

## 1. Pendahuluan

Salah satu platform microblogging paling populer, Twitter, memungkinkan pengguna untuk mengirim dan menerima tweet [1]. Sebagai layanan yang memungkinkan pengguna untuk mempublikasikan apa pun, Twitter akan menampilkan berbagai topik [1]. Selain itu, popularitas situs media sosial seperti Twitter dan Facebook membuat sejumlah besar data terus menerus dihasilkan sehingga menimbulkan tantangan besar dalam pemrosesan data yang efisien [2]. Pesan aktif Twitter dibatasi hanya 280 karakter [1], sehingga memberikan batasan pada pengguna untuk bertukar informasi dan memaksa pengguna menggunakan berbagai strategi seperti bahasa gaul dan singkatan untuk memadatkan lebih banyak informasi [3]. Karena hal tersebut, pengguna perlu memahami konteks dari informasi yang diterima, seperti mengetahui topik dari apa yang sedang dibahas, sehingga diperlukan klasifikasi topik pada permasalahan ini. Oleh karena itu, Twitter dapat digunakan sebagai penelitian klasifikasi topik. Penelitian klasifikasi topik telah dilakukan sebelumnya, seperti penelitian klasifikasi topik menggunakan fitur ekspansi fastText [1], yang mencoba menggunakan fastText sebagai fitur ekspansi dengan menggunakan Gradient Boosted Decision Tree sebagai model klasifikasi. penelitian dengan fitur ekspansi word2vec [4], dan penelitian pendekatan data ekstensif pada klasifikasi topik [5]. Berdasarkan uraian masalah sebelumnya, penelitian, dan sepanjang pengetahuan peneliti, belum ada penelitian yang menerapkan *hybrid deep learning* untuk topik klasifikasi dikombinasikan dengan fitur ekspansi. Keterbatasan pesan di Twitter menimbulkan kosa kata yang berbeda dan membuat sulit untuk menentukan topik pada tweet [3]. Oleh karena itu, penelitian ini menggunakan Glove sebagai fitur ekspansi untuk mengatasi masalah tersebut. Pemilihan Glove di penelitian ini secara konsisten didasarkan pada Glove lebih unggul dibandingkan dengan Word2vec [6]. Sehingga dalam penelitian ini, akan dilakukan klasifikasi topik menggunakan hybrid deep learning dengan fitur ekspansi Glove. penelitian ini menggunakan model Bi-LSTM dan CNN sebagai klasifikasi topik model. Model Bi-LSTM dipilih berdasarkan sifat pemrosesannya yang cepat pada data multi-label dan fleksibilitasnya dalam menyesuaikan bahasa yang beragam [7]. CNN model dipilih berdasarkan penelitian [8] yang menunjukkan jaringan saraf CNN menjadi lebih akurat daripada konvensional pembelajaran mesin untuk mengklasifikasikan teks dari microblogging. Data latih dan uji yang digunakan dalam penelitian ini adalah data tweet berbahasa Indonesia. Pada data pelatihan yang semakin kompleks, pendekatan hybrid merupakan model potensial untuk mengurangi kesalahan [9].

Kontribusi utama dari penelitian ini meliputi menggabungkan variasi N-gram sebagai fitur ekstraksi TF-IDF dalam model klasifikasi topik, menggunakan fitur ekspansi Glove untuk mengatasi perbedaan kata dengan korpus tweet berbahasa Indonesia dan gabungan antara data tweet dan korpus berita menggunakan Bi-LSTM, CNN, dan *hybrid*. Terakhir, peneliti mengusulkan menggunakan metode hybrid deep learning dalam klasifikasi topik dengan dua variasi urutan penggabungan yaitu hybrid BiLSTM-CNN dan CNN-BiLSTM. Menggabungkan dua atau lebih banyak model, metode hybrid mengungguli model tunggalnya dengan tipe data yang berbeda [9]. Penelitian ini bertujuan untuk mengembangkan model klasifikasi topik dengan performansi terbaik dan melihat efek menggunakan fitur ekspansi dalam model Bi-LSTM, CNN, dan Hybrid dari kedua model klasifikasi topik tersebut.

Susunan penulisan setelah pendahuluan adalah terstruktur sebagai berikut. Bab 2 membahas studi terkait yang mendukung atau berkaitan pada penelitian ini. Bab 3 adalah metodologi penelitian. Bagian ini menjelaskan metodologi yang digunakan dalam penelitian ini. Bab 4 menjelaskan hasil dan analisis; terakhir adalah kesimpulan penelitian.