BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah mendorong lonjakan penggunaan media sosial di Indonesia, dan lonjakan ini secara bersamaan turut memperbesar skala peredaran berita bohong atau hoaks di ruang digital. Pada Januari 2022, jumlah pengguna media sosial di Indonesia tercatat mencapai 191 juta orang, meningkat 12,35% dari 170 juta pengguna pada tahun sebelumnya, dan Twitter/X menjadi salah satu platform yang paling sering menjadi lokasi penyebaran informasi yang belum terverifikasi kebenarannya (Sinapoy et al., 2026). Kementerian Komunikasi dan Informatika Republik Indonesia (Kemkominfo) turut mencatat bahwa terdapat sekitar 800.000 situs penyebar hoaks yang beroperasi di Indonesia, disertai hasil survei yang menunjukkan bahwa 38% hingga 61,1% masyarakat pedesaan dan 45,3% hingga 79,6% masyarakat perkotaan pernah menjadi korban informasi palsu (Simanjuntak et al., 2024). Data-data tersebut menegaskan bahwa penyebaran hoaks bukan lagi persoalan yang terbatas pada satu kelompok masyarakat tertentu, melainkan telah menjangkau seluruh lapisan sosial secara masif, terus-menerus, dan sulit dibendung hanya dengan metode verifikasi manual.

Permasalahan ini semakin krusial apabila melihat besarnya jumlah pengguna media sosial di Indonesia, yang hingga Januari 2024 mencapai 185,3 juta orang dari total 278,7 juta penduduk, dengan 49,9% populasi atau sekitar 139 juta orang memiliki akun media sosial dan 75% pengguna internet setidaknya satu platform (Hamdikatama, 2025). Besarnya kecepatan penyebaran konten di *platform* seperti Twitter/X menyulitkan verifikasi fakta secara manual dan *real-time*, sehingga pemanfaatan teknologi berbasis pemrosesan bahasa alami dan kecerdasan buatan dinilai menjadi alternatif yang lebih cepat dan efisien dibandingkan metode *fact-checking* konvensional (Pujiono et al., 2024). Karakteristik penyebaran hoaks dinamis, berjejaring, dan melibatkan interaksi antarpengguna seperti *retweet* dan

balasan inilah yang menjadi fenomena mendasar yang melatarbelakangi perlunya pendekatan deteksi hoaks yang tidak hanya mampu memahami konteks bahasa semantik, tetapi juga mampu menangkap pola relasi antarpengguna serta perubahan struktur interaksi dari waktu ke waktu di media sosial Twitter/X.

Penyebaran hoaks di Indonesia terus meningkat seiring pesatnya adopsi media sosial. Data Kementerian Komunikasi dan Informatika mencatat lebih dari 10.000 kasus hoaks dilaporkan dalam lima tahun terakhir, diperparah oleh rendahnya literasi informasi masyarakat Indonesia yang menempati peringkat 71 dari 81 negara dalam survei PISA 2022 (Amin et al., 2025). Kondisi ini semakin krusial pada isu-isu sensitif seperti kesehatan publik; tercatat lebih dari 2.300 laporan hoaks terkait Covid-19 beredar di masyarakat, dengan misinformasi yang diperkirakan berkontribusi pada 800 kematian di seluruh dunia (Suadaa et al., 2021). Sebagian besar informasi tersebut menyebar melalui platform media sosial yang bersifat masif dan cepat. Twitter/X sebagai platform dengan karakteristik penyebaran informasi yang sangat cepat, bersifat *real-time*, dan melibatkan interaksi antarpengguna secara masif menjadikannya salah satu kanal penyebaran hoaks, sehingga dibutuhkan sistem deteksi yang mampu mengimbangi laju dan skala penyebaran tersebut. Kompleksitas ini diperparah oleh kecepatan viralitas konten di Twitter/X yang seringkali melampaui kapasitas verifikasi manual, sehingga hoaks dapat menyebar luas dan memengaruhi opini publik sebelum sempat diklarifikasi oleh pihak berwenang.

Namun, pendekatan deteksi hoaks yang ada saat ini umumnya masih bersifat statis dan hanya menganalisis konten teks secara independen tanpa mempertimbangkan dinamika penyebaran maupun perubahan pola bahasa dari waktu ke waktu. Riadi et al., (2025) menunjukkan bahwa model *IndoBERT* yang dilatih pada data periode 2018–2023 mengalami penurunan akurasi dari kisaran 99,67–99,89% pada data sezaman menjadi hanya 95,45–95,87% saat diuji pada data 2025, akibat pergeseran kosakata (*vocabulary drift*) dan konsep (*concept drift*) yang tidak dapat diantisipasi oleh model statis. Di sisi lain, Trisna et al., (2025) mengungkap bahwa model berbasis teks murni cenderung menghasilkan *recall* yang lebih rendah dibanding presisi akibat ketidakseimbangan kelas data,

sehingga berpotensi melewatkan sejumlah besar hoaks aktual di lapangan. Kedua temuan ini mengindikasikan model berbasis representasi teks saja belum memadai untuk mendeteksi hoaks yang terus berevolusi dan menyebar melalui struktur jaringan sosial yang kompleks, sehingga diperlukan pendekatan yang mampu menangkap informasi semantik sekaligus struktur relasi antarpengguna dan dimensi temporal penyebaran informasi secara dinamis.

Penyebaran hoaks di media sosial, khususnya Twitter/X, terus meningkat dan mengancam kepercayaan publik serta stabilitas sosial-politik di Indonesia, diperparah oleh karakteristik bahasa Indonesia yang beragam seperti slang, campur bahasa, dan variasi ekspresi, serta sifat platform yang *real-time* dan melibatkan interaksi secara masif sehingga menyulitkan proses verifikasi manual secara cepat, tepat, dan akurat. Kondisi ini semakin nyata dari keterbatasan pendekatan berbasis teks semata, di mana model *BERT* hanya mencapai akurasi 76% dan bahkan turun menjadi 63% pada varian *BERT Multilingual* akibat kompleksitas linguistik berita berbahasa Indonesia (Hanum et al., 2024). Persoalan ini semakin krusial pada *dataset* Indonesia *False News* yang timpang, yakni 3.465 berita hoaks berbanding hanya 766 berita *valid* dari total 4.231 artikel, sehingga model berbasis teks murni seperti *BERT-only* bias ekstrem terhadap kelas mayoritas dengan akurasi rata-rata pada validasi silang 5-*fold* hanya 71% dan *F1-score* 0,59, sementara kombinasi *hybrid BERT–GNN* yang memanfaatkan struktur relasi antarentitas post, pengguna, dan sumber berita terbukti mampu meningkatkan akurasi hingga 90,48% dengan *F1-score* 0,89 dan ROC-AUC 0,92 (Khairunnisa et al., 2026). Namun demikian, graf yang dibangun pada model tersebut masih bersifat statis dan hanya merepresentasikan relasi antarentitas pada satu titik waktu tertentu saja, belum mengakomodasi dimensi temporal penyebaran informasi, padahal pola *retweet* dan propagasi temporal terbukti menjadi indikator kuat untuk deteksi dini hoaks di Twitter/X. Kondisi inilah yang mendorong perlunya pengembangan model *hybrid BERT–GNN* dengan graf temporal dinamis yang mampu mengintegrasikan pemahaman semantik kontekstual, struktur relasional antarentitas, dan dinamika penyebaran informasi secara lebih akurat dan adaptif.

Beberapa penelitian terdahulu menunjukkan bahwa upaya deteksi hoaks berbahasa Indonesia masih menyisakan sejumlah celah signifikan yang belum terpecahkan secara komprehensif. Sinapoy et al., (2026) membuktikan *IndoBERT* mencapai akurasi 92,07% mengungguli LSTM (87,54%), namun hanya diuji pada *dataset* kecil bertema tunggal (Covid-19) dan menunjukkan ketidakstabilan performa antar *fold*. Amin et al., (2025) mencatat akurasi tinggi 97,74% menggunakan CNN 1D, namun distribusi *dataset* belum mempertimbangkan kategori berita sehingga generalisasi lintas topik terbatas. Trisna et al., (2025) menemukan BiLSTM mencapai *F1-score* 0,90, tetapi *dataset* yang sangat *imbalanced* (13.913 berbanding 1.360) dan fokus tunggal pada isu Presiden Jokowi membatasi generalisasinya. Hamdikatama, (2025) menunjukkan *Stacking Ensemble* unggul dengan akurasi 82% dan ROC-AUC 0,77, namun waktu komputasinya mencapai 63,72 detik sehingga kurang layak untuk konteks *real-time* seperti Twitter/X. Temuan Khairunnisa et al., (2026) membuktikan integrasi *BERT–GNN* mencapai akurasi 90,48% dan F1-score 0,89, mengungguli *BERT-only* (70,78%) maupun *GNN-only* (85,35%), namun graf yang dibangun masih bersifat statis, hanya merepresentasikan relasi antarentitas pada satu titik waktu tanpa mengakomodasi dinamika temporal penyebaran informasi. Berdasarkan sintesis tersebut, tampak jelas belum ada penelitian yang secara khusus mengintegrasikan pemahaman semantik kontekstual berbasis *BERT* dengan struktur graf temporal dinamis yang mampu menangkap pola *retweet* dan evolusi interaksi antarpengguna dari waktu ke waktu, sehingga penelitian ini berupaya mengisi kesenjangan tersebut melalui pengembangan model *hybrid BERT–GNN* dengan graf temporal dinamis untuk deteksi hoaks di Twitter/X.

Berdasarkan kesenjangan tersebut, penelitian ini mengembangkan model *hybrid BERT–GNN* yang diperkaya graf temporal dinamis untuk mendeteksi hoaks berbahasa Indonesia pada *platform* Twitter/X secara adaptif. Berbeda dari pendekatan *hybrid* sebelumnya yang membangun graf statis pada satu titik waktu, penelitian ini merepresentasikan relasi antara *node post* dan *user* sebagai graf kumulatif yang berevolusi melalui enam *snapshot* per lima hari, sehingga mampu menangkap dinamika penyebaran informasi secara lebih responsif terhadap

perubahan pola interaksi di lapangan. Model *Hybrid-Dinamis* memadukan *embedding IndoBERT* pada fitur *node post* dengan *encoder GraphSAGE* dua lapis untuk menangkap struktur jaringan sosial secara bersamaan. Tujuan penelitian ini adalah merancang, mengimplementasikan, dan mengevaluasi model *Hybrid-Dinamis* berbasis *graf* temporal, membandingkan performanya dengan model *BERT-only* dan *GNN-only* menggunakan skema *time-aware split*, serta menganalisis kontribusi dimensi temporal terhadap peningkatan akurasi, presisi, *recall*, *F1-score*, dan ROC-AUC dalam mendeteksi hoaks di Twitter/X.

1.2 Rumusan Masalah

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

1. Bagaimana rancangan dan implementasi model *hybrid BERT-GNN* dengan graf temporal dinamis untuk mendeteksi hoaks berbahasa Indonesia di Twitter/X, dibandingkan pendekatan berbasis teks murni (*BERT-Only*) dan berbasis struktur graf saja (*GNN-Only*)?
2. Bagaimana performa model *hybrid BERT-GNN* dengan graf temporal dinamis dibandingkan model *BERT-only* dan *GNN-only* berdasarkan metrik akurasi, presisi, recall, *F1-score*, dan ROC-AUC, termasuk tren performanya pada evaluasi per *snapshot* (*Snapshot* 5 dan 6)?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang dan mengimplementasikan model *hybrid BERT-GNN* dengan *graf* temporal dinamis untuk deteksi hoaks berbahasa Indonesia di Twitter/X, mulai dari *crawling* data, *preprocessing*, *auto-labeling*, hingga konstruksi graf berbasis *snapshot incremental*.
2. Membandingkan performa model *hybrid BERT-GNN* terhadap model *BERT-only* dan *GNN-only* melalui pengujian metrik akurasi, presisi, *recall*, *F1-score*, ROC-AUC, dan uji signifikansi *bootstrap paired t-test*, guna mengukur kontribusi dimensi temporal terhadap efektivitas deteksi hoaks secara dini.

1.4 Manfaat Penelitian

Manfaat Penelitian ini diharapkan dapat memberikan kontribusi secara praktis dan akademis. Adapun manfaat dari penelitian ini secara praktis yaitu:

1. Memberikan alternatif sistem lebih adaptif bagi *platform* media dan pengelola konten (media berita, moderator, tim komunikasi publik) untuk mengidentifikasi informasi menyesatkan di Twitter/X secara lebih dini.
2. Menjadi rujukan bagi lembaga terkait penanganan hoaks (Kemkominfo, lembaga *fact-checking*) dalam evolusi interaksi antarpengguna, bukan sekadar analisis konten statis.
3. Menjadi dasar pengembangan alat bantu verifikasi informasi yang dapat digunakan masyarakat umum untuk menilai kredibilitas suatu tweet sebelum menyebarkannya lebih lanjut.

Adapun manfaat dalam hal akademis dari penelitian ini diantaranya adalah sebagai berikut:

1. Memperkaya ilmu komputer, khususnya bidang NLP dan *Graph Neural Network*, mengenai integrasi representasi semantik kontekstual (*BERT*) dengan struktur *graf* temporal dinamis (*GraphSAGE*) untuk deteksi hoaks.
2. Menjadi rujukan pembanding penelitian selanjutnya yang mengembangkan model *hybrid* berbasis *graf* dinamis pada domain bahasa Indonesia, khususnya yang menyasar keterbatasan *graf* statis pada penelitian terdahulu..
3. Memberikan kontribusi metodologis berupa skema *time-aware split* dan evaluasi per *snapshot incremental* yang dapat diadopsi pada riset deteksi misinformasi lain di media sosial.

1.5 Ruang Lingkup dan Batasan Masalah

Agar penelitian lebih terarah dan tidak menyimpang dari tujuan yang telah ditetapkan, maka ruang lingkup dan batasan masalah penelitian ini meliputi:

1. Penelitian berfokus deteksi hoaks berbasis teks tweet berbahasa Indonesia di platform Twitter/X, tanpa mencakup konten multimedia (gambar, video).
2. Data dikumpulkan melalui *crawling* harian selama satu bulan penuh (Agustus 2026), dengan total 10.393 data tweet yang dibatasi pada satu tema, yaitu

politik. Pengumpulan data menggunakan 12 kata kunci pencarian, yaitu "demo 27 Agustus", "RUU Perampasan Aset", "Merebut Kemerdekaan", "harga BBM naik", "harga sembako", "Makan Bergizi Gratis", "MBG", "Koperasi Desa Merah Putih", "korupsi KKN", "hukuman mati koruptor", "pergantian Kapolri", dan "gaji kepala daerah", sehingga penelitian ini tidak mencakup tema lain di luar politik.

3. Pelabelan data (FAKTA/HOAX) dilakukan secara otomatis (*auto-labeling*) berdasarkan kecocokan entitas dan *similarity embedding IndoBERT* terhadap klaim dari empat portal referensi (Antara Aceh, Detik, Kompas, CNN), divalidasi sebagian secara manual dengan *Cohen's Kappa*, sehingga label bersifat biner tanpa kategori misinformasi seperti *satire* atau *clickbait*.
4. Sistem membangun dan membandingkan tiga model, yaitu *BERT-Only*, *GNN-Only*, dan *Hybrid-Dinamis* berbasis *GraphSAGE* dengan graf temporal snapshot kumulatif per 5 hari, tanpa membandingkan dengan arsitektur graf lain di luar *GraphSAGE* (misalnya GAT).
5. *Edge graf* dibatasi, relasi *user↔post* dan *post↔post*, dengan skema *time-aware split* tetap (*Snapshot* 1–3 *training*, *Snapshot* 4 validasi, *Snapshot* 5–6 *testing*), tanpa skema *random split* atau *k-fold cross-validation* konvensional.
6. Evaluasi dibatasi pada metrik akurasi, presisi/*recall*/*F1-score*, ROC-AUC, tren performa per *snapshot* (5 dan 6), serta uji *bootstrap paired t-test*, tanpa mencakup uji *adversarial*, *interpretability* (*SHAP/LIME*), efisiensi komputasi, atau generalisasi ke platform lain (Facebook, TikTok).
7. Selain *dashboard* untuk admin, sistem juga dilengkapi fitur cek fakta untuk pengguna umum, di mana pengguna dapat menempelkan teks lalu memperoleh hasil berupa label FAKTA atau HOAX beserta skor kemiripan dan sumber artikelnya. Fitur ini menggunakan mekanisme *auto-labeling* yang sama dengan proses pelabelan data (pencocokan terhadap artikel Antara Aceh, Detik, Kompas, dan CNN Indonesia), bukan prediksi dari model *BERT-Only*, *GNN-Only*, maupun *Hybrid-Dinamis*, dan hanya menerima masukan berupa teks.