

1. Pendahuluan

Latar Belakang

Di era digital ini, perkembangan teknologi informasi dan komunikasi yang pesat telah memberikan banyak manfaat di semua aspek sosial. Salah satu perkembangan teknologi informasi dan komunikasi yaitu dengan adanya media sosial yang digunakan untuk bertukar informasi terkini dan memberikan kesempatan kepada pengguna untuk menghasilkan sebuah konten [1]. Salah satu media sosial yang banyak diminati oleh berbagai kalangan saat ini yaitu Twitter.

Twitter adalah layanan *microblog* populer yang menarik lebih dari 500 juta pengguna dan menghasilkan lebih dari 340 juta *tweet* setiap harinya [2]. *Tweet* yang diposting dapat berupa tulisan, gambar atau video yang setiap *tweet* nya dibatasi hanya 280 karakter saja [3]. Tidak heran jika Twitter banyak dimanfaatkan untuk berbagai keperluan seperti, mendapatkan informasi terbaru tentang suatu produk, isu sosial, politik, selebriti dan banyak lagi [3]. Berbagai manfaat yang diberikan oleh Twitter telah menjadikannya media sosial yang efektif dan efisien dalam menyampaikan informasi dengan cepat [4]. Selain itu, pengguna Twitter dapat mempublikasikan kembali informasi tersebut dengan cara *re-tweet* atau memposting ulang *tweet*, menyukai postingan *tweet*, mengutip *tweet* dan juga dapat menjadi pengikut dari pengguna lain atau diikuti oleh banyak pengguna [3]. Sehingga informasi tersebut dapat tersebar lagi di Twitter. Pemodelan difusi informasi yang berkembang ini sangat penting agar dapat dipahami penyebarannya dan dapat mengendalikannya dengan baik [5].

Pada penelitian sebelumnya, menerapkan pembelajaran mesin menggunakan metode pengklasifikasian *random forest* dengan tiga jenis fitur: berbasis pengguna, berbasis waktu dan berbasis konten [5]. Penelitian serupa mengenai prediksi *retweet* dengan fitur berbasis sosial, berbasis konten, berbasis *tweet* dan berbasis pengguna dengan membandingkan metode klasifikasi *decision tree* J48, SVM dan regresi logistik [6]. Berdasarkan pernyataan diatas penulis tertarik untuk melakukan penelitian tentang bagaimana prediksi pengguna Twitter yang tertarik untuk melakukan *retweet* suatu konten isu yang sedang berkembang. Penelitian penulis berfokus pada penyebaran informasi di media sosial atau lebih tepatnya untuk memprediksi apakah suatu *tweet* akan mendapatkan *retweet* atau tidak. Fitur yang dipakai pada penelitian tugas akhir ini menggunakan dua model yaitu fitur pengguna dan konten. Metode yang digunakan adalah Naïve Bayes yang memiliki kesalahan paling sedikit dibandingkan dengan algoritma klasifikasi lainnya [7]. Metode ini memiliki perhitungan matematik dasar yang sangat kuat dan dalam efisiensi klasifikasinya juga stabil serta dapat mengolah data dalam jumlah yang besar dengan menghasilkan akurasi yang tinggi [7].

Topik dan Batasan

Berdasarkan latar belakang yang telah disampaikan sebelumnya, maka rumusan masalah yang dapat diambil yaitu memprediksi apakah suatu *tweet* akan mendapatkan *retweet* atau tidak dengan fitur berbasis konten dan berbasis pengguna. Serta melihat performansi dan akurasi dari metode Naïve Bayes dalam prediksi *retweet*. Sedangkan batasan masalah dari penelitian ini adalah penelitian ini menggunakan data *tweet* pengguna Twitter yang diambil pada bulan April hingga Mei 2021.

Tujuan

Tujuan dari penelitian tugas akhir ini adalah membangun sistem prediksi *retweet* berbasis konten dan berbasis pengguna menggunakan metode klasifikasi Naïve Bayes.