

Fahmirullah Abdillah 2023, **Klasterisasi dan Geovisualisasi *Tweet* Penyebaran Penyakit Menular Langsung (Studi Kasus COVID-19)**, Skripsi ini dibawah bimbingan Ira Puspitasari, S.T., M.T., Ph.D., dan Drs. Eto Wuryanto, DEA., Program Studi S1 Sistem Informasi. Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya

---

## ABSTRAK

Layanan media sosial *microblog* seperti Twitter menghasilkan aliran besar dalam penyebaran informasi terhadap suatu kejadian. Maka dari itu media sosial dapat dimanfaatkan sebagai bentuk pengawasan penyebaran penyakit menular langsung, seperti kasus pandemi COVID-19. Analisis klaster digunakan dalam penelitian ini untuk mengelompokkan gejala yang dialami oleh penderita COVID-19 secara daring melalui media sosial. Penelitian ini bertujuan untuk menciptakan hasil klaster yang optimal dalam pengawasan penyebaran penyakit menular langsung. Metode yang digunakan dalam membangun luaran dari penelitian ini adalah metode algoritma pengelompokan berbasis kepadatan *Density-Based Spatial Cluster of Application with Noise* (DBSCAN) dan algoritme augmentasi *Ordering Points to Identify the Clustering Structure* (OPTICS), serta *Silhouette Coefficient* untuk uji validasi luaran klaster yang terbentuk, yang diimplementasikan di bahasa pemrograman Python. Luaran dari penelitian ini adalah hasil klaster yang membentuk terhadap *tweet* terkait penyebaran penyakit menular langsung studi kasus COVID-19. Data *tweet* yang sudah diakuisisi akan dilakukan *preprocessing*, kemudian dilakukan pembobotan *term* yang selanjutnya akan melalui proses analisis metode *cluster*. Hasil terbaik dari analisis yang dilakukan kedua metode, didapatkan algoritma OPTICS memberikan hasil optimal dengan parameter *xi score* = 0,05 dan *minpts* = 10, dan uji validasi *silhouette coefficient* sebesar 0,6508317895. Analisis ini menghasilkan 6 klaster dan *term*, yaitu klaster 1 membentuk *term* sakit kepala, klaster 2 membentuk *term* diare, klaster 3 membentuk *term* pilek, klaster 4 membentuk *term* batuk, klaster 5 membentuk *term* covid, dan klaster 6 membentuk *term* demam.

**Kata Kunci:** Twitter, *Preprocessing*, *Clustering*, DBSCAN, OPTICS, *Density-based Algorithm*, *silhouette coefficient*.

Fahmirullah Abdillah 2023, **Klasterisasi dan Geovisualisasi *Tweet* Penyebaran Penyakit Menular Langsung (Studi Kasus COVID-19)**, Skripsi ini dibawah bimbingan Ira Puspitasari, S.T., M.T., Ph.D., dan Drs. Eto Wuryanto, DEA., Program Studi S1 Sistem Informasi. Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya

---

## ABSTRACT

Microblogging social media services such as Twitter generate a large flow of information dissemination towards an event. Therefore, social media can be used as a form of monitoring the spread of direct infectious diseases, such as the case of the COVID-19 pandemic. Cluster analysis was used in this study to group symptoms experienced by COVID-19 sufferers online through social media. This study aims to create optimal cluster results in monitoring the spread of direct infectious diseases. The methods used in building the output of this research are the density-based clustering algorithm method Density-Based Spatial Cluster of Application with Noise (DBSCAN) and the augmentation algorithm Ordering Points to Identify the Clustering Structure (OPTICS), as well as Silhouette Coefficient for validation tests of the output of the cluster formed, which is implemented in the Python programming language. The output of this study is the result of clusters that form tweets related to the spread of infectious diseases directly COVID-19 case studies. Tweet data that has been acquired will be preprocessed, then weighted terms which will then go through the cluster method analysis process. The best results from the analysis carried out by both methods, the OPTICS algorithm obtained optimal results with parameters  $xi\ score = 0.05$  and  $minpts = 10$ , and silhouette coefficient validation test of 0.6508317895. This analysis produced 6 clusters and terms, namely cluster 1 formed the term headache, cluster 2 formed the term diarrhea, cluster 3 formed the term cold, cluster 4 formed the term cough, cluster 5 formed the term covid, and cluster 6 formed the term fever.

**Keywords:** Twitter, *Preprocessing*, *Clustering*, DBSCAN, OPTICS, *Density-based Algorithm*, *silhouette coefficient*.