

## INTISARI

*Twitter* merupakan salah satu situs *microblogging* memungkinkan penggunanya untuk menulis tentang berbagai opini, komentar, dan berita yang membahas isu-isu yang terjadi pada saat ini. Banyak pengguna yang melakukan *posting* pendapat mereka akan sebuah produk atau layanan yang mereka gunakan. Hal tersebut dapat digunakan sebagai sumber data untuk menilai sentimen pada *Twitter*. Cara pengelompokan emosi secara otomatis dapat digunakan, salah satunya menggunakan *Agglomerative Hierarchical Clustering*. Tujuan dari penelitian ini adalah membangun sistem yang secara otomatis mampu mengelompokkan emosi setiap *tweet*, dan mengetahui tingkat akurasi pengelompokan.

Tahapan proses dimulai dari *preprocessing*, terdapat beberapa proses yaitu *tokenizing*, *stopword*, *stemming*, pembobotan kata, serta normalisasi, selanjutnya dapat diolah menggunakan *Agglomerative Hierarchical Clustering*. Proses *clustering* terdiri dari, menghitung matriks jarak antar data, mencari jarak terdekat, menggabungkan menjadi satu *cluster*, memperbaharui matriks hingga semua data menjadi satu *cluster*. Setelah itu melakukan perhitungan akurasi menggunakan *confusion matrix*. Selanjutnya untuk melihat kesesuaian sistem yang dibuat, maka dimasukkan data baru yang diproses dengan sistem, lalu dapat menentukan data tergolong salah satu jenis emosi.

Dari penelitian yang telah dilakukan, didapatkan total data *tweet* sebanyak 500 data serta jumlah *cluster* terbagi menjadi lima yaitu cinta, marah, sedih, senang, dan takut. Hasil penelitian analisis sentimen pada *twitter* dapat berjalan dengan baik dengan akurasi 81,6% untuk jumlah frekuensi kata unik maksimal 85 dan kata unik minimal 2 dengan melakukan normalisasi menggunakan *Z-Score*, perhitungan jarak menggunakan *Cosine Similarity* serta metode *AHC Average Linkage*.

Kata Kunci : *Tweet, Agglomerative Hierarchical Clustering, Cluster, Cosine similarity, Euclidean Distance, Z-Score, Min-Max, Confusion Matrix*

## ABSTRACT

Twitter is one of the microblogging sites allowing its users to write about various opinions, comments, and news discussing issues that are happening at the moment. Many users post their opinions on a product or service that they use. This phenomenon can be used as data source to assess sentiment on Twitter. Automatic grouping of emotions can be used, one of them using Agglomerative Hierarchical Clustering. The aims of this research were to build a system that automatically able to group the emotions of each tweet, and to find out accuracy level of grouping.

The steps of the process started from preprocessing, there were several processes, included tokenizing, stopword, stemming, word weighting, and normalization, then the data processed using Agglomerative Hierarchical Clustering.. Clustering process consisted of calculating the distance matrix between data, finding the closest distance, combining into one cluster, updating the matrix until all the data became one cluster. After that, calculating the accuracy using confusion matrix. To see the suitability of the system created, then inserted new data which were processed with the system so that the data could be determined to particular type of emotion.

The results of this research showed that there were 500 data tweet in total and the number of clusters were divided into five which were love, angry, sad, happy, and afraid. The result of sentimental analysis on twitter could run well with an accuracy of 81.6% for the maximum frequency of unique words 85 and unique word of at least 2 by normalizing using Z-Score, Cosine Similarity distance calculation and AHC Average Linkage method.

*Keyword : Tweet, Agglomerative Hierarchical Clustering, Cluster, Cosine similarity, Euclidean Distance, Z-Score, Min-Max, Confusion Matrix*