

## 1. Pendahuluan

### Latar Belakang

Seiring berkembang pesatnya teknologi informasi, yang memudahkan dalam mendapatkan informasi melalui media sosial, dimana media sosial telah mendapat banyak perhatian dari masyarakat. Media sosial tidak hanya menyediakan beberapa informasi berita tetapi media sosial seperti Twitter menyediakan berbagai informasi yang diperlukan mulai dari informasi Kesehatan, Pendidikan, Olahraga, Politik, Makanan, dan Keuangan. Berdasarkan fakta yang ada dengan menggunakan situs media sosial masyarakat dapat berbagi informasi maupun perasaan yang mereka miliki untuk mengekspresikan kesukaan dan minat mereka terhadap suatu topik yang mendukung berbagai interaksi sosial dan penyebaran informasi diantara pengguna.

Banyaknya informasi yang tersebar melalui media sosial twitter menjadi sebuah tantangan bagi peneliti sehingga informasi yang tersedia dapat memaksimalkan pengaruh penyebarannya sehingga dapat dimanfaatkan secara lebih efektif. Tidak hanya membantu dalam mempelajarinya saja, tetapi juga dapat dimanfaatkan untuk memecahkan masalah penyebaran informasi yang ada pada media sosial Twitter. Berikut terdapat penelitian mengenai sistem penyebaran informasi pada media sosial yang sudah lebih dahulu dilakukan yaitu Sebagian besar sistem yang dihasilkan dapat mempelajari penyebaran informasi melalui media sosial berdasarkan model probabilitasnya. Pada penelitian [1] fitur yang digunakan berupa *Topic Information, User Activities, Network Connection, Similarity Between User*, dan *Content-Based Similarity*. Menurut penelitian [2] mereka mengusulkan metode yang berfokus pada model linear threshold (LT) untuk menghasilkan probabilitas antara pengguna. Mereka juga menunjukkan Teknik untuk memprediksi waktu yang diinginkan oleh pengguna sesuai dengan waktu yang diinginkan. Pada penelitian [3] dimana fitur yang digunakan yaitu user-based, content-based dan content similarity dan menghasilkan pemodelan yang. Pada penelitian [4] mengenai penyebaran informasi di twitter meneliti mengenai relasi *retweet* dari *user* pada sekumpulan data yang besar, dan menemukan bahwa *user* dengan *followers* yang besar memiliki peran yang besar.

Sistem difusi informasi twitter merupakan sistem penyebarluasan informasi pada media sosial twitter yang dilakukan secara meluas. Sistem ini ditujukan untuk meningkatkan pengetahuan dari segala bidang yang mana dalam penelitian ini lebih khusus membahas bidang kesehatan, antara lembaga dengan pengguna twitter. Sasaran yang hendak dicapai dari sistem ini adalah dapat memprediksi suatu informasi akan disebar atau tidak dengan penambahan penggunaan fitur *network*, kesamaan antara konten dan berbasis konten, serta melihat seberapa berpengaruh penambahan fitur tersebut terhadap prediksi sistem difusi informasi.

Sistem digunakan metode Bernoulli dan Gaussian *Naïve Bayes*. Metode klasifikasi ini dipilih karena dianggap memiliki tingkat akurasi yang lebih baik, efisien dan stabil [5] dibanding model klasifikasi yang lain serta menghasilkan nilai untuk melakukan prediksi berdasarkan fitur yang akan digunakan yaitu *Network Feature, Content Similarity* dan *Content Based Feature*. Menggunakan data dari media sosial twitter berbahasa indonesia sebagai datasetnya.

### Topik dan Batasan

Berdasarkan latar belakang yang telah dibuat, didapatkan rumusan masalah yang ada yaitu Bagaimana cara mendapatkan fitur *Network Feature, Content Similarity* dan *Content Based*, serta bagaimana cara menerapkan metode Bernoulli dan Gaussian *Naïve Bayes* dalam memprediksi suatu tweet akan disebarluaskan. Batasan masalah dari tugas akhir ini yaitu Fitur yang akan digunakan berupa *Network Feature, Content Similarity* dan *Content Based*. Metode yang digunakan hanya metode Bernoulli dan Gaussian *Naïve Bayes*. Serta data yang digunakan dari Twitter dengan mengambil topik tentang vaksinasi covid-19.

### Tujuan

Tujuan dari penelitian ini adalah membangun sistem yang dapat memprediksi penyebaran informasi di twitter menggunakan metode Bernoulli dan Gaussian *Naïve Bayes* yang menerapkan beberapa fitur seperti *Network Feature, Content Similarity*, dan *Content Based Feature*.

### Organisasi Tulisan

Organisasi tulisan yang terdapat pada proposal ini yaitu pendahuluan, studi terkait, system yang dibangun, evaluasi, kesimpulan dan daftar Pustaka. Pada pendahuluan merupakan penjelasan lebih detail dari abstrak lebih utama lagi mendeskripsikan mengenai latar belakang penjelasan atau mengidentifikasi topik atau masalah erta batasan, adapula tujuan, serta metode penelitian. Untuk studi terkait berisikan teori atau studi literatur dimana berkaitan sesuai topik tugas akhir yang dilakukan. Sistem yang dibangun dapat menjelaskan dan mendeskripsikan rancangan dan system atau produk yang dihasilkan. Evaluasi yaitu berupa evaluasi pengujian yang didapat dan analisis dari pengujian yang dilakukan. Kesimpulan yaitu menjelaskan hasil akhir yang didapat dari hasil pengujian yang dilakukan serta analisa hasil. Daftar Pustaka berisi literatur yang membantu dalam pengerjaan.