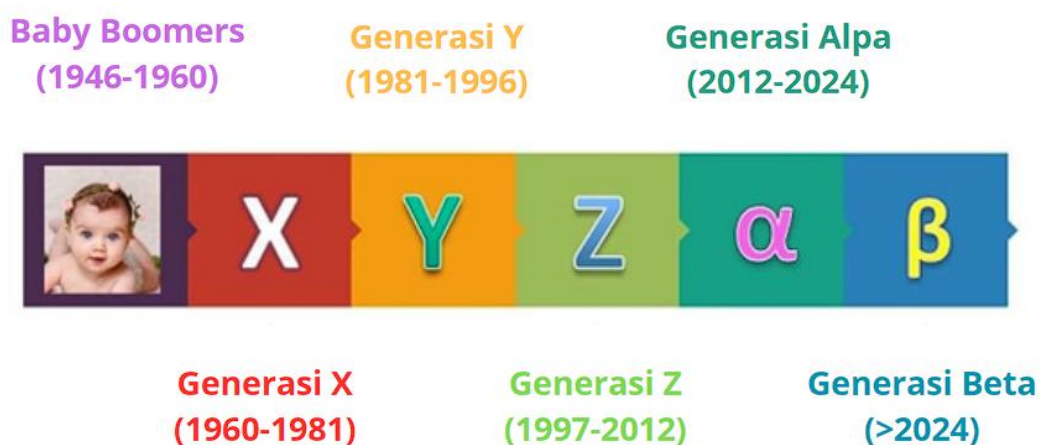


BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi *digital* telah membawa perubahan cukup besar dalam cara manusia berkomunikasi. *Platform* media sosial seperti *TikTok*, *Instagram*, dan *YouTube* menjadi wadah utama bagi individu, khususnya Generasi Z, untuk mengekspresikan diri, berinteraksi, dan membentuk opini publik. Generasi Z, yang lahir antara tahun 1997 hingga 2012, dikenal sebagai *digital natives* yang sangat mahir dalam menggunakan media sosial sebagai sarana utama komunikasi. Mereka sering menggunakan bahasa informal, singkatan, emoji, meme, serta berbagai ekspresi khas dalam interaksi daring, menciptakan pola bahasa dan gaya komunikasi yang unik dan dinamis [1].



Gambar 1. 1 Generasi

Pemahaman terhadap pola komunikasi Generasi Z menjadi penting untuk memahami dinamika sosial dalam ekosistem *digital* serta implikasinya dalam berbagai aspek kehidupan, termasuk pemasaran, pendidikan, dan interaksi sosial. Mengingat karakteristik komunikasi *Gen Z* yang khas dan terus berkembang seiring dengan *tren* media sosial, diperlukan pendekatan analisis berbasis kecerdasan buatan untuk mengidentifikasi pola linguistik mereka secara lebih akurat.

Salah satu metode yang digunakan dalam penelitian ini adalah pemodelan berbasis *Natural Language Processing* (NLP) dengan menggunakan *IndoBERTtweet*, sebuah model *deep learning* yang dilatih khusus untuk memahami bahasa Indonesia dalam konteks media sosial. *IndoBERTtweet* mampu

mengenalistruktur kalimat informal, penggunaan *slang*, emotikon, serta nuansa emosional yang umum dalam teks media sosial, sehingga efektif untuk menganalisis pola bahasa Generasi Z [2].

Dengan adanya perbedaan cukup besar dalam pola bahasa antar generasi, diperlukan pendekatan analisis berbasis kecerdasan buatan untuk mengidentifikasi karakteristik linguistik yang khas dari masing-masing kelompok. Salah satu metode yang dapat digunakan adalah pemodelan berbasis *Natural Language Processing* (NLP) dengan menggunakan *IndoBERTweet*, sebuah model *deep learning* yang telah dilatih secara spesifik untuk memahami bahasa Indonesia dalam konteks media sosial. *IndoBERTweet* mampu mengenali struktur kalimat, makna kata, serta nuansa emosional dalam teks yang digunakan oleh pengguna media sosial, sehingga dapat digunakan untuk mengklasifikasikan pola bahasa dan gaya komunikasi setiap generasi [3].

IndoBERTweet dikembangkan berdasarkan *IndoBERT*, model bahasa berbasis *BERT* (*Bidirectional Encoder Representations from Transformers*) yang dikembangkan khusus untuk bahasa Indonesia. *IndoBERT* sendiri merupakan adaptasi dari model *BERT* yang dilatih menggunakan korpus teks berbahasa Indonesia untuk meningkatkan pemahaman terhadap struktur linguistik dan konteks lokal [4]. *IndoBERTweet* lahir sebagai varian dari *IndoBERT* yang difokuskan pada data dari media sosial, terutama *Twitter*, dengan mempertimbangkan ciri khas bahasa yang lebih santai, penggunaan singkatan, emotikon, serta elemen khas lain yang sering muncul dalam percakapan daring. Dengan pelatihan berbasis data media sosial, *IndoBERTweet* mampu menangkap nuansa bahasa yang lebih dinamis dan kontekstual, menjadikannya alat yang efektif untuk menganalisis komunikasi lintas generasi [5].

Pemanfaatan *IndoBERTweet* dalam penelitian ini memiliki relevansi yang kuat mengingat media sosial memiliki format komunikasi yang unik, seperti penggunaan emotikon, singkatan, serta kata-kata *slang* yang terus berkembang. Model ini mampu menangkap kompleksitas bahasa yang digunakan dalam komentar-komentar di *TikTok*, *Instagram*, dan *YouTube*, yang masing-masing memiliki karakteristik komunikasi berbeda. *TikTok*, misalnya, cenderung lebih interaktif dan mengandalkan *tren visual* serta *audio*, sehingga komentarnya sering kali bersifat

singkat, mengandung ekspresi unik, atau bahkan menggunakan bahasa campuran. *Instagram* lebih berorientasi pada estetika dan *storytelling* [4], sehingga komentar di *Platform* ini dapat lebih deskriptif dan personal. Sementara itu, *YouTube* sebagai *Platform* berbasis video panjang memungkinkan adanya diskusi yang lebih mendalam, sehingga komentar cenderung lebih panjang dan berisi argumentasi yang lebih kompleks.

Penelitian sebelumnya telah menunjukkan pentingnya *preprocessing* dalam pengembangan model *deep learning* untuk pemrosesan bahasa alami. mengeksplorasi pengaruh tahapan *preprocessing*, seperti penghapusan *stopwords* dan *stemming*, terhadap kinerja model *IndoBERT* dan *IndoBERTtweet* dalam mendeteksi emosi pada komentar akun berita *Instagram*. Hasil penelitian mereka menunjukkan bahwa tahapan *preprocessing* justru dapat menurunkan kinerja model dalam hal akurasi deteksi emosi. Model yang tidak melewati tahapan *remove stopwords* dan *stemming* menghasilkan akurasi yang lebih tinggi, dengan 92,54% untuk *IndoBERTtweet* dan 88,81% untuk *IndoBERT*. Temuan ini memperlihatkan bahwa meskipun tahapan *preprocessing* sering dianggap penting untuk meningkatkan kualitas data, dalam beberapa kasus, penghapusan elemen bahasa tertentu seperti *stopwords* dan penerapan *stemming* dapat mengurangi kemampuan model dalam menangkap nuansa emosional yang terdapat dalam teks. Penelitian ini memberikan wawasan baru terkait strategi *preprocessing* dalam analisis teks di media sosial, khususnya untuk aplikasi deteksi emosi [6].

Penelitian terkait lainnya juga menunjukkan pentingnya analisis opini publik dalam transisi menuju kendaraan listrik. Mengidentifikasi opini publik terhadap kendaraan listrik di Indonesia dengan menggunakan data komentar di *YouTube*, menerapkan model pemodelan topik *BERTopic* yang memanfaatkan *IndoBERTtweet* untuk *embedding*. Hasil penelitian ini memberikan wawasan penting tentang bagaimana masyarakat merespons kebijakan transisi kendaraan listrik yang digagas oleh pemerintah Indonesia. Mereka menemukan bahwa topik-topik yang dihasilkan dari data komentar memiliki koherensi tinggi, dengan nilai berkisar antara 0,6 hingga 1, serta nilai *diversity* yang sangat tinggi (0,95838), menunjukkan bahwa opini publik mengenai kendaraan listrik sangat beragam dan mencakup berbagai aspek baik dari segi ekonomi, teknologi, maupun dampak

lingkungan. Penelitian ini memperlihatkan bahwa meskipun ada dorongan untuk beralih ke kendaraan listrik, masih terdapat keberagaman pendapat yang mencakup tantangan-tantangan seperti ketergantungan pada pembangkit listrik berbahan bakar batu bara dan kurangnya infrastruktur yang memadai untuk mendukung transisi tersebut [7].

Penelitian terkait lainnya juga menunjukkan pentingnya pemanfaatan model *deep learning* dalam mendeteksi ujaran kebencian di media sosial. Penelitian ini melakukan penelitian mengenai deteksi ujaran kebencian dalam bahasa Indonesia menggunakan *IndoBERTweet* yang digabungkan dengan lapisan *Recurrent Neural Network* (RNN) *BiLSTM*. Dalam penelitian ini, mereka mengembangkan model yang dapat mendeteksi ujaran kebencian pada *Platform Twitter* dengan tingkat akurasi yang tinggi, mencapai 93,7%. Model *IndoBERTweet*, yang telah dilatih khusus untuk bahasa Indonesia, memberikan peluang besar untuk meningkatkan performa klasifikasi teks, terutama dalam konteks deteksi ujaran kebencian. Dengan tambahan lapisan *BiLSTM*, model ini dapat memproses informasi secara lebih mendalam dan mengidentifikasi pola-pola kompleks dalam teks yang mengandung ujaran kebencian. *Dataset* yang digunakan dalam penelitian ini adalah *dataset* publik yang telah diberi label sebagai ujaran kebencian dan *non-ujaran* kebencian, yang kemudian diterapkan pada model untuk meningkatkan akurasi deteksi [8].

Penelitian ini Bertujuan untuk menganalisis bagaimana pola bahasa dan gaya komunikasi berbeda antar generasi dalam komentar di tiga *Platform* media sosial tersebut. Dengan menggunakan *IndoBERTweet*, penelitian ini akan mengkategorikan perbedaan bahasa berdasarkan aspek seperti panjang teks, penggunaan kata *slang*, tingkat formalitas, serta ekspresi emosional. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam bidang linguistik *digital*, pemasaran media sosial, serta analisis *tren* komunikasi generasi di era *digital*.

Lebih jauh, pemahaman mengenai pola bahasa dan gaya komunikasi tiap generasi dapat menjadi acuan bagi berbagai sektor, termasuk perusahaan, pemerintah, dan lembaga pendidikan dalam menyusun strategi komunikasi yang efektif. Bagi perusahaan, hasil penelitian ini dapat membantu dalam merancang

kampanye pemasaran yang lebih sesuai dengan target *audiens*. Pemerintah dapat menggunakan temuan ini untuk menyusun kebijakan komunikasi publik yang lebih inklusif dan efektif. Sementara itu, lembaga pendidikan dapat memahami cara terbaik dalam berinteraksi dengan generasi muda, baik dalam konteks pembelajaran daring maupun penyebaran informasi.

Oleh karena itu, saya memilih judul ini dengan harapan bahwa penelitian ini dapat memberikan pemahaman yang lebih komprehensif mengenai pola bahasa dan gaya komunikasi Generasi Z di media sosial, serta bagaimana faktor-faktor seperti perkembangan teknologi, *tren digital*, dan budaya internet mempengaruhi cara mereka berinteraksi. Dengan adanya penelitian ini, diharapkan berbagai pihak, seperti akademisi, praktisi pemasaran, pengembang teknologi, dan pembuat kebijakan, dapat memahami karakteristik komunikasi Generasi Z dan menyesuaikan pendekatan mereka dalam berbagai bidang. Dalam dunia bisnis, hasil penelitian ini dapat menjadi panduan bagi perusahaan dalam merancang strategi pemasaran *digital* yang lebih efektif dan sesuai dengan preferensi Generasi Z sebagai target *audiens* utama. Sementara itu, dalam bidang pendidikan, wawasan mengenai gaya komunikasi Generasi Z dapat membantu institusi pendidikan dalam mengembangkan metode pembelajaran yang lebih interaktif, kreatif, dan relevan bagi generasi muda. Tidak hanya itu, penelitian ini diharapkan juga dapat berkontribusi dalam pengembangan teknologi kecerdasan buatan, seperti *chatbot* dan asisten virtual, yang mampu beradaptasi dengan pola bahasa khas Generasi Z. Dengan demikian, penelitian ini diharapkan tidak hanya berkontribusi dalam ranah akademis dan sosial, tetapi juga memiliki dampak nyata dalam pengembangan strategi komunikasi, pemasaran, kebijakan publik, serta teknologi berbasis kecerdasan buatan yang lebih inklusif dan responsif terhadap kebutuhan generasi *digital* masa kini.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Permasalahan utama dalam penelitian ini adalah tantangan dalam mengidentifikasi dan menganalisis pola bahasa dan gaya komunikasi Generasi Z berdasarkan komentar di media sosial. Hal ini terutama disebabkan oleh tidak adanya indikator eksplisit yang menunjukkan usia pengguna dalam komentar di

Platform seperti *TikTok*, *Instagram*, dan *YouTube*. Meskipun Generasi Z memiliki ciri khas dalam berkomunikasi, seperti penggunaan bahasa informal, *slang*, emoji, serta ekspresi kreatif lainnya, karakteristik ini sering kali muncul secara halus dan bervariasi antar individu. Oleh karena itu, dibutuhkan pendekatan yang efektif berbasis *Natural Language Processing* (NLP) untuk memahami dan mengklasifikasikan pola komunikasi Generasi Z hanya melalui teks komentar tanpa bergantung pada data demografis eksplisit. Tantangan ini menuntut pengembangan sistem analisis otomatis yang mampu menangkap nuansa bahasa *Gen Z* secara akurat di berbagai *Platform* media sosial.

1.2.2 Metode Usulan

Untuk mengatasi masalah tersebut, penelitian ini mengusulkan penggunaan model *IndoBERTweet* untuk melakukan klasifikasi komentar berdasarkan pola bahasa Generasi Z. Model *IndoBERTweet* akan dilatih menggunakan *dataset* komentar yang telah dikumpulkan dan diproses untuk memahami konteks linguistik serta karakteristik komunikasi khas Generasi Z di media sosial. Dengan kemampuan *IndoBERTweet* dalam menangkap hubungan kata dalam konteks dua arah serta memahami bahasa informal, penggunaan *slang*, dan ekspresi khas media sosial, model ini diharapkan mampu mengidentifikasi pola bahasa yang digunakan oleh Generasi Z secara akurat pada *Platform* seperti *TikTok*, *Instagram*, dan *YouTube*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan yang telah dijelaskan sebelumnya serta metode yang diusulkan, muncul pertanyaan sebagai berikut:

1. Bagaimana mengklasifikasikan pola bahasa Generasi Z berdasarkan komentar di *TikTok*, *YouTube*, dan *Instagram*?
2. Bagaimana performa model *IndoBERTweet* dalam mengklasifikasikan pola bahasa Generasi Z dengan menggunakan variasi hyperparameter yang berbeda?

1.3 Tujuan Penelitian

Berdasarkan pertanyaan penelitian yang telah dijelaskan sebelumnya serta pertanyaan yang diusulkan, muncul tujuan penelitian sebagai berikut:

1. Menganalisis kinerja model *IndoBERTweet* dalam mengklasifikasikan komentar berdasarkan pola bahasa Generasi Z, serta mengukur tingkat akurasi dalam konteks data media sosial berbahasa Indonesia.
2. Mengidentifikasi dan mengklasifikasikan pola bahasa yang digunakan oleh Generasi Z dalam komentar di media sosial, khususnya *TikTok*, *YouTube*, dan *Instagram*, berdasarkan karakteristik linguistik seperti panjang komentar, penggunaan *slang*, tingkat formalitas, dan ekspresi khas media sosial.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam beberapa aspek. Secara akademis, penelitian ini berkontribusi pada pengembangan kajian linguistik *digital* dan pemrosesan bahasa alami (*Natural Language Processing*), khususnya dalam memahami pola bahasa Generasi Z di media sosial berbahasa Indonesia. Temuan dari penelitian ini juga dapat menjadi referensi bagi studi lanjutan yang tertarik pada karakteristik linguistik generasi muda dalam konteks komunikasi *digital*. Secara praktis, hasil penelitian ini memberikan wawasan kepada praktisi di bidang pemasaran *digital*, teknologi, dan komunikasi publik mengenai preferensi bahasa Generasi Z, sehingga dapat dimanfaatkan dalam penyusunan strategi komunikasi yang lebih efektif, seperti perancangan konten media sosial atau pengembangan *chatbot* yang sesuai dengan gaya bahasa mereka. Dari sisi teknologi, penelitian ini turut mendorong penerapan model *IndoBERTweet* dalam berbagai aplikasi berbasis kecerdasan buatan yang mampu memahami serta merespons komentar atau percakapan daring secara lebih relevan dengan karakteristik *Gen Z*.

1.5 Batasan masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya dan metode yang diusulkan, batasan masalah meliputi:

1. Penelitian ini hanya fokus pada analisis komentar di tiga *Platform* media sosial, yaitu *TikTok*, *Instagram*, dan *YouTube*.
2. Analisis komentar terbatas pada Generasi Z

3. Pendekatan yang digunakan untuk klasifikasi komentar hanya mengandalkan model *IndoBERTtweet* dan tidak mempertimbangkan faktor lain seperti data demografi pengguna.
4. *Dataset* komentar yang digunakan dalam penelitian ini diambil pada Juni 2025 dan tidak diperbarui secara otomatis untuk mencakup data komentar terbaru.
5. Fokus penelitian hanya pada klasifikasi pola bahasa, tanpa membahas lebih dalam mengenai topik atau konten spesifik dalam komentar.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Dalam bagian ini akan dibahas mengenai sub bab diantaranya latar belakang, perumusan masalah, tujuan, dan manfaat, batasan masalah, pertanyaan penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Pada bagian ini disajikan dasar teori dan metode yang digunakan sebagai dasar dan alat untuk mengatasi masalah serta uraian beberapa hal yang terkait dalam penelitian.

BAB III METODE USULAN

Pada bagian ini berisi langkah-langkah yang akan dilakukan pada penelitian seperti tahapan penelitian, variabel penelitian, teknik pengambilan data, metode yang akan mengolah data dengan berisi langkah-langkahnya, dan flowchart penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari proses penelitian yang telah dilakukan, baik berupa data mentah, hasil pemrosesan, maupun *visualisasi* seperti grafik atau tabel. Penjelasan setiap hasil disertai pembahasan yang mengaitkan temuan dengan teori atau penelitian sebelumnya. Pembahasan dilakukan secara kritis dan objektif untuk mengungkap apakah hasil mendukung hipotesis atau menunjukkan hal baru.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan, dirumuskan secara ringkas dan sesuai dengan tujuan awal penelitian. Selain itu, disertakan juga saran-saran yang ditujukan kepada peneliti selanjutnya maupun pihak-pihak yang

berkepentingan, sebagai masukan untuk pengembangan riset lebih lanjut atau penerapan hasil penelitian.

DAFTAR PUSTAKA

Bab berisi tentang keterangan mengenai sumber rujukan yang digunakan dalam penyusunan Tugas Akhir. Pada penelitian ini daftar pustaka disusun secara otomatis dengan menggunakan aplikasi mendeley.

DAFTAR LAMPIRAN

Bagian ini mencakup lampiran yang berisi data dan informasi pendukung yang relevan dengan penelitian