

BAB I

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[2]. 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[3]. TF-IDF menjadi salah satu metode pembobotan kata yang sering digunakan. Namun, TF-IDF mengabaikan label kelas sehingga efektivitasnya menjadi berkurang. Metode pembobotan kata alternatif seperti *Term Frequency–Inverse Gravity Moment* (TF-IGM) yang memanfaatkan label kelas dapat digunakan untuk memperbaiki hal itu[1].

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[4]. Penelitian yang dilakukan oleh Fira Sukmanisa, Yuita Arum Sari, dan Imam Cholissodin (2021) pada riset

yang berjudul “Deteksi Emosi pada Tweet Berbahasa Indonesia tentang Pembelajaran Jarak Jauh menggunakan *K-Nearest Neighbor* dengan Pembobotan Kata *Term Frequency-Inverse Gravity Moment*” menemukan bahwa pembobotan kata yang tidak menggunakan label kelas seperti TF-IDF memiliki kecenderungan akurasi yang menurun dibandingkan dengan pembobotan kata dengan label kelas seperti TF-IGM yang tidak menggunakan koefisien[1].

Saat ini masyarakat secara masif dengan tanpa keberatan membagikan pikiran, pandangan, dan pendapatnya di media sosial[5]. Salah satu data yang beredar dalam penyebaran pikiran, pandangan, dan pendapat adalah data teks. Teks yang merupakan bagian dari bahasa memiliki unsur emotif[6]. Deteksi emosi dapat memberi berbagai manfaat seperti pengingat akan pentingnya privasi pribadi, sebagai pendukung dalam pembuatan Keputusan, dan juga eksplorasi masalah psikologis[7]. Seringkali seseorang mengalami gejala depresi dan kecemasan karena keadaan jiwa yang apatis dan menekan perasaan[8]. Deteksi emosi membantu pengenalan diri dengan lebih baik sehingga dapat mendekatkan permasalahan terhadap solusi. Dengan mengenal diri sendiri, perasaan inferioritas dan kesedihan akan berkurang sebagaimana firman Allah Swt.: “*Janganlah kamu (merasa) lemah dan jangan (pula) bersedih hati, padahal kamu paling tinggi (derajatnya) jika kamu orang-orang mukmin.*” (Q.S. Ali Imran[3]: 139).

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)[9] dan Facebook[10]. 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. Menjamurnya platform media sosial dan penelitian terhadapnya dapat mendukung salah satu SDGs yaitu No. 9: Industri, Inovasi, dan Infrastruktur[11].

Deteksi emosi dapat dilakukan dengan menggunakan beragam algoritma. Seperti algoritma-algoritma *Machine Learning* berupa *Naïve Bayes*, *Random Forest*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbors*. Algoritma

Random Forest adalah salah satu algoritma yang memiliki nilai akurasi yang tinggi[12]. Namun dibandingkan dengan itu, metode *Deep Learning* memiliki nilai akurasi yang lebih akurat[13]. 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*)[14] yang juga memiliki potensi kemampuan memahami emosi di dalam teks. Model pengembangan dari BERT yang khusus untuk penggunaan bahasa Indonesia adalah IndoBERT[15], [16].

Dalam melakukan deteksi emosi, penelitian Robert Plutchik yang mengidentifikasi kompleksitas emosi ke dalam Roda Emosi Plutchik[17] 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. Roda Emosi Plutchik dapat mewadahi 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 dari teks pada unggahan pengguna Threads yang terbilang merupakan platform baru dengan bantuan model IndoBERT. TF-IGM dikenal karena pendekatan barunya terkait pengukuran *class distinguishing power* terhadap *term*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka dapat dirumuskan masalah berupa:

1. Bagaimana menerapkan TF-IGM dalam deteksi emosi teks pada platform Threads menggunakan model IndoBERT?
2. Bagaimana kinerja TF-IGM setelah diterapkan untuk mendeteksi emosi teks di platform Threads menggunakan model IndoBERT?

1.3 Tujuan

Sebagai upaya memecahkan persoalan dalam latar belakang dan rumusan masalah, tujuan dalam melakukan penelitian ini adalah untuk:

1. Menerapkan TF-IGM dalam mendeteksi emosi pada teks di platform Threads dengan model IndoBERT.
2. Mengetahui kinerja TF-IGM dalam mendeteksi emosi teks di platform Threads dengan model IndoBERT.

1.4 Manfaat

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

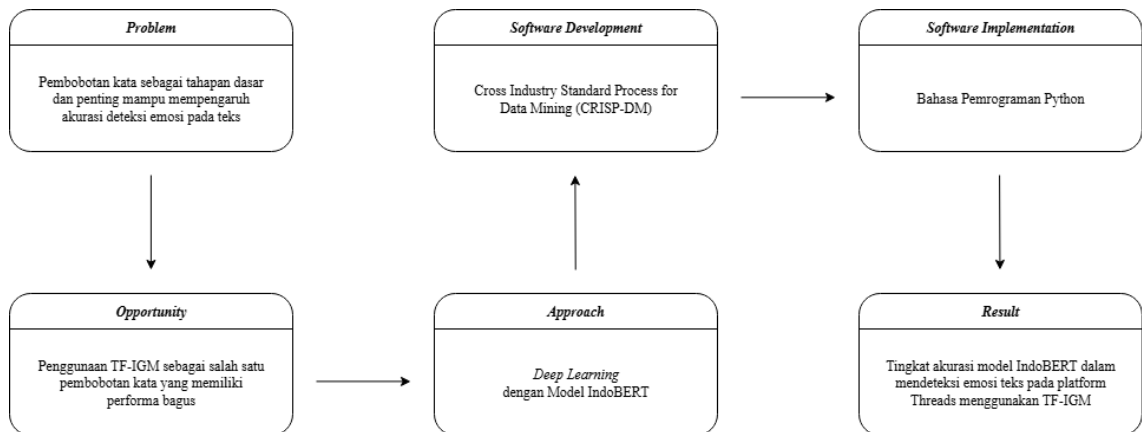
1.5 Batasan Masalah

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

1. Data diperoleh dari unggahan teks pengguna platform Threads dengan visibilitas publik;
2. Pengklasifikasian emosi berdasarkan Roda Emosi Plutchik dengan jumlah delapan emosi berupa: marah, antisipasi, jijik, takut, senang, sedih, terkejut, dan percaya;
3. Penggunaan bahasa terbatas untuk bahasa Indonesia.

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. *Problem* mencoba menetapkan masalah yang ada untuk kemudian dibahas dan dihasilkan satu pemecahan masalah baru.
2. *Opportunity* berupa hipotesis menawarkan solusi yang dapat diterapkan dalam pemecahan masalah.
3. *Approach* adalah metode yang dapat dilakukan untuk menempuh hasil jawaban dari hipotesis yang diajukan.
4. *Software Development* merupakan bagian dari metode dengan standar baku yang dapat digunakan untuk keperluan penelitian akademis.
5. *Software Implementation* menyebutkan bahan dan/atau alat yang digunakan dalam menerapkan metodologi penelitian.
6. *Result* ialah keluaran berdasarkan penelitian yang sedang dikaji yakni pengaruh TF-IGM dalam meningkatkan akurasi deteksi emosi teks pada platform Threads menggunakan model IndoBERT.

1.7 Metodologi Penelitian

Penelitian ini menggunakan tiga pendekatan yang terdiri dari metode pengumpulan data, metode pembangunan program, dan metode pengujian. Uraian mengenai ketiga metode tersebut adalah sebagai berikut:

1.7.1 Metode Pengumpulan Data

Deteksi emosi teks pada platform Threads dimaksudkan terhadap akun yang dibuka secara publik oleh pengguna media sosial tersebut. Unggahan pengguna akan dikumpulkan melalui proses *scraping* menggunakan library *playwright*.

1.7.2 Metode Pembangunan Program

Dalam membangun program, penelitian ini menggunakan metode Proses Standar Lintas Industri untuk Penambangan Data atau *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang umum digunakan dalam riset *data mining*. Metode ini digunakan untuk kebutuhan pengembangan program dan acuan dalam pelaksanaan penelitian.

1.7.3 Metode Pengujian

Tahap evaluasi dilakukan dengan *Confusion Matrix* yang dapat menampilkan visualisasi hasil akurasi deteksi dengan menggunakan TF-IGM dengan hasil akurasi deteksi tanpa pembobotan.

1.8 Sistematika Penulisan

Penulisan ini ditulis dengan sistematika yang terdiri dari lima bagian, yang disusun sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan kerangka pemikiran penelitian, serta sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini berisi hasil penelitian terdahulu berupa *the state of the art*. Dasar-dasar teori yang menjadi landasan penelitian juga dibahas dengan jelas dan tersusun, didasarkan pada teori-teori dan metode terkini.

BAB III METODOLOGI PENELITIAN

Bab ini berisi metode berupa tahap-tahap yang akan dilaksanakan seperti pengumpulan dan pemrosesan data serta simpulan hipotetis.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi pembahasan hasil setelah menempuh metode penelitian. Hasil yang dibahas adalah mengenai Pengaruh *Term Frequency-Inverse Gravity Moment* (TF-IGM) dalam Meningkatkan Akurasi Deteksi Emosi dari Teks pada Unggahan Pengguna Threads Menggunakan IndoBERT.

BAB V SIMPULAN DAN SARAN

Bab ini berisi penutup yang terdiri atas kesimpulan dan saran perbaikan atau rekomendasi untuk penelitian selanjutnya.

