, foto dan komentar sehingga tetap terhubung satu sama lain membentuk pola komunikasi baru dimasyarakat modern. Media sosial bukan lagi sekedar hiburan tapi menjadi ruang diskusi publik yang aktif dalam membentuk opini sosial [1].

Peran besar media sosial dalam mempengaruhi dan membentuk opini publik dimasa kini semakin signifikan. Kemudahan akses membuat media sosial menjadi dasar terbentuknya berbagai gerakan nasional dan tren budaya yang mendorong perubahan dalam asturan bersosial secara massal. Berawal dari sebuah postingan yang dikomentari sehingga memunculkan beragam pandangan lain. Oleh karena itu, media sosial berperan besar dalam membentuk persepsi publik terhadap isu tertentu [2].

Instagram merupakan salah satu *platform* media sosial yang paling populer didunia [3]. Instagram menjadi salah satu penentu arah perhatian publik. Berbagai isu viral ada yang berasal dari konten Instagram. Konten viral menyoroti berbagai isu atau topik, tren yang sedang marak, berbagi pengalaman atau kebiasaan di kehidupan sehari hari pengguna. Konten yang disajikan pun menggugah berbagai reaksi bagi masyarakat yang diimplementasikan dengan berbagai jenis komentar [4]. Berbagai isu viral kerap muncul dan berkembang melalui Instagram, menjadikannya sebagai salah satu penentu arah perhatian publik terhadap isu-isu tertentu.

Salah satu konten viral Instagram yang menuai banyak reaksi dari pengguna adalah konten *meme* [5] yang memparodikan bagaimana perubahan emosi saat menjelang menstruasi (PMS) pada wanita melalui sebuah potongan film. Tingginya interaksi pengguna menunjukkan bahwa topik menstruasi khususnya perubahan emosi

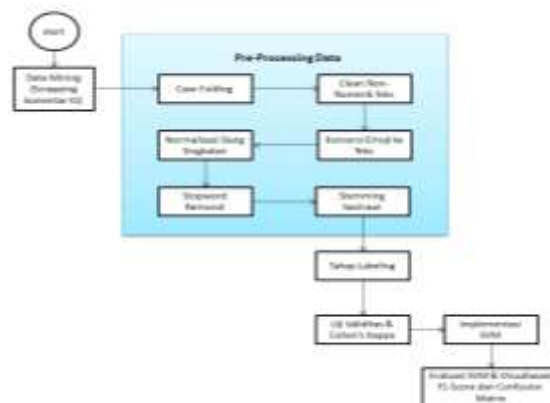
menjelang menstruasi mendapatkan perhatian besar dari publik. Namun disisi lain, topik PMS ini masih dianggap tabu, diremehkan, atau disalahpahami dalam kehidupan sosial sehari-hari [6]. Kondisi tersebut menunjukkan adanya kebutuhan untuk memahami bagaimana persepsi dan sikap publik terhadap isu perubahan emosi menjelang menstruasi dalam ruang digital. Media sosial dapat menjadi sarana, khususnya wanita muda dapat mengekspresikan, memvalidasi atau mencari dukungan terkait pengalaman emosional mereka secara kolektif di *platform* sosial media. Oleh karena itu, analisis terhadap pengguna di Instagram menjadi tujuan peneliti agar dapat memberikan gambaran mengenai apakah lingkungan cenderung bersifat suportif, netral atau justru menghakimi terhadap isu PMS. Hasil analisis ini berpotensi menjadi masukan bagi program edukasi kesehatan dalam upaya mengatasi stigma atau misinformasi dari PMS.

Komentar pada sosial media, khususnya Instagram umumnya ditulis menggunakan bahasa informal, kaya akan *slang*, bahasa gaul, emoji atau variasi penulisan yang tidak baku [7]. Karakteristik bahasa ini menjadi tantangan dalam proses analisis sentimen. Terdapat dua jenis metode analisis sentimen yang paling sering digunakan yakni *lexicon based* dan *machine learning* [8]. Penelitian sebelumnya masih terbatas dalam mengkaji analisis sentimen terhadap isu persepsi publik dengan pendekatan *lexicon based* dalam hal ini objeknya menggunakan data komentar Twitter berbahasa informal, dimana model yang dihasilkan memiliki keterbatasan dalam mengenali konteks sarkasme dan negasi bahasa informal [9] sehingga, diperlukan pendekatan komputasional lain yang lebih mampu menangani kompleksitas bahasa tersebut dalam menganalisis persepsi publik. Beberapa penelitian secara umum menggunakan *Naive bayes* dan Support Vector Machine (SVM) [10]. Penelitian analisis sentimen oleh Tetrian (2023) yang mengkomparasikan metode *Naive bayes* dan SVM pada RUU Kesehatan di Twitter menyatakan bahwa metode SVM memiliki akurasi yang lebih tinggi dengan nilai 87% [11]. Sejalan dengan penelitian Raditya dan Hasbi (2025) yang mengklasifikasikan konten Instagram dengan metode SVM, mendapatkan nilai akurasi yang lebih tinggi dengan nilai 96% [12]. Nanda dan Arief (2025) melakukan perbandingan analisis sentimen antara metode *Naive bayes* dan SVM menghasilkan temuan bahwa algoritma SVM lebih akurat 89,23% dari *Naive bayes* dengan akurasi 73,14% [13]. Kemampuan SVM dalam memisahkan kelas sentimen menjadikan model handal dalam menganalisis sentimen di media sosial. Meskipun begitu, dibutuhkan algoritma pendekatan yang mendukung tahap-tahap analisis menggunakan SVM agar lebih akurat, tahap yang krusial adalah *stemming*. Penelitian yang dilakukan oleh Daffa dkk.(2024), membandingkan 3 jenis algoritma pada proses *stemming* pada *Naive bayes* dan SVM, hasil menunjukkan metode SVM dengan *stemming* Sastrawi menghasilkan akurasi tertinggi 91% [14].

Berdasarkan permasalahan dan tinjauan penelitian terdahulu, penelitian ini bertujuan untuk menganalisis sentimen komentar Instagram terhadap konten yang membahas perubahan emosi menjelang menstruasi menggunakan algoritma SVM dengan *stemming* Sastrawi. Algoritma SVM dipilih karena terbukti memiliki kemampuan dalam mengklasifikasikan teks secara efektif pada data berdimensi tinggi. Penelitian ini diharapkan dapat memberikan kontribusi dalam pemahaman publik terhadap isu PMS media sosial yang menjadi dasar dalam pengembangan strategi edukasi dan komunikasi kesehatan yang berbasis data.

## 2. Metode Penelitian

Penelitian dilakukan dengan rangkaian langkah terstruktur untuk mengevaluasi sentimen analisis pada isu perubahan emosi menjelang menstruasi (PMS) menggunakan metode SVM dengan pendekatan algoritma *stemming* Sastrawi. Gambar 1 menampilkan tahap eksperimen yang digunakan pada penelitian analisis sentimen ini.



Gambar 1. Tahapan Analisis Sentimen

## 2.1 Data Mining

Penelitian ini diawali dengan tahap pengumpulan data melalui proses *scraping* komentar menggunakan pihak ketiga pada *platform* media sosial Instagram yang berkaitan dengan konten *reels* perubahan emosi menjelang menstruasi (PMS) dengan tetap memperhatikan kebijakan privasi dan etika penggunaan data.

## 2.2 Pre-processing Data

Tahap *Case folding* dilakukan dengan menyeragamkan semua teks komentar menjadi huruf kecil atau *lowercase* sehingga setiap kata mendapat perlakuan yang sama dari model yang dibuat. Data kemudian masuk ke tahap pembersihan karakter bukan teks dilakukan dengan mengolah data hasil *case folding* untuk text *pre-processing* sehingga data yang akan digunakan lebih bersih dan berkualitas. Bersih disini diartikan dengan karakter yang tidak relevan untuk sentimen akan dieliminasi, seperti link, angka, tanda baca, *username* instagram, spasi, angka atau huruf yang tidak perlu/berlebihan/ganda seperti "tttakutt" menjadi "takut" karena bagian ini akan menjadi *noise* dan tidak membawa nilai sentimen.

Keunikan komentar Instagram yang seringkali menggunakan emoji sebagai bentuk ekspresi emosi juga ditangani. Perubahan symbol emosi menjadi representasi teks dilakukan secara manual. Hal ini dilakukan untuk mengantisipasi kesalahpahaman makna emoji, mengingat fenomena emoji oleh pengguna Instagram yang kadang mengartikan emoji senyum lebar sebagai sebuah sarkasme atau sindiran halus [15], menjadi perlu untuk dibuatkan kamus emoji dalam bentuk teks agar dapat dibaca oleh model.

Tahap normalisasi kemudian dilakukan dengan mengubah bahasa informal menjadi bahasa baku. Penambahan kamus *slang* buatan sendiri yang secara spesifik dibuat berdasarkan karakteristik kata pada komentar. Penambahan *slang* buatan sendiri dilakukan secara manual dengan memperhatikan keadaan data komentar. Misalnya, kata "bgt" akan diganti menjadi "banget", "gegara" ke "gara-gara" atau dalam kasus lain jika ditemukan kata yang tidak senonoh istilah "njir" dinormalisasi menjadi "sial" mengingat karakteristik data komentar Instagram yang unik menjadikan tahap ini krusial.

Pemecahan teks menjadi unit-unit kata dilakukan setelah normalisasi. Tahap ini menungkingkan tiap komentar hanya berisi list kata. Data komentar selanjutnya akan dihapus berdasarkan kata-kata yang sangat umum dan tidak berpengaruh. Misalnya, kata hubung "di", "untuk" kecuali, jika ditemukan kata negatif seperti "tidak" yang harus dipertahankan melalui tahap *stopword Removal* kata tersebut akan dieliminasi jika tidak memiliki bobot informasi semantik. Selanjutnya ke proses *stemming*, yakni mengembalikan kata berimbuhan ke kata dasar menggunakan algoritma Sastrawi dengan teknik Kamus *lookup* sehingga model yang dibuat dapat terjaga konsistensi fiturnya. *Stemming* dilakukan menggunakan *respository* Sastrawi berbasis Python dengan tujuan mengubah kata berimbuhan menjadi bentuk dasarnya sebagai contoh mengubah kata "terasa" menjadi "rasa", "bersedih" menjadi "sedih".

Tahap penerjemahan dilakukan dengan mengubah data komentar ke dalam bahasa Indonesia, agar analisis lebih berfokus pada hasil dalam bahasa Indonesia. Penggunaan bahasa inggris dan Indonesia secara bersamaan menjadi alasan dilakukan tahap penerjemahan menggunakan pustaka *Deep Translator* agar memastikan seluruh korpus data dalam bahasa Indonesia sebelum dilakukan analisis sentimen.

## 2.3 Pelabelan Data

Tahap pelabelan dilakukan untuk memisahkan kategori sentimen positif, negatif dan netral sebagai basis kebenaran (*Ground Truth*). Adanya batasan sumber daya, validasi pelabelan akan dilakukan melalui prosedur validasi silang. Validasi silang [10] dilakukan menggunakan LLM (*Large Language Model*) [10] dengan teknik labeling otomatis lalu membandingkan dengan hasil pelabelan manual [16]. Kemudian, Prosedur verifikasi pakar akan dilakukan oleh peneliti utama dengan melakukan pelabelan mandiri dengan panduan kodifikasi yang berdasar pada referensi kesehatan terkait perubahan emosi saat menstruasi [17]. Berikut pemetaan label.

Tabel 1. Pemetaan Kriteria Pelabelan

Kategori Label	Kriteria	Dasar Temuan Jurnal
Negatif	Marah, Frustrasi, Iritabilitas, Nyeri fisik, panggul, kram dan perubahan mood (mood swing)	Marah dan Iritabilitas (81.4%), Depresi (73.2%), Nyeri fisik/kram (80.2%)
Positif	<i>Copying Mechanism</i> (upaya penanganan, edukasi, perilaku yang mendukung pemulihan, humor	Kelaparan (Konteks kepuasan makan), tips mengurangi stress
Netral	Perubahan perilaku umum tanpa emosi yang kuat atau sekedar informasi actual	Perubahan hormon dan libido (43.5%), perubahan pola tidur/sosial

Validitas konten dalam proses pelabelan dilakukan dengan merujuk pada kriteria simptomatis yang dipetakan pada tabel yang sesuai dengan dipublikasikan dalam studi literatur kesehatan dan bukan sekedar intuisi subjektif. Penentuan label “Negatif” didasarkan pada gejala psiko-emosional seperti iritabilitas dan kemarahan yang mencapai 81.4%, serta nyeri fisik/kram sebesar 80.2% . Objektivitas pelabelan tetap terjaga meskipun dilakukan secara mandiri oleh peneliti karena merujuk pada data klinis.

Tabel 2. Nilai Validitas *Cohen's Kappa*

Nilai Kappa	Kekuatan Kesepakatan	Persentase Data yang Dapat Diandalkan
0 - 0,20	Tidak ada	0–4%
0,21 - 0,39	Minimal	4–15%
0,40 - 0,59	Lemah	15–35%
0,60 - 0,79	Sedang	35–63%
0,80 - 0,90	Kuat	64–81%
Di atas 90	Hampir Sempurna	82–100%

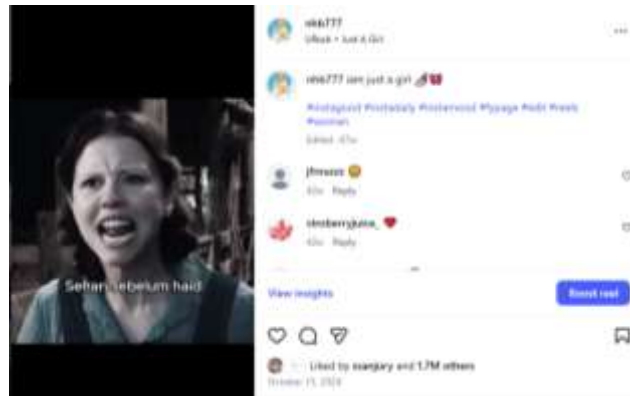
Penjaminan validitas hasil anotasi dilakukan dengan mengambil teknik *random sampling* sebesar 10% sampel. Pengambilan jumlah sampel disesuaikan dengan standar pengembangan dataset modern yang menekankan pada *Data-Centric AI* [18]. Tingkat reabilitas *annotator* diukur dengan melihat hasil verifikasi menggunakan Uji *Cohen's Kappa* [19]. Tabel 2 menampilkan nilai validitas dari *Cohen's Kappa* [20]. Kesepakatan yang mengkonfirmasi label otomatis menggunakan LLM dapat diandalkan untuk analisis lebih lanjut dan memastikan objektivitas klasifikasi diatas nilai 0.60 untuk menghindari bias agar data analisis valid digunakan. Selanjutnya, data ditransformasikan kedalam bentuk numeric menggunakan metode ekstraksi fitur TF-IDF (*Term Frequency-Inverse Document Frequency*) [21]. Metode ini akan memberi bobot pada setiap kata berdasarkan intensitas kemunculannya dalam dataset sehingga kata-kata kunci yang merepresentasikan emosi dapat terdeteksi lebih akurat.

#### 2.4 Implementasi dan Evaluasi SVM

Penerapan algoritma SVM dengan kernel linear dilakukan di tahap pemodelan. Kernel linear dianggap mampu dalam menangani data teks dengan dimensi fitur yang tinggi. Dataset akan dibagi menjadi dua bagian, yakni data latih (*training data*) sebesar 80% untuk membangun model dan data uji (*testing data*) sebesar 20% untuk mengukur performa model disesuaikan dengan standar umum penggunaan dataset yang memadai [22]. Tahap terakhir berupa evaluasi hasil menggunakan *Confusion Matrix* [10]. Performa model kemudian diukur berdasarkan parameter akurasi, *precision*, *recall* dan *F1-Score* [10] untuk mengetahui sejauh mana algoritma SVM mampu mengelompokkan sentimen komentar terkait perubahan emosi menjelang menstruasi secara tepat, Klasifikasi dengan Support Vector Machine Setelah pelabelan kelas sentimen, data kemudian dibagi menjadi dua, yaitu data training dan data testing. Data training digunakan untuk mengklasifikasikan sedangkan data testing digunakan untuk mengevaluasi hasil pengklasifikasian dari data training.

### 3. Hasil dan Diskusi

Pengumpulan data komentar menggunakan teknik *scraping* komentar pada akun Instagram @nhb777 sebagai sumber utama dalam penelitian ini. Penelitian ini menganalisis sentimen terhadap tingginya interaksi dan viralitas konten isu perubahan emosi saat PMS dengan studi kasus pada unggahan video *reels meme* yang dipublikasikan pada tanggal 15 Oktober 2024. Pemilihan konten didasarkan pada jumlah interaksi dan respon yang tinggi dari pengguna yang merepresentasikan beragam ekspresi emosional. Setelah tahap preprocessing selesai, langkah [23]berikutnya adalah pelabelan kelas sentimen Proses ini menggunakan library TextBlob untuk menerjemahkan data komentar ke dalam bahasa Inggris sebelum dilakukan labeling Labeling dibagi menjadi tiga kelas, yaitu sentimen positif, negatif, dan netral. [24] Klasifikasi dengan Support Vector Machine Setelah pelabelan kelas sentimen, data kemudian dibagi menjadi dua, yaitu data training dan data testing. Data training digunakan untuk mengklasifikasikan sedangkan data testing digunakan untuk mengevaluasi hasil pengklasifikasian dari data training.



Gambar 2. Konten Reels meme isu PMS pada akun @nhb777

Sebanyak 2762 data komentar didapatkan dari proses *Scraping* menggunakan aplikasi pihak ketiga. Konten isu PMS yang dijadikan objek penelitian ini masih dapat diakses secara *real-time*, sehingga kemungkinan data komentar akan masih tetap bertambah. Namun, keterbatasan aplikasi yang digunakan menyebabkan proses *scraping* hanya memperoleh 2762 dari 2839 data komentar. Oleh karena itu, seluruh data yang berhasil diperoleh ditetapkan sebagai data yang akan dianalisis. Tahap analisis dilakukan menggunakan aplikasi *real-time Google Collab* sebagai *platform* bahasa Python dan sebelum dianalisis lebih lanjut.



Gambar 3. Hasil *Scraping* Komentar Instagram

Tahap *pre-processing* dimulai dari *Case folding*, penghapusan URL/*username/hashtag*, normalisasi huruf ganda, penanganan emoji dengan deskripsi teks, normalisasi *slang*, penghapusan karakter non-alfanumerik, tokenisasi *stemming* dan penerjemah bahasa. *Case folding* dilakukan dengan membuat semua data komentar yang awalnya kombinasi huruf besar dan kecil, kemudian disamakan menjadi huruf kecil. Penerapan tahap *pre-processing* membuat data yang awalnya berjumlah 2762 berkurang menjadi 2755. Pengurangan 7 data disebabkan proses *pre-processing*. Proses awal dari *pre-processing* dilakukan dengan pembersihan komentar dari URL, *username*, *hashtag* dan mengubah huruf yang berulang 3 kali atau lebih menjadi 2 huruf, contohnya komentar “Diam kamuuu” menjadi “Diam kamu”.

text_komentar	komentar_bersih	text_setelah_terjemah
Pas selesai marah gw baru ngeh "kok gw marah *engak jelas' sih tdi 🤔"	selesai marah baru sadar marah jelas risih	selesai marah baru sadar marah jelas rasih
Dan lu bayangin haid 2 kali dalam sebulan rusuhnya kek gimana	lu bayangin haid kali bulan rusuh gimana	lu bayangin haid kali bulan ruuh gimana
Biasa ny H-1 haid itu pasti lgi rajin <sup>2</sup> nya ibadah 🙏🙏	biasa ny h haid rajin ibadahmaaf	biasa ny h haid rajin beribadahmaaf
🙏🙏	berat maaf	berat maaf
Realll 🤔🤔	reallertawa dukung	reallertawa dukung
❤️	suka	suka
Rlllllll 🤔🤔🤔🤔	rlertawatertawatertawa	rlertawatertawatertawa
kebalikkkkkkk 🤔	kebalikertawa	kebalikertawa
Diam kamuuu 🤔🤔🤔🤔	diam kamu sangat marah sangat marah sangat marah sangat marah	diam kamu sangat marah sangat marah sangat marah sangat marah
❤️	suka	suka

Gambar 4. Komentar setelah *Pre-processing*

Pada gambar 4, penanganan emoji juga dilakukan dengan mengganti emoji dengan deskripsi teks. Penggunaan emoji pada sosial media, tidak selalu berarti digunakan sesuai dengan nama emojinya sehingga, perlu melihat lebih dalam tentang kebiasaan dan tren orang-orang yang ada disosial media dalam menggunakan emoji.

Misalnya, emoji “menangis” secara umum dapat diartikan sedih namun dalam bersosial media beberapa pengguna mengartikan sebaliknya, dianggap sebagai hal yang sangat lucu atau humor yang membuat tertawa. Gambar 5 menampilkan jenis emoji yang digunakan untuk mendeskripsikan perasaan pengguna pada komentar berdasarkan emoji yang paling sering muncul

```
text = text.replace('😂', 'sangat lucu').replace('😄', 'tertawa').replace('🙇', 'maaf')
).replace('❤️', 'suka').replace('👍', 'membenarkan').replace('😞', 'sedih').replace('😭', 'sayang')
).replace('👉', 'perilaku wanita').replace('😡', 'sangat marah').replace('👏', 'mendukung')
).replace('😬', 'keberatan').replace('😏', 'risih').replace('😜', 'bercanda').replace('👉', 'sedikit').replace('👉', 'umum')
).replace('👍', 'bagus').replace('😲', 'terharu').replace('😄', 'senang').replace('😞', 'lelah').replace('😄', 'tertawa keras')
).replace('👉', 'aku').replace('😞', 'malu kecewa').replace('👉', 'jengkel').replace('😞', 'menerima').replace('😞', 'berpikir')
).replace('🙇', 'sama').replace('😄', 'sangat suka').replace('😞', 'malu').replace('😄', 'suka').replace('👉', 'menyayangi')
).replace('😐', 'netral').replace('👉', 'sangat buruk').replace('😞', 'sarkas')
```

Gambar 5. Penanganan emoji pada *pre-processing*

Tahap *Noise Removal* dilakukan dengan menghapus karakter non-alfanumerik, spasi dan karakter yang dianggap tidak memiliki nilai informatif pada sentimen. Penggunaan *Google Translator* diterapkan untuk menghapus seluruh karakter non-alfanumerik agar model hanya memproses fitur teks murni. Selanjutnya, penanganan terhadap variasi tertawa digital seperti *haha*, *hihi*, *hehe*, *wkwk* dan *awowkok*. Tahap ini dilakukan dengan tujuan mereduksi fitur dan menjaga konsistensi bobot term pada TF-IDF, sehingga algoritma SVM dapat mengenali ekspresi emosi positif dengan lebih efektif.

Proses Normalisasi *slang* dan singkatan dilakukan dengan mengubah semua *slang* singkat yang ditemukan untuk diganti dengan bahasa yang lebih baku menggunakan *slang* Kamus manual. Sebanyak 230 *slang* manual ditambahkan berdasarkan *slang* singkat umum yang ada pada komentar. Selanjutnya, tahap Tokenisasi menghasilkan komentar yang telah diubah menjadi unit-unit kata dan selanjutnya kata-kata yang masih memiliki imbuhan diubah menjadi kata dasar melalui proses *stemming* menggunakan kamus Sastrawi dan sebanyak 28 kata list stopword Indonesia diaplikasikan untuk menghilangkan imbuhan dan kata sambung yang tidak diperlukan.

Pada gambar 4 terdapat proses akhir dalam *pre-processing* yakni penerjemahan semua komentar kedalam bahasa Indonesia. Keunikan komentar Instagram dimana bahasa yang digunakan pengguna adalah gabungan bahasa indonesia dan inggris menjadi penting untuk dilakukan penerjemahan untuk menyeragamkan bahasa dalam data komentar sebelum dilakukan pelabelan. Penerjemahan komentar dilakukan menggunakan deep translate dari *Google Translator* dari fitur bawaan bahasa pemrograman Python.

text_komentar	komentar_bersih	text_setelah_terjemah
ya gitula	iya gitula	ya begitulah
Di luar kendali 😞	luar kendali	di luar kendali
@nufushytii_Bene banget.pas semangat2 nya eh besokan nya dapet	benar bangetpas semangat eh besok haid	benar bangetpas semangat eh besok haid
lh malah kebalikanya	malah kebalikanya	malah kebalikanya
@kianaddc ril krk gitu kah	nyata krk	krk nyata
🙇	maaf	maaf
😄😄	sangat lucutertawa	sangat lucutertawa
FRRRRR 😞😞	firtawatertawa	Firtawatertawa
@_arya032		
hari kedua sensi lagi 😞	hari dua marah lagitertawa	hari dua marah lagitertawa

Gambar 6. Hasil *Pre-processing*

Pada gambar 6 seluruh tahapan pembersihan dan normalisasi data telah dilakukan, namun ditemukan sejumlah komentar kosong yang hanya berisi satu karakter yang tidak memiliki makna semantik. Komentar kosong tersebut dihapus untuk mengurangi *noise* dan meningkatkan kualitas data agar dapat digunakan dalam proses labeling. Setelah proses ini, jumlah data akhir yang digunakan dalam penelitian ini menjadi 1201 komentar, berkurang sebanyak 1554 data komentar dari data awal.

Proses labeling dilakukan dengan membandingkan hasil labeling manual dan LLM. Pelabelan ribuan komentar secara manual memakan waktu yang banyak dan menjadi keterbatasan sumber daya dalam penelitian ini menjadi alasan penggunaan Asisten AI dalam hal ini LLM dan validasi dari sisi manusia (peneliti) menjadi strategi peneliti. Pelabelan awal dilakukan dengan menggunakan LLM jenis ChatGPT dengan *prompt* terstruktur yang berisi kategori, aturan anotasi dan format *output* yang konsisten agar kualitas anotasi LLM terjamin. Peneliti sebagai *annotator* pada pelabelan manual dilakukan dengan berdasar pada analisis kriteria dari literatur kesehatan [17] untuk menjamin objektivitas interpretasi emosi dalam komentar. Kategori pelabelan disesuaikan dengan hasil pemetaan pada jurnal dimana dibagi menjadi tiga kategori *Positif*, *Negatif* dan *Netral*.

Akurasi: 89.57%  
Skor Cohen's Kappa: 0.8513

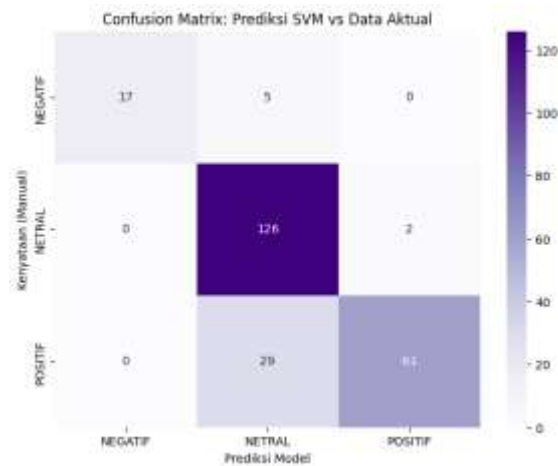
Gambar 7. Hasil Uji Reabilitas *Cohen's Kappa*

Uji Kesepakatan diukur menggunakan *Cohen's Kappa* dengan mengambil sample 10%. Pelabelan otomatis menghasilkan kategori "Netral" dengan 696 komentar, 441 komentar untuk kategori "Positif" dan 93 komentar pada kategori "Negatif". Hasil tersebut, menunjukkan bahwa terdapat imbalanced atau tidak seimbang nya jumlah pengkategorian, sehingga untuk mengatasinya peneliti mengambil tindakan 93 populasi untuk tiap kategori. Tiap kategori diambil populasinya sebanyak 93 komentar yang kemudian diambil 10% data acak sebagai uji kesepakatan. Hasil uji menghasilkan score *Cohen's Kappa* 0,8513 yang menunjukkan tingkat realibilitas "Kuat" dengan akurasi 89,57%. Hasil ini mengkonfirmasi bahwa hasil pelabelan otomatis dapat diandalkan untuk analisis lebih lanjut.

```
--- EVALUASI MODEL SVM ---  
Akurasi SVM: 85.00%  
  
Laporan Klasifikasi:  
precision recall f1-score support  
NEGATIF 1.00 0.77 0.87 22  
NETRAL 0.79 0.98 0.88 128  
POSITIF 0.97 0.68 0.80 90  
accuracy 0.85 240  
macro avg 0.92 0.81 0.85 240  
weighted avg 0.87 0.85 0.85 240
```

Gambar 8. Hasil Evaluasi Model SVM

Hasil pengujian Algoritma SVM ditunjukkan pada gambar 8 menghasilkan akurasi sebesar 85%. Kelas Negatif menunjukkan performa nilai *precision* sangat tinggi (1,00) yang mengindikasikan model klasifikasi sentimen Negatif jarang melakukan kesalahan. Kelas Netral dengan nilai *recall* tinggi (0,98) yang menunjukkan komentar didominasi narasi dan diskusi informasi dari pengguna sementara kelas *Positif* dengan nilai *recall* rendah (0,68) mengindikasikan sentimen positif dalam komentar cenderung bersifat tersirat dan ambigu. Hasil analisis sentimen ini menunjukkan bahwa mayoritas komentar berada pada kategori netral sebagai indikasi bahwa interaksi pengguna didominasi pernyataan yang naratif dan informative. Bentuk dukungan dan empati pada sentimen positif dan sentimen negatif dengan lebih sedikit namun jelas dapat dideteksi oleh model. Hasil akurasi model dengan nilai 85%, menunjukkan bahwa hasil klasifikasi sentimen dinilai cukup andal dalam menggambarkan kecenderungan opini pengguna pada topik yang dianalisis.

Gambar 9. *Confusion Matrix* Hasil Evaluasi Model SVM

Gambar 9 menampilkan *Confusion Matrix* dari data *testing* yang menunjukkan bahwa model teridentifikasi berhasil mengklasifikasikan komentar Negatif sebanyak 17 data dengan 5 data yang disalah klasifikasikan sebagai Netral. Meskipun komentar Netral dapat diklasifikasikan sebanyak 126 data, terdapat 2 data yang salah prediksi sebagai sentimen Positif. Komentar Positif mampu diklasifikasikan sebanyak 61 data komentar dengan 29 data yang salah diprediksikan sebagai sentimen Netral. Model SVM menunjukkan performa yang baik dalam membedakan sentimen ekstrem antara Negatif dan Positif. Kesalahan prediksi sebagian besar terjadi pada sentimen Positif dan Netral yang mengindikasikan bahwa ekspresi sentimen dalam data komentar yang dianalisis bersifat tersirat dan kontekstual.

#### 4. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa analisis sentimen terhadap komentar Instagram pada konten *meme* yang membahas perubahan emosi menjelang menstruasi (PMS) mencerminkan persepsi publik diruang digital. Tahap-tahap yang dilakukan mulai dari proses *scraping*, proses *pre-processing* dengan sastrawi pada berbagai aspek komentar, pelabelan LLM dengan kombinasi validasi manusia dan terakhir pada penerapan Algoritma SVM. Penelitian yang dilakukan terbukti dapat mengklasifikasikan sentimen komentar kedalam tiga kategori yakni Negatif, Netral dan Positif secara objektif dan konsisten. Hasil analisis mengindikasikan bahwa mayoritas komentar pengguna berada pada sisi Netral dimana interaksi publik terhadap isu PMS di Instagram didominasi pernyataan naratif, informatif dan validasi fakta yang terjadi dengan respon tanpa ekspresi emosi yang kuat. Jumlah sentimen Negatif yang lebih sedikit menunjukkan keluhan emosional, iritabilitas dan pegalaman fisik yang relevan dengan gejala PMS sesuai dengan literatur kesehatan. Sentimen Positif mengindikasikan dukungan, empati, humor serta bentuk koping pengguna sebagai bentuk dari respon isu perubahan menjelang menstruasi. Pendekatan yang dilakukan pada pelabelan data komentar menggunakan pelabelan semi-otomatis dengan validasi manusia dapat dikatakan efektif dan efisien dalam pembangunan dataset analisis sentimen berskala besar. Temuan ini dibuktikan dengan hasil uji realibilitas anotasi dengan *Cohen's Kappa* yang menunjukkan tingkat kesepakatan yang kuat antara pelabelan otomatis berbasis LLM dan pelabelan manual dengan tetap menjaga objektivitas data. Evaluasi model SVM dengan akurasi sebesar 85% menunjukkan bahwa Algoritma SVM dengan ekstraksi TF-IDF dan *stemming* Sastrawi memiliki performa yang dianggap cukup mampu dalam mengklasifikasikan sentimen komentar media sosial yang bersifat informal dan kontekstual. *Confusion Matrix* menunjukkan bahwa model mampu membedakan dengan jelas sentimen Negatif dan Positif. Kesalahan klasifikasi sebagian besar terjadi pada sentimen Netral terhadap sentimen Positif karena mencerminkan komentar yang mengekspresikan emosi pengguna media sosial yang bersifat tersirat dan ambigu. Penelitian ini membuktikan bahwa analisis sentimen berbasis SVM, efektif dan dapat diterapkan dalam memahami dinamika persepsi publik terhadap isu sensitif seperti PMS di media sosial. Adapun tindak lanjut pengembangan penelitian selanjutnya dari analisis sentimen ini, disarankan memperluas sumber data, menggunakan metode analisis teks yang lebih konseptual serta penguatan kolaborasi antara manusia dan AI (LLM). Hasil penelitian diharapkan dapat menjadi dasar pengembangan strategi edukasi kesehatan, kampanye literasi emosional dan dalam upaya pengurangan stigma perubahan emosi menjelang menstruasi dengan pendekatan komunikasi yang lebih empatik dan berbasis data.

## Referensi

- [1] Formadiksi UM, 'Hari Media Sosial: Mengapa Media Sosial Menjadi Pilar Utama Komunikasi Modern?', Formadiksi UM. Accessed: Jan. 28, 2026. [Online]. Available: <https://formadiksi.um.ac.id/hari-media-sosial-mengapa-media-sosial-menjadi-pilar-utama-komunikasi-modern/>
- [2] S. Suhendra and F. Selly Pratiwi, 'Peran Komunikasi Digital dalam Pembentukan Opini Publik: Studi Kasus Media Sosial', *Iapa Proceedings Conference*, pp. 293–315, 2024, doi: 10.30589/proceedings.2024.1059.
- [3] L. A. Abdillah, *Peranan Media Sosial Modern*, 1st ed. Palembang: Bening Media Publishing, 2022.
- [4] R. L. Alamin, R. P. Hermoyo, M. Pendidikan, S. Indonesia, and F. P. Komunikasi, 'Fenomena Kesantunan Berbahasa Di Media Sosial Instagram', *Journal of Innovative and Creativity*, vol. 6, no. 1, pp. 3840–3850, 2026, doi: <https://doi.org/10.31004/joecy.v6i1.7068>.
- [5] nasional, 'BPPTIK Kementerian Komunikasi dan Informatika RI', Bpptik Komdigi. Accessed: Jan. 28, 2026. [Online]. Available: <https://bpptik.komdigi.go.id/Publikasi/detail/sejarah-java%0Ahttps://bpptik.kominfo.go.id/Publikasi/detail/logo-dan-identitas-visual#>
- [6] Y. W. Hasibuani, 'Pendidikan Haid Menurut Fikih Dan Mitos Tabu', *Jurnal Insan Cendekia*, vol. 4, no. 2, pp. 85–95, 2023, doi: 10.54012/jurnalinsancendekia.v4i2.267.
- [7] V. P. E. Sari, A. Hermawan, L. Sa'diyah, and S. Hadi, 'Jurnal Cakrawala', *Jurnal Cakrawala Indonesia*, vol. 10, no. 1, pp. 266–276, 2025, doi: <https://doi.org/10.55678/jci.v10i1.1958>.
- [8] T. Walasary, 'Survey Paper tentang Analisis Sentimen', *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, pp. 201–206, 2022, doi: 10.24002/konstelasi.v2i1.5378.
- [9] S. M. Z. A. Difyah, E. A. Winanto, and P. A. Jusia, 'Jurnal Processor Analisis Sentimen Persepsi Publik terhadap Tagar # KaburAjaDulu di', *Jurnal Processor*, vol. 20, no. 2, pp. 223–233, 2025, doi: <https://doi.org/10.33998/processor.2025.20.2.2542>.
- [10] Budi Raharjo, *Pembelajaran Mesin (Machine Learning)*. Semarang, 2021.
- [11] T. Widyanto, I. Ristiana, and A. Wibowo, 'Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter', *SINTECH (Science and Information Technology) Journal*, vol. 6, no. 3, pp. 147–161, 2023, doi: 10.31598/sintechjournal.v6i3.1433.
- [12] R. Firmansyah and H. Firmansyah, 'Klasifikasi Data Instagram dengan Support Vector Machine Instagram Data Classification with Support Vector Machine', *JICN : Jurnal Intelek dan Cendekiawan Nusantara*, vol. 2, no. 6, 2026.
- [13] N. A. Maulana and D. Darwis, 'Perbandingan Metode SVM dan Naïve Bayes untuk Analisis Sentimen pada Twitter tentang Obesitas di Kalangan Gen Z', *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, pp. 655–666, 2025, doi: 10.52436/1.jpti.691.
- [14] Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan, 'Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z', *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 16–25, 2024, doi: 10.52158/jacost.v5i1.715.
- [15] Yuliana Khoirun Nisa, 'Analisis Semiotika terhadap Penggunaan Emoji dalam Komunikasi Daring oleh Remaja', *Jurnal Ilmiah Research Student*, vol. 2, no. 2, pp. 852–859, 2025, doi: 10.61722/jirs.v2i2.5876.
- [16] X. Wang, H. Kim, S. Rahman, K. Mitra, and Z. Miao, 'Human-LLM Collaborative Annotation Through Effective Verification of LLM Labels', *Conference on Human Factors in Computing Systems - Proceedings*, vol. 303, no. Mei, pp. 1–21, 2024, doi: 10.1145/3613904.3641960.
- [17] M. O. Ojezele, A. T. Eduviere, E. A. Adedapo, and T. K. Wool, 'Mood Swing during Menstruation: Confounding Factors and Drug Use', *Ethiop. J. Health Sci.*, vol. 32, no. 4, pp. 681–688, 2022, doi: 10.4314/ejhs.v32i4.3.
- [18] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, 'Data and its (dis)contents: A survey of dataset development and use in machine learning research', *Patterns*, vol. 2, no. 11, p. 100336, 2021, doi: 10.1016/j.patter.2021.100336.
- [19] M. Harri, Suharto, T. Kurnia Fitriani, and I. Kurniawan Subagja, *Pendekatan Kualitatif Kuantitatif dan Campuran*, 1st ed. Purwokerto Selatan, Banyumas, Jawa Tengah: CV. Pena Persada, 2022.
- [20] M. L. McHugh, 'Lessons in biostatistics interrater reliability : the kappa statistic', *Biochemica Medica*, vol. 22, no. 3, pp. 276–282, 2012.

- [21] Y. Yunitasari, *Teori dan Implementasi Analisis Sentimen Menggunakan Python*, 1st ed. Madiun Jawa Timur: UNIPMA Press, 2023.
- [22] B. Farnham, S. Tokyo, B. Boston, F. Sebastopol, and T. Beijing, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems SECOND EDITION*, 2nd ed. United States of America: O'Reilly Media, Inc., 2019.
- [23] F. Andriansyah and P. Astuti, 'Analisis Sentimen Komentar Konten Edukatif Di Instagram Dengan Metode Naïve Bayes Dan Support Vector Machine', 2025. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/imtechno>
- [24] S. Rahmawati Hakim, M. Alfa Rizki, N. Fitri, Y. A. Rizkie, and R. Nooraeni, 'Analisis Sentimen Pengguna Instagram Terhadap Kebijakan Kemdikbud Mengenai Bantuan Kuota Internet dengan Metode Support Vector Machine (SVM)', *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 8, no. 2, 2020, pp. 15-24, doi : 10.24252/msa.v8i2.16795