SISTEM IMPLEMENTASI ANALISIS KLASTERISASI DAN SENTIMEN DATA UNTUK MENGETAHUI TOPIK YANG PALING BANYAK DIBAHAS MENGGUNAKAN METODE K-MEANS (STUDI KASUS DATA TWITTER DAN YOUTUBE TENTANG PPKM)

Tesis
untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi

Dikky Nayosa
30000320410010

SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2023

HALAMAN PERSETUJUAN

Tesis dengan judul :

SISTEM IMPLEMENTASI ANALISIS KLASTERISASI DAN SENTIMEN DATA UNTUK MENGETAHUI TOPIK YANG PALING BANYAK DIBAHAS MENGGUNAKAN METODE K-MEANS (STUDI KASUS DATA TWITTER DAN YOUTUBE TENTANG PPKM)

Oleh:
Dikky Nayosa
30000320410010

Telah dilakukan pembimbingan tesis dan dinyatakan layak untuk mengikuti ujian tesis pada Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang,
Menyetujui,

Pembimbing I	Pembimbing II
Dr. Rahmat Gernowo, MSI	Dr.Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPU, ASEAN Eng.
NIP.196511231994031003	NIP.197007272000121001

PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS

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

Nama	:	Dikky Nayosa
NIM	:	30000320410010
Program Studi	:	Magister Sistem Informasi
Program	:	Sekolah Pascasarjana
Jenis Karya	:	Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

SISTEM IMPLEMENTASI ANALISIS KLASTERISASI DAN SENTIMEN DATA UNTUK MENGETAHUI TOPIK YANG PALING BANYAK DIBAHAS MENGGUNAKAN METODE K-MEANS (STUDI KASUS DATA TWITTER DAN YOUTUBE TENTANG PPKM)

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang
Pada tanggal : 7 Juni 2023
Yang menyatakan

Dikky Nayosa
NIM 30000320410010

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 7 Juni 2023

Dikky Nayosa

SISTEM IMPLEMENTASI ANALISIS KLASTERISASI DAN SENTIMEN DATA UNTUK MENGETAHUI TOPIK YANG PALING BANYAK DIBAHAS MENGGUNAKAN METODE K-MEANS (STUDI KASUS DATA TWITTER DAN YOUTUBE TENTANG PPKM)

ABSTRAK

Pada tahun 2021, pemerintah menerapkan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang menarik perhatian publik saat ini. Pemberlakuan PPKM ini mendapat tanggapan yang beragam dari masyarakat. Twitter dan Youtube adalah platform media sosial yang digunakan oleh pengguna untuk menyampaikan pendapat secara langsung. Konten yang terdapat di dalam Twitter dan Youtube juga sangat beragam. Dalam konteks ini, dibutuhkan suatu metode pendeteksian topik secara otomatis, seperti metode *Mini Batch K-means Clustering*, yang dapat mempermudah pengguna dalam mengakses informasi. Penelitian ini menggunakan pendekatan *Mini Batch*, di mana hanya sekelompok kecil data yang digunakan dalam proses pengelompokan (*clustering*). Hasil penelitian menunjukkan bahwa untuk data tweet dengan kata kunci PPKM, diperoleh 12 kelompok cluster berdasarkan pengujian menggunakan *Sum of Squared Error* (SSE). Setelah proses pengelompokan data, hasil clustering akan divisualisasikan menggunakan *Word Cloud*, dan sistem akan menampilkan persentase kata-kata yang sesuai dengan *Word Cloud* tersebut.

Kata kunci : Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), clustering, K-Means, Mini Batch K-Means Clustering

IMPLEMENTATION SYSTEM FOR CLUSTER ANALYSIS AND DATA SENTIMENT USING THE K-MEANS METHOD TO DETERMINE THE MOST DISCUSSED TOPICS

ABSTRACT

The Community Activities Restrictions Enforcement (CARE) government rule, that is currently a matter of public concern, will be enforced in 2021. Community Activities Restrictions Enforcement (CARE) highlights the community's pros and cons. Twitter and YouTube are social media that facilitate direct user expression. The material offered on social media platforms Twitter and YouTube is likewise quite diversified. Therefore, an automatic approach for topic detection is required, such as Mini Batch K-means Clustering, which facilitates user access to information. This study employs the Mini Batch method, which utilizes just a limited set of data for the clustering procedure. Based on testing with the Sum of Squared Error, this study's clustering results for tweet data including the phrase Community Activities Restrictions Enforcement (CARE) produced 12 cluster groups. The clustering results will be represented using Word Cloud, and the system will display the percentage of words based on Word Cloud.

Keywords : Community Activities Restrictions Enforcement (CARE), clustering, K-Means, Mini Batch K-Means Clustering

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
tf	Suatu kata atau *term* terdapat dalam dokumen atau tidak
Ln	Fungsi *inverse*
n	Jumlah dokumen
W	Bobot
K	Jumlah klaster
Xi	Data ke i
Xn	Data ke n
Ck	Centroid klaster
μ_k	Titik awal klaster
$\in$	Anggota dari

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
TF	*Term* frekuensi
IDF	*Inverse* dokumen frekuensi
SSE	*Sum of Squared Errors*
PSBB	Pembatasan Sosial Berskala Besar
PPKM	Pemberlakuan Pembatasan Kegiatan Masyarakat
CSV	*Comma Separated Values*

URL	*Uniform Resource Locators*