

**KLASTERISI OPINI MASYARAKAT PADA *TWITTER*
TERHADAP PELAYANAN BPJS SELAMA MASA PANDEMI
MENGUNAKAN *K-MEANS* DAN *WARD***

SKRIPSI



JUNDAN BAROFIL HIBBAN

**PROGRAM STUDI S1 SISTEM INFORMASI
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA
2022**

KLASTERISASI OPINI MASYARAKAT PADA *TWITTER* TERHADAP PELAYANAN BPJS SELAMA MASA PANDEMI MENGUNAKAN *K-MEANS* DAN *WARD*

SKRIPSI



JUNDAN BAROFIL HIBBAN
NIM. 081511633023

PROGRAM STUDI S1 SISTEM INFORMASI
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA
2022

**KLASTERISASI OPINI MASYARAKAT PADA *TWITTER*
TERHADAP PELAYANAN BPJS SELAMA MASA PANDEMI
MENGUNAKAN *K-MEANS* DAN *WARD***

SKRIPSI

**Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer Bidang
Sistem Informasi Pada Fakultas Sains Dan Teknologi Universitas Airlangga**

Oleh :

Jundan Barofil Hibban

NIM: 081511633023

Disetujui Oleh :

Pembimbing I,



Taufik, S.T., M.Kom.
NIP. 197101042008121001

Pembimbing II,



Purbandini, S.Si., M.Kom.
NIP. 197107122008122001

LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Klasterisasi Opini Masyarakat Pada *Twitter* Terhadap Pelayanan BPJS Selama Masa Pandemi Menggunakan *K-Means* dan *Ward*

Penyusun : Jundan Barofil Hibban

NIM : 081511633023

Pembimbing I : Taufik ST, M.Kom.

Pembimbing II : Purbandini, S.Si, M.Kom.

Tanggal Ujian : 22 Juli 2022

Disetujui Oleh

Pembimbing I



Taufik ST, M.Kom.
NIP. 197101042008121001

Pembimbing II

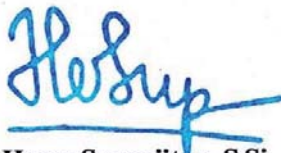


Purbandini, S.Si, M.Kom.
NIP. 197107122008122001

Mengetahui,

Ketua

Departemen Matematika
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Herry Suprajitno, S.Si., M.Si.
NIP. 196804041994031020

Koordinator

Program Studi S1 Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Rimuljo Hendradi, S.Si., M.Si.
NIP. 197102111997021001

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, tetapi tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun, dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah. Dokumen skripsi ini merupakan hak milik Universitas Airlangga.

SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama : Jundan Barofil Hibban
NIM : 081511633023
Program Studi : S-1 Sistem Informasi
Fakultas : Sains dan Teknologi
Jenjang : Sarjana (S1)

Menyatakan bahwa saya tidak melakukan kegiatan plagiat dalam penulisan skripsi saya yang berjudul :

“KLASTERISASI OPINI MASYARAKAT PADA *TWITTER* TERHADAP
PELAYANAN BPJS SELAMA MASA PANDEMI MENGGUNAKAN K-MEANS
DAN *WARD*”

Apabila suatu saat nanti terbukti melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, Juli 2022



Jundan Barofil Hibban

NIM. 081511633023

KATA PENGANTAR

Puji syukur kepada kehadiran Allah SWT karena berkat rahmat-Nya penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Klasterisasi Opini Masyarakat Terhadap Pelayanan BPJS Selama Masa Pandemi Menggunakan *K-Means* dan *Ward*”. Penulis tak lupa berterima kasih kepada Taufik ST, M.Kom. selaku dosen pembimbing I dan Purbandini, S.Si, M.Kom. selaku dosen pembimbing II, yang selalu sabar dan ikhlas membimbing dan memberikan ilmunya kepada penulis. Penulis juga berterima kasih kepada orang tua dan keluarga yang selalu mendukung penelitian penulis, orang terdekat dan teman-teman penulis yang telah memberikan semangat, bantuan, dan dukungannya kepada penulis.

Penulis menyadari bahwa dalam penyusunan skripsi ini terdapat kekurangan, untuk itu penulis senantiasa bersedia dan terbuka dalam menerima saran serta kritik yang bersifat membangun demi kesempurnaan skripsi ini. Semoga skripsi ini dapat memberikan manfaat dan wawasan yang berguna, Aamiin

Surabaya, Juli 2022

Penulis



Jundan Barofil

UCAPAN TERIMAKASIH

Segala puji syukur kehadiran Tuhan Yang Maha Esa atas curahan anugerah dan karunia-Nya sehingga skripsi dengan judul **Klasterisasi Opini Masyarakat pada Twitter Terhadap Pelayanan BPJS Selama Masa Pandemi Menggunakan K-Means dan Ward** ini dapat diselesaikan.

Dalam pelaksanaan dan penyusunan skripsi ini, penulis banyak menemui kendala, namun dengan adanya bantuan dari berbagai pihak, akhirnya penelitian ini dapat terselesaikan. Oleh karena itu, penulis tidak lupa mengucapkan terima kasih kepada:

1. Allah SWT yang senantiasa memberikan segala berkah, rahmat, dan karunia-Nya dan Rasulullah Muhammad SAW yang telah menjadi suri tauladan sehingga penulis dapat terus menjalani roda kehidupan dengan sebaik-baiknya.
2. Indarwati dan Arif Hidayatullah, selaku orang tua penulis yang telah memberikan dukungan penuh dalam bentuk doa, kasih sayang, dan materi sekaligus menjadi semangat dan motivasi bagi penulis dalam menyelesaikan studi dan skripsi dengan baik.
3. Ira Puspitasari, S.T., M.T., Ph.D. selaku dosen wali yang dengan sabar membantu penulis dalam menyelesaikan studi S1 Sistem Informasi Universitas Airlangga.
4. Taufik, S.T., M.Kom., selaku dosen pembimbing I yang dengan sabar memahami, membimbing, membantu, dan memberikan ilmu selama penulisan hingga skripsi ini terselesaikan.
5. Purbandini, S.Si., M.Kom., selaku dosen pembimbing II yang dengan sabar memahami, membimbing, membantu, dan memberikan ilmu selama penulisan hingga skripsi ini terselesaikan.
6. Cendra Devayana Putra dan Ahmad Alif Robit Al Hazmi yang telah membantu dengan berbagi ilmu pengetahuan selama penyusunan skripsi.
7. Nimas Azimatur Rohmah, yang telah memberi bantuan secara moral dan semangat selama penyusunan skripsi ini.

8. Muhammad Adam Bahtiar, Mohammad Agung Nugroho, Abdul Aziz Al Khakim, Mohammad Ichwan Satryo dan seluruh teman-teman S1 Sistem Informasi Universitas Airlangga angkatan 2015 yang telah mendukung satu sama lain dan membantu dengan berbagi ilmu pengetahuan selama penulisan skripsi.
9. Teman-teman serta berbagai pihak yang telah memberikan bantuan dan informasi selama proses penulisan skripsi.

Jundan Barofil Hibban. 2022. **KLASTERISASI OPINI MASYARAKAT PADA TWITTER TERHADAP PELAYANAN BPJS SELAMA MASA PANDEMI MENGGUNAKAN *K-MEANS* DAN *WARD***. Skripsi ini di bawah bimbingan Taufik, S.T., M.Kom. dan Purbandini, S.Si., M.Kom. S1 Sistem Informasi, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRAK

Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) merupakan Badan Hukum Publik yang bertugas menyelenggarakan jaminan Kesehatan Nasional bagi seluruh rakyat Indonesia terutama bagi Aparatur Sipil Negara dan Rakyat kurang mampu. Sejak 27 Desember 2019 jumlah penduduk Indonesia yang menggunakan BPJS Kesehatan sebanyak 224,1 Juta atau 83% dari total penduduk Indonesia. BPJS memiliki peran yang sangat besar dalam dunia kesehatan di Indonesia. Terlebih selama terjadinya pandemi covid 19. Hal ini tentunya menghasilkan beberapa opini rakyat akan peran BPJS selama pandemi di Indonesia. Dalam penelitian ini akan mining opini masyarakat yang dilakukan menggunakan selenium untuk scrapping data dari twitter. Pada proses Analisa sentiment akan dibutuhkan proses word embedding. *Word embedding* bertugas memetakan setiap kata dalam dokumen ke dalam dense *vector*, di mana sebuah *vector* merepresentasikan proyeksi kata di dalam ruang *vector*. Saat ini proses *word embedding* yang populer merupakan *Word2Vec* dan *GloVe*. Penelitian ini bertujuan mengetahui bagaimana opini masyarakat terhadap bpjs Kesehatan selama pandemic dan membandingkan algoritma kombinasi *word embedding* dan klasterisasi mana yang paling optimal jika diaplikasikan terhadap dataset teks dari twitter. Berdasarkan hasil penelitian dapat diketahui bahwa algoritma *Word2Vec* + *K-means* dinilai lebih unggul. Opini masyarakat diklasterkan menjadi 3 yaitu tentang Website BPJS dengan sentimen 40% positif dan 60% Negatif, Pandemic Corona dengan sentimen 30% positif dan 70% Negatif lalu yang terakhir tentang tarif Iuran BPJS Kesehatan dengan sentiment 31% Positif dan 69% negatif.

Kata Kunci: *BPJS Kesehatan, Analisa Sentimen, Word2Vec, GloVe, K-Means, Ward, SSE, DBI*

Jundan Barofil Hibban. 2022. **KLASTERISASI OPINI MASYARAKAT PADA TWITTER TERHADAP PELAYANAN BPJS SELAMA MASA PANDEMI MENGGUNAKAN *K-MEANS* DAN *WARD***. Skripsi ini di bawah bimbingan Taufik, S.T., M.Kom. dan Purbandini, S.Si., M.Kom. S1 Sistem Informasi, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRACT

Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) is a Public Legal Entity in charge of administering National Health insurance for all Indonesian citizen, especially for State Civil apparatuses and underprivileged citizen. Since December 27, 2019, the number of Indonesians citizen who use BPJS Kesehatan is 224.1 million, or 83% of the total population of Indonesia. BPJS has a very big role in the world of health in Indonesia. Especially during the covid 19 pandemic. This of course resulted in some people's opinions about the role of BPJS during the pandemic in Indonesia. In this study, public opinion mining will be carried out using selenium for scrapping data from Twitter. In the sentiment analysis process, a *word embedding* process will be needed. *Word embedding* is in charge of mapping every word in the document into a dense vector, where a vector represents the projection of words in the vector space. Currently, the most popular *word embedding* processes are *Word2Vec* and *GloVe*. This study aims to find out how public opinion is on the BPJS Kesehatan during the pandemic and compare the combination of *word embedding* and clustering algorithms, which is the most optimal when applied to text datasets from Twitter. Based on the results of the study, it can be seen that the *Word2Vec* + K – means algorithm is considered superior. Public opinion is clustered into 3, namely the BPJS Website with 40% positive and 60% negative sentiments, the Corona Pandemic with 30% positive and 70% negative sentiments and the last is the BPJS Kesehatan subscription rates with 31% Positive and 69% negative sentiments.

Keywords: BPJS Kesehatan, Sentiment Analysis, *Word2Vec*, *GloVe*, K-Means, Ward, SSE, DBI

DAFTAR ISI

KLASTERISASI OPINI MASYARAKAT PADA <i>TWITTER</i>	i
LEMBAR PENGESAHAN NASKAH SKRIPSI.....	iii
PEDOMAN PENGGUNAAN SKRIPSI	iv
SURAT PERNYATAAN TENTANG ORISINALITAS	v
KATA PENGANTAR	vi
UCAPAN TERIMA KASIH.....	vii
ABSTRAK.....	ix
ABSTRACT.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan.....	5
1.4 Manfaat.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 BPJS Kesehatan.....	7
2.2 <i>Web Scraping</i>	8
2.3 <i>Preprocessing Data</i>	8
2.4 <i>Word embedding</i>	16
2.5 Metode Klasterisasi	19

2.6	Pemilihan Metode Optimal	21
2.7	<i>Lexicon-Based Lebeling</i>	23
2.8	Analisa Hasil Klasterisasi.....	23
2.9	<i>Python Libraries</i>	24
2.10	Penelitian Terdahulu.....	26
BAB III METODE PENELITIAN.....		29
3.1	Waktu Penelitian	29
3.2	Objek Penelitian	29
3.3	Tahap Penelitian	30
BAB IV PEMBAHASAN DAN HASIL		40
4.1	Pemilihan <i>Keyword</i>	41
4.2	<i>Scraping Twitter</i>	42
4.3	Praproses Data.....	50
4.4	Word Embedding	56
4.5	Klasterisasi	63
4.6	Pemilihan Metode Optimal	68
4.7	Analisa dan Interpretasi hasil Klaster.....	69
BAB V KESIMPULAN DAN SARAN.....		73
5.1	Kesimpulan.....	73
5.2	Saran.....	73
DAFTAR PUSTAKA		75
LAMPIRAN.....		79
Lampiran 1: Daftar Kata Stopword.....		79
Lampiran 2: Daftar Kata Normalisasi		81

DAFTAR TABEL

Tabel 2. 1 Kombinasi yang tidak dibolehkan.....	14
Tabel 2. 2 Aturan pemengalan kompleks (Sumber: Tahitoe, 2011)	15
Tabel 2. 3 Range nilai dan interpretasi Silhouette Klaster.....	22
Tabel 3. 1 Atribut Data yang digunakan	29
Tabel 3. 2 Contoh dari hasil Scapping Twitter.	31
Tabel 3. 3 Contoh Hasil Case folding	32
Tabel 3. 4 Penghapusan symbol khusus pada Tweet.	32
Tabel 3. 5 Hasil proses penghapusan Angka dan Tanda baca	33
Tabel 3. 6 Hasil proses penghapusan White Space dan Single Character	33
Tabel 3. 7 Hasil proses tokenisasi	34
Tabel 3. 8 Hasil Proses Penghapusan Stopword	34
Tabel 3. 9 Contoh daftar Kata Non Standart Word handling.....	35
Tabel 3. 10 Hasil Proses Non Standart Word Handling	35
Tabel 3. 11 Hasil Proses Stemming	35
Tabel 4. 1 Karakteristik Dataset sebelum proses praproses.	40
Tabel 4. 2 Range Jumlah Kata dalam 1 Tweet.....	40
Tabel 4. 3 Range Jumlah Karakter pada 1 Tweet	40
Tabel 4. 4 Karakteristik Dataset Setelah Proses Praproses.	56
Tabel 4. 5 Range Jumlah kata pada 1 Tweet Setelah Proses Praproses.....	56
Tabel 4. 6 Range Jumlah Karakter pada 1 Tweet Setelah Proses Praproses.....	57
Tabel 4. 7 Parameter K - means	65
Tabel 4. 8 Parameter Ward.....	66
Tabel 4. 9 Komparasi SSE , Silhouette, dan DBI antar Algoritma.....	68
Tabel 4. 10 Hasil InSet Klaster 1	70
Tabel 4. 11 Hasil InSet Klaster 2	70
Tabel 4. 12 Hasil InSet Klaster 3	70

DAFTAR GAMBAR

Gambar 2. 1 Langkah Praprocessing Data. (Hidayatullah, 2016).....	9
Gambar 2. 2 Arsitektur Skip-gram (Tomas Mikolov, 2013)	10
Gambar 2. 3 Contoh Hasil Perhitungan Similarity	11
Gambar 2. 4 One Word Context CBOW (Rong, 2016)	17
Gambar 2. 5 Multiple Context Word CBOW (Rong, 2016)	17
Gambar 2. 6 Skip-Gram (Rong, 2016).....	18
Gambar 2. 7 Dendrogram Hirarki Klastering.....	20
Gambar 2. 8 Knee Point Pada Jumlah Klaster Optimal	22
Gambar 3. 1 Alur Penelitian.....	30
Gambar 3. 2 Kata stopwords yang sudah didefinisikan.	34
Gambar 3. 3 Potongan isi file Word Vectors <i>FastText</i>	36
Gambar 3. 4 Alur kerja Algoritma <i>GloVe</i>	37
Gambar 3. 5 vector kata <i>GloVe</i>	37
Gambar 3. 6 Komparasi hasil <i>GloVe</i> (a) dan <i>Word2Vec</i> (b)......	37
Gambar 4. 1 Diagram Pie untuk Distribusi Kata dan Karakter yang terdapat pada Tweet.....	41
Gambar 4. 2 Contoh tweet dengan Keyword “BPJS”	41
Gambar 4. 3 Tweet dengan Keyword " Layanan BPJS ".	42
Gambar 4. 4 Proses import Library.....	42
Gambar 4. 5 Proses link ke Twitter dan memasukan search term	44
Gambar 4. 6 Halaman Search pada Twitter	44
Gambar 4. 7 Proses scraping Tweet yang ditampilkan pada halaman chrome.....	45
Gambar 4. 8 Script Autoscroll	45
Gambar 4. 9 Code untuk menyimpan hasil scrape ke csv.....	46
Gambar 4. 10 Raw Data hasil Scrapinig.	46
Gambar 4. 11 Proses Penggabungan hasil Scraping	47
Gambar 4. 12 Regular Expression yang digunakan untuk kombinasi file.....	47
Gambar 4. 13 Proses penggabungan dan penyimpanan file Scraping	48
Gambar 4. 14 Tweet Yang berisikan Handle, Username dan Timestamp	48

Gambar 4. 15 Proses Penghapusan username dan handle.....	48
Gambar 4. 16 Output proses penghapusan Username dan Handle	49
Gambar 4. 17 Proses Penghapusan Mention.....	49
Gambar 4. 18 Hasil Output dari proses Penghapusan Mention	49
Gambar 4. 19 Proses penghapusan tweet duplikat.....	50
Gambar 4. 20 Proses dan Output Case Folding	51
Gambar 4. 21 Proses Penghapusan simbol khusus Tweet.	51
Gambar 4. 22 Output Proses pembersihan Khusus di Twitter	51
Gambar 4. 23 Proses dan Output Penghapusan Angka dan tanda Baca	52
Gambar 4. 24 Proses dan Output Penghapusan White Space dan Single Character	52
Gambar 4. 25 Proses dan Output dari Tokenisasi	53
Gambar 4. 26 Contoh Kata Stopword.....	53
Gambar 4. 27 Proses dan Output dari Stopword Removal.	54
Gambar 4. 28 contoh List kata yang dinormalisasi.....	54
Gambar 4. 29 Proses dan Output Non Standart Word Handling	55
Gambar 4. 30 Proses Stemming.....	55
Gambar 4. 31 Output proses Stemming	56
Gambar 4. 32 Diagram Pie untuk Distribusi Kata dan Karakter yang terdapat pada Tweet setelah Proses Praproses.	57
Gambar 4. 33 Cloning Fasttext dari github.....	57
Gambar 4. 34 Proses akusisi Text Vector untuk <i>Word2Vec</i>	57
Gambar 4. 35 Script Praprocessing & Pad Sequence	58
Gambar 4. 36 Script Proses Word Embedding	59
Gambar 4. 37 Proses menyimpan Document Matrix ke file bermformat .npy	59
Gambar 4. 38 Clonning Model <i>GloVe</i>	60
Gambar 4. 39 Proses pelatihan <i>GloVe</i>	61
Gambar 4. 40 Konversi vector <i>GloVe</i> ke <i>Word2Vec</i>	62
Gambar 4. 41 Input Model yang sudah Dikonversi	62
Gambar 4. 42 Proses Embedding <i>GloVe</i>	63
Gambar 4. 43 Menjadinan Term Document Matrix menjadi 2 Dimensi.	64

Gambar 4. 44 Script Dimension Reduction	64
Gambar 4. 45 SSE <i>GloVe</i> + K - Means.....	65
Gambar 4. 46 SSE <i>Word2Vec</i> + K - Means.....	66
Gambar 4. 47 Silhouette <i>Word2Vec</i> + Ward.....	67
Gambar 4. 48 Silhouette <i>GloVe</i> + Ward	67
Gambar 4. 49 Komparasi DBI antar kombinasi Algoritma	68
Gambar 4. 50 Wordcloud Klaster 1	71
Gambar 4. 51 Wordcloud Klaster 2	71
Gambar 4. 52 Wordcloud Klaster 3.	72
Gambar 4. 53 Wordcloud Klaster 2.	72