

1. Pendahuluan

Latar Belakang

Di antara berbagai sistem *microblogging*, Twitter adalah layanan yang paling populer sejauh ini. Karena kemudahannya dalam berbagi informasi secara real-time, Twitter banyak mempengaruhi wacana publik di masyarakat. Di Twitter, banyak informasi dibagikan melalui jejaring sosialnya yang terstruktur, akan tetapi hanya sedikit yang mengetahui tentang bagaimana dan mengapa informasi tertentu menyebar lebih luas daripada yang lain (Bongwon Suh, 2010).

Secara struktural, retweet sendiri merupakan fitur Twitter yang setara dengan fitur *forward email* yang mana pengguna memposting pesan yang awalnya diposting oleh orang lain[1]. Melakukan *retweet* membawa orang baru kedalam sebuah topik tertentu, mengundang mereka untuk terlibat tanpa mengundang mereka secara langsung[1].

Pada paper acuan penulis beranggapan bahwa *Retweet* adalah mekanisme kunci untuk penyebaran informasi di Twitter. Peneliti sebelumnya percaya akan pentingnya untuk mengeksplorasi bagaimana retweet berfungsi untuk memahami bagaimana informasi disebarkan di Twitter jaringan dan untuk memahami mengapa tweet tertentu menyebar lebih banyak luas dari yang lain[2]. Maka dari itu peneliti sebelumnya gunakan 74 Juta tweet pada paper acuan sebagai data untuk mengidentifikasi faktor faktor yang mempengaruhi *retweet rate*. Berdasarkan hasil penelitian terdapat berbagai macam *feature* yang ada dalam data, ditemukan berbagai macam hal menarik yang dapat maupun tidak mempengaruhi *Retweetability* dari sebuah *tweet*[2]. Principal Component Analysis (PCA), Generalized Linear Model (GLM), Regular Expression Method, Feature Retweet Method digunakan penulis dalam penelitiannya untuk membantu mengkonversi data menjadi hasil prediksi[2]. Diharapkan penelitian ini juga dapat membantu mengetahui factor yang mempengaruhi penyebaran informasi, terutama dalam media social twitter.

Dalam penelitian ini akan digunakan *feature user-based* dalam data set hari pemilihan presiden amerika serikat yang berjumlah hampir 400 ribu data oleh Chris Albon dari website kaggle[11]. Akan digunakan 1000 data teratas sebagai data latih dan data uji menggunakan *Support Vector Machine* di bantu oleh k-fold cross validation untuk membangun sistem pada penelitian ini.

Topik dan Batasan

Sesuai dengan apa yang sudah disampaikan pada Latar Belakang, rumusan masalah dalam penelitian ini adalah memprediksi apakah suatu tweet akan di retweet atau tidak berdasarkan *feature user-based* yang dipilih dalam data. Dilihat juga performansi dan akurasi dari *Support Vector Machine* sebagai metode klasifikasi yang dipilih. Batasan masalah dari penelitian ini merupakan sebagian data yang digunakan dan *Feature User-Based* yang dipakai.

Tujuan

Tujuan dari penelitian ini adalah untuk membangun sebuah sistem yang diharapkan dapat memprediksi *retweet* berdasarkan *feature user-based* dengan menggunakan metode *Support Vector Machine*.

Organisasi Tujuan

Akan dijelaskan pada bagian bagian selanjutnya mengenai studi terkait penelitian ini pada bagian kedua. Pada bagian ketiga akan dipaparkan sistem yang dibangun dalam penelitian ini pada dan juga gambaran besar system. Pada bagian keempat evaluasi dari penelitian. Dan yang terakhir merupakan kesimpulan dan juga saran untuk penelitian selanjutnya.