

SKRIPSI

**ANALISIS SENTIMEN TERHADAP PANDEMI COVID-19 DI
INDONESIA MENGGUNAKAN K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE**



ARYA SURYA SATRIA

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

IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

SKRIPSI

**ANALISIS SENTIMEN TERHADAP PANDEMI COVID-19 DI
INDONESIA MENGGUNAKAN K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE**



ARYA SURYA SATRIA

081711633026

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

**ANALISIS SENTIMEN TERHADAP PANDEMI COVID-19 DI
INDONESIA MENGGUNAKAN K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE**

SKRIPSI

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

Oleh :

Arya Surya Satria

NIM: 081711633026

Disetujui Oleh :

Pembimbing I,



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

Pembimbing II,



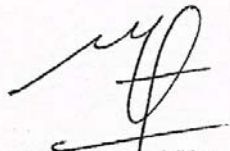
Purbandini, S.Si., M.Kom.
NIP. 197107122008122001

LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Analisis Sentimen Terhadap Pandemi Covid-19 di Indonesia
Menggunakan *K-Nearest Neighbor* Dan *Support Vector Machine*
Penyusun : Arya Surya Satria
NIM : 081711633026
Pembimbing I : Taufik, S.T., M.Kom.
Pembimbing II : Purbandini, S.Si., M.Kom.
Tanggal Ujian : 29 Juli 2022

Disetujui Oleh,

Pembimbing I,



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

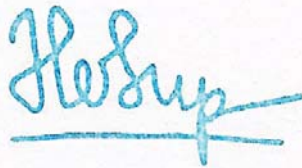
Pembimbing II,



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

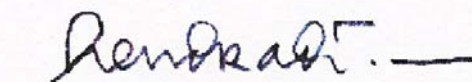
Mengetahui,

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

SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama : Arya Surya Satria
NIM : 081711633026
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 :

“ANALISIS SENTIMEN TERHADAP PANDEMI COVID-19 DI INDONESIA MENGGUNAKAN K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE”

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



Arya Surya Satria

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.

Dokumen skripsi ini merupakan hak milik Universitas Airlangga

KATA PENGANTAR

Puji syukur saya panjatkan atas kehadiran Allah SWT yang telah melimpahkan anugerah-Nya sehingga saya dapat menyelesaikan penyusunan tugas akhir skripsi yang berjudul “**Analisis Sentimen Terhadap Pandemi Covid-19 di Indonesia Menggunakan K-Nearest Neighbor dan Support Vector Machine**”. Berkat rahmat-Nya, penyusun dapat mengerjakan penyusunan dengan lancar sampai pada tahap ini.

Terima kasih juga penyusun ucapkan kepada seluruh pihak yang telah mendukung maupun ikut membantu secara langsung dalam pengerjaan skripsi ini hingga dapat terselesaikan dengan baik. Ucapan tersebut ditujukan oleh penyusun pada:

1. Allah SWT yang senantiasa memberikan berkah dan rahmat-Nya sehingga penelitian ini dapat berjalan lancar dan memberi kekuatan pada penyusun untuk menyelesaikan tugas akhir skripsi ini.
2. Orang tua yaitu Ibu Kurnia Rachmawati dan Bapak Herjanto Maryono dan keluarga yang selalu mendukung dan memberi penyusun semangat serta fasilitas sehingga dapat menyelesaikan skripsi ini dengan baik.
3. Taufik, S.ST., M.Kom. selaku dosen pembimbing I yang selalu meluangkan waktunya untuk memberi bimbingan pada anak bimbingnya hampir setiap minggunya dan memberi pengarahan dengan baik sehingga penyusun dapat menyelesaikan penelitian ini.
4. Purbandini, S.Si., M.Kom. selaku dosen pembimbing II dan dosen wali yang sudah meluangkan waktunya untuk memberi bimbingan pada anak bimbingannya meskipun beliau sibuk serta memberikan ilmu dan arahan kepada penyusun dari mahasiswa baru hingga penelitian skripsi ini selesai.
5. Seluruh dosen Program Studi S1 Sistem Informasi Universitas Airlangga yang telah memberi ilmu akademik maupun non-akademik yang bermanfaat selama perkuliahan.

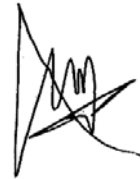
6. Naomi Jasmine, Dya Wahyu Anggraini, Adinda Novira, Jodie Raihan Z.F, selaku Mahasiswa Psikologi dan Ilmu Komunikasi yang membantu dalam pelabelan data dalam penelitian ini.
7. Teman-teman mahasiswa WSJF atau Program Studi S1 Sistem Informasi Universitas Airlangga angkatan 2017 yang telah menemani penyusun dari awal perkuliahan hingga saat ini, yang telah berbagi cerita dan pengalaman serta berkembang bersama sebagai mahasiswa.
8. Teman-teman ‘Hago’ dan ‘W’ yang telah menemani, mendukung dan berbagi cerita dan pengalaman di masa perkuliahan maupun saat pengerjaan penelitian ini.
9. Kakak-kakak tingkat yang telah memberi ilmu serta saran kepada penyusun, serta adik-adik tingkat yang telah menerima penyusun sebagai senior dalam hal organisasi maupun sebagai mahasiswa.
10. Ramdhan, Kevin, Jodie, Reza, Bachtiar, Virgha, Faiz, Nimas, Cera selaku teman-teman dari SMP dan SMA yang tidak pernah bosan untuk mendukung dan mengingatkan selama proses pengerjaan penelitian ini.
11. Anisa Haq, Wulan, Lia, Ilma, Meralda, Nashrul, Nadia, May, Bagas, Ulfah, Lila, Lintang, Ajeng, Mufidah, Rima selaku rekan dari Admin Geliat Airlangga yang telah memberi ruang dan kesempatan pada penyusun untuk menambah pengalaman dalam hal baru.
12. Teman-teman dan seluruh pihak yang telah membantu, memberikan semangat serta dukungan selama ini hingga penelitian ini terselesaikan dengan baik.
13. Fahira Firdausi, yang sudah menemani dan membantu dalam penyusunan, merapikan dalam pelabelan data, selalu mengingatkan tenggat akhir, dan memberikan semangat dan motivasi dalam pengerjaan untuk tugas akhir skripsi ini.
14. Diri sendiri yang telah berusaha keras dan tidak mudah menyerah dalam mengerjakan penelitian ini.

Akhir kata, penyusun menyadari jika terdapat kekurangan pada skripsi ini, oleh karena itu, kritik dan saran akan sangat bermanfaat. Penyusun berharap,

semoga penelitian ini kelak dapat selesai dengan baik dan menjadi sumber ilmu bagi seluruh praktisi maupun akademisi untuk masa depan yang lebih baik.

Surabaya, Juli 2022

Penyusun

A handwritten signature in black ink, consisting of stylized, overlapping loops and lines, positioned above the name Arya Surya Satria.

Arya Surya Satria

Arya Surya Satria. 2022. **Analisis Sentimen Terhadap Pandemi Covid-19 di Indonesia Menggunakan *K-Nearest Neighbor* dan *Support Vector Machine***. 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

Semenjak Indonesia ditemukan kasus pada bulan maret tahun 2020, kurva penyebaran covid meningkat secara signifikan di Indonesia. Hal tersebut menimbulkan dampak-dampak pada pandemi covid-19 ini, seperti frustrasi, kebosanan, kecemasan, dan ketakutan yang tidak terkontrol. Di Twitter terdapat banyak opini terkait dengan dampak dari pandemi covid-19 ini, terutama pada saat hari-hari besar dan *event* yang ada. Diperlukannya analisis sentimen untuk melihat bagaimana respon masyarakat Indonesia pada pandemi covid-19 ini. Penelitian ini bertujuan untuk mengetahui mengklasifikasikan dan menganalisa pola sentimen pada opini publik mengenai pandemi covid-19 serta mendapatkan algoritma terbaik untuk pengklasifikasian dengan metode *K-Nearest Neighbor* dan *Support Vector Machine*. Pengumpulan data dilakukan dengan menggunakan data Twitter terkait dengan pandemi covid-19 dengan kata kunci tertentu. Tahapan selanjutnya yaitu *pre-processing* yang meliputi *case Folding*, *tokenizing*, *stopword removal*, dan *stemming*. Lalu dilanjutkan merubah data teks menjadi matriks TF-IDF nantinya digunakan untuk proses klasifikasi *K-Nearest Neighbor* dan *Support Vector Machine*. Setelah itu dilakukan pelabelan data menggunakan *Indonesia Sentimen Lexicon* serta *Crowdsourced Labelling* untuk perbandingannya. Data diklasifikasikan menjadi tiga kelas, yaitu positif, negatif, dan netral. Hasil pengklasifikasian didapatkan label positif dengan sejumlah 2.850 data *tweet*, label negatif sejumlah 8.015 data *tweet*, dan label netral sejumlah 629 data *tweet*. Penerapan dari klasifikasi *K-Nearest Neighbor* menghasilkan akurasi 83,9% dan *Support Vector Machine* menghasilkan akurasi 76,9%. Respon masyarakat Indonesia mengenai dampak pandemi covid-19 ini didominasi oleh respon negatif. Klasifikasi dengan *K-Nearest Neighbor* memperoleh akurasi tertinggi dibandingkan dengan *Support Vector Machine*.

Kata Kunci: Dampak Pandemi Covid-19, Klasifikasi, *K-Nearest Neighbor* (K-NN), *Support Vector Machine* (SVM), TF-IDF

Arya Surya Satria. 2022. **Sentiment Analysis on Covid-19 Pandemic in Indonesia using K-Nearest Neighbor and Support Vector Machine**. This thesis is under the guidance of Taufik, S.T., M.Kom. and Purbandini, S.Si., M.Kom. Undergraduate Information Systems, Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya.

ABSTRACT

Since the covid case was discovered in Indonesia in March 2020, the curve of Covid spread in Indonesia has increased significantly. This affects Covid-19 pandemics such as frustration, boredom, fear and uncontrollable anxiety. There are many opinions on Twitter about the impact of the Covid-19 pandemic, especially during major holidays and events. Sentiment analysis is needed to see how Indonesian people react to the Covid-19 pandemic. This study aims to determine the classification and analysis of public sentiment patterns regarding the Covid-19 pandemic and to obtain the best algorithms for classification using the K-Nearest Neighbor and Support Vector Machine method. Collecting data is done using Twitter data related to the Covid-19 pandemic with specific keyword. The next step is preprocessing including case *folding*, tokenization, stopword removal, and stemming. Convert the text data to a TF-IDF matrix. It is used for K-Nearest Neighbor and Support Vector Machines classification processes. Labelling the data for comparison using Indonesia Sentiment Lexicon and Crowdsourcing Labeling. The data is divided into three classes: positive, negative and neutral. The classification resulted in a positive label with a total of 2,850 tweet data, a negative label with a total of 8,015 tweet data, and a neutral label with a total of 629 tweet data. Using the K-nearest neighbor classification gives 83.9% accuracy, and the support vector machine gives 76.9% accuracy. The reaction of Indonesian people to the impacts of the Covid-19 pandemic has been dominated by negative reactions. Classification by the K-Nearest Neighbor method provides the highest accuracy compared to Support Vector Machines.

Keywords: *Classification, Impacts of COVID-19 Pandemic, K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), TF-IDF*

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN NASKAH SKRIPSI.....	iii
SURAT PERNYATAAN TENTANG ORISINALITAS	iv
PEDOMAN PENGGUNAAN SKRIPSI	v
KATA PENGANTAR	vi
ABSTRAK.....	ix
ABSTRACT.....	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	4
1.5. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	6
2.1. <i>Microblogging</i>	6
2.2. Covid-19	7
2.3. <i>Data Mining</i>	7
2.4. Analisis Sentimen.....	8
2.5. <i>Preprocessing</i>	9
2.6. <i>Crowdsourced Labeling</i>	10
2.7. <i>Lexicon-Based Labeling</i>	10
2.8. Pembobotan TF-IDF.....	11
2.9. <i>K-Nearest Neighbor</i>	11
2.10. <i>Support Vector Machine</i>	14

2.11. <i>Confussion Matrix</i>	16
2.12. <i>K-Fold Cross Validation</i>	18
2.13. Python.....	19
BAB III METODE PENELITIAN	22
3.1. Pengumpulan Data	23
3.2. <i>Pre-processing</i>	23
3.3. Pelabelan Data.....	25
3.4. Pembobotan TF-IDF.....	26
3.5. Analisis Data	27
3.6. Evaluasi Model.....	29
BAB IV HASIL DAN PEMBAHASAN	30
4.1. Pengumpulan Data	30
4.2. <i>Pre-processing</i>	32
4.3. Pelabelan Data.....	35
4.4. Pembobotan TF-IDF.....	40
4.5. Analisis Data	43
4.6. Pembahasan dan Evaluasi Model	54
BAB V KESIMPULAN DAN SARAN.....	75
5.1. Kesimpulan.....	75
5.2. Saran	75
DAFTAR PUSTAKA	77
LAMPIRAN.....	82
Lampiran 1 <i>Source Code</i> untuk Pengumpulan Data Twitter	82
Lampiran 2 <i>Source code</i> untuk Pre-Processing Data	84
Lampiran 3 Hasil akurasi dari setiap percobaan klasifikasi metode K-NN	87

DAFTAR GAMBAR

Gambar 2.1 Impementasi K-NN.....	12
Gambar 2.2 <i>Support vector Machine</i>	16
Gambar 2.3 5-Fold Cross Validation.....	19
Gambar 3.1 Tahapan Penelitian.....	22
Gambar 3.2 Alur Pelabelan Data.....	26
Gambar 3.3 Alur klasifikasi dengan metode K-NN	28
Gambar 3.4. Alur klasifikasi dengan metode SVM.....	29
Gambar 4.1 Plot Perbandingan Pelabelan Data antara Crowdsourced dengan Inset Lexicon.....	38
Gambar 4.2 Grafik Sentimen Masyarakat Pada Tiap Periode.....	39
Gambar 4.3 Plot Jumlah Sentimen pada Dampak Pandemi Covid-19 pada Tiap Periode	40
Gambar 4.4 Matrik Hasil Perhitungan <i>Term Frequency</i> (TF).....	41
Gambar 4.5 Matrik Hasil Perhitungan <i>Inverse Document Frequency</i> (IDF)	42
Gambar 4.6 Matrik Hasil Perhitungan TF-IDF	43
Gambar 4.7 Plot Prediksi Periode Pilkada.....	55
Gambar 4.8 Hasil Pengukuran Metode K-NN Dataset Periode Pilkada	57
Gambar 4.9 Plot Prediksi Dataset Periode Natal dan Tahun Baru	58
Gambar 4.10 Hasil klasifikasi Metode K-NN Dataset Periode Natal dan Tahun Baru	59
Gambar 4.11 Plot Prediksi Dataset Periode Vaksinasi Pertama.....	60
Gambar 4.12 Hasil Klasifikasi Metode K-NN Dataset Periode Vaksinasi Pertama	61
Gambar 4.13 Plot Prediksi Dataset Periode Lebaran	62
Gambar 4.14 Hasil Klasifikasi Metode K-NN Dataset Periode Lebaran	63
Gambar 4.15 Plot Prediksi Periode Pilkada.....	64
Gambar 4.16 Hasil Klasifikasi Metode SVM Dataset Periode Pilkada	65
Gambar 4.17 Plot Prediksi Dataset Periode Natal dan Tahun Baru	66

Gambar 4.18 Hasil Klasifikasi Metode SVM Dataset Periode Natal dan Tahun Baru	67
Gambar 4.19 Plot Prediksi Dataset Periode Vaksinasi Pertama.....	68
Gambar 4.20 Hasil Klasifikasi Metode SVM Dataset Periode Vaksinasi Pertama	69
Gambar 4.21 Plot Prediksi Dataset Periode Lebaran	70
Gambar 4.22 Hasil Klasifikasi Metode SVM Dataset Periode Lebaran	71

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	17
Tabel 3.1 Contoh Case Folding	23
Tabel 3.2 Contoh Tokenization	24
Tabel 3.3 Contoh Stopwords Removal.....	24
Tabel 3.4 Contoh Stemming	25
Tabel 4.1 Contoh Atribut pada Scrapping Data Twitter.....	31
Tabel 4.2 Contoh Data Hasil Seleksi Atribut	31
Tabel 4.3 Contoh Hasil Case Folding	32
Tabel 4.4 Contoh Hasil Tokenizing.....	33
Tabel 4.5 Contoh Hasil Stopwords Removal	34
Tabel 4.6 Contoh Hasil Stemming	35
Tabel 4.7 Hasil Pelabelan dari Crowdsouced Labelling	36
Tabel 4.8 Jumlah Hasil Pelabelan dari Crowdsourced Labelling.....	36
Tabel 4.9 Hasil Pelabelan InSet Lexicon	37
Tabel 4.10 Jumlah Hasil Pelabelan InSet Lexicon.....	38
Tabel 4.11 Jumlah Hasil Pelabelan Total dengan InSet Lexicon.....	39
Tabel 4.12 Hasil Perhitungan term frequency	43
Tabel 4.13 Parameter K-NN.....	44
Tabel 4.14 Perhitungan 10-Fold Periode Pilkada.....	46
Tabel 4.15 Perhitungan 10-Fold Periode Natal dan Tahun Baru	47
Tabel 4.16 Perbandingan 10-Fold Periode Vaksinasi Pertama	48
Tabel 4.17 Perbandingan 10-Fold Periode Lebaran.....	49
Tabel 4.18 10-Fold Terbaik dari Masing-Masing Periode (K-NN)	50
Tabel 4.19 Parameter Support Vector Machine	51
Tabel 4.20 Perbandingan 10- <i>Fold</i> Periode Pilkada.....	52
Tabel 4.21 Perbandingan 10- <i>Fold</i> Periode Natal dan Tahun Baru	52
Tabel 4.22 Perbandingan 10- <i>Fold</i> Periode Vaksinasi Pertama	52
Tabel 4.23 Perbandingan 10- <i>Fold</i> Periode Lebaran.....	52
Tabel 4.24 10- <i>Fold</i> Terbaik dari Masing-Masing Periode (SVM).....	53

Tabel 4.25 Hasil Perhitungan <i>Confusion Matrix</i> Metode K-NN Dataset Periode Pilkada.....	57
Tabel 4.26 Hasil Perhitungan <i>Confusion Matrix</i> Metode K-NN Periode Natal dan Tahun Baru.....	59
Tabel 4.27 Hasil Perhitungan <i>Confusion Matrix</i> Metode K-NN Periode Vaksinasi Pertama.....	61
Tabel 4.28 Hasil Perhitungan <i>Confusion Matrix</i> Metode K-NN Dataset Periode Lebaran.....	63
Tabel 4.29 Hasil Perhitungan <i>Confusion Matrix</i> Metode SVM Dataset Periode Pilkada.....	65
Tabel 4.30 Hasil Perhitungan <i>Confusion Matrix</i> Metode SVM Dataset Periode Natal dan Tahun Baru	67
Tabel 4.31 Hasil Perhitungan <i>Confusion Matrix</i> Metode SVM Dataset Periode Vaksinasi Pertama	69
Tabel 4.32 Hasil Perhitungan <i>Confusion Matrix</i> Metode SVM Dataset Periode Lebaran.....	71
Tabel 4.33 Perbandingan Akurasi dari K-NN dan SVM tiap Periode	71
Tabel 1 10-Fold Periode Pilkada	87
Tabel 2 10-Fold Periode Natal dan Tahun Baru.....	89
Tabel 3 10-Fold Periode Vaksinasi Pertama	91
Tabel 4 10-Fold Periode Lebaran.....	93