

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah memberikan dampak yang signifikan terhadap perubahan pola aktivitas masyarakat dalam kehidupan sehari-hari. Kehadiran internet, perangkat pintar (*smart devices*), dan media sosial menciptakan ekosistem digital yang semakin dinamis, tidak hanya sebagai sarana memperoleh informasi tetapi juga dalam membentuk opini publik dan interaksi sosial [1]. Menurut laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pada tahun 2025 jumlah pengguna internet di Indonesia telah mencapai sekitar 229,4 juta jiwa dengan tingkat penetrasi sebesar 80,66% dari total penduduk [2]. Temuan tersebut mengindikasikan bahwa media sosial memiliki peran strategis dalam membentuk persepsi masyarakat terhadap layanan publik maupun produk media.

Transformasi digital berdampak signifikan terhadap industri media, termasuk televisi yang kini tidak hanya mengandalkan siaran konvensional, tetapi juga memanfaatkan media digital sebagai sarana memperluas dan memaksimalkan jangkauan *audiens* [3]. Kehadiran media sosial seperti YouTube dan Instagram menjadi kanal penting bagi televisi dalam menyebarkan konten serta membangun interaksi dengan masyarakat secara *real-time* [4]. Dengan demikian, media sosial berfungsi tidak hanya sebagai sarana penyebaran informasi, tetapi juga sebagai ruang diskusi publik di mana *audiens* dapat memberikan komentar, kritik, maupun apresiasi terhadap program yang ditayangkan [5].

Jawapos Media Televisi (JTV) menjadi salah satu stasiun televisi lokal yang beradaptasi dengan perubahan perilaku *audiens* di era media digital. Didirikan di Surabaya pada tanggal 8 November 2001, hadir sebagai bagian dari Jawa Pos Group, salah satu perusahaan media besar di Indonesia. Sebagai televisi lokal pertama di Indonesia yang beroperasi penuh dalam industri penyiaran lokal dengan siaran yang tidak hanya konvensional, tetapi juga memiliki karakter lokal kuat melalui penggunaan bahasa daerah seperti Jawa Timuran, Suroboyoan dan Madura.

Identitas lokalnya diperkuat dengan program-program yang mengangkat kearifan budaya Jawa Timur, baik dari sisi konten, gaya penyajian, maupun keterlibatan masyarakat setempat [6]. Dengan visi yang mengusung konsep “*Lokal, Nakal, Massal*”, JTV berupaya menyajikan program-program yang informatif, edukatif, dan menghibur serta menjadi perekat budaya di tengah keberagaman masyarakat Jawa Timur.

Dalam menghadapi perkembangan era digital, JTV tidak hanya mengandalkan siaran televisi konvensional, melainkan juga memanfaatkan media daring, seperti media sosial dan platform digital lainnya dalam memperluas jangkauan *audiens*, mempererat interaksi dengan pemirsa, serta menjaga relevansi di tengah pesatnya perubahan industri media [7]. Pertumbuhan interaksi penonton melalui media sosial membuat komentar di platform YouTube dan Instagram JTV berfungsi sebagai indikator signifikan dalam mengukur kualitas tayangan serta sejauh mana program berhasil menarik perhatian publik. Bagi pemirsa, komentar berfungsi sebagai sarana untuk menyampaikan pendapat, kritik, maupun apresiasi yang dapat memengaruhi persepsi *audiens* lain terhadap program. Seperti yang teramati dalam studi persepsi publik terhadap program “Pojok Kampung” di kanal YouTube JTV Rek, di mana respon *audiens* baik positif maupun negatif berdampak pada cara pandang mereka terhadap konten [8]. Sementara bagi pihak media, komentar *audiens* berperan sebagai masukan strategis guna mengevaluasi konten dan memperbaiki mutu tayangan, serta memperkuat hubungan dengan pemirsa, sebagaimana terlihat dalam studi penerapan metode *optimasi feature selection* terhadap komentar media sosial terkait transisi siaran TV digital yang menunjukkan bahwa komentar positif dan negatif memberikan wawasan mengenai persepsi penonton terhadap kualitas siaran [9].

Volume komentar yang terus bertambah dari waktu ke waktu menjadikannya indikator penting dalam mengukur keterlibatan dan citra JTV di ruang digital. Namun, karakteristik komentar yang sangat beragam, mulai dari penggunaan bahasa campuran (Indonesia dan Jawa), gaya bahasa informal, hingga kata *slang* atau singkatan membuat analisis manual menjadi tidak efisien dan berpotensi menimbulkan bias dalam penilaian. Hal ini mendukung penelitian yang

dilakukan oleh [10] yang menyatakan bahwa komentar media sosial tidak hanya menggunakan bahasa Indonesia, tetapi juga campuran bahasa lain serta emotikon, sehingga pengecekan secara manual membutuhkan waktu yang lama dan kurang efektif.

Pendapat masyarakat di media sosial, khususnya YouTube maupun Instagram pada dasarnya diungkapkan melalui teks pendek yang dapat dikompilasi dan dianalisis. Proses pengolahan data tekstual ini dikenal sebagai analisis sentimen yang merupakan bagian dari *text mining* dengan tujuan utama mengekstraksi informasi dari data tidak terstruktur dalam jumlah besar dengan menerapkan identifikasi pola seperti *text classification* dan *information extraction* [11]. Selain digunakan untuk mengidentifikasi sentimen positif dan negatif, data tekstual pada media sosial juga dapat dimanfaatkan untuk mengenali emosi yang lebih spesifik, seperti kebahagiaan (*joy*), kemarahan (*anger*), kesedihan (*sadness*), dan ketakutan (*fear*). Informasi emosi memberikan pemahaman yang lebih mendalam mengenai respons audiens terhadap suatu konten karena mampu menggambarkan bentuk reaksi yang muncul secara lebih rinci dibandingkan analisis sentimen umum. Penelitian [12] menyatakan bahwa pemilihan kategori emosi dalam klasifikasi teks perlu disesuaikan dengan karakteristik data, tujuan penelitian, dan kemudahan proses anotasi. Oleh karena itu, penelitian ini menggunakan lima kategori emosi, yaitu *joy*, *anger*, *sadness*, *fear*, dan netral. Kategori tersebut dipilih karena dinilai lebih relevan untuk merepresentasikan variasi respons *audiens* pada komentar media sosial JTV, sedangkan kategori *disgust* dan *surprise* tidak digunakan karena kemunculannya relatif terbatas serta lebih sulit diidentifikasi secara konsisten pada proses anotasi. Untuk mengidentifikasi emosi tersebut secara otomatis pada data komentar dalam jumlah besar, diperlukan pendekatan klasifikasi teks yang mampu memahami konteks dan keterkaitan informasi yang terkandung dalam komentar.

Analisis teks pada media sosial telah berkembang luas seiring meningkatnya volume data digital dan pemanfaatan berbagai metode *machine learning* yang digunakan untuk klasifikasi. Penelitian kini telah bergerak ke arah klasifikasi *multitopic text classification* di mana satu komentar atau dokumen dapat terdiri lebih dari satu topik sekaligus [13] [14]. Pendekatan *multitopic* muncul karena pada

kenyataannya teks yang ditulis oleh pengguna media sosial sering kali bersifat kompleks dan dapat memuat lebih dari satu makna atau konteks dalam satu kalimat. Misalnya, opini yang mencakup aspek layanan, emosi, dan pengalaman pengguna secara bersamaan. Oleh karena itu, diperlukan metode klasifikasi yang mampu menangkap keterkaitan antar topik dan mengenali hubungan semantik antar label [15]. Penelitian [16] menunjukkan bahwa penerapan metode *multitopic* berbasis *machine learning* dan *deep learning* mampu meningkatkan akurasi prediksi dibandingkan klasifikasi tunggal karena model tidak hanya mengenali satu topik dominan, tetapi juga memahami hubungan semantik antar topik.

Untuk mendukung proses klasifikasi *multitopic* yang mampu mengenali keterkaitan antar topik dan hubungan semantis antar label, diperlukan representasi teks yang mampu menangkap makna laten dari kata dan konteksnya. *Latent Semantic Analysis* (LSA) merupakan salah satu metode yang digunakan untuk mengekstraksi hubungan semantik antar kata melalui matriks term-dokumen dan reduksi dimensi menggunakan *Singular Value Decomposition* (SVD) [17]. Dengan representasi yang lebih padat dalam ruang semantik, LSA memungkinkan model untuk memahami keterkaitan antar kata dan topik meskipun tidak secara eksplisit muncul bersama dalam teks [18]. Teknik ini sangat penting dalam analisis ulasan emosi pada media sosial yang sering bersifat ambigu dan kontekstual. Di sisi lain, kombinasinya dengan pendekatan *Stacking Ensemble* yang menggabungkan beberapa model dasar dan *meta-learner* terbukti meningkatkan kinerja klasifikasi dibandingkan model tunggal [19]. Oleh karena itu, diperlukan penerapan pendekatan *ensemble* yang mampu mengombinasikan kekuatan beberapa model secara bersamaan untuk memperoleh hasil prediksi yang lebih optimal.

Meskipun pendekatan *multitopic* telah terbukti mampu meningkatkan kualitas klasifikasi teks, performa hasil prediksi sangat bergantung pada kemampuan model dasar (*base learners*) dalam mengenali pola data yang kompleks dan beragam. *Ensemble learning* merupakan strategi yang efektif untuk mengoptimalkan performa model dengan mengombinasikan beberapa algoritma pembelajaran untuk menghasilkan model yang lebih kuat dan stabil dibandingkan model tunggal [20]. Konsep dasar dari *ensemble learning* adalah memanfaatkan

keunggulan tiap model dasar dan menutupi kelemahan antar model, sehingga hasil akhir yang diperoleh lebih akurat dan generalisasi model menjadi lebih baik [21].

Salah satu bentuk *ensemble learning* yang banyak digunakan adalah *Stacking Ensemble*, yaitu metode yang mengombinasikan beberapa model dasar (*base learners*) seperti *Support Vector Machine* (SVM), *Random Forest*, dan *Extreme Gradient Boosting* (XGBoost) untuk menghasilkan prediksi akhir yang lebih akurat [22]. Pada metode ini, keluaran dari model dasar digunakan sebagai masukan bagi model tingkat lanjut (*meta-learner*) seperti *Logistic Regression*. Model ini bertugas mengenali pola keunggulan dan kelemahan dari masing-masing *base learners* sehingga mampu menghasilkan generalisasi yang lebih baik.

Dalam penelitian oleh [23] diterapkan pendekatan *ensemble learning* untuk analisis sentimen ulasan pengguna aplikasi Google Play Store, khususnya pada aplikasi Shopee dan Zoom. Teknik *ensemble* seperti *Random Forest* dan *Boosting* dibandingkan dengan algoritma klasifikasi tunggal seperti *Naïve Bayes*. Hasil penelitian menunjukkan bahwa *Random Forest* unggul memberikan performa terbaik dengan akurasi sebesar 94,15% pada ulasan aplikasi Zoom dan 80,69% pada ulasan Shopee, serta *Support Vector Machine* (SVM) juga menunjukkan hasil yang kompetitif. Penelitian ini menegaskan bahwa model *ensemble* cenderung lebih efektif dibandingkan algoritma tunggal, meskipun penerapannya masih terbatas pada klasifikasi sentimen ulasan ke dalam kategori positif dan negatif. Selanjutnya penelitian oleh [24] menggunakan metode *Stacking Ensemble* untuk menganalisis ulasan pengguna terhadap salah satu platform layanan pinjaman online AdaKami dari Google Play Store. Penelitian ini menggunakan tiga model klasifikasi utama, yaitu *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM), kemudian hasil prediksi digabungkan dengan *Logistic Regression* sebagai *meta-learner*. Penelitian ini menunjukkan bahwa model *Stacking Ensemble* menghasilkan kinerja paling optimal dengan tingkat akurasi 87,05%, melampaui model individu. Selain itu, penelitian ini juga mengintegrasikan metode LIME (*Local Interpretable Model-agnostic Explanations*) untuk menjelaskan faktor-faktor kata yang memengaruhi klasifikasi positif maupun negatif.

Lebih lanjut, penelitian oleh [25] menganalisis komentar YouTube terkait reaksi veteran militer Amerika Serikat terhadap pengambilalihan Afghanistan oleh Taliban dengan metode *Stacking Ensemble*. Empat algoritma digunakan sebagai *base learner* yaitu *Logistic Regression*, SVM, KNN, dan CART, yang kemudian digabungkan menggunakan *Logistic Regression* sebagai *meta-learner*. Hasil penelitian menunjukkan bahwa model *Stacking Ensemble* menghasilkan tingkat akurasi lebih tinggi sebesar 75,3% dibandingkan model tunggal dengan *Logistic Regression* sebagai algoritma individu terbaik sebesar 74,6%. Studi ini memperkuat bukti bahwa *Stacking Ensemble* mampu meningkatkan kinerja klasifikasi sentimen, meskipun peningkatan akurasi relatif kecil dan klasifikasi masih terbatas pada kategori positif, negatif, dan netral.

Berdasarkan penelitian sebelumnya, metode ensemble seperti *Stacking Ensemble* terbukti efektif dalam meningkatkan performa klasifikasi teks, khususnya pada analisis sentimen. Namun, sebagian besar penelitian masih berfokus pada klasifikasi polaritas umum, seperti positif, negatif, dan netral, serta belum banyak mengintegrasikan hasil klasifikasi emosi dengan analisis topik pada komentar media sosial. Oleh karena itu, penelitian ini mengembangkan pendekatan klasifikasi emosi pada komentar *audiens* media sosial PT. Jawapos Media Televisi (JTV) menggunakan metode *Stacking Ensemble*, serta menerapkan *Latent Semantic Analysis* (LSA) untuk mengekstraksi representasi topik yang muncul dalam komentar. Melalui pendekatan tersebut, penelitian ini tidak hanya mengidentifikasi kecenderungan emosi *audiens*, tetapi juga memberikan pemahaman yang lebih komprehensif mengenai topik-topik utama yang dibahas dalam komentar. Hasil klasifikasi emosi dan analisis topik tersebut kemudian diimplementasikan dalam antarmuka interaktif berbasis web menggunakan Streamlit sebagai *Graphical User Interface* (GUI). Implementasi ini bertujuan untuk membantu pihak JTV dalam memvisualisasikan hasil klasifikasi emosi secara lebih mudah dan dinamis, sehingga dapat mendukung proses pemantauan tanggapan *audiens* serta pengambilan keputusan berbasis data.

1.2. Rumusan Masalah

Dengan mempertimbangkan latar belakang, fokus permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan metode *Stacking Ensemble* dalam melakukan klasifikasi emosi pada komentar media sosial PT. Jawapos Media Televisi?
2. Bagaimana performa model klasifikasi emosi yang dihasilkan berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*?
3. Bagaimana penerapan metode *Latent Semantic Analysis* (LSA) dalam mengekstraksi representasi topik dari komentar media sosial PT. Jawapos Media Televisi?
4. Bagaimana penerapan GUI Streamlit untuk memvisualisasikan hasil klasifikasi emosi pada ulasan media sosial PT. Jawapos Media Televisi?

1.3. Batasan Masalah

Agar penelitian ini memiliki ruang lingkup yang jelas dan dapat dilaksanakan secara efektif, beberapa batasan diterapkan sebagai berikut:

1. Penelitian ini menggunakan data berupa ulasan atau komentar yang diperoleh melalui proses *web scraping* dari platform media sosial resmi JTV dengan rentang waktu 01 Januari 2024 - 31 Agustus 2025.
2. Data yang dianalisis berupa teks berbahasa Indonesia yang memungkinkan adanya penggunaan bahasa campuran, seperti bahasa Indonesia dan bahasa Jawa, serta penggunaan gaya bahasa informal yang umum digunakan pada media sosial.
3. Representasi fitur untuk proses klasifikasi emosi dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), sedangkan metode *Latent Semantic Analysis* (LSA) digunakan untuk mengekstraksi representasi topik laten dari komentar media sosial.
4. Penelitian ini difokuskan pada klasifikasi emosi komentar pengguna dengan kategori emosi yang merupakan hasil penyederhanaan dari lima kelas awal, yaitu *joy*, *anger*, *sadness*, *fear*, dan netral menjadi tiga kelas utama, yaitu *joy*, *negative*, dan netral.

5. Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui kemampuan model dalam melakukan klasifikasi emosi.
6. Bahasa pemrograman yang digunakan adalah Python dengan berbagai pustaka yang mendukung penerapan algoritma *machine learning* dan pemrosesan bahasa alami (NLP).

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menjawab rumusan masalah serta memberikan arahan yang lebih jelas bagi penelitian ini. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Menerapkan metode *Stacking Ensemble* dalam melakukan klasifikasi emosi pada komentar media sosial PT. Jawapos Media Televisi (JTV).
2. Mengevaluasi performa model klasifikasi emosi yang dihasilkan berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.
3. Menerapkan metode *Latent Semantic Analysis* (LSA) dalam mengekstraksi representasi topik dari komentar media sosial PT. Jawapos Media Televisi.
4. Mengimplementasikan *Graphical User Interface* (GUI) berbasis Streamlit untuk memvisualisasikan hasil klasifikasi emosi pada komentar media sosial PT. Jawapos Media Televisi.

1.5. Manfaat Penelitian

Hasil dari penelitian ini diharapkan mampu memberikan manfaat baik secara teoritis maupun praktis:

1.5.1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan bidang *Natural Language Processing* (NLP), khususnya dalam klasifikasi emosi dan analisis topik pada media sosial berbahasa Indonesia. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi dalam penerapan metode *Stacking*

Ensemble dan *Latent Semantic Analysis* (LSA) pada data teks media sosial yang bersifat informal dan multibahasa.

1.5.2. Manfaat Praktis

Bagi mahasiswa dan peneliti, penelitian ini dapat menjadi referensi dalam penerapan *preprocessing*, ekstraksi fitur, analisis topik, dan klasifikasi emosi berbasis *machine learning* pada data media sosial.

Bagi akademisi, penelitian ini diharapkan dapat mendukung pengembangan penelitian lanjutan di bidang *Natural Language Processing* (NLP), khususnya terkait klasifikasi emosi dan metode ensemble pada teks berbahasa Indonesia.

Bagi industri media, hasil penelitian ini dapat dimanfaatkan untuk memahami persepsi *audiens* terhadap konten yang disajikan melalui analisis emosi dan topik komentar media sosial. Selain itu, implementasi antarmuka berbasis Streamlit diharapkan dapat membantu visualisasi hasil analisis secara lebih interaktif dan mudah dipahami.

Halaman ini sengaja dikosongkan