

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Pembobotan kata merupakan tahapan penting dan dapat menjadi permasalahan dasar karena mampu mempengaruhi akurasi deteksi teks[1]. Secara evolutif pembobotan kata dapat bermula dari *Binary Term Weighting* yang hanya akan memberikan nilai 1 jika suatu kata ada dalam suatu dokumen dan 0 jika kata tersebut tidak ada. *Binary Term Wighting* tidak menghiraukan jumlah kemunculan kata (*term frequency*) dalam suatu dokumen. Hal itu diperbaiki oleh *Raw Term Frequency* yang menghitung jumlah kemunculan kata atau *Term Frequency* (TF). Permasalahan kembali muncul karena bobot dari jumlah kemunculan kata dapat menjadi tidak terkendali besarannya. *Log Frequency Weighting* menjadi solusi untuk mengendalikan peningkatan bobot. Hal yang sama dapat berlaku bukan semata pada jumlah kemunculan kata dalam suatu dokumen, melainkan pada jumlah dokumen yang memuat kemunculan kata atau disebut *Document Frequency*. Untuk mencari kata yang spesifik yang berarti jarang muncul dalam banyak dokumen, maka digunakanlah *Term Frequency* yang dipadukan dengan *Inverse Document Frequency* (IDF). Kata yang sering muncul di banyak dokumen dianggap kurang informatif, sehingga TF-IDF dapat menjadi solusi. TF-IDF menjadi salah satu metode pembobotan kata yang sering digunakan. Namun, TF-IDF mengabaikan label kelas sehingga efektivitasnya menjadi berkurang[1]. Metode pembobotan kata alternatif seperti *Term Frequency–Inverse Gravity Moment* (TF-IGM) yang memanfaatkan label kelas dapat digunakan untuk memperbaiki hal itu.

Label kelas akan mampu mengoptimalkan proses pembobotan kata. Nilai akurasi suatu deteksi emosi terhadap teks akan meningkat dengan pembobotan kata yang optimal. Penggunaan label kelas juga sangat penting bagi proses seleksi fitur agar terhindar dari hasil deteksi yang menyesatkan[2].

Saat ini masyarakat secara masif dengan tanpa keberatan membagikan pikiran, pandangan, dan pendapatnya di media sosial[3]. Salah satu data yang banyak beredar dalam penyebaran pikiran, pandangan, dan pendapat adalah data teks. Teks yang merupakan bagian dari bahasa memiliki unsur emotif[4]. Deteksi emosi dapat memberi berbagai manfaat seperti pengingat akan pentingnya privasi pribadi dan sebagai pendukung dalam pembuatan Keputusan juga eksplorasi masalah psikologis[5].

Berbagai platform dapat dijadikan wadah dalam menampung data teks. Platform populer yang banyak dijadikan penelitian untuk deteksi emosi berbasis teks adalah aplikasi Twitter (kemudian bernama X)[6] dan Facebook[7]. Saat ini terdapat aplikasi media sosial baru yang dikeluarkan oleh Meta (sebelumnya, Facebook) bernama Threads. Platform Threads mulai dirilis pada tanggal 5 Juli 2023. Sementara, X pada tanggal 15 Juli 2006 dan Facebook pada tanggal 4 Februari 2004.

Deteksi emosi dapat dilakukan dengan menggunakan beragam algoritma. Seperti algoritma Naïve Bayes, algoritma Random Forest, algoritma Support Vector Machine (SVM), dan algoritma K-Nearest Neighbors. Algoritma Random Forest adalah salah satu algoritma yang memiliki nilai akurasi yang tinggi[8] di antara algoritma-algoritma *machine learning* lainnya. Namun dibandingkan dengan itu, metode *deep learning* memiliki nilai akurasi yang lebih akurat[9]. Model-model *deep learning* yang mampu menangkap konteks dalam semantik memungkinkan hubungan antar kata tertangkap dengan baik. Seperti model BERT (*Bidirectional Encoder Representations from Transformers*) yang juga memiliki potensi kemampuan memahami emosi di dalam teks. Model pengembangan dari BERT yang khusus untuk penggunaan bahasa Indonesia adalah IndoBERT[10], [11].

Dalam melakukan klasifikasi emosi, penelitian Robert Plutchik yang mengidentifikasi kompleksitas emosi ke dalam *Plutchik's Wheel of Emotions*[12] digunakan sebagai acuan ilmiah dan praktis untuk proses deteksi emosi. Roda emosi Plutchik membagi emosi pada delapan kategori utama yang berlawanan antara positif dan negatif. Setiap kategori memiliki subkelompok yang

menggambarkan keadaan berdasarkan konteks yang terjadi. Plutchik's Wheel of Emotions dapat memudahkan pemilihan jumlah emosi yang akan diidentifikasi dengan akurat.

Penelitian ini berfokus pada pengaruh *Term Frequency-Inverse Gravity Moment* (TF-IGM) dalam meningkatkan akurasi deteksi emosi teks pada platform Threads yang terbilang baru dengan bantuan model IndoBERT. TF-IGM dikenal karena pendekatan barunya terkait pengukuran *class distinguishing power* terhadap *term*.

1.2. Rumusan Masalah

Pembobotan kata sebagai proses akhir dari *preprocessing* digunakan untuk identifikasi istilah yang penting atau frekuensi kemunculan terbanyak suatu kata pada sekumpulan dokumen. Pembobotan kata menggunakan TF-IGM memberikan hasil yang lebih efektif dan stabil. Hasil dari proses pembobotan kata digunakan salah satunya untuk deteksi emosi teks yang telah banyak dilakukan di platform Twitter (X) dan Facebook. Berbagai metode seperti Naive Bayes, Random Forest, Support Vector Machines, K-Nearest Neighbor, dan LSTM-GloVe adalah yang populer digunakan. Namun, model-model *deep learning* terbukti lebih unggul. Salah satu model *deep learning* yang khusus digunakan untuk pemrosesan data bahasa Indonesia adalah IndoBERT. Berdasarkan latar belakang yang telah dipaparkan, maka dapat dirumuskan masalah berupa bagaimana pembobotan kata TF-IGM mampu mempengaruhi tingkat akurasi deteksi emosi teks pada platform Threads menggunakan model IndoBERT?

1.3. Batasan Masalah

Gambaran lebih jelas dapat dilihat dengan memberikan batasan-batasan masalah pada penelitian ini sebagai berikut:

1. Data diperoleh dari teks postingan pada platform Threads;
2. Pengklasifikasian emosi berdasarkan *Plutchik's Wheel of Emotions*;
3. Penggunaan bahasa terbatas untuk bahasa Indonesia.

1.4. Tujuan

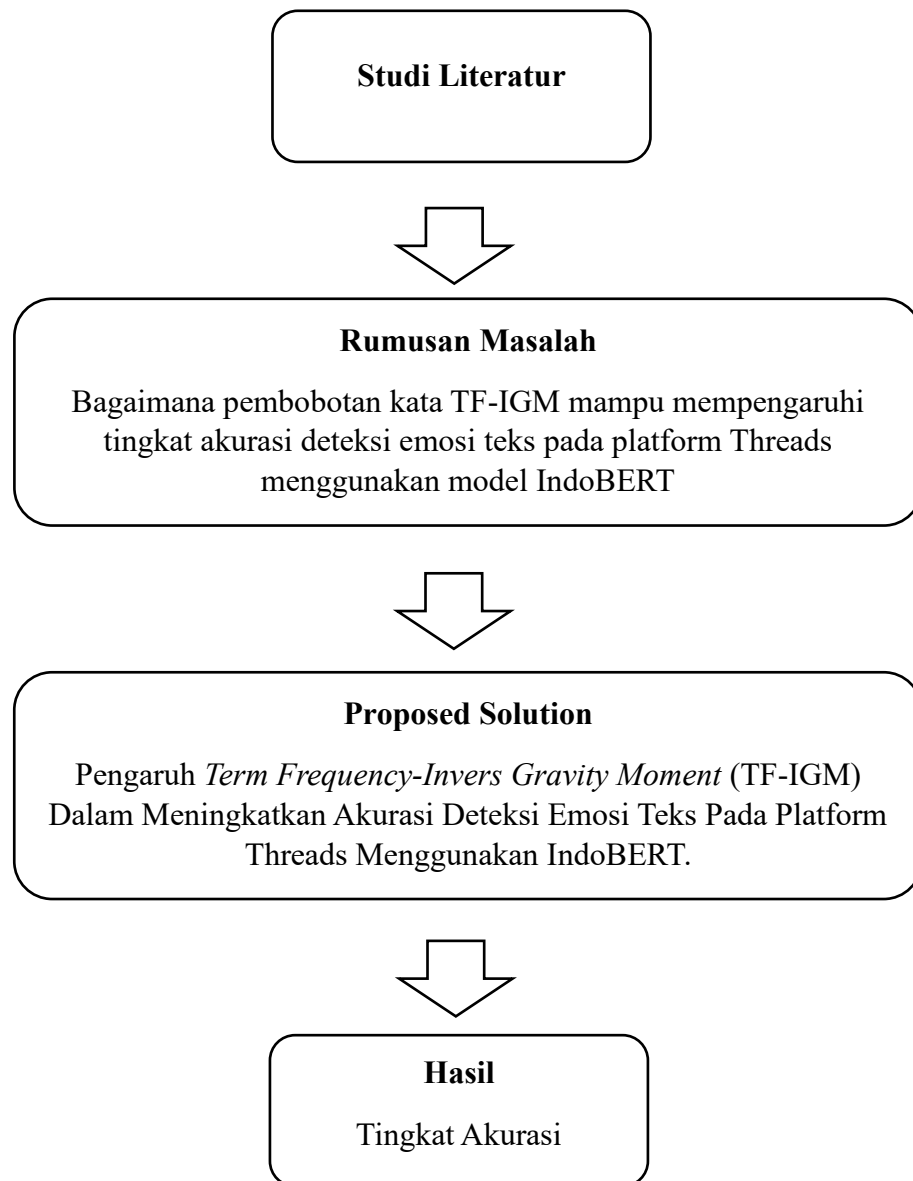
Tujuan dalam melakukan penelitian ini adalah untuk mengetahui pengaruh TF-IGM dalam deteksi emosi pada teks di platform yang belum banyak dijadikan bahan penelitian seperti Threads, dengan menggunakan model IndoBERT.

1.5. Manfaat

1. Menunjukkan hasil akurasi deteksi emosi pada teks pengguna Threads;
2. Mengetahui performa model IndoBERT yang menggunakan pembobotan kata TF-IGM;
3. Penelitian ini diharapkan dapat memberi manfaat dalam perkembangan ilmu Teknik Informatika.

1.6. Kerangka Pemikiran

Penelitian ini memiliki kerangka pemikiran sebagaimana diwakili dalam Gambar 1.1.



Gambar 1.1. Kerangka Pemikiran

Gambar 1.1. memperlihatkan kerangka pemikiran dalam penelitian ini. Kerangka pemikiran tersebut terdiri dari berikut:

1. Studi literatur adalah kegiatan mempelajari hasil dari penelitian-penelitian sebelumnya yang terkait dengan judul penelitian ini.

2. Rumusan masalah mencoba menetapkan masalah yang ada berdasarkan studi literatur untuk kemudian dibahas untuk kemudian menghasilkan satu pemecahan masalah baru.
3. Solusi yang ditawarkan adalah hasil studi dan analisis peneliti untuk kemajuan ilmu pengetahuan.
4. Hasil merupakan keluaran berdasarkan penelitian yang sedang dikaji yakni pengaruh TF-IGM dalam meningkatkan akurasi deteksi emosi teks pada platform Threads menggunakan model IndoBERT.