

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi dan komunikasi telah menjadikan media sosial sebagai salah satu ruang utama bagi masyarakat dalam melakukan interaksi digital. Melalui media sosial, pengguna dapat menyampaikan pendapat, berbagi pengalaman, memberikan tanggapan, serta mengekspresikan emosi dan gaya komunikasi secara bebas[1]. Dalam komunikasi digital, penyampaian pesan tidak hanya bergantung pada makna literal suatu kalimat, tetapi juga dipengaruhi oleh konteks, pilihan kata, gaya bahasa, serta maksud pragmatik yang terkadang tidak disampaikan secara langsung. Kondisi tersebut menyebabkan teks digital memiliki karakteristik yang kompleks karena pengguna dapat menyampaikan pesan melalui berbagai bentuk komunikasi, seperti humo dan serius, maupun bentuk ekspresi tidak langsung lainnya[2].

Keragaman gaya komunikasi tersebut menjadi tantangan dalam proses analisis teks, terutama dalam membedakan teks yang memiliki karakteristik humor dan serius[3]. Teks humor tidak selalu ditandai dengan penggunaan kata-kata lucu secara eksplisit, sedangkan teks serius dapat menggunakan struktur bahasa yang ringan, santai, atau menyerupai percakapan sehari-hari. Dalam kajian pragmatik, makna suatu ujaran tidak hanya ditentukan oleh susunan kata, tetapi juga dipengaruhi oleh konteks penggunaan, maksud penutur, serta interpretasi penerima pesan[4]. Oleh karena itu, sebuah kalimat yang memiliki struktur bahasa serupa dapat menghasilkan makna berbeda apabila digunakan dalam konteks komunikasi yang berbeda.

Meningkatnya aktivitas masyarakat dalam menggunakan media sosial menyebabkan jumlah data teks digital terus mengalami pertumbuhan. Setiap hari, pengguna menghasilkan berbagai bentuk teks berupa komentar, opini, ulasan, dan tanggapan dalam jumlah besar. Kondisi tersebut menyebabkan proses analisis teks secara manual menjadi kurang efektif karena membutuhkan waktu dan sumber daya

yang besar. Oleh karena itu, diperlukan pendekatan otomatis berbasis metode pemrosesan bahasa alami yang mampu mengolah dan memahami informasi dalam teks digital secara lebih efektif, termasuk dalam melakukan klasifikasi gaya komunikasi[5].

Namun, metode analisis teks berbasis NLP masih menghadapi tantangan dalam memahami teks yang memiliki karakteristik ambigu, kontekstual, dan bersifat pragmatik[6]. Model klasifikasi tidak hanya perlu mengenali keberadaan kata tertentu, tetapi juga harus mampu memahami pola frasa, hubungan antarkata, serta urutan kata dalam suatu kalimat. Permasalahan tersebut menjadi semakin kompleks pada klasifikasi gaya komunikasi humor dan serius karena kedua kategori tersebut tidak selalu dapat dibedakan hanya berdasarkan kata tertentu, melainkan membutuhkan pemahaman terhadap konteks penyampaian pesan[7].

Beberapa penelitian sebelumnya telah menerapkan metode *deep learning* dalam analisis teks media sosial. Pengembangan model hibrida IChOA-CNN-LSTM untuk pengenalan emosi pada data media sosial dengan kategori *joy*, *anger*, *sadness*, *fear*, dan *neutral* serta memperoleh akurasi sebesar 97,8%[8]. Penelitian tersebut menunjukkan bahwa kombinasi CNN dan LSTM mampu meningkatkan kemampuan klasifikasi teks melalui penggabungan proses ekstraksi fitur lokal oleh CNN dan pemahaman hubungan urutan kata oleh LSTM. Hasil penelitian tersebut menunjukkan bahwa arsitektur CNN-LSTM memiliki kemampuan dalam memahami pola bahasa dan konteks teks pada data media sosial[9].

Selain itu, Penerapan model CNN-LSTM untuk klasifikasi *cyberbullying* pada media sosial dengan beberapa kategori kelas, seperti *non-cyberbullying*, *insult*, *threat*, *hate speech*, dan *harassment*, serta memperoleh akurasi sebesar 72,9%[10]. Penelitian tersebut menunjukkan bahwa model CNN-LSTM mampu menangani permasalahan klasifikasi teks dengan karakteristik bahasa yang beragam dan memiliki konteks yang kompleks[11]. Kedua penelitian tersebut menjadi dasar bahwa kombinasi CNN dan LSTM memiliki potensi untuk diterapkan pada berbagai permasalahan klasifikasi teks berbasis media sosial.

Meskipun penelitian berbasis CNN-LSTM telah berkembang dalam berbagai bidang analisis teks, sebagian besar penelitian terdahulu masih berfokus

pada klasifikasi sentimen, pengenalan emosi, maupun deteksi perilaku negatif seperti *cyberbullying*. Penelitian tersebut belum secara khusus membahas klasifikasi gaya komunikasi berdasarkan karakter penyampaian pesan, seperti humor dan serius[12]. Padahal, gaya komunikasi humor dan serius memiliki karakteristik linguistik yang berbeda karena tidak hanya dipengaruhi oleh penggunaan kata, tetapi juga oleh konteks, pola frasa, urutan kata, serta maksud komunikasi. Hal tersebut menunjukkan adanya peluang penelitian untuk menerapkan model *Deep Learning* dalam memahami variasi gaya komunikasi pada teks media sosial[13].

Berdasarkan permasalahan tersebut, penelitian ini melakukan klasifikasi gaya komunikasi humor dan serius pada media sosial X menggunakan model Hibrida CNN-LSTM[14]. CNN digunakan untuk mengekstraksi fitur lokal dari teks, seperti pola kata, kombinasi kata, dan frasa yang memiliki karakteristik tertentu. Sementara itu, LSTM digunakan untuk memahami hubungan urutan kata dan konteks kalimat melalui kemampuannya dalam mempertahankan informasi dari kata sebelumnya. Kombinasi kedua metode tersebut diharapkan mampu menghasilkan representasi teks yang lebih baik sehingga meningkatkan kemampuan model dalam membedakan teks humor dan serius[15].

Penelitian ini juga melakukan perbandingan performa antara model CNN tunggal dan model Hibrida CNN-LSTM untuk mengetahui pengaruh penerapan arsitektur hibrida terhadap hasil klasifikasi. Evaluasi performa dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Melalui perbandingan tersebut[12]. Penelitian ini dapat mengetahui model yang memiliki kemampuan lebih baik dalam melakukan klasifikasi gaya komunikasi humor dan serius pada media sosial X.

Berdasarkan uraian tersebut, penelitian mengenai klasifikasi gaya komunikasi humor dan serius pada media sosial X penting dilakukan karena masih terdapat keterbatasan penelitian yang secara khusus membahas pemahaman gaya penyampaian pesan dalam teks digital. Penelitian ini memiliki perbedaan dengan penelitian sebelumnya yang lebih banyak berfokus pada klasifikasi sentimen, emosi, atau perilaku negatif. Dengan menerapkan model Hibrida CNN-LSTM,

penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode *Natural Language Processing* (NLP), khususnya dalam memahami variasi gaya komunikasi pada teks digital berbahasa Indonesia[16].

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini yaitu.

- a) Bagaimana klasifikasi gaya komunikasi humor dan serius pada media sosial X ?
- b) Bagaimana hasil perbandingan performa antara model Hibrida CNN-LSTM dengan model CNN tunggal dalam meningkatkan akurasi klasifikasi gaya komunikasi di media sosial X?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan, penelitian ini memiliki tujuan sebagai berikut:

- a) Mengklasifikasi gaya komunikasi humor dan serius pada media sosial di X
- b) Menganalisis perbandingan performa antara model Hibrida CNN-LSTM dengan model CNN tunggal dalam meningkatkan akurasi klasifikasi gaya komunikasi di media sosial X

1.4 Batasan Masalah

Untuk mencapai tujuan tersebut, dan untuk menghindari pembahasan yang melebar dan melenceng dari masalah, penulis membatasi masalah yang diteliti sebagai berikut;

- a) Penelitian ini menggunakan dataset teks dari Hugging Face sebagai sumber data untuk analisis gaya komunikasi. Dataset yang digunakan adalah SmSA (*Sentiment Analysis*) dari koleksi benchmark IndoNLU, yang diunduh dari repositori resmi IndoNLP di GitHub, yaitu IndoNLP/indonlu.
- b) Penelitian ini mengklasifikasikan gaya komunikasi (humor dan serius) dari dataset *Hugging Face*, tanpa memasukkan kategori tambahan.

Preprocessing data terbatas pada pembersihan teks, tokenisasi, dan normalisasi kata, sesuai standar NLP, sebelum digunakan dalam model LSTM-CNN.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Akademis

- a) Memberikan kontribusi pada pengembangan ilmu Pemrosesan Bahasa Alami (*Natural Language Processing/NLP*), khususnya dalam klasifikasi teks dan analisis gaya komunikasi.
- b) Menjadi referensi bagi penelitian selanjutnya yang ingin mengembangkan model Hibrida CNN-LSTM atau metode klasifikasi serupa.

2. Manfaat Praktis

- a) Membantu perusahaan, peneliti, atau pengembang sistem dalam menganalisis interaksi pengguna di media sosial secara lebih akurat.
- b) Mendukung pengembangan sistem yang mampu mengidentifikasi dan memahami gaya komunikasi pengguna, sehingga bermanfaat untuk strategi komunikasi, pemasaran digital, atau deteksi opini publik.

3. Manfaat Sosial

- a) Mempermudah pemahaman terhadap tren komunikasi online, sehingga dapat meningkatkan kualitas interaksi digital yang lebih sehat dan efektif