

**ANALISIS SENTIMEN MASYARAKAT TERHADAP PRABOWO  
SUBIANTO MENGGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

Disusun Oleh  
**NABILA TASYA AMALIA**  
**H92219052**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA**

**2023**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : NABILA TASYA AMALIA

NIM : H92219052

Program Studi : Matematika

Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "ANALISIS SENTIMEN MASYARAKAT TERHADAP PRABOWO SUBIANTO MENGGUNAKAN *SUPPORT VECTOR MACHINE*".

Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 2 Juli 2023

Yang menyatakan,



NABILA TASYA AMALIA  
NIM. H92219052

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : NABILA TASYA AMALIA

NIM : H92219052

Judul skripsi : ANALISIS SENTIMEN MASYARAKAT TERHADAP  
PRABOWO SUBIANTO MENGGUNAKAN *SUPPORT*  
*VECTOR MACHINE*

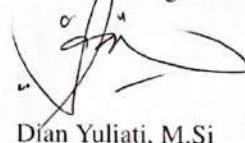
telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



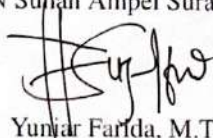
Nurissaiyah Ummuha, M.Kom  
NIP. 199011022014032004

Pembimbing II



Dian Yulianti, M.Si  
NIP. 198707142020122015

Mengetahui,  
Ketua Program Studi Matematika  
UIN Sunan Ampel Surabaya



Yuniar Farida, M.T  
NIP. 197905272014032002

## PENGESAHAN TIM PENGUJI SKRIPSI

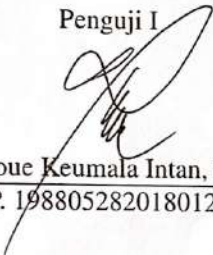
Skripsi oleh

Nama : NABILA TASYA AMALIA  
NIM : H92219052  
Judul Skripsi : ANALISIS SENTIMEN MASYARAKAT TERHADAP  
PRABOWO SUBIANTO MENGGUNAKAN *SUPPORT  
VECTOR MACHINE*

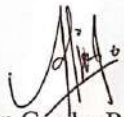
Telah dipertahankan di depan Tim Penguji  
pada tanggal 5 Juli 2023

Mengesahkan,  
Tim Penguji


Penguji I

  
Putroue Keumala Intan, M.Si  
NIP. 198805282018012001

Penguji II

  
Dr. Dian Candra Rini Novitasari, M.Kom  
NIP. 198511242014032001

Penguji III

  
Nurissaidah Ulinnuha, M.Kom  
NIP. 199011022014032004

Penguji IV

  
Dian Yuliati, M.Si  
NIP. 198707142020122015

Mengetahui,

Dean Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya

  
Hamdani, M.Pd  
NIP. 196507312000031002



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA  
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax 031-8413300  
E-Mail: [perpus@uinsby.ac.id](mailto:perpus@uinsby.ac.id)

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : NABILA TASYA AMALIA  
NIM : H92219052  
Fakultas/Jurusan : SAINS DAN TEKNOLOGI / MATEMATIKA  
E-mail address : nabilatasyamalia@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Skripsi  Tesis  Disertasi  Lain-lain (.....)

yang berjudul :

ANALISIS SENTIMEN MASYARAKAT TERHADAP  
PRABOWO SUBIANTO MENGGUNAKAN SUPPORT  
VECTOR MACHINE

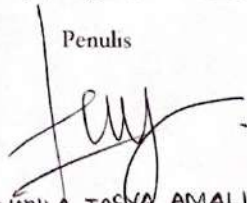
beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 5 JULI 2023

Penulis

  
(NABILA TASYA AMALIA)

*nama terang dan tanda tangan*

## ABSTRAK

### ANALISIS SENTIMEN MASYARAKAT TERHADAP PRABOWO SUBIANTO MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Partai Gerakan Indonesia Raya (GERINDRA) mengusulkan Prabowo Subianto sebagai calon presiden dalam dua periode PEMILU karena dianggap sebagai sosok yang memiliki kemampuan dan kesanggupan dalam perjuangan partai. Adanya penelitian ini diharapkan mampu melakukan analisis sentimen masyarakat terhadap Prabowo Subianto sehingga hasil dari analisis dapat digunakan masyarakat untuk lebih mengenal dan menilai sosok Prabowo Subianto. Sentimen diambil dari media sosial twitter yang selanjutnya diberi label positif dan negatif menggunakan lexicon based. Setiap kata pada sentimen dilakukan pembobotan kata menggunakan TF-IDF. Hasil dari pembobotan dan labelling, di klasifikasi menggunakan *Support Vector Machine* (SVM). SVM ialah *machine learning algorithm* yang dapat digunakan dalam klasifikasi maupun regresi. Evaluasi dari klasifikasi SVM dibuktikan dengan hasil dari *confusion matrix*. Dari penelitian ini diperoleh hasil performa sistem yaitu akurasi sebesar 94,72%, recall sebesar 98,15%, dan presisi sebesar 96,02%. Berdasarkan data twitter yang diperoleh, pada tahun 2022 masyarakat cenderung menilai Prabowo secara positif daripada negatif.

**Kata kunci:** Analisis Sentimen, Prabowo Subianto, *Lexicon Based*, *Support Vector Machine*

## ABSTRACT

### SENTIMENT ANALYSIS OF PRABOWO SUBIANTO USING SUPPORT VECTOR MACHINE

Partai Gerakan Indonesia Raya ((GERINDRA) proposed Prabowo Subianto as a presidential candidate in two election periods because he was considered a figure who had the ability and capability in the party's struggle. This research is expected to be able to analyze public sentiment towards Prabowo Subianto so that the results of the analysis can be used by the public to better recognize and assess the figure of Prabowo Subianto. Sentiments were taken from twitter social media which were then labeled positive and negative using lexicon based. Each word in the sentiment was weighted using TF-IDF. The results of weighting and labeling are classified using Support Vector Machine (SVM). SVM is a machine learning algorithm that can be used in classification and regression. Evaluation of SVM classification is proven by the results of the confusion matrix. In this study, the system performance results were obtained accuracy of 94.72%, recall of 98.15%, and precision of 96.02%. Based on the twitter data obtained, positive sentiment of Prabowo is much more than negative sentiment.

**Keywords:** Sentiment Analysis, Prabowo Subianto, Lexicon Based, Support Vector Machine

UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	i
<b>LEMBAR PERSETUJUAN PEMBIMBING</b>	ii
<b>PENGESAHAN TIM PENGUJI SKRIPSI</b>	iii
<b>HALAMAN PERNYATAAN KEASLIAN</b>	iv
<b>MOTTO</b>	v
<b>HALAMAN PERSEMBAHAN</b>	vi
<b>KATA PENGANTAR</b>	vii
<b>DAFTAR ISI</b>	ix
<b>DAFTAR TABEL</b>	xi
<b>DAFTAR GAMBAR</b>	xii
<b>ABSTRAK</b>	xiii
<b>ABSTRACT</b>	xiv
<b>I PENDAHULUAN</b>	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	8
1.3. Tujuan Penelitian	8
1.4. Batasan Masalah	9
1.5. Sistematika Penulisan	9
<b>II TINJAUAN PUSTAKA</b>	11
2.1. Analisis Sentimen	11
2.2. Twitter	12
2.3. <i>Text Mining</i>	12
2.4. <i>Pre-Processing Text</i>	13
2.4.1. <i>Clean Text</i>	13
2.4.2. <i>Case Folding</i>	14
2.4.3. <i>Deleting Stopword</i>	14
2.4.4. <i>Tokenisasi</i>	14



2.5. <i>Lexicon Based</i>	15
2.6. <i>Term Frequency Inverse Document Frequency</i>	16
2.7. <i>K-fold Cross Validation</i>	17
2.8. <i>Support Vector Machine</i>	18
2.8.1. SVM pada Data <i>Linear Separable</i>	21
2.8.2. SVM pada Data <i>Non-Linear Separable</i>	23
2.8.3. Kernel	26
2.9. Evaluasi Sistem	27
2.10. Pemimpin dalam Perspektif Islam	29
2.10.1. <i>Shidiq</i>	30
2.10.2. <i>Amanah</i>	31
2.10.3. <i>Tabligh</i>	32
2.10.4. <i>Fathonah</i>	32
<b>III METODE PENELITIAN</b>	<b>34</b>
3.1. Jenis Penelitian	34
3.2. Sumber Data	34
3.3. Tahapan Penelitian	34
<b>IV HASIL DAN PEMBAHASAN</b>	<b>39</b>
4.1. <i>Crawling Data Twitter</i>	39
4.2. <i>Pre-Processing Text</i>	41
4.3. Pelabelan Data dan Seleksi Fitur	43
4.4. <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	44
4.5. Klasifikasi SVM	46
4.6. Evaluasi Sistem	51
4.7. Visualisasi dan Analisa Hasil	53
4.8. Kesesuaian Hasil Penelitian dengan Islam	65
<b>V PENUTUP</b>	<b>68</b>
5.1. Kesimpulan	68
5.2. Saran	68
<b>DAFTAR PUSTAKA</b>	<b>70</b>

## DAFTAR TABEL

2.1 Tabel <i>confusion matrix</i> . . . . .	28
4.1 <i>Sample data</i> sentimen Prabowo Subianto . . . . .	40
4.2 Contoh tahapan <i>clean text</i> . . . . .	41
4.3 Contoh setelah dilakukan <i>case folding</i> . . . . .	42
4.4 Contoh setelah dilakukan <i>deleting stopword</i> . . . . .	42
4.5 Tokenisasi Kata . . . . .	43
4.6 Pelabelan Dokumen . . . . .	43
4.7 Pelabelan Kata . . . . .	44
4.8 Perhitungan <i>Term Frequency</i> . . . . .	45
4.9 Tabel Hasil Perhitungan TF-IDF . . . . .	46
4.10 Sampel Data Perhitungan Kernel . . . . .	47
4.11 Sampel <i>Support Vector</i> . . . . .	48
4.12 Hasil Evaluasi . . . . .	50
4.13 Hasil <i>Confusion Matrix Kernel Linear</i> dengan $k = 10$ . . . . .	51
4.14 Hasil Evaluasi Terbaik . . . . .	53

UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR GAMBAR

2.1	Arsitektur <i>K-fold Cross Validation</i>	18
2.2	Arsitektur SVM	19
3.1	Diagram alir	35
4.1	Plot Sebaran Sentimen	40
4.2	Grafik Perbandingan Jumlah Sentimen Positif dan Negatif	54
4.3	Sentimen Positif terhadap Prabowo Subianto	55
4.4	sampel <i>tweet</i> positif dengan keyword "pertahanan"	55
4.5	sampel <i>tweet</i> positif dengan keyword "ganjar"	56
4.6	sampel <i>tweet</i> positif dengan keyword "alutsista"	57
4.7	sampel <i>tweet</i> positif dengan keyword "alutsista"	57
4.8	sampel <i>tweet</i> positif dengan keyword "perdamaian"	57
4.9	sampel <i>tweet</i> positif dengan keyword "teratas"	58
4.10	sampel <i>tweet</i> positif dengan keyword "teratas"	58
4.11	<i>Fishbone</i> Diagram Sentimen Positif	59
4.12	Sentimen Negatif terhadap Prabowo Subianto	61
4.13	sampel <i>tweet</i> negatif dengan keyword "gagal"	62
4.14	sampel <i>tweet</i> negatif dengan keyword "gagal"	62
4.15	sampel <i>tweet</i> negatif dengan keyword "dihina"	63
4.16	<i>Fishbone</i> Diagram Sentimen Negatif	63

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Indonesia merupakan negara demokratis dimana bentuk atau mekanisme pemerintahan diwujudkan berdasarkan kedaulatan rakyat. Warga Negara Indonesia mempunyai hak yang sama dalam diambilnya suatu keputusan yang mampu merubah kelangsungan hidup mereka (Van Veelen and Van Der Horst, 2018). Sistem demokrasi ialah diizinkanannya keikutsertaan warga negara, baik secara langsung maupun tidak langsung dalam perumusan, pengembangan, serta pembuatan hukum (Zulfirman and Manurung, 2018). Sebagaimana telah tercantum dalam pembukaan Undang-Undang Dasar 1945 yang berbunyi. Bentuk negara demokrasi juga dijelaskan pada pasal 1 ayat 2 Undang-Undang Dasar 1945 bahwa, "Kedaulatan berada di tangan rakyat dan dilaksanakan menurut Undang-Undang Dasar." Berdasarkan Undang-Undang tersebut, maka rakyat Indonesia dapat memberikan aspirasi untuk kemajuan Indonesia (Yuvens, 2016). Aspirasi rakyat dapat diwujudkan dengan ikut serta memberikan suaranya pada Pemilihan Umum (PEMILU).

Presiden di Indonesia dipilih oleh rakyat melalui PEMILU yang diselenggarakan setiap 5 tahun sekali (Santoso and Budhiati, 2021). Pasangan calon presiden yang dapat bergabung dalam PEMILU merupakan usulan dari partai politik atau gabungan partai politik yang memperoleh sekurang-kurangnya 15% jumlah kursi di Dewan Perwakilan Rakyat (DPR) atau 20% dari perolehan

suara sah nasional dalam PEMILU anggota DPR sebagaimana telah disebutkan dalam UU Nomor 23 tahun 2003 pasal 5 ayat 4 (Zulfirman and Manurung, 2018).

Partai politik mengusulkan pasangan calon presiden dengan melihat dari beberapa aspek. Salah satu aspek yang menjadi penilaian yaitu dengan melihat sentimen masyarakat terhadap tokoh yang akan diusulkan dalam PEMILU (Farisa, 2021). Tokoh yang memiliki sentimen positif diharapkan dapat menjadi pemimpin yang baik. Partai Gerakan Indonesia Raya (GERINDRA) mengusulkan Prabowo Subianto sebagai calon presiden dalam dua periode PEMILU karena dianggap sebagai sosok yang memiliki kemampuan dan kesanggupan dalam perjuangan partai. Arief Puyuno menyatakan Prabowo adalah sosok yang negarawan dan tidak gila jabatan (Aditya, 2021). Pada PEMILU 2014 dan 2019, persentase suara yang didapatkan oleh Prabowo tidak kalah banyak daripada Jokowi. Persentase hasil suara PEMILU menunjukkan selisih yang tidak lebih dari 10% suara. Hal ini dapat diartikan bahwa sebagian besar masyarakat Indonesia berada di pihak Prabowo (Farisa, 2022).

Prabowo gagal dalam dua periode PEMILU, akan tetapi dilantik menjadi Menteri Pertahanan Indonesia dalam kabinet presiden Joko Widodo pada tahun 2019. Kegagalan Prabowo dalam PEMILU diduga karena isu pelanggaran HAM yang dilakukan pada tahun 1997-1998. Prabowo diduga terlibat dalam penculikan dan alasan dibalik hilangnya para aktivis pada masa itu (A43, 2022). Muhammad Qodari selaku Direktur Eksekutif Indobarometer menyatakan bergabungnya Prabowo dalam pemerintahan Jokowi dapat menjadi modal awal untuk menarik hati masyarakat apabila Prabowo menyalonkan diri kembali pada PEMILU 2024 mendatang (Ekawati, 2019). Prabowo menyatakan bahwa dirinya setuju untuk menjadi bagian dari kabinet Jokowi karena ia ingin membuat Indonesia maju dan

menjadi bagian dari pemerintahan yang sekarang juga termasuk dalam memajukan Indonesia (Simanjuntak, 2021). Prabowo dilantik menjadi Menteri Pertahanan menyebabkan beberapa pendukung tidak setuju, akan tetapi beberapa lainnya sangat mendukung keputusannya (Redaksi, 2019).

Pada bulan Maret 2023, presiden RI Jokowi mengunjungi panen raya padi di Desa Lajer, Jawa Tengah. Jokowi melakukan kunjungan bersama dengan Prabowo dan Ganjar serta melakukan swafoto bersama dengan para petani. Swafoto yang dilakukan Jokowi, Ganjar, dan Prabowo menjadi topik perbincangan utama baik di media digital maupun media sosial. Hal tersebut dianggap sebagai dukungan Jokowi kepada Prabowo dan Ganjar untuk mencalonkan diri pada PILPRES 2024 (Novianto, 2023). Menurut Ahmad Riza Patria sebagai politisi GERINDRA, partai telah menetapkan Prabowo Subianto sebagai calon presiden di PILPRES 2024 (Rozie, 2023). Elektabilitas yang dimiliki Prabowo Subianto pada tahun 2023 menurut Lembaga Survei Nasional (LSN) menunjukkan Prabowo unggul dengan 38,5% daripada Ganjar dan Anies yang hanya 32,8% dan 21,9%. Pada survei Litbang Kompas di basis pemilih Nahdlatul Ulama (NU) juga menunjukkan keunggulan elektabilitas Prabowo yang mencapai 25,8%. Elektabilitas Prabowo juga menjadi yang tertinggi pada survei Indikator Politik Indonesia sebesar 38% (Redaksi CNBC Indonesia, 2023).

Berdasarkan hasil survei yang dilakukan oleh lembaga survei *poltracking*, sosok pemimpin yang diinginkan masyarakat adalah peduli pada rakyat, jujur, bisa dipercaya dan bersih dari korupsi serta berpengalaman (Febriyan, 2022). Kriteria sosok pemimpin sebagaimana telah dicantumkan dalam survei, diharapkan mampu menjalankan hukum secara adil. Sebagaimana disebutkan dalam Al-qur'an Q.S. Sad ayat 26,

يٰدَاوُد اِنَّا جَعَلْنَاكَ خَلِيفَةً فِي الْاَرْضِ فَاحْكُم بَيْنَ النَّاسِ بِالْحَقِّ وَلَا تَتَّبِعِ الْهَوٰى فَيُضِلَّكَ عَنْ سَبِيْلِ  
 اللّٰهِ اِنَّ الَّذِيْنَ يَصِلُوْنَ عَنْ سَبِيْلِ اللّٰهِ لَهُمْ عَذَابٌ شَدِيْدٌۢۤ اِمْا نَسُوْا يَوْمَ الْحِسَابِ

artinya: “Wahai Daud, sesungguhnya kami menjadikanmu khalifah (penguasa) di bumi. Maka, berilah keputusan (perkara) di antara manusia dengan hak dan janganlah mengikuti hawa nafsu karena akan menyesatkan engkau dari jalan Allah. Sesungguhnya orang-orang yang sesat dari jalan Allah akan mendapat azab yang berat, karena mereka melupakan hari Perhitungan.”

Pada ayat tersebut dijelaskan bahwa sosok pemimpin diharapkan mampu mengayomi, melindungi, bahkan memberikan keadilan bagi masyarakat luas. Hal ini seperti yang diinginkan oleh masyarakat sebagaimana hasil survei yang dilakukan oleh lembaga survei *Poltracking*. Adapun pada hadis Bukhari dijelaskan bahwa setiap pemimpin akan dimintai pertanggung jawaban.

كُلُّكُمْ رَاعٍ وَكُلُّكُمْ مَسْئُوْلٌ عَنْ رَعِيَّتِهِ

Artinya: *Kamu sekalian adalah pemimpin dan semua kamu akan diminta pertanggung jawaban tentang kepemimpinannya.*

Hadis diatas menjelaskan bahwa setiap pemimpin akan dimintai pertanggung jawaban atas apa yang dikerjakannya. Hal ini didukung dengan kaidah ke-6 yang berbunyi,

تَصَرَّفُ الْاِمَامِ عَلٰى الرَّعِيَّةِ مَنُوْطٌ بِالْمَصْلَحَةِ

artinya: ”Tindakan pemimpin (imam) terhadap rakyatnya harus dikaitkan dengan kemaslahatan”

Kaidah ini merupakan acuan para pemimpin dalam mengambil kebijakan-kebijakan yang berkaitan dengan rakyat. Para pemimpin diharapkan mempertimbangkan kemaslahatan rakyat dalam menentukan kebijakan.

Tokoh dengan sentimen positif dinilai dapat memiliki peluang yang cukup besar untuk memenangkan PEMILU. Sentimen-sentimen publik dapat diperoleh melalui hasil survei, website dan media sosial (Leelawat et al., 2022). Penelitian oleh Arif menggunakan data sentimen masyarakat melalui website pegipegi.com. Arif menggunakan data sentimen masyarakat sebagai dasar untuk menilai maskapai penerbangan dengan 1100 ulasan. Data sentimen masyarakat digunakan oleh arif karena data sentimen dapat mewakili aspirasi dari pengguna maskapai Negara et al. (2020). Penelitian Sangeetha tahun 2023 menjelaskan tentang *review* produk pada platform amazon. Data yang digunakan adalah 25.000 ulasan produk yang mendapatkan *rating* bintang lima. Sangeetha menggunakan data tersebut karena data ulasan mewakili penilaian dari produk-produk yang mendapatkan *rating* bintang lima. Data yang digunakan Sangeetha hanyalah ulasan pembeli yang memberikan ratingnya untuk produk tertentu Sangeetha and Kumaran (2022). Data sentimen melalui twitter juga digunakan pada penelitian Miftahul tahun 2023 Miftahul melakukan penelitian untuk mengukur sentimen masyarakat terhadap vaksin COVID-19. Data yang digunakan ialah 42.796 tweet dalam 12 hari. Menurut Miftahul, twitter merupakan platform yang dapat digunakan oleh semua orang dan data pada twitter sangatlah banyak dibandingkan dengan platform-platform lainnya Qorib et al. (2023).

Data twitter juga digunakan pada penelitian yang lain. Veny menggunakan



data sentimen sosial media twitter dalam penelitiannya. Veny menggunakan 3,744 tweet sentimen tentang kampanye anti LGBT di Indonesia. Data sentimen twitter digunakan oleh Veny karena twitter merupakan media sosial yang memuat hingga ribuan sentimen setiap harinya secara bebas [Fitri et al. \(2019\)](#). Penelitian lain dilakukan oleh Natt tahun 2022 menjelaskan tentang penilaian sentimen masyarakat mengenai sektor pariwisata di Thailand pada saat pandemi COVID-19. Data yang digunakan adalah data twitter sebanyak 150.580 opini. Data sentimen twitter digunakan oleh Natt karena twitter menyediakan banyak sekali opini yang sesuai dengan topik yang diinginkan [Leelawat et al. \(2022\)](#). Penelitian yang serupa juga dilakukan oleh Siru Liu pada tahun 2021 tentang sentimen masyarakat terhadap vaksin COVID-19 selama 3 bulan. Menurut Siru Liu, twitter merupakan platform penting yang mengumpulkan opini publik dan secara luas membahas tentang studi-studi kesehatan masyarakat. Penelitian Siru Liu menunjukkan bahwa data yang digunakan mencapai lebih dari 2.000.000 tweet atau sentimen. Hal ini membuktikan bahwa data twitter relevan dengan topik-topik tertentu, mudah diakses dan berjumlah banyak [Liu and Liu \(2021\)](#). Berdasarkan beberapa penelitian yang telah dilakukan, dapat dilihat bahwa media sosial twitter memiliki banyak sentimen masyarakat dengan akses yang mudah dibandingkan dengan media lainnya, maka pada penelitian ini menggunakan sentimen masyarakat melalui data twitter.

Sentimen twitter dengan jumlah banyak data membutuhkan suatu sistem otomatis yang bisa melakukan klasifikasi sentimen positif dan negatif. Penelitian Junadhi tentang klasifikasi sentimen masyarakat terhadap pembelajaran daring. Junadhi menggunakan *K-Nearest Neighbor (KNN) Algorithm* dan akurasi yang dihasilkan mencapai 56%. Akurasi tersebut menunjukkan sistem yang dibuat oleh

algoritma tersebut kurang baik [Anam \(2022\)](#). Penelitian Anam terhadap analisis sentimen interaksi masyarakat dengan pemerintahan menggunakan metode yang berbeda. Anam menggunakan *decision tree* dan hasil akurasi yang didapatkan mencapai 62.28% [Anam et al. \(2021\)](#). Penelitian serupa juga dilakukan oleh Abdul pada tahun 2019. Penelitian Abdul menggunakan sentimen twitter tentang ulasan masyarakat terhadap maskapai penerbangan. Metode klasifikasi yang dilakukan Abdul menggunakan *naive bayes* dan hasil yang didapatkan sebesar 76.56% [Rahat et al. \(2019\)](#).

Penelitian oleh Abdelrahman tentang analisis sentimen twitter terhadap maskapai penerbangan Amerika Serikat dengan membandingkan beberapa metode. Metode yang digunakan yaitu *Support Vector Machine (SVM)*, *logistic legression*, *random forest*, *xgboost*, *decision tree*, *naive bayes* dan KNN. Hasil terbaik didapatkan pada SVM dengan akurasi sebesar 83.31%, sedangkan untuk metode lainnya rata-rata menghasilkan 75%. Menurut Abdelrahman, SVM dapat memberikan desain model yang dapat melakukan klasifikasi banyaknya data twitter [Saad \(2020\)](#). Penelitian lain membandingkan SVM, *naive bayes*, KNN, dan terhadap analisis sentimen masyarakat. Hasil yang diperoleh yaitu SVM sebagai metode klasifikasi yang paling baik daripada *naive bayes* dan KNN. Hasil akurasi menunjukkan bahwa SVM mencapai 90.81%, sedangkan untuk *naive bayes* dan KNN hanya sebesar 79.20% dan 62.10%. Peneliti juga menyebutkan bahwa SVM mampu mengolah data berdimensi tinggi tanpa mengalami penurunan performa [Pamungkas and Kharisudin \(2021\)](#). Penelitian lainnya tentang klasifikasi sentimen menggunakan SVM juga dilakukan oleh Gunawan. Gunawan melakukan analisis sentimen terhadap calon gubernur jawa barat. Hasil akurasi yang didapat Gunawan sebesar 93.03% [Gunawan et al. \(2020\)](#). Penelitian serupa juga dilakukan oleh Arsi

dengan sentimen wacana pemindahan ibu kota dilakukan menggunakan SVM. Akurasi yang dihasilkan dari penelitian Arsi sebesar 96.68% [Arsi and Waluyo \(2021\)](#). Styawati juga melakukan penelitian pada sentimen masyarakat terhadap program kartu pekerja. Styawati menggunakan SVM sebagai metode penelitiannya dan menghasilkan akurasi sebesar 98.67% [Styawati et al. \(2021\)](#).

Berdasarkan beberapa penelitian yang telah dilakukan, dapat ditunjukkan bahwa SVM menunjukkan kinerja yang baik dengan akurasi yang tinggi selama beberapa tahun terakhir. Berdasarkan paparan tersebut, maka perlu dilakukan analisis sentimen terhadap Prabowo Subianto dengan menggunakan SVM. Adanya penelitian ini diharapkan mampu melakukan analisis sentimen masyarakat terhadap Prabowo subianto sehingga hasil dari analisis dapat digunakan masyarakat untuk lebih mengenal dan menilai sosok Prabowo Subianto.

## 1.2. Rumusan Masalah

1. Bagaimana hasil pelabelan sentimen masyarakat menggunakan *lexicon based* terhadap Prabowo Subianto?
2. Bagaimana hasil akurasi, *recall*, dan presisi *Support Vector Machine* pada klasifikasi sentimen Prabowo Subianto?

## 1.3. Tujuan Penelitian

1. Mengetahui sentimen positif dan negatif Prabowo Subianto menggunakan *lexicon based*.
2. Mengetahui hasil akurasi, *recall*, dan presisi *Support Vector Machine* pada klasifikasi sentimen Prabowo Subianto.

#### 1.4. Batasan Masalah

1. Data yang digunakan adalah opini masyarakat terhadap Prabowo Subianto yang didapatkan melalui media sosial twitter berbahasa Indonesia pada tahun 2022.
2. Pelabelan data hanya dilakukan berdasarkan probabilitas kata tanpa memperhatikan aturan *semiotika*
3. Klasifikasi dibagi dalam dua kelas, yaitu sentimen negatif dan sentimen positif.

#### 1.5. Sistematika Penulisan

Pada sistematika penelitian ini, isi dari penelitian terdiri dari lima bab yang telah dirangkum, yaitu:

1. BAB I PENDAHULUAN yang berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah dan sistematika penulisan.
2. BAB II TINJAUAN PUSTAKA yang memaparkan beberapa teori yang mengacu pada penelitian yang sudah ada sebelumnya baik jurnal maupun buku pendukung penelitian. Tinjauan pustaka penelitian ini berisi pemimpin dalam perspektif islam, opini masyarakat pada media sosial twitter sebagai sentimen, beberapa teori analisis sentimen dan data mining, tahapan *pre-processing*, pelabelan *lexicon based*, ekstraksi fitur TF-IDF, dan tahapan SVM.
3. BAB III METODE PENELITIAN yang memaparkan jenis penelitian, sumber data penelitian, variabel dan subjek apa saja yang digunakan dalam penelitian,

alur mengelola data tersebut hingga rumusan masalah dari penelitian dapat terselesaikan.

4. BAB IV HASIL DAN PEMBAHASAN yang memaparkan hasil penelitian yang dilakukan untuk membuktikan kebenaran rumusan masalah, serta dijelaskan proses pada setiap tahapan yang dilakukan.
5. BAB V PENUTUP menguraikan kesimpulan penelitian dari hasil yang didapat dan memberikan saran kepada peneliti-peneliti selanjutnya yang akan menggunakan data atau metode pada penelitian ini.



UIN SUNAN AMPEL  
S U R A B A Y A

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Analisis Sentimen

Analisis sentimen yaitu proses analisis penggunaan teks untuk mendapatkan berbagai sumber data yang berasal dari internet maupun berbagai platform jejaring sosial. Analisis sentimen ialah salah satu bidang dari *Natural Language Processing* (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Dengan menggunakan analisis sentimen, informasi akan menjadi lebih terstruktur (Yang et al., 2020).

Analisis sentimen dilakukan dengan mengelompokkan beberapa teks yang terdapat dalam suatu kalimat. Kemudian menganalisis apakah teks tersebut termasuk kedalam sentimen positif, negatif, atau netral. Analisis sentimen erat hubungannya dengan *text mining* yang bertujuan untuk menganalisis pendapat, sentimen, serta emosi seseorang terhadap suatu layanan, produk, aktivitas, bahkan individu tertentu (Xu et al., 2019).

Tujuan dari analisis sentimen yaitu untuk mengetahui suatu opini maupun perilaku seseorang terhadap suatu topik tertentu. Oleh karena itu, hal ini dapat membantu mempermudah dalam pekerjaan manusia. Sebagai contoh, sebuah perusahaan produksi yang ingin mengetahui minat pasar saat ini dapat dilihat dengan menggunakan analisis sentimen. Adapun *review* produk suatu brand tertentu juga dapat dilihat melalui analisis sentimen. (Yadav and Vishwakarma, 2020).

## 2.2. Twitter

Twitter adalah salah satu platform jejaring sosial milik suatu perusahaan di Amerika yaitu Twitter, Inc. yang diresmikan pada Maret 2006 (Dang et al., 2020). Jumlah pengguna twitter di Indonesia mencapai 18,450,000 pengguna pada tahun 2022. Jumlah tersebut setara dengan 4.23 persen dari total pengguna twitter di dunia yang mencapai 436,000,000 pengguna. Terdapat beberapa istilah dasar dalam twitter yaitu, *tweet*, *mention*, *hashtag*, dan *retweet* (Bayu, 2022).

*Tweet* merupakan opini atau curahan yang ditulis kemudian diposting oleh pengguna twitter. *Mention* adalah suatu cara untuk menghubungkan tweet kepada pengguna tertentu. *Mention* digunakan untuk menandai seseorang dalam suatu *tweet*, atau membalas *tweet* tertentu. Istilah lainnya yaitu *hashtag* atau simbol pagar yang ditulis didepan suatu topik tanpa spasi. *Hashtag* digunakan untuk mencari topik tertentu dengan sangat mudah. *Hashtag* akan menjadi *tranding* topik ketika *hashtag* tersebut sering disebutkan dalam banyak *tweet*. Hal tersebut sangat memudahkan untuk mencari topik-topik hangat terkini. *Retweet* adalah memposting kembali *tweet* seseorang. *Retweet* digunakan untuk menyampaikan kembali suatu *tweet* dengan menggunakan akun pribadi kepada pengikut atau teman dalam twitter (Phan et al., 2020).

## 2.3. Text Mining

*Text Mining* merupakan proses identifikasi pola yang bermakna dengan mengubah teks yang tidak terstruktur menjadi format terstruktur, dengan mengubah suatu data menjadi informasi yang dapat dipahami oleh sistem. *Text mining* melakukan proses klasifikasi secara otomatis berdasarkan sentimen, topik, dan suatu maksud tertentu. *Text mining* menggabungkan beberapa gagasan, seperti

statistik, linguistik, dan *machine learning* untuk membuat model pembelajaran dari data latih yang kemudian dapat memprediksi hasil pada informasi baru berdasarkan pengalaman sebelumnya (Qaiser and Ali, 2018).

Pemrosesan bahasa alami digunakan secara otomatis dalam *text mining* untuk mengumpulkan informasi mendalam dari teks yang tidak terstruktur. *Text mining* memproses kategori teks berdasarkan sentimen, topik, dan maksud dengan mengubah data menjadi pengetahuan yang dapat dipahami komputer secara otomatis. Adapun metode yang dapat digunakan dalam *text mining* yaitu, *Support Vector Machine*, *Naive Bayes Classifier*, dan beberapa algoritma *learning* lainnya (Kim and Gil, 2019).

#### **2.4. Pre-Processing Text**

*Pre-Processing text* merupakan tahapan yang dilakukan untuk mendapatkan dokumen teks yang sesuai untuk klasifikasi. Dokumen teks yang ada biasanya tidak memiliki struktur yang jelas, sehingga sulit untuk memproses informasi yang dikandungnya secara langsung. Pada tahap ini juga dapat menyampaikan pokok bahasan suatu teks dengan akurat. Tahapan ini meliputi *clean text*, *case folding*, *deleting stopwords*, dan tokenisasi (Liu et al., 2018).

##### **2.4.1. Clean Text**

*Clean Text* adalah proses menghilangkan kata-kata yang tidak bermakna, seperti *html*, *username*, *hashtag*, *web*, *emoji*, *tanda baca*, *angka* dan *simbol-simbol*. Oleh karena itu, *clean text* digunakan untuk menghapus semua karakter atau kata yang tidak memiliki makna dan hanya menyisihkan kata-kata bermakna (Villavicencio et al., 2021).



#### **2.4.2. Case Folding**

*Case folding* digunakan untuk memudahkan analisis, dengan cara mengubah semua huruf kapital menjadi huruf kecil. *Case folding* dilakukan untuk memastikan bahwa kata yang terdapat di awal kalimat memiliki arti yang sama dengan kata yang ada di tengah maupun di akhir kalimat (Fransiska et al., 2020).

#### **2.4.3. Deleting Stopword**

*Stopword* adalah kumpulan dari berbagai kata yang tidak memiliki makna. Sebagian besar *tweet* yang ditulis oleh berbagai pengguna mengandung kata-kata yang terdapat dalam *stopword*, seperti kata imbuhan, kata penghubung, dan kata ganti. Contohnya, yang, di, ke, dari, dia dan lain sebagainya. Kata-kata tersebut tidak bisa dijadikan suatu acuan dalam pengelompokan teks sehingga perlu dihapus (Al Amrani et al., 2018).

#### **2.4.4. Tokenisasi**

Tokenisasi merupakan proses membagi dokumen teks menjadi token-token atau bagian tertentu yang dapat berupa kalimat atau paragraf. Tokenisasi digunakan untuk memisahkan kata per kata agar bisa dilakukan pembobotan untuk setiap katanya. Contohnya pada kalimat "Jokowi telah berperan penting dalam pembangunan negara yang baik" diubah menjadi "Jokowi, telah, berperan, penting, dalam, pembangunan, negara, yang, baik" yang menghasilkan 9 token (Al Amrani et al., 2018).

## 2.5. *Lexicon Based*

*Lexicon based* merupakan metode dalam pemrosesan bahasa alami yang berfokus pada analisis dan pemahaman makna kata-kata dalam teks dengan memanfaatkan leksikon atau kamus kata-kata. Pendekatan ini bergantung pada informasi yang terkait dengan kata-kata dalam leksikon untuk menginterpretasikan makna teks. Melalui kamus *lexicon*, pendekatan berbasis *lexicon* dapat mengenali sinonim, antonim, dan hubungan semantik antara kata-kata, serta membangun representasi makna yang lebih komprehensif (Yang et al., 2020).

*Lexicon based* telah digunakan dalam berbagai aplikasi pemrosesan bahasa alami, termasuk pemahaman teks, analisis sentimen, dan terjemahan mesin. Metode yang umum digunakan dalam pendekatan ini melibatkan analisis kata-kata dalam teks dan pencocokan dengan entri dalam leksikon untuk memperoleh informasi makna (Mustaqim et al., 2020). Dengan menggunakan informasi leksikon, sistem dapat mengidentifikasi makna kata-kata dan memahami hubungan semantik dalam konteks yang diberikan. Terdapat beberapa algoritma dalam *lexicon based*, salah satunya yaitu *Valence Aware Dictionary and sEntiment Reasoner* (VADER) (Borg and Boldt, 2020).

VADER merupakan algoritma analisis sentimen yang populer dalam pemrosesan bahasa alami. Algoritma ini dikembangkan untuk mengidentifikasi sentimen positif, negatif, atau netral dalam teks. VADER menggunakan pendekatan *lexicon based*, di mana ia mengandalkan sebuah *lexicon* yang mengaitkan kata-kata dengan skor sentimen. *Lexicon* dalam VADER menyediakan skor sentimen untuk ribuan kata-kata dalam bahasa Inggris, dan setiap kata diberi skor berdasarkan sejauh mana kata tersebut cenderung memiliki sentimen positif atau negatif (Abiola et al., 2023).

VADER sering digunakan dalam berbagai aplikasi seperti analisis sosial media, analisis ulasan produk, survei kepuasan pelanggan, dan analisis sentimen teks lainnya. Kelebihan VADER adalah kemampuannya untuk mengatasi beberapa tantangan dalam analisis sentimen, seperti kata-kata yang ambigu atau frasa dengan konstruksi gramatikal yang kompleks. Dalam kamus *lexicon* VADER disertai juga dengan bobot pada masing-masing kata. Setelah menganalisis teks, VADER akan menghasilkan skor compound. Skor compound didapat dari penjumlahan skor *valence* dari setiap kata dalam *lexicon* kemudian dinormalisasi dengan rentang -1 sampai dengan 1. Skor sentimen positif memiliki nilai antara 0 hingga 1, dengan nilai yang lebih tinggi menunjukkan sentimen positif yang lebih kuat. Skor sentimen negatif memiliki nilai antara -1 hingga 0, dengan nilai yang lebih rendah menunjukkan sentimen negatif yang lebih kuat. Skor sentimen netral bernilai 0 (Elbagir and Yang, 2019).

## 2.6. *Term Frequency Inverse Document Frequency*

Analisis data teks memerlukan pembobotan pada setiap kata. Hal ini dikarenakan kata yang ada pada satu kalimat dengan kalimat lainnya bisa saja bermakna berbeda. Pembobotan kata pada data teks dapat dilakukan dengan TF-IDF (*Term Frequency Inverse Document Frequency*). TF-IDF adalah metode pengukuran statistik yang mengevaluasi seberapa relevan suatu kata dengan suatu dokumen dalam kumpulan dokumen (Riego and Villarba, 2023). Perhitungannya dilakukan dengan mengalikan 2 buah matriks yang berisi banyaknya suatu kata yang muncul pada suatu dokumen, serta invers frekuensi dokumen dari kata pada satu set dokumen. Fungsi dari TF-IDF sendiri yaitu untuk analisis teks secara otomatis, dan sangat berguna dalam pembobotan kata dalam algoritma *machine learning* untuk *Natural Language Processing* (NLP) (Kim and Gil, 2019).

*Term Frequency (TF)* berperan dalam menghitung frekuensi kemunculan suatu kata pada dokumen. Sedangkan *Inverse Document Frequency (IDF)* berperan dalam menghitung seberapa sering atau jarang sebuah kata muncul pada seluruh dokumen. Hasil dari perkalian TF dan IDF adalah skor suatu kata dalam dokumen. Semakin tinggi skor yang didapat, maka semakin relevan kata tersebut dalam dokumen tertentu (Gomes et al., 2023). Perhitungan TF-IDF seperti pada persamaan (2.1) & (2.2):

$$w_{i,j} = TF_{i,j} \times IDF \quad (2.1)$$

$$IDF = \log \left( \frac{N}{df_i} \right) \quad (2.2)$$

Dimana,

$w_{i,j}$  = bobot kata ( $i$ ) pada dokumen ( $j$ )

$TF_{i,j}$  = banyak kata ( $i$ ) yang muncul pada dokumen ( $j$ )

$df_i$  = banyaknya dokumen yang mengandung kata ( $i$ )

Dengan,

$i = 1, 2, 3, \dots, p$  (jumlah variabel atau kata)

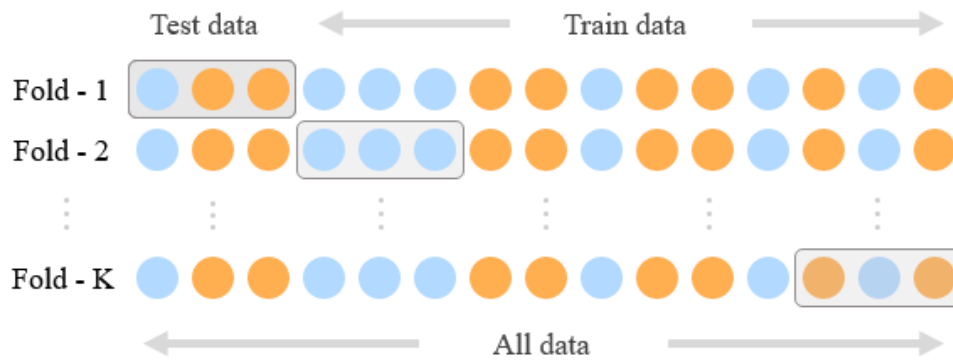
$j = 1, 2, 3, \dots, N$  (jumlah data atau dokumen)

$N$  = Jumlah seluruh data

## 2.7. K-fold Cross Validation

*Cross validation* merupakan suatu metode dari teknik data mining yang bertujuan untuk mengevaluasi kinerja model atau algoritma dengan data dipisahkan menjadi dua subset. Metode ini disebut juga dengan *k-fold cross validation* dimana data latih dan data uji digunakan secara berulang sebanyak  $k$  kali untuk satu model dengan parameter yang sama (Zhang et al., 2023). *Cross*

*validation* dapat bekerja dengan cepat dengan pengambilan sampel yang lebih terstruktur. Pada pengujian, data latih dan data uji akan diambil dengan data yang berbeda dengan percobaan atau iterasi sebelumnya (Li et al., 2023). Pada metode ini dilakukan secara acak dengan membagi data menjadi  $k$  lipatan sehingga akurasi rata-rata yang dihasilkan tidak konstan. Proses *training* digunakan dengan  $k - 1$  *fold*, untuk menghitung nilai eror klasifikasi digunakan data uji pada *fold* yang tersisa (Anandan and Manikandan, 2023). Pada Gambar 2.1 merupakan arsitektur dari *k-fold cross validation*.



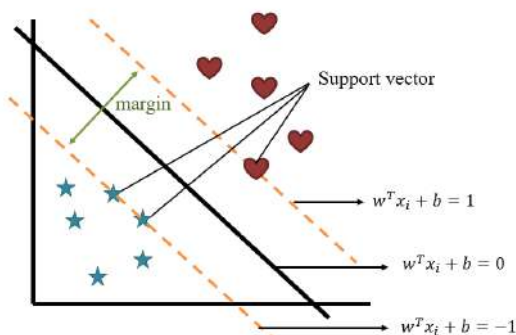
Gambar 2.1 Arsitektur *K-fold Cross Validation*

## 2.8. Support Vector Machine

*Support Vector Machine* (SVM) ialah *machine learning algorithm* yang dapat digunakan dalam klasifikasi maupun regresi. Untuk pertama kalinya, SVM digunakan dalam klasifikasi teks oleh (Thosten Joachims, 1998). SVM telah digunakan di berbagai bidang (bioinformatika, pencarian informasi, visi komputer, keuangan, dll.) (Cengiz et al., 2023). SVM menggunakan asumsi bahwa fungsi-fungsi linier ada dalam fitur ruang dengan beberapa dimensi dan menggabungkan bias pembelajaran dari teori pembelajaran statistik (Singh and

(Singh, 2023). SVM merupakan pendekatan yang relatif baru dibandingkan dengan teknik lain pada umumnya, metode ini bekerja lebih baik pada berbagai bidang aplikasi. Fungsi dan parameter kernel yang digunakan selama proses transisi SVM akan berdampak signifikan pada tingkat akurasi model yang dihasilkan (Zendejboudi et al., 2018).

Teknik klasifikasi statistik pada SVM memaksimalkan margin antara dua kelas dengan mencari *hyperplane* yang optimal. Hal tersebut dianggap sebagai strategi klasifikasi teks yang paling efektif karena dapat menangani vektor fitur dengan banyak dimensi tak terhingga. Algoritma SVM melakukan *plotting* setiap titik data sebagai titik dalam ruang n-dimensi, di mana n adalah jumlah fitur dan nilai setiap fitur adalah nilai koordinat tertentu (Al Amrani et al., 2018).



Gambar 2.2 Arsitektur SVM

SVM bertujuan untuk mengidentifikasi *hyperplane* yang optimal dalam ruang input yang membagi dua kelas. Pada Gambar (2.2) terdapat anggota dari dua kelas yang dipisahkan oleh *hyperplane* yaitu kelas +1 dan -1 dengan kelas -1 berwarna biru, sedangkan kelas +1 berwarna merah. Gambar a menunjukkan *hyperplane* pemisah sesuai dengan kelas. Margin adalah pemisah antara *hyperplane* terdekat dari setiap kelas. Data yang terdapat di dalam bidang

pembatas disebut dengan *support vector* (Wang et al., 2023). *Hyperplane* yang optimal dapat dihitung dengan mengukur margin dengan mencari titik maksimal (Yatsugi et al., 2023). Misalkan terdapat data latih sebanyak  $N$  data dengan  $x_i = [x_1, x_2, \dots, x_n]$  adalah vektor baris di ruang dimensi  $n$  dan  $t_i = \{-1, 1\}$  adalah nilai target dalam tiap baris vektor, dengan dua kelas yang berbeda yaitu,  $R_1(t_i = +1)$  dan  $R_2(t_i = -1)$ . Penyelesaian umum SVM menggunakan dapat dilihat pada persamaan (2.3):

$$y(w) = w^T x + b \quad (2.3)$$

dimana,  $w$  merupakan parameter bobot,  $b$  adalah bias, dan  $x$  adalah input vektor yang termasuk dalam kelas  $R_1$  jika  $y(x) > 0$  dan termasuk dalam kelas  $R_2$  jika  $y(x) < 0$ . Sehingga, SVM memanfaatkan *hyperplane* dalam mengkategorikan data ke dalam dua kelas yang berbeda, yaitu  $R_1$  dan  $R_2$ . Setiap kelas mempunyai garis pembatas yang sejajar dengan *hyperplane*, yang secara efektif memisahkan dua kelas seperti pada persamaan (2.4) dan (2.5):

$$w^T x_i + b \geq 1, t_i = 1 \quad (2.4)$$

$$w^T x_i + b \geq -1, t_i = -1 \quad (2.5)$$

atau persamaan (2.4) dan (2.5) dapat dituliskan menjadi:

$$t_i(w^T x_i + b) - 1 \geq 0 \quad (2.6)$$

Jarak antara garis pembatas dengan *hyperplane* disebut dengan margin. Garis pembatas dan margin dalam SVM ditentukan berdasarkan letak *support*

*vector*. *Support vector* adalah titik-titik data yang letaknya paling mendekati *hyperplane*. Jarak normal antara *support vector* dengan *hyperplane* seperti pada persamaan (2.7) (Bishop and Nasrabadi, 2006):

$$\frac{w^T x}{\|w\|} = -\frac{b}{\|w\|} \quad (2.7)$$

### 2.8.1. SVM pada Data *Linear Separable*

Asumsikan seluruh data terklasifikasi secara linier, maka digunakan *hard margin* untuk menghitung jarak dari garis pembatas terhadap *hyperplane*. *Hard margin* didapatkan dengan memaksimalkan nilai  $\frac{1}{\|w\|}$  yang sama dengan meminimalkan nilai  $\|w\|^2$ . Sehingga dapat dirumuskan masalah optimasi pada data linier yaitu (Bishop and Nasrabadi, 2006):

$$\arg \min_{w,b} \frac{1}{2} \|w\|^2 \quad (2.8)$$

$$s.t. \quad t_i(w^T x_i + b) - 1 \geq 0, \quad i = 1, 2, \dots, N \quad (2.9)$$

Persamaan (2.8) dan (2.9) akan lebih mudah diselesaikan dengan menggunakan fungsi pengali *Lagrange*  $\alpha$ , dimana  $\alpha_i \geq 0$ . Akibatnya, masalah optimasi seperti pada persamaan (2.10):

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i \{t_i(w^T x_i + b) - 1\} \quad (2.10)$$



hasil turunan dari persamaan (2.10) dengan memperhatikan dua kondisi, yaitu:

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^N \alpha_i t_i x_i = 0 \rightarrow w = \sum_{i=1}^N \alpha_i t_i x_i \quad (2.11)$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = - \sum_{i=1}^N \alpha_i t_i = 0 \rightarrow \sum_{i=1}^N \alpha_i t_i = 0 \quad (2.12)$$

substitusikan persamaan (2.11) dan (2.12) ke persamaan (2.10):

$$L(w, b, \alpha) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^N \alpha_i \{ t_i (w^T x_i + b) - 1 \}$$

$$\tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i t_i \alpha_j t_j k(x_i, x_j) \quad (2.13)$$

sehingga, didapatkan masalah dual dengan memaksimalkan persamaan (2.13) dengan memperhatikan kendala sebagai berikut:

$$\alpha_i \geq 0, i = 1, 2, \dots, N$$

$$\sum_{i=1}^N \alpha_i t_i = 0$$

dimana,  $k(x_i, x_j)$  adalah fungsi kernel linier.

Bentuk optimasi diatas harus memenuhi kondisi *Karush-Kuhn-Tucker* (KKT) sebagai berikut:

$$\alpha_i \geq 0 \quad (2.14)$$

$$t_i (w^T x_i + b) - 1 \geq 0 \quad (2.15)$$

$$\alpha_i \{ t_i (w^T x_i + b) - 1 \} \geq 0 \quad (2.16)$$

sehingga, untuk setiap data latih dengan nilai  $\alpha_i > 0$  serta  $t_i (w^T x_i + b) = 1$  disebut dengan *support vector*. Hasil dari fungsi keputusan hanya dipengaruhi oleh *support*

vector.

Selanjutnya, menentukan kelas dari data uji berdasarkan nilai fungsi keputusan seperti pada persamaan (2.17):

$$y(x) = \sum_{i=1}^{N_S} \alpha_i t_i x_i^T x + b \quad (2.17)$$

dimana,  $x_i$  merupakan *support vector* dan  $N_S$  merupakan jumlah *support vector*.

### 2.8.2. SVM pada Data *Non-Linear Separable*

Data yang tidak dapat diklasifikasikan secara linier, maka digunakan *soft margin* dengan menambahkan variabel *slack*  $\xi_i \geq 0$ ,  $i = 1, 2, \dots, n$ . Sehingga bentuk primal dari masalah optimasi pada data *non-linear* yaitu:

$$\begin{aligned} \arg \min_{w,b,\xi} & \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^N \xi_i \right) \\ \text{s.t. } & t_i (w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, N \end{aligned}$$

dimana  $C$  adalah parameter yang mengontrol batas antara kesalahan klasifikasi dan memaksimalkan margin.

Selain menambahkan variabel *slack*, perlu dilakukan transformasi data ke dimensi yang lebih tinggi. Hal ini dilakukan dengan memanfaatkan fungsi kernel sehingga data dapat dipisahkan secara linier pada dimensi yang lebih tinggi. Fungsi kernel memetakan data  $x_i$  dengan menggunakan fungsi  $\phi(x_i)$ , sehingga persamaan model linier yang digunakan sebagai *hyperplane* seperti pada

persamaan (2.18) (Bishop and Nasrabadi, 2006):

$$y(x) = w^T \phi(x) + b \quad (2.18)$$

akibatnya, masalah optimasi dari *soft margin* menjadi:

$$\arg \min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (2.19)$$

dengan kendala,

$$t_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (2.20)$$

Selanjutnya, meminimumkan persamaan (2.19) dengan kendala (2.20) dengan  $\xi_i \geq 0$ . Fungsi *Lagrange* yang digunakan seperti pada persamaan (2.21):

$$L(w, b, \xi, \alpha, \mu) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i + \sum_{i=1}^N \alpha_i \{1 - \xi_i - t_i(w^T \phi(x_i) + b)\} - \sum_{i=1}^N \mu_i \xi_i \quad (2.21)$$

dimana,  $\{\alpha_i \geq 0\}$  dan  $\{\mu_i \geq 0\}$  merupakan *Lagrange multiplier*.

Diberikan kondisi KKT sebagai berikut:

$$\alpha_i \geq 0 \quad (2.22)$$

$$t_i(w^T \phi(x_i) + b) - 1 + \xi_i \geq 0 \quad (2.23)$$

$$\alpha_i \{t_i(w^T \phi(x_i) + b) - 1 + \xi_i\} = 0 \quad (2.24)$$

$$\mu_i \geq 0 \quad (2.25)$$

$$\xi_i \geq 0 \quad (2.26)$$

$$\mu_i \xi_i = 0 \quad (2.27)$$

dimana,  $i = 1, 2, \dots, N$ .

Selanjutnya, mengoptimalkan nilai  $w, b$  dan  $\{\xi_i\}$  dengan menggunakan definisi pada persamaan (2.18), sehingga diperoleh:

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i t_i \phi(x_i) \quad (2.28)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i t_i = 0 \quad (2.29)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \alpha_i = C - \mu_i \quad (2.30)$$

Eliminasikan  $w, b$  dan  $\{\xi_i\}$  pada fungsi *Lagrange* dengan menggunakan persamaan (2.28), (2.29) dan (2.30), sehingga didapatkan masalah dualnya sebagai berikut:

$$\max \tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j (\phi(x_i)^T \phi(x_j)) \quad (2.31)$$

dengan kendala,

$$0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^N \alpha_i t_i = 0$$

Masalah dual pada persamaan (2.31) memenuhi kondisi KKT (2.22) sampai dengan (2.27).

Menentukan nilai bias pada persamaan (2.18) perlu diperhatikan bahwa *support vector* yang mana  $0 < \alpha_i < C$  dengan  $\xi_i = 0$  sehingga  $t_i(w^T \phi(x_i) + b) = 1$  dan memenuhi persamaan (2.32).

$$t_i \left( \sum_{j \in S} \alpha_j t_j K(x_i, x_j) + b \right) = 1 \quad (2.32)$$

dimana,  $S$  adalah himpunan indeks dari *support vector*.

Selanjutnya, menentukan kelas dari data uji berdasarkan nilai fungsi keputusan seperti pada persamaan (2.33):

$$y(x) = \sum_{i=1}^{N_S} \alpha_i t_i K(x_i, x_j) + b \quad (2.33)$$

dimana,  $x_i$  adalah *support vector* dan  $N_S$  adalah jumlah *support vector*.

### 2.8.3. Kernel

*Separating hyperplane* pada SVM tidak selalu dapat memisahkan semua data, meskipun *slack* variabel telah ditambahkan. Metode kernel dapat digunakan untuk memungkinkan SVM dapat diterapkan pada data-data yang terpisah secara *non linear* (Pande et al., 2023). Kernel berfungsi untuk memetakan data-data ke

ruang dimensi yang lebih tinggi dengan fungsi pemetaan  $\phi(x)$ , dengan perkalian fungsinya yaitu

$$(\phi(x_i) \cdot \phi(x_j)) = K(x_i, x_j)$$

Terdapat beberapa kernel yang digunakan dalam SVM, diantaranya

1. Kernel *Linear*

$$K(x_i, x_j) = x_i^T x_j \quad (2.34)$$

2. Kernel *Polynomial*

$$K(x_i, x_j) = (\gamma x_i^T \cdot x_j + r)^d, \gamma > 0 \quad (2.35)$$

3. Kernel *Radial Basis Function* (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (2.36)$$

## 2.9. Evaluasi Sistem

Semua kumpulan data harus dapat diklasifikasikan secara akurat oleh sistem klasifikasi, akan tetapi sangat sedikit sistem yang dapat melakukan hingga mencapai 100% nilai akurasi. Hal ini disebabkan karena tidak semua sistem yang dibuat selalu berjalan dengan benar. Seringkali terdapat kesalahan pada sistem tersebut. Evaluasi sistem berguna untuk menentukan seberapa efektif metode tersebut untuk mengategorikan kumpulan data tertentu atau seberapa baik sistem tersebut telah dibangun (Townsend, 1971).

*Confusion matrix* merupakan metode yang seringkali digunakan untuk melakukan evaluasi sistem klasifikasi. Definisinya secara umum adalah sebuah *matrix* yang digunakan untuk membandingkan hasil klasifikasi dengan data asli. *Confusion matrix* dapat direpresentasikan dalam bentuk tabel seperti pada Tabel (2.1) (Xu et al., 2020).

Tabel 2.1 Tabel *confusion matrix*

		Nilai Aktual	
		Sentimen Positif	Sentimen Negatif
Nilai Prediksi	Sentimen Positif	<i>True Positive</i>	<i>False Positive</i>
	Sentimen Negatif	<i>False Negative</i>	<i>True Negative</i>

Terdapat 4 jenis data dalam *confusion matrix* yaitu, *true positive*, *false positive*, *true negative*, dan *false negative*. *True positive* merupakan banyaknya data yang termasuk dalam kelas positif yang kemudian diklasifikasikan sebagai kelas positif. Berbeda dengan *true positive*, *True negative* merupakan banyaknya data yang merupakan kelas negatif yang kemudian diklasifikasikan sebagai kelas negatif. Hasil klasifikasi yang benar termasuk kedalam *true positive* dan *true negative*. *False positive* merupakan data yang sebenarnya merupakan kelas negatif akan tetapi diklasifikasikan sebagai kelas positif. Sedangkan data yang sebenarnya kelas positif akan tetapi diklasifikasikan sebagai kelas negatif disebut dengan *false negative* (Chicco et al., 2021).

Hasil evaluasi dari *confusion matrix* terhadap sistem klasifikasi sentimen yang digunakan ada 3 jenis yaitu, akurasi, *recall*, dan presisi. Akurasi menghasilkan persentase yang merupakan perbandingan antara banyaknya data yang diklasifikasikan secara benar yang dibagi oleh jumlah seluruh data.

Sedangkan perbandingan antara banyaknya *true positive* dibagi oleh jumlah semua data yang diklasifikasikan positif disebut sebagai presisi. *Recall* merupakan perbandingan antara banyaknya *true positive* dibagi oleh semua data yang berkategori positif (Hong and Oh, 2021).

Formula yang digunakan dalam menghitung akurasi, *recall*, dan presisi seperti pada persamaan (2.37), (2.38), dan (2.39).

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.37)$$

$$\text{Recall} = \frac{TP}{FN + TP} \times 100\% \quad (2.38)$$

$$\text{Presisi} = \frac{TP}{FP + TP} \times 100\% \quad (2.39)$$

Dimana,

$TP = \text{True Positive}$

$TN = \text{True Negative}$

$FP = \text{False Positive}$

$FN = \text{False Negative}$

## 2.10. Pemimpin dalam Perspektif Islam

Sepanjang sejarah peradaban islam, sudah tidak asing lagi dengan istilah pemimpin. Dalam islam, pemimpin disebut sebagai *khalifah*, sebagaimana tercantum pada Q.S. Al-Baqarah ayat 30,

وَإِذْ قَالَ رَبُّكَ لِلْمَلٰئِكَةِ اِنِّيْ جَاعِلٌ فِي الْاَرْضِ خَلِيْفَةًۭۙ قَالُوْۤا اَتَجْعَلُ فِيْهَا مَنْ يُفْسِدُ فِيْهَا وَيَسْفِكُ  
الدِّمَآءَۙ وَنَحْنُ نُسَبِّحُ بِحَمْدِكَ وَنُقَدِّسُ لَكَۙ قَالَ اِنِّيْۤ اَعْلَمُ مَا لَا تَعْلَمُوْنَ ﴿ۓ﴾



artinya: ”(Ingatlah) ketika Tuhanmu berfirman kepada para malaikat, “Aku hendak menjadikan khalifah di bumi.” Mereka berkata, “Apakah Engkau hendak menjadikan orang yang merusak dan menumpahkan darah di sana, sedangkan kami bertasbih memuji-Mu dan menyucikan nama-Mu?” Dia berfirman, “Sesungguhnya Aku mengetahui apa yang tidak kamu ketahui.””

Pada masa-masa jahiliyah, Allah SWT mengutus nabi dan rasul sebagai rahmat bagi umat manusia. Karena pada masa tersebut, manusia masih menyembah berhala, oleh karenanya tujuan utama nabi dan rasul diutus oleh Allah SWT yaitu untuk memimpin umat manusia kembali ke jalan yang benar. Sebagaimana Allah SWT berfirman dal Q.S. At-Taubah ayat 128,

لَقَدْ جَاءَكُمْ رَسُولٌ مِّنْ أَنْفُسِكُمْ عَزِيزٌ عَلَيْهِ مَا عَدْتُمْ حَرِيصٌ عَلَيْكُمْ بِالْمُؤْمِنِينَ رَءُوفٌ رَّحِيمٌ ﴿١٢٨﴾

artinya: ”Sungguh, benar-benar telah datang kepadamu seorang rasul dari kaummu sendiri. Berat terasa olehnya penderitaan yang kamu alami, sangat menginginkan (keimanan dan keselamatan) bagimu, dan (bersikap) penyantun dan penyayang terhadap orang-orang mukmin.”

Sebagai seorang khalifah, seorang Nabi dan rasul yang diutus Allah SWT mempunyai 4 sifat wajib. Sifat wajib tersebut juga harus ada dalam diri seorang pemimpin. Sifat wajib bagi nabi dan rasul yaitu, *shidiq*, *amanah*, *tabligh*, dan *fathonah*.

### 2.10.1. *Shidiq*

Shidiq merupakan sifat wajib yang memiliki arti benar dan jujur. Hal ini memiliki arti bahwa perkataan nabi dan rasul adalah benar, baik ketika

menyampaikan wahyu dari Allah SWT maupun ketika mengeluarkan perkataan duniawi. Seorang pemimpin juga harus bersifat jujur, agar apa yang disampaikan dapat dipercaya oleh masyarakat. Sifat shidiq telah tercantum dalam Q.S. Maryam ayat 41,

وَأَذْكُرُ فِي الْكِتَابِ إِبْرَاهِيمَ ۗ إِنَّهُ كَانَ صِدِّيقًا نَبِيًّا ﴿٤١﴾

artinya: "Ceritakanlah (Nabi Muhammad, kisah) Ibrahim di dalam Kitab (Al-Qur'an)! Sesungguhnya dia adalah seorang yang sangat benar dan membenarkan lagi seorang nabi."

### 2.10.2. Amanah

Sifat kedua yang wajib dimiliki seorang nabi dan rasul yaitu amanah. Amanah artinya dapat dipercaya. Seorang pemimpin harus bisa amanah, karena segala sesuatu yang diperbuatnya kelak harus dipertanggung jawabkan. Sifat ini seperti pada Q.S. An-Nisa ayat 58,

﴿ إِنَّ اللَّهَ يَأْمُرُكُمْ أَنْ تُؤَدُّوا الْأَمَانَاتِ إِلَىٰ أَهْلِهَا وَإِذَا حَكَمْتُمْ بَيْنَ النَّاسِ أَنْ تَحْكُمُوا بِالْعَدْلِ إِنَّ اللَّهَ نِعِمَّا يَعِظُكُمْ بِهِ إِنَّ اللَّهَ كَانَ سَمِيعًا بَصِيرًا ﴾ ﴿٥٨﴾

artinya: "Sesungguhnya Allah menyuruh kamu menyampaikan amanah kepada pemiliknya. Apabila kamu menetapkan hukum di antara manusia, hendaklah kamu tetapkan secara adil. Sesungguhnya Allah memberi pengajaran yang paling baik kepadamu. Sesungguhnya Allah Maha Mendengar lagi Maha Melihat."

### 2.10.3. *Tabligh*

Sifat ketiga yaitu *tabligh* atau menyampaikan. Tugas utama seorang pemimpin yaitu menyampaikan hal-hal yang baik dan benar. Sebagaimana disebutkan pada Q.S. Al-Maidah ayat 67,

﴿ يَا أَيُّهَا الرَّسُولُ بَلِّغْ مَا أُنزِلَ إِلَيْكَ مِنْ رَبِّكَ وَإِنْ لَمْ تَفْعَلْ فَمَا بَلَّغْتَ رِسَالَتَهُ وَاللَّهُ يَعْصِمُكَ مِنَ النَّاسِ إِنَّ اللَّهَ لَا يَهْدِي الْقَوْمَ الْكَافِرِينَ ﴾

artinya: "Wahai Rasul, sampaikanlah apa yang diturunkan Tuhanmu kepadamu. Jika engkau tidak melakukan (apa yang diperintahkan itu), berarti engkau tidak menyampaikan risalah-Nya. Allah menjaga engkau dari (gangguan) manusia. Sesungguhnya Allah tidak memberi petunjuk kepada kaum yang kafir."

### 2.10.4. *Fathonah*

Sifat wajib yang terakhir adalah *fatonah*. *Fatonah* artinya pintar, cerdas, dan cerdas. Seorang pemimpin haruslah berakal, agar apa yang disampaikannya dapat dipertanggung jawabkan. Pemimpin juga harus cerdas dalam mengambil segala keputusan dan bisa membedakan antara yang benar dan salah. Seperti dalam Q.S. Asy-Syuara ayat 23 ketika Fir'aun bertanya kepada Nabi Musa,

قَالَ فِرْعَوْنُ وَمَا رَبُّ الْعَالَمِينَ

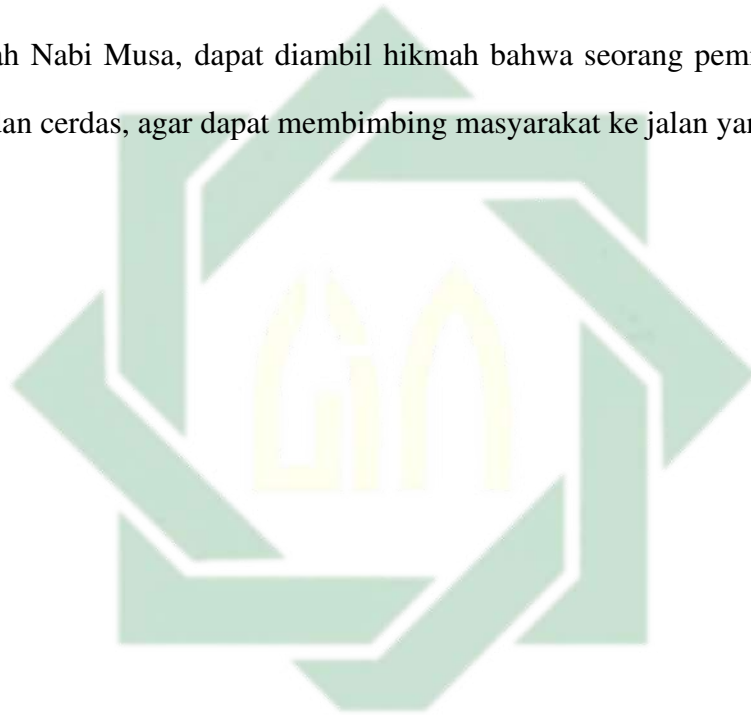
artinya: "Fir'aun berkata, "Siapa Tuhan semesta alam itu?""

Kemudian Nabi Musa menjawab dengan kecerdasannya pada Q.S. Asy-Syuara ayat 26:

قَالَ رَبُّكُمْ وَرَبُّ آبَائِكُمُ الْأَوَّلِينَ

artinya: "Dia (Musa) berkata, "(Dia) Tuhanmu dan Tuhan nenek moyangmu terdahulu."

Dari kisah Nabi Musa, dapat diambil hikmah bahwa seorang pemimpin haruslah berakal dan cerdas, agar dapat membimbing masyarakat ke jalan yang benar.



UIN SUNAN AMPEL  
S U R A B A Y A

## BAB III

### METODE PENELITIAN

#### 3.1. Jenis Penelitian

Penelitian ini termasuk dalam penelitian kuantitatif. Analisis terhadap sentimen dilakukan setelah mendapatkan klasifikasi model tertinggi dan akurasi pengujian. Penelitian ini termasuk dalam penelitian analitik dengan menggunakan data sekunder yang dikumpulkan tanpa melakukan penelitian lapangan secara langsung.

#### 3.2. Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder berupa sentimen pengguna twitter terhadap calon Prabowo Subianto. Data diperoleh menggunakan Twitter API pada tahun 2022. *Crawling* data digunakan untuk mengumpulkan data dengan menggunakan kata kunci tertentu. Pada penelitian ini menggunakan kata "Prabowo" dan "Prabowo Subianto". Data yang dikumpulkan adalah data teks berupa *tweets* dari berbagai akun twitter. Banyaknya data yang diperoleh yaitu sebanyak 34906 *tweets* dari 1 Januari sampai 31 Desember 2022.

#### 3.3. Tahapan Penelitian

Tahapan pada penelitian ini dijelaskan seperti pada Gambar 3.1 yang bertujuan agar memperjelas jalannya penelitian. Terdapat 4 tahapan penting yaitu, *pre-processing text*, ekstraksi fitur, klasifikasi, dan visualisasi.



**Gambar 3.1 Diagram alir**

Adapun beberapa tahapan pada penelitian ini diantaranya:

### 1. *Crawling* Data Twitter

Tahapan awal penelitian yaitu pengumpulan data dengan cara *crawling* data twitter dengan menginputkan Prabowo Subianto untuk menjadi kata kunci yang akan dicari. Hasil *crawling* disimpan dalam format .csv

### 2. Pre-Processing

Data yang telah disimpan kemudian diolah dengan 4 tahapan yaitu, *clean text*, *case folding*, *deleting stopward*, dan *tokenisasi*.

#### (a) *Clean Text*

*Tweet* yang telah disimpan merupakan data sekunder yang masih terdapat banyak kata atau simbol yang tidak memiliki makna, seperti

emoji, tanda baca, angka dan url. Emoji, tanda baca, angka dan url yang ada pada tweet dihapuskan karena tidak bermakna.

(b) *Case Folding*

Kata-kata didalam *tweet* diubah menjadi huruf kecil agar memudahkan dalam mengolah kata untuk ekstraksi fitur.

(c) *Deleting Stopward*

*Deleting stopward* atau penghapusan kata-kata yang tidak bermakna merupakan tahap selanjutnya dalam *pre-processing*. Pada setiap tweet cenderung mengandung *stopward* atau kata yang tidak bermakna. Contohnya, yang, dan, jika, dia, yang merupakan kata penghubung dan kata ganti.

(d) Tokenisasi

Tokenisasi berperan dalam membagi dokumen teks yang berupa paragraf atau kalimat menjadi bagian kata per kata.

### 3. Pelabelan Data

Pelabelan kata dan dokumen dilakukan dengan memanfaatkan function *VaderSentimentAnalysis* pada software MATLAB. Terdapat dua label kelas, yaitu positif dan negatif

### 4. Penghapusan Kata Netral

Pada setiap sentimen yang telah terlabel sebagai positif dan negatif, dilakukan penghapusan kata netral pada setiap sentimen. Penghapusan dilakukan untuk meminimalkan fitur.

### 5. Ekstraksi Fitur

Sentimen yang sudah diberi label kemudian dilakukan pembobotan pada setiap kata dengan TF-IDF menggunakan persamaan (2.1) dan (2.2).

#### 6. Klasifikasi dengan SVM

Hasil dari pembobotan kata yang sudah berupa matriks dilakukan klasifikasi dengan SVM menggunakan persamaan (2.17) dan (2.33).

#### 7. Perhitungan Evaluasi Sistem

Setelah didapatkan hasil klasifikasi, maka dilakukan perhitungan evaluasi sistem untuk mengetahui seberapa baik sistem tersebut bekerja menggunakan *confusion matrix* dengan mengacu pada akurasi, *recall* dan presisi menggunakan persamaan (2.37) sampai dengan persamaan (2.39)

#### 8. Visualisasi dan Analisis Hasil

Setelah hasil didapatkan dari proses klasifikasi SVM, dilakukan pembuatan *wordcloud* untuk analisis dan visualisasi untuk menentukan kosakata yang sesuai dengan informasi yang ingin ditampilkan.

#### 9. Skenario Uji Coba

Data yang akan digunakan untuk klasifikasi, dilakukan uji coba dengan *k-fold cross validation*. Penelitian oleh Hidayat tentang analisis sentimen menggunakan uji coba *k-fold cross validation* dengan  $k=3$  menghasilkan akurasi cukup baik yaitu sebesar 85% [Hidayat and Sulistiyono \(2022\)](#).

Penelitian lain oleh Hanmastiana tentang analisis sentimen menggunakan uji coba *k-fold* dengan  $k$  sebanyak 3,5,7, dan 9. Hasil yang diperoleh Hanmastiana yaitu akurasi terbaik pada  $k=7$  dengan akurasi sebesar 76,93% [Hanmastiana et al. \(2021\)](#). Analisis sentimen menggunakan uji coba *k-fold cross validation* juga dilakukan oleh Fazrin dengan  $k=10$ . Penelitian Fazrin



menghasilkan akurasi yang baik yaitu sebesar 85.22% [Fazrin et al. \(2022\)](#). Berdasarkan beberapa uji coba yang telah dilakukan pada penelitian sebelumnya, maka jumlah  $k$  yang digunakan pada penelitian ini yaitu 3, 5, 7, 9, 10 dan 15.



UIN SUNAN AMPEL  
S U R A B A Y A

## BAB IV

### HASIL DAN PEMBAHASAN

Hasil yang didapatkan dari penelitian ini dijelaskan pada bab ini dari proses yang pertama sampai proses yang terakhir.

#### 4.1. *Crawling* Data Twitter

Crawling data pada media sosial twitter merupakan langkah pertama yang dilakukan dalam penelitian ini. Pada tahap ini, pencarian data twitter dilakukan yang berhubungan dengan tokoh publik yang juga sedang menjabat sebagai menteri pertahanan Indonesia yaitu Prabowo Subianto. Crawling dilakukan dengan menggunakan kata kunci Prabowo dan Prabowo Subianto, sehingga semua *tweet* yang berhubungan dengan Prabowo dan Prabowo Subianto akan termuat dalam pencarian ini.

Pencarian data twitter yang berhubungan dengan menteri pertahanan Indonesia, didapatkan data sebanyak 34906 *tweet* yang memuat data twitter selama tahun 2022. Data tersebut merupakan data yang terdiri dari dua kategori yaitu, *tweet* positif dan negatif. Pada tabel [4.1](#) merupakan sampel hasil *crawling* data twitter yang berhubungan dengan Prabowo Subianto.

Tabel 4.1 *Sample data sentimen Prabowo Subianto*

<i>Time</i>	<i>Tweet</i>
2022-09-08	Betul banget, Prabowo Subianto betul keren profesionalnya, Saya dukung dan semoga dia bisa memimpin Indonesia ini.
2022-08-08	Riwayat buruk keikutsertaan Prabowo Subianto dalam dua pemilihan presiden membuat ia tak layak menjadi calon pemimpin negeri ini. Gerindra bisa memilih calon lain yang mumpuni.

Data yang didapatkan merupakan sentimen masyarakat terhadap Prabowo Subianto pada tahun 2022. Pada Gambar 4.1, terlihat bahwa pada bulan Desember 2022 topik mengenai Prabowo paling banyak dibicarakan. Pada bulan tersebut, masyarakat tertarik untuk membahas pertemuan Prabowo sebagai Menteri Pertahanan dengan YM Sebastien Lecornu, Menteri Angkatan Bersenjata Republik Prancis. Pertemuan tersebut membahas kerja sama pertahanan antara Indonesia dengan Prancis dalam hal pendidikan dan pelatihan, perkembangan terbaru kerja sama kawasan Indo-Pasifik, serta modernisasi alutsista TNI (Biro Humas Setjen Kemhan, 2022b).



Gambar 4.1 Plot Sebaran Sentimen

Sentimen masyarakat terhadap Prabowo cenderung naik dari bulan Januari sampai Desember 2022.

#### 4.2. Pre-Processing Text

Data twitter yang didapatkan selanjutnya dilakukan pre-proses teks. Pada pre-proses teks ini terdiri dari beberapa tahapan diantaranya, *clean text*, *case folding*, *deleting stopword*, dan tokenisasi.

Tahap pertama yang dilakukan yaitu *clean text*. *Clean text* menghapus semua html, *username*, *hashtag*, web, emoticon, dan simbol-simbol karena hal-hal tersebut tidak memiliki makna dalam proses pengklasifikasian dokumen. Angka dan tanda baca juga dilakukan penghapusan pada tahapan ini. Pada tabel 4.2 merupakan hasil dari tahapan *clean text*.

**Tabel 4.2 Contoh tahapan *clean text***

<i>Tweet setelah clean text</i>
Betul banget Prabowo Subianto betul keren profesionalnya Saya dukung dan semoga dia bisa memimpin Indonesia ini
Riwayat buruk keikutsertaan Prabowo Subianto dalam dua pemilihan presiden membuat ia tak layak menjadi calon pemimpin negeri ini. Gerindra bisa memilih calon lain yang mumpuni

Tahapan selanjutnya adalah *case folding*. Pada tahap ini, semua huruf yang ada pada dokumen diubah ke dalam bentuk yang sama, yaitu *lower case* agar memudahkan untuk pengklasifikasian dokumen. Pada tabel 4.3 merupakan hasil dari *case folding*.

**Tabel 4.3 Contoh setelah dilakukan *case folding***

Tweet setelah <i>case folding</i>
betul banget prabowo subianto betul keren profesionalnya saya dukung dan semoga dia bisa memimpin indonesia ini
riwayat buruk keikutsertaan prabowo subianto dalam dua pemilihan presiden membuat ia tak layak menjadi calon pemimpin negeri ini gerindra bisa memilih calon lain yang mumpuni

Tahapan selanjutnya yaitu *deleting stopword*. Pada tahap ini, semua kata yang terkandung dalam dokumen *stopword* dihapus. Kata-kata yang mengandung *stopword* tidak dapat dijadikan sebagai ciri pengklasifikasian pada dokumen tertentu, oleh karena itu kata-kata yang terkandung dalam *stopword* dihapus. Pada tabel 4.4 merupakan hasil dari *deleting stopword*.

**Tabel 4.4 Contoh setelah dilakukan *deleting stopword***

Tweet setelah <i>deleting stopword</i>
banget prabowo subianto keren profesional dukung semoga dia memimpin indonesia
riwayat buruk ikutserta prabowo subianto pemilihan presiden tak layak calon pemimpin negeri gerindra memilih calon mumpuni

Tahapan terakhir dalam pre-proses teks yaitu tokenisasi. Pada tahap tokenisasi, dilakukan partisi pada setiap kata dalam dokumen. Dokumen yang awalnya merupakan kalimat, diubah menjadi kata per kata. Hal ini dilakukan agar mempermudah pembobotan untuk masing-masing kata. Pada tabel 4.5 merupakan hasil dari tokenisasi.

**Tabel 4.5 Tokenisasi Kata**

<i>Tweet</i> setelah proses tokenisasi
banget - prabowo - subianto - keren - profesional - dukung - semoga - memimpin - indonesia
riwayat - buruk - ikutserta - prabowo - subianto - pemilihan - presiden - tak - layak - calon - pemimpin - negeri - gerindra - memilih - calon - mumpuni

### 4.3. Pelabelan Data dan Seleksi Fitur

Langkah selanjutnya yaitu pelabelan data dan seleksi fitur. Pada tahap ini, data yang telah dipisahkan menjadi beberapa token diberi label pada setiap dokumennya. Data yang diperoleh diberi label secara berbeda. Terdapat 2 label yaitu positif dan negatif. Jumlah dokumen pada penelitian ini dihitung dalam jumlah yang sangat banyak, oleh karena itu pelabelan dokumen dilakukan menggunakan *VaderSentimentAnalysis* yang merupakan *function* dari software MATLAB. Hal ini dapat memudahkan dalam pelabelan dan tidak membutuhkan waktu yang lama untuk pelabelan dokumen dalam jumlah yang sangat besar.

Pelabelan menggunakan *VaderSentimentAnalysis* menghasilkan nilai sentimen pada setiap dokumen. Dokumen dengan nilai  $> 0$  merupakan sentimen positif, sedangkan untuk dokumen dengan nilai  $< 0$  merupakan sentimen negatif seperti pada tabel .

**Tabel 4.6 Pelabelan Dokumen**

Dokumen	Nilai Sentimen	Label Kelas
banget - prabowo - subianto - keren - profesional - dukung - semoga - dia - memimpin - indonesia	0.7884	Positif
riwayat - buruk - ikutserta - prabowo - subianto - pemilihan - presiden - tak - layak - calon - pemimpin - negeri - gerindra - memilih - calon - mumpuni	-0.5423	Negatif

Selanjutnya dilakukan seleksi fitur. Banyaknya fitur pada data penelitian membuat sistem yang akan dibuat menjadi berat dan membutuhkan waktu yang lama. Fitur-fitur pada penelitian ini dilakukan pelabelan kata per kata seperti pada Tabel 4.7.

**Tabel 4.7 Pelabelan Kata**

Fitur	Nilai Fitur	Label Fitur
keren	0.3182	Positif
buruk	-0.5423	Negatif
dukung	0.4019	Positif
saran	0	Netral
pinang	0	Netral
pemimpin	0	Netral

Fitur atau kata yang memiliki label netral dihapuskan sehingga fitur yang tersisa yaitu kata positif dan negatif yang berjumlah 4874 fitur. fitur-fitur yang telah diseleksi kemudian digunakan dalam klasifikasi sistem.

#### **4.4. Term Frequency Inverse Document Frequency (TF-IDF)**

Pembobotan kata dalam penelitian ini menggunakan perhitungan TF-IDF. Perhitungan bobot ini digunakan untuk menentukan hasil sentimen. Setelah dilakukan tokenisasi, setiap kata yang terdapat pada dokumen dilakukan perhitungan TF seperti pada Tabel 4.8.

Tabel 4.8 Perhitungan *Term Frequency*

Doc.	<i>Term Frequency</i>						
	partai	sandiaga	terbendung	senjata	salah	...	pemerintah
1	1	0	0	0	0	...	0
2	1	0	0	0	0	...	0
3	3	1	0	0	0	...	0
4	0	0	1	0	0	...	0
5	1	0	0	0	0	...	0
6	0	0	0	1	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
34906	0	0	0	0	1	...	1

Seperti pada persamaan 2.1, nilai yang didapatkan dari TF pada tiap-tiap fitur dikalikan dengan nilai IDF yang menghasilkan bobot pada masing-masing fitur. Berikut ini merupakan sampel perhitungan TF-IDF:

$$w_{1,1}^{partai} = 1 \times \log\left(\frac{34906}{2654}\right) = 1,1205 \qquad w_{3,1}^{partai} = 3 \times \log\left(\frac{34906}{2654}\right) = 3,3614$$

$$w_{3,2}^{sandiaga} = 1 \times \log\left(\frac{34906}{218}\right) = 2,2044 \qquad w_{4,3}^{terbendung} = 1 \times \log\left(\frac{34906}{3}\right) = 4,0658$$

$$w_{6,4}^{senjata} = 1 \times \log\left(\frac{34906}{528}\right) = 1,8203 \qquad w_{34906,5}^{salah} = 1 \times \log\left(\frac{34906}{735}\right) = 1,6766$$



Tabel 4.9 Tabel Hasil Perhitungan TF-IDF

Doc.	TF-IDF						
	partai	sandiaga	terbendung	senjata	salah	...	pemerintah
1	1,1205	0	0	0	0	...	0
2	1,1205	0	0	0	0	...	0
3	3,3614	2,2044	0	0	0	...	0
4	0	0	4,0658	0	0	...	0
5	1,1277	0	0	0	0	...	0
6	0	0	0	1,8203	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
34906	0	0	0	0	1,6766	...	1,7327

#### 4.5. Klasifikasi SVM

Proses klasifikasi dilakukan dengan data latih dan data uji pada sentimen masyarakat terhadap Prabowo Subianto menggunakan SVM. Data latih digunakan dalam tahapan *training* yang berguna untuk mendapatkan model sistem. Data uji pada tahapan *testing* digunakan untuk menguji seberapa akurat sistem yang telah didapatkan pada proses *training*. Pada penelitian ini digunakan 34906 data sentimen yang telah diproses dengan seleksi fitur dan pelabelan dokumen. Fitur-fitur tersebut merupakan kata-kata yang terkandung dalam seluruh dokumen yang berlabel positif dan negatif. Fitur yang didapatkan sebanyak 4874 kata. Fitur-fitur tersebut yang digunakan sebagai parameter input dalam klasifikasi SVM.

Data sentimen dilakukan pembagian menjadi dua bagian, yaitu data uji dan data latih. Pada penelitian ini menggunakan *k-fold cross validation* dalam

pembagian data latih dan data uji. *K-fold* yang digunakan sebanyak 3,5,7,9,10 dan 15. Terdapat 2 kelas pada proses klasifikasi yang memisahkan data secara biner, yaitu kelas positif dan kelas negatif. Kernel yang digunakan pada penelitian ini adalah kernel *linear*, RBF, dan *polynomial*.

Berikut ini merupakan sampel perhitungan proses klasifikasi SVM menggunakan kernel *polynomial* sebanyak 4 data dengan Tabel 4.10 merupakan sampel data yang digunakan dalam perhitungan beserta dengan label kelasnya.

**Tabel 4.10 Sampel Data Perhitungan Kernel**

$x_1$	$x_2$	$x_3$	$x_4$	$y$
1,0905	0	0	0	1
3,0992	1,0331	0	0	1
0	0	1,4325	0	-1
0	0	0	1,1012	-1

Beberapa parameter *input* yang digunakan untuk membantu dalam proses klasifikasi SVM diantaranya,

Sampel data *training* :  $x = x_1, x_2, x_3, x_4$

Label data *training* :  $y = y_1, y_2, y_3, y_4 \forall 0 < y < 0$

Jenis kernel : *polynomial*

Parameter kernel :  $\gamma = 1.8, r = 0.0, d = 3$

Langkah selanjutnya yaitu melakukan perhitungan matriks kernel dengan memaksimalkan matriks data *training*  $X$ . Matriks kernel didapatkan dengan perkalian matriks *training* dengan *transpose* dari *support vector*. *Support vector*

didapatkan dari model SVM yang terbentuk yang ditunjukkan pada Tabel 4.11.

**Tabel 4.11 Sampel Support Vector**

-0,2612	-0,0528	-0,0066	-0,0662
-0,2612	-0,0528	-0,0066	13,1123
2,7778	-0,0528	-0,0066	-0,0662
-0,2612	-0,0528	-0,0066	-0,0662

Pada persamaan (2.35) terdapat perkalian *dot vector* antara data *training* dengan *transpose support vector* sebagai berikut:

$$X^T \times X = \begin{pmatrix} -0,2612 & -0,2612 & 2,7778 & -0,2612 \\ -0,0528 & -0,0528 & -0,0528 & -0,0528 \\ -0,0066 & -0,0066 & -0,0066 & -0,0066 \\ -0,0662 & 13,1123 & -0,0662 & -0,0662 \end{pmatrix} \times \begin{pmatrix} 1,0905 & 0 & 0 & 0 \\ 3,0992 & 1,0331 & 0 & 0 \\ 0 & 0 & 1,4325 & 0 \\ 0 & 0 & 0 & 1,1012 \end{pmatrix}$$

Hasil perkalian matriks diatas dapat digunakan untuk menghitung nilai matriks kernel  $K$  dengan contoh perhitungan sebagai berikut:

$$x_1^T \cdot x_1 = [-0,2612 \ -0,0528 \ -0,0066 \ -0,0662] \cdot [1,0905 \ 3,0992 \ 0 \ 0] = -0,4484$$

$$\begin{aligned} K(x_1, x_1) &= (1,8(-0,4484) + 0,0)^3 \\ &= -0,5258 \end{aligned}$$

$$x_1^T \cdot x_2 = [-0,2612 \ -0,0528 \ -0,0066 \ -0,0662] \cdot [0 \ 1,0331 \ 0 \ 0] = -0,0545$$

$$\begin{aligned} K(x_1, x_2) &= (1,8(-0,0545) + 0,0)^3 \\ &= -9,4407 \times 10^{-8} \end{aligned}$$

$$x_1^T \cdot x_3 = [-0,2612 \quad -0,0528 \quad -0,0066 \quad -0,0662] \cdot [0 \quad 0 \quad 1,4325 \quad 0] = -0,0094$$

$$\begin{aligned} K(x_1, x_3) &= (1,8(-0,0094) + 0,0)^3 \\ &= -4,8439 \times 10^{-11} \end{aligned}$$

⋮

$$x_4^T \cdot x_4 = [-0,2612 \quad -0,0528 \quad -0,0066 \quad -0,0662] \cdot [0 \quad 0 \quad 0 \quad 1,1012] = -0,0729$$

$$\begin{aligned} K(x_4, x_4) &= (1,8(0,0729) + 0,0)^3 \\ &= -2,2594 \times 10^{-8} \end{aligned}$$

Berdasarkan perhitungan yang dilakukan, diperoleh matriks kernel  $K$  yaitu:

$$K = \begin{pmatrix} -0,5258 & -9,4407 \times 10^{-8} & -4,8439 \times 10^{-11} & -2,2594 \times 10^{-8} \\ -0,5258 & -9,4407 \times 10^{-8} & -4,8439 \times 10^{-11} & 17557,2274 \\ 137,2348 & -9,4407 \times 10^{-8} & -4,8439 \times 10^{-11} & -2,2594 \times 10^{-8} \\ -0,5258 & -9,4407 \times 10^{-8} & -2,2594 \times 10^{-8} & -2,2594 \times 10^{-8} \end{pmatrix}$$

Matriks kernel  $K$  digunakan untuk mencari nilai pengali *lagrange* seperti pada Persamaan [2.31](#), dimana persamaan [2.31](#) merupakan persamaan dualitas yang sulit diselesaikan secara manual, oleh karena itu dibutuhkan *software* yang dapat membantu menyelesaikan persamaan tersebut. Hasil klasifikasi berupa akurasi (Acc), *recall* (Rec), dan presisi (Pre) pada sentimen masyarakat terhadap Prabowo Subianto pada 2022 menggunakan SVM berdasarkan kernel dapat dilihat pada Tabel [4.12](#)

Tabel 4.12 Hasil Evaluasi

Kernel	K-Fold	Hasil Evaluasi		
		Acc (%)	Rec (%)	Pre (%)
<i>Linear</i>	3	94,35	98	95,76
	5	94,47	98,1	95,79
	7	94,54	98,17	95,81
	9	94,68	98,13	95,99
	10	94,72	98,15	96,02
	15	94,67	98,19	95,93
RBF	3	93,6	98,67	94,4
	5	93,91	98,7	94,69
	7	93,9	98,68	94,7
	9	93,96	98,6	94,83
	10	93,93	98,64	94,76
	15	93,92	98,58	94,8
<i>Polynomial</i>	3	94,34	97,65	96,07
	5	60,99	58,55	96,2
	7	94,58	97,59	96,37
	9	94,55	97,57	96,35
	10	94,62	97,6	96,41
	15	94,7	97,62	96,47

Pada Tabel 4.12, proses *learning* yang dilakukan didasarkan pada data sentimen dengan pembagian *k-fold* = 3,5,7,9,10,15 dan menggunakan tiga jenis kernel pada SVM yaitu kernel *linear*, *polynomial*, dan RBF. Hasil *learning* terbaik

dari proses klasifikasi SVM yaitu pada kernel *linear* dengan pembagian 10 *k-fold* yang menghasilkan akurasi sebesar 94,72%, *recall* sebesar 98,15%, dan presisi sebesar 96,02%. Penelitian yang dilakukan oleh Praghakusuma tentang analisis sentimen terhadap KPK menggunakan uji coba kernel pada SVM menghasilkan akurasi terbaik pada kernel *linear* sebesar 83,06% [Praghakusma and Charibaldi \(2021\)](#). Uji coba kernel pada SVM juga dilakukan oleh Muttaqin pada penelitiannya mengenai analisis sentimen aplikasi gojek. Penelitian Muttaqin menunjukkan bahwa kernel *linear* menghasilkan akurasi terbaik yaitu 87,98% [Muttaqin and Kharisudin \(2021\)](#). Kernel *linear* menghasilkan akurasi terbaik karena data teks merupakan data *linear*. Pada penelitian ini, nilai akurasi 94,72% persen termasuk dalam kategori sangat baik.

#### 4.6. Evaluasi Sistem

Hasil terbaik dari masing-masing kernel dilakukan evaluasi dengan menggunakan *confusion matrix*. Evaluasi dilakukan untuk mendapatkan model terbaik berdasarkan analisa hasil dari *confusion matrix*. Terdapat tiga analisa hasil dari *confusion matrix*, yaitu akurasi, *recall*, dan presisi menggunakan persamaan [2.37](#), [2.38](#), dan [2.39](#). Pada Tabel [4.13](#) merupakan hasil dari *confusion matrix* berdasarkan kernel terbaik.

**Tabel 4.13 Hasil Confusion Matrix Kernel Linear dengan  $k = 10$**

		Nilai Aktual	
		Sentimen Positif	Sentimen Negatif
Nilai Prediksi	Sentimen Positif	30535	1267
	Sentimen Negatif	576	2528

Confusion matrix tersebut menghasilkan nilai TP, TN, FP, dan FN yang digunakan untuk menghitung nilai evaluasi, yaitu akurasi, *recall*, dan presisi. Adapun perhitungan nilai akurasi sebagai berikut:

$$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\ &= \frac{30535 + 2528}{30535 + 2528 + 1367 + 576} \times 100\% \\ &= 94,72\% \end{aligned}$$

Hasil dari perhitungan akurasi menunjukkan bahwa banyaknya data yang terklasifikasi benar dengan data aktual oleh sistem yang telah dibuat sangat baik. Selanjutnya dilakukan perhitungan nilai *recall* sebagai berikut:

$$\begin{aligned} \text{recall} &= \frac{TP}{FN + TP} \times 100\% \\ &= \frac{30535}{576 + 30535} \times 100\% \\ &= 98,15\% \end{aligned}$$

Hasil dari perhitungan *recall* menunjukkan banyaknya data yang terklasifikasi benar pada kelas positif. Recall mengukur kinerja sistem dalam pengklasifikasian kelas positif. Selanjutnya dilakukan perhitungan nilai presisi sebagai berikut:

$$\begin{aligned} \text{presisi} &= \frac{TP}{FP + TP} \times 100\% \\ &= \frac{30535}{1267 + 30535} \times 100\% \\ &= 96,02\% \end{aligned}$$

Hasil dari perhitungan presisi menunjukkan banyaknya data yang terklasifikasi benar pada kelas negatif. Presisi mengukur kinerja sistem dalam pengklasifikasian kelas negatif. Pada Tabel 4.14 merupakan hasil dari perhitungan nilai akurasi, *recall*, dan presisi pada kernel *linear* dengan pembagian 10 *k-fold*. Hasil perhitungan tersebut merupakan hasil terbaik pada penelitian ini.

**Tabel 4.14 Hasil Evaluasi Terbaik**

Hasil Evaluasi		
Akurasi	<i>Recall</i>	Presisi
94,72%	98,15%	96,02%

#### 4.7. Visualisasi dan Analisa Hasil

Pada penelitian ini, dilakukan dua tahapan dalam visualisasi dan analisa hasil. Tahapan pertama yaitu visualisasi sentimen menggunakan grafik *wordcloud*. Grafik *wordcloud* merupakan grafik yang menunjukkan kata yang terkandung dalam suatu dokumen. Adapun penentuan ukuran kata dalam *wordcloud* dapat dihitung pada persamaan (4.1):

$$Z_i = \begin{cases} \frac{f_{max}(t_i - t_{min})}{t_{max} - t_{min}}, & t_i > t_{min} \\ 1, & otherwise \end{cases} \quad (4.1)$$

dimana:

$Z_i$  : Nilai ukuran kata ke-i

$f_{max}$  : Nilai ukuran kata terbesar

$f_{min}$  : Nilai ukuran kata terkecil

$t_{max}$  : Jumlah kata terbanyak

$t_{min}$  : Jumlah kata paling sedikit

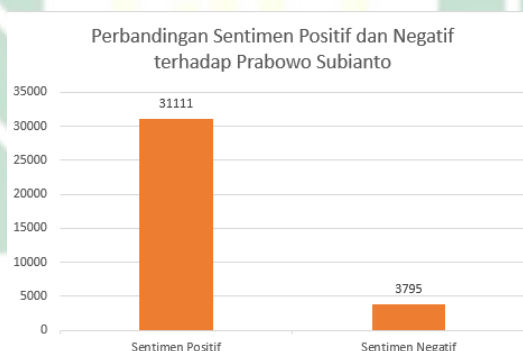


Berikut ini contoh perhitungan *wordcloud* pada kata ”partai”:

$$Z_{\text{partai}} = \frac{10(2645 - 3)}{5393 - 3} \quad (4.2)$$

$$= 4,9017 \quad (4.3)$$

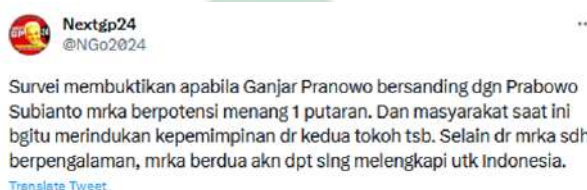
Tahapan kedua yaitu menampilkan perbandingan jumlah sentimen negatif dan positif seperti pada Gambar 4.2. Dapat dilihat pada Gambar 4.2 menunjukkan bahwa sentimen terhadap prabowo lebih banyak positif daripada negatif. Hal ini dapat menunjukkan bahwa selama Prabowo menjabat sebagai Menteri Pertahanan Indonesia pada tahun 2022, sebagian besar masyarakat Indonesia mendukung dan memberikan komentar positifnya terhadap Prabowo.



Gambar 4.2 Grafik Perbandingan Jumlah Sentimen Positif dan Negatif



Selanjutnya, terlihat juga nama Ganjar yang merupakan gubernur Jawa Tengah yang dikabarkan akan mencalonkan diri sebagai presiden Indonesia pada PEMILU 2024 (Tim DetikJateng, 2023). Sebagian besar warga twitter mendukung jika Ganjar menjadi wakil presiden mendampingi Prabowo, akan tetapi ada juga yang tidak setuju. *Tweet* positif yang mendukung Prabowo jika berpasangan dengan Ganjar dapat dilihat pada Gambar (4.5)



**Gambar 4.5 sampel *tweet* positif dengan *keyword* "ganjar"**

Pada Desember 2022, Prabowo berhasil mendatangkan 2 Alat Utama Sistem Senjata (ALUTSISTA) canggih yaitu pesawat interim Falcon 7X dan 8X TNI AU (Biro Humas Setjen Kemhan, 2022b). Hal ini menjadikan kata "alutsista" banyak dibahas dan menunjukkan bahwa kata tersebut dalam penelitian ini mempunyai sentimen yang positif terhadap Prabowo Subianto seperti pada Gambar (4.6) dan (4.7). Menurut lulutami1 dalam cuitan *tweet*nya "Keren, misi Prabowo Subianto untuk memodernisasi alutsista TNI memakai diplomasi pertahanan dengan negara2 sahabat." menyatakan sentimen positifnya terhadap Prabowo. Kayyinov dalam *tweet*nya juga menanggapi positif kinerja Prabowo, "Nyata banget, dari menjabat sebagai Menhan mulai 2019, Prabowo Subianto beri perhatian serius terhadap alutsista TNI".

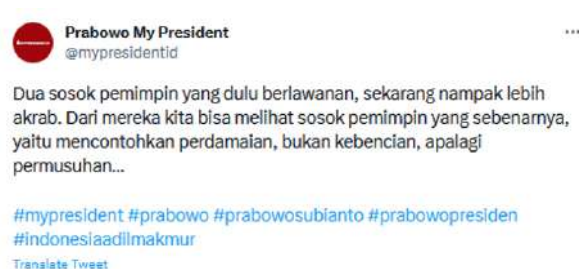


**Gambar 4.6 sampel tweet positif dengan keyword "alutsista"**



**Gambar 4.7 sampel tweet positif dengan keyword "alutsista"**

Selanjutnya, kata "perdamaian" juga disebut pada sentimen positif Prabowo, "Dua sosok pemimpin yang dulu berlawanan, sekarang nampak lebih akrab. Dari mereka kita bisa melihat sosok pemimpin yang sebenarnya, yaitu mencontohkan perdamaian, bukan kebencian, apalagi permusuhan" seperti pada Gambar (4.8). Sentimen positif tersebut membahas pertemuan Prabowo dan presiden Joko Widodo mengenai penyerahan Daftar Isian Pelaksanaan Anggaran (DIPA) Tahun Anggaran 2023 (Biro Humas Setjen Kemhan, 2022a). Hal ini mendapatkan respon positif dari masyarakat twitter.



**Gambar 4.8 sampel tweet positif dengan keyword "perdamaian"**

Kata "teratas" juga memiliki frekuensi yang cukup tinggi, kata ini dibahas dalam sentimen mengenai elektabilitas Prabowo. jpnncom dalam *tweetnya* juga menyebutkan elektabilitas Prabowo teratas dan terus meningkat. Menurut lembaga survei, Survei dan Polling Indonesia (SPIN) pada awal Desember 2022, elektabilitas Prabowo teratas dengan 31,8% (Hermawan, 2022). *Tweet* yang membahas elektabilitas Prabowo teratas dapat dilihat pada Gambar (4.9) dan (4.10).



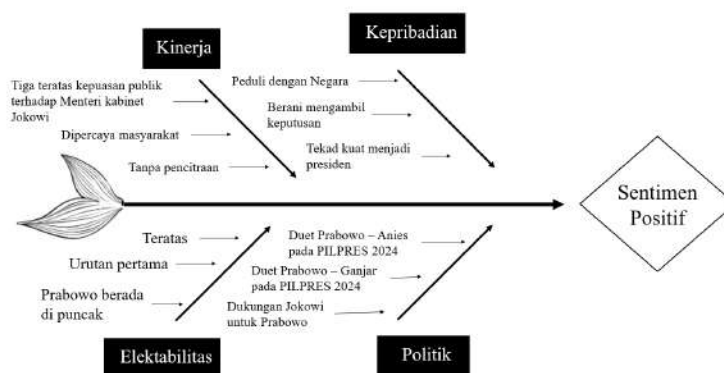
Gambar 4.9 sampel *tweet* positif dengan *keyword* "teratas"



Gambar 4.10 sampel *tweet* positif dengan *keyword* "teratas"

Setelah itu, terlihat juga kata-kata positif yang sering digunakan para pendukung Prabowo seperti terbaik, kuat, favorit, baik, tulus, adil, dan lain sebagainya yang menunjukkan bahwa kata-kata tersebut dalam penelitian ini

mempunyai sentimen yang positif terhadap Prabowo Subianto. Berdasarkan sentimen-sentimen positif yang telah didapatkan, juga dapat dilakukan visualisasi dengan menggunakan *fishbone* diagram seperti pada Gambar (4.11).



**Gambar 4.11 Fishbone Diagram Sentimen Positif**

Diagram *fishbone* untuk sentimen positif di atas menyoroti pendapat masyarakat mengenai kinerja, kepribadian, elektabilitas, dan politik yang berhubungan dengan Prabowo Subianto. Adapun analisis terkait diagram *fishbone* adalah sebagai berikut:

### 1. Kinerja

Masyarakat menilai kinerja Prabowo sangatlah baik dan tanpa pencitraan seperti pada Gambar 4.6 dan 4.7. Menurut survei yang telah dilakukan, Prabowo mendapatkan urutan teratas dalam kepuasan publik terhadap para menteri pada kabinet Jokowi seperti pada Gambar 4.9. Hal ini membuktikan bahwa Prabowo dipercaya oleh masyarakat selama menjabat menjadi Menteri Pertahanan Indonesia.

### 2. Kepribadian

Dalam hal kepribadian, beberapa masyarakat menilai bahwa Prabowo sosok yang peduli dengan Negara. Prabowo juga dinilai berani mengambil

keputusan tanpa adanya pencitraan untuk kepentingan PILPRES. Kinerja Prabowo yang sangat baik dinilai juga menjadi tekad kuat Prabowo untuk menjadi presiden di PILPRES mendatang.

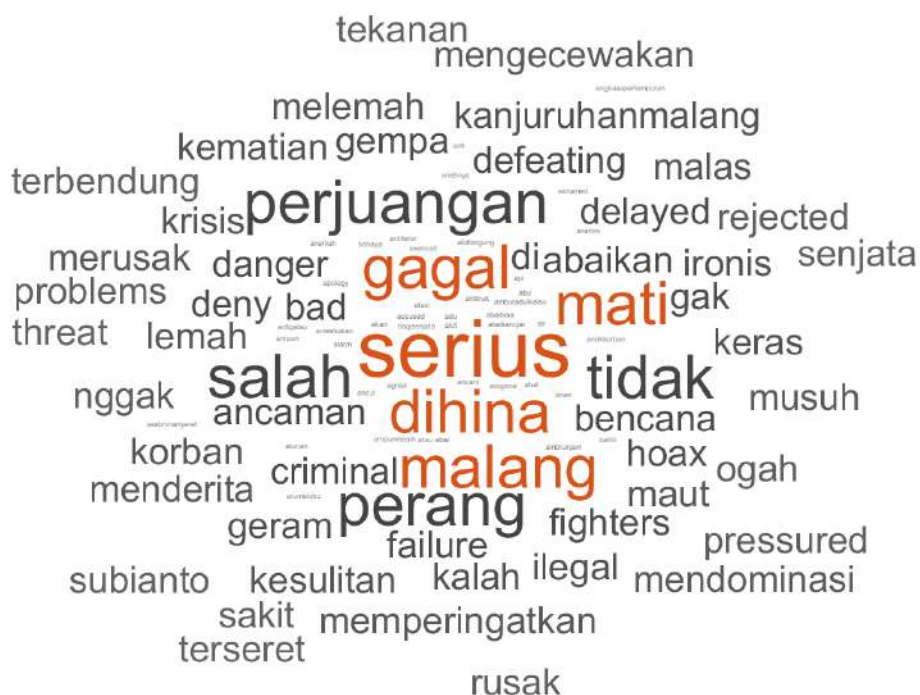
### 3. Elektabilitas

Menurut SPIN, elektabilitas Prabowo teratas dan menguat seperti pada Gambar 4.10. Survei Indikator Politik Indonesia juga melakukan survei elektabilitas dan hasilnya adalah Prabowo menempati urutan pertama.

### 4. Politik

Dalam segi politik, Prabowo dipasangkan dengan Anies pada PILPRES 2024. Ada juga masyarakat yang ingin Prabowo berpasangan dengan Ganjar pada PILPRES 2024 seperti pada Gambar 4.5. Jokowi mendukung Prabowo jika Prabowo maju dalam PILPRES 2024.

Berdasarkan hasil analisis dan diagram *fishbone* sentimen positif yang dilakukan, pada tahun 2022 masyarakat cenderung menilai Prabowo secara positif berdasarkan kinerja selama menjabat menjadi Menteri Pertahanan, seperti keberhasilan Prabowo dalam pengembangan ALUTSISTA TNI dengan mendatangkan dua pesawat tempur dari Prancis (Biro Humas Setjen Kemhan, 2022b). Prabowo juga mendapatkan urutan teratas dalam survei kepuasan publik terhadap para menteri di kabinet Jokowi (Egeham, 2022). Adapun elektabilitas Prabowo di akhir tahun 2022 dalam sentimen positif mengalami kenaikan dan menempati urutan teratas (Hermawan, 2022). Hasil dari penelitian ini hanya memaparkan sentimen positif Prabowo dengan waktu yang terbatas yaitu satu tahun, sedangkan Prabowo menjabat sebagai Menteri Pertahanan selama 5 tahun.



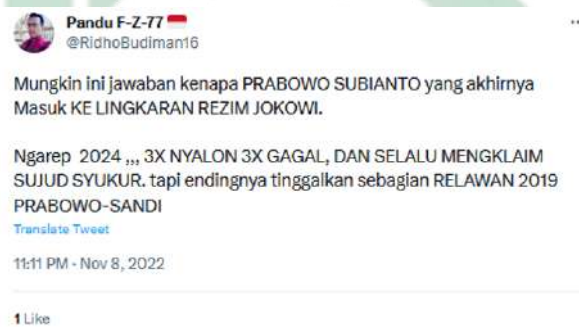
**Gambar 4.12 Sentimen Negatif terhadap Prabowo Subianto**

Selanjutnya, pada Gambar 4.12 menunjukkan bahwa kata-kata yang mempunyai frekuensi tinggi adalah gagal, serius, dihina, mati, salah, dll. Hal ini mengindikasikan bahwa kata-kata tersebut menghasilkan tweet yang mengandung sentimen negatif. Tweet bersentimen negatif terhadap Prabowo Subianto sebagian besar menyinggung kegagalan Prabowo dalam tiga kali PEMILU presiden. Tweet dengan kata "gagal" dapat dilihat pada Gambar 4.13 dan 4.14.





Gambar 4.13 sampel *tweet* negatif dengan *keyword* "gagal"



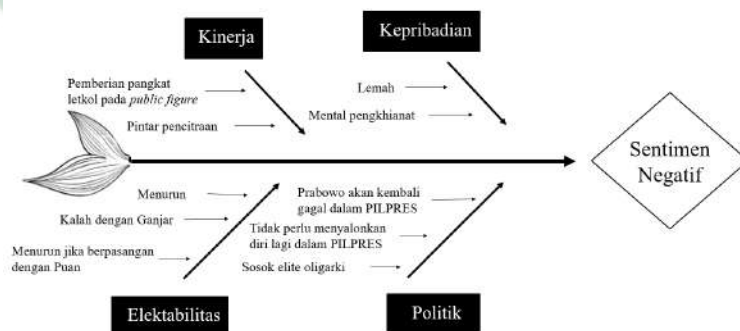
Gambar 4.14 sampel *tweet* negatif dengan *keyword* "gagal"

Kata "dihina" juga sering dibicarakan dalam sentimen negatif Prabowo. Dapat dilihat pada Gambar (4.15) bahwa Prabowo dihina dengan sebutan "macan mengeong" oleh Edy yang merupakan mantan politikus Partai Keadilan Sejahtera.



Gambar 4.15 sampel *tweet* negatif dengan *keyword* "dihina"

Berdasarkan sentimen-sentimen negatif yang telah didapatkan, juga dapat dilakukan visualisasi dengan menggunakan *fishbone* diagram seperti pada Gambar (4.16).



Gambar 4.16 *Fishbone* Diagram Sentimen Negatif

Diagram *fishbone* untuk sentimen negatif di atas menyoroti pendapat masyarakat mengenai kinerja, kepribadian, elektabilitas, dan politik yang berhubungan dengan Prabowo Subianto. Adapun analisis terkait diagram *fishbone* adalah sebagai berikut:

### 1. Kinerja

Menurut mantan asisten Prabowo, Prabowo diduga pintar pencitraan saja.

Oleh karena itu, mantan asisten tersebut mendukung Jokowi. Pemberian

pangkat letkol yang dilakukan Prabowo kepada *public figure* juga menyebabkan sentimen negatif. Hal tersebut dianggap pencitraan untuk menaikkan elektabilitas Prabowo (Dewi, 2022).

## 2. Kepribadian

Prabowo diduga lemah dan sebagian besar masyarakat sudah muak. Adapun *tweet* yang menyebutkan bahwa Prabowo memiliki mental pengkhianat.

## 3. Elektabilitas

Pada akhir tahun 2022, elektabilitas Prabowo menurun. Adapun menurut masyarakat, elektabilitas Prabowo menurun sejak deklarasi Anies sebagai bakal calon presiden pada November 2022 (Annur, 2022). Masyarakat juga berpendapat elektabilitas Prabowo menurun jika berpasangan dengan Puan. Elektabilitas Ganjar teratas mengalahkan Prabowo.

## 4. Politik

Menurut pakar politik, Prabowo akan kembali gagal jika mencalonkan diri kembali pada PILPRES 2024. Adapun beberapa masyarakat tidak ingin Prabowo untuk mencalonkan diri lagi dalam PILPRES 2024. Prabowo juga disebut elite oligarki dan berada dalam kendali Megawati.

Berdasarkan hasil analisis dan diagram *fishbone* sentimen negatif yang dilakukan, pada tahun 2022 masyarakat cenderung menilai Prabowo secara negatif berdasarkan kinerja selama menjabat menjadi Menteri Pertahanan. Alasan mengenai kegagalan Prabowo dalam tiga kali PEMILU tidak disebutkan dalam sentimen negatif yang diambil pada tahun 2022, seperti isu pelanggaran HAM yang dilakukan Prabowo pada tahun 1997-1998, isu pengkhianatan Prabowo terhadap Soeharto pada tahun 1997 (A43, 2022). Kekecewaan para pendukung

Prabowo ketika Prabowo dilantik sebagai Menteri Pertahanan oleh Jokowi yang saat itu merupakan rivalnya juga tidak dibahas dalam sentimen negatif twitter pada tahun 2022 (Redaksi, 2019). Sentimen negatif mengenai Prabowo pada tahun 2022 juga tidak membahas hal-hal yang dapat menjatuhkan Prabowo pada PEMILU 2024.

#### 4.8. Kesesuaian Hasil Penelitian dengan Islam

Hasil dari penelitian ini, SVM mampu melakukan klasifikasi pada sentimen masyarakat terhadap Prabowo Subianto. Klasifikasi yang didapatkan yaitu 30535 data sentimen terklasifikasi sebagai sentimen positif dan 2528 data sentimen terklasifikasi sebagai sentimen negatif. Klasifikasi dilakukan untuk membedakan data antara sentimen positif dan negatif. Klasifikasi sebenarnya juga terdapat dalam islam. Sebagaimana pada Q.S. Fatir ayat 32:

ثُمَّ أَوْرَثْنَا الْكِتَابَ الَّذِينَ اصْطَفَيْنَا مِنْ عِبَادِنَا فَمِنْهُمْ ظَالِمٌ لِنَفْسِهِ وَمِنْهُمْ مُقْتَصِدٌ وَمِنْهُمْ سَابِقٌ بِالْخَيْرَاتِ بإِذْنِ اللَّهِ ذَلِكَ هُوَ الْفَضْلُ الْكَبِيرُ

artinya: "Kemudian, Kitab Suci itu Kami wariskan kepada orang-orang yang Kami pilih di antara hamba-hamba Kami. Lalu, di antara mereka ada yang menzalimi diri sendiri, ada yang pertengahan, dan ada (pula) yang lebih dahulu berbuat kebaikan dengan izin Allah. Itulah (dianugerahkannya kitab suci adalah) karunia yang besar."

Ayat tersebut menerangkan bahwa Allah SWT membagi hamba-Nya menjadi tiga golongan, yaitu yang menzalimi diri sendiri, pertengahan, dan yang lebih dahulu berbuat kebaikan. Pembagian golongan ini termasuk ke dalam hal klasifikasi.

Dapat dilihat pada Gambar 4.3 terdapat kata yang mencerminkan sosok pemimpin dalam islam yaitu amanah atau dapat dipercaya sesuai dengan firman Allah Q.S. An-Nisa ayat 58 yaitu:

﴿ إِنَّ اللَّهَ يَأْمُرُكُمْ أَنْ تُؤَدُّوا الْأَمَانَاتِ إِلَىٰ أَهْلِهَا وَإِذَا حَكَمْتُمْ بَيْنَ النَّاسِ أَنْ تَحْكُمُوا بِالْعَدْلِ إِنَّ اللَّهَ نِعِمَّا يَعِظُكُمْ بِهِ إِنَّ اللَّهَ كَانَ سَمِيعًا بَصِيرًا ﴾

artinya: "Sesungguhnya Allah menyuruh kamu menyampaikan amanah kepada pemiliknya. Apabila kamu menetapkan hukum di antara manusia, hendaklah kamu tetapkan secara adil. Sesungguhnya Allah memberi pengajaran yang paling baik kepadamu. Sesungguhnya Allah Maha Mendengar lagi Maha Melihat."

Sentimen positif Prabowo juga menyebutkan kata "adil". Hal ini telah dijelaskan dalam Q.S. Sad ayat 26 bahwasanya seorang pemimpin haruslah berlaku adil, sebagaimana firman Allah,

يَا دَاوُدَ إِنَّا جَعَلْنَاكَ خَلِيفَةً فِي الْأَرْضِ فَاحْكُم بَيْنَ النَّاسِ بِالْحَقِّ وَلَا تَتَّبِعِ الْهَوَىٰ فَيُضِلَّكَ عَنْ سَبِيلِ اللَّهِ إِنَّ الَّذِينَ يَضِلُّونَ عَنْ سَبِيلِ اللَّهِ لَهُمْ عَذَابٌ شَدِيدٌ بِمَا نَسُوا يَوْمَ الْحِسَابِ

artinya: "Wahai Daud, sesungguhnya Kami menjadikanmu khalifah (penguasa) di bumi. Maka, berilah keputusan (perkara) di antara manusia dengan hak dan janganlah mengikuti hawa nafsu karena akan menyesatkan engkau dari jalan Allah. Sesungguhnya orang-orang yang sesat dari jalan Allah akan mendapat azab yang berat, karena mereka melupakan hari Perhitungan."

Pada Q.S. Al-Qashash ayat 26 disebutkan bahwa seorang pemimpin harus memiliki dua sifat yaitu, *al-Qowiyy* dan *al-Amin* yang artinya kuat dan dapat

dipercaya. Kuat yang dimaksud yaitu kuat secara fisik maupun secara mental. Menurut Imam Malik, apabila ada 2 calon pemimpin, kurang alim dan cukup berani atau cukup alim dan kurang berani, maka pilihlah pemimpin yang kurang alim dan cukup berani. Pemimpin yang kurang alim hanya akan mencelakakan dirinya, sedangkan pemimpin yang kurang berani akan mencelakakan banyak orang. Pada masa Nabi Muhammad SAW, Rasul tidak memberikan jabatan kepada sahabatnya, Abu Dzar al-Ghifari karena sahabat tersebut cukup alim akan tetapi tidak memiliki keberanian. Kata kuat dan dipercaya termasuk kedalam sentimen positif Prabowo Subianto.

Gambar 4.3 juga terlihat kata "baik". Kata tersebut sesuai dengan pemimpin yang telah di prediksi oleh Rasulullah. Pemimpin yang baik akan mendatangkan kebaikan, baik bagi yang memimpin maupun yang dipimpin. Sebagaimana Rasulullah SAW bersabda,

"Sepeninggalku nanti ada pemimpin-pemimpin yang akan memimpin kalian, pemimpin yang baik akan memimpin dengan kebaikannya dan pemimpin yang fajir akan memimpin kalian dengan kefajirannya. Maka dengarlah dan taatilah mereka pada perkara-perkara yang sesuai dengan kebenaran saja. Apabila mereka berbuat baik maka kebaikannya adalah bagimu dan untuk mereka, jika mereka berbuat buruk maka bagimu (untuk tetap berbuat baik) dan bagi mereka (keburukan mereka)." (HR. Bukhari Muslim)

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil percobaan klasifikasi menggunakan SVM pada sentimen masyarakat terhadap Prabowo Subianto dengan kelas sentimen positif dan sentimen negatif, maka diperoleh kesimpulan dan saran sebagai berikut:

1. Berdasarkan data yang diperoleh dari twitter, masyarakat Indonesia cenderung menilai Prabowo Subianto secara positif daripada negatif. Hal ini dibuktikan dengan jumlah tweet bersentimen positif sebanyak 31111 sedangkan sentimen negatif sebanyak 3795 tweet.
2. Berdasarkan beberapa uji coba yang dilakukan, didapatkan hasil terbaik yaitu pada kernel linear dengan nilai akurasi 96.7%, recall 98,15%, dan presisi 96,02% dengan pembagian kfold sebesar  $k = 10$ .

#### 5.2. Saran

Penelitian ini masih terdapat banyak kekurangan. Oleh karena itu, penulis menyarankan beberapa hal berikut sebagai bahan perbaikan untuk penelitian dan pengembangan di masa depan:

1. Pada proses *crawling* data twitter, data yang diperoleh merupakan tweet mengenai Prabowo selama 1 tahun. Sentimen yang didapatkan dalam 1 tahun dapat digunakan untuk mengenal Prabowo dalam waktu yang terbatas.

Hal ini disebabkan karena *crawling* data twitter saat ini berbayar. Oleh karena itu, *crawling* sebaiknya dilakukan selama Prabowo menjabat menjadi Menteri Pertahanan Indonesia agar lebih mengenal dan menilai kinerja Prabowo selama menjadi Menteri Pertahanan atau juga bisa dilakukan ketika terjadi *event* tertentu yang menyebabkan banyaknya opini publik menanggapi *event* tersebut.

2. Pada proses labelling sentimen, terdapat beberapa kata positif yang terlabel negatif. Seperti kata "serius" yang terdapat dalam sentimen "Prabowo sosok yang serius dalam bekerja". Sentimen tersebut tergolong dalam sentimen negatif, akan tetapi faktanya sentimen tersebut termasuk kedalam sentimen positif. Hal ini dikarenakan labelling dilakukan dengan perbedaan bahasa yaitu bahasa inggris. Oleh karena itu, proses labelling data sebaiknya dilakukan oleh ahli dengan memperhatikan kaidah-kaidah bahasa indonesia.
3. Labelling hanya dilakukan berdasarkan probabilitas kata-kata pada setiap dokumen tanpa memperhatikan aturan semiotika. Akibatnya, dua kalimat yang bermakna sama akan terlabel berbeda, seperti "kamu jahat" dengan "kamu tidak baik". Kedua kalimat tersebut bermakna sama yaitu negatif, akan tetapi tanpa semiotika, kata "tidak" termasuk dalam stopword yang berarti kata tersebut dihapus dari *corpus* dokumen. Oleh karena itu, pada penelitian selanjutnya sebaiknya memperhatikan semiotika agar sentimen dapat terlabel dengan baik dan benar.
4. Klasifikasi yang telah dilakukan menggunakan SVM hanya pada dua kelas saja dengan fitur terbatas. Hal ini dikarenakan fitur yang terlalu banyak membuat SVM tidak *compatible* dengan data yang diberikan. Oleh karena itu, klasifikasi sebaiknya dilakukan dengan metode lain yang dapat



melakukan klasifikasi pada data teks dalam jumlah yang besar, seperti menggunakan *deep learning*.

5. Pembobotan kata yang dilakukan menggunakan TF-IDF seringkali menghasilkan nilai yang sama pada tiap kata di dokumen yang berbeda. Hal ini dapat membuat sistem mengalami kesulitan dalam klasifikasi. Oleh karena itu, pembobotan kata dilakukan dengan metode lainnya seperti, *Long-Short Term Memory (LSTM)*, *Gated Recurrent Unit (GRU)*, atau metode pembobotan terbaru lainnya.
6. Hasil analisis dan diagram *fishbone* menunjukkan bahwa Prabowo dinilai secara positif dan negatif berdasarkan kinerja sebagai Menteri Pertahanan Republik Indonesia dalam waktu satu tahun saja tanpa adanya bahasan mengenai kegagalan Prabowo dalam tiga kali PILPRES serta sedikit sekali membahas mengenai PEMILU 2024. Hasil penelitian ini juga tidak dapat digunakan sebagai acuan untuk mengenal baik Prabowo Subianto. Hal ini dikarenakan pengambilan data yang dilakukan pada penelitian ini terbatas. Oleh karena itu, pengambilan data sebaiknya dilakukan selama 5 tahun selama Prabowo menjabat sebagai Menteri Pertahanan atau beberapa bulan sebelum PEMILU 2024 jika penelitian yang dilakukan berbasis *event*.

## DAFTAR PUSTAKA

- A43 (2022). Prabowo Pasti Gagal di 2024? Diakses pada 20 Juli 2023, dari <https://www.pinterpolitik.com/in-depth/prabowo-pasti-gagal-di-2024/>
- Abiola, O., Abayomi-Alli, A., Tale, O. A., Misra, S., and Abayomi-Alli, O. (2023). Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser. *Journal of Electrical Systems and Information Technology*, 10(1):1–20.
- Aditya, R. N. (2021). Arief Poyuono: Prabowo Tak Pernah Pikir Pilpres, Hanya Bantu Jokowi di Sektor Pertahanan. Diakses pada 27 Juni 2023, dari <https://nasional.kompas.com/read/2021/12/19/16554761/arief-poyuono-prabowo-tak-pernah-pikir-pilpres-hanya-bantu-jokowi-di-sektor>
- Al Amrani, Y., Lazaar, M., and El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127:511–520.
- Anam, M. K. (2022). Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 11(3):216–225.
- Anam, M. K., Pikir, B. N., Firdaus, M. B., Erlinda, S., and Agustin, A. (2021). Penerapan Na ve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(1):139–150.

- Anandan, B. and Manikandan, M. (2023). Machine learning approach with various regression models for predicting the ultimate tensile strength of the friction stir welded AA 2050-T8 joints by the K-Fold cross-validation method. *Materials Today Communications*, 34:105286.
- Annur, C. M. (2022). Survei Indikator: Elektabilitas Anies Naik Pasca Deklarasi Capres Nasdem. Diakses pada 20 Juli 2023, dari <https://databoks.katadata.co.id/datapublish/2022/12/02/survei-indikator-elektabilitas-anies-naik-pasca-deklarasi-capres-nasdem>
- Arsi, P. and Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1):147.
- Bayu, D. (2022). Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022. Diakses pada 1 Maret 2023, dari <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>
- Biro Humas Setjen Kemhan (2022a). Menhan Prabowo Menerima DIPA Tahun 2023 dari Presiden Jokowi. Diakses pada 11 Juli 2023, dari <https://www.kemhan.go.id/2022/12/28/menhan-prabowo-tinjau-pesawat-interim-falcon-7x-dan-8x-tni-au.html>
- Biro Humas Setjen Kemhan (2022b). Menhan Prabowo Tinjau Pesawat Interim Falcon 7X dan 8X TNI AU. Diakses pada 11 Juli 2023, dari <https://www.kemhan.go.id/2022/12/28/menhan-prabowo-tinjau-pesawat-interim-falcon-7x-dan-8x-tni-au.html>
- Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

- Borg, A. and Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications*, 162:113746.
- Cengiz, A., Budak, M., YA MUR, N., and BALAIK, F. (2023). Comparison between random forest and support vector machine algorithms for LULC classification. *International Journal of Engineering and Geosciences*, 8(1):1–10.
- Chicco, D., Totsch, N., and Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData mining*, 14(1):1–22.
- Dang, N. C., Moreno-Garcia, M. N., and De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3):483.
- Dewi, S. (2022). Pengamat: Deddy Corbuzier Sengaja Dijadikan Influencer demi Prabowo. Diakses pada 20 Juli 2023, dari <https://www.idntimes.com/news/indonesia/santi-dewi/pengamat-deddy-corbuzier-sengaja-dijadikan-influencer-demi-prabowo>
- Egeham, L. (2022). Survei Poltracking: Kepuasan Menteri Kabinet Jokowi, Prabowo-Erick Thohir Tertinggi. Diakses pada 20 Juli 2023, dari <https://www.liputan6.com/news/read/5148395/survei-poltracking-kepuasan-menteri-kabinet-jokowi-prabowo-erick-thohir-tertinggi>.
- Ekawati, A. (2019). Prabowo Masuk Kabinet Bisa Jadi Modal Kuat Untuk 2024. Diakses pada 27 Juni 2023, pada <https://www.dw.com/id/pengamat-prabowo-masuk-kabinet-bisa-jadi-modal-kuat-untuk-2024/a-50946823>
- Elbagir, S. and Yang, J. (2019). Twitter sentiment analysis using natural

language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 122, page 16.

Farisa, F. C. (2021). Jadi Syarat Ajukan Capres, Bagaimana Awalnya Presidential Threshold Ada? Diakses pada 11 Juli 2023, pada <https://nasional.kompas.com/read/2021/12/17/16330081/jadi-syarat-ajukan-capres-bagaimana-awalnya-presidential-threshold-ada>.

Farisa, F. C. (2022). Jejak Prabowo di Tiga Pemilu Presiden: 2009, 2014, dan 2019. Diakses pada 27 Juni 2023, pada <https://nasional.kompas.com/read/2022/08/15/06150071/jejak-prabowo-di-tiga-pemilu-presiden-2009-2014-dan-2019>.

Fazrin, F., Pratiwi, O. N., and Andreswari, R. (2022). Comparison of K-Nearest Neighbor and Logistic Regression Algorithms on Sentiment Analysis of Covid-19 Vaccination on Twitter with Vader And Textblob Labeling. In *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, pages 39–44. IEEE.

Febriyan (2022). Survei Pilpres 2024, Ini 3 Karakter Pemimpin yang Diinginkan Masyarakat. Diakses pada 11 Juli 2023, pada <https://nasional.tempo.co/read/1600270/survei-pilpres-2024-ini-3-karakter-pemimpin-yang-diinginkan-masyarakat>

Fitri, V. A., Andreswari, R., and Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naive Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161:765–772.

- Fransiska, S., Rianto, R., and Gufroni, A. I. (2020). Sentiment Analysis Provider by. U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*, 7(2):203–212.
- Gomes, L., da Silva Torres, R., and Cortes, M. L. (2023). BERT-and TF-IDF-based feature extraction for long-lived bug prediction in FLOSS: a comparative study. *Information and Software Technology*, 160:107217.
- Gunawan, D., Riana, D., Ardiansyah, D., Akbar, F., and Alfarizi, S. (2020). Komparasi Algoritma Support Vector Machine Dan Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023. V (1), 135138.
- Hanmastiana, I. M., Warsito, B., Rahmawati, R., Yasin, H., and Kartikasari, P. (2021). Classification of Public Opinion on Social Media Twitter concerning the Education in Indonesia Using the K-Nearest Neighbors (K-NN) Algorithm and K-Fold Cross Validation. *Statistika*, 21(2):99–106.
- Hermawan, B. (2022). Survei SPIN: Elektabilitas Prabowo Masih Tertinggi di Akhir 2022. Diakses pada 11 Juli 2023, pada <https://news.republika.co.id/berita/rnnwsm354/survei-spin-elektabilitas-prabowo-masih-tertinggi-di-akhir-2022>
- Hidayat, M. R. and Sulistiyono, M. (2022). Comparison of Accuracy and Time Of Naive Bayes Algorithm with Support Vector Machine Algorithm in Twitter Sentiment Analysis of Peduli Lindungi Application. In *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, pages 172–176. IEEE.

- Hong, C. S. and Oh, T. G. (2021). TPR-TNR plot for confusion matrix. *Communications for Statistical Applications and Methods*, 28(2):161–169.
- Kim, S.-W. and Gil, J.-M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-centric Computing and Information Sciences*, 9(1):1–21.
- Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K., and Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, 8(10):e10894.
- Li, J., Gao, F., Lin, S., Guo, M., Li, Y., Liu, H., Qin, S., and Wen, Q. (2023). Quantum k-fold cross-validation for nearest neighbor classification algorithm. *Physica A: Statistical Mechanics and its Applications*, 611:128435.
- Liu, C.-z., Sheng, Y.-x., Wei, Z.-q., and Yang, Y.-Q. (2018). Research of text classification based on improved TF-IDF algorithm. In *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, pages 218–222. IEEE.
- Liu, S. and Liu, J. (2021). Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine*, 39(39):5499–5505.
- Mustaqim, T., Umam, K., and Muslim, M. A. (2020). Twitter text mining for sentiment analysis on governments response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm. In *Journal of Physics: Conference Series*, volume 1567, page 32024. IOP Publishing.
- Muttaqin, M. N. and Kharisudin, I. (2021). Analisis sentimen aplikasi gojek

- menggunakan support vector machine dan k nearest neighbor. *Unnes Journal of Mathematics*, pages 22–27.
- Negara, A. B. P., Muhandi, H., and Putri, I. M. (2020). Analisis sentimen maskapai penerbangan menggunakan metode naive bayes dan seleksi fitur information gain. *J. Teknol. Inf. dan Ilmu Komput*, 7(3).
- Novianto, R. D. (2023). Jokowi Foto Bareng Prabowo dan Ganjar di Sawah, Sinyal Pilpres 2024? Diakses pada 27 Juni 2023, dari <https://nasional.sindonews.com/read/1042279/12/jokowi-foto-bareng-prabowo-dan-ganjar-di-sawah-sinyal-pilpres-2024-1678345386>
- Pamungkas, F. S. and Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. In *PRISMA, Prosiding Seminar Nasional Matematika*, volume 4, pages 628–634.
- Pande, C. B., Kushwaha, N. L., Orimoloye, I. R., Kumar, R., Abdo, H. G., Tolche, A. D., and Elbeltagi, A. (2023). Comparative assessment of improved SVM method under different kernel functions for predicting multi-scale drought index. *Water Resources Management*, 37(3):1367–1399.
- Phan, H. T., Tran, V. C., Nguyen, N. T., and Hwang, D. (2020). Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model. *IEEE Access*, 8:14630–14641.
- Praghakusma, A. Z. and Charibaldi, N. (2021). Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus: Komisi Pemberantasan Korupsi). *Jurnal Sarjana Teknik Informatika ISSN*, 2338(5197):33.



- Qaiser, S. and Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1):25–29.
- Qorib, M., Oladunni, T., Denis, M., Ososanya, E., and Cotae, P. (2023). COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Systems with Applications*, 212:118715.
- Rahat, A. M., Kahir, A., and Masum, A. K. M. (2019). Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages 266–270. IEEE.
- Redaksi, T. (2019). Prabowo Resmi Jadi Menhan, Netizen Terbelah. Diakses pada 11 Juli 2023, dari <https://www.cnbcindonesia.com/tech/20191023123657-37-109372/prabowo-resmi-jadi-menhan-netizen-terbelah>
- Redaksi CNBC Indonesia (2023). 5 Survei Terbaru Capres 2024: Prabowo Vs Ganjar Vs Anies. Diakses pada 27 Juni 2023, dari <https://www.cnbcindonesia.com/news/20230614043018-4-445669/5-survei-terbaru-capres-2024-prabowo-vs-ganjar-vs-anies>
- Riego, N. C. R. and Villarba, D. B. (2023). Utilization of Multinomial Naive Bayes Algorithm and Term Frequency Inverse Document Frequency (TF-IDF Vectorizer) in Checking the Credibility of News Tweet in the Philippines. *arXiv preprint arXiv:2306.00018*.
- Rozie, F. (2023). Gerindra Pastikan Prabowo Subianto Tetap Akan Maju Jadi Capres, Bukan Cawapres. Diakses pada 11 Juli 2023, dari

<https://www.liputan6.com/pemilu/read/5272465/gerindra-pastikan-prabowo-subianto-tetap-akan-maju-jadi-capres-bukan-cawapres>

Saad, A. I. (2020). Opinion mining on US Airline Twitter data using machine learning techniques. In *2020 16th International Computer Engineering Conference (ICENCO)*, pages 59–63. IEEE.

Sangeetha, J. and Kumaran, U. (2022). Comparison of Sentiment Analysis on Online Product Reviews Using Optimised RNN-LSTM with Support Vector Machine. *Webology*, 19(1):3883–3898.

Santoso, T. and Budhiati, I. (2021). *Pemilu di Indonesia: kelembagaan, pelaksanaan, dan pengawasan*. Sinar Grafika.

Simanjuntak, J. (2021). Bersedia Masuk Kabinet Jokowi, Prabowo: Kita Sama-sama Mengabdikan untuk Merah Putih. Diakses pada 27 Juni 2023, dari <https://nasional.sindonews.com/read/454110/12/bersedia-masuk-kabinet-jokowi-prabowo-kita-sama-sama-mengabdikan-untuk-merah-putih-1623553605>

Singh, J. and Singh, S. (2023). Support vector machine learning on slurry erosion characteristics analysis of Ni-and Co-alloy coatings. *Surface Review and Letters*, page 2340006.

Styawati, S., Hendrastuty, N., and Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3):150–155.

Tim DetikJateng (2023). Sah! Megawati Umumkan Ganjar Pranowo Jadi Capres

2024. Diakses pada 11 Juli 2023, dari <https://www.detik.com/jateng/berita/d-6684499/sah-megawati-umumkan-ganjar-pranowo-jadi-capres-2024>.

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 9(1):40–50.

Van Veelen, B. and Van Der Horst, D. (2018). What is energy democracy? Connecting social science energy research and political theory. *Energy Research & Social Science*, 46:19–28.

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J.-H., and Hsieh, J.-G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naive bayes. *Information*, 12(5):204.

Wang, H., Li, G., and Wang, Z. (2023). Fast SVM classifier for large-scale classification problems. *Information Sciences*, 642:119136.

Xu, G., Meng, Y., Qiu, X., Yu, Z., and Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. *Ieee Access*, 7:51522–51532.

Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., and Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China Life Sciences*, 63(3):457–460.

Yadav, A. and Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6):4335–4385.

Yang, L., Li, Y., Wang, J., and Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, 8:23522–23530.

- Yatsugi, K., Pandarakone, S. E., Mizuno, Y., and Nakamura, H. (2023). Common Diagnosis Approach to Three-Class Induction Motor Faults Using Stator Current Feature and Support Vector Machine. *IEEE Access*, 11:24945–24952.
- Yuvens, D. A. (2016). Penerapan Pasal 1 ayat (2) Kitab Undang-Undang Hukum Pidana dalam Perspektif Kontemporer. *Jurnal Konstitusi*, 13(4):743–765.
- Zendeboudi, A., Baseer, M. A., and Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of cleaner production*, 199:272–285.
- Zhang, L., Li, C., Chen, L., Chen, D., Xiang, Z., and Pan, B. (2023). A Hybrid Forecasting Method for Anticipating Stock Market Trends via a Soft-Thresholding De-noise Model and Support Vector Machine (SVM). *World Basic and Applied Sciences Journal*, 13(2023):597–602.
- Zulfirman, Z. and Manurung, R. S. (2018). Pembukaan UUD 1945: Analisis Nilai Politik dan Hukum Indonesia. *Jurnal IUS Kajian Hukum dan Keadilan*, 6(1):72–89.

UIN SUNAN AMPEL  
S U R A B A Y A