

1. Pendahuluan

Akses terhadap media sosial telah menjadi kebutuhan primer dari setiap orang. Itu dikarenakan adanya kebutuhan informasi, hiburan, pendidikan, dan akses pengetahuan dari seluruh dunia [1]. Perkembangan pada teknologi informasi membuat komunikasi tidak lagi terbatas untuk menghubungkan antar dua orang, namun bisa lebih dari dua orang. Media sosial adalah suatu tempat dimana penggunanya dapat merepresentasikan dirinya kepada dunia, mengungkapkan detail wawasan pribadi kedalam kehidupan mereka [2]. Salah satu contoh media sosial adalah Twitter. Twitter adalah layanan *microblogging* yang mempunyai pengguna lebih dari 41 juta, Twitter memungkinkan penggunanya melakukan *tweet* tentang topik apa pun, dengan mempunyai batasan karakter sejumlah 140 karakter [3]. Karena adanya batasan pada Twitter, maka setiap *tweet* yang ditulis oleh pengguna mempunyai banyak variasi singkatan kata, dan tidak menggunakan tata bahasa yang benar. Banyaknya variasi membuat *tweet* pada Twitter sulit dipahami [4]. Twitter mempunyai fitur yang disebut *trending topic*, *trending topic* sendiri adalah daftar topik yang sering dibicarakan orang, dan daftar *trending topic* sendiri selalu berubah seiring waktu, mengikuti kejadian besar yang sedang terjadi [5]. *Trending topic* menjadi hal yang penting, karena dalam keadaan darurat seperti kecelakaan, bencana alam, dan terorisme, *trending topic* memberikan laporan yang lebih cepat dibandingkan media konvensional [4]. Karena itu *trending topic* menjadi unik untuk diteliti, dengan data yang banyak, dan konten yang singkat. Untuk mengetahui informasi dan berita yang sedang *trending*, pengguna harus melakukan klik terhadap *hashtag*, dan membaca setiap *tweet* yang muncul untuk mengetahui informasi atau berita yang akurat pada *trending topic*. Karena itu perlu melakukan klasifikasi topik yang ada pada *trending topic* kedalam kategori umum dengan akurasi tinggi untuk pengambilan informasi yang lebih baik [6]. Pengklasifikasian teks memiliki banyak fungsi dan banyak metode. Masing-masing metode memiliki kelebihan dan kelemahannya sendiri, baik dari sisi keakuratan, waktu proses, dimensi yang ideal untuk diproses dan juga beberapa *factor* yang lainnya bisa mempengaruhi [7]. Pada tugas akhir ini metode yang digunakan adalah metode klasifikasi *Support Vector Machine* (SVM) dan menggunakan metode pembobotan TF-IDF. Alasan kenapa menggunakan metode klasifikasi SVM adalah karena pada SVM ada beberapa fungsi kernel untuk data yang bersifat *non-linear* [8]. Dan kenapa pada penelitian ini menggunakan metode pembobotan TF-IDF, karena pembobotan TF-IDF terkenal efisien, mudah dipahami dan hasilnya lebih akurat [7]. Dengan tujuan ada pengaruh pembobotan TF-IDF terhadap akurasi dari setiap kernel yang digunakan pada penelitian ini dalam menganalisa *tweet* pada *trending topic* di Twitter.

Batasan masalah dalam pengerjaan topik tugas akhir ini adalah hanya menggunakan 12 kategori. Ciri karakteristik untuk setiap kategori untuk menentukan suatu data *tweet* berdasar dikategori mana, ditentukan oleh penulis, beserta anggota kelompok yang lain. Total data yang digunakan pada penelitian ini berjumlah 77793. Data yang digunakan pada penelitian ini berasal dari *tweet* berbahasa Indonesia dari media *social* Twitter. *Tweet* tersebut diambil dari bulan Juli 2019 sampai bulan Agustus 2019, dengan jumlah *tweet* yang berbeda setiap harinya.