

**ANALISIS SENTIMEN CALON PRESIDEN INDONESIA  
2024 DI FACEBOOK MENGGUNAKAN METODE  
*SUPPORT VECTOR MACHINE (SVM)***



Disusun Oleh:

N a m a : Fardhan Saifullah Fattah  
NIM : 19523202

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA**

**2023**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**ANALISIS SENTIMEN CALON PRESIDEN INDONESIA  
2024 DI FACEBOOK MENGGUNAKAN METODE  
*SUPPORT VECTOR MACHINE (SVM)***

**TUGAS AKHIR**



Yogyakarta, 2 November 2023

Pembimbing,

( Chanifah Indah Ratnasari, S.Kom., M.Kom. )

HALAMAN PENGESAHAN DOSEN PENGUJI

**ANALISIS SENTIMEN CALON PRESIDEN INDONESIA  
2024 DI FACEBOOK MENGGUNAKAN METODE  
*SUPPORT VECTOR MACHINE (SVM)***

**TUGAS AKHIR**

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 2 November 2023

Tim Penguji

Chanifah Indah Ratnasari, S.Kom., M.Kom.

**Anggota 1**

Arrie Kurniawardhani, S.Si., M.Kom.

**Anggota 2**

Beni Suranto, S.T., M.Soft.Eng.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia

( Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D)

**HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR**

Yang bertanda tangan di bawah ini:

Nama : Fardhan Saifullah Fattah

NIM : 19523202

Tugas akhir dengan judul:

**ANALISIS SENTIMEN CALON PRESIDEN INDONESIA  
2024 DI FACEBOOK MENGGUNAKAN METODE  
*SUPPORT VECTOR MACHINE (SVM)***

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 2 November 2023



Handwritten signature of Fardhan Saifullah Fattah.

( Fardhan Saifullah Fattah)

## HALAMAN PERSEMBAHAN

Assalamualaikum warahmatullahi wabarakatuh, dengan rendah hati dan bersyukur, saya mengucapkan terima kasih kepada Allah SWT atas izin, karunia, dan berkat-Nya yang telah memberikan kelancaran dalam menyelesaikan Tugas Akhir ini tepat waktu. Semoga ini menjadi langkah awal saya dalam mencapai cita-cita dan berkontribusi secara positif bagi masyarakat, Aamiin.

Sebagai ungkapan rasa terima kasih atas cinta, perhatian, dan dukungan yang tak terhingga, penulis ingin mengabdikan laporan tugas akhir ini untuk Ibu Zuriyati Enika dan Bapak Herry Noverizal. Mereka selalu memberikan dukungan moril dan materiel, doa, serta restu kepada penulis untuk kelancaran segala urusannya.

Penulis ingin menyampaikan rasa terima kasih kepada Bapak Chanifah Indah Ratnasari, S.Kom., M.Kom. atas bimbingan dan arahannya yang sangat berharga dalam menyusun tugas akhir ini. Penulis juga ingin mengucapkan terima kasih kepada rekan-rekan sejurusan di Program Studi Informatika 2019 yang telah menjadi bagian berharga dalam perjalanan perkuliahan, mendukung dalam berbagai situasi, dan saling membantu selama masa kuliah.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

**HALAMAN MOTO**

*"You can change your wife, your politics, your religion. But never, never can you change your favourite football team."*

(Eric Cantona)

*"Kebahagiaan sesungguhnya adalah menjaga persahabatan tetap utuh untuk selamanya."*

(Hatake Kakashi)

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Dengan rendah hati dan penuh rasa syukur kehadiran Allah Swt. atas segala limpahan kasih, karunia, dan kehendak-Nya sehingga Tugas Akhir dengan judul Analisis Sentimen Calon Presiden Indonesia 2024 di Facebook Menggunakan Metode *Support Vector Machine* (SVM) dapat diselesaikan dengan baik. Selesaiannya Tugas Akhir ini tidak terlepas dari bantuan, bimbingan dan doa dari berbagai pihak. Pada kesempatan ini ingin disampaikan terima kasih kepada semua pihak yang telah membantu dalam pembuatan karya ini, ucapan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Kedua orang tua saya, yang selalu memberikan semangat, dukungan, dan memahami penulis selama ini.
2. Bapak Prof. Fathul Wahid, S.T.,M.Sc.,Ph.D selaku Rektor Universitas Islam Indonesia.
3. Bapak Hari Purnomo, Prof., Dr., Ir., M.T., IPU, ASEAN.Eng selaku Dekan Fakultas Teknologi Industri.
4. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc, selaku Ketua Jurusan Informatika Universitas Islam Indonesia
5. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Kepala Prodi Informatika.
6. Ibu Chanifah Indah Ratnasari, S.Kom., M.Kom. selaku dosen pembimbing yang telah meluangkan waktu selama proses bimbingan.
7. Seluruh dosen serta staf pengajar Program Studi Informatika Universitas Islam Indonesia.
8. Adik saya Fikram Arief yang saat ini sedang menjalani pendidikan di Gontor.
9. Teman-teman saya tercinta: Riko, Reizi, Imam, Rafi, Alfa, Alafta, Rafiq, Ghilman, Jasmine, Sallu, Apriza dan yang saya tidak bisa sebutkan satu persatu yang banyak memberikan motivasi, bantuan, serta selalu membuat hari penulis berwarna dan bersemangat dalam menyelesaikan Tugas Akhir ini.
10. Teman-teman Santos saya tersayang: Ghazi, Abiel, Rio, Fiqi, serta teman-teman lainnya yang selalu menjadi bagian penyemangat selama penulisan berlangsung.

## SARI

Pemilihan Presiden Indonesia 2024 telah menjadi topik yang hangat diperbincangkan dalam kehidupan sehari-hari, salah satunya pada *platform* media sosial Facebook. Semua orang memiliki kebebasan untuk mengemukakan pendapat atau menyuarakan opini mereka mengenai calon Presiden Indonesia 2024. Hal ini menghasilkan beragam pendapat, pendapat tersebut ada yang mengandung sentimen positif ataupun negatif. Penelitian ini bertujuan untuk mengklasifikasi sentimen positif dan negatif pada data komentar postingan Facebook masing-masing calon presiden. Dalam proses analisis sentimen, terdapat beberapa langkah yang terdiri dari *scraping data*, *preprocessing data*, pelabelan data, dan pengklasifikasian data dengan menggunakan metode *Support Vector Machine* (SVM). Data yang digunakan adalah komentar yang diambil dari halaman Facebook tiap bakal calon presiden dari tanggal 1 Juni 2023 sampai dengan 31 Juli 2023. Sentimen positif tertinggi dari Prabowo Subianto terdapat pada topik 'lain-lain' (cerita tentang masa lalu, demo dan resep memasak, berbagi berbagai tip, dan menonton pertandingan sepak bola) dengan nilai sentimen positif sebesar 88,09%. Anies Baswedan memiliki sentimen positif tertinggi pada topik 'melaksanakan ibadah haji' dengan nilai sentimen positif sebesar 91,70%, dan Ganjar Pranowo mendapati sentimen positif tertinggi pada topik 'lain-lain' sebesar 86,56%.

Kata Kunci: analisis sentimen, *support vector machine*, calon presiden 2024, Facebook, SVM.

## GLOSARIUM

|                               |   |
|-------------------------------|---|
| Klasifikasi                   | salah satu model pada <i>machine learning</i> yang dapat membedakan dua atau lebih data berdasarkan kelas.                          |
| <i>Support Vector Machine</i> | metode pada <i>supervised learning</i> yang dapat digunakan untuk klasifikasi <i>Machine Learning</i> dan regresi.                  |
| <i>Preprocessing</i>          | tahapan pengolahan data yang digunakan untuk mengubah data menjadi format tertentu yang berguna, efisien, dan konsisten.            |
| <i>Library</i>                | kumpulan modul terkait berisi kumpulan kode yang dapat digunakan berulang kali dalam program yang berbeda.                          |
| <i>DataFrame</i>              | tabel atau data tabular dengan <i>array</i> dua dimensi yang juga merupakan salah satu fungsi dari <i>library Pandas</i> di Python. |

## DAFTAR ISI

|   |      |
|---|------|
| HALAMAN JUDUL .....                             | i    |
| HALAMAN PENGESAHAN DOSEN PEMBIMBING .....       | ii   |
| HALAMAN PENGESAHAN DOSEN PENGUJI.....           | iii  |
| HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....    | iv   |
| HALAMAN PERSEMBAHAN .....                       | v    |
| HALAMAN MOTO .....                              | vi   |
| KATA PENGANTAR .....                            | vii  |
| SARI .....                                      | viii |
| GLOSARIUM.....                                  | ix   |
| DAFTAR ISI .....                                | x    |
| DAFTAR TABEL .....                              | xi   |
| DAFTAR GAMBAR.....                              | xii  |
| BAB I PENDAHULUAN.....                          | 1    |
| 1.1 Latar Belakang .....                        | 1    |
| 1.2 Rumusan Masalah .....                       | 3    |
| 1.3 Batasan Masalah.....                        | 3    |
| 1.4 Tujuan Penelitian.....                      | 3    |
| 1.5 Manfaat Penelitian.....                     | 3    |
| 1.6 Sistematika Penulisan.....                  | 4    |
| BAB II LANDASAN TEORI.....                      | 5    |
| 2.1 Penelitian Terdahulu.....                   | 5    |
| 2.2 Facebook .....                              | 8    |
| 2.3 Analisis Sentimen.....                      | 9    |
| 2.4 Machine Learning.....                       | 9    |
| 2.5 <i>Support Vector Machine</i> .....         | 11   |
| 2.6 <i>Confusion Matrix</i> .....               | 12   |
| BAB III METODOLOGI PENELITIAN .....             | 15   |
| 3.1 Tahap Penelitian .....                      | 15   |
| 3.2 Pengumpulan Data .....                      | 16   |
| 3.3 <i>Data Preprocessing</i> .....             | 16   |
| 3.4 Pelabelan Kelas Sentimen .....              | 17   |
| 3.5 <i>Split Data Train dan Data Test</i> ..... | 17   |
| 3.6 <i>Feature Extraction</i> .....             | 18   |
| 3.7 Klasifikasi Menggunakan SVM .....           | 18   |
| BAB IV HASIL DAN PEMBAHASAN .....               | 20   |
| 4.1 Pengumpulan Data .....                      | 20   |
| 4.2 <i>Data Preprocessing</i> .....             | 26   |
| 4.4 <i>Split Data Train dan Data Test</i> ..... | 36   |
| 4.5 <i>Feature Extraction</i> .....             | 38   |
| 4.6 Klasifikasi Menggunakan SVM .....           | 39   |
| BAB V KESIMPULAN DAN SARAN .....                | 45   |
| 5.1 Kesimpulan.....                             | 45   |
| 5.2 Saran.....                                  | 46   |
| DAFTAR PUSTAKA .....                            | 47   |
| LAMPIRAN .....                                  | 50   |

## DAFTAR TABEL

|   |    |
|---|----|
| Tabel 4.1 Jumlah Postingan dan Komentar.....  | 21 |
| Tabel 4.2 Contoh Postingan Calon Presiden.....  | 22 |
| Tabel 4.3 Pembagian Data Ganjar Pranowo Berdasarkan Topik Bahasan .....                         | 25 |
| Tabel 4.4 Pembagian Data Anies Baswedan Berdasarkan Topik Bahasan.....                          | 25 |
| Tabel 4.5 Pembagian Data Prabowo Subianto Berdasarkan Topik Bahasan .....                       | 25 |
| Tabel 4.6 Contoh data hasil <i>scraping</i> yang akan dilakukan <i>preprocessing</i> .....      | 26 |
| Tabel 4.7 Contoh komentar setelah dilakukan <i>case folding</i> .....                           | 27 |
| Tabel 4.8 Contoh komentar setelah dilakukan <i>tokenizing</i> .....                             | 28 |
| Tabel 4.9 Contoh kata yang ada di kamus normalisasi .....                                       | 29 |
| Tabel 4.10 Contoh komentar setelah dilakukan normalisasi.....                                   | 30 |
| Tabel 4.11 Contoh komentar setelah dilakukan <i>stopword removal</i> .....                      | 32 |
| Tabel 4.12 Contoh Kata Positif dan Negatif .....  | 33 |
| Tabel 4.13 Hasil Pelabelan Sentimen dengan Lexicon pada Masing-Masing Topik Calon Presiden..... | 35 |
| Tabel 4.14 Pembagian jumlah <i>train data</i> dan <i>test data</i> Ganjar Pranowo.....          | 37 |
| Tabel 4.15 Pembagian jumlah <i>train data</i> dan <i>test data</i> Anies Baswedan.....          | 38 |
| Tabel 4.16 Pembagian jumlah <i>train data</i> dan <i>test data</i> Prabowo Subianto .....       | 38 |
| Tabel 4.17 Perbandingan Hasil Klasifikasi SVM Ganjar Pranowo .....                              | 40 |
| Tabel 4.18 Perbandingan Hasil Klasifikasi SVM Anies Baswedan.....                               | 40 |
| Tabel 4.19 Perbandingan Hasil Klasifikasi SVM Prabowo Subianto .....                            | 41 |
| Tabel 4.20 Hasil dari <i>confusion matrix</i> Ganjar Pranowo .....                              | 41 |
| Tabel 4.21 Hasil dari <i>confusion matrix</i> Anies Baswedan .....                              | 42 |
| Tabel 4.22 Hasil dari <i>confusion matrix</i> Prabowo Subianto.....                             | 42 |
| Tabel 4.23 Sentimen Positif Tertinggi .....   | 43 |
| Tabel 4.24 Sentimen Negatif Tertinggi.....  | 44 |

**DAFTAR GAMBAR**

|  |    |
|--|----|
| Gambar 2.1 Analisis Sentimen.....  | 9  |
| Gambar 2.2 Alur algoritma <i>supervised machine learning</i> .....           | 10 |
| Gambar 3.1 Tahap Penelitian .....  | 15 |
| Gambar 3.2 Tahapan dari Pra-pemrosesan data.....                             | 16 |
| Gambar 4.1 Pengambilan Data di <i>apify</i> .....                            | 20 |
| Gambar 4.2 Contoh hasil dari <i>scrapping</i> menggunakan <i>apify</i> ..... | 26 |
| Gambar 4.3 kode program <i>case folding</i> .....                            | 27 |
| Gambar 4.4 Kode program <i>tokenizing</i> .....                              | 28 |
| Gambar 4.5 Kode program proses normalisasi .....                             | 30 |
| Gambar 4.6 Kode program untuk menginstal <i>library sastrawi</i> .....       | 31 |
| Gambar 4.7 Kode program proses <i>stopword removal</i> .....                 | 32 |
| Gambar 4.8 Kode program proses pelabelan data.....                           | 34 |
| Gambar 4.9 Kode program untuk pembagian data .....                           | 37 |
| Gambar 4.10 Kode program SVM .....   | 35 |

## BAB I PENDAHULUAN

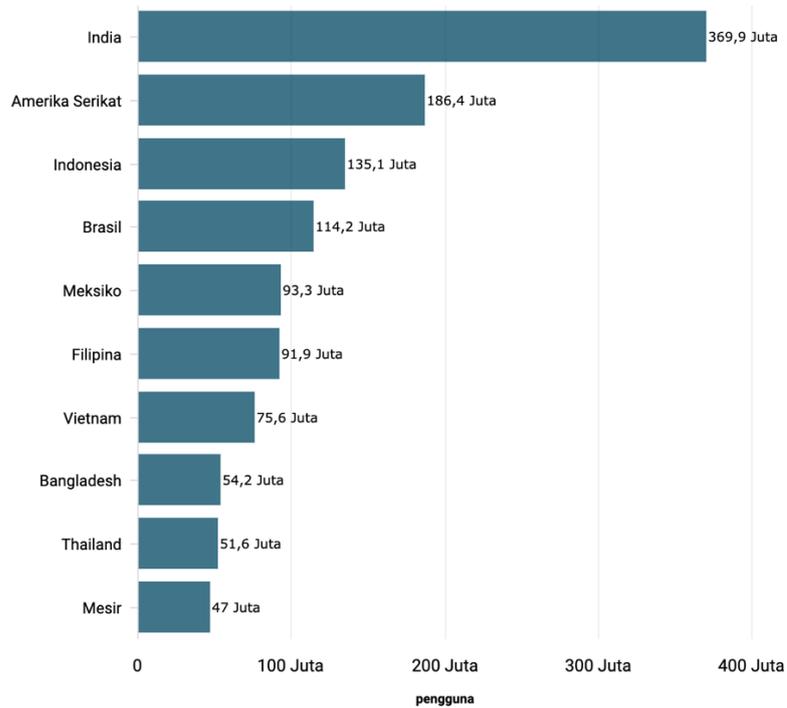
### 1.1 Latar Belakang

Ciri utama sebuah negara yang demokratis adalah tingkat partisipasi masyarakat yang tinggi dalam pelaksanaan pemilihan umum. Hal ini dikarenakan keterlibatan politik dari warga negara adalah elemen kunci dalam struktur pemerintahan demokratis (Liando, 2016). Meskipun pemilihan presiden (pilpres) 2024 masih beberapa bulan lagi, gejolak ketegangan dalam masyarakat tentang calon yang layak memimpin Indonesia untuk lima tahun ke depan sudah mulai terlihat (Sari, 2018). Ada beberapa nama yang telah diusung berbagai partai untuk maju sebagai calon presiden 2024 yaitu, Prabowo Subianto, Ganjar Pranowo, dan Anies Baswedan.

Pemilihan presiden di Indonesia merupakan salah satu momen penting dalam demokrasi. Pada tahun 2024, Indonesia akan menggelar pilpres kembali. Dalam setiap penyelenggaraan pilpres, masyarakat secara aktif menyuarakan berbagai pendapat mengenai calon-calon presiden yang akan bersaing dalam pilpres 2024 (Sabily et al., 2019). Salah satu wadah yang sering digunakan oleh masyarakat untuk mengungkapkan pandangan mereka adalah kolom komentar dari *page* tiap bakal calon presiden pada *platform* Facebook. Sentimen atau pandangan masyarakat terhadap postingan calon presiden (*capres*) di Facebook dapat digunakan untuk melihat apakah memperoleh pendapat positif atau negatif, juga dapat digunakan untuk melihat postingan terkait hal apa (*topik*) dari masing-masing *capres* yang memperoleh pendapat positif dan negatif terbanyak.

Sosial media menjadi salah satu sarana yang dibutuhkan masyarakat untuk berkomunikasi, berekspresi, dan juga bertukar informasi. Salah satu sosial media yang cukup digandrungi adalah Facebook. Saat ini, Facebook menjadi salah satu media sosial yang sangat populer di Indonesia bahkan di dunia. Pengguna Facebook mencapai 1,8 miliar dan sekitar 800 juta pengguna aktif yang menghabiskan waktu sekitar 40 menit sehari mengakses Facebook (Utami et al., 2021). Facebook bukan hanya sekadar jejaring pertemanan di dunia maya, tetapi juga memberikan wadah bagi publik untuk dapat mengekspresikan pendapat mereka secara luas dengan bebas (Salman & Widiyanesti, 2022). Indonesia menempati peringkat ke-3 sebagai pengguna Facebook terbanyak di dunia dengan 135,05 juta orang dan sekitar 95% di antaranya berusia di atas 17 tahun seperti yang ditunjukkan pada Gambar 1.1 (Databoks, 2023). Facebook menjadi salah satu *platform* yang digunakan oleh bakal calon presiden untuk meningkatkan

tingkat kepopuleran mereka. Di dalam Facebook terdapat banyak opini dan informasi yang akan digunakan sebagai alat penentu kebijakan dengan menggunakan *text mining*.



Gambar 1.1 Grafik pengguna Facebook di dunia

Sumber: [databoks.katadata.co.id](http://databoks.katadata.co.id)

Analisis sentimen merupakan cabang studi yang mengkaji pendapat, pandangan, evaluasi, emosi, sikap, atau penilaian individu terhadap berbagai entitas seperti individu, produk, masalah, organisasi, peristiwa, atau topik tertentu (Sabily et al., 2019). Dalam penelitian ini, analisis sentimen digunakan karena pemilihan presiden sering kali memunculkan beragam pendapat dari masyarakat. Hasil dari analisis sentimen akan menghasilkan kesimpulan apakah sentimen tersebut bersifat positif atau negatif. Penelitian ini juga dapat berguna untuk memprediksi tingkat dukungan bagi bakal calon presiden yang akan bersaing dalam pemilihan presiden tahun 2024. Selain itu, penelitian ini juga bertujuan untuk melihat postingan terkait hal apa dari masing-masing capres yang memperoleh pendapat positif dan negatif terbanyak.

Dalam menganalisis opini dalam bentuk komentar melalui media sosial seperti Facebook, diperlukan suatu metode perhitungan untuk mengklasifikasikan komentar-komentar tersebut. Salah satu metode yang dapat digunakan adalah *Support Vector Machine* (SVM). SVM adalah teknik yang digunakan untuk mengelompokkan data dengan mencari *hyperplane* teroptimal dari *dataset* (Lukmana et al., 2019).

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat dirumuskan masalahnya sebagai berikut:

- a. Topik apa untuk masing-masing calon presiden yang memiliki nilai sentimen positif tertinggi dan negatif tertinggi?
- b. Siapakah di antara calon presiden yang memiliki sentimen positif tertinggi?
- c. Berapa nilai *precision* dari metode *support vector machine* (SVM) dalam mengklasifikasi komentar pengguna pada Facebook *page* tiap calon presiden?

## 1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

- a. Data yang diambil berasal dari Facebook *page* Prabowo Subianto, Ganjar Pranowo, dan Anies Baswedan dari tanggal 1 Juni 2023 sampai tanggal 31 Juli 2023.
- b. Data diolah menggunakan bahasa pemrograman *python*.
- c. Metode yang digunakan di penelitian ini adalah *Support Vector Machine* (SVM).

## 1.4 Tujuan Penelitian

Tujuan dari dilakukannya penelitian ini adalah:

- a. Mengetahui sentimen publik terhadap tiap calon presiden.
- b. Mengetahui topik postingan dari masing-masing capres yang memiliki sentimen positif terbanyak.
- c. Mengetahui efektivitas dan akurasi metode *support vector machine* (SVM) dalam mengklasifikasi komentar pengguna Facebook pada Facebook *page* tiap calon presiden.

## 1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah:

- a. Membantu pihak yang membutuhkan atau yang terkait seperti partai politik/tim sukses untuk mengevaluasi opini publik terhadap postingan pada Facebook *page* tiap capres, yang salah satunya dapat digunakan untuk merumuskan program kerja atau hal terkait.

- b. Membantu sebagai rujukan untuk penelitian mendatang dengan topik atau tema yang berkaitan.

## 1.6 Sistematika Penulisan

Sistematika penulisan dalam penyusunan tugas akhir ini terdiri dari beberapa bab, yang mencakup gambaran dari keseluruhan masalah dan penyelesaiannya. Berikut sistematika penulisan yang terbagi dalam 5 bab:

### BAB I PENDAHULUAN

Pada bab ini berisi pembahasan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### BAB II LANDASAN TEORI

Pada bab ini berisi pembahasan mengenai konsep dan prinsip dasar untuk memecahkan masalah yang ada di penelitian. Terdapat juga uraian tentang hasil penelitian yang pernah dilakukan sebelumnya oleh peneliti lain yang memiliki kemiripan dengan penelitian yang dilakukan.

### BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang tahapan penelitian yang akan dilakukan seperti pengambilan data, *preprocessing*, pelabelan data, pembuatan *data train* dan *data test*, dan klasifikasi menggunakan SVM.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini akan membahas tentang pengimplementasian metode dan pembahasan hasilnya.

### BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir yang akan membahas kesimpulan dan saran terhadap penelitian yang telah dilakukan pada tugas akhir.

## BAB II LANDASAN TEORI

### 2.1 Penelitian Terdahulu

Pada bagian ini, peneliti akan memaparkan beberapa penelitian sebelumnya yang relevan dengan penelitian ini yang diambil dari Google Scholar dengan *keyword* "analisis sentimen" dari tahun 2016 hingga tahun 2023. Dalam konteks ini, peneliti mengambil enam penelitian tentang analisis sentimen dari penelitian terdahulu. Penelitian terdahulu merupakan hal yang penting karena dapat digunakan sebagai referensi dan kajian untuk melihat penelitian-penelitian sejenis yang sudah dilakukan, sehingga dapat mengambil ilmu dari hal tersebut, serta untuk mengetahui *state-of-the-art* dari penelitian yang sedang dikerjakan

Berikut ini merupakan beberapa penelitian terdahulu yang penulis jadikan referensi. Penelitian pertama dilaksanakan oleh Tineges et al. (2020) yang berjudul “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter dengan Metode Klasifikasi *Support Vector Machine* (SVM)”. Studi kasus yang digunakan pada penelitian tersebut adalah *tweet* yang diambil dari akun *twitter* @indihome dari tanggal 16 Maret 2020 sampai 22 Maret 2020. Penelitian tersebut bertujuan untuk mengetahui tingkat kepuasan pengguna layanan Indihome yang diambil di Twitter. Dari hasil pengujian, didapatkan tingkat akurasi sebesar 87%.

Kedua, penelitian yang dilakukan oleh Arsi & Waluyo (2021) yang berjudul “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM)”. Dengan dilakukannya penelitian ini bertujuan untuk menerapkan algoritma SVM pada data sentimen dari Twitter yang berkaitan dengan isu pemindahan ibu kota untuk mengevaluasi sejauh mana akurasi, *recall*, dan *precision* dapat diukur dalam analisis sentimen. Adapun data *training* dan *testing* yang digunakan menggunakan skala perbandingan 9:1. Hasil analisis terhadap *tweets* perpindahan ibu kota dari Twitter sebanyak 1.236 *twit* (404 positif dan 832 negatif) menggunakan SVM, didapatkan akurasi sebesar 96%

Penelitian yang dilakukan oleh Rachmat C & Lukito (2016) yang berjudul “Klasifikasi Sentimen Komentar Politik dari Facebook *Page* Menggunakan Naive Bayes”. Studi kasus yang dijadikan fokus pada riset tersebut adalah status dan komentar terhadap Facebook *Page* calon presiden Republik Indonesia pada Pemilu tahun 2014. Langkah-langkah penelitian melibatkan data sebanyak 68 status (dengan total 3400 komentar) selama masa kampanye, yang kemudian diolah melalui proses pra-pemrosesan seperti tokenisasi, *stemming*, dan pembobotan token. Selanjutnya, dilakukan tahap klasifikasi, dan evaluasi kinerja model menggunakan *confusion*

*matrix*. Hasil pengujian dan evaluasi menunjukkan bahwa pendekatan menggunakan metode Naive Bayes menghasilkan akurasi klasifikasi sentimen lebih dari 83%.

Keempat, penelitian oleh Salam et al. (2018) dengan judul "Analisis Sentimen Data Komentar Sosial Media Facebook dengan *K-Nearest Neighbor* (Studi Kasus pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)". Hasil penelitian ini yaitu menggunakan 800 data yang dibagi menjadi 300 data *training* dan 500 data *testing*. Kemudian dari 300 data *training* dibagi lagi menjadi 100 data *training* positif, 100 data *training* negatif, dan 100 data netral. Setelah 6 kali diuji coba dengan menggunakan metode *K-Nearest Neighbor* didapatkan akurasi tertinggi sebesar 79,21%.

Kelima, penelitian Hendriyanto et al. (2022) dengan judul "*Sentiment Analysis of MOLA Application Reviews on Google Play Store using Support Vector Machine Algorithm*". Analisis sentimen terhadap aplikasi MOLA di *Google Play Store* dilakukan dengan menggunakan algoritma *Support Vector Machine* (SVM) untuk menggali pandangan pengguna terhadap aplikasi tersebut. Penelitian ini menerapkan metode *Knowledge Discovery in Database* (KDD). Data yang digunakan terdiri dari 520 ulasan aplikasi MOLA, yang terbagi menjadi 312 ulasan positif dan 208 ulasan negatif. Hasil terbaik ditemukan pada skenario 1 (90:10) dengan penggunaan *kernel* RBF (*Radial Basis Function*). Hasil dari eksperimen ini menunjukkan tingkat akurasi sebesar 92,31%, presisi 96,3%, *recall* 89,66%, dan nilai *f1-score* mencapai 92,86%. Hal ini mengindikasikan bahwa SVM dengan konfigurasi ini mampu memberikan hasil analisis sentimen yang sangat baik terhadap ulasan pengguna terkait aplikasi MOLA di *Google Play Store*.

Keenam, penelitian oleh Fitriyyah et al. (2019) berjudul "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial *Twitter* Menggunakan Metode Naive Bayes". Tujuan utama dari penelitian ini adalah untuk mengevaluasi implementasi metode Naive Bayes dalam mengklasifikasikan sentimen pengguna *Twitter* ke dalam dua kelas (negatif dan positif) dan tiga kelas (negatif, positif, dan netral). Hasil penelitian menunjukkan bahwa pengujian dilakukan dengan menggunakan tiga kategori dan dua kategori sentimen untuk setiap pasangan calon presiden. Hasil pengujian tiga kategori untuk pasangan calon presiden 01 dan 02 menghasilkan akurasi masing-masing sebesar 64,6% dan 58%. Sementara itu, pengujian dua kategori untuk pasangan calon presiden 01 dan 02 menghasilkan akurasi masing-masing sebesar 77,7% dan 88%. Calon presiden nomor urut dua mencapai performansi tertinggi dengan nilai *f-measure* sebesar 0,88. Rekap dari penelitian terdahulu yang sudah dibahas di atas ditunjukkan pada Tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

| <b>Penulis</b>             | <b>Judul</b>  | <b>Metode</b>                 | <b>Hasil</b> |
|----------------------------|---|-------------------------------|--------------|
| Tineges et al. (2020)      | Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi <i>Support Vector Machine</i> (SVM)".                           | <i>Support Vector Machine</i> | 87%          |
| Arsi & Waluyo, (2021)      | Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)".                                       | <i>Support Vector Machine</i> | 96%          |
| Rachmat C & Lukito, (2016) | Klasifikasi Sentimen Komentar Politik dari Facebook <i>Page</i> Menggunakan Naive Bayes   | <i>Naive Bayes</i>            | 83%          |
| Salam et al. (2018)        | Analisis Sentimen Data Komentar Sosial Media Facebook dengan <i>K-Nearest Neighbor</i> (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia | <i>K-Nearest Neighbor</i>     | 79,21%       |
| Hendriyanto et al. (2022)  | <i>Sentiment Analysis of MOLA Application Reviews on Google Play Store Using Support Vector Machine Algorithm</i>   | <i>Support Vector Machine</i> | 92%          |
| Fitriyyah et al. (2019)    | Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial <i>Twitter</i> Menggunakan Metode Naive Bayes   | <i>Naive Bayes</i>            | 88%          |

Berdasarkan pada sejumlah penelitian yang tercantum dalam Tabel 2.1, hasil yang nilai akurasi dan performansinya cenderung lebih tinggi dari metode lainnya adalah *Support Vector Machine*. Dengan ini penulis akan melakukan penelitian dengan menerapkan metode *Support Vector Machine*.

## 2.2 Facebook

Facebook merupakan *platform* media sosial yang saat ini sangat populer. Sebagai media sosial, Facebook menjadi wadah yang digunakan untuk berinteraksi dan berkomunikasi. Keberadaan media sosial ini memberikan beragam kemudahan dalam menjalankan proses komunikasi. Facebook memberikan manfaat yang signifikan dalam menjalin hubungan komunikasi dengan berbagai lapisan masyarakat. Fitur unggulan yang disediakan oleh Facebook untuk interaksi adalah perbaruan status. Perbaruan status menjadi salah satu opsi yang dapat dimanfaatkan untuk berbagi informasi mengenai aktivitas yang sedang dijalani. Setelah status berhasil diposting, teman-teman yang terhubung dapat dengan mudah memberikan tanggapan dalam bentuk komentar. Inilah yang membentuk komunikasi melalui *platform* Facebook menggunakan tulisan sebagai media komunikasi utama (Setiawan & Zyuliantina, 2020).

Facebook didirikan oleh Mark Zuckerberg bersamarekannya sekamar dan sesama mahasiswa ilmu komputer yaitu Eduardo Saverin, Dustin Moskovitz, dan Chris Hughes. Awalnya, keanggotaan di situs web ini dibatasi hanya untuk mahasiswa Harvard, kemudian diperluas untuk mencakup perguruan tinggi lain di Boston, Ivy League, dan Universitas Stanford. Pada bulan Januari 2011, jumlah pengguna aktif Facebook mencapai lebih dari 600 juta. Data statistik tentang pengguna menunjukkan bahwa Amerika Serikat menduduki peringkat teratas dengan jumlah pengguna sekitar 146 juta dan tingkat penetrasi sebesar 47,25 persen. Di tingkat global, Jepang berada di peringkat 53. Namun, cukup mengejutkan bahwa jumlah pengguna Facebook di Indonesia menduduki peringkat kedua dengan jumlah sekitar 33 juta pengguna, diikuti oleh Inggris di peringkat ketiga dengan 27 juta pengguna. Tingkat penetrasi Facebook di Indonesia mencapai 13,96 persen, sementara di Inggris mencapai 44,1 persen (Patria & Yulianto, 2011).

Facebook menyediakan berbagai fitur di antaranya yaitu sebagai sarana mengakses berita, video, promosi produk, atau untuk pelaku bisnis, menonton film, serta berpartisipasi dalam grup. Selain itu juga terdapat beberapa pengaruh yang dapat diperoleh dari penggunaan Facebook, seperti menjadi sarana promosi, wadah diskusi, alat interaksi sosial, memudahkan berhubungan dengan teman, mempererat hubungan dengan kenalan, memperluas jaringan sosial, serta memfasilitasi pertukaran ide dan informasi secara efektif (Fitriana et al., 2020).

### 2.3 Analisis Sentimen

Analisis sentimen adalah proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna mendapatkan informasi mengenai sentimen yang terkandung dalam suatu pernyataan pendapat. Pengaruh besar dan manfaat yang dihasilkan dari analisis sentimen telah mendorong perkembangan pesat dalam penelitian dan aplikasi yang berfokus pada analisis sentimen (Buntoro, 2017).

Ada beberapa tantangan yang dihadapi dalam melakukan analisis sentimen. Tantangan yang pertama adalah kata-kata pendapat yang dianggap positif dalam satu situasi bisa jadi dianggap negatif dalam situasi yang berbeda. Tantangan kedua adalah bahwa orang tidak selalu mengungkapkan pendapat dengan cara yang sama. Sebagian besar pengolahan teks tradisional mengandalkan asumsi bahwa perbedaan kecil antara dua bagian teks tidak akan mengubah makna secara sangat signifikan. Namun, dalam analisis sentimen, pernyataan "gambar tersebut bagus" memiliki makna yang sangat berbeda dari "gambar tersebut tidak bagus". Orang dapat memiliki pernyataan yang kontradiktif dalam pendapat mereka. Sebagian besar ulasan akan mengandung komentar positif dan negatif, yang dapat diatasi dengan menganalisis kalimat satu per satu (Vinodhini & Chandrasekaran, 2012). Gambar 2.1 menunjukkan ilustrasi dari analisis sentimen.



Gambar 0.1 Analisis Sentimen

Sumber: (Amalopo, 2021)

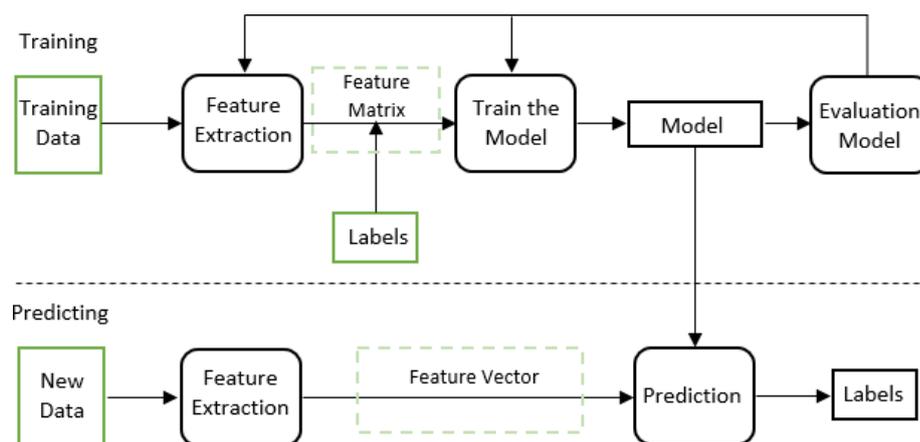
### 2.4 Machine Learning

*Machine learning* (ML) merupakan cabang ilmu yang terus berkembang dalam memahami pola dan teori komputasi dalam kecerdasan buatan (*AI/Artificial Intelligence*). ML melibatkan pembelajaran dan pengembangan algoritma yang mampu memahami dan melakukan prediksi terhadap kumpulan data. Proses ini melibatkan model yang menggunakan

input untuk menghasilkan prediksi atau keputusan berdasarkan data, bukan mengikuti instruksi program statis (Ramadhanty, 2021). *Machine Learning* memiliki beberapa jenis, yaitu:

### 1. *Supervised Learning*

*Supervised learning* adalah metode *machine learning* yang fokus pada pengenalan pola yang menghubungkan input dengan *output* berdasarkan contoh pasangan input-*output*. Ini menciptakan fungsi dari data pelatihan yang telah diberi label, yang terdiri dari serangkaian sampel pelatihan. Dalam *supervised machine learning*, digunakan algoritma yang memerlukan bantuan eksternal. *Dataset* dibagi menjadi dua bagian, yaitu *train dataset* dan *test dataset*. Algoritma mempelajari berbagai jenis pola dari *dataset* pelatihan dan kemudian mengaplikasikannya pada *dataset* pengujian untuk melakukan prediksi atau klasifikasi (Batta, 2018). Skema kerja algoritma *supervised machine learning* diilustrasikan pada Gambar 2.2.



Gambar 0.2 Alur algoritma *supervised machine learning*

### 2. *Unsupervised Learning*

Dalam jenis pembelajaran *unsupervised learning*, sistem diberikan sejumlah sampel input tetapi tanpa adanya *output* yang ditentukan. Karena tidak ada keluaran yang diharapkan, tugasnya adalah untuk mengelompokkan data sedemikian rupa sehingga algoritma dapat membedakan dengan tepat antara kelompok data yang berbeda. *Unsupervised learning* terfokus pada pencarian pola atau struktur tersembunyi dalam data yang tidak memiliki label. Lebih lanjut, *unsupervised learning* dibagi menjadi dua jenis masalah utama, yaitu *clustering* (pengelompokan) dan asosiasi. Masalah *clustering*

digunakan untuk mengidentifikasi kelompok yang ada dalam data, seperti mengelompokkan pelanggan berdasarkan pola pembelian mereka. Sementara itu, masalah asosiasi mencari aturan yang menggambarkan hubungan antara sebagian besar data yang ada, seperti orang yang membeli produk A juga cenderung membeli produk B (Roihan et al., 2020).

## 2.5 *Support Vector Machine*

*Support Vector Machine* (SVM) merupakan teknik yang awalnya dikembangkan oleh Boser, Guyon, dan Vapnik, dan pertama kali dipresentasikan pada tahun 1992 dalam acara *Annual Workshop on Computational Learning Theory*. Prinsip dasar dari SVM sebenarnya adalah penggabungan yang seimbang dari konsep-konsep komputasi yang telah ada selama beberapa dekade sebelumnya. Sebagai contoh, gagasan tentang kernel telah diperkenalkan oleh Aronszajn pada tahun 1950, begitu pula dengan aspek-aspek pendukung lainnya. Namun, hingga tahun 1992, belum ada usaha konkret untuk mengintegrasikan komponen-komponen ini menjadi satu kesatuan yang kokoh (Kamal, 2021).

Beberapa jenis kernel yang sering digunakan dalam metode *Support Vector Machine* meliputi:

### a. *Kernel Linear*

Kernel ini bekerja dengan menghitung produk titik (*dot product*) dari dua vektor data dalam ruang fitur asli. Ini berarti itu mencoba untuk menemukan garis linear yang memisahkan dua kelas. Jika garis linear tersebut ada, maka SVM akan menemukannya. Berikut rumus dari *linear kernel* dinyatakan pada persamaan (2.1).

$$K(X_i, X_j) = (X_i \cdot X_j) \quad (2.1)$$

### b. *Kernel Polynomial*

*Kernel polynomial* adalah jenis fungsi kernel yang digunakan ketika data tidak dapat dipisahkan secara linear. Biasanya, metode kernel ini efektif digunakan pada situasi ketika *dataset* pelatihan telah dinormalisasi. Berikut rumus dari *polynomial kernel* dinyatakan pada persamaan (2.2).

$$K(X_i, X_j) = (X_i \cdot X_j + 1)^d \quad (2.2)$$

### c. *Kernel Radial Basis Function (RBF)*

*Kernel RBF* bekerja dengan menghitung jarak antara titik data dalam ruang fitur. Ini mengubah data ke dalam representasi yang tak terhingga dimensi dengan menghitung jarak ke pusat titik data tertentu. Kernel ini sangat fleksibel dan cocok untuk masalah klasifikasi yang tidak linier. Berikut rumus dari *polynomial kernel* dinyatakan pada persamaan (2.3).

$$K(X_i, X_j) = \exp(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2) \quad (2.3)$$

### d. *Kernel Sigmoid*

*Kernel sigmoid* digunakan saat data tidak dapat dipisahkan secara linear. Ini adalah pengembangan dari jaringan saraf tiruan dan menampilkan sifat kurva berbentuk "S". Kernel ini digunakan terutama dalam kasus klasifikasi biner yang melibatkan dua kelas. Berikut rumus dari *sigmoid kernel* dinyatakan pada persamaan (2.4).

$$K(X_i, X_j) = \tanh(\gamma(X_i \cdot X_j) + c) \quad (2.4)$$

Keterangan:

- $K(X_i, X_j)$  : fungsi *kernel*
- $X_i \cdot X_j$  : *dot product* vektor  $X_i$  dan  $X_j$
- $d$  : derajat *polynomial*
- $\gamma$  : parameter *gamma*
- $c$  : koefisien

## 2.6 *Confusion Matrix*

*Confusion Matrix* merupakan suatu teknik yang sering digunakan untuk mengukur tingkat akurasi dalam konteks *data mining*. *Confusion matrix* diilustrasikan dalam bentuk tabel yang mengindikasikan jumlah data uji yang berhasil diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan dengan kesalahan, yang diperlukan untuk menghitung metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* untuk mengevaluasi kinerja model klasifikasi (Rahman et al., 2017). Tabel 2.2 merupakan tabel dari *Confusion matrix*.

Tabel 2. 2 Tabel *Confusion Matrix*

| Label   | Positif                    | Negatif                    |
|---------|----------------------------|----------------------------|
| Positif | <i>True Positive (TP)</i>  | <i>False Positive (FP)</i> |
| Negatif | <i>False Negative (FN)</i> | <i>True Negative (TN)</i>  |

Dengan merujuk tabel *confusion matrix* tersebut:

- True Positive (TP)* merujuk pada situasi di mana model berhasil memprediksi dengan benar suatu *instance* sebagai positif saat memang benar-benar positif.
- True Negative (TN)* merujuk pada situasi di mana model berhasil memprediksi dengan benar suatu *instance* sebagai negatif saat memang benar-benar negatif.
- False Positive (FP)* merujuk pada situasi di mana model secara tidak tepat memprediksi suatu *instance* sebagai positif ketika sebenarnya negatif.
- False Negative (FN)* merujuk pada situasi di mana model secara tidak tepat memprediksi suatu *instance* sebagai negatif ketika sebenarnya positif.

Ada beberapa metrik yang dapat digunakan untuk mengukur kinerja model klasifikasi *machine learning*, termasuk akurasi (*accuracy*), presisi (*precision*), dan *recall*. Model klasifikasi yang telah dibuat adalah suatu proses pemetaan dari baris data ke hasil prediksi kelas untuk data tersebut. Klasifikasi yang menghasilkan dua label keluaran disebut sebagai klasifikasi biner, yang sering kali direpresentasikan dengan  $\{0,1\}$ ,  $\{\text{ya, tidak}\}$ ,  $\{+1,-1\}$ , atau  $\{\text{positif; negatif}\}$  untuk setiap data masukan yang dianalisis.

a. *Accuracy*

Akurasi mengukur sejauh mana model mampu memprediksi dengan benar seluruh kelas data, termasuk kelas positif dan negatif. Rumus akurasi dapat ditemukan dalam persamaan berikut (2.5):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.5)$$

b. *Precision*

Presisi mengukur sejauh mana prediksi positif model adalah akurat, yaitu berapa persen dari semua prediksi positif yang tepat. Rumus presisi dapat ditemukan dalam persamaan berikut (2.6):

$$Precision = \frac{TP}{TP + FP} \quad (2.6)$$

c. *Recall*

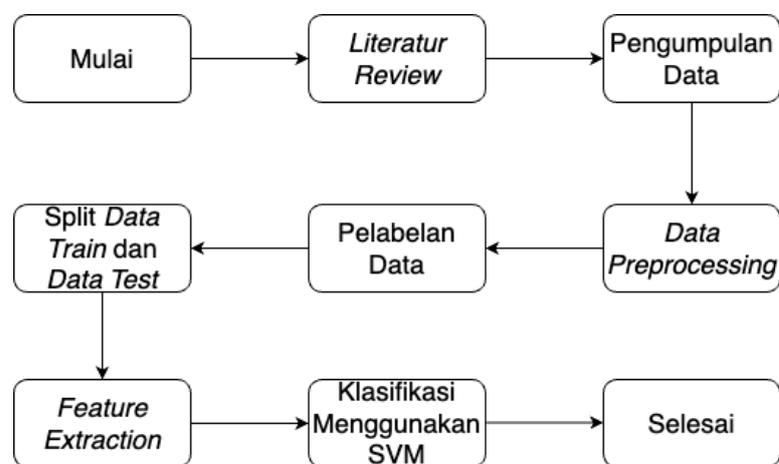
*Recall* mengukur sejauh mana model dapat mengenali semua data positif yang sebenarnya ada. Dengan kata lain, ini mengukur berapa persen dari seluruh data positif yang berhasil diidentifikasi oleh model. Rumus *recall* dapat ditemukan dalam persamaan berikut (2.7):

$$Recall = \frac{TN}{TN + FP} \quad (2.7)$$

## BAB III METODOLOGI PENELITIAN

### 3.1 Tahap Penelitian

Gambar 3.1 menunjukkan urutan langkah-langkah yang dilakukan selama penelitian. Seluruh proses analisis dan representasi visual data dilakukan setelah data dikumpulkan dan melewati tahap prapemrosesan. Dalam penelitian ini, tahapan prapemrosesan hingga tahap evaluasi menggunakan bahasa pemrograman *Python*.



Gambar 3.1 Tahap Penelitian

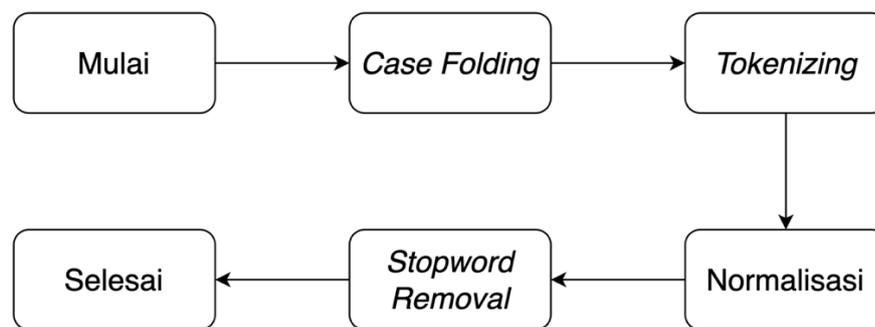
Proses penelitian dimulai dengan langkah awal pengumpulan data ulasan dari halaman Facebook masing-masing calon presiden yang akan digunakan sebagai *dataset*. Kemudian, *dataset* komentar dari postingan Facebook masing-masing calon presiden yang telah terkumpul akan melewati serangkaian tahapan *preprocessing*, termasuk penyalarsan huruf, tokenisasi, formalisasi, dan *stopword removal*. Setelah data berhasil dibersihkan melalui tahap *preprocessing*, langkah selanjutnya adalah pelabelan data, di mana setiap komentar akan dikelompokkan menjadi kategori positif atau negatif. Setelah proses pelabelan selesai, data akan menjalani tahap klasifikasi dengan menggunakan metode *Support Vector Machine* (SVM). Selanjutnya, hasil dari tahap klasifikasi akan dievaluasi dalam tahap pengujian untuk mengevaluasi performa sistem yang telah dibangun.

### 3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari halaman Facebook milik Prabowo Subianto, Anies Baswedan, dan Ganjar Pranowo. Pengambilan data dilakukan dengan cara menyalin tautan yang akan diambil, kemudian proses pengambilan data akan berlangsung secara otomatis melalui *platform* Apify. Lalu, data akan disimpan dalam format CSV di direktori yang telah ditentukan sebelumnya. Setelah data berhasil diambil, data akan dikelompokkan berdasarkan topik bahasannya.

### 3.3 Data Preprocessing

Pra-pemrosesan data atau *data preprocessing* merupakan tahap krusial dalam analisis teks yang memiliki tujuan untuk membersihkan dan menyiapkan data teks sebelum dilakukan pemrosesan lebih lanjut. Beberapa langkah yang terlibat dalam tahap pra-pemrosesan ditunjukkan pada Gambar 3.2.



Gambar 3.2 Tahapan dari Pra-pemrosesan data

*Preprocessing* bertujuan untuk mencegah kerugian hilangnya informasi, redundansi data, dan ketidakkonsistenan data. Dalam penelitian ini, terdapat beberapa langkah *preprocessing* yang akan dilaksanakan seperti berikut:

#### a. Case Folding

*Case Folding* adalah tahap di mana semua huruf dalam teks diubah menjadi huruf kecil atau huruf besar, sehingga tidak ada lagi masalah yang timbul akibat perbedaan kapitalisasi dalam analisis teks.

b. *Tokenizing*

Proses tokenisasi adalah tahap di mana teks dibagi menjadi unit-unit yang lebih kecil seperti kata, frasa, atau kalimat. Tujuannya adalah untuk memfasilitasi pengolahan teks selanjutnya, seperti menerapkan analisis statistik atau pembuatan model bahasa.

c. *Normalisasi*

Normalisasi adalah tindakan mengganti atau memperbaiki kata-kata yang disingkat atau salah eja. Penggantian kata ini bertujuan untuk menghindari peningkatan jumlah dimensi kata yang tidak perlu. Peningkatan dimensi kata dapat terjadi jika kata-kata yang salah eja atau disingkat tidak diperbaiki, meskipun sebenarnya kata-kata tersebut memiliki arti yang sama. Namun, tanpa perbaikan, kata-kata tersebut akan dianggap sebagai entitas yang berbeda dalam proses pembuatan matriks.

d. *Stopword Removal*

*Stopword* pada dasarnya adalah kata-kata yang sangat umum dan tidak memiliki nilai informasi yang signifikan, sehingga dapat diabaikan atau dihapus. Dalam proses ini, kata-kata *stopword* yang ada dalam data komentar akan dihapus, sehingga data hanya akan mengandung informasi yang relevan dan bermakna (Pradana & Hayaty, 2019).

### 3.4 Pelabelan Kelas Sentimen

Setelah menyelesaikan tahap *preprocessing*, langkah berikutnya adalah tahap pelabelan untuk analisis sentimen yang dilakukan oleh penulis menggunakan *Python*. Tahap pelabelan dilaksanakan setelah tahap *preprocessing* karena data yang akan diberi label sudah dalam keadaan bersih. Tujuan dari pelabelan adalah untuk mengelompokkan atau mengklasifikasikan data yang ada ke dalam kategori-kategori berdasarkan polaritasnya. Data-data yang telah melewati proses *preprocessing* akan ditempatkan ke dalam salah satu dari dua kategori yang tersedia, yaitu positif dan negatif.

### 3.5 *Split Data Train dan Data Test*

Pembuatan *data train* dan *data test* memiliki peran yang sangat penting dalam melatih performa mesin pembelajaran. *Data train* dan *data test* akan berdampak pada tingkat akurasi yang dapat dicapai oleh mesin pembelajaran. Pada tahap pembelajaran (*learning*), sebagian dari data yang telah memiliki label kelasnya (*training set*) digunakan untuk mengembangkan model. *Dataset* pelatihan dibuat dengan mengelompokkan seluruh data yang berasal dari

komentar tiap calon presiden ke dalam dua kategori, yaitu komentar positif dan komentar negatif.

### 3.6 *Feature Extraction*

*Feature extraction* adalah proses identifikasi dan ekstraksi fitur-fitur penting dari data mentah untuk membangun representasi yang lebih sederhana dan bermakna. Tujuannya adalah meningkatkan efisiensi dan kinerja model atau mempersiapkan data untuk analisis lebih lanjut. Fitur-fitur ini dapat mencakup atribut-atribut kritis atau representasi yang lebih relevan, dan proses ini sering melibatkan teknik matematika atau statistik seperti pemilihan fitur, transformasi, atau ekstraksi fitur yang lebih kompleks. *Feature extraction* sangat penting dalam konteks pembelajaran mesin dan analisis data untuk memastikan bahwa model dapat memahami dan memproses informasi secara optimal.

### 3.7 **Klasifikasi Menggunakan SVM**

Proses klasifikasi melibatkan penggunaan data pelatihan dan data pengujian dalam pembuatan model *machine learning*. Model *machine learning* dibangun melalui proses pelatihan menggunakan *data train*, yang kemudian dapat diuji dengan menggunakan *data test*. Pengujian model dilakukan untuk menilai tingkat akurasi model dan seberapa baik model dapat melakukan klasifikasi yang benar. Dalam pemilihan metode *machine learning*, ada beberapa algoritma yang dapat digunakan, dan dalam penelitian ini, metode algoritma yang digunakan adalah *Support Vector Machine* (SVM). SVM merupakan algoritma *supervised learning*, yang berarti algoritma ini membutuhkan data pelatihan yang sudah diberi label (label kelas atau keluaran yang diinginkan) untuk mengembangkan model. SVM memanfaatkan data pelatihan ini untuk mempelajari batasan keputusan yang memisahkan kelas atau kelompok data yang berbeda. Kemudian, model SVM dapat digunakan untuk memprediksi label kelas data baru berdasarkan apa yang telah dipelajari dari data pelatihan.

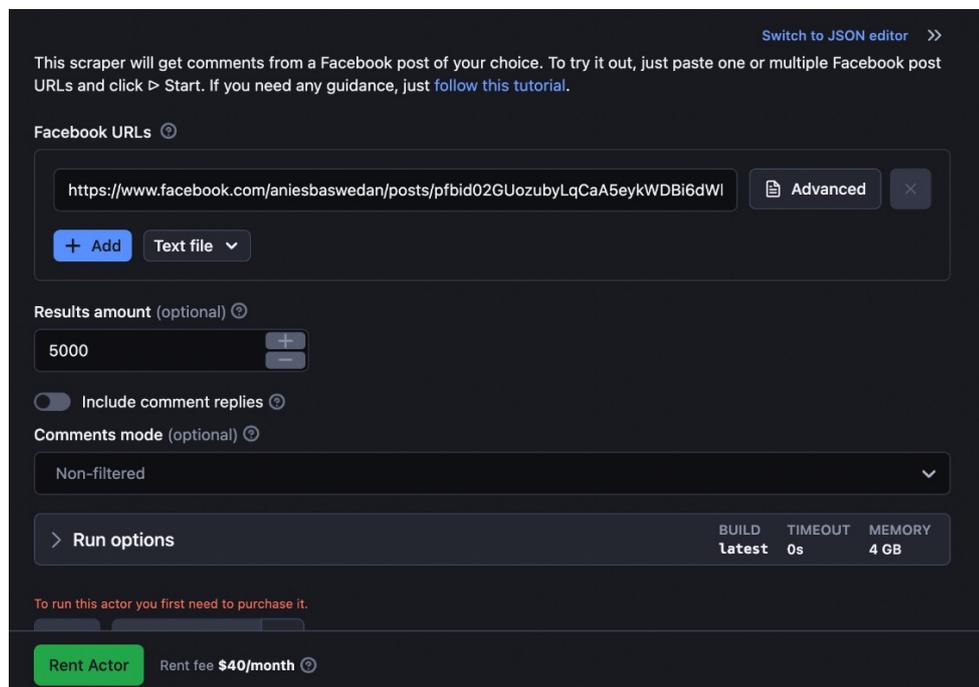
Ada beberapa metrik yang dapat digunakan untuk mengukur kinerja model klasifikasi, termasuk akurasi (*accuracy*), presisi (*precision*), dan *recall*. Model klasifikasi yang telah dibuat merupakan suatu proses pemetaan dari baris data ke hasil prediksi kelas untuk data tersebut. Klasifikasi yang menghasilkan dua label keluaran disebut sebagai klasifikasi biner, yang sering kali direpresentasikan dengan  $\{0,1\}$ ,  $\{\text{ya, tidak}\}$ ,  $\{+1,-1\}$ , atau  $\{\text{positif; negatif}\}$  untuk setiap data masukan yang dianalisis.

Dalam proses evaluasi klasifikasi, terdapat empat skenario yang mungkin terjadi sebagai hasil dari klasifikasi data. Jika data positif dan diprediksi sebagai positif, ini akan dianggap sebagai *true positive*, sementara jika data positif tetapi diprediksi sebagai negatif, ini akan dianggap sebagai *false negative*. Di sisi data negatif, jika diprediksi sebagai negatif, ini akan dianggap sebagai *true negative*, tetapi jika diprediksi sebagai positif, ini akan dianggap sebagai *false positive* (Fawcett, 2006).

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Pengumpulan Data

*Scraping* merupakan metode yang digunakan dalam proses pengambilan data. Data diambil dengan cara membuka *page* tiap calon presiden, meng-*copy* tautan yang akan diambil, setelah itu proses pengambilan data akan dilakukan secara otomatis menggunakan *platform* Apify. Berikut ini adalah instruksi untuk mengambil komentar dari halaman Facebook calon presiden Indonesia tahun 2024 yang ditampilkan pada Gambar 4.1.



Gambar 4.1 Pengambilan Data di *apify*

Data komentar yang akan diambil melalui proses *scraping* adalah komentar dari periode mulai 1 Juni 2023 hingga 31 Juli 2023 dari akun Facebook Anies Baswedan, Ganjar Pranowo, dan Prabowo Subianto. Berikut pada Tabel 4.1 merupakan total data postingan dari setiap *page* Facebook calon presiden Indonesia tahun 2024 yang menunjukkan bahwa terdapat 124 postingan dengan lebih dari 39.000 komentar pada *page* Ganjar Pranowo. Sedangkan Anies Baswedan hanya memiliki 64 postingan dengan lebih dari 27.000 komentar dan Prabowo Subianto memiliki 89 postingan dengan lebih dari 105.000 komentar di dalamnya.

Tabel 4.1 Jumlah Postingan dan Komentar

| <b>Nama</b>      | <b>Jumlah Postingan</b> | <b>Komentar</b> |
|------------------|-------------------------|-----------------|
| Ganjar Pranowo   | 124                     | 39.265          |
| Anies Baswedan   | 64                      | 27.340          |
| Prabowo Subianto | 89                      | 105.635         |

Salah satu area penting yang perlu diperhatikan adalah potensi bias terkait keterkaitan antara jumlah pengikut seorang pengguna dengan preferensi politiknya terhadap calon presiden tertentu. Jumlah pengikut di *platform* media sosial tidak selalu mencerminkan kecenderungan politik yang sebenarnya, karena seseorang dapat mengikuti orang lain atas alasan yang tidak berhubungan dengan politik, seperti hiburan atau popularitas.

Setelah data komentar berhasil diambil, lalu data dikelompokkan berdasarkan isi bahasan dari tiap postingan yang ada. Dengan cara ini, penulis dapat memahami apa yang sedang dibahas dalam setiap postingan dan bagaimana komentar-komentar tersebut berhubungan dengan topik tersebut. Untuk mengelompokkan postingan menjadi sebuah topik bahasan, diperlukan setidaknya 5 postingan yang memiliki bahasan sejenis. Apabila kurang dari 5 postingan, maka akan diklasifikasikan sebagai topik lain-lain. Ganjar Pranowo membuat 124 postingan yang kemudian dikelompokkan menjadi 6 topik postingan, yaitu "kebijakan gubernur" yang terdiri dari 18 postingan dan 5.088 komentar, "bertemu masyarakat" terdiri dari 37 postingan dengan 6.871 komentar, "bertemu publik figur" terdiri dari 25 postingan dengan 14.808 komentar, "pembangunan infrastruktur" terdiri dari 8 postingan dengan 1.890 komentar, "menghadiri acara" terdiri dari 14 postingan dengan 6.998 komentar, dan "lain-lain" terdiri dari 22 postingan dengan 3.610 komentar. Topik bahasan "lain-lain" Ganjar Pranowo terdiri dari kumpulan topik yang tidak termasuk ke dalam topik lainnya, seperti "ucapan hari besar", "cerita tentang masa lalu", "tip LDR", "menonton pertandingan sepak bola", "menerima penghargaan", dan "tip untuk perantau". Anies Baswedan memiliki 64 postingan yang dibagi menjadi 5 topik postingan yaitu, "bertemu masyarakat" terdiri dari 14 postingan dengan 4.140 komentar, "bertemu publik figur" terdiri dari 14 postingan dengan 9.533 komentar, "melaksanakan ibadah haji" terdiri dari 7 postingan dengan 6.765 komentar, "menghadiri acara" terdiri dari 20 postingan dengan 5.708 komentar, "lain-lain" terdiri dari 9 postingan

dengan 1.194 komentar. Topik bahasan "lain-lain" Anies Baswedan terdiri dari "ucapan hari besar", "menerima penghargaan", "demo dan resep memasak", dan "cerita tentang pahlawan". Prabowo Subianto memiliki 89 postingan yang dibagi menjadi 5 topik postingan, yaitu "bertemu masyarakat" terdiri dari 19 postingan dengan 18.660 komentar, "bertemu publik figur" terdiri dari 26 postingan dengan 33.658 komentar, "menghadiri acara" terdiri dari 29 postingan dengan 34.727 komentar, "masa-masa di TNI" terdiri dari 5 postingan dengan 5.999 komentar, "lain-lain" terdiri dari 10 postingan dengan 12.591 komentar. Topik bahasan "lain-lain" Prabowo Subianto terdiri dari "cerita tentang masa lalu", "cerita tentang pahlawan", dan "melaksanakan shalat Eid". Tabel 4.2 merupakan contoh postingan yang diambil dari tiap calon presiden. Untuk topik "bertemu masyarakat" kata-kata yang digunakan adalah "sambutan masyarakat", "diskusi bersama warga", "menyapa masyarakat", topik "bertemu publik figur" kata-kata yang digunakan adalah "bertemu", "mengundang", "mampir", topik "menghadiri acara" kata-kata yang digunakan adalah "menghadiri acara", "menghadiri syukuran", "hadir di acara", "menghadiri", topik "lain-lain" kata-kata yang digunakan adalah "selamat hari", "kulineran", "selamat hari lahir", topik "kebijakan gubernur" kata-kata yang digunakan adalah "menindaklanjuti", "intruksikan", "program gubernur", topik "pembangunan infrastruktur" kata-kata yang digunakan adalah "anggaran", "diperbaiki", "bantuan keuangan", topik "melaksanakan ibadah haji" kata-kata yang digunakan adalah "haji", topik "masa-masa di TNI" kata-kata yang digunakan adalah "saat di TNI", "saat di akabri". Rekap topik postingan tiap calon presiden ditunjukkan pada Tabel 4.3, Tabel 4.4, dan Tabel 4.5.

Tabel 4.2 Contoh Postingan Calon Presiden

| Topik Bahasan      | Contoh Postingan  |
|--------------------|---|
| Bertemu Masyarakat | <div data-bbox="555 1496 608 1547"></div> <p data-bbox="616 1496 916 1547"><b>Ganjar Pranowo</b> is in <b>Banjarmasin</b>.<br/>28 July · 🌐 ...</p> <p data-bbox="555 1559 1193 1610">Terima kasih atas sambutan hangatnya saudara-saudaraku di Banjarmasin. Senyum manis adik ini melekat sampai saya pulang.</p> <div data-bbox="555 1641 608 1693"></div> <p data-bbox="616 1641 879 1693"><b>Anies Baswedan</b> is in <b>Jakarta</b>.<br/>18 July · 🌐 ...</p> <p data-bbox="555 1704 1222 1776">Alhamdulillah, kembali dari haji kami sudah niatkan untuk bisa kembali bersilaturahmi dengan warga Kampung Tanah Merah dan Kampung Akuarium, Jakarta Utara.</p> <p data-bbox="555 1787 1230 1859">Alhamdulillah apa yg diikhtiarkan untuk di tingkat Jakarta berhasil kita jalankan. Insya Allah kita akan tingkatkan lagi untuk di seluruh Indonesia agar warga bisa hidup tenang sampai generasi yang akan datang.</p> <p data-bbox="555 1870 1238 1942">Perjuangan kita belum selesai, ayo kita tuntaskan sama-sama. Mari kita kirimkan pesan pada semua orang. Sampaikan bahwa keadilan itu tidak hanya diucapkan saat upacara, tapi direalisasikan dalam kehidupan nyata.</p> |

|                      |  |
|----------------------|--|
|                      | <p> <b>Prabowo Subianto</b> <span>✓</span><br/>24 July · 🌐</p> <p>Bersama <a href="#">Presiden Joko Widodo</a> mengunjungi, menyapa masyarakat dan para pedagang pasar, sekaligus menyalurkan bantuan langsung tunai di Pasar Bululawang, Malang, Jawa Timur.</p> <p>Bahagia melihat senyum dan tawa rakyat, tugas kita sebagai seorang pemimpin yaitu menghadirkan senyum di wajah rakyat, wong cilik iso gemuyu.</p>  |
| Bertemu Publik Figur | <p> <b>Ganjar Pranowo</b> <span>✓</span><br/>31 July · 🌐</p> <p>Terima kasih para Purnawirawan TNI Polri yg sudah mengundang saya.</p> <p>Kita hanya berbagi cerita, bagaimana menjadi satu Indonesia ke depannya. Karena tantangan kita adalah perang baru di masa depan, yg jelas bisa dihadapi jika kita mengesampingkan perbedaan.</p> <p>Kita semua satu, untuk Indonesia.</p> <p> <b>Anies Baswedan</b> is at <b>Pangandaran, Jawa Barat, Indonesia.</b><br/>26 July · Pangandaran · 🌐</p> <p>Kenapa naik pickup sama Bu Susi? Jadi begini, sesudah videonya bu Susi nyetir sendiri pickup buat antar cucu-cucunya berenang jadi viral, siapa sih yang nggak mau merasakan disetirin bu Susi? Saya pun request, mau coba juga. Memang jago sih. Bonusnya, percakapan seru sepanjang jalan!</p> <p> <b>Prabowo Subianto</b> <span>✓</span><br/>31 July · 🌐</p> <p>Terima kasih sudah mampir dan makan siang di kantor Kementerian Pertahanan Republik Indonesia, Mas <a href="#">Raffi Ahmad</a> !</p>   |
| Menghadiri Acara     | <p> <b>Ganjar Pranowo</b> is in <b>Wonosobo.</b><br/>15 July · 🌐</p> <p>Sejarah dan komitmen GP Ansor terhadap Republik ini tidak perlu lagi dipertanyakan.</p> <p>Orientasi perjuangannya bukan sekadar berdasarkan program, tapi juga berdasar pada spirit keagamaan yang dipadu dengan spirit kemanusiaan.</p> <p>Kemah Bakti GP Ansor dan Banser ini jadi ruang penguat antar anggota. Agar GP Ansor dan Banser selalu tegak lurus untuk menjaga kiai sekaligus menjaga negeri.</p> <p>Sukses selalu rekan-rekan GP Ansor dan Banser Kab. Wonosobo. Semoga kita selalu diberi kekuatan untuk terus berbuat kebaikan.</p> <p> <b>Anies Baswedan</b> is at <b>Kuningan City.</b><br/>31 July · Jakarta · 🌐</p> <p>Hadir di acara Menjadi Manusia, acara anak muda bicara tentang mental health. Sebagai anggota Gen X, saya belajar banyak.</p> <p>Mendukung mental health anak muda itu bukan cuma mendukung ketika sudah jadi masalah. Tapi yang penting juga, meringankan beban pikiran sejak awal.</p> <p>Bantu sediakan hunian terjangkau, transportasi publik nyaman dan murah, mudah dapat pendidikan dan pekerjaan, jaga kualitas udara, jamin kebebasan berpendapat, sediakan ruang publik buat healing dan interaksi, bahkan bantu pengobatan orangtuanya yang kesulitan. Intinya: jaga iklim dan suasana yang mendukung kesehatan mental, bikin lebih tenang.</p> <p>Apa lagi masalah yang dihadapi teman-teman dan bagaimana pemerintah bisa membantu? Mari diskusi.</p> <p>Bonus: foto terakhir ketemu sobat lama. 😊</p> <p> <b>Prabowo Subianto</b> was live.<br/>30 July · 🌐</p> <p>Menghadiri Milad Partai Bulan Bintang Yang Ke 25</p> |

|                           |  |
|---------------------------|--|
| Lain-lain                 | <p> <b>Ganjar Pranowo</b> is in Semarang. 24 July · 🌐 ...</p> <p>Kita semua punya kewajiban untuk menyediakan ruang, memberi ruang, memahami, serta menciptakan rasa aman untuk anak-anak.</p> <p>Jika itu semua sudah terwujud, maka dunia ini akan menjadi tempat yg baik dan nyaman untuk tumbuh kembang generasi selanjutnya.</p> <p>Terima kasih, pada semua yg sudah mewujudkan itu di Jateng. Alhamdulillah, tiga kali berturut Jateng menerima predikat Provinsi Layak Anak. Kita jaga terus komitmennya.</p> <p>Selamat Hari Anak Nasional 2023!</p> <p> <b>Anies Baswedan</b> is in Indonesia. 23 July · 🌐 ...</p> <p>Selamat Hari Anak Nasional.</p> <p> <b>Prabowo Subianto</b> 24 July · 🌐 ...</p> <p>Selamat hari lahir ke-25 Partai Kebangkitan Bangsa. Terus bergerak melayani Indonesia.</p> <p>Stadion Manahan, 23 Juli 2023.</p> |
| Kebijakan Gubernur        | <p> <b>Ganjar Pranowo</b> 27 July · 🌐 ...</p> <p>Sekolah Dilarang Menahan Ijazah</p> <p>Sejak saya buka posko pengaduan pungutan sekolah. Sudah ada 842 aduan yang langsung kami tindaklanjuti ke pihak sekolah.</p> <p>Dari jumlah itu sebanyak 323 aduan sudah selesai dan 120 aduan berprogres utk diambil langkah penyelesaian.</p> <p>Dari 842 aduan itu ternyata tidak semua terkait pungutan. Ada juga soal ditahannya ijazah oleh sekolah karena siswa menunggak pembayaran.</p> <p>Pada kasus ini saya instruksikan seluruh sekolah untuk memutihkan tunggakan dan langsung menyerahkan ijazah. Karena ijazah adalah hak siswa.</p> <p>Jika sampai hari ini masih ada ijazah yang ditahan silahkan dilaporkan. Tentu saja yang bisa langsung saya eksekusi hanya sekolah SMA/SMK/SLB Negeri. Selebihnya adalah kewenangan temen-temen bupati/walikota atau Kemenag.</p> <p>Mohon doa semoga kami bisa bekerja dengan baik demi pendidikan yang lebih baik untuk anak-anak kita.</p>                          |
| Pembangunan Infrastruktur | <p> <b>Ganjar Pranowo</b> is in Temanggung. 28 July · 🌐 ...</p> <p>Jalan ruas Jumo-Muntung Temanggung sudah diperbaiki, tadinya bergelombang, sekarang sudah jd bagus.</p> <p>Ini adalah hasil bantuan keuangan ke Temanggung yg berjalan baik. Kira-kira anggarannya Rp5 miliar dan hasilnya bagus, mulus.</p> <p>PRnya ya kita tindak tegas truk-truk ODOL biar jalannya masa gunanya bisa lbh panjang lg. Krn salah satu penyebab jalan berlubang adalah truk yg muatannya berlebih.</p>   |
| Melaksanakan Ibadah Haji  | <p> <b>Anies Baswedan</b> 14 July · 🌐 ...</p> <p>Alhamdulillah, Rabu dini hari kemarin kami tiba di tanah air kami tiba di Tanah Air, setelah sejak tanggal 22 Juni memenuhi panggilan berhaji, menjadi salah satu dari jutaan tamu Allah SWT di tanah suci.</p> <p>Kami bersama Fery, Mutiara, Ali, Mikail, Kaisar, dan Ismail sungguh merasakan perjalanan spiritual yang sangat luar biasa. Ibadah haji ini selain menjadi momentum menjalankan perintah Allah SWT, juga memberikan ruang interaksi yang berkualitas bagi kami.</p> <p>Kami selalu bersama-sama, dalam dua minggu terakhir. Sesuatu yang selama ini menjadi hal yang langka kami rasakan, di tengah kesibukan masing-masing.</p>   |

|                  |   |
|------------------|---|
| Masa-masa di TNI |  <b>Prabowo Subianto</b> ✓<br>21 July · 🌐 <p>Sejak daftar di AKABRI, pada waktu itu, saya diambil sumpah sebagai prajurit. Sejak itu sudah diajarkan oleh pengasuh, pemimpin saya</p> <p>“Kau anak-anak muda, sadarkah kau? Bahwa kau masuk tentara, kau sudah teken mati”</p> <p>Nilai-nilai itu yang diajarkan dari hari pertama saya masuk tentara dan hingga saat ini masih tertanam di dalam diri saya, bahwa bagi seorang prajurit tidak ada kata pensiun. Bagi seorang prajurit selama nafas dikandung badan, kesetiaan hanya kepada rakyat dan bangsa Indonesia. Bahwa saya ingin selalu berbakti, berjuang dan melindungi segenap tumpah darah Indonesia.</p> <p>Prajurit bangsa akan selalu menjadi seorang prajurit, dan pada saatnya dibutuhkan, kita harus mengorbankan jiwa dan raga untuk rakyat, bangsa Indonesia.</p> |
|------------------|---|

Tabel 4.3 Pembagian Data Ganjar Pranowo Berdasarkan Topik Bahasan

| Topik Bahasan             | Jumlah Postingan | Jumlah Komentar |
|---------------------------|------------------|-----------------|
| Kebijakan Gubernur        | 18               | 5.088           |
| Bertemu Masyarakat        | 37               | 6.871           |
| Bertemu Publik Figur      | 25               | 14.808          |
| Pembangunan Infrastruktur | 8                | 1.890           |
| Menghadiri Acara          | 22               | 6.998           |
| Lain-lain                 | 22               | 3.610           |

Tabel 4.4 Pembagian Data Anies Baswedan Berdasarkan Topik Bahasan

| Topik Bahasan            | Jumlah Postingan | Jumlah Komentar |
|--------------------------|------------------|-----------------|
| Bertemu Masyarakat       | 14               | 4.140           |
| Bertemu Publik Figur     | 14               | 9.533           |
| Melaksanakan Ibadah Haji | 7                | 6.765           |
| Menghadiri Acara         | 20               | 5.708           |
| Lain-lain                | 9                | 1.194           |

Tabel 4.5 Pembagian Data Prabowo Subianto Berdasarkan Topik Bahasan

| Topik Bahasan        | Jumlah Postingan | Jumlah Komentar |
|----------------------|------------------|-----------------|
| Bertemu Masyarakat   | 19               | 18.660          |
| Bertemu Publik Figur | 26               | 33.658          |
| Masa-masa di TNI     | 5                | 5.999           |
| Menghadiri Acara     | 29               | 34.727          |
| Lain-lain            | 10               | 12.592          |

Hasil implementasi pengambilan *dataset* menggunakan Apify ditunjukkan pada Gambar 4.2. Terlihat bahwa data tersebut memiliki masalah seperti adanya karakter yang tidak relevan dan kata-kata yang tidak memiliki penggunaan yang jelas. Oleh karena itu, diperlukan tahap *preprocessing* untuk membersihkan dan mengatur data tersebut.

Mari kita lakukan perubahan bersama rakyat untuk kemajuan bangsa Indonesia 🇮🇩 .  
 Muantap... Semangat wujudkan perubahan  
 . Aamiin  
 Bismillah..semoga Allah SWT meridhoi dan memudahkan setiap ihtiar Pak Anies Baswedan  
 Aamiin  
 Sebaiknya memang hanya memilih partai pengusungnya bapak Anies Baswedan  
 Mantap idola km mga lancar trus sehat ya Pk Salam Restorasi tuk Indonesia ARB yang terbaik  
 Salam restorasi...  
 Bersatu.. Berjuang.. Menang.. !!!  
 Semoga slalu terbuka jalan kemudahan untuk setiap niat baik nya. 🙏🙏  
 Assalamu'alaikum..  
 Salam dan Do'a kami  
 dari Sumatera Barat..  
 Barakallahu fikum pak Anies Baswedan 🙏🙏🇮🇩🇲🇳  
 Semoga dimudahkan dan mendapat rahmat Allah Swt 🙏  
 Alhamdulillah , tetap semangat  
 Semangat pa,Anis,,,,,semoga setiap langkah perjuangan senantiasa di ridhoi oleh Allah SWT,  
 Barakallah bang Anies Baswedan, walaupun rute sulit, Insha Allah akan dimudahkan Allah SWT  
 Semoga perjuangan pk Anies yg di dukung oleh partai nasdem , dan di sertai oleh do'a kita semua , tercapai dg baik , lancar dan dg penuh kemudahan , aamiin yra .  
 Ttp p. anies yg trbaik cerdas, jujur, tegas, amanah, berprestasi

Gambar 4.2 Contoh hasil dari *scraping* menggunakan Apify

## 4.2 Data Preprocessing

Proses *preprocessing* dilakukan untuk menghindari kehilangan informasi dalam data, mengurangi redundansi data, dan mengatasi ketidakkonsistenan dalam data yang telah diambil melalui proses *scraping*. Hasil dari pengambilan data melalui *scraping* sering kali mengandung ketidakkonsistenan, seperti penggunaan gaya penulisan yang tidak standar, singkatan teks, dan simbol yang tidak relevan. Oleh karena itu, tahapan *preprocessing* diperlukan untuk menjadikan teks lebih konsisten. Contoh komentar dari data yang telah diambil dan akan masuk ke dalam tahap *preprocessing* dapat ditunjukkan pada Tabel 4.6.

Tabel 4.6 Contoh data hasil *scraping* yang akan dilakukan *preprocessing*

| Nomor | Komentar                                |
|-------|---|
| 1     | BAGUSSS PAK Ganjar Pranowo              |
| 2     | Semoga menjadi Pemimpin yang menyatukan |
| 3     | Prabowo selalu di hati                  |

### a. Case Folding

Langkah awal dalam tahap *preprocessing* pada penelitian ini adalah melakukan *case folding*, yang bertujuan untuk mengubah semua huruf menjadi huruf kecil. Proses *case folding* ini diimplementasikan menggunakan fungsi *Python* yang disebut "lower()". Hasil dari penerapan *case folding* ini dapat dilihat dalam Tabel 4.7. Sementara itu, penerapan *case folding* dapat diamati melalui penggunaan fungsi "lower()" yang ditunjukkan pada Gambar 4.3.

Tabel 4.7 Contoh komentar setelah dilakukan *case folding*

| Komentar Asli                         | Komentar setelah Case Folding         |
|---------------------------------------|---------------------------------------|
| BAGUSSS PK Ganjar Pranowo             | bagusss pk ganjar pranowo             |
| Semoga menjadi Pemimpin yg menyatukan | semoga menjadi pemimpin yg menyatukan |
| Prabowo selalu di hati                | prabowo selalu di hati                |

```
#mengubah huruf casefolding
def lowercase():
    data = df['clean_bersih'].str.lower()
    return data
results = uppercase()
print(results)
```

Gambar 4.3 Kode program *case folding*

Kode pada Gambar 4.3 adalah sebuah fungsi *Python* bernama `lowercase()` yang digunakan untuk mengubah teks dalam kolom `'clean_bersih'` dari sebuah *DataFrame* `df` ke huruf kecil (*lowercase*). Fungsi ini menggunakan fungsi `str.lower()` untuk melakukan transformasi ini, dan hasilnya disimpan dalam variabel `'data'`. Kemudian, nilai `'data'` dikembalikan sebagai hasil fungsi, dan hasilnya disimpan dalam variabel `'results'`. Akhirnya, hasilnya dicetak ke layar. Kode ini berguna untuk menormalisasi teks dalam *DataFrame* `df` agar semua kata berada dalam huruf kecil untuk konsistensi pemrosesan data.

#### b. *Tokenizing*

Tahap kedua adalah proses *tokenizing*, yang merupakan langkah untuk menguraikan teks atau dokumen menjadi unit-unit terkecil yang disebut "token". Token dapat berupa kata-kata, frasa, atau karakter terpisah, tergantung pada tingkat detail yang diinginkan. Kode pada Gambar 4.4 menggunakan *library Natural Language Toolkit* (NLTK) untuk melakukan tokenisasi teks dalam sebuah *DataFrame* `df`. Pertama, `"\w+"` mengidentifikasi kata-kata yang terdiri dari huruf, angka, dan garis bawah. Lalu, `"$[0-9]+\S+"` mengidentifikasi angka yang dimulai dengan tanda dolar, dan diikuti oleh satu atau lebih digit (0-9), diakhiri oleh karakter selain spasi (dengan `\"S+\"`). `"|"` digunakan untuk menggabungkan dua pola yang berbeda. Selanjutnya, tokenisasi diterapkan pada kolom `'clean_bersih'` dari *DataFrame* `df`. Hasil tokenisasi disimpan dalam kolom baru `'tokenize'` dalam *DataFrame* `df`. Akhirnya, *DataFrame* yang telah dimodifikasi dicetak

ke layar. Dengan demikian, kode ini melakukan tokenisasi teks dalam kolom 'clean\_bersih' dan menyimpan hasilnya dalam kolom baru 'tokenize' untuk pemrosesan teks lebih lanjut. Tabel 4.8 memperlihatkan hasil implementasi *tokenizing* tersebut.

```
from nltk.tokenize import RegexpTokenizer

# Create RegexpTokenizer
regex = RegexpTokenizer(r'\w+|${0-9}+\s+')

# Apply tokenization using lambda function
df['tokenize'] = df['clean_bersih'].apply(lambda x:
regex.tokenize(x))
print(df)
```

Gambar 4.4 Kode program *tokenizing*

Tabel 4.8 Contoh komentar setelah dilakukan *tokenizing*

| Komentar Sebelum                      | Komentar setelah Tokenizing                         |
|---------------------------------------|---|
| bagusss pk ganjar pranowo             | "bagusss", "pk", "ganjar", "pranowo"                |
| semoga menjadi pemimpin yg menyatukan | "semoga", "menjadi", "pemimpin", "yg", "menyatukan" |
| prabowo selalu di hati                | "prabowo", "selalu", "di", "hati"                   |

### c. Normalisasi

Normalisasi adalah langkah dalam mengubah kata-kata yang tidak sesuai ke bentuk standar sesuai yang terdapat dalam Kamus Besar Bahasa Indonesia (KBBI). Pada tahap ini, penulis menggunakan kamus khusus yang disimpan dalam format *file* CSV yang bernama *normalisasi.csv*. Kamus *normalisasi.csv* diambil dari GitHub pada tautan <https://github.com/denny-saryanto/analisis-sentimen-naive-bayes/blob/main/dataset/preprocessing/normalisasi.csv>. Kamus tersebut saat ini berisi 233 kata. Tujuan kamus ini adalah untuk mengoreksi bentuk kata yang tidak sesuai dengan KBBI, seperti kata-kata *slang word*, singkatan, dan kesalahan penulisan kata (*typo*).

Beberapa contoh kata normalisasi ditunjukkan pada Tabel 4.9. Kolom "sebelum" dalam tabel tersebut menampilkan contoh kata-kata yang kurang baku dan slang word yang sering digunakan di Facebook. Kemudian, kata-kata ini akan diubah menjadi bentuk yang lebih standar dalam kolom "sesudah". Proses normalisasi juga mencakup perubahan

kata-kata singkatan seperti mengganti kata 'gaje' dengan bentuk lengkapnya, yakni "tidak jelas".

Tabel 4.9 Contoh kata yang ada di kamus normalisasi

| Sebelum | Sesudah     |
|---------|-------------|
| yg      | yang        |
| ttp     | tetap       |
| bpk     | bapak       |
| msyrkt  | masyarakat  |
| diem    | diam        |
| gaje    | tidak jelas |

Kode pada Gambar 4.5 digunakan untuk melakukan normalisasi kata dalam sebuah *DataFrame* `df` dengan mengacu pada kamus normalisasi yang disimpan dalam file CSV "normalisasi.csv". Pertama, data normalisasi dibaca dari file CSV ke dalam *DataFrame* `normalisasi\_kata`. Kemudian, kode membuat sebuah *array* `normalisasi\_kata\_dict` yang berisi pasangan kata asli dan kata yang telah dinormalisasi. Selanjutnya, sebuah fungsi `normalized\_term(document)` didefinisikan, yang digunakan untuk mengganti kata-kata dalam dokumen dengan kata yang telah dinormalisasi berdasarkan kata yang terdapat dalam `normalisasi\_kata\_dict`. Hasilnya dicetak dengan menggunakan `df.head()`. Kode ini berguna untuk mengganti kata-kata dalam teks dengan bentuk yang telah dinormalisasi sesuai dengan daftar kata normalisasi yang ada dalam *file* "normalisasi.csv".

```

normalisasi_kata = pd.read_csv("normalisasi.csv")

normalisasi_kata_dict = {}

for index, row in normalisasi_kata.iterrows():
    if row[0] not in normalisasi_kata_dict:
        normalisasi_kata_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalisasi_kata_dict[term] if term in
normalisasi_kata_dict else term for term in document]

df['normalisasi'] = df['tokenize'].apply(normalized_term)
df.head()

```

Gambar 4.6 Kode program proses normalisasi

Tabel 4.10 menampilkan hasil dari proses normalisasi, di mana kata-kata tidak baku, singkatan, dan kata-kata slang telah dikurangi atau dihilangkan sepenuhnya. Juga di kasus lain terdapat beberapa kata yang tidak mengalami normalisasi yang memadai, seperti 'wkwkwk', 'kannnn'. Situasi ini terjadi dikarenakan banyaknya variasi dalam penggunaan kata-kata oleh pengguna Facebook yang berbeda-beda, dan karena kata-kata tersebut belum terdapat dalam kamus karena belum pernah tercatat penggunaannya. Sebagai contoh, kata 'wkwk' dapat ditulis hanya dengan satu huruf 'k' di akhir atau mungkin ditambahi dengan banyak huruf 'k'. Oleh karena itu, kata-kata yang tidak mengalami normalisasi akan ditambahkan secara manual untuk memperkaya kamus normalisasi, sehingga semua kata dalam kalimat nantinya dapat dinormalisasikan dengan benar. Total data normalisasi setelah penambahan manual menjadi 234 data. Adapun daftar data lengkap yang digunakan untuk normalisasi terdapat pada lampiran A.

Tabel 4.10 Contoh komentar setelah dilakukan normalisasi

| Komentar Sebelum                                       | Komentar setelah Normalisasi                             |
|--|--|
| "bagusss", "pk", "ganjar",<br>"pranowo"                | "bagusss", "pak", "ganjar", "pranowo"                    |
| "semoga", "menjadi", "pemimpin",<br>"yg", "menyatukan" | "semoga", "menjadi", "pemimpin",<br>"yang", "menyatukan" |
| "prabowo", "selalu", "di", "hati"                      | "prabowo", "selalu", "di", "hati"                        |

#### d. *Stopword Removal*

*Stopword Removal* adalah proses penyaringan atau pemilihan kata dalam dokumen dengan tujuan untuk mengurangi jumlah kata dalam korpus dengan menggunakan *stopwords*. *Stopwords* adalah kumpulan kata-kata yang cenderung kurang informatif atau memiliki dampak yang kurang signifikan secara keseluruhan dan sering muncul dalam dokumen. *Stopwords* ini biasanya terdiri dari kata-kata seperti kata ganti orang, kata penghubung, kata seruan, kata pertanyaan, dan kata-kata lainnya yang tidak memiliki makna penting dalam penentuan topik dokumen. Contoh kata-kata yang termasuk dalam proses *stopword removal* ini adalah sebagai berikut:

1. Kata hubung: jika, dan, namun, dan sebagainya.
2. Kata depan: antara, di, ke, dan sebagainya.

Proses ini dilakukan dengan memanfaatkan *library* yang terdapat dalam *Python* yang dikenal sebagai *sastrawi*. Sebelum *library sastrawi* digunakan, diperlukan penginstalan *library* ini dengan perintah "pip," sebagaimana yang diperlihatkan dalam Gambar 4.6.

```
pip install Sastrawi
```

Gambar 4.7 Kode program untuk menginstal *library sastrawi*

Pada tahap ini, penghapusan *stopword* dilakukan menggunakan fungsi bernama "menghapusKata(text)". Fungsi ini menerima sebuah *string* sebagai input dan bertujuan untuk menghapus kata-kata *stopword*. Fungsi *menghapusKata* terlihat dalam Gambar 4.7. Untuk menghilangkan *stopword*, digunakan dua sumber kata *stopword*, yaitu sumber dari *library Sastrawi* dan daftar kata yang ditambahkan secara manual. Untuk sumber *stopword* dari *library Sastrawi*, digunakan fungsi "StopWordRemoverFactory()" dan "create\_stop\_word\_remover()" untuk membuat penghapus *stopword*.

Setelah itu, *stopword* dari *library Sastrawi* disimpan dalam variabel bernama "stopword," sementara *stopword* yang dibuat secara manual disimpan dalam variabel "more\_stopwords." Kedua daftar kata *stopword* ini kemudian digabungkan menjadi satu dan hasilnya disimpan dalam variabel yang disebut "stopwords". Selanjutnya, dalam fungsi "menghapusKata," dilakukan proses tokenisasi untuk setiap data dengan menggunakan fungsi "word\_tokenize" yang tersedia dalam *library NLTK*.

```
#menghapus kata
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory

factory = StopWordRemoverFactory()
more_stopword =
['kab','kec','d','ds','bok','p','pulah','ehhh','sok','bar','to','esd
orem','wan','elso','hahahaha','wasweswos','amp','ajig','bajingan','b
ajinguk','bacot','mencret','hmmm','njirr',
'hahahahaha','goblokk','taik','bgd','dancuk','bangsaaadd','bacodnya
','fvck','taek','asu','matamu','brengek','anjay','peleeerrr','bgcd'
,'jir','fuck','eek','cuk','genjik','prix','uhuyyy','haq','anjing²',
```

```
'ckckckkkk','hmmmm','ttt','fkg','m³','kss','weh','cup','uwow','wkwke
kekekekesemka','oiii','bor','imm','hehe','dex','haha'
'wkwkwkl','hoho','hoooooooo','owh','wkwkw','wkwkwk','Wkwkwkwkwk',
'wkwkwkwkwk','upsssttt','ckck','wii','jancoooooook','cot','ohh']

stopword = factory.create_stop_word_remover()
stopwords = factory.get_stop_words()+ more_stopword

def menghapusKata(text):
    word_tokens = word_tokenize(text)
    filtered_sentence = [w for w in word_tokens if not w in stopwords]
    return ' '.join(filtered_sentence)
df['hasilremove'] = df['normalisasi'].apply(menghapusKata)
df.head()
```

Gambar 4.8 Kode program proses *stopword removal*

Tabel 4.11 menunjukkan hasil dari tahap *stopword removal*. Hasil ini berbeda dari tabel sebelumnya yang masih belum diolah dan belum dinormalisasi, karena Tabel 4.10 hanya memuat informasi yang relevan dan penting. Proses ini mengakibatkan kata-kata yang tidak berarti dihapus karena diidentifikasi sebagai *stopwords*.

Tabel 4.11 Contoh komentar setelah dilakukan *stopword removal*

| Komentar Sebelum   | Komentar setelah <i>stopword removal</i>         |
|--|--|
| "bagusss", "pak", "ganjar",<br>"pranowo"                 | "bagusss", "pak", "ganjar", "pranowo"            |
| "semoga", "menjadi", "pemimpin",<br>"yang", "menyatukan" | "semoga", "menjadi", "pemimpin",<br>"menyatukan" |
| "prabowo", "selalu", "di", "hati"                        | "prabowo", "selalu", "hati"                      |

### 4.3 Pelabelan Kelas Sentimen

Setelah menyelesaikan tahap *preprocessing*, langkah selanjutnya adalah melakukan pelabelan data. Pada tahap ini, penulis juga mencari hasil representasi yang diinginkan dari *corpus*. Proses pelabelan sentimen dilakukan dengan metode penghitungan otomatis menggunakan kamus yang telah ada. Proses pelabelan tersebut dilakukan dengan pembobotan *lexicon based* yang mempertimbangkan skor sentimen secara otomatis (Ramadhanty, 2021). Proses perhitungan skor sentimen dilakukan dengan pencarian jumlah kata yang mengandung

sentimen positif dan jumlah kata yang mengandung sentimen negatif. Skor sentimen kemudian dihitung dengan mengurangkan jumlah polaritas kata-kata dengan sentimen negatif dari jumlah polaritas kata-kata dengan sentimen positif.

Proses pelabelan terbagi menjadi dua kategori sentimen, yakni positif dan negatif, menggunakan metode *scoring*. Tidak digunakannya kategori netral karena kelas netral dapat mengakibatkan penurunan akurasi (Fitriyyah et al., 2019). Penentuan kategori sentimen dilakukan dengan memanfaatkan sekumpulan kata dalam bahasa Indonesia yang terdiri dari kata-kata positif dan kata-kata negatif. Kumpulan kata-kata positif disimpan dalam suatu format *csv* dengan nama "positive.csv", sedangkan kumpulan kata-kata negatif disimpan dengan nama "negative.csv". Dalam penelitian ini, kamus yang digunakan diambil dari tautan berikut: <https://github.com/fajri91/InSet/>. Sebagai hasilnya, penulis mendapatkan 3.609 kata-kata dengan nilai positif dan 6.609 kata-kata dengan nilai negatif (Koto & Rahmaningtyas, 2017). Contoh dari kata-kata positif dan kata-kata negatif terdapat pada Tabel 4.12. Gambar 4.8 menunjukkan langkah yang dilakukan untuk pelabelan kelas sentimen positif dan negatif menggunakan *Python*.

Tabel 4.12 Contoh Kata Positif dan Negatif

| Contoh Kata Positif | Contoh Kata Negatif |
|---------------------|---------------------|
| kekaguman           | yaudah              |
| kasih sayang        | ngga                |
| kehangatan          | bodoh               |
| makmur              | sepuh               |
| ditunggu            | gangguan jiwa       |

```
import csv

# Baca file CSV leksikon positif
positive_lexicon = set()
with open('positive.csv', 'r', newline='', encoding='utf-8') as
positive_file:
    csv_reader = csv.reader(positive_file)
    for row in csv_reader:
        positive_lexicon.add(row[0])

# Baca file CSV leksikon negatif
negative_lexicon = set()
with open('negative.csv', 'r', newline='', encoding='utf-8') as
negative_file:
    csv_reader = csv.reader(negative_file)
    for row in csv_reader:
        negative_lexicon.add(row[0])
```

```

# Baca dan label data CSV
csv_file_path = 'aniesjuli.csv'

with open(csv_file_path, 'r', newline='', encoding='utf-8') as csv_file:
    csv_reader = csv.reader(csv_file)
    data = list(csv_reader)

for row in data:
    text = row[0] # Misalnya, teks berada di kolom pertama
    sentiment_score = 0

    words = text.split()

    for word in words:
        if word in positive_lexicon:
            sentiment_score += 1
        elif word in negative_lexicon:
            sentiment_score -= 1

    if sentiment_score >= 0:
        row.append('positif')
    elif sentiment_score <= 0:
        row.append('negatif')

# Tulis kembali ke file CSV
output_csv_file_path = 'aniesjulilabeled.csv'

with open(output_csv_file_path, 'w', newline='', encoding='utf-8') as
output_csv_file:
    csv_writer = csv.writer(output_csv_file)
    csv_writer.writerows(data)

```

Gambar 4.9 Kode program proses pelabelan data

Kode pada Gambar 4.10 adalah proses pelabelan data. Pertama, kode membaca dua *file* CSV yang berisi leksikon kata-kata positif dan negatif, kemudian menyimpannya dalam dua set yang disebut `positive\_lexicon` dan `negative\_lexicon`. Selanjutnya, kode membaca data dari file CSV `aniesjuli.csv`, memproses setiap baris teks dalam file tersebut dengan menghitung skor sentimen berdasarkan kata-kata yang ada dalam leksikon positif dan negatif. Hasil sentimen ditambahkan ke setiap baris data sebagai label 'positif' atau 'negatif'. Akhirnya, data yang telah diberi label sentimen disimpan kembali ke file CSV baru dengan nama `aniesjulilabeled.csv`. Dengan demikian, kode ini digunakan untuk mengkategorikan sentimen teks berdasarkan kata-kata yang terdapat dalam leksikon positif dan negatif, dan menyimpan hasilnya dalam file CSV baru.

Dalam penggunaan kumpulan kata-kata bahasa Indonesia ini, *Python* secara otomatis akan melabeli dokumen dengan menghitung skor yang didapatkan dari selisih jumlah kata positif dan jumlah kata negatif dalam sebuah komentar. Kata-kata akan diberi nilai positif (+1) jika kata tersebut sesuai dengan kamus positif, atau nilai negatif (-1) jika kata tersebut sesuai dengan kamus negatif. Setelah kata-kata memiliki nilai, perhitungan akan dilakukan dengan mengurangkan jumlah nilai kata yang sesuai pada kamus positif dengan jumlah nilai kata yang sesuai dengan kamus negatif. Jika skor suatu kalimat adalah 0 atau lebih, maka kalimat tersebut akan dinyatakan sebagai sentimen positif, sedangkan jika skor kurang dari 0, kalimat tersebut akan dinyatakan sebagai sentimen negatif. Tabel 4.13 adalah hasil perbandingan jumlah data hasil pelabelan kelas sentimen yang diperoleh.

Tabel 4.13 Hasil Pelabelan Sentimen dengan Lexicon pada Masing-Masing Topik Calon Presiden

| Nama Topik                | Ganjar Pranowo     |                   | Anies Baswedan    |                   | Prabowo Subianto   |                   |
|---------------------------|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|
|                           | Sentimen Positif   | Sentimen Negatif  | Sentimen Positif  | Sentimen Negatif  | Sentimen Positif   | Sentimen Negatif  |
| Bertemu Masyarakat        | 5.802<br>(84,44%)  | 1.069<br>(15,55%) | 3.573<br>(86,30%) | 567<br>(13,69%)   | 15.908<br>(85,25%) | 2.752<br>(14,74%) |
| Bertemu Publik Figur      | 12.122<br>(81,86%) | 2.686<br>(18,13%) | 8.224<br>(86,26%) | 1.309<br>(13,73%) | 29.374<br>(87,27%) | 4.284<br>(12,72%) |
| Menghadiri Acara          | 5.363<br>(76,63%)  | 1.635<br>(23,36%) | 4.952<br>(86,75%) | 756<br>(13,24%)   | 30.497<br>(87,81%) | 4.230<br>(12,18%) |
| Lain - Lain               | 3.125<br>(86,56%)  | 485<br>(13,44%)   | 1.047<br>(87,68%) | 147<br>(12,31%)   | 11.092<br>(88,09%) | 1.499<br>(11,90%) |
| Kebijakan Gubernur        | 4.161<br>(81,78%)  | 927<br>(18,21%)   | -                 | -                 | -                  | -                 |
| Pembangunan Infrastruktur | 1.628<br>(86,13%)  | 262<br>(13,86%)   | -                 | -                 | -                  | -                 |
| Melaksanakan Ibadah haji  | -                  | -                 | 6.204<br>(91,70%) | 561<br>(8,29%)    | -                  | -                 |
| Masa-masa di TNI          | -                  | -                 | -                 | -                 | 5.167<br>(86,13%)  | 832<br>(13,86%)   |

Untuk topik "bertemu masyarakat" calon presiden yang memiliki sentimen positif tertinggi adalah Anies Baswedan sebesar 86,30%, topik "bertemu publik figur" yang memiliki sentimen positif tertinggi adalah Prabowo subianto sebesar 86,26%, topik "menghadiri acara" calon presiden yang memiliki sentimen positif tertinggi adalah Prabowo subianto sebesar 87,81%, dan untuk topik "lain-lain" calon presiden yang memiliki sentimen positif tertinggi adalah Prabowo subianto sebesar 88,09%,

#### 4.4 *Split Data Train dan Data Test*

Setelah label diberikan pada data komentar, langkah berikutnya adalah melakukan pembagian data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih algoritma hingga membentuk suatu model, sementara data pengujian digunakan untuk menguji performa model yang telah dibentuk oleh data pelatihan ketika dihadapkan pada data baru. Berdasarkan Fitriyyah et al. (2019), data komentar yang telah diberi label sentimen akan dibagi dengan perbandingan 80% untuk data pelatihan dan 20% data pengujian untuk mendapatkan akurasi yang terbaik.

Data dibagi menjadi empat variabel yang berbeda dengan bantuan fungsi "train\_test\_split", yaitu `x_train`, `x_test`, `y_train`, dan `y_test`. `x_train` berisi teks yang digunakan untuk melatih model, sementara `x_test` berisi teks yang digunakan untuk menguji model setelah pelatihan. Di sisi lain, `y_train` dan `y_test` adalah target atau label yang sesuai dengan data tersebut. `y_train` berisi label-label yang menentukan sentimen atau kategori yang akan diprediksi oleh model selama pelatihan, sementara `y_test` berisi label-label yang digunakan untuk mengukur kinerja model pada data uji. Proses pembagian *dataset* ini memisahkan data yang akan digunakan untuk pelatihan dan pengujian. Variabel "df", yang sebelumnya berisi seluruh data, dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian, dengan perbandingan 80:20.

```
import numpy as np
import joblib

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report
```

```

x_train, x_test, y_train, y_test = train_test_split(
    df['text'].values,
    df['positif'].values,
    test_size=0.2
)

print(f"jumlah x train adalah {len(x_train)}")
print(f"jumlah y train adalah {len(y_train)}")
print(f"jumlah x test adalah {len(x_test)}")
print(f"jumlah y test adalah {len(y_test)}")

vect = CountVectorizer()
vect.fit(x_train)
data_train = vect.transform(x_train)
data_test = vect.transform(x_test)

```

Gambar 4.11 Kode program untuk pembagian data

Kode pada Gambar 4.12 adalah implementasi pembuatan model klasifikasi menggunakan *library Scikit-Learn*. Pertama, kode mengimpor *library* yang diperlukan seperti *NumPy*, *joblib*, dan berbagai modul *Scikit-Learn*. Selanjutnya, data dari kolom "text" dan 'positif' dari *DataFrame* `df` dibagi menjadi data pelatihan (x\_train dan y\_train) dan data pengujian (x\_test dan y\_test) dengan rasio 80:20 menggunakan `train\_test\_split`. Kode kemudian mencetak jumlah data dalam set pelatihan dan pengujian. Ini adalah langkah awal dalam mempersiapkan data untuk pelatihan dan pengujian model klasifikasi. Pada tahap selanjutnya, model klasifikasi menggunakan metode *Support Vector Machine* dilatih dengan data pelatihan. Hasil dari pembagian *train data* dan *test data* untuk masing-masing calon presiden dapat dilihat pada Tabel 4.14 hingga Tabel 4.16.

Tabel 4.14 Pembagian jumlah *train data* dan *test data* Ganjar Pranowo

| Nama Topik           | Sentimen   | Jumlah |
|----------------------|------------|--------|
| Kebijakan Gubernur   | Data Latih | 4.070  |
|                      | Data Uji   | 1.018  |
| Bertemu Masyarakat   | Data Latih | 5.496  |
|                      | Data Uji   | 1.375  |
| Bertemu Publik Figur | Data Latih | 11.846 |
|                      | Data Uji   | 2.962  |

|                           |            |       |
|---------------------------|------------|-------|
| Pembangunan Infrastruktur | Data Latih | 1.512 |
|                           | Data Uji   | 378   |
| Menghadiri Acara          | Data Latih | 5.598 |
|                           | Data Uji   | 1.400 |
| Lain-lain                 | Data Latih | 2.888 |
|                           | Data Uji   | 722   |

Tabel 4.15 Pembagian jumlah *train data* dan *test data* Anies Baswedan

| Nama Topik               | Sentimen   | Jumlah |
|--------------------------|------------|--------|
| Bertemu Masyarakat       | Data Latih | 3.312  |
|                          | Data Uji   | 828    |
| Bertemu Publik Figur     | Data Latih | 7.627  |
|                          | Data Uji   | 1.907  |
| Melaksanakan Ibadah haji | Data Latih | 5.412  |
|                          | Data Uji   | 1.353  |
| Menghadiri Acara         | Data Latih | 4.566  |
|                          | Data Uji   | 1.142  |
| Lain-lain                | Data Latih | 955    |
|                          | Data Uji   | 239    |

Tabel 4.16 Pembagian jumlah *train data* dan *test data* Prabowo Subianto

| Nama Topik           | Sentimen   | Jumlah Data |
|----------------------|------------|-------------|
| Bertemu Masyarakat   | Data Latih | 14.928      |
|                      | Data Uji   | 3.732       |
| Bertemu Publik Figur | Data Latih | 26.926      |
|                      | Data Uji   | 6.732       |
| Masa-masa di TNI     | Data Latih | 4.799       |
|                      | Data Uji   | 1.200       |
| Menghadiri Acara     | Data Latih | 27.781      |
|                      | Data Uji   | 6.946       |
| Lain-lain            | Data Latih | 10.072      |
|                      | Data Uji   | 2.519       |

#### 4.5 Feature Extraction

*Feature extraction* adalah proses di mana informasi relevan dan bermanfaat diidentifikasi dan diekstrak dari data mentah. Tujuan adalah untuk menyederhanakan representasi data, mempertahankan informasi yang paling penting, dan meningkatkan kualitas data untuk penggunaan lebih lanjut, seperti dalam pembelajaran mesin atau analisis data (Salau & Jain, 2019). Kode pada Gambar 4.13 adalah implementasi *feature extraction* dimana objek

`CountVectorizer` digunakan untuk mengubah teks menjadi vektor fitur yang dapat digunakan oleh model klasifikasi. Data latihan dan data uji diubah menjadi vektor fitur menggunakan metode `transform`. Vektor fitur ini akan berisi informasi tentang frekuensi kata-kata dalam teks.

#### 4.6 Klasifikasi Menggunakan SVM

Proses klasifikasi dalam penelitian ini melibatkan penggunaan algoritma *Support Vector Machine* (SVM). Prinsip kerja SVM yaitu dengan mempelajari pola dari *data training* yang telah ditentukan sebelumnya. Algoritma SVM akan menganalisis pola yang ada dalam *data training* tersebut, kemudian melakukan pelatihan untuk mengklasifikasikan karakteristik dari setiap kelas atau kategori dalam data. Tahapan ini merupakan bagian dari ilmu *machine learning* (ML) yang dikenal dengan sebutan *supervised machine learning*. Hasil dari pelatihan algoritma SVM akan menghasilkan suatu model, yang selanjutnya diuji menggunakan *data testing*. Langkah ini bertujuan untuk mengukur tingkat akurasi yang diberikan oleh model tersebut dan juga untuk melihat sejauh mana kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, jenis *kernel* yang digunakan meliputi *kernel Linear*, *Polinomial*, *Radial Basis Function* (RBF), dan *Sigmoid*. Lalu hasil akurasi dari *kernel* terbaik akan digunakan di tahap *confusion matrix*. Gambar 4.10 adalah implementasi pemodelan model *Support Vector Machine* (SVM) dengan *kernel linear* pada Python. Pertama, penulis membuat model SVM dengan *kernel linear* dan mengatur nilai *seed* acak (`random\_state`) agar hasilnya dapat direproduksi. Selanjutnya, model dilatih menggunakan data pelatihan (`data\_train`) dan label pelatihan (`y\_train`) untuk mempelajari pola dalam data. Setelah pelatihan, model digunakan untuk melakukan prediksi pada data uji (`data\_test`) dan hasil prediksi disimpan dalam variabel `prediction`. Akhirnya, penulis menggunakan metode `accuracy\_score` untuk mengukur akurasi model, yaitu sejauh mana model berhasil dalam memprediksi dengan benar, dengan membandingkan prediksi dengan label sebenarnya pada data uji (`y\_test`). Hasil perbandingan akurasi dari setiap *kernel* yang telah diujikan tersaji dalam Tabel 4.17-Tabel 4.19.

```
model = SVC(kernel='linear', random_state=10)
model.fit(data_train, y_train)
prediction = model.predict(data_test)
prediction
```

```
accuracy_score(y_test, prediction)
```

Gambar 4.10 Kode program SVM

Tabel 4.17 Perbandingan Hasil Klasifikasi SVM Ganjar Pranowo

| <b>Nama Topik</b>         | <b>RBF(%)</b> | <b><i>Sigmoid</i>(%)</b> | <b><i>Linear</i>(%)</b> | <b><i>Polynomial</i>(%)</b> |
|---------------------------|---------------|--------------------------|-------------------------|-----------------------------|
| Kebijakan Gubernur        | 82,4          | 82,1                     | 86,2                    | 80,9                        |
| Bertemu Masyarakat        | 86,0          | 83,9                     | 86,3                    | 83,4                        |
| Bertemu Publik Figur      | 86,0          | 81,7                     | 87,3                    | 81,6                        |
| Pembangunan Infrastruktur | 83,5          | 82,0                     | 85,0                    | 83,5                        |
| Menghadiri Acara          | 84,1          | 80,5                     | 84,7                    | 78,5                        |
| Lain-Lain                 | 86,1          | 85,0                     | 88,7                    | 85,0                        |

Tabel 4.18 Perbandingan Hasil Klasifikasi SVM Anies Baswedan

| <b>Nama Topik</b>        | <b>RBF(%)</b> | <b><i>Sigmoid</i>(%)</b> | <b><i>Linear</i>(%)</b> | <b><i>Polynomial</i>(%)</b> |
|--------------------------|---------------|--------------------------|-------------------------|-----------------------------|
| Bertemu Masyarakat       | 87,5          | 84,5                     | 88,0                    | 86,3                        |
| Bertemu Publik Figur     | 87,0          | 85,0                     | 88,2                    | 86,2                        |
| Melaksanakan Ibadah haji | 91,9          | 89,8                     | 91,9                    | 91,7                        |
| Menghadiri Acara         | 85,8          | 85,6                     | 87,4                    | 85,9                        |
| Lain-Lain                | 86,6          | 86,1                     | 84,5                    | 85,7                        |

Tabel 4.19 Perbandingan Hasil Klasifikasi SVM Prabowo Subianto

| <b>Nama Topik</b>    | <b>RBF(%)</b> | <b>Sigmoid(%)</b> | <b>Linear(%)</b> | <b>Polynomial(%)</b> |
|----------------------|---------------|-------------------|------------------|----------------------|
| Bertemu Masyarakat   | 88,3          | 83,5              | 89,2             | 84,7                 |
| Bertemu Publik Figur | 90,5          | 86,2              | 91,4             | 87,8                 |
| Masa-masa di TNI     | 88,5          | 87,3              | 88,9             | 87,9                 |
| Menghadiri Acara     | 90,5          | 86,3              | 91,4             | 87,9                 |
| Lain-Lain            | 88,8          | 86,5              | 90,5             | 87,6                 |

Evaluasi model adalah langkah penting untuk menilai sejauh mana model yang telah dibuat mencapai tingkat performa yang diinginkan. Evaluasi model dapat dilakukan dengan mempertimbangkan sejumlah metrik seperti akurasi model, *f1-score*, dan penggunaan *confusion matrix*. Penggunaan *confusion matrix* sangat berguna dalam menghitung akurasi dengan lebih detail dan tepat. Dalam Tabel 4.14 hingga Tabel 4.16, dapat diamati bahwa dari perbandingan empat *kernel* yang diujikan, *kernel linear* rata-rata menunjukkan tingkat akurasi tertinggi. Oleh karena itu, *kernel linear* akan menjadi pilihan utama untuk proses evaluasi. Tabel 4.20 hingga Tabel 4.22 merupakan hasil dari *confusion matrix*.

Tabel 4.20 Hasil dari *confusion matrix* Ganjar Pranowo

| <b>Nama Topik</b>         | <b>Precision (%)</b> |                | <b>Recall (%)</b> |                | <b>f1-Score(%)</b> |                |
|---------------------------|----------------------|----------------|-------------------|----------------|--------------------|----------------|
|                           | <b>Negatif</b>       | <b>Positif</b> | <b>Negatif</b>    | <b>Positif</b> | <b>Negatif</b>     | <b>Positif</b> |
| Kebijakan Gubernur        | 68                   | 89             | 52                | 94             | 59                 | 92             |
| Bertemu Masyarakat        | 60                   | 90             | 46                | 94             | 52                 | 92             |
| Bertemu Publik Figur      | 67                   | 91             | 60                | 93             | 64                 | 92             |
| Pembangunan Infrastruktur | 59                   | 87             | 27                | 96             | 37                 | 91             |
| Menghadiri Acara          | 69                   | 89             | 61                | 92             | 65                 | 90             |
| Lain-Lain                 | 67                   | 91             | 46                | 96             | 55                 | 94             |

Tabel 4.21 Hasil dari *confusion matrix* Anies Baswedan

| Nama Topik               | <i>Precision (%)</i> |         | <i>Recall (%)</i> |         | <i>f1-Score(%)</i> |         |
|--------------------------|----------------------|---------|-------------------|---------|--------------------|---------|
|                          | Negatif              | Positif | Negatif           | Positif | Negatif            | Positif |
| Bertemu Masyarakat       | 54                   | 93      | 50                | 94      | 52                 | 93      |
| Bertemu Publik Figur     | 61                   | 92      | 51                | 94      | 55                 | 93      |
| Melaksanakan Ibadah haji | 51                   | 95      | 39                | 97      | 44                 | 96      |
| Menghadiri Acara         | 61                   | 90      | 39                | 96      | 48                 | 93      |
| Lain-Lain                | 62                   | 90      | 44                | 95      | 51                 | 92      |

Tabel 4.22 Hasil dari *confusion matrix* Prabowo Subianto

| Nama Topik           | <i>Precision (%)</i> |         | <i>Recall (%)</i> |         | <i>f1-Score(%)</i> |         |
|----------------------|----------------------|---------|-------------------|---------|--------------------|---------|
|                      | Negatif              | Positif | Negatif           | Positif | Negatif            | Positif |
| Bertemu Masyarakat   | 62                   | 90      | 44                | 95      | 51                 | 92      |
| Bertemu Publik Figur | 69                   | 94      | 57                | 96      | 62                 | 95      |
| Masa-masa di TNI     | 56                   | 93      | 53                | 94      | 55                 | 94      |
| Menghadiri Acara     | 69                   | 94      | 57                | 96      | 63                 | 95      |
| Lain-Lain            | 63                   | 94      | 55                | 96      | 59                 | 95      |

Dalam tahap evaluasi, penulis memanfaatkan metode *confusion matrix* untuk menghitung nilai *precision*, *recall*, dan *f1-score*. *Precision* mengukur sejauh mana prediksi positif atau negatif dari model adalah benar, sementara *recall* mengukur kemampuan model dalam mengidentifikasi semua nilai positif atau negatif yang sebenarnya. *F1-score* adalah nilai tengah antara *precision* dan *recall*.

Hasil *precision confusion matrix* sentimen positif tertinggi dari Prabowo Subianto ada pada topik "lain-lain" dan "menghadiri acara" dengan akurasi sebesar 94% dan *precision confusion matrix* sentimen negatif tertinggi di topik "menghadiri acara" dan "bertemu publik figur" dengan akurasi mencapai 69%. Anies Baswedan memiliki hasil *precision confusion matrix* sentimen positif tertinggi pada topik "melaksanakan ibadah haji" dengan akurasi sebesar 95% dan *precision confusion matrix* sentimen negatif tertinggi pada topik "lain-lain" dengan akurasi sebesar 62%. Sementara Ganjar Pranowo mendapati hasil *precision confusion matrix*

sentimen positif tertinggi pada topik "lain-lain" dan "bertemu publik figur" dengan akurasi sebesar 91% dan *precision confusion matrix* sentimen negatif tertinggi di topik "menghadiri acara" dengan akurasi sebesar 69%.

Pada penelitian ini, peneliti menyadari masih ada kekurangan seperti nilai *precision* dan *recall* yang jauh lebih rendah pada *confusion matrix* sentimen negatif jika dibandingkan dengan nilai *precision* dan *recall* dari *confusion matrix* sentimen positif dikarenakan ketidakseimbangan dalam jumlah sampel antara kelas positif dan negatif dalam data. Sebagai akibatnya, model cenderung lebih baik dalam mengenali sentimen positif karena memiliki lebih banyak sampel untuk belajar.

Tabel 4.23 dan Tabel 4.24 menunjukkan sentimen positif tertinggi dan sentimen negatif tertinggi. Di antara ketiga calon presiden yang memiliki sentimen positif tertinggi adalah Anies Baswedan pada topik "melaksanakan ibadah haji" dengan nilai sentimen positif sebesar 91,70% dan yang memiliki sentimen negatif tertinggi adalah Ganjar Pranowo pada topik "menghadiri acara" dengan nilai sentimen negatif sebesar 23,36%.

Tabel 4.23 Sentimen Positif Tertinggi

| Calon Presiden   | Topik Bahasan             | Jumlah (%) |
|------------------|---------------------------|------------|
| Anies Baswedan   | Melaksanakan Ibadah Haji  | 91,70%     |
| Prabowo Subianto | Lain-Lain                 | 88,09%     |
| Prabowo Subianto | Menghadiri Acara          | 87,81%     |
| Anies Baswedan   | Lain-Lain                 | 87,68%     |
| Prabowo Subianto | Bertemu Publik Figur      | 87,27%     |
| Anies Baswedan   | Menghadiri Acara          | 86,75%     |
| Ganjar Pranowo   | Lain-Lain                 | 86,56%     |
| Anies Baswedan   | Bertemu Masyarakat        | 86,30%     |
| Anies Baswedan   | Bertemu Publik Figur      | 86,26%     |
| Prabowo Subianto | Masa-Masa di TNI          | 86,13%     |
| Ganjar Pranowo   | Pembangunan Infrastruktur | 86,13%     |
| Prabowo Subianto | Bertemu Masyarakat        | 85,25%     |
| Ganjar Pranowo   | Bertemu Masyarakat        | 84,44%     |
| Ganjar Pranowo   | Bertemu Publik Figur      | 81,86%     |
| Ganjar Pranowo   | Kebijakan Gubernur        | 81,78%     |
| Ganjar Pranowo   | Menghadiri Acara          | 76,63%     |

Tabel 4.24 Sentimen Negatif Tertinggi

| <b>Calon Presiden</b> | <b>Topik Bahasan</b>      | <b>Jumlah (%)</b> |
|-----------------------|---------------------------|-------------------|
| Ganjar Pranowo        | Menghadiri Acara          | 23,36%            |
| Ganjar Pranowo        | Kebijakan Gubernur        | 18,21%            |
| Ganjar Pranowo        | Bertemu Publik Figur      | 18,13%            |
| Ganjar Pranowo        | Bertemu Masyarakat        | 15,55%            |
| Prabowo Subianto      | Bertemu Masyarakat        | 14,74%            |
| Ganjar Pranowo        | Pembangunan Infrastruktur | 13,86%            |
| Prabowo Subianto      | Masa-Masa di TNI          | 13,86%            |
| Anies Baswedan        | Bertemu Publik Figur      | 13,73%            |
| Anies Baswedan        | Bertemu Masyarakat        | 13,69%            |
| Ganjar Pranowo        | Lain-Lain                 | 13,44%            |
| Anies Baswedan        | Menghadiri Acara          | 13,24%            |
| Prabowo Subianto      | Bertemu Publik Figur      | 12,72%            |
| Anies Baswedan        | Lain-Lain                 | 12,31%            |
| Prabowo Subianto      | Menghadiri Acara          | 12,18%            |
| Prabowo Subianto      | Lain-Lain                 | 11,90%            |
| Anies Baswedan        | Melaksanakan Ibadah Haji  | 8,29%             |

## BAB V

### KESIMPULAN DAN SARAN

Bab ini akan menggambarkan rangkuman dari hasil penelitian ini, berdasarkan evaluasi atas proses, algoritma, dan penerapan sistem yang telah dibuat. Saran akan mengulas catatan dan perbaikan yang diperlukan untuk penelitian ini dalam waktu mendatang.

#### 5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan mengenai Analisis Sentimen Calon Presiden Indonesia 2024 di Facebook Menggunakan Metode *Support Vector Machine* (SVM), didapatkan beberapa kesimpulan sebagai berikut:

- a. Hasil sentimen positif tertinggi dari Ganjar Pranowo ada pada topik "lain-lain" ("ucapan hari besar", "cerita tentang masa lalu", "tip LDR", "menonton pertandingan sepak bola", "menerima penghargaan", dan "tip untuk perantau") dan sentimen negatif tertinggi di topik "menghadiri acara". Anies Baswedan memiliki hasil sentimen tertinggi pada topik "melaksanakan ibadah haji" dan sentimen negatif tertinggi pada topik "bertemu dengan publik figur". Sementara Prabowo Subianto mendapati hasil sentimen tertinggi pada topik "lain-lain" ("cerita tentang masa lalu", "cerita tentang pahlawan", dan "melaksanakan shalat Eid") dan sentimen negatif tertinggi di topik "bertemu masyarakat".
- b. Di antara ketiga calon presiden yang memiliki sentimen positif tertinggi tertinggi adalah Anies Baswedan pada topik "melaksanakan ibadah haji".
- c. Dalam evaluasi performa empat algoritma *Support Vector Machine*, ditemukan bahwa nilai *precision* terbaik untuk terdapat pada *kernel linear*, dengan nilai sentimen positif tertinggi dari Ganjar Pranowo ada pada topik "bertemu dengan publik figur" dan "lain-lain" dengan nilai akurasi sebesar 91% dan sentimen negatif tertinggi pada topik "menghadiri acara" dengan nilai akurasi sebesar 69%. Anies Baswedan memiliki hasil sentimen positif tertinggi pada topik "melaksanakan ibadah haji" dengan nilai akurasi sebesar 95% dan sentimen negatif tertinggi pada topik "lain-lain" dengan nilai akurasi sebesar 62%. Sementara Prabowo Subianto mendapati hasil sentimen positif tertinggi pada topik "bertemu publik figur", "menghadiri acara", dan "lain-lain" dengan nilai akurasi sebesar 94% dan sentimen

negatif tertinggi di topik "menghadiri acara" dan "bertemu publik figur" dengan nilai akurasi sebesar 69%.

## 5.2 Saran

Harapannya adalah bahwa penelitian ini dapat menjadi rujukan atau sumber referensi bagi penelitian yang sejenis di masa mendatang. Beberapa saran untuk penelitian lanjutan adalah sebagai berikut:

- a. Mengambil informasi dari *platform* sosial media lainnya sebagai sumber data.
- b. Untuk penelitian berikutnya, disarankan untuk mempertimbangkan penggunaan algoritma klasifikasi lainnya sebagai perbandingan untuk menentukan algoritma yang memiliki kinerja paling baik.

## DAFTAR PUSTAKA

- Arsi, P., & 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. <https://doi.org/10.25126/jtiik.0813944>
- Batta, M. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 18(8), 381–386. <https://doi.org/10.21275/ART20203995>
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1), 32–41. <https://doi.org/10.31284/j.integer.2017.v2i1.95>
- Diki Hendriyanto, M., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm. *Journal of Information Technology and Computer Science (INTECOMS)*, 5(1), 1–7.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fitriana, D. E., Yanto, A. F., & Budiman, J. (2020). Analisis user experience ( UX ) fitur marketplace facebook. *Jurnal Ekonomi Dan Teknik Informatika*, 8(2), 47–66. <http://e-journal.polsa.ac.id/index.php/jneti/article/view/151>
- Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(3), 279. <https://doi.org/10.26418/jp.v5i3.34368>
- Kamal, W. W. (2021). Analisis Sentimen Ulasan Produk Skincare Menggunakan Metode Support Vector Machine (Studi Kasus: Forum Female Daily). 65.
- Koto, F., & Rahmaningtyas, G. Y. (2017). InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs. *2017 International Conference on Asian Language Processing (IALP)*, 391–394.
- Liando, D. M. (2016). Pemilu dan Partisipasi Politik Masyarakat (Studi Pada Pemilihan Anggota Legislatif Dan Pemilihan Presiden Dan Calon Wakil Presiden Di Kabupaten Minahasa Tahun 2014). *Jurnal LPPM Bidang EkoSosBudKum*, 3(2), 14–28.
- Lukmana, D. T., Subanti, S., & Susanti, Y. (2019). Analisis Sentimen Terhadap Calon Presiden

- 2019 Dengan Support Vector Machine Di Twitter. *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT, 2002*, 154–160.
- Patria, L., & Yulianto, K. (2011). Pemanfaatan Facebook untuk Menunjang Kegiatan Belajar Mengajar Online Secara Mandiri. *Prosiding Seminar Nasional FMIPA-UT*. <http://repository.ut.ac.id/id/eprint/2305>
- Pradana, A. W., & Hayaty, M. (2019). The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(3), 375–380. <https://doi.org/10.22219/kinetik.v4i4.912>
- Rachmat C, A., & Lukito, Y. (2016). Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes. *Jurnal Informatika Dan Sistem Informasi Universitas Ciputra*, 02(02), 26–34. <https://github.com/sastrawi/sastrawi>
- Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/jifo.v11i1.a5452>
- Ramadhanty, D. R. (2021). *Implementasi Algoritma Support Vector Machine Pada Analisis Sentimen Data Twitter*. 1–83. <https://dspace.uui.ac.id/handle/123456789/36015>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sabily, A. F., Adikara, P. P., & Fauzi, M. A. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(5), 4204–4209. <http://j-ptiik.ub.ac.id>
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia). *Prosiding SINTAK*, 480–486.
- Salau, A. O., & Jain, S. (2019). Feature Extraction: A Survey of the Types, Techniques, Applications. *2019 International Conference on Signal Processing and Communication, ICSC 2019*, 158–164. <https://doi.org/10.1109/ICSC45622.2019.8938371>
- Salman, F. M., & Widiyanesti, S. (2022). *Implementasi Sentiment Analysis Dalam Penyebaran Informasi Vaksinasi COVID-19 Menggunakan Metode Naive Bayes Di Facebook*. *Implementation Of Sentiment Analysis In Distribution Of COVID-19 Vaccination Information Using Naive Bayes Method On Facebook*. 9(4), 1833–1843.

- Sari, I. P. (2018). Keberpihakan Media Dalam Pemilihan Presiden 2014. *Jurnal Penelitian Komunikasi*, 21(1), 73–86. <https://doi.org/10.20422/jpk.v21i1.492>
- Setiawan, K. E. P., & Zyuliantina, W. (2020). Analisis Kesalahan Berbahasa Indonesia Pada Status Dan Komentar Di Facebook. *Tabasa: Jurnal Bahasa, Sastra Indonesia, Dan Pengajarannya*, 1(1), 96–109. <https://doi.org/10.22515/tabasa.v1i1.2605>
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 4(3), 650. <https://doi.org/10.30865/mib.v4i3.2181>
- Utami, N. A., Maharani, W., & Atastina, I. (2021). Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method. *Procedia Computer Science*, 179(2020), 177–184. <https://doi.org/10.1016/j.procs.2020.12.023>
- Vinodhini, G., & Chandrasekaran, R. (2012). Sentiment Analysis and Opinion Mining : A Survey International Journal of Advanced Research in Sentiment Analysis and Opinion Mining : A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(6), 283–292.

## LAMPIRAN

### A. Data Normalisasi

| nomor | before    | after       | nomor | before     | after       |
|-------|-----------|-------------|-------|------------|-------------|
| 1     | yng       | yang        | 117   | kluar      | keluar      |
| 2     | knp       | kenapa      | 118   | mayan      | lumayan     |
| 3     | knpa      | kenapa      | 119   | dipake     | dipakai     |
| 4     | sampe     | sampai      | 120   | teredia    | tersedia    |
| 5     | smpai     | sampai      | 121   | rekomen    | rekomendasi |
| 6     | smpe      | sampai      | 122   | buanyak    | banyak      |
| 7     | sy        | saya        | 123   | kyknya     | sepertinya  |
| 8     | sya       | saya        | 124   | trs        | terus       |
| 9     | lu        | kamu        | 125   | ap         | apa         |
| 10    | kmu       | kamu        | 126   | yaaaaa     | ya          |
| 11    | jngn      | jangan      | 127   | uda        | udah        |
| 12    | jgn       | jangan      | 128   | semuaa     | semua       |
| 13    | tdk       | tidak       | 129   | ktnya      | katanya     |
| 14    | tdak      | tidak       | 130   | sampek     | sampai      |
| 15    | bener     | benar       | 131   | bukak      | buka        |
| 16    | bner      | benar       | 132   | anying     | anjing      |
| 17    | krn       | karena      | 133   | yaampunn   | yaampun     |
| 18    | krna      | karena      | 134   | dprd       | daripada    |
| 19    | td        | tadi        | 135   | egk        | tidak       |
| 20    | mlam      | malam       | 136   | ttp        | tetap       |
| 21    | gw        | saya        | 137   | kesehata   | kesehatan   |
| 22    | gue       | saya        | 138   | pakk       | bapak       |
| 23    | jangn     | jangan      | 139   | tlp        | telepon     |
| 24    | kgk       | kagak       | 140   | mangat     | semangat    |
| 25    | bal       | bali        | 141   | mangaaattt | semangat    |
| 26    | ajaaaaa   | aja         | 142   | sempet     | sempat      |
| 27    | ntn       | nonton      | 143   | sklh       | sekolah     |
| 28    | lgi       | lagi        | 144   | klsnya     | kelasnya    |
| 29    | lgi       | lagi        | 145   | iniii      | ini         |
| 30    | yaaaa     | yang        | 146   | padet      | padat       |
| 31    | gblg      | goblok      | 147   | ruma       | rumah       |
| 32    | pemrintah | pemerintah  | 148   | pake       | pakai       |
| 33    | ttp       | tetap       | 149   | tetep      | tetap       |
| 34    | bodetabek | jabodetabek | 150   | pagiii     | pagi        |
| 35    | asw       | asu         | 151   | grgr       | gara-gara   |
| 36    | udh       | udah        | 152   | gegara     | gara-gara   |
| 37    | kmrn      | kemarin     | 153   | ngak       | tidak       |
| 38    | ekonom    | ekonomi     | 154   | skuy       | ayo         |
| 39    | junjunga  | junjungan   | 155   | skuuy      | ayo         |
| 40    | istil     | istilah     | 156   | kuy        | ayo         |
| 41    | janua     | januari     | 157   | skrg       | sekarang    |

|    |             |            |     |            |            |
|----|-------------|------------|-----|------------|------------|
| 42 | perna       | pernah     | 158 | takde      | tidak ada  |
| 43 | sm          | sama       | 159 | mainn      | main       |
| 44 | bsok        | besok      | 160 | aplg       | apalagi    |
| 45 | pssbb       | psbb       | 161 | utk        | untuk      |
| 46 | ak          | aku        | 162 | slamat     | selamat    |
| 47 | prna        | pernah     | 163 | cus        | pergi      |
| 48 | bangat      | banget     | 164 | masi       | masih      |
| 49 | bln         | bulan      | 165 | sbg        | sebagai    |
| 50 | hmpir       | hampir     | 166 | tsb        | tersebut   |
| 51 | gaprna      | gapernah   | 167 | bangeetttt | banget     |
| 52 | ktemu       | ketemu     | 168 | daaann     | dan        |
| 53 | yauda       | yaudah     | 169 | pling      | paling     |
| 54 | emg         | emang      | 170 | tpi        | tapi       |
| 55 | jkt         | jakarta    | 171 | plann      | rencana    |
| 56 | cvd         | covid      | 172 | blm        | belum      |
| 57 | msih        | masih      | 173 | bs         | bisa       |
| 58 | bnrn        | beneran    | 174 | rese       | ganggu     |
| 59 | smpe        | sampai     | 175 | jgn        | juga       |
| 60 | psb         | psbb       | 176 | pdhl       | padahal    |
| 61 | gua         | gue        | 177 | sampe      | sampai     |
| 62 | kenceeeeeng | kencang    | 178 | ttup       | tutup      |
| 63 | smpai       | sampai     | 179 | gada       | tidak ada  |
| 64 | tgl         | tanggal    | 180 | moga       | semoga     |
| 65 | nant        | nanti      | 181 | pnya       | punya      |
| 66 | mles        | malas      | 182 | rame       | ramai      |
| 67 | pny         | punya      | 183 | bsk        | besok      |
| 68 | syg         | sayang     | 184 | yamg       | yang       |
| 69 | thn         | tahun      | 185 | mnurut     | menurut    |
| 70 | tng         | tangerang  | 186 | sdh        | sudah      |
| 71 | tmpt        | tempat     | 187 | akn        | akan       |
| 72 | anjg        | anjing     | 188 | ttp        | tetap      |
| 73 | bdg         | bandung    | 189 | dtg        | datang     |
| 74 | anj         | anjing     | 190 | slama      | selamat    |
| 75 | karna       | karena     | 191 | disuru     | disuruh    |
| 76 | lagiii      | lagi       | 192 | dimulau    | dimulai    |
| 77 | sby         | surabaya   | 193 | cape       | capek      |
| 78 | mlg         | malang     | 194 | dket       | dekat      |
| 79 | hihhhhhh    | hih        | 195 | dkt        | dekat      |
| 80 | perhati     | perhatikan | 196 | skrg       | sekarang   |
| 81 | setop       | stop       | 197 | dkmana     | dimana     |
| 82 | liha        | lihat      | 198 | blg        | bilang     |
| 83 | boosss      | bos        | 199 | pamarentah | pemerintah |
| 84 | sblmnya     | sebelumnya | 200 | bole       | boleh      |
| 85 | hppy        | happy      | 201 | laksanakan | laksanakan |
| 86 | slese       | selesai    | 202 | ttapi      | tetapi     |

|     |                |                 |       |           |             |
|-----|----------------|-----------------|-------|-----------|-------------|
| 87  | tpi            | tapi            | 203   | klo       | kalau       |
| 88  | bgiin          | bagiin          | 204   | gimano    | gimana      |
| 89  | dket           | deket           | 205   | aktivts   | aktivitas   |
| 90  | kyk            | seperti         | 206   | sekarng   | sekarang    |
| 91  | stress         | stres           | 207   | maren     | kemarin     |
| 92  | yekan          | yakan           | 208   | sebellll  | sebal       |
| 93  | mlihat         | melihat         | 209   | dfahami   | dipahami    |
| 94  | bsk            | besok           | 210   | adlh      | adalah      |
| 95  | beroperasi     | beroperasi      | 211   | mnghdpi   | menghadapi  |
| 96  | bagaiman       | bagaimana       | 212   | msyrkt    | masyarakat  |
| 97  | inget          | ingat           | 213   | nnti      | nanti       |
| 98  | kmna           | kemana          | 214   | sndri     | sendiri     |
| 99  | orng           | orang           | 215   | warg      | warga       |
| 100 | gais           | guys            | 216   | dIm       | dalam       |
| 101 | pemirsaaahh    | pemirsa         | 217   | pemrintah | pemerintah  |
| 102 | eaaa           | ya              | 218   | konta     | kontak      |
| 103 | banged         | banget          | 219   | mrk       | mereka      |
| 104 | bangett        | banget          | 220   | drmh      | dirumah     |
| 105 | jmlah          | jumlah          | 221   | diem      | diam        |
| 106 | penu           | penuh           | 222   | tugad     | tugas       |
| 107 | mengkawatirkan | mengkhawatirkan | 223   | bnget     | banget      |
| 108 | siappaaa       | siapa           | 224   | skrgkan   | sekarang    |
| 109 | yaaakk         | yakan           | 225   | tsbt      | tersebut    |
| 110 | bange          | banget          | 226   | ga        | tidak       |
| 111 | yodah          | yaudah          | 227   | gaje      | tidak jelas |
| 112 | pgn            | pengen          | 228   | bpk       | bapak       |
| 113 | besa           | besar           | 229   | mg        | moga        |
| 114 | ape            | apa             | 230   | kesygan   | kesayangan  |
| 115 | sekaliny       | sekalinya       | 231   | dg        | dengan      |
| 116 | rmh            | rumah           | 232   | pk        | bapak       |
| 233 |                | kpn             | kapan |           |             |

