

TESIS

**ANALISIS SENTIMEN TERHADAP VAKSIN COVID-19
BERDASARKAN OPINI MASYARAKAT PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE DAN NAIVE BAYES**



Disusun oleh:

Nama : Frizka Fitriana
NIM : 20.52.1304
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

TESIS

**ANALISIS SENTIMEN TERHADAP VAKSIN COVID-19
BERDASARKAN OPINI MASYARAKAT PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE DAN NAIVE BAYES**

**SENTIMENT ANALYSIS OF COVID-19 VACCINE BASED ON PUBLIC
OPINION ON SOCIAL MEDIA TWITTER USING VECTOR MACHINE
SUPPORT ALGORITHM AND NAIVE BAYES**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Fritzka Fitriana
NIM : 20.52.1304
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PENGESAHAN

ANALISIS SENTIMEN TERHADAP VAKSIN COVID-19 BERDASARKAN OPINI MASYARAKAT PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES

SENTIMENT ANALYSIS OF COVID-19 VACCINE BASED ON PUBLIC OPINION ON SOCIAL MEDIA TWITTER USING VECTOR MACHINE SUPPORT ALGORITHM AND NAIVE BAYES

Dipersiapkan dan Disusun oleh

Frizka Fitriana

20.52.1304

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Jumat, 04 Februari 2022

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 04 Februari 2022

Rektor

Prof. Dr. M. Suvanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN TERHADAP VAKSIN COVID-19 BERDASARKAN OPINI MASYARAKAT PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES

ANALISIS SENTIMEN TERHADAP VAKSIN COVID-19 BERDASARKAN OPINI MASYARAKAT PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES

Dipersiapkan dan Disusun oleh

Frizka Fitriana

20.52.1304

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Jumat, 04 Februari 2022

Pembimbing Utama

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Anggota Tim Penguji

Prof. Dr. Kusriani, M.Kom
NIK. 190302106

Pembimbing Pendamping

Hanif Al Fatta, M.Kom
NIK. 190302096

Alva Hendi Muhammad, S.T.,
M.Eng., Ph.D.
NIK. 190302493

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer.

Yogyakarta, 4 Februari 2022
Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Frizka Fitriana
NIM : 20.52.1304
Konsentrasi : Informaties Technopreneurship

Menyatakan bahwa Tesis dengan judul berikut:

Analisis Sentimen Terhadap Vaksin Covid-19 Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine Dan Naive Bayes.

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, S. Si, M.Kom
Dosen Pembimbing Pendamping : Haniif Al Fata, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 15 Februari 2022

Yang Menyatakan



Frizka Fitriana

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamin tesis ini saya persembahkan dan tujuannya kepada:

1. Allah SWT atas ridho-Nya yang telah memberikan kekuatan dan karunia sehingga penulis dapat menyelesaikan tesis ini dengan baik.
2. Kedua orang tua Bapak Meli Eriadi dan Ibu Yuli Mardiana terima kasih banyak karena tak pernah lelah memberikan doa, nasehat dan dukungan yang tiada hentinya.
3. Bapak M. Suyanto, Prof., Dr., MM. selaku Rektor Universitas Amikom Yogyakarta.
4. Untuk Prof. Dr. Ema Utami, S. Si, M.kom dan Bapak Hanif Al Fatta, M.Kom selaku dosen pembimbing, terimakasih telah memberikan kritik dan saran yang membangun dalam penyusunan tesis ini.
5. Untuk Adik Silvia Dwi Putri dan Seluruh keluarga besar penulis, terimakasih sudah mendukung dan selalu mengirim doa untuk kelancaran penulisan tesis ini.
6. Teman-teman kelas MTI Angkatan 24 yang selama 3 semester telah berjuang dan belajar bersama. Hendrik, Mala, Mbak Cristiani, seluruh teman sekelas yang tidak bisa saya sebut satu-satu.
7. Serta semua pihak yang telah membantu tersusunnya tesis ini yang tidak dapat saya sebutkan satu persatu.

HALAMAN MOTTO

- ❖ Jangan menyerah sebelum dicoba.
- ❖ Lakukan semua hal dengan Bismillah dan Berdoa.
- ❖ Dan selalulah meminta Doa kepada orang tua setiap hendak memilih semua keputusan dalam hidup mu.
- ❖ “Ketika Anda melihat seseorang yang telah diberikan kekayaan dan keindahan melebihi Anda, lihatlah orang-orang yang masih mampu bersyukur meski dalam kekurangan.” (HR Muslim)
- ❖ “Dan jika kamu menghitung-hitung nikmat Allah, niscaya kamu tak dapat menentukan jumlahnya. Sesungguhnya Allah benar-benar Maha Pengampun lagi Maha Penyayang”. (QS. An-Nahl/ 16:18)
- ❖ “Dan hanya kepada Tuhanmulah (Allah SWT), hendaknya kamu berharap”. (Qs Al Insyirah: 8)
- ❖ “Berdoalah (mintalah) kepadaKu (Allah SWT), Pastilah aku kabulkan untukmu”. (QS. Al-Mukmin : 60)

KATA PENGANTAR

Puji dan syukur kepada Tuhan Allah Yang Maha Esa yang telah memberikan karunia dan rahmatNya sehingga penulis dapat menyelesaikan penelitian tesis dengan baik. Pada kesempatan ini penulis hendak mengucapkan terima kasih kepada pihak-pihak yang telah mendukung selama mengerjakan penelitian tesis, antara lain :

1. Prof. Dr. Ema Utami, S.Si., M.Kom dan Hanif Al Fatta, M.Kom yang penuh kesabaran membimbing, memberi saran terbaik dan motivasi untuk menyelesaikan tesis ini.
2. Direktur Program Studi Pascasarjana Universitas AMIKOM, Ketua Program Studi Pascasarjana, para dosen program studi S2 Informatics Techopreneurship dan staf admisi MTI AMIKOM.
3. Kedua orangtua, yang selalu memberikan dukungan, semangat dan doa yang tak terbatas untuk kelancaran dan selesainya semua tugas dan tesis.
4. Teman – teman MTI Universitas AMIKOM Yogyakarta atas persahabatan, semangat dan dukungan dalam bentuk apapun.

Akhirnya kritik dan saran sangat penulis harapkan untuk perbaikan dimasa mendatang. Semoga hasil penelitian ini dapat memberikan manfaat dan pengetahuan bagi banyak pihak.

Yogyakarta,

Penulis

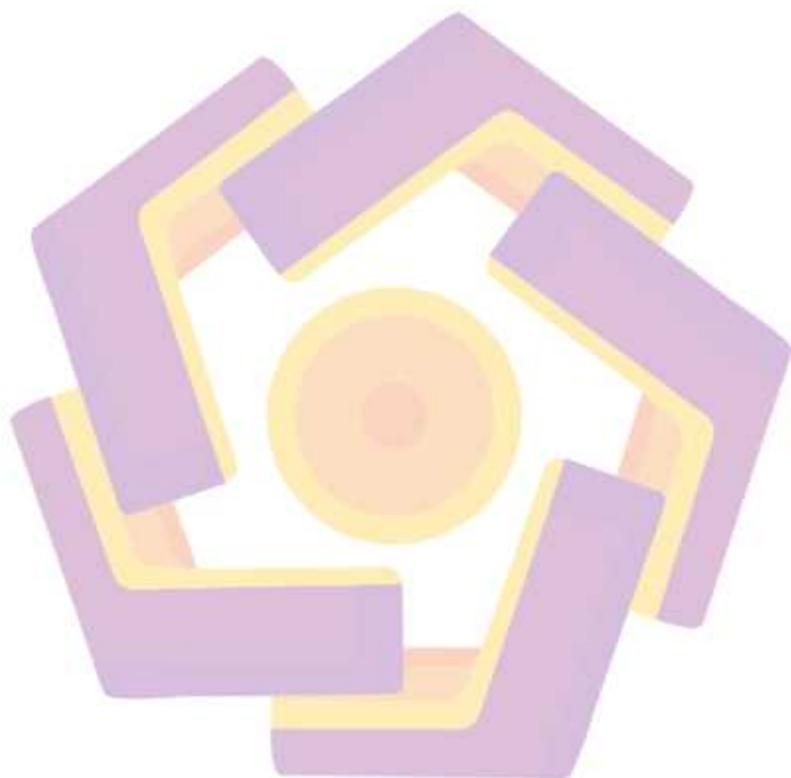
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xvi
INTISARI	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.1. Rumusan Masalah	11
1.2. Batasan Masalah	11
1.3. Tujuan Penelitian	12
1.4. Manfaat Penelitian	13
BAB II TINJAUAN PUSTAKA	14
2.1. Tinjauan Pustaka	14
2.1. Keaslian Penelitian	19
2.2. Landasan Teori	23

2.2.1. Analisis Sentimen.....	23
2.2.2. Twitter API.....	24
2.2.3. Naive Bayes	25
2.2.4. Support Vector Machine (SVM).....	26
2.2.5. Metode Pengujian Confusion Matrix	28
BAB III METODE PENELITIAN	31
3.1. Metodologi Penelitian.....	31
3.1.1. Jenis, Sifat, dan Pendekatan Peneltiain	31
3.1.2. Metode Pengumpulan Data	32
3.1.3. Metode Analisis Data.....	33
3.2. Alur Penelitian.....	34
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	39
4.1. Diagnosing.....	39
4.1.1. Studi Literatur	39
4.1.2. Pengumpulan Data dan Informasi	39
4.1.3. Requirements Definition	40
4.2. Action Taking.....	40
4.2.1. Pemodelan Algoritma.....	40
4.2.1.1. Naive Bayes.....	40
4.2.1.2. Support Vector Machine.....	41
4.2.2. Pemodelan Proses.....	42
4.2.3. Pemodelan Data (Analisis Data)	46

4.2.4. Pemodelan Sistem	54
4.3 Evaluation (System Testing)	58
4.3.1. Pengujian Naive Bayes	58
4.3.1.1. Pengujian Akurasi Naive Bayes	58
4.3.1.2. Pengujian Presisi <i>Naive Bayes</i>	62
4.3.1.3. Pengujian Recall <i>Naive Bayes</i>	66
4.3.1.4. Pengujian Waktu training <i>Naive Bayes</i>	70
4.3.1.5. Pengujian F1 score <i>Naive Bayes</i>	74
4.3.1.6. Pengujian MAE <i>Naive Bayes</i>	78
4.3.2. Pengujian Support Vector Machine	82
4.3.2.1. Pengujian Akurasi Support Vector Machine	82
4.3.2.2. Pengujian Presisi Support Vector Machine	87
4.3.2.3. Pengujian Recall Support Vector Machine	91
4.3.2.4. Pengujian Waktu training <i>Support Vector Machine</i>	95
4.3.2.5. Pengujian F1 score Support Vector Machine	99
4.3.2.6. Pengujian MAE SVM.....	103
4.4. <i>Specifying learning</i> (Representasi Hasil)	107
4.4.1. Analisis Perbandingan Hasil Pengujian	107
4.4.2. Penarikan Kesimpulan	112
BAB V PENUTUP	113
5.1. Kesimpulan.....	113

5.2. Saran	114
DAFTAR PUSTAKA	115
LAMPIRAN.....	cxix



DAFTAR TABEL

Tabel 2.1 Matriks literatur review dan posisi penelitian Analisis Sentimen Twitter Vaksin covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes.....	19
Tabel 4.1 Hasil Akurasi <i>Naive Bayes</i> Subset 1 (200 Data)	58
Tabel 4.2 Hasil Akurasi <i>Naive Bayes</i> Subset 2 (400 Data)	59
Tabel 4.3 Hasil Akurasi <i>Naive Bayes</i> Subset 3 (600 Data)	59
Tabel 4.4 Hasil Akurasi <i>Naive Bayes</i> Subset 4 (800 Data)	60
Tabel 4.5 Hasil Akurasi <i>Naive Bayes</i> Subset 5 (1000 Data)	61
Tabel 4.6 Hasil Presisi <i>Naive Bayes</i> Subset 1 (200 Data)	62
Tabel 4.7 Hasil Presisi <i>Naive Bayes</i> Subset 2 (400 Data)	63
Tabel 4.8 Hasil Presisi <i>Naive Bayes</i> Subset 3 (600 Data)	63
Tabel 4.9 Hasil Presisi <i>Naive Bayes</i> Subset 4 (800 Data)	64
Tabel 4.10 Hasil Presisi <i>Naive Bayes</i> Subset 5 (1000 Data)	65
Tabel 4.11 Hasil Recall <i>Naive Bayes</i> Subset 1 (200 Data)	66
Tabel 4.12 Hasil Recall <i>Naive Bayes</i> Subset 2 (400 Data)	67
Tabel 4.13 Hasil Recall <i>Naive Bayes</i> Subset 3 (600 Data)	67
Tabel 4.14 Hasil Recall <i>Naive Bayes</i> Subset 4 (800 Data)	68
Tabel 4.15 Hasil Recall <i>Naive Bayes</i> Subset 5 (1000 Data)	69
Tabel 4.16 Hasil Waktu training <i>Naive Bayes</i> Subset 1 (200 Data)	70
Tabel 4.17 Hasil Waktu training <i>Naive Bayes</i> Subset 2 (400 Data)	71
Tabel 4.18 Hasil Waktu training <i>Naive Bayes</i> Subset 3 (600 Data)	71
Tabel 4.19 Hasil Waktu training <i>Naive Bayes</i> Subset 4 (800 Data)	72

Tabel 4.20 Hasil Waktu training <i>Naive Bayes</i> Subset 5 (1000 Data).....	73
Tabel 4.21 Hasil F1 score <i>Naive Bayes</i> Subset 1 (200 Data)	74
Tabel 4.22 Hasil F1 score <i>Naive Bayes</i> Subset 2 (400 Data)	75
Tabel 4.23 Hasil F1 score <i>Naive Bayes</i> Subset 3 (600 Data)	76
Tabel 4.24 Hasil F1 score <i>Naive Bayes</i> Subset 4 (800 Data)	76
Tabel 4.25 Hasil F1 score <i>Naive Bayes</i> Subset 5 (1000 Data)	77
Tabel 4.26 Hasil MAE <i>Naive Bayes</i> Subset 1 (200 Data)	78
Tabel 4.27 Hasil MAE <i>Naive Bayes</i> Subset 2 (400 Data)	79
Tabel 4.28 Hasil MAE <i>Naive Bayes</i> Subset 3 (600 Data)	80
Tabel 4.29 Hasil MAE <i>Naive Bayes</i> Subset 4 (800 Data)	80
Tabel 4.30 Hasil MAE <i>Naive Bayes</i> Subset 5 (1000 Data)	81
Tabel 4.31 Hasil Akurasi <i>Support Vector Machine</i> Subset 1 (200 Data)	83
Tabel 4.32 Hasil Akurasi <i>Support Vector Machine</i> Subset 2 (400 Data)	83
Tabel 4.33 Hasil Akurasi <i>Support Vector Machine</i> Subset 3 (600 Data)	84
Tabel 4.34 Hasil Akurasi <i>Support Vector Machine</i> Subset 4 (800 Data)	85
Tabel 4.35 Hasil Akurasi <i>Support Vector Machine</i> Subset 5 (1000 Data)	85
Tabel 4.36 Hasil Presisi <i>Support Vector Machine</i> Subset 1 (200 Data)	87
Tabel 4.37 Hasil Presisi <i>Support Vector Machine</i> Subset 2 (400 Data)	87
Tabel 4.38 Hasil Presisi <i>Support Vector Machine</i> Subset 3 (600 Data)	88
Tabel 4.39 Hasil Presisi <i>Support Vector Machine</i> Subset 4 (800 Data)	89
Tabel 4.40 Hasil Presisi <i>Support Vector Machine</i> Subset 5 (1000 Data)	89
Tabel 4.41 Hasil Recall <i>Support Vector Machine</i> Subset 1 (200 data)	91
Tabel 4.42 Hasil Recall <i>Support Vector Machine</i> Subset 2 (400 data)	91
Tabel 4.43 Hasil Recall <i>Support Vector Machine</i> Subset 3 (600 data)	92

Tabel 4.44 Hasil Recall <i>Support Vector Machine</i> Subset 4 (800 data)	93
Tabel 4.45 Hasil Recall <i>Support Vector Machine</i> Subset 5 (1000 data)	93
Tabel 4.46 Hasil Waktu training <i>Support Vector Machine</i> Subset 1 (200 Data).....	95
Tabel 4.47 Hasil Waktu training <i>Support Vector Machine</i> Subset 2 (400 Data).....	95
Tabel 4.48 Hasil Waktu training <i>Support Vector Machine</i> Subset 3 (600 Data).....	96
Tabel 4.49 Hasil Waktu training <i>Support Vector Machine</i> Subset 4 (800 Data).....	97
Tabel 4.50 Hasil Waktu training <i>Support Vector Machine</i> Subset 5 (1000 Data).....	97
Tabel 4.51 Hasil F1 score <i>Support Vector Machine</i> Subset 1 (200 Data)	99
Tabel 4.52 Hasil F1 score <i>Support Vector Machine</i> Subset 2 (400 Data)	100
Tabel 4.53 Hasil F1 score <i>Support Vector Machine</i> Subset 3 (600 Data)	100
Tabel 4.54 Hasil F1 score <i>Support Vector Machine</i> Subset 4 (800 Data)	101
Tabel 4.55 Hasil F1 score <i>Support Vector Machine</i> Subset 5 (1000 Data)	102
Tabel 4.56 Hasil MAE <i>SVM</i> Subset 1 (200 Data)	103
Tabel 4.57 Hasil MAE <i>SVM</i> Subset 2 (400 Data)	104
Tabel 4.58 Hasil MAE <i>SVM</i> Subset 3 (600 Data)	105
Tabel 4.59 Hasil MAE <i>SVM</i> Subset 4 (800 Data)	105
Tabel 4.60 Hasil MAE <i>SVM</i> Subset 5 (1000 Data)	106
Tabel 4.61 Hasil Perbandingan Tingkat Performa	108

DAFTAR GAMBAR

Gambar 1.1 Jumlah vaksin seluruh negara yang sudah diberikan.....	3
Gambar 1.2 Jumlah vaksin indonesia yang sudah diberikan.....	4
Gambar 3.1. Metode Penelitian.....	35
Gambar 3.2. Alur Impelementasi Algoritma.....	36
Gambar 4.1. Model Arsitektur Arsitektur Naive Bayes.....	41
Gambar 4.2. Model Arsitektur SVM.....	42
Gambar 4.3. Usecase Diagram.....	43
Gambar 4.4. Activity Diagram SVM.....	44
Gambar 4.5. Activity Diagram Naive Bayes.....	45
Gambar 4.6. Class Diagram Controler.....	46
Gambar 4.7. Antarmuka Dataset.....	55
Gambar 4.8. Antarmuka Training Data.....	55
Gambar 4.9. Antarmuka Hasil Perbandingan.....	56
Gambar 4.10. Antarmuka Input Parameter Prediksi.....	57
Gambar 4.11. Antarmuka Hasil Prediksi.....	57
Gambar 4.12 Grafik Akurasi <i>Naive Bayes</i>	62
Gambar 4.13 Grafik Presisi <i>Naive Bayes</i>	66
Gambar 4.14 Grafik Recall <i>Naive Bayes</i>	70
Gambar 4.15 Grafik Waktu training <i>Naive Bayes</i>	74
Gambar 4.16 Grafik F1 score <i>Naive Bayes</i>	78
Gambar 4.17 Grafik MAE <i>Naive Bayes</i>	82
Gambar 4.18 Grafik Akurasi <i>Support Vector Machine</i>	86

Gambar 4.19 Grafik Presisi <i>Support Vector Machine</i>	90
Gambar 4.20 Grafik Recall <i>Support Vector Machine</i>	94
Gambar 4.21 Grafik Waktu training <i>Support Vector Machine</i>	99
Gambar 4.22 Grafik F1 score <i>Support Vector Machine</i>	103
Gambar 4.23 Grafik MAE <i>SVM</i>	107
Gambar 4.24 Perbandingan Performa Akurasi	108
Gambar 4.25 Perbandingan Performa Presisi	109
Gambar 4.26 Perbandingan Performa Recall	109
Gambar 4.27 Perbandingan Performa Waktu Training	110
Gambar 4.28 Perbandingan Performa F1-Score	111
Gambar 4.29 Perbandingan Performa MAE.....	111



INTISARI

Wabah penyakit baru yang disebabkan oleh virus corona (2019-nCoV) atau COVID-19 resmi dinyatakan sebagai pandemi global oleh Organisasi Kesehatan Dunia (WHO) pada 11 Maret 2020. Padahal pusat penyebaran virus di akhir tahun 2019 berada di kota wuhan, cina, kini virus tersebut telah menyebar ke seluruh masyarakat dunia dengan total lebih dari 41,5 juta kasus dan jumlah kematian lebih dari 1,1 juta per 23 oktober 2020. Kondisi tersebut berdampak langsung berdampak pada jutaan dan bahkan semuanya. komunitas dunia.

Mengingat penyebaran COVID-19 yang begitu cepat, salah satu cara yang sangat mungkin dilakukan untuk mencegah penyebaran virus ini adalah dengan mengembangkan vaksin. Keberadaan vaksinasi bagi masyarakat Indonesia dinilai kontroversial sehingga mengundang banyak kalangan untuk memberikan pendapat, aspirasi dan kritiknya, namun keterbatasan waktu dan tempat menghalangi aspirasi masyarakat untuk tersampaikan. Salah satu media yang banyak digunakan oleh masyarakat untuk memberikan pendapat tentang suatu hal adalah media sosial twitter.

Penelitian ini bertujuan untuk menganalisis respon masyarakat terhadap vaksinasi menjadi respon positif dan negatif. Selanjutnya opini publik juga akan dikelompokkan menggunakan metode yang dapat digunakan untuk mewujudkannya yaitu dengan menggunakan analisis sentimen dalam penelitian ini menggunakan Support Vector Machine Algorithm (SVM) dan Naive Bayes

Kata kunci—Covid-19, Twitter, Naive Bayes, SVM, Analisis Sentimen

ABSTRACT

The outbreak of a new disease caused by the corona virus (2019-nCoV) or COVID-19 was officially declared a global pandemic by the World Health Organization (WHO) on March 11, 2020. Even though the center of the spread of the virus at the end of 2019 was in the city of Wuhan, China, now The virus has spread throughout the world community with a total of more than 41.5 million cases and a death toll of more than 1.1 million as of October 23, 2020. These conditions have a direct impact on millions and even all of them. world community.

Given the rapid spread of COVID-19, one very possible way to prevent the spread of this virus is to develop a vaccine. The existence of vaccination for the people of Indonesia is considered controversial so that it invites many people to provide their opinions, aspirations and criticisms; but the limited time and place hinders the aspirations of the people to be conveyed. One of the media that is widely used by the public to provide opinions about a matter is social media twitter.

This study aims to analyze the public response to vaccination into positive and negative responses. Furthermore, public opinion will also be grouped using methods that can be used to make it happen, namely by using sentiment analysis in this study using the Support Vector Machine Algorithm (SVM) and Naive Bayes

Keywords—Covid-19, Twitter, Naïve Bayes, SVM, Sentiment Analysis.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Wabah penyakit baru yang diakibatkan oleh virus korona (2019-nCoV) atau yang biasa disebut dengan COVID-19 ditetapkan secara resmi sebagai pandemi global oleh World Health Organization (WHO) pada tanggal 11 Maret 2020 lalu. Meskipun pusat penyebaran virus tersebut pada akhir tahun 2019 lalu berada di Kota Wuhan, China, kini virus tersebut telah tersebar menjangkit ke seluruh masyarakat dunia dengan jumlah kasus sebanyak lebih dari 41,5 juta kasus dan jumlah kematian sebanyak lebih dari 1,1 juta jiwa per tanggal 23 Oktober 2020. Kondisi demikian memberikan dampak langsung kepada jutaan bahkan seluruh masyarakat dunia, sebagai akibat dari diberlakukannya protokol kesehatan yang harus ditetapkan pada seluruh aspek kegiatan, mulai dari pembatasan sosial hingga lockdown total sehingga menghambat seluruh kegiatan masyarakat. Efek lanjutan dari COVID-19 ini berpotensi membawa tantangan besar bagi sistem kesehatan dunia dan memiliki konsekuensi yang luas pada ekonomi global jika penyebaran virus tidak dikendalikan secara efektif.

Melihat pesatnya penyebaran COVID-19 dan bahaya yang akan muncul jika tidak segera ditangani, salah satu cara yang sangat mungkin untuk mencegah penyebaran virus ini adalah dengan mengembangkan vaksin. Vaksin tidak hanya melindungi mereka yang divaksinasi tetapi juga masyarakat luas dengan mengurangi penyebaran penyakit dalam populasi.

Dalam menyikapi hal tersebut, Pemerintah Indonesia juga turut aktif dalam rencana kegiatan vaksinasi yang akan diberikan kepada masyarakatnya. Dalam rangka melakukan percepatan penanganan penyebaran COVID-19 di Indonesia, Pemerintah Republik Indonesia telah mengeluarkan vaksinasi untuk masyarakat Indonesia pada akhir tahun 2020. Menteri Kesehatan Budi Gunadi Sadikin, menyebutkan bahwa rencana vaksinasi di Indonesia akan dilakukan dalam dua periode. Hal tersebut sudah dikonsultasikan kepada Indonesian Technical Advisory Group on Immunization (ITAGI) yang bertugas memberikan nasehat /advice kepada Menteri Kesehatan. Periode pertama akan dimulai pada Januari sampai dengan April 2021.

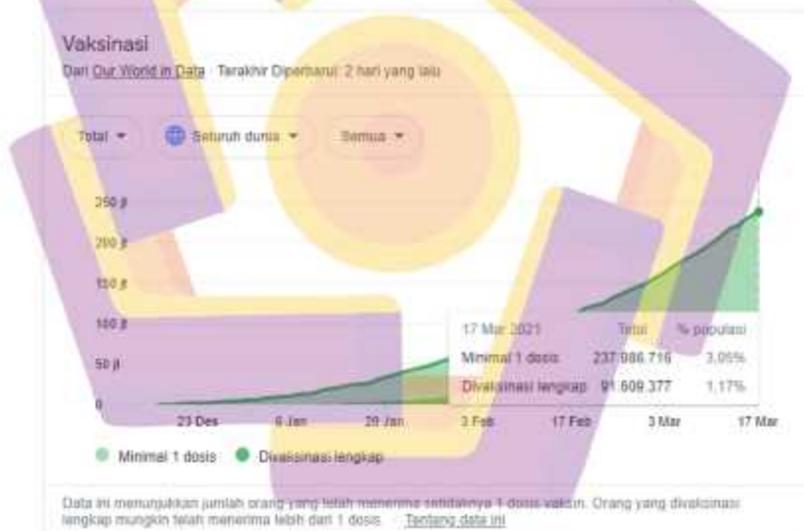
Pada periode pertama target yang akan mendapatkan vaksinasi adalah tenaga kesehatan dengan jumlah 1,3 juta orang, petugas publik 17,4 juta yaitu petugas yang sulit menjaga jarak secara efektif dan penduduk lanjut usia di atas usia 60 tahun sebanyak 21,5 juta. Periode kedua adalah periode April 2021 – Maret 2022 dengan jumlah penerima vaksin adalah 63,90 juta masyarakat dengan risiko penularan tinggi yang dikategorikan menurut kategori tempat tinggal atau kelas ekonomi dan sosial. Selanjutnya diikuti, sebanyak 77,4 juta masyarakat umum dengan pendekatan kluster sesuai ketersediaan vaksin Nareza, M.(2021).

Berdasarkan Keputusan Menteri Kesehatan Nomor H.K.01.07/Menkes/9860 /2020 tentang Penetapan jenis Vaksin untuk Pelaksanaan Vaksinasi Corona Virus Disease (Covid 19) diketahui bahwa telah ditetapkan enam jenis vaksin untuk proses vaksinasi di Indonesia. Adapun jenisnya adalah vaksin yang produksi oleh

P.T. Bio Farma (persero), Astra Zeneca, China National Pharmaceutical Group Corporation (Sinopharm), Moderna, Pfizer- BioNTech, dan Sinovac Biotech Ltd.

Pemerintah telah mendatangkan vaksin Covid 19 sebanyak 3 juta dari produsen Sinovac dari China dalam dua tahap, yaitu sebanyak 1,20 juta dosis sebagai tahap pertama pada 06 Desember 2020 dan tahap ke dua adalah sebesar 1,80 juta dosis pada 31 Desember 2020 (Nareza, M., 2021) .

Diberitakan oleh Tim Our World in Data pada tanggal 17 maret 2021 jumlah data vaksin yang sudah diberikan di seluruh negara dan Indonesia sebagai berikut :



Gambar 1.1 Jumlah vaksin seluruh negara yang sudah diberikan



Gambar 1.2 Jumlah vaksin Indonesia yang sudah diberikan

Dengan adanya vaksinasi untuk masyarakat Indonesia tersebut dinilai kontroversial sehingga mengundang banyak kalangan untuk memberikan pendapatnya, aspirasi dan kritikan, namun keterbatasan waktu dan ruang membuat aspirasi masyarakat tidak tersampaikan. Masyarakat bebas memberikan respon dan opininya di berbagai media. Salah satu media yang banyak digunakan oleh masyarakat untuk memberikan pendapatnya terhadap sesuatu adalah media sosial. Media sosial kini seolah merupakan suatu hal yang wajib dimiliki oleh seluruh masyarakat. Berdasarkan data dari Global Digital Statistic “Digital, Social & Mobile in 2019” di We Are Social (2019), pada tahun 2019 jumlah pengguna media sosial di Indonesia yaitu berjumlah lebih dari 150 juta pengguna. Salah satu media sosial yang paling banyak digunakan oleh masyarakat Indonesia adalah

media sosial twitter, yang mencakup lebih dari 52 persen dari total pengguna media sosial di Indonesia. Hal tersebut menunjukkan adanya peluang sumber data yang sangat besar yang dapat dimanfaatkan untuk menghasilkan suatu pengetahuan yang bermanfaat.

Disisi lain jejaring sosial merupakan salah satu tempat yang cocok untuk mengungkapkan segala hal yang ada, salah satunya adalah twitter, twitter adalah jejaring sosial yang memungkinkan pengguna mengirim karakter hingga 140 karakter yang sering disebut tweet atau kicauan. Pendapat tersebut haruslah dipertimbangkan sebagai bahan evaluasi sehingga vaksinasi yang akan dilakukan dapat berjalan dengan baik. Dengan memanfaatkan data dari media sosial twitter, penelitian ini bertujuan untuk menganalisis respon masyarakat terhadap vaksinasi dengan cara mengklasifikasikan respon tersebut ke dalam respon positif dan negatif. Selanjutnya juga akan dilakukan pengelompokan opini masyarakat menggunakan suatu metode yang dapat digunakan untuk mewujudkannya yaitu dengan menggunakan analisis sentiment yang di penelitian ini menggunakan Algoritma Support Vector Machine dan Naive Bayes untuk mengetahui topik pembicaraan yang sering dibahas oleh masyarakat terkait dengan adanya vaksinasi tersebut.

Penelitian yang dilakukan oleh Muhammad Syarifuddin ini untuk menganalisis sentimen opini publik mengenai covid-19 pada twitter menggunakan metode naïve bayes dan knn. Penelitian ini dimulai dari pencarian data di twitter menggunakan API Twitter, kemudian didapatkan kumpulan data tweet yang disebut data set sebanyak 1098 tweets, selanjutnya data tweet tentang COVID-19

perlu di klasifikasi apakah bersifat opini positif ataupun negatif, setelah data diklasifikasi tahap selanjutnya yaitu Data cleansing, penulis membuang karakter yang dapat menghambat proses olah data mining, dan menghapus tweet yang sama (double), serta penghapusan data yang memiliki nilai null, dilanjutkan dengan proses preprocessing, yaitu mengubah beberapa kata asing ke Bahasa Indonesia dan membuang beberapa karakter yang tidak diperlukan, dilanjutkan dengan pembobotan terms, yaitu menggunakan TIF-IDF dan GINI INDEX, yang merupakan algoritma dalam ilmu komputer yang digunakan dalam menghitung bobot dalam setiap kata yang umum digunakan, yang terdiri model klasifikasi Naïve Bayes dan KNN yaitu, model klasifikasi data dalam data mining yang menggunakan aplikasi rapidminer. Dilanjut dengan data testing, dilakukan dengan aplikasi rapidminer sehingga dapat diketahui secara jelas berapa nilai accuracy, presicion, recall, dan AUC dari setiap model klasifikasi, dan yang terakhir dilakukan perbandingan dari model klasifikasi Naïve Bayes dan KNN mana yang paling akurat dalam mengklasifikasikan data tweet COVID-19. Hasil penelitian dari kedua metode tersebut metode Naïve Bayes merupakan model klasifikasi yang paling akurat dibandingkan dengan KNN dengan nilai accuracy sebesar 63.21% (Muhammad, 2020).

Pada penelitian yang dilakukan oleh Rian Tineges, Agung Triayudi dan Ira Diana Sholihati ini berfokus pada Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). Penelitian ini dimulai dari pengumpulan data dengan melakukan crawling melalui API Twitter dengan menggunakan modul tweepy pada bahasa pemrograman

Python. Setelah proses pengumpulan data dikarenakan dataset tersebut termasuk dalam unstructured data (data tidak terstruktur). Sebelum dilakukan analisis lebih lanjut, terlebih dahulu dataset melalui proses text preprocessing untuk menghilangkan serta mengatasi noisy data agar hasil perhitungan optimal. Adapun alur proses text preprocessing antara lain, case folding, filtering, tokenizing, dan stopword removal. Setelah data bersih maka dilakukan proses pelabelan secara manual. Hasil yang didapat dari proses-proses tersebut yaitu dengan menerapkan metode Support Vector Machine untuk analisis sentimen terhadap layanan Indihome berdasarkan data tweet hasil evaluasi menggunakan Confusion Matrix, didapat akurasi sebesar 87% dengan ketepatan antara hasil prediksi dengan data sebenarnya (precision) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (recall) sebesar 95%, tingkat kesalahan semua data yang diprediksi (error rate) sebesar 13%, sedangkan untuk nilai perbandingan rata-rata precision dan recall (f1- score) adalah sebesar 90% (Rian, 2020).

Dalam penelitiannya (Ratino, 2020) menggunakan metode Support Vector Machine dan Naïve Bayes dengan penambahan Particle Swarm Optimization (PSO) untuk menghasilkan akurasi terbaik dalam menganalisa komentar Instagram terhadap informasi berkaitan dengan COVID-19. Dalam penelitian ini pengumpulan data mengambil dari Instagram pada akun World Health Organization (WHO) dengan postingan berkisar tentang covid-19. Dan hasil dari penelitian analisis sentimen terhadap data komentar Instagram mengenai Covid-19 yang dilakukan dengan Support Vector Machine dengan nilai akurasi 80,23% dan Naïve Bayes dengan nilai akurasi 78,02%. Dan hasil Support Vector Machine

dengan Particle Swarm Optimization dengan nilai akurasi 81,16% dan Naïve Bayes dengan Particle Swarm Optimization dengan memperoleh nilai akurasi 79,07%.

Untuk Mengetahui opini positif atau negatif suatu peristiwa atau isu maka (Pristiyono et al., 2021) melakukan analisis sentimen dengan merayapi data Twitter dengan kata kunci 'Vaccine Covid-19'. Dalam penelitian tersebut menggunakan metode naïve bayes. Hasil pengukuran sentimen dengan lebih dari 3,4 ribu tweet negatif, lebih dari 2,4 ribu tweet positif, dan sisa 301 tweet dianggap netral selama periode seminggu. Sedangkan pada minggu kedua dan ketiga Januari 2021. Analisis pada periode tersebut menunjukkan 39% sentimen positif, 56% sentimen negatif, dan 1% sentimen menguntungkan. Pendapat negatif pada periode tersebut muncul karena masyarakat tidak percaya bahwa vaksin itu aman.

Dalam penelitiannya Ardianne Luthfika Fairuz, Rima Dias Ramadhani dan Nia Annisa Ferani Tanjung akan dibahas mengenai sentimen analisis masyarakat terhadap pandemi COVID-19 pada media sosial Twitter. Pada penelitian ini dipilih Naïve Bayes dan K-Nearest Neighbor sebagai algoritma yang digunakan dalam mengklasifikasi komentar di media sosial Twitter. Hasil kinerja kedua algoritma yang digunakan dalam penelitian, yakni Naïve Bayes dan K-Nearest Neighbor, keduanya bekerja dengan baik. Nilai akurasi dari kedua metode yang digunakan untuk Naïve Bayes didapatkan nilai akurasi sebesar 85%, sementara untuk K-Nearest Neighbor didapatkan akurasi terbaik sebesar 82% (Ardianne et al, 2021).

Penelitian terdahulu oleh Cahyo Prianto dan Nisa Hanum Harani yang menggunakan algoritma naïve bayes proses pengambilan datanya dilakukan dengan menggunakan twitter API yang terhubung dengan Rstudio. Dan hasil yang didapat dari penelitian analisis sentiment ini yaitu pada data tweet Maret 2020, sentimen positif 51% dan sentimen negatif 49%. Dengan tingkat akurasi 0,7586, spesifisitas 0,6667, prevalensi 0,5862. Sedangkan data tweet Juni 2020 menunjukkan, 59% sentimen negatif dan 41% sentimen positif. Dengan tingkat akurasi 0,6486, spesifisitas 0,6111, prevalensi 0,5135 (Cahyo et al, 2020).

Pada penelitiannya Samsir, Ambiyar, Unung Verawardina, Firman Edi dan Ronal Watrianthos dilakukan penelitian yaitu analisis sentimen pembelajaran daring pada twitter di masa pandemi covid-19. Dalam penelitian tersebut menggunakan metode naïve bayes. Berdasarkan hasil percobaan didapatkan hasil bahwa metode penelitian ini menerapkan algoritma Naïve Bayes untuk klasifikasi sentimen. Analisis sentimen dalam study ini menggunakan data Twitter dengan kata kunci 'pembelajaran daring', 'vaksin', 'belajar', 'covid', 'daring', dan tagar #BelajarDariRumah yang difilter dengan kata kunci 'covid' dan 'rumah' pada twit-twit dalam bahasa Indonesia pada minggu pertama November 2020. Penelitian pada periode tersebut menunjukkan 30% sentimen positif, 69% sentimen negatif, dan 1% netral. Persepsi negatif dihasilkan karena ketidakpuasan masyarakat terhadap pembelajaran daring. Beberapa twit menunjukkan kekecewaan dengan kata 'stres' dan 'malas' merupakan kata yang memiliki frekuensi tinggi dalam percakapan pada periode tersebut. Berdasarkan hasil ini pembelajaran daring belum maksimal diterapkan di Indonesia pada masa

pandemi yang terlihat dari tingginya kekecewaan public pada awal November 2020 (Samsir, 2021).

Pada penelitian yang dilakukan oleh Deni Gunawan, Dwiza Riana, Dian Ardiansyah, Fajar Akbar dan Salman Alfarizi ini menggunakan metode Support Vector Machine dan Naïve bayes yang dipadu dengan Algoritma Genetika yang akan mengoptimasi parameter Support Vector Machine. Penelitian ini dimulai dari dengan mengumpulkan masing-masing data, dan data dikelompokkan menjadi positif dan negatif dengan total data sebanyak 2643 data tweet . Model yang diuji akan menghasilkan nilai accuracy, precision, recall dan AUC dari setiap algoritma. Hasil pengujian data tweet mengenai calon gubernur jawa barat periode 2018-2023 dengan Algoritma Support Vector Machine menghasilkan rata-rata akurasi 92,61% dengan AUC 0,950, Algoritma Naïve Bayes menghasilkan rata-rata akurasi 93,29% dengan AUC 0,525, Algoritma Support Vector Machine berbasis Genetic Algorithm menghasilkan rata-rata akurasi 93,03% dengan AUC 0,869 dan Algoritma Naive Bayes berbasis Genetic Algorithm menghasilkan rata-rata akurasi 92,85% dengan AUC 0,543. Dari hasil di atas dapat disimpulkan bahwa model algoritma Support Vector Machine berbasis Genetic Algorithm adalah model algoritma terbaik dalam penelitian ini dan dapat memberikan hasil terbaik dalam pengujian dan pengklasifikasian analisis sentiment (deni et al, 2020).

Berdasarkan latar belakang yang diangkat algoritma naïve bayes dikenal mampu melakukan klasifikasi dengan baik berdasarkan penalaran probabilitas yang algoritma tersebut miliki, akan tetapi algoritma lain yaitu Support Vector

Machine (SVM) juga merupakan salah algoritma machine learning yang dikenal cukup baik dalam melakukan klasifikasi berdasarkan pembobotan yang diproses dalam algoritma tersebut. Oleh karena itu pada penelitian ini akan dilakukan peneliti yaitu malakukan perbandingan model algoritma naïve bayes dan support vector machine dalam melakukan analisis sentimen pada data postingan twitter berdasarkan kata kunci vaksin covid-19.

1.2. Rumusan Masalah

Dari latar belakang tersebut, rumusan masalah yang akan dibahas pada penelitian ini, antara lain :

- a. Berapakah nilai performa dari model algoritma algoritma support vector machine dan navie bayes berdasarkan nilai peforma berdasarkan evaluasi akurasi, presisi, recall, fl score, waktu training dan MAE ?
- b. Berapakah nilai analisis tren kenaikan atau penurunan data pada performa algoritma naïve bayes dan algoritma support vector machine terhadap jumlah record dataset ?

1.3. Batasan Masalah

Batasan masalah dari penelitian adalah sebagai berikut :

- a. Menggunakan dataset yang berasal dari data postingan twitter berdasarkan kata kunci vaksin covid-19 dengan total data 1000 data.
- b. Dataset yang digunakan merupakan data berformat text yang didapatkan melalui media API Twitter.

- c. Dataset yang digunakan bersumber dari postingan twitter dengan komposisi data yang terbagi menjadi 3 bagian yaitu positif, negatif dan netral.
- d. Jumlah pembagian dataset untuk data training dan testing adalah 20% untuk testing dan 80% untuk training (Pradito et al, 2013).
- e. Dataset yang digunakan adalah data postingan pendapat masyarakat indonesia mengenai vaksin covid-19 tanpa memiliki kencerungan dengan golongan atau instansi manapun.
- f. Fokus penelitian adalah membandingkan algoritma naive bayes dan algoritma support vector machine untuk klasifikasi.
- g. Analisis data dilakukan melalui pre-prosesing menggunakan teknik proses case folding, tokenizing, filtering, dan stemming.
- h. Algoritma yang digabungkan adalah algoritma naive bayes dan algoritma support vector machine untuk klasifikasi.
- i. Model pengukuran validasi performa model algoritma klasifikasi menggunakan metode atau teknik K-Fold Cross-Validation
- j. Nilai performa yang diukur atau dihitung adalah performa akurasi, presisi, recall, f1 score, waktu training dan MAE.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan sebelumnya, tujuan penelitian ini, antara lain :

- a. Mengetahui nilai performa berdasarkan hasil penelitian dari implementasi algoritma naive bayes dan algoritma support vector machine berdasarkan

nilai performa akurasi, presisi, fl score, recall, waktu training dan MAE terhadap dataset twitter vaksin covid-19.

- b. Mengetahui analisis tren kenaikan/penurunan data data pada performa algoritma naïve bayes dan algoritma support vector machine terhadap jumlah record dataset.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini, antara lain :

- a. Dapat memberikan pengetahuan baru tentang sentimen analisis menggunakan data postingan twitter melalui kata kunci vaksin covid-19.
- b. Dapat memberikan rekomendasi algoritma sebagai landasan terapan pengembangan sentimen analisis dengan menggunakan data pada twitter.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Dalam melakukan penelitian perlu diberikan rujukan atau tinjauan pustaka sebagai acuan dasar penelitian, sehingga penulis melakukan komparasi dengan penelitian dibidang lain terkait metode yang akan digunakan. Salah satunya adalah penelitian yang di lakukan oleh Muhammad Syarifuddin yang melakukan analisis sentimen opini publik mengenai covid-19 pada twitter menggunakan metode naïve bayes dan knn menggunakan Open API Twitter, anlsisi sentimen dilakukan dengan menggunakan dataset sebanyak 1098 tweets yang akan diklasifikasikan menjadi opini positif ataupun negatif, setelah data diklasifikasi tahap selanjutnya yaitu Data cleansing, penulis membuang karakter yang dapat menghambat proses olah data mining, dan menghapus tweet yang sama (double), serta penghapusan data yang memiliki nilai null , dilanjut dengan proses preprocessing untuk kemudian dilakukan analisis sentimen oleh algoritma knn dan naïve bayes. Hasil penelitian dari kedua metode tersebut metode Naïve Bayes mendapatkan nilai akurasi 72,68% merupakan model klasifikasi yang paling akurat dibandingkan dengan KNN dengan nilai akurasi sebesar 63.21%.

Pada penelitian yang dilakukan oleh Rian Tineges, Agung Triayudi dan Ira Diana Sholihati ini berfokus pada Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). Penelitian ini dimulai dari pengumpulan data dengan melakukan crawling melalui

API Twitter dengan menggunakan modul tweepy pada bahasa pemrograman Python. Setelah proses pengumpulan data dikarenakan dataset tersebut termasuk dalam unstructured data (data tidak terstruktur). Sebelum dilakukan analisis lebih lanjut, terlebih dahulu dataset melalui proses text preprocessing untuk menghilangkan serta mengatasi noisy data agar hasil perhitungan optimal. hasil evaluasi menggunakan Confussion Matrix, didapat akurasi sebesar 87% dengan ketepatan antara hasil prediksi dengan data sebenarnya (precision) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (recall) sebesar 95%, tingkat kesalahan semua data yang diprediksi (error rate) sebesar 13%, sedangkan untuk nilai perbandingan rata-rata precision dan recall (f1- score) adalah sebesar 90%.

Penelitiannya (Ratino, 2020) mencoba menerapkan metode Support Vector Machine dan Naïve Bayes dengan penambahan Particle Swarm Optimization (PSO) untuk menghasilkan akurasi terbaik dalam menganalisa komentar Instagram terhadap informasi berkaitan dengan COVID-19. Dalam penelitian ini pengumpulan data mengambil dari Instagram pada akun World Health Organization (WHO) dengan postingan berkisar tentang covid-19. Dan hasil dari penelitian analisis sentimen terhadap data komentar Instagram mengenai Covid-19 yang dilakukan dengan Support Vector Machine dengan nilai akurasi 80,23% dan Naïve Bayes dengan nilai akurasi 78,02%. Dan hasil Support Vector Machine dengan Particle Swarm Optimization dengan nilai akurasi 81,16% dan Naïve Bayes dengan Particle Swarm Optimization dengan nilai akurasi 79,07%.

Untuk Mengetahui opini positif atau negatif suatu peristiwa atau isu maka (Pristiyono et al., 2021) melakukan analisis sentimen dengan merayapi data Twitter dengan kata kunci 'Vaccine Covid-19. Dalam penelitian tersebut menggunakan metode naïve bayes. Hasil pengukuran sentimen dengan lebih dari 3,4 ribu tweet negatif, lebih dari 2,4 ribu tweet positif, dan sisa 301 tweet dianggap netral selama periode seminggu. Sedangkan pada minggu kedua dan ketiga Januari 2021. Analisis pada periode tersebut menunjukkan 39% sentimen positif, 56% sentimen negatif, dan 1% sentimen menguntungkan. Pendapat negatif pada periode tersebut muncul karena masyarakat tidak percaya bahwa vaksin itu aman.

Dalam penelitiannya Ardianne Luthfika Fairuz, Rima Dias Ramadhani dan Nia Annisa Ferani Tanjung akan dibahas mengenai sentimen analisis masyarakat terhadap pandemi COVID-19 pada media sosial Twitter. Pada penelitian ini akan dipilih Naïve Bayes dan K-Nearest Neighbor sebagai algoritma yang akan digunakan untuk mengklasifikasi komentar pada media sosial Twitter. Hasil dari penelitian ini kinerja kedua algoritma yang digunakan dalam penelitian, yakni Naïve Bayes dan K-Nearest Neighbor, keduanya bekerja dengan baik. Nilai akurasi dari kedua metode yang digunakan untuk Naïve Bayes didapatkan nilai akurasi sebesar 85%, sementara untuk K-Nearest Neighbor didapatkan akurasi terbaik sebesar 82%.

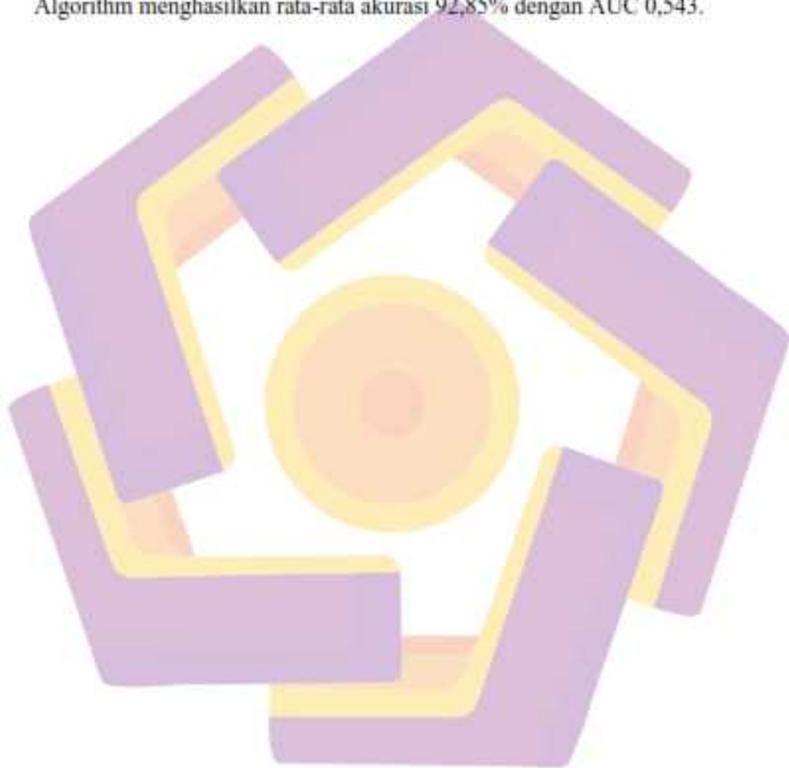
Cahyo Prianto dan Nisa Hanum Harani mencoba melakukan sentimen analisis dengan menggunakan algoritma naïve bayes proses pengambilan datanya dilakukan dengan menggunakan twitter API yang terhubung dengan Rstudio. Dan

hasil yang didapat dari penelitian analisis sentiment ini yaitu pada data tweet Maret 2020, sentimen positif 51% dan sentimen negatif 49%. Dengan tingkat akurasi 0,7586, spesifisitas 0,6667, prevalensi 0,5862. Sedangkan data tweet Juni 2020 menunjukkan, 59% sentimen negatif dan 41% sentimen positif. Dengan tingkat akurasi 0,6486, spesifisitas 0,6111, prevalensi 0,5135.

Pada penelitiannya Samsir, Ambiyar, Unung Verawardina, Firman Edi dan Ronal Watrianthos dilakukan penelitian yaitu analisis sentimen pembelajaran daring pada twitter di masa pandemi covid-19. Dalam penelitian tersebut menggunakan metode naïve bayes. Berdasarkan hasil percobaan didapatkan hasil bahwa metode penelitian ini menerapkan algoritma Naïve Bayes untuk klasifikasi sentimen. Analisis sentimen dalam study ini menggunakan data Twitter dengan kata kunci 'pembelajaran daring', 'vaksin', 'belajar', 'covid', 'daring', dan tagar #BelajarDariRumah yang difilter dengan kata kunci 'covid' dan 'rumah' pada twit-twit dalam bahasa Indonesia pada minggu pertama November 2020. Penelitian pada periode tersebut menunjukkan 30% sentimen positif, 69% sentimen negatif, dan 1% netral. Persepsi negatif dihasilkan karena ketidakpuasan masyarakat terhadap pembelajaran daring.

Deni Gunawan, Dwiza Riana, Dian Ardiansyah, Fajar Akbar dan Salman Alfarizi ini menggunakan metode Support Vector Machine dan Naïve bayes yang dipadu dengan Algoritma Genetika yang akan mengoptimasi parameter Support Vector Machine. Penelitian ini dimulai dari dengan mengumpulkan masing-masing data, dan data dikelompokkan menjadi positif dan negatif dengan total data sebanyak 2643 data tweet. Algoritma Support Vector Machine menghasilkan

rata-rata akurasi 92,61% dengan AUC 0,950, Algoritma Naïve Bayes menghasilkan rata-rata akurasi 93,29% dengan AUC 0,525, Algoritma Support Vector Machine berbasis Genetic Algorithm menghasilkan rata-rata akurasi 93,03% dengan AUC 0,869 dan Algoritma Naive Bayes berbasis Genetic Algorithm menghasilkan rata-rata akurasi 92,85% dengan AUC 0,543.



2.2.Keaslian Penelitian

Tabel 2.1 Matriks literatur review dan posisi penelitian Analisis Sentimen Twitter Vaksin covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn	Muhammad Syarifuddin, Inti Nusa Mandiri. 2020	Penelitian ini mencoba melakukan analisis sentimen opini publik mengenai covid-19 pada twitter menggunakan metode naive bayes dan knn	Pada penelitian ini menghasilkan kesimpulan dari kedua metode tersebut metode Naive Bayes mendapatkan nilai akurasi 72,68% merupakan model klasifikasi yang paling akurat dibandingkan dengan KNN dengan nilai akurasi sebesar 63,21%	Perlu adanya parameter pengujian lain selain akurasi serta perlu dilakukan validasi pengujian menggunakan beberapa fold atau cluster pengujian	Pada penelitian yang akan dilakukan menggunakan 6 variabel pengujian yaitu akurasi, presisi, recall, f1 score, waktu training dan MAE. Pengujian juga dilakukan dengan menggunakan 5 fold untuk mendapatkan hasil yang lebih valid
2	Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)	Rian Tingeges, Agung Triayudi, Iru Diana Sholihati. Jurnal Media Informatika Budidarma. 2020	Pada penelitian ini berfokus pada analisis sentimen terhadap layanan indihome berdasarkan twitter dengan metode klasifikasi support vector machine (SVM).	evaluasi menggunakan confusion matrix, didapat akurasi sebesar 87% dengan ketepatan antara hasil prediksi dengan data sebenarnya (precision) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (recall) sebesar 95%, tingkat kesalahan semua data yang diprediksi (error rate)	Pada penelitian ini pengujian hanya dilakukan dengan menggunakan 1 iterasi pengujian serta tidak adanya perbandingan hasil pengujian untuk mengevaluasi hasil performa algoritma	Penelitian yang akan dilakukan mencoba melakukan validasi pengujian dengan 5 kali iterasi atau fold untuk mendapatkan nilai performa terbaik kemudian menerapkan 2 algoritma yaitu naive bayes dan SVM untuk dilakukan perbandingan performa

Tabel 2.1 Matriks literatur review dan posisi penelitian

Analisis Sentimen Twitter Vaksin covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				sebesar 13%, sedangkan untuk nilai perbandingan rata-rata precision dan recall (f1-score) adalah sebesar 90%		
3	Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naive Bayes	Ratino, Noor Hafidz, Sita Anggraen, Windu Gata, Jurnal JUPITER. 2020.	Pada penelitian ini mencoba menerapkan metode Support Vector Machine dan Naive Bayes dengan penambahan Particle Swarm Optimization (PSO) untuk menghasilkan akurasi terbaik dalam menganalisa komentar Instagram terhadap informasi berkaitan dengan COVID-19	Hasil dari penelitian analisis sentimen terhadap data komentar Instagram mengenai Covid-19 yang dilakukan dengan Support Vector Machine dengan nilai akurasi 80,23% dan Naive Bayes dengan nilai akurasi 78,02%. Dan hasil Support Vector Machine dengan Particle Swarm Optimization dengan nilai akurasi 81,16% dan Naive Bayes dengan Particle Swarm Optimization dengan nilai akurasi 79,07%.	Pengujian nilai performa hanya menggunakan 1 kali iterasi pengujian dan 1 parameter pengujian	Penelitian ini akan mencoba melakukan perbandingan dan analisis performa dengan menggunakan 5 kali iterasi untuk validasi pengujian serta penggunaan 6 parameter pengujian yaitu akurasi, presisi, recall, f1 score, waktu training dan MAE.
4	Sentiment analysis of COVID-19 vaccine in Indonesia using Naive	Pristiyono, Mulkan Ritonga, Muhammad Ali Al	Pada penelitian ini mencoba melakukan analisis sentimen	Pada penelitian memiliki hasil pengukuran sentimen dengan lebih dari 3,4 ribu tweet negatif, lebih dari 2,4	Pengujian dilakukan hanya untuk mengetahui nilai positif, negatif dan	Penelitian ini mencoba menerapkan dan membandingkan analisis sentimen dengan.

Tabel 2.1 Matriks literatur review dan posisi penelitian

Analisis Sentimen Twitter Vaksin covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Bayes Algorithm	Ihsan, Agus Anjar, Fauziah Hanum Rambe. Annual Conference on Computer Science and Engineering Technology (AC2SET). 2020.	dengan merayapi data Twitter dengan kata kunci 'Vaccine Covid-19	ribu tweet positif, dan sisa 301 tweet dianggap netral selama periode seminggu. Sedangkan pada minggu kedua dan ketiga Januari 2021. Analisis pada periode tersebut menunjukkan 39% sentimen positif, 56% sentimen negatif, dan 1% sentimen menguntungkan	netral tanpa adanya pengujian peforma terhadap algoritma yang digunakan	menggunakan algoritma naive bayes dan SVM, selain untuk mengetahui nilai positif, negatif dan netral penelitian ini juga mencoba menerapkan 6 parameter pengujian dengan akurasi, presisi, recall, f1score, waktu training dan MAE untuk mengukur peforma dari algoritma yang digunakan.
5	Analisis Sentimen Covid-19 Sebagai Pandemi Media Sosial	Cahyo Prianto, Nisa Hanum Harani, Jurnal Internasional Sistem Informasi & Teknologi. 2020	Pada penelitian ini mencoba analisis sentimen pembelajaran daring pada twitter di masa pandemi covid-19 menggunakan metode naive bayes.	hasil yang didapat dari penelitian analisis sentiment ini yaitu pada data tweet Maret 2020, sentimen positif 51% dan sentimen negatif 49%. Dengan tingkat akurasi 0,7586, spesifisitas 0,6667, prevalensi 0,5862. Sedangkan data tweet Juni 2020 menunjukkan, 59% sentimen negatif dan 41% sentimen positif. Dengan tingkat akurasi 0,6486, spesifisitas 0,6111.	Pengujian hanya dilakukan dengan menggunakan 1 kali iterasi untuk validasi pengujian, tidak adanya perbandingan untuk mengukur peforma yang didapatkan	Penelitian ini mencoba menerapkan dan membandingkan analisis sentimen dengan menggunakan algoritma naive bayes dan SVM, validasi pengujian juga dilakukan dengan menggunakan 5 literasi untuk validasi pengukuran algoritma

Tabel 2.1 Matriks literatur review dan posisi penelitian

Analisis Sentimen Twitter Vaksin covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				prevalensi 0,5135		
6	Komparasi Algoritma Support Vector Machine Dan Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023	Deni, G., Dwiza, R., Dian, A., Fajar, A., Salma, A. . Jurnal Teknik Komputer AMIK BSI, 2020	Pada penelitian ini mencoba melakukan Komparasi algoritma support vector machine dan naive bayes dengan algoritma genetika pada analisis sentimen calon gubernur jabar 2018-2023	Penelitian ini mendapatkan nilai akurasi dari kedua metode yang digunakan untuk Naive Bayes didapatkan nilai akurasi sebesar 85%, sementara untuk K-Nearest Neighbor didapatkan akurasi terbaik sebesar 82%.	Pada penelitian ini hanya menggunakan 1 parameter pengujian yaitu akurasi dan dilakukan dengan 1 kali iterasi pengujian	Pada penelitian ini akan dilakukan pengujian peforma dengan parameter pengujian akurasi, presisi, recall, waktu training dan MAE yang dilakukan melalui 5 iterasi untuk validasi hasil pengujian

2.3. Landasan Teori

2.3.1. Analisis Sentimen

Analisis sentimen merupakan sebuah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi seseorang terhadap suatu barang, organisasi, orang, masalah konkrit, peristiwa. Ada beberapa penamaan terhadap studi ini, yaitu analisis sentimen, penambangan opini (opinion mining), ekstraksi opini (opinion extraction), penambangan sentimen (sentiment mining) dimana semuanya sekarang berada dibawah ranah analisis sentimen atau penambangan opini. Dalam dunia industri kata "analisis sentimen" sering digunakan tetapi di dunia pendidikan kata "analisis sentimen" dan "opinion mining" sering digunakan. Kata "analisis sentimen" pertama muncul tahun 2003 oleh Nasukawa dan Yi, sementara "opinion mining" muncul pada tahun 2003 oleh Dave. Lawrence dan Pennock. (Liu, 2012)

Meskipun linguistik dan pemrosesan bahasa natural (Natural Language Processing) disingkatnya NLP mempunyai sejarah lama, hanya sedikit penelitian sudah dilakukan tentang opini orang dan sentimen sebelum tahun 2000. Semenjak itu bidangnya sudah menjadi sangat aktif sebagai area penelitian. Ada beberapa alasan mengapa demikian, pertama memiliki penerapan yang luas, hampir disemua bidang. Analisis sentimen pada ranah sekitar industri juga sudah maju akibat adanya perkembangan dari penggunaan komersial yang menyediakan motivasi kuat untuk penelitian. Kedua, menyediakan banyak masalah penelitian yang menantang yang belum dipelajari sebelumnya.

2.3.2. Twitter API

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh Twitter Inc., yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan Tweets (Twitter, 2013). Mikroblog adalah salah satu jenis alat komunikasi covid dimana pengguna dapat memperbarui status tentang mereka yang sedang memikirkan dan melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu. Tweets adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. Tweets bisa dilihat secara publik, namun pengirim dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Pengguna dapat melihat Tweets pengguna lain yang dikenal dengan sebutan pengikut (follower).

Ada beberapa cara untuk mendapatkan sekumpulan data tweet pada Twitter yang salah satunya dengan menggunakan Twitter API. Berdasarkan pada halaman situs resmi Twitter, sebuah API (Application Programming Interface) merupakan salah satu cara computer "berbicara" satu sama lain sehingga dapat memesan dan mengantar informasi. Agar dapat memesan informasi maka pengguna harus memperbolehkan aplikasi perangkat lunak mengakses sebuah informasi spesifik, seperti nomor telepon dimana setiap pengguna memiliki nomor yang unik.

Untuk dapat mengakses Twitter API maka pengguna harus terlebih dahulu mendaftarkan aplikasinya. Setelah mendaftar maka pengguna dapat

mengakses bermacam-macam tweet dari suatu akun spesifik dengan menggunakan suatu kata kunci tertentu

2.3.3. Naive Bayes

Naive Bayes adalah sebuah Naive dengan dua penafsiran berbeda, dalam penafsiran Bayes, Naive ini menyatakan seberapa jauh derajat kepercayaan subjektif harus berubah secara rasional ketika ada petunjuk baru. Dalam penafsiran frekuents, Naive ini menjelaskan representasi *invers* probabilitas dua kejadian (Kusumadewi, 2003).

Naive Bayes ditemukan oleh Reverend Thomas Bayes pada abad ke-18. Naive ini merupakan dasar dari statistika Bayes dan dapat diterapkan dalam banyak bidang seperti sains, rekayasa, ilmu ekonomi, teori permainan kedokteran, hukum, dan lain sebagainya.

Evidence Tunggal (E) dan hipotesis tunggal (H)

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Keterangan :

$P(H|E)$ = probabilitas hipotesis H terjadi jika evidence E terjadi

$P(E|H)$ = probabilitas munculnya evidence E jika hipotesis H terjadi

$P(H)$ = probabilitas hipotesis H tanpa memandang evidence apapun

$P(E)$ = probabilitas evidence E tanpa memandang apapun

1. Evidence tunggal (E) dan hipotesis ganda (H_1, H_2, \dots, H_n)

$$P(H_i|E) = \frac{P(E|H) \times P(H)}{\sum_{k=1}^n P(E|H) \times P(H_k)}$$

Keterangan

$P(H_i|E)$ = probabilitas hipotesis H_i benar terjadi jika diberikan evidence E

$P(E_i|H)$ = probabilitas munculnya evidence E jika diketahui hipotesis H_i Benar

$P(E_i|H)$ = hipotesis H_i benar terjadi H_i (menurut hasil sebelumnya tanpa memandang evidence apapun

n = jumlah hipotesis yang mungkin.

2. Evidence ganda dan hipotesis ganda

$$P(H_i|E_1E_2...E_m) = \frac{P(E_1|H_i) \times P(E_2|H_i) \times \dots \times P(E_m|H_i) \times P(H_i)}{\sum_{k=1}^n P(E_1E_2...E_m|H_k) \times P(H_k)}$$

Akan tetapi pengaplikasian tersebut tidak mungkin karena harus mengetahui semua probabilitas bersyarat dari semua kombinasi, maka persamaan tersebut diganti dengan persamaan :

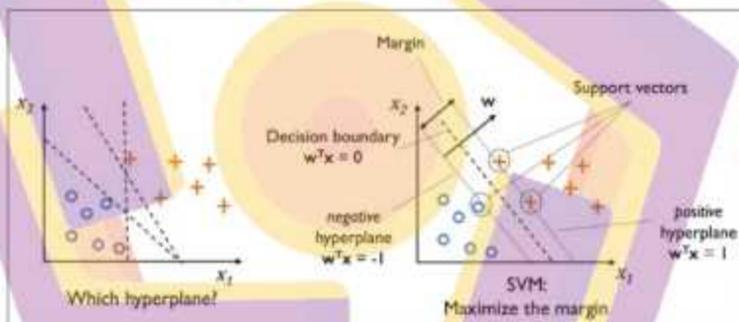
$$P(H_i|E_1E_2...E_m) = \frac{P(E_1|H_i) \times P(E_2|H_i) \times \dots \times P(E_m|H_i) \times P(H_i)}{\sum_{k=1}^n P(E_1|H_k) \times P(E_2|H_k) \times \dots \times P(E_m|H_k) \times P(H_k)}$$

2.3.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya.

SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear.

SVM digunakan untuk mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas. Hyperplane adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas. Dalam 2-D fungsi yang digunakan untuk klasifikasi antar kelas disebut sebagai line whereas, fungsi yang digunakan untuk klasifikasi antar kelas dalam 3-D disebut plane similarly, sedangkan fungsi yang digunakan untuk klasifikasi di dalam ruang kelas dimensi yang lebih tinggi disebut hyperplane.



Gambar 2.1 Hyperplane memisahkan dua kelas positif (+1) dan negatif(-1)

Hyperplane yang ditemukan SVM diilustrasikan seperti Gambar 1 posisinya berada ditengah-tengah antara dua kelas, artinya jarak antara hyperplane dengan objek-objek data berbeda dengan kelas yang berdekatan (terluar) yang diberi tanda bulat kosong dan positif. Dalam SVM objek data terluar yang paling dekat dengan hyperplane disebut support vector. Objek yang disebut support vector paling sulit diklasifikasikan dikarenakan posisi yang hampir tumpang

tindih (overlap) dengan kelas lain. Mengingat sifatnya yang kritis, hanya support vector inilah yang diperhitungkan untuk menemukan hyperplane yang paling optimal oleh SVM.

2.3.5. Metode Pengujian Confusion Matrix

Dalam mengevaluasi performance algoritma dari Machine Learning (ML), kita menggunakan acuan Confusion Matrix. Confusion Matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma ML. Berdasarkan Confusion Matrix, kita bisa menentukan Accuracy, Precision, Recall, F1 score, Waktu training dan MAE. Dalam tulisan ini, saya berbagi tentang pengukuran performance hasil algoritma ML. Tulisan ini berdasarkan tulisan ini yang kemudian saya tulis dengan bahasa sendiri untuk meningkatkan pemahaman (serta dilengkapi dengan studi kasus, dengan menggunakan contoh kasus prediksi mahasiswa Drop Out (DO) untuk memudahkan memahami Accuracy, Precision, Recall, F1 score, Waktu training dan MAE.

Pengukuran Performance

Accuracy

Merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data. Akurasi menjawab pertanyaan "Berapa persen mahasiswa yang benar diprediksi DO dan Tidak DO dari keseluruhan mahasiswa"

$$\text{Akurasi} = (TP + TN) / (TP + FP + FN + TN)$$

pada contoh kasus di atas, Akurasi = $(4+3) / (4+2+1+3) = 7/10 = 70\%$

Precision

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Precision menjawab pertanyaan “Berapa persen mahasiswa yang benar DO dari keseluruhan mahasiswa yang diprediksi DO?”

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

pada contoh kasus di atas, $\text{Precision} = 4 / (4+2) = 4/6 = 67\%$.

Recall (Sensitifitas)

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Recall menjawab pertanyaan “Berapa persen mahasiswa yang diprediksi DO dibandingkan keseluruhan mahasiswa yang sebenarnya DO”.

$$\text{Recall} = (\text{TP}) / (\text{TP} + \text{FN})$$

pada contoh kasus di atas $\text{Recall} = 4 / (4+1) = 4/5 = 80\%$.

F1 Score

F1 Score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Waktu Training (Times)

Merupakan rasio pengukuran waktu dimulainya trainin hingga selesai. Waktu training mengukur performa seberapa cepat algoritma dapat menyelesaikan masalah perhitungan.

$$\text{Waktu Training} = \text{Start (t start)} - \text{Finish (t end)}$$

Mean Absolute Error (MAE)

MAE (Mean Absolute Error) adalah rata-rata selisih mutlak nilai sebenarnya (aktual) dengan nilai prediksi (peramalan). MAE digunakan untuk mengukur keakuratan suatu model statistik dalam melakukan prediksi atau peramalan.

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i|$$

dimana:

n adalah ukuran sampel

A_i adalah nilai data aktual ke- i

F_i adalah nilai data peramalan ke- i

Karena pada rumus MAE di atas terdapat tanda mutlak ($| \ |$), maka nilai MAE akan selalu bernilai positif.

BAB III

METODE PENELITIAN

3.1. Metodologi Penelitian

Penelitian ini melakukan pengujian tingkat akurasi yang tertinggi menggunakan metode naïve bayes dan support vector machine dengan jumlah data set yang sama. Pengujian ini bertujuan untuk mengetahui metode yang lebih akurat dan tepat dalam melakukan sentimen analisis pada data twitter melalui kata kunci vaksin covid-19.

3.1.1. Jenis, Sifat, dan Pendekatan Penelitian

Adapun jenis, sifat dan pendekatan penelitian yang akan dilakukan pada penelitian ini sebagai berikut :

1. Jenis Penelitian Eksperimen

Penelitian ini memperoleh data dari postingan twitter berdasarkan kata kunci vaksin covid-19 dan membandingkan tingkat akurasi metode naïve bayes dan support vector machine menggunakan data tersebut. Data yang diperoleh dari twitter diambil dari data tahun 2021 berkisar 1000 data melalui kata vaksin covid-19 akan diuji untuk mendapatkan nilai performa tertinggi.

2. Sifat Penelitian Deskriptif

Tujuan dari penelitian ini untuk mengetahui perbandingan tingkat akurasi, presisi, recall, F1 score, waktu training dan MAE dengan melakukan sentimen analisis pada data twitter melalui kata vaksin

covid-19 berjumlah 1000 data menggunakan algoritma naïve bayes dan support vector machine, dan untuk mengetahui apakah proses pre-processing pada pengolahan dataset menjadi kunci baik atau buruknya suatu proses klasifikasi serta jumlah dataset yang diolah menjadi kedalam validasi subset sebanyak 5 kali perulangan yang akan berpengaruh terhadap performa kedua algoritma.

3. Pendekatan Penelitian Kuantitatif

Pada penelitian ini menggunakan pendekatan kuantitatif yang nantinya hasil dari penelitian ini merupakan informasi-informasi berupa angka dan diagram hasil dari eksperimen perbandingan metode naïve bayes dan SVM yang dilakukan. Pengumpulan data dilakukan melalui hasil eksperimen yang kemudian data tersebut dilakukan perbandingan dan analisis seperti dibuatkannya tabel dan diagram untuk melihat perbandingan metode mana yang paling baik dalam melakukan sentimen analisis menggunakan data twitter.

3.1.2. Metode Pengumpulan Data

Penelitian ini menggunakan tahap eksperimen sebagai metode pengumpulan data, namun selain itu peneliti juga mengumpulkan data awal sebagai bahan referensi melalui pengambilan data twitter melalui kata kunci vaksin covid-19. Tahap eksperimen dilakukan dengan melakukan perbandingan model algoritma *naïve bayes* dan *support vector machine* sebagai metode yang akan dibandingkan, kemudian hasil perbandingan kedua model algoritma tersebut di evaluasi berdasarkan tingkat akurasi, presisi, recall, F1 Score, waktu training

dan MAE untuk dilihat hasil perbandingannya. Proses pengumpulan data dimulai dari penerapan API twitter untuk mendapatkan parameter-parameter yang akan diteliti, dan selanjutnya akan dilanjutkan proses pengamambilan data pada objek yang akan diteliti, dalam kasus ini peneliti mengambil data data twitter melalui kata kunci vaksin covid-19 yang akan digunakan sebagai dataset, dataset yang akan digunakan merupakan data selama vaksinasi covid-19 di tahun 2021.

3.1.3. Metode Analisis Data

Metode analisis data pada penelitian ini yaitu membandingkan hasil eksperimen mulai dari awal sampai akhir. Eksperimen dimulai dari melakukan proses pengambilan data melalui API Twitter yang kemudian dilakukan proses seleksi data atau preprocessing melalui tahap case folding dengan mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (huruf kecil atau lowercase), tokenizing merupakan tahap melakukan pemotongan string input berdasarkan tiap kata yang menyusunnya, filtering yang merupakan tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpan kata penting), dan stemming untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, dan juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

Penarikan kesimpulan dilakukan berdasarkan hasil evaluasi pada kedua model algoritma yang terpilih dengan menggunakan pengukuran nilai nilai performa akurasi, presisi, recall, F1 score, waktu training dan MAE. Hasil

pengukuran tersebut akan dijadikan acuan atau pedoman dalam menentukan hasil atau analisis sentimen pada penelitian ini

3.2. Alur Penelitian

Secara garis besar, alur penelitian yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Diagnosing

Pada tahapan ini peneliti melakukan studi pustaka dengan mencari dan membaca jurnal, buku, makalah, dan laporan penelitian lainnya yang terkait dengan obyek yang akan diteliti. Selanjutnya melakukan studi lapangan dengan melakukan crawling data melalui API Twitter untuk mendapatkan data primer.

2. Action Planning

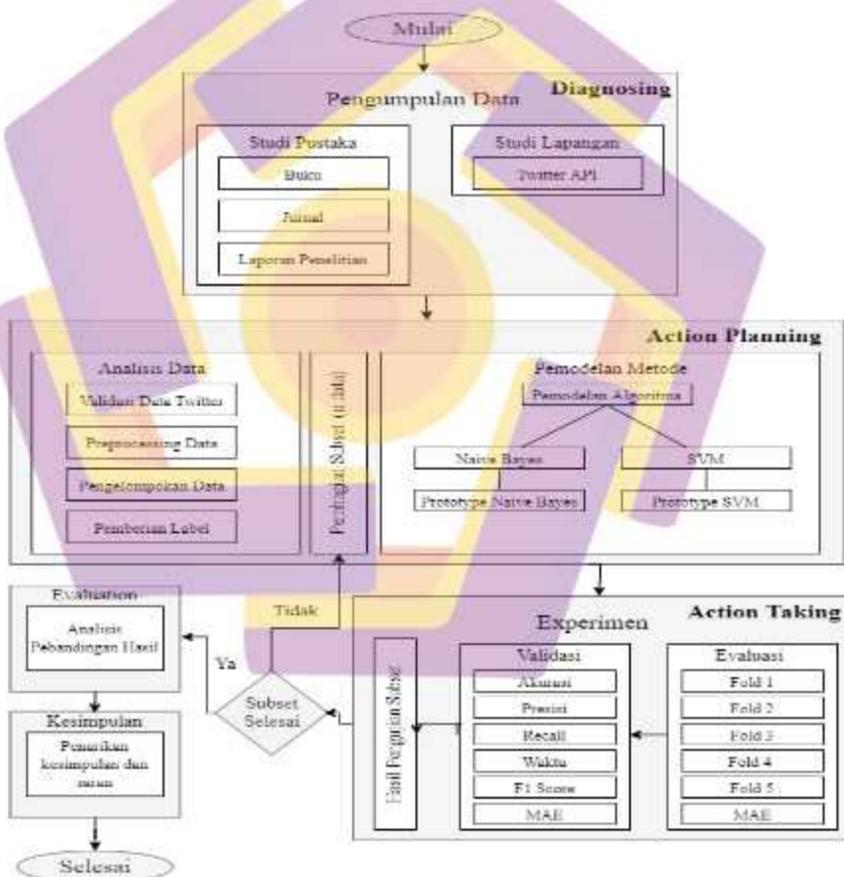
Pada tahapan ini peneliti melakukan analisis data, setelah data didapatkan lalu dilakukan verifikasi data dalam hal ini ada beberapa tahapan pre-processing untuk mendapatkan nilai atau bobot dari setiap data. Selanjutnya membuat pemodelan data dan algoritma kedua metode dengan data yang tersedia dan membuat prototype sistem analisis sentimen dengan satu sistem satu metode, sehingga terdapat dua prototype.

3. Action Taking

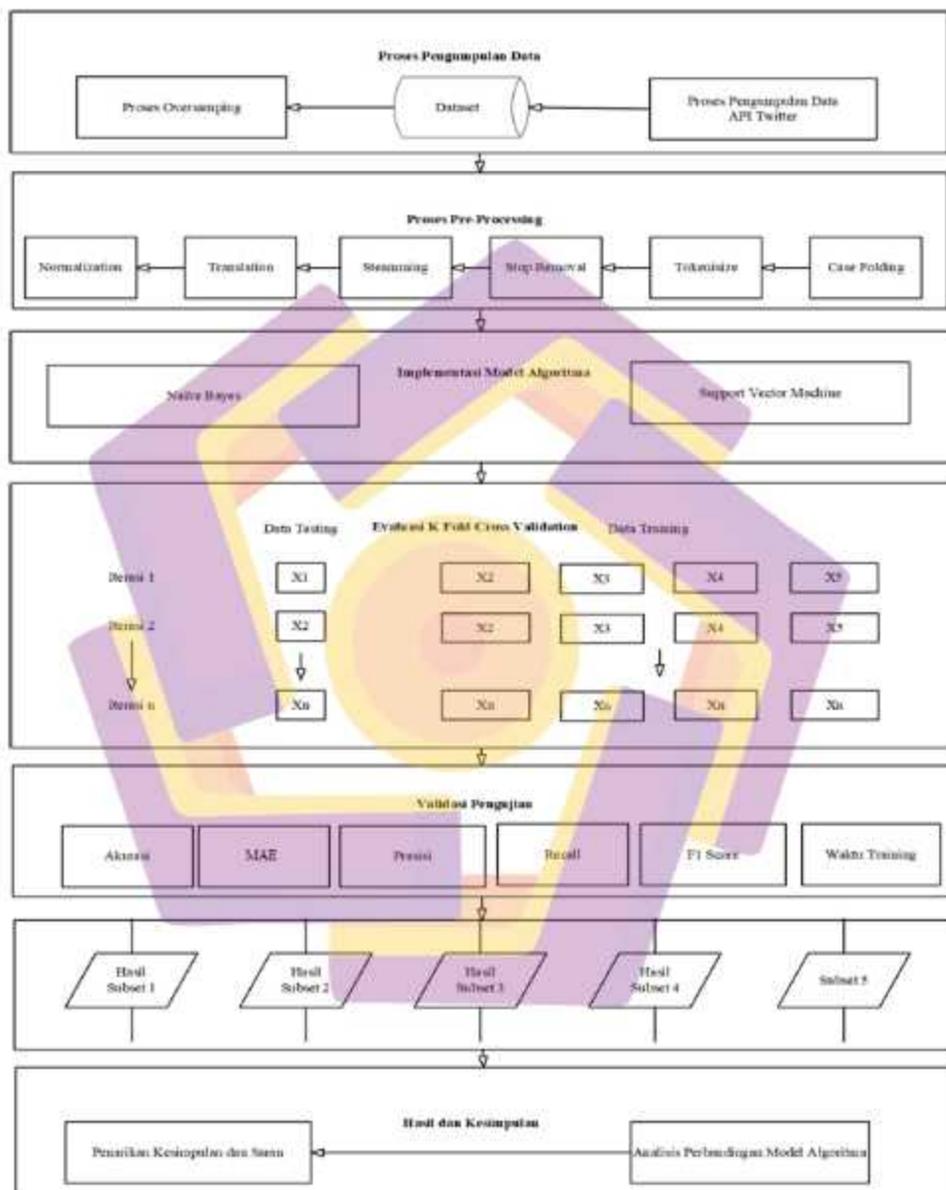
Pada tahapan ini peneliti melakukan eksperimen pengujian tingkat performa melalui pengujian serta membandingkan performa kedua metode melalui parameter pengujian akurasi, presisi, recall, f1 score, waktu training dan MAE untuk mendapatkan nilai performa terbaik antara kedua algoritma.

4. Evaluation

Pada tahapan yang terakhir, peneliti melakukan evaluasi hasil perbandingan dua metode yang digunakan yaitu naïve bayes dan SVM. Selanjutnya dihitung hasil yang sesuai untuk mendapatkan prosentase performa pada masing-masing metode. Tahap terakhir diperoleh kesimpulan, metode mana yang paling baik untuk dalam melakukan sentimen analisis data twitter melalui kata vaksin covid-19.



Gambar 3.1. Metode Penelitian



Gambar 3.2. Alur Implementasi Algoritma

Secara garis besar, alur implementasi algoritma yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Proses Pengumpulan Data

Proses pengumpulan data didapatkan melalui API Twitter untuk mengambil data postingan melalui kata vaksin covid-19, setelah data didapatkan maka data tersebut akan menjadi dataset awal untuk dilakukan proses over sampling, over sampling akan membentuk data kedalam pola yang telah ditentukan dan format yang sesuai untuk proses pre-processing

2. Pre-Processing

Pada tahapan ini peneliti melakukan proses pre-processing dimana data akan sortir kedalam proses case folding, tokenize, stop removal, stemming, translation dan normalization, setelah data tersebut sudah di proses maka data tersebut siap untuk di proses ke dalam algoritma.

3. Implementasi Algoritma

Pada tahapan ini peneliti melakukan implementasi algoritma, dimana algoritma akan digunakan untuk melakukan olah data dan melakukan sentimen analisis, algoritma yang digunakan dalam penelitian ini adalah algoritma naïve bayes dan SVM.

4. Evaluasi K-Fold Cross Validation

Pada tahapan ini peneliti validasi untuk pengujian algoritma, dimana dataset akan diatur melalui skenario 5 subset dan di poses sebanyak 5 kali perulangan untuk mendapatkan validasi pengujian.

5. Validasi Pengujian

Pada tahapan ini peneliti validasi dilakukan untuk mengukur performa algoritma melalui parameter akurasi, presisi, recall, f1 score, waktu training dan MAE yang sudah diproses melalui pembagian 5 subset dan 5 kali perulangan.

6. Hasil dan Kesimpulan

Pada tahapan ini peneliti akan diolah hasil untuk ditarik sebuah kesimpulan performa algoritma, dimana pada tahap ini algoritma naïve bayes dan SVM akan dilakukan analisis hasil dan perbandingan untuk mengukur setiap algoritma yang digunakan.



BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Diagnosing

4.1.1 Studi Literatur

Dalam menyelesaikan laporan penelitian ini harus sesuai dengan metode penelitian yang telah dibuat sebelumnya, studi literatur merupakan langkah awal yang dilakukan dengan mempelajari arsip/buku, jurnal, makalah dan laporan penelitian sebelumnya sebagai pedoman/landasan dalam menyusun penelitian ini. Studi literatur terdapat pada BAB II Landasan Teori pada penelitian ini, studi literatur memberikan penjelasan mengenai teori-teori yang dapat membantu dalam menyelesaikan penelitian ini.

4.1.2 Pengumpulan Data dan Informasi

Pada tahap pengumpulan data dan informasi ini dilakukan kegiatan pengumpulan data dan informasi yang dibutuhkan dalam penelitian ini, dalam memenuhi kegiatan tersebut perlu dilakukan beberapa tahapan yaitu:

a. Observasi

Setelah menentukan instrumen-instrumen yang dibutuhkan, selanjutnya peneliti melakukan observasi untuk mengetahui sentiment berkaitan dengan vaksin covid-19 pada media sosial Twitter.

b. Dokumentasi

Pada tahap ini peneliti melakukan dokumentasi melalui hasil pada dokumentasi yang telah dilakukan sebelumnya, hasil dokumentasi dari kegiatan

tersebut akan dijadikan laporan dan alur/atauran dari penelitian yang akan dilakukan.

4.1.3 Requirements Definition

Pada tahap *Requirements definition* atau definisi persyaratan ini dilakukan untuk mengetahui alur/proses analisis sentiment tentang vaksin covid-19 pada twitter yang berjalan guna menemukan kebutuhan dan kelemahan dalam sistem yang sedang berjalan saat ini, dalam kegiatan tersebut melalui beberapa tahapan yang akan dilakukan.

Pada tahap ini akan dilakukan identifikasi kebutuhan data yang akan digunakan dalam penelitian ini, dalam hal ini data yang akan digunakan adalah data sentiment atau postingan pada sosial media twitter berkaitan dengan hastag vaksin covid-19 berlangsung. Untuk detail kebutuhan data yang akan digunakan akan dijelaskan pada sub bab analisis data.

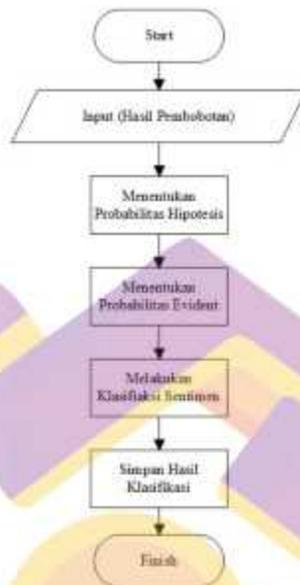
4.2. Action Taking

4.2.1. Pemodelan Algoritma

Pada tahap ini pemodelan metode untuk setiap model algoritma yaitu *Support Vector Machine* dan *Naive Bayes* akan dilakukan. Untuk detail tentang pemodelan metode akan dijelaskan pada sub bab berikut.

4.2.1.1. Naive Bayes

Pada tahap ini akan dilakukan pemodelan model algoritma dengan melakukan pembuatan model atau arsitektur Naive Bayes dengan konfigurasi seperti pada Gambar 4.1.:

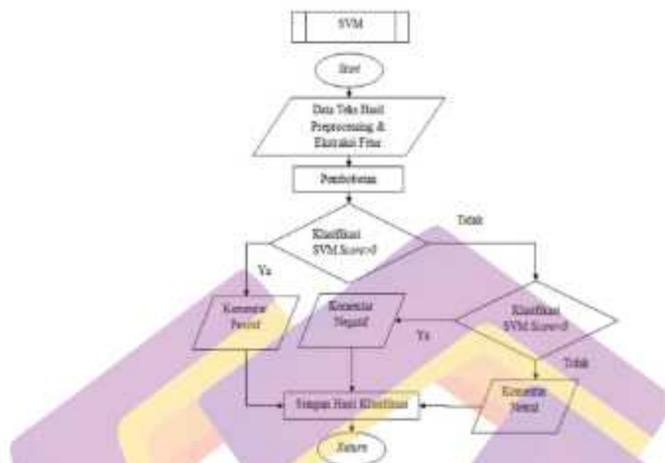


Gambar 4.1. Model Arsitektur Arsitektur Naive Bayes

Pada Gambar 4.1 akan dilakukan proses analisis menggunakan Naïve Bayes, setelah dokumen dalam bentuk matriks kata-dokumen dan telah diberi pembobotan TF-IDF, maka proses selanjutnya Naïve Bayes menganalisis, yaitu setiap dokumen akan diberi tanda positif, negatif dan netral. Kemudian akan dilakukan proses Naïve Bayes untuk menghitung nilai bobot antara data uji dengan semua data latih pada dokumen artikel dengan menggunakan metode probabilitas.

4.2.1.2. Support Vector Machine

Pada tahap ini akan dilakukan pemodelan model algoritma dengan melakukan pembuatan model atau arsitektur Support Vector Machine dengan konfigurasi seperti pada Gambar 4.2:



Gambar 4.2. Model Arsitektur SVM

Pada Gambar 4.2 akan dilakukan proses analisis menggunakan SVM dimulai mengubah text menjadi data vektor. Vektor dalam penelitian ini memiliki dua komponen yaitu dimensi (word id) dan bobot. Bobot ini adalah nilai tf-idf, tujuan dari model ruang vektor digunakan untuk memberikan setiap kata dalam dokumen sebuah ID (dimensi) dan sebuah bobot berdasarkan seberapa penting keberadaannya dalam dokumen (posisi dokumen dalam dimensi itu). SVM mencoba untuk menemukan garis yang terbaik membagi tiga kelas, dan kemudian mengklasifikasikan dokumen uji berdasarkan di sisi mana dari garis tersebut mereka muncul.

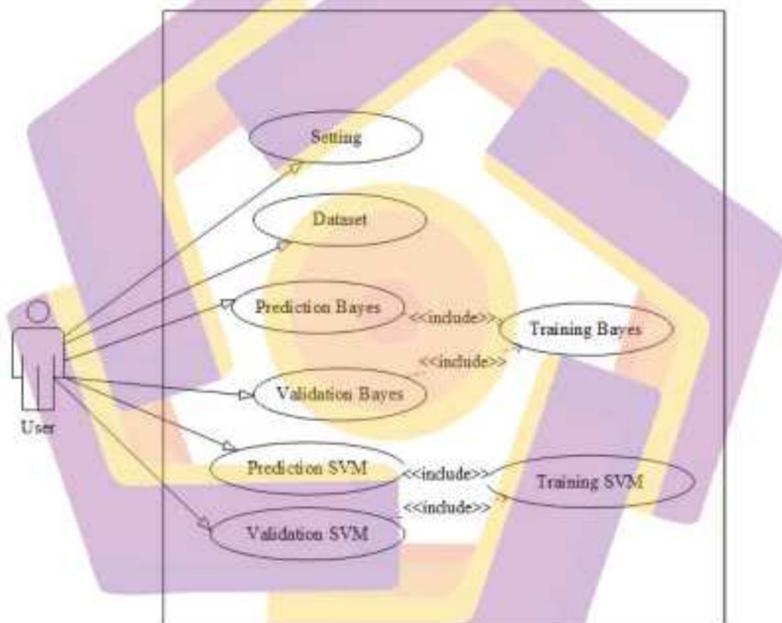
4.2.2. Pemodelan Proses

Pada tahap ini akan dilakukan pemodelan proses dengan menggunakan UML atau *Unified Modelling Language* untuk membantu dalam pendeskripsian dan desain sistem perangkat lunak yang akan digunakan dalam penelitian ini.

Pemodelan UML yang akan digunakan dalam penelitian ini adalah usecase diagram, activity diagram dan class diagram seperti berikut:

a. *Usecase Diagram*

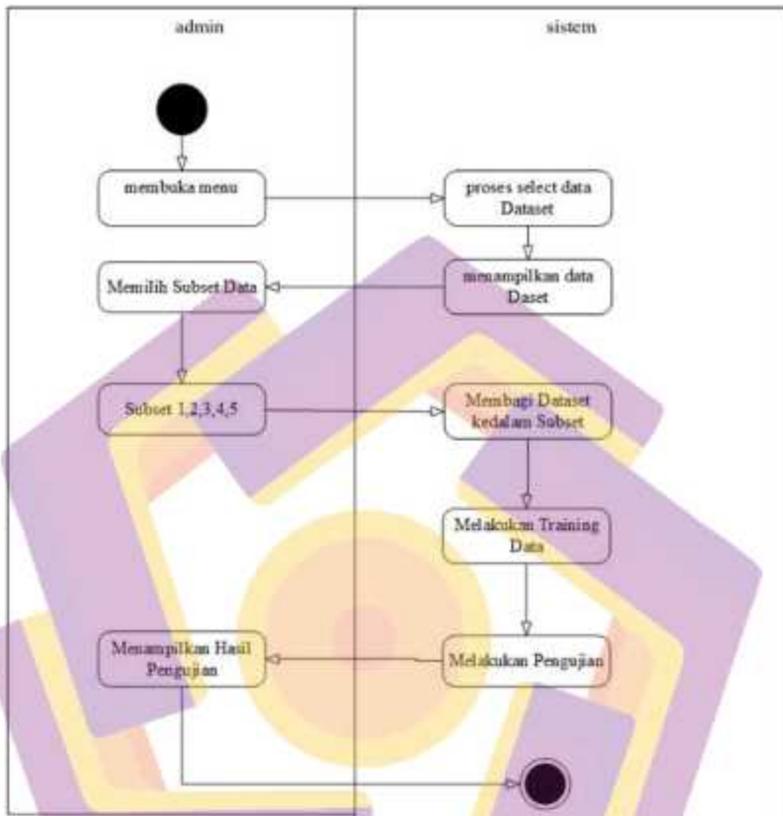
Usecase diagram merupakan gambaran dari kebutuhan fungsional dari sistem yang akan dibuat dan diteliti, usecase diagram dalam penelitian ini dapat dilihat pada Gambar 4.3.



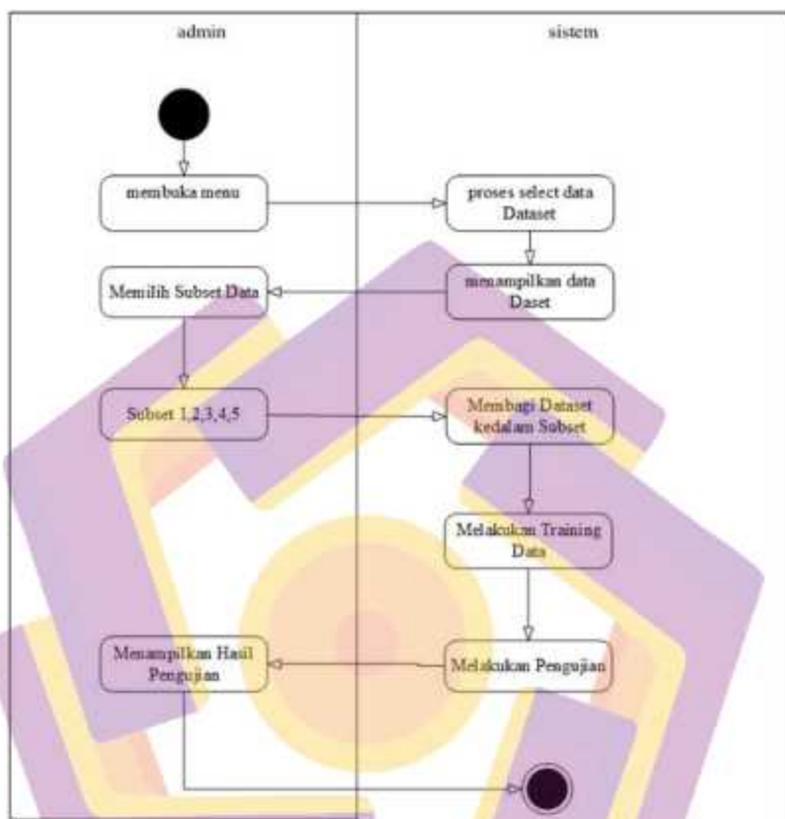
Gambar 4.3. Usecase Diagram

b. *Activity Diagram*

Activity diagram merupakan gambaran proses atau alur dari setiap menu atau fungsi yang akan dibuat dan diteliti, activity diagram dalam penelitian ini dapat dilihat pada Gambar 4.4 dan Gambar 4.5.



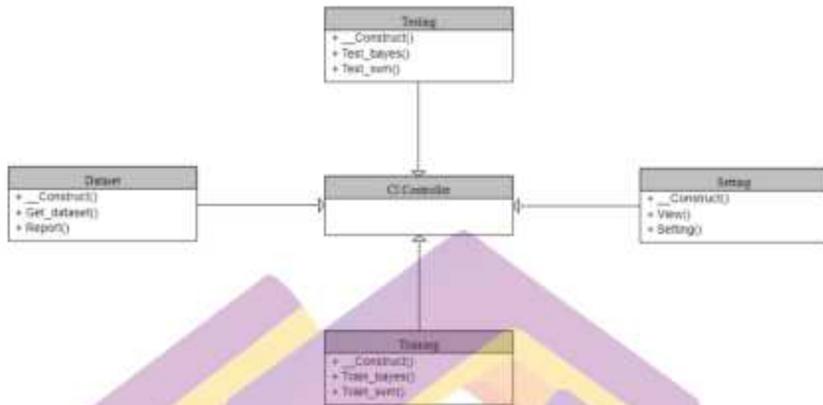
Gambar 4.4. Activity Diagram SVM



Gambar 4.5. Activity Diagram Naive Bayes

c. *Class Diagram*

Class diagram merupakan gambaran kelas dan fungsi yang akan dibuat pada implementasi pemrograman yang akan dilakukan (*Coding*), Class diagram dalam penelitian ini dapat dilihat pada Gambar 4.6.



Gambar 4.6. Class Diagram Controller

4.2.3. Pemodelan Data (Analisis Data)

Pada tahap analisis data akan dilakukan 3 tahap analisis yaitu pengambilan data dari twitter dengan parameter vaksin covid-19, pre-processing data serta pelabelan yang akan dilakukan oleh pakar untuk menentukan label sentiment analisis.

Tahap pertama akan dilakukan pengambilan data dari twitter yaitu melalui metode scraping dengan menggunakan selenium untuk library pengambilan data seperti berikut:

a. Inisialisasi Selenium

Pada tahap ini akan dilakukan inisialisasi paket library yang akan digunakan untuk melakukan scraping data yaitu melalui perintah berikut:

```

!pip install selenium
!apt-get update # to update ubuntu to correctly run apt install
!apt install chromium-chromedriver
# !cp /usr/lib/chromium-browser/chromedriver /usr/bin
from selenium import webdriver
  
```

```

from selenium.webdriver.common.keys import Keys
from selenium.webdriver.common.by import By
from time import sleep
import csv
# driver=webdriver.Chrome(path)
# sys.path.insert(0,usr/lib/chromium-browser/chromedriver')
chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument('--headless')
chrome_options.add_argument('--no-sandbox')
chrome_options.add_argument('--disable-dev-shm-usage')
driver
webdriver.Chrome('chromedriver',chrome_options=chrome_options)
# driver.set_page_load_timeout(1000)
loadPageTimot = 10000
driver.set_page_load_timeout(loadPageTimot)
# return driver

```

b. GetTweets

Pada tahap ini akan dilakukan inialisasi pengambilan postingan atau tweet yang akan digunakan untuk melakukan scraping data yaitu melalui perintah berikut:

```

def getTweets(driver, totalTweet, secondToSleep):
    import time
    resultTweet = []
    currentDocHeight = 0

    while(len(resultTweet) < totalTweet):
        currentTweetOnPage = getTweet(driver)
        for t in currentTweetOnPage:
            if t not in resultTweet:
                resultTweet.append(t)

        docHeight = scrollDown(driver, secondToSleep)
        print(currentDocHeight)
        print(docHeight)
        if currentDocHeight == docHeight:
            break

        currentDocHeight = docHeight

    return resultTweet

```

c. GetKeyword

Pada tahap ini akan dilakukan scraping atau pengambilan postingan atau tweet sesuai dengan keyword yang telah ditentukan yaitu melalui perintah berikut:

```
def getTweetByKeyword(driver, keyword, total):
url='https://twitter.com/search?q=vaksin%20covid&src=typed_query'.format(key
word)
    driver.get(url)
    sleep(10)
    return getTweets(driver, total, 10)
```

d. GetFile

Pada tahap ini akan export data hasil scraping kedalam format excel yaitu melalui perintah berikut:

```
import pandas as pd
def saveToExcel(tweets, name):
    pd.DataFrame.from_dict(tweet).to_excel("{} .xlsx".format(name))

for scrapping in "key".split(",") :
    tweet = getTweetByKeyword(driver, scrapping.lstrip().rstrip(), 100)
    saveToExcel(tweet, scrapping)
    sleep(10)
```

Pada tahap selanjutnya akan dilakukan pemodelan data dilakukan melalui beberapa tahapan analisis data yaitu case folding, Tokenizing, Filtering (Stopword Removal), Stemming dan Normalization dengan detail seperti berikut:

d. Case folding

Pada tahap ini akan dilakukan salah satu bentuk text preprocessing yang paling sederhana dan efektif meskipun sering diabaikan. Tujuan dari case folding untuk mengubah semua huruf dalam dokumen menjadi huruf kecil.

Hanya huruf 'a' sampai 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter, dengan detail seperti berikut.

Mengubah text menjadi lowercase, pada tahap ini semua text yang terinput akan diubah kedalam bentuk lowercase dengan contoh sebagai berikut

```
# Input
kalimat = "Vaksin covid-19 ke 2 ini sangat menyeramkan!!!"

lower_case = kalimat.lower()
print(lower_case)

# output
kalimat = "vaksin covid-19 ke 2 ini sangat menyeramkan!!!"
```

Menghapus angka, pada tahap ini semua angka yang terinput akan diubah dihapus dengan contoh sebagai berikut

```
# Input
kalimat = "vaksin covid-19 ke 2 ini sangat menyeramkan!!!"

hasil = re.sub(r"\d+", "", kalimat)
print(hasil)

# output
kalimat = "vaksin covid- ke ini sangat menyeramkan!!!"
```

Menghapus tanda baca, pada tahap ini semua tanda baca yang terinput akan diubah dihapus dengan contoh sebagai berikut

```
# Input
kalimat = "vaksin covid-19 ke ini sangat menyeramkan!!!"

hasil = kalimat.translate(str.maketrans("", "", string.punctuation))
print(hasil)

# output
kalimat = "vaksin covid ke ini sangat menyeramkan"
```

e. *Tokenizing*

Pada tahap ini akan dilakukan proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam NLP, token diartikan sebagai "kata" meskipun tokenize juga dapat dilakukan pada paragraf maupun kalimat, dengan detail seperti berikut:

```
# Input
kalimat = " vaksin covid-19 ke ini sangat menyeramkan "

pisah = kalimat.split()
print(pisah)

# output
kalimat = ["vaksin", ["covid"], ["ke"], ["ini"], ["sangat"],
["menyeramkan"]];
```

f. *Filtering (Stopword Removal)*

Filtering adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpan kata penting).

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam bahasa Indonesia adalah "yang", "dan", "di", "ini", dll. Makna di balik penggunaan stopwords yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya.

Detail proses Filtering (Stopword Removal) adalah seperti berikut:

```
# Input
kalimat = " vaksin covid-19 ke ini sangat menyeramkan "

tokens = word_tokenize(kalimat)
```

```

listStopword = set(stopwords.words('indonesian'))

removed = []
for t in tokens:
    if t not in listStopword:
        removed.append(t)

# output
kalimat = "["'vaksin', ['covid'], ['sangat'], ['menyeramkan']";

```

g. *Stemming*

Stemming merupakan proses menghilangkan infleksi kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan akar kata (root word). Misalnya kata “mendengarkan”, “dengarkan”, “didengarkan” akan ditransformasi menjadi kata “dengar”. Dengan contoh seperti berikut:

```

# Input
kalimat = " vaksin covid-19 ke ini sangat menyeramkan "

factory = StemmerFactory()
stemmer = factory.create_stemmer()

hasil = stemmer.stem(kalimat)

# output
kalimat = "["'vaksin', ['covid'], ['sangat'], ['seram]";

```

h. *Normalization*

Normalization merupakan tahap hasil dari proses text preprocessing akan dinormalkan kedalam bentuk array pada tabel di sebuah database. Data array tersebut yang nantinya akan digunakan sebagai data training dan testing untuk sistem dalam melakukan klasifikasi.

Setelah didapatkan data yang sudah ternormalisasi maka selanjutnya adalah tahap pelabelan data yang akan dilakukan oleh pakar Bahasa yaitu Ummi Nur Jannah S.Pd., Hanifa Hafiza,MPd dan Sherly miranti, S.sos. pada tahap ini

akan dilakukan labeling data untuk menentukan kalimat apakah termasuk kalimat positif, negatif ataupun netral, seperti contoh kalimat berikut:

```
# Input
kalimat = "vaksin covid-19 ke ini sangat menyheramkan "

# klaksifikasi
kalimat = "negatif";
```

Berdasarkan analisis sentiment yang telah dilakukan oleh Ummi Nur Jannah S.Pd, Ummi Nur Jannah S.Pd., Hanifa Hafiza, M.Pd selaku pakar Bahasa dan Sherly miranti, S.sos. selaku pakar psikologis maka diperoleh labeling kalimat untuk dataset sentiment analisis vaksin covid-19 yaitu negatif 311 data, positif 443 data, netral 246 data.

i. Algoritma TF-IDF (Term Frequency – Inverse Document Frequency)

Algoritma TF-IDF (Term Frequency – Inverse Document Frequency) adalah salah satu algoritma yang dapat digunakan untuk menganalisa hubungan antara sebuah frase/kalimat dengan sekumpulan dokumen.

Inti utama dari algoritma ini adalah melakukan perhitungan nilai TF dan nilai IDF dari sebuah setiap kata kunci terhadap masing-masing dokumen. Nilai TF dihitung dengan rumus $TF = \text{jumlah frekuensi kata terpilih} / \text{jumlah kata}$ dan nilai IDF dihitung dengan rumus $IDF = \log(\text{jumlah dokumen} / \text{jumlah frekuensi kata terpilih})$. Selanjutnya adalah melakukan perkalian antara nilai TF dan IDF untuk mendapatkan jawaban akhir.

Contoh kalimat “Vaksin kita suka”

Hitung nilai TF dengan rumus

$TF = \text{jumlah frekuensi kata terpilih} / \text{jumlah kata}$

```

For j As Integer = 0 To daftarKataQuery.Count - 1
  Dim frekuensi As Integer = 0
  For k As Integer = 0 To jumlahKataPerDokumen(i).Count - 1
    If daftarKataQuery(j) = jumlahKataPerDokumen(i)(k).kata
  Then
    frekuensi = jumlahKataPerDokumen(i)(k).jumlah
  Exit For
  End If
  Next
  TF(i).Add(frekuensi / daftarKata.Length)
Next

```

Hitung IDF dengan rumus

$IDF = \log(\text{jumlah dokumen} / \text{jumlah frekuensi kata terpilih})$

```

Dim frekuensiDokumen(daftarKataQuery.Count - 1) As Integer
For i As Integer = 0 To daftarDokumen.Count - 1
  For j As Integer = 0 To daftarKataQuery.Count - 1
    If TF(i)(j) > 0 Then frekuensiDokumen(j) += 1
  Next
Next

For j As Integer = 0 To daftarKataQuery.Count - 1
  IDF(j) = Math.Log10(daftarDokumen.Count /
frekuensiDokumen(j))
Next

```

Nilai jawaban dilakukan dengan cara melakukan perkalian antara nilai TF dan nilai IDF

Tampilkan hasil jawaban akhir pada layer

```

For i As Integer = 0 To daftarDokumen.Count - 1
    For j As Integer = 0 To daftarKataQuery.Count - 1
        Console.WriteLine("Nilai TF-IDF untuk kata " &
        daftarKataQuery(j).PadRight(8) & " terhadap dokumen " & (i + 1) & " adalah " &
        (TF(i)(j) * IDF(j)).ToString("F4"))
    Next
Next

```

Hasil akhir adalah:

```

Nilai TF-IDF untuk kata vaksin terhadap dokumen 1 adalah 0.000
Nilai TF-IDF untuk kata kita terhadap dokumen 1 adalah 0.002
Nilai TF-IDF untuk kata suka terhadap dokumen 1 adalah 0.043
.....
Nilai TF-IDF untuk kata vaksin terhadap dokumen 1000 adalah 0.075
Nilai TF-IDF untuk kata kita terhadap dokumen 1000 adalah 0.081
Nilai TF-IDF untuk kata suka terhadap dokumen 1000 adalah 0.030

```

4.2.4. Pemodelan Sistem

Pada tahap ini desain antar muka sistem akan dirancang sebagai acuan pembuatan sistem yang akan dibuat nanti, adapun beberapa rancangan utama sistem ini adalah sebagai berikut:

a. Antarmuka Pengelolaan Dataset

Pada menu ini pengguna dapat memasukan dataset dengan cara menginput atau mengimport melalui API Twitter. Desain antarmuka pengelolaan dataset dapat dilihat pada Gambar 4.7.

Kata 1	Kata 2	Kata 3	Kata 4
Kata 1	Kata 2	Kata 3	Kata 4
Kata 1	Kata 2	Kata 3	Kata 4
Kata 1	Kata 2	Kata 3	Kata 4
Kata 1	Kata 2	Kata 3	Kata 4

Gambar 4.7. Antarmuka Dataset

b. Antarmuka Pengelolaan Data Training

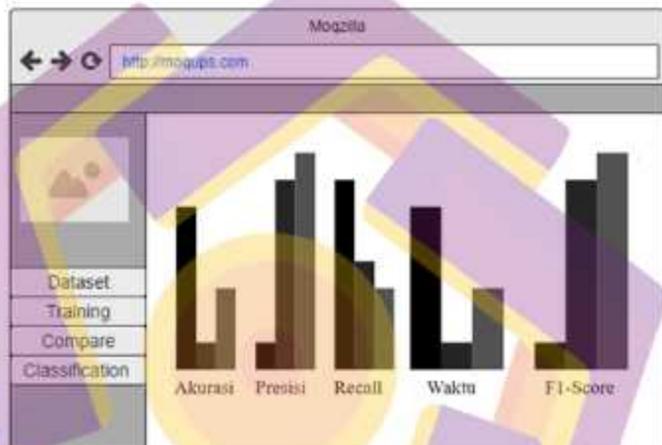
Pada menu ini pengguna dapat melakukan training data, terdapat 2 pilihan metode algoritma dalam training data yaitu SVM dan Naive Bayes. Desain antarmuka pengelolaan data training dapat dilihat pada Gambar 4.8.

Algoritma	Akurasi	Presisi	Recall	Waktu	F1 Score
Bayes	Akurasi	Presisi	Recall	Waktu	F1 Score
SVM	Akurasi	Presisi	Recall	Waktu	F1 Score

Gambar 4.8. Antarmuka Training Data

c. Antarmuka Hasil Perbandingan

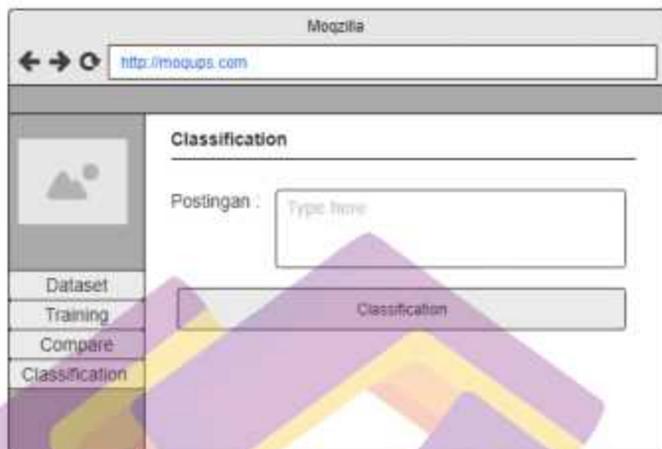
Pada menu ini pengguna dapat melihat hasil perbandingan tingkat akurasi, presisi, recall, waktu training, F-1 Score dan MAE dari model algoritma SVM dan Naive Bayes. Desain antarmuka hasil perbandingan dapat dilihat pada Gambar 4.9.



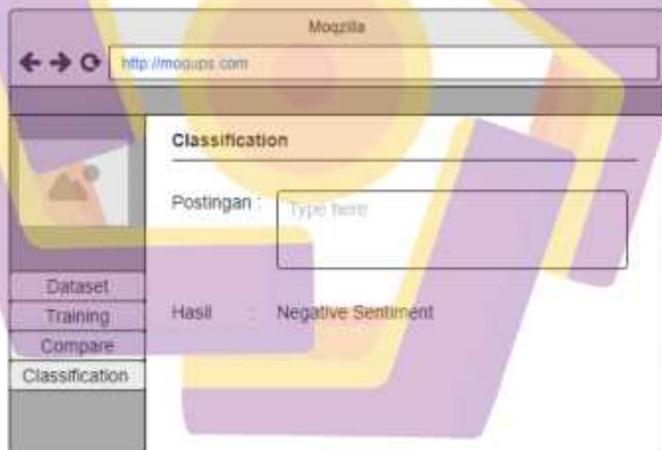
Gambar 4.9. Antarmuka Hasil Perbandingan

d. Antarmuka Prediksi

Pada menu ini pengguna dapat melakukan prediksi melalui input beberapa parameter seperti pada Gambar 4.10 dengan model algoritma SVM atau Naive Bayes. Untuk hasil dari prediksi tersebut dapat dilihat pada Gambar 4.11.



Gambar 4.10. Antarmuka Input Parameter Prediksi



Gambar 4.11. Antarmuka Hasil Prediksi

4.3 Evaluation (System Testing)

4.3.1. Pengujian Naive Bayes

4.3.1.1. Pengujian Akurasi Naive Bayes

Pengujian pertama dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.1.

Tabel 4.1 Hasil Akurasi *Naive Bayes* Subset 1 (200 Data)

Subset	Akurasi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	72	74	77	78	75	75,20

Hasil pengujian Tabel 4.1 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 72%, Fold ke-2 mendapatkan hasil 74%, Fold ke-3 mendapatkan hasil 77%, Fold ke-4 mendapatkan hasil 78% dan Fold ke-5 mendapatkan hasil 75%, dengan demikian nilai rata-rata rata-rata akurasi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 75,20%.

Pengujian kedua dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.2.

Tabel 4.2 Hasil Akurasi *Naive Bayes* Subset 2 (400 Data)

Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
2	75	76	78	78	76	76,60

Hasil pengujian Tabel 4.2 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 75%, Fold ke-2 mendapatkan hasil 76%, Fold ke-3 mendapatkan hasil 78%, Fold ke-4 mendapatkan hasil 78% dan Fold ke-5 mendapatkan hasil 76%, dengan demikian nilai rata-rata rata-rata akurasi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 76,60%.

Pengujian ketiga dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.3.

Tabel 4.3 Hasil Akurasi *Naive Bayes* Subset 3 (600 Data)

Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
3	75	79	81	80	79	78,80

Hasil pengujian Tabel 4.3 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 75%,

Fold ke-2 mendapatkan hasil 79%, Fold ke-3 mendapatkan hasil 81%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 79%, dengan demikian nilai rata-rata akurasi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 78,80%.

Pengujian keempat dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.4.

Tabel 4.4 Hasil Akurasi *Naive Bayes* Subset 4 (800 Data)

Subset	Akurasi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	76	78	80	82	82	79,60

Hasil pengujian Tabel 4.4 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 76%, Fold ke-2 mendapatkan hasil 78%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 82% dan Fold ke-5 mendapatkan hasil 82%, dengan demikian nilai rata-rata akurasi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 79,60%.

Pengujian kelima dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.5.

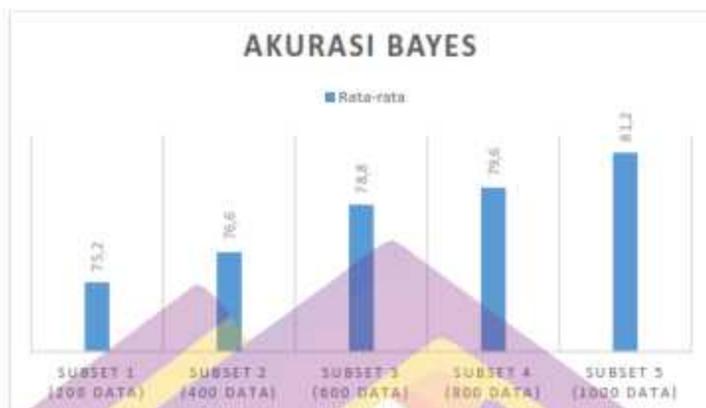
Tabel 4.5 Hasil Akurasi *Naive Bayes* Subset 5 (1000 Data)

Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
5	78	88	84	80	80	81,20

Hasil pengujian Tabel 4.5 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 78%, Fold ke-2 mendapatkan hasil 88%, Fold ke-3 mendapatkan hasil 84%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 80%, dengan demikian nilai rata-rata rata-rata akurasi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 81,20%.

Berdasarkan hasil lima percobaan di atas maka diperoleh hasil akurasi pada penerapan algoritma bayes adalah subset ke-1 adalah 75,20%, subset ke-2 adalah 76,60%, subset ke-3 adalah 78,80%, subset ke-4 adalah 79,60% dan subset ke-5 adalah 81,20%, maka berdasarkan hasil percobaan akurasi paling tinggi didapat pada percobaan subset ke-5 dengan nilai akurasi sebesar 81,20%.

Berdasarkan gambar 4.12 menunjukkan bahwa performa akurasi akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik akurasi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.12.



Gambar 4.12 Grafik Akurasi *Naive Bayes*

4.3.1.2. Pengujian Presisi *Naive Bayes*

Pengujian kedua dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.6.

Tabel 4.6 Hasil Presisi *Naive Bayes* Subset 1 (200 Data)

Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	71	76	75	74	73	73,80

Hasil pengujian Tabel 4.6 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 71%, Fold ke-2 mendapatkan hasil 76%, Fold ke-3 mendapatkan hasil 75%, Fold ke-4 mendapatkan hasil 74% dan Fold ke-5

mendapatkan hasil 73%, dengan demikiran nilai rata-rata rata-rata presisi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 73,80%.

Pengujian kedua dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.7.

Tabel 4.7 Hasil Presisi Naive Bayes Subset 2 (400 Data)

Subset	Presisi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
2	72	76	77	75	76	75,20

Hasil pengujian Tabel 4.7 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 72%, Fold ke-2 mendapatkan hasil 76%, Fold ke-3 mendapatkan hasil 77%, Fold ke-4 mendapatkan hasil 75% dan Fold ke-5 mendapatkan hasil 76%, dengan demikiran nilai rata-rata rata-rata presisi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 75,20%.

Pengujian ketiga dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.8.

Tabel 4.8 Hasil Presisi Naive Bayes Subset 3 (600 Data)

Subset	Presisi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
3	74	77	79	76	75	76,20

Hasil pengujian Tabel 4.8 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 74%, Fold ke-2 mendapatkan hasil 77%, Fold ke-3 mendapatkan hasil 79%, Fold ke-4 mendapatkan hasil 76% dan Fold ke-5 mendapatkan hasil 75%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 76,20%.

Pengujian keempat dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.9.

Tabel 4.9 Hasil Presisi *Naive Bayes* Subset 4 (800 Data)

Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	76	82	80	76	78	78,40

Hasil pengujian Tabel 4.9 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 76%, Fold ke-2 mendapatkan hasil 82%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 76% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 78,40%.

Pengujian kelima dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.10.

Tabel 4.10 Hasil Presisi *Naive Bayes* Subset 5 (1000 Data)

Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	76	86	80	77	78	79,40

Hasil pengujian Tabel 4.10 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 76%, Fold ke-2 mendapatkan hasil 86%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 77% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata presisi pada penerapan algoritma *Naive Bayes* sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 79,40%.

Berdasarkan hasil lima percobaan di atas maka diperoleh hasil presisi pada penerapan algoritma bayes adalah subset ke-1 adalah 73,80%, subset ke-2 adalah 75,20%, subset ke-3 adalah 76,20%, subset ke-4 adalah 78,40% dan subset ke-5 adalah 79,40%, maka berdasarkan hasil percobaan presisi paling tinggi didapat pada percobaan subset ke-5 dengan nilai presisi sebesar 79,40%.

Berdasarkan gambar 4.13 menunjukkan bahwa performa presisi akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik presisi pada penerapan algoritma *Naive Bayes* sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.13.



Gambar 4.13 Grafik Presisi *Naive Bayes*

4.3.1.3. Pengujian Recall *Naive Bayes*

Pengujian ketiga dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.11.

Tabel 4.11 Hasil Recall *Naive Bayes* Subset 1 (200 Data)

Substet	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	71	73	77	81	71	74,60

Hasil pengujian Tabel 4.11 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 71%, Fold ke-2 mendapatkan hasil 73%, Fold ke-3 mendapatkan hasil 77%, Fold ke-4 mendapatkan hasil 81% dan Fold ke-5 mendapatkan hasil 71%, dengan demikian nilai rata-rata rata-rata recall pada

penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 74,60%.

Pengujian kedua dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.12.

Tabel 4.12 Hasil Recall *Naive Bayes* Subset 2 (400 Data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	73	73	78	82	72	75,60

Hasil pengujian Tabel 4.7 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 73%, Fold ke-2 mendapatkan hasil 73%, Fold ke-3 mendapatkan hasil 78%, Fold ke-4 mendapatkan hasil 82% dan Fold ke-5 mendapatkan hasil 72%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 75,60%.

Pengujian ketiga dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.13.

Tabel 4.13 Hasil Recall *Naive Bayes* Subset 3 (600 Data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	74	75	81	82	75	77,40

Hasil pengujian Tabel 4.13 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 74%, Fold ke-2 mendapatkan hasil 75%, Fold ke-3 mendapatkan hasil 81%, Fold ke-4 mendapatkan hasil 82% dan Fold ke-5 mendapatkan hasil 75%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 77,40%.

Pengujian keempat dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.14.

Tabel 4.14 Hasil Recall *Naive Bayes* Subset 4 (800 Data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	77	75	80	84	76	78,40

Hasil pengujian Tabel 4.14 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 77%, Fold ke-2 mendapatkan hasil 75%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 84% dan Fold ke-5 mendapatkan hasil 76%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 78,40%.

Pengujian kelima dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.15.

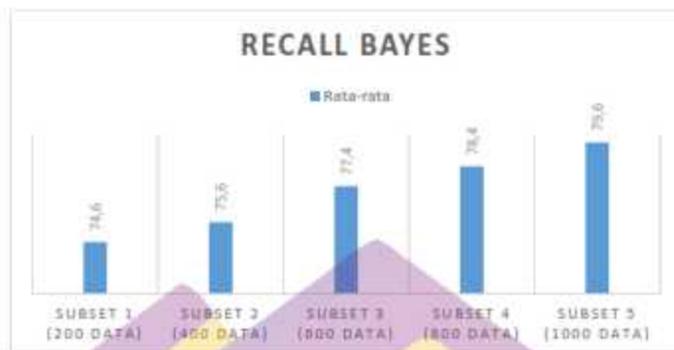
Tabel 4.15 Hasil Recall *Naive Bayes* Subset 5 (1000 Data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	77	76	81	86	77	79,60

Hasil pengujian Tabel 4.15 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 77%, Fold ke-2 mendapatkan hasil 76%, Fold ke-3 mendapatkan hasil 81%, Fold ke-4 mendapatkan hasil 86% dan Fold ke-5 mendapatkan hasil 77%, dengan demikian nilai rata-rata recall pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 79,60%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil recall pada penerapan algoritma bayes adalah subset ke-1 adalah 74,60%, subset ke-2 adalah 75,60%, subset ke-3 adalah 77,40%, subset ke-4 adalah 78,40% dan subset ke-5 adalah 79,60%, maka berdasarkan hasil percobaan recall paling tinggi didapat pada percobaan subset ke-5 dengan nilai recall sebesar 79,60%.

Berdasarkan gambar 4.14 menunjukkan bahwa performa recall akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik recall pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.14.



Gambar 4.14 Grafik Recall *Naive Bayes*

4.3.1.4. Pengujian Waktu training *Naive Bayes*

Pengujian keempat dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.16.

Tabel 4.16 Hasil Waktu training *Naive Bayes* Subset 1 (200 Data)

Substet	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	10	7	10	8	10	9,00

Hasil pengujian Tabel 4.16 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 10 detik, Fold ke-2 mendapatkan hasil 7 detik, Fold ke-3 mendapatkan hasil 10 detik, Fold ke-4 mendapatkan hasil 8 detik dan Fold ke-5 mendapatkan hasil 10 detik, dengan demikian nilai rata-rata rata-rata waktu

training pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 9 detik.

Pengujian kedua dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.17.

Tabel 4.17 Hasil Waktu training *Naive Bayes* Subset 2 (400 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	10	8	10	9	12	9,80

Hasil pengujian Tabel 4.17 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 10 detik, Fold ke-2 mendapatkan hasil 8 detik, Fold ke-3 mendapatkan hasil 10 detik, Fold ke-4 mendapatkan hasil 9 detik dan Fold ke-5 mendapatkan hasil 12 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 9,80 detik.

Pengujian ketiga dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.18.

Tabel 4.18 Hasil Waktu training *Naive Bayes* Subset 3 (600 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	11	9	12	9	12	10,60

Hasil pengujian Tabel 4.18 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 11 detik, Fold ke-2 mendapatkan hasil 9 detik, Fold ke-3 mendapatkan hasil 12 detik, Fold ke-4 mendapatkan hasil 9 detik dan Fold ke-5 mendapatkan hasil 12 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 10,60 detik.

Pengujian keempat dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.19.

Tabel 4.19 Hasil Waktu training *Naive Bayes* Subset 4 (800 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	13	10	12	11	13	11,80

Hasil pengujian Tabel 4.19 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 13 detik, Fold ke-2 mendapatkan hasil 10 detik, Fold ke-3 mendapatkan hasil 12 detik, Fold ke-4 mendapatkan hasil 11 detik dan Fold ke-5 mendapatkan hasil 13 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan

algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 11,80 detik.

Pengujian kelima dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.20.

Tabel 4.20 Hasil Waktu training *Naive Bayes* Subset 5 (1000 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	14	10	13	12	14	12,60

Hasil pengujian Tabel 4.20 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 14 detik, Fold ke-2 mendapatkan hasil 10 detik, Fold ke-3 mendapatkan hasil 13 detik, Fold ke-4 mendapatkan hasil 12 detik dan Fold ke-5 mendapatkan hasil 14 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 12,60%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil waktu training pada penerapan algoritma bayes adalah subset ke-1 adalah 9 detik, subset ke-2 adalah 9,80 detik, subset ke-3 adalah 10,60 detik, subset ke-4 adalah 11,80 detik dan subset ke-5 adalah 12,60 detik, maka berdasarkan hasil percobaan waktu training paling cepat didapat pada percobaan subset ke-1 dengan nilai waktu training sebesar 9 detik.

Berdasarkan gambar 4.15 menunjukkan bahwa performa waktu training akan mengalami penurunan pada setiap subset atau penambahan data. Oleh karena itu grafik waktu training pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.15.



Gambar 4.15 Grafik Waktu training *Naive Bayes*

4.3.1.5. Pengujian F1 score *Naive Bayes*

Pengujian pertama dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.21.

Tabel 4.21 Hasil F1 score *Naive Bayes* Subset 1 (200 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	72	74	76	80	72	74,80

Hasil pengujian Tabel 4.21 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1

mendapatkan hasil 72%, Fold ke-2 mendapatkan hasil 74%, Fold ke-3 mendapatkan hasil 76%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 72%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 74,80%.

Pengujian kedua dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.22.

Tabel 4.22 Hasil F1 score *Naive Bayes* Subset 2 (400 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	72	74	80	83	73	76,40

Hasil pengujian Tabel 4.22 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 72%, Fold ke-2 mendapatkan hasil 74%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 83% dan Fold ke-5 mendapatkan hasil 73%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 76,40%.

Pengujian ketiga dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.23.

Tabel 4.23 Hasil F1 score *Naive Bayes* Subset 3 (600 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	75	77	82	84	78	79,20

Hasil pengujian Tabel 4.23 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 75%, Fold ke-2 mendapatkan hasil 77%, Fold ke-3 mendapatkan hasil 82%, Fold ke-4 mendapatkan hasil 84% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata rata-rata fl score pada penerapan algoritma *Naive Bayes* sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 79,20%.

Pengujian keempat dilakukan untuk nilai tingkat performa fl score dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.24.

Tabel 4.24 Hasil F1 score *Naive Bayes* Subset 4 (800 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	78	76	82	86	78	80,00

Hasil pengujian Tabel 4.24 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 78%,

Fold ke-2 mendapatkan hasil 76%, Fold ke-3 mendapatkan hasil 82%, Fold ke-4 mendapatkan hasil 86% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 80,00%.

Pengujian kelima dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.25.

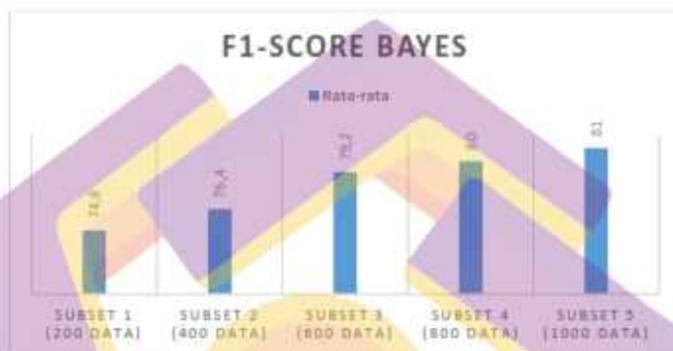
Tabel 4.25 Hasil F1 score *Naive Bayes* Subset 5 (1000 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	79	77	83	88	78	81,00

Hasil pengujian Tabel 4.25 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 79%, Fold ke-2 mendapatkan hasil 77%, Fold ke-3 mendapatkan hasil 83%, Fold ke-4 mendapatkan hasil 88% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 81,60%.

Berdasarkan hasil lima percobaan di atas maka diperoleh hasil f1 score pada penerapan algoritma bayes adalah subset ke-1 adalah 74,80%, subset ke-2 adalah 76,40%, subset ke-3 adalah 79,20%, subset ke-4 adalah 80,00% dan subset ke-5 adalah 81,60%, maka berdasarkan hasil percobaan f1 score paling tinggi didapat pada percobaan subset ke-5 dengan nilai f1 score sebesar 81,60%.

Berdasarkan gambar 4.16 menunjukkan bahwa performa F1 score akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik f1 score pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.16.



Gambar 4.16 Grafik F1 score *Naive Bayes*

4.3.1.6. Pengujian MAE *Naive Bayes*

Pengujian kedua dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.26.

Tabel 4.26 Hasil MAE *Naive Bayes* Subset 1 (200 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	0,0132	0,0129	0,0139	0,0133	0,0136	0,01398

Hasil pengujian Tabel 4.26 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui

metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,0132, Fold ke-2 mendapatkan hasil 0,0119, Fold ke-3 mendapatkan hasil 0,0129, Fold ke-4 mendapatkan hasil 0,0133 dan Fold ke-5 mendapatkan hasil 0,0136, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 0,01298

Pengujian kedua dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.27.

Tabel 4.27 Hasil MAE *Naive Bayes* Subset 2 (400 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	0,0142	0,0135	0,0138	0,0140	0,0140	0,01390

Hasil pengujian Tabel 4.27 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,0142, Fold ke-2 mendapatkan hasil 0,0135, Fold ke-3 mendapatkan hasil 0,0138, Fold ke-4 mendapatkan hasil 0,0140 dan Fold ke-5 mendapatkan hasil 0,0140, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 0,01390.

Pengujian ketiga dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.28.

Tabel 4.28 Hasil MAE *Naive Bayes* Subset 3 (600 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	0,0140	0,0133	0,0137	0,0138	0,0128	0,01352

Hasil pengujian Tabel 4.28 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,0140, Fold ke-2 mendapatkan hasil 0,0133, Fold ke-3 mendapatkan hasil 0,0137, Fold ke-4 mendapatkan hasil 0,0138 dan Fold ke-5 mendapatkan hasil 0,0128, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma *Naive Bayes* sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 0,01352.

Pengujian keempat dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.29.

Tabel 4.29 Hasil MAE *Naive Bayes* Subset 4 (800 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	0,0139	0,0130	0,0129	0,0135	0,0139	0,01344

Hasil pengujian Tabel 4.29 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 4 dengan

jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,0139, Fold ke-2 mendapatkan hasil 0,0130, Fold ke-3 mendapatkan hasil 0,0129, Fold ke-4 mendapatkan hasil 0,0135 dan Fold ke-5 mendapatkan hasil 0,0139, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 0,01344%.

Pengujian kelima dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.30.

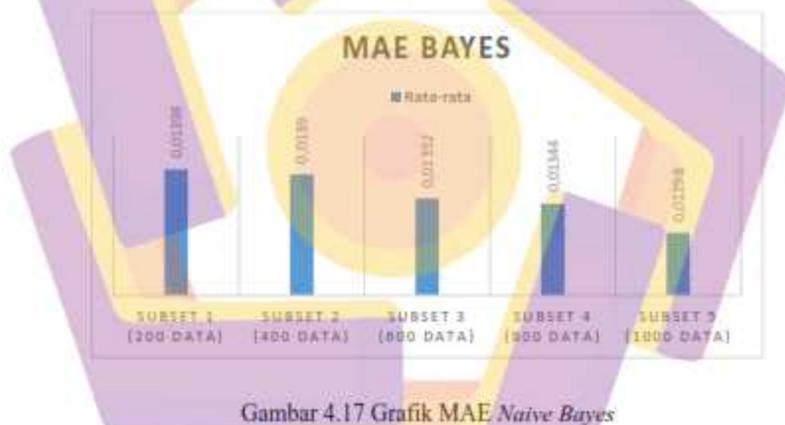
Tabel 4.30 Hasil MAE *Naive Bayes* Subset 5 (1000 Data)

Subset	MAE (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
5	0.0132	0.0119	0.0129	0.0133	0.0136	0.01298

Hasil pengujian Tabel 4.30 merupakan pengujian dengan menggunakan algoritma bayes yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,0132, Fold ke-2 mendapatkan hasil 0,0119, Fold ke-3 mendapatkan hasil 0,0129, Fold ke-4 mendapatkan hasil 0,0133 dan Fold ke-5 mendapatkan hasil 0,0136, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 0,01298.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil MAE pada penerapan algoritma bayes adalah subset ke-1 adalah 0,01398, subset ke-2 adalah 0,01390, subset ke-3 adalah 0,01352, subset ke-4 adalah 0,01344% dan subset ke-5 adalah 0,01298, maka berdasarkan hasil percobaan MAE paling baik didapat pada percobaan subset ke-5 dengan nilai MAE sebesar 0,01298.

Berdasarkan gambar 4.17 menunjukkan bahwa performa MAE akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik MAE pada penerapan algoritma Naive Bayes sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.17.



Gambar 4.17 Grafik MAE Naive Bayes

4.3.2. Pengujian Support Vector Machine

4.3.2.1. Pengujian Akurasi Support Vector Machine

Pengujian pertama dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.31.

Tabel 4.31 Hasil Akurasi *Support Vector Machine* Subset 1 (200 Data)

Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
1	78	80	78	79	81	79,40

Hasil pengujian Tabel 4.31 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 78%, Fold ke-2 mendapatkan hasil 80%, Fold ke-3 mendapatkan hasil 78%, Fold ke-4 mendapatkan hasil 79% dan Fold ke-5 mendapatkan hasil 81%, dengan demikian nilai rata-rata akurasi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 79,40%.

Pengujian kedua dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.32.

Tabel 4.32 Hasil Akurasi *Support Vector Machine* Subset 2 (400 Data)

Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
2	78	82	80	80	83	80,60

Hasil pengujian Tabel 4.32 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 78%,

Fold ke-2 mendapatkan hasil 82%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 83%, dengan demikian nilai rata-rata rata-rata akurasi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 76,60%.

Pengujian ketiga dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.33.

Tabel 4.33 Hasil Akurasi *Support Vector Machine* Subset 3 (600 Data)

Subset	Akurasi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	79	82	80	81	84	80,80

Hasil pengujian Tabel 4.33 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 79%, Fold ke-2 mendapatkan hasil 82%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 81% dan Fold ke-5 mendapatkan hasil 84%, dengan demikian nilai rata-rata rata-rata akurasi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 80,80%.

Pengujian keempat dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.34.

Tabel 4.3 Hasil Akurasi *Support Vector Machine* Subset 4 (800 Data)

Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
4	80	86	83	83	86	83,60

Hasil pengujian Tabel 4.34 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 86%, Fold ke-3 mendapatkan hasil 83%, Fold ke-4 mendapatkan hasil 83% dan Fold ke-5 mendapatkan hasil 86%, dengan demikian nilai rata-rata rata-rata akurasi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 83,60%.

Pengujian kelima dilakukan untuk nilai tingkat performa akurasi dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.35.

Tabel 4.35 Hasil Akurasi *Support Vector Machine* Subset 5 (1000 Data)

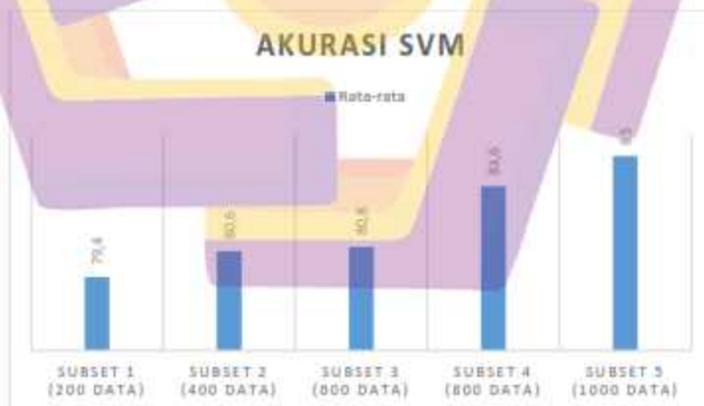
Subset	Akurasi (100%)					
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	Rata-rata
5	80	88	85	84	88	85,00

Hasil pengujian Tabel 4.35 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 88%, Fold ke-3 mendapatkan hasil 85%, Fold ke-4

mendapatkan hasil 84% dan Fold ke-5 mendapatkan hasil 88%, dengan demikiran nilai rata-rata rata-rata akurasi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 85,00%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil akurasi pada penerapan algoritma svm adalah subset ke-1 adalah 79,40%, subset ke-2 adalah 80,60%, subset ke-3 adalah 80,80%, subset ke-4 adalah 83,60% dan subset ke-5 adalah 85,00%, maka berdasarkan hasil percobaan akurasi paling tinggi didapat pada percobaan subset ke-5 dengan nilai akurasi sebesar 85,00%.

Berdasarkan gambar 4.18 menunjukkan bahwa performa akurasi akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik akurasi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.18.



Gambar 4.18 Grafik Akurasi *Support Vector Machine*

4.3.2.2. Pengujian Presisi Support Vector Machine

Pengujian kedua dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.36.

Tabel 4.36 Hasil Presisi Support Vector Machine Subset 1 (200 Data)

Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	78	79	80	80	83	80,00

Hasil pengujian Tabel 4.36 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 78%, Fold ke-2 mendapatkan hasil 79%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 83%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 80,00%.

Pengujian kedua dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.37.

Tabel 4.37 Hasil Presisi Support Vector Machine Subset 2 (400 Data)

Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	78	80	80	80	83	80,20

Hasil pengujian Tabel 4.37 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 78%, Fold ke-2 mendapatkan hasil 80%, Fold ke-3 mendapatkan hasil 80%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 83%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 80,20%.

Pengujian ketiga dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.38.

Tabel 4.38 Hasil Presisi *Support Vector Machine* Subset 3 (600 Data)

Substet	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	80	83	82	81	84	82,00

Hasil pengujian Tabel 4.38 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 83%, Fold ke-3 mendapatkan hasil 82%, Fold ke-4 mendapatkan hasil 81% dan Fold ke-5 mendapatkan hasil 84%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 82,20%.

Pengujian keempat dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.39.

Tabel 4.39 Hasil Presisi *Support Vector Machine* Subset 4 (800 Data)

Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	80	85	83	80	85	82,60

Hasil pengujian Tabel 4.39 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 85%, Fold ke-3 mendapatkan hasil 83%, Fold ke-4 mendapatkan hasil 80% dan Fold ke-5 mendapatkan hasil 85%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 82,60%.

Pengujian kelima dilakukan untuk nilai tingkat performa presisi dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.40.

Tabel 4.40 Hasil Presisi *Support Vector Machine* Subset 5 (1000 Data)

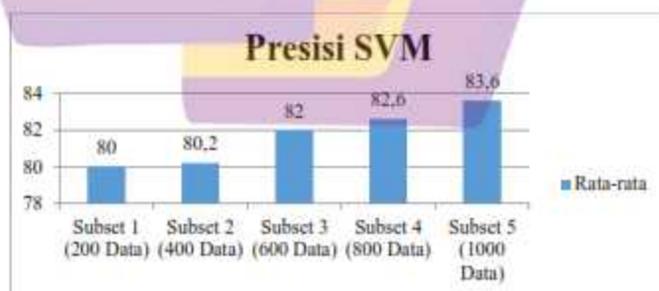
Subset	Presisi (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	81	86	84	82	85	83,60

Hasil pengujian Tabel 4.40 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui

mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 81%, Fold ke-2 mendapatkan hasil 86%, Fold ke-3 mendapatkan hasil 84%, Fold ke-4 mendapatkan hasil 82% dan Fold ke-5 mendapatkan hasil 85%, dengan demikian nilai rata-rata rata-rata presisi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 83,60%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil presisi pada penerapan algoritma svm adalah subset ke-1 adalah 80,00%, subset ke-2 adalah 80,20%, subset ke-3 adalah 82,00%, subset ke-4 adalah 82,60% dan subset ke-5 adalah 83,60%, maka berdasarkan hasil percobaan presisi paling tinggi didapat pada percobaan subset ke-5 dengan nilai presisi sebesar 83,60%.

Berdasarkan gambar 4.19 menunjukkan bahwa performa presisi akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik presisi pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.19.



Gambar 4.19 Grafik Presisi *Support Vector Machine*

4.3.2.3. Pengujian Recall Support Vector Machine

Pengujian ketiga dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.41.

Tabel 4.41 Hasil Recall *Support Vector Machine* Subset 1 (200 data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	79	78	82	79	78	79,20

Hasil pengujian Tabel 4.41 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 79%, Fold ke-2 mendapatkan hasil 78%, Fold ke-3 mendapatkan hasil 82%, Fold ke-4 mendapatkan hasil 79% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 79,20%.

Pengujian kedua dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.42.

Tabel 4.42 Hasil Recall Support Vector Machine Subset 2 (400 data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	79	80	83	79	78	79,80

Hasil pengujian Tabel 4.42 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 79%, Fold ke-2 mendapatkan hasil 80%, Fold ke-3 mendapatkan hasil 83%, Fold ke-4 mendapatkan hasil 79% dan Fold ke-5 mendapatkan hasil 78%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 79,80%.

Pengujian ketiga dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.43.

Tabel 4.43 Hasil Recall *Support Vector Machine* Subset 3 (600 data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	80	81	85	81	80	81,40

Hasil pengujian Tabel 4.3 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 81%, Fold ke-3 mendapatkan hasil 85%, Fold ke-4 mendapatkan hasil 81% dan Fold ke-5 mendapatkan hasil 80%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 81,40%.

Pengujian keempat dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.44.

Tabel 4.44 Hasil Recall *Support Vector Machine* Subset 4 (800 data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	80	83	86	82	83	83,60

Hasil pengujian Tabel 4.44 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 83%, Fold ke-3 mendapatkan hasil 86%, Fold ke-4 mendapatkan hasil 82% dan Fold ke-5 mendapatkan hasil 83%, dengan demikian nilai rata-rata rata-rata recall pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 83,60%.

Pengujian kelima dilakukan untuk nilai tingkat performa recall dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.45.

Tabel 4.45 Hasil Recall *Support Vector Machine* Subset 5 (1000 data)

Subset	Recall (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	82	84	87	83	84	84,00

Hasil pengujian Tabel 4.45 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui

mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 82%, Fold ke-2 mendapatkan hasil 84%, Fold ke-3 mendapatkan hasil 87%, Fold ke-4 mendapatkan hasil 83% dan Fold ke-5 mendapatkan hasil 84%, dengan demikian nilai rata-rata recall pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 84,00%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil recall pada penerapan algoritma svm adalah subset ke-1 adalah 79,20%, subset ke-2 adalah 79,80%, subset ke-3 adalah 81,40%, subset ke-4 adalah 83,60% dan subset ke-5 adalah 84,00%, maka berdasarkan hasil percobaan recall paling tinggi didapat pada percobaan subset ke-5 dengan nilai recall sebesar 84,00%.

Berdasarkan gambar 4.20 menunjukkan bahwa performa recall akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik recall pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.20.



Gambar 4.20 Grafik Recall Support Vector Machine

4.3.2.4. Pengujian Waktu training *Support Vector Machine*

Pengujian keempat dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.46.

Tabel 4.46 Hasil Waktu training *Support Vector Machine* Subset 1 (200 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	28	38	41	50	46	40,60

Hasil pengujian Tabel 4.46 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 28 detik, Fold ke-2 mendapatkan hasil 38 detik, Fold ke-3 mendapatkan hasil 41 detik, Fold ke-4 mendapatkan hasil 50 detik dan Fold ke-5 mendapatkan hasil 46 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 40,60 detik.

Pengujian kedua dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.47.

Tabel 4.47 Hasil Waktu training *Support Vector Machine* Subset 2 (400 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	29	38	43	51	48	41,80

Hasil pengujian Tabel 4.47 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 29 detik, Fold ke-2 mendapatkan hasil 38 detik, Fold ke-3 mendapatkan hasil 43 detik, Fold ke-4 mendapatkan hasil 51 detik dan Fold ke-5 mendapatkan hasil 48 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 41,80 detik.

Pengujian ketiga dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.48.

Tabel 4.48 Hasil Waktu training *Support Vector Machine* Subset 3 (600 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	32	40	45	52	50	43,80

Hasil pengujian Tabel 4.48 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 32 detik, Fold ke-2 mendapatkan hasil 40 detik, Fold ke-3 mendapatkan hasil 45 detik, Fold ke-4 mendapatkan hasil 52 detik dan Fold ke-5 mendapatkan hasil 50 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan

algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 43,80 detik.

Pengujian keempat dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.49.

Tabel 4.49 Hasil Waktu training *Support Vector Machine* Subset 4 (800 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	35	43	49	55	54	47,20

Hasil pengujian Tabel 4.49 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 35 detik, Fold ke-2 mendapatkan hasil 43 detik, Fold ke-3 mendapatkan hasil 49 detik, Fold ke-4 mendapatkan hasil 55 detik dan Fold ke-5 mendapatkan hasil 54 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 47,20 detik.

Pengujian kelima dilakukan untuk nilai tingkat performa waktu training dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.50.

Tabel 4.50 Hasil Waktu training *Support Vector Machine* Subset 5 (1000 Data)

Subset	Waktu (Detik)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	39	48	51	58	56	50,40

Hasil pengujian Tabel 4.50 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 39 detik, Fold ke-2 mendapatkan hasil 48 detik, Fold ke-3 mendapatkan hasil 51 detik, Fold ke-4 mendapatkan hasil 58 detik dan Fold ke-5 mendapatkan hasil 56 detik, dengan demikian nilai rata-rata rata-rata waktu training pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 50,40%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil waktu training pada penerapan algoritma svm adalah subset ke-1 adalah 40,60 detik, subset ke-2 adalah 41,80 detik, subset ke-3 adalah 43,80 detik, subset ke-4 adalah 47,20 detik dan subset ke-5 adalah 50,40 detik, maka berdasarkan hasil percobaan waktu training paling cepat didapat pada percobaan subset ke-1 dengan nilai waktu training sebesar 40,60 detik.

Berdasarkan gambar 4.21 menunjukkan bahwa performa waktu training akan mengalami penurunan pada setiap subset atau penambahan data. Oleh karena itu grafik waktu training pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.21.



Gambar 4.21 Grafik Waktu training *Support Vector Machine*

4.3.2.5. Pengujian F1 score *Support Vector Machine*

Pengujian pertama dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.51.

Tabel 4.51 Hasil F1 score *Support Vector Machine* Subset 1 (200 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	80	79	84	81	80	80,80

Hasil pengujian Tabel 4.51 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 80%, Fold ke-2 mendapatkan hasil 79%, Fold ke-3 mendapatkan hasil 84%, Fold ke-4 mendapatkan hasil 81% dan Fold ke-5 mendapatkan hasil 80%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Support Vector

Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 80,80%.

Pengujian kedua dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.52.

Tabel 4.52 Hasil F1 score *Support Vector Machine* Subset 2 (400 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	81	81	87	82	80	82,20

Hasil pengujian Tabel 4.52 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 81%, Fold ke-2 mendapatkan hasil 81%, Fold ke-3 mendapatkan hasil 87%, Fold ke-4 mendapatkan hasil 82% dan Fold ke-5 mendapatkan hasil 80%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 82,20%.

Pengujian ketiga dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.53.

Tabel 4.53 Hasil F1 score *Support Vector Machine* Subset 3 (600 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	83	82	88	83	84	84,00

Hasil pengujian Tabel 4.53 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 83%, Fold ke-2 mendapatkan hasil 82%, Fold ke-3 mendapatkan hasil 88%, Fold ke-4 mendapatkan hasil 83% dan Fold ke-5 mendapatkan hasil 84%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 84,00%.

Pengujian keempat dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.54.

Tabel 4.54 Hasil F1 score *Support Vector Machine* Subset 4 (800 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	84	86	88	84	86	85,60

Hasil pengujian Tabel 4.54 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 4 dengan jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 84%, Fold ke-2 mendapatkan hasil 86%, Fold ke-3 mendapatkan hasil 88%, Fold ke-4 mendapatkan hasil 84% dan Fold ke-5 mendapatkan hasil 86%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Support Vector

Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 85,60%.

Pengujian kelima dilakukan untuk nilai tingkat performa f1 score dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.55.

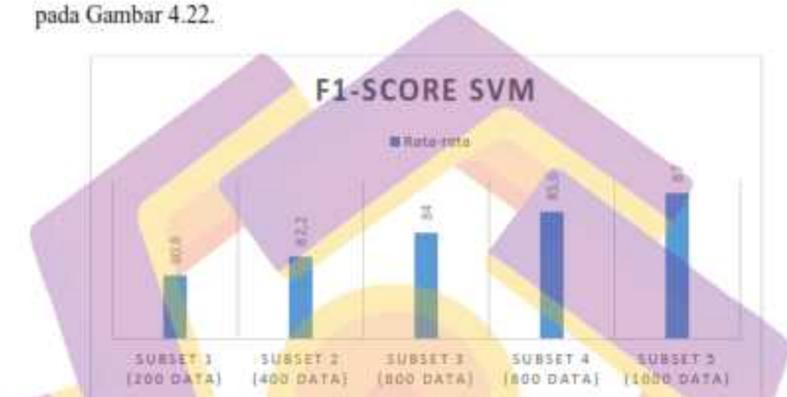
Tabel 4.55 Hasil F1 score *Support Vector Machine* Subset 5 (1000 Data)

Subset	F1 score (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	88	88	87	86	86	87,00

Hasil pengujian Tabel 4.55 merupakan pengujian dengan menggunakan algoritma svm yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 88%, Fold ke-2 mendapatkan hasil 88%, Fold ke-3 mendapatkan hasil 87%, Fold ke-4 mendapatkan hasil 86% dan Fold ke-5 mendapatkan hasil 86%, dengan demikian nilai rata-rata rata-rata f1 score pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 87,00%.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil f1 score pada penerapan algoritma svm adalah subset ke-1 adalah 80,80%, subset ke-2 adalah 82,20%, subset ke-3 adalah 84,00%, subset ke-4 adalah 85,60% dan subset ke-5 adalah 87,00%, maka berdasarkan hasil percobaan f1 score paling tinggi didapat pada percobaan subset ke-5 dengan nilai f1 score sebesar 87,00%.

Berdasarkan gambar 4.22 menunjukkan bahwa performa f1 score akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik f1 score pada penerapan algoritma Support Vector Machine sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.22.



Gambar 4.22 Grafik F1 score *Support Vector Machine*

4.3.2.6. Pengujian MAE SVM

Pengujian kedua dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-1 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.56.

Tabel 4.56 Hasil MAE *SVM* Subset 1 (200 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
1	0,00148	0,00139	0,00142	0,00144	0,00146	0,001438

Hasil pengujian Tabel 4.56 merupakan pengujian dengan menggunakan algoritma SVM yang dilakukan melalui skenario pengujian pada subset 1 dengan jumlah dataset sebesar 200 dataset, kemudian dilakukan validasi melalui

mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,00148, Fold ke-2 mendapatkan hasil 0,00139, Fold ke-3 mendapatkan hasil 0,00142, Fold ke-4 mendapatkan hasil 0,00144 dan Fold ke-5 mendapatkan hasil 0,00146, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma SVM sistem klasifikasi sentimen vaksin covid-19 pada subset ke-1 adalah 0,001438.

Pengujian kedua dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-2 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.57.

Tabel 4.57 Hasil MAE SVM Subset 2 (400 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
2	0,00146	0,00137	0,00140	0,00143	0,00142	0,001416

Hasil pengujian Tabel 4.57 merupakan pengujian dengan menggunakan algoritma SVM yang dilakukan melalui skenario pengujian pada subset 2 dengan jumlah dataset sebesar 400 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,00146, Fold ke-2 mendapatkan hasil 0,00137, Fold ke-3 mendapatkan hasil 0,00140, Fold ke-4 mendapatkan hasil 0,00143 dan Fold ke-5 mendapatkan hasil 0,00142, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma SVM sistem klasifikasi sentimen vaksin covid-19 pada subset ke-2 adalah 0,001416.

Pengujian ketiga dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-3 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.58.

Tabel 4.58 Hasil MAE SVM Subset 3 (600 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
3	0,00143	0,00135	0,00139	0,00141	0,00141	0,001398

Hasil pengujian Tabel 4.58 merupakan pengujian dengan menggunakan algoritma SVM yang dilakukan melalui skenario pengujian pada subset 3 dengan jumlah dataset sebesar 600 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,00143, Fold ke-2 mendapatkan hasil 0,00135, Fold ke-3 mendapatkan hasil 0,00139, Fold ke-4 mendapatkan hasil 0,00141 dan Fold ke-5 mendapatkan hasil 0,00141, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma SVM sistem klasifikasi sentimen vaksin covid-19 pada subset ke-3 adalah 0,001398.

Pengujian keempat dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-4 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.59.

Tabel 4.59 Hasil MAE SVM Subset 4 (800 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
4	0,00141	0,00131	0,00134	0,00138	0,00140	0,001368

Hasil pengujian Tabel 4.59 merupakan pengujian dengan menggunakan algoritma SVM yang dilakukan melalui skenario pengujian pada subset 4 dengan

jumlah dataset sebesar 800 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,00141, Fold ke-2 mendapatkan hasil 0,00131, Fold ke-3 mendapatkan hasil 0,00134, Fold ke-4 mendapatkan hasil 0,00138 dan Fold ke-5 mendapatkan hasil 0,00140, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma SVM sistem klasifikasi sentimen vaksin covid-19 pada subset ke-4 adalah 0,001368.

Pengujian kelima dilakukan untuk nilai tingkat performa MAE dengan menggunakan konfigurasi subset ke-5 yang telah dilakukan sebelumnya dengan hasil seperti pada Tabel 4.60.

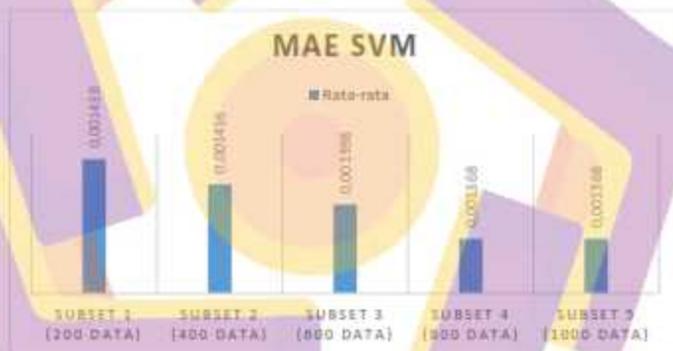
Tabel 4.60 Hasil MAE SVM Subset 5 (1000 Data)

Subset	MAE (100%)					Rata-rata
	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	
5	0,00138	0,00127	0,00131	0,00135	0,00136	0,001334

Hasil pengujian Tabel 4.60 merupakan pengujian dengan menggunakan algoritma SVM yang dilakukan melalui skenario pengujian pada subset 5 dengan jumlah dataset sebesar 1000 dataset, kemudian dilakukan validasi melalui mekanisme pembagian data training 80% dan testing 20% melalui metode k-Fold Cross Validation dengan jumlah 5 Fold, pada Fold ke-1 mendapatkan hasil 0,00138, Fold ke-2 mendapatkan hasil 0,00127, Fold ke-3 mendapatkan hasil 0,00131, Fold ke-4 mendapatkan hasil 0,00135 dan Fold ke-5 mendapatkan hasil 0,00136, dengan demikian nilai rata-rata rata-rata MAE pada penerapan algoritma SVM sistem klasifikasi sentimen vaksin covid-19 pada subset ke-5 adalah 0,001334.

Berdasarkan hasil lima percobaan diatas maka diperoleh hasil MAE pada penerapan algoritma SVM adalah subset ke-1 adalah 0,001438, subset ke-2 adalah 0,001416, subset ke-3 adalah 0,001398, subset ke-4 adalah 0,001368 dan subset ke-5 adalah 0,001334, maka berdasarkan hasil percobaan MAE paling baik didapat pada percobaan subset ke-5 dengan nilai MAE sebesar 0,001334.

Berdasarkan gambar 4.23 menunjukkan bahwa performa MAE akan mengalami peningkatan pada setiap subset atau penambahan data. Oleh karena itu grafik MAE pada penerapan algoritma SVM sistem klasifikasi sentimen vaksin covid-19 secara berturut-turut dari Fold 1 s.d. 5 seperti pada Gambar 4.23.



Gambar 4.23 Grafik MAE SVM

4.4. Specifying learning (Representasi Hasil)

4.4.1. Analisis Perbandingan Hasil Pengujian

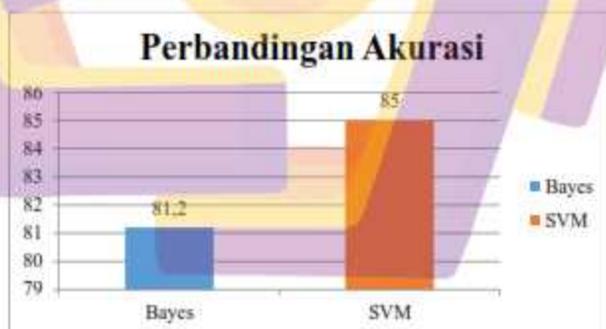
Pada bagian ini hasil percobaan dari implementasi *Naive Bayes* serta *Support vector machine* akan dilakukan perbandingan. Hasil yang akan dibandingkan merupakan rata-rata terbaik dari keseluruhan percobaan pengujian dengan parameter Akurasi, Presisi, Recall, Waktu training dan MAE yang telah

dilakukan pada sub bab sebelumnya. Hasil nilai rata-rata performa pada percobaan terbaik dengan menggunakan dua algoritma dapat dilihat pada Tabel 4.61.

Tabel 4.61 Hasil Perbandingan Tingkat Performa

No	Algoritma	Akurasi (%)	Presisi(%)	Recall (%)	Waktu (Detik)	F1-Score (%)	MAE
1	BAYES	81,20	79,40	79,60	9,00	81	0,01298
2	SVM	85	83,60	84	40,60	87	0,001334
	Selisih	3,80	4,20	4,40	31,60	6,00	0,01165

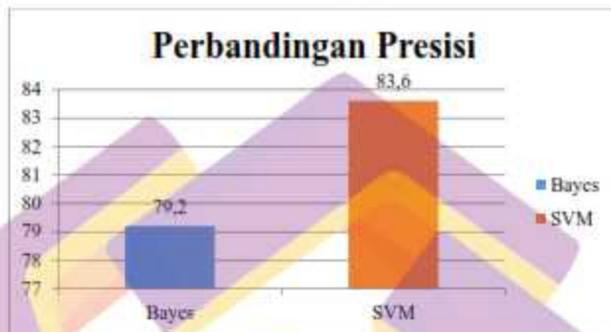
Berdasarkan hasil pada Tabel 4.61 dapat ditarik suatu kesimpulan bahwa nilai performa untuk akurasi model algoritma *Support vector machine* memiliki tingkat akurasi yang lebih baik dengan nilai akurasi sebesar 85% dengan selisih 3,80% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 81,20% pada konfigurasi dataset sebanyak 1000 data. Adapun grafik dari perbandingan nilai akurasi pada Gambar 4.24.



Gambar 4.24 Perbandingan Performa Akurasi

Kemudian untuk nilai tingkat performa presisi *Support vector machine* memiliki lebih baik dengan nilai presisi sebesar 83,60% dengan selisih

4,20% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 79,40% pada konfigurasi dataset sebanyak 1000 data. Adapun grafik dari perbandingan nilai presisi pada Gambar 4.25.



Gambar 4.25 Perbandingan Performa Presisi

Kemudian untuk nilai tingkat performa recall *Support vector machine* memiliki lebih baik dengan nilai presisi sebesar 84% dengan selisih 4,40% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 79,60% pada konfigurasi dataset sebanyak 1000 data. Adapun grafik dari perbandingan nilai recall pada Gambar 4.26.



Gambar 4.26 Perbandingan Performa Recall

Kemudian untuk nilai tingkat performa waktu training *Naive Bayes* memiliki lebih baik dengan nilai waktu sebesar 9,00 detik dengan selisih 31,60 detik jika dibandingkan dengan model algoritma *Support vector machine* dengan nilai yang lebih rendah sebesar 40,60 detik pada konfigurasi dataset sebanyak 1000 data. Adapun grafik dari perbandingan nilai waktu training pada Gambar 4.27.



Gambar 4.27 Perbandingan Performa Waktu Training

Kemudian untuk nilai tingkat performa f1-score *Support vector machine* memiliki lebih baik dengan nilai presisi sebesar 87% dengan selisih 6,00% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 81% pada konfigurasi dataset sebanyak 1000 data. Adapun grafik dari perbandingan nilai F1-Score pada Gambar 4.28.



Gambar 4.28 Perbandingan Performa F1-Score

Kemudian untuk nilai tingkat performa MAE *Support vector machine* memiliki lebih baik dengan nilai MAE sebesar 0,001334 dengan selisih 0,01165 jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 0,01298 pada konfigurasi dataset sebanyak 1000 data. Adapun grafik dari perbandingan nilai MAE pada Gambar 4.29.



Gambar 4.29 Perbandingan Performa MAE

4.4.2. Penarikan Kesimpulan

Berdasarkan hasil pengujian pada semua skenario maka didapatkan suatu temuan bahwa tingkat performa algoritma SVM lebih baik pada parameter pengujian nilai tingkat performa akurasi dengan nilai 85%, performa presisi dengan nilai 83,60%, performa recall dengan nilai 84% serta performa waktu training dengan nilai 40,60 detik, performa f1 score dengan nilai 87% jika dibandingkan dengan naïve bayes dengan tingkat performa akurasi dengan nilai 81,20%, performa presisi dengan nilai 79,40%, performa recall dengan nilai 79,60%, serta performa f1 score dengan nilai 81%, akan tetapi untuk parameter pengujian waktu training algoritma naïve bayes lebih unggul dengan nilai waktu training tercepat 9,00 detik jika dibandingkan SVM dengan waktu training 40,60 detik, untuk parameter pengujian MAE algoritma SMV lebih unggul dengan nilai MAE terbaik 0,001334 jika dibandingkan Naïve Bayes dengan MAE 0,01298.

BAB V

PENUTUP

5.1. Kesimpulan

Setelah melalui tahap pengujian tingkat performa dengan parameter akurasi, presisi, recall, waktu Training, F1-Score dan MAE pada model algoritma Naive Bayes serta Support vector machine untuk klasifikasi sentiment mengenai vaksin covid-19, maka dapat diambil beberapa kesimpulan antara lain:

1. Model algoritma *Naive Bayes* memiliki nilai tingkat performa akurasi dengan nilai 81,20%, performa presisi dengan nilai 79,40%, performa recall dengan nilai 79,60%, performa waktu training dengan nilai 9,00 detik serta performa f1 score dengan nilai 81% dan nilai MAE pada subset 5 dengan nilai 0,01298 dengan jumlah 1000 data. Sedangkan pada model algoritma Support vector machine memiliki nilai tingkat performa akurasi dengan nilai 85%, performa presisi dengan nilai 83,60%, performa recall dengan nilai 84% serta performa waktu training dengan nilai 40,60 detik, performa f1 score dengan nilai 87% dan nilai MAE dengan nilai 0,001334 pada subset 5 dengan jumlah 1000 data.
2. Pengaruh jumlah *records* terhadap model algoritma yang digunakan memiliki pengaruh baik atau tren kenaikan terhadap nilai akurasi, presisi, recall, F1 score dan MAE dimana nilai akurasi, presisi, recall, F1 score dan MAE akan meningkat jika jumlah *records* semakin

bertambah, akan tetapi berdampak buruk atau tren penurunan pada kecepatan waktu training dimana waktu training akan melambat jika jumlah *records* semakin bertambah.

5.2. Saran

Saran yang dapat digunakan oleh peneliti-peneliti selanjutnya untuk meningkatkan hasil kesimpulan pada penelitian ini, adalah sebagai berikut :

1. Pada penelitian ini tahap *preprocessing* dilakukan secara manual dan pada penelitian selanjutnya tahap *preprocessing* dapat dilakukan dengan mengimplementasikan teknik *feature selection* menggunakan teknik atau algoritma tertentu.
2. Pada penelitian selanjutnya dapat dilakukan pengujian dengan mengimplementasikan model-model algoritma machine learning lain serta pendekatan statistik ataupun *squential* lain.

DAFTAR PUSTAKA

PUSTAKA BUKU

Santoso, B. 2007. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.

Yogyakarta: Graha Ilmu

Sulaeman, M. (1998). Ilmu Budaya Dasar Suatu Pengantar. Bandung: Rafika Aditama.

Kusumadewi, Sri. 2003. Artificial Intellegence Teknik dan Aplikasi. Yogyakarta : Graha Ilmu.

Kementrian Kesehatan RI. 2020. Penetapan Jenis Vaksin Untuk Pelaksanaan Vaksinasi Corona Virus Disease 2019 (Covid-19). Kementrian Kesehatan RI Nomor HK.01.07/Menkes/9860/2020.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

Syarifuddin, M. 2020. "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan KNN", Inti Nusa Mandiri, Vol. 15. NO. 1 Agustus 2020

Rian, T., Agung, T., Ira, D. 2020. "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)". Jurnal Media Informatika Budidarma. Volume 4, Nomor 3, Juli 2020, Page 650-658.

- Ratino, Noor, H., Sita A., Windu, G. 2020. "Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes". Jurnal JUPITER Vol. 12 No. 2 Bulan 10 Tahun 2020, Hal 1 - 11
- Pristiyono, Mulkan, R., Muhammad, A., Agus, A, Fauziah, H. 2020. " Analisis sentimen vaksin COVID-19 di Indonesia menggunakan Algoritma Naïve Bayes". Konferensi Tahunan tentang Ilmu Komputer dan Teknologi Rekayasa (AC2SET). doi: 10.1088 / 1757-899X / 1088/L/012045.
- Ardianne, L., F., Rima, D., R., Nia, A., F. 2021. "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter". Jurnal Dinda. Vol.1x No.1 (2021) Feb 2021
- Cahyo, P., Nisa, H. 2020. "Analisis Sentimen Covid-19 Sebagai Pandemi Media Sosial". Jurnal Internasional Sistem Informasi & Teknologi . Vol. 4, Tidak. 1, (2020), hlm.509-517
- Deni, G., Dwiza, R., Dian, A., Fajar, A., Salma, A. 2020. "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023". Jurnal Teknik Komputer AMIK BSI. Volume VI No.1 Januari 2020.
- M. Wino Adi Putra, Susanti, Erlin, Herwin. 2020. "Analisis Sentimen Dompok Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier". IT Journal Research and Development (ITJRD). Vol.5, No.1, Agustus 2020,
- Muh. Ali Masnun, Eny Sulistyowati, Irfa Ronaboyd. 2021. "Pelindungan Hukum Atas Vaksin Covid-19 Dan Tanggung Jawab Negara Pemenuhan Vaksin

- Dalam Mewujudkan Negara Kesejahteraan". Jurnal Ilmu Hukum. Volume 17 Nomor 1 Februari 2021
- Samsir, Ambiyar, Unung Verawardina, Firman Edi, Ronal Watrianthos . 2021. "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes". Jurnal Media Informatika Budidarma. Volume 5, Nomor 1, Januari 2021, Page 157-163
- Marulak Pardede. 2021. "Aspek Hukum Kekarantinaan Kesehatan Dan Perlindungan Konsumen Dalam Penanggulangan Pandemi Covid-19 (Legal Aspects Of Health Quarantine And Consumer Protection In Treatment Of Covid-19 Pandemic)". Jurnal Penelitian Hukum De Jure. Volume 21 Nomor 1, Maret 2021
- Rochani Nani Rahayu, Sensusiyati . 2021. "Vaksin Covid 19 Di Indonesia : Analisis Berita Hoax". Intelektiva : Jurnal Ekonomi, Sosial & Humaniora. VOL.2 NO. 07 - FEBRUARY 2021
- Fajar Fathur Rachman, Setia Pramana. 2020. "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter". ndonesian of Health Information Management Journal. Vol.8, No.2, Desember 2020, p.100-109
- Praditio Aditya Nugraha, Ristu Saptono, Meiyanto Eko Sulistyو. 2013. "Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan". JURNAL ITSMART Vol 2. No 2. Desember 2013 ISSN : 2301-7201

PUSTAKA ELEKTRONIK

Nareza, M. (2021). Jangan Termakan Isu, Ini Fakta Penting Vaksin COVID-19.

[https:// www.alodokter.com/](https://www.alodokter.com/), diakses 25 Januari 2021

World Health Organization (WHO). 2021. Coronavirus disease (COVID-19)

pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, diakses 25 Januari 2021

Our World In Data. 2021. Coronavirus (COVID-19) Vaccinations.

<https://ourworldindata.org/covid-vaccinations>, diakses 25 Januari 2021

Global Digital Statistic. 2021. Digital, Social & Mobile in 2019.

<https://wearesocial.com/blog/2021/01/digital-use-around-the-world-in-January-2020>, diakses 25 Januari 2021



LAMPIRAN

Text	Pakar 1 Ummi Nur Jannah S.pd	Pakar 2 Hanifa Hafiza, MPD	Pakar 3 Sherly Miranti, S.sos	Sistem
Alhamdulillah,,gak rugi gw nentang di vaksin..gw bilang juga apa??	Negatif	Negatif	Negatif	Negatif
Didesa saya mulai diberlakukan hal gila kayak gini jika tidak mau vaksin.	Negatif	Negatif	Negatif	Negatif
Sya bersumpah celakalah celakalah celakalah semua nakes di dunia ini yg dimaksud dr. Tifa d bawah ini. Celaka malapetaka.musibah,marabah aya dll. Celaka pnyakit imbas dri vaksin mulai dri cardiovascular disorder,myocarditis,cancer dll. CELAKALAH KALIAN SEMUA! TUNGGU KAMI DI AKHIRAT	Negatif	Negatif	Negatif	Negatif
Ramai mati sebab covid. Sebab lambat dapat vaksin. Sebab kekangan kewangan. Mana duit negara?	Negatif	Negatif	Negatif	Negatif
Selama hampir 2 tahun, mereka menjanjikan anda basmi covid dengan larangan bepergian, penguncian, masker, dan vaksin. Sekarang Fauci dengan santai mengatakan tentu saja kita tidak bisa memberantasnya. Ini sudah sangat jelas sejak awal, namun anda sangat suka dibohongi	Negatif	Negatif	Negatif	Negatif

Salom selamat pagi selamat penuh berkah, selamat beraktivitas. God bless all. Salam sehat, Tetap jalankan PROKES, Ayo vaksin covid-19.	Positif	Positif	Positif	Positif
Yuk segera vaksin Covid-19, bagi yang belum divaksin.... vaksin aman dan efektif	Positif	Positif	Positif	Positif
Covid-19 belum tentu ada virusnya. Vaksin C-19 sudah pasti ada virusnya.	Negatif	Negatif	Negatif	Negatif
Fokus pemerintah pada 2022 pengendalian covid-19 Vaksin Lawan Covid19	Positif	Positif	Positif	Positif
Menteri Koordinator bidang Maritim dan Investasi (Menko Marves) Luhut Binsar Pandjaitan menyebut tak cukup waktu cepat untuk bisa mengetahui efek varian Omicron terhadap vaksin Covid-19. Dia memperkirakan membutuhkan waktu 1-2 minggu.	Netral	Netral	Netral	Netral
Ayo Ikuti Vaksinasi Gratis !!! JADILAH SALAH SATU AGEN YG Mendukung PERCEPATAN VAKSINASI NASIONAL DEMI MEWUJUDKAN INDONESIA BEBAS COVID-19 Tempat : GERALI VAKSIN PRESISI POLRES TTS Hari/Tgl : SABTU (27 NOVEMBER 2021) Sasaran : REMAJA & MASYARAKAT UMUM (VAKSIN TAHAP 1 & TAHAP 2)	Positif	Positif	Positif	Positif

Video Podcast Deddy Corbuzier Beberapa Waktu Lalu Tentang Efek Vaksin Covid-19 Bagi Kehidupan Masyarakat Sekarang Semua Terbukti. #DudungBaliho	Netral	Netral	Netral	Netral
Jangan takut atau ragu, vaksin covid-19 aman dan halal. Ayo Disiplin Prokes	Positif	Positif	Positif	Positif
Vaksin covid-19 aman dan halal digunakan di semua kalangan. Ayo Disiplin Prokes	Positif	Positif	Positif	Positif
Giak Vaksin Covid 19 di Desa Jawa Tengah	Netral	Positif	Netral	Netral
Ayo Bersama Vaksin Cegah Penyebaran Covid-19	Positif	Positif	Positif	Positif
Jngn ada varian covid baru lagi yak. Ayoo sehat sehat. #ayovaksin #indonesiasehat	Netral	Netral	Netral	Netral
Kasihannya yg divaksin sinovac gak boleh masuk, siapa yg bertanggung jawab dgn vaksin sinovac.	Negatif	Negatif	Negatif	Negatif
wkwkw maaf oot tapi gue blm vaksin pas kabar sinovac gabisa buat ke luar negri tapi gue masi 17 woi nangis	Negatif	Negatif	Negatif	Negatif
OHMNON JANGAN GITU ANJIR GAKUAT VAKSIN GUE CUMA SINO VAC	Negatif	Negatif	Negatif	Negatif
LUPA VAKSIN AJA MASIH SINO VAC ANYING	Negatif	Negatif	Negatif	Negatif
wait gue baru sadar walopun mereka ngadain konser di saudi but vaksin gue kan sinovac	Negatif	Negatif	Negatif	Negatif
MAJU LO SINI GA TAKUT GW UDH VAKSIN SINO VAC	Positif	Positif	Positif	Positif

ProtectHealth ingin memaklumkan bahawa pada masa ini Pusat Pemberian Vaksin (PPV) yg menawarkan Sinovac sbg pilihan jenis vaksin bagi penerimaan dos penggalak adalah terhad. Kebanyakan PPV yang terlibat dlm pemberian dos penggalak buat masa sekarang menawarkan vaksin jenis Pfizer	Netral	Netral	Netral	Netral
Ade gw moga ga kenapa napa.matanya pasca vaksin sinovac 2 dos Takut buta aja	Negatif	Negatif	Negatif	Negatif
vaksin sinovac menangis liat kabarr konser di seoul sksksksksk duitnya pun belum terlihatsee u soon tannies	Negatif	Negatif	Negatif	Negatif
Maju lo jim sini gua udah vaksin sinovac #PARKJIMIN #JIMIN	Positif	Positif	Positif	Positif
Menteri Kesehatan Budi Gunadi Sadikin meminta masyarakat jangan ragu untuk divaksinasi menggunakan vaksin selain Sinovac yang stoknya mulai menipis. Vaksin Lindungi Keluarga	Positif	Positif	Positif	Positif

<p>Gerai Vaksin Presisi Polres Kubu Raya.</p> <p>Vaksinasi Covid 19 dosis 1 dan 2.</p> <p>Bertempat di Polsek Sungai Ambawang yang dilaksanakan pada hari Sabtu, 4 Desember 2021 pukul 08.00 WIB s/d Selesai dengan syarat sebagai berikut:</p> <p>Vaksin yang digunakan adalah SINOVAC.</p> <p>@pontianakvaksinasi</p>	Positif	Positif	Positif	Positif
<p>Bhabinkamtibmas Kalurahan Palbapang laksanakan pengamanan giat vaksinasi masal dosis kedua bagi warga Palbapang. Adapun masyarakat yg berhasil di vaksinasi sejumlah 317 orang dari 325 undangan dengan jenis vaksin sinovac.</p>	Positif	Positif	Positif	Positif
<p>Ayo datang ke sentra vaksinasi di PMI Kota Tangerang. Ada PFIZER dan SINOVAC loh. #vaksin #palangmerahindonesia #pmikotatangerang</p>	Positif	Positif	Positif	Positif
<p>Kak Yoongi jidatan hiks</p> <p>Maju lo kak gue udh vaksin</p> <p>Tapi vaksin gue sinovac jadi gampang mleyot. GWS Nay</p>	Negatif	Negatif	Negatif	Negatif
<p>Masyarakat tidak perlu ragu untuk vaksinasi covid-19 anak 6-12 tahun termasuk menggunakan vaksin sinovac. Vaksin Selamatkan Keluarga</p>	Positif	Positif	Positif	Positif

BPOM menyatakan aman vaksin sinovac untuk diberikan pada anak 6-12 tahun	Positif	Positif	Positif	Positif
Dipanggil lagi utk uji klinis vaksin booster (Pfizer, AZ, Sinovac). Bismillah. Semoga bermanfaat utk masyarakat umum.	Positif	Positif	Positif	Positif
vaksin lain juga sama bagusnya dengan sinovac	Positif	Positif	Positif	Positif
Kegiatan Pelaksanaan Vaksin masal sinovac dan AstraZeneka bagi para *PARA PELAJAR PONPES MANGARAJA PANUSUNAN AKHIR HASIBUAN DAN MASYARAKAT DESA PARGARUTAN JAE DI KECAMATAN ANGKOLA TIMUR KABUPATEN TAPANULI SELATAN.* dengan tema *AYO IKUTI VAKSINASI KESEHATAN PULIH EKONOMI BANGKIT*	Netral	Positif	Netral	Netral
Pemerintah efikasi vaksin sinovac 96,4% terhadap covid-19 varian non alva #NKRIPulihDenganVaksinasi	Positif	Positif	Positif	Positif
Bharaka Tri Febrianto S.Kep anggota Batalyon C Pelopor Sat Brimob Polda Kalbar yang BKO Urkes Polres Sintang, melaksanakan vaksinasi di setiap penjuru Kabupaten Sintang. Demi Mencegah penularan covid19. Jenis vaksin yang disalurkan yaitu Vaksin Sinovac Dosis 1 dan Dosis 2.	Netral	Netral	Netral	Netral

Bhabinkamtibmas Desa Kalangsari Aiptu Toyiban Melaksanakan Pengamanan Pelaksanaan Kegiatan Vaksinasi Covid-19 Serentak Indonesia Untuk umum Usia 12 tahun ke Atas Dosis 1 dan 2, dengan Vaksin Sinovac, bertempat di Kantor Puskesmas Kalangsari Kec Kec Rengasdengklok Kab Karawang.	Netral	Netral	Netral	Netral
anjir nyesel knp vaksin gue sinovac ya sial	Negatif	Negatif	Negatif	Negatif
Ada beberapa negara yg ga ngebolehkan vaksin sinovac buat masuk ke negaranya gitu	Negatif	Negatif	Negatif	Negatif
Halo min @contactap2 Mau tanya apa benar kalo vaksin sinovac tidak bisa naik pesawat ke Malaysia ya? Padahal saya udah vaksin dosis 2 sinovac :(Negatif	Negatif	Negatif	Negatif
.....
MAJU SINI LO BERDUA, GUE UDAH VAKSIN 2X. t-tapi sinovac	Negatif	Negatif	Negatif	Negatif