



**Pengelompokan (*Clustering*) Data Tweet Pada Analisis Tanggapan
Masyarakat Di Twitter (X) Terhadap Polusi Udara di Jakarta
Menggunakan Algoritma *K-Means***

LAPORAN TUGAS AKHIR



**GOKLAS GEBRYEL SIAGIAN
41520010216**

**PROGRAM STUDI TEKNIK INFORMATIKA
/FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2024**



**Pengelompokan (*Clustering*) Data Tweet Pada Analisis Tanggapan
Masyarakat Di Twitter (X) Terhadap Polusi Udara di Jakarta
Menggunakan Algoritma *K-Means***

LAPORAN TUGAS AKHIR

GOKLAS GEBRYEL SIAGIAN

41520010216

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2024

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan dibawah ini:

Nama : Goklas Gebryel Siagian
NIM : 41520010216
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Pengelompokan (*Clustering*) Data *Tweet* Pada Analisis Tanggapan Masyarakat Di Twitter (X) Terhadap Polusi Udara Di Jakarta Menggunakan Algoritma *K-Means*

Menyatakan bahwa Laporan Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 20 Juni 2024



Goklas Gebryel Siagian

HALAMAN PENGESAHAN

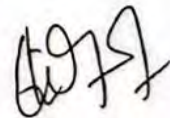
Laporan Skripsi ini diajukan oleh:

Nama : Goklas Gebryel Siagian
NIM : 41520010216
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Pengelompokan (*Clustering*) Data *Tweet* Pada Analisis Tanggapan Masyarakat Di Twitter (X) Terhadap Polusi Udara Di Jakarta Menggunakan Algoritma *K-Means*

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Umniy Salamah, S.Kom, MMSI
NIDN : 0306098104
Ketua Penguji : Dr. Hadi Santoso, S.Kom., M.Kom
NIDN : 0225067701
Penguji 1 : Siti Maesaroh, S.Kom., MTI
NIDN : 0413059003
Penguji 2 : Wawan Gunawan, S.Kom., MT
NIDN : 0424108104



Jakarta, 19 Juni 2024

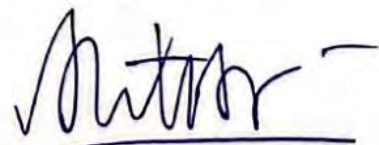
Mengetahui,

Dekan



Dr. Bambang Jokonowo, S.Si., MTI

Ketua Program Studi



Dr. Hadi Santoso, S.Kom., M.Kom

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT, Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Prof. Dr. Andi Adriansyah, M.Eng, selaku Rektor Universitas Mercu Buana
2. Bapak Dr. Bambang Jekonowo, S.Si., MTI, selaku Dekan Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Dr. Hadi Santoso, S.Kom., M.Kom, selaku Ketua Program Studi Teknik Informatika.
4. Umniy Salamah, S.Kom, MMSI, selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan laporan skripsi ini.
5. Siti Maesaroh, S.Kom., MTI dan Wawan Gunawan, S.Kom., MT selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Kedua orang tua dan keluarga, yang senantiasa selalu memberikan kasih sayang, doa, motivasi, nasihat serta dukungan yang begitu besar.
7. Dan teman – teman Teknik Informatika Universitas Mercu Buana Menteng yang telah mensupport saya sehingga saya berhasil menyelesaikan penelitian ini.

Akhir kata, saya berharap Allah SWT, Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu

Jakarta, 20 Juni 2024

Goklas Gebryel Siagian

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Goklas Gebryel Siagian
NIM : 41520010216
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Pengelompokan (*Clustering*) Data *Tweet* Pada Analisis Tanggapan Masyarakat Di Twitter (X) Terhadap Polusi Udara Di Jakarta Menggunakan Algoritma *K-Means*

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan)

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 20 Juni 2024

Yang menyatakan,



Goklas Gebryel Siagian

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*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian	3
BAB II.....	4
TINJAUAN PUSTAKA	4
2.1 Penelitian Terdahulu.....	4
2.2 Teori Pendukung	23
BAB III.....	28
METODE PENELITIAN.....	28
3.1 Jenis Penelitian	28
3.2 Tahapan Penelitian	28
BAB IV.....	34
PEMBAHASAN	34
4.1 Dataset.....	34
4.2 <i>Pre-Processing</i> Data.....	35
4.3 <i>Term Weighting</i>	38

4.4 Analisis Algoritma.....	39
BAB V.....	47
KESIMPULAN DAN SARAN	47
5.1 Kesimpulan	47
5.2 Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN.....	52
Lampiran 1 Kartu Asistensi	52
Lampiran 2 Curriculum Vitae	53
Lampiran 3 Sertifikat BNSP.....	54
Lampiran 4 Surat Pernyataan HAKI	55
Lampiran 5 Form Revisi Dosen Penguji	57
Lampiran 6 Halaman Persetujuan	59



UNIVERSITAS
MERCU BUANA

DAFTAR TABEL

Tabel 2. 1 Daftar Penelitian terdahulu.....	4
Tabel 4. 1 Contoh Hasil Proses TF-IDF	38
Tabel 4. 2 Hasil Jumlah <i>Tweets</i> Setiap <i>Cluster</i>	41
Tabel 4. 3 Contoh Hasil <i>Cluster</i> setiap <i>Tweets</i>	42



DAFTAR GAMBAR

Gambar 3. 1 Flowchart Tahapan Penelitian	28
Gambar 4. 1 Sampel data mentah.....	34
Gambar 4. 2 Atribut.....	35
Gambar 4. 3 <i>Cleaning</i>	36
Gambar 4. 4 <i>Case folding</i> and Tokenization	36
Gambar 4. 5 Filtering atau Stopword Removal.....	37
Gambar 4. 6 <i>Stemming</i> Data.....	38
Gambar 4. 7 Hasil <i>WordCloud</i> dari TF-IDF.....	39
Gambar 4. 8 Penentuan jumlah <i>Cluster</i> dengan metode <i>Elbow</i>	40
Gambar 4. 9 Hasil <i>Cluster Plot Text Tweets</i>	42
Gambar 4. 10 <i>WordCloud Cluster 1</i>	44
Gambar 4. 11 <i>WordCloud Cluster 2</i>	45
Gambar 4. 12 <i>WordCloud Cluster 3</i>	45

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi.....	52
Lampiran 2 Curriculum Vitae	53
Lampiran 3 Sertifikat BNSP.....	54
Lampiran 4 Surat Pernyataan HAKI	55
Lampiran 5 Form Revisi Dosen Penguji	57
Lampiran 6 Halaman Persetujuan	59

