

Topic Classification Using the Long Short-Term Memory (LSTM) Method with FastText Feature Expansion on Twitter

Bella Adriani Putri
School of Computing
Telkom University
Bandung, Indonesia

bellaadrp@student.telkomuniversity.ac.id

Erwin Budi Setiawan
School of Computing
Telkom University
Bandung, Indonesia

erwinbudisetiawan@telkomuniversity.ac.id

Abstrak—Twitter berfungsi sebagai jaringan sosial di mana orang dapat terhubung melalui pengiriman pesan pendek, yang dikenal sebagai *tweet*. Keterbatasan jumlah karakter sering menyebabkan variasi dalam penggunaan kata, sehingga membuat *tweet* sulit untuk dipahami tanpa klasifikasi yang tepat. Dalam penelitian ini, ekspansi fitur digunakan untuk mengatasi masalah ini. Tujuan utama dari penelitian ini adalah untuk mengevaluasi pengaruh dari ekspansi fitur *FastText* dalam konteks klasifikasi topik pada *tweet* berbahasa Indonesia. Penelitian ini menggunakan *FastText* untuk mengurangi ketidakcocokan kosa kata dalam klasifikasi topik dengan menghitung *similarity* antar kata dalam korpus. Oleh karena itu, hasil yang diharapkan dari penelitian ini adalah *FastText* dapat meningkatkan performansi pada klasifikasi topik. Tahap ekstraksi fitur melibatkan pembobotan kata menggunakan TF-IDF yang harus dilakukan sebelum menguji ekspansi fitur. Penelitian ini juga menggunakan *Long Short-Term Memory* (LSTM) sebagai metode klasifikasi karena LSTM cenderung bekerja lebih baik dalam menggunakan sel memori dibandingkan dengan jaringan saraf rekuren pada umumnya. Hasil terbaik dari penelitian ini mencapai akurasi 95,30% menggunakan metode LSTM dan ekspansi fitur dengan *FastText* pada *top similarity* 5 yang dibangun menggunakan korpus Berita. Penelitian ini membuktikan bahwa metode LSTM menggunakan *FastText* dapat meningkatkan nilai akurasi model dalam klasifikasi topik di Twitter.

Keywords—*tweet*, klasifikasi topik, ekspansi fitur, *FastText*, *Long Short-Term Memory* (LSTM)