

**PENERAPAN ALGORITMA KNN PADA ANALISIS SENTIMEN
DAMPAK EKONOMI AKIBAT COVID DI MEDIA SOSIAL TWITTER**

SKRIPSI



disusun oleh

Laila Nur Aini

18.11.1980

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENERAPAN ALGORITMA KNN PADA ANALISIS SENTIMEN
DAMPAK EKONOMI AKIBAT COVID DI MEDIA SOSIAL TWITTER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Laila Nur Aini

18.11.1980

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA KNN PADA ANALISIS SENTIMEN DAMPAK EKONOMI AKIBAT COVID DI MEDIA SOSIAL TWITTER

yang dipersiapkan dan disusun oleh

Laila Nur Aini

18.11.1980

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 06 April 2022

Dosen Pembimbing,

Anna Baita, M. Kom

NIK. 190302290

PENGESAHAN

SKRIPSI

PENERAPAN ALGORITMA KNN PADA ANALISIS SENTIMEN DAMPAK EKONOMI AKIBAT COVID DI MEDIA SOSIAL TWITTER

yang dipersiapkan dan disusun oleh

Laila Nur Aini

18.11.1980

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 April 2022

Susunan Dewan Penguji

Nama Penguji

Yuli Astuti, M.Kom
NIK. 190302146

Hartatik, S.T., M.Cs.
NIK. 190302232

Anna Baita, M.Kom
NIK. 190302290

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 April 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 17 Mei 2022



Laila Nur Aini

NIM. 18.11.1980

MOTTO

"Musuh yang paling berbahaya di atas dunia ini adalah penakut dan bimbang.

Teman yang paling setia, hanyalah keberanian dan keyakinan yang teguh."

(Andrew Jackson)

"Bersikaplah kukuh seperti batu karang yang tidak putus-putus-nya dipukul ombak. Ia tidak saja tetap berdiri kukuh, bahkan ia menenteramkan amarah ombak dan gelombang itu."

(Marcus Aurelius)

"Kamu tidak harus menjadi hebat untuk memulai, tetapi kamu harus mulai untuk menjadi hebat."

(Zig Ziglar)

PERSEMBAHAN

Puji dan Syukur kepada Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah dan karunia-Nya, atas kemudahan, petunjuk dan kekuatan yang diberikan-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Penerapan Algoritma KNN pada Analisis Sentimen Dampak Ekonomi Akibat Covid di Media Sosial Twitter”, dan diajukan sebagai salah satu syarat untuk mendapat gelar Sarjana pada jurusan Informatika, Universitas Amikom Yogyakarta. Karya tulis ini penulis persembahkan kepada :

1. Pertama saya ucapkan terimakasih kepada bapak saya Mardi, yang sudah menyiapkan segala kebutuhan financial saya sampai detik ini dan telah mempersiapkan semuanya dari jauh jauh hari. Beserta doa dan dukungan yang selalu beliau berikan kepada saya.
2. Kedua saya ucapkan terimakasih kepada ibu saya Nur Istatik yang tanpa henti mengingatkan saya untuk tetap semangat dalam menjalani semua ini. Dukungan beliau sangat penting untuk kelancaran *study* saya.
3. Terimakasih kepada adik saya Muhammad Faisal Habibi yang selalu memberikan semangat untuk mengerjakan skripsi ini.
4. Serta dukungan teman-teman yang selalu mengiringi langkah saya.

KATA PENGANTAR

Puji dan Syukur kepada Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah dan karunia-Nya, atas kemudahan, petunjuk dan kekuatan yang diberikan-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Penerapan Algoritma KNN pada Analisis Sentimen Dampak Ekonomi Akibat Covid di Media Sosial Twitter”, dan diajukan sebagai salah satu syarat untuk mendapat gelar Sarjana pada jurusan Informatika, Universitas Amikom Yogyakarta. Dalam penyusunannya, penulis memperoleh banyak bantuan dari berbagai pihak, oleh karena itu penulis mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan fakultas ilmu komputer Universitas Amikom Yogyakarta.
3. Ibu Anna Baita, M.Kom sebagai dosen pembimbing yang telah memberikan kontribusi berupa bimbingan, motivasi, kritik dan saran atas terselesaikannya skripsi ini.
4. Bapak Bayu Setiaji, M.Kom sebagai dosen wali yang telah memberikan saran atas terselesaikannya skripsi ini.
5. Semua Dosen Fakultas ilmu Komputer Universitas Amikom Yogyakarta, yang tidak bisa penulis sebutkan satu persatu. Terimakasih atas ilmunya, semoga bisa penulis amalkan dikemudian hari.

6. Kedua orang tua beserta keluarga tercinta yang senantiasa mendoakan dan memberi dukungan selalu.
7. Teman teman yang membantu baik secara tidak langsung hingga tugas skripsi ini dapat diselesaikan dengan sebaik-baiknya.
8. Semua pihak yang telah membantu dalam penyusunan tugas skripsi ini yang tidak dapat penulis sampaikan satu persatu.

Tentu saja tidak ada kata selain terima kasih sebesar-besarnya kepada seluruh pihak yang terlibat dalam skripsi ini. Demikian pula dengan skripsi ini. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Maka dari itu, semua kritik dan saran yang bersifat membangun sangat penulis harapkan dan tentu saja penulis terima dengan lapang dada. Dengan begitu, akan menjadi suatu pelajaran berharga untuk penulis di kemudian hari.

Yogyakarta, 12 Mei 2022

Penulis

Laila Nur Aini

NIM 18.11.1980

DAFTAR ISI

JUDUL	ii
LEMBAR PERSETUJUAN.....	iii
LEMBAR PENGESAHAN	iv
PERNYATAAN.....	v
MOTTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR RUMUS	xviii
INTISARI.....	xix
ABSTRACT.....	xx
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan.....	4
1.5 Manfaat Penulisan.....	4
1.6 Metode Penelitian.....	5
1.6.1 Metode Pengumpulan Data.....	5

1.6.2	Metode Analisis	6
1.6.3	Metode Implementasi.....	6
1.6.4	Metode Pengujian	7
1.7	Sistematika Penulisan	7
BAB II LANDASAN TEORI.....		9
2.1	Tinjauan Pustaka	9
2.2	Dasar Teori.....	13
2.2.1	Dampak Ekonomi Akibat Covid.....	13
2.2.2	Twitter.....	15
2.2.3	Text Mining.....	17
2.2.4	Analisis Sentimen	17
2.2.5	Flowchart	18
2.2.6	Python	18
2.2.7	Library Python	18
2.2.8	Google Colaboratory.....	19
2.2.9	Preprocessing	20
2.2.10	TF-IDF	21
2.2.11	SMOTE	21
2.2.12	K-Fold Cross Validation	22
2.2.13	K-Nearest Neighbor (KNN).....	23
2.2.14	Confusion Matrix	25
BAB III ANALISIS DAN PERANCANGAN		28
3.1	Analisis Kebutuhan	28

3.1.1 Analisis Kebutuhan Fungsional	28
3.1.2 Analisis Kebutuhan Non Fungsional	28
3.2 Arsitektur Umum	29
3.3 Perancangan Sistem	30
3.3.1 Pengumpulan Data (Crawling Data).....	30
3.3.2 Labelisasi.....	31
3.3.3 Preprocessing	32
3.3.3.1 Cleaning	33
3.3.3.2 Case Folding	34
3.3.3.3 Tokenizing.....	35
3.3.3.4 Stopword Removal.....	36
3.3.3.5 Stemming	36
3.3.4 TF-IDF	37
3.3.5 Splitting Data	43
3.3.6 SMOTE	44
3.3.7 K-Fold Cross Validation	44
3.3.8 Implementasi KNN	45
3.3.9 Evaluasi	49
BAB IV IMPLEMENTASI DAN PEMBAHASAN	50
4.1 Deskripsi Implementasi.....	50
4.2 Implementasi Connecting Google Colab dengan Drive.....	51
4.3 Implementasi Proses Crawling Data Twitter	52
4.4 Implementasi Proses Labelling	55

4.5 Implementasi Proses Preprocessing Data.....	56
4.5.1 Cleaning	56
4.5.2 Case Folding	57
4.5.3 Tokenizing.....	58
4.5.4 Stopword Removal.....	58
4.5.5 Stemming	59
4.6 Visualisasi Modeling kata Setelah dilakukan Preprocessing	59
4.6.1 Visualisasi kata yang sering muncul dalam analisis sentimen....	60
4.6.2 Visualisasi Jumlah Sentimen Positif, Netral, dan Negatif	61
4.6.3 Visualisasi Sentimen Positif yang sering muncul	62
4.6.4 Visualisasi Sentimen Netral yang sering muncul	62
4.6.5 Visualisasi Sentimen Negatif yang sering muncul.....	63
4.7 Implementasi Proses Pembobotan kata	64
4.8 Implementasi Splitting Data.....	64
4.9 Implementasi SMOTE	67
4.10 K-Fold Cross Validation	68
4.11 Implementasi Algoritma K-Nearest Neighbor	69
4.12 Evaluasi	72
BAB V PENUTUP.....	77
5.1 Kesimpulan	77
5.2 Saran.....	77
DAFTAR PUSTAKA	79
LAMPIRAN	81

DAFTAR TABEL

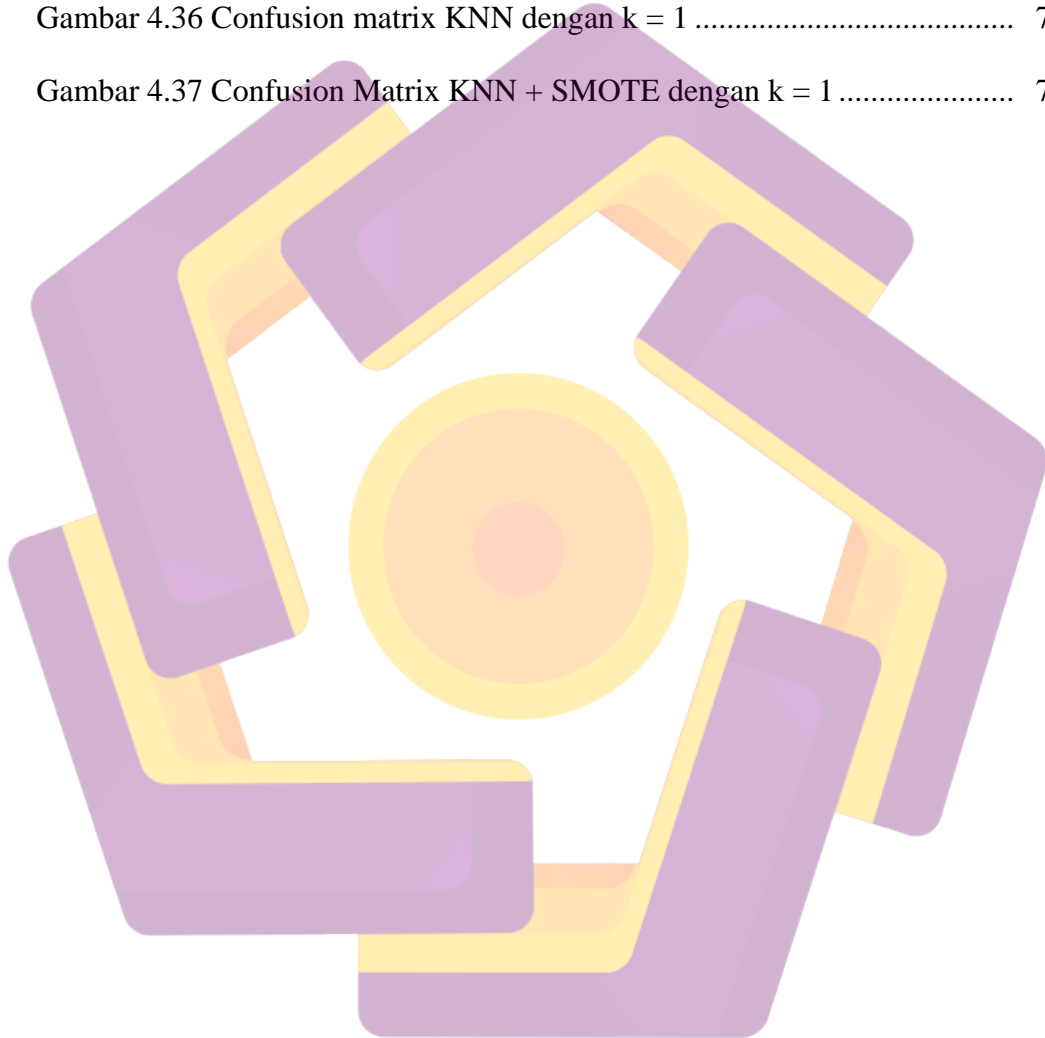
Tabel 2.1 Perbandingan Penelitian Sebelumnya.....	11
Tabel 3.1 Contoh labeling pada teks tweet	32
Tabel 3.2 Proses Cleaning Data	33
Tabel 3.3 Proses Case Folding.....	34
Tabel 3.4 Proses Tokenization	35
Tabel 3.5 Proses Stopword Removal	36
Tabel 3.6 Proses Stemming.....	37
Tabel 3.7 Contoh data untuk proses TF-IDF	38
Table 3.8 Kumpulan Kata	39
Tabel 3.9 Perhitungan TF dan DF.....	39
Tabel 3.10 Perhitungan IDF.....	41
Tabel 3.11 Perhitungan TF-IDF.....	42
Tabel 3.12 Perhitungan KNN.....	46
Tabel 3.13 Perhitungan Euclidean Distance	46
Tabel 3.14 mengurutkan jarak K.....	47
Tabel 3.15 Kategori K.....	48
Tabel 3.16 Hasil perhitungan KNN	48
Tabel 4.1 Perbandingan hasil Splitting Data dengan Algoritma KNN	65
Tabel 4.2 Perbandingan hasil Splitting Data dengan Algoritma KNN+SMOTE	65
Tabel 4.3 Hasil Pengujian KNN	72
Tabel 4.4 Hasil Pengujian KNN + SMOTE.....	72

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja Cross Validation	23
Gambar 2.2 Metode KNN untuk menentukan tetangga terdekat.....	25
Gambar 2.3 Model Confusion matrix	26
Gambar 3.1 Arsitektur Umum	29
Gambar 3.2 Proses Crawling Data	31
Gambar 3.3 Proses Preprocessing Data	33
Gambar 3.4 Flowchart TF-IDF	38
Gambar 3.5 Cara kerja K-Fold Cross Validation dengan 10-fold	45
Gambar 3.6 Tahapan Proses KNN	45
Gambar 4.1 Code Connecting Drive.....	51
Gambar 4.2 Code lokasi penyimpanan file pada drive	51
Gambar 4.3 Instalasi Library Python	52
Gambar 4.4 Autentikasi API Twitter	53
Gambar 4.5 Code field data yang akan diambil	54
Gambar 4.6 Code Pengambilan Data	55
Gambar 4.7 Hasil Crawling Sebelum dilakukan Labelling	55
Gambar 4.8 Output Setelah dilakukan Labelling.....	56
Gambar 4.9 Code Cleaning Data	57
Gambar 4.10 Code Case Folding	57
Gambar 4.11 Code Tokenizing	58
Gambar 4.12 Code Stopword Removal	58

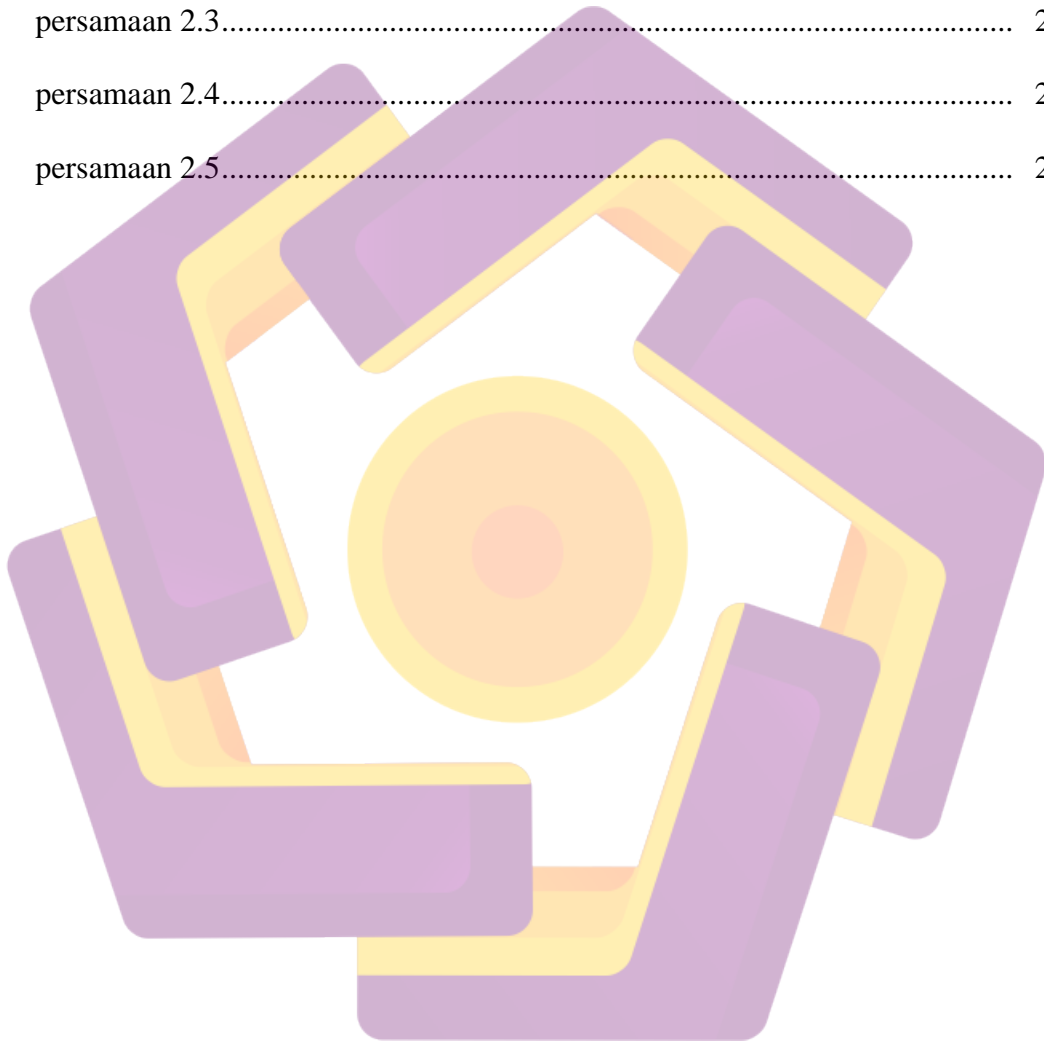
Gambar 4.13 Code Stemming.....	59
Gambar 4.14 Code Modelling kata yang sering muncul	60
Gambar 4.15 Output Modelling Kata yang sering Muncul.....	60
Gambar 4.16 Code Jumlah Data Sentimen	61
Gambar 4.17 Output Jumlah Data Sentimen	61
Gambar 4.18 Code Sentimen Positif.....	62
Gambar 4.19 Output Sentimen Positif	62
Gambar 4.20 Code Sentimen Netral	62
Gambar 4.21 Output Sentimen Netral.....	63
Gambar 4.22 Code Sentimen Negatif	63
Gambar 4.23 Output Sentimen Negatif.....	64
Gambar 4.24 Code Pembobotan Kata / TF-IDF	64
Gambar 4.25 Grafik Perbandingan Persentase.....	66
Gambar 4.26 Code Splitting Data	67
Gambar 4.27 Code SMOTE Data	67
Gambar 4.28 Hasil SMOTE Data	68
Gambar 4.29 (a) 5-fold Validation; (b) 10-fold Validation	68
Gambar 4.30 (a) Hasil rata-rata 5 -Fold Validation; (b) Hasil rata-rata 10-Fold Validation.....	69
Gambar 4.31 (a) Code mencari nilai k pada KNN; (b) Code mencari nilai k pada KNN+SMOTE.....	69
Gambar 4.32 (a) Output nilai k pada KNN; (b) Output nilai k pada KNN+SMOTE.....	70

Gambar 4.33 (a) Code KNN; (b) Code KNN+SMOTE.....	70
Gambar 4.34 (a) Hasil perhitungan akurasi KNN;	
(b) hasil perhitungan KNN+SMOTE.....	71
Gambar 4.35 Grafik perbandingan nilai KNN dan KNN+SMOTE.....	73
Gambar 4.36 Confusion matrix KNN dengan $k = 1$	74
Gambar 4.37 Confusion Matrix KNN + SMOTE dengan $k = 1$	75



DAFTAR RUMUS

Persamaan 2.1	21
Persamaan 2.2	22
persamaan 2.3.....	24
persamaan 2.4.....	27
persamaan 2.5.....	27



INTISARI

Pada akhir tahun 2019 Dunia digemparkan adanya virus baru yang mematikan, yaitu virus corona atau covid-19 (*CoronaVirus Disease*). Virus ini ditemukan pertama kali di Wuhan, China. Virus ini dapat menyebar dengan cepat, bahkan sudah menyebar hampir ke seluruh negara di Dunia, termasuk Indonesia. Pertama kali Indonesia melaporkan adanya kasus Covid-19 pada 2 Maret 2020. Covid-19 mengakibatkan berbagai dampak pada beberapa sektor, baik dampak Positif maupun dampak Negatif.

Sektor yang paling terpengaruh yaitu sektor ekonomi karena sektor ekonomi merupakan sektor yang paling penting dalam kehidupan. Pada sektor ekonomi mengalami banyak kerugian yang menimbulkan kenaikan harga barang pokok dan diikuti dengan kenaikan harga gas LPG, BBM, dll. Selain itu juga cukup baik mengalami kenaikan dalam memperkenalkan produk umkm dalam negeri kepada masyarakat. Hal ini menimbulkan berbagai sentimen dari masyarakat terkait covid-19 di Indonesia melalui media sosial Twitter. Twitter merupakan salah satu media sosial yang mudah penggunaannya untuk menyebarkan informasi secara *real time*. Banyaknya sentimen masyarakat mengenai dampak covid mendorong penulis untuk melakukan penelitian analisis sentimen dengan algoritma KNN (*K-Nearest Neighbor*) dalam melakukan klasifikasi data twitter.

Algoritma KNN digunakan untuk mengklasifikasikan data twitter menjadi tiga kategori, yaitu sentimen positif, sentimen netral, dan sentimen negatif. Berdasarkan hasil implementasi algoritma knn memiliki akurasi sebesar 74%. Kemudian penulis melakukan upaya untuk menaikkan tingkat akurasi dengan menggunakan metode SMOTE dan menghasilkan nilai akurasi sebesar 75,42%. Dengan jumlah opini lebih condong ke positif dan topik paling dibicarakan yaitu ekonomi.

Kata Kunci : K-Nearest Neighbor (KNN), Analisis Sentimen, Twitter, Dampak Covid, Klasifikasi.

ABSTRACT

At the end of 2019 the world was shocked by a new deadly virus, namely the corona virus or covid-19 (Coronavirus Disease). This virus was first discovered in Wuhan, China. This virus can spread quickly, it has even spread to almost all countries in the world, including Indonesia. For the first time, Indonesia reported a case of Covid-19 on March 2, 2020. Covid-19 had various impacts on several sectors, both positive and negative impacts.

The sector that is most affected is the economic sector because the economic sector is the most important sector in life. In the economic sector, there are many losses that cause an increase in the price of basic goods and is followed by an increase in the price of LPG, fuel, etc. In addition, it is also quite good to experience an increase in introducing domestic UMKM products to the public. This has raised various sentiments from the public regarding Covid-19 in Indonesia through Twitter social media. Twitter is one of the social media that is easy to use to disseminate information in real time. The number of public sentiments regarding the impact of covid has prompted the author to conduct sentiment analysis research with the KNN (K-Nearest Neighbor) in classifying twitter data.

The KNN algorithm is used to classify Twitter data into three categories, namely positive sentiment, neutral sentiment, and negative sentiment. Based on the results of the implementation of the KNN algorithm, it has an accuracy of 74%. Then the author made an effort to increase the level of accuracy by using the SMOTE method and resulted in an accuracy value of 75.42%. With the number of opinions more inclined to positive and the most discussed topic is the economy.

Keywords: *K-Nearest Neighbor (KNN), Sentiment Analysis, Twitter, Impact of Covid, Classification.*