

1. Pendahuluan

Latar Belakang

Akhir – akhir ini peran media sosial dalam menyebarkan informasi sangatlah penting. Keanekaragaman sifat atau perilaku manusia tidak hanya dapat dilihat melalui interaksi secara langsung, tetapi juga dapat dilihat melalui media sosial berdasarkan aktivitas penggunanya. Platform media sosial memungkinkan pengguna untuk memberikan informasi mengenai data pengenalan pribadi mereka, saling berbagi informasi, dan juga dengan memungkinkan penggunanya untuk membuat konten seperti cuitan (*tweet*), gambar, tautan dan banyak hal lain. Data tersebut kemudian dapat diolah untuk mengidentifikasi perilaku pengguna dan menganalisis karakteristik pengguna berdasarkan topik tertentu [1].

Salah satu *platform* atau wadah media sosial yang sangat populer untuk dilakukan penelitian adalah Twitter. Twitter mulai dikenal pada tahun 2007 dan merupakan layanan yang dapat menghubungkan pengguna dengan memberikan rekomendasi berita atau topik yang sedang hangat diperbincangkan dalam bentuk cuitan. Twitter ini sendiri menggunakan sistem *microblogging* yang memungkinkan penggunanya untuk membuat dan membaca cuitan hingga 140 karakter di setiap cuitannya. Serta memiliki total lebih dari 15.7 juta pengguna aktif [2].

Teknik pra-pemrosesan atau *preprocessing* data sangat perlu dilakukan sebelum data diolah dalam metode *clustering* atau metode klasifikasi seperti *Support Vector Machine (SVM)*, *Deep Learning (DL)* dan *Naïve Bayes (NB)* untuk mengontrol bentuk data yang seringkali tidak terstruktur dan sulit untuk dikelola. Tahap ini berperan penting dalam akurasi prediksi dan mengoptimalkan waktu komputasi pada sistem [3].

Ada banyak cara untuk mengidentifikasi perilaku pengguna pada media sosial tertentu secara akurat [4]. Teknik *clustering* merupakan teknik yang banyak digunakan untuk menemukan struktur *cluster* yang mirip satu sama lain dalam suatu kumpulan data. Beberapa potensi penelitian yang dapat dikembangkan dan diimplementasikan adalah penggunaan metode *K-Means Clustering* dan metode *Gaussian Mixture Model Clustering (GMM)* [5]. Algoritma *K-Means* sangat umum digunakan karena sederhana dan mudah untuk diimplementasikan, namun algoritma ini juga memiliki kekurangan yaitu sulit untuk memberikan probabilitas yang diinginkan secara optimal jika dihadapkan pada pola dataset yang besar dan *non-circular*. Sedangkan pada metode GMM terdapat varians yang memungkinkan metode ini untuk melakukan *clustering* jika diberikan pola dataset yang besar dan bervariasi [6].

Penggunaan algoritma *Expectation – Maximization* dalam *clustering* merupakan hal yang umum digunakan untuk memperkirakan probabilitas maksimum dengan variabel laten (variabel yang diamati dari data). Cara penggunaan algoritma ini adalah dengan menginisialisasi pendugaan model parameter awal, dan selanjutnya akan melakukan tahapan ekspektasi dengan mengestimasi nilai variabel laten dan terakhir dengan melakukan tahapan maksimalisasi dengan mengestimasi nilai baru yang akan digunakan sebagai model parameter baru [7], [8].

Pada studi lain [9] yang menganalisis perilaku pengguna adalah penelitian *platform* berbasis web Kassi menggunakan metode *Social Network Analysis (SNA)*. Kassi adalah forum pertukaran barang atau jasa di kampus yang disediakan oleh University of Finland. Condor dan NodeXL merupakan alat SNA yang digunakan. Condor digunakan untuk menganalisis dinamika evolusi jejaring sosial sedangkan NodeXL digunakan karena dapat bekerja dengan baik untuk menghasilkan visualisasi yang detail seperti bentuk, warna, ukuran dan lain-lainnya. Itu juga menggunakan metode sentralitas untuk menentukan perilaku pengguna.

Salah satu topik yang ramai diperbincangkan di media sosial adalah berkaitan dengan politik. Topik ini merupakan topik yang cukup sensitif karena banyaknya pro dan kontra dari opini publik. Oleh karena itu, penelitian ini memilih topik seputar politik untuk mengidentifikasi apakah perilaku pengguna terhadap topik tersebut positif atau negatif. Untuk mengetahui perilaku pengguna dapat dilihat berdasarkan *mood* dari cuitan pengguna itu sendiri, kemudian dapat ditandai positif jika mengandung ungkapan dukungan atau pujian dan dapat ditandai negatif jika mengandung ungkapan hujatan [10].

Topik dan Batasannya

Berdasarkan penjelasan pada latar belakang yang sudah dituliskan, maka dapat disimpulkan bahwa topik yang diambil pada penelitian ini berfokus kepada tipe *tweet* politik. Dimana data – data *tweet* ini akan selanjutnya dikumpulkan untuk diolah menggunakan teknik pengelompokan GMM. Hasil akhir yang diinginkan berupa analisis *cluster* yang nantinya akan dapat menentukan apakah sebuah *cluster* bersifat positif atau negatif.

Batasan – batasan yang dimiliki adalah pengerjaan *code* sistem akan dilakukan menggunakan bahasa *Python* dan data yang dikumpulkan akan disimpan dalam format “.csv” dan data akan diambil menggunakan API Twitter.

Tujuan

Tujuan dari dilakukannya penelitian ini adalah untuk mengetahui implementasi metode *Gaussian Mixture Model Clustering* dalam mengidentifikasi perilaku pengguna berdasarkan tipe *tweet* di Twitter, mengetahui hasil pengelompokan data dan karakteristik dari setiap cluster yang ada dan menganalisis hasil akhir *clustering* untuk menentukan jenis *user behavior*.

Organisasi Tulisan

Pada penelitian ini terdapat lima tahapan yang harus dilalui yaitu: pendahuluan, studi terkait, sistem yang akan dibangun, evaluasi dan kesimpulan. Pada tahapan pertama yaitu pendahuluan berisi tentang penjelasan latar belakang dan penguraian masalah, tujuan, Batasan, ruang lingkup serta proses yang nantinya akan dilakukan. Pada tahapan kedua yaitu studi terkait berisi mengenai penjelasan dan contoh dari studi – studi terkait dengan topik penelitian ini yang sebelumnya sudah dilakukan, hal ini dijadikan sebagai acuan penelitian. Pada tahapan ketiga yaitu sistem yang akan dibangun berisi langkah dan alur sistem pengerjaan yang digunakan untuk penelitian. Pada tahapan keempat yaitu evaluasi berisi tentang penjelasan apakah hasil pengujian dan sistem yang dibangun sudah sesuai dengan tujuan awal dan apakah hasil akhir yang didapatkan sesuai atau tidak. Pada tahapan terakhir yaitu kesimpulan berisi mengenai rangkuman dari seluruh penelitian yang dilakukan dan saran yang diberikan untuk pengembangan selanjutnya.