

Alisya

by - -

Submission date: 11-Nov-2023 05:09AM (UTC+1100)

Submission ID: 2224010549

File name: Draft_Skripsi_3_181023.docx (1.06M)

Word count: 3369

Character count: 21922

**SENTIMEN ANALISIS PADA TWITTER
MENGUNAKAN NAÏVE BAYES DAN LSTM UNTUK
PEMILU PRESIDEN 2024**

SKRIPSI SARJANA

23

Skripsi sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari
Fakultas Teknologi Komunikasi dan Informatika

Oleh

Alisya Mutia Mantika

207064516054



16

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2023/2024

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunianya sehingga penulis dapat menyelesaikan skripsi dengan judul “SENTIMEN ANALISIS PADA TWITTER MENGGUNAKAN NAÏVE BAYES DAN LSTM UNTUK PEMILU PRESIDEN 2024” sebagai salah satu syarat kelulusan Program Studi Sarjana Fakultas Teknologi Komunikasi dan Informatika.

Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, Bapak Dr. Agung Triayudi, S.Kom., M.Kom. dan Bu Rima Tamara Aldisa, S.Kom., M.Kom. yang telah meluangkan banyak waktu, tenaga, pikiran, arahan, motivasi serta memaklumi kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi. Penulis juga mengucapkan banyak terima kasih kepada:

1. Ayah dan Bunda selaku orang tua penulis yang telah banyak memberi dukungan dan menyemangati penulis.
2. Seluruh dosen pengajar di Program Studi Informatika maupun dosen di Program Studi lain yang membantu penulis .
3. Teman-teman seangkatan dan sehimpuan berbagai angkatan yang telah membantu dan mendukung.
4. Seluruh keluarga besar penulis yang tidak bisa disebutkan satu persatu tetapi tidak mengurangi rasa Terimakasih dan Syukur penulis.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuannya dengan hal yang jauh lebih baik dari apa yang diterima penulis. Dalam penulisan skripsi ini penulis paham dan sadar akan berbagai kesalahan maupun kekeliruan yang secara sengaja maupun tidak sengaja dilakukan, maka dari itu kritik dan saran akan sangat membantu penulis. Semoga Skripsi ini memberikan manfaat di bidang Informatika.

Jakarta, Oktober 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	6 ii
DAFTAR ISI	iii
BAB I PENDAHULUAN	v
1.1 Latar Belakang	v
1.2 Identifikasi Masalah	vi
1.3 Tujuan Penelitian	vii
1.4 Manfaat Penelitian	vii
1.5 Batasan Masalah	vii
BAB II TINJAUAN PUSTAKA	viii
2.1 Kajian Teori	viii
2.2.1 Sentimen Analisis	viii
2.2.2 Machine Learning	viii
2.2.3 Python	ix
2.2.4 Naïve Bayes	ix
2.2.5 Long Short-Time Memory	ix
2.2.6 Twitter Harvest	x
2.2.7 Confusion Matrix	xi
2.2.8 Crawling Data	xi
2.2.9 Data Preprocessing	xii
2.2 Studi Literatur	xii
BAB III Metode Penelitian	xv
3.1 Waktu Penelitian	xv
3.2 Alat dan Bahan	xv
3.3 Fokus Penelitian	xv
3.4 Sumber Data	xvi
3.5 Tahapan Penelitian	xvi
3.5.1 Identifikasi Masalah	xvi
3.5.2 Studi Literatur	xvi
3.5.3 Perancangan Program	xvi
3.5.4 Pengumpulan Data	xvii
3.5.5 Metode Pengujian	xvii
DAFTAR PUSTAKA (MENDELEY)	xxi



28
BAB I
PENDAHULUAN

1.1 Latar Belakang

24
Indonesia sebagai negara yang berpaham demokrasi sesuai dengan Pasal 1 ayat 2 UUD 1945 yang menyatakan bahwa “Kedaulatan berada di tangan rakyat dan dilaksanakan menurut Undang-Undang Dasar” (Silalahi & Hartanto, 2023). Pada tahun 2024, Indonesia akan melaksanakan pemilihan presiden yang diadakan setiap 5 tahun sekali yang menjunjung tinggi demokrasi. Pemilihan presiden merupakan kegiatan politik yang bersifat demokrasi karena suara atau pilihan rakyat menjadi penentu lewat pemungutan suara yang dilaksanakan secara simultan (Delvika et al., 2023).

37
Pemilihan tokoh calon presiden dan wakil presiden pastinya akan mempertimbangkan opini dan respon dari masyarakat terhadap setiap aspek seperti gagasan, visi misi, program yang akan dilaksanakan, masalah yang akan dipecahkan, dan lainnya. Dari opini tersebut akan tercipta survei lewat popularitas atau kecenderungan masyarakat. Dari opini masyarakat itulah tercipta pro dan kontra terkait calon presiden dan wakil presiden yang menjadi acuan dalam memilih kandidat yang tepat (Al-Areef & Saputra S, 2023).

Berkampanye adalah aktifitas meyakinkan dan mengumpulkan pendukung kandidat dengan memberikan atau menawarkan gagasan, visi, misi, program, dan untuk diri seorang kandidat. Menjelang pesta demokrasi pemilihan presiden, para kandidat sudah mulai berunjuk diri melalui media sosial sebelum masa berkampanye dimulai.

Teknologi dan media sosial yang terus menerus berkembang juga akan berdampak kepada penggunaanya yang terus menerus bertambah (Syafii, 2023). Masyarakat Indonesia memanfaatkan media sosial sebagai media mendapatkan informasi, berkampanye, dan beropini secara public. Karena cenderung menjadi wadah masyarakat berkeluh kesah dan mengungkapkan pendapat, media sosial juga bisa mempengaruhi pandangan publik (Humam & Laksito, 2023).

31
Twitter menjadi salah satu media sosial yang kerap digunakan untuk mengemukakan pendapat, berkampanye, mendapatkan informasi terkini,

membahas topik yang sedang hangat, dan berdiskusi secara objektif karena *Twitter* membatasi penulisan *tweets* hanya 280 kata(Vindua & Zailani, 2023). Indonesia berada di peringkat ke 5 pengguna media sosial *Twitter* terbanyak di dunia, atau sekitar 14,8 Juta pengguna. Data di media sosial *Twitter* berupa teks bisa diolah menjadi informasi yang berguna(Wenando et al., 2020). Salah satu topik yang saat ini kerap dibahas adalah masalah pemilihan presiden, baik tentang politik maupun aktifitas kandidatnya. Media sosial *Twitter* sebagai penyedia data opini universal dan juga menjadi sarana untuk ikut terlibat aktif dalam demokrasi serta mendukung capres pilihan (Zain et al., 2023). Karena *Twitter* media yang menampung aspirasi atau opini masyarakat tak jarang *tweets* atau konten yang bersifat negatif atau berisi cacian, makian maupun kata kata kasar.

Sentimen analisis adalah proses klasifikasi atau Analisa opini, sentiment dan emosi individu yang dituangkan kedalam bentuk teks. Sentimen analisis dilakukan untuk menentukan apakah teks tersebut cenderung berkonotasi negative atau positif (Asmara et al., 2020). Lazimnya sentimen analisis dimanfaatkan untuk kepentingan politik, pemerintahan, Pendidikan, bisnis dan lainnya (Muzaki & Witanti, 2021).

17 Algoritma yang digunakan adalah algoritma *Naïve bayes* dan *LSTM*. Algoritma *Naïve Bayes* berfungsi untuk mengklasifikasikan data uji dengan mencari nilai probabilitas tertinggi(Juanita, 2020). Naive bayes kebanyakan digunakan untuk memprediksi peluang dari sebuah data. Algoritma *Long Short-Term Memory (LSTM)* berfungsi untuk menganalisa data dan mengidentifikasi data yang ada sehingga dapat diklasifikasikan. LSTM merupakan salah satu bagian dari algoritma RNN yang dikembangkan untuk mengatasi masalah ketergantungan jangka panjang. Algoritma LSTM sudah terbukti memiliki keakuratan tertinggi diangka 78% dibandingkan algoritma lainnya.(Srinivas et al., 2021).

25

1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut terdapat identifikasi masalah yang akan di Bahasa pada penelitian ini, ialah:

- Kegiatan berkampanye masih sering menggunakan kata-kata kasar dan saling menjatuhkan kandidat lainnya.

- Media social sering digunakan untuk mencaci, memaki, dan menjatuhkan satu sama lain.
- Masih banyak masyarakat yang pasif dalam demokrasi
- Semakin dekat dengan hari pemilu semakin gencar juga para pendukung kandidat tertentu sehingga mencoba segala cara untuk menjatuhkan atau membuat rumor tentang kandidat lain.

35

1.3 Tujuan Penelitian

Tujuan Penelitian ini ialah:

1. Menampilkan data yang relevan dengan pemilu 2024
2. Mengklasifikasikan sentiment analisis dari setiap kandidat calon presiden 2024
3. Melihat respon masyarakat terhadap setiap kandidat calon presiden 2024
4. Menghasilkan klasifikasi terbaik dalam menganalisa sentiment Masyarakat pada pemilu 2024

1.4 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Menghasilkan data negatif, positif dan netral dari opini masyarakat sehingga bisa dijadikan acuan dalam mengadakan kampanye atau pemilihan umum tanpa menggunjing atau mengadu domba kandidat presiden 2024.
2. Untuk memberikan data dari kampanye atau kandidat yang lebih unggul
3. Untuk melihat partisipasi masyarakat pada pemilu 2024

36

1.5 Batasan Masalah

Penelitian ini dibatasi pada:

1. Penelitian ini menggunakan data pada *Twitter*
2. Data yang diambil dari *Twitter* berupa *tweets*/konten yang berkaitan dengan calon presiden atau dengan hashtag yang berkaitan.
3. Penelitian ini menggunakan algoritma *Naïve Bayes* dan *LSTM* untuk mengolah data.
4. Penelitian ini menggunakan bahasa pemrograman *Python*.
5. Penelitian ini menggunakan perangkat lunak *Google Colab*.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Teori

2.2.1 Sentimen Analisis

Sentimen analisis adalah proses klasifikasi atau analisa opini, sentiment dan emosi individu yang dituangkan kedalam bentuk teks. Sentimen analisis dilakukan untuk menentukan apakah teks tersebut cenderung berkonotasi negatif atau positif (Asmara et al., 2020). Lazimnya sentimen analisis dimanfaatkan untuk kepentingan politik, pemerintahan, Pendidikan, bisnis dan lainnya (Muzaki & Witanti, 2021).

Sentimen analisis dipergunakan untuk mengakumulasi opini masyarakat akan suatu isu sehingga data opini masyarakat tersebut dapat diklasifikasikan dan dimanfaatkan sesuai dengan kebutuhan. Sentimen analisis juga dimanfaatkan untuk memantau pandangan masyarakat terhadap aktivitas politik seperti mengamati sikap masyarakat terhadap kandidat calon presiden (Al-Areef & Saputra S, 2023). Sentimen analisis mengklasifikasikan data dalam beberapa bagian menurut kategori positif, negatif atau netral (Buntoro et al., 2021).

2.2.2 Machine Learning

Machine learning ialah cabang kecerdasan buatan (AI) dan ilmu computer (*Computer Science*) yang berfokus pada penggunaan data dan algoritma untuk meniru cara manusia belajar, dan secara bertahap meningkatkan akurasi. Melalui penggunaan metode statistik, algoritma dilatih untuk membuat klasifikasi atau prediksi, dan untuk mengungkap wawasan penting dalam proyek penambangan data (*Data Mining*).

Algoritma pembelajaran mesin menggunakan metode komputasi untuk “mempelajari” informasi langsung dari data tanpa bergantung pada persamaan yang telah ditentukan sebagai model. Algoritma secara

adaptif meningkatkan kinerjanya seiring dengan meningkatnya jumlah sampel yang tersedia untuk pembelajaran.

2.2.3 Python

Python dibuat oleh Guido van Rossum dan dirilis pada tahun 1991. Python merupakan Bahasa pemrograman yang berfungsi untuk pengembangan situs web dan perangkat lunak, otomatisasi tugas, analisis data, dan visualisasi data. Python memungkinkan analisis data untuk melakukan perhitungan statistik yang kompleks, membuat visualisasi data, dan algoritma machine learning.

2.2.4 Naïve Bayes

Algoritma *Naïve Bayes* merupakan algoritma klasifikasi berbentuk probabilistik dan statistik yang berasumsi setiap atribut bersifat bebas (*independence*) atau ciri dari sebuah kelas tidak ada hubungannya dengan kelas lainnya (Fikri et al., 2020).

Berikut persamaan dari teorema Bayes (Fikri et al., 2020):

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

A, B = events

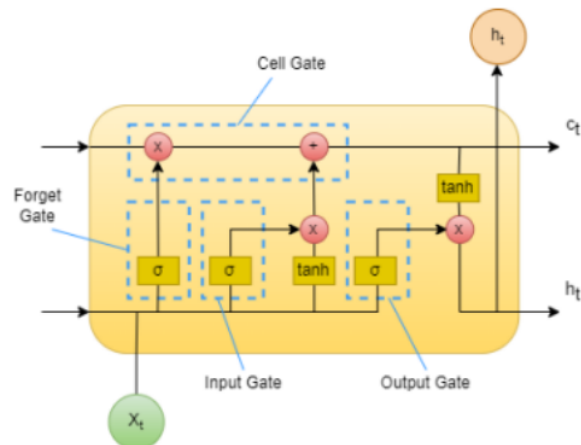
P(A|B) = probabilitas A jika B adalah True

P(B|A) = probabilitas B jika A adalah True

P(A), P(B) = probabilitas kejadian A dan B

2.2.5 Long Short-Time Memory

Algoritma LSTM ialah algoritma sejenis RNN yang digunakan untuk mempelajari kondisi jangka panjang (Srinivas et al., 2021). LSTM mempunyai tiga jenis gate, yaitu Input gate, Forget gate dan Output gate. Algoritma LSTM menyimpan informasi dalam jangka waktu yang lama sehingga digunakan untuk memproses, memperkirakan, dan mengklasifikasikan berdasarkan data.



Gambar 2.1.5 - .

2.2.6 Twitter Harvest

Tweet Harvest merupakan alat baris perintah (command-line tool) yang menggunakan Playwright untuk mengikis (Scrape) tweet dari hasil pencarian Twitter berdasarkan kata kunci dan rentang tanggal tertentu. Tweet yang diambil disimpan dalam file CSV. Skrip twitter harvest hanya untuk tujuan Pendidikan. Twitter melarang pengguna yang tidak diautentikasi melakukan pencarian atau pencarian lanjutan.

Untuk menggunakan skrip ini, Anda harus memiliki akun Twitter yang valid dan mendapatkan Token Akses, yang dapat diperoleh dengan masuk ke Twitter di browser Anda dan mengekstrak cookie auth_token. Tweet Harvest akan membuka browser Chromium dan menavigasi ke halaman pencarian Twitter. Ini kemudian akan memasukkan parameter pencarian Anda dan mengikis tweet yang dihasilkan. Tweet akan disimpan dalam file CSV di direktori bernama tweets-data di direktori kerja saat ini.

Crawl data twitter dengan Tweet harvest akan menimbulkan error jika data yang diambil lebih dari rate limit perharinya (Helmi Satria, 2023).

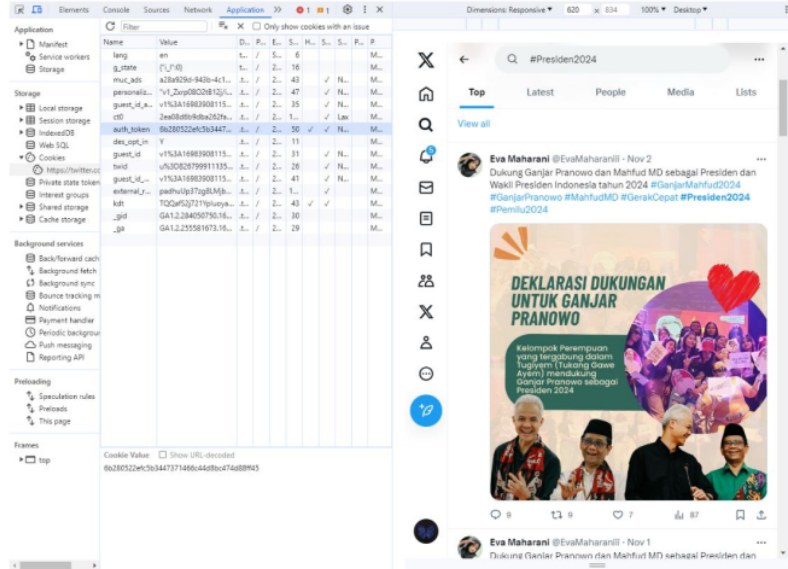
2.2.7 Confusion Matrix

Confusion Matrix merupakan pengukuran performa algoritma klasifikasi machine learning yang akan mengukur Precision, Precision, Accuracy dan F1 Score. Evaluasi performa dilakukan dengan confusion matrix yang mengukur tingkat True Positive (TP rate), tingkat True Negative (TN rate), tingkat False Positive (FP rate), dan tingkat False Negative (FN rate).

	Predicted Negatif	Predicted Positif	Predicted Netral
Aktual Negatif	True Negatif(TN)	False Positif (FP)	False Positif (FP)
Aktual Positif	True Negatif (TN)	True Positif(TP)	False Negatif (TN)
Aktual Netral	False Negatif (FN)	False Positif (FP)	True Negatif (TN)

2.2.8 Crawling Data

Dalam pengambilan atau crawling data twitter, penelitian ini menggunakan library tweet harvest lewat *Google Colaboratory* sebagai platform cloud computing yang mirip seperti *Jupyter Notebook*. Untuk crawling data pada twitter juga diperlukan token autentikasi (authentication token). Token autentikasi berguna untuk mengakses twitter yang kita sehingga tweet harvest dapat melakukan crawling data lewat akun twitter kita. Pada penelitian ini keyword yang digunakan untuk mengambil data ialah GanjarPranowo, AniesBaswedan dan Prabowo.



2.2.9 Data Preprocessing

Data Preprocessing ialah proses mengolah data asli (Raw Data) menjadi data yang mudah dipahami atau mudah diklasifikasikan (Fathir et al., 2023). Text pre-processing diperlukan untuk analisis text mining melalui beberapa tahapan yaitu: casefolding, tokenizing, filtering, stemming, tagging, dan analysis (Samsir et al., 2021).

2.2 Studi Literatur

Pada penelitian (Humam & Laksito, 2023), sentiment analisis menggunakan algoritma LSTM kemungkinan besar menghasilkan data yang tidak seimbang. Diperlukan metode oversampling untuk menyeimbangkan nilai akurasi. Metode random oversampling menghasilkan akurasi sebesar 0.82 pada epoch 25, sedangkan non-oversampling menghasilkan akurasi sebesar 0.61 pada epoch 50.

Pada penelitian (Firdlous et al., 2023), Algoritma LSTM melakukan evaluasi model dengan *confusion matrix* yang menghasilkan akurasi 78% dengan sentimen positif sebesar 52,2%, sentiment negatif sebanyak 37%, dan sentiment netral sebesar 10,8% dari 858 data dengan perbandingan 80% data train dan 20% data uji.

Pada penelitian (Hidayatullah et al., 2021), berdasarkan pengujian data pemilihan presiden 2019, algoritma deep neaural network menghasilkan akurasi terbaik dibandingkan dengan tiga algoritma machine learning traditional (SVM, Logistic Regression dan Naïve Bayes Multinomial Naïve Bayes). Berdasarkan penelitian ini, Bidirectional LSTM lebih unggul dibandingkan metode deep learning lainnya dengan akurasi 84.60%.

Pada penelitian (Muzaki & Witanti, 2021) , hasil dari data sentiment analisis data twitter tentang Pilkada 2020 ditengah pandemic covid-19 adalah dengan perbandingan 70% data training dan 30% data uji, algoritma Naïve bayes menghasilkan nilai akurasi sebesar 92,2%. Dari hasil akurasi tersebut menunjukkan dari data yang diambil terdapat lebih dari 50% hasil sentiment negative.

Pada penelitian (Srinivas et al., 2021), proses sentiment analisis menggunakan algoritma LSTM dengan 1.6 juta data twitter menghasilkan nilai akurasi sebesar 87%. Nilai akurasi tersebut merupakan nilai akurasi tertinggi dibandingkan metode lainnya yaitu metode CNN dan Neural network.

Pada penelitian (Fathir et al., 2023), penelitian ini menggunakan algoritma Naïve Bayes, Random Forest dan SVM. Dibandingkan algoritma lain yang menggunakan SMOTE, Algoritma Naïve Bayes memiliki nilai ke akurasian terendah yaitu sebesar 84,44% sedangkan nilai akurasi algoritma Random Forest sebesar 90,78% dan nilai akurasi algoritma SVM sebesar 90,70%.

Penelitian ini berbeda dengan penelitian lain karena menggunakan Algoritma Naïve Bayes dan LSTM untuk menghasilkan data sentiment analisis dan juga melihat algoritma mana yang lebih unggul akurasinya. Algoritma LSTM yang menggunakan data time series atau data jangka waktu yang panjang. Penelitian ini juga menggunakan tweet harvest yang berbeda dengan jurnal atau penelitian sebelumnya yang menggunakan

library tweepy maupun snsrape untuk mengambil atau crawling data
Twitter Api (Al-Areef & Saputra S, 2023).

BAB III Metode Penelitian

22 3.1 Waktu Penelitian

No	Kegiatan	Waktu Pelaksanaan																			
		Oktober				November				Desember				Januari				Februari			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Penentuan Topik																				
2	Studi Literatur																				
3	Pengumpulan Data																				
4	Perancangan Algoritma																				
6	Seminar Proposal																				
7	Revisi																				
8	Perancangan Sistem Lanjutan																				
9	Pengujian Sistem																				
10	Penyusunan Laporan																				
11	Sidang Akhir																				

3.2 Alat dan Bahan

Laptop Merk Acer Swift SF114-34

Spesifikasi :

- Sistem Operasi Windows 11 64 bit
- Processor Intel® Pentium® Silver N6000 @ 1.10GHz 1.11 GHz
- Ram 4GB
- Microsoft Excel 2023
- Google Colaboratory Python
- Data set Twitter

40 3.3 Fokus Penelitian

Penelitian ini berfokus pada menentukan algoritma mana yang lebih unggul diantara algoritma Naïve Bayes dan LSTM dengan menganalisis sentiment Masyarakat Indonesia dengan menganalisa karakteristik (tingkat akurasi, presisi dan recall) data tweets masyarakat Indonesia tentang kandidat calon presiden Indonesia sehingga dapat membantu atau bermanfaat pada pelaksanaan kampanye dan penyelenggaraan pemilu 2024.

3.4 Sumber Data

Data yang didapatkan diambil dengan library tweet harvest yang mengambil 512 data tweets berkaitan dengan Ganjar Pranowo, 509 data tweets berkaitan dengan Anies Baswedan dan 552 data tweets berkaitan dengan Prabowo Subianto. Data Tweets tersebut diambil pada bulan November 2023 dengan 12 Atribut yang meliputi created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorite_count, lang, user_id, conversation_id_str, username, dan tweet_url.

3.5 Tahapan Penelitian

Penelitian ini dilakukan dengan tahapan *flowchart* berikut:

Gambar 3.5 Flowchart Tahapan Penelitian 1

3.5.1 Identifikasi Masalah

Pemilu masih sering dijadikan ajang menjatuhkan dan mengadu domba antar kandidat presiden. Twitter menjadi media diskusi, mengeluarkan aspirasi dan berkampanye bagi Masyarakat Indonesia. Dalam berkampanye, pengguna media social masih

3.5.2 Studi Literatur

Pada Tahapan studi literatur, penelitian ini menggunakan referensi penelitian terdahulu sebagai teori pendukung dan juga sebagai bahan acuan untuk mempermudah pembuatan laporan penelitian.

3.5.3 Perancangan Program

Penelitian ini menggunakan Bahasa pemrograman Python dalam perancangan program dan menggunakan Google Colaboratory untuk mengeksekusi atau menjalankan program yang dibuat. Dengan Google Collab, program ini dibuat menggunakan Algoritma Machine Learning Klasifikasi Naïve Bayes dan LSTM. Program ini berfungsi untuk mempelajari pola dari data yang tersedia dengan algoritma untuk

menghasilkan output sehingga output tersebut bisa menjadi acuan dalam menentukan algoritma mana yang lebih tinggi akurasi.

3.5.4 Pengumpulan Data

Data yang didapatkan diambil dengan library tweet harvest yang mengambil 512 data tweets berkaitan dengan Ganjar Pranowo, 509 data tweets berkaitan dengan Anies Baswedan dan 552 data tweets berkaitan dengan Prabowo Subianto. Data Tweets tersebut diambil pada bulan November 2023 dengan 12 Atribut yang meliputi created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorite_count, lang, user_id, conversation_id_str, username, dan tweet_url.

Created_at	id_str	full_text	Quote_count	Reply_count	retweet_count	Favorite_count	lang	user_id	Conversation_id_str	username	tweet_url
Tue Nov 07 05:04:48 +0000 2023	1,72E+18	@Korandri @aniesbaswedan @moekahfidah Beniran blum tau? Kalo Anies bukannya petagas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, nou partai tapi sebelum lahir pun sudah Nahdliyyin. Kalo blum tau, rajin2 baca bro. https://t.co/UnEaZYQ9v	0	0	0	0	in	1,23E+18	1,72E+18	Bentir16	https://twitter.com/Bentir16/status/172175558393557487
Tue Nov 07 05:02:38 +0000 2023	1,72E+18	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Bka bahas @OldTraffordNews	0	0	0	0	in	1,44E+18	1,72E+18	choymarkochoy	https://twitter.com/choymarkochoy/status/1721755013746471308
Tue Nov 07 05:02:43 +0000 2023	1,72E+18	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarrahfud2024 #GanjarMahfud #jagademokrasi #paktelur #makpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo https://t.co/Zu4GQ8aamu	0	0	0	0	in	1,54E+18	1,72E+18	Ketik_salah24	https://twitter.com/Ketik_salah24/status/1721755034290106622

Gambar 3.2 – Tabel Data Twitter

3.5.5 Metode Pengujian

3.5.5.1 Confusion Matrix

Confusion Matrix merupakan pengukuran performa algoritma klasifikasi machine learning yang akan mengukur Precision, Precision, Accuracy dan F1 Score. Confusion Matrix tidak hanya memungkinkan penghitungan keakuratan pengklasifikasi, baik akurasi global maupun berdasarkan kelas, namun juga membantu menghitung metrik penting lainnya yang sering digunakan pengembang untuk mengevaluasi model mereka.

		Predicted Values	
		POSITIVE	NEGATIVE
Actual Values	POSITIVE	TP (True positive)	FN (False negative)
	NEGATIVE	FP (False positive)	TN (True negative)

5

Ada 4 istilah pada Confusion Matrix yaitu: *true positive (TP)*, *false positive (FP)*, *true negative (TN)*, dan *false negative (FN)*.

HASIL DAN PEMBAHASAN

4.1 Data Crawling dengan Tweet Harvest

4.2 Data Preprocessing

Pada penelitian ini data processing dilakukan untuk membersihkan atribut full_text dengan bantuan library python. Data processing yang dilakukan ialah menarik hashtag, membersihkan text (dari URL, Mention, dan emoji atau symbol lainnya), menghapus stopwords, mengubah text menjadi lowercase atau huruf kecil, dan tanda baca. Sehingga data asli atribut full_text (gambar 3.3) berubah menjadi (gambar 3.4) setelah dilakukan data preprocessing.

No	Full_text
1	Repost dari Bapak @hary_tanoe https://t.co/APgz6uppgiâ€ #GanjarPranowo #MahfudMD #HaryTanoesoedibjo #TuanGuruBajang #AngelaTanoesoedibjo #JoniSupriyanto #AhmadRofiq #PartaiPerindo #UntukIndonesiaSejahtera
2	Gaspol datang ke Pesta Rakyat Ganjar Pranowo yang berlangsung di Cilacap !! Berlangsung pada Jumat, 17 November 2023 #SahabatGanjar #GanjarPranowo #Ganjar #GenerasiGotongRoyong #GanjarPresiden #GanjarUntukSemua #GanjarMahfud2024 https://t.co/ocXNYqsMu0
3	Ratusan Pekerja Migran Indonesia (PMI) di kota Riyadh Saudi Arabia deklarasi dukung kepada calon Presiden @ganjarpranowo dan calon Wakil Presiden @mohmahfudmd pada Pemilu tahun 2024. @PDI_Perjuangan https://t.co/j3E8NMIWUP @GunRomli

Gambar 3.3 – Data Asli Atribut full_text

No	Full_text
1	repost dari bapak
2	gaspol datang ke pesta rakyat ganjar pranowo yang berlangsung di cilacap berlangsung pada jumat november
3	ratusan pekerja migran indonesia pmi di kota riyadh saudi arabia deklarasi dukung kepada calon presiden dan calon wakil presiden pada pemilu tahun

Gambar 3.4 – Data setelah preprocessing

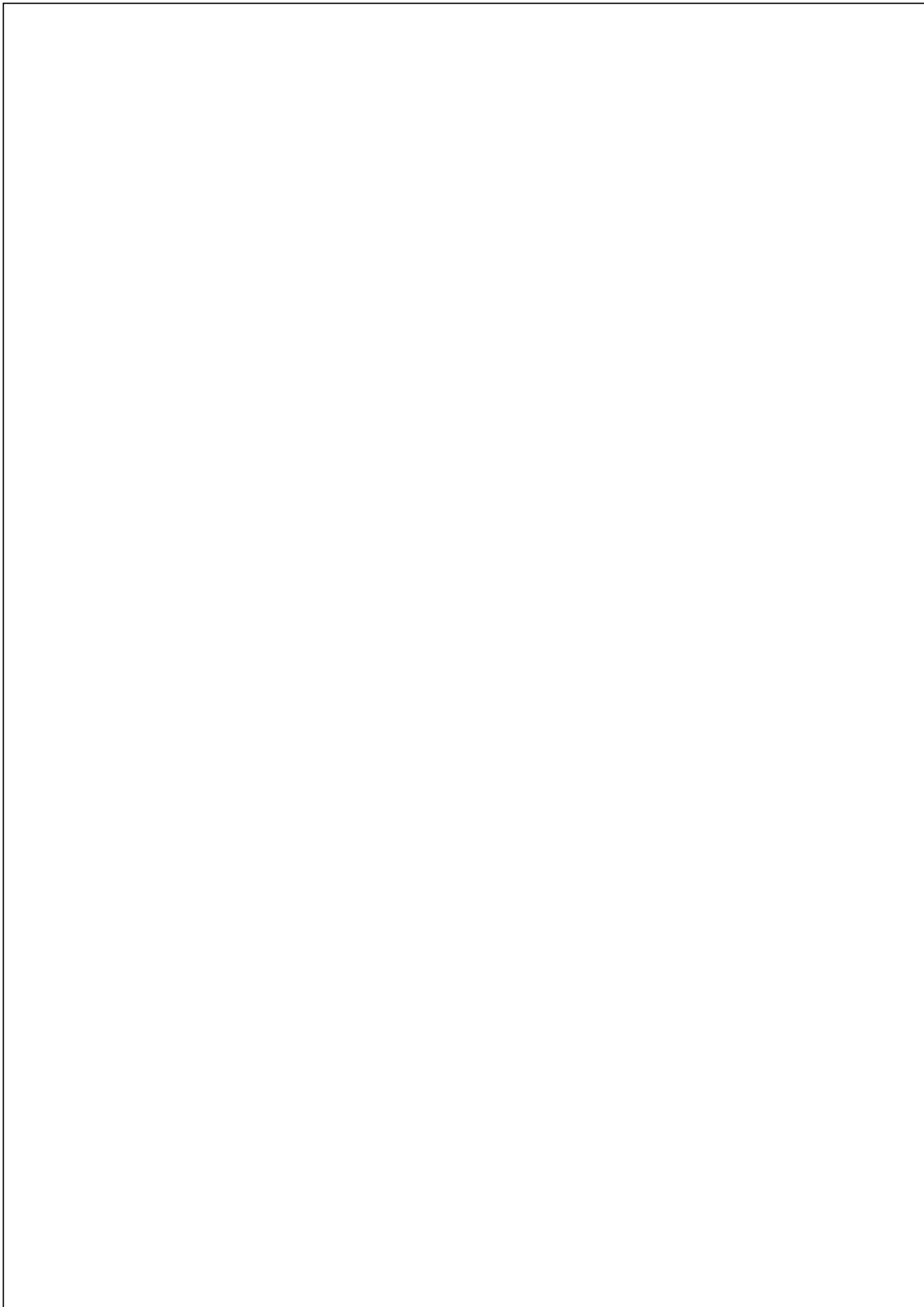
Data Labeling

Perhitungan Manual dengan Algoritma Naïve Bayes

DAFTAR PUSTAKA

- Al-Areef, M. H., & Saputra S, K. (2023). Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 22(2), 270. <https://doi.org/10.53513/jis.v22i2.8680>
- Asmara, R., Ardiansyah, M. F., & Anshori, M. (2020). Analisa Sentiment Masyarakat terhadap Pemilu 2019 berdasarkan Opini di Twitter menggunakan Metode Naive Bayes Classifier. *INOVTEK Polbeng - Seri Informatika*, 5(2), 193. <https://doi.org/10.35314/isi.v5i2.1095>
- Buntoro, G., Arifin, R., Syaifuddiin, G., Selamat, A., Krejcar, O., & Hamido, F. (2021). THE IMPLEMENTATION OF THE MACHINE LEARNING ALGORITHM FOR THE SENTIMENT ANALYSIS OF INDONESIA'S 2019 PRESIDENTIAL ELECTION. *IIUM Engineering Journal*, 22(1), 78–92. <https://doi.org/10.31436/iiumej.v22i1.1532>
- Delvika, B., Apriana, Abror, N., & Gurning, U. R. (2023). Perbandingan Algoritma NBC dan C4.5 Dalam Analisa Sentimen Pemilihan Presiden 2024 Pada Twitter. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 1(1), 41–48.
- Fathir, Hariyadi, M. A., & Miftachul A, Y. (2023). ANALISIS SENTIMEN ARTIKEL BERITA PEMILU BERBASIS METODE KLASIFIKASI. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 4(2), 485–493. <https://doi.org/10.35870/jimik.v4i2.220>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Firdlous, D. A., Andrian, R., & Widodo, S. (2023). Sentiment Analysis Public Twitter on 2024 Election using the Long Short Term Memory Model. *SISTEMASI*, 12(1), 52. <https://doi.org/10.32520/stmsi.v12i1.2145>
- Helmi Satria. (2023, October 18). *Cara crawl data twitter lebih dari 1500 tweets - 18 Oct*. Youtube. <https://youtu.be/1uOm95nMcuk>
- Hidayatullah, A. F., Cahyaningtyas, S., & Hakim, A. M. (2021). Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset. *IOP Conference Series: Materials Science and Engineering*, 1077(1), 012001. <https://doi.org/10.1088/1757-899X/1077/1/012001>

- Humam, C., & Laksito, A. D. (2023). Implementasi Aplikasi Sentimen Pada Data Twitter Jelang Pemilu 2024. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(2), 105–112. <https://doi.org/10.30591/jpit.v8i2.5051>
- Juanita, S. (2020). Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3), 552. <https://doi.org/10.30865/mib.v4i3.2140>
- Muzaki, A., & Witanti, A. (2021). SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER. *Jurnal Teknik Informatika (Jutif)*, 2(2), 101–107. <https://doi.org/10.20884/1.jutif.2021.2.2.51>
- Samsir, Irmayani, D., Edi, F., Harahap, J. M., Jupriaman, Rangkuti, R. K., Ulya, B., & Watrianthos, R. (2021). Naives Bayes Algorithm for Twitter Sentiment Analysis. *Journal of Physics: Conference Series*, 1933(1), 012019. <https://doi.org/10.1088/1742-6596/1933/1/012019>
- Silalahi, W., & Hartanto, A. (2023). Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024. *JRST (Jurnal Riset Sains Dan Teknologi)*, 7(2), 245. <https://doi.org/10.30595/jrst.v7i2.18133>
- Srinivas, A. C. M. V., Satyanarayana, Ch., Divakar, Ch., & Sirisha, K. P. (2021). Sentiment Analysis using Neural Network and LSTM. *IOP Conference Series: Materials Science and Engineering*, 1074(1), 012007. <https://doi.org/10.1088/1757-899X/1074/1/012007>
- Syafii, M. I. (2023). SENTIMEN ANALISIS PADA MEDIA SOSIAL MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER (NBC). *Jurnal Teknologi Pintar*, 3(2).
- Vindua, R., & Zailani, A. U. (2023). Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python. *JURIKOM (Jurnal Riset Komputer)*, 10(2), 479. <https://doi.org/10.30865/jurikom.v10i2.5945>
- Wenando, F. A., Hayami, R., & Anggrawan, A. J. (2020). ANALISIS SENTIMEN PADA PEMERINTAHAN TERPILIH PADA PILPRES 2019 DITWITTER MENGGUNAKAN ALGORITME NAÏVEBAYES. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 7(1), 101–106. <https://doi.org/10.33330/jurteksi.v7i1.851>
- Zain, H. H., Awangga, R. M., & Rahayu, W. I. (2023). Perbandingan Model Svm, Knn Dan Naïve Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024. *JIMPS: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 8(3).



ORIGINALITY REPORT

26%

SIMILARITY INDEX

24%

INTERNET SOURCES

9%

PUBLICATIONS

13%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Southville International School and Colleges Student Paper	2%
2	rp.sith.itb.ac.id Internet Source	1%
3	repository.usbypkp.ac.id Internet Source	1%
4	sistemasi.ftik.unisi.ac.id Internet Source	1%
5	etheses.uin-malang.ac.id Internet Source	1%
6	Submitted to vitka Student Paper	1%
7	digilib.unila.ac.id Internet Source	1%
8	Submitted to UIN Sunan Gunung Djati Bandung Student Paper	1%
9	Submitted to STT PLN	

10

ojs.trigunadharma.ac.id

Internet Source

1 %

11

Submitted to Chester College of Higher Education

Student Paper

1 %

12

Saptari Wijaya Mulia, Sujiharno Sujiharno, Arief Wibowo. "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI FREKUENSI TUNAI PADA MESIN ATM DI MASA TRANSISI PEMBATAHAN SOSIAL BERSKALA BESAR (PSBB) PANDEMI COVID-19", SINTECH (Science and Information Technology) Journal, 2021

Publication

1 %

13

www.ekrut.com

Internet Source

1 %

14

gitlab.gti-ia.upv.es

Internet Source

1 %

15

ejournal.poltektegal.ac.id

Internet Source

1 %

16

journal.lembagakita.org

Internet Source

1 %

17

e-jurnal.lppmunsera.org

Internet Source

1 %

18	www.wartaekonomi.co.id Internet Source	1 %
19	Submitted to University of Southampton Student Paper	1 %
20	www.samuelpasaribu.com Internet Source	1 %
21	Submitted to Universitas Raharja Student Paper	<1 %
22	Submitted to Universitas Negeri Jakarta Student Paper	<1 %
23	anzdoc.com Internet Source	<1 %
24	download.garuda.ristekdikti.go.id Internet Source	<1 %
25	repo.undiksha.ac.id Internet Source	<1 %
26	text-id.123dok.com Internet Source	<1 %
27	ecampus.pelitabangsa.ac.id Internet Source	<1 %
28	repositori.usu.ac.id Internet Source	<1 %
29	repository.uph.edu Internet Source	<1 %

30	adoc.pub Internet Source	<1 %
31	blog.nolimit.id Internet Source	<1 %
32	jutif.if.unsoed.ac.id Internet Source	<1 %
33	riniinggriani.blogspot.com Internet Source	<1 %
34	eproceeding.itenas.ac.id Internet Source	<1 %
35	lib.unnes.ac.id Internet Source	<1 %
36	pt.scribd.com Internet Source	<1 %
37	Rengga Asmara, Muhammad Febrian Ardiansyah, Muhammad Anshori. "Analisa Sentiment Masyarakat terhadap Pemilu 2019 berdasarkan Opini di Twitter menggunakan Metode Naive Bayes Classifier", INOVTEK Polbeng - Seri Informatika, 2020 Publication	<1 %
38	es.scribd.com Internet Source	<1 %
39	garuda.kemdikbud.go.id Internet Source	<1 %

40

repository.iainpare.ac.id

Internet Source

<1 %

41

repository.ub.ac.id

Internet Source

<1 %

42

repository.unpkediri.ac.id

Internet Source

<1 %

43

ejurnalunsam.id

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On

Alisya

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23
