

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyebaran informasi di era digital berlangsung sangat cepat, didorong oleh pesatnya perkembangan teknologi informasi. Media sosial, seperti Twitter (X), telah menjadi episentrum dari percepatan arus informasi ini. Hal ini ditunjukkan dengan penetrasi platform ini yang sangat tinggi di Indonesia, di mana berdasarkan data yang dirilis oleh iniliah.com pada Januari 2024, pengguna X (Twitter) di Indonesia mencapai 25,2 juta atau setara dengan 8,8% dari total populasi. Fenomena kecepatan penyebaran ini, yang tercermin dari seberapa cepat sebuah unggahan di-*retweet*, telah menjadi objek kajian penting untuk memetakan lanskap informasi digital.

Karena volume data yang dihasilkan setiap hari sangat besar, mengidentifikasi topik politik secara manual merupakan tugas yang tidak efisien dan memakan waktu. Lebih lanjut, data Twitter tidak hanya mencakup informasi teks tetapi juga hubungan antar pengguna melalui *mention*, dan *reply*, dan pola penyebaran informasi berubah tergantung pada waktu *posting*. Oleh karena itu, mengklasifikasikan topik politik hanya berdasarkan informasi teks sangat sulit. Akibatnya, diperlukan pendekatan yang dapat secara bersamaan memanfaatkan berbagai karakteristik data Twitter untuk mencapai proses klasifikasi yang lebih akurat.

Penelitian sebelumnya telah menggunakan teknik pembelajaran mesin dan pembelajaran mendalam untuk mengklasifikasikan data di media sosial. Namun, sebagian besar penelitian masih berfokus pada representasi data tunggal, misalnya, hanya menggunakan informasi teks atau memanfaatkan hubungan antar pengguna. Pendekatan ini gagal menangkap sepenuhnya kompleksitas data Twitter karena tidak dapat mengintegrasikan hubungan semantik antar *tweet*, pola interaksi pengguna, dan dinamika penyebaran informasi dari waktu ke waktu. Akibatnya,

hasil klasifikasi yang dihasilkan memiliki keterbatasan dalam merepresentasikan hubungan antar data secara komprehensif.

Sebuah studi komprehensif oleh Wang *et al.* (2024) mengkaji penerapan *Graph Neural Networks* (GNN) untuk klasifikasi teks. Inti dari studi ini adalah observasi bahwa data teks dapat dimodelkan secara alami ke dalam grafik untuk merepresentasikan fitur global. Dua pendekatan konstruksi grafik diidentifikasi: (1) *Corpus-level Graph*, yang secara efektif menangkap informasi struktural global, dan (2) *Document-level Graph*, yang secara eksplisit merepresentasikan hubungan antar kata dalam sebuah dokumen. Efektivitas GNN terletak pada kemampuannya untuk mengatasi keterbatasan model pembelajaran mendalam sekuensial, yang dianggap kurang mampu menangani hubungan kompleks dan hubungan kata kontekstual. Studi ini menyimpulkan bahwa GNN, meskipun berasal dari ranah data *non-Euclidean*, menunjukkan kinerja yang sangat menjanjikan di bidang klasifikasi teks.

Berdasarkan uraian diatas, dapat disimpulkan bahwa menganalisis penyebaran informasi di media sosial seperti Twitter memerlukan pendekatan yang mampu menangkap hubungan antar pengguna, isi pesan, dan dinamika waktu secara bersamaan. Karena struktur data Twitter bersifat *graf*, *Graph Neural Network* (GNN) merupakan metode yang tepat untuk mengkaji hubungan tersebut. Lebih lanjut, pendekatan *Multi-View* memungkinkan integrasi berbagai jenis informasi mulai dari konten dan interaksi sosial hingga aspek temporal sehingga analisis yang dihasilkan lebih akurat dan komprehensif. Oleh karena itu, penelitian ini berjudul "**Klasifikasi Topik di Twitter Menggunakan *Graph Neural Network* (GNN) *Multi-View***" dengan tujuan memodelkan dan memprediksi penyebaran informasi politik di Twitter melalui integrasi tiga jenis *graf*: *Semantic Content Graph*, *User Interaction Graph*, dan *Temporal Propagation Graph*.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka yang menjadi rumusan masalah dalam penelitian ini adalah :

1. Bagaimana membangun representasi data Twitter dalam bentuk tiga jenis grafik: *semantic content graph*, *user interaction graph*, dan *temporal*

propagation graph, sehingga dapat digunakan sebagai input untuk model (GNN)?

2. Bagaimana merancang dan mengimplementasikan model *Graph Neural Network* (GNN) yang dapat mengintegrasikan ketiga jenis grafik tersebut untuk mengklasifikasikan topik politik di Twitter secara akurat?
3. Bagaimana hasil klasifikasi topik politik yang diperoleh dari model GNN *Multi-View* dapat digunakan untuk memetakan pola penyebaran dan keterkaitan antar pengguna di Twitter?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini Adalah sebagai berikut:

1. Membangun representasi data Twitter dalam bentuk tiga jenis graf, yaitu *Semantic Content Graph*, *User Interaction Graph*, dan *Temporal Propagation Graph*, sebagai dasar untuk analisis hubungan antar pengguna dan penyebaran informasi.
2. Merancang dan mengimplementasikan model *Multi-View Graph Neural Network* (GNN) yang mampu mengintegrasikan ketiga jenis *graf* tersebut guna melakukan klasifikasi topik politik di Twitter secara akurat dan komprehensif.
3. Menganalisis hasil klasifikasi topik politik yang diperoleh dari model GNN *Multi-View* untuk memetakan pola penyebaran dan keterkaitan antar pengguna dalam diskursus politik di Twitter.

1.4 Manfaat Penulisan

Manfaat yang didapat dari penelitian ini antara lain :

1. Memberikan wawasan baru mengenai implementasi *Graph Neural Network* (GNN) dalam analisis media sosial.
2. Menjadi referensi bagi peneliti dibidang *machine learning*, *data mining*, dan analisis jejaring media sosial.
3. Meningkatkan pemahaman terhadap kualitas dan efektivitas klasifikasi menggunakan GNN .

1.5 Ruang Lingkup Penelitian

Penelitian ini memiliki batasan masalah tertentu antara lain :

1. Penelitian ini difokuskan pada penerapan metode *Graph Neural Network* (GNN) dengan pendekatan *Multi-View* untuk melakukan klasifikasi topik politik pada platform media sosial Twitter.
2. Data yang digunakan dalam penelitian ini berasal dari cuitan (*tweet*) berbahasa Indonesia yang mengandung topik politik. Topik politik dalam penelitian ini dibatasi pada isu pemerintah, tokoh politik, kebijakan publik, dan partai politik. Data ini dikumpulkan melalui Twitter API dalam rentang waktu tahun 2020-2025 agar hasil analisis bersifat representatif.
3. Penelitian ini, peneliti menggunakan tiga representasi *graph* sebagai *input* untuk model GNN *multi-view*. *Semantic Content Graph* dibentuk melalui proses ekstraksi *embedding* menggunakan IndoBERT, dan hubungan antar tweet dibangun berdasarkan nilai kesamaan kosinus. *User Interaction Graph* dibentuk dari hubungan antar interaksi pengguna dalam bentuk *mention* dan *reply*, dan menghasilkan representasi *node* dengan 21 fitur. *Temporal Propagation Graph* dibangun menggunakan informasi waktu *created_at* dan *conversation_id* dan merepresentasikan hubungan antar *tweet* yang telah menyebar dari waktu ke waktu dengan 29 fitur.
4. Proses pelatihan dan evaluasi model dilakukan menggunakan 1.677 data berlabel yang dibagi menjadi data *training* (63,9%), *validation* (16,0%), dan *testing* (20,0%). Sementara itu, 13.454 data tanpa label tidak digunakan pada proses pelatihan dan evaluasi, tetapi digunakan sebagai data inferensi (prediksi) setelah model selesai dilatih untuk mengklasifikasikan *tweet* ke dalam kategori Politik atau Non-Politik.
5. Evaluasi kinerja model dilakukan berdasarkan metrik klasifikasi, seperti *accuracy*, *precision*, *recall*, dan *F1-score*, tanpa membandingkan performa dengan metode lain di luar model GNN *Multi-View*.
6. Output dari penelitian ini berupa model GNN *Multi-View* dan hasil analisis klasifikasi topik politik di Twitter, yang diharapkan dapat menggambarkan

pola penyebaran serta keterkaitan antar pengguna dalam diskursus politik daring.

