

BAB I

PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup penelitian, serta sistematika penulisan pada skripsi Klasifikasi Pengguna *Fake* dan *Real* pada Media Sosial Twitter Menggunakan Model Multimodal Berbasis LSTM dan MLP dengan Pendekatan *Late Fusion*.

1.1. Latar Belakang

Media sosial merupakan sebuah ruang publik yang memiliki peran penting dalam penyebaran informasi dan interaksi antarpengguna. Media sosial tidak hanya digunakan untuk berbagi pengalaman pribadi, tetapi juga dapat membentuk opini dan persepsi publik terhadap berbagai isu sosial, politik, dan budaya (Fachrurrazi, 2025). Salah satu contohnya adalah media sosial X (sebelumnya Twitter), yang memiliki sekitar 586 juta pengguna per Januari 2025 (DataReportal, 2025). Dengan jumlah pengguna yang sangat besar, informasi menyebar dengan cepat dan menjangkau banyak orang dalam waktu yang singkat (Sharma dkk., 2018).

Dibalik besarnya jumlah pengguna media sosial X, sejumlah penelitian menunjukkan bahwa tidak semua akun yang terlibat dalam interaksi di X merupakan akun pengguna asli. Penelitian oleh Varol dkk. (2017) memperkirakan bahwa sekitar 9% hingga 15% akun aktif di X menunjukkan karakteristik sebagai akun bot atau akun palsu. Akun bot atau akun palsu merujuk pada akun yang tidak merepresentasikan identitas pengguna asli dan umumnya dikendalikan secara otomatis untuk tujuan tertentu. Dalam konteks penelitian media sosial, istilah bot sering digunakan untuk menggambarkan akun otomatis (*automated accounts*) dan termasuk dalam kategori akun tidak otentik (*inauthentic accounts*) (Varol dkk., 2017).

Penelitian Cresci dkk. (2017) menunjukkan akun bot terus berkembang dan mampu meniru perilaku akun pengguna asli melalui konsep *social spambots*, yaitu akun bot yang memiliki pola interaksi dan aktivitas menyerupai pengguna asli, seperti metadata profil yang lengkap dan pola posting yang tampak wajar. Dalam penelitian ini, istilah akun *fake* atau akun palsu merujuk pada akun bot yaitu *social spambots*, sebagaimana sebelumnya telah didefinisikan dalam dataset Cresci-2017, sehingga cakupan penelitian ini tidak mencakup seluruh akun palsu lain yang dijalankan oleh manusia. Perkembangan akun palsu menimbulkan tantangan dalam proses deteksi, karena pendekatan konvensional yang mengandalkan satu jenis fitur belum mampu menangkap pola perilaku akun *social spambots* (Cresci dkk., 2017).

Isu terkait akun palsu di X kembali menjadi sorotan publik setelah akuisisi X oleh Elon Musk pada tahun 2022, di mana salah satu tujuannya adalah mengurangi aktivitas dari akun palsu (*Associated Press*, 2022). Namun, penelitian terbaru oleh Hickey dkk. (2025) menyatakan bahwa jumlah akun palsu di X tidak berkurang, bahkan diperkirakan bertambah

Berbagai penelitian telah dilakukan untuk mendeteksi akun palsu di media sosial X. Dalam konteks pembelajaran mesin dan *deep learning*, modalitas merujuk pada jenis atau sumber informasi yang digunakan oleh model untuk melakukan analisis atau klasifikasi. Penelitian deteksi akun palsu awalnya banyak dilakukan dari satu modalitas atau unimodal, yaitu deteksi menggunakan satu jenis sumber informasi, misalnya pendekatan berbasis metadata akun seperti informasi profil pengguna (jumlah pengikut, jumlah mengikuti, atau umur akun), atau aktivitas posting *tweet* (frekuensi *tweet*, rasio interaksi pengguna).

Salah satu penelitian secara unimodal yang dilakukan oleh Daouadi dkk. (2019) dalam mendeteksi akun palsu pada media sosial X (Twitter) menggunakan modalitas metadata profil akun pengguna dan aktivitas posting *tweet*. Penelitian dilakukan menggunakan salah satu algoritma *machine learning*, yaitu *Deep Forest* dengan hasil akurasi mencapai 97,55%. Selain itu, penelitian lebih lanjut dilakukan oleh Mohanapriya dkk. (2023) menggunakan algoritma *Feed-Forward Neural Network* (FFNN) yang menghasilkan nilai evaluasi lebih tinggi dibandingkan model *neural network* lain. Hasil nilai akurasi, *precision*, dan *recall* yang didapatkan berada di atas 93%. Kedua hasil penelitian tersebut menunjukkan performa yang baik dalam mengklasifikasikan akun palsu dari akun pengguna asli.

Berdasarkan penelitian berbasis unimodal menggunakan metadata akun oleh Daouadi dkk. (2019), pendekatan unimodal yang dilakukan pada penelitian tersebut masih memiliki keterbatasan. Penelitian tersebut mengabaikan konten linguistik dari *tweet* dengan alasan ketergantungan bahasa (*language-dependent*). Akibatnya, informasi semantik dalam teks *tweet* tidak digunakan, sehingga informasi dari aspek linguistik belum sepenuhnya dimanfaatkan.

Untuk menangani keterbatasan *language-dependent*, Wei dan Nguyen (2019) melakukan penelitian dengan pendekatan unimodal teks *tweet* untuk mendeteksi *tweet* dihasilkan oleh akun palsu atau pengguna asli. Penelitian dilakukan menggunakan arsitektur *deep learning* yang terdiri dari *embedding layer*, tiga lapisan *Bidirectional Long Short-Term Memory* (BiLSTM), dan sebuah *fully connected layer* serta representasi kata berbasis *word embedding* dari konten *tweet*. Penelitian tersebut menunjukkan bahwa informasi semantik yang tersimpan dalam teks *tweet* mampu membedakan akun palsu dari pengguna manusia dengan akurasi sebesar 92,9%.

Meskipun demikian, penelitian Ilias dan Roussaki (2021) melakukan penelitian dengan membandingkan dua teknik terpisah untuk fitur metadata akun menggunakan *classifier machine learning* dan *deep learning* (BiLSTM dan GloVe *embedding*) untuk teks *tweet*. Hasil yang didapatkan menyatakan bahwa pendekatan yang hanya mengandalkan konten teks *tweet* tanpa informasi tentang profil akun menghasilkan *F1-score* sebesar 86,43% dan akurasi 82,61%. Sedangkan untuk tingkat pengguna, pendekatan berhasil mencapai akurasi sebesar 99,06% dan *F1-score* sebesar 98,86%. Oleh karena itu, penelitian mulai dilakukan menggunakan pendekatan multimodal dengan menggabungkan beberapa modalitas informasi secara bersamaan seperti yang dilakukan oleh Kudugunta dan Ferrara (2018).

Penelitian oleh Kudugunta dan Ferrara (2018) dilakukan secara multimodal dengan menggabungkan konten teks *tweet* dan metadata akun melalui arsitektur *Contextual LSTM* mampu meningkatkan performa deteksi dengan nilai AUC yang mencapai lebih dari 96%. Selain itu, penelitian Lin dkk. (2024) membandingkan pendekatan multimodal berbasis teks dan gambar dengan metode *early fusion* dan *late fusion* dalam deteksi berita palsu. Hasil penelitian menunjukkan bahwa pendekatan *early fusion* menghasilkan akurasi sekitar 87%, sedangkan pendekatan *late fusion* mampu meningkatkan akurasi menjadi 90%, dengan peningkatan akurasi sebesar 3% dan *F1-score* sebesar 4%. Hasil tersebut menunjukkan peningkatan performa pada pendekatan *late fusion* pada penelitian yang berbasis multimodal.

Penelitian skripsi ini mengusulkan model multimodal *deep learning* yang menggabungkan fitur metadata akun dan konten teks *tweet* untuk klasifikasi pengguna *fake* dan *real* pada media sosial X. Berbeda dari penelitian Kudugunta dan Ferrara (2018) yang menggabungkan kedua modalitas dalam satu arsitektur tunggal, penelitian skripsi ini menggunakan arsitektur dua cabang terpisah supaya masing-masing jenis modalitas dapat diproses sesuai dengan karakteristik datanya lalu dilakukan penggabungan di akhir pemodelan atau disebut pendekatan *late fusion*.

Penelitian yang dilakukan pada skripsi ini menggunakan dataset Cresci-2017 baik pada tingkat pengguna (*user*) maupun pada tingkat *tweet*. Dataset *tweet* yang bersifat sekuensial diproses menggunakan *Long Short-Term Memory* (LSTM) untuk memproses hubungan antarkata dalam suatu urutan *tweet* secara sekuensial. Sedangkan fitur metadata akun yang berupa data numerik diproses menggunakan *Multi-Layer Perceptron* (MLP) untuk memodelkan hubungan nonlinier antar fitur. Representasi dari kedua cabang tersebut kemudian digabungkan melalui pendekatan *late fusion* untuk menghasilkan klasifikasi akhir.

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana membangun model klasifikasi pengguna akun *fake* dan *real* pada media sosial Twitter menggunakan pendekatan multimodal *deep learning* berbasis LSTM dan MLP dengan teknik penggabungan *late fusion*?
2. Bagaimana perbandingan performa antara pendekatan unimodal dan multimodal dalam mendeteksi pengguna akun *fake* dan *real* pada dataset Cresci-2017?

1.3. Tujuan Penelitian

Tujuan penelitian dirumuskan untuk menjawab rumusan masalah yang telah ditetapkan. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan dan mengevaluasi model klasifikasi pengguna akun *fake* dan *real* pada media sosial Twitter menggunakan pendekatan multimodal *deep learning* berbasis LSTM dan MLP.
2. Menganalisis perbandingan performa antara pendekatan unimodal dan multimodal dalam mendeteksi akun palsu berdasarkan dataset Cresci-2017.

1.4. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah mampu memberikan kontribusi dalam pengembangan metode deteksi akun palsu pada media sosial Twitter berbasis pendekatan multimodal dan menjadi referensi bagi penelitian selanjutnya dalam pengembangan model klasifikasi akun menggunakan kombinasi metadata dan teks.

1.5. Ruang Lingkup Penelitian

Penelitian ini memiliki beberapa batasan agar penelitian lebih terarah dan tidak keluar konteks penelitian. Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini berfokus pada pengembangan model klasifikasi pengguna akun *fake* dan *real* pada media sosial Twitter menggunakan pendekatan multimodal *deep learning*.
2. Penelitian hanya dilakukan pada *platform* media sosial Twitter dan tidak mencakup deteksi akun *fake* pada *platform* lain seperti Instagram, TikTok, atau Facebook.
3. Dataset yang digunakan dalam penelitian ini adalah dataset Cresci-2017 yang mencakup dua kelas akun, yaitu akun bot otomatis (*social spambots*) dan akun pengguna asli (*genuine accounts*).
4. Modalitas dalam penelitian ini meliputi metadata akun dan konten teks *tweet*.

5. Bahasa *tweet* yang digunakan dalam penelitian ini adalah bahasa Inggris, tanpa mempertimbangkan gaya bahasa seperti majas atau bentuk ekspresi linguistik lainnya.
6. Akun yang digunakan dalam penelitian ini mencakup akun individual, dan tidak mencakup akun komunitas maupun akun bisnis.
7. Dalam pendekatan *deep learning*, modalitas metadata akun diproses menggunakan *Multi-Layer Perceptron* (MLP), sedangkan modalitas teks *tweet* diproses menggunakan *Long-Short Term Memory* (LSTM).

1.6. Sistematika Penulisan

Berikut adalah sistematika penulisan skripsi yang disesuaikan dalam beberapa bab bahasan untuk memudahkan pembaca dalam memahami skripsi dengan urut dan jelas.

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup penelitian, serta sistematika penulisan pada skripsi Klasifikasi Pengguna Fake dan Real pada Media Sosial Twitter Menggunakan Model Multimodal Berbasis LSTM dan MLP dengan Pendekatan Late Fusion.

BAB II LANDASAN TEORI

Bab ini menyajikan teori-teori yang mendukung penelitian skripsi Klasifikasi Pengguna *Fake* dan *Real* pada Media Sosial Twitter Menggunakan Model Multimodal Berbasis LSTM dan MLP dengan Pendekatan *Late Fusion* yang terdiri dari beberapa sub bab, meliputi *state of the art*, media sosial X, akun *fake* dan *real*, dataset Cresci-2017, *data preprocessing* (*data cleaning, feature creation, encoding, feature selection, imbalanced data handling, scaling*), *text preprocessing* (*text normalization, text cleaning, tokenization dan padding, word embedding* (GloVe)), algoritma klasifikasi *machine learning, deep learning*, LSTM, MLP, pendekatan multimodal, *late fusion*, serta metrik evaluasi model klasifikasi.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan metodologi penelitian dalam Klasifikasi Pengguna *Fake* dan *Real* pada Media Sosial Twitter Menggunakan Model Multimodal Berbasis LSTM dan MLP. Tahapan penelitian meliputi pengumpulan dataset, *data preprocessing* yang terdiri dari *data cleaning, feature creation, split dataset, data preprocessing* pada tingkat pengguna

(*user-level*), serta *data preprocessing* pada tingkat *tweet* (*tweet-level*), dilanjutkan dengan proses pemodelan, serta evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dan analisis yang telah dilakukan sesuai dengan metodologi penelitian Klasifikasi Pengguna *Fake* dan *Real* pada Media Sosial Twitter Menggunakan Model Multimodal Berbasis LSTM dan MLP dengan Pendekatan *Late Fusion*.

BAB V PENUTUP

Bab ini berisi kesimpulan yang didapatkan dari hasil penelitian serta saran yang dapat dilakukan untuk penelitian selanjutnya.