

## 1. Pendahuluan

Perkembangan peradaban manusia berdampak pada berkembangnya teknologi dan informasi. Salah satu yang terlihat dengan jelas adalah perkembangan Internet yang memudahkan manusia melampaui batas-batas yang tak terhingga serta sebagai media penyebaran informasi yang sangat cepat. Internet yang dinilai memberikan kemudahan dalam penyebaran informasi memberikan peluang untuk media massa dan platform-platform jejaring sosial berkembang[1]. Salah satu platform media sosial yang paling banyak digunakan masyarakat adalah media sosial Twitter.

Twitter merupakan salah satu jejaring sosial dan *mikroblog* daring yang memungkinkan penggunaanya dalam mengirimkan serta membaca pesan yang berbasis teks singkat hingga 140 karakter, namun sekarang diperbesar hingga 280 karakter[2]. Twitter telah menjadi salah satu portal informasi terbesar yang menyediakan platform dengan kemudahan, cepat dan terpercaya bagi pengguna untuk berbagi apapun yang terjadi di sekitar mereka dengan teman dan pengikut lainnya. Twitter awalnya dibuat oleh *Jack Dorsey*, *Biz Stone*, dan *Evan Williams* tahun 2006 sebagai sebuah layanan SMS (*Short Message Service*) yang bertujuan untuk berkomunikasi dalam kelompok kecil. Sesuai dengan namanya, tak heran jika Twitter dijuluki “pesan singkat internet”[3].

Keterbatasan karakter pada Twitter jelas membatasi seseorang dalam menyampaikan informasi dalam satu *tweet*[4]. Karakter yang sangat sedikit membuat orang harus memutar otak dalam menyampaikan informasi dengan sangat pendek sehingga akibatnya orang menggunakan singkatan dan kosakata baru yang akhirnya menimbulkan bias makna dari informasi aslinya[4]. Pesan *tweet* yang hanya dibatasi 140 karakter sangatlah pendek, tidak selalu benar, jarang menggunakan tata bahasa yang benar, serta banyak variasi kata yang tujuannya untuk bertukar pesan semaksimal mungkin dengan karakter seminimal mungkin[4]. Namun hal ini akhirnya membuat orang terpaksa menggunakan *emoticon*, singkatan, bahasa gaul internet, serta simbol-simbol tertentu yang akhirnya menyebabkan kesalahan tanggap pada informasi yang hendak disampaikan[4]. Dalam mengatasi permasalahan ini, salah satu cara yang bisa dilakukan adalah melakukan ekspansi kata pada *tweet*. Hal ini sudah dilakukan Twitter dengan memperbanyak karakter hingga dua kali lipat yang awalnya 140 karakter menjadi 280 karakter. Ekspansi ini merupakan proses semantik dalam memperbanyak teks asli menjadi teks yang berukuran lebih besar[5].

Berdasarkan penjabaran tersebut, penelitian ini bertujuan untuk meneliti mengenai ekspansi kata pada *tweet* menggunakan *Word2vec* dengan Metode *Gradient Boosted Decision Tree* untuk melakukan klasifikasinya. Metode *Gradient Boosted Decision Tree* dipilih karena memiliki keunggulan dalam *mem-break down* proses pengambilan keputusan yang kompleks menjadi lebih sederhana, sehingga pengambil keputusan dapat menginterpretasikan solusi dari permasalahan yang ada, serta mempermudah eksplorasi data dan pemodelan, jadi sangat bagus jika dijadikan untuk langkah awal dalam proses pemodelan dan dijadikan sebagai model akhir dari penerapan dengan teknik yang lain. *Gradient Boosted Decision Tree* merupakan salah satu metode klasifikasi yang menggunakan struktur pohon, yang setiap *node* merepresentasikan atribut serta pada cabangnya merepresentasikan nilai dari suatu atribut, sedangkan daunnya digunakan untuk merepresentasikan kelas. *Word2vec* mencoba untuk mengaitkan makna antara dua kata yang kemudian menggambarkan kemiripan relasi kedua kata tersebut. Penelitian dimaksudkan untuk meneliti pula mengenai ekspansi fitur pada *tweet* menggunakan *Word2vec*.

Penelitian ini membahas klasifikasi topik pada akun media sosial menggunakan *Gradient Boosted Decision Tree*. Pada penelitian ini juga dibahas tentang pembobotan kata yang menggunakan teknik pembobotan *Term Frequency – Inverse Document Frequency* (TF IDF) serta Ekspansi Fitur dengan corpus similarity *Word2Vec*. Hasil akhir yang akan dicapai pada penelitian ini adalah untuk mengetahui pengaruh penerapan *baseline* dengan menggunakan metode *Gradient Boosted Decision Tree* dalam klasifikasi topik pada media sosial Twitter, pengaruh penerapan pembobotan kata pada teknik TF-IDF dengan menggunakan metode *Gradient Boosted Decision Tree* dalam klasifikasi topik pada media sosial Twitter, serta pengaruh penerapan ekspansi fitur dengan menggunakan metode *Gradient Boosted Decision Tree* dalam klasifikasi topik pada media sosial Twitter.