

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROGRAM  
PEMERINTAH INDONESIA PADA APLIKASI TWITTER MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE***

**SKRIPSI**

Oleh:  
**HELMI ZULFAN FANANI**  
NIM. 19650125



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROGRAM  
PEMERINTAH INDONESIA PADA APLIKASI TWITTER  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE***

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:  
**HELMI ZULFAN FANANI**  
NIM. 19650125

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**HALAMAN PERSETUJUAN**


**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROGRAM  
PEMERINTAH INDONESIA PADA APLIKASI TWITTER  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE***

**SKRIPSI**

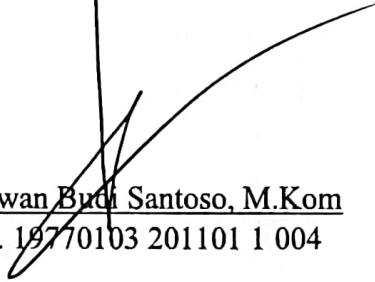
Oleh:  
**HELMI ZULFAN FANANI**  
**NIM. 19650125**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 29 November 2023

Pembimbing I,


  
Dr. Cahyo Crysdiyan  
NIP. 19740424 200901 1 008

Pembimbing II,

  
Dr. Irwan Budi Santoso, M.Kom  
NIP. 19770103 201101 1 004

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## HALAMAN PENGESAHAN

### KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROGRAM PEMERINTAH INDONESIA PADA APLIKASI TWITTER MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*

#### SKRIPSI

Oleh:  
**HELMI ZULFAN FANANI**  
NIM. 19650125

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 11 Desember 2023

#### Susunan Dewan Penguji

Ketua Penguji : Dr. M. Amin Hariyadi, M.T  
NIP. 19670018 200501 1 001

Anggota Penguji I : Fajar Rohman Hariri, M.Kom  
NIP. 19890515 201801 1 001

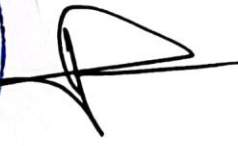
Anggota Penguji II : Dr. Cahyo Crysdi  
NIP. 19740424 200901 1 008

Anggota Penguji III : Dr. Irwan Budi Santoso, M.Kom  
NIP. 19770103 201101 1 004

(  )  
(  )  
(  )  
(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Helmi Zulfan Fanani

NIM : 19650125

Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Klasifikasi Sentimen Masyarakat Terhadap Program  
Pemerintah Indonesia Pada Aplikasi Twitter Menggunakan  
Metode Support Vector Machine.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 11 Desember 2023

Yang membuat pernyataan,



Helmi Zulfan Fanani

NIM. 19650125

## **MOTTO**

*“Life is never flat, go up and never stop”*

*“Hidup ini seperti pensil yang lama lama akan habis,  
tetapi akan meninggalkan tulisan yang indah dalam  
kehidupan”*

## **HALAMAN PERSEMBAHAN**

Saya persembahkan karya ini kepada:  
Ayahku Nuri Arif, S.Pd dan Ibuku Aliyah  
Yang telah mengantarkan aku hingga titik ini

Saudaraku,  
Haris Budi Erwanto, Alfian Edi Masyhudi, Andi Syahrial Lathifi  
Yang telah memberi bimbingan serta nasihat

Sahabat-sahabatku,  
Semua anggota grup ahlussunnah wal jamaah  
Yang telah memberikan kebahagiaan selama perkuliahan

Teman seperjuangan,  
Teknik Informatika angkatan 2019  
Sukses selalu untuk kita semua

## KATA PENGANTAR

*Assalamualaikum Wr. Wb.*

Alhamdulillah Wasyukurillah segala puji bagi Allah Subhanahu Wa Ta'ala atas segala rahmat dan karunia-Nya sehingga penulis dapat diberikan Kesehatan dan kelancaran dalam menyelesaikan skripsi yang berjudul “Klasifikasi Sentimen Masyarakat Terhadap Program Pemerintah Indonesia Pada Aplikasi Twitter Menggunakan Metode *Support Vector Machine*”. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad Sallallahu ‘Alaihi wa Sallam. Dan semoga kita sekalian mendapatkan syafaatnya di hari akhir kelak, *Aamiin*.

Rasa syukur dan terima kasih yang sebesar-besarnya penulis ucapkan kepada semua pihak-pihak yang senantiasa memberikan dukungan kepada penulis untuk dapat berproses dalam pengerjaan skripsi ini. Ucapan ini penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPM, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak Dr. Cahyo Crysdiand dan Bapak Dr. Irwan Budi Santoso, M.Kom selaku Dosen Pembimbing I dan Dosen Pembimbing II yang dengan segala arahan,



saran serta masukan yang diberikan sangat berarti dalam setiap langkah dalam proses pengerjaan skripsi ini.

5. Bapak Dr. M. Amin Hariyadi, M.T dan Bapak Fajar Rohman Hariri, M.Kom selaku dosen Penguji I dan dosen Penguji II yang telah menguji serta memberikan masukan sehingga penulis dapat menghasilkan skripsi yang baik.
6. Segenap Dosen, Admin, Laboran dan jajaran pada Program Studi Teknik Informatika yang telah memberikan bimbingan dan bantuan selama studi.
7. Ibu, Ayah, serta 3 saudara yang selalu memberikan semangat untuk terus berproses, dan rangkaian doa yang terbaik selalu dicurahkan sehingga dapat menjadikan lancar dan selesainya skripsi ini dengan baik.
8. Tata Farindra Ayuningtyas selaku teman, sahabat, dan orang istimewa yang selalu menemani dan mendukung serta bangga terhadap apa yang dilakukan oleh penulis.
9. Saudara Himpunan Mahasiswa Teknik Informatika (HIMATIF ENCODER) yang telah membantu penulis dalam berproses dan memberikan berbagai pengalaman yang sangat berharga.
10. Sahabat-sahabati ku “Ahlussunnah Wal Jama’ah” yang selalu tanggap dalam segala Kondisi. Terima kasih atas segala ruang dan waktunya.

Dalam penyusunan skripsi ini masih terdapat kekurangan dan harapannya semoga skripsi ini dapat memberikan manfaat kepada pembaca.

*Wassalamualaikum Wr. Wb.*

Malang, 11 Desember 2023

Penulis

## DAFTAR ISI

<b>HALAMAN PENGAJUAN</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN</b> .....	<b>v</b>
<b>MOTTO</b> .....	<b>vi</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>vii</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR TABEL</b> .....	<b>xii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>ABSTRAK</b> .....	<b>xiv</b>
<b>ABSTRACT</b> .....	<b>xv</b>
مستخلص البحث .....	<b>xvi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah .....	5
1.3 Tujuan Penelitian .....	6
1.4 Batasan Masalah.....	6
1.5 Manfaat Penelitian .....	6
<b>BAB II STUDI PUSTAKA</b> .....	<b>8</b>
2.1 Klasifikasi Sentimen .....	8
2.2 TF-IDF .....	11
2.3 Support Vector Machine .....	13
<b>BAB III METODE PENELITIAN</b> .....	<b>17</b>
3.1 Pengumpulan Data .....	17
3.2 Desain Sistem.....	18
3.2.1. <i>Preprocessing</i> .....	19
3.2.2. TF-IDF.....	27
3.2.3. Support Vector Machine .....	29
<b>BAB IV UJI COBA DAN PEMBAHASAN</b> .....	<b>34</b>
4.1 Skenario Uji Coba .....	34
4.2 Hasil Uji Coba.....	38
4.2.1 Hasil Uji Coba-1 .....	39
4.2.2 Hasil Uji Coba-2 .....	44
4.2.3 Hasil Uji Coba-3 .....	46
4.2.4 Hasil Uji Coba-4 .....	47
4.2.5 Hasil Uji Coba-5 .....	48
4.3 Pembahasan.....	52
<b>BAB V KESIMPULAN DAN SARAN</b> .....	<b>59</b>

5.1 Kesimpulan .....	59
5.2 Saran .....	60

**DAFTAR PUSTAKA**

**LAMPIRAN**

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkait Mengenai Klasifikasi Sentimen.....	11
Tabel 2.2 Penelitian Terkait Mengenai TF-IDF .....	12
Tabel 2.3 Penelitian Terkait Mengenai <i>Support Vector Machine</i> .....	16
Tabel 3.1 Pelabelan Data.....	18
Tabel 3.2 Contoh <i>Emoji Conversion</i> .....	21
Tabel 3.3 Contoh <i>Cleaning</i> .....	22
Tabel 3.4 Contoh <i>Case Folding</i> .....	23
Tabel 3.5 Contoh <i>Spell Checker</i> .....	23
Tabel 3.6 Contoh <i>Tokenizing</i> .....	24
Tabel 3.7 Contoh <i>Stopword Removal</i> .....	25
Tabel 3.8 Contoh <i>Stemming</i> .....	26
Tabel 3.9 Perhitungan <i>Term Frequency</i> .....	28
Tabel 3.10 Perhitungan <i>Inverse Document Frequency</i> .....	28
Tabel 4.1 <i>Confussion Matrix</i> .....	37
Tabel 4.2 Sampel Data Penelitian .....	38
Tabel 4.3 Hasil Akurasi Uji Coba 1 .....	43
Tabel 4.4 Hasil Uji Coba 2.....	45
Tabel 4.5 Hasil Uji Coba 3.....	46
Tabel 4.6 Hasil Uji <i>10-Fold Cross Validation</i> .....	47
Tabel 4.7 Hasil Uji Reduksi 1 .....	48
Tabel 4.8 Hasil Uji Reduksi 2 .....	49
Tabel 4.9 Hasil Uji Reduksi 3 .....	49
Tabel 4.10 Hasil Uji Reduksi 4 .....	50
Tabel 4.11 Hasil Uji Reduksi 5 .....	50
Tabel 4.12 Hasil Uji Reduksi Dimensi.....	51

## DAFTAR GAMBAR

Gambar 3.1 Desain Sistem .....	19
Gambar 3.2 Proses <i>Preprocessing</i> .....	20
Gambar 3.3 <i>Flowchart Cleaning</i> .....	21
Gambar 3.4 <i>Flowchart Case Folding</i> .....	22
Gambar 3.5 <i>Flowchart Tokenizing</i> .....	24
Gambar 3.6 <i>Flowchart Stopword Removal</i> .....	25
Gambar 3.7 <i>Flowchart Stemming</i> .....	26
Gambar 3.8 Alur TF-IDF .....	27
Gambar 3.9 Visualisasi data .....	29
Gambar 3.10 <i>Flowchart</i> proses pelatihan .....	32
Gambar 4.1 Visualisasi <i>10-fold cross validation</i> .....	36
Gambar 4.2 <i>Confusion Matrix</i> Rasio 6:4 .....	40
Gambar 4.3 <i>Confusion Matrix</i> Rasio 7:3 .....	41
Gambar 4.4 <i>Confusion Matrix</i> Rasio 8:2 .....	42
Gambar 4.5 <i>Confusion Matrix</i> Rasio 9:1 .....	43
Gambar 4.6 <i>Confusion Matrix</i> Uji Coba tanpa <i>Spell checker &amp; emoji conversion</i> .....	44
Gambar 4.7 <i>Confusion Matrix</i> Uji Coba <i>Spell checker &amp; emoji conversion</i> .....	45

## ABSTRAK

Fanani, Helmi Zulfan. 2023. **Klasifikasi Sentimen Masyarakat Terhadap Program Pemerintah Indonesia Pada Aplikasi Twitter Menggunakan Metode Support Vector Machine**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crys dian. (II) Irwan Budi Santoso, M.Kom.

**Kata kunci:** *Klasifikasi Sentimen, SVM, Program Pemerintah, Twitter.*

Program pemerintah adalah sebuah inisiatif atau Tindakan yang dicanangkan oleh pemerintah yang ditujukan kepada masyarakat baik berdampak langsung maupun tidak langsung. Adanya program pemerintah ini tentu menimbulkan beragam tanggapan masyarakat, ada masyarakat yang mendukung dan ada pula yang tidak mendukung. Topik ini sangat menarik untuk diteliti, karena ada berbagai sentimen positif dan negatif terkait program pemerintah Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program pemerintah Indonesia pada aplikasi twitter menggunakan algoritma *Support Vector Machine*. Dari penelitian ini didapatkan 1000 data sentimen yang terdiri dari 536 sentimen positif dan 464 sentimen negatif. Dari penelitian ini menghasilkan bahwa algoritma Support Vector Machine dapat bekerja dengan baik pada proses klasifikasi sentimen. Nilai akurasi terbaik didapatkan dari hasil uji coba menggunakan rasio perbandingan data negatif dan positif dalam data testing 1:9 dengan akurasi sebesar 91%. Selain itu didapatkan hasil bahwa terdapat beberapa aspek yang mempengaruhi performa sistem seperti rasio data training dan testing, penggunaan fitur pada tahap *preprocessing*, rasio data positif dan negatif pada data testing, serta pemilihan data yang digunakan untuk training dan testing.

## ABSTRACT

Fanani, Helmi Zulfan. 2023. **The Classification of Community Sentiment Towards Indonesian Government Programs Regarding the Twitter Application Using the Support Vector Machine Method.** Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Cahyo Crysdiyan. (II) Irwan Budi Santoso, M.Kom.

Government programs are initiatives or actions initiated by the government, and it is provided for the community either directly or indirectly. The existence of this government program certainly causes a variety of public responses. Some people support the program, and some do not. This topic is very interesting to be investigated since there are many various positive and negative sentiments related to the Indonesian government program. This study aims to analyze public sentiment towards the Indonesian government program regarding the twitter application involving the *Support Vector Machine* algorithm. This study obtained 1000 sentiment data consisting of 536 positive sentiments and 464 negative sentiments. Based on this study, it is found out that the *Support Vector Machine* algorithm can work well in the sentiment classification process. The best accuracy value was obtained from the test results using the ratio of negative and positive data in 1:9 data testing with an accuracy of 91%. In addition, the results showed that there were several aspects that affected the performance of the system, such as the ratio of training and testing data, the use of features at the *preprocessing* stage, the ratio of positive and negative data on testing data, and the selection of data used for training and testing.

**Keywords:** *Sentiment Classification, SVM, Government Program, Twitter.*

## مستخلص البحث

فناي، حلمي زلفان. 2023. تصنيف المشاعر العامة تجاه برامج الحكومة الإندونيسية على تطبيق تويتر باستخدام خوارزمية آلة المتجة الداعم (*Support Vector Machine*). البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. جاهيو كريسديان. المشرف الثاني: إيروان بودي سانتوسو، الماجستير.

**الكلمات الرئيسية:** تصنيف المشاعر، SVM، البرنامج الحكومي، تويتر.

البرنامج الحكومي هو مبادرة أو عمل تطلقه الحكومة يستهدف المجتمع سواء بشكل مباشر أو غير مباشر. من المؤكد أن وجود هذا البرنامج الحكومي يسبب مجموعة متنوعة من الاستجابات العامة، فبعض الناس يؤيدون والبعض الآخر يرفضون. هذا الموضوع مثير للاهتمام للغاية للبحث، لأن هناك العديد من المشاعر الإيجابية والسلبية المتعلقة ببرامج الحكومة الإندونيسية. يهدف هذا البحث إلى تحليل المشاعر العامة تجاه برامج الحكومة الإندونيسية على تطبيق تويتر باستخدام خوارزمية آلة المتجة الداعم *Support Vector Machine*. من هذا البحث، تم الحصول على 1000 بيانات المشاعر تتكون من 536 مشاعر إيجابية و 464 مشاعر سلبية. نتج عن هذا البحث أن خوارزمية *Support Vector Machine* يمكن أن تعمل بشكل جيد في عملية تصنيف المشاعر. يتم الحصول على أفضل قيمة دقة من نتائج الاختبار باستخدام نسبة مقارنة البيانات السلبية والإيجابية في بيانات الاختبار 1:9 بدقة 91%. بالإضافة إلى ذلك، وجد أن هناك العديد من الجوانب التي تؤثر على أداء النظام مثل نسبة بيانات التدريب والاختبار، واستخدام الميزات في مرحلة المعالجة المسبقة، ونسبة البيانات الإيجابية والسلبية في بيانات الاختبار، واختيار البيانات المستخدمة للتدريب والاختبار.



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Sentimen masyarakat adalah perasaan atau pandangan yang dimiliki oleh masyarakat terhadap suatu hal tertentu. Sentimen masyarakat diukur terhadap program pemerintah yang dikomunikasikan melalui media sosial. Sentimen masyarakat dapat diklasifikasikan menjadi dua kategori utama, yaitu positif dan negatif (Bagus Sasmita dkk., 2022). Sentimen positif menunjukkan bahwa masyarakat menganggap program tersebut baik dan dapat memberikan manfaat, sedangkan sentimen negatif menunjukkan bahwa masyarakat menganggap program tersebut buruk dan tidak dapat memberikan manfaat. Oleh karena itu, sentimen masyarakat dapat dijadikan acuan untuk penilaian suatu kebijakan atau program yang dibuat oleh suatu instansi atau pemerintah.

Program pemerintah hakikatnya adalah suatu program yang dibuat untuk masyarakat baik secara langsung maupun tidak langsung. Program pemerintah dapat juga diartikan sebagai sebuah kebijakan publik dikarenakan program pemerintah disusun dan dirancang dengan sasaran tujuannya adalah publik. Kebijakan publik ini memiliki tujuan untuk mencapai kesejahteraan masyarakat melalui peraturan yang dibuat serta dapat memperoleh nilai-nilai yang diperlukan oleh warga negara untuk meningkatkan kualitas hidup (Rosdiana dkk., 2019). Dalam penelitian ini menggunakan data sentimen program Pemerintah Indonesia dalam bidang Pendidikan. Bidang Pendidikan dipilih karena bidang tersebut merupakan salah satu bidang yang sangat penting dan strategis dalam pembangunan

dan perubahan dalam suatu negara (Ghoni dkk., 2020). Beberapa waktu belakang pemerintah Indonesia membuat berbagai program baru di bidang Pendidikan untuk mengatasi masalah Pendidikan di masa pandemi hingga masa pasca pandemi baik untuk tingkat dasar maupun jenjang perkuliahan.

Untuk mewujudkan program yang baik demi kesejahteraan umum, pemerintah perlu melakukan perbaikan terhadap kualitas program yang akan dicanangkan. Hal tersebut perlu dilakukan karena tidak semua program mendapatkan dukungan dari masyarakat, dikarenakan masyarakat merasa dirugikan dengan program tersebut. Dalam melakukan perbaikan terhadap program tersebut pemerintah membutuhkan *feedback* dari masyarakat yang merupakan sumber informasi utama, dikarenakan tujuan program tersebut diperuntukkan untuk masyarakat (Rosdiana dkk, 2019). Salah satu cara untuk mendapatkan umpan balik dari masyarakat adalah dengan menggunakan sosial media. Salah satu contoh media sosial yang populer dalam media untuk mengekspresikan diri adalah aplikasi twitter, masyarakat menggunakan twitter untuk menuliskan opininya terhadap suatu objek.

Program yang dicanangkan oleh pemerintah Indonesia sangat berdampak terhadap kehidupan masyarakat, oleh karena itu pemerintah diharapkan dapat memberikan pelayanan berupa program yang terbaik. Hal ini disebutkan dalam Al-Qur'an surah Ali 'Imran ayat 110 berikut:

كُنْتُمْ خَيْرَ أُمَّةٍ أُخْرِجَتْ لِلنَّاسِ تَأْمُرُونَ بِالْمَعْرُوفِ وَتَنْهَوْنَ عَنِ الْمُنْكَرِ وَتُؤْمِنُونَ بِاللَّهِ ۗ وَلَوْ ءَامَنَ أَهْلُ الْكِتَابِ لَكَانَ خَيْرًا لَهُمْ ۚ مِنْهُمْ الْمُؤْمِنُونَ وَأَكْثَرُهُمُ الْفَاسِقُونَ

*“Kamu (umat Islam) adalah umat terbaik yang dilahirkan untuk manusia (selama) kamu menyuruh (berbuat) yang makruf, mencegah dari yang mungkar, dan*

*beriman kepada Allah. Seandainya Ahlulkitab beriman, tentulah itu lebih baik bagi mereka. Di antara mereka ada yang beriman dan kebanyakan mereka adalah orang-orang fasik” (Q.S Ali ‘Imran: 110).*

Menurut tafsir Jalalain, ayat ini menunjukkan bahwa umat Muslim adalah umat terbaik yang Allah telah mengutus untuk kepentingan seluruh manusia. Mereka memiliki tanggung jawab untuk memerintah kebaikan (amar ma'ruf) dan mencegah kejahatan (nahi munkar) (Al-Mahalli & As-Suyuthi, n.d.). Ini adalah bagian dari keimanan mereka kepada Allah. Dalam ayat tersebut dijelaskan bahwa setiap kita memiliki tanggung jawab untuk memerintah kebaikan dan mencegah keburukan. Hal ini berarti pemerintah yang bertanggung jawab untuk mengajak masyarakatnya kepada kebaikan dan mencegah kemungkaran dengan memberikan program yang baik. Sedangkan masyarakat bertanggung jawab untuk menyuruh kebaikan dan mencegah kemungkaran dengan memberikan tanggapannya terhadap program yang diberikan pemerintah, dengan itu diharapkan dapat terbentuk program yang baik sesuai dengan kebutuhan masyarakat.

Hal tersebut juga selaras dengan apa yang dijelaskan dalam Al-Qur'an surah Al-A'raf ayat 199 sebagai berikut:

خُذِ الْعَفْوَ وَأْمُرْ بِالْعُرْفِ وَأَعْرِضْ عَنِ الْجَاهِلِينَ

*“Jadilah pemaaf dan suruhlah orang mengerjakan yang makruf, serta jangan pedulikan orang-orang yang bodoh.” (QS. Al-A'raf: 199)*

Menurut Ibnu Katsir, ayat ini mengajarkan pentingnya memberikan nasihat yang baik dan konstruktif kepada pemimpin atau pemerintah ketika mereka

melakukan kesalahan atau bertindak tidak adil (Dr. Abdullah, 2004a). Hal tersebut demi terwujudnya program baik untuk kesejahteraan masyarakat.

Opini yang diberikan masyarakat melalui aplikasi twitter tersebut terdiri dari ulasan yang bersifat mendukung adanya program dan menentang adanya program (Prasetyo dkk, 2021). Jumlah penilaian/opini masyarakat tersebut semakin lama akan meningkat secara signifikan sehingga untuk mengetahui sebuah ulasan tergolong dalam sentimen positif, negatif, ataupun netral akan membutuhkan banyak waktu apabila dilakukan secara manual. Untuk mengatasi hal tersebut, akan dibuat sistem komputerisasi untuk mengkategorikan sentimen tersebut dengan menggunakan teknik pembelajaran mesin. Untuk melakukan analisis sentimen, berbagai teknik machine learning dapat dimanfaatkan, salah satunya adalah metode support vector machine (SVM).

Support Vector Machine adalah teknik klasifikasi yang dapat memprediksi kelas berdasarkan pola yang ditemukan melalui hasil proses pelatihan. Hyperplane adalah pola yang dibuat oleh algoritma SVM yang berfungsi sebagai garis pemisah kelas (Faiq dkk., 2022). Meskipun SVM adalah pendekatan yang lebih baru daripada teknik lainnya, SVM bekerja lebih baik di berbagai bidang aplikasi seperti bioinformatika, pengenalan tulisan tangan, kategorisasi teks, dan bidang lainnya. SVM memiliki keunggulan dibandingkan teknik klasifikasi lainnya karena dapat mencapai pemisahan linier pada data nonlinier berdimensi tinggi dengan memilih fungsi yang sesuai. Manfaat inilah yang mendorong penulis untuk memilih SVM sebagai metode klasifikasi untuk penelitian ini.

Penelitian dengan judul Klasifikasi Sentimen Masyarakat terhadap Program Pemerintah Indonesia pada Aplikasi Twitter diajukan oleh penulis menggunakan metode Support Vector Machine berdasarkan penjelasan latar belakang dan permasalahan. Diharapkan dengan dilakukannya penelitian ini akan tercipta suatu sistem yang dapat digunakan untuk mengukur bagaimana persepsi masyarakat Indonesia terhadap program-program Pemerintah Indonesia. Selain itu, untuk menilai kinerja model dan keakuratan hasil klasifikasi yang ditemukan dalam penelitian ini, selanjutnya dilakukan pengukuran untuk menilai kinerja model menggunakan metrik nilai kinerja seperti akurasi, *recall*, presisi, dan *f-measure*. Rasio data yang diklasifikasikan dengan benar, baik positif maupun negatif, terhadap jumlah total data dalam sampel dapat digunakan untuk mengukur tingkat akurasi.

## 1.2 Pernyataan Masalah

Pernyataan masalah dalam adalah sebagai berikut:

1. Seberapa besar nilai akurasi, *recall*, presisi dan *f-measure* yang diperoleh dari klasifikasi sentiment masyarakat terhadap program pemerintah Indonesia pada aplikasi twitter menggunakan metode Support Vector Machine?
2. Apa saja aspek yang mempengaruhi performa dalam klasifikasi sentimen masyarakat terhadap program Pemerintah Indonesia pada aplikasi Twitter menggunakan metode *Support Vector Machine*?

### 1.3 Tujuan Penelitian

Tujuan dari penelitian yang akan dilakukan ini yaitu sebagai berikut:

1. Mengukur nilai akurasi, *recall*, presisi dan *f-measure* yang diperoleh dari klasifikasi sentiment masyarakat terhadap program pemerintah Indonesia pada aplikasi twitter menggunakan metode Support Vector Machine.
2. Menganalisa aspek apa saja yang mempengaruhi performa dalam klasifikasi sentimen masyarakat terhadap program Pemerintah Indonesia pada aplikasi Twitter menggunakan metode *Support Vector Machine*.

### 1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Program Pemerintah Indonesia pada bidang Pendidikan pada periode 2019-2024.
2. Data yang digunakan merupakan data primer berupa *tweet* atau komentar masyarakat di aplikasi twitter tentang program Pemerintah Indonesia pada bidang Pendidikan yaitu kampus merdeka, kurikulum merdeka, program Indonesia pintar, dan platform merdeka mengajar.

### 1.5 Manfaat Penelitian

Setelah melakukan klasifikasi sentiment masyarakat terhadap program pemerintah Indonesia pada aplikasi twitter menggunakan metode support vector machine diharapkan dapat memberikan manfaat sebagai berikut:

1. Dapat dijadikan bahan rujukan penelitian mengenai klasifikasi sentiment masyarakat menggunakan metode Support Vector Machine.

2. Memberikan informasi mengenai rekomendasi terhadap program pemerintah di bidang Pendidikan kepada yang berkepentingan untuk bahan penelitian selanjutnya dalam bidang intelligence dan sains.
3. Membantu Pemerintah Negara Indonesia dalam melakukan evaluasi terhadap program kerja yang telah dicanangkan untuk meningkatkan kualitas program kerjanya demi terwujudnya kesejahteraan umum yang dalam hal ini program pemerintah pada bidang Pendidikan.

## **BAB II**

### **STUDI PUSTAKA**

Pada studi pustaka peneliti akan membahas tentang penelitian terdahulu yang terkait dengan topik penelitian dan landasan teori yang digunakan sebagai acuan untuk melakukan penelitian.

#### **2.1 Klasifikasi Sentimen**

Opini masyarakat dapat diungkapkan secara online melalui beberapa platform media sosial seperti Twitter, Instagram, Tiktok, Facebook ataupun platform lainnya. Melalui opini tersebut dapat diperoleh perbedaan pemikiran, pendapat dan sikap dari setiap orang terhadap suatu masalah. Opini yang diungkapkan dapat berupa dukungan (pro), tidak berpihak (netral), dan menentang (kontra), opini tersebut tidak jarang diungkapkan oleh masyarakat melalui postingan dari akun pemerintah yang membahas mengenai program/kebijakan yang diterapkan seperti pada komentar postingan instagram atau komentar pada *tweet* di Twitter (Astuti dkk., 2022). Klasifikasi dilakukan dengan kaidah teks *mining* yang menggolongkan muatan teks berdasarkan pada kesamaan teksnya. Dengan menerapkan klasifikasi pada *tweets* di Twitter dapat membantu dalam mengekstraksi informasi dari sebuah masalah atau kejadian karena telah dikelompokkan berdasarkan kategorinya (Satrio & Fauzi, 2019).

Faiq dkk. (2022) melakukan klasifikasi terhadap sentiment masyarakat terhadap kebijakan pemerintah di media sosial twitter menggunakan metode *support vector machine* dengan ekspansi fitur GloVe. Peneliti menggunakan data



*tweet* sebanyak 16.597 *tweet* berbahasa Indonesia dengan beberapa topik kebijakan pemerintah. Pada tahap *word embedding* GloVe dilakukan proses representasi kata pada setiap dokumen dengan membentuk *corpus* dengan membuat matriks kemunculan yang menampilkan relasi antar kata dan juga seberapa sering munculnya kata dalam dokumen. Tahap tersebut dilakukan untuk mempermudah dalam tahap proses dalam pelatihan data yang jumlahnya besar. Hasil dari penelitian ini menunjukkan hasil pengujian menggunakan *feature expansion* dapat menghasilkan hasil yang berbeda sesuai dengan penggunaan corpus kata dan ukuran fitur, penggunaan pembobotan TF-IDF dan *feature expansion* dapat meningkatkan nilai akurasi hingga 4,77% dengan nilai yang didapatkan sebesar 79,52% dan nilai F-1 Score sebesar 0,7942.

Naz dkk. (2019) melakukan penelitian yang melakukan klasifikasi pada data sentiment di sosial media twitter dengan menggunakan empat fitur N-Gram yang berbeda dan juga tiga skema pembobotan yang berbeda. Dalam penelitian ini peneliti ingin menemukan kombinasi fitur dan pembobotan mana yang memiliki nilai akurasi yang terbaik. Proses N-Gram digunakan untuk memprediksi kata selanjutnya dari kata N-1 sebelumnya. Sedangkan proses pembobotan bertujuan untuk menghitung tingkat pentingnya fitur dalam suatu dokumen dengan mengubah kata menjadi vector kata yang memiliki bobot. Penggunaan *feature extraction* dapat mempengaruhi tingkat akurasi dalam proses klasifikasi secara signifikan. Dibuktikan dengan hasil yang didapatkan pada fitur unigram *feature extraction* TF mendapatkan akurasi 79,1% dan selanjutnya 80,0% pada TF-IDF dan 81,0 pada Biner.

Hal ini juga dibuktikan dengan penelitian yang dilakukan oleh Abd dkk. (2020) yang melakukan klasifikasi pada artikel menggunakan *feature extraction* yang berbeda. Penelitian ini diawali dengan pelabelan data dokumen menjadi tiga label, lalu dilanjutkan *preprocessing* meliputi normalization, stop words, dan stemming. Penggunaan *feature extraction* menggunakan dua teknik yaitu Term Frequency (TF) dan Term Frequency-Inverse Document Frequency (TF-IDF). Dengan menggunakan kernel linear dan pembagian data *training* dan data *testing* menggunakan rasio 7:3 hasil akurasi, presisi, recall, dan F1-score yang diperoleh menggunakan *feature extraction* TF berturut-turut adalah 91,93; 91,0; 92,33; dan 91,33. Sedangkan hasil yang diperoleh dengan menggunakan *feature extraction* TF-IDF menunjukkan peningkatan yang signifikan pada *performance* dengan nilai akurasi, presisi, recall, F1-score berturut-turut sebesar 95,16; 95,0; 95,66; dan 95,0.

Dalam klasifikasi sentimen, ada beberapa tahap yang penting dilakukan agar mendapatkan hasil klasifikasi dengan tingkat akurasi yang tinggi dan waktu yang cepat. Tahap *preprocessing* termasuk dalam tahapan yang perlu dilakukan, karena data yang awalnya mentah dapat dibersihkan melalui berbagai proses di tahap ini sehingga data menjadi bersih dan lebih optimal saat dilakukan pembobotan dan pelatihan model (Faiq dkk, 2022). Selain itu, pemilihan metode ekstraksi fitur juga dapat mempengaruhi performa dalam klasifikasi sentiment. Abd dkk., (2020) dalam penelitiannya melakukan klasifikasi sentiment dengan menggunakan dua ekstraksi fitur yang berbeda yaitu Term Frequency (TF) dan Term Frequency-Inverse Document Frequency (TF-IDF) yang hasilnya menunjukkan bahwa TF-IDF unggul nilai presentasi sebesar 3,23% dibandingkan dengan TF.

Tabel 2.1 Penelitian Terkait Mengenai Klasifikasi Sentimen

No.	Pengarang (tahun)	Judul	Metode	Hasil
1	Faiq dkk. (2022)	Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media sosial Twitter	<i>Feature expansion word embedding GloVe &amp; support vector machine</i>	Metode SVM dikombinasikan dengan <i>feature expansion</i> dapat meningkatkan proses klasifikasi dengan hasil akurasi sebesar 4,77% yaitu dengan nilai akurasi 79,52% dan nilai F-1 Score 0,7942
2	Naz dkk. (2019)	Sentiment Classification on Twitter Data Using Support Vector Machine	<i>support vector machine</i>	Metode SVM dapat melakukan klasifikasi dengan baik saat dilakukan perbandingan dengan beberapa <i>feature extraction</i> , dengan akurasi menggunakan TF 79,1% dan TF-IDF 80,0% dan 81,0% pada Biner.
3	Abd dkk. (2020)	Classifying Political Arabic Articles Using Support Vector Machine with Different Feature Extraction	<i>support vector machine &amp; TF-IDF</i>	Percobaan klasifikasi metode SVM menggunakan 2 <i>feature extraction</i> menghasilkan nilai yang berbeda, dengan menggunakan TF memperoleh akurasi sebesar 91% sedangkan menggunakan TF-IDF menghasilkan kenaikan nilai yang cukup signifikan yakni sebesar 95%.

## 2.2 TF-IDF

Dalam proses klasifikasi, data yang sebelumnya telah melalui proses preprocessing yang berbentuk text tidak akan dapat dibaca dan diproses oleh komputer, oleh karena itu perlu dilakukan proses yang dapat mengubah data teks

menjadi data yang berupa angka agar dapat dibaca dan diproses oleh sistem (Yerzi & Sibaroni, 2021). Pada penelitian ini proses pembobotan kata dilakukan menggunakan metode pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*) yang akan memberikan nilai pada *term* dengan melakukan perhitungan nilai kemunculan *term* tersebut lalu melakukan perhitungan invers document frequency (IDF) dan mengkalikan nilai TF dan IDF (Wijaya, 2021).

Penggunaan pembobotan TF-IDF pada analisis sentiment pengguna twitter terhadap polemik pesepakbolaan Indonesia dengan metode KNN berhasil menghasilkan akurasi sebesar 79,99% dan error rate 20,01%. Setiap kata akan diberikan bobot atau nilai dan hasilnya akan terdapat nilai yang lebih kecil jika suatu kata sering muncul pada semua dokumen dan sebaliknya nilai akan lebih besar jika suatu kata tidak sering muncul dalam semua dokumen (Septian dkk., 2019).

TF-IDF dikombinasikan dengan *Support Vector Machine* dalam penelitian dengan judul analisis sentiment *review* film. TF-IDF dianggap mempunyai proporsi kepentingan yang sesuai dengan menggunakan jumlah kemunculan dalam teks dan dokumen. Data yang telah melalui *preprocessing* dilakukan pembobotan TF-IDF dan dioptimalkan oleh algoritma *support vector machine*. Hasil kombinasi TF-IDF dan Algoritma SVM mendapatkan akurasi 85% (Gifari dkk., 2022).

Tabel 2.2 Penelitian Terkait Mengenai TF-IDF

No.	Pengarang (tahun)	Judul	Metode	Hasil
1	Septian dkk., (2019)	Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF	<i>K-Nearest Neighbor &amp; TF-IDF</i>	Penggunaan TF-IDF dapat menghasilkan bobot yang sesuai karena didapatkan dari kemunculan kata dalam dokumen, hasilnya dikombinasikan dengan

		dan K-Nearest Neighbor		metode K-Nearest Neighbor mendapatkan akurasi sebesar 79,99% dan error rate 20,01%.
2	Gifari dkk., (2022)	Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine	<i>support vector machine &amp; TF-IDF</i>	Kombinasi TF-IDF dengan SVM memberikan hasil yang baik pada klasifikasi data <i>review</i> film dengan menghasilkan nilai akurasi sebesar 85%.

### 2.3 Support Vector Machine

Support Vector Machine (SVM) binomial adalah salah satu algoritma machine learning yang digunakan untuk melakukan klasifikasi biner, yaitu klasifikasi yang hanya memiliki dua kelas yang mungkin (Nanda dkk., 2022). SVM binomial memanfaatkan konsep hyperplane untuk memisahkan kedua kelas dengan margin yang maksimal. Hyperplane dalam SVM binomial adalah suatu garis atau permukaan yang memisahkan dua kelas dalam ruang dimensi yang lebih tinggi (Noor, 2018). SVM binomial mencari hyperplane yang memisahkan kedua kelas dengan margin yang maksimal. SVM binomial menggunakan teknik kernel untuk memetakan data ke dalam ruang dimensi yang lebih tinggi, sehingga memungkinkan pembentukan hyperplane yang lebih baik. Kernel yang umum digunakan dalam SVM binomial adalah kernel linier dan kernel non-linier seperti kernel RBF (Radial Basis Function). Dalam SVM binomial, setiap data dilabeli dengan nilai +1 atau -1, dan kemudian dilatih pada model SVM. Data yang memiliki label +1 akan ditempatkan di satu sisi hyperplane, sedangkan data dengan label -1 akan ditempatkan di sisi lainnya. Data yang berada pada hyperplane disebut dengan support vector, dan merupakan titik penting dalam pembentukan model SVM.

Untuk masalah klasifikasi teks, Support Vector Machine adalah teknik klasifikasi yang sangat cepat dan efisien. yang dalam hal geometri, dapat dilihat sebagai fitur space hyperplane membagi poin yang mewakili kasus positif dari kategori yang mewakili keadaan negatif dalam klasifikasi biner (Rahman Isnain dkk., 2021). Klasifikasi SVM memiliki keunggulan yang signifikan dalam pendekatan teoretisnya yang memperbaiki masalah *overfitting* yang memungkinkannya berfungsi dengan baik, oleh karena itu pengklasifikasi ini dipilih selama pelatihan sebagai hyperplane tunggal yang membagi kasus positif yang diketahui dari kasus negatif. (Yerzi & Sibaroni, 2021).

Himawan & Eliyani (2021) pada penelitiannya membuktikan bahwa Support Vector Machine adalah metode terbaik dalam melakukan pengklasifikasikan data tweet masyarakat terhadap pemerintah DKI Jakarta. hasil tersebut didapatkan dalam penelitiannya yang membandingkan 3 algoritma klasifikasi yang merupakan supervised learning, yaitu random forest classifier, naive bayes classifier, dan support vector machine. Pengujian data dilakukan dengan *cross validation* dan hasil menunjukkan akurasi pada masing-masing model klasifikasi yaitu. *Linear SVM* sebesar 77,58%, *Multinomial naïve bayes* sebesar 75,22%, dan *random forest classifier* sebesar 75,81%%. Hal itu menunjukkan bahwa metode SVM memiliki keunggulan dalam akurasi dibandingkan dengan metode lainnya seperti *Naïve bayes* dan *Random forest*.

Nyoto & Ruldeviyani (2022) dalam penelitiannya juga melakukan perbandingan menggunakan 3 metode pada klasifikasi sentimen masyarakat terhadap program sumur resapan di DKI Jakarta untuk mengetahui akurasi dan

performa dari masing-masing metode yaitu Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor. Data yang telah melalui *preprocessing* selanjutnya dilakukan pengujian menggunakan *cross validation* yang akan mencari tahu performa terbaik dari suatu model algoritma dengan percobaan sebanyak  $k$  kali. Hasil pengujian menyatakan SVM adalah metode terbaik dengan tingkat akurasi sebesar 93,32% dengan  $k = 8$ , diikuti oleh Naïve Bayes dengan akurasi sebesar 90,52% dengan  $k = 8$ , dan KNN sebesar 80,63% dengan  $k = 7$ . Pada saat yang sama SVM mendapatkan nilai presisi 91,16%, *recall* 95,94, dan *f-measure* 93,49%. Naïve Bayes mencapai presisi 86,23%, *recall* 96,45%, dan *f-measure* 91,05%. Dan KNN mencapai preisi 76,70%, *recall* 87,9%, dan *f-measure* 81,95%.

Hal yang sama juga didapatkan pada penelitian Firmansyah dkk., (2021) yang melakukan analisis sentimen pada ulasan aplikasi PeduliLindungi dengan menggunakan 2 algoritma klasifikasi yaitu Naïve Bayes dan Support Vector Machine. Data yang diperoleh dari ulasan di Google Play Store diklasifikasikan ke dalam 2 kelas, positif dan negatif. Hasilnya metode Support Vector Machine menjadi yang terbaik dibandingkan metode Naïve Bayes dengan nilai akurasi sebesar 84,33% dan presisi 87,38%, sedangkan Naïve Bayes mendapatkan nilai akurasi sebesar 81,16% dan presisi 86,15%.

Penelitian sebelumnya membuktikan bahwa metode Support Vector Machine adalah metode yang dapat melakukan klasifikasi dengan sangat baik. Sebab itu dalam penelitian ini akan digunakan algoritma Support Vector Machine binomial untuk menyelesaikan masalah klasifikasi sentimen masyarakat terhadap program

Pemerintah Indonesia pada aplikasi twitter menggunakan metode support vector machine.

Tabel 2.3 Penelitian Terkait Mengenai *Support Vector Machine*

No.	Pengarang (tahun)	Judul	Metode	Hasil
1	Himawan & Eliyani (2021)	Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi	<i>Support Vector Machine, Naïve Bayes, dan Random Forest Classifier</i>	Menghasilkan bahwa SVM adalah metode terbaik untuk klasifikasi pada penelitian. dengan pengujian yang sama linear SVM memperoleh akurasi 77,58%, sedangkan multinomial naive bayes sebesar 75,22% dan random forest sebesar 75,81%.
2	Nyoto & Ruldeviyani (2022)	Infiltration Wells Program in Jakarta: Twitter Sentiment Analysis Rebecca	<i>Support Vector Machine, Naïve Bayes, dan K-Nearest Neighbor</i>	Metode SVM memiliki akurasi terbaik dibandingkan 2 metode lainnya yaitu naive bayes dan k-nearest neighbor. Dengan metode SVM memperoleh nilai akurasi sebesar 93,32%, sedangkan naive bayes sebesar 90,52% dan KNN sebesar 80,23%.
3	Firmansyah dkk., (2021)	A Comparison of Support Vector Machine and Naïve Bayes Classifier in Binary Sentiment Reviews for PeduliLindungi Application Isal	<i>Support Vector Machine &amp; Naïve Bayes</i>	Metode SVM mendapatkan nilai akurasi lebih tinggi dibandingkan dengan naive bayes, dengan nilai akurasi SVM sebesar 84,33%, sedangkan naive bayes sebesar 81,16%.



## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Pengumpulan Data**

Data yang digunakan pada penelitian ini adalah data primer dan sekunder. Untuk data primer didapatkan dari *tweet* pada aplikasi Twitter. Pengambilan data dilakukan dengan cara *scraping* terhadap data *tweet* yang berkaitan dengan program Pemerintah Indonesia dibidang Pendidikan. Untuk pengambilan data akan diambil 1.000 data ulasan program pemerintah dibidang Pendidikan seperti kampus merdeka, kurikulum merdeka, program Indonesia pintar, dan platform merdeka mengajar yang diambil pada tanggal 27 Februari 2023. Data sentimen yang dipilih hanya data sentimen yang berupa teks sehingga sentimen yang hanya menampilkan foto saja tidak digunakan. Selain itu sentimen harus bernilai positif atau negatif. Karena proses klasifikasi akan mengklasifikasikan sentimen menjadi 2 kelas yaitu positif dan negatif. Pelabelan data sentimen tersebut dilakukan oleh ahli bahasa dan ahli di bidang Pendidikan, Prof. Triyo Supriyatno, Ph.D. Kemudian data di simpan dalam format *XLSX (Extensible Stylesheet Language XML)*. Tabel 3.1 menunjukkan contoh data sentimen masyarakat yang digunakan sebagai sampel data dalam penelitian ini.

Selain itu, untuk data sekunder yang digunakan dalam penelitian ini adalah berdasarkan dari studi literatur terhadap penelitian-penelitian yang berhubungan dengan topik pada penelitian ini, yaitu mengenai klasifikasi sentimen masyarakat terhadap program pemerintah pada aplikasi twitter menggunakan metode *support*

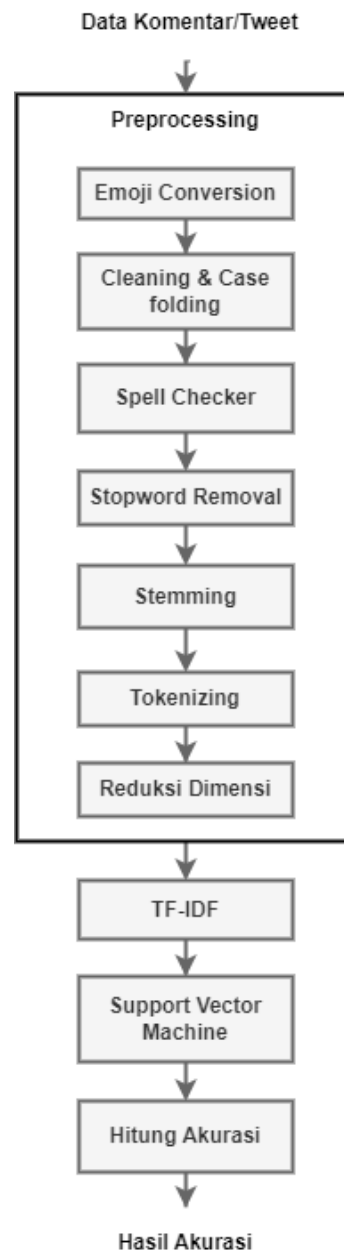
*vector machine*. Data ini digunakan sebagai pembanding atau pendukung dalam melakukan penelitian ini.

Tabel 3.1 Pelabelan Data

Dokumen	Komentar/Tweet	Kelas
D1	@xyvwaffa @sbmptnfess Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah. Kayak magang bersertifikat gitu gitu. Tapi tergantung kaka nya juga si	Positif
D2	@whosdisbi @caratstalk Mending cari cari magang sendri di luar kampus, atau ikut program magang dari kampus merdeka	Positif
D3	Lucu syekali 🤔🤔 lolos program kampus MERDEKA tpi rasa romusha wkwwkw 🤔	Negatif
D4	dan cukup ini kali terakhir aku ikut program kampus merdeka	Negatif
D5	Problematika kurikulum merdeka, guru belum siap dan murid pun kewalahan project. 🤔	Negatif

### 3.2 Desain Sistem

Tahapan sistem yang akan dibangun dan dijadikan acuan dalam penelitian ini digambarkan pada Gambar 3.1 Untuk tercapainya penelitian, peneliti membangun sistem menggunakan bahasa pemrograman python. Data sentiment masyarakat yang telah didapatkan akan diproses melalui beberapa langkah yaitu *preprocessing*, TF-IDF, *Support Vector Machine* sehingga hasilnya sentimen tersebut akan terklasifikasi berdasarkan kelas dalam label.

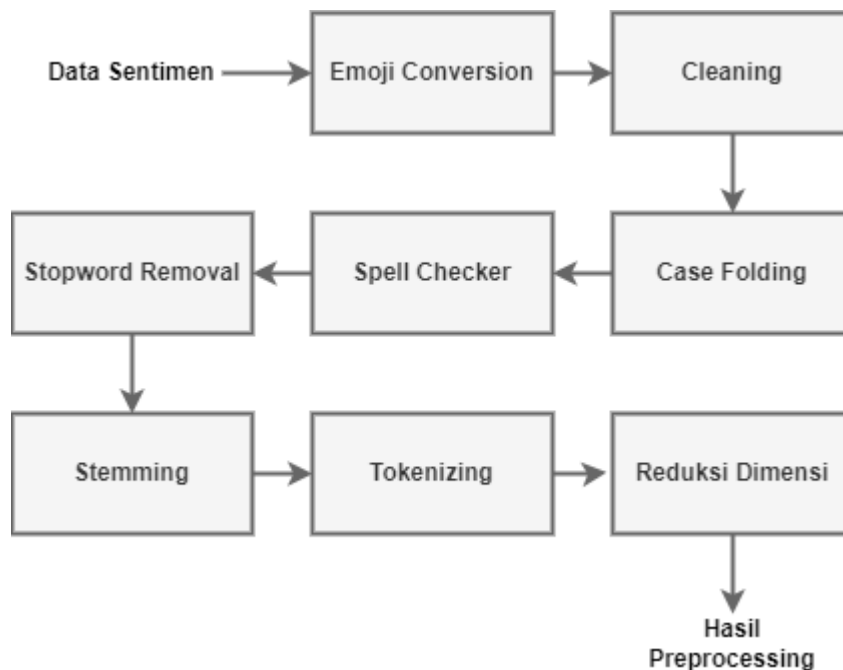


Gambar 3.1 Desain Sistem

### 3.2.1. *Preprocessing*

Data yang telah berhasil dikumpulkan perlu melalui tahap preprocessing agar nantinya dapat siap digunakan di tahap selanjutnya. Kebutuhan studi yang

dilakukan dapat disesuaikan pada setiap tingkat preprocessing. Dalam penelitian ini, proses yang dilakukan meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Gambaran dari alur *preprocessing* yang akan dilakukan dalam penelitian ini dapat dijelaskan pada Gambar 3.2.



Gambar 3.2 Proses *Preprocessing*

### 3.2.1.1 *Emoji Conversion*

Emoji conversion adalah suatu proses yang bertujuan untuk mengubah teks yang mengandung simbol emoji menjadi bentuk teks yang dapat dipahami oleh mesin atau sistem (Wolny, 2016). Tujuan utama dari emoji conversion adalah untuk menyederhanakan analisis teks yang memanfaatkan penggunaan emoji. Dengan melakukan konversi ini, informasi yang terkandung dalam teks emoji dapat lebih mudah diinterpretasikan dan diproses oleh algoritma atau mesin pembelajaran. Contoh proses emoji conversion dapat ditampilkan pada Tabel 3.2.

Tabel 3.2 Contoh *Emoji Conversion*

Sebelum <i>Emoji Conversion</i>	@xyvwaffa @sbmptnfess Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah. Kayak magang bersertifikat gitu. Tapi tergantung kaka nya juga si 😊
Setelah <i>Emoji Conversion</i>	@xyvwaffa @sbmptnfess Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah. Kayak magang bersertifikat gitu. Tapi tergantung kaka nya juga si :grinning face

### 3.2.1.2 *Cleaning*

Tahap yang pertama dalam *preprocessing* yang dilakukan adalah *cleaning*. Proses *cleaning* akan menghapus karakter selain huruf yang ada pada data inputan.

*Flowchart* proses *cleaning* ditampilkan pada Gambar 3.3 dibawah.

Gambar 3.3 *Flowchart Cleaning*

Pada proses *Cleaning*, karakter berupa tanda baca, tautan, emoji, dll. yang tidak diperlukan akan dihilangkan sehingga setelah dilakukan proses *cleaning*, data hanya berupa teks. Contoh proses *Cleaning* ditampilkan pada Tabel 3.3.

Tabel 3.3 Contoh *Cleaning*

Sebelum <i>Cleaning</i>	@xyvwaffa @sbmptnfess Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah. Kayak magang bersertifikat gtu gtu. Tapi tergantung kaka nya juga si :grinning face
Setelah <i>Cleaning</i>	Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah Kayak magang bersertifikat gtu gtu Tapi tergantung kaka nya juga si grinning face

### 3.2.1.3 Case Folding

Pada tahap *case folding*, data inputan yang hanya terdiri dari sebuah teks perlu diseragamkan untuk menghindari *case sensitive* dimana huruf kapital dan huruf kecil akan menimbulkan perbedaan makna. *Flowchart* proses *case folding* ditampilkan pada Gambar 3.4.

Gambar 3.4 *Flowchart Case Folding*

Tahap *case folding* berfungsi untuk menyeragamkan data, seluruh teks yang menggunakan huruf kapital seperti pada awalan kata, capslock, penyebutan nama tempat dll. yang terdapat huruf kapitalnya akan diubah menjadi huruf kecil (*lowercase*). Contoh proses *case folding* ditampilkan pada Tabel 3.4.

Tabel 3.4 Contoh *Case Folding*

Sebelum <i>Case Folding</i>	Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah Kayak magang bersertifikat gtu gtu Tapi tergantung kaka nya juga si grinning face
Setelah <i>Case Folding</i>	soalnya lebih enak kampus merdeka kak bisa ikut program program dari pemerintah kayak magang bersertifikat gtu gtu tapi tergantung kaka nya juga si grinning face

### 3.2.1.4 *Spell Checker*

Spell checker merupakan suatu mekanisme yang digunakan untuk memverifikasi keakuratan ejaan dalam sebuah teks (Yanfi dkk., 2022). Penerapan spell checker menjadi sangat relevan dalam konteks media sosial seperti Twitter, di mana sentimen yang disampaikan cenderung bersifat informal. Karakter informal ini seringkali disertai dengan ejaan kata yang tidak standar dan bervariasi. Maka dari itu, penggunaan spell checker digunakan untuk mengoreksi ejaan yang salah. Tujuan dari penggunaan spell checker ini adalah untuk mengurangi jumlah fitur atau kata kunci yang memiliki arti serupa, namun mungkin diekspresikan dengan ejaan yang berbeda. Contoh proses *spell checker* ditampilkan pada Tabel 3.5.

Tabel 3.5 Contoh *Spell Checker*

Sebelum <i>Spell Checker</i>	soalnya lebih enak kampus merdeka kak bisa ikut program program dari pemerintah kayak magang bersertifikat gtu gtu tapi tergantung kaka nya juga si grinning face
Setelah <i>Spell Checker</i>	soalnya lebih enak kampus merdeka kak bisa ikut program program dari pemerintah kayak magang bersertifikat gitu gitu tapi tergantung kaka nya juga si grinning face

### 3.2.1.5 *Tokenizing*

Pada tahap tokenizing, data sentimen yang masih berupa kalimat dipecah menjadi satuan kata atau token. *Flowchart* proses *tokenizing* ditampilkan pada Gambar 3.5.



Gambar 3.5 Flowchart Tokenizing

Pada tahap tokenizing, data dari proses case folding sebelumnya berupa teks, akan dipotong-potong menjadi sebuah token atau satuan kata. Contoh proses tokenizing ditampilkan pada Tabel 3.6.

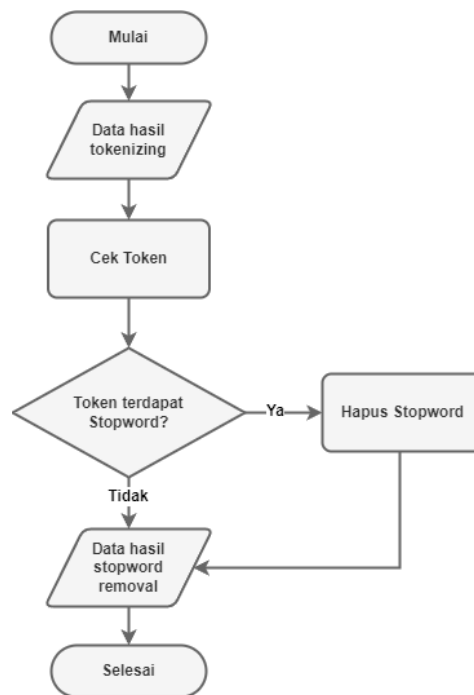
Tabel 3.6 Contoh *Tokenizing*

Sebelum <i>Tokenizing</i>	soalnya lebih enak kampus merdeka kak bisa ikut program program dari pemerintah kayak magang bersertifikat gitu gitu tapi tergantung kaka nya juga si grinning face
Setelah <i>Tokenizing</i>	“soalnya”, “lebih”, “enak”, “kampus”, “merdeka”, “kak”, “bisa”, “ikut”, “program”, “program”, “dari”, “pemerintah”, “kayak”, “magang”, “bersertifikat”, “gitu”, “gitu”, “tapi”, “tergantung”, “kaka”, “nya”, “juga”, “si”, “grinning”, “face”

### 3.2.1.6 Stopword Removal

Pada proses stopword removal, kata-kata yang cenderung tidak penting dan tidak ada hubungannya dengan dokumen seperti konjungsi dan preposisi akan dihilangkan.





Gambar 3.6 Flowchart Stopword Removal

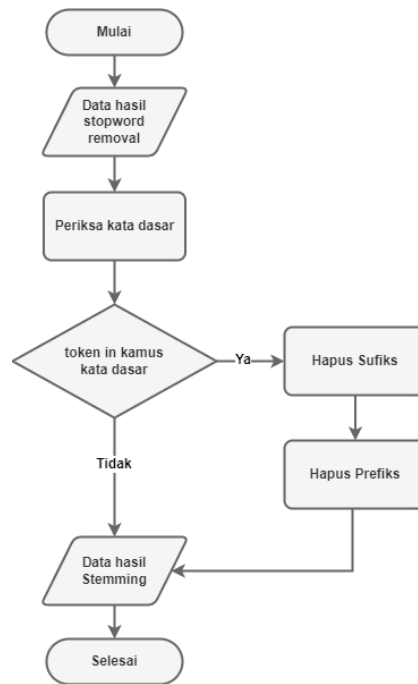
Pada tahap stopword removal, kata dari proses tokenizing akan dihilangkan jika kata tersebut termasuk dalam kata stopword. Contoh proses stopword removal ditampilkan pada Tabel 3.7.

Tabel 3.7 Contoh Stopword Removal

Sebelum Stopword Removal	“soalnya”, “lebih”, “enak”, “kampus”, “merdeka”, “kak”, “bisa”, “ikut”, “program”, “program”, “dari”, “pemerintah”, “kayak”, “magang”, “bersertifikat”, “gitu”, “gitu”, “tapi”, “tergantung”, “kaka”, “nya”, “juga”, “si”, “grinning”, “face”
Setelah Stopword Removal	“kampus”, “merdeka”, “kak”, “program”, “program”, “pemerintah”, “kayak”, “magang”, “bersertifikat”, “gitu”, “gitu”, “tergantung”, “kaka”, “si”, “grinning”, “face”

### 3.2.1.7 Stemming

Proses terakhir yaitu *stemming*, data sentimen yang telah menjadi kata akan diubah ke dalam bentuk kata dasarnya. Flowchart proses *stemming* ditampilkan pada Gambar 3.7.



Gambar 3.7 Flowchart Stemming

Pada tahap ini, jika kata yang mengandung sufiks dan prefiks, maka akan dihapus sufiks dan prefiksnya. Contoh proses stemming dapat ditampilkan pada Tabel 3.8.

Tabel 3.8 Contoh Stemming

Sebelum Stemming	“kampus”, “merdeka”, “kak”, “program”, “program”, “pemerintah”, “kayak”, “magang”, “bersertifikat”, “gitu”, “gitu”, “tergantung”, “kaka”, “si”, “grinning”, “face”
Setelah Stemming	“kampus”, “merdeka”, “kak”, “program”, “program”, “perintah”, “kayak”, “magang”, “sertifikat”, “gitu”, “gitu”, “gantung”, “kaka”, “si”, “grinning”, “face”

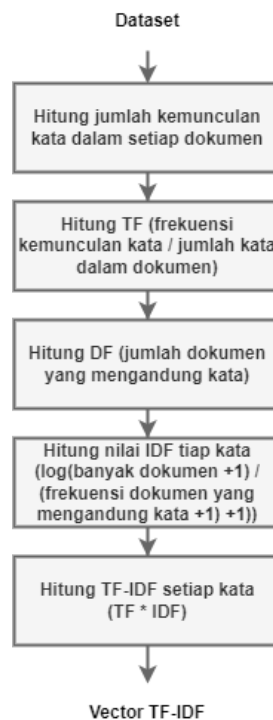
### 3.2.1.8 Reduksi Dimensi

Reduksi dimensi merupakan salah satu tahap dalam analisis data yang bertujuan untuk mengurangi jumlah fitur yang relevan atau signifikan dalam suatu dataset. Tujuannya adalah untuk menghilangkan variabel yang kurang informatif atau tidak memberikan kontribusi yang signifikan terhadap tujuan klasifikasi,

sehingga dapat meningkatkan efisiensi komputasi dan menghindari overfitting. Dengan melakukan reduksi dimensi, kita dapat memperoleh representasi data yang lebih fokus dan relevan.

### 3.2.2. TF-IDF

Peneliti memilih pembobotan kata dalam penelitian ini dengan menggunakan pendekatan TF-IDF. Himpunan kata tersebut akan menjadi term pada langkah selanjutnya setelah data berhasil menyelesaikan tahap preprocessing. Kata "*term frequency*" dalam *Term Frequency Inverse Document Frequency* (TF-IDF) mengacu pada seberapa sering suatu istilah muncul dalam sebuah teks. Algoritme yang dikenal sebagai IDF, atau *Inverse Document Frequency*, menentukan kemungkinan terbalik untuk menemukan sebuah kata dalam sebuah teks. Alur dari TF-IDF dapat ditunjukkan pada Gambar 3.8 berikut:



Gambar 3.8 Alur TF-IDF

Untuk mendapatkan nilai dari TF-IDF dapat dilakukan dengan melakukan perkalian antara nilai TF dan IDF seperti dalam Persamaan 3.1, 3.2 dan 3.3 (Khan dkk., 2020).

$$tf_{ij} = \frac{f_{ij}}{f_j} \quad (3.1)$$

Tabel 3.9 Perhitungan *Term Frequency*

Dokumen	Jumlah kata	Kata	Frekuensi	TF
D1	14	Program	2	2/14
D2	10		1	1/10
D3	9		1	1/9
D4	3		1	1/3
D5	7		0	0/7

Pada Persamaan 3.1  $f_{ij}$  adalah jumlah kata  $i$  dalam dokumen  $j$ , sedangkan  $f_j$  merupakan jumlah kata dalam dokumen  $j$ .

$$idf_i = \left( \log \left( \frac{N + 1}{df_i + 1} \right) + 1 \right) \quad (3.2)$$

Tabel 3.10 Perhitungan *Inverse Document Frequency*

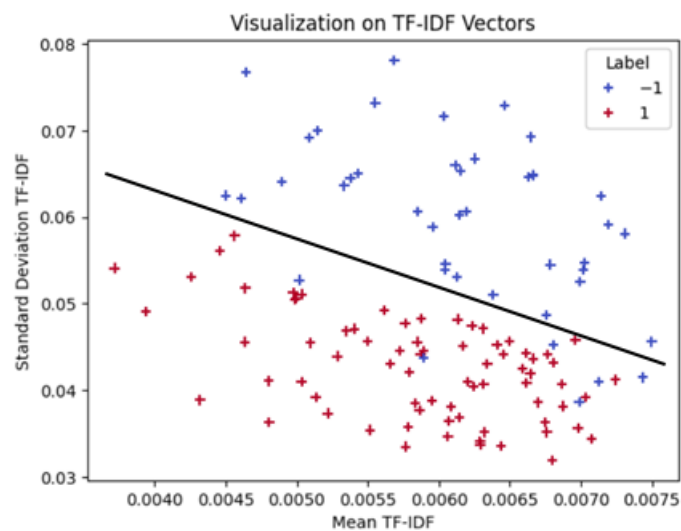
Dokumen	Kata	Frekuensi	IDF
D1	Program	2	$\left( \log \left( \frac{5 + 1}{4 + 1} \right) + 1 \right) = 1.079181$
D2		1	
D3		1	
D4		1	
D5		0	

Pada Persamaan 3.2  $N$  adalah jumlah banyaknya dokumen yang digunakan untuk data latih, sedangkan  $df_i$  merupakan jumlah dokumen yang mengandung kata  $i$ . Selanjutnya untuk menghitung nilai TF-IDF dihitung dengan Persamaan 3.3 berikut:

$$TFIDF_{ij} = tf_{ij} \times idf_i \quad (3.3)$$

### 3.2.3. Support Vector Machine

Dalam metode SVM (Support Vector Machine), sebaran data atau distribusi data sangat mempengaruhi kinerja dan hasil dari model. SVM merupakan metode klasifikasi yang mencoba menemukan batas keputusan yang optimal untuk memisahkan dua kelas data. Sebaran data atau distribusi data pada svm juga menentukan penggunaan model *hyperplane* SVM yang akan digunakan. Sebaran data pada penelitian ini dapat ditampilkan pada Gambar 3.9.



Gambar 3.9 Visualisasi data

Dari sebaran data pada Gambar 3.9 dapat ditetapkan model *hyperplane* yang digunakan pada penelitian ini adalah *linear separable*. Model *hyperplane* SVM *linear separable* digunakan dalam penelitian ini karena pada sebaran data positif dan negatif dapat dipisahkan secara linear, selain itu pada penelitian ini sistem hanya membagi komentar teks menjadi 2 kelas yaitu klasifikasi positif dan negatif.

Pada model *Support Vector Machine* biner, pelabelan data dilakukan dengan variable label  $y_i = 1$  dan  $y_i = -1$  yang menandakan kelas dokumen termasuk dalam

kelas positif atau negatif. Fungsi tujuan svm digunakan untuk meminimalkan kesalahan klasifikasi (Chen, 2009). Fungsi tujuan ditunjukkan pada Persamaan 3.4.

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^n \zeta_i \quad (3.4)$$

$$\text{Subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$

$$\zeta_i \geq 0, i = 1, \dots, n$$

Batasan pertama memastikan bahwa jarak antara titik data dan hyperplane tidak boleh lebih kecil daripada margin. Batasan kedua memastikan bahwa tidak ada titik data yang terlalu dekat dengan hyperplane. Data dipisahkan menggunakan *linear hyperplane* dengan menghitung nilai parameter dari setiap dokumen yang digunakan menghasilkan hasil klasifikasi untuk setiap dokumen dalam model linier, menggunakan Persamaan 3.5.

$$f(x) = \sum_{i \in SV} y_i \alpha_i K(x_i, x) + b \quad (3.5)$$

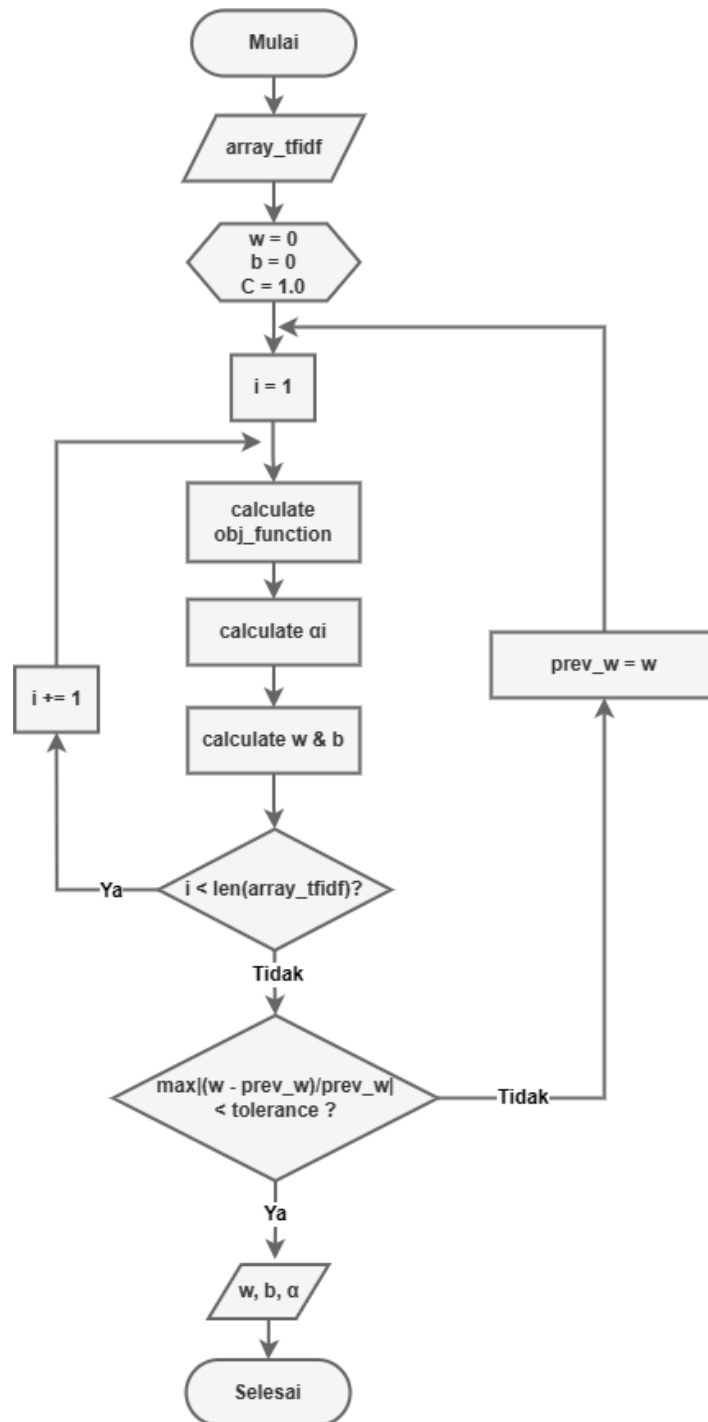
Nilai  $\alpha_i$  adalah nilai alfa ke-i, sedangkan  $x_i$  merupakan nilai dari parameter pada setiap fitur yang didapatkan pada tahap vektorisasi TF-IDF, nilai  $y_i$  adalah data nilai label data ke-i, dan  $b$  adalah bias. Dari Persamaan 3.4 menghasilkan kemungkinan sebagai berikut:

- a. Ketika  $f(x) < 0$  maka  $y_i = -1$
- b. Ketika  $f(x) > 0$  maka  $y_i = +1$

Selanjutnya dapat dilihat penurunan nilai bobot dalam setiap iterasinya yang dapat disebut konvergensi. Selanjutnya proses training akan berlanjut hingga telah dipenuhi kriteria yang ditetapkan, seperti pada Persamaan 3.6.

$$\max \left\{ \left| \frac{w^t - w^{t-1}}{w^t} \right| \right\} < tolerance \quad (3.6)$$

Batas toleransi juga dapat ditentukan oleh peneliti sesuai dengan kebutuhan. Dalam penelitian ini nilai batas toleransi ditetapkan sebesar 0,001. Implementasi sistem yang telah dirancang menggunakan data *training* dan *testing* merupakan tahap selanjutnya. *Input* data *testing* akan digunakan untuk menguji model klasifikasi SVM yang telah dibuat oleh sistem dengan menggunakan nilai parameter fitur dari data *training*. Berdasarkan nilai  $f(x)$ , sistem diimplementasikan untuk mengidentifikasi kelas data testing yang dihasilkan dari Persamaan 3.5. Hasilnya teks akan masuk ke kelas negatif jika nilai  $f(x) < 0$ . Sebaliknya, jika  $f(x) > 0$  teks tersebut masuk sebagai anggota kelas positif. Alur proses pelatihan dapat ditampilkan pada Gambar 3.10.



Gambar 3.10 Flowchart proses pelatihan

Tahap selanjutnya adalah tahap implementasi sistem yang telah dibuat, yang mana hasil pengujian sistem berupa penentuan kelas berdasarkan nilai yang



dihasilkan dari Persamaan 3.5. Data testing yang ditentukan sesuai dengan skenario uji coba akan diprediksi dengan perbandingan dari hasil pembelajaran sistem terhadap pola penentuan kelas.

## **BAB IV**

### **UJI COBA DAN PEMBAHASAN**

Pada bab ini berisi tentang hasil dari pengujian dari desain sistem yang telah dirancang sebelumnya, meliputi preprocessing, TF-IDF, proses klasifikasi menggunakan metode support vector machine, dan evaluasi menggunakan confusion matrix untuk melihat hasil akurasi, presisi, recall, dan f-measure dari hasil percobaan. Serta pembahasan dari hasil uji coba dan integrasinya dengan nilai-nilai islam.

#### **4.1 Skenario Uji Coba**

Uji coba dilakukan pada sistem yang telah dibangun dengan mengimplementasikan data training untuk membuat model klasifikasi dengan metode Support Vector Machine, sedangkan data testing digunakan untuk menguji kinerja sistem dari model yang telah dibuat tersebut. Pengujian dilakukan beberapa kali untuk memperoleh akurasi terbaik pada proses pengklasifikasian sentiment dan menemukan aspek apa saja yang dapat mempengaruhi performa sistem. Pengujian pertama dilakukan dengan pembagian rasio data. Pengujian ini dilakukan untuk menentukan rasio terbaik dari berbagai pembagian data yang telah ditentukan yaitu sebesar 6:4, 7:3, 8:2 dan 9:1 pada data training dan testing. Dengan adanya pembagian rasio pembagian data akan dilihat rasio mana yang memiliki akurasi tertinggi. Selanjutnya pengujian kedua dilanjutkan dengan menggunakan *spell checker* dan *emoji conversion* pada data yang akan diproses untuk klasifikasi. Penggunaan *spell checker* dan *emoji conversion* bertujuan untuk melihat pengaruh

penggunaan proses tersebut pada tahap *preprocessing* terhadap performa sistem dalam proses klasifikasi.

Pengujian ketiga dilakukan dengan pembagian rasio sentiment positif dan sentiment negatif dalam data testing menggunakan rasio 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1. Pengujian ini dilakukan untuk melihat apakah perbandingan rasio jumlah sentiment positif dan negatif dalam data testing memiliki pengaruh terhadap hasil akurasi sistem dalam proses klasifikasi. Dalam pengujian ini rasio yang digunakan dalam pembagian data training dan testing adalah sebesar 8:2.

Pengujian keempat dilakukan pengujian menggunakan metode *k-fold cross validation*. Metode tersebut digunakan untuk validasi model pada data yang membagi data menjadi sebanyak k subset yang berisi data sama besar, dimana salah satu subset akan digunakan sebagai data *testing* dan subset lainnya akan digunakan sebagai *training*. Proses akan terus dilakukan sebanyak k yang digunakan, dan hasil evaluasi pada setiap iterasi akan diambil rata-ratanya untuk mendapat nilai akurasi rata-rata dari model data yang digunakan. Pada penelitian ini digunakan k sebanyak 10 sehingga data akan dibagi dan dilakukan perulangan sebanyak 10. Alur dari 10-fold cross validation dapat divisualisasikan dalam Gambar 4.1 berikut.

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10
iterasi-1	testing	training	training	training	training	training	training	training	training	training
iterasi-2	training	testing	training	training	training	training	training	training	training	training
iterasi-3	training	training	testing	training	training	training	training	training	training	training
iterasi-4	training	training	training	testing	training	training	training	training	training	training
iterasi-5	training	training	training	training	testing	training	training	training	training	training
iterasi-6	training	training	training	training	training	testing	training	training	training	training
iterasi-7	training	training	training	training	training	training	testing	training	training	training
iterasi-8	training	training	training	training	training	training	training	testing	training	training
iterasi-9	training	training	training	training	training	training	training	training	testing	training
iterasi-10	training	training	training	training	training	training	training	training	training	testing

Gambar 4.1 Visualisasi 10-fold cross validation

Pengujian kelima dilakukan pengujian menggunakan fitur reduksi dimensi pada dataset yang digunakan. Pada tahap ini akan dilakukan penghapusan fitur atau set kata yang dirasa tidak memiliki kontribusi yang signifikan. Pada pengujian ini dilakukan reduksi dimensi terhadap set kata yang hanya muncul satu kali, dua kali, dan tiga kali dalam total dokumen. Pengujian ini dilakukan untuk melihat apakah penggunaan fitur reduksi dimensi memiliki pengaruh terhadap akurasi dalam proses klasifikasi.

Evaluasi dalam penelitian ini dilakukan menggunakan model *Confusion Matrix* yang membandingkan kelas data yang sebenarnya dengan data kelas hasil prediksi. Hasilnya akan didapatkan nilai performa dari sistem dengan parameter

akurasi, presisi, *recall*, dan *f-measure*. Tabel dari confusion matrix dapat ditunjukkan pada Tabel 4.1 berikut.

Tabel 4.1 *Confussion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	TP ( <i>True Positive</i> )	FN ( <i>False Negative</i> )
Negatif	FP ( <i>False Positive</i> )	TN ( <i>True Negative</i> )

Perhitungan nilai parameter seperti akurasi, presisi, *recall*, dan *f-measure* dilakukan menggunakan Persamaan berikut.

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (4.1)$$

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4.3)$$

$$F - Measure = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (4.4)$$

Dengan keterangan sebagai berikut,

TP: banyaknya sentimen positif yang diprediksi dengan benar ke dalam kelas positif.

FP: banyaknya sentimen negatif yang salah prediksi ke dalam kelas positif.

FN: banyaknya sentimen positif yang salah prediksi ke dalam kelas negatif.

TN: banyaknya sentimen negatif yang diprediksi dengan benar ke dalam kelas negatif.

## 4.2 Hasil Uji Coba

Dalam penelitian ini, data yang digunakan berasal dari sumber data sosial media Twitter dengan melakukan proses pengumpulan data melalui teknik data *scraping*. Data ini berkaitan dengan topik-topik seperti "kampus merdeka," "kurikulum merdeka," "program Indonesia pintar," dan "platform merdeka," yang menghasilkan sejumlah total 1000 data. Seluruh data tersebut kemudian diberi label sebagai data positif atau negatif. Dalam hal ini, terdapat 536 data yang diberi label positif dan 464 data yang diberi label negatif. Dengan mengacu pada jumlah data dalam setiap kelas, dapat disimpulkan bahwa komentar positif menyumbang sekitar 53,6% dari total data, sementara komentar negatif sekitar 46,4%. Sebagai ilustrasi lebih lanjut, beberapa sampel data dari penelitian ditampilkan pada Tabel 4.2.

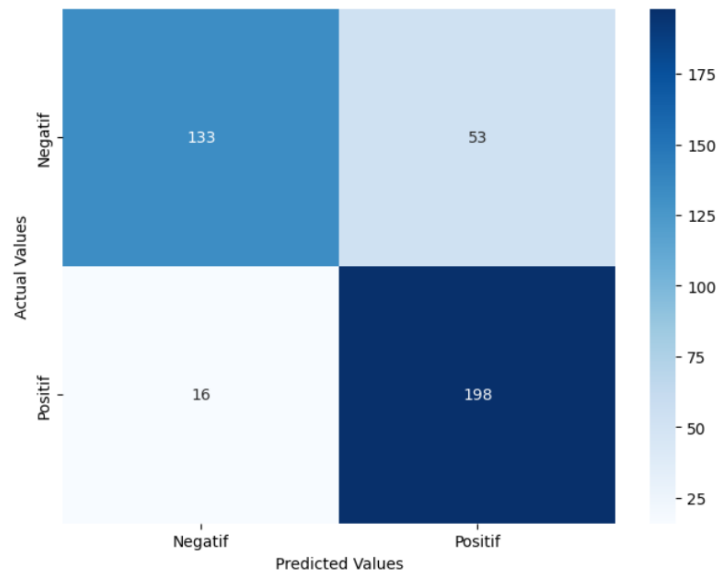
Tabel 4.2 Sampel Data Penelitian

No.	Sentimen	Label
1.	Sekarang udh ad program kampus merdeka, sks bisa diganti ama pengalaman magang yg lebih worth it dri pd bem..	1
2.	Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah. Kayak magang bersertifikat gitu gitu. Tapi tergantung kaka nya juga si	1
3.	Plus lebih pada milih ngejar program iisma mbkm dan program kampus merdeka lainnya yang experiencenya lebih relevan sama dunia setelah perkuliahan	1
4.	Sekarang ada program kampus merdeka, dan itu lebih worth it sih	1
5.	Lucu syekali 🤔🤔 lolos program kampus MERDEKA tpi rasa romusha wkwkwk 🤔	-1
6.	Kenapa si tiap ak ikut program kampus merdeka ada aja problem yg gede(π~~π)	-1
7.	Kayaknya kita bisa sama2 sepakat program kampus merdeka MSIB Batch 4 ini yang paling buruk diantara batch2 sebelumnya. Mulai dari SR SPTJM, timeline, kebijakan yang berubah2, sampe simulasi survey yang gabisa di akses. Speechless parah	-1
8.	Ikut program kampus merdeka, lumayan dapat uang jajan sama konversi sks ntar	1
9.	ambil aja, ikut program kampus merdeka tuh seru tau gw aja smpe ketagihan wkwkw	1
10.	kadang kepengen magang gituu tapi terhalang sama program kampus merdeka 🤔	-1
11.	pengen sesekali dapet kesempatan ikut program kampus merdeka	1
12.	Program Kampus Merdeka dan Permendikbud PPKS menurut gue adalah salah dua output beliau yang paling mantep selama menjabat sih.	1

13.	Percuma ada program kampus merdeka, daerah gue kagak ada perusahaan yg ikut programnya. Jadi kalau mau magang harus keluar provinsi dulu, yg ada bokek.	-1
14.	Tapi gue belajar metopen kualifikasi di semester 5 sih, berkat program pertukaran kampus merdeka wkwkwk	1
15.	Program Kampus Merdeka sukses nih harus dilanjutkan programnya dengan Program Fresh Graduate Merdeka 😊	1
16.	jujur kadang sedih kebijakan kampus merdeka yg bolehin cuma ikut 2 program aja. soalnya aku belum coba yang magang 🙄	-1
17.	Apadaya kampus gue tidak mendukung program dari kampus merdeka katanya kurikulum angkatan gue beda sama mereka 🙄🙄🙄🙄🙄🙄🙄	-1
18.	Apa kabar yg pada ikut program kampus merdeka dapet B wkwk 20sks B sangat menurunkan IPK	-1
19.	Mbkm gini banget ya tuhann, katanya program kampus merdeka tapi kok aku merasa terjajah ☹️	-1
20.	sekali ikut program Kampus Merdeka bakalan ketagihan 😊	1
21.	Coba join program magang merdeka/program kampus merdeka lainnya nder. Lumayan itu bisa konversi sks jg	1
22.	program kampus merdeka ini masih harus banyak revisinya :(	-1
23.	Program kampus merdeka ini semakin hari makin keren2	1
24.	Program Kampus Merdeka ngebuat gue banyak belajar. Dari setiap program Kemendikbudristek ngebuat mahasiswa bisa eksplora diri bareng"	1
25.	program kampus merdeka ini lebih banyak program yg bisa dibilang gagal daripada yg berjalan sesuai dg kesepakatan diawal 😊	-1

#### 4.2.1 Hasil Uji Coba-1

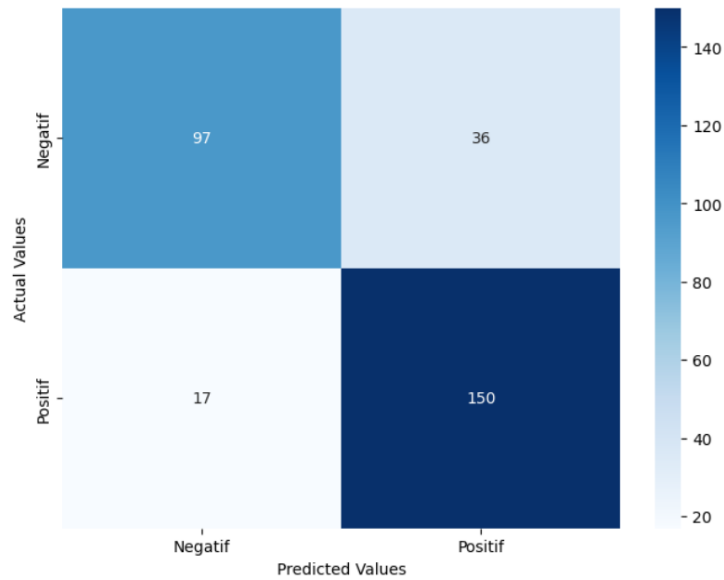
Pada Uji Coba pertama dilakukan pengujian pada variasi rasio data, dengan tujuan menemukan perbandingan yang optimal. Rasio yang digunakan adalah 6:4, 7:3, 8:2, dan 9:1 untuk pembagian antara data pelatihan dan data pengujian. Dalam skenario pembagian data 6:4, mendapatkan hasil pembagian data yang terdiri dari 600 data pelatihan dan data pengujian terdiri dari 400 data, dengan 214 di antaranya adalah sentimen positif dan 186 lainnya adalah sentimen negatif. Hasil dari uji coba rasio 6:4 ditunjukkan pada tabel *confusion matrix* pada Gambar 4.2.



Gambar 4.2 *Confusion Matrix* Rasio 6:4

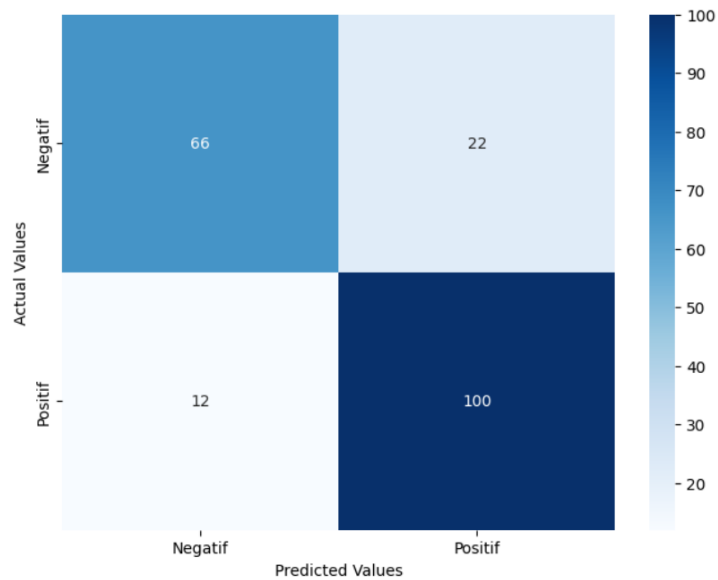
Berdasarkan hasil uji coba menggunakan perbandingan 6:4 mendapatkan tingkat akurasi sebesar 82,75%, presisi sebesar 78,8%, *recall* sebesar 92,5%, dan *f-measure* sebesar 85%. Selanjutnya Dalam skenario pembagian data 7:3, mendapatkan hasil pembagian data yang terdiri dari 700 data pelatihan dan 300 data pengujian. Dalam data pelatihan, terdapat 369 data sentimen positif dan 331 data sentimen negatif. Sementara dalam data pengujian, terdapat 167 data sentimen positif dan 133 data sentimen negatif. Hasil dari uji coba rasio 7:3 ditunjukkan pada tabel *confusion matrix* pada Gambar 4.3.





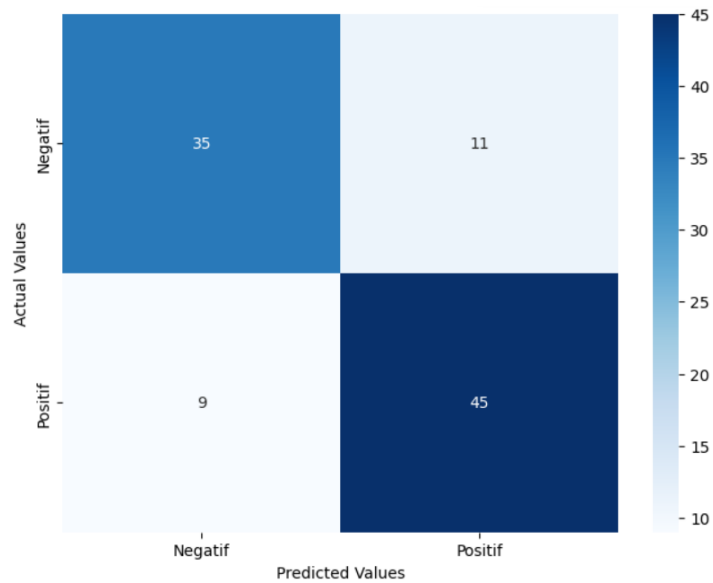
Gambar 4.3 *Confusion Matrix* Rasio 7:3

Berdasarkan hasil uji coba menggunakan perbandingan 7:3 didapatkan tingkat akurasi sebesar 82,3%, presisi sebesar 80,6%, *recall* sebesar 89,8%, dan *f-measure* sebesar 84,9%. Selanjutnya Dalam skenario pembagian data 8:2, mendapatkan hasil pembagian data yang terdiri dari 800 data pelatihan dan 200 data pengujian. Dalam data pelatihan, terdapat 424 data sentimen positif dan 376 data sentimen negatif. Sementara dalam data pengujian, terdapat 112 data sentimen positif dan 88 data sentimen negatif. Hasil dari uji coba rasio 8:2 ditunjukkan pada tabel *confusion matrix* pada Gambar 4.4.



Gambar 4.4 *Confusion Matrix* Rasio 8:2

Berdasarkan hasil uji coba menggunakan perbandingan 8:2 didapatkan tingkat akurasi sebesar 83%, presisi sebesar 81,9%, *recall* sebesar 89,2%, *f-measure* sebesar 85,4%. Selanjutnya Dalam skenario pembagian data 9:1, mendapatkan hasil pembagian data yang terdiri dari 900 data pelatihan dan 100 data pengujian. Dalam data pelatihan, terdapat 482 data sentimen positif dan 418 data sentimen negatif. Sementara dalam data pengujian, terdapat 54 data sentimen positif dan 46 data sentimen negatif. Hasil dari uji coba rasio 9:1 ditunjukkan pada tabel *confusion matrix* pada Gambar 4.5.



Gambar 4.5 *Confusion Matrix* Rasio 9:1

Berdasarkan hasil uji coba menggunakan perbandingan 9:1 didapatkan tingkat akurasi sebesar 80%, presisi sebesar 80,3%, recall sebesar 83,3%, dan f-measure sebesar 81,8%. Dari hasil pengujian dengan variasi rasio data didapatkan hasil akurasi yang berbeda. Perbedaan hasil uji coba pertama dengan melakukan variasi rasio data dapat ditunjukkan pada Tabel 4.3.

Tabel 4.3 Hasil Akurasi Uji Coba 1

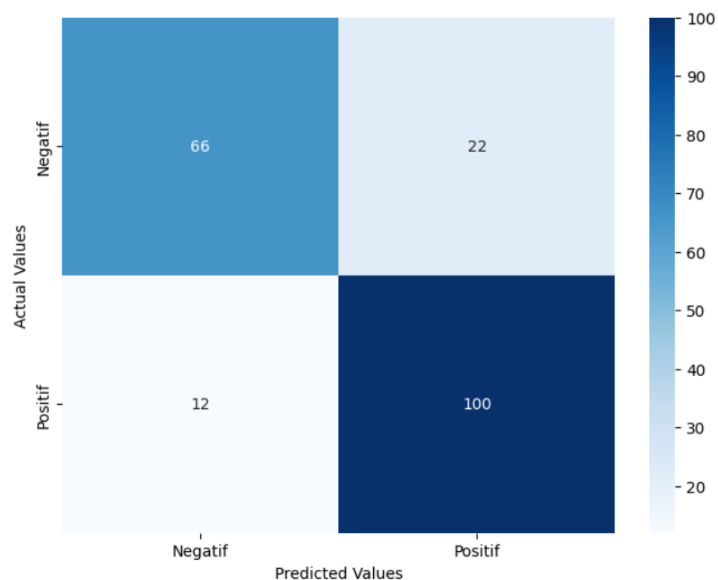
Rasio Pembagian data	Hasil Akurasi
6:4	82,75%
7:3	82,3%
8:2	83%
9:1	80%

Dari Tabel 4.3 dapat disimpulkan bahwa uji coba dengan menggunakan variasi rasio yang berbeda dapat mempengaruhi hasil akurasi. Selain itu hasil akurasi tertinggi didapatkan pada rasio 8:2 dengan tingkat akurasi 79%. Hal ini membuktikan bahwa proses klasifikasi sentimen dapat optimal pada pembagian rasio 80% untuk data *training* dan 20% untuk data *testing*. Dari hasil yang

diperoleh, maka pada pengujian selanjutnya akan menggunakan rasio 8:2 pada pembagian data pelatihan dan pengujian, yang mana akan terdapat 800 data pelatihan dan 200 data pengujian.

#### 4.2.2 Hasil Uji Coba-2

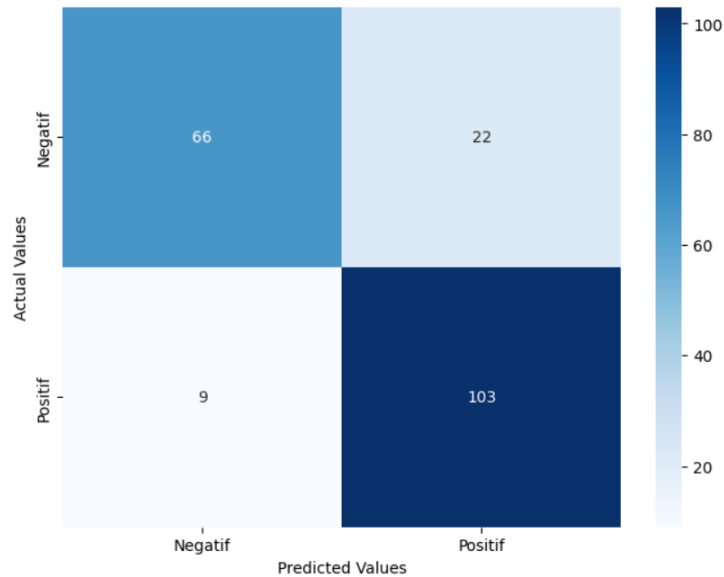
Pada Uji Coba Kedua dilakukan pengujian menggunakan rasio pembagian data 8:2, pada percobaan ini dilakukan untuk membuktikan bagaimana pengaruh penggunaan fitur *spell checker* dan *emoji conversion* pada tahap *preprocessing* terhadap performa sistem yang telah dibangun dan tingkat akurasi pada proses klasifikasi. Pada Gambar 4.6 Menunjukkan visualisasi dari hasil uji coba tanpa menggunakan *spell checker* dan *emoji conversion*.



Gambar 4.6 *Confusion Matrix* Uji Coba tanpa *Spell checker* & *emoji conversion*

Berdasarkan Gambar 4.6 diperoleh hasil perhitungan confusion matrix untuk pengujian tanpa menggunakan spell checker dan emoji conversion pada tahap preprocessing dengan nilai Akurasi sebesar 83%, Presisi 81,9%, Recall 89,2%, dan

F-Measure 85,4%. Selanjutnya pada Gambar 4.7 Menunjukkan visualisasi dari hasil uji coba dengan menggunakan *spell checker* dan *emoji conversion*.



Gambar 4.7 Confusion Matrix Uji Coba *Spell checker* & *emoji conversion*

Berdasarkan Gambar 4.7 diperoleh hasil perhitungan confusion matrix untuk pengujian dengan menggunakan *spell checker* dan *emoji conversion* pada tahap *preprocessing* dengan nilai Akurasi sebesar 84,5%, Presisi 82,4%, Recall 91,9%, dan F-Measure 86,9%. Sehingga perbedaan hasil antara uji coba penggunaan *spell checker* dan *emoji conversion* dengan yang tidak menggunakan *spell checker* dan *emoji conversion* dapat ditampilkan pada Tabel 4.4.

Tabel 4.4 Hasil Uji Coba 2

Skema Percobaan	Hasil Akurasi
Tanpa <i>spell checker</i> & <i>emoji conversion</i>	83%
<i>Spell checker</i> & <i>emoji conversion</i>	84,5%

Berdasarkan hasil uji coba 2 dengan menggunakan *spell checker* dan *emoji conversion* untuk tambahan pada tahap *preprocessing* didapatkan hasil akurasi sebesar 84,5%. Dibandingkan dengan tanpa menggunakan *spell checker* dan *emoji conversion* nilai

akurasi yang didapatkan tidak sebesar seperti saat menggunakan fitur tersebut yakni dengan perbedaan 1,5% pada nilai akurasi. Hal tersebut berarti membuktikan bahwa penggunaan *spell checker* dan *emoji conversion* memiliki pengaruh terhadap kinerja sistem dalam proses klasifikasi.

### 4.2.3 Hasil Uji Coba-3

Pengujian dilakukan pada rasio perbandingan antara data positif dan negatif pada data testing dengan jumlah data testing sebanyak 200 dikarenakan rasio data training dan testing yang digunakan adalah 8:2 sesuai dengan rasio terbaik pada uji coba sebelumnya. Hasil pengujian dapat ditampilkan pada Tabel 4.5.

Tabel 4.5 Hasil Uji Coba 3

<b>Rasio Data Test</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F-Measure</b>
1 : 9	91%	96%	93%	95%
2 : 8	88%	92%	92%	92%
3 : 7	86%	87%	92%	90%
4 : 6	84%	82%	93%	87%
5 : 5	83%	78%	92%	84%
6 : 4	81%	71%	91%	80%
7 : 3	80%	62%	88%	73%
8 : 2	77%	46%	90%	61%
9 : 1	76%	31%	100%	48%

Berdasarkan hasil pengujian pada rasio perbandingan antara data negatif dan positif pada data testing didapatkan bahwa nilai akurasi tertinggi dicapai pada rasio negatif dan positif 1:9 yang memperoleh nilai akurasi sebesar 91% dan memperoleh nilai presisi, recall, dan f-measure sebesar 96%, 93%, dan 95%. Sedangkan nilai akurasi terendah dicapai pada rasio negatif dan positif 9:1 yang memperoleh nilai akurasi sebesar 76%. Hasil pengujian ini mengindikasikan bahwa sistem lebih akurat dalam klasifikasi sentimen positif dari pada negatif. Hal tersebut juga dibuktikan dengan rata-rata presisi yang lebih rendah daripada rata-rata recall,

yang berarti bahwa hasil pengujian memiliki kesalahan prediksi positif (FN) yang lebih sedikit dibandingkan kesalahan prediksi negatif (FP).

#### 4.2.4 Hasil Uji Coba-4

Selanjutnya dilakukan pengujian dengan validasi silang menggunakan Teknik k-fold cross validation dengan nilai k yaitu 10. Oleh karena itu, akan dilakukan proses iterasi sebanyak 10 kali untuk membagi data pelatihan dan pengujian. Dalam setiap fold dalam pengujian ini memiliki perbandingan data positif dan negatif yang sama. Hasil pengujian Teknik 10-fold cross validation dapat ditampilkan pada Tabel 4.6.

Tabel 4.6 Hasil Uji 10-Fold Cross Validation

Fold ke-	Akurasi	Presisi	Recall	F-Measure
1	81%	76%	94%	84%
2	84%	81%	90%	85%
3	79%	73%	94%	82%
4	81%	76%	94%	84%
5	84%	80%	92%	86%
6	82%	77%	94%	85%
7	80%	76%	90%	82%
8	84%	80%	92%	85%
9	73%	68%	92%	78%
10	72%	69%	83%	75%
<b>Rata-rata</b>	<b>79,9%</b>	<b>76%</b>	<b>91,9%</b>	<b>83,1%</b>

Berdasarkan hasil pengujian dengan menggunakan metode 10-fold cross-validation, ditemukan bahwa nilai akurasi rata-rata yang diperoleh adalah sebesar 79,9%. Hasil akurasi tertinggi didapatkan sebesar 84% pada fold kedua, kelima, dan kedelapan. Selain akurasi, beberapa metrik lainnya juga memberikan hasil yang signifikan pada fold ini. Hasil pengujian ini mengindikasikan bahwa penerapan validasi silang dengan menggunakan metode 10-fold cross-validation memiliki dampak positif pada kinerja sistem klasifikasi sentimen yang sedang diuji. Hasil

yang sangat baik pada fold kedua, kelima, dan kedelapan menunjukkan bahwa model memiliki kemampuan yang tinggi dalam mengklasifikasikan sentimen dengan benar, khususnya dalam hal akurasi, presisi, recall, dan F-measure. Selain itu percobaan ini membuktikan bahwa pemilihan data training dan data testing dapat mempengaruhi akurasi sistem dalam proses klasifikasi.

#### 4.2.5 Hasil Uji Coba-5

Pengujian dilakukan dengan mereduksi dimensi fitur atau set kata yang muncul dalam jumlah yang ditentukan dalam total seluruh dokumen. Pengujian dilakukan dengan tiga tahapan yaitu reduksi set kata yang hanya muncul kurang dari atau sama dengan satu kali, dua kali, tiga kali, empat kali, dan lima kali dalam total seluruh dokumen dalam dataset. Selain itu dalam setiap fold pengujian ini menggunakan perbandingan data positif dan negatif yang seimbang seperti pengujian sebelumnya. Hasil uji coba dapat ditampilkan pada Tabel 4.7. 4.8. 4.9. 4.10. dan 4.11.

Tabel 4.7 Hasil Uji Reduksi 1

<b>Fold ke-</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F-Measure</b>
1	80%	75%	94%	83%
2	86%	83%	92%	87%
3	79%	74%	92%	82%
4	84%	79%	94%	86%
5	85%	81%	92%	86%
6	83%	78%	94%	85%
7	83%	79%	92%	85%
8	84%	81%	90%	85%
9	74%	69%	92%	79%
10	75%	71%	86%	78%
<b>Rata-rata</b>	<b>81,3%</b>	<b>77,4%</b>	<b>92,3%</b>	<b>84,1%</b>

Hasil pengujian dengan reduksi set kata yang muncul kurang atau sama dengan 1 kali dalam total seluruh dokumen mendapatkan rata-rata akurasi sebesar



81,3%. Didapatkan hasil berupa peningkatan dari percobaan sebelumnya tanpa reduksi yang memperoleh rata-rata akurasi sebesar 79,9%.

Tabel 4.8 Hasil Uji Reduksi 2

<b>Fold ke-</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F-Measure</b>
1	80%	75%	94%	83%
2	85%	81%	92%	86%
3	79%	74%	92%	82%
4	84%	79%	94%	86%
5	86%	83%	92%	87%
6	84%	79%	94%	86%
7	84%	80%	92%	85%
8	87%	84%	92%	88%
9	76%	71%	92%	80%
10	72%	69%	84%	76%
<b>Rata-rata</b>	<b>81,7%</b>	<b>77,9%</b>	<b>92,3%</b>	<b>84,4%</b>

Hasil pengujian dengan reduksi set kata yang muncul kurang atau sama dengan 2 kali dalam total seluruh dokumen mendapatkan rata-rata akurasi sebesar 81,7%. Didapatkan hasil berupa peningkatan dari percobaan sebelumnya dengan reduksi set kata yang muncul kurang dari atau sama dengan 1 kali yang memperoleh rata-rata akurasi sebesar 81,3%.

Tabel 4.9 Hasil Uji Reduksi 3

<b>Fold ke-</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F-Measure</b>
1	81%	76%	94%	84%
2	86%	83%	92%	87%
3	79%	74%	92%	82%
4	84%	80%	92%	86%
5	84%	81%	90%	85%
6	84%	79%	94%	86%
7	83%	80%	90%	84%
8	88%	85%	92%	89%
9	76%	71%	92%	80%
10	76%	73%	86%	79%
<b>Rata-rata</b>	<b>82,1%</b>	<b>78,6%</b>	<b>91,9%</b>	<b>84,6%</b>

Hasil pengujian dengan reduksi set kata yang muncul kurang atau sama dengan 3 kali dalam total seluruh dokumen mendapatkan rata-rata akurasi sebesar 82,1%. Didapatkan hasil berupa peningkatan dari percobaan sebelumnya dengan

reduksi set kata yang muncul kurang dari atau sama dengan 2 kali yang memperoleh rata-rata akurasi sebesar 81,7%.

Tabel 4.10 Hasil Uji Reduksi 4

<b>Fold ke-</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F-Measure</b>
1	81%	76%	94%	84%
2	84%	82%	91%	86%
3	80%	75%	92%	83%
4	84%	80%	92%	86%
5	84%	82%	91%	86%
6	84%	80%	94%	86%
7	83%	80%	90%	85%
8	87%	85%	90%	88%
9	76%	72%	90%	80%
10	77%	73%	88%	80%
<b>Rata-rata</b>	<b>82%</b>	<b>78,6%</b>	<b>91,5%</b>	<b>84,5%</b>

Hasil pengujian dengan reduksi set kata yang muncul kurang atau sama dengan 4 kali dalam total seluruh dokumen mendapatkan rata-rata akurasi sebesar 82%. Didapatkan hasil berupa penurunan dari percobaan sebelumnya dengan reduksi set kata yang muncul kurang dari atau sama dengan 3 kali yang memperoleh rata-rata akurasi sebesar 82,1%.

Tabel 4.11 Hasil Uji Reduksi 5

<b>Fold ke-</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F-Measure</b>
1	82%	77%	94%	85%
2	85%	83%	91%	87%
3	79%	75%	91%	82%
4	84%	81%	92%	86%
5	84%	83%	89%	86%
6	82%	78%	92%	85%
7	82%	80%	88%	84%
8	87%	86%	90%	88%
9	75%	71%	88%	79%
10	78%	74%	88%	81%
<b>Rata-rata</b>	<b>81,8%</b>	<b>78,8%</b>	<b>90,6%</b>	<b>84,2%</b>

Hasil pengujian dengan reduksi set kata yang muncul kurang atau sama dengan 5 kali dalam total seluruh dokumen mendapatkan rata-rata akurasi sebesar 81,8%. Didapatkan hasil berupa penurunan dari percobaan sebelumnya dengan

reduksi set kata yang muncul kurang dari atau sama dengan 4 kali yang memperoleh rata-rata akurasi sebesar 82%. Hasil seluruh pengujian fitur reduksi dimensi dapat ditampilkan pada Tabel 4.12.

Tabel 4.12 Hasil Uji Reduksi Dimensi

Skema Percobaan	Hasil Akurasi
Reduksi Dimensi 1	81,3%
Reduksi Dimensi 2	81,7%
Reduksi Dimensi 3	82,1%
Reduksi Dimensi 4	82%
Reduksi Dimensi 5	81,8%

Berdasarkan hasil pengujian dengan reduksi dimensi, ditemukan bahwa penggunaan fitur reduksi dimensi memiliki pengaruh terhadap sistem dalam proses klasifikasi, dibuktikan dengan meningkatnya nilai akurasi dalam pengujian fitur reduksi dimensi ini. Hasil kfold cross validation tanpa menggunakan reduksi dimensi didapatkan nilai akurasi rata-rata sebesar 79,9% dibandingkan dengan saat menggunakan reduksi dimensi dengan menghapus set kata yang muncul kurang atau sama dengan 1 kali didapatkan peningkatan akurasi rata-rata sebesar 81,3%. Pada pengujian dengan menghapus set kata yang muncul kurang atau sama dengan 2 kali dan 3 kali juga mengalami peningkatan akurasinya yakni menjadi 81,7% untuk reduksi set kata yang muncul 2 kali dan 82,1% untuk reduksi set kata yang muncul 3 kali. Sedangkan pada pengujian dengan menghapus set kata yang kurang atau sama dengan 4 kali dan 5 kali mengalami penurunan akurasinya yakni sebesar 82% dan 81,8%. Oleh karena itu, dapat disimpulkan bahwa reduksi set kata yang muncul kurang atau sama dengan 3 kali dari total seluruh dokumen adalah yang terbaik pada pengujian ini. Hasil pengujian ini mengindikasikan bahwa penerapan fitur reduksi dimensi memiliki dampak positif terhadap pada kinerja sistem dalam

proses klasifikasi dibuktikan dengan peningkatan rata-rata akurasi dibandingkan dengan tanpa menggunakan fitur, namun penghapusan fitur tersebut juga harus disesuaikan dengan kondisi data karena jika terlalu banyak fitur penting yang direduksi juga dapat menurunkan performa sistem dalam klasifikasi.

### 4.3 Pembahasan

Dari hasil uji coba yang telah dilakukan terhadap 1000 data sentiment masyarakat yang diperoleh dari proses scraping data pada media sosial twitter, dapat diketahui bahwa data tersebut sebelum dapat di proses oleh sistem harus dilakukan pengolahan terlebih dahulu pada tahap *preprocessing* dengan melalui tahapan seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Proses tersebut sangat membantu untuk sistem dapat mengolah data dengan baik, oleh karena itu tahap *preprocessing* adalah salah satu faktor yang menentukan performa sistem berjalan dengan baik atau tidak. Setelah tahap *preprocessing* selanjutnya tahap pembobotan adalah juga salah satu tahap yang sangat mempengaruhi sistem klasifikasi sentimen. Dalam penelitian ini pembobotan menggunakan TF-IDF yang akan merubah setiap kata/token menjadi sebuah nilai/angka sehingga dapat diproses oleh sistem.

Dari hasil uji coba 1 didapatkan hasil performa yang berbeda dalam evaluasi modelnya. Sebagaimana yang telah ditampilkan pada Tabel 4.3 hasil perbandingan dari beberapa skenario pengujian dengan membandingkan beberapa rasio pembagian data didapatkan hasil pengujian terbaik memperoleh akurasi sebesar 83% pada rasio pembagian data 8:2. Maka dapat dikatakan bahwa pada uji coba pertama pembagian data 800 untuk data *training* dan 200 untuk data *testing*

memiliki performa yang lebih baik. Oleh karena itu dapat disimpulkan bahwa penggunaan rasio pembagian data adalah aspek yang mempengaruhi performa dalam proses klasifikasi sentimen masyarakat terhadap program Pemerintah Indonesia pada aplikasi Twitter menggunakan metode *Support Vector Machine*.

Selanjutnya pada uji coba 2 menunjukkan hasil uji coba dengan melakukan perbedaan fitur pada tahap *preprocessing* yang membandingkan penggunaan spell checker dan emoji conversion dengan tanpa menggunakannya. Seperti yang diketahui sebelumnya pada tahap preprocessing terdapat beberapa proses seperti cleaning, case folding, tokenizing, stopword removal, dan juga stemming. Pengujian kedua dilakukan untuk melihat apakah penambahan fitur spell checker dan emoji conversion dapat membantu meningkatkan performa sistem, atau penambahan tersebut tidak memiliki pengaruh terhadap sistem. Dari hasil pengujian yang telah dijelaskan pada Tabel 4.4 dapat dilihat bahwa penggunaan spell checker dan emoji conversion dapat meningkatkan performa sistem dalam melakukan klasifikasi. Dibuktikan dengan peningkatan tingkat akurasi sebesar 1,5% dari sebelumnya yang tidak menggunakan spell checker dan emoji conversion.

Selanjutnya pada pengujian rasio perbandingan data negatif dan positif pada data testing dilakukan untuk melihat pengaruh sistem dalam proses pengklasifikasian menggunakan rasio data negatif dan positif yang berbeda pada data testing. Proses ini menggunakan perbandingan data training dan testing 8:2 yang berarti terdapat 200 data testing. Hasil uji coba ini didapatkan bahwa pencapaian tertinggi didapatkan pada rasio negatif dan positif sebesar 1:9 yakni

sebesar 91%. Hasil sebaliknya didapatkan pada rasio negatif dan positif sebesar 9:1 yang memperoleh hasil terendah dengan nilai akurasi sebesar 76%. Hal tersebut membuktikan bahwa terdapat pengaruh yang signifikan terhadap akurasi yang dipengaruhi oleh penggunaan perbandingan data negatif dan positif dalam data testing. Oleh karena itu dapat disimpulkan bahwa perbandingan data negatif dan positif pada data testing merupakan aspek yang mempengaruhi performa dalam proses klasifikasi sentimen.

Pada tahap validasi silang menggunakan metode k-fold cross-validation diterapkan untuk mengevaluasi kinerja model. Proses k-fold cross-validation dalam penelitian ini berguna untuk memastikan konsistensi performa model terhadap variasi data yang berbeda. Proses k-fold cross-validation akan memanfaatkan seluruh dataset secara bergantian sebagai data uji dan data pelatihan. Hal ini bertujuan untuk mencapai representasi yang lebih akurat dari sebaran data dan memastikan bahwa model tidak hanya mampu mengidentifikasi pola data spesifik, melainkan juga dapat menghasilkan performa yang maksimal dalam proses klasifikasi. Hasil uji coba dengan menggunakan 10-fold cross-validation dapat ditemukan dalam Tabel 4.6, yang menunjukkan bahwa pencapaian tertinggi dalam hal akurasi terjadi dengan nilai akurasi mencapai 84%. Dan terdapat perbedaan nilai akurasi yang cukup signifikan pada setiap fold dalam percobaan ini. Hasil ini membuktikan bahwa pembagian dan pemilihan data untuk penggunaan data dalam *training* dan *testing* yang dalam penelitian ini menggunakan k-fold cross-validation memiliki pengaruh signifikan terhadap kinerja sistem dalam proses klasifikasi sentimen masyarakat.

Selanjutnya pengujian kelima menunjukkan hasil uji coba dengan melakukan perbedaan fitur pada tahap *preprocessing* yang membandingkan penggunaan fitur reduksi dimensi dengan tanpa menggunakannya. Pengujian dilakukan dengan 10-fold cross-validation dan hasilnya menunjukkan bahwa sebelum penggunaan fitur rata-rata akurasi adalah sebesar 79,9% dan setelah menerapkan fitur reduksi dimensi terhadap set kata yang muncul hanya satu kali dalam total dokumen dalam dataset, hasilnya didapatkan peningkatan nilai rata-rata akurasi yakni sebesar 81,3%. Selanjutnya pengujian dilakukan dengan reduksi dimensi terhadap set kata yang muncul kurang atau sama dengan dua kali dan tiga kali, didapatkan bahwa terdapat peningkatan nilai rata-rata akurasi yakni menjadi 81,7% setelah melalui reduksi dimensi terhadap set kata yang muncul kurang dari atau sama dengan dua kali dan meningkat kembali menjadi 82,1% setelah melalui reduksi dimensi terhadap set kata yang muncul kurang atau sama dengan tiga kali. Namun, pada percobaan reduksi set kata yang muncul kurang dari atau sama dengan empat kali dan lima kali terdapat penurunan nilai akurasi yakni sebesar 82% dan 81,8%. Hasil ini membuktikan bahwa penggunaan fitur reduksi dimensi juga berpengaruh terhadap proses klasifikasi dan dapat disimpulkan bahwa tahap *preprocessing* data adalah aspek yang mempengaruhi performa dalam proses klasifikasi sentimen masyarakat terhadap program Pemerintah Indonesia pada aplikasi Twitter menggunakan metode *Support Vector Machine*.

Sistem ini dirancang dengan tujuan untuk memberikan kemampuan klasifikasi sentimen masyarakat terhadap program pemerintah Indonesia secara efektif dan akurat. Dalam konteks manfaat penelitian, keberadaan sistem ini

diharapkan dapat mengurangi risiko kesalahan dalam pemahaman dan interpretasi informasi yang sering kali mengakibatkan kesalahan dalam merancang strategi kebijakan masa depan. Sistem ini juga diharapkan dapat menjadi alat bantu dalam upaya pencarian kebenaran data yang mendalam dan berdasarkan fakta. Dengan demikian, implementasi sistem ini dapat memberikan kontribusi positif dalam meningkatkan kualitas pengambilan keputusan dan perencanaan strategis di tingkat pemerintah. Sebagaimana Allah subhanahu wa ta'ala berfirman dalam surah Al – Baqarah ayat 42:

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

*“Jangan kalian mencampur kebenaran dengan kebatilan. Jangan juga kalian menyembunyikan kebenaran. Padahal kalian menyadarinya.”* (Surah Al – Baqarah ayat 42).

Dalam tafsir Ibnu Katsir, ayat ini mengajarkan umat Islam untuk tidak mencampurkan antara yang benar (haq) dengan yang salah (bathil). Ini adalah perintah agar orang-orang beriman untuk selalu berpegang pada kebenaran dan jangan berusaha untuk menutupinya atau menggantikannya dengan kebohongan atau kesalahan (Dr. Abdullah, 2004b). Oleh karena itu, hal ini selaras dengan penelitian ini yang diharapkan dapat membantu menghindari campur aduk antara respon (sentiment masyarakat) yang positif dan negatif, sehingga kebenaran data yang ditemukan dapat dipahami dan digunakan untuk perencanaan strategis yang lebih baik. Perencanaan yang strategis dari pemerintah nantinya akan membawakan kesejahteraan bagi masyarakatnya.

Dari hasil penelitian yang dilakukan diharapkan dapat menjadi sarana bagi kita sebagai umat muslim terutama pemerintah (pemimpin) dan masyarakat untuk



mengajak pada kebaikan dan mencegah kemungkaran. Terkait hal ini dijelaskan dalam surah Ali ‘Imran ayat 110 yang berbunyi:

كُنْتُمْ خَيْرَ أُمَّةٍ أُخْرِجَتْ لِلنَّاسِ تَأْمُرُونَ بِالْمَعْرُوفِ وَتَنْهَوْنَ عَنِ الْمُنْكَرِ وَتُؤْمِنُونَ بِاللَّهِ وَلَوْ ءَامَنَ أَهْلُ الْكِتَابِ لَكَانَ خَيْرًا لَهُمْ مِمَّنْهُمْ الْمُؤْمِنُونَ وَأَكْثَرُهُمُ الْفَاسِقُونَ

*“Kamu (umat Islam) adalah umat terbaik yang dilahirkan untuk manusia (selama) kamu menyuruh (berbuat) yang makruf, mencegah dari yang mungkar, dan beriman kepada Allah. Seandainya Ahlulkitab beriman, tentulah itu lebih baik bagi mereka. Di antara mereka ada yang beriman dan kebanyakan mereka adalah orang-orang fasik” (Q.S Ali ‘Imran: 110).*

Ayat ini adalah salah satu ayat yang menunjukkan kedudukan mulia dan tanggung jawab besar yang dimiliki oleh umat Islam. Dalam tafsirnya, Ibnu Katsir menjelaskan bahwa ayat ini menegaskan bahwa umat Islam adalah umat terbaik yang dipilih oleh Allah untuk menjadi panutan dan penyebar kebaikan bagi seluruh manusia (Dr. Abdullah, 2004c). Ayat ini juga menunjukkan bahwa umat Islam memiliki tanggung jawab moral dan sosial untuk memerintah kebaikan (amar ma'ruf) dan mencegah kemungkaran (nahi munkar) dalam masyarakat. Dalam hal ini pemerintah (pemimpin) dan masyarakat turut bertanggung jawab dalam mengajak kepada kebaikan dan mencegah keburukan dalam memberikan program yang baik sebagai pewujudan program pemerintah yang dapat mensejahterakan masyarakat. Sedangkan masyarakat dapat mengajak kebaikan dan mencegah keburukan dengan memberikan tanggapan dan masukan terhadap program yang telah diberikan pemerintah agar terwujudnya kesejahteraan seluruh masyarakat.

Hal ini pula dijelaskan dalam hadist dari Abu Hurairah r.a., Rasulullah SAW bersabda:

رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ يَقُولُ مَنْ رَأَى مِنْكُمْ مُنْكَرًا فَلْيَعْبِرْهُ بِيَدِهِ فَإِنْ لَمْ يَسْتَطِعْ فَلْيَسَانِهِ فَإِنْ لَمْ يَسْتَطِعْ فَيَقْلِبْهُ وَذَلِكَ أضعفُ الإِيمَانِ

“Barangsiapa di antara kamu yang melihat suatu kemungkaran, maka hendaklah ia mengubahnya dengan tangannya. Jika tidak mampu, maka hendaklah dengan lisannya. Jika tidak mampu, maka hendaklah dengan hatinya, dan yang demikian itu adalah selemah-lemah iman.” (Sahih Muslim).

Dalam hadis ini menjelaskan kepada umat Islam tentang tanggung jawab mereka untuk mencegah munkar (kemungkaran) dan mendukung kebaikan dengan cara yang sesuai dengan kemampuan masing-masing orang. Jika seseorang melihat tindakan yang salah, ia harus mencoba mengubahnya. Jika tidak mampu untuk melakukannya secara fisik, maka dengan lisan dan jika bahkan itu tidak mungkin, maka dengan hati yang merasa tidak setuju. Ini menunjukkan bahwa seseorang yang memiliki iman yang kuat harus selalu melakukan hal tersebut, dan hal ini adalah salah satu bagian penting dari tanggung jawab sosial dalam Islam.

Setelah dilakukannya penelitian klasifikasi sentiment masyarakat terhadap program pemerintah Indonesia pada aplikasi twitter menggunakan metode *Support Vector Machine* ini diharapkan dapat bermanfaat bagi beberapa pihak seperti pemerintah Indonesia yang khususnya pada bidang Pendidikan untuk dapat melakukan evaluasi terhadap program yang telah dicanangkan agar dapat diperbaiki demi terwujudnya program Pendidikan yang bermanfaat bagi masyarakat Indonesia, dan bagi akademisi yang melakukan penelitian terkait klasifikasi sentiment atau penelitian dengan metode *Support Vector Machine*.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dataset yang digunakan dalam penelitian ini didapatkan dengan cara scraping data pada aplikasi twitter yang menghasilkan 1000 data untuk diproses dengan jumlah data sentimen positif sebanyak 536 data dan data sentimen negatif sebanyak 464 data. Uji coba dilakukan sebanyak 5 kali dengan skenario yang berbeda agar melihat perbedaan serta pengaruhnya terhadap performa klasifikasi. Pengujian pertama dilakukan dengan perbandingan rasio data training dan data testing, didapatkan rasio 8:2 menjadi rasio terbaik dengan mendapatkan hasil akurasi sebesar 83%. Pengujian kedua dilakukan pengujian dengan penggunaan fitur *spell checker* dan *emoji conversion* pada tahap *preprocessing* dengan menggunakan rasio data training dan testing sebesar 8:2 didapatkan hasil akurasi sebesar 84,5%. Selanjutnya pengujian ketiga dilakukan pengujian terhadap rasio perbandingan data sentimen negatif dan positif dalam data testing dengan menggunakan rasio data training dan testing 8:2 yang berarti terdapat 200 data testing, didapatkan hasil akurasi tertinggi didapatkan pada perbandingan data negatif dan positif 1:9 dengan nilai akurasi sebesar 91% dan akurasi terendah didapatkan pada perbandingan data negatif dan positif 9:1 dengan nilai akurasi 76%. Pengujian keempat dilakukan pengujian menggunakan metode *k-fold cross validation* dengan perbandingan data positif dan negatif yang sama pada tiap fold nya, didapatkan hasil rata-rata akurasi sebesar 79,9%. Kemudian pengujian kelima dilakukan pengujian dengan menggunakan fitur reduksi dimensi pada tahap *preprocessing* dengan pengujian

menggunakan *k-fold cross validation*, hasil terbaik didapatkan pada percobaan dengan menghapus set kata yang hanya muncul kurang atau sama dengan tiga kali dalam total seluruh dokumen dataset didapatkan hasil akurasi rata-rata sebesar 82,1%. Oleh karena itu, dapat disimpulkan bahwa terdapat beberapa aspek yang mempengaruhi performa dalam proses klasifikasi seperti rasio perbandingan data training dan data testing, penggunaan fitur pada tahap *preprocessing*, rasio data positif dan negatif pada data testing, serta pemilihan data training dan data testing.

## 5.2 Saran

Berdasarkan hasil pengujian yang telah dilakukan, dalam penelitian ini terdapat beberapa hal yang dapat diperbaiki untuk mencapai hasil yang maksimal.

Berikut adalah saran untuk dilakukan pada penelitian selanjutnya:

1. Melakukan penambahan dataset yang akan digunakan dalam penelitian dan menggunakan data dari sosial media yang lain seperti Instagram, facebook, dan lain-lain agar data lebih bervariasi.
2. Menambahkan berbagai fungsi pada tahap *preprocessing* agar data yang akan diolah oleh sistem dapat lebih bersih. Karena hal tersebut sangat berpengaruh terhadap hasil performa sistem.
3. Menggunakan metode pembobotan kata yang berbeda dan menggunakan algoritma klasifikasi yang berbeda selain *support vector machine* untuk menjadi perbandingan dan mendapatkan hasil performa yang terbaik.

## DAFTAR PUSTAKA

- Abd, D. H., Sadiq, A. T., & Abbas, A. R. (2020). Classifying Political Arabic Articles Using Support Vector Machine with Different Feature Extraction. *Communications in Computer and Information Science, 1174 CCIS*, 79–94. [https://doi.org/10.1007/978-3-030-38752-5\\_7](https://doi.org/10.1007/978-3-030-38752-5_7)
- Al-Mahalli, J., & As-Suyuthi, J. (n.d.). *Tafsir Jalalain* (A. Medan (trans.); Jilid 1). Sinar Baru Algesindo.
- Astuti, F., Candra, R. M., & Agustian, S. (2022). *Klasifikasi Sentimen Masyarakat Terhadap Pemerintah Terkait Penerapan Kebijakan New Normal Menggunakan Metode K-Nearest Neighbor*. 5(3), 531–538.
- Bagus Sasmita, A., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(3), 1208–1214. <http://j-ptiik.ub.ac.id>
- Chen, H. F. (2009). In silico log p prediction for a large data set with support vector machines, radial basis neural networks and multiple linear regression. *Chemical Biology and Drug Design*, 74(2), 142–147. <https://doi.org/10.1111/j.1747-0285.2009.00840.x>
- Dr. Abdullah. (2004a). *Tafsir Ibnu Katsir* (M. A. G. E.M (trans.); Jilid 3).
- Dr. Abdullah. (2004b). *Tafsir Ibnu Katsir* (M. A. G. E.M (trans.); Jilid 1). Pustaka Imam Asy-Syafi'i.
- Dr. Abdullah. (2004c). *Tafsir Ibnu Katsir* (M. A. G. E.M (trans.); Jilid 2). Pustaka Imam Asy-Syafi'i.
- Faiq, M., Putro, A., & Setiawan, E. B. (2022). Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media Sosial Twitter. *E-Proceeding of Engineering*, 9(1), 54–66.
- Firmansyah, I., Asnawi, M. H., Hasanah, S. A., Novian, R., & Pravitasari, A. A. (2021). A Comparison of Support Vector Machine and Naïve Bayes Classifier in Binary Sentiment Reviews for PeduliLindungi Application. *2021 International Conference on Artificial Intelligence and Big Data Analytics, ICAIBDA 2021*, 18, 140–145. <https://doi.org/10.1109/ICAIBDA53487.2021.9689771>
- Ghoni Khoirul Abror et al. 2020. (2020). *Jurnal Ilmiah Administrasi Publik (JIAP) Implementasi Program Corporate Social Responsibility di Bidang Pendidikan*. 6(1), 71–81.
- Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of*

- Himawan, R. D., & Eliyani, E. (2021). Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 7(1), 58. <https://doi.org/10.26418/jp.v7i1.41728>
- Khan, A., Gul, M. A., Uddin, M. I., Ali Shah, S. A., Ahmad, S., Al Firdausi, M. D., & Zaindin, M. (2020). Summarizing Online Movie Reviews: A Machine Learning Approach to Big Data Analytics. *Scientific Programming*, 2020. <https://doi.org/10.1155/2020/5812715>
- Nanda, R., Haerani, E., Gusti, S. K., & Ramadhani, S. (2022). Klasifikasi Berita Menggunakan Metode Support Vector Machine. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 5(2), 269–278. <https://doi.org/10.32672/jnkti.v5i2.4193>
- Naz, S., Sharan, A., & Malik, N. (2019). Sentiment Classification on Twitter Data Using Support Vector Machine. *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018*, 676–679. <https://doi.org/10.1109/WI.2018.00-13>
- Noor, A. (2018). Perbandingan Algoritma Support Vector Machine Biasa dan Support Vector Machine berbasis Particle Swarm Optimization untuk Prediksi Gempa Bumi. *Jurnal Humaniora Teknologi*, 4(1), 31–37. <https://doi.org/10.34128/jht.v4i1.37>
- Nyoto, R. L. V., & Ruldeviyani, Y. (2022). Infiltration Wells Program in Jakarta: Twitter Sentiment Analysis. *2022 1st International Conference on Information System and Information Technology, ICISIT 2022*, 352–357. <https://doi.org/10.1109/ICISIT54091.2022.9872911>
- Prasetyo, H. D., Pramiyati, T., & Isnainiyah, I. N. (2021). Sentimen Analisis Pengguna Twitter Terhadap Kebijakan Merdeka Belajar Menggunakan Algoritma Naive Bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA), April*, 559–568.
- Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jdmsi*, 2(1), 31–37. <https://t.co/NfhnfMjtXw>
- Rosdiana, R., Eddy, T., Zawiyah, S., & Muhammad, N. Y. U. (2019). Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar. In *Proceeding SNTEI (Issue June 2020, pp. 87–93)*.
- Satrio, R. H., & Fauzi, M. A. (2019). *Klasifikasi Tweets Pada Twitter Menggunakan Metode K-Nearest Neighbour (K-NN) Dengan Pembobotan TF-IDF*. 3(8), 2548–2964. <http://j-ptiik.ub.ac.id>
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen

Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent System and Computation*, 1(1), 43–49. <https://doi.org/10.52985/insyst.v1i1.36>

Wijaya, R. H. (2021). *Perbandingan Feature Selection Chi-Square Dan Query Expansion Ranking (Qer) Pada Analisis Sentimen Terkait Revitalisasi Monas* .... 3(4), 401–410. <https://eprints.umm.ac.id/72008/>

Wolny, W. (2016). Emotion analysis of twitter data that use emoticons and emoji ideograms. *25th International Conference on Information Systems Development, ISD 2016, February*, 476–483.

Yanfi, Y., Gaol, F. L., Soewito, B., & Warnars, H. L. H. S. (2022). Spell Checker for the Indonesian Language: ExtensiveReview. *International Journal of Emerging Technology and Advanced Engineering*, 12(5), 1–7. [https://doi.org/10.46338/ijetae0522\\_01](https://doi.org/10.46338/ijetae0522_01)

Yerzi, F. S., & Sibaroni, Y. (2021). Analisis Sentimen Terhadap Kebijakan Pemerintah Dalam Menangani Covid-19 Dengan Pendekatan Lexicon Based. *E-Proceeding of Engineering*, 8(5), 11354–11366.

# **LAMPIRAN**



## LAMPIRAN

### Lampiran I

**Tabel Dataset dan Label**

No.	Teks	Label
1	@leejenouns @collegemenfess di univku semua program kampus merdeka bisa konversi 20 sks kak	P
2	@cheezemeanie @caratstalk Honestly, kalau kamu jadi anggota doang, bukan pengurus organisasi, gak terlalu mendongkrak track ” -	P
3	Program Merdeka Belajar Kampus Merdeka yang diselenggarakan sesuai dengan kebutuhan yang bertujuan untuk memberikan ” -	P
4	Pada akhirnya mahasiswa lebih realistis, memilih program kampus merdeka atau magang sejenisnya untuk investasi ” -	P
5	Sekarang udh ad program kampus merdeka, sks bisa diganti ama pengalaman magang yg lebih worth it dri pd bem..	P
6	@collegemenfess menurutku karena skrg banyak program yang lebih menjanjikan kaya program MSIB kampus merdeka. ” -	P
7	@Jeruksunkist304 @collegemenfess Iyapp, coba aja dulu. Pokoknya punya plan, dan yang terpenting usahain punya ” -	P
8	@xyvwaffa @sbmptnfess Soalnya lebih enak kampus merdeka kak bisa ikut program program dari Pemerintah. Kayak magang ” -	P
9	Ormawa in general ga menarik dibandingkan byk program pemerintah/kampus yg mungkin lebih "menguntungkan" ” -	P
10	@collegemenfess Lebih realistis, kayaknya km program magang atau kampus merdeka yang dipandang lebih berguna dan relevan ” -	P
11	@Lactococcuss @collegemenfess plus lebih pada milih ngejar program iisma mbkm dan program kampus merdeka lainnya yang experiencenya lebih relevan sama dunia setelah perkuliahan	P
12	@collegemenfess Sekarang ada program kampus merdeka, dan itu lebih worth it sih	P

13	@collegemenfess Kalo gue masih mahasiswa saat ada program kampus merdeka, gue keknya lebih memilih ikutan kampus ” -	P
14	@collegemenfess Karena ada Kampus Merdeka yang membuka peluang memudahkan internship sana-sini. Dengan adanya ” -	P
15	@collegemenfess Mungkin karena udah ada kampus merdeka, program langsung dari pemerintah plus dapet uang saku dan sertifikat juga. Jadi mereka lebih memilih ke program tersebut seperti acu hehe 🤔	P
16	@collegemenfess Kampus merdeka membebaskan mahasiswa terjun ke industri. Dr organisasi embel2 menunjang karir, ya ” -	P
17	@collegemenfess Karena sekarang ada program kampus merdeka, jadi mahasiswa lebih tertarik buat cari pengalaman di luar sehingga nggak tertarik sama organisasi kampus	P
18	@heiranyuri @raymndjo Aku juga bakal selalu rekomen kegiatan Kampus Merdeka. Karena aku udah 2x ikutan program itu dan ” -	P
19	Anyway, lebih bagus ikut program Kampus Merdeka dah lu.	P
20	Sumpah program kampus merdeka bener2 kegiatan full bgt ajg Teko jam 8 - 11 on boarding edufarmers ” -	N
21	@worksfess ada dari 2020an keknya. kebijakan ugm per 2023 semua mahasiswa harus ikut setidaknya 1 program kampus ” -	P
22	@verotrinov iyaaa udah lebih banyak yang mikirin prospek ke depannya dan pengalaman program kampus merdeka dianggap lebih menarik untuk companies dibandingkan di ormawa ☐	P
23	@worksfess seringnya kebutuhan company banyak yg ga keluar dikampus, yes memang harus banyak belajar sendiri. tp skrng ada program kampus merdeka ko, bisa jadi jalan buat ngembangin kompetensi sesuai kebutuhan perusahaan	P
24	@collegemenfess Nder kalo kamu ga tertarik masuk organisasi Coba ikut program kampus merdeka Lumayan sih untuk nambah relasi dan pengalaman syaratnya juga ga belibet	P
25	Lucu syekali 🤔🤔 lolos program kampus MERDEKA tpi rasa romusha wkwkwk 🤔 <a href="https://t.co/LJb6sf3Lmu">https://t.co/LJb6sf3Lmu</a>	N
...		
976	selamat #HariGuruNasional mari kita memberikan terobosan untuk #SerentakBerinovasi lewat platform Merdeka Mengajar, Puncak HGN 2022 <a href="https://t.co/Z35vmyz0QM">https://t.co/Z35vmyz0QM</a>	P

977	Selamat #HariGuruNasional terima kasih atas jasa jasa mu guru... Ayo #SerentakBerinovasi lewat platform Merdeka Mengajar, Puncak HGN 2022 <a href="https://t.co/PjGJKQicuu">https://t.co/PjGJKQicuu</a>	P
978	Selamat #HariGuruNasional. Banyak inovasi di dunia pendidikan. Semua dilakukan utk kemajuan bangsa. Banyak platform di ” -	P
979	selamat #HariGuruNasional sudah banyak upaya yang dilakukan pak Nadiem untuk ” -	P
980	Selamat #HariGuruNasional saatnya dunia pendidikan #SerentakBerinovasi melalui Platform Merdeka Mengajar (PMM) Puncak HGN 2022 <a href="https://t.co/YEC1hSjtK7">https://t.co/YEC1hSjtK7</a>	P
981	Platform Merdeka Mengajar dibangun untuk menunjang penerapan Kurikulum Merdeka agar dapat membantu guru dalam mendapatkan referensi, inspirasi, dan pemahaman dalam menerapkan Kurikulum Merdeka. <a href="https://t.co/aOrFZo7w8S">https://t.co/aOrFZo7w8S</a>	P
982	mabok platform merdeka mengajar	N
983	@kelin_ciucul Aksi nyata platform merdeka mengajar (PMM) itu, Bu. 😊	P
984	Dalam rangka memperingati Hari Guru Nasional pada tanggal 25 November 2022, kami ingin berbagi cerita mengenai guru yang ” -	P
985	Produk-produk dalam platform Merdeka Mengajar juga mendorong akses pengembangan kualitas guru karena membantu para guru menerapkan pembelajaran paradigma baru <a href="https://t.co/zfwLcpwc8j">https://t.co/zfwLcpwc8j</a>	P
986	Apresiasi untuk Mendikbudristek @nadiemmakarim berkat dengan inovasi2nya @Kemdikbud_RI telah meluncurkan Platform ” -	P
987	@Kemdikbud_RI meluncurkan Platform Merdeka Mengajar (PMM) pada 2022 untuk mendongkrak kualitas mengajar para ” -	P
988	Belajar memahami platform merdeka mengajar. #merdekamengajar #pmm #kurikulummerdeka <a href="https://t.co/IGfZEH2ZEW">https://t.co/IGfZEH2ZEW</a>	P
989	@thebensraum @pujanggalaijo @Puthutea Program yang bagus menurut saya: - platform merdeka mengajar : aplikasi impian saya dimana guru bisa berbagi dan mencari referensi.	P
990	@Puthutea ARKAS, Rapor Pendidikan, Platform Merdeka Mengajar, Tanya BOS bener bener membantu, meskipun di masa transisi ini struggle nya dapat banget tapi yakin bisa berguna suatu saat nanti untuk pendidikan di Indonesia 🇮🇩	P

991	Kepala lumayan pusing kerjain modul <sup>2</sup> platform merdeka mengajar 🤖	N
992	@ardiishere @toramichan @kpertiwi29 Untungnya website belajar-id dan platform merdeka mengajar sudah jadi dan ” -	P
993	Nah, melalui transformasi teknologi yang dilakukan Kemendikbudristek bersama GovTech Edu, saat ini ” -	P
994	@piki_blunder @Androngehe @AREAJULID Boleh dikoreksi kalo salah ya kak. Tapi setiap pembelajaran yg diberikan itu standarnya sesuai kurikulum. ” -	P
995	Wah, semakin optimis ya penggunaan platform ini meningkat bagi para guru. Jadi jangan pernah lupa untuk selalu berbagi ” -	P
996	Mana disuruh yg di platform merdeka mengajar harus dah selesai semua nonton materi+ngerjain post test nya 🤖	N
997	@Sutarno90739398 di Platform Merdeka Mengajar banyak bgt lho pak tutorial belajar Kurikulum Merdeka, gratis pulaaaa	P
998	@Sutarno90739398 padahal Kemdikbud udah nyediain berbagai platform pendampingan Kurikulum Merdeka, salah satunya platform Merdeka Mengajar	P
999	@BadWord123_ @thankyoupiggy @schfess Kurikulum Merdeka ini bisa digunakan oleh sekolah manapun, ga harus sekolah ” -	P
1000	Wajib install platform merdeka mengajar 🤖 dan ios belum ada 🤖	N

## Lampiran II

### Perbandingan Label Aktual dan Hasil Prediksi

Keterangan:

N : Negatif

P : Positif

Tabel label aktual dan prediksi rasio 8:2

Data	Actual	Predict	Data	Actual	Predict	Data	Actual	Predict
1.	P	N	68.	P	P	135.	P	P
2.	N	P	69.	P	P	136.	P	P
3.	N	N	70.	P	N	137.	P	P
4.	N	N	71.	N	P	138.	N	P
5.	N	N	72.	P	N	139.	P	P
6.	N	N	73.	N	N	140.	P	P
7.	N	N	74.	P	P	141.	P	P
8.	N	N	75.	N	N	142.	P	P
9.	P	P	76.	P	P	143.	N	P
10.	N	P	77.	N	N	144.	P	P
11.	P	P	78.	P	P	145.	P	P
12.	P	P	79.	P	P	146.	P	P
13.	N	N	80.	N	N	147.	P	P
14.	P	P	81.	N	N	148.	P	P
15.	P	P	82.	P	P	149.	N	N
16.	P	P	83.	P	P	150.	N	N
17.	N	P	84.	P	P	151.	N	N
18.	N	P	85.	N	N	152.	P	P
19.	P	P	86.	N	N	153.	P	P
20.	N	N	87.	P	P	154.	N	N
21.	P	N	88.	N	P	155.	P	P
22.	P	P	89.	P	P	156.	P	P
23.	N	N	90.	N	N	157.	N	N
24.	P	N	91.	P	N	158.	P	P
25.	P	P	92.	N	N	159.	N	P
26.	P	P	93.	N	N	160.	P	P
27.	N	P	94.	P	P	161.	N	N
28.	P	P	95.	P	P	162.	P	P
29.	P	P	96.	N	N	163.	P	P
30.	P	P	97.	P	P	164.	N	N
31.	N	N	98.	N	P	165.	P	P
32.	N	P	99.	P	P	166.	N	N
33.	N	N	100.	P	P	167.	N	N
34.	N	N	101.	P	P	168.	N	N

35.	N	N	102.	N	N	169.	P	P
36.	P	P	103.	P	P	170.	N	N
37.	N	N	104.	N	P	171.	N	N
38.	N	N	105.	N	N	172.	N	N
39.	P	P	106.	N	N	173.	P	P
40.	P	P	107.	P	N	174.	P	P
41.	N	N	108.	P	P	175.	P	N
42.	P	P	109.	N	P	176.	P	P
43.	N	N	110.	P	P	177.	P	P
44.	P	P	111.	N	N	178.	N	P
45.	P	P	112.	P	P	179.	P	P
46.	P	P	113.	P	P	180.	N	N
47.	P	P	114.	P	P	181.	N	N
48.	N	N	115.	P	P	182.	P	N
49.	N	N	116.	P	N	183.	N	N
50.	N	P	117.	N	P	184.	N	N
51.	P	P	118.	P	P	185.	P	P
52.	P	P	119.	N	N	186.	N	P
53.	P	P	120.	N	P	187.	N	N
54.	N	N	121.	P	P	188.	N	N
55.	P	P	122.	N	N	189.	P	P
56.	N	P	123.	P	P	190.	P	P
57.	N	N	124.	P	P	191.	P	P
58.	P	P	125.	P	P	192.	N	N
59.	P	P	126.	N	N	193.	P	P
60.	N	N	127.	N	P	194.	P	P
61.	N	N	128.	N	N	195.	N	P
62.	P	P	129.	P	P	196.	N	N
63.	P	N	130.	P	P	197.	P	P
64.	P	P	131.	P	P	198.	P	P
65.	N	N	132.	P	P	199.	N	N
66.	P	P	133.	P	P	200.	P	P
67.	P	N	134.	N	N			

Tabel label aktual dan prediksi rasio 8:2 dengan spell checker & emoji conversion

Data	Actual	Predict	Data	Actual	Predict	Data	Actual	Predict
1.	P	N	68.	P	P	135.	P	P
2.	N	P	69.	P	P	136.	P	P
3.	N	P	70.	P	P	137.	P	P
4.	N	N	71.	N	P	138.	N	P
5.	N	N	72.	P	N	139.	P	P
6.	N	N	73.	N	N	140.	P	P

7.	N	N	74.	P	P	141.	P	P
8.	N	N	75.	N	N	142.	P	P
9.	P	P	76.	P	P	143.	N	P
10.	N	P	77.	N	N	144.	P	P
11.	P	P	78.	P	P	145.	P	P
12.	P	P	79.	P	P	146.	P	P
13.	N	N	80.	N	N	147.	P	P
14.	P	P	81.	N	N	148.	P	P
15.	P	P	82.	P	P	149.	N	N
16.	P	P	83.	P	P	150.	N	N
17.	N	P	84.	P	P	151.	N	N
18.	N	P	85.	N	N	152.	P	P
19.	P	P	86.	N	N	153.	P	P
20.	N	N	87.	P	P	154.	N	N
21.	P	N	88.	N	N	155.	P	P
22.	P	P	89.	P	P	156.	P	P
23.	N	N	90.	N	N	157.	N	N
24.	P	N	91.	P	N	158.	P	P
25.	P	P	92.	N	N	159.	N	P
26.	P	P	93.	N	N	160.	P	P
27.	N	P	94.	P	P	161.	N	N
28.	P	P	95.	P	P	162.	P	P
29.	P	P	96.	N	N	163.	P	P
30.	P	P	97.	P	P	164.	N	N
31.	N	N	98.	N	P	165.	P	P
32.	N	P	99.	P	P	166.	N	N
33.	N	N	100.	P	P	167.	N	N
34.	N	N	101.	P	P	168.	N	N
35.	N	N	102.	N	N	169.	P	P
36.	P	P	103.	P	P	170.	N	N
37.	N	N	104.	N	P	171.	N	N
38.	N	N	105.	N	N	172.	N	N
39.	P	P	106.	N	N	173.	P	P
40.	P	P	107.	P	P	174.	P	P
41.	N	N	108.	P	P	175.	P	N
42.	P	P	109.	N	P	176.	P	P
43.	N	N	110.	P	P	177.	P	P
44.	P	P	111.	N	N	178.	N	P
45.	P	P	112.	P	P	179.	P	P
46.	P	P	113.	P	P	180.	N	N
47.	P	P	114.	P	P	181.	N	N
48.	N	N	115.	P	P	182.	P	P
49.	N	N	116.	P	N	183.	N	N
50.	N	P	117.	N	P	184.	N	N

51.	P	P	118.	P	P	185.	P	P
52.	P	P	119.	N	N	186.	N	N
53.	P	P	120.	N	P	187.	N	N
54.	N	N	121.	P	P	188.	N	N
55.	P	P	122.	N	N	189.	P	P
56.	N	P	123.	P	P	190.	P	P
57.	N	N	124.	P	P	191.	P	P
58.	P	P	125.	P	P	192.	N	N
59.	P	P	126.	N	N	193.	P	P
60.	N	N	127.	N	P	194.	P	P
61.	N	N	128.	N	N	195.	N	P
62.	P	P	129.	P	P	196.	N	N
63.	P	N	130.	P	P	197.	P	P
64.	P	P	131.	P	P	198.	P	P
65.	N	N	132.	P	P	199.	N	N
66.	P	P	133.	P	P	200.	P	P
67.	P	N	134.	N	P			

Tabel label aktual dan prediksi rasio data negatif dan positif 1:9 pada data testing

Data	Actual	Predict	Data	Actual	Predict	Data	Actual	Predict
1.	N	N	68.	P	N	135.	P	P
2.	N	P	69.	P	P	136.	P	P
3.	N	N	70.	P	P	137.	P	P
4.	N	N	71.	P	N	138.	P	P
5.	N	N	72.	P	P	139.	P	P
6.	N	P	73.	P	P	140.	P	P
7.	N	N	74.	P	P	141.	P	N
8.	N	N	75.	P	P	142.	P	P
9.	N	N	76.	P	P	143.	P	P
10.	N	N	77.	P	P	144.	P	P
11.	N	N	78.	P	P	145.	P	P
12.	N	N	79.	P	P	146.	P	P
13.	N	P	80.	P	P	147.	P	P
14.	N	N	81.	P	P	148.	P	P
15.	N	N	82.	P	P	149.	P	N
16.	N	P	83.	P	P	150.	P	P
17.	N	N	84.	P	P	151.	P	P
18.	N	P	85.	P	P	152.	P	P
19.	N	P	86.	P	P	153.	P	P
20.	N	N	87.	P	P	154.	P	P
21.	P	P	88.	P	P	155.	P	P
22.	P	P	89.	P	P	156.	P	P
23.	P	P	90.	P	P	157.	P	P



24.	P	P	91.	P	P	158.	P	P
25.	P	P	92.	P	P	159.	P	P
26.	P	P	93.	P	P	160.	P	P
27.	P	P	94.	P	P	161.	P	P
28.	P	P	95.	P	P	162.	P	P
29.	P	P	96.	P	P	163.	P	P
30.	P	P	97.	P	P	164.	P	P
31.	P	P	98.	P	P	165.	P	P
32.	P	P	99.	P	P	166.	P	P
33.	P	P	100.	P	P	167.	P	P
34.	P	P	101.	P	P	168.	P	N
35.	P	P	102.	P	P	169.	P	P
36.	P	P	103.	P	P	170.	P	P
37.	P	P	104.	P	P	171.	P	P
38.	P	P	105.	P	P	172.	P	P
39.	P	P	106.	P	P	173.	P	P
40.	P	P	107.	P	P	174.	P	P
41.	P	N	108.	P	P	175.	P	P
42.	P	P	109.	P	P	176.	P	P
43.	P	P	110.	P	P	177.	P	P
44.	P	N	111.	P	P	178.	P	P
45.	P	N	112.	P	P	179.	P	N
46.	P	P	113.	P	P	180.	P	P
47.	P	P	114.	P	P	181.	P	P
48.	P	P	115.	P	P	182.	P	P
49.	P	P	116.	P	N	183.	P	P
50.	P	P	117.	P	P	184.	P	P
51.	P	N	118.	P	P	185.	P	P
52.	P	P	119.	P	P	186.	P	P
53.	P	P	120.	P	P	187.	P	P
54.	P	P	121.	P	P	188.	P	P
55.	P	P	122.	P	P	189.	P	P
56.	P	P	123.	P	P	190.	P	P
57.	P	P	124.	P	P	191.	P	P
58.	P	P	125.	P	P	192.	P	P
59.	P	P	126.	P	P	193.	P	P
60.	P	P	127.	P	P	194.	P	P
61.	P	P	128.	P	P	195.	P	P
62.	P	N	129.	P	P	196.	P	P
63.	P	P	130.	P	P	197.	P	P
64.	P	P	131.	P	P	198.	P	P
65.	P	P	132.	P	P	199.	P	P
66.	P	P	133.	P	P	200.	P	P
67.	P	P	134.	P	P			

Tabel label aktual dan prediksi fold-5 pada uji 10-fold cross validation

<b>Data</b>	<b>Actual</b>	<b>Predict</b>	<b>Data</b>	<b>Actual</b>	<b>Predict</b>	<b>Data</b>	<b>Actual</b>	<b>Predict</b>
<b>1.</b>	P	P	<b>35.</b>	N	N	<b>69.</b>	N	N
<b>2.</b>	P	P	<b>36.</b>	P	N	<b>70.</b>	N	N
<b>3.</b>	P	P	<b>37.</b>	N	N	<b>71.</b>	P	N
<b>4.</b>	P	P	<b>38.</b>	N	N	<b>72.</b>	N	P
<b>5.</b>	P	P	<b>39.</b>	N	N	<b>73.</b>	P	P
<b>6.</b>	N	P	<b>40.</b>	N	N	<b>74.</b>	N	N
<b>7.</b>	P	P	<b>41.</b>	N	N	<b>75.</b>	P	P
<b>8.</b>	P	P	<b>42.</b>	N	N	<b>76.</b>	P	P
<b>9.</b>	P	P	<b>43.</b>	N	N	<b>77.</b>	P	P
<b>10.</b>	N	P	<b>44.</b>	P	P	<b>78.</b>	P	P
<b>11.</b>	P	P	<b>45.</b>	P	N	<b>79.</b>	P	P
<b>12.</b>	P	P	<b>46.</b>	N	N	<b>80.</b>	P	P
<b>13.</b>	N	P	<b>47.</b>	N	N	<b>81.</b>	P	P
<b>14.</b>	N	P	<b>48.</b>	N	N	<b>82.</b>	N	P
<b>15.</b>	P	P	<b>49.</b>	N	N	<b>83.</b>	N	N
<b>16.</b>	P	P	<b>50.</b>	N	N	<b>84.</b>	P	P
<b>17.</b>	P	P	<b>51.</b>	N	N	<b>85.</b>	P	P
<b>18.</b>	N	P	<b>52.</b>	N	N	<b>86.</b>	P	P
<b>19.</b>	P	P	<b>53.</b>	N	N	<b>87.</b>	P	P
<b>20.</b>	N	P	<b>54.</b>	P	P	<b>88.</b>	P	P
<b>21.</b>	N	P	<b>55.</b>	N	N	<b>89.</b>	P	P
<b>22.</b>	P	P	<b>56.</b>	N	N	<b>90.</b>	N	P
<b>23.</b>	P	P	<b>57.</b>	P	P	<b>91.</b>	P	P
<b>24.</b>	P	P	<b>58.</b>	P	P	<b>92.</b>	P	P
<b>25.</b>	P	P	<b>59.</b>	N	N	<b>93.</b>	P	P
<b>26.</b>	N	P	<b>60.</b>	N	N	<b>94.</b>	P	P
<b>27.</b>	P	P	<b>61.</b>	N	N	<b>95.</b>	P	P
<b>28.</b>	N	P	<b>62.</b>	N	N	<b>96.</b>	P	P
<b>29.</b>	P	P	<b>63.</b>	N	N	<b>97.</b>	P	P
<b>30.</b>	N	N	<b>64.</b>	N	N	<b>98.</b>	P	P
<b>31.</b>	N	N	<b>65.</b>	N	N	<b>99.</b>	P	P
<b>32.</b>	N	N	<b>66.</b>	P	P	<b>100.</b>	P	P
<b>33.</b>	N	N	<b>67.</b>	P	P			
<b>34.</b>	P	N	<b>68.</b>	N	N			

Tabel label aktual dan prediksi uji reduksi dimensi set kata kurang atau sama dengan 3 kali kemunculan kata fold-8

<b>Data</b>	<b>Actual</b>	<b>Predict</b>	<b>Data</b>	<b>Actual</b>	<b>Predict</b>	<b>Data</b>	<b>Actual</b>	<b>Predict</b>
<b>1.</b>	P	P	<b>35.</b>	N	N	<b>69.</b>	N	N
<b>2.</b>	P	P	<b>36.</b>	N	N	<b>70.</b>	P	P
<b>3.</b>	P	P	<b>37.</b>	P	P	<b>71.</b>	N	N
<b>4.</b>	N	P	<b>38.</b>	P	N	<b>72.</b>	N	N
<b>5.</b>	P	P	<b>39.</b>	N	N	<b>73.</b>	N	N
<b>6.</b>	P	P	<b>40.</b>	N	N	<b>74.</b>	N	P
<b>7.</b>	P	P	<b>41.</b>	N	N	<b>75.</b>	N	P
<b>8.</b>	P	P	<b>42.</b>	N	N	<b>76.</b>	N	N
<b>9.</b>	P	P	<b>43.</b>	N	N	<b>77.</b>	N	P
<b>10.</b>	P	P	<b>44.</b>	N	N	<b>78.</b>	P	P
<b>11.</b>	P	P	<b>45.</b>	N	N	<b>79.</b>	P	P
<b>12.</b>	P	P	<b>46.</b>	N	N	<b>80.</b>	P	P
<b>13.</b>	N	N	<b>47.</b>	N	N	<b>81.</b>	P	P
<b>14.</b>	P	P	<b>48.</b>	N	N	<b>82.</b>	P	P
<b>15.</b>	P	P	<b>49.</b>	N	N	<b>83.</b>	N	N
<b>16.</b>	P	P	<b>50.</b>	P	P	<b>84.</b>	P	P
<b>17.</b>	P	P	<b>51.</b>	P	P	<b>85.</b>	P	P
<b>18.</b>	P	P	<b>52.</b>	N	N	<b>86.</b>	N	N
<b>19.</b>	P	P	<b>53.</b>	N	N	<b>87.</b>	P	P
<b>20.</b>	P	P	<b>54.</b>	N	N	<b>88.</b>	P	P
<b>21.</b>	P	P	<b>55.</b>	N	N	<b>89.</b>	P	P
<b>22.</b>	N	P	<b>56.</b>	P	N	<b>90.</b>	P	P
<b>23.</b>	N	P	<b>57.</b>	P	N	<b>91.</b>	P	P
<b>24.</b>	P	P	<b>58.</b>	N	N	<b>92.</b>	P	P
<b>25.</b>	P	P	<b>59.</b>	P	N	<b>93.</b>	N	N
<b>26.</b>	P	P	<b>60.</b>	N	N	<b>94.</b>	P	P
<b>27.</b>	P	P	<b>61.</b>	N	N	<b>95.</b>	P	P
<b>28.</b>	P	P	<b>62.</b>	N	N	<b>96.</b>	P	P
<b>29.</b>	N	N	<b>63.</b>	N	N	<b>97.</b>	P	P
<b>30.</b>	P	P	<b>64.</b>	N	N	<b>98.</b>	P	P
<b>31.</b>	N	N	<b>65.</b>	P	P	<b>99.</b>	N	P
<b>32.</b>	N	N	<b>66.</b>	N	N	<b>100.</b>	P	P
<b>33.</b>	N	P	<b>67.</b>	N	N			
<b>34.</b>	N	N	<b>68.</b>	N	N			