

ABSTRAK

Nama Mahasiswa : Goklas Gebryel Siagian
NIM : 41520010216
Program Studi : Teknik Informatika
Judul Proposal Penelitian : Pengelompokan (*Clustering*) Data *Tweet* Pada Analisis Tanggapan Masyarakat Di Twitter (X) Terhadap Polusi Udara Di Jakarta Menggunakan Algoritma *K-Means*
Pembimbing : Umniy Salamah, S.Kom, MMSI

Polusi udara merupakan masalah lingkungan yang serius di kota-kota besar seperti Jakarta, yang dapat menyebabkan berbagai masalah kesehatan dan lingkungan. Dengan memanfaatkan data mining dan teknik clustering, penelitian ini mengelompokkan data tweet berdasarkan kesamaan konten untuk mengidentifikasi pola opini dan sentimen masyarakat terhadap polusi udara. Data tweet dikumpulkan dan diproses melalui beberapa tahap, termasuk pembersihan data, tokenisasi, dan penghapusan stopword. Algoritma K-Means kemudian diterapkan untuk mengelompokkan tweet-tweet tersebut menjadi beberapa kluster. Hasil clustering dievaluasi menggunakan metrik Silhouette Score untuk memastikan kualitas kluster yang dihasilkan. Hasil penelitian menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan tweet berdasarkan kesamaan konten, menghasilkan beberapa kluster yang menggambarkan variasi tanggapan masyarakat. Kluster-kluster ini memberikan wawasan tentang bagaimana masyarakat merespons isu polusi udara di Jakarta, yang dapat digunakan untuk perbaikan kebijakan lingkungan dan peningkatan kualitas udara. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam literatur ilmiah terkait analisis teks dan lingkungan, serta menjadi referensi bagi penelitian dan pengembangan metode analisis clustering di masa depan.

Kata Kunci: *Clustering, Data mining, K-Means, Polusi udara, Twitter*

ABSTRACT

Name : Goklas Gebryel Siagian
NIM : 41520010216
Study Program : Teknik Informatika
Title Research Proposal : Pengelompokan (*Clustering*) Data *Tweet* Pada Analisis Tanggapan Masyarakat Di Twitter (X) Terhadap Polusi Udara Di Jakarta Menggunakan Algoritma *K-Means*
Supervisor : Umniy Salamah, S.Kom, MMSI

Air pollution is a serious environmental problem in big cities like Jakarta, which can cause various health and environmental problems. Utilising data mining and clustering techniques, this study groups tweet data based on content similarity to identify patterns of public opinion and sentiment towards air pollution. The tweet data is collected and processed through several stages, including data cleaning, tokenisation, and stopword removal. The K-Means algorithm is then applied to group the tweets into clusters. The clustering results are evaluated using the Silhouette Score metric to ensure the quality of the resulting clusters. The results show that the K-Means algorithm is effective in clustering tweets based on content similarity, resulting in several clusters that illustrate the variation of public responses. These clusters provide insights into how people respond to the issue of air pollution in Jakarta, which can be used for environmental policy improvement and air quality improvement. This research is expected to make a significant contribution to the scientific literature related to text analysis and the environment, as well as a reference for future research and development of clustering analysis methods.

Keywords: *Clustering, Data mining, K-Means, Air pollution, Twitter*