

**ANALISIS TENDENSI PORTAL BERITA TENTANG
VAKSINASI COVID-19 DI INDONESIA
MENGUNAKAN METODE
K-NEAREST NEIGHBOR**

SKRIPSI

**Sebagai Syarat Dalam Menyelesaikan Program Studi Strata Satu (S-1)
Program Studi Jurusan Teknik Informatika**



Disusun Oleh :

**NAMA : AMBAR DWI SAPUTRA
NPM : 2017470080**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JAKARTA
2021**

UNIVERSITAS MUHAMMADIYAH JAKARTA

FAKULTAS TEKNIK – JURUSAN TEKNIK INFORMATIKA

LEMBAR PERSETUJUAN

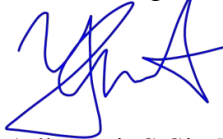
**ANALISIS TENDENSI PORTAL BERITA TENTANG
VAKSINASI COVID-19 DI INDONESIA
MENGUNAKAN METODE
*K-NEAREST NEIGHBOR***

NAMA : AMBAR DWI SAPUTRA
NPM : 2017470080
JURUSAN : TEKNIK INFORMATIKA

Skripsi ini telah disetujui pada tanggal,

Oleh

Pembimbing Utama



(Yana Adharani, S.Si., M.KOM)

Mengetahui

Ketua Prodi Teknik Informatika

(Popy Meilina, S.T., M.KOM)

UNIVERSITAS MUHAMMADIYAH JAKARTA

FAKULTAS TEKNIK – JURUSAN TEKNIK INFORMATIKA

TANDA BUKTI PERSETUJUAN PEMBIMBING SKRIPSI

Pada Semester Genap

Tahun Akademik 2020/2021

Yang bertanda tangah dibawah ini Pembimbing Skripsi menyetujui bahwa :

NAMA : AMBAR DWI SAPUTRA
NPM : 2017470080

Judul TA

ANALISIS TENDENSI PORTAL BERITA TENTANG VAKSINASI COVID-19
DI INDONESIA MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

Dimulai bulan, tahun : Maret 2021

Selesai bulan, tahun : Juli 2021

Untuk ikut serta Ujiuan Sidang Strata Satu (S1) yang diselenggarakan oleh
Program Studi Teknik Informatika Universitas Muhammadiyah Jakarta.

Jakarta, 4 Agustus 2021

Pembimbing Utama



(Yana Adharani, S.Si., M.KOM)

UNIVERSITAS MUHAMMADIYAH JAKARTA
FAKULTAS TEKNIK – JURUSAN TEKNIK INFORMATIKA


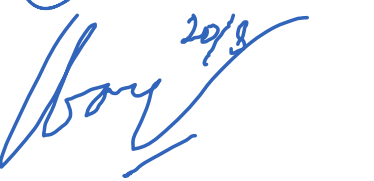

LEMBAR PENGESAHAN

**ANALISIS TENDENSI PORTAL BERITA TENTANG
VAKSINASI COVID-19 DI INDONESIA
MENGUNAKAN METODE
*K-NEAREST NEIGHBOR***

NAMA : AMBAR DWI SAPUTRA
NPM : 2017470080
JURUSAN : TEKNIK INFORMATIKA

Skripsi ini telah diuji pada tanggal, 4 Agustus 2021

Oleh Penguji :

1. Popy Meilina, S.T., M.KOM :  B/S
2. Sitti Nurbaya Ambo, M.M.S.I :  20/8
3. Yana Adharani, S.SI., M.KOM : 

UNIVERSITAS MUHAMMADIYAH JAKARTA

FAKULTAS TEKNIK – JURUSAN TEKNIK INFORMATIKA

LEMBAR PERNYATAAN

Bersama ini Saya Menyatakan Bahwa isi yang terkandung dalam Skripsi ini dengan judul :

**ANALISIS TENDENSI PORTAL BERITA TENTANG
VAKSINASI COVID-19 DI INDONESIA
MENGUNAKAN METODE
*K-NEAREST NEIGHBOR***

Adalah murni merupakan hasil penelitian dan pemikiran saya sendiri.

Demikian pernyataan ini saya buat dan siap menerima konsekuensi apapun dimasa yang akan datang, apabila ternyata Skripsi ini merupakan salinan ataupun contoh karya-karya yang telah dibuat/terbitkan sebelum tanggal Skripsi ini.

Jakarta, 28 Juli 2021



(Ambar Dwi Saputra)

UNIVERSITAS MUHAMMADIYAH JAKARTA

FAKULTAS TEKNIK – JURUSAN TEKNIK INFORMATIKA

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini, saya :

Nama : AMBAR DWI SAPUTRA

NIM : 2017470080

Program Studi : Teknik Informatika

Jenjang : Strata Satu (S1) Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Studi Teknik Informatika **FT-UMJ Hak Bebas Royalti NonEksklusif (Non-exclusive Royalti-Free Right)** atas karya ilmiah saya yang berjudul : “ANALISIS TENDENSI PORTAL BERITA TENTANG VAKSINASI COVID-19 DI INDONESIA MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak FT-UMJ berhak menyimpan, mengalih-media atau bentuk-kan, mengelolanya dalam pangkalan data (database), mendistribusikannya dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari

Saya selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut. Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak FT- UMJ, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 21 Juli 2021



Yang Menyatakan
(Ambar Dwi Saputra)





UNIVERSITAS MUHAMMADIYAH JAKARTA FAKULTAS
TEKNIK-PRODI TEKNIK INFORMATIKA






DAFTAR PRESENSI BIMBINGAN SKRIPSI

**ANALISIS TENDENSI PORTAL BERITA TENTANG
VAKSINASI COVID-19 DI INDONESIA MENGGUNAKAN
METODE *K-NEAREST NEIGHBOR***

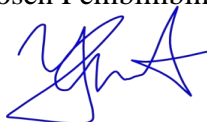
Nama : Ambar Dwi Saputra
NPM : 2017470080
Program Studi : Teknik Informatika

Dosen Pembimbing : Yana Adharani, S.SI., M.KOM

No	Tanggal	Catatan Dosen Pembimbing	Paraf
1	18 Maret 2021	Pengajuan BAB I : Latar Belakang, Judul, Identifikasi Masalah, Rumusan Masalah, Manfaat & Tujuan, Batasan Masalah, Metodologi Penelitian, Sistematika Penulisan	
2	27 Maret 2021	Revisi BAB I : Perbaikan sitasi pada latar belakang masalah, Perbaikan Identifikasi Masalah, Perbaikan Rumusan Masalah, Menambahkan Batasan Masalah tentang portal berita apa saja yang digunakan Pengajuan BAB II : Landasan Teori tentang Vaksin Covid19, Text Processing, Text Mining	
3	14 April 2021	Revisi BAB II : Penambahan latar belakang covid 19, perbaikan kalimat yang rancu atau ambigu	
4	26 April 2021	Pengajuan BAB III : Tahapan Metode Penelitian, Tahap Text Preprocessing, Pembobotan TF-IDF	

No	Tanggal	Catatan Dosen Pembimbing	Paraf
5	15 Juni 2021	Revisi Hasil Seminar Proposal : Perbaikan Identifikasi Masalah, Rumusan Masalah, penjelasan metode pengumpulan data, penjelasan pembagian data <i>training</i> dan data <i>testing</i> yang digunakan, penambahan evaluasi model yang digunakan	
6	12 Juli 2021	Revisi BAB III : Diagram metode penelitian. Pengajuan BAB IV : Hasil Pengujian	
7	24 Juli 2021	Revisi BAB III : Pengajuan desain tampilan pada menu dashboard, data master, text preprocessing, hasil klasifikasi dan visualisasi. Revisi BAB IV : Pengajuan hasil dari tampilan pada menu dashboard, data master, text preprocessing, hasil klasifikasi dan visualisasi. Pengajuan hasil pengujian dan pembahasan hasil. Demo Aplikasi	
8	27 Juli 2021	Pengajuan BAB V : Kesimpulan dan Saran Revisi Abstrak Revisi BAB I : Perbaikan Latar Belakang, Identifikasi Masalah, Rumusan Masalah Revisi BAB IV : Pembahasan hasil ujicoba, penjelasan <i>word cloud</i> .	
9	28 Juli 2021	ACC Keseluruhan BAB	

Dosen Pembimbing



(Yana Adharani, S.SI., M.KOM)

ABSTRACT

COVID-19 is one of the diseases that caused a pandemic in 223 countries in the world with the number of victims until July 27, 2021 as many as 194,608,040 confirmed positive and 4,170,155 victims died. In Indonesia alone, as of July 27, 2021, there were 4,170,155 confirmed positive cases with 86,835 deaths. One way to suppress the increase and spread of COVID-19 cases is by administering vaccines or vaccinations. Related to this, there are various news about the COVID-19 vaccine on online news portals. The news provided has positive, negative or neutral tendencies that can affect the public's view of the COVID-19 vaccine. Based on this, in this study, a classification was carried out on the tendency of news portals to report on the COVID-19 vaccine using the K-Nearest Neighbor (KNN) method. The trial was carried out on 1000 data with 5 combinations of training and testing data 50 times for different k values. Data labeling is done manually and automatically using sentiment_id. The test results on data with manual labeling showed the highest accuracy level of 78.50% at k = 6. For data with automatic labeling, the highest accuracy was obtained at 93% at k = 6. The highest accuracy level was obtained for the use of training data by 80% and testing data 20%. The level of accuracy that is not optimal is caused by the amount of data in each class is not balanced and there are labels on the training and testing data that are not accurate. The application can identify dominant words, both in positive, negative and neutral news for the identified class.

Keywords: classification , COVID-19, K-Nearest Neighbor, news portal, vaccines, vaccinations

ABSTRAK

COVID-19 salah satu yang penyakit yang mengakibatkan pandemi di 223 negara di dunia dengan jumlah korban sampai dengan 27 Juli 2021 sebanyak 194.608.040 terkonfirmasi positif dan 4.170.155 korban meninggal. Di Indonesia sendiri per tanggal 27 Juli 2021 kasus yang terkonfirmasi positif sebanyak 4.170.155 dengan korban meninggal 86.835. Salah satu cara menekan pertambahan dan penyebaran kasus COVID-19 yaitu dengan melakukan pemberian vaksin atau vaksinasi. Terkait hal tersebut terdapat berbagai pemberitaan mengenai vaksin COVID-19 di portal berita online. Pemberitaan yang diberikan ada yang memiliki tendensi positif, negatif maupun netral yang dapat memengaruhi pandangan masyarakat terhadap pemberian vaksin COVID-19. Berdasarkan hal tersebut dalam penelitian ini dilakukan klasifikasi terhadap tendensi pemberitaan portal berita terhadap vaksin COVID-19 dengan menggunakan metode *K-Nearest Neighbor* (KNN). Uji coba dilakukan terhadap 1000 data dengan 5 kombinasi data *training* dan *testing* sebanyak 50 kali untuk dengan nilai k yang berbeda. Pelabelan data dilakukan secara manual dan otomatis menggunakan `sentistrentn_id`. Hasil uji coba pada data dengan pelabelan manual menunjukkan tingkat *accuracy* tertinggi sebesar 78,50% pada $k = 6$. Untuk data dengan pelabelan otomatis diperoleh *accuracy* tertinggi sebesar 93% pada $k = 6$. Tingkat *accuracy* tertinggi diperoleh untuk penggunaan data *training* sebesar 80% dan data *testing* 20%. Tingkat *accuracy* yang belum optimal diakibatkan karena jumlah data pada setiap kelas tidak berimbang dan terdapat pelabelan pada data *training* maupun *testing* yang tidak akurat. Aplikasi dapat mengidentifikasi kata-kata yang dominan, baik dalam pemberitaan positif, negatif maupun netral untuk kelas yang teridentifikasi.

Kata Kunci: klasifikasi, COVID-19, *K-Nearest Neighbor*, portal berita, vaksin, vaksinasi

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas Rahmat dan Hidayah-Nya, penulis dapat menyelesaikan Laporan Skripsi dengan judul “Analisis Tendensi Portal Berita Tentang Vaksinasi Covid-19 Di Indonesia Menggunakan Metode K-Nearest Neighbor”. Skripsi ini dibuat sebagai salah satu syarat kelulusan akademis Program Studi Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jakarta. Berkat doa, dorongan dan arahan dari berbagai pihak, sehingga semua kesulitan dan hambatan yang penulis hadapi dapat terselesaikan dengan baik.

Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian Skripsi ini, yaitu:

1. Bapak Irfan Purnawan, S.T., M. Chem.Eng., selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jakarta.
2. Ibu Popy Meilina S.T., M.Kom., selaku Ketua Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jakarta
3. Ibu Yana Adharani, S.SI., M.KOM, selaku Dosen Pembimbing yang meluangkan waktunya untuk membimbing penulis.
4. Ibu Retnani Latifah, M.KOM selaku Dosen Teknik Informatika yang telah memberikan masukan dan arahan kepada penulis.
5. Ibu, Bapak, keluarga dan teman-teman yang telah memberikan semangat, dukungan serta do'a kepada penulis.
6. Seluruh staf dan pengajar Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jakarta.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulisan dan penyusunan Skripsi ini.

Penulis tentunya masih ada kekurangan dalam penyusunan Skripsi ini. Untuk itu, penulis mengharapkan saran dan kritik yang membangun demi kesempurnaan penelitian selanjutnya.

Akhir kata semoga Skripsi ini dapat bermanfaat bagi pembaca dan terkhusus bagi penulis untuk pengetahuan dan penelitian selanjutnya.

Bekasi, 28 Juli 2021

A handwritten signature in black ink, appearing to be 'Ambar Dwi Saputra'.

Ambar Dwi Saputra

DAFTAR ISI

HALAMAN COVER.....	i
LEMBAR PERSETUJUAN.....	ii
TANDA BUKTI PERSETUJUAN PEMBIMBING SKRIPSI	iii
LEMBAR PENGESAHAN	iv
LEMBAR PERNYATAAN	v
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
DAFTAR PRESENSI BIMBINGAN SKRIPSI	viii
ABSTRACT.....	x
ABSTRAK	xi
KATA PENGANTAR	xii
DAFTAR ISI.....	xiv
BAB I PENDAHULUAN	22
1.1 Latar Belakang Masalah	22
1.2 Identifikasi Masalah	25
1.3 Rumusan Masalah	25
1.4 Batasan Masalah.....	26
1.5 Tujuan Penelitian.....	27
1.6 Manfaat Penelitian.....	27
1.7 Metode Penelitian.....	27
1.8 Sistematika Penulisan.....	28
BAB II LANDASAN TEORI	30
2.1 Vaksin COVID-19.....	30
2.2 <i>Text Mining</i>	32

2.3	<i>Text Preprocessing</i>	32
2.4	Pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	35
2.5	Sentimen Analisis	36
2.6	Klasifikasi <i>K-Nearest Neighbor</i> (KNN)	37
2.7	<i>Confusion Matrix</i>	39
BAB III METODOLOGI PENELITIAN		43
3.1	Metode Penelitian	43
3.2	Pengumpulan Data	46
3.3	<i>Preprocessing</i>	47
3.3.1	<i>Case Folding</i>	48
3.3.2	<i>Tokenizing</i>	49
3.3.3	<i>Filtering</i>	51
3.3.4	<i>Stemming</i>	53
3.4	Pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	55
3.5	Implementasi <i>K-Nearest Neighbor</i> (KNN)	58
3.6	Tahap Pengujian	65
3.7	Desain Tampilan	67
3.7.1	Desain Tampilan Dashboard	67
3.7.2	Desain Tampilan Data Master	68
3.7.3	Desain Tampilan <i>Text Preprocessing</i>	70
3.7.4	Desain Tampilan Hasil Klasifikasi	74
3.7.5	Desain Tampilan Visualisasi	75
BAB IV HASIL DAN PEMBAHASAN		76
4.1	Tampilan Hasil	76
4.1.1	Tampilan Hasil <i>Dashboard</i>	76

4.1.2	Tampilan Hasil Data Master	77
4.1.3	Tampilan Hasil <i>Text Preprocessing</i>	79
4.1.4	Tampilan Hasil Klasifikasi.....	83
4.1.5	Tampilan Hasil Visualisasi	84
4.2	Hasil Pengujian.....	85
4.2.1	Pelabelan Manual	85
4.2.2	Pelabelan Otomatis <i>sentistrength_id</i>	100
4.2.3	<i>Word Cloud</i>	113
4.3	Pembahasan Hasil.....	118
BAB V KESIMPULAN DAN SARAN.....		124
5.1	Kesimpulan.....	124
5.2	Saran	125
DAFTAR PUSTAKA		126
LAMPIRAN.....		130

DAFTAR TABEL

Tabel 1.1 Contoh Berita Positif dan Negatif	23
Tabel 1.2 Data Portal Berita yang Digunakan	26
Tabel 3.1 Dokumen Hasil Pengumpulan Data	47
Tabel 3.2 Contoh <i>Case Folding</i>	49
Tabel 3.3 Contoh <i>Tokenizing</i>	50
Tabel 3.4 Contoh <i>Filtering</i>	52
Tabel 3.5 Contoh <i>Stemming</i>	54
Tabel 3.6 Hasil Pembobotan <i>Term Frequency (TF)</i>	56
Tabel 3.7 Hasil Pembobotan <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	57
Tabel 3.8 Hasil Pembobotan <i>Term Frequency (TF)</i> Data Uji	61
Tabel 3.9 Hasil Pembobotan <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> Data Uji	62
Tabel 3.10 Hasil Perhitungan <i>Euclidean Distance</i> Data Uji	64
Tabel 3.11 Pengurutan Dokumen Berdasarkan Jarak <i>Euclidean</i>	65
Tabel 3.12 Tabel Skenario Percobaan	66
Tabel 4.1 Hasil Nilai k Skenario Data <i>Training</i> 70% dan Data <i>Testing</i> 30% Menggunakan Pelabelan Manual	87
Tabel 4.2 Hasil Nilai k Skenario Data <i>Training</i> 60% dan Data <i>Testing</i> 40% Menggunakan Pelabelan Manual	89
Tabel 4.3 Hasil Nilai k Skenario Data <i>Training</i> 65% dan Data <i>Testing</i> 35% Menggunakan Pelabelan Manual	91
Tabel 4.4 Hasil Nilai k Skenario Data <i>Training</i> 75% dan Data <i>Testing</i> 25% Menggunakan Pelabelan Manual	94
Tabel 4.5 Hasil Nilai k Skenario Data <i>Training</i> 80% dan Data <i>Testing</i> 20% Menggunakan Pelabelan Manual	97
Tabel 4.6 Tabel Hasil Skenario Percobaan Menggunakan Pelabelan Manual....	100
Tabel 4.7 Hasil Nilai k Skenario Data <i>Training</i> 70% dan Data <i>Testing</i> 30% Menggunakan Otomatis	102

Tabel 4.8 Hasil Nilai k Skenario Data <i>Training</i> 60% dan Data <i>Testing</i> 40% Menggunakan Pelabelan Otomatis.....	104
Tabel 4.9 Hasil Nilai k Skenario Data <i>Training</i> 65% dan Data <i>Testing</i> 35% Menggunakan Pelabelan Otomatis.....	106
Tabel 4.10 Hasil Nilai k Skenario Data <i>Training</i> 75% dan Data <i>Testing</i> 25% Menggunakan Pelabelan Otomatis.....	108
Tabel 4.11 Hasil Nilai k Skenario Data <i>Training</i> 80% dan Data <i>Testing</i> 20% Menggunakan Pelabelan Otomatis.....	110
Tabel 4.12 Tabel Hasil Skenario Percobaan Menggunakan Pelabelan Otomatis	112

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja Vaksin (WHO, 2020).....	31
Gambar 2.2 Tahap <i>Tokenizing</i>	33
Gambar 2.3 Tahap <i>Filtering</i>	34
Gambar 2.4 Tahap <i>Stemming</i>	35
Gambar 2.5 Ilustrasi Algoritma KNN (Navlani, 2018).....	38
Gambar 2.6 <i>Confusion Matriks</i> (Mohajon, 2020).....	40
Gambar 3.1 Metode Penelitian.....	44
Gambar 3.2 Proses <i>Web Scraping</i> (Alifian Adexe, 2021)	46
Gambar 3.3 Tahap <i>Preprocessing</i>	48
Gambar 3.4 Tahapan Implementasi metode KNN.....	58
Gambar 3.5 Desain Tampilan <i>Dashboard</i>	67
Gambar 3.6 Desain Tampilan Hasil <i>Crawling</i>	68
Gambar 3.7 Desain Tampilan Hasil <i>Labeling</i>	69
Gambar 3.8 Desain Tampilan <i>Form Data</i>	70
Gambar 3.9 Desain Tampilan <i>Case Folding</i>	70
Gambar 3.10 Desain Tampilan <i>Tokenizing</i>	71
Gambar 3.11 Desain Tampilan <i>Filtering</i>	72
Gambar 3.12 Desain Tampilan <i>Stemming</i>	73
Gambar 3.13 Desain Tampilan Hasil Klasifikasi.....	74
Gambar 3.14 Desain Tampilan Visualisasi.....	75
Gambar 4.1 Hasil Tampilan <i>Dashboard</i>	76
Gambar 4.2 Input File CSV	77
Gambar 4.3 Hasil Tampilan <i>Crawling</i>	77
Gambar 4.4 Hasil Tampilan <i>Labeling</i>	78
Gambar 4.5 Tampilan Hasil <i>Form Data</i>	79
Gambar 4.6 Hasil Tampilan <i>Case Folding</i>	79
Gambar 4.7 Hasil Tampilan <i>Tokenizing</i>	80
Gambar 4.8 Hasil Tampilan <i>Filtering</i>	81
Gambar 4.9 Hasil Tampilan <i>Stemming</i>	82

Gambar 4.10 Tampilan Hasil Klasifikasi.....	83
Gambar 4.11 Tampilan Hasil Visualisasi.....	84
Gambar 4.12 Hasil Pengujian $k = 1$	85
Gambar 4.13 Grafik Nilai k Skenario Data <i>Training</i> 70% dan Data <i>Testing</i> 30% Menggunakan Pelabelan Manual	88
Gambar 4.14 Grafik Nilai k Skenario Data <i>Training</i> 60% dan Data <i>Testing</i> 40% Menggunakan Pelabelan Manual	90
Gambar 4.15 Grafik Nilai k Skenario Data <i>Training</i> 65% dan Data <i>Testing</i> 35% Menggunakan Pelabelan Manual	92
Gambar 4.16 Grafik Nilai k Skenario Data <i>Training</i> 75% dan Data <i>Testing</i> 25% Menggunakan Pelabelan Manual	95
Gambar 4.17 Grafik Nilai k Skenario Data <i>Training</i> 80% dan Data <i>Testing</i> 20% Menggunakan Pelabelan Manual	99
Gambar 4.18 Grafik Nilai k Skenario Data <i>Training</i> 70% dan Data <i>Testing</i> 30% Menggunakan Pelabelan Otomatis.....	103
Gambar 4.19 Grafik Nilai k Skenario Data <i>Training</i> 60% dan Data <i>Testing</i> 40% Menggunakan Pelabelan Otomatis.....	105
Gambar 4.20 Grafik Nilai k Skenario Data <i>Training</i> 65% dan Data <i>Testing</i> 35% Menggunakan Pelabelan Otomatis.....	107
Gambar 4.21 Grafik Nilai k Skenario Data <i>Training</i> 75% dan Data <i>Testing</i> 25% Menggunakan Pelabelan Otomatis.....	109
Gambar 4.22 Grafik Nilai k Skenario Data <i>Training</i> 80% dan Data <i>Testing</i> 20% Menggunakan Pelabelan Otomatis.....	111
Gambar 4.23 <i>Word Cloud</i> Tendensi Positif Pelabelan Manual	113
Gambar 4.24 <i>Word Cloud</i> Tendensi Negatif Pelabelan Manual.....	114
Gambar 4.25 <i>Word Cloud</i> Tendensi Netral Pelabelan Manual.....	115
Gambar 4.26 <i>Word Cloud</i> Tendensi Positif Pelabelan Otomatis.....	116
Gambar 4.27 <i>Word Cloud</i> Tendensi Netral Pelabelan Otomatis	117
Gambar 4.30 <i>Confusion Matrix</i> $k = 6$ Pelabelan Manual.....	119
Gambar 4.31 <i>Confusion Matrix</i> $k = 6$ Menggunakan Pelabelan Otomatis	120

DAFTAR LAMPIRAN

Lampiran 1 Hasil Klasifikasi Labeling Dataset Manual	130
Lampiran 2 Hasil Klasifikasi Labeling Dataset Otomatis.....	144

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pandemi Coronavirus atau COVID-19 yang terjadi di Indonesia sejak ditemukannya kasus positif di masyarakat pertama kali pada Maret 2020 tahun lalu belum ada tanda bahwa pandemi ini akan segera berakhir. Bertambahnya kasus harian yang semakin melonjak menandakan bahwa virus ini dapat dengan mudah menular dari orang yang positif terkena COVID-19 ke orang lainya. Menurut data pada website *www.covid19.go.id* sampai dengan 27 Juli 2021 tercatat virus ini telah mengakibatkan pandemi di 223 negara di dunia dengan jumlah korban sebanyak 194.608.040 terkonfirmasi positif dan 4.170.155 korban meninggal, di Indonesia kasus yang terkonfirmasi positif sebanyak 3.239.936 dengan korban meninggal 86.835 jiwa. Virus Corona dapat menular melalui kontak fisik dengan orang yang tertular COVID-19. Namun, penularan juga bisa terjadi tanpa kontak fisik seperti dengan menularkan tetesan kecil air yang mengandung virus kepada orang lain saat batuk, bersin atau bernafas. Dalam menyikapi hal tersebut Pemerintah Indonesia melakukan berbagai cara untuk memutus mata rantai penyebaran virus corona di antaranya dengan melakukan pembatasan-pembatasan kegiatan bermasyarakat yang menimbulkan kerumunan, dan melakukan 5M (Memakai masker, Mencuci tangan dengan sabun, Menjaga jarak, Menjauhi kerumunan dan Membatasi mobilitas).

Melihat pesatnya penyebaran COVID-19 dan bertambahnya kasus harian orang yang positif virus corona maka diperlukan suatu cara untuk mencegah penyebaran virus ini agar tidak meluas, cara tersebut yaitu dengan melakukan vaksinasi. Pemberian vaksin dapat mencegah angka penyebaran virus corona karena vaksin dapat meningkatkan sistem imunitas tubuh untuk melawan virus corona, dengan demikian diharapkan angka kematian akibat COVID-19 juga akan menurun. Dalam pelaksanaan vaksinasi untuk mencegah virus corona Presiden

Joko Widodo meresmikan Peraturan Presiden (PP) Nomor 99 Tahun 2020 tentang pengadaan vaksin dan pelaksanaan vaksinasi dalam rangka penanggulangan pandemi corona virus disease 2019 (COVID-19). Untuk tahap 1 pemerintah menggunakan vaksin yang di produksi oleh perusahaan biofarmasi asal China, yaitu vaksin corona Sinovac, uji kinis dilakukan oleh tim dari Fakultas Kedokteran Universitas Padjajaran. Dari hasil uji klinis tersebut menunjukkan efikasi vaksin Sinovac sebesar 65,3 persen , hasil tersebut sudah sesuai dengan persyaratan WHO di mana minimal efikasi vaksin adalah 50 persen (Farisa, 2021). Setelah melakukan jalan panjang vaksin COVID-19 Sinovac akhirnya BPOM memberikan izin penggunaan darurat atau *Emergency Use Authorization* (EUA), dengan terbitnya EUA, vaksin Sinovac secara resmi boleh digunakan untuk vaksinasi secara massal (Farisa, 2021).

Terkait hal tersebut berbagai media memberitakan vaksin dengan segala judul berita yang diberikan, terdapat pemberitaan positif dan pemberitaan negatif tentang vaksin, salah satu judul berita tersebut adalah sebagai berikut.

Tabel 1.1 Contoh Berita Positif dan Negatif

Judul Berita Positif	Judul Berita Negatif
Akhirnya, MUI Pastikan Vaksin Corona Sinovac Suci dan Halal	Sudah Divaksin, Bupati dan Wakil Bupati Ciamis Tetap Terpapar Covid-19
AstraZeneca: Vaksin COVID-19 Kami Tidak Mengandung Produk Turunan Babi atau Hewan Lain	Pembekuan Darah Usai Divaksin AstraZeneca Muncul Lagi
Kemendes jamin vaksin COVID-19 AstraZeneca aman	Pemprov Bali tanggapi berita tukang jahit meninggal usai divaksin
Vaksin Sinovac 94 persen efektif cegah COVID-19 bergejala	Pembekuan Darah, 4 Orang Meninggal Usai Divaksin AstraZeneca

Masyarakat yang mendapatkan informasi tentang kegiatan vaksinasi tersebut melalui portal berita memberikan berbagai macam tanggapannya terhadap rencana pemerintah melakukan vaksinasi. Informasi yang diberikan portal berita ke masyarakat dapat membuat berbagai macam opini dan respon terhadap vaksin. Seperti yang ungkapkan oleh Dr dr Kohar Hari Santoso, Direktur RSSA Malang dan Ketua Tim Tracing Gugus Tugas COVID-19 Jawa Timur, Media sangat membantu program imunisasi MR di tahun 2017 dengan menyebarkan sosialisasi dan edukasi positif, upaya ini sangat efektif membuat masyarakat bersedia untuk diimunisasi (Suteja, 2021). Dengan banyaknya pemberitaan positif dan edukasi atau kampanye melalui portal berita kepada masyarakat tentang vaksinasi COVID-19 diharapkan agar target pemberian vaksin COVID-19 tercapai, sedangkan dengan banyaknya pemberitaan negatif dapat menghambat percepatan dalam pemberian vaksin COVID-19. Berdasarkan hal tersebut maka perlu dilakukan analisis terhadap tendensi atau kecenderungan pemberitaan portal berita online mengenai vaksin COVID-19, karena dapat mempengaruhi persepsi masyarakat dan tingkat keberhasilan program pemberian vaksin.

Beberapa penelitian untuk menganalisis tendensi atau sentimen pada portal berita maupun sosial media telah banyak dilakukan. Pada penelitian yang dilakukan Sadida et al.(2017) dengan judul “Perancangan Sistem Analisis Sentimen Masyarakat Pada Sosial Media Dan Portal Berita”, didapatkan hasil informasi dari berupa grafik sentimen positif, negatif dan netral dari media online. Pada penelitian lain yang dilakukan oleh Salam et al.(2018) dengan judul “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)” hasil dari penelitian ini adalah yang didapatkan dari implementasi metode KNN ini cukup baik dengan uji coba sebanyak 6 kali. Rata-rata *accuracy* tertinggi adalah 79.21% sedangkan *accuracy* terendah adalah 70.3%.

K-Nearest Neighbor (KNN) adalah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi

sebelumnya, metode ini termasuk dalam supervised learning, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN (Informatikalogi, 2017).

Berdasarkan latar belakang tersebut, maka dalam penelitian ini diimplementasikan metode *K-Nearest Neighbor* untuk mengetahui tendensi atau kecenderungan pemberitaan portal berita terhadap vaksinasi COVID-19 dengan judul **“ANALISIS TENDENSI PORTAL BERITA TENTANG VAKSINASI COVID-19 DI INDONESIA MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*”**.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dapat diidentifikasi masalah dalam penelitian ini sebagai berikut :

- a. Terdapat pemberitaan mengenai vaksin COVID-19 oleh portal media online yang memiliki tendensi positif, negatif maupun netral.
- b. Pemberitaan terhadap pemberian vaksinasi dapat mempengaruhi keberhasilan program vaksinasi itu sendiri, sehingga perlu diketahui kata-kata apa saja yang berpengaruh terhadap tendensi pemberitaan vaksinasi COVID-19 pada portal berita online.

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka dirumuskan permasalahan dalam penelitian ini sebagai berikut :

- a. Bagaimana kecenderungan pemberitaan portal berita online terhadap vaksinasi COVID-19 di Indonesia?
- b. Bagaimana mengklasifikasikan tendensi positif, negatif maupun netral tentang pemberitaan vaksin COVID-19 dengan menggunakan metode *K-Nearest Neighbor*?
- c. Bagaimana mengetahui kata-kata apa saja yang dominan terhadap tendensi pemberitaan portal media online terhadap vaksin COVID-19.

1.4 Batasan Masalah

a. Input

Data yang digunakan untuk penelitian ini diambil dari portal berita online yang terkait dengan pemberian vaksin COVID-19 dalam rentang waktu bulan Januari 2021 hingga Mei 2021, yaitu sebagai berikut:

Tabel 1.2 Data Portal Berita yang Digunakan

Portal Berita	Jumlah	Bulan
DetikHealth	216 Berita	Januari
Kompas	187 Berita	Februari
Liputan 6	209 Berita	Maret
CNBC Indonesia	175 Berita	April
Antaraneews	215 Berita	Mei

b. Proses

Pemrosesan data dibagi menjadi 3 tahap, yaitu:

- Tahap *Preprocessing*:
 1. *Case Folding* : Proses merubah data menjadi huruf kecil.
 2. *Tokenizing* : Proses pemisahan teks menjadi kata.
 3. *Filtering* : Proses eliminasi kata-kata yang kurang penting.
 4. *Stemming* : Proses pemecahan kata-kata berimbuhan menjadi kata dasar.
- Tahap pembobotan kata dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).
- Tahap klasifikasi tendensi pemberitaan portal media dengan menggunakan metode *K-Nearest Neighbor* (KNN):
 1. Menentukan jumlah tetangga terdekat (k)
 2. Menghitung kuadrat jarak euclidean objek terhadap data *training* yang diberikan. Data *training* merupakan data pemberitaan mengenai vaksin COVID-19 dari berbagai portal berita yang telah diberi label kelas positif, negatif maupun netral

3. Mengurutkan hasil perhitungan kuadrat Euclidean Distance dari kecil ke besar (*ascending*).
4. Memilih sebanyak k objek terdekat dan melakukan pengklasifikasian berita berdasarkan jumlah tetangga terdekat.

c. Output

- Output berupa hasil klasifikasi tendensi portal berita yang dibagi menjadi 3 jenis yaitu tendensi berita positif, negatif maupun netral.
- Hasil klasifikasi akan divisualisasikan berbentuk grafik untuk menunjukkan presentasi jumlah berita positif, negatif dan netral.

1.5 Tujuan Penelitian

Tujuan dari dilakukannya penelitian ini adalah untuk melihat tendensi atau kecenderungan portal berita dalam memberikan informasi tentang vaksinasi kepada masyarakat dan tingkat *accuracy* klasifikasi yang diperoleh menggunakan metode *K-Nearest Neighbor*.

1.6 Manfaat Penelitian

1. Manfaat yang didapat dari penelitian ini diharapkan memberikan gambaran tendensi atau kecenderungan portal berita dalam memberitakan informasi vaksinasi COVID-19.
2. Dengan melihat kecenderungan portal berita tentang informasi vaksinasi COVID-19 dapat diketahui apa saja kata-kata atau isu yang memberikan pemberitaan positif, negatif maupun netral.

1.7 Metode Penelitian

Metode penelitian yang dilakukan dalam analisis tendensi portal berita tentang vaksinasi covid-19 di Indonesia adalah sebagai berikut :

1. Pengumpulan Data

Pada tahap ini, dilakukan proses *crawling* terhadap 5 portal berita yaitu Antaranews 215 data, CNBC Indonesia 175 data, DetikHealth 216 data, Liputan 6 209 data Kompas 187 data untuk mengumpulkan data dan informasi yang akan digunakan untuk proses analisis dan

klasifikasi tendensi portal berita tentang vaksinasi COVID-19 menggunakan metode *K-Nearest Neighbor* (KNN).

2. *Preprocessing*

Tahap *preprocessing* dilakukan seleksi data yang diambil pada portal berita agar data yang akan digunakan menjadi lebih terstruktur. Tahap *preprocessing* pada penelitian ini menggunakan beberapa tahapan, yaitu: *case folding*, *tokenizing*, *filtering* dan *stemming*.

3. Pembobotan kata

Pembobotan kata dilakukan dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

4. Klasifikasi berita dengan menggunakan metode *K-Nearest Neighbor* (KNN)

Data yang telah melalui *preprocessing* akan di klasifikasi sesuai kelasnya untuk menentukan polaritas dari *text* tersebut apakah pemberitaan vaksinasi COVID-19 tersebut termasuk dalam berita yang positif atau berita negatif.

5. Pengujian

Pengujian dilakukan untuk mengukur tingkat ketepatan atau *accuracy* yang dihasilkan oleh metode *K-Nearest Neighbor* (KNN). Pada pengujian penelitian ini adalah menggunakan *confusion matrix*.

1.8 Sistematika Penulisan

Untuk mengetahui secara ringkas permasalahan yang sedang dibahas, maka sistematika penulisan Tugas Akhir analisis tendensi portal berita tentang vaksinasi COVID-19 di indonesia menggunakan metode k-nearest neighbor adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini membahas mengenai Latar Belakang, Identifikasi Masalah, Rumusan Masalah, Batasan Masalah, Tujuan dan Manfaat Penelitian serta Sistematika Penulisan.

BAB II LANDASAN TEORI

Bab ini berisi definis dari teori-teori dan yang berkaitan dengan teknik *preprocessing*, metode pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF), pengenalan *K-Nearest Neighbor* sebagai metode klasifikasi dan pendukung dalam pemecahan masalah.

BAB III METODE PENELITIAN

Bab ini berisi pembahasan langkah-langkah yang dilakukan dalam proses penelitian berkaitan dengan pengumpulan data, meliputi proses *case folding*, *tokenizing*, *filtering* dan *stemming*. Bab ini juga membahas proses pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan menggunakan data yang sudah melalui *preprocessing* serta proses analisis dari metode *K-Nearest Neighbor*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan penjabaran hasil penelitian berdasarkan hasil yang di dapat, serta perhitungan *accuracy* terhadap metode *K-Nearest Neighbor* (KNN).

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian yang sudah dilakukan serta saran-saran yang mungkin diperlukan untuk pengembangan penelitian selanjutnya.

BAB II

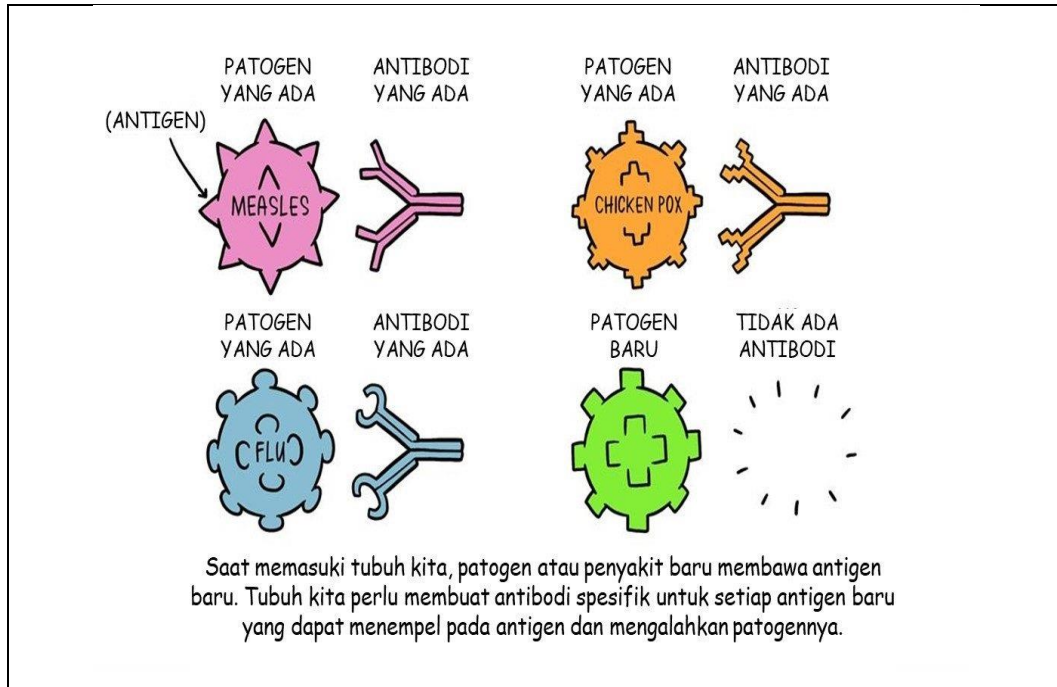
LANDASAN TEORI

2.1 Vaksin COVID-19

COVID-19 adalah penyakit menular yang disebabkan oleh jenis coronavirus yang baru ditemukan di Wuhan, Tiongkok pada bulan Desember 2019. Coronavirus adalah suatu kelompok virus yang dapat menyebabkan penyakit pada hewan dan manusia. Penyakit karena virus ini menyebabkan infeksi pada saluran pernafasan manusia seperti batuk dan pilek (WHO, 2020).

Pada tanggal 11 Maret 2020, Organisasi Kesehatan Dunia (WHO) menetapkan wabah COVID-19 sebagai pandemi global, dikarenakan jumlah kasus di luar China yang meningkat tiga belas kali lipat dan bertambahnya jumlah negara yang terkena dampak COVID-19 (Kompas, 2020). Pandemi global COVID-19 melahirkan problematika baru bagi negara, khususnya mengenai bagaimana upaya negara untuk mencegah dan menghentikan penyebaran virus ini agar tidak semakin meluas, penggunaan vaksin sosial seperti kebijakan pembatasan sosial (social distancing) dan lockdown pun dilakukan oleh negara-negara sebagai respons atas situasi darurat ini. (Valerisha & Putra, 2020). Penggunaan vaksin sosial juga dinilai juga belum cukup untuk penanggulangan penyebaran Covid-19, maka dari itu berbagai negara sedang melakukan riset dan berlomba-lomba untuk dapat menghasilkan vaksin.

Secara ilmu medis vaksin adalah suatu zat yang sengaja dibuat untuk merangsang pembentukan kekebalan tubuh dari penyakit tertentu, sehingga bisa mencegah terjangkit dari penyakit tertentu tersebut. Dalam vaksin mengandung antigen yang sama dengan antigen penyakit tersebut, namun antigen yang ada di dalam vaksin tersebut sudah dikendalikan atau dilemahkan (Ratriani, 2020).



Gambar 2.1 Cara Kerja Vaksin (WHO, 2020)

Vaksin bekerja dengan melatih sistem kekebalan untuk mengenali dan memerangi patogen, baik virus maupun bakteri. Untuk melakukannya, molekul tertentu dari patogen harus dimasukkan ke dalam tubuh guna memicu respons imun. Molekul tersebut disebut dengan antigen, yang ada di semua virus dan bakteri. Dengan menyuntikkan antigen ke dalam tubuh, sistem kekebalan akan belajar mengenalinya. Sebagai pelindung tubuh, sistem kekebalan akan menyerang, memproduksi antibodi, serta mengingatnya jika suatu saat bakteri atau virus tersebut muncul kembali. Jika di kemudian hari muncul, sistem kekebalan otomatis akan mengenali antigen dan menyerang secara agresif sebelum patogen menyebar yang menyebabkan penyakit (Fadli, 2020).

Vaksin yang digunakan oleh pemerintah Republik Indonesia adalah vaksin Sinovac, vaksin dari china ini sudah mendapatkan izin pengunaandarurat *Emergency Use Authorization* (EUA) dari Badan Pengawas Obat dan Makanan (BPOM). Vaksin Sinovac adalah vaksin berjenis *inactivated vaccine* atau virus mati, secara singkat *inactivated vaccine* adalah vaksin menggunakan versi lemah atau inaktivasi dari virus untuk memancing respons imun. Vaksin *inactivated*

memerlukan beberapa dosis dari waktu ke waktu untuk mendapatkan imunitas berkelanjutan terhadap penyakit (CNN, 2021).

2.2 *Text Mining*

Text mining merupakan proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen (Soepomo, 2014). *Text mining* mencoba memecahkan masalah kelebihan informasi (*information overload*) dengan menggunakan teknik dari bidang ilmu terkait, *Text mining* dapat dipandang sebagai perluasan dari data mining atau *Knowledge Discovery in Database* (KDD), yang bertujuan untuk menemukan pola-pola menarik dari basis data berskala besar (Yulian, 2018). *Text mining* bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik.

2.3 *Text Preprocessing*

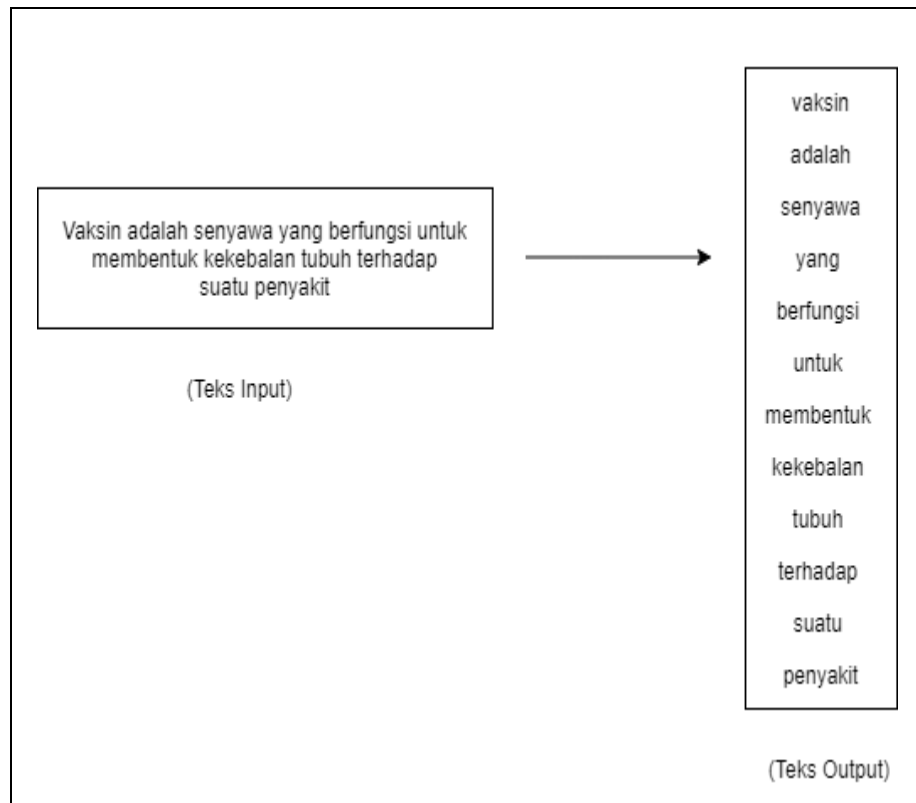
Tahap *Text Preprocessing* adalah mengolah sumber data baik yang terstruktur, terstruktur sebagian dan yang tidak terstruktur dari beberapa sumber. *Text Preprocessing* bertujuan untuk mengolah data yang masih bermacam – macam untuk dijadikan sebuah data teratur yang dapat dikenai atau diterapkan beberapa metode text mining yang ada (H, 2015). Pada proses *text preprocessing* terdapat 4 tahap yaitu meliputi *case folding*, *tokenizing*, *filtering*, dan *stemming* (E. Nugroho, 2011) :

a. *Case Folding*

Case folding adalah tahap mengubah huruf kapital menjadi huruf kecil. Hanya huruf A sampai Z yang dapat diterima karakter selain huruf dihilangkan. Terdapat juga kata-kata tertentu yang harus sesuai dengan kaidah yang tidak bisa dilakukan. Sebagai contoh apabila user ingin mendapatkan informasi tentang “VAKSIN” lalu mengetik “Vaksin”, “VakSin”, atau “ VaksiN” maka akan tetap diberikan hasil retrieval yang sama yaitu “vaksin”.

b. *Tokenizing*

Tahap *tokenizing* atau parsing adalah tahap pemisahan string input berdasarkan tiap kata yang menyusunnya. Contoh tahap tokenisasi seperti Gambar 2.2 berikut :



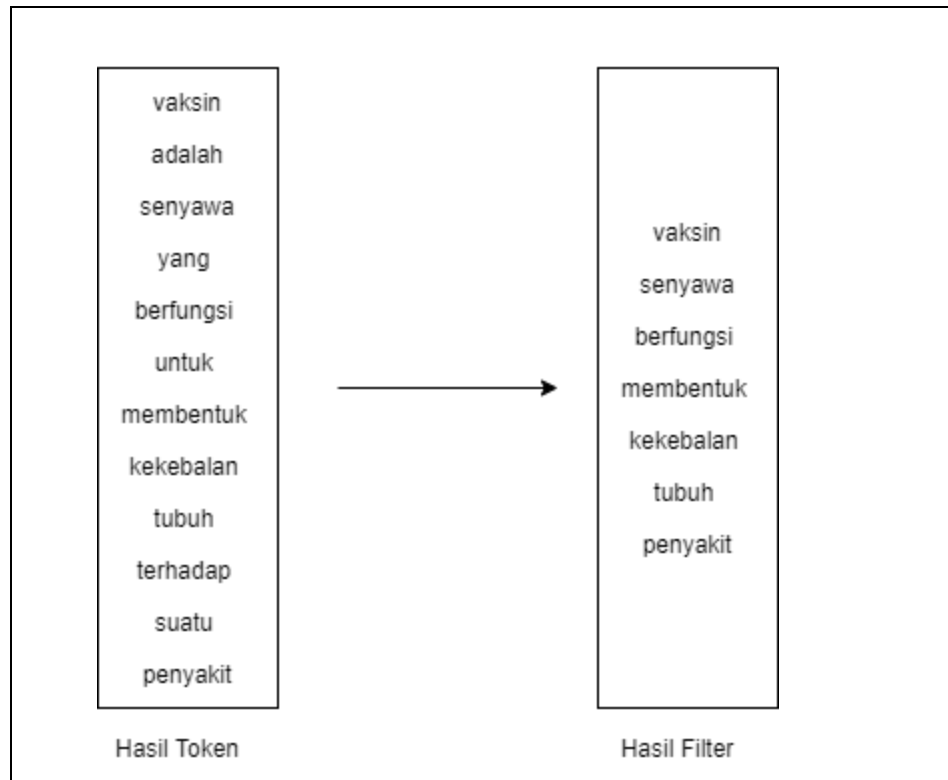
Gambar 2.2 Tahap *Tokenizing*

Pada tahap tokenisasi akan memecah suatu kumpulan karakter ke dalam satuan kata. Misalnya, karakter spasial seperti input, tab, dan spasi dianggap sebagai pengolah kata. Namun, untuk karakter tunggal ('), waktu (.), Titik koma (;), dua arti (:), atau lainnya, dapat menentukan beberapa posisi sebagai pembatas kata.

c. *Filtering*

Filtering adalah tahap pengambilan kata-kata yang penting setelah hasil dari token, atau biasa disebut pengeliminasian sebuah kata-kata sesuai dengan kaidahnya. Dalam pengeliminasian dapat menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist

(menyimpan kata penting). Stoplist atau stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya. Contoh tahap *filtering* adalah seperti Gambar 2.3 berikut ini :



Gambar 2.3 Tahap *Filtering*

d. *Stemming*

Stemming merupakan tahap untuk memecah suatu varian-varian kata atau kata berimbuhan menjadi kata dasar sesuai dengan kata yang di proses. Tahap *stemming* dapat dilihat seperti gambar berikut



Gambar 2.4 Tahap Stemming

2.4 Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah suatu proses untuk melakukan transformasi data dari bentuk teks ke dalam data numerik untuk dilakukan pembobotan pada tiap kata atau fitur. TF-IDF ini adalah ukuran statistik yang digunakan untuk menentukan kata-kata penting dalam dokumen. *Term Frequency* (TF) adalah frekuensi kemunculan kata atau term pada suatu dokumen. *Inverse Document Frequency* (IDF) adalah *inverse* dari nilai DF. Hasil dari pembobotan kata menggunakan TF-IDF menghasilkan suatu nilai yang apabila bobot nilai tersebut besar maka semakin banyak kata yang muncul pada suatu dokumen, apabila bobot nilai tersebut kecil maka semakin kecil kata yang muncul di suatu dokumen (Septian et al., 2018).

Tahapan pembobotan TF-IDF adalah

$$W = TF \times IDF$$

(Persamaan 1)

Dimana:

W = Bobot term atau kata.

TF = Banyaknya kemunculan suatu kata dalam suatu dokumen.

IDF = inverse frekuensi dokumen, dimana

$$\text{IDF} = \log \left(\frac{N}{n_t} \right) \quad (\text{Persamaan 2})$$

N = Jumlah semua dokumen.

n_t = Jumlah dokumen yang mengandung term.

Cara yang lain untuk menghitung nilai IDF adalah dengan menggunakan library python sklearn TfidfVectorizer, terdapat perbedaan dalam perhitungan IDF seperti sebelumnya, persamaan nilai IDF menggunakan TfidfVectorizer adalah sebagai berikut.

$$\text{IDF} = \ln \left(\frac{N}{n_t} \right) + 1 \quad (\text{Persamaan 3})$$

2.5 Sentimen Analisis

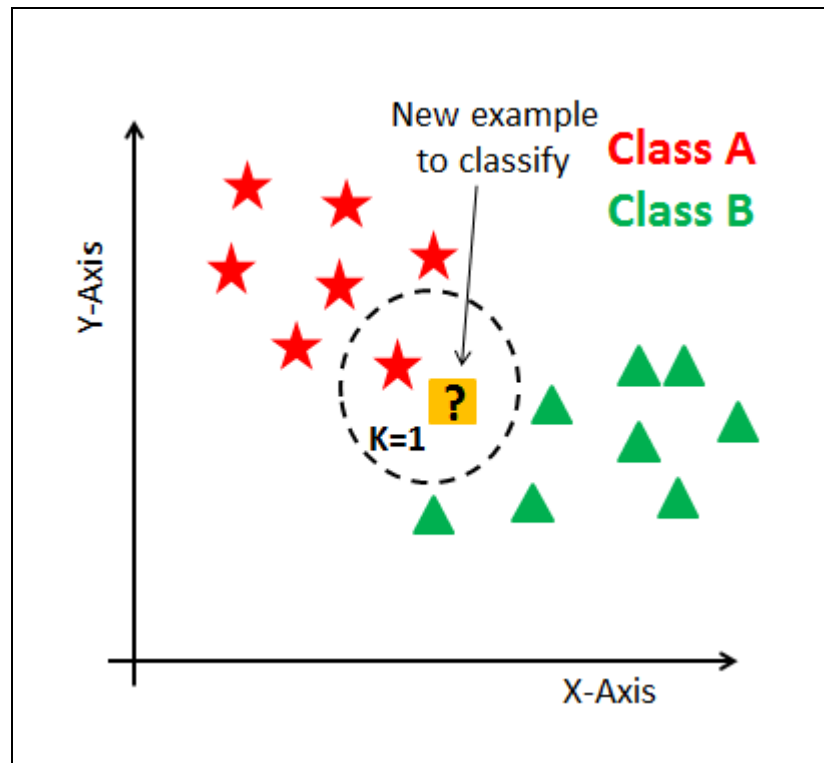
Analisis Sentimen adalah studi yang menganalisis suatu pendapat atau opini-opini orang untuk memberikan penilaian terhadap suatu entitas produk, layanan, peristiwa dan yang lainnya (Liu, 2012). Analisis sentimen telah berkembang menjadi salah satu penelitian aktif dalam *natural language processing*. Dalam beberapa tahun terakhir, penggunaan analisis sentimen juga berkembang dengan digunakannya pada kegiatan industri (Liu, 2012). Analisis sentimen dapat memberikan klasifikasi atas kalimat atau dokumen terhadap sesuatu, nilai yang biasa digunakan pada analisis sentimen adalah positif dan

negatif, dimana dengan nilai tersebut dapat dijadikan parameter dalam pengambilan keputusan (Haryanto et al., 2018).

Banyaknya penelitian mengenai analisis sentimen disebabkan karena analisis sentimen memberikan solusi atas pemecahan suatu masalah dalam pengambilan keputusan. Berkembangnya media sosial dan pertumbuhan teknologi juga turut serta berkembangnya penelitian tentang analisis sentimen ini. Jejak pendapat yang dilakukan orang-orang dikumpulkan dan dianalisis untuk diambil kesimpulan. Contoh penggunaan analisis sentimen adalah apabila *customer* ingin membeli suatu produk maka tidak perlu lagi bertanya kepada keluarga atau teman dekat untuk melihat bagaimana penilaian atas produk yang ingin dibeli, karena sudah banyak tanggapan atau opini pengguna lain tentang produk tersebut. Tak hanya itu analisis sentimen juga dapat digunakan untuk mengamati tanggapan masyarakat terhadap isu suatu permasalahan, sehingga dapat digunakan untuk mengetahui respon politik (Liu, 2012).

2.6 Klasifikasi *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor (KNN) adalah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya, metode ini termasuk dalam supervised learning, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN (Informatikalogi, 2017). KNN ini merupakan metode yang sederhana untuk pengklasifikasian tanpa harus melakukan perhitungan secara kompleks, oleh sebab itu KNN ini juga biasa disebut *lazy learning* (Septian et al., 2018). Ilustrasi algoritma KNN dapat dilihat pada Gambar berikut :



Gambar 2.5 Ilustrasi Algoritma KNN (Navlani, 2018)

Untuk menghitung jarak antar fitur dalam KNN digunakan perhitungan jarak *Euclidean Distance*. *Euclidean distance* adalah perhitungan jarak antara 2 buah titik dalam *Euclidean space*. Euclidean berkaitan dengan teori Pythagoras dan biasanya digunakan untuk 1, 2 dan 3 dimensi. Tetapi juga sederhana apabila diterapkan pada dimensi yang lebih tinggi (Fauzan, 2019). Misalkan diketahui objek **a** dan objek **b** memiliki atribut sebanyak **d**, maka kedekatan jarak antara objek **a** dan objek **b** dapat dihitung dengan menggunakan Euclidean Distance sebagai berikut:

$$D(a, b) = \sqrt{\sum_{i=1}^d (a_i - b_i)^2}$$

(Persamaan 4)

Dimana D adalah jarak kedua titik, **a** adalah koordinat titik akhir dan **b** adalah koordinat titik awal.

Tahapan langkah untuk menggunakan Algoritma KNN adalah sebagai berikut :

1. Menentukan jumlah tetangga terdekat (k)
2. Menghitung kuadrat jarak *euclidean* objek terhadap data *training* yang diberikan.
3. Mengurutkan hasil perhitungan kuadrat *Euclidean Distance* dari kecil ke besar (*ascending*).
4. Ambil sebanyak k objek terdekat dan klasifikasikan berdasarkan nilai tetangga terdekat.
5. Hasil prediksi kategori objek ditentukan berdasarkan nilai tetangga terdekat yang paling banyak.

Peneletian sebelumnya yang membahas tentang analisis sentimen menggunakan metode klasifikasi K-Nearest Neighbor telah banyak dilakukan, pada penelitian yang dilakukan Ernawati dan Wati (2018) dengan judul “Penerapan Algoritma *K Nearest Neighbors* Pada Analisis Sentimen *Review* Agen Travel”, dengan *accuracy* sebesar 87% menghasilkan klasifikasi yang berhubungan dengan sentimen yaitu *Fast, Good, Great, Bad, Cancel* dan *Wait*. Dalam penelitian lain yang dilakukan oleh Salam et al. (2018) dengan judul “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan *K-Nearest Neighbor* (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)”, didapatkan hasil yang cukup baik dengan uji coba sebanyak 6 kali, dengan rata-rata *accuracy* tertinggi adalah 79.21% dan *accuracy* terendah adalah 70.3%.

2.7 *Confusion Matrix*

Confusion matrix adalah suatu metode yang biasa digunakan untuk melakukan perhitungan tingkat *accuracy* pada data mining (Edukasi et al., 2018). *Confusion matrix* dapat memberikan informasi perbandingan hasil dari klasifikasi yang dilakukan oleh model dengan hasil klasifikasi sebenarnya (K. S. Nugroho, 2019). *Confusion matrix* mempunyai kombinasi nilai prediksi dan nilai aktual yang berbeda, dapat dilihat pada Gambar 2.6 berikut.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Gambar 2.6 Confusion Matriks (Mohajon, 2020)

- **True Positive (TP)**
Merupakan jumlah data yang bernilai positif yang diprediksi benar sebagai positif.
- **True Negative (TN)**
Merupakan jumlah data yang bernilai negatif dan diprediksi benar sebagai negatif.
- **False Postive (FP)**
Merupakan jumlah data yang bernilai negatif tetapi diprediksi sebagai data positif..
- **False Negative (FN)**
Merupakan jumlah data yang bernilai positif tetapi diprediksi sebagai data negatif.

Ada beberapa perhitungan untuk melihat seberapa kinerja model yang telah dibuat, beberapa *performance metrics* populer yang umum dan sering digunakan yaitu *accuracy*, *precision*, dan *recall*.

- ***Accuracy***

Accuracy mengacu pada keakuratan model yang dibuat dapat mengklasifikasikan dengan benar. *Accuracy* adalah jarak kedekatan antara nilai prediksi dan nilai aktual (sebenarnya) (K. S. Nugroho, 2019) . Nilai *accuracy* dapat diperoleh dengan persamaan 4 .

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{(Persamaan 5)}$$

- ***Precision***

Precision mengacu pada tingkat keakuratan antara data yang diminta dan hasil prediksi yang diberikan oleh model. *Precision* adalah rasio prediksi positif yang benar terhadap prediksi positif secara keseluruhan (K. S. Nugroho, 2019). Nilai *precision* dapat diperoleh dengan persamaan 5.

$$precision = \frac{TP}{TP + FP} \quad \text{(Persamaan 6)}$$

- ***Recall***

Recall merupakan keberhasilan model dalam menemukan kembali sebuah informasi. *Recall* adalah rasio antara prediksi yang benar positif dibandingkan dengan keseluruhan data positif yang sebenarnya (K. S. Nugroho, 2019). Nilai *Recall* dapat diperoleh dengan persamaan 6.

$$\mathit{recall} = \frac{TP}{TP + FN}$$

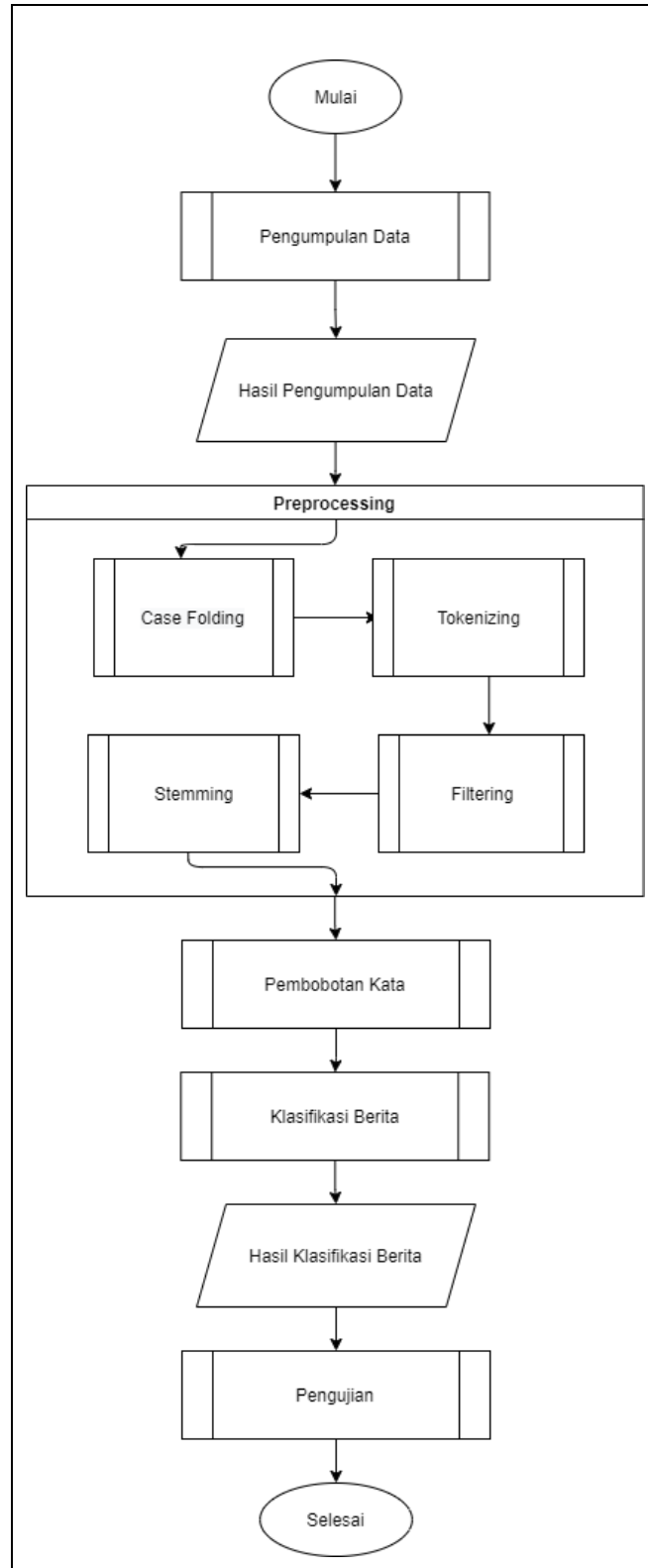
(Persamaan 7)

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Pada bab ini dijelaskan bagaimana tahapan penelitian dan implementasi metode yang digunakan analisis sentimen pada portal berita tentang vaksinasi COVID-19. Adapun tahapan yang dibahas pada bab ini yaitu tahap pengumpulan data, tahap *text preprocessing*, tahap pembobotan kata dan tahap implementasi metode yang digunakan yaitu metode *K-Nearest Neighbor* (KNN). Untuk diagram alir proses untuk metode penelitian seperti Gambar 3.1 sebagai berikut :



Gambar 3.1 Metode Penelitian

Flowchart diatas menggambarkan urutan proses secara mendetail dan hubungan antara suatu proses (intruksi) dengan proses lainnya dengan deskripsi sebagai berikut

1. Pengumpulan Informasi

Proses *crawling* dilakukan terhadap berbagai portal berita untuk mengumpulkan data dan informasi yang akan digunakan untuk proses analisis. Pengumpulan data yang digunakan berkaitan dengan berita kegiatan vaksinasi COVID-19 di Indonesia. Data yang dikumpulkan tadi akan dibagi untuk menjadi data *training* dan data *testing*.

2. *Text Preprocessing*

Data yang berhasil dikumpulkan selanjutnya pada portal berita diseleksi agar data yang akan digunakan menjadi lebih terstruktur Tahap *text preprocessing* pada penelitian ini menggunakan beberapa tahapan, diantaranya: *case folding*, *tokenizing*, *filtering* dan *stemming*.

3. Pembobotan kata

Data yang sudah melalui tahap preprocessing selanjutnya dilakukan pembobotan kata dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

4. Implementasi Klasifikasi *K-Nearest Neighbor* (KNN)

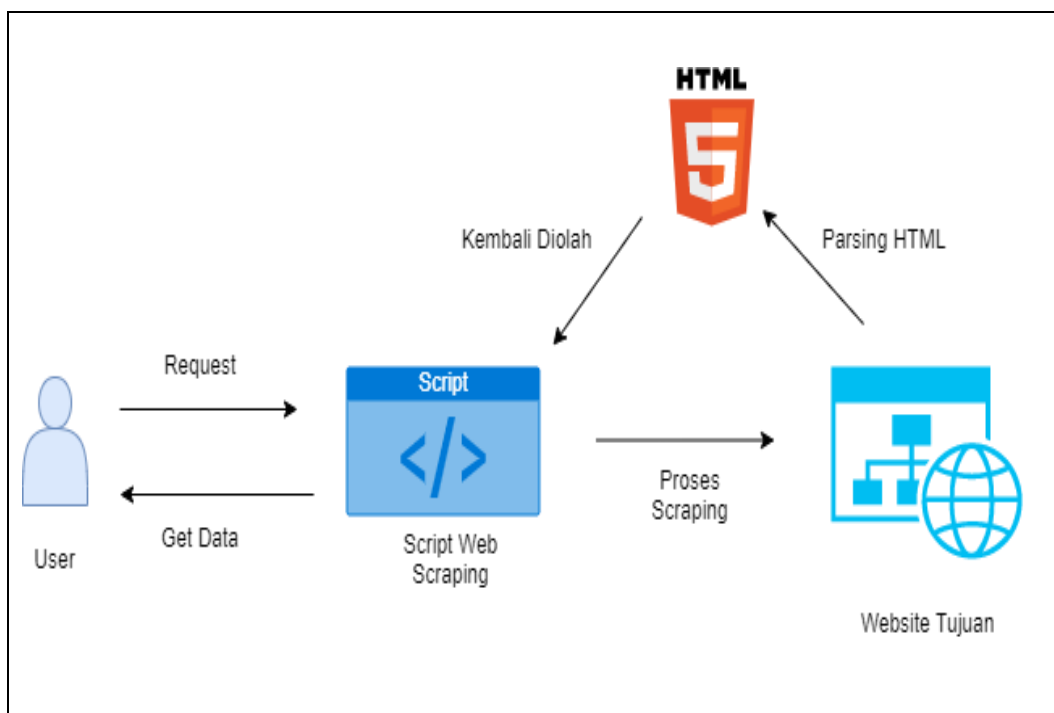
Data yang telah melalui pre-processing dan sudah dilakukan pembobotan kata atau *term* akan di klasifikasi sesuai kelasnya untuk menentukan polaritas dari *text* tersebut apakah *text* tersebut termasuk dalam berita yang positif ,negatif maupun netral terhadap vaksin COVID-19.

5. Pengujian

Pengujian dilakukan untuk mengukur tingkat ketepatan atau *accuracy* yang dihasilkan oleh metode *K-Nearest Neighbor* (KNN). Pada pengujian penelitian ini adalah menggunakan metode *confusion matrix*s.

3.2 Pengumpulan Data

Pengumpulan data yang digunakan untuk menganalisis sentimen berupa informasi yang terdapat pada portal berita tentang vaksinasi COVID-19 di Indonesia. Data yang dikumpulkan dari berita vaksinasi COVID-19 di Indonesia yaitu dicari dengan berbagai keyword yaitu **vaksin covid-19**, **vaksin corona**, **vaksin** dan **vaksinasi**. Website portal berita yang digunakan di antaranya detiknews.com, antranews.com, kompas.com, Liputan6.com dan cnbcindonesia.com. Teknik yang digunakan untuk pengumpulan data yaitu dengan menggunakan *web scraping* dimana data didapatkan dengan pengambilan *title* berita pada web yang digunakan untuk analisis tendensi. Proses *web scraping* dapat dilihat pada Gambar 3.2 berikut.



Gambar 3.2 Proses Web Scraping (Alifian Adexe, 2021)

Hasil data yang dikumpulkan dengan teknik scraping adalah 1000 data, alokasi atau pembagian data *training* dan data *testing* akan dilakukan dengan 5 skenario pembagian data berbeda untuk melihat hasil model yang terbaik yaitu 60%:40% , 65%:35%, 70%:30%, 75%:25% dan 80%:20%. Contoh hasil dari pengumpulan data menggunakan *web scraping* ditunjukkan pada tabel 3.1 sebagai berikut

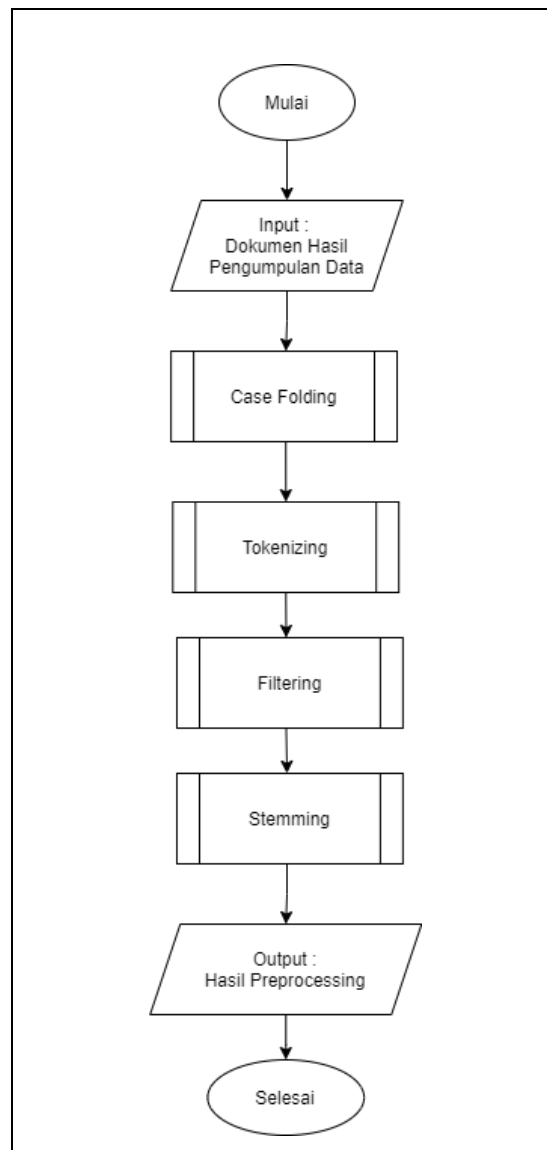
Tabel 3.1 Dokumen Hasil Pengumpulan Data

Title	Sumber	Label
Pakar Imunisasi: Insyaallah Vaksin Tetap Aman Meski Berpuasa	detikHealth	1
Inggris Catat 168 Kasus Pembekuan Darah Usai Vaksinasi AstraZeneca	detikHealth	-1
Vaksinasi Pedagang, Tumbuhkan Rasa Aman Belanja di Pasar	Cnbcindonesia	1
Covid Bisa Dijinakkan Hitungan Bulan, Ekonomi Siap Tancap Gas	Cnbcindonesia	1
Ratusan lansia antusias divaksin COVID-19 saat malam hari	Antaranews	1

Kolom label memberikan keterangan klasifikasi berita tersebut, pemberian label 1 pada title berita menjelaskan bahwa berita tersebut merupakan tendensi berita positif, label 0 untuk tendensi berita netral dan label -1 untuk tendensi berita negatif. Pemberian label pada pengumpulan data yang dilakukan dengan cara manual dengan melihat konteks kalimat secara keseluruhan.

3.3 Preprocessing

Tahap selanjutnya setelah melakukan pengumpulan data yaitu adalah tahap *text preprocessing*, dimana tahap ini merupakan hal yang penting untuk tahap selanjutnya. Pada tahap ini adalah menghilangkan atau mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi. Data yang dikumpulkan dengan teknik web scraping masih merupakan data mentah dan perlu adanya proses agar dokumen yang dihasilkan berkualitas harapannya mempermudah dalam proses klasifikasi. Tahapan *text preprocessing* terdiri dari beberapa tahapan yang dapat dilihat pada Gambar 3.3 berikut.



Gambar 3.3 Tahap Preprocessing

3.3.1 Case Folding

Tahap *case folding* adalah tahap perubahan huruf kapital pada suatu dokumen menjadi huruf kecil. Tujuannya untuk menghilangkan pengulangan data yang berbeda pada hurufnya saja. Pada penelitian ini akan dicek isi title berita yang didapat dari pengumpulan data informasi tentang vaksinasi COVID19, apabila terdapat karakter huruf kapital maka akan diubah kebentuk *lowercase* atau huruf kecil. Berikut gambaran proses dari *case folding* yang dihasilkan terdapat pada tabel 3.2.

Tabel 3.2 Contoh Case Folding

<i>Title/Judul</i>	<i>Hasil Case Folding</i>
Pakar Imunisasi: Inshaallah Vaksin Tetap Aman Meski Berpuasa	pakar imunisasi: insyaallah vaksin tetap aman meski berpuasa
Inggris Catat 168 Kasus Pembekuan Darah Usai Vaksinasi AstraZeneca	inggris catat 168 kasus pembekuan darah usai vaksinasi astrazeneca
Vaksinasi Pedagang, Tumbuhkan Rasa Aman Belanja di Pasar	vaksinasi pedagang, tumbuhkan rasa aman belanja di pasar
Covid Bisa Dijinakkan Hitungan Bulan, Ekonomi Siap Tancap Gas	covid bisa dijinakkan hitungan bulan, ekonomi siap tancap gas
Ratusan lansia antusias divaksin COVID-19 saat malam hari	ratusan lansia antusias divaksin covid-19 saat malam hari

3.3.2 Tokenizing

Tahap *tokenizing* merupakan pemisahan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal. Kata dalam dokumen yang dimaksud adalah kata yang dipisah oleh spasi. Sehingga hasil dari proses ini merupakan kata tunggal yang nantinya akan di perlukan untuk pembobotan kata. Tahap *tokenizing* ini dilakukan setelah melalui tahap *case folding*, dimana dokumen yang berisi title berita yang sudah di ubah ke huruf kecil akan di pisahkan tiap katanya. Untuk memaksimalkan agar mendapatkan data yang baik maka proses *tokenizing* ini juga terdapat proses *cleansing* dimana adanya penghapusan kata yang termasuk link, URL, angka, tanda baca, kata yang tidak sesuai ASCII dan lainnya. Berikut gambaran proses dari *tokenizing* yang dihasilkan terdapat pada tabel 3.3.

Tabel 3.3 Contoh Tokenizing

Hasil <i>Case Folding</i>	Hasil Tokenizing
pakar imunisasi: insyaallah vaksin tetap aman meski berpuasa	pakar imunisasi insyaallah vaksin tetap aman meski berpuasa
inggris catat 168 kasus pembekuan darah usai vaksinasi astrazeneca	inggris catat kasus pembekuan darah usai vaksinasi astrazeneca
vaksinasi pedagang, tumbuhkan rasa aman belanja di pasar	vaksinasi pedagang tumbuhkan rasa aman belanja di pasar

Hasil <i>Case Folding</i>	Hasil Tokenizing
covid bisa dijinakkan hitungan bulan, ekonomi siap tancap gas	covid bisa dijinakkan hitungan bulan ekonomi siap tancap gas
ratusan lansia antusias divaksin covid- 19 saat malam hari	ratusan lansia antusias divaksin covid saat malam hari

3.3.3 *Filtering*

Tahap *filtering* adalah tahap pengambilan kata-kata yang penting setelah hasil dari *tokenizing*, atau biasa disebut pengeliminasian sebuah kata-kata sesuai dengan kaidahnya. Pada proses ini terdapat penghapusan *stopword*, yaitu kata yang tidak deskriptif. Nantinya hasil title berita yang sudah di *tokenizing* di proses kembali dan dilakukan proses penghapusan *stopword*, apabila pada title berita terdapat kata yang termasuk *stopword* maka kata tersebut akan dibuang atau dihapus. Pada proses *filtering stopwords removal* ini digunakan library python *stemmer Sastrawi* yang membantu menghilangkan kata-kata yang tidak deskriptif, selain itu untuk menghasilkan data yang lebih baik atau terstruktur pada

saat pembuatan model klasifikasi, ditambahkan juga daftar *stopword* secara manual. Daftar *stopword* yang ditambahkan adalah kata vaksin, divaksin, vaksinasi, corona, covid dan divaksinasi. Berikut gambaran proses dari *filtering* yang dihasilkan terdapat pada tabel 3.4.

Tabel 3.4 Contoh *Filtering*

Hasil Tokenizing	Hasil <i>Filtering</i> (Remove Stopword)
pakar imunisasi insyaallah vaksin tetap aman meski berpuasa	pakar imunisasi insyaallah tetap aman berpuasa
inggris catat kasus pembekuan darah usai vaksinasi astrazeneca	inggris catat pembekuan darah astrazeneca
vaksinasi pedagang tumbuhkan rasa aman belanja di pasar	pedagang tumbuhkan aman belanja pasar

Hasil Tokenizing	Hasil <i>Filtering</i> (Remove Stopword)
covid bisa dijinakkan hitungan bulan ekonomi siap tancap gas	dijinakkan hitungan ekonomi tancap gas
ratusan lansia antusias divaksin covid saat malam hari	ratusan lansia antusias malam

3.3.4 *Stemming*

Tahap terakhir dari *preprocessing* adalah *stemming*. Pada tahap ini akan dilakukan pemecahan suatu varian-varian kata atau kata berimbuhan menjadi kata dasar sesuai dengan kata yang di proses. Pada proses *stemming* ini digunakan kamus sesuai kaidah bahasa indonesia menggunakan library python Sastrawi, dimana library tersebut menerapkan Algoritma Nazief dan Adriadini. Berikut gambaran proses dari *stemming* menggunakan dokumen yang sudah melalui proses *filtering* sebelumnya, dihasilkan seperti pada tabel 3.5.

Tabel 3.5 Contoh Stemming

Hasil <i>Filtering</i> (Remove Stopword)	Hasil <i>Stemming</i>
pakar imunisasi insyaallah tetap aman berpuasa	pakar imunisasi insyaallah tetap aman puasa
inggris catat pembekuan darah astrazeneca	inggris catat beku darah astrazeneca
pedagang tumbuhkan aman belanja pasar	dagang tumbuh aman belanja pasar
dijinakkan hitungan ekonomi tancap gas	jinak hitung ekonomi tancap gas
ratusan lansia antusias malam	ratus lansia antusias malam

3.4 Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Pada tahap ini diberikan pembobotan tiap kata atau term sebelum melakukan klasifikasi dengan dokumen hasil *stemming*. Dalam perhitungan pembobotan TF-IDF penelitian kali ini akan digunakan library python sklearn `TfidfVectorizer`. Langkah pertama yang dilakukan adalah menentukan *Term Frequency* (TF) yaitu menghitung kemunculan *term* pada dokumen. Hasil dari yang didapat adalah seperti pada tabel 3.6 sebagai berikut:

Tabel 3.6 Hasil Pembobotan *Term Frequency* (TF)

Term	TF				
	Berita 1	Berita 2	Berita 3	Berita 4	Berita 5
aman	1	0	1	0	0
antusias	0	0	0	0	1
astrazeneca	0	1	0	0	0
beku	0	1	0	0	0
belanja	0	0	1	0	0
catat	0	1	0	0	0
dagang	0	0	1	0	0
darah	0	1	0	0	0
ekonomi	0	0	0	1	0
gas	0	0	0	1	0
hitung	0	0	0	1	0
imunisasi	1	0	0	0	0
inggris	0	1	0	0	0
insyaallah	1	0	0	0	0
jinak	0	0	0	1	0
lansia	0	0	0	0	1
malam	0	0	0	0	1
pakar	1	0	0	0	0
pasar	0	1	0	0	0
puasa	1	0	0	0	0
ratus	0	0	0	0	1
tancap	0	0	0	1	0
tetap	1	0	0	0	0
tumbuh	0	0	1	0	0

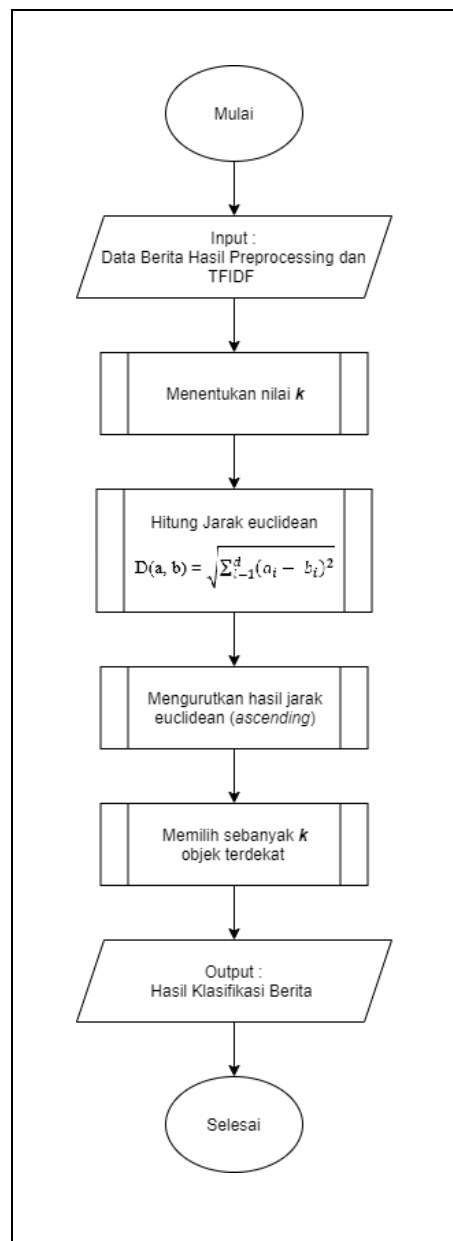
Setelah menghitung *Term Frequency* (TF) tahap selanjutnya yaitu menghitung *Inverse Document Frequency* (IDF), hasil dari perhitungan terdapat pada tabel 3.7 berikut:

Tabel 3.7 Hasil Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Term	DF	D/df	IDF	W = TF IDF				
				Berita 1	Berita 2	Berita 3	Berita 4	Berita 5
aman	2	2,5000	1,9163	1,9163	0,0000	1,9163	0,0000	0,0000
antusias	1	5,0000	2,6094	0,0000	0,0000	0,0000	0,0000	2,6094
astrazeneca	1	5,0000	2,6094	0,0000	2,6094	0,0000	0,0000	0,0000
beku	1	5,0000	2,6094	0,0000	2,6094	0,0000	0,0000	0,0000
belanja	1	5,0000	2,6094	0,0000	0,0000	2,6094	0,0000	0,0000
catat	1	5,0000	2,6094	0,0000	2,6094	0,0000	0,0000	0,0000
dagang	1	5,0000	2,6094	0,0000	0,0000	2,6094	0,0000	0,0000
darah	1	5,0000	2,6094	0,0000	2,6094	0,0000	0,0000	0,0000
ekonomi	1	5,0000	2,6094	0,0000	0,0000	0,0000	2,6094	0,0000
gas	1	5,0000	2,6094	0,0000	0,0000	0,0000	2,6094	0,0000
hitung	1	5,0000	2,6094	0,0000	0,0000	0,0000	2,6094	0,0000
imunisasi	1	5,0000	2,6094	2,6094	0,0000	0,0000	0,0000	0,0000
inggris	1	5,0000	2,6094	0,0000	2,6094	0,0000	0,0000	0,0000
insyaallah	1	5,0000	2,6094	2,6094	0,0000	0,0000	0,0000	0,0000
jinak	1	5,0000	2,6094	0,0000	0,0000	0,0000	2,6094	0,0000
lansia	1	5,0000	2,6094	0,0000	0,0000	0,0000	0,0000	2,6094
malam	1	5,0000	2,6094	0,0000	0,0000	0,0000	0,0000	2,6094
pakar	1	5,0000	2,6094	2,6094	0,0000	0,0000	0,0000	0,0000
pasar	1	5,0000	2,6094	0,0000	2,6094	0,0000	0,0000	0,0000
puasa	1	5,0000	2,6094	2,6094	0,0000	0,0000	0,0000	0,0000
ratus	1	5,0000	2,6094	0,0000	0,0000	0,0000	0,0000	2,6094
tancap	1	5,0000	2,6094	0,0000	0,0000	0,0000	2,6094	0,0000
tetap	1	5,0000	2,6094	2,6094	0,0000	0,0000	0,0000	0,0000
tumbuh	1	5,0000	2,6094	0,0000	0,0000	2,6094	0,0000	0,0000

3.5 Implementasi *K-Nearest Neighbor* (KNN)

Pada tahap ini adalah implementasi metode yang digunakan, yaitu dengan menggunakan algoritma KNN, untuk memberi gambaran bagaimana proses klasifikasi berita dengan menggunakan metode KNN, mulai dari memasukkan data berita hingga menampilkan hasil akhir dari klasifikasi berita dapat dilihat pada Gambar 3.4 berikut:



Gambar 3.4 Tahapan Implementasi metode KNN

Penjelasan secara proses secara mendetail implementasi klasifikasi dengan menggunakan metode KNN pada Gambar 3.4 adalah sebagai berikut:

1. Input atau masukkan data berita hasil *preprocessing* dan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF).
2. Menentukan jumlah tetangga terdekat (k) yang digunakan untuk membuat klasifikasi berita.
3. Menghitung kuadrat jarak euclidean objek dengan rumus (Persamaan 4) terhadap data *training* yang diberikan. Data *training* merupakan data pemberitaan mengenai vaksin Covid-19 dari berbagai portal berita yang telah diberi label kelas positif, negatif dan netral.
4. Mengurutkan hasil perhitungan kuadrat Euclidean Distance dari kecil ke besar (*ascending*).
5. Memilih sebanyak k objek terdekat dan melakukan pengklasifikasian berita berdasarkan jumlah tetangga terdekat.
6. Hasil output berupa tampilan klasifikasi berita berdasarkan jumlah tetangga terdekat

Hasil prediksi kategori objek ditentukan berdasarkan nilai tetangga terdekat yang paling banyak. Sebagai contoh diberikan suatu data uji yang nantinya akan di klasifikasi ke dalam sentimen yang ada. Data uji yang diberika adalah suatu title berita tentang vaksinasi COVID-19, adapun data ujinya adalah sebagai berikut.

“Vaksinasi menggunakan Sinovac terbukti aman”

Dengan data uji tersebut tahap selanjutnya yaitu *text preprocessing*, title tersebut melalui proses *case folding*, *tokenizing*, *filtering stopword*, dan *stemming*, adapun tahap-tahapnya adalah sebagai berikut:

1. *Case Folding*

Dokumen title berita diubah ke dalam huruf kecil menjadi.

“vaksinani menggunakan sinovac terbukti aman”

2. *Tokenizing*

Setelah diubah ke huruf kecil, dilakukan proses *tokenizing*.

[vaksinani, menggunakan, sinovac, terbukti, aman]

3. *Filtering*

Setelah melakukan *tokenizing* selanjutnya tahap *filtering stopwords removal*.

[sinovac, terbukti, aman]

4. *Stemming*

Tahap terakhir title berita yang sudah melalui *stopword removal* adalah proses *stemming*

[novac, bukti, aman]

Setelah melakukan tahap *text preprocessing* selanjutnya menghitung pembobotan kata *Term Frequency-Inverse Document Frequency* (TF-IDF) pada data uji. Untuk tahap pertama adalah pembobotan TF, Hasil pembobotan TF menggunakan data uji dapat dilihat pada tabel 3.8 berikut.

Tabel 3.8 Hasil Pembobotan *Term Frequency* (TF) Data Uji

Term	TF					
	Berita 1	Berita 2	Berita 3	Berita 4	Berita 5	Berita 6 (Data Uji)
aman	1	0	1	0	0	1
antusias	0	0	0	0	1	0
astrazeneca	0	1	0	0	0	0
beku	0	1	0	0	0	0
belanja	0	0	1	0	0	0
catat	0	1	0	0	0	0
dagang	0	0	1	0	0	0
darah	0	1	0	0	0	0
ekonomi	0	0	0	1	0	0
gas	0	0	0	1	0	0
hitung	0	0	0	1	0	0
imunisasi	1	0	0	0	0	0
inggris	0	1	0	0	0	0
insyaallah	1	0	0	0	0	0
jinak	0	0	0	1	0	0
lansia	0	0	0	0	1	0
malam	0	0	0	0	1	0
pakar	1	0	0	0	0	0
pasar	0	1	0	0	0	0
puasa	1	0	0	0	0	0
ratus	0	0	0	0	1	0
tancap	0	0	0	1	0	0
tetap	1	0	0	0	0	0
tumbuh	0	0	1	0	0	0
sinovac	0	0	0	0	0	1
bukti	0	0	0	0	0	1

Tahap selanjutnya adalah menghitung *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil perhitungan TF-IDF menggunakan data uji yang diberikan seperti pada tabel 3.9 berikut.

Setelah mendapatkan nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dari setiap kata atau term, kemudian menghitung jarak Euclidean Distance antar dokumen dengan (Persamaan 4) berdasarkan nilai TF-IDF yang ada pada tabel TF-IDF. Perhitungan dilakukan dengan cara nilai TF-IDF pada data uji dikurangi nilai TF-IDF pada seluruh data atau dokumen yang ada dari B1(Berita 1) sampai B5(Berita 5), kemudian dikuadratkan lalu dijumlahkan semua kata-kata yang sudah dikuadratkan. Berikut hasil Euclidean Distance antar dokumen menggunakan data uji seperti pada tabel 3.10.

Tabel 3.10 Hasil Perhitungan Euclidean Distance Data Uji

(B6, B5)	(B6, B4)	(B6, B3)	(B6, B2)	(B6, B1)
4,4042	4,4042	0,0000	4,4042	0,0000
6,8092	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	6,8092	0,0000
0,0000	0,0000	0,0000	6,8092	0,0000
0,0000	0,0000	6,8092	0,0000	0,0000
0,0000	0,0000	0,0000	6,8092	0,0000
0,0000	0,0000	6,8092	0,0000	0,0000
0,0000	0,0000	0,0000	6,8092	0,0000
0,0000	6,8092	0,0000	0,0000	0,0000
0,0000	6,8092	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	0,0000	6,8092
0,0000	0,0000	0,0000	6,8092	0,0000
0,0000	0,0000	0,0000	0,0000	6,8092
0,0000	6,8092	0,0000	0,0000	0,0000
6,8092	0,0000	0,0000	0,0000	0,0000
6,8092	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	0,0000	6,8092
0,0000	0,0000	0,0000	6,8092	0,0000
0,0000	0,0000	0,0000	0,0000	6,8092
6,8092	0,0000	0,0000	0,0000	0,0000
0,0000	6,8092	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	0,0000	6,8092
0,0000	0,0000	6,8092	0,0000	0,0000
6,8092	6,8092	6,8092	6,8092	6,8092
6,8092	6,8092	6,8092	6,8092	6,8092
45,2592	52,0683	34,0458	58,8775	47,6642

Setelah didapatkan hasil perhitungan antar dokumen antara data uji dan seluruh dokumen, maka langkah selanjutnya mengurutkan hasil dari angka yang paling kecil ke yang paling besar seperti pada tabel 3.11 berikut.

Tabel 3.11 Pengurutan Dokumen Berdasarkan Jarak Euclidean

(DU, Dn)	Hasil perhitungan Jarak
(B6, B3)	34,0458
(B6, B5)	45,2592
(B6, B1)	47,6642
(B6, B4)	52,0683
(B6, B2)	58,8775

Kemudian menentukan jumlah tetangga terdekat, untuk pengujian dipilih tetangga terdekat sejumlah 3, berdasarkan hasil tersebut maka hasil klasifikasi label pada data uji seperti pada B3(Berita 3) yaitu positif.

“Vaksinani menggunakan Sinovac terbukti aman”	Berita Positif
---	----------------

3.6 Tahap Pengujian

Untuk menghasilkan performa model yang maksimal pada pengujian penelitian ini dilakukan pengukuran tingkat ketepatan atau *accuracy* yang dihasilkan oleh metode *K-Nearest Neighbor* (KNN). Pengujian akan dilakukan dengan 5 skenario pembagian data *training* dan data *testing* yang yaitu 60%:40% , 65%:35%, 70%:30%, 75%:25% dan 80%:20%, setiap skenario pembagian data akan dilakukan percobaan sebanyak 50 kali dengan nilai k yang berbeda yaitu 1 sampai 50, kemudian akan di uji menggunakan *Confusion Matrix* pada setiap percobaannya. *Confusion matrix* dapat memberikan informasi perbandingan hasil dari klasifikasi yang dilakukan oleh model dengan hasil klasifikasi sebenarnya. Pada tahap ini akan melakukan pengujian *performance metrics* yaitu *accuracy*,

precision dan *recall*. Berikut skenario percobaan yang akan dilakukan terdapat pada Tabel 3.12 berikut.

Tabel 3.12 Tabel Skenario Percobaan

Data <i>Training</i>	Data <i>Testing</i>	Nilai <i>k</i> terbaik	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
60%	40%				
65%	35%				
70%	30%				
75%	25%				
80%	20%				

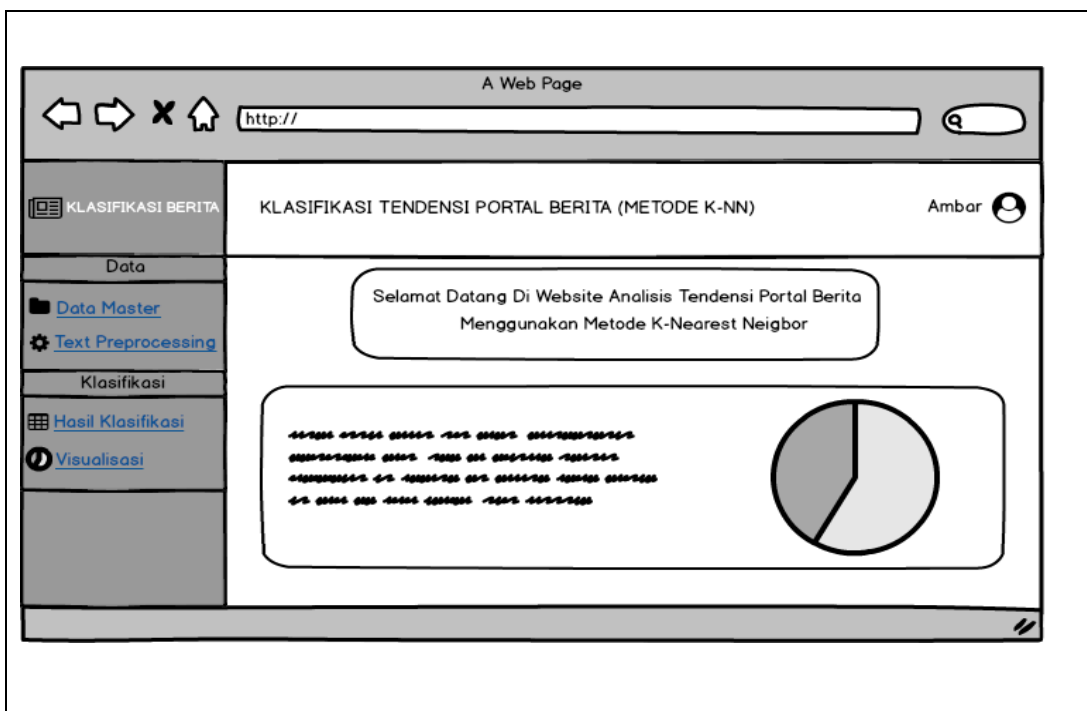
Setelah mendapatkan model pembagian data yang paling baik dengan melihat nilai *accuracy* tertinggi menggunakan hasil *labeling* secara manual, selanjutnya dilakukan pengujian pelabelan otomatis menggunakan program python `sentistrength_id` yang berdasarkan list atau daftar kata positif dan negatif dengan percobaan skenario pembagian data seperti 3.12. Pengujian dilakukan untuk membandingkan dan melihat hasil *labeling* yang digunakan model dalam memberikan klasifikasi.

3.7 Desain Tampilan

Pada tahap ini dilakukan sebuah perancangan desain tampilan yang akan digunakan untuk proses klasifikasi, perancangan desain tampilan dilakukan untuk memberikan visualisasi atau gambaran jelas yang mudah dipahami terhadap klasifikasi tendensi portal berita. Rancangan tampilan yang digunakan pada visualisasi klasifikasi tendensi portal berita adalah menggunakan aplikasi web.

3.7.1 Desain Tampilan Dashboard

Tampilan dashboard adalah halaman utama dari web aplikasi dan memberikan informasi singkat tentang data yang digunakan dari hasil klasifikasi. Pada tampilan dashboard akan menampilkan jumlah data, jumlah sentimen positif, jumlah sentimen netral, dan jumlah sentimen negatif dalam bentuk pie *chart* yang dapat dilihat pada Gambar 3.5 berikut.



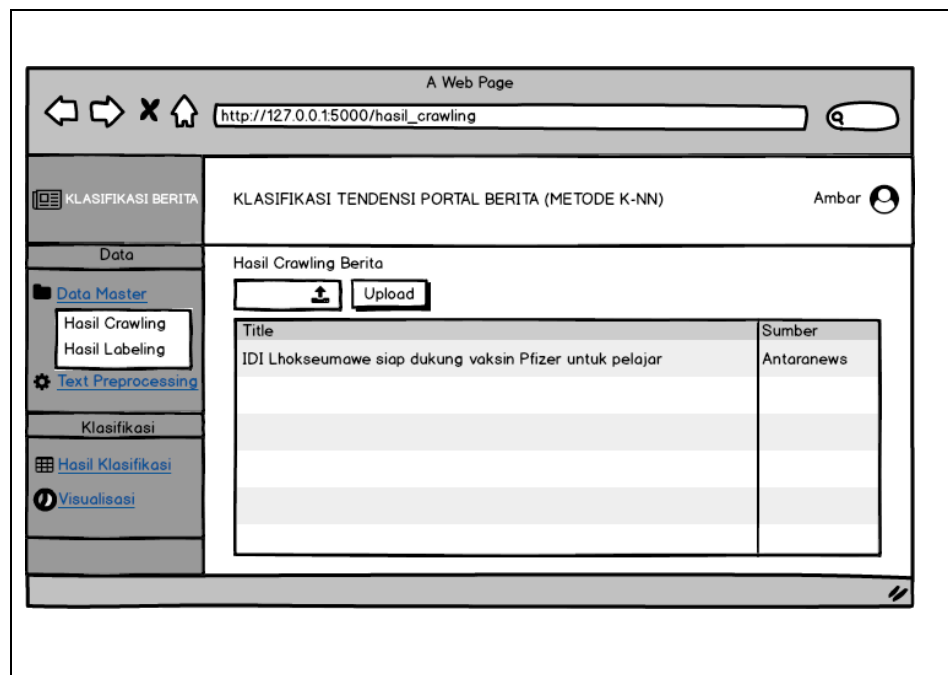
Gambar 3.5 Desain Tampilan Dashboard

3.7.2 Desain Tampilan Data Master

Data Master adalah tampilan yang memberikan informasi isi semua data yang digunakan, pada tampilan data master akan dibagi menjad 2 menu, yaitu menu Hasil *Crawling* dan Hasil *Labeling*.

a. Desain Tampilan Hasil *Crawling*

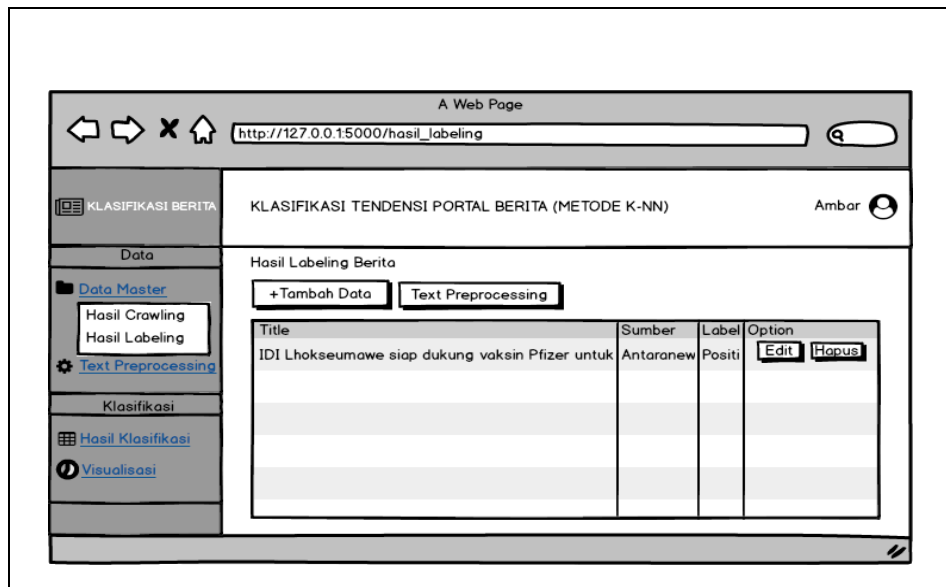
Tampilan aplikasi pada menu Hasil *Crawling* terdapat 2 kolom yaitu kolom title dan kolom sumber. Berita yang sudah di *crawling* akan di tampilkan di tabel tersebut. Terdapat juga form input upload file berupa data berita yang bersumber dari file CSV, karena untuk *crawling* berita dilakukan menggunakan metode lain terpisah dari web aplikasi yang dibuat. Desain tampilan menu Hasil *Crawling* dapat dilihat pada Gambar 3.6 berikut.



Gambar 3.6 Desain Tampilan Hasil *Crawling*

b. Desain Tampilan Hasil *Labeling*

Tampilan Hasil *Labeling* menampilkan tiap-tiap berita beserta sumbernya dan label tendensi berita positif, negatif maupun netral. Pemberian label dilakukan secara manual, data hasil upload CSV yang dilakukan di menu Hasil *Crawling* nantinya akan otomatis juga tampil di menu Hasil *Labeling*. Terdapat juga tombol submit yang berfungsi sebagai langkah selanjutnya dari analisis tendensi portal berita yaitu tombol proses *text preprocessing* yang mencakup proses *case folding*, *tokenizing*, *filterin*, dan *stemming*. Desain tampilan Hasil *Labeling* dapat dilihat pada Gambar 3.7.



Gambar 3.7 Desain Tampilan Hasil *Labeling*

Pada Gambar 3.7 terlihat button atau tombol hapus yang berfungsi menghapus data yang tidak digunakan. Untuk menambah data baru atau mengubah data yang sudah ada disediakan fungsi tambah dan edit data, form data akan tampil yang nantinya akan memperbarui isi dari data CSV. Desain tampilan form data dapat dilihat pada Gambar 3.8 berikut.

Form Data

Title/Judul

Sumber

Label

Gambar 3.8 Desain Tampilan Form Data

3.7.3 Desain Tampilan *Text Preprocessing*

Tampilan *text preprocessing* memberikan sebuah informasi data berita yang diberikan label sebelumnya kemudian melalui proses *text preprocessing*. Dalam tampilan *text preprocessing* terdapat 4 menu yang menampilkan hasil dari *case folding*, *tokenizing*, *filterin*, dan *stemming*.

a. Tampilan *Case Folding*

Desain tampilan web pada menu *case folding* yang menampilkan tabel hasil proses perubahan title berita dari huruf kapital menjadi huruf kecil, dapat dilihat dari Gambar 3.9 berikut.

A Web Page

http://127.0.0.1:5000/hasil_labeling

KLASIFIKASI BERITA

KLASIFIKASI TENDENSI PORTAL BERITA (METODE K-NN) Ambar

Data

Hasil Labeling Berita

Text Preprocessing

Title	Sumber	Label
IDI Lhokseumawe siap dukung vaksin Pfizer untuk pelajar	Antaranews	Positif

Hasil Crawling

Hasil Labeling

Text Preprocessing

Klasifikasi

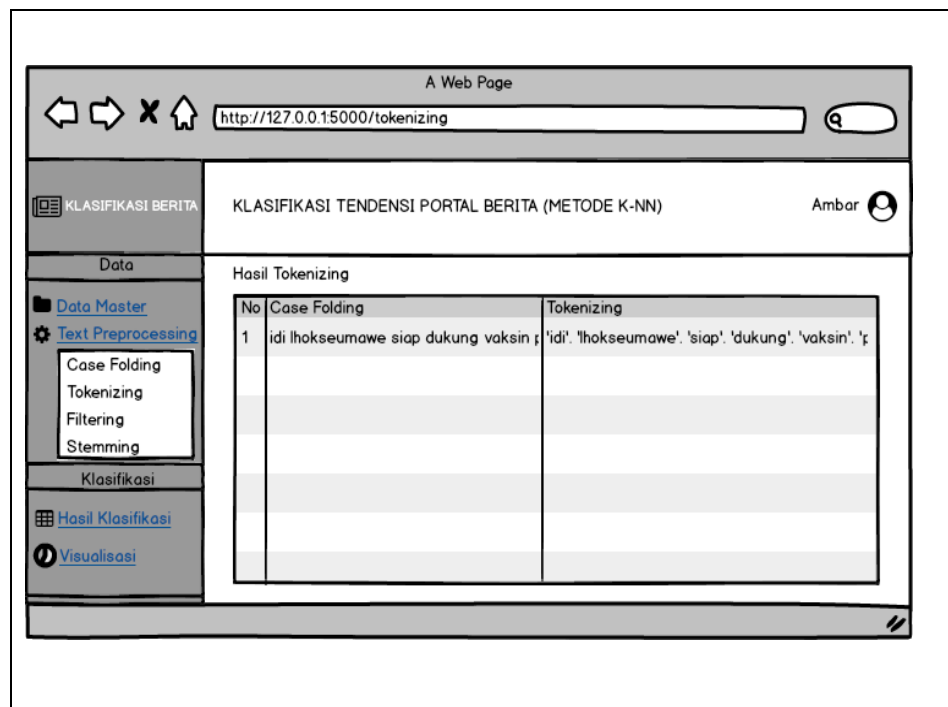
Hasil Klasifikasi

Visualisasi

Gambar 3.9 Desain Tampilan *Case Folding*

b. Tampilan *Tokenizing*

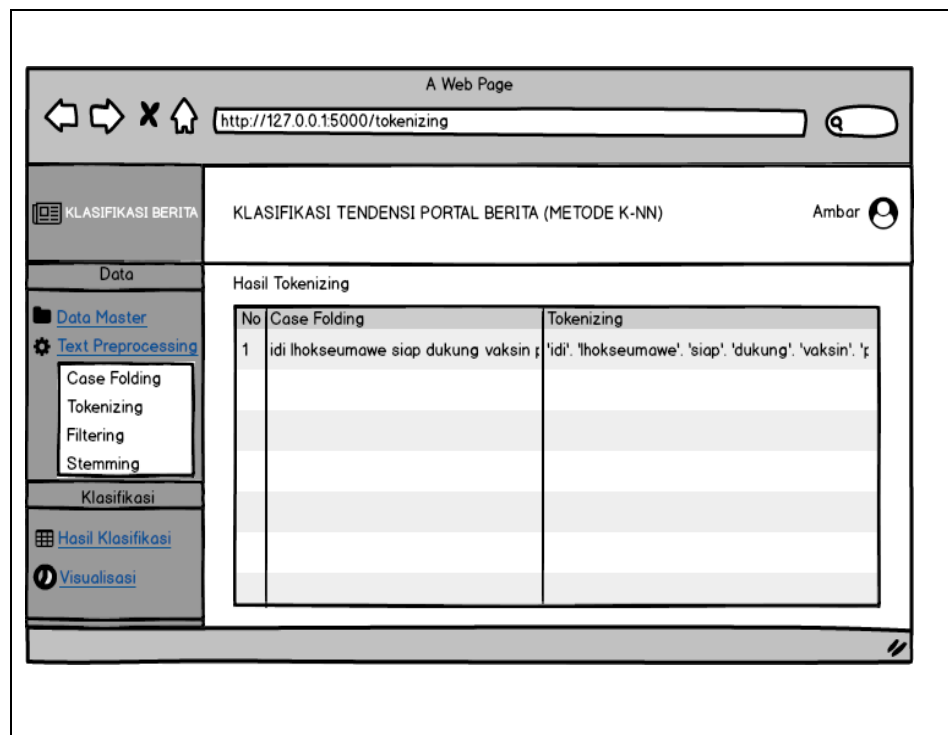
Desain tampilan web pada menu *tokenizing* menampilkan tabel yang berisi 3 kolom. Kolom 1 untuk pemberian nomer berita, kolom 2 berisi hasil proses perubahan title berita setelah melalui proses *case folding*, kemudian kolom 3 adalah proses *tokenizing* dimana proses ini adalah pemisahan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal, dapat dilihat dari Gambar 3.10 berikut.



Gambar 3.10 Desain Tampilan *Tokenizing*

c. Tampilan *Filtering*

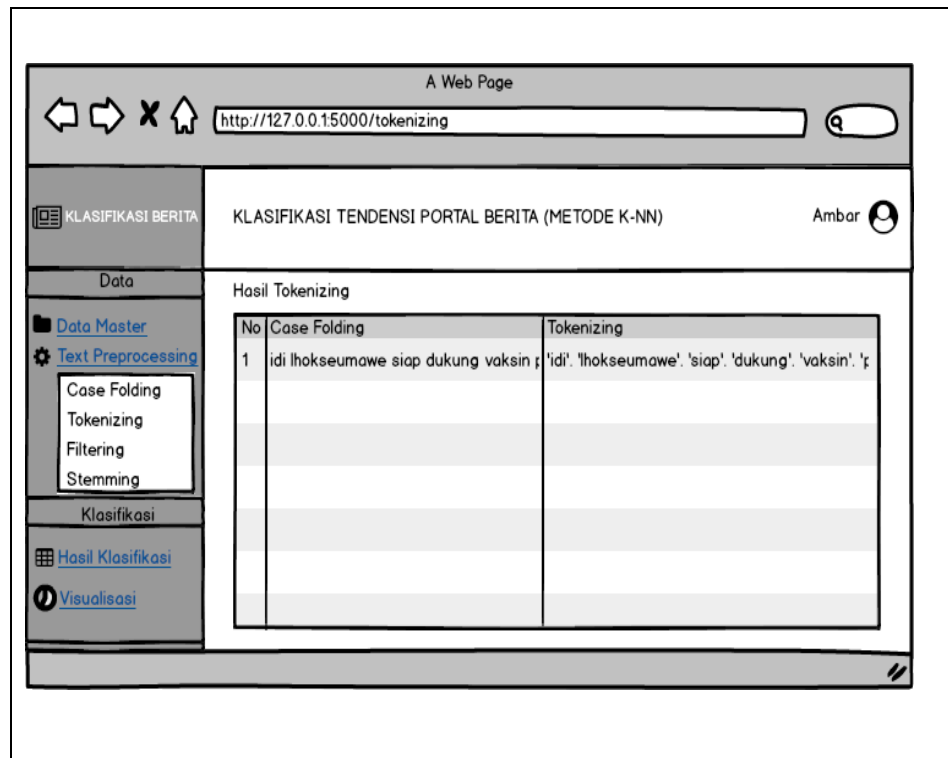
Desain tampilan web pada menu *filtering* menampilkan tabel yang berisi 3 kolom. Kolom 1 untuk pemberian nomer berita, kolom 2 berisi hasil pemisahan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal pada saat proses *tokenizing*, kemudian kolom 3 adalah proses *filtering* dimana proses ini terdapat penghapusan *stopword*, yaitu kata yang tidak deskriptif, dapat dilihat dari Gambar 3.11 berikut.



Gambar 3.11 Desain Tampilan *Filtering*

d. Tampilan *Stemming*

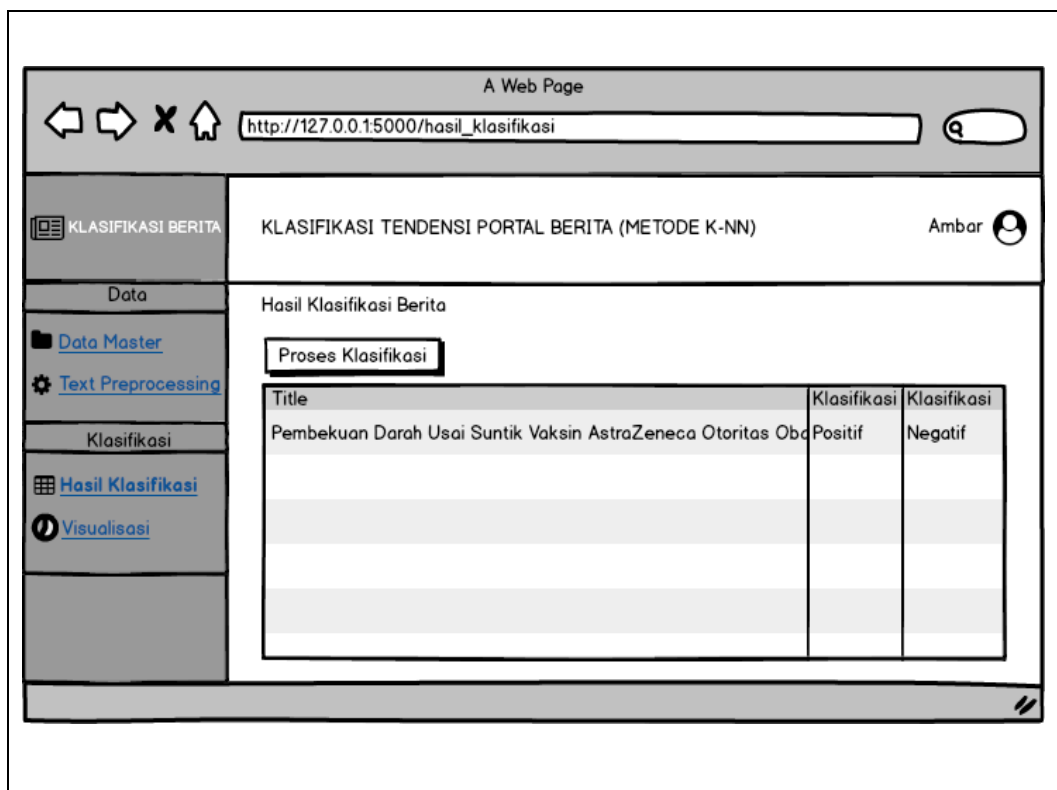
Desain tampilan web pada menu *stemming* menampilkan tabel yang berisi 3 kolom. Kolom 1 untuk pemberian nomer berita, kolom 2 berisi hasil penghapusan *stopword* atau kata yang tidak deskriptif pada saat proses *filtering*, kemudian kolom 3 adalah proses *stemming* dimana proses pemecahan suatu varian-varian kata atau kata berimbuhan menjadi kata dasar, dapat dilihat dari Gambar 3.12 berikut.



Gambar 3.12 Desain Tampilan *Stemming*

3.7.4 Desain Tampilan Hasil Klasifikasi

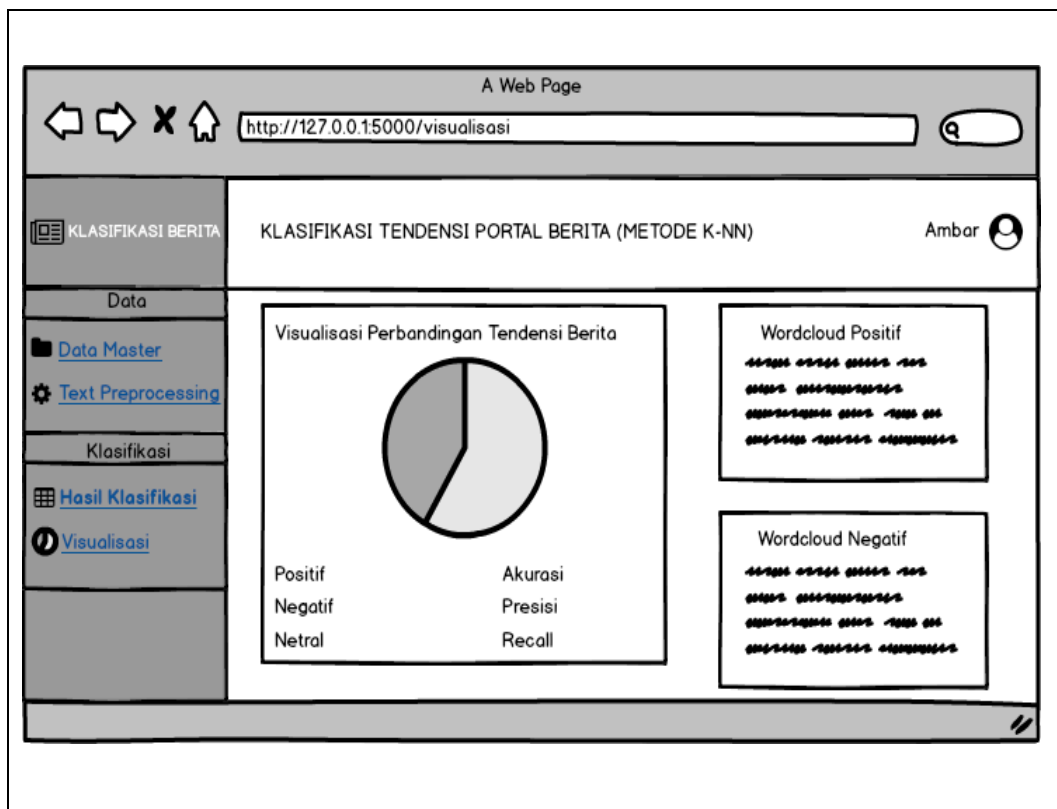
Desain tampilan menu Hasil Klasifikasi memberikan sebuah informasi hasil dari proses klasifikasi berita yang dilakukan menggunakan metode K-Nearest Neighbor. Informasi tabel yang diberikan berisi title berita yang diberikan klasifikasi manual dan hasil klasifikasi dari sistem. Terdapat juga tombol submit Proses Klasifikasi untuk melakukan klasifikasi berita dari hasil *text preprocessing* sebelumnya. Desain tampilan menu Hasil Klasifikasi dapat dilihat pada Gambar 3.13 berikut.



Gambar 3.13 Desain Tampilan Hasil Klasifikasi

3.7.5 Desain Tampilan Visualisasi

Desain tampilan menu Visualisasi memberikan sebuah gambaran yang hasil dari proses klasifikasi berita menggunakan metode K-Nearest Neighbor. Terdapat pie chart yang menggambarkan perbandingan tendensi berita yang di dapat dari hasil data tes saat di melakukan klasifikasi, serta menampilkan evaluasi dari model digunakan. Menu Visualisasi juga memberikan wordcloud atau kata yang sering muncul tiap tendensi berita yang di uji. Desain tampilan menu Visualisasi dapat dilihat pada Gambar 3.14 berikut.



Gambar 3.14 Desain Tampilan Visualisasi

BAB IV

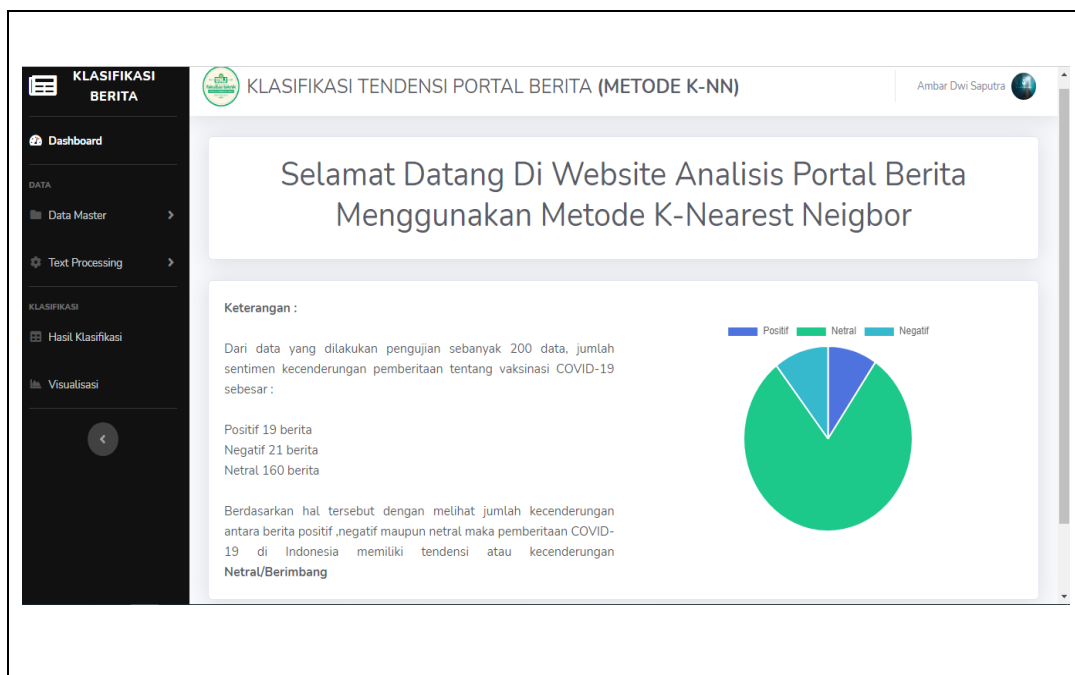
HASIL DAN PEMBAHASAN

4.1 Tampilan Hasil

Tampilan hasil merupakan wujud dari rancangan web aplikasi yang telah dibangun. Pengkodean dilakukan untuk membuat tampilan yang bertujuan memberikan visualisasi atau gambaran jelas yang mudah dipahami terhadap klasifikasi tendensi portal berita, hasil dari pengkodean tampilan adalah sebagai berikut.

4.1.1 Tampilan Hasil *Dashboard*

Dari hasil perancangan desain tampilan yang telah dilakukan, hasil aplikasi web untuk menampilkan *dashboard* berfungsi sebagai halaman utama dan informasi hasil klasifikasi meliputi jumlah data, jumlah sentimen positif, jumlah sentimen netral, dan jumlah sentimen negatif, tampilan hasil dapat terlihat seperti pada Gambar 4.1.



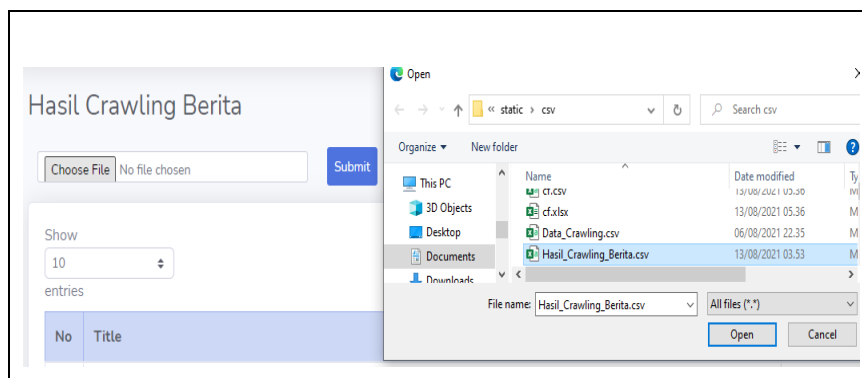
Gambar 4.1 Hasil Tampilan *Dashboard*

4.1.2 Tampilan Hasil Data Master

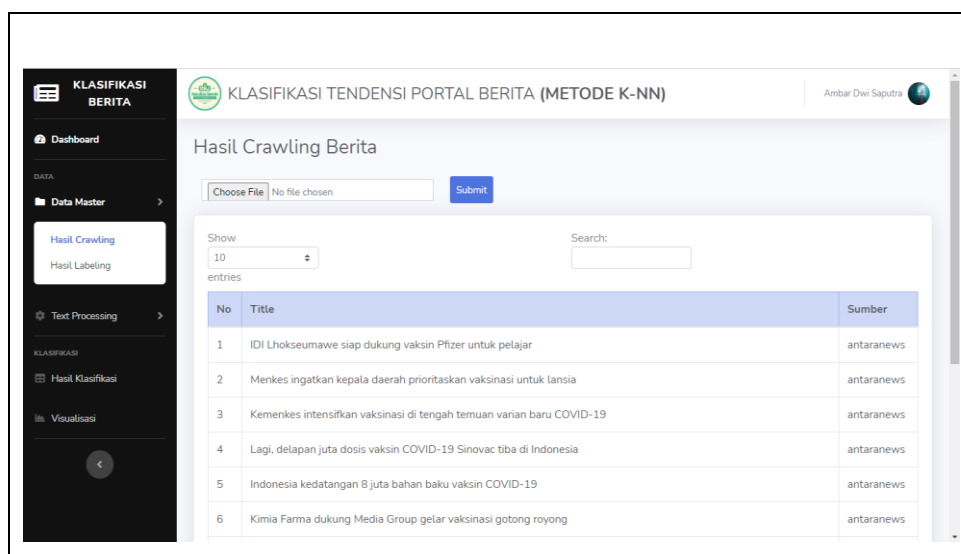
Dari hasil perancangan desain tampilan yang telah dilakukan, hasil aplikasi web menu Data Master yang berfungsi memberikan informasi isi semua data yang digunakan, pada tampilan data master akan dibagi menjad 2 menu, yaitu menu tampilan Hasil *Crawling* dan Hasil *Labeling*.

a. Hasil Tampilan *Crawling*

Tampilan web pada menu Hasil *Crawling* terdapat 2 kolom yaitu kolom *title* dan kolom sumber. Pada tabel tersebut berita yang sudah di *crawling* akan di input atau upload kemudian di tampilkan menggunakan file yang bersumber dari *Comma Separated Values (CSV)*. Tampilan Input file CSV dan Hasil *Crawling* yang bersumber dari file CSV dapat dilihat pada Gambar 4.2 dan 4.3.



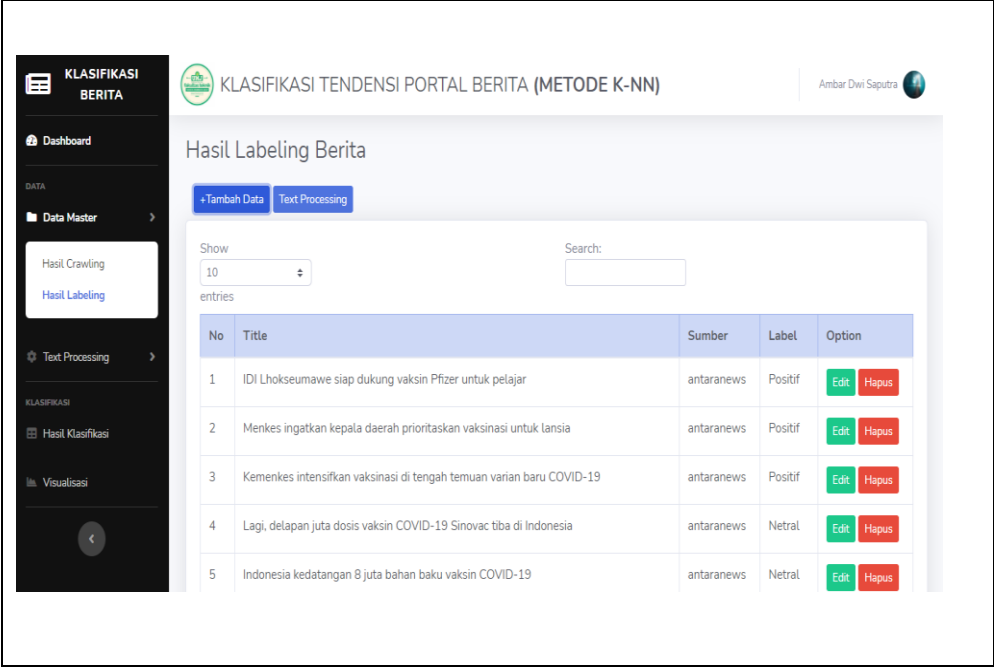
Gambar 4.2 Input File CSV



Gambar 4.3 Hasil Tampilan *Crawling*

b. Hasil Tampilan *Labeling*

Tampilan web pada menu Hasil *Labeling* menampilkan tiap-tiap berita beserta sumbernya dan label tendensi berita positif, negatif maupun netral. Data hasil upload CSV yang dilakukan di menu Hasil *Crawling* nantinya akan otomatis juga tampil di menu Hasil *Labeling*. Terdapat juga tombol submit yang berfungsi sebagai langkah selanjutnya dari analisis tendensi portal berita yaitu tombol proses *text preprocessing* yang mencakup proses *case folding*, *tokenizing*, *filtering*, dan *stemming*. Tampilan Hasil *Labeling* dapat dilihat pada Gambar 4.4 berikut.



No	Title	Sumber	Label	Option
1	IDI Lhokseumawe siap dukung vaksin Pfizer untuk pelajar	antaranews	Positif	Edit Hapus
2	Menkes ingatkan kepala daerah prioritaskan vaksinasi untuk lansia	antaranews	Positif	Edit Hapus
3	Kemendes intensifkan vaksinasi di tengah temuan varian baru COVID-19	antaranews	Positif	Edit Hapus
4	Lagi, delapan juta dosis vaksin COVID-19 Sinovac tiba di Indonesia	antaranews	Netral	Edit Hapus
5	Indonesia kedatangan 8 juta bahan baku vaksin COVID-19	antaranews	Netral	Edit Hapus

Gambar 4.4 Hasil Tampilan *Labeling*

Data yang digunakan untuk proses klasifikasi dapat ditambahkan, diubah atau dihapus. Berikut ini tampilan *form* data untuk menambahkan data atau edit data yang sudah ada dapat dilihat pada Gambar 4.5 berikut.

The image shows a 'Form Data' window with the following fields and buttons:

- Title:** Input field containing 'Data Baru'.
- Sumber:** Input field containing 'Detik'.
- Label:** Dropdown menu showing 'Positif'.
- Buttons:** 'Close' (grey) and 'Simpan' (blue).

Gambar 4.5 Tampilan Hasil Form Data

4.1.3 Tampilan Hasil *Text Preprocessing*

Dari hasil perancangan desain tampilan yang telah dilakukan, hasil aplikasi web menu *Text Preprocessing* menampilkan 4 menu yang terdiri dari *case folding*, *tokenizing*, *filterin*, dan *stemming*.

a. Hasil Tampilan *Case Folding*

Hasil tampilan web pada menu *case folding* yang menampilkan tabel hasil proses perubahan title berita dari huruf kapital menjadi huruf kecil, hasil tampilan dapat dilihat dari Gambar 4.6 berikut

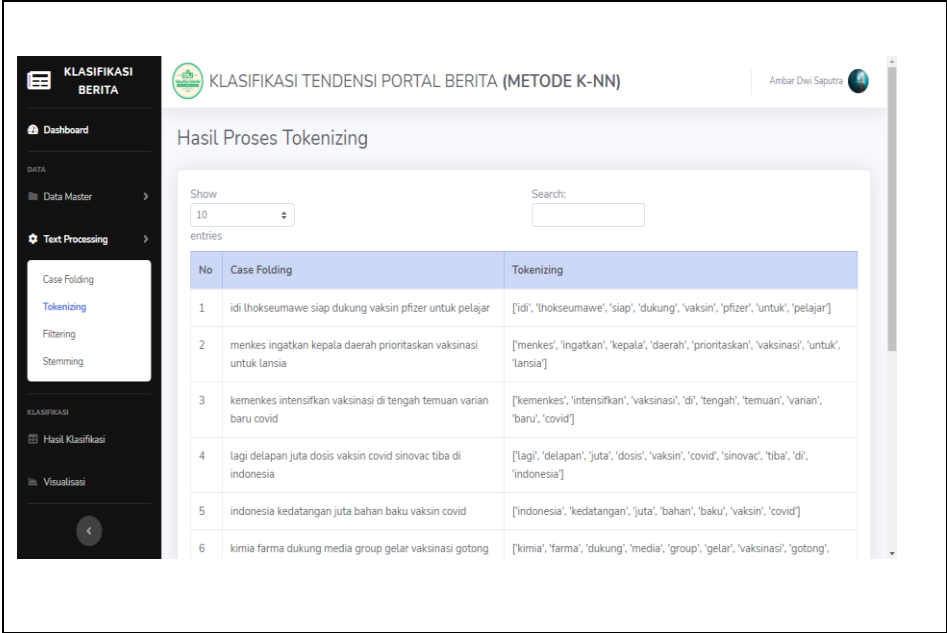
The screenshot shows a web application interface with a sidebar menu and a main content area. The sidebar menu includes 'Dashboard', 'DATA', 'Data Master', 'Text Processing', and 'KLASIFIKASI BERITA'. The 'Text Processing' menu is expanded, showing 'Case Folding', 'Tokenizing', 'Filtering', and 'Stemming'. The main content area displays the 'Hasil Proses Case Folding' table.

No	Title	Case Folding
1	IDI Lhokseumawe siap dukung vaksin Pfizer untuk pelajar	idi lhokseumawe siap dukung vaksin pfizer untuk pelajar
2	Menkes ingatkan kepala daerah prioritaskan vaksinasi untuk lansia	menkes ingatkan kepala daerah prioritaskan vaksinasi untuk lansia
3	Kemenkes intensifkan vaksinasi di tengah temuan varian baru COVID	kemenkes intensifkan vaksinasi di tengah temuan varian baru covid
4	Lagi delapan juta dosis vaksin COVID Sinovac tiba di Indonesia	lagi delapan juta dosis vaksin covid sinovac tiba di indonesia
5	Indonesia kedatangan juta bahan baku vaksin COVID	indonesia kedatangan juta bahan baku vaksin covid
6	Kimia Farma dukung Media Group gelar vaksinasi gotong royong	kimia farma dukung media group gelar vaksinasi gotong royong

Gambar 4.6 Hasil Tampilan *Case Folding*

b. Hasil Tampilan *Tokenizing*

Hasil tampilan web pada menu *tokenizing* hasil proses perubahan title berita setelah melalui proses *case folding* dan menampilkan hasil proses *tokenizing* dimana proses ini adalah pemisahan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal, hasil tampilan dapat dilihat dari Gambar 4.7 berikut.

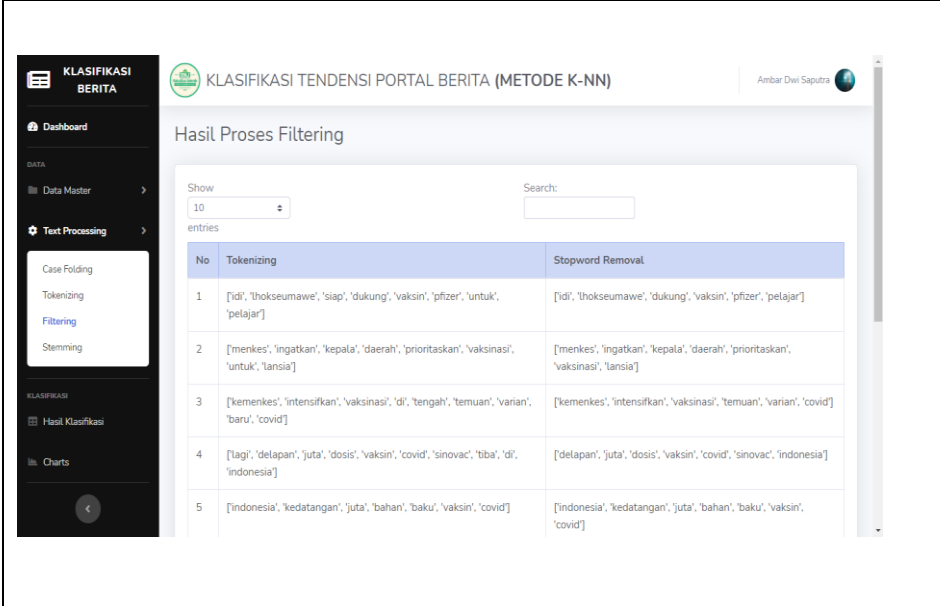


No	Case Folding	Tokenizing
1	idi lhokseumawe siap dukung vaksin pfizer untuk pelajar	['idi', 'lhokseumawe', 'siap', 'dukung', 'vaksin', 'pfizer', 'untuk', 'pelajar']
2	menkes ingatkan kepala daerah prioritaskan vaksinasi untuk lansia	['menkes', 'ingatkan', 'kepala', 'daerah', 'prioritaskan', 'vaksinasi', 'untuk', 'lansia']
3	kemenkes intensifkan vaksinasi di tengah temuan varian baru covid	['kemenkes', 'intensifkan', 'vaksinasi', 'di', 'tengah', 'temuan', 'varian', 'baru', 'covid']
4	lagi delapan juta dosis vaksin covid sinovac tiba di indonesia	['lagi', 'delapan', 'juta', 'dosis', 'vaksin', 'covid', 'sinovac', 'tiba', 'di', 'indonesia']
5	indonesia kedatangan juta bahan baku vaksin covid	['indonesia', 'kedatangan', 'juta', 'bahan', 'baku', 'vaksin', 'covid']
6	kimia farma dukung media group gelar vaksinasi gotong	['kimia', 'farma', 'dukung', 'media', 'group', 'gelar', 'vaksinasi', 'gotong']

Gambar 4.7 Hasil Tampilan *Tokenizing*

c. Hasil Tampilan *Filtering*

Hasil tampilan web pada menu *filtering* menampilkan pemisahan kata title berita setelah melalui proses *tokenizing* dan menampilkan hasil proses *filtering* dimana proses ini terdapat penghapusan *stopword*, yaitu kata yang tidak deskriptif, hasil tampilan dapat dilihat dari Gambar 4.8 berikut.



No	Tokenizing	Stopword Removal
1	[idi, 'thokseumawe', 'siap', 'dukung', 'vaksin', 'pfizer', 'untuk', 'pelajar']	[idi, 'thokseumawe', 'dukung', 'vaksin', 'pfizer', 'pelajar']
2	[menkes, 'ingatkan', 'kepala', 'daerah', 'prioritaskan', 'vaksinasi', 'untuk', 'lansia']	[menkes, 'ingatkan', 'kepala', 'daerah', 'prioritaskan', 'vaksinasi', 'lansia']
3	[kemenkes, 'intensifikan', 'vaksinasi', 'di', 'tengah', 'temuan', 'varian', 'baru', 'covid']	[kemenkes, 'intensifikan', 'vaksinasi', 'temuan', 'varian', 'covid']
4	[lagi, 'delapan', 'juta', 'dosis', 'vaksin', 'covid', 'sinovac', 'tiba', 'di', 'indonesia']	[delapan, 'juta', 'dosis', 'vaksin', 'covid', 'sinovac', 'indonesia']
5	[indonesia, 'kedatangan', 'juta', 'bahan', 'baku', 'vaksin', 'covid']	[indonesia, 'kedatangan', 'juta', 'bahan', 'baku', 'vaksin', 'covid']

Gambar 4.8 Hasil Tampilan *Filtering*

d. Hasil Tampilan *Stemming*

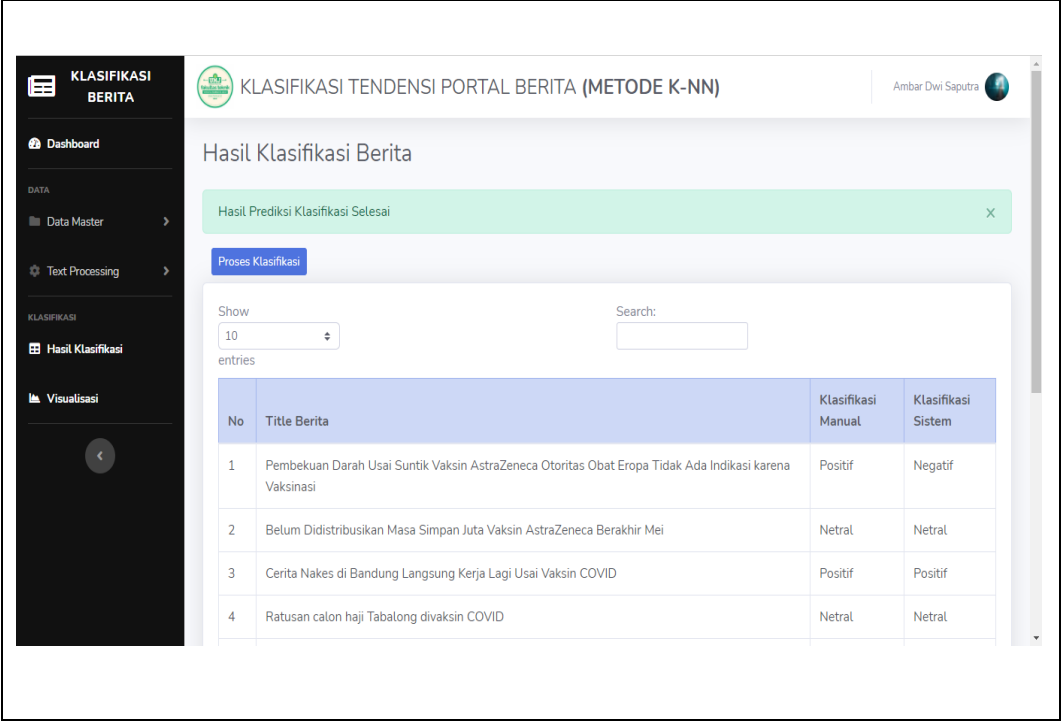
Hasil tampilan web pada menu *stemming* menampilkan hasil penghapusan *stopword* atau kata yang tidak deskriptif pada saat proses *filtering* dan hasil proses *stemming* dimana proses pemecahan suatu varian-varian kata atau kata berimbuhan menjadi kata dasar, hasil tampilan dapat dilihat dari Gambar 4.9 berikut.

No	Stopword Removal	Stemming
1	[idi, 'lhokseumawe', 'dukung', 'vaksin', 'pfizer', 'pelajar]	[idi, 'lhokseumawe', 'dukung', 'vaksin', 'pfizer', 'ajar]
2	[menkes, 'ingatan', 'kepala', 'daerah', 'prioritaskan', 'vaksinasi', 'lansia']	[menkes, 'ingat', 'kepala', 'daerah', 'prioritas', 'vaksinasi', 'lansia']
3	[kemenkes, 'intensifikan', 'vaksinasi', 'temuan', 'varian', 'covid']	[kemenkes, 'intensif', 'vaksinasi', 'temu', 'varian', 'covid']
4	[delapan, 'juta', 'dosis', 'vaksin', 'covid', 'sinovac', 'indonesia']	[delapan, 'juta', 'dosis', 'vaksin', 'covid', 'sinovac', 'indonesia']
5	[indonesia, 'kedatangan', 'juta', 'bahan', 'baku', 'vaksin', 'covid']	[indonesia, 'datang', 'juta', 'bahan', 'baku', 'vaksin', 'covid']
6	[kimia, 'farma', 'dukung', 'media', 'group', 'gelar', 'vaksinasi', 'gotong', 'royong']	[kimia, 'farma', 'dukung', 'media', 'group', 'gelar', 'vaksinasi', 'gotong', 'royong']

Gambar 4.9 Hasil Tampilan *Stemming*

4.1.4 Tampilan Hasil Klasifikasi

Tampilan web menu Hasil Klasifikasi Berita yang sudah melalui proses klasifikasi menggunakan metode K-Nearest Neighbor. terdapat tabel yang diberikan berisi title berita yang diberikan klasifikasi manual dan hasil klasifikasi dari sistem. Terdapat tombol submit Proses Klasifikasi yang berfungsi sebagai perintah untuk memulai klasifikasi berita dari hasil *text preprocessing* sebelumnya. Tampilan Hasil Klasifikasi dapat dilihat pada Gambar 4.10 berikut.



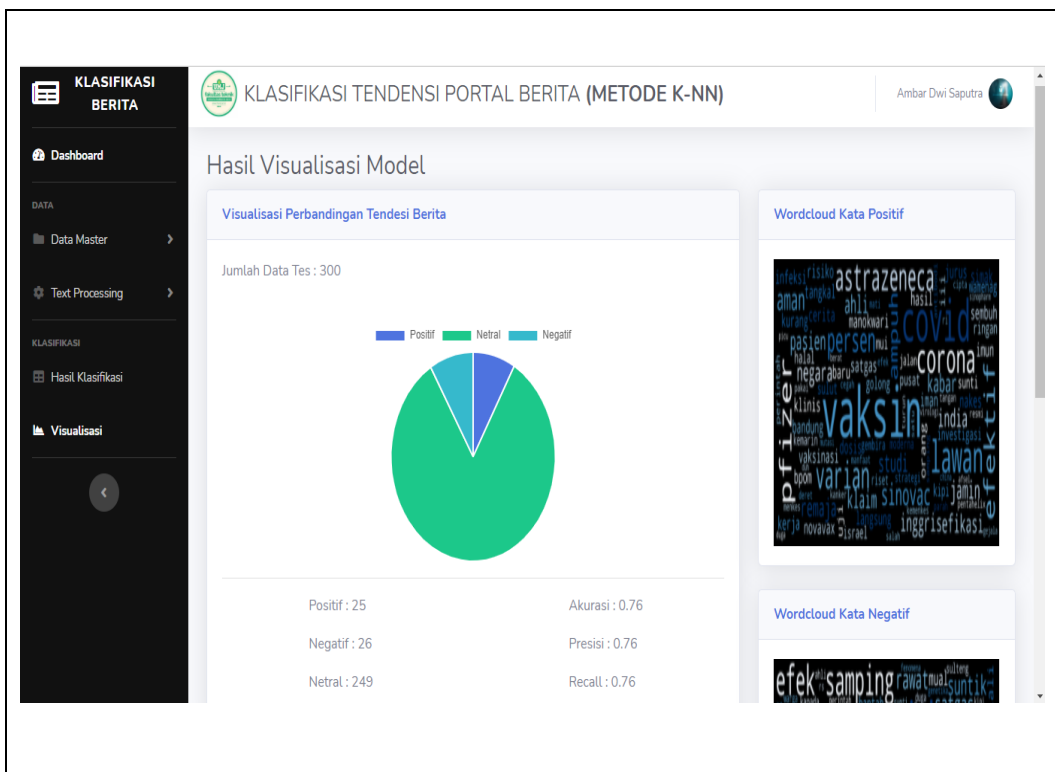
The screenshot displays the 'Hasil Klasifikasi Berita' (News Classification Results) page. The page title is 'KLASIFIKASI TENDENSI PORTAL BERITA (METODE K-NN)' and the user is 'Ambar Dwi Saputra'. The page shows a table with 4 rows of news titles and their manual and system classifications. The system classifications are: 1. Positif, 2. Netral, 3. Positif, 4. Netral.

No	Tittle Berita	Klasifikasi Manual	Klasifikasi Sistem
1	Pembekuan Darah Usai Suntik Vaksin AstraZeneca Otoritas Obat Eropa Tidak Ada Indikasi karena Vaksinasi	Positif	Negatif
2	Belum Didistribusikan Masa Simpan Juta Vaksin AstraZeneca Berakhir Mei	Netral	Netral
3	Cerita Nakes di Bandung Langsung Kerja Lagi Usai Vaksin COVID	Positif	Positif
4	Ratusan calon haji Tabalong divaksin COVID	Netral	Netral

Gambar 4.10 Tampilan Hasil Klasifikasi

4.1.5 Tampilan Hasil Visualisasi

Tampilan menu Visualisasi pada saat memberikan informasi gambaran yang berbentuk pie chart dari hasil proses klasifikasi berita menggunakan metode K-Nearest Neighbor. Pie chart tersebut menggambarkan perbandingan tendensi berita yang di dapat dari hasil data tes saat di melakukan klasifikasi, serta menampilkan evaluasi dari model digunakan yaitu nilai *accuracy*, *precision* dan *recall*. Menu Visualisasi juga memberikan wordcloud atau kata yang sering muncul tiap tendensi berita yang di uji. Tampilan hasil Visualisasi dapat dilihat pada Gambar 4.11 berikut.



Gambar 4.11 Tampilan Hasil Visualisasi

4.2 Hasil Pengujian

Untuk mengetahui model yang dibangun mempunyai *accuracy* dan tingkat ketepatan maka dibutuhkan suatu pengujian dalam implementasi klasifikasi berita menggunakan metode K-Nearest Neighbor. Pada pengujian penelitian ini adalah jumlah data yang digunakan sebesar 1000 data yang didapatkan dari hasil *scrapping* web pada portal berita online. Data yang dikumpulkan kemudian diberi label secara manual dan otomatis untuk melihat perbandingan hasil *labeling* serta klasifikasi berita.

4.2.1 Pelabelan Manual

Data yang yang dikumpulkan dibagi ke dalam data *training* sebesar 70% dan data *testing* 30% , kemudian diberikan pelabelan secara manual dengan klasifikasi 3 kelas yaitu -1 negatif, 0 netral dan 1 positif. Hasil dari pelabelan tersebut kemudian digunakan untuk proses *text processing* meliputi *case folding* ,*tokenizing*, *filtering* dan *stemming* untuk mendapatkan data yang baik dan lebih terstruktur. Hasil dari *text processing* selanjutnya akan di proses untuk membuat model dan menentukan klasifikasi tendensi pada portal berita. Berikut ini tampilan 10 data pertama hasil pengujian klasifikasi dengan nilai $k = 1$ seperti pada gambar 4.12 berikut.

No	Title Berita	Klasifikasi Manual	Klasifikasi Sistem
1	Wamenkes Jelaskan Alasan Pasien COVID-19 Tak Dapat Jatah Vaksin	Netral	Netral
2	Penderita Kanker Diperbolehkan Ikut Vaksinasi Covid-19, Ini Syaratnya	Positif	Netral
3	UPDATE 10 Februari: Kasus Aktif Covid-19 di Indonesia Tercatat 168.416	Netral	Netral
4	Vaksinasi Covid-19 Tahap Dua Dimulai Besok, Grobogan Dapat Jatah 12.000 Dosis	Netral	Netral
5	Menkes Disuntik Vaksin Dosis Kedua: Rasanya Jadi Pengen Makan Banyak	Netral	Positif
6	UPDATE 19 Februari: 668.914 Nakes Sudah Divaksinasi Dosis Kedua, 1.191.031 Dosis Pertama	Netral	Netral
7	UPDATE 25 Februari: 853.745 Tenaga Kesehatan Sudah Divaksinasi Dosis Kedua, 1.461.920 Dosis Pertama	Netral	Netral
8	Tak Ragu Disuntik Vaksin Sinovac, dr Tirta: Sudah Terjamin Uji Klinis	Positif	Netral
9	Vaksinasi COVID-19 ASN Jabar Dimulai, 570 Orang Sudah Disuntik Vaksin Corona	Netral	Negatif
10	Penghentian sementara vaksin AstraZeneca CTMAV547	Netral	Netral

Showing 1 to 10 of 300 entries

Previous 1 2 3 4 5 ... 30 Next

Gambar 4.12 Hasil Pengujian $k = 1$

Setelah melakukan pengujian dengan nilai $k = 1$ didapat *accuracy* sebesar 69,67%, untuk melihat performa model yang dibuat maka model diuji dengan 5 skenario pembagian data *training* dan data *testing* yang berbeda seperti pada Tabel 3.12. Setiap skenario dilakukan percobaan sebanyak 50 kali dengan jumlah tetangga terdekat(k) dimulai dari $k = 1$ sampai dengan $k = 50$. Tiap pengujian dalam iterasi k dievaluasi model tersebut menggunakan *confusion matrix* dengan melihat nilai *accuracy*, *precision* dan *recall*. Dengan melihat hasil nilai-nilai tersebut nantinya dapat mengetahui alokasi pembagian data *training* dan *testing* yang optimal untuk menentukan jumlah tetangga terdekat atau nilai k yang memiliki performa paling baik untuk model menggunakan *K-Nearest Neighbor*.

a. Skenario Data Training 70% dan Data Testing 30% Menggunakan Pelabelan Manual

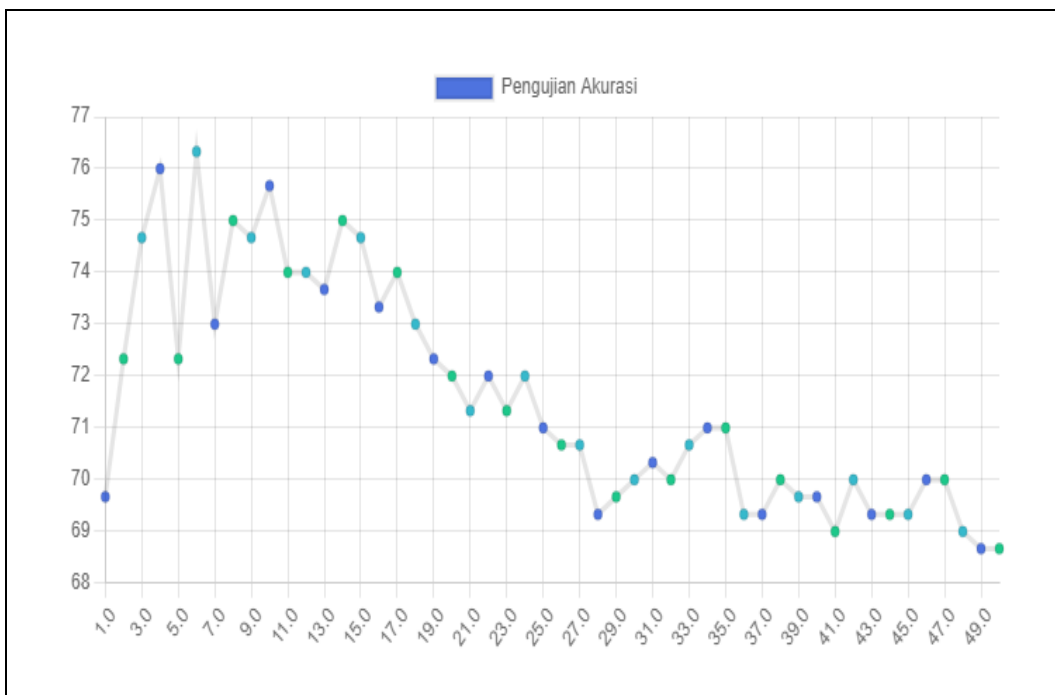
Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 700 data *training* dan 300 data *testing* dengan uji performa model sebanyak $k = 50$ pada klasifikasi berita Vaksinasi Berita COVID-19 seperti pada Tabel 4.1 berikut.

Tabel 4.1 Hasil Nilai k Skenario Data *Training* 70% dan Data *Testing* 30% Menggunakan Pelabelan Manual

Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l	Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l
1	69,67	59,33	58,93	26	71,00	64,61	48,97
2	72,33	63,57	59,11	27	70,67	64,84	48,27
3	74,67	64,38	61,93	28	70,67	64,90	48,63
4	76,00	68,81	61,27	29	69,33	61,86	45,99
5	72,33	63,03	56,71	30	69,67	62,74	46,52
6	76,33	69,29	60,72	31	70,00	63,53	46,78
7	73,00	63,77	56,60	32	70,33	65,02	47,22
8	75,00	68,13	57,99	33	70,00	64,20	46,69
9	74,67	67,69	57,02	34	70,67	66,60	47,12
10	75,67	69,50	59,67	35	71,00	71,53	46,94
11	74,00	67,25	57,47	36	71,00	70,81	46,67
12	74,00	67,49	56,32	37	69,33	64,40	43,76
13	73,67	67,50	57,04	38	69,33	64,33	44,38
14	75,00	71,07	57,99	39	70,00	70,15	43,84
15	74,67	69,38	58,17	40	69,67	66,73	43,31
16	73,33	66,52	54,45	41	69,67	65,56	43,31
17	74,00	69,37	55,42	42	69,00	64,49	41,99
18	73,00	65,84	55,17	43	70,00	68,19	43,48
19	72,33	65,99	52,60	44	69,33	66,00	42,16
20	72,00	65,17	52,07	45	69,33	66,00	42,16
21	71,33	64,09	51,01	46	69,33	66,00	42,16
22	72,00	64,40	51,36	47	70,00	70,70	43,48
23	71,33	63,98	50,66	48	70,00	70,70	43,48
24	72,00	66,60	51,09	49	69,00	67,54	41,63

25	69,67	59,33	58,93	50	68,67	66,19	41,10
----	-------	-------	-------	----	-------	-------	-------

Dari pengujian yang dilakukan dengan 50 kali untuk menentukan jumlah tetangga terdekat yang optimal dan *confusion matrix* dengan melihat nilai *accuracy*, *precision* dan *recall* maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 6 dengan tingkat *accuracy* sebesar 76,33% yang terlihat pada Tabel 4.1 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.11 berikut.



Gambar 4.13 Grafik Nilai k Skenario Data *Training* 70% dan Data *Testing* 30% Menggunakan Pelabelan Manual

Terlihat pada Gambar 4.13 saat data di uji sebanyak 50 kali jumlah $k = 6$ menunjukkan *accuracy* tertinggi, iterasi selanjutnya nilai k pada *accuracy* bergerak fluktuatif dan mengalami penurunan dan tidak mengalami kenaikan yang signifikan melebihi iterasi sebelumnya.

b. Skenario Data *Training* 60% dan Data *Testing* 40% Menggunakan Pelabelan Manual

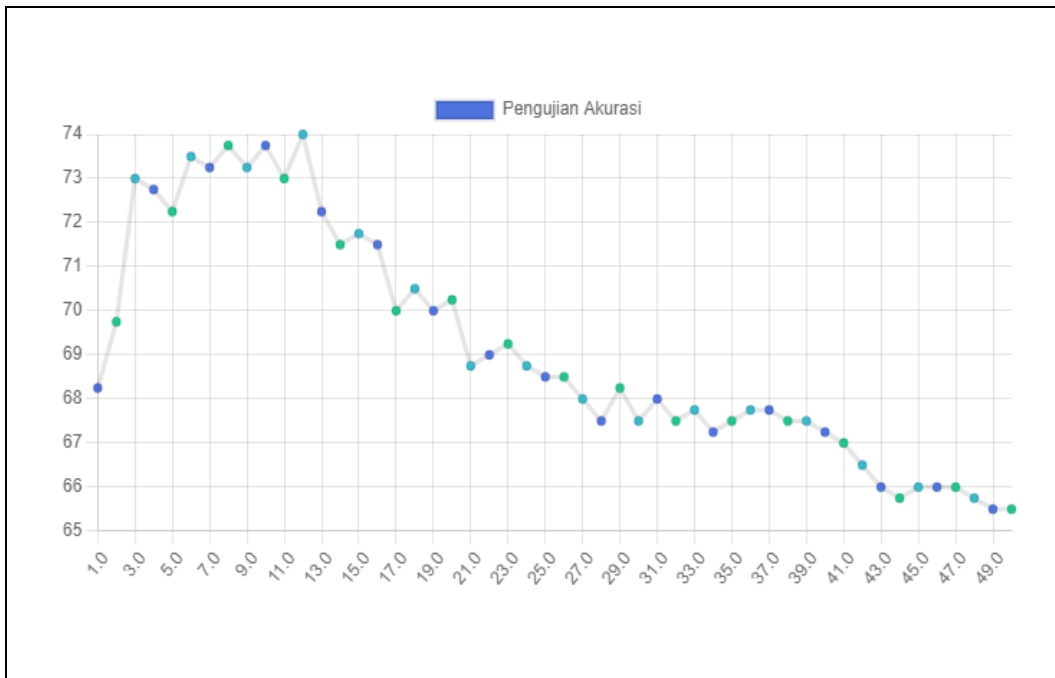
Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 600 data *training* dan 400 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.2 berikut.

Tabel 4.2 Hasil Nilai k Skenario Data *Training* 60% dan Data *Testing* 40% Menggunakan Pelabelan Manual

Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l	Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l
1	68.25	58.74	57.94	26	68.50	64.96	46.16
2	69.75	63.26	58.66	27	68,00	65.27	44.98
3	73,00	63.70	61.71	28	67.50	66.74	44.06
4	72.75	64.73	58.58	29	68.25	69.07	44.99
5	72.25	64.57	58.97	30	67.50	65.60	43.68
6	73.50	66.71	59.37	31	68,00	68.25	44.08
7	73.25	66.23	57.93	32	67.50	67.80	43.16
8	73.75	67.49	56.36	33	67.75	68.78	43.30
9	73.25	66,00	56.50	34	67.25	67.17	42.50
10	73.75	67.14	56.64	35	67.50	68.29	42.63
11	73,00	66.82	55.20	36	67.75	69.55	42.77
12	74,00	69.47	55.87	37	67.75	70.11	42.50
13	72.25	66.89	53.09	38	67.50	67.73	41.98
14	71.50	64.74	52.31	39	67.50	67.73	41.98
15	71.75	65.79	52.30	40	67.25	71.01	41.46
16	71.50	66.52	52.04	41	67,00	71.94	40.94
17	70,00	62.10	49.68	42	66.50	72.02	39.90
18	70.50	63.07	50.07	43	66,00	70.24	39.10
19	70,00	62.38	49.29	44	65.75	68.99	38.58
20	70.25	62.69	49.55	45	66,00	70.24	39.10
21	68.75	60.59	46.67	46	66,00	67.52	39.10

22	69,00	63.26	46.55	47	66,00	67.52	39.23
23	69.25	63.63	47.34	48	65.75	67.91	38.44
24	68.75	62.58	46.30	49	65.50	66.26	38.05
25	68.50	63.37	45.90	50	65.50	66.26	38.05

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 60% dan data *testing* 40% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 12 dengan tingkat *accuracy* sebesar 74,00% yang terlihat pada Tabel 4.2 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.14 berikut.



Gambar 4.14 Grafik Nilai k Skenario Data *Training* 60% dan Data *Testing* 40% Menggunakan Pelabelan Manual

Terlihat pada Gambar 4.14 saat data di uji sebanyak 50 kali pada jumlah $k = 12$ menunjukkan *accuracy* tertinggi, pada iterasi selanjutnya nilai *accuracy* mengalami penurunan dan tidak mengalami kenaikan yang signifikan melebihi iterasi sebelumnya.

c. Skenario Data *Training* 65% dan Data *Testing* 35% Menggunakan Pelabelan Manual

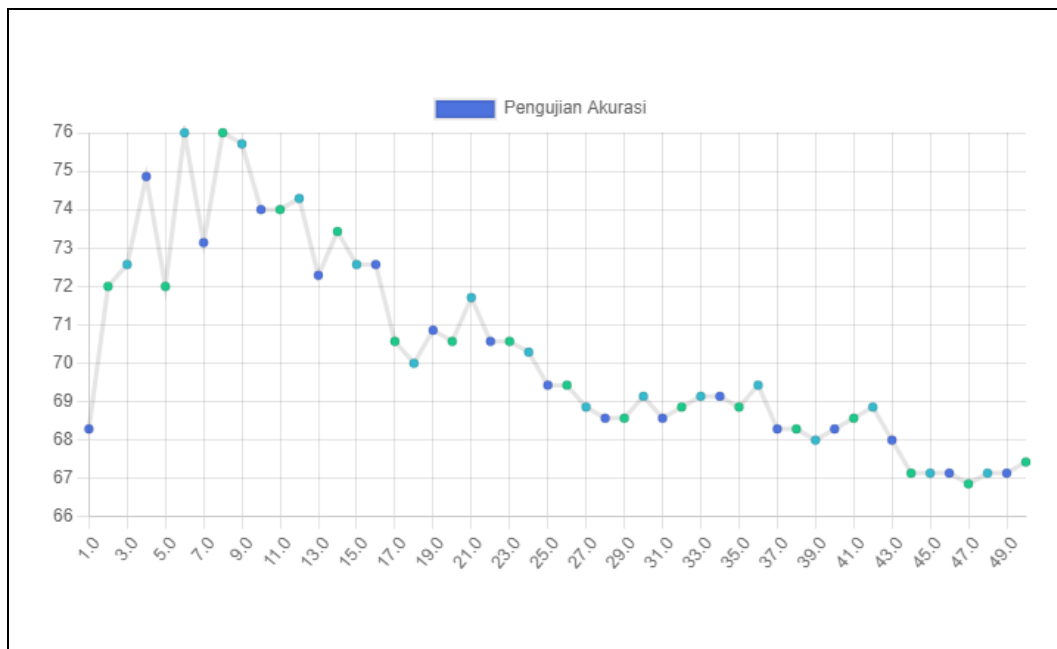
Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 650 data *training* dan 350 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.3 berikut.

Tabel 4.3 Hasil Nilai k Skenario Data *Training* 65% dan Data *Testing* 35% Menggunakan Pelabelan Manual

Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l	Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l
1	68.29	58.77	57.25	26	69.43	63.71	47.31
2	72,00	66.26	60.38	27	68.86	63.03	46.25
3	72.57	62.88	61.15	28	68.57	64.66	45.64
4	74.86	68.34	61.13	29	68.57	63.92	45.03
5	72,00	62.83	58.11	30	69.14	66.93	45.94
6	76,00	69.97	61.73	31	68.57	65.03	44.58
7	73.14	63.90	57.33	32	68.86	68.37	45.03
8	76,00	70.64	60.81	33	69.14	70.49	44.87
9	75.71	70.26	59.60	34	69.14	72.69	44.72
10	74,00	67.16	57.79	35	68.86	69.23	44.72
11	74,00	66.99	56.87	36	69.43	71.30	45.33
12	74.29	67.54	57.33	37	68.29	67.39	43.50
13	72.29	64.79	53.07	38	68.29	68.42	43.20
14	73.43	69.13	54.59	39	68,00	66.19	42.75
15	72.57	65.88	53.07	40	68.29	70.81	43.05
16	72.57	66.69	53.38	41	68.57	73.75	43.20
17	70.57	62.66	50.79	42	68.86	72.26	43.81
18	70,00	59.71	49.73	43	68,00	69.92	42.14
19	70.86	63.55	50.80	44	67.14	68.99	40.62
20	70.57	61.98	50.50	45	67.14	68.99	40.62
21	71.71	65.38	52.31	46	67.14	71.50	40.47

22	70.57	64.11	49.12	47	66.86	70.52	40.02
23	70.57	64.09	49.43	48	67.14	73.15	40.17
24	70.29	62.64	48.97	49	67.14	73.15	40.17
25	69.43	62.57	47.15	50	67.43	74.14	40.77

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 65% dan data *testing* 35% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 6 dan 8 dengan tingkat *accuracy* sebesar 76,00% yang terlihat pada Tabel 4.3 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.15 berikut.



Gambar 4.15 Grafik Nilai k Skenario Data *Training* 65% dan Data *Testing* 35% Menggunakan Pelabelan Manual

Terlihat pada Gambar 4.15 saat data di uji sebanyak 50 kali terdapat 2 nilai k berbeda yang mempunyai nilai *accuracy* tertinggi yaitu pada jumlah $k = 6$ dan $k = 8$.

d. Skenario Data *Training* 25% dan Data *Testing* 75% Menggunakan Pelabelan Manual

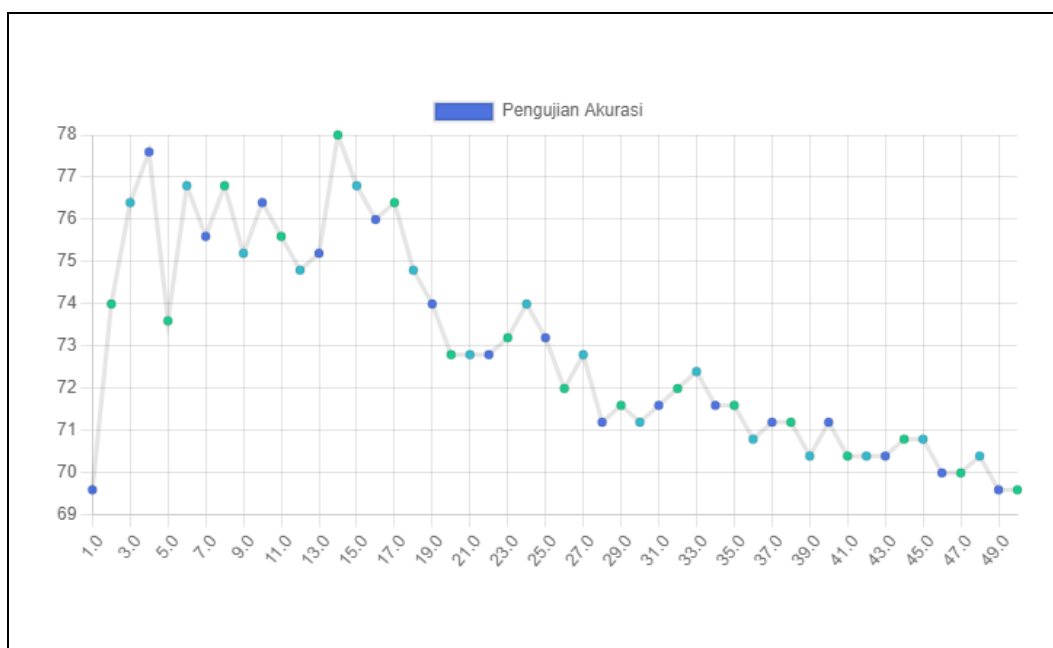
Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 250 data *training* dan 750 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.4 berikut.

Tabel 4.4 Hasil Nilai k Skenario Data *Training* 75% dan Data *Testing* 25% Menggunakan Pelabelan Manual

Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l	Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l
1	69.60	57.58	56.35	26	72,00	64.95	47.30
2	74,00	66.47	60.49	27	72.80	66.92	48.98
3	76.40	65.82	63.33	28	71.20	62.72	45.63
4	77.60	68.96	62.19	29	71.60	63.56	46.64
5	73.60	62.34	57.51	30	71.20	62.96	46.09
6	76.80	68.28	60.39	31	71.60	63.80	47.10
7	75.60	66.88	58.51	32	72,00	67.08	47.43
8	76.80	69.97	59.45	33	72.40	67.83	48.44
9	75.20	66.41	56.56	34	71.60	66.28	45.95
10	76.40	69.37	60.06	35	71.60	68.15	46.76
11	75.60	67.84	58.73	36	70.80	66.22	45.08
12	74.80	66.79	56.71	37	71.20	70.30	45.75
13	75.20	68.31	57.84	38	71.20	70.17	44.82
14	78,00	74.20	60.52	39	70.40	67.43	42.33
15	76.80	72.01	58.64	40	71.20	68.69	44.35
16	76,00	73.26	56.03	41	70.40	66.82	43.14
17	76.40	72.59	56.70	42	70.40	64.45	42.67
18	74.80	68.13	54.84	43	70.40	64.45	42.67
19	74,00	67.01	52.82	44	70.80	66.40	43.34
20	72.80	62.76	50.82	45	70.80	69.54	43.34
21	72.80	62.76	50.82	46	70,00	67.95	41.32
22	72.80	64.41	50.01	47	70,00	66.89	41.66
23	73.20	65.27	51.49	48	70.40	68.84	42.33
24	74,00	67.10	52.70	49	69.60	64.48	40.99

25	73.20	66.12	50.68	50	69.60	64.48	40.99
----	-------	-------	-------	----	-------	-------	-------

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 75% dan data *testing* 25% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 14 dengan tingkat *accuracy* sebesar 78,00% yang terlihat pada Tabel 4.4 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.16 berikut.



Gambar 4.16 Grafik Nilai k Skenario Data *Training* 75% dan Data *Testing* 25% Menggunakan Pelabelan Manual

Terlihat pada Gambar 4.16 saat data di uji sebanyak 50 kali pada jumlah $k = 14$ menunjukkan *accuracy* tertinggi, pada iterasi selanjutnya nilai *accuracy* mengalami penurunan dan tidak mengalami kenaikan yang signifikan melebihi iterasi sebelumnya.

e. Skenario Data *Training* 20% dan Data *Testing* 80% Menggunakan Pelabelan Manual

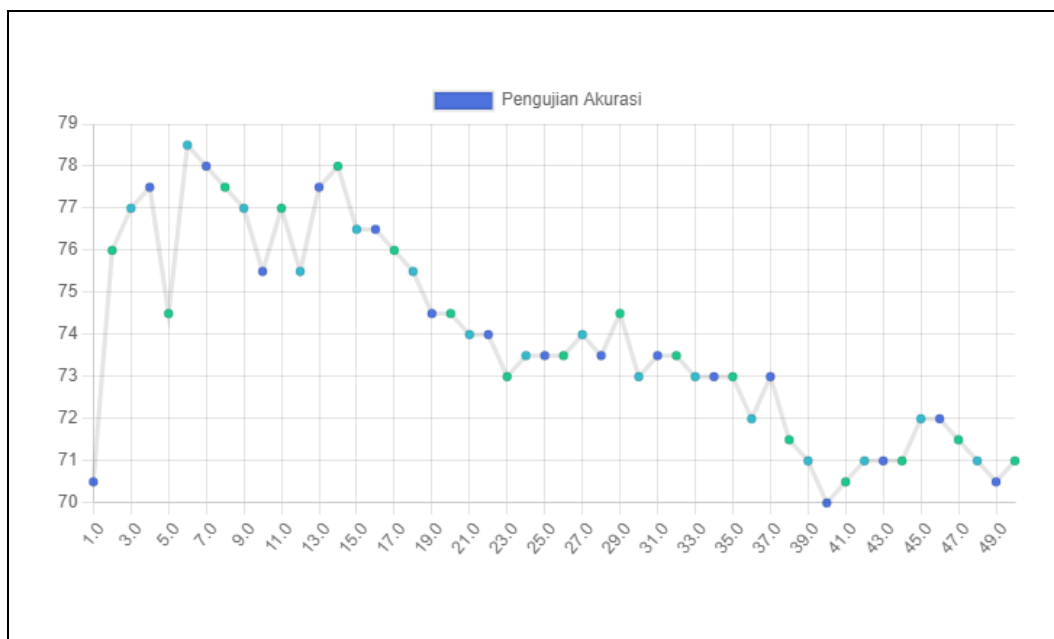
Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 200 data *training* dan 800 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.4 berikut.

Tabel 4.5 Hasil Nilai k Skenario Data *Training* 80% dan Data *Testing* 20% Menggunakan Pelabelan Manual

Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l	Jumlah k terdeka t	<i>Accurac</i> y	<i>Precissio</i> n	<i>Recal</i> l
1	70.50	58.24	55.94	26	73.50	64.18	48.54
2	76,00	68.48	61.29	27	74,00	65.18	49.77
3	77,00	64.49	60.78	28	73.50	64.18	48.54
4	77.50	69.25	60.37	29	74.50	66.06	51
5	74.50	63.08	58.56	30	73,00	63.02	48.29
6	78.50	70.57	60.20	31	73.50	64.02	49.53
7	78,00	70.42	59.62	32	73.50	64.28	49.19
8	77.50	69.84	58.40	33	73,00	62.60	48.62
9	77,00	68.85	57.49	34	73,00	62.60	48.62
10	75.50	66.67	55.78	35	73,00	62.60	48.62
11	77,00	69.96	57.82	36	72,00	60.45	46.16
12	75.50	67.37	54.45	37	73,00	62.75	47.63
13	77.50	70.84	57.73	38	71.50	58.61	44.26
14	78,00	71.80	57.66	39	71,00	58.23	43.03
15	76.50	69.31	54.94	40	70,00	54.23	41.23
16	76.50	71.17	54.94	41	70.50	55.19	41.47
17	76,00	67.76	54.04	42	71,00	59.20	41.72
18	75.50	68.14	52.81	43	71,00	59.20	41.72
19	74.50	65.69	51.66	44	71,00	59.20	41.72
20	74.50	65.16	51.34	45	72,00	62.83	44.19
21	74,00	64.81	49.77	46	72,00	62.83	44.19
22	74,00	64.74	50.43	47	71.50	61.78	42.95
23	73,00	62.05	48.62	48	71,00	60.45	41.72
24	73.50	63.75	48.87	49	70.50	58.74	40.48

25	73.50	63.75	48.87	50	71,00	60.57	40.73
----	-------	-------	-------	----	-------	-------	-------

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 80% dan data *testing* 20% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 6 dengan tingkat *accuracy* sebesar 78,50% yang terlihat pada Tabel 4.5 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.17 berikut.



Gambar 4.17 Grafik Nilai k Skenario Data *Training* 80% dan Data *Testing* 20% Menggunakan Pelabelan Manual

Terlihat pada Gambar 4.17 saat data di uji sebanyak 50 kali pada jumlah $k = 6$ menunjukan *accuracy* tertinggi, pada iterasi selanjutnya nilai *accuracy* mengalami penurunan dan tidak mengalami kenaikan yang signifikan melebihi iterasi sebelumnya.

Setelah melakukan percobaan nilai k sebanyak 50 kali dengan pembagian data yang berbeda seperti pada Tabel 3.12 didapatkan hasil terbaik yaitu yang mempunyai nilai *accuracy* tertinggi pada setiap skenario percobaan seperti pada Tabel 4.6 sebagai berikut.

Tabel 4.6 Tabel Hasil Skenario Percobaan Menggunakan Pelabelan Manual

<i>Data Training</i>	<i>Data Testing</i>	Nilai k terbaik	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
60%	40%	12	74,00	69.47	55.87
65%	35%	6	76,00	69.97	61.73
		8	76,00	70.64	60.81
70%	30%	6	76,33	69,29	60,72
75%	25%	14	78,00	74.20	60.52
80%	20%	6	78.50	70.57	60.20

Berdasarkan skenario pembagian data yang dilakukan didapatkan hasil untuk performa model yang paling baik dengan melihat nilai *accuracy* tertinggi yaitu pada pembagian jumlah data *training* 80% dan data *testing* 20% dengan nilai *accuracy* sebesar 78,50% pada $k = 6$. Hasil Klasifikasi menggunakan model dengan performa yang paling baik berdasarkan pelabelan dataset secara manual dapat dilihat pada Lampiran 1.

4.2.2 Pelabelan Otomatis *sentistrength_id*

Pada percobaan sebelumnya pelabelan dataset dilakukan secara manual kemudian diujikan untuk melihat performa model yang paling baik, selanjutnya model kembali diujikan dengan pelabelan secara otomatis menggunakan program python *sentistrength_id*, *sentistrength_id* adalah program deteksi untuk menghitung kekuatan sentimen pada kalimat atau kata Bahasa Indonesia. Pengujian ini dilakukan sebagai perbandingan hasil klasifikasi suatu dataset yang diberi label manual dengan hasil klasifikasi yang diberi label otomatis

berdasarkan daftar kata positif dan negatif pada program python `sentistrengt_id`. Dengan pelabelan otomatis model akan di uji kembali sama seperti Tabel 3.12, setiap skenario dilakukan percobaan sebanyak 50 kali dengan jumlah tetangga terdekat(k) dimulai dari $k = 1$ sampai dengan $k = 50$.

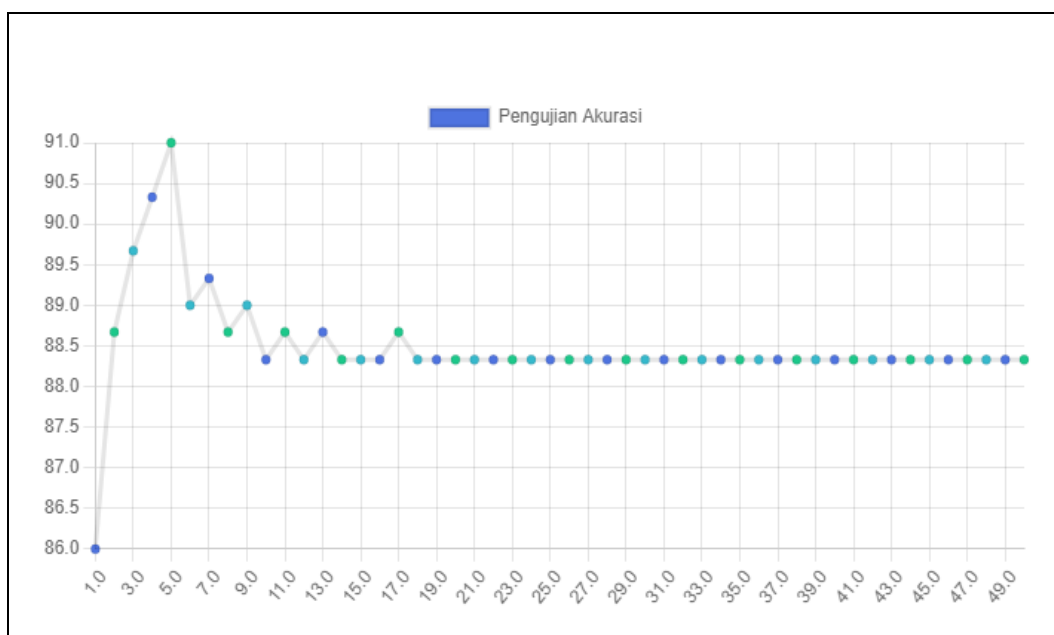
a. Skenario Data *Training* 70% dan Data *Testing* 30% Menggunakan Pelabelan Otomatis

Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 700 data *training* dan 300 data *testing* dengan uji performa model sebanyak $k = 50$ pada klasifikasi berita Vaksinasi Berita COVID-19 seperti pada Tabel 4.7 berikut.

Tabel 4.7 Hasil Nilai k Skenario Data *Training* 70% dan Data *Testing* 30% Menggunakan Otomatis

Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
1	86,00	50.10	49.21	26	88.33	29.44	33.33
2	88.67	63.52	47.36	27	88.33	29.44	33.33
3	89.67	67.17	50.59	28	88.33	29.44	33.33
4	90.33	92.04	47.99	29	88.33	29.44	33.33
5	91,00	76.75	50.14	30	88.33	29.44	33.33
6	89,00	62.98	35.48	31	88.33	29.44	33.33
7	89.33	63.08	36.56	32	88.33	29.44	33.33
8	88.67	62.88	34.41	33	88.33	29.44	33.33
9	89,00	62.98	35.48	34	88.33	29.44	33.33
10	88.33	29.44	33.33	35	88.33	29.44	33.33
11	88.67	62.88	34.41	36	88.33	29.44	33.33
12	88.33	29.44	33.33	37	88.33	29.44	33.33
13	88.67	62.88	34.41	38	88.33	29.44	33.33
14	88.33	29.44	33.33	39	88.33	29.44	33.33
15	88.33	29.44	33.33	40	88.33	29.44	33.33
16	88.33	29.44	33.33	41	88.33	29.44	33.33
17	88.67	62.88	34.41	42	88.33	29.44	33.33
18	88.33	29.44	33.33	43	88.33	29.44	33.33
19	88.33	29.44	33.33	44	88.33	29.44	33.33
20	88.33	29.44	33.33	45	88.33	29.44	33.33
21	88.33	29.44	33.33	46	88.33	29.44	33.33
22	88.33	29.44	33.33	47	88.33	29.44	33.33
23	88.33	29.44	33.33	48	88.33	29.44	33.33
24	88.33	29.44	33.33	49	88.33	29.44	33.33
25	88.33	29.44	33.33	50	88.33	29.44	33.33

Dari pengujian yang dilakukan dengan 50 kali untuk menentukan jumlah tetangga terdekat yang optimal dan *confusion matrix* dengan melihat nilai *accuracy*, *precision* dan *recall* maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 5 dengan tingkat *accuracy* sebesar 91,00% yang terlihat pada Tabel 4.7 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.18 berikut.



Gambar 4.18 Grafik Nilai k Skenario Data *Training* 70% dan Data *Testing* 30% Menggunakan Pelabelan Otomatis

Terlihat pada Gambar 4.18 saat data di uji sebanyak 50 kali jumlah $k = 5$ menunjukkan *accuracy* tertinggi, iterasi selanjutnya nilai k pada *accuracy* bergerak fluktuatif dan mengalami penurunan, pada $k = 18$ nilai *accuracy* bergerak stagnan atau tidak mengalami kenaikan dan penurunan.

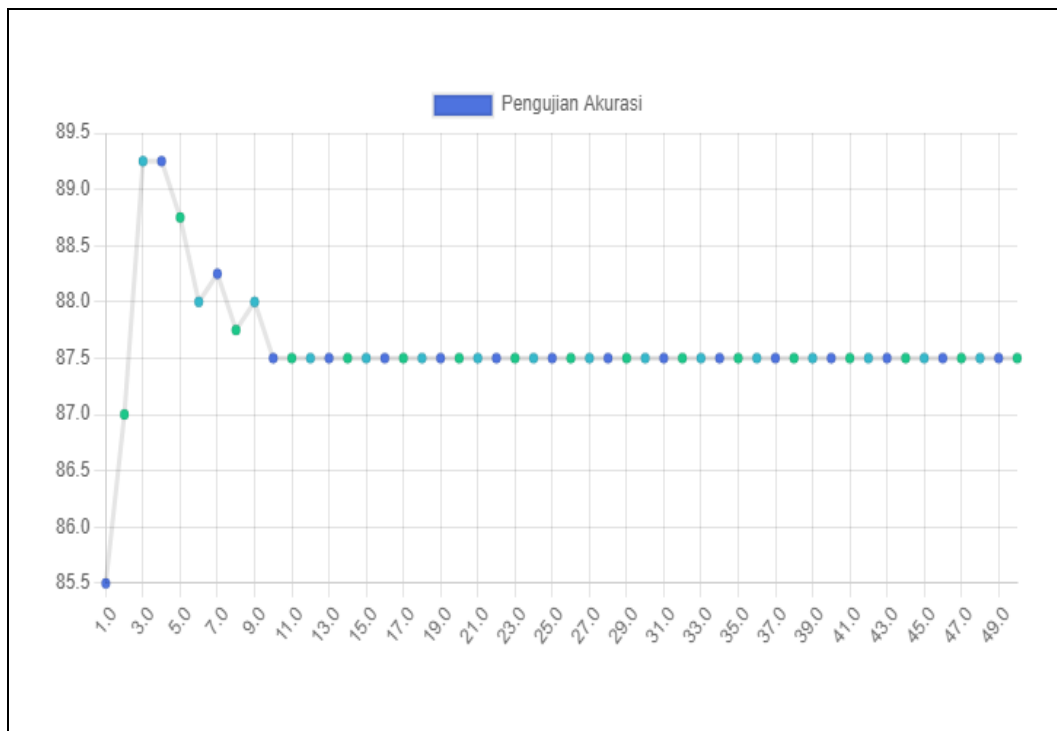
b. Skenario Data *Training* 60% dan Data *Testing* 40% Menggunakan Pelabelan Otomatis

Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 600 data *training* dan 400 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.8 berikut.

Tabel 4.8 Hasil Nilai k Skenario Data *Training* 60% dan Data *Testing* 40% Menggunakan Pelabelan Otomatis

Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
1	85.50	51.86	46.23	26	87.50	29.17	33.33
2	87,00	59.24	43.23	27	87.50	29.17	33.33
3	89.25	74.55	48.39	28	87.50	29.17	33.33
4	89.25	96.35	44.81	29	87.50	29.17	33.33
5	88.75	71.41	41.72	30	87.50	29.17	33.33
6	88,00	54.45	35.68	31	87.50	29.17	33.33
7	88.25	56.12	36.49	32	87.50	29.17	33.33
8	87.75	62.57	34.15	33	87.50	29.17	33.33
9	88,00	62.65	34.96	34	87.50	29.17	33.33
10	87.50	29.17	33.33	35	87.50	29.17	33.33
11	87.50	29.17	33.33	36	87.50	29.17	33.33
12	87.50	29.17	33.33	37	87.50	29.17	33.33
13	87.50	29.17	33.33	38	87.50	29.17	33.33
14	87.50	29.17	33.33	39	87.50	29.17	33.33
15	87.50	29.17	33.33	40	87.50	29.17	33.33
16	87.50	29.17	33.33	41	87.50	29.17	33.33
17	87.50	29.17	33.33	42	87.50	29.17	33.33
18	87.50	29.17	33.33	43	87.50	29.17	33.33
19	87.50	29.17	33.33	44	87.50	29.17	33.33
20	87.50	29.17	33.33	45	87.50	29.17	33.33
21	87.50	29.17	33.33	46	87.50	29.17	33.33
22	87.50	29.17	33.33	47	87.50	29.17	33.33
23	87.50	29.17	33.33	48	87.50	29.17	33.33
24	87.50	29.17	33.33	49	87.50	29.17	33.33
25	87.50	29.17	33.33	50	87.50	29.17	33.33

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 60% dan data *testing* 40% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 3 dan 4 dengan tingkat *accuracy* sebesar 89,25% yang terlihat pada Tabel 4.8 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.19 berikut.



Gambar 4.19 Grafik Nilai k Skenario Data *Training* 60% dan Data *Testing* 40% Menggunakan Pelabelan Otomatis

Terlihat pada Gambar 4.19 saat data di uji sebanyak 50 kali pada jumlah $k = 3$ dan $k = 4$ menunjukkan *accuracy* tertinggi, iterasi selanjutnya nilai k pada *accuracy* bergerak fluktuatif dan mengalami penurunan, pada $k = 10$ nilai *accuracy* bergerak stagnan atau tidak mengalami kenaikan dan penurunan.

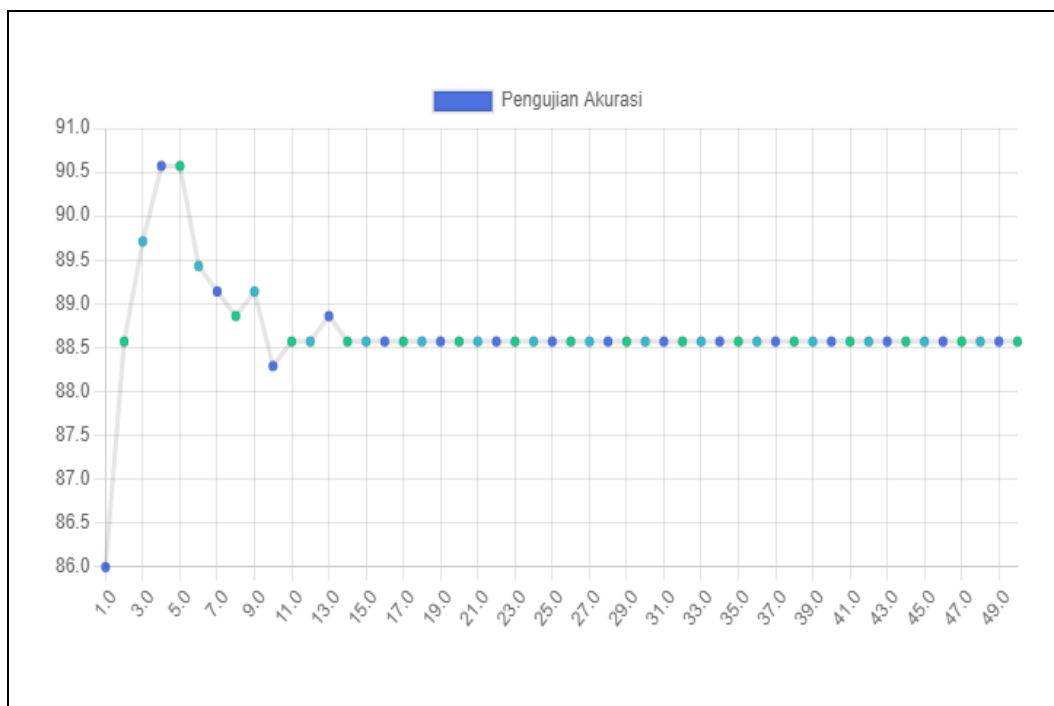
c. Skenario Data *Training* 65% dan Data *Testing* 35% Menggunakan Pelabelan Otomatis

Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 650 data *training* dan 350 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.9 berikut.

Tabel 4.9 Hasil Nilai k Skenario Data *Training* 65% dan Data *Testing* 35% Menggunakan Pelabelan Otomatis

Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
1	86,00	50.18	45.14	26	88.57	29.52	33.33
2	88.57	59.91	43.40	27	88.57	29.52	33.33
3	89.71	67.10	46.54	28	88.57	29.52	33.33
4	90.57	96.79	44.16	29	88.57	29.52	33.33
5	90.57	76.13	45.06	30	88.57	29.52	33.33
6	89.43	63.20	36.36	31	88.57	29.52	33.33
7	89.14	54.77	36.26	32	88.57	29.52	33.33
8	88.86	62.94	34.34	33	88.57	29.52	33.33
9	89.14	63.03	35.35	34	88.57	29.52	33.33
10	88.29	29.51	33.23	35	88.57	29.52	33.33
11	88.57	29.52	33.33	36	88.57	29.52	33.33
12	88.57	29.52	33.33	37	88.57	29.52	33.33
13	88.86	62.94	34.34	38	88.57	29.52	33.33
14	88.57	29.52	33.33	39	88.57	29.52	33.33
15	88.57	29.52	33.33	40	88.57	29.52	33.33
16	88.57	29.52	33.33	41	88.57	29.52	33.33
17	88.57	29.52	33.33	42	88.57	29.52	33.33
18	88.57	29.52	33.33	43	88.57	29.52	33.33
19	88.57	29.52	33.33	44	88.57	29.52	33.33
20	88.57	29.52	33.33	45	88.57	29.52	33.33
21	88.57	29.52	33.33	46	88.57	29.52	33.33
22	88.57	29.52	33.33	47	88.57	29.52	33.33
23	88.57	29.52	33.33	48	88.57	29.52	33.33
24	88.57	29.52	33.33	49	88.57	29.52	33.33
25	88.57	29.52	33.33	50	88.57	29.52	33.33

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 65% dan data *testing* 35% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 4 dan 5 dengan tingkat *accuracy* sebesar 90,57 yang terlihat pada Tabel 4.9 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.20 berikut.



Gambar 4.20 Grafik Nilai k Skenario Data *Training* 65% dan Data *Testing* 35% Menggunakan Pelabelan Otomatis

Terlihat pada Gambar 4.20 saat data di uji sebanyak 50 kali pada jumlah $k = 4$ dan $k = 5$ menunjukkan *accuracy* tertinggi, iterasi selanjutnya nilai k pada *accuracy* bergerak fluktuatif dan mengalami penurunan, pada $k = 14$ nilai *accuracy* bergerak stagnan atau tidak mengalami kenaikan dan penurunan.

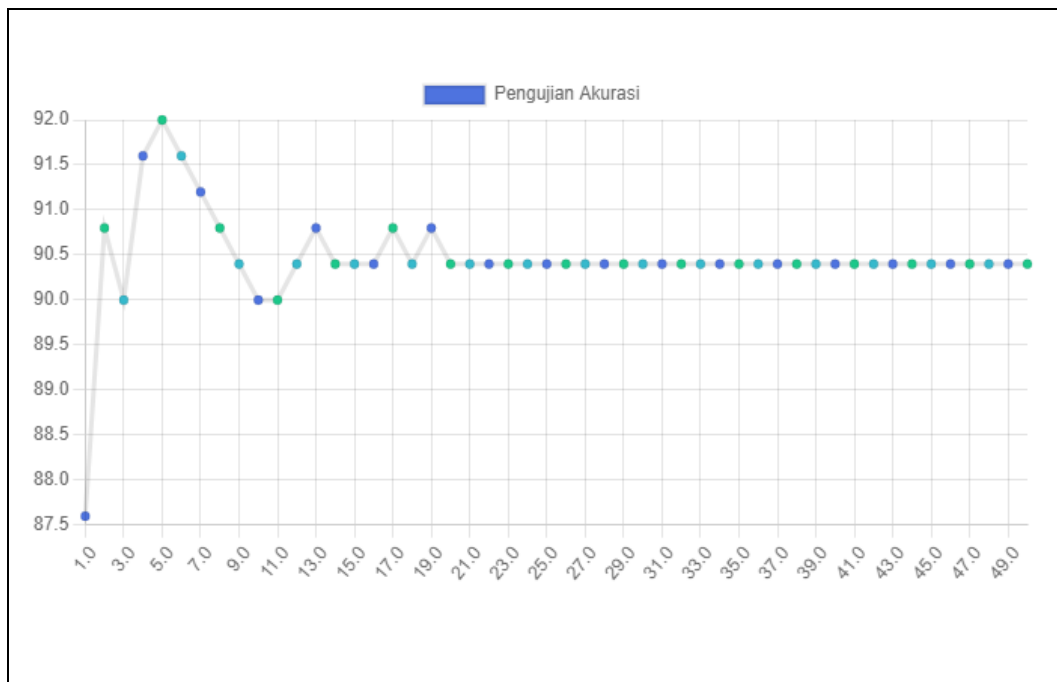
d. Skenario Data *Training* 25% dan Data *Testing* 75% Menggunakan Pelabelan Otomatis

Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 250 data *training* dan 750 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.10 berikut.

Tabel 4.10 Hasil Nilai k Skenario Data *Training* 75% dan Data *Testing* 25% Menggunakan Pelabelan Otomatis

Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	Jumlah k terdekat	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
1	87.60	50.44	49.60	26	90.40	30.13	33.33
2	90.80	64.29	49.26	27	90.40	30.13	33.33
3	90,00	61.49	48.97	28	90.40	30.13	33.33
4	91.60	86.28	48.04	29	90.40	30.13	33.33
5	92,00	74.20	48.19	30	90.40	30.13	33.33
6	91.60	63.83	38.33	31	90.40	30.13	33.33
7	91.20	55.49	38.19	32	90.40	30.13	33.33
8	90.80	63.59	35,00	33	90.40	30.13	33.33
9	90.40	46.91	34.85	34	90.40	30.13	33.33
10	90,00	30.12	33.19	35	90.40	30.13	33.33
11	90,00	41.34	34.71	36	90.40	30.13	33.33
12	90.40	30.13	33.33	37	90.40	30.13	33.33
13	90.80	63.59	35,00	38	90.40	30.13	33.33
14	90.40	30.13	33.33	39	90.40	30.13	33.33
15	90.40	46.91	34.85	40	90.40	30.13	33.33
16	90.40	30.13	33.33	41	90.40	30.13	33.33
17	90.80	63.59	35,00	42	90.40	30.13	33.33
18	90.40	30.13	33.33	43	90.40	30.13	33.33
19	90.80	63.59	35,00	44	90.40	30.13	33.33
20	90.40	30.13	33.33	45	90.40	30.13	33.33
21	90.40	30.13	33.33	46	90.40	30.13	33.33
22	90.40	30.13	33.33	47	90.40	30.13	33.33
23	90.40	30.13	33.33	48	90.40	30.13	33.33
24	90.40	30.13	33.33	49	90.40	30.13	33.33
25	90.40	30.13	33.33	50	90.40	30.13	33.33

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 75% dan data *testing* 25% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 5 dengan tingkat *accuracy* sebesar 92,00% yang terlihat pada Tabel 4.10 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.21 berikut.



Gambar 4.21 Grafik Nilai k Skenario Data *Training* 75% dan Data *Testing* 25% Menggunakan Pelabelan Otomatis

Terlihat pada Gambar 4.21 saat data di uji sebanyak 50 kali pada jumlah $k = 5$ menunjukkan *accuracy* tertinggi, pada $k = 21$ nilai *accuracy* bergerak stagnan atau tidak mengalami kenaikan dan penurunan.

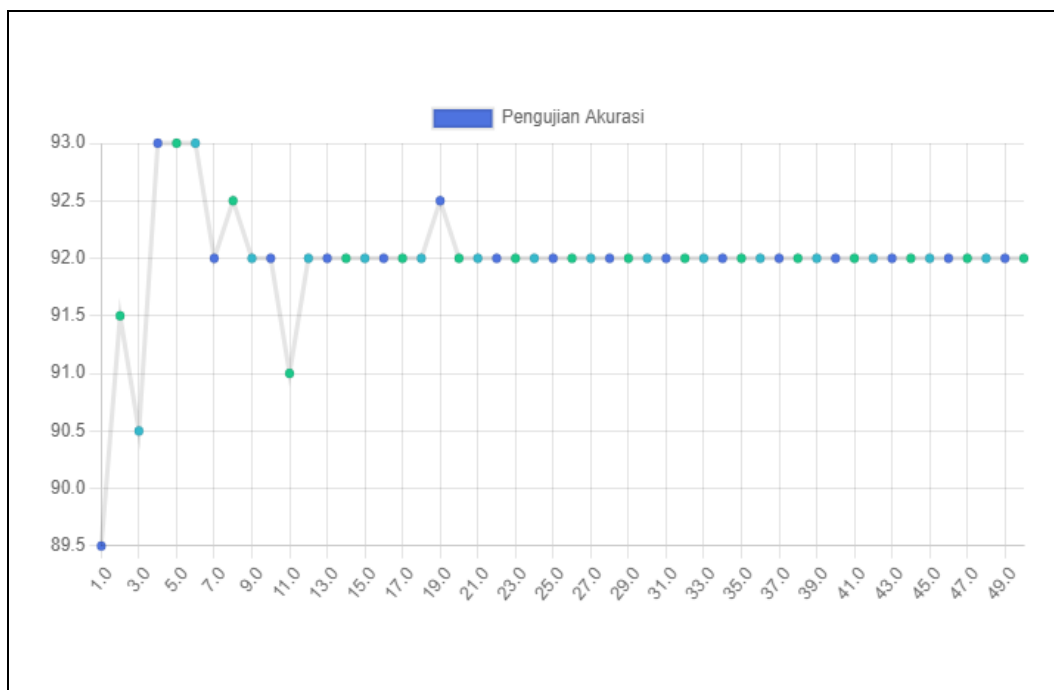
e. Skenario Data *Training* 20% dan Data *Testing* 80% Menggunakan Pelabelan Otomatis

Berikut ini adalah hasil salah satu skenario pembagian data yang terdiri dari 200 data *training* dan 800 data *testing* dengan uji performa model sebanyak $k = 50$ seperti pada Tabel 4.11 berikut.

Tabel 4.11 Hasil Nilai k Skenario Data *Training* 80% dan Data *Testing* 20% Menggunakan Pelabelan Otomatis

Jumlah k terdekat	<i>Accuracy</i>	<i>Precission</i>	<i>Recall</i>	Jumlah k terdekat	<i>Accuracy</i>	<i>Precission</i>	<i>Recall</i>
1	89.50	43.26	44.34	26	92,00	30.67	33.33
2	91.50	56.09	40.30	27	92,00	30.67	33.33
3	90.50	43.56	39.94	28	92,00	30.67	33.33
4	93,00	56.12	40.84	29	92,00	30.67	33.33
5	93,00	56.12	40.84	30	92,00	30.67	33.33
6	93,00	64.31	38.46	31	92,00	30.67	33.33
7	92,00	47.62	38.10	32	92,00	30.67	33.33
8	92.50	64.15	35.90	33	92,00	30.67	33.33
9	92,00	47.47	35.72	34	92,00	30.67	33.33
10	92,00	30.67	33.33	35	92,00	30.67	33.33
11	91,00	30.64	32.97	36	92,00	30.67	33.33
12	92,00	30.67	33.33	37	92,00	30.67	33.33
13	92,00	30.67	33.33	38	92,00	30.67	33.33
14	92,00	30.67	33.33	39	92,00	30.67	33.33
15	92,00	30.67	33.33	40	92,00	30.67	33.33
16	92,00	30.67	33.33	41	92,00	30.67	33.33
17	92,00	30.67	33.33	42	92,00	30.67	33.33
18	92,00	30.67	33.33	43	92,00	30.67	33.33
19	92.50	64.15	35.90	44	92,00	30.67	33.33
20	92,00	30.67	33.33	45	92,00	30.67	33.33
21	92,00	30.67	33.33	46	92,00	30.67	33.33
22	92,00	30.67	33.33	47	92,00	30.67	33.33
23	92,00	30.67	33.33	48	92,00	30.67	33.33
24	92,00	30.67	33.33	49	92,00	30.67	33.33
25	92,00	30.67	33.33	50	92,00	30.67	33.33

Dari pengujian yang dilakukan dengan 50 kali untuk skenario data *training* 80% dan data *testing* 20% maka dihasilkan jumlah tetangga terdekat atau nilai k yang paling optimal adalah 4, 5 dan 6 dengan tingkat *accuracy* sebesar 93,00% yang terlihat pada Tabel 4.11 diatas. Untuk melihat plot data berdasarkan tingkat *accuracy* yang terjadi tiap iterasinya dapat dilihat pada Gambar 4.22 berikut.



Gambar 4.22 Grafik Nilai k Skenario Data *Training* 80% dan Data *Testing* 20% Menggunakan Pelabelan Otomatis

Terlihat pada Gambar 4.22 saat data di uji sebanyak 50 kali pada jumlah $k = 12$, $k = 8$ dan $k = 9$ menunjukan *accuracy* tertinggi, pada $k = 21$ nilai *accuracy* bergerak stagnan atau tidak mengalami kenaikan dan penurunan

Setelah melakukan percobaan nilai k sebanyak 50 kali dengan pembagian data yang berbeda seperti pada Tabel 3.12 didapatkan hasil terbaik yaitu yang mempunyai nilai *accuracy* tertinggi pada setiap skenario percobaan seperti pada Tabel 4.12 sebagai berikut.

Tabel 4.12 Tabel Hasil Skenario Percobaan Menggunakan Pelabelan Otomatis

<i>Data Training</i>	<i>Data Testing</i>	Nilai <i>k</i> terbaik	<i>Accuracy</i>	<i>Precission</i>	<i>Recall</i>
60%	40%	3	89.25	74.55	48.39
		4	89.25	96.35	44.81
65%	35%	4	90,57	96,79	44,16
		5	90,57	76,13	45,06
70%	30%	5	91,00	76.75	50.14
75%	25%	5	92,00	74.20	48.19
80%	20%	6	93.00	56.12	40.84
		8	93,00	56.12	40.84
		9	93.00	64.31	38.56

Berdasarkan skenario pembagian data yang dilakukan didapatkan hasil untuk performa model yang paling baik dengan melihat nilai *accuracy* tertinggi yaitu pada pembagian jumlah data *training* 80% dan data *testing* 20% dengan nilai *accuracy* sebesar 93,00% pada $k = 6$, $k = 8$ dan $k = 9$, karena memiliki nilai *accuracy* tertinggi yang sama pada k berbeda maka dipilih salah satu nilai k untuk dijadikan pembahasan yaitu pada nilai $k = 6$. Klasifikasi menggunakan model dengan performa yang paling baik berdasarkan pelabelan dataset secara otomatis dapat dilihat pada Lampiran 2.

4.2.3 Word Cloud

Word Cloud berguna untuk melihat frekuensi dari sebuah kata yang sering muncul. *Word Cloud* memberikan hasil gambaran yang jelas dan informatif sehingga mudah dipahami. Pada pengujian kali ini akan ditampilkan hasil *dari word cloud* berdasarkan klasifikasi tendensi portal berita dengan model yang memiliki *accuracy* tertinggi yaitu $k = 6$.

4.2.3.1 Word Cloud Pelabelan Manual

A. Word Cloud Tendensi Positif

Pada hasil pengujian kali ini didapatkan sebuah kata-kata yang yang sering muncul dan termasuk kata-kata positif. Kata yang memiliki ukuran besar menandakan frekuensi kemunculan tinggi. Kata-kata yang muncul pada tendensi berita positif adalah “astrazeneca”, “sinovac”, “pfizer”, “persen”, “antibodi”, “efektif”, ”halal” dan lainnya, untuk detail hasil *word cloud* tendensi kata positif dapat dilihat pada gambar 4.23 berikut.

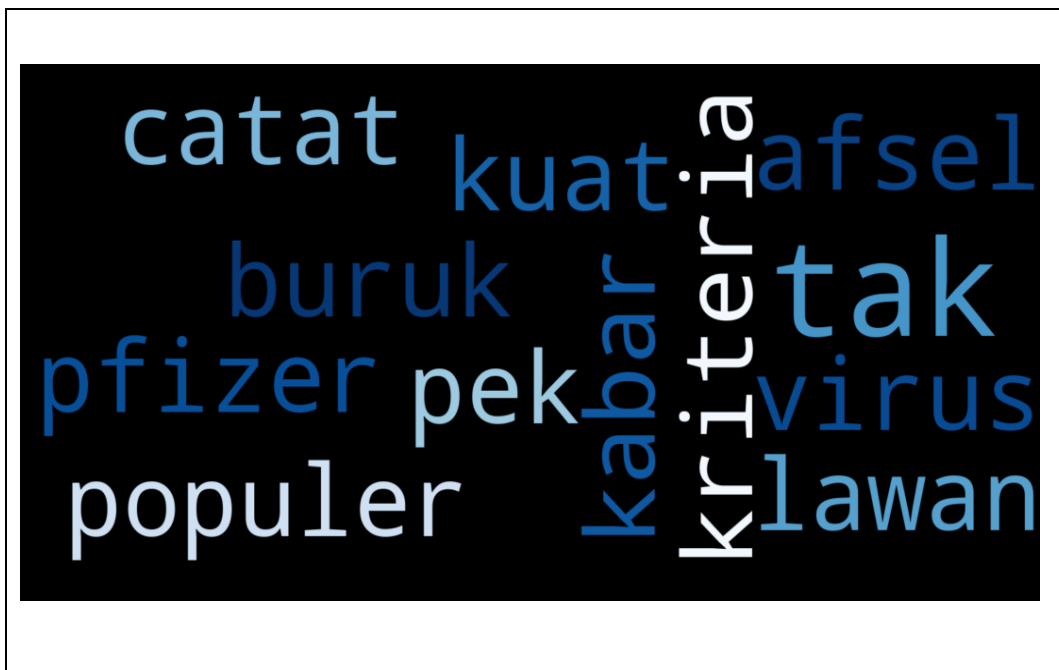


Gambar 4.23 Word Cloud Tendensi Positif Pelabelan Manual

4.2.3.2 *Word Cloud* Pelabelan Otomatis

A. *Word Cloud* Tendensi Positif

Pada hasil pengujian menggunakan pelabelan otomatis didapatkan sebuah kata-kata yang yang sering muncul dan termasuk kata-kata positif. Kata yang memiliki ukuran besar menandakan frekuensi kemunculan tinggi. Kata-kata yang muncul pada tendensi berita positif adalah “kriteria”, “pfizer”, “populer”, “kuat”, “kriteria”, ”lawan” dan lainnya, untuk detail hasil *word cloud* tendensi kata positif dapat dilihat pada gambar 4.26 berikut.



Gambar 4.26 *Word Cloud* Tendensi Positif Pelabelan Otomatis

B. *Word Cloud* Tendensi Negatif

Word cloud untuk tendensi negatif pada pelabelan otomatis tidak ada dikarenakan pada hasil klasifikasi tidak terdapat berita yang masuk kedalam kelas negatif.

C. *Word Cloud* Tendensi Netral

Kata-kata yang sering muncul menggunakan pelabelan otomatis pada tendensi berita netral adalah “astrazeneca”, “sinovac”, “menkes”, “lansia”, “mulai”, “indonesia”, “terima”, “kemenkes”, “aman”, “halal”, “beku darah”, “efek samping” dan lainnya, untuk detail hasil *word cloud* tendensi kata netral dapat dilihat pada gambar 4.27 berikut.



Gambar 4.27 *Word Cloud* Tendensi Netral Pelabelan Otomatis

4.3 Pembahasan Hasil

Uji coba dilakukan menggunakan 5 skenario pembagian data *testing* dan data *training* yang diberikan label secara manual, kemudian setiap skenario dilakukan percobaan sebanyak 50 kali dengan nilai k yang berbeda yaitu 1 sampai 50 seperti pada Tabel 3.12. Berdasarkan hasil uji coba seperti Tabel 4.6 nilai *accuracy* tertinggi yaitu pada pembagian data *training* 80% data dan data *testing* 20% . Dengan hasil uji coba yang tertera pada Gambar 4.17 dan Tabel 4.5 nilai *accuracy* bergerak fluktuatif dengan nilai tertinggi pada $k = 6$. Untuk $k = 7$ sampai $k = 50$ pergerakan nilai *accuracy* masih fluktuatif, akan tetapi tidak lebih tinggi dari $k = 6$ yaitu sebesar 78,50%. Hal tersebut mengindikasikan bahwa semakin tinggi nilai k yang digunakan belum tentu memberikan nilai *accuracy* yang lebih tinggi.

Dengan tingkat *accuracy* sebesar 78,50% hasil klasifikasi yang berhasil diprediksi dengan benar adalah 157 dari 200 data, sementara itu terjadi kesalahan prediksi sebesar 43 data. Jika dilihat pada tabel Lampiran 1, kurangnya tingkat *accuracy* pada data dengan pelabelan manual diakibatkan karena ketidakmampuan model dalam melakukan prediksi terhadap kelas positif dan negatif. Hal ini terlihat saat terdapat judul berita yang bertendensi positif atau negatif rata-rata model memprediksi netral. Contoh pada berita “Ketersediaan vaksin COVID-19 terjaga”, secara informasi itu merupakan berita tendensi positif karena ketersediaan akan vaksin terjaga tetapi sistem memberikan hasil prediksi klasifikasi netral, selanjutnya pada berita “Studi Chile: Suntikan Pertama Sinovac Tak Lindungi dari Covid” secara informasi itu merupakan berita tendensi negatif karena vaksin Sinovac yang digunakan tidak melindungi dari corona sehingga akan menimbulkan keraguan pada masyarakat, tetapi secara prediksi yang diberikan oleh model itu merupakan tendensi netral. Penjelasan lebih detail dapat

dilihat berdasarkan evaluasi model menggunakan *confusion matrix* pada $k = 6$ seperti pada Gambar 4.30 berikut.

	precision	recall	f1-score	support
Negatif	0.67	0.52	0.58	27
Netral	0.82	0.96	0.89	136
Positif	0.63	0.32	0.43	37
accuracy			0.79	200
macro avg	0.71	0.60	0.63	200
weighted avg	0.76	0.79	0.76	200

Gambar 4.28 Confusion Matrix $k = 6$ Pelabelan Manual

Terlihat pada Gambar 4.30 nilai positif maupun negatif untuk *precision* dan *recall* mempunyai nilai yang cukup kecil, berbanding terbalik dengan netral yang mempunyai nilai *precision* dan *recall* yang cukup tinggi. Hal tersebut mengindikasikan tingkat *accuracy* untuk prediksi kelas positif dan negatif masih kurang. Kurangnya tingkat *accuracy* dapat diakibatkan karena tidak imbangnya kelas atau *imbalanced data*, dari total data yang dikumpulkan dan diberi label manual, dimana jumlah kelas positif 222 data, kelas negatif 150 data sedangkan netral 628 data. Perbedaan jumlah label antar kelas yang signifikan membuat model sulit untuk memprediksi hasil yang diujikan. Disamping itu kurangnya tingkat *accuracy* dapat disebabkan karena data yang di klasifikasikan hanya berdasarkan judul tanpa mengikutsertakan isi berita.

Berbeda dari hasil pelabelan manual, untuk dataset yang diberi label secara otomatis menggunakan program *sentistrength_id* berdasarkan *list* atau daftar kata yang mengandung positif dan negatif menghasilkan nilai *accuracy* yang lebih

tinggi yaitu sebesar 93% pada $k = 6$ dengan pembagian data 80% *training* dan 20% *testing*, walaupun memiliki nilai *accuracy* yang cukup tinggi namun hasil *labeling* otomatis dengan menggunakan *sentistrength_id* masih kurang akurat. *Labeling* otomatis menggunakan *sentistrength_id* menghasilkan label dominan netral, hal tersebut dapat dilihat pada hasil *labeling* otomatis di Lampiran 2. Sebagai contoh pada berita dengan judul “Penderita Kanker Diperbolehkan Ikut Vaksinasi Covid-19, Ini Syaratnya” seharusnya masuk kedalam kelas positif karena penderita kanker diperbolehkan mengikuti vaksinasi dengan syarat tertentu, tetapi dari hasil label otomatis yang dihasilkan berita tersebut diberi label netral. Dari Hasil *labeling* yang dominan netral tersebut apabila model melakukan *training* dan *testing* akan didapatkan hasil klasifikasi yang dominan netral juga seperti pada Lampiran 2. Penjelasan lebih detail untuk evaluasi *confusion matrix* hasil klasifikasi menggunakan pelabelan otomatis dapat dilihat pada Gambar 4.31 berikut.

	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	3
Netral	0.93	1.00	0.96	184
Positif	1.00	0.15	0.27	13
accuracy			0.93	200
macro avg	0.64	0.38	0.41	200
weighted avg	0.92	0.93	0.90	200

Gambar 4.29 Confusion Matrix $k = 6$ Menggunakan Pelabelan Otomatis

Terlihat pada Gambar 4.31 nilai positif maupun negatif untuk *precision* dan *recall* rata-rata sangat kecil, bahkan untuk nilai *precision* dan *recall* untuk kelas negatif hanya 0 karena tidak ada hasil prediksi yang masuk kedalam kelas negatif. Dari hasil pelabelan otomatis yang dominan netral terlihat jelas bahwa

nilai *precision* dan *recall* untuk kelas netral sangat tinggi. Hal tersebut mengindikasikan prediksi yang diberikan model cenderung dominan netral. Hal ini dapat diakibatkan karena tidak imbangnya kelas atau imbalanced data, dimana dari total data yang dikumpulkan dan diberi label otomatis jumlah kelas positif 90 data, kelas negatif 20 data sedangkan netral 890 data. Dari hasil analisis secara manual masih banyak pelabelan *sentistrength_id* yang tidak akurat sehingga dapat mengakibatkan ketidakakuratan dalam memprediksi data. Dengan demikian tingkat *accuracy* tinggi pada hasil klasifikasi data dengan pelabelan otomatis tidak merepresentasikan hasil yang sebenarnya dikarenakan tidak akuratnya hasil *labeling* secara otomatis pada data *training* maupun *testing*.

Hasil visualisasi *text* dalam bentuk *word cloud* memberikan sebuah gambaran kata yang sering atau paling banyak muncul dalam pemberitaan di portal berita online tentang vaksinasi COVID-19. Dari hasil *word cloud* kata tendensi berita positif pada Gambar 4.23 kata yang sering muncul di *title* atau judul portal berita online ditandai dengan kata yang ukurannya lebih besar diantaranya yaitu “astrazeneca”, “sinovac”, “pfizer”, “efektif”, “lawan”, “antibodi”, “persen”, “halal”, “kebal”, “varian”, “sembuh” dan “baru”. Hal ini menunjukkan portal berita telah memberikan informasi tentang vaksin COVID-19 mulai dari nama vaksin yang beredar di masyarakat hingga efektifitas penggunaan vaksin terhadap kekebalan tubuh. Disamping itu informasi kehalalan vaksin COVID-19 juga menjadi salah satu isu yang merupakan tendensi positif dalam pemberitaan. Pemberian informasi pada portal berita yang bertendensi positif diharapkan dapat meningkatkan antusiasme masyarakat dalam melakukan vaksinasi COVID-19.

Word cloud juga menampilkan visualisasi kata yang sering muncul pada tendensi berita negatif tentang vaksinasi COVID-19. Dari hasil *word cloud* kata tendensi berita negatif pada Gambar 4.24 kata yang sering muncul di *title* atau judul portal berita online meliputi “beku”, “darah”, “efek”, “samping”, “kanker”, “usai”, “positif”, “suntik”, “astrazeneca”. Hal ini menunjukkan bahwa vaksin yang beredar dan digunakan saat ini belum sepenuhnya berkerja dengan baik,

munculnya kata “efek”, “samping”, “beku”, “darah” juga menjelaskan bahwa vaksin dapat menimbulkan efek negatif yang akan membuat masyarakat mengalami kejadian ikutan pasca imunisasi (KIPI) setelah melakukan vaksinasi COVID-19. Pemberitaan inilah yang membuat masyarakat akhirnya ragu serta tidak antusias melakukan vaksinasi bahkan menolak program tersebut.

Hasil *word cloud* untuk tendensi berita netral menunjukkan bahwa kata yang sering muncul “indonesia”, “menkes”, “distribusi”, “juta”, “mulai”, “terima”. Jika dilihat dari kata yang muncul pada tendensi berita netral terlihat kata-kata tersebut tidak terindikasi sebagai berita yang bertendensi positif maupun negatif.

Berdasarkan hasil *word cloud* menggunakan pelabelan manual kata “astrazeneca” selalu muncul di setiap tendensi berita positif, negatif maupun netral. Hal tersebut mengindikasikan bahwa pemberitaan mengenai vaksin astrazeneca memiliki pemberitaan yang berimbang.

Word cloud pada pelabelan otomatis untuk tendensi kelas positif didapatkan hasil kata yang sering muncul yaitu “kriteria”, “pfizer”, “populer”, “kuat”, “lawan”. Adanya kata “kuat” dan “lawan” mengindikasikan bahwa salah satu isu berita yaitu tentang efektifitas vaksin yang kuat dan dapat melawan COVID-19. Berbeda dari hasil *word cloud* untuk tendensi negatif, pada pelabelan otomatis tidak ada *word cloud* yang dihasilkan karena tidak adanya hasil klasifikasi yang masuk kedalam kelas negatif. Sementara itu *word cloud* untuk tendensi netral menghasilkan kata yang sering muncul yaitu “astrazeneca”, “sinovac”, “menkes”, “lansia”, “mulai”, “indonesia”, “terima”, “kemenkes”, “aman”, “halal”, “beku darah”, “efek samping”, “resiko” dan lainnya. Jika dilihat dari kata yang muncul pada tendensi berita netral terlihat kata-kata tersebut ada yang sebenarnya lebih cocok untuk masuk kedalam kata pada kelas positif, seperti “aman” dan “halal”, serta ada yang lebih cocok untuk masuk kedalam kata pada kelas negatif, seperti “beku darah”, “efek samping”, “resiko”.

Berdasarkan hasil ujicoba dapat disimpulkan bahwa metode *K-Nearest Neighbor* dapat digunakan untuk mengklasifikasikan tendensi berita tentang

vaksinasi COVID-19, dengan tingkat *accuracy* yang berbeda antara pelabelan manual dan pelabelan otomatis menggunakan *sentistrength_id*. Disamping itu aplikasi juga dapat menampilkan *word cloud* dari setiap kelas yang teridentifikasi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian analisis tendensi portal berita menggunakan metode K-Nearest Neighbor, dapat ditarik kesimpulan sebagai berikut :

1. Metode *K-Nearest Neighbor* dapat digunakan untuk mengklasifikasikan tendensi pada portal berita tentang vaksinasi COVID-19 dengan tingkat *accuracy* tertinggi sebesar 78,50% pada $k = 6$ menggunakan pelabelan manual. Sementara itu pada data dengan pelabelan otomatis dengan *sentistrength_id* diperoleh *accuracy* sebesar 93% pada $k = 6$. Adapun data uji coba yang digunakan sebanyak 1000 data dengan 80% data *training* dan 20% data *testing*.
2. Tingkat *accuracy* berada di kisaran 78,50% pada pelabelan manual dapat diakibatkan karena adanya *imbalanced* data yang digunakan serta berita yang di klasifikasikan hanya berdasarkan judul tanpa mengikutsertakan isi berita.
3. Walaupun tingkat *accuracy* pada pelabelan otomatis mencapai 93% tetapi tidak merepresentasikan hasil yang sebenarnya dikarenakan tidak akuratnya hasil labeling secara otomatis pada data *training* maupun *testing*. Kata yang muncul pada tendensi berita netral ada yang lebih cocok untuk masuk kedalam kata pada kelas positif, seperti “aman” dan “halal”, serta ada yang lebih cocok untuk masuk kedalam kata pada kelas negatif, seperti ”beku darah”, ”efek samping”, ”resiko”.
4. Pada pelabelan otomatis model tidak dapat mengidentifikasi kelas negatif dikarenakan jumlah data yang *imbalanced* untuk setiap kelas.
5. Dari hasil klasifikasi yang dilakukan dalam penelitian ini tentang tendensi portal berita terhadap pemberitaan vaksinasi COVID-19, metode ini dapat

mengidentifikasi kata-kata yang dominan, baik dalam pemberitaan positif, negatif maupun netral untuk kelas yang teridentifikasi.

6. Berdasarkan hasil *word cloud* pada pelabelan manual terdapat kata yang muncul, baik di pemberitaan positif, negatif maupun netral yaitu kata “astrazeneca”. Hal tersebut mengindikasikan bahwa pemberitaan mengenai vaksin astrazeneca adalah berimbang.

5.2 Saran

Pada penelitian ini masih memiliki beberapa kekurangan dan keterbatasan. Oleh karena itu, ada beberapa hal yang perlu diperhatikan untuk pengembangan selanjutnya agar bisa memberikan informasi lebih baik :

1. Pelabelan manual dilakukan oleh pakar atau ahli bahasa agar dapat memberikan *labeling* lebih baik.
2. Untuk meningkatkan nilai *accuracy* dari hasil klasifikasi sebaiknya menggunakan data dengan jumlah yang berimbang untuk setiap kelas.
3. Klasifikasi dilakukan tidak hanya berdasarkan judul akan tetapi memperhitungkan isi berita.
4. Adanya penelitian terkait dengan menggunakan metode yang berbeda atau dibandingkan dengan metode klasifikasi lainnya.

DAFTAR PUSTAKA

- Alifian Adexe. (2021). *Belajar Dasar Web Scrapping Bersama Python*.
[Https://Www.Anbidev.Com/](https://www.anbidev.com/). <https://www.anbidev.com/python-web-scrapping/>
- CNN. (2021). *Mengenal Vaksin Sinovac yang Dipakai RI Basmi Covid-19*.
[Www.Cnnindonesia.Com](http://www.cnnindonesia.com).
<https://www.cnnindonesia.com/teknologi/20201230175916-199-588012/mengenal-vaksin-sinovac-yang-dipakai-ri-basmi-covid-19>
- Edukasi, J., Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). *Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes*. 4(2), 113–118.
- Ernawati, S., & Wati, R. (2018). *Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel*. VI(1).
- Fadli, R. (2020). *Begini Cara Kerja Vaksin Mencegah Virus di Dalam Tubuh*.
[Www.Halodoc.Com](http://www.halodoc.com). <https://www.halodoc.com/artikel/begini-cara-kerja-vaksin-mencegah-virus-di-dalam-tubuh>
- Farisa, F. C. (2021). *Jalan Panjang Vaksin Covid-19 Sinovac di RI: Dari Uji Klinis hingga Dapat Izin Penggunaan*. [Www.Kompas.Com](http://www.kompas.com).
<https://nasional.kompas.com/read/2021/01/12/08364271/jalan-panjang-vaksin-covid-19-sinovac-di-ri-dari-uji-klinis-hingga-dapat?page=all>
- Fauzan, A. (2019). *Mengukur Jarak Euclidean: Teori dan Implementasi*

Menggunakan Java. [Www.Kitainformatika.Com.](http://www.kitainformatika.com)
<https://www.kitainformatika.com/2019/10/mengukur-jarak-euclidean-teori-dan.html>

H, A. T. J. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Informatika UPGRIS, 1*, 1–9.

Haryanto, D. J., Muflikhah, L., & Fauzi, M. A. (2018). Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya, 2*(9), 2909–2916.

Informatikalogi. (2016). *Text Preprocessing.* [Www.Informatikalogi.Com.](http://www.informatikalogi.com)
<https://informatikalogi.com/text-preprocessing/>

Informatikalogi. (2017). *Algoritma K-Nearest Neighbor (KNN).* [Www.Informatikalogi.Com.](http://www.informatikalogi.com) <https://informatikalogi.com/algoritma-KNN-k-nearest-neighbor/>

Kompas. (2020). *WHO Resmi Sebut Virus Corona Covid-19 sebagai Pandemi Global.* [Www.Kompas.Com.](http://www.kompas.com)
<https://www.kompas.com/sains/read/2020/03/12/083129823/who-resmi-sebut-virus-corona-covid-19-sebagai-pandemi-global?page=all>

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers.

Mohajon, J. (2020). *Confusion Matrix for Your Multi-Class Machine Learning*

Model. [Www.Towardsdatascience.Com.
https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826](https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826)

Navlani, A. (2018). *KNN Classification using Scikit-learn.* Www.Datacamp.Com.
<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>

Nugroho, E. (2011). *Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp.* Universitas Brawijaya Malang.

Nugroho, K. S. (2019). *Confusion Matrix untuk Evaluasi Model pada Supervised Learning.* Www.Medium.Com. <https://medium.com/@ksnugroho/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>

Ratriani, V. (2020). *Apa itu vaksin dan bagaimana cara kerjanya?* Www.Kesehatan.Kontan.Co.Id. <https://kesehatan.kontan.co.id/news/apa-itu-vaksin-dan-bagaimana-cara-kerjanya?page=all>

Sadida, R., Azkia, M. R., Candra, B. P., Rezeki, N., & Rendy, M. O. C. (2017). Perancangan Sistem Analisis Sentimen Masyarakat pada Sosial Media dan Portal Berita. *Semina Nasional Teknologi Informasi Dan Multimedia 2017, 2015*, 2–7.

Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus

- Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia). *Prosiding SINTAK*, 480–486.
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2018). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor. *Journal Of Intelligent Systems And Computaion*, 43–49.
- Soepomo, P. (2014). Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes. *Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes*, 2(3), 73–83.
<https://doi.org/10.12928/jstie.v2i3.2877>
- Suteja, J. (2021). *Pemerintah Rangkul Media Massa Tingkatkan Kesadaran Masyarakat soal Vaksin Covid-19*. [Www.Beritasatu.Com](http://www.beritasatu.com).
<https://www.beritasatu.com/kesehatan/701843/pemerintah-rangkul-media-massa-tingkatkan-kesadaran-masyarakat-soal-vaksin-covid19>
- Valerisha, A., & Putra, M. A. (2020). Pandemi Global Covid-19 Dan Problematika Negara-Bangsa: Transparansi Data Sebagai Vaksin Socio-Digital? *Jurnal Ilmiah Hubungan Internasional*, 0(0), 131–137.
<https://doi.org/10.26593/jihi.v0i0.3871.131-137>
- WHO. (2020). *Tanya Jawab: Bagaimana cara kerja vaksin?* [Www.Who.Int](http://www.who.int).
<https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public>
- Yulian, E. (2018). Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung. *Jurnal Matematika*

LAMPIRAN

Lampiran 1 Hasil Klasifikasi *Labeling* Dataset Manual

Title	Label_Manual	Hasil_Prediksi
Wamenkes Jelaskan Alasan Pasien COVID-19 Tak Dapat Jatah Vaksin	Netral	Netral
Penderita Kanker Diperbolehkan Ikut Vaksinasi Covid-19, Ini Syaratnya	Positif	Negatif
UPDATE 10 Februari: Kasus Aktif Covid-19 di Indonesia Tercatat 168.416	Netral	Netral
Vaksinasi Covid-19 Tahap Dua Dimulai Besok, Grobogan Dapat Jatah 12.000 Dosis	Netral	Netral
Menkes Disuntik Vaksin Dosis Kedua: Rasanya Jadi Pengen Makan Banyak	Netral	Netral
UPDATE 19 Februari: 668.914 Nakes Sudah Divaksinasi Dosis Kedua, 1.191.031 Dosis Pertama	Netral	Netral
UPDATE 25 Februari: 853.745 Tenaga Kesehatan Sudah Divaksinasi Dosis Kedua, 1.461.920 Dosis Pertama	Netral	Netral
Tak Ragu Disuntik Vaksin Sinovac, dr Tirta: Sudah Terjamin Uji Klinis	Positif	Netral
Vaksinasi COVID-19 ASN Jabar Dimulai, 570 Orang Sudah Disuntik Vaksin Corona	Netral	Netral
Penghentian sementara vaksin AstraZeneca CTMAV547	Netral	Netral

Indonesia Terima Kedatangan Vaksin COVID-19 Tahap Ketujuh	Netral	Netral
14.815.666 jiwa penduduk RI telah menjalani vaksinasi dosis pertama	Netral	Netral
IPB Siap Bantu Uji Praklinis Vaksin Covid-19 Merah Putih	Netral	Netral
Penjelasan Kemenkes soal Vaksin Covid-19 Sinovac Kedaluwarsa 25 Maret 2021	Netral	Netral
Kasus Aktif Menurun, Wamenkes Sebut Bukti Vaksin Covid-19 Kurangi Risiko Sakit Parah	Positif	Positif
Vaksin AstraZeneca, Ma'ruf Amin: Bukan Soal Halal atau Tidak Halal, yang Penting Boleh	Positif	Netral
Kapan Vaksin Merah Putih Dipakai Lawan Covid? Ini Kata BPOM	Netral	Netral
Menkes Deg-degan Stok Vaksin Sinovac Hanya 7 Juta Dosis untuk April 2021	Negatif	Negatif
Agar Pandemi Tak Berlanjut hingga 10 Tahun, Ini 3 Saran IDI untuk Pemerintah	Netral	Netral
Heboh China Akui Kemanjuran Vaksinnya Rendah, Faktanya?	Negatif	Negatif
Komentar Pakar UGM soal MUI Tetapkan Vaksin Corona Sinovac Halal	Positif	Netral
GTPP Denpasar : 361 terpapar COVID-19 masih dirawat di RS	Netral	Negatif
Vaksinasi Covid-19 RI Tahap 3 Mulai Juni, Molor?	Netral	Netral
4 Negara Pemakai Vaksin Sinovac Buatan China	Netral	Netral

Bali targetkan dua pekan tuntas 500 ribu vaksinasi dosis pertama	Netral	Netral
OJK-IKKNB-Kemenkes Gelar Vaksinasi Covid-19 Khusus Lansia	Netral	Netral
Dinkes Gunung Kidul gunakan AstraZeneca untuk vaksinasi	Netral	Netral
Sempat Positif Covid-19, Wali Kota Depok M Idris Terima Suntikan Vaksin	Netral	Netral
Jokowi: Kiai-Kiai di Jawa Timur Siap Diberi Vaksin AstraZeneca	Netral	Netral
Wamenag: MUI Pusat Tidak Sebut Vaksin AstraZeneca Halal, Tapi Boleh Digunakan	Positif	Positif
Menkes Akui Laju Vaksinasi Covid Melambat, Ini Biang Keroknya	Negatif	Negatif
Kakesdam: Vaksin AstraZeneca Batch CTMAV547 tak ditarik tapi ditunda	Netral	Netral
Wapres: Bendera Merah Putih Berkibar di Luar Negeri Hanya karena 2 Hal	Netral	Netral
Menkes Pertimbangkan Jalur Vaksin COVID-19 Mandiri, Ini Pendapat Pakar	Netral	Netral
Nah Lho! Brasil Tolak Vaksin Covid-19 India	Netral	Netral
Vaksin AstraZeneca kembali tiba di Indonesia	Netral	Netral
Gara-gara Baca Berita Hoaks, Warga Takut Divaksin Covid-19, Lalu Sembunyi di Hutan	Negatif	Netral
Syarat Dapat E-Tiket Vaksinasi Covid-19	Netral	Netral
Data Sementara Uji Klinis III di AS: Efikasi Vaksin COVID-19 AstraZeneca 79 Persen	Positif	Netral
Penerima vaksin lengkap warga Indonesia	Netral	Netral

bertambah 25.318 jiwa		
Pakai Masker Usai Divaksin Bisa Lindungi Nyawa Ribuan Orang	Positif	Netral
Kilas NusAntara Edisi COVID-19	Netral	Netral
Pembekuan Darah, Regulator AS: Setop Pakai Vaksin Covid J&J	Negatif	Negatif
Vaksin Covid-19 Sinovac Batch 1 Kedaluwarsa 25 Maret 2021	Netral	Netral
Blak-blakan Ketua Fraksi PAN Mau Disuntik Vaksin Terawan	Netral	Netral
WHO: Dunia di Titik Kritis Pandemi, Mutasi Covid Makin Ganas	Negatif	Netral
RI Tunda Distribusi Vaksin AstraZeneca, Kemenkes Ungkap Alasannya	Netral	Netral
Daftar 7 Vaksin COVID-19 yang Beredar di Indonesia	Netral	Netral
Kemenkes: 132 Ribu Lebih Nakes Sudah Divaksin COVID-19, 20.154 Berhalangan	Netral	Netral
AstraZeneca: Tak Ada Bukti Vaksin COVID-19 Tingkatkan Risiko Pembekuan Darah	Positif	Negatif
Tak Langsung Kebal, Butuh Waktu Bangun Antibodi Setelah Suntik Vaksin COVID-19	Netral	Positif
Kemenkes: Vaksin AstraZeneca Didistribusikan ke 7 Provinsi, Tidak Ada KIPI Berat Ditemukan	Positif	Netral
Menkes Bicara Vaksin Terawan & Pengawasan Tak Lagi di BPOM	Netral	Netral
Daftar Petinggi Negara yang Sudah Vaksin COVID-19, Joe Biden hingga Raja Salman	Netral	Netral
Ketersediaan vaksin COVID-19 terjaga	Positif	Netral

BPOM: Vaksin AstraZeneca bisa digunakan kembali	Netral	Netral
Vaksin Covid-19 Picu Efek Samping Mirip Kanker Payudara, Berbahayakah?	Negatif	Negatif
Pemkab Pamekasan vaksin 768 lansia	Netral	Netral
RI Resmi Pakai Vaksin Sinopharm Lawan Covid, Efikasinya 78%	Positif	Netral
Merasa Bebas Setelah Vaksin, Wanita di AS Dinyatakan Positif COVID-19	Negatif	Netral
Pembekuan Darah, Penyuntikan Vaksin Johnson & Johnson Disetop	Negatif	Positif
Jubir tegaskan vaksin gotong royong gratis	Positif	Netral
Momen Dokter Kepresidenan Gemetar Saat Menyuntik Jokowi	Netral	Netral
Vaksin COVID-19 telah diterima 12.995.710 orang di Indonesia	Netral	Netral
7.450 Lansia di Banjarmasin Jalani Vaksinasi Covid-19	Netral	Netral
Menkes: Setelah Vaksinasi, Kasus Covid-19 Tenaga Kesehatan Menurun	Positif	Netral
Kemenkes: Efek Samping Disuntik Vaksin Covid-19 Nyeri dan Gatal-gatal	Negatif	Negatif
Klinis Fase 3 di Turki: Efikasi Vaksin COVID-19 Sinovac 83,5 Persen	Positif	Positif
Pekanbaru gelar vaksinasi massal lagi kini di delapan titik	Netral	Netral
Komentar Jokowi Usai Terima Vaksin COVID-19: Nggak Terasa Sama Sekali	Positif	Netral
Warga Moskow diawasi 380 ribu CCTV selama pandemi	Netral	Netral

Terpopuler Sepekan: Catat 16 Kriteria yang Tak Boleh Divaksin Corona!	Netral	Netral
Kemenkes Imbau Calon Penerima Vaksin Corona Jujur saat Skrining	Netral	Netral
UPDATE 10 Februari: 969.546 Tenaga Kesehatan Disuntik Vaksin Covid-19 Dosis Pertama	Netral	Netral
Ahli: Vaksin Nusantara Bukan Inovasi Anak Bangsa, Tapi AS	Netral	Netral
Jangan Takabur, Orang yang Divaksin Masih Bisa Kena Covid	Negatif	Netral
Vaksinasi Covid-19 untuk Lansia di Kalsel Baru Dimulai pada Maret 2021	Netral	Netral
Pro dan kontra risiko AstraZeneca	Netral	Netral
Hasil Akhir Uji Vaksin Sinovac 78 Persen Efektif, 100 Persen Cegah Kasus Berat	Positif	Positif
Beberapa Anggota DPR Disuntik Vaksin Nusantara Hari Ini	Netral	Netral
Suntikan Pertama Vaksin Pfizer di Israel 85 Persen Efektif, Studi Jelaskan	Positif	Positif
Total 14,8 juta penduduk Indonesia telah divaksinasi COVID-19	Netral	Netral
Resmi! RI Kedatangan 6 Juta Bulk Vaksin Covid-19 Sinovac	Netral	Netral
Gubernur Riau harap vaksinasi massal di Dumai selesai sepekan	Netral	Netral
Jokowi Hapus Kemenristek, Gimana Nasib Vaksin Made in RI?	Netral	Netral
Bio Farma Sebut Pfizer Ogah Dituntut Bila Vaksin Corona Buatannya Bermasalah	Negatif	Netral

Indonesia Kini Punya 38 Juta Dosis Vaksin Covid-19 dari Sinovac	Netral	Netral
Jokowi: Tak Mudah Dapatkan Vaksin Covid-19, Rebutan 215 Negara	Netral	Netral
Para Profesor Tenaga Kesehatan yang Sudah Lansia Mulai Terima Vaksin Covid-19	Netral	Netral
Kabar Buruk! Vaksin Pfizer Tak Kuat Lawan Virus Corona Afsel	Negatif	Positif
Presiden Brasil Sebut Covid Senjata Biologis, Sindir Siapa?	Netral	Netral
Viral Anjuran Minum Air Kelapa untuk Netralkan Vaksin COVID-19, Ini Faktanya	Netral	Netral
18 Tahun ke Atas Jadi Prioritas, Ini Batas Usia Vaksin COVID-19	Netral	Netral
Mantap! 6 Juta Vaksin Sinovac & 400 Ribu Sinopharm Tiba di RI	Positif	Netral
Ibu Hamil Positif Corona Tak Selalu Lahirkan Bayi dengan Antibodi Covid-19, Mengapa?	Positif	Positif
Kemenkes: Sangat Mungkin Bupati Sleman Divaksin COVID-19 Saat Masa Inkubasi	Netral	Negatif
Usai Vaksinasi Nakes Lansia, secara Paralel Lansia Lain Akan Diberi Vaksin Covid-19	Netral	Netral
Kemarin, literasi digital sampai penggunaan AstraZeneca dilanjutkan	Netral	Netral
Satgas: Tingkat kesembuhan pasien COVID-19 Mataram naik signifikan	Positif	Positif
Mulai Uji Klinis Vaksin COVID-19 pada Anak Usia 6 Bulan hingga 11 Tahun	Netral	Netral
Masuk Kategori Lanjut Usia, Bupati Magetan	Netral	Netral

Suprawoto Terima Vaksin Covid-19		
Studi Chile: Suntikan Pertama Sinovac Tak Lindungi dari Covid	Negatif	Netral
Menkes Sebut Ketersediaan Vaksin Covid-19 Terancam Akibat Embargo	Negatif	Netral
Sempat Bikin Heboh, Begini 'Ending' Vaksin Nusantara Terwan	Netral	Netral
Pemilik Kondisi Kesehatan Khusus Ini Disebut Dapat Menerima Vaksin Covid-19	Netral	Netral
	Netral	Netral
Tips agar Lancar Jalani Vaksinasi COVID-19	Netral	Netral
5.124.948 Orang Telah Terima Vaksin Covid-19	Netral	Netral
Kemenkes: 80,8 persen publik bersedia terima vaksin COVID-19	Positif	Netral
Kedaluwarsa Akhir Mei 2021, Menkes Budi Akui Masa Simpan Vaksin AstraZeneca Pendek	Netral	Netral
Disuntik di Abu Dhabi, WNI Ungkap Efek Samping Vaksin COVID-19	Negatif	Negatif
Menkes Budi: Vaksinasi COVID-19 Pakai Vaksin AstraZeneca Mulai Pekan Depan	Netral	Netral
Parlemen Ingatkan Pemerintah Soal Pengeluaran EUA Vaksin AstraZeneca Sesuai Prosedur	Netral	Netral
Lakukan Uji Klinis Awal, Sinovac Sebut Vaksin COVID-19 CoronaVac Aman untuk Anak	Positif	Netral
Dinkes Madiun jamin keamanan vaksin AstraZeneca	Positif	Positif

Banyak Keluhan Ngantuk Usai Vaksin COVID-19, Ini Penjelasan Tim Riset	Negatif	Negatif
Tim pakar: Maksimalkan vaksinasi saat penyekatan larangan mudik	Netral	Netral
Jumat 12 Maret 2021, 116.860 Orang Disuntik Vaksin Covid-19	Netral	Netral
Tak Cuma Jokowi, Ini Daftar Penerima Vaksin Pertama	Netral	Netral
Imuwan Sebut Varian Corona P1 Brasil Bermutasi, Lebih Ganas	Negatif	Negatif
Sebelum Diuji ke Anak, Keamanan dan Manfaat Vaksin COVID-19 pada Dewasa Harus Dipastikan	Positif	Netral
Lima Terobosan Menkes Tangani Covid dan Program Vaksinasi	Netral	Netral
Vaksinasi Covid-19 Tahap Kedua Dimulai 17 Februari, Bertahap, Prioritas Jawa-Bali	Netral	Netral
Fix! Vaksinasi Gotong Royong Pakai Vaksin Sinopharm & Sputnik	Netral	Netral
Resmi! Uni Eropa Gugat AstraZeneca Gegara Vaksin Covid	Netral	Netral
BPOM Sebut Efikasi Vaksin Sinovac 65,3 Persen, Ini Artinya	Netral	Positif
Beredar Daftar Artis Penerima Vaksin Pertama, Ini Kata Jubir Vaksinasi	Netral	Netral
Ilmuwan Oxford Siapkan Vaksin COVID-19 Versi Baru untuk Atasi Mutasi Corona	Netral	Netral
Peneliti: Vaksin COVID-19 Pfizer Kurang Efektif bagi Penderita Obesitas	Negatif	Netral
Diuji ke Anak, Keamanan dan Manfaat	Netral	Netral

Vaksin COVID-19 pada Dewasa Harus Dipastikan		
Jelang Vaksinasi COVID-19, Menkes Ingatkan Puskesmas yang Tak Punya Kulkas	Netral	Netral
Regulator Obat Uni Eropa Restui Penggunaan Vaksin COVID-19 Johnson and Johnson	Netral	Netral
Alasan Anang Hermansyah Mau Disuntik Vaksin Nusantara Terawan	Netral	Netral
Jokowi: 4,6 Juta Dosis Vaksin AstraZeneca Tiba di Indonesia Maret 2021	Netral	Netral
Alasan Sleman Pilih dr Tirta Jadi Penerima Vaksin COVID-19 Pertama	Netral	Netral
KPC-PEN : 8 juta vaksin Sinovac kembali datang Selasa	Netral	Netral
Dinkes Palembang vaksin disabilitas dan orang gangguan jiwa	Netral	Netral
Tak Hanya Pandemi, Kominfo Ingatkan Ancaman Infodemic Covid	Netral	Netral
Humaniora kemarin, varian baru COVID-19 hingga pembajakan buku	Negatif	Positif
Pemberian vaksin AstraZeneca dilanjutkan	Netral	Netral
Seluruh pengajar di Surabaya divaksin jelang sekolah tatap muka	Netral	Netral
Nah Loh! AstraZeneca Digugat Eropa Gegara Vaksin Covid-19	Netral	Netral
Ganjar Belum Mau Langsung Sanksi Penolak Vaksin Covid-19	Positif	Netral
Warga Bumi Harus Suntik Ulang Vaksin Covid Setiap Tahun?	Netral	Netral

Lewati Pandemi Flu saat Kecil, Nenek 104 Tahun di AS Disuntik Vaksin COVID-19	Netral	Netral
BPOM RI: Vaksin AstraZeneca yang Masuk Indonesia Sesuai Standar Mutu Global	Positif	Netral
Wamenkes Puji Penerima Vaksin COVID-19 Berusia 104 di Bogor	Positif	Positif
AS Kembali Izinkan Vaksin Covid-19 Johnson & Johnson Beredar	Netral	Netral
Vaksin AstraZeneca Boleh Digunakan, Kemenkes: Tak Ada Alasan Ragu Vaksinasi COVID-19	Positif	Positif
Bambang Brodjonegoro Pamer N29 Hingga Drone di Hannover Messe	Netral	Netral
Tak Sabar Ingin Divaksin, Gubernur Banten: Supaya Bisa Leluasa ke Mana-mana	Positif	Netral
Total penerima vaksin COVID-19 capai 13.699.393 orang	Netral	Netral
Satgas Covid-19: Vaksin AstraZeneca Tidak Terindikasi Menyebabkan Pembekuan Darah	Positif	Negatif
Tim Riset: Pasca 3 Bulan Suntik, Antibodi Vaksin Sinovac Masih di Atas 90 Persen	Positif	Positif
Jubir Wapres: Jangan Lagi Persoalkan Halal Haram Vaksin AstraZeneca, yang Penting Dbolehkan	Positif	Positif
Moderna Kaji Kemungkinan Muncul Alergi Usai Disuntik Vaksin Corona Buatannya	Negatif	Negatif
Besok, Sejumlah Anggota DPR Disuntik Vaksin Nusantara Terawan	Netral	Netral
Vaksinasi Tahap Dua: 7,5 Juta Dosis Vaksin Segera Didistribusikan	Netral	Netral

Kemenkes: Realisasi 1 juta vaksinasi per hari bergulir Juni	Netral	Netral
Indonesia Jadi Negara ke-4 Terbanyak Vaksinasi yang Mencapai 10 Juta Vaksin	Netral	Netral
WNI di UEA Ceritakan Pengalaman Dapat Vaksin COVID-19 Dosis Kedua	Netral	Netral
Tingginya Angka Kematian Jadi Alasan BPOM Bolehkan Lansia Disuntik Vaksin Sinovac	Netral	Netral
UPDATE 23 Februari: 789.966 Tenaga Kesehatan Disuntik Vaksin Dosis Kedua, 1.269.905 Dosis Pertama	Netral	Netral
Pendaftaran Vaksinasi Covid-19 Lansia di Denpasar Dimulai, Hanya Bisa Secara Online	Netral	Netral
Menkes Budi Gunadi: Jumlah Vaksin Jadi Tantangan Terbesar Vaksinasi COVID-19	Netral	Netral
Vaksinasi Covid-19 Tahap Dua, Pedagang Malioboro hingga Beringharjo Akan Disuntik Vaksin	Netral	Netral
Resmi! Denmark Setop Penggunaan Vaksin Covid-19 AstraZeneca	Netral	Netral
Wapres Jadi Pelopor Kelompok Umum, Vaksinasi Covid-19 Lansia Dimulai...	Netral	Netral
Bambang Brodjonegoro Beberkan Karya Anak Bangsa Lawan Covid	Netral	Netral
BPOM Buka-bukaan Soal Tahapan Uji Vaksin Covid-19 di RI	Netral	Netral
Usai Disuntik Vaksin Sputnik, Presiden Ini Positif Covid-19	Negatif	Negatif
Ahli Geriatri UI Jelaskan Kematian Lansia di	Negatif	Positif

Norwegia Usai Vaksin Pfizer		
Indocement vaksin 1.000 karyawan cegah penularan COVID-19	Netral	Positif
BPOM Setujui Vaksin AstraZeneca Dipakai untuk Vaksinasi Covid-19 Massal	Netral	Netral
Pendistribusian Vaksin Sinovac Dilengkapi Barcode Hingga GPS	Netral	Netral
Duh, 30 Orang Alami Pembekuan Darah Usai Divaksin AstraZeneca	Negatif	Negatif
Efek Samping Vaksin AstraZeneca di RI, Bukan Pembekuan Darah	Positif	Negatif
Catat! Vaksinasi Mandiri Dimulai Minggu Ketiga Mei 2021	Netral	Netral
Memahami Upaya Pemenuhan Ketersediaan dan Distribusi Vaksin COVID-19	Netral	Netral
Vaksin Covid-19 Anhui Diuji Klinis, Butuh 4.000 Relawan, Lansia Boleh Daftar	Netral	Netral
Viral Gambar Saya Siap Divaksin, Template Bisa Cari di Sini	Netral	Netral
UPDATE: 736.710 Tenaga Kesehatan Terima Vaksin Covid-19 Dosis Kedua	Netral	Netral
Belajar Tatap Muka Segera Dimulai, Vaksin Covid Anak Tak Ada	Netral	Netral
Drama Vaksin COVID-19, Profesor-Perawat Sama-sama Groggi Suntik Presiden	Netral	Netral
Positif COVID-19 Usai Vaksin, Kondisi Kesehatan Sekda Bandung Membaik	Positif	Negatif
Pekerja distributor barang perlu diprioritaskan vaksin gotong royong	Netral	Netral
Syarat Baru Perjalanan RI & Masa	Netral	Netral

Berlaku PCR-Antigen-GeNose		
Duh! Kasus Covid India Terus Meroket, Pasokan Vaksin Terancam	Negatif	Negatif
Tinjau Vaksinasi di Halmahera Utara, Jokowi Pastikan Distribusi Vaksin Merata	Positif	Netral
Baru 411 Nakes Padang Panjang yang Disuntik Vaksin Covid-19 dari Target 1.460 Orang	Netral	Netral
Infografis Sertifikat Vaksin Covid-19 Jadi Syarat Bepergian?	Netral	Netral
Satgas COVID-19: Penundaan Distribusi Vaksin AstraZeneca Hanya Sementara	Netral	Netral
[POPULER SAINS] Keamanan Vaksin Covid-19 Lansia Gejala Long Covid Bertahan 9 Bulan	Netral	Netral
Indonesia terima batch ketiga vaksin AstraZeneca dari fasilitas COVAX	Netral	Netral
Mengenal Varian Corona 'Ganas' India, Seberapa Menakutkan?	Negatif	Netral
Menkes: Antibodi Terbentuk 2-3 Minggu Setelah Vaksin COVID-19	Positif	Netral
Jangan Salah, Ini Alasan Muncul Efek Samping Vaksin Covid-19	Negatif	Negatif
Satgas COVID-19 pusat monitor pelaksanaan PPKM Mikro di Kota Madiun	Netral	Netral
UPDATE 27 Februari: 1.616.165 Tenaga Kesehatan Disuntik Vaksin Covid-19 Dosis Pertama	Netral	Netral
80 persen CJH Aceh sudah divaksin meski belum pasti berangkat	Netral	Netral

Lampiran 2 Hasil Klasifikasi *Labeling* Dataset Otomatis

Title	Label_Otomatis	Hasil_Prediksi
Wamenkes Jelaskan Alasan Pasien COVID-19 Tak Dapat Jatah Vaksin	Netral	Netral
Penderita Kanker Diperbolehkan Ikut Vaksinasi Covid-19, Ini Syaratnya	Netral	Netral
UPDATE 10 Februari: Kasus Aktif Covid-19 di Indonesia Tercatat 168.416	Netral	Netral
Vaksinasi Covid-19 Tahap Dua Dimulai Besok, Grobogan Dapat Jatah 12.000 Dosis	Netral	Netral
Menkes Disuntik Vaksin Dosis Kedua: Rasanya Jadi Pengen Makan Banyak	Netral	Netral
UPDATE 19 Februari: 668.914 Nakes Sudah Divaksinasi Dosis Kedua, 1.191.031 Dosis Pertama	Netral	Netral
UPDATE 25 Februari: 853.745 Tenaga Kesehatan Sudah Divaksinasi Dosis Kedua, 1.461.920 Dosis Pertama	Netral	Netral
Tak Ragu Disuntik Vaksin Sinovac, dr Tirta: Sudah Terjamin Uji Klinis	Netral	Netral
Vaksinasi COVID-19 ASN Jabar Dimulai, 570 Orang Sudah Disuntik Vaksin Corona	Netral	Netral
Penghentian sementara vaksin AstraZeneca CTMAV547	Netral	Netral
Indonesia Terima Kedatangan Vaksin COVID-19 Tahap Ketujuh	Netral	Netral
14.815.666 jiwa penduduk RI telah menjalani	Netral	Netral

vaksinasi dosis pertama		
IPB Siap Bantu Uji Praklinis Vaksin Covid-19 Merah Putih	Netral	Netral
Penjelasan Kemenkes soal Vaksin Covid-19 Sinovac Kedaluwarsa 25 Maret 2021	Netral	Netral
Kasus Aktif Menurun, Wamenkes Sebut Bukti Vaksin Covid-19 Kurangi Risiko Sakit Parah	Netral	Netral
Vaksin AstraZeneca, Ma'ruf Amin: Bukan Soal Halal atau Tidak Halal, yang Penting Boleh	Netral	Netral
Kapan Vaksin Merah Putih Dipakai Lawan Covid? Ini Kata BPOM	Netral	Netral
Menkes Deg-degan Stok Vaksin Sinovac Hanya 7 Juta Dosis untuk April 2021	Netral	Netral
Agar Pandemi Tak Berlanjut hingga 10 Tahun, Ini 3 Saran IDI untuk Pemerintah	Netral	Netral
Heboh China Akui Kemanjuran Vaksinnya Rendah, Faktanya?	Netral	Netral
Komentar Pakar UGM soal MUI Tetapkan Vaksin Corona Sinovac Halal	Netral	Netral
GTPP Denpasar : 361 terpapar COVID-19 masih dirawat di RS	Netral	Netral
Vaksinasi Covid-19 RI Tahap 3 Mulai Juni, Molor?	Netral	Netral
4 Negara Pemakai Vaksin Sinovac Buatan China	Netral	Netral
Bali targetkan dua pekan tuntas 500 ribu vaksinasi dosis pertama	Netral	Netral
OJK-IKKNB-Kemenkes Gelar Vaksinasi	Netral	Netral

Covid-19 Khusus Lansia		
Dinkes Gunung Kidul gunakan Astrazeneca untuk vaksinasi	Netral	Netral
Sempat Positif Covid-19, Wali Kota Depok M Idris Terima Suntikan Vaksin	Netral	Netral
Jokowi: Kiai-Kiai di Jawa Timur Siap Diberi Vaksin AstraZeneca	Netral	Netral
Wamenag: MUI Pusat Tidak Sebut Vaksin AstraZeneca Halal, Tapi Boleh Digunakan	Netral	Netral
Menkes Akui Laju Vaksinasi Covid Melambat, Ini Biang Keroknya	Netral	Netral
Kakesdam: Vaksin AstraZeneca Batch CTMAV547 tak ditarik tapi ditunda	Netral	Netral
Wapres: Bendera Merah Putih Berkibar di Luar Negeri Hanya karena 2 Hal	Netral	Netral
Menkes Pertimbangkan Jalur Vaksin COVID-19 Mandiri, Ini Pendapat Pakar	Netral	Netral
Nah Lho! Brasil Tolak Vaksin Covid-19 India	Positif	Netral
Vaksin AstraZeneca kembali tiba di Indonesia	Netral	Netral
Gara-gara Baca Berita Hoaks, Warga Takut Divaksin Covid-19, Lalu Sembunyi di Hutan	Netral	Netral
Syarat Dapat E-Tiket Vaksinasi Covid-19	Netral	Netral
Data Sementara Uji Klinis III di AS: Efikasi Vaksin COVID-19 AstraZeneca 79 Persen	Netral	Netral
Penerima vaksin lengkap warga Indonesia bertambah 25.318 jiwa	Netral	Netral
Pakai Masker Usai Divaksin Bisa Lindungi Nyawa Ribuan Orang	Netral	Netral

Kilas NusAntara Edisi COVID-19	Netral	Netral
Pembekuan Darah, Regulator AS: Setop Pakai Vaksin Covid J&J	Netral	Netral
Vaksin Covid-19 Sinovac Batch 1 Kedaluwarsa 25 Maret 2021	Netral	Netral
Blak-blakan Ketua Fraksi PAN Mau Disuntik Vaksin Terawan	Netral	Netral
WHO: Dunia di Titik Kritis Pandemi, Mutasi Covid Makin Ganas	Netral	Netral
RI Tunda Distribusi Vaksin AstraZeneca, Kemenkes Ungkap Alasannya	Netral	Netral
Daftar 7 Vaksin COVID-19 yang Beredar di Indonesia	Netral	Netral
Kemenkes: 132 Ribu Lebih Nakes Sudah Divaksin COVID-19, 20.154 Berhalangan	Netral	Netral
AstraZeneca: Tak Ada Bukti Vaksin COVID-19 Tingkatkan Risiko Pembekuan Darah	Netral	Netral
Tak Langsung Kebal, Butuh Waktu Bangun Antibodi Setelah Suntik Vaksin COVID-19	Netral	Netral
Kemenkes: Vaksin AstraZeneca Didistribusikan ke 7 Provinsi, Tidak Ada KIPI Berat Ditemukan	Netral	Netral
Menkes Bicara Vaksin Terawan & Pengawasan Tak Lagi di BPOM	Netral	Netral
Daftar Petinggi Negara yang Sudah Vaksin COVID-19, Joe Biden hingga Raja Salman	Netral	Netral
Ketersediaan vaksin COVID-19 terjaga	Positif	Netral
BPOM: Vaksin AstraZeneca bets CTMAV547 dapat digunakan kembali	Netral	Netral
Vaksin Covid-19 Picu Efek Samping Mirip	Netral	Netral

Kanker Payudara, Berbahayakah?		
Pemkab Pamekasan vaksin 768 lansia	Netral	Netral
RI Resmi Pakai Vaksin Sinopharm Lawan Covid, Efikasinya 78%	Netral	Netral
Merasa Bebas Setelah Vaksin, Wanita di AS Dinyatakan Positif COVID-19	Netral	Netral
Pembekuan Darah, Penyuntikan Vaksin Johnson & Johnson Disetop	Netral	Netral
Jubir tegaskan vaksin gotong royong gratis	Netral	Netral
Momen Dokter Kepresidenan Gemetar Saat Menyuntik Jokowi	Netral	Netral
Vaksin COVID-19 telah diterima 12.995.710 orang di Indonesia	Netral	Netral
7.450 Lansia di Banjarmasin Jalani Vaksinasi Covid-19	Netral	Netral
Menkes: Setelah Vaksinasi, Kasus Covid-19 Tenaga Kesehatan Menurun	Netral	Netral
Kemenkes: Efek Samping Disuntik Vaksin Covid-19 Nyeri dan Gatal-gatal	Netral	Netral
Klinis Fase 3 di Turki: Efikasi Vaksin COVID-19 Sinovac 83,5 Persen	Netral	Netral
Pekanbaru gelar vaksinasi massal lagi kini di delapan titik	Netral	Netral
Komentar Jokowi Usai Terima Vaksin COVID-19: Nggak Terasa Sama Sekali	Netral	Netral
Warga Moskow diawasi 380 ribu CCTV selama pandemi	Netral	Netral
Terpopuler Sepekan: Catat 16 Kriteria yang Tak Boleh Divaksin Corona!	Positif	Positif
Kemenkes Imbau Calon Penerima Vaksin	Netral	Netral

Corona Jujur saat Skrining		
UPDATE 10 Februari: 969.546 Tenaga Kesehatan Disuntik Vaksin Covid-19 Dosis Pertama	Netral	Netral
Ahli: Vaksin Nusantara Bukan Inovasi Anak Bangsa, Tapi AS	Netral	Netral
Jangan Takabur, Orang yang Divaksin Masih Bisa Kena Covid	Netral	Netral
Vaksinasi Covid-19 untuk Lansia di Kalsel Baru Dimulai pada Maret 2021	Netral	Netral
Pro dan kontra risiko AstraZeneca	Negatif	Netral
Hasil Akhir Uji Vaksin Sinovac 78 Persen Efektif, 100 Persen Cegah Kasus Berat	Netral	Netral
Beberapa Anggota DPR Disuntik Vaksin Nusantara Hari Ini	Netral	Netral
Suntikan Pertama Vaksin Pfizer di Israel 85 Persen Efektif, Studi Jelaskan	Netral	Netral
Total 14,8 juta penduduk Indonesia telah divaksinasi COVID-19	Netral	Netral
Resmi! RI Kedatangan 6 Juta Bulk Vaksin Covid-19 Sinovac	Positif	Netral
Gubernur Riau harap vaksinasi massal di Dumai selesai sepekan	Netral	Netral
Jokowi Hapus Kemenristek, Gimana Nasib Vaksin Made in RI?	Netral	Netral
Bio Farma Sebut Pfizer Ogah Dituntut Bila Vaksin Corona Buatannya Bermasalah	Netral	Netral
Indonesia Kini Punya 38 Juta Dosis Vaksin Covid-19 dari Sinovac	Netral	Netral
Jokowi: Tak Mudah Dapatkan Vaksin Covid-	Netral	Netral

19, Rebutan 215 Negara		
Para Profesor Tenaga Kesehatan yang Sudah Lansia Mulai Terima Vaksin Covid-19	Netral	Netral
Kabar Buruk! Vaksin Pfizer Tak Kuat Lawan Virus Corona Afsel	Positif	Positif
Presiden Brasil Sebut Covid Senjata Biologis, Sindir Siapa?	Netral	Netral
Viral Anjuran Minum Air Kelapa untuk Netralkan Vaksin COVID-19, Ini Faktanya	Netral	Netral
18 Tahun ke Atas Jadi Prioritas, Ini Batas Usia Vaksin COVID-19	Netral	Netral
Mantap! 6 Juta Vaksin Sinovac & 400 Ribu Sinopharm Tiba di RI	Positif	Netral
Ibu Hamil Positif Corona Tak Selalu Lahirkan Bayi dengan Antibodi Covid-19, Mengapa?	Netral	Netral
Kemenkes: Sangat Mungkin Bupati Sleman Divaksin COVID-19 Saat Masa Inkubasi	Netral	Netral
Usai Vaksinasi Nakes Lansia, secara Paralel Lansia Lain Akan Diberi Vaksin Covid-19	Netral	Netral
Kemarin, literasi digital sampai penggunaan AstraZeneca dilanjutkan	Netral	Netral
Satgas: Tingkat kesembuhan pasien COVID-19 Mataram naik signifikan	Positif	Netral
Mulai Uji Klinis Vaksin COVID-19 pada Anak Usia 6 Bulan hingga 11 Tahun	Netral	Netral
Masuk Kategori Lanjut Usia, Bupati Magetan Suprawoto Terima Vaksin Covid-19	Netral	Netral
Studi Chile: Suntikan Pertama Sinovac Tak Lindungi dari Covid	Netral	Netral

Menkes Sebut Ketersediaan Vaksin Covid-19 Terancam Akibat Embargo	Netral	Netral
Sempat Bikin Heboh, Begini 'Ending' Vaksin Nusantara Terwan	Netral	Netral
Pemilik Kondisi Kesehatan Khusus Ini Disebut Dapat Menerima Vaksin Covid-19	Netral	Netral
100.000 Lansia Akan Divaksin Covid-19, Penerima Dites Naiki 10 Anak Tangga	Netral	Netral
Tips agar Lancar Jalani Vaksinasi COVID-19	Netral	Netral
5.124.948 Orang Telah Terima Vaksin Covid-19	Netral	Netral
Kemenkes: 80,8 persen publik bersedia terima vaksin COVID-19	Netral	Netral
Kedaluwarsa Akhir Mei 2021, Menkes Budi Akui Masa Simpan Vaksin AstraZeneca Pendek	Netral	Netral
Disuntik di Abu Dhabi, WNI Ungkap Efek Samping Vaksin COVID-19	Netral	Netral
Menkes Budi: Vaksinasi COVID-19 Pakai Vaksin AstraZeneca Mulai Pekan Depan	Netral	Netral
Parlemen Ingatkan Pemerintah Soal Pengeluaran EUA Vaksin AstraZeneca Sesuai Prosedur	Netral	Netral
Lakukan Uji Klinis Awal, Sinovac Sebut Vaksin COVID-19 CoronaVac Aman untuk Anak	Netral	Netral
Dinkes Madiun jamin keamanan vaksin AstraZeneca	Netral	Netral
Banyak Keluhan Ngantuk Usai Vaksin COVID-19, Ini Penjelasan Tim Riset	Netral	Netral

Tim pakar: Maksimalkan vaksinasi saat penyekatan larangan mudik	Netral	Netral
Jumat 12 Maret 2021, 116.860 Orang Disuntik Vaksin Covid-19	Netral	Netral
Tak Cuma Jokowi, Ini Daftar Penerima Vaksin Pertama	Netral	Netral
Imuwan Sebut Varian Corona P1 Brasil Bermutasi, Lebih Ganas	Netral	Netral
Sebelum Diuji ke Anak, Keamanan dan Manfaat Vaksin COVID-19 pada Dewasa Harus Dipastikan	Netral	Netral
Lima Terobosan Menkes Tangani Covid dan Program Vaksinasi	Netral	Netral
Vaksinasi Covid-19 Tahap Kedua Dimulai 17 Februari, Bertahap, Prioritas Jawa-Bali	Netral	Netral
Fix! Vaksinasi Gotong Royong Pakai Vaksin Sinopharm & Sputnik	Positif	Netral
Resmi! Uni Eropa Gugat AstraZeneca Negara Vaksin Covid	Positif	Netral
BPOM Sebut Efikasi Vaksin Sinovac 65,3 Persen, Ini Artinya	Netral	Netral
Beredar Daftar Artis Penerima Vaksin Pertama, Ini Kata Jubir Vaksinasi	Netral	Netral
Ilmuwan Oxford Siapkan Vaksin COVID-19 Versi Baru untuk Atasi Mutasi Corona	Netral	Netral
Peneliti: Vaksin COVID-19 Pfizer Kurang Efektif bagi Penderita Obesitas	Netral	Netral
Diuji ke Anak, Keamanan dan Manfaat Vaksin COVID-19 pada Dewasa Harus Dipastikan	Netral	Netral

Jelang Vaksinasi COVID-19, Menkes Ingatkan Puskesmas yang Tak Punya Kulkas	Netral	Netral
Regulator Obat Uni Eropa Restui Penggunaan Vaksin COVID-19 Johnson and Johnson	Netral	Netral
Alasan Anang Hermansyah Mau Disuntik Vaksin Nusantara Terawan	Netral	Netral
Jokowi: 4,6 Juta Dosis Vaksin AstraZeneca Tiba di Indonesia Maret 2021	Netral	Netral
Alasan Sleman Pilih dr Tirta Jadi Penerima Vaksin COVID-19 Pertama	Netral	Netral
KPC-PEN : 8 juta vaksin Sinovac kembali datang Selasa	Netral	Netral
Dinkes Palembang vaksin disabilitas dan orang gangguan jiwa	Negatif	Netral
Tak Hanya Pandemi, Kominfo Ingatkan Ancaman Infodemic Covid	Netral	Netral
Humaniora kemarin, varian baru COVID-19 hingga pembajakan buku	Negatif	Netral
Pemberian vaksin AstraZeneca dilanjutkan	Netral	Netral
Seluruh pengajar di Surabaya divaksin jelang sekolah tatap muka	Netral	Netral
Nah Loh! AstraZeneca Digugat Eropa Gegara Vaksin Covid-19	Positif	Netral
Ganjar Belum Mau Langsung Sanksi Penolak Vaksin Covid-19	Netral	Netral
Warga Bumi Harus Suntik Ulang Vaksin Covid Setiap Tahun?	Netral	Netral
Lewati Pandemi Flu saat Kecil, Nenek 104 Tahun di AS Disuntik Vaksin COVID-19	Netral	Netral

BPOM RI: Vaksin AstraZeneca yang Masuk Indonesia Sesuai Standar Mutu Global	Netral	Netral
Wamenkes Puji Penerima Vaksin COVID-19 Berusia 104 di Bogor	Netral	Netral
AS Kembali Izinkan Vaksin Covid-19 Johnson & Johnson Beredar	Netral	Netral
Vaksin AstraZeneca Boleh Digunakan, Kemenkes: Tak Ada Alasan Ragu Vaksinasi COVID-19	Netral	Netral
Bambang Brodjonegoro Pamer N29 Hingga Drone di Hannover Messe	Netral	Netral
Tak Sabar Ingin Divaksin, Gubernur Banten: Supaya Bisa Leluasa ke Mana-mana	Netral	Netral
Total penerima vaksin COVID-19 capai 13.699.393 orang	Netral	Netral
Satgas Covid-19: Vaksin AstraZeneca Tidak Terindikasi Menyebabkan Pembekuan Darah	Netral	Netral
Tim Riset: Pasca 3 Bulan Suntik, Antibodi Vaksin Sinovac Masih di Atas 90 Persen	Netral	Netral
Jubir Wapres: Jangan Lagi Persoalkan Halal Haram Vaksin AstraZeneca, yang Penting Dbolehkan	Netral	Netral
Moderna Kaji Kemungkinan Muncul Alergi Usai Disuntik Vaksin Corona Buatannya	Netral	Netral
Besok, Sejumlah Anggota DPR Disuntik Vaksin Nusantara Terawan	Netral	Netral
Vaksinasi Tahap Dua: 7,5 Juta Dosis Vaksin Segera Didistribusikan	Netral	Netral
Kemenkes: Realisasi 1 juta vaksinasi per hari bergulir Juni	Netral	Netral

Indonesia Jadi Negara ke-4 Terbanyak Vaksinasi yang Mencapai 10 Juta Vaksin	Netral	Netral
WNI di UEA Ceritakan Pengalaman Dapat Vaksin COVID-19 Dosis Kedua	Netral	Netral
Tingginya Angka Kematian Jadi Alasan BPOM Bolehkan Lansia Disuntik Vaksin Sinovac	Netral	Netral
UPDATE 23 Februari: 789.966 Tenaga Kesehatan Disuntik Vaksin Dosis Kedua, 1.269.905 Dosis Pertama	Netral	Netral
Pendaftaran Vaksinasi Covid-19 Lansia di Denpasar Dimulai, Hanya Bisa Secara Online	Netral	Netral
Menkes Budi Gunadi: Jumlah Vaksin Jadi Tantangan Terbesar Vaksinasi COVID-19	Netral	Netral
Vaksinasi Covid-19 Tahap Dua, Pedagang Malioboro hingga Beringharjo Akan Disuntik Vaksin	Netral	Netral
Resmi! Denmark Setop Penggunaan Vaksin Covid-19 AstraZeneca	Positif	Netral
Wapres Jadi Pelopor Kelompok Umum, Vaksinasi Covid-19 Lansia Dimulai...	Netral	Netral
Bambang Brodjonegoro Beberkan Karya Anak Bangsa Lawan Covid	Netral	Netral
BPOM Buka-bukaan Soal Tahapan Uji Vaksin Covid-19 di RI	Netral	Netral
Usai Disuntik Vaksin Sputnik, Presiden Ini Positif Covid-19	Netral	Netral
Ahli Geriatri UI Jelaskan Kematian Lansia di Norwegia Usai Vaksin Pfizer	Netral	Netral
Indocement vaksin 1.000 karyawan cegah	Netral	Netral

penularan COVID-19		
BPOM Setujui Vaksin Astrazeneca Dipakai untuk Vaksinasi Covid-19 Massal	Netral	Netral
Pendistribusian Vaksin Sinovac Dilengkapi Barcode Hingga GPS	Netral	Netral
Duh, 30 Orang Alami Pembekuan Darah Usai Divaksin Astrazeneca	Netral	Netral
Efek Samping Vaksin AstraZeneca di RI, Bukan Pembekuan Darah	Netral	Netral
Catat! Vaksinasi Mandiri Dimulai Minggu Ketiga Mei 2021	Positif	Netral
Memahami Upaya Pemenuhan Ketersediaan dan Distribusi Vaksin COVID-19	Netral	Netral
Vaksin Covid-19 Anhui Diuji Klinis, Butuh 4.000 Relawan, Lansia Boleh Daftar	Netral	Netral
Viral Gambar Saya Siap Divaksin, Template Bisa Cari di Sini	Netral	Netral
UPDATE: 736.710 Tenaga Kesehatan Terima Vaksin Covid-19 Dosis Kedua	Netral	Netral
Belajar Tatap Muka Segera Dimulai, Vaksin Covid Anak Tak Ada	Netral	Netral
Drama Vaksin COVID-19, Profesor-Perawat Sama-sama Grogi Suntik Presiden	Netral	Netral
Positif COVID-19 Usai Vaksin, Kondisi Kesehatan Sekda Bandung Membaik	Netral	Netral
Pekerja distributor barang perlu diprioritaskan vaksin gotong royong	Netral	Netral
Syarat Baru Perjalanan RI & Masa Berlaku PCR-Antigen-GeNose	Netral	Netral
Duh! Kasus Covid India Terus Meroket,	Positif	Netral

Pasokan Vaksin Terancam		
Tinjau Vaksinasi di Halmahera Utara, Jokowi Pastikan Distribusi Vaksin Merata	Netral	Netral
Baru 411 Nakes Padang Panjang yang Disuntik Vaksin Covid-19 dari Target 1.460 Orang	Netral	Netral
Infografis Sertifikat Vaksin Covid-19 Jadi Syarat Bepergian?	Netral	Netral
Satgas COVID-19: Penundaan Distribusi Vaksin AstraZeneca Hanya Sementara	Netral	Netral
[POPULER SAINS] Keamanan Vaksin Covid-19 Lansia Gejala Long Covid Bertahan 9 Bulan	Netral	Netral
Indonesia terima batch ketiga vaksin AstraZeneca dari fasilitas COVAX	Netral	Netral
Mengenal Varian Corona 'Ganas' India, Seberapa Menakutkan?	Netral	Netral
Menkes: Antibodi Terbentuk 2-3 Minggu Setelah Vaksin COVID-19	Netral	Netral
Jangan Salah, Ini Alasan Muncul Efek Samping Vaksin Covid-19	Netral	Netral
Satgas COVID-19 pusat monitor pelaksanaan PPKM Mikro di Kota Madiun	Netral	Netral
UPDATE 27 Februari: 1.616.165 Tenaga Kesehatan Disuntik Vaksin Covid-19 Dosis Pertama	Netral	Netral
80 persen CJH Aceh sudah divaksin meski belum pasti berangkat	Netral	Netral

