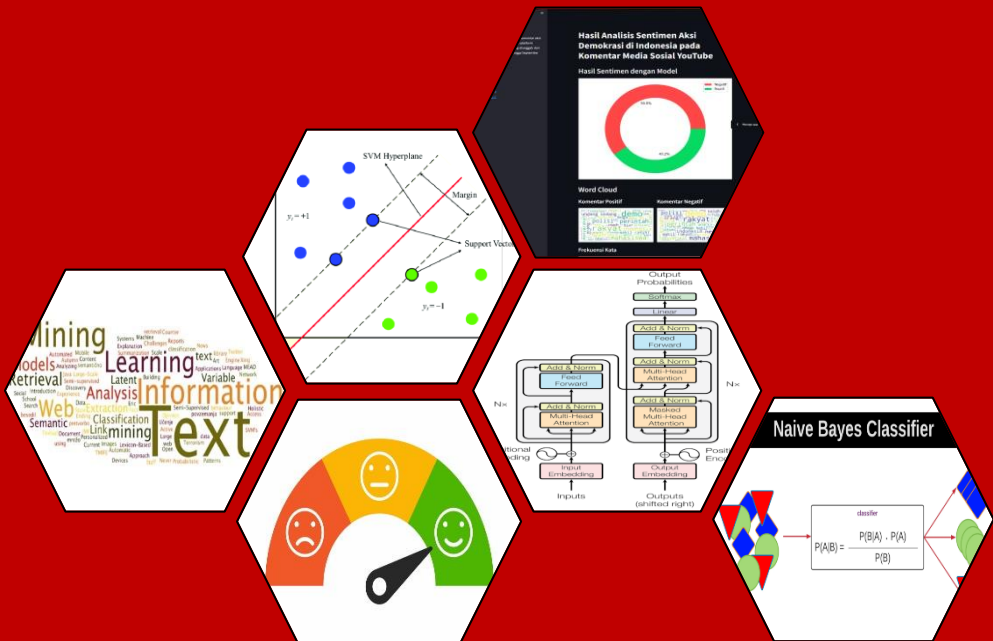


**ANALISIS SENTIMEN MASYARAKAT TERHADAP  
AKSI DEMONSTRASI DI INDONESIA PADA MEDIA  
SOSIAL *YOUTUBE* MENGGUNAKAN METODE  
INDOBERT, *NAÏVE BAYES*, DAN  
*SUPPORT VECTOR MACHINE***



**FAIZAH MAPPANYOMPA**

**H071201045**

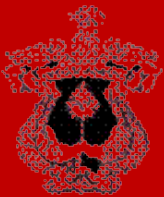
**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2024**



**ANALISIS SENTIMEN MASYARAKAT TERHADAP  
AKSI DEMONSTRASI DI INDONESIA PADA MEDIA  
SOSIAL *YOUTUBE* MENGGUNAKAN METODE  
INDOBERT, *NAÏVE BAYES*, DAN  
*SUPPORT VECTOR MACHINE***

**FAIZAH MAPPANYOMPA**

**H071201045**



**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP  
AKSI DEMONSTRASI DI INDONESIA PADA MEDIA  
SOSIAL *YOUTUBE* MENGGUNAKAN METODE  
INDOBERT, *NAÏVE BAYES*, DAN  
*SUPPORT VECTOR MACHINE***

FAIZAH MAPPANYOMPA

H071201045

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Sistem Informasi

pada

**PROGRAM STUDI SISTEM INFORMASI  
DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERESITAS HASANUDDIN  
MAKASSAR**

**2024**

# SKRIPSI

## ANALISIS SENTIMEN MASYARAKAT TERHADAP AKSI DEMONSTRASI DI INDONESIA PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN METODE INDOBERT, NAÏVE BAYES, DAN SUPPORT VECTOR MACHINE

**FAIZAH MAPPANYOMPA**

**H071201045**

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Sistem Informasi pada  
31 Mei 2024

dan dinyatakan telah memenuhi syarat kelulusan

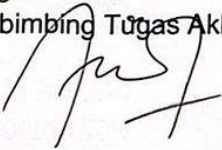
pada

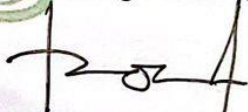
Program Studi Sistem Informasi  
Departemen Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin  
Makassar

Mengesahkan:

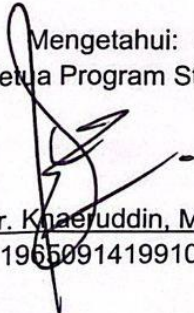
Pembimbing Tugas Akhir,

Pembimbing Pendamping,

  
A. Muh. Amil Siddik, S.Si., M.Si  
NIP. 199110032019031015

  
Rozalina Amran, S.T., M.Eng  
NIDN. 199102242019044001

Mengetahui:  
Ketua Program Studi,

  
Dr. Khaeruddin, M.Sc.  
NIP. 196509141991031003



## PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Analisis Sentimen Masyarakat Terhadap Aksi Demonstrasi di Indonesia pada Media Sosial *Youtube* Menggunakan Metode *IndoBERT*, *Naïve Bayes*, dan *Support Vector Machine*" adalah benar karya saya dengan arahan dari pembimbing A. Muh. Amil Siddik, S.Si., M.Si sebagai Pembimbing Utama dan Rozalina Amran, S.T., M.Eng sebagai Pembimbing Pendamping. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 6-Juni-2024



74053ALX199141687  
FAIZAH MAPPANYOMPA  
H071201045



## UCAPAN TERIMA KASIH

Puji dan syukur kita panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala pertolongan, rahmat, dan kasih sayang-Nya yang diberikan penulis hingga saat ini. Shalawat serta salam senantiasa tercurahkan kepada baginda Nabi Muhammad *Shallallahu 'Alaihi Wa Salam*. Alhamdulillah, berkat nikmat dan pertolongan yang diberikan oleh-Nya, penulis dapat menyelesaikan skripsi ini dengan judul "**ANALISIS SENTIMEN MASYARAKAT TERHADAP AKSI DEMONSTRASI DI INDONESIA PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN METODE INDOBERT, NAÏVE BAYES, DAN SUPPORT VECTOR MACHINE**".

Skripsi ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar Strata Satu (S1) di Universitas Hasanuddin Fakultas Matematika dan Ilmu Pengetahuan Alam Departemen Matematika Prodi Sistem Informasi. Dengan skripsi ini, penulis berharap dapat berkontribusi dan memberikan manfaat bagi perkembangan ilmu pengetahuan dan teknologi khususnya pada bidang *natural language processing* atau pemrosesan bahasa alami. Semoga hasil penelitian pada skripsi ini dapat dijadikan referensi demi perkembangan yang lebih baik.

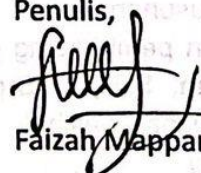
Perjalanan dalam menyelesaikan penulisan skripsi tentu banyak tantangan yang telah dihadapi. Namun, berkat kehendak-Nya penulis berhasil menyelesaikan skripsi ini. Oleh karena itu, pada kesempatan ini, izinkan penulis mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta, Ayahanda **Mappanyompa Rukka** dan Ibunda **Tati Herawati** yang telah membesarkan, mendidik, memberikan kasih sayang dan dukungan penuh sehingga penulis dapat menyelesaikan studi. Tanpa doa dan restu mereka, penulisan skripsi mungkin tidak dapat terselesaikan. Penulis juga mengucapkan terima kasih kepada pihak lainnya yang telah memberikan dukungan baik secara psikis atau materiil, bimbingan, dan motivasi dalam proses penyusunan skripsi ini:

1. Rektor Universitas Hasanuddin, Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M. Sc.**, beserta jajarannya.
2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Bapak **Dr. Eng. Amiruddin, S.Si., M.Si.**, beserta jajarannya.
3. Ketua Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Bapak **Dr. Firman, S.Si., M.Si.**, beserta jajarannya.
4. Ketua Program Studi Sistem Informasi Bapak **Dr. Khaeruddin, M.Sc.**, beserta jajarannya.
5. Dosen pembimbing utama Bapak **A. Muh. Amil Siddik, S.Si., M.Si** yang telah memberikan arahan, bimbingan, dan panduan selama proses penyusunan skripsi.
6. Dosen pembimbing pertama sekaligus penasehat akademik Ibu **Rozalina Amran, S.T., M.Eng.** yang telah meluangkan waktu untuk memberikan konsultasi, nasihat, dan memberikan arahan selama proses penyusunan skripsi.

7. Dosen penguji pertama Bapak **Dr. Eng. Armin Lawi, S.Si., M.Eng.**, yang bersedia meluangkan waktu untuk menguji, memberikan saran, dan masukan berharga untuk memperbaiki skripsi ini.
8. Dosen penguji kedua Bapak **Jeriko Gormantara, S.Si., M.Si.** yang bersedia meluangkan waktu untuk menguji selama seminar proposal, seminar hasil, dan sidang skripsi serta memberikan segala saran untuk perbaikan skripsi yang lebih baik.
9. **Kakak-kakak penulis** tersayang, dr. Fatimah Mappanyompa, M.Biomed dan Muhammad Mappanyompa, S.Kom., M.Eng., yang telah memberikan semangat untuk menyelesaikan pengerjaan skripsi dan memenuhi kebutuhan finansial penulis selama menjadi mahasiswa.
10. Tim *developer* sekaligus teman seperjuangan masa kuliah **Galaxy Dev**, Kakak Azhar Tawakkal, Haerul, Mamet, Ojan, dan Iman. Terima kasih sudah menjadi sahabat belajar, berbagi informasi, memberikan semangat, serta tempat berkeluh kesah selama perkuliahan.
11. Teman-teman mentor, *cohort*, dan tim di **Bangkit Academy 2024 Batch 1** yang memberikan penulis kesempatan sebagai mentor kelas *Machine Learning-05*. Umpan yang positif dan pengembangan kompetensi diri selama menjadi mentor membantu penulis tetap semangat dalam menyelesaikan skripsi.
12. **Nilam Maghfira** teman seperjuangan kuliah yang menjadi teman belajar dan membantu penulis dalam mengurus administrasi skripsi dan beasiswa S2 saat penulis masih melaksanakan KKNT di Bantaeng.
13. Sahabat TK hingga SMA **Sekolah Indonesia Makkah**, Jamilah, Doa, Fedda, Nun, Naiyay, dan Fatim yang menjadi teman belajar *online* selama pengerjaan skripsi.
14. Teman-teman **KKNT 111 ITTG Bantaeng 2 Posko 3 Paten**, yaitu amalia, meisya, widiya, kurniawan, muhaimin, nadhil, dan thamrin atas kerja samanya sehingga dapat menyelesaikan program ini sebagai salah satu syarat kelulusan.
15. Seluruh teman-teman **Sistem Informasi Angkatan 2020** yang telah menjadi teman perjuangan dari masa kuliah daring hingga akhir perkuliahan di Universitas Hasanuddin.

Penulis menyadari bahwa skripsi ini jauh dari kata sempurna, oleh karena itu dengan segala kerendahan hati, penulis ingin meminta maaf dan berikan kritik dan saran yang membangun. Akhir kata, penulis berharap skripsi ini memberikan manfaat bagi penulis dan pembaca, Amin.

Penulis,



Faizah Mappanyompa

## ABSTRAK

FAIZAH MAPPANYOMPA. **Analisis Sentimen Masyarakat Terhadap Aksi Demonstrasi di Indonesia pada Media Sosial *Youtube* Menggunakan Metode IndoBERT, *Naïve Bayes*, dan *Support Vector Machine*** (dibimbing oleh A. Muh. Amil Siddik dan Rozalina Amran).

**Latar belakang.** Pesatnya perkembangan media sosial, khususnya YouTube, memberikan masyarakat untuk berinteraksi dan partisipasi aktif dalam diskusi. YouTube tidak hanya menjadi platform berbagi informasi, tetapi tempat bagi masyarakat untuk menyampaikan komentar terkait isu-isu yang ada, seperti aksi demonstrasi. Analisis sentimen terhadap komentar yang terdapat pada platform YouTube menjadi kunci untuk memahami dan pengambilan keputusan yang tepat. **Tujuan.** Penelitian ini bertujuan mengetahui hasil klasifikasi sentimen, membandingkan kinerja model, dan merancang dashboard hasil analisis sentimen dari aksi demonstrasi di Indonesia pada platform YouTube. **Metode.** Algoritma yang digunakan yaitu IndoBERT, *Naïve Bayes*, dan *Support Vector Machine*. **Hasil.** Hasil klasifikasi menunjukkan metode SVM mencapai nilai akurasi sebesar 71.31%, presisi 74.4%, recall 74.03%, dan f-1 score 74.31%. Metode *Naïve Bayes* mencapai nilai akurasi sebesar 70.47%, presisi 70.97%, recall 79.65%, dan f-1 score 75.0%. Sementara metode IndoBERT mencapai nilai akurasi sebesar 78.67%, presisi 87.16%, recall 82.25%, dan f-1 score 84.58%. **Kesimpulan.** Berdasarkan nilai metrik tersebut, metode IndoBERT menunjukkan kinerja yang lebih unggul dibandingkan dua model lainnya. Penelitian ini menegaskan bahwa dalam analisis sentimen masyarakat terhadap aksi demonstrasi di Indonesia melalui platform *YouTube*, metode IndoBERT menunjukkan kinerja yang lebih baik.

**Kata Kunci:** Analisis Sentimen, IndoBERT, *Naïve Bayes*, *Support Vector Machine*, *YouTube*



## ABSTRACT

FAIZAH MAPPANYOMPA. **Analysis of Public Sentiment towards Demonstrations in Indonesia on Youtube Social-Media Using IndoBERT, Naïve Bayes, and Support Vector Machine Methods** (supervised by A. Muh. Amil Siddik and Rozalina Amran)

**Background.** The rapid development of social media, especially YouTube, has allowed people to interact and actively participate in discussions. YouTube is not only an information-sharing platform, but a place for people to comments related to existing issues, such as demonstrations. Sentiment analysis of comments on the YouTube platform is key to understanding and making the right decisions. **Aim.** This research aims to determine the results of sentiment classification, compare model performance, and design a dashboard of sentiment analysis results from demonstrations in Indonesia on the YouTube platform. **Methods.** The algorithms used are IndoBERT, Naïve Bayes, and Support Vector Machine. **Results.** The classification results show that SVM achieved an accuracy of 71.31%, precision 74.4%, recall 74.03%, and f-1 score 74.31%. Naïve Bayes achieved an accuracy of 70.47%, precision 70.97%, recall 79.65%, and f-1 score 75.0%. While the IndoBERT achieved an accuracy of 78.67%, precision 87.16%, recall 82.25% and f-1 score 84.58%. **Conclusion.** Based on the metrics value, the IndoBERT method shows superior performance compared to the other two models. This research confirms that in analyzing public sentiment towards demonstrations in Indonesia through the YouTube platform, the IndoBERT method shows better performance.

**Keywords:** Sentiment Analysis, IndoBERT, Naïve Bayes, Support Vector Machine, YouTube

## DAFTAR ISI

HALAMAN JUDUL .....	I
PERNYATAAN PENGAJUAN .....	II
HALAMAN PENGESAHAN .....	III
PERNYATAAN KEASLIAN SKRIPSI .....	IV
UCAPAN TERIMA KASIH .....	V
ABSTRAK .....	VII
ABSTRACT .....	VIII
DAFTAR ISI .....	IX
DAFTAR TABEL .....	XI
DAFTAR GAMBAR .....	XII
DAFTAR RUMUS .....	XIII
DAFTAR LAMPIRAN .....	XIV
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Teori .....	3
1.6.1 Analisis Sentimen .....	3
1.6.2 BERT .....	4
1.6.3 IndoBERT .....	7
1.6.4 <i>Naïve Bayes</i> .....	7
1.6.5 Support Vector Machine .....	8
1.6.6 Perbandingan Model .....	8
1.6.7 Metrics Evaluation .....	9
1.6.8 Python .....	10
1.6.9 YouTube .....	11
1.6.10 PyTorch .....	12
1.6.11 WordCloud .....	12
1.6.12 Streamlit .....	12
1.6.13 Google Colaboratory .....	13
1.6.14 Visual Studio Code .....	13
1.6.15 Aksi Demonstrasi di Indonesia .....	13
BAB II METODE PENELITIAN .....	15
2.1 Sumber Data .....	15
2.1.1 Metode Pelabelan Data .....	15
2.2 Waktu dan Tempat Penelitian .....	16
2.2.1 Waktu Penelitian .....	16
2.2.2 Tempat Penelitian .....	16

2.3 Tahapan Penelitian .....	16
2.3.1 Tahapan Penelitian Menggunakan IndoBERT .....	19
2.3.2 Tahapan Penelitian Menggunakan <i>Naïve Bayes</i> .....	20
2.3.3 Tahapan Penelitian Menggunakan SVM .....	21
2.4 Rancangan Dashboard pada Streamlit .....	22
2.5 Instrumen Penelitian .....	25
2.5.1 Perangkat Keras ( <i>Hardware</i> ) .....	25
2.5.2 Perangkat Lunak ( <i>Software</i> ) .....	26
BAB III HASIL DAN PEMBAHASAN .....	27
3.1 <i>Data Collection</i> .....	27
3.2 <i>Data Labelling</i> .....	28
3.3 <i>Data Pre-Processing</i> .....	29
3.3.1 <i>Drop Unrelated Comment</i> .....	29
3.3.2 <i>Remove Emoji</i> .....	29
3.3.3 <i>Change Mistype and Slang Words</i> .....	30
3.3.4 <i>Remove Numbers, Punctuation, and Special Characters</i> .....	31
3.3.5 <i>Case Folding</i> .....	32
3.3.6 <i>Remove Stopwords</i> .....	32
3.3.7 <i>Stemming with Sastrawi</i> .....	33
3.4 <i>Tokenization</i> .....	34
3.5 <i>Feature Extraction (TF-IDF)</i> .....	35
3.6 <i>Exploratory Data Analysis</i> .....	38
3.6.1 Hasil Sentimen Positif dan Negatif .....	38
3.6.2 Frekuensi Kata .....	39
3.6.3 <i>Word Cloud</i> .....	39
3.7 <i>Data Modelling</i> .....	41
3.7.1 Implementasi <i>Support Vector Machine</i> .....	41
3.7.2 Implementasi <i>Naïve Bayes</i> .....	42
3.7.3 Implementasi IndoBERT .....	42
3.8 Model Evaluation .....	45
3.9 <i>Model Deployment</i> .....	48
BAB IV KESIMPULAN DAN SARAN .....	52
4.1 Kesimpulan .....	52
4.2 Saran .....	52
DAFTAR PUSTAKA .....	53
LAMPIRAN .....	56

## DAFTAR TABEL

Nomor Urut	Halaman
1. Perbandingan Model .....	8
2. Confusion Matrix.....	9
3. Sumber Data Komentar Aksi Demonstrasi di Indonesia.....	15
4. Tabel Waktu Penelitian.....	16
5. Tabel Spesifikasi Perangkat Keras.....	26
6. Tabel Perangkat Lunak .....	26
7. Komentar Yang Tidak Relevan.....	29
8. Hasil Penerapan Penghapusan Emotikon .....	30
9. Hasil Perubahan Kata Salah dan Bahasa Gaul .....	31
10. Hasil penghapusan angka, tanda-tanda baca, dan karakter khusus .....	31
11. Tabel Hasil <i>Case Folding</i> .....	32
12. Hasil Penghapusan <i>Stopwords</i> .....	33
13. Hasil <i>Stemming</i> dengan Sastrawi.....	33
14. Jumlah Komentar Sebelum dan Sesudah Pembersihan .....	33
15. Hasil Tokenisasi pada Model SVM dan <i>Naïve Bayes</i> .....	34
16. Hasil Tokenisasi pada Model IndoBERT .....	35
17. Contoh Tabel Dokumen.....	35
18. Contoh Tabel TF.....	36
19. Contoh Tabel IDF .....	36
20. Contoh Tabel TF-IDF.....	37



## DAFTAR GAMBAR

Nomor Urut	Halaman
1. Arsitektur <i>Transformers</i> (Vaswani <i>et al.</i> , 2017) .....	5
2. Representasi Input <i>BERT</i> (Devlin <i>et al.</i> , 2018) .....	6
3. Tahapan Pre-Training dan Fine-Tuning <i>BERT</i> (Devlin <i>et al.</i> , 2018) .....	6
4. <i>Flowchart</i> Tahapan Penelitian .....	18
5. Alur Penelitian algoritma IndoBERT .....	19
6. Alur Penelitian algoritma <i>Naïve Bayes</i> .....	20
7. Alur Penelitian algoritma SVM .....	21
8. Hasil Sentimen Positif dan Negatif .....	22
9. Kata-kata pada komentar positif dan negatif .....	23
10. Frekuensi Kata-Kata pada Komentar .....	23
11. Komparasi Akurasi Setiap Model .....	24
12. Tentang Dataset dan Sumber Dataset .....	24
13. <i>Dashboard</i> Hasil Analisis Sentimen Aksi Demokrasi di Indonesia .....	25
14. Hasil pengambilan data menggunakan <i>YouTube Data API</i> .....	28
15. Hasil pelabelan pada dataset .....	28
16. 3 <i>Key</i> dan <i>Value</i> dari Kata-Kata Salah dan Gaul .....	30
17. <i>Word Vector</i> Frekuensi Setiap Kata di Dokumen .....	36
18. Contoh <i>Word Vector</i> yang sudah terbobot .....	38
19. Grafik Lingkaran Hasil Sentimen Positif dan Negatif .....	38
20. Grafik Batang Hasil 10 Kata Teratas .....	39
21. <i>Word Cloud</i> Umum .....	40
22. <i>Word Cloud</i> Positif .....	40
23. <i>Word Cloud</i> Negatif .....	41
24. Pemodelan SVM .....	42
25. Pemodelan <i>Naïve Bayes</i> .....	42
26. Inisialisasi <i>Common Function</i> .....	43
27. Penerapan Kelas <i>EntailmentDataset</i> dan <i>EntailmentDataLoader</i> .....	44
28. <i>Training History</i> Data Latih dan Validasi .....	44
29. Hasil <i>Confusion Matrix SVM</i> .....	45
30. Hasil <i>Confusion Matrix Naïve Bayes</i> .....	46
31. Hasil <i>Confusion Matrix IndoBERT</i> pada data uji .....	47
32. Hasil <i>Confusion Matrix IndoBERT</i> pada data validasi .....	47
33. Contoh <i>Script</i> Visualisasi <i>WordCloud</i> pada <i>File Python</i> .....	49
34. Dependensi dari <i>requirement.txt</i> .....	50
35. Proses <i>deploy dashboard</i> .....	50
36. Hasil <i>dashboard</i> analisis sentimen aksi demonstrasi di Indonesia menggunakan <i>Streamlit</i> yang sudah diterapkan ke web .....	51

**DAFTAR RUMUS**

Nomor Urut	Halaman
1. Teorema Bayes .....	7
2. Metrics Evaluation: Accuracy.....	9
3. Metrics Evaluaton: Precision.....	9
4. Metrics Evaluation: Recall.....	9
5. Metrics Evaluation: F-1 Score.....	9

**DAFTAR LAMPIRAN**

Nomor Urut	Halaman
1. <i>Source code</i> pengambilan data menggunakan <i>YoutubeDataAPI</i> .....	<b>56</b>
2. <i>Source Code Data Cleaning</i> .....	<b>57</b>
3. <i>Source Code Data Modelling</i> .....	<b>58</b>
4. <i>Source Code Model Deployment</i> .....	<b>59</b>
5. <i>Data Source</i> .....	<b>60</b>

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Media sosial menjadi bagian penting bagi masyarakat pada zaman modern ini. Hadirnya media sosial memudahkan masyarakat dalam mengakses berbagai informasi yang tersedia. Media sosial merupakan tempat dimana para pengguna dapat berpartisipasi, berbagi, menciptakan isi jejaring sosial serta forum di dunia virtual (Qadri, 2020). Selain itu, media sosial memudahkan pengguna dalam berpartisipasi aktif dalam sebuah diskusi. Pengguna dapat saling terhubung dan berinteraksi satu sama lain tanpa dibatasi ruang dan waktu. Salah satu media sosial yang memenuhi kriteria tersebut adalah *YouTube*.

*YouTube* merupakan layanan media sosial berbagi video terbesar di dunia. Pada bulan Januari tahun 2023, pengguna *YouTube* di Indonesia menempati peringkat keempat di dunia sebesar 139 juta pengguna (Sadya, 2023). Bagi masyarakat Indonesia, *YouTube* menjadi salah satu platform untuk berbagi pendapat, menyebarkan informasi, serta pandangan terkait isu yang ada. Isu terkait sosial dan politik merupakan isu yang sering muncul dalam ruang publik di Indonesia, contohnya aksi demonstrasi di Indonesia.

Video-video terkait demonstrasi seringkali diunggah pada kanal berita di *YouTube*. Dalam video tersebut, beragam komentar yang disampaikan masyarakat mulai dari yang mendukung, menentang, dan tidak menyetujui kedua pihak. Untuk memahami setiap opini yang diutarakan, penting untuk melakukan analisis sentimen terhadap isu aksi demonstrasi yang disampaikan melalui komentar pada platform *YouTube*.

Analisis sentimen menjadi topik yang populer dikarenakan pesatnya perkembangan media sosial dan peningkatan opini publik. Meningkatnya kebutuhan untuk memahami opini publik, data teks yang dihasilkan memainkan peran penting untuk diproses dengan tujuan untuk dianalisis sampai tahap pengambilan keputusan yang tepat (Tan *et al.*, 2023). Analisis sentimen atau dikenal dengan *opinion mining* merupakan proses pengolahan data tekstual untuk melihat informasi sentimen positif, netral atau negatif yang terkandung dalam suatu kalimat (Rozi *et al.*, 2013). Adanya opini dan sentimen publik, menjadikan analisis sentimen sebagai alat penting untuk memahami kecenderungan opini masyarakat dari berbagai bidang seperti hiburan, bisnis, politik, dan lainnya.

Penelitian sebelumnya terkait analisis sentimen aksi demokrasi di Indonesia pernah dilakukan oleh (Made Tara Okta Adriana *et al.*, 2023) menggunakan data yang diambil dari Twitter dari rentang Januari 2010 sampai Mei 2022. Hasil penelitian menunjukkan bahwa masyarakat memiliki sentimen negatif lebih besar dibandingkan sentimen positif dan sentimen netral. Sentimen negatif sebesar 54,98%, diikuti sentimen positif sebesar 15,86%, dan 29,15% untuk masyarakat yang bersikap netral



dari total 2.036.786 data *tweets*. Penelitian tersebut mendapatkan hasil akurasi yang lebih baik ketika menggunakan algoritma *Support Vector Machine* (SVM) dibandingkan algoritma *Random Forest* yaitu sebesar 80,3% dengan skala data latih dan uji 90:10. Penelitian yang dilakukan oleh (Nikmah *et al.*, 2022) mengenai analisis sentimen pelecehan seksual di media sosial *Twitter* menunjukkan bahwa akurasi dengan model SVM dan *Naïve Bayes* lebih unggul dibandingkan dengan model yang terbaru. Akurasi tersebut sebesar 86.54% untuk SVM, 85.45% untuk *Naïve Bayes*, dan 84.62% untuk *Long-Short Term Memory* (LSTM). Penelitian sebelumnya yang dilakukan oleh (Fransiscus & Girsang, 2022) mengenai sentimen analisis Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) saat COVID-19 di Indonesia, menunjukkan bahwa *fine-tuning* IndoBERT memiliki nilai *precision* sebesar 86%, diikuti dengan *Multinomial Naïve Bayes* sebesar 83% dan SVM sebesar 84%.

Hasil penemuan sebelumnya menunjukkan bahwa IndoBERT merupakan model terbaru dengan performa yang tinggi yang dirilis pada tahun 2020. Namun, terdapat beberapa penelitian yang menunjukkan bahwa SVM dan *Naïve Bayes* merupakan metode yang lebih baik dari segi akurasi dibandingkan model yang lebih baru seperti LSTM (1997) untuk mengklasifikasikan opini publik. *BERT* merupakan metode yang efektif bagi analisis sentimen dikarenakan algoritma ini dikembangkan untuk NLP dan dapat diaplikasikan untuk memproses data yang besar. Oleh karena itu, penulis memilih metode *BERT*, spesifiknya IndoBERT sebagai salah satu model yang akan dibandingkan dengan *Naïve Bayes* dan SVM.

Berdasarkan uraian di atas, penulis bermaksud untuk mendalami analisis sentimen terkait aksi demonstrasi pada platform media sosial lain, yaitu *YouTube*. Solusi yang penulis ajukan menggunakan tiga metode yaitu *IndoBERT*, *Naïve Bayes*, dan SVM dengan tujuan membandingkan akurasi kinerja masing-masing model. Dengan penelitian ini, diharapkan informasi yang dihasilkan memberikan manfaat untuk memahami opini publik, pembuatan keputusan yang tepat, dan mengembangkan strategi serta memperbaiki kebijakan bagi pemerintah untuk mengatasi permasalahan terkait isu aksi demonstrasi di Indonesia.

## 1.2 Rumusan Masalah

Bedasarkan latar belakang masalah di atas, maka rumusan masalah pada penelitian ini sebagai berikut:

1. Bagaimana hasil sentimen aksi demonstrasi di Indonesia pada media sosial *YouTube* menggunakan algoritma IndoBERT, *Naïve Bayes*, dan *Support Vector Machine*?
2. Bagaimana perbandingan kinerja antara ketiga model dalam mengklasifikasikan sentimen tentang aksi demonstrasi di Indonesia pada media sosial *YouTube*?
3. Bagaimana *dashboard* hasil sentimen dari aksi demonstrasi di Indonesia dapat ditampilkan dalam bentuk *website*?

### 1.3 Batasan Masalah

Agar pembahasan tidak terlalu luas, diperlukan batasan-batasan masalah sebagai tolak ukur untuk mencapai target. Beberapa batasan masalah tersebut yaitu:

1. Data yang digunakan pada penelitian ini adalah data komentar dengan 5 video dari rentang Oktober 2020 hingga September 2022.
2. Data yang digunakan pada penelitian ini berasal dari tiga kanal *YouTube*, masing-masing 3 video dari CNN Indonesia, 1 video dari BBC News Indonesia, dan 1 video dari Official iNews.
3. Klasifikasi sentimen akan dibagi menjadi sentimen positif dan sentimen negatif.
4. Pembagian data latih dan uji dengan perbandingan 80:20.
5. Uji kriteria klasifikasi menggunakan *metrics evaluation* dengan nilai *accuracy*, *precision*, *recall*, dan *f-1 score*.

### 1.4 Tujuan Penelitian

Adapun tujuan penelitian kali ini sebagai berikut:

1. Merancang hasil klasifikasi sentimen aksi demonstrasi di Indonesia pada media sosial *YouTube* menggunakan metode IndoBERT, *Naïve Bayes*, dan SVM.
2. Merancang perbandingan kinerja antara ketiga model tersebut untuk analisis sentimen aksi demonstrasi di Indonesia pada media sosial *YouTube*.
3. Merancang *dashboard* hasil analisis sentimen dari aksi demonstrasi di Indonesia yang ditampilkan dalam bentuk *website*.

### 1.5 Manfaat Penelitian

Manfaat yang diharapkan pada penelitian ini sebagai berikut:

1. Memahami sentimen publik terkait aksi demonstrasi di Indonesia menggunakan metode IndoBERT, *Naïve Bayes*, dan SVM.
2. Memahami sentimen publik untuk memperbaiki kebijakan yang dibuat untuk mengatasi permasalahan terkait isu aksi demonstrasi di Indonesia.
3. Mengidentifikasi opini publik mengenai isu yang mendorong aksi demonstrasi di Indonesia.

### 1.6 Teori

#### 1.6.1 Analisis Sentimen

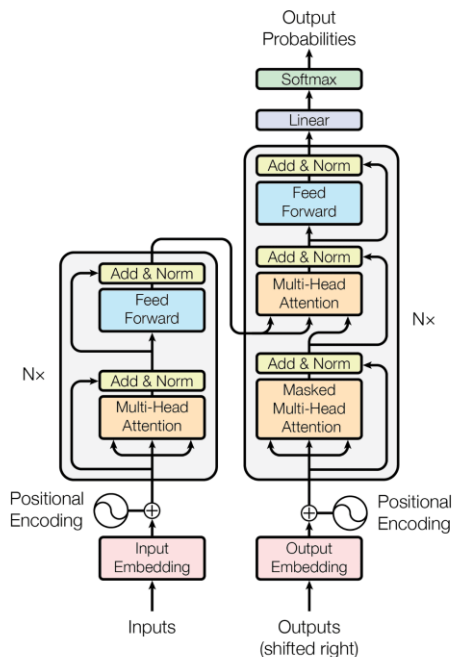
Analisis sentimen, yang juga dikenal dengan sebutan *opinion mining* merupakan sub bidang *Natural Language Processing* (NLP) yang menganalisis suatu kalimat untuk menentukan sentimen atau opini seseorang. Analisis ini bertujuan untuk

mengetahui sentimen, perspektif, evaluasi, emosi, terkait isu dan topik pada suatu individu, organisasi, atau aktivitas (Riza Kurniawanda & Adline Twince Tobing, 2022). Menurut (Liu *et al.*, 2005) klasifikasi sentimen dapat dikategorikan menjadi sentimen positif atau negatif. Sentimen positif merupakan kalimat yang diucapkan dengan kata-kata baik dan tidak mengandung unsur serangan. Sedangkan sentimen negatif merupakan kalimat yang mengandung unsur menyinggung, ejekan, dan tidak mendukung suatu pihak. Pengklasifikasian sentimen ini dapat membantu dalam mengevaluasi kinerja atau produk serta dalam hal pengambilan keputusan (Made Tara Okta Adriana *et al.*, 2023).

### 1.6.2 BERT

*Bidirectional Encoder Representations from Transformers* atau disingkat dengan BERT merupakan model representasi bahasa yang dikembangkan oleh Google pada tahun 2018. BERT merupakan *pre-trained* model yang artinya telah dilatih dengan representasi dua arah (*bidirectional*) dari teks yang tidak berlabel dengan cara mengkondisikan konteks sebuah teks dari kiri dan kanan layer. Hasilnya, model *pre-trained* BERT dapat dilakukan *fine-tuning* dengan tambahan satu kondisi *output layer* saja (Devlin *et al.*, 2018).

BERT memiliki arsitektur yang terdiri atas beberapa tumpukan *layer encoder* atau dikenal dengan *multi-layer bidirectional Transformer* (Vaswani *et al.*, 2017). *Transformer* memiliki dua mekanisme yang digunakan yaitu *encoder* dan *decoder*. *Encoder* digunakan untuk membaca input. Sedangkan, *decoder* digunakan sebagai *output* untuk menentukan prediksi. *Transformer* ini juga menggunakan metode *self-attention* yang mempelajari hubungan antara input dan output dalam kata. Arsitektur dari *transformer* dapat dilihat pada gambar 1.

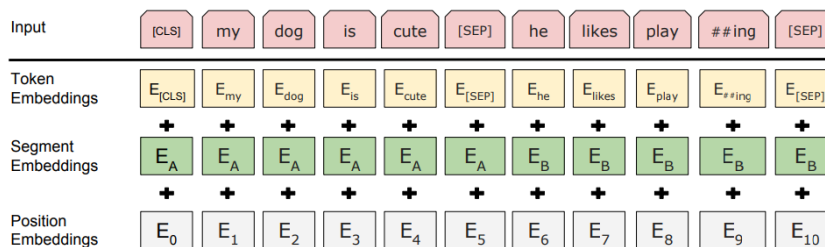


Gambar 1 Arsitektur *Transformers* (Vaswani et al., 2017)

Setiap lapisan *encoder* memiliki dua sub *layer*, yaitu *Multi-Head Self-Attention* dan *Fully Connected Feed-Forward Network*. *Multi-Head* artinya menjalankan beberapa mekanisme *self-attention* secara paralel. Setiap input *sequence* akan melewati lapisan *self-attention* yang menentukan berapa bobot untuk ditambahkan pada setiap token. Output dari *attention* ini akan dimasukkan sebagai input di *feed-forward network*. Input *sequence* merupakan *sequence* input token ke BERT berupa satu atau dua kalimat yang digabungkan bersama.

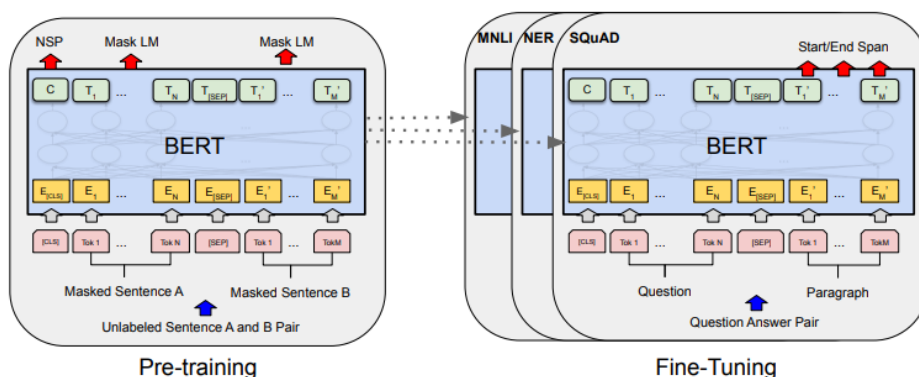
Sebelum memahami *sequence* input token lebih jauh, terdapat teknik yang digunakan pada BERT, yaitu *WordPiece embeddings* yang menggunakan sebanyak 30.000 kosakata (*vocabulary*) token (Wu et al., 2016). Setiap *sequence* input diawali dengan token khusus yaitu ([CLS]) atau token *classification*. Kemudian, menambahkan token khusus ([SEP]) atau token *seperated* pada akhir kalimat untuk membedakan setiap kalimat. Setiap token ditambahkan *segment embedding* untuk membedakan apakah kata tersebut dari kalimat A atau kalimat B. Untuk setiap token yang telah diberikan, tambahkan *positional embedding* untuk posisi token pada kalimat. Dari rangkaian proses ini, maka inputan dari *encoder* BERT merupakan jumlah dari segmen token, segmen *embedding*, dan *positional embedding*. Representasi input ini dapat dilihat pada gambar 2.





Gambar 2 Representasi Input BERT (Devlin et al., 2018)

Terdapat dua tahapan utama dalam BERT yaitu *pre-training* dan *fine-tuning*. *Pre-train* BERT menggunakan dua tugas *unsupervised* yaitu *Masked Language Modelling* (MLM) dan *Next Sentence Prediction* (NSP). Dalam penelitian BERT, MLM menyamarkan (*mask*) sekitar 15 persen dari input token *WordPiece* secara acak untuk diganti dengan token [MASK], setelah itu memprediksi token *vocabulary* yang telah disamarkan. NSP mempelajari hubungan antar kalimat untuk menentukan apakah kalimat pertama merupakan lanjutan kalimat kedua. Sedangkan, *Fine-Tuning* merupakan mekanisme *self-attention* yang memungkinkan BERT memodelkan tugas lain baik *single text* atau *pair of text* dengan menukar input dan output yang sesuai. *Fine-Tuning* mengambil *pre-trained* model (yang telah dilatih) dan melatihnya lebih lanjut untuk pada dataset untuk menyesuaikan tugas spesifik seperti sentimen analisis atau *question-answering*. Menurut (Devlin et al., 2018) *hyperparameter* yang optimal digunakan saat *fine-tuning* yaitu *batch size* sebesar 16 dan 32, optimizer yang digunakan adalah Adam (*learning rate* 5e-5, 3e-5, dan 2e-5) pada *dev set* (*validation set*) dan panjang epoch 2, 3, atau 4. Ilustrasi *pre-training* dan *fine-tuning* dapat dilihat pada gambar 3.



Gambar 3 Tahapan Pre-Training dan Fine-Tuning BERT (Devlin et al., 2018)

### 1.6.3 IndoBERT

IndoBERT (*Indonesia Bidirectional Encoder Representations from Transformers*) merupakan model *transformer-based pre-trained* dari model *BERT* untuk pengolahan teks bahasa Indonesia. IndoBERT dilatih menggunakan 220 juta kata dari tiga sumber utama: Wikipedia Indonesia (74 juta kata), artikel berita dari Kompas, Tempo, Liputan 6 (total 55 juta kata), dan web korpus Indonesia (90 juta kata). Selain itu IndoBERT telah dievaluasi menggunakan dataset IndoLEM (*Indonesian Language Understanding Evaluation and Benchmarking*). Model ini dilatih sebanyak 180 *epochs* dengan hasil akhir yang mirip dengan *English BERT*. Model IndoBERT pernah diuji sebelumnya dan mencapai performa yang tinggi pada tugas seperti *part-of-speech tagging*, *Named Entity Recognition (NER)*, termasuk analisis sentimen (Koto *et al.*, 2020)

### 1.6.4 Naïve Bayes

Algoritma *Naïve Bayes* merupakan teknik pengklasifikasian berdasarkan Teorema Bayes dengan asumsi independen antar fitur, artinya fitur yang digunakan untuk klasifikasi tidak bergantung satu sama lain dengan label atau target kelas yang diberikan (Hlaing Moe *et al.*, 2018). *Naïve Bayes* merupakan algoritma yang menggunakan *word* dan *feature frequency* dari suatu dokumen teks untuk menghitung probabilitas dari kelas sentimen. Teks dokumen tersebut dapat direpresentasikan sebagai sekumpulan kata-kata yang belum terurut dan hanya menyimpan frekuensinya dalam dokumen (Jurafsky & Martin, 2000). Sebagai contoh ketika merepresentasikan sebuah kalimat “Maju terus mahasiswa demi rakyat Indonesia.” dan “Panjang umur perjuangan !!!! Hidup Mahasiswa, Hidup Rakyat Indonesia”, maka yang dicatat berupa kata “mahasiswa”, “hidup”, “rakyat” dan “indonesia” sebanyak 2 kali dan sisa kata pada kalimat tersebut sebanyak 1 kali.

Untuk menghitung klasifikasi tersebut terdapat perhitungan statistik Teorema Bayes yang digunakan untuk menghitung probabilitas dari suatu kejadian. Teorema ini memiliki persamaan sebagai berikut:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (1)$$

Keterangan:

$x$  : kelas negatif atau positif

$y$  : dokumen teks

$P(x|y)$  : peluang dokumen teks x diklasifikasikan ke dalam kelas y

$P(y|x)$  : peluang dokumen teks y diklasifikasikan ke dalam kelas x

$P(x)$  : peluang kelas x

$P(y)$  : peluang dokumen y

### 1.6.5 Support Vector Machine

*Support Vector Machine* (SVM) merupakan salah satu algoritma *supervised learning* yang digunakan untuk klasifikasi atau regresi. Algoritma SVM bekerja dengan menggunakan *decision boundary* atau batas keputusan untuk memetakan titik data menjadi beberapa kelas sehingga membentuk sebuah model linier atau *hyperplane* yang optimal dalam mengklasifikasikan data (Putri *et al.*, 2022).

### 1.6.6 Perbandingan Model

Berikut tabel perbandingan ketiga model berdasarkan pembahasan di atas.

**Tabel 1** Perbandingan Model

Fitur	SVM	Naïve Bayes	IndoBERT
Jenis	<i>Supervised Learning</i>	<i>Supervised Learning</i>	<i>Neural Network (Transformers)</i>
Metode	Menggunakan <i>decision boundary</i> (batas keputusan) untuk memetakan titik data menjadi beberapa kelas.	Mengklasifikasikan data berdasarkan probabilitas.	Merepresentasikan setiap kata dan hubungannya dengan kata yang lain.
Kelebihan	Efektif untuk dataset berskala kecil.	Cepat dalam melatih data dan fitur yang memiliki asumsi independen.	Baik dalam performa klasifikasi teks dan tugas semantik.
Kekurangan	Komputasi tinggi untuk dataset yang besar. Memerlukan parameter tuning yang tepat (Rahman Isnain <i>et al.</i> , 2021).	Sebab asumsi yang independen, memungkinkan kinerja buruk pada data kontinu.	Membutuhkan komputasi yang tinggi.

### 1.6.7 Metrics Evaluation

*Metrics Evaluation* atau evaluasi metrik merupakan metrik yang digunakan untuk mengukur kinerja dari suatu model, dalam hal ini klasifikasi. Metrik ini memberikan gambaran seberapa baik model dapat melakukan generalisasi dan membandingkan tingkat akurasi model yang berbeda (Chauhan, 2020). Evaluasi model dapat dilihat melalui tabel *confusion matrix*.

*Confusion Matrix* merupakan sebuah tabel perbandingan nilai sebenarnya dengan dengan nilai yang diprediksi. Tabel tersebut menunjukkan jumlah dari *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) (Nikmah et al., 2022).

**Tabel 2** Confusion Matrix

<b>Confusion Matrix</b>		<b>Predicted Values</b>	
		<b>Negative</b>	<b>Positive</b>
<b>Actual Values</b>	<b>Negative</b>	TN	FP
	<b>Positive</b>	FN	TP

Keterangan:

*True Positive* (TP) : kelas yang memprediksi positif dan benar sebagai positif

*True Negative* (TN): kelas yang memprediksi negatif dan benar sebagai negatif

*False Positive* (FP): kelas yang memprediksi positif dan salah sebagai negatif

*False Negative* (FN) : kelas yang memprediksi negatif dan salah sebagai positif

*Confusion matrix* ini didapatkan dari nilai evaluasi metrik yaitu *accuracy*, *precision*, dan *recall*. Adapun penjelasan masing-masing metrik sebagai berikut:

1. *Accuracy* merupakan evaluasi metrik yang memprediksi jumlah benar dibagi dengan seluruh nilai prediksi

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (2)$$

2. *Precision* merupakan evaluasi metrik yang memprediksi klasifikasi data positif benar dibagi dengan data yang terklasifikasi positif.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

3. *Recall* merupakan evaluasi metrik yang memprediksi klasifikasi data positif benar dibagi dengan keseluruhan data yang benar positif.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

4. *F-1 Score* merupakan evaluasi metrik yang yang menggabungkan perhitungan antara nilai presisi dan recall.

$$\text{F1 Score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (5)$$

### 1.6.8 Python

*Python* merupakan bahasa pemrograman populer, berorientasi objek, *open-source*, dan kinerjanya yang tinggi. Sering kali *python* digunakan oleh *developer* karena bahasanya yang mudah digunakan dalam bekerja. *Python* menjadi bahasa pilihan banyak *developer* dalam mengembangkan aplikasi dari mobile hingga web, proses *machine learning*, *data science*, dan visualisasi data. Selain itu, *python* juga dikenal dengan koleksi *library* dan *module* yang sangat luas. *Library python* merupakan kode pra-tulis yang menyediakan berbagai fungsi seperti analisis data, *machine learning*, dan masih banyak lagi. Dengan *library* ini, *developer* dapat membangun aplikasi secara efisien dan menggunakan kode yang dapat digunakan kembali (Rao, 2023). Pada penelitian kali ini, terdapat beberapa *library* populer yang akan membantu dalam pemrosesan data seperti *Matplotlib*, *re*, *Pandas*, *Natural Language Toolkit* (NLTK), *Scikit-Learn*, dan *Sastrawi*.

#### 1.6.8.1 Matplotlib

*Matplotlib* merupakan sebuah *library* interaktif untuk menghasilkan visualisasi grafik, bagan, plot berkualitas tinggi dalam beberapa format. Beberapa visualisasi tersebut berupa plot garis, histogram, grafik batang, grafik pai, kurva, dan sebagainya (Nederkoorn, 2020).

#### 1.6.8.2 re

*re* merupakan sebuah modul yang sudah tersedia dalam *library* standar *python*. *Library* ini berfungsi sebagai *regular expression* yang digunakan untuk mengubah dan memanipulasi teks (Kuchling, n.d.). Selain itu, *library* ini banyak digunakan untuk proses pembersihan data dan beberapa tugas NLP.

#### 1.6.8.3 Pandas

*Pandas* merupakan salah satu *library python* yang digunakan untuk bekerja dengan kumpulan data di *machine learning*. Kegunaan *pandas* dapat memanipulasi dan melakukan analisis data. Selain itu, *pandas* menyediakan berbagai metode yang sudah tersedia, seperti menggabungkan dua atau lebih data, mengkombinasikan data, menyaring data, hingga fungsi untuk *time series* (Rao, 2023).

#### 1.6.8.4 Natural Language Toolkit

*Natural Language Toolkit* atau disingkat dengan *NLTK* merupakan sebuah *framework* dan *library* berbasis *python* yang digunakan untuk bekerja pemrosesan data. *NLTK* menyediakan lebih dari 50 kumpulan korpus dengan pemrosesan teks untuk klasifikasi, tokenisasi, *stemming*, *tagging* dan tugas *NLP* lainnya (*NLTK:: Natural Language Toolkit*, 2023).

#### 1.6.8.5 Scikit-Learn

*Scikit-learn* merupakan *library* yang menyediakan algoritma untuk *machine learning* dan beberapa tugas *data mining*. *Scikit-learn* dapat melakukan prediksi analisis data untuk *supervised* dan *unsupervised learning* seperti regresi, klasifikasi, *Decision Trees*, *Random Forest*, *XGBoost*, *SVM*, *Naïve Bayes*, dan juga metrik evaluasi (Mahapatra, 2023). Pada penelitian ini, model *SVM* dan *Naïve Bayes* yang akan digunakan berasal dari *library scikit-learn*.

#### 1.6.8.6 Sastrawi

Sastrawi merupakan *library python* untuk mengurangi kata-kata dalam bahasa Indonesia ke bentuk dasarnya. Teknik ini dikenal dengan *stemming*. *Stemming* digunakan untuk mengurangi kepadatan kata dan meningkatkan prediksi dan pelatihan kerja *machine learning* (Sastrawi, 2016). Sebagai contoh jika kita memiliki kalimat 'Perekonomian Indonesia sedang dalam pertumbuhan yang membanggakan' maka output *stemming* yang dihasilkan dari sastrawi adalah 'ekonomi indonesia sedang dalam tumbuh yang bangga'.

#### 1.6.9 YouTube

*YouTube* merupakan layanan situs atau *web* berbagi video terbesar yang memiliki misi untuk memberi kebebasan untuk menyampaikan pendapat dan menunjukkan dunia kepada setiap orang. Kebebasan dapat diartikan sebagai kebebasan dalam berekspresi, mendapatkan informasi, menyampaikan pendapat, dan tempat berkarya (Aji, 2022). *YouTube* memiliki banyak topik mulai dari hiburan, pendidikan, politik, olahraga, dan berbagai informasi lainnya. Sebagai contoh KompasTV, CNN Indonesia, dan tvOneNews merupakan kanal di *YouTube* yang menyajikan konten berita seputar nasional serta internasional. Banyak individu, perusahaan, maupun organisasi menggunakan *YouTube* untuk mendapatkan penghasilan melalui konten yang diunggah.

### 1.6.9.1 YouTube Data API

*YouTube Data API (Application Programming Interface)* merupakan alat untuk mengambil data dari *YouTube* termasuk data komentar pengguna di berbagai video. Hal ini dapat digunakan untuk analisis sentimen dan memantau interaksi pengguna. *YouTube* API memungkinkan pengambilan data komentar lebih struktur dan otomatis, serta memudahkan *developer* dalam mengambil berbagai data lainnya seperti analisis konten atau pelacakan tren saat ini.

### 1.6.10 PyTorch

*PyTorch* atau *torch* dikembangkan oleh grup *Meta AI Research* pada tahun 2016. *PyTorch* merupakan *framework open-source deep learning* yang dapat melakukan komputasi dinamis, pembuatan model secara fleksibel hingga arsitektur *neural network* yang kompleks seperti *computer vision (CV)* dan *NLP*. *PyTorch* memiliki implementasi model yang disebut dengan *PyTorch-Transformers*, yaitu sebuah library untuk melatih model NLP seperti *IndoBERT*.

### 1.6.11 WordCloud

*WordCloud* merupakan visualisasi kumpulan kata dalam bentuk awan. *WordCloud* menghitung frekuensi dari setiap teks yang dihasilkan. Teks tersebut ditampilkan dalam bentuk awan dimana ukuran teks merepresentasikan frekuensi dan seberapa penting teks tersebut (Mueller, 2022). Jika teks tersebut memiliki ukuran yang lebih besar dari teks yang lainnya, maka teks tersebut memiliki frekuensi yang paling banyak disebut. Untuk menampilkan visualisasi tersebut, *WordCloud* pada penelitian ini diambil melalui library *python* yang tersedia di *GitHub* milik Mueller.

### 1.6.12 Streamlit

*Streamlit* merupakan sebuah *framework open-source* untuk *machine learning* dan *data science* menggunakan bahasa pemrograman *Python*. *Streamlit* dapat digunakan sebagai kerangka untuk membangun aplikasi *web*. Integrasi model sentimen analisis ke *streamlit* dapat menyediakan fitur interaktif yang mudah untuk menganalisis sentimen. Dalam hal ini merancang *dashboard* atau ringkasan hasil sentimen analisis beserta akurasi secara *real-time*. Kemampuan *streamlit* dalam mengkombinasikan model *machine learning* dengan pengembangan *web* menjadikan *streamlit* alat yang ideal untuk *deploy* model (Mhadhbi, 2021).

### 1.6.13 Google Colaboratory

*Google Colaboratory* atau sering disebut dengan *Colab* merupakan sebuah *Jupyter Notebook* interaktif berbasis cloud yang menyediakan GPU dan CPU secara gratis. *Colab* menyediakan fitur untuk menulis kode dan *rich text* dalam satu dokumen seperti menambahkan gambar, tautan, HTML, LaTeX, dan lainnya. *Colab* banyak digunakan untuk tugas *data science* dan *machine learning* mulai dari mengirimkan data, analisis data, melatih model, evaluasi model, hingga visualisasi model hanya dalam beberapa baris kode. *Google Colab* dijalankan dalam *server google cloud* yang artinya GPU dan (*Tensor Processing Unit*) TPU disediakan oleh google (Lacok, n.d.).

### 1.6.14 Visual Studio Code

*Visual Studio Code* atau disingkat dengan *VSCode* merupakan sebuah *code editor* ringan yang mendukung banyak bahasa pemrograman, salah satunya yaitu bahasa pemrograman *python*. *VSCode* memiliki banyak fitur seperti *debugging*, *syntax highlighting*, *snippets*, *code factoring*, dan lainnya. Selain itu, *VSCode* menyediakan fitur ekstensi gratis yang dapat digunakan oleh pengguna untuk kecepatan dan fleksibilitas dalam pengerjaan kode (Lacok, n.d.). Dalam penelitian ini, *VSCode* akan digunakan sebagai *code editor* untuk mengembangkan aplikasi *web streamlit*.

### 1.6.15 Aksi Demonstrasi di Indonesia

Demokrasi berasal dari kata *demos* yang artinya rakyat dan *krator* yang artinya pemerintahan. Maka demokrasi dapat diartikan sebagai pemerintahan dari rakyat, oleh rakyat, dan untuk rakyat (Fakultas Hukum UMSU, 2023). Pasal 1 ayat 3 UU Nomor 9 Tahun 1998 menjelaskan bahwa demonstrasi adalah kegiatan yang dilakukan oleh seseorang atau lebih untuk mengeluarkan pikiran dengan lisan, tulisan, dan sebagainya secara demonstratif di muka umum. Indonesia merupakan negara hukum dan demokrasi yang menjunjung tinggi hak kebebasan berekspresi dan berpendapat. Negara Indonesia juga dapat dikatakan sebagai kiblat negara demokrasi di wilayah Asia. Hal ini ditunjukkan dengan masuknya negara Indonesia dalam *Global Democracy Ranking* yang membuktikan Indonesia memiliki potensi luar biasa dalam memimpin negara demokrasi. Memberikan kesempatan kepada rakyat dalam menentukan kebijakan yang dikeluarkan oleh pemerintah merupakan bentuk demokrasi yang ideal (Fithriana *et al.*, n.d.).

Demonstrasi dan hak kebebasan berpendapat di media sosial memiliki kaitan erat. Seringkali informasi mengenai demonstrasi diunggah di media sosial sehingga rakyat atau pengguna sosial media tersebut dapat menyampaikan pendapat atau opini miliknya. Namun, tidak jarang aksi demonstrasi tersebut dilakukan dengan keributan



seperti pembakaran, membahayakan nyawa seseorang, bahkan menjadi aksi *narcisstic* di media sosial (Syahrul Jiwandono, 2020).

Bedasarkan permasalahan di atas, peneliti ingin menganalisis sentimen masyarakat terhadap aksi demonstrasi di Indonesia melalui komentar-komentar pengguna yang diutarakan di media sosial *YouTube*.