

## 1. Pendahuluan

Informasi menjadi suatu hal yang dibutuhkan di kalangan masyarakat. Penyebaran informasi sangat cepat dilakukan melalui berbagai media. Media sosial adalah salah satu dari media penyampaian informasi yang saat ini kerap digunakan. Informasi sendiri memiliki peranan penting dalam kehidupan karena dapat mempengaruhi tindakan dari masyarakat.

Dalam kehidupan sehari-hari banyak hal yang terjadi mulai dari fenomena yang baik adanya, hingga insiden yang buruk. Di era digital setiap kejadian dapat diketahui informasinya dengan sangat cepat karena perkembangan media sosial di kalangan masyarakat. Peristiwa-peristiwa yang menjadi topik hangat akan semakin banyak dibicarakan melalui media sosial yang kerap digunakan masyarakat seperti Twitter.

Media sosial Twitter memungkinkan pengguna mengunggah cuitan (*tweet*) dan berinteraksi dengan pengguna lain melalui fitur seperti *reply* dan *retweet*. Kemudahan interaksi yang dibangun dalam Twitter menjadikan berbagai macam informasi dapat menyebar tanpa diketahui fakta dan kebenarannya. Hal ini mengakibatkan munculnya penyebaran rumor di Twitter.

Akun Twitter memiliki banyak fitur yang merepresentasikan pengguna tersebut, seperti jumlah *follower*, jumlah *tweet*, akun terverifikasi atau bukan, *tweet*, jumlah *favorite*, dan seterusnya. Untuk identifikasi akun pada Twitter banyak dilakukan untuk deteksi seperti bot atau *spammers*[1] juga *buzzer*[2]. Klasifikasi dilakukan terhadap fitur yang ada dari dataset Twitter dengan fitur profil *user*, *social network*, dan teks *tweet*. Pada deteksi *buzzer* dilakukan dengan melakukan ekspansi fitur properti akun berupa nilai mean, kuartil, dan *range* dari fitur yang ada seperti *follower*, *following*, dan lain-lain.

Support Vector Machine (SVM) merupakan salah satu metode yang digunakan untuk permasalahan klasifikasi. SVM berfokus pada supervised learning untuk klasifikasi biner dengan ide untuk menggambarkan garis yang akan membagi data menjadi dua kelas[3]. Model ini akan mengklasifikasikan data ke dalam target tertentu berdasarkan variabel fitur yang telah dipelajari pada data *train* (latih). Dalam mengidentifikasi akun penyebar rumor digunakan fitur dari data Twitter berdasarkan *tweet* dan akun pengguna.

Berbeda dengan bot atau *spammers*, untuk *rumor spreaders* membutuhkan fitur yang dapat mengidentifikasi akun tersebut termasuk penyebar atau non-penyebar. Pada tugas akhir ini, penulis berfokus pada identifikasi 'kemungkinan' penyebar rumor di Twitter. Penulis mendefinisikan penyebar rumor sebagai pengguna yang kerap kali mengunggah *tweet* rumor, sehingga 'kemungkinan' besar pengguna tersebut dapat menjadi penyebar rumor.

Penambahan fitur yang dimaksudkan tersebut mengakibatkan perlunya data pengguna yang mengunggah *tweet* beberapa kali untuk melihat intensitasnya. Maka penulis menggunakan PHEME dataset yang terdiri dari sembilan insiden yang terjadi antara 2014 hingga 2015[4]. Terdapat lima insiden yang memiliki jumlah dan perbandingan data yang dapat digunakan untuk identifikasi penyebar rumor di Twitter, yaitu i) *Charlie hebdo*, ii) *Ottawa shooting*, iii) *Germanwings crash*, iv) *Sydney siege*, dan v) *Ferguson*.

PHEME dataset diubah menjadi dataset penyebar rumor dengan ekstraksi *tweet*, akun user, dan menghitung nilai intensitas (seberapa sering pengguna mengunggah *tweet* rumor). Pelabelan yang dilakukan pada pendekatan ini adalah *near ground-truth* yang hanya akan mengidentifikasi akun yang 'mungkin' adalah penyebar rumor karena terbatasnya pendekatan yang sesuai.

Pada tugas akhir ini akan dilakukan pemodelan yang dapat mengidentifikasi 'kemungkinan' penyebar rumor pada Twitter menggunakan model klasifikasi SVM dan ekstraksi level fitur dengan batasan data Twitter yang digunakan dalam bahasa Inggris dan fitur percakapan, *user*, dan *tweet*. Penelitian tugas akhir ini bertujuan untuk membuat sistem untuk mengidentifikasi 'kemungkinan' penyebar rumor pada Twitter serta mengevaluasi model dengan hasil *performance measure* yang terdiri dari nilai akurasi, precision, recall, dan F1-score. Sehingga model tersebut dapat dikatakan mampu mengidentifikasi 'kemungkinan' penyebar rumor pada Twitter.