

## 1. Pendahuluan

Media sosial menjadi salah satu sarana bagi masyarakat untuk mengungkapkan keinginan, menyalurkan minat dan bakat, atau menyebarkan informasi. Media sosial telah menjadi bentuk kepribadian seseorang. Penyebaran informasi atau difusi informasi yang terjadi selama periode ini memiliki dampak yang lebih besar melalui media sosial daripada metode penyebaran tradisional.

Salah satu media sosial yang berperan sangat penting dalam penyebaran informasi adalah Twitter. Di Twitter, pengguna dapat memberikan informasi berupa teks, video, atau gambar. Jumlah pengguna Twitter tumbuh pesat, menurut data dari statista, ada sekitar 16,32 juta pengguna aktif Twitter di Indonesia pada Juli 2021 [1]. Twitter juga sangat populer di kalangan developer karena mudahnya mendapatkan data yang dibutuhkan developer [2]. Proses penyebaran informasi yang terjadi di Twitter berpindah antar pengguna melalui fungsi *retweet*, sehingga semakin banyak *retweet* yang tersedia, semakin meluas informasi yang tersedia, berdasarkan besarnya pengaruh.

Penelitian sebelumnya membahas mengenai prediksi kebiasaan dalam *retweet* dengan menggunakan model Bayes dalam implementasinya berdasarkan dari perilaku *retweet* pengguna [3]. Dan dalam penelitian lainnya yang membahas tentang prediksi difusi informasi di Twitter menggunakan beberapa fitur, yaitu *user-based*, *content-based* dan *time-based* dan diterapkan ke dalam beberapa model klasifikasi, diantaranya *Naive Baiyes*, *Support Vector Machine*, dan *Random Forest* [4]. *Random Forest* dipilih karena memiliki hasil yang terbaik secara konsisten [3].

Dalam penelitian ini model yang dibangun bertujuan untuk melakukan prediksi terhadap sebuah *tweet* akan mendapat *retweet* atau tidak. Metode klasifikasi yang digunakan adalah *random forest* dengan berfokus terhadap *feature user-based* [4] sebagai vektor ciri yang akan digunakan dan akan melalui tahap *preprocessing* dan juga *k-fold cross validation*.

Batasan masalah pada penelitian ini adalah menggunakan data *tweet* terakhir pengguna Twitter yang diambil secara acak pada bulan Juli 2021. Tujuan penelitian ini adalah membangun sistem prediksi *retweet* berbasis pengguna menggunakan metode klasifikasi *random forest*. Bagian selanjutnya dari penelitian ini adalah bagian 2 yang membahas studi terkait penelitian yang telah dilakukan sebelumnya. Pada bagian 3 dibahas teori yang berkaitan dengan teori terkait perancangan sistem yang dibangun. Selanjutnya pada bagian 4 akan dijelaskan evaluasi penelitian terkait hasil pengujian dan analisis hasil pengujian. Pada bagian 5 kemudian akan dibahas kesimpulan dan saran untuk penelitian yang akan datang.