

**ANALISIS SENTIMEN PERANG HAMAS ISRAEL TERHADAP
KEMERDEKAAN PALESTINA BERDASARKAN RESPONS
PENGGUNA MEDIA SOSIAL X DI INDONESIA
MENGUNAKAN METODE *NAIVE BAYES CLASSIFIER***

SKRIPSI

Diajukan untuk Memenuhi Tugas Akhir dan Melengkapi Syarat
Guna Memperoleh Gelar Sarjana Strata Satu (S-1) dalam Teknologi
Informasi



Diajukan oleh :

**NOVA RIO REDONDO
NIM : 2008096028**

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO SEMARANG
TAHUN 2024**

PERNYATAAN KEASLIAN

Yang bertandatangan dibawah ini:

Nama : Nova Rio Redondo
NIM : 2008096028
Prodi : Teknologi Informasi

Bahwa skripsi yang berjudul:

**Analisis Sentimen Perang Hamas Israel Terhadap
Kemedekaan Palestina Berdasarkan Respons
Pengguna Media Sosial X Di Indonesia Menggunakan
Metode Naive Bayes Classifier**

Secara keseluruhan adalah hasil
penelitian/karya saya sendiri,kecuali bagian
tertentu yang dirujuk sumbernya.

Semarang, 9 Agustus 2024

Pembuat Pernyataan,



Nova Rio Redondo

NIM : 2008096028



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka Ngaliyan Semarang
Telp.024-7601295 Fax.7615387

PENGESAHAN

Naskah skripsi berikut ini:

Judul : Analisis Sentimen Perang Hamas Israel Terhadap
Kemerdekaan Palestina Berdasarkan Respons
Pengguna Media Sosial X Di Indonesia
Menggunakan Metode *Naive Bayes Classifier*

Penulis : Nova Rio Redondo
NIM : 2008096028
Jurusan : Teknologi Informasi

Telah diujikan dalam sidang tugas akhir oleh Dewan Penguji
Fakultas Sains dan Teknologi UIN Walisongo dan dapat
diterima sebagai salah satu syarat memperoleh gelar sarjana
dalam Teknologi Informasi.

Semarang, 27 September 2024

DEWAN PENGUJI

Penguji I

Hery Mustofa, M.Kom.
NIP. 198703172019031007

Penguji II

Adzhal Arwani Mahfudh, M.Kom.
NIP. 199107032019031006

Penguji III

Dr. Wenty Dwi Yuniarti, S.Pd., M.Kom.
NIP. 197706222006042005

Penguji IV

Mokhammad Ikhlil Mustofa, M.Kom.
NIP. 198808072019031010

Pembimbing I

Dr. Khothibul Umam, S.T., M.Kom.
NIP. 197908272011011007

Pembimbing II

Adzhal Arwani Mahfudh, M.Kom.
NIP. 199107032019031006

NOTA DINAS

Semarang, 16 Agustus 2024

Yth. Ketua Program Studi Teknologi Informasi

Fakultas Sains dan Teknologi

UIN Walisongo Semarang

Assalamua'alaikum Warahmatullahi Wabarakatuh.

Dengan ini memberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul : ANALISIS SENTIMEN PERANG HAMAS ISRAEL TERHADAP KEMERDEKAAN PALESTINA BERDASARKAN RESPONS PENGGUNA MEDIA SOSIAL X DI INDONESIA MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

Nama : Nova Rio Redondo

NIM : 2008096028

Jurusan : Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Program Studi Teknologi Informasi dan Fakultas Sains dan Teknologi UIN Walisongo Semarang untuk diujikan dalam Sidang Munaqosyah.

Wassalamua'alaikum Warahmatullahi Wabarakatuh.

Pembimbing I



Dr. Khothibul Umam S.T., M. Kom.
NIP. 19790827 201101 1 007

NOTA DINAS

Semarang, 16 Agustus 2024

Yth. Ketua Program Studi Teknologi Informasi

Fakultas Sains dan Teknologi

UIN Walisongo Semarang

Assalamua'alaikum Warahmatullahi Wabarakatuh.

Dengan ini memberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul : ANALISIS SENTIMEN PERANG HAMAS
ISRAEL TERHADAP KEMERDEKAAN
PALESTINA BERDASARKAN RESPON
PENGGUNA MEDIA SOSIAL X DI INDONESIA
MENGUNAKAN METODE *NAIVE BAYES
CLASSIFIER*

Nama : Nova Rio Redondo

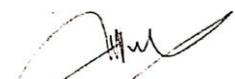
NIM : 2008096028

Jurusan : Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Program Studi Teknologi Informasi dan Fakultas Sains dan Teknologi UIN Walisongo Semarang untuk diujikan dalam Sidang Munaqosyah.

Wassalamua'alaikum Warahmatullahi Wabarakatuh.

Pembimbing II



Adzhal Arwani M., M. Kom.
NIP. 19910703 201903 1 006

LEMBAR PERSEMBAHAN

Puji dan Syukur Penulis ucapkan kepada Allah SWT, laporan tugas akhir ini dapat terselesaikan dengan baik. Karya tulis ini penulis persembahkan untuk:

1. Kedua orang tua tercinta yang selalu memberikan doa dan semangat.
2. Saudara dan saudari penulis yang selalu membantu dan mendoakan, serta selalu memberikan dukungan moral dan material.
3. Segenap dosen Jurusan Teknologi Informasi yang selalu membimbing dan mengarahkan penulis.
4. Serta teman-teman Teknologi Informasi 2020.

Almamater Universitas Islam Negeri Walisongo Semarang.

MOTTO

Barangsiapa bertakwa kepada Allah, niscaya Dia akan
memberikan jalan keluar baginya.

(QS. At-Talaq: 2)

Ilmu tanpa amal adalah kegilaan, dan amal tanpa ilmu adalah
kesia-siaan.

(Umar bin Khattab)

Life is filled with suffering, but because of that, it's also
beautiful.

(Kyojuro Rengoku)

ABSTRAK

Perang Hamas dengan Israel memiliki dampak yang luar biasa pada kesetabilan keamanan dunia. Perang Hamas dengan Israel menimbulkan kontroversi, hal tersebut memicu pro dan kontra di masyarakat. Era digital sekarang, banyak orang menyampaikan pendapat mereka melalui media sosial, termasuk di media sosial X. Penelitian ini menggunakan analisis sentimen untuk mengklasifikasikan reaksi pengguna media sosial X terhadap kejadian tersebut ke dalam dua kelas sentimen, yaitu positif dan negatif.

Data komentar *tweet* yang didapat menggunakan kata kunci "palestina merdeka hamas israel". Proses *crawling data* dilakukan menggunakan *tweet harvest*. Data awal diproses melalui *preprocessing* kemudian dibobotkan menggunakan *TF-IDF*, dan diklasifikasikan dengan metode naive bayes classifier. Hasilnya menunjukkan bahwa sentimen positif mendominasi dengan 74,5%, dan sentimen negatif sebesar 24,5%. Dari tiga skenario *split validation*, validasi data terbaik menggunakan pembagian 80:20. Dari validasi tersebut didapatkan nilai performa dari model *naive bayes classifier* yang baik dengan tingkat akurasi sebesar 81,69%, *precision* 83,44%, *recall* 81,69%, dan *f1-score* 78,74%.

Kata Kunci: analisis sentimen, perang hamas israel, media sosial X, *naive bayes classifier*.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada Nabi Muhammad SAW yang telah menjadi pelopor peradaban Islam yang Kaffah serta salam untuk keluarga dan para sahabat-sahabat Rasulullah SAW.

Skripsi yang berjudul "Analisis Sentimen Perang Hamas Israel Terhadap Kemerdekaan Palestina Berdasarkan Respons Pengguna Media Sosial X Di Indonesia Menggunakan Metode *Naive Bayes Classifier*" ini dapat diselesaikan dan disusun guna memenuhi salah satu syarat dalam memperoleh gelar sarjana dalam bidang komputer Fakultas Sains dan Teknologi Universitas Islam Negeri Walisongo Semarang.

Proses penyusunan skripsi ini tidak lepas dari doa, bantuan, bimbingan, motivasi dan peran dari banyak pihak. Sehingga penulis mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. Nizar, M.Ag., selaku rektor UIN Walisongo Semarang.
2. Bapak Prof. Dr. H. Musahadi, M.Ag., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Walisongo Semarang.

3. Bapak Dr. Khothibul Umam S.T., M.Kom., selaku Kepala Program Studi Teknologi Informasi.
4. Bapak Dr. Khothibul Umam S.T., M.Kom., dan Adzhal Arwani Mahfudh M.Kom., selaku Dosen Pembimbing yang telah bersedia meluangkan waktu, tenaga dan pemikirannya dalam membantu penulis menyelesaikan skripsi.
5. Bapak/Ibu Dosen Jurusan Teknologi Informasi Fakultas Sains dan Teknologi UIN Walisongo Semarang yang telah sabar dan ikhlas memberi ilmu yang disampaikan kepada penulis.
6. Kedua orang tua dan saudara yang telah memberikan do'a secara moral dan material, dukungan, motivasi dan nasihatnya.
7. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi hingga selesainya skripsi ini.

Semoga kebaikan semuanya menjadi amal ibadah yang diterima dan mendapat pahala yang berlimpah dari Allah SWT. Aamiin. Atas segala kekurangan dan kelemahan dalam skripsi ini penulis mengharapkan saran dan kritik yang membangun. Semoga karya tulis yang sederhana ini dapat menjadi bacaan yang bermanfaat dan dapat dikembangkan bagi peneliti-peneliti selanjutnya.

DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN.....	ii
LEMBAR PENGESAHAN	iii
NOTA PEMBIMBING.....	iv
NOTA PEMBIMBING.....	v
LEMBAR PERSEMBAHAN	vi
MOTTO	vii
ABSTRAK.....	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL.....	xvi
BAB I.....	1
A. Latar Belakang.....	1
B. Rumusan Masalah	5
C. Tujuan Penelitian	6
D. Batasan Masalah	6
E. Manfaat Penelitian	7
BAB II	9
A. Analisis Sentimen.....	9
B. Media Sosial X (Twitter)	10

C. <i>Naive Bayes Classifier</i>	13
D. <i>Text Mining</i>	14
E. Python	15
F. <i>Crawling Data</i>	16
G. <i>Text Preprocessing</i>	17
H. <i>TF-IDF</i>	19
I. <i>Split Validation Data</i>	21
J. Tahap Evaluasi	21
K. Kajian Penelitian Terkait	24
BAB III	30
A. Metode Pengumpulan Data	30
1. Studi Pustaka	30
2. Studi Lapangan	31
B. Perangkat Penelitian	31
1. Kebutuhan Perangkat Keras:	32
2. Kebutuhan Perangkat Lunak.....	32
C. Alur Pengerjaan Penelitian	33
D. Uraian Metodologi.....	34
1. Pengambilan Data X (Twitter)	34
2. <i>Preprocessing</i>	36
3. Ekstraksi Fitur.....	41
4. Klasifikasi <i>Naive Bayes</i>	42
5. Uji Model.....	43
6. Evaluasi Model	43
7. Visualisasi	44

BAB IV	46
A. Pengambilan Data Media Sosial X.....	46
B. <i>Preprocessing</i>.....	51
C. Ekstrasi Fitur.....	64
D. Metode Naive Bayes Classifier.....	71
E. Uji Model	73
F. Evaluasi Model	76
G. Visualisasi.....	86
BAB V.....	92
A. Kesimpulan	92
B. Saran.....	93
DAFTAR PUSTAKA.....	95
DAFTAR LAMPIRAN.....	100
RIWAYAT HIDUP	120

DAFTAR GAMBAR

Gambar 2. 1 Tampilan Beranda Media Sosial X.....	12
Gambar 3. 1 <i>Flowchart</i> Alur Penelitian.....	35
Gambar 3. 2 Hasil <i>Crawling Data Tweet</i>	37
Gambar 3. 3 Contoh Data Yang Sama.....	40
Gambar 4. 1 Tampilan Google Colab 2024.....	47
Gambar 4. 2 <i>Source Code Crawling Data</i>	48
Gambar 4. 3 Proses <i>Crawling Data</i>	49
Gambar 4. 4 Hasil <i>Crawling Data</i>	49
Gambar 4. 5 Data Setelah Diberi Sentimen	50
Gambar 4. 6 <i>Source Code Case Folding</i>	52
Gambar 4. 7 Hasil <i>Case Folding</i>	52
Gambar 4. 8 <i>Source Code Cleansing</i>	53
Gambar 4. 9 Hasil <i>Cleansing</i>	54
Gambar 4. 10 <i>Source Code Remove Duplicate</i>	54
Gambar 4. 11 Hasil <i>Remove Duplicate</i>	55
Gambar 4. 12 <i>Source Code Normalization</i>	56
Gambar 4. 13 Hasil <i>Normalization</i>	57
Gambar 4. 14 <i>Source Code</i> Memasukkan kbba	57
Gambar 4. 15 Hasil <i>Normalization</i> Bagian Dua.....	58
Gambar 4. 16 <i>Instal nlp-id</i>	58
Gambar 4. 17 <i>Source Code Stopword Removal</i>	59
Gambar 4. 18 <i>Output Stopword Removal</i>	59
Gambar 4. 19 <i>Instal Library Sastrawi</i>	60
Gambar 4. 20 Impor Kelas <i>Stemmer Factory</i>	60
Gambar 4. 21 <i>Source Code Stemming</i>	60
Gambar 4. 22 Hasil <i>Stemming</i>	61
Gambar 4. 23 <i>Source Code Tokenization</i>	62
Gambar 4. 24 Setelah Proses <i>Preprocessing</i>	63
Gambar 4. 25 Jumlah Sentimen.....	64
Gambar 4. 26 <i>Split Validation Data</i>	65

Gambar 4. 27 Pembobotan Kata dengan TF-IDF	66
Gambar 4. 28 Hasil Pembobotan Kata TF-IDF	67
Gambar 4. 29 <i>Source Code Import sklearn</i>	72
Gambar 4. 30 Proses <i>Naive Bayes Classifier</i>	72
Gambar 4. 31 Hasil Akurasi Ketiga Skenario	73
Gambar 4. 32 <i>Source Code</i> Memunculkan <i>Confusion Matrix</i> ...	74
Gambar 4. 33 <i>Confusion Matrix</i> Ketiga Skenario	74
Gambar 4. 34 <i>Source Code</i> Perhitungan Naive Bayes.....	83
Gambar 4. 35 Hasil Perhitungan Skenario 1	84
Gambar 4. 36 Hasil Perhitungan Skenario 2.....	84
Gambar 4. 37 Hasil Perhitungan Skenario 3.....	85
Gambar 4. 38 Presentase Distribusi Sentimen	87
Gambar 4. 39 <i>Wordcloud</i> Hasil <i>Crawling Data</i>	88
Gambar 4. 40 <i>Wordcloud</i> Setelah <i>Preprocessing</i>	89
Gambar 4. 41 <i>Wordcloud</i> Sentimen Positif	90
Gambar 4. 42 <i>Wordcloud</i> Sentimen Negatif.....	90

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i> Dua Kelas.....	22
Tabel 2. 2 Kajian Penelitian Terkait	24
Tabel 3. 1 Kebutuhan Perangkat Keras	32
Tabel 3. 2 Kebutuhan Perangkat Lunak	32
Tabel 3. 3 Contoh Pelabelan Sentimen	36
Tabel 3. 4 Penerapan <i>Case Folding</i>	37
Tabel 3. 5 Penerapan Proses <i>Cleansing</i>	38
Tabel 3. 6 penerapan Proses <i>Normalization</i>	39
Tabel 3. 7 Penerapan Proses <i>Stopword Removal</i>	39
Tabel 3. 8 Penerapan Proses <i>Stemming</i>	40
Tabel 3. 9 Penerapan Proses <i>Tokenization</i>	41
Tabel 4. 1 Perhitungan TF	68
Tabel 4. 2 Perhitungan IDF.....	68
Tabel 4. 3 Perkalian TF dan IDF.....	69
Tabel 4. 4 Hasil TF-IDF	70
Tabel 4. 5 Perbandingan Data Latih dan Uji	71
Tabel 4. 6 <i>Confusion Matrix</i>	73
Tabel 4. 7 <i>Confusion Matrix</i> Data Uji.....	74
Tabel 4. 8 Hasil Akurasi Ketiga Skenario	76
Tabel 4. 9 <i>Confusion Matrix</i> Skenario 1	77
Tabel 4. 10 <i>Confusion Matrix</i> Skenario 2	78
Tabel 4. 11 <i>Confusion Matrix</i> Skenario 3	80
Tabel 4. 12 Hasil Performa Kelas.....	82
Tabel 4. 13 Hasil Performa <i>Naive Bayes Classifier</i>	85

BAB I

PENDAHULUAN

A. Latar Belakang

Perang antara pejuang Hamas melawan tentara Israel memang sudah sangat sering terjadi, serangan terbaru adalah pada tanggal 7 Oktober 2023. Hamas yang ingin merebut kembali tanah miliknya melalui jalur militer selalu mendapat perlawanan yang serius dari tentara Israel. Perang antara Hamas dan Israel selalu menyita perhatian dunia. Perang merupakan permusuhan antara negara yang terjadi di dalam wilayah suatu negara yang dilakukan dengan menggunakan sarana kekuatan bersenjata. Dalam hukum internasional, suatu negara yang berperang dianggap menuruti aturan hukum internasional manakala dua negara atau lebih menyatakan secara resmi bahwa di antara mereka terjadi kondisi permusuhan (Plano, 2011).

Hamas atau yang biasa disebut *Harakat Al Muqawwamat Al Islamiyah* adalah gerakan perlawanan Islam dari kelompok militer yang mayoritas bermazhab sunni dan berkembang menjadi partai politik di Palestina (Idatul & Fadhil, 2019). Perang Hamas dengan Israel memiliki dampak buruk terutama di wilayah Timur Tengah.

Perang yang terjadi antara Hamas melawan Israel ini menimbulkan banyak sekali pendapat pro dan kontra di

Indonesia bahkan sampai dunia. Hamas yang menduduki wilayah Jalur Gaza memilih berjuang memerdekakan Palestina dengan angkat senjata melawan Israel, sedangkan Fatah atau yang dikenal dengan nama Gerakan Pembebasan Palestina yang menguasai daerah Tepi Barat sungai Yordan (West Bank) lebih memilih jalur diplomasi untuk memerdekakan Palestina. Impian tentang sebuah kemerdekaan warga Palestina sampai sekarang ini masih belum terwujud.

Karena terdapat banyak sekali pro dan kontra terhadap perang tersebut, akibat hal tersebut munculah beberapa sentimen di berbagai media sosial termasuk X (Twitter), untuk saat ini, era Internet telah mengubah cara orang mengekspresikan pandangan, opini serta pendapat mereka (Sonawanse, 2016). Bahkan tokoh-tokoh terkenal juga ikut menyuarakan kejadian tersebut, terlebih menyoroti hal kemanusiaan yang terjadi dalam perang tersebut, bahkan diberbagai platform media sosial, salah satunya adalah X (Twitter) tagar *#freepalestine* sempat menjadi yang teratas beberapa hari setelah terjadinya perang tersebut.

Media sosial X sendiri merupakan salah satu media yang penggunaanya paling ramai menyuarakan suatu kejadian atau isu yang beredar terkait perang yang terjadi di Jalur Gaza dan hal yang berkaitan dengan kemanusiaan yang terjadi di Jalur Gaza. Data yang biasa diambil dari media sosial X adalah *tweet* dari penggunaanya. Pengguna media sosial X biasanya membagikan

sebuah postingan atau membagikan sebuah pesan singkat terkait aktivitas sehari-hari antar pengguna ke pengguna lainnya yang disebut dengan *tweet* (H. Dhery et al., 2023).

Banyak sekali media yang menyiarkan konflik yang berada di Jalur Gaza setiap terjadi konflik, namun kali ini konflik antara Hamas dan Israel merupakan yang terlama dan terparah, bahkan masyarakat dunia juga ikut menyoroti dengan pandangannya masing-masing. Media saat ini memang sangat menyoroti hal tersebut. Karena dengan media, seseorang dapat secara bebas mengutarakan ide, pendapat, opini dalam bentuk teks atau kalimat (Fajar et al., 2018).

Fenomena meluasnya beragam pandangan dan pendapat yang disampaikan melalui platform media sosial seperti X memungkinkan dari kita untuk mengkategorikan sentimen yang muncul, sehingga dapat dengan mudah menganalisis apakah mayoritas masyarakat berpendapat positif atau negatif terkait dengan peperangan yang terjadi antara Hamas dengan Israel pada saat ini. Dikarenakan data dari X (Twitter) tidak terstruktur dan banyak memuat noise sehingga dibutuhkan proses *text mining* agar data menjadi lebih terstruktur. *Text mining* memiliki peran penting dalam bidang *data mining*.

Salah satu metode yang bisa digunakan untuk menganalisis sentimen adalah dengan menggunakan metode algoritma *naive bayes classifier*. *Naive bayes classifier*

merupakan algoritma klasifikasi yang sederhana, efisiensi, komputasi tinggi, dan memiliki akurasi klasifikasi yang baik terutama untuk data berdimensi tinggi (Mahfudh & Mustofa, 2019).

Analisis sentimen terhadap perang Hamas dengan Israel yang berpengaruh terhadap kemerdekaan Palestina sangat penting untuk diperhatikan, terlebih penduduk Palestina yang berada di Jalur Gaza saat ini sedang mengalami kesusahan yang sangat luar biasa akibat gempuran tentara Israel. Indonesia sebagai negara dengan jumlah umat muslim terbanyak di dunia tidak boleh tinggal diam, Indonesia juga harus tegas mengupayakan perdamaian dan membantu saudara sesama muslim yang ada di Palestina.

Sesuai dalam Al-Qur'an surah Al-Anfal ayat 73. Allah SWT, berfirman:

كَبِيرٌ وَفَسَادُ الْأَرْضِ فِي فِتْنَةٍ تَكُنْ تَفْعَلُوهُ إِلَّا بَعْضٌ أَوْلِيَاءَ بَعْضُهُمْ كَفَرُوا وَالَّذِينَ

Artinya: *“Adapun orang-orang yang kafir, sebagian mereka menjadi pelindung bagi sebagian yang lain. Jika kamu para muslimin tidak melaksanakan apa yang telah diperintahkan Allah itu (untuk saling melindungi), niscaya akan terjadi kekacauan di muka bumi dan kerusakan yang besar.”* (QS Al-Anfal 8: Ayat 73).

Ayat tersebut menerangkan bahwa sebagian orang-orang kafir saling melindungi umat kafir lainnya. Contohnya seperti Amerika Serikat dan Israel. Kita sebagai umat muslim juga harus saling melindungi satu sama lain. Sebab, jika umat muslim tidak saling melindungi, niscaya bumi akan mengalami

kekacauan yang lebih besar, akibat ulah orang-orang kafir.

Selain itu sebagai umat muslim kita juga harus tabayun. dengan mengumpulkan berbagai opini dari masyarakat kemudian menganalisisnya sehingga mendapatkan hasil yang berguna bagi masyarakat dan mengetahui kebenarannya.

Oleh karena itu penulis melakukan penelitian analisis sentimen terhadap perang yang terjadi antara Hamas dan Israel terhadap kemerdekaan Palestina dengan dua pelabelan sentimen, positif dan negatif menggunakan algoritma naive bayes classifier dengan data- data diambil dari media sosial X.

B. Rumusan Masalah

Berdasarkan uraian dari latar belakang, maka rumusan masalah yang menjadi fokus pada penelitian ini adalah sebagai berikut:

1. Bagaimana implementasi algoritma *naive bayes classifier* dalam membantu analisis sentimen perang antara Hamas dan Israel terhadap kemerdekaan Palestina pada respons pengguna media sosial X (Twitter)?
2. Bagaimana performa yang diberikan oleh algoritma *naive bayes classifier* pada analisis sentimen perang antara Hamas dan Israel terhadap kemerdekaan Palestina pada respons pengguna media sosial X (Twitter)?

C. Tujuan Penelitian

Dilihat dari rumusan masalah yang telah diberikan, penelitian ini dilakukan dengan tujuan:

1. Menerapkan algoritma *naive bayes classifier* untuk menganalisis tanggapan positif dan negatif pada respons pengguna media sosial X yang berkaitan dengan perang Hamas dengan Israel terhadap kemerdekaan Palestina.
2. Mengetahui performa dari hasil perhitungan algoritma *naive bayes classifier* untuk klasifikasi tanggapan positif dan negatif perang Hamas dengan Israel terhadap kemerdekaan Palestina pada respons pengguna media sosial X.

D. Batasan Masalah

Batasan penelitian merujuk pada pembatasan atau ruang lingkup yang ditentukan oleh peneliti. Batasan dari pada penelitian ini adalah:

1. Analisis sentimen dilakukan dengan menggunakan metode *naive bayes classifier*.
2. Data diambil dari media sosial X (Twitter) berbahasa Indonesia.
3. Pengelompokan data sentimen terbagi menjadi sentimen negatif dan positif.

4. Kata kunci (*keyword*) pencarian data pada twitter adalah “palestina merdeka hamas israel”.
5. Jumlah data yang digunakan dari media sosial X berjumlah 1574 data.
6. Pengambilan data *tweet* dilakukan pada 1 Januari 2024.
7. Data yang didapat dari X memiliki rentang waktu antara 8 Oktober 2023 – 31 Desember 2023.
8. Penelitian menggunakan bahasa pemrograman python. Menggunakan *software google colab*.

E. Manfaat Penelitian

Manfaat dari hasil penelitian ini dibagi menjadi dua, yaitu berdasarkan manfaat secara teoritis dan manfaat secara praktis.

1. Manfaat Teoritis:

- a. Penelitian ini dapat berkontribusi pada literatur yang membahas tentang analisis sentimen media sosial X yang berkaitan dengan perang Hamas-Israel dengan menggunakan metode *naive bayes classifier*.

2. Manfaat Praktis

- a. Dapat mengetahui jumlah respons atau tanggapan tentang perang Hamas dan Israel terhadap kemerdekaan Palestina yang lebih cenderung ke hal yang positif atau negatif.
- b. Dengan memanfaatkan analisis sentimen, pemerintah dapat menjadi lebih responsif terhadap isu-isu internasional yang penting bagi masyarakat Indonesia, sehingga meningkatkan kepercayaan dan dukungan publik terhadap kebijakan luar negeri mereka.
- c. Analisis sentimen dari respons pengguna media sosial di Indonesia dapat membantu pemerintah untuk memahami pandangan dan sikap masyarakat terkait konflik antara Hamas dan Israel serta isu kemerdekaan Palestina.

BAB II

LANDASAN PUSTAKA

A. Analisis Sentimen

Analisis sentimen disebut juga sebagai *opinion mining* yang berguna dalam pengelolaan bahasa alami, komputasi linguistik, dan *text mining*. Tujuan dari analisis sentimen adalah untuk menganalisis sentimen masyarakat seperti sikap, opini, emosi, dll. terhadap beberapa elemen seperti, produk, individu, topik, organisasi, dan layanan (Sonawanse, 2016).

Analisis sentimen adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas tersebut dapat mewakili individu (Syakuro, 2017). Analisis sentimen adalah salah satu cabang sebuah penelitian *text mining*, yang berkaitan dengan bidang yang lebih luas seperti pengolahan data kegiatan tertentu (A. Samad, 2013).

Pada dasarnya analisis sentimen terdiri dari beberapa langkah diantaranya yaitu *crawling data*, *pre-processing*, *feature selection*, *classification*, dan *evaluation*. Analisis sentimen dapat mengubah data tidak beraturan menjadi data yang tersusun rapi. Manfaat adanya analisis sentimen yaitu sebagai evaluasi dan ide pada berbagai bidang. Analisis sentimen dapat menganalisis suatu pernyataan, kejadian, dan komentar yang kontroversi. Hasil dari analisis sentimen juga

dapat menjadi sebuah gambaran bagi perusahaan, *public figure*, dan pemerintahan untuk menentukan langkah selanjutnya (Natasuwarna, 2020).

Analisis sentimen juga dapat digunakan untuk menganalisa pendapat bahkan sampai emosi terhadap suatu entitas. Analisis sentimen atau *opinion mining* adalah bidang studi yang menganalisa pendapat, pandangan, evaluasi, penilaian, sikap, dan emosi terhadap entitas dan aspek-aspeknya yang diekspresikan melalui teks (Zhang L., & Liu B., 2016).

B. Media Sosial X (Twitter)

Dalam zaman digital, penggunaan media sosial semakin meluas, tidak hanya di Indonesia tetapi juga di seluruh dunia. Media sosial merupakan platform daring yang memungkinkan interaksi efektif dan efisien baik secara individu maupun dalam kelompok, dengan cakupan yang tak terbatas. Pandangan lain mengatakan bahwa media sosial adalah media yang dapat meng-*support* interaksi sosial serta mengubah pandangan dari komunikasi secara pasif menjadi interaktif dengan transformasinya digitalisasi (Faiq, et al., 2022).

Media sosial adalah platform media yang memfokuskan pada eksistensi pengguna yang memfasilitasi mereka dalam beraktifitas maupun berkolaborasi. Karena itu media sosial dapat dilihat sebagai fasilitator *online* yang menguatkan

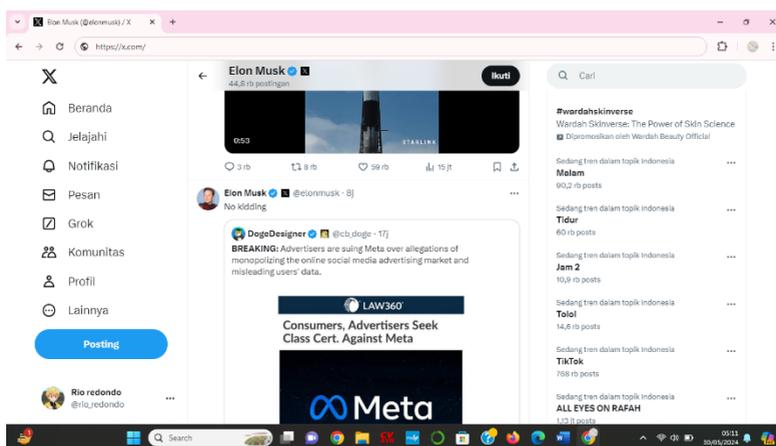
hubungan antar pengguna sekaligus sebuah ikatan sosial (Nasrullah., 2015).

Media sosial X atau lebih akrab disapa Twitter adalah platform media sosial yang memfasilitasi pertukaran ide dan opini. Pengguna media sosial X memiliki kemampuan untuk mengirim dan menerima pesan berupa *tweet*, yang dapat berupa teks, gambar, atau video. Salah satu perbedaan utama media sosial X dengan media sosial lainnya adalah batasan karakter untuk setiap pesan di Twitter, *tweet* di media sosial X dibatasi hanya sampai 280 karakter, sementara media sosial lain tidak memiliki batasan karakter serupa.

Twitter mempunyai kelebihan yaitu jangkauan yang luas, dapat menjangkau publik figur, media promosi lebih luas, banyak jaringan, dan lebih mudah diukur kemampuannya. Berikut fitur yang ada pada Twitter (I. Taufik dan S.A.Pamungkas, 2018):

1. *Tranding topic* adalah fitur yang menampilkan topik atau pembahasan teratas berupa hastag yang banyak dibicarakan pengguna.
2. *Hastag* adalah fitur yang dapat mengelompokkan *tweet* atau pesan.
3. *Retweet* adalah fitur untuk membagikan *tweet* dari pengguna lain.
4. *Following* adalah fitur untuk menghubungkan antar pengguna atau sering disebut teman.

Media sosial X memiliki tampilan menu utama yang sederhana dan mudah dipahami. Media sosial X sekarang ini juga dikenal dengan sebutan “X”. Twitter baru-baru ini mengalami perubahan tampilan dan logo, bahkan alamat URL (*Uniform Resource Locator*) yang tadinya *twitter.com* telah berubah menjadi *x.com*. Setelah Twitter secara resmi diakuisi oleh Elon Musk, Twitter mengalami semakin banyak perubahan didalamnya, termasuk juga fitur premium yang ada dalam media sosial X saat ini. Berikut merupakan tampilan beranda media sosial X terbaru.



Gambar 2. 1 Tampilan Beranda Media Sosial X

Media sosial X hanya sebagai salah satu media sosial yang memiliki banyak jumlah pengguna aktif di Indonesia. Pada awal Januari 2023 menurut situs terkemuka tentang media sosial yaitu *wearesocial* (*wearesocial.com*) tercatat sebanyak 167 juta pengguna atau sekitar 60,4% dari

populasi negara Indonesia. Diperkirakan jumlah pengguna media sosial X akan bertambah setiap tahunnya, mengikuti jumlah kenaikan pengguna media sosial di Indonesia. Bukan hanya X saja, melainkan jumlah seluruh pengguna media sosial di Indonesia akan bertambah setiap tahunnya.

C. *Naïve Bayes Classifier*

Naïve Bayes Classifier adalah teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (Yuniarti et al., 2020).

Naive bayes adalah salah satu metode klasifikasi yang sangat populer dan sering digunakan karena tingkat akurasi yang tinggi. Banyak penelitian telah dilakukan untuk mengkaji penggunaan algoritma ini dalam konteks klasifikasi. Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasian probabilitas. *Naive bayes classifier* tidak bisa mendeteksi gambar, tetapi hanya bisa mendeteksi numerik. Untuk menghitung probabilitas dalam metode ini menggunakan pendekatan teorema bayes (Prabowo & Wiguna, 2021).

Naive bayes classifier memiliki rumus umum seperti dibawah ini.

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)}$$

Keterangan:

X = Data dengan *class* yang belum diketahui

Y = Hipotesis data merupakan *class* spesifik

$P(Y|X)$ = Probabilitas hipotesis Y berdasar kondisi X

$P(Y)$ = Probabilitas hipotesis Y

$P(X|Y)$ = Probabilitas hipotesis X berdasar kondisi Y

$P(X)$ = Probabilitas hipotesis X

D. Text Mining

Text mining merupakan tahapan proses dari analisis dalam data yang berupa teks biasanya digunakan dalam klasifikasi dokumen tekstual dimana dokumen-dokumen tersebut akan diklasifikasikan sesuai dengan topik dokumen yang ada (Darwis et al., 2021).

Definisi lain dari *text mining* atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal seperti itu yang membedakannya dengan *data mining* dimana data mining mengolah data yang sifatnya terstruktur (Jamil, 2017).

Text mining bisa dijelaskan sebagai suatu metode untuk menemukan informasi baru yang sebelumnya tidak diketahui oleh komputer, dengan cara otomatis mengekstrak informasi dari berbagai sumber. Proses ini secara krusial melibatkan penggabungan informasi yang berhasil diekstraksi dari berbagai sumber yang berbeda.

E. Python

Python merupakan bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Bahasa pemrograman python merupakan bahasa pemrograman yang bersifat *open source*. Python mendukung pemrograman berorientasi objek dan tersedia pada banyak sistem operasi seperti Windows, Linux, UNIX, Amigo, Mac OS, dll (Srinath, 2017).

Dalam bahasa python, terdapat istilah *library* yang mengacu pada kumpulan modul atau paket yang telah dikembangkan sebelumnya dan dapat diimpor ke dalam kode python untuk memberikan fungsionalitas tambahan. Didalam python terdapat banyak *library* yang membantu *text preprocessing* seperti *text folding*, *tokenezing*, *convert emoticon*, *stemming* dan lain-lain sebagaimana dalam analisis sentimen diperlukan adanya pemrosesan teks (Syah, H., & Witanti, A., 2022).

Dengan *sintaks* yang mirip dengan bahasa Inggris dan struktur yang intuitif, python memungkinkan programmer untuk mengekspresikan ide-ide mereka dengan jelas dan efisien. Bahasa pemrograman python khusus dirancang untuk memudahkan programmer dalam membuat program baik itu dari efesiensi waktu, mudah dalam pengembangan dan kompalibilitasnya dengan sistem. Python dapat digunakan

untuk membuat aplikasi *stand alone* (berdiri sendiri) atau pemrograman *script*. (A, Nur and T, Kurniawan., 2018).

F. *Crawling Data*

Crawling data merupakan tahap dalam penelitian yang bertujuan untuk mengumpulkan atau mengunduh data dari suatu *database* (George et al., 2014).

Crawling data dalam media sosial x dapat dilakukan dengan berbagai macam cara. Untuk penelitian ini penulis menggunakan metode *tweet-harvest*. *Crawling data* pada penelitian ini fokus kepada kata kunci “palestina merdeka hamas israel”. *Tweet-harvest* merupakan *tools* yang digunakan untuk melakukan *crawling data* pada media sosial Twitter dengan menggunakan *Application Programming Interface* (API). (S. A. Putra and A. Wijaya. 2023)

Tweet-harvest ini merupakan alat untuk mengambil atau *crawl* data dari media sosial X (Twitter), jadi pengguna bisa mendapatkan info yang lebih lengkap tentang topik yang sedang dibicarakan dalam bentuk *tweet/thread* atau komentar. Cara melakukan *crawling data* ialah dengan membuat program dengan memasukkan kata kunci untuk mencari *tweet* sesuai yang kita inginkan.

Sedangkan untuk metode *snsrape* saat ini sudah tidak bisa digunakan lagi, karena Twitter sudah mengubah API dan fitur mereka. X juga telah membatasi pencarian data tanpa

masuk ke situs web X, jika pengguna keluar dari akun X di situs web X, pengguna tidak akan bisa melakukan pencarian disana.

G. Text Preprocessing

Text preprocessing merupakan tahapan yang sangat krusial. *Text preprocessing* merupakan tahapan paling penting karena pada proses ini dilakukan perubahan data yang masih mentah menjadi data sesuai anjuran proses *mining* (Fikri et al., 2020).

Data yang belum diolah atau data tidak bersih kategorinya seperti format data yang tidak beraturan, data kosong, tipe data yang berbeda-beda, adanya atribut yang tidak penting, dan lain sebagainya semuanya harus diolah pada proses *preprocessing*. Semakin bersih pra proses yang dilakukan, maka kemungkinan besar hasil data tersebut semakin akurat (Irwansyah Saputra, 2022).

Tahap *text preprocessing* pada penelitian ini terdiri dari tujuh tahap, diantaranya adalah *case folding*, *cleaning*, *remove duplicate*, *normalizaion*, *stopword removal*, *stemming*, dan *tokenization*. Proses tersebut dilakukan secara berurutan dari awal hingga akhir.

1. Case Folding

Tahap dimana huruf kapital yang terdapat pada data semuanya diubah menjadi huruf kecil. Karakter yang dirubah menjadi huruf kecil adalah semua karakter huruf, mulai dari huruf "a" sampai dengan huruf "z".

2. *Cleansing*

Cleansing adalah proses yang bertujuan untuk membersihkan informasi dalam suatu kumpulan data, dengan menghilangkan kata-kata yang tidak relevan, tidak memiliki makna, atau memiliki potensi memengaruhi sentimen, seperti elemen HTML, tautan, nama pengguna (*mention*), dan tagar (*hashtag*).

3. *Remove Duplicate*

Proses ini bertujuan untuk menghapus data sentimen yang muncul secara berulang dengan teks yang sama persis, untuk nantinya hanya diambil satu data saja.

4. *Normalization*

Bertujuan untuk mengubah kata yang tidak baku menjadi baku, serta mengubah kata yang salah agar bisa dipahami.

5. *Stopword Removal*

Pada langkah ini, dilakukan penghapusan atau eliminasi daftar kata umum yang tidak memiliki informasi yang relevan. *Stopword* adalah tahapan untuk menghapus kata yang sering muncul tapi tidak memiliki arti penting dan maknanya tidak berpengaruh pada sistem, seperti „oh“, „di“, „pada“, dan sebagainya (Taufiqurrahman et al., 2021).

6. *Stemming*

Tahap *stemming* merupakan proses yang esensial untuk mengurangi jumlah indeks yang beragam pada satu set data, sehingga kata-kata yang memiliki akhiran (suffix) atau awalan (prefix) akan disederhanakan kembali ke bentuk dasarnya.

7. *Tokenization*

Pada langkah ini, dilakukan pembagian teks suatu kalimat menjadi beberapa bagian kata. Tokenisasi adalah pemecahan data set ke dalam bentuk token atau potongan kata agar mempermudah tahapan selanjutnya. Contohnya kalimat “saya ingin makan” dipenggal menjadi ['saya', 'ingin', 'makan'] (Baihaqi et al., 2020).

H. *TF-IDF*

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah suatu proses untuk melakukan transformasi data dari data tekstual ke dalam data numerik untuk dilakukan pembobotan pada tiap kata atau fitur (Septian, J. A., 2019). TF-IDF ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen. Dalam analisis teks dengan pembelajaran mesin, algoritma TF-IDF membantu mengurutkan data ke dalam kategori, serta mengekstrak kata kunci.

Jika terdapat sebuah kalimat "Perang antara Hamas dan Israel akan berakhir lama", pemahaman kita terhadap makna kata dan kalimat didasarkan pada pengetahuan semantik terlihat mudah. Namun, dalam konteks pemrograman, seperti Python, interpretasi kalimat semacam ini menjadi lebih kompleks. Bahasa pemrograman lebih efisien dalam memahami dan memproses data tekstual ketika direpresentasikan dalam bentuk nilai numerik.

Term Frequency (TF) merupakan jumlah kemunculan suatu kata dalam setiap dokumen, sementara *Inverse Document Frequency* (IDF) adalah nilai kebalikan dari jumlah dokumen yang mengandung kata tersebut. Pada proses perhitungan TF dan IDF akan dikalikan sehingga menghasilkan nilai bobot dari kata tersebut (Fikri et al., 2020). Berikut merupakan rumus persamaan TFIDF.

$$TFIDF(d,t) = TF(d,t) \cdot IDF(t)$$

Keterangan:

d = Dokumen

t = Kata

TF(d,t) = Jumlah kata pada setiap dokumen

Rumus TF(d,t)

$$TF(d,t) = \frac{\text{Jumlah kata t pada dokumen d}}{\text{Total kata pada dokumen d}} \quad (2.2)$$

Rumus IDF(t)

$$\text{IDF}(t) = \log \frac{\text{Total dokumen}}{\text{Jumlah dokumen yang mengandung kata } t} + 1 \quad (2.3)$$

I. *Split Validation Data*

Split validation data merupakan teknik validasi yang membagi dua data bagian secara acak, bagian data pertama digunakan sebagai data latih (*training*) dan bagian data kedua digunakan sebagai data uji (*testing*) (Turmudi Zy et al., 2021).

Split validation data adalah metode validasi yang memisahkan data secara acak menjadi dua bagian, di mana satu bagian digunakan sebagai data latih dan bagian lainnya sebagai data uji. Dengan menerapkan validasi pemisahan, dilakukan eksperimen pelatihan berdasarkan rasio pemisahan yang telah ditetapkan sebelumnya.

Data latih (*training*) adalah data yang akan dipakai dalam melakukan proses pembelajaran (*learning*) sedangkan data uji (*testing*) adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian (Turmudi Zy et al., 2021).

J. Tahap Evaluasi

Tahap evaluasi digunakan untuk menilai kesesuaian model dengan tujuan yang diharapkan. Evaluasi dilakukan dengan melihat nilai akurasi dan tabel *confusion matrix* (Yuniarti et al., 2020).

Secara umum tabel *confusion matrix* (matriks klasifikasi atau tabel kontigensi) digunakan untuk menentukan performa model *naive bayes* untuk mengklasifikasikan data, yang diterapkan untuk mengevaluasi hasil pengujian pada sistem dalam analisis sentimen. (Apriani et al., 2019). Berikut merupakan contoh tabel *confusion matrix* dua kelas.

Tabel 2. 1 *Confusion Matrix* Dua Kelas

		<i>True Class</i>	
		Positif	Negatif
<i>Predicted Class</i>	Positif	TP	FP
	Negatif	FN	TN

Berikut merupakan arti dari keempat nilai acuan pada tabel *confusion matrix*:

1. TP (*True Positif*) = Data yang diprediksi positif dan faktanya data itu positif (Sesuai).
2. FP (*False Positif*) = Data yang diprediksi positif dan faktanya data itu negatif (Tidak Sesuai).
3. FN (*False Negatif*) = Data yang diprediksi negatif dan faktanya data itu positif (Tidak Sesuai).
4. TN (*True Negatif*) = Data yang diprediksi negatif dan faktanya data itu negatif (Sesuai).

Confusion matrix bertujuan untuk menghitung berbagai *performance metrix*, sekaligus untuk mengukur kinerja model yang telah dibuat. Pada penelitian ini *performance metrix* digunakan untuk mengukur: *accuracy*, *precision*, *recall*, dan *F1*

score. *Confusion matrix* berisi berbagai performa yang dapat diukur seperti akurasi, presisi, *recall*, spesifikasi, dan *F1 Score* untuk mengetahui seberapa baik kinerja dari pemodelan yang telah dilakukan sebelumnya (Irwansyah Saputra, 2022).

Akurasi menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar. Maka, akurasi merupakan rasio prediksi benar dengan keseluruhan data. Dengan kata lain, akurasi merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya). Nilai akurasi dapat diperoleh dengan persamaan (2.1).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.1)$$

Presisi adalah ukuran sejauh mana sistem memberikan jawaban yang tepat sesuai dengan informasi yang diminta oleh pengguna. Penilaian nilai *precision* dapat dilakukan dengan merujuk pada persamaan (2.2)

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2.2)$$

Recall merupakan seberapa besar tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Untuk menghitung nilai *recall* dapat dilihat pada persamaan (2.3)

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.3)$$

F1 Score merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan (Irwansyah Saputra, 2022). Untuk menghitung *f1 score* dapat dilihat pada parameter (2.4)

$$f1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (2.4)$$

K. Kajian Penelitian Terkait

Terdapat lima kajian penelitian terkait sebelumnya, di bawah ini merupakan tabel kajian penelitian terkait.

Tabel 2. 2 Kajian Penelitian Terkait

No	Pustaka	Objek dan Metode	Hasil
1.	(Louis, M. C., 2023)	Analisis sentimen terhadap konflik Rusia-Ukraina di media sosial Twitter menggunakan algoritma <i>naïve bayes</i> .	Hasil analisa sentimen pada penelitian ini yang mengambil 414 data <i>tweet</i> , dengan rincian yang diperoleh berupa 146 data sentimen positif dan 268 data sentimen negatif menghasilkan tingkat presisi sebesar 96,91% .
2.	(Fikri, Mujaddid Izzul, et al., 2020)	Analisis sentimen tentang Universitas Muhammdiyah Malang (UMM) pada Twitter	Hasil perbandingan kedua metode <i>naïve bayes</i> dan <i>Support Vector Machine</i> (SVM) pada penelitian ini

		dengan metode <i>naive bayes</i> dan <i>Support Vector Machine</i> (SVM).	menunjukkan <i>Naive bayes</i> memiliki akurasi yang lebih tinggi yaitu 73,65% sedangkan metode <i>Support Vector Machine</i> (SVM) memiliki akurasi 70,20%.
3.	(Samsir, et al., 2021)	Analisi sentimen terhadap pembelajaran daring pada masa pandemi Covid 19 pada media sosial twitter dengan menggunakan media sosial <i>naive bayes</i> .	Hasil pengujian analisa sentimen dilakukan dengan memperoleh data sebesar 12.906 tweet dengan hasil sentimen 69% opini negatif, 30% opini positif dan 1% opini netral. Dengan hasil tingkat presisi pada metode <i>naive bayes</i> sebesar 97.15% .
4.	(Zidan, M., 2022).	Analisis sentimen kenaikan harga Bahan Bakar Minyak (BBM) berdasarkan respon pengguna media sosial twitter di Indonesia menggunakan metode <i>naive bayes</i> .	Penelitian tersebut mengambil sebanyak 1500 data tweet. Berdasarkan hasil performa pada metode <i>naive bayes</i> , penelitian tentang BBM ini menghasilkan nilai akurasi sebesar 81% , dan hasil presisinya sebesar 83%.
5.	(Darwis et al., 2021).	Penerapan algoritma <i>naive bayes</i> terhadap	Jumlah data dalam penelitian tersebut berjumlah 1179 data

	opini <i>review</i> data twitter BMKG Nasional.	tweet yang dikelompokan menjadi tiga sentimen yaitu positif, negatif, dan netral. Hasil uji akurasi pada metode <i>naïve bayes</i> untuk klasifikasi pada penelitian ini yaitu 69.97% .
--	---	--

Pada bagian kajian pustaka ini, akan diulas penelitian-penelitian sebelumnya dengan maksud untuk meningkatkan kevalidan pelaksanaan penelitian ini. Terdapat sebuah penelitian yang terkait dengan analisis sentimen dengan menggunakan metode *naïve bayes classifier*.

Penelitian pertama yang menjadi kajian pustaka adalah penelitian analisis sentiment terhadap konflik Rusia-Ukraina di media sosial X (Twitter) dengan menggunakan metode *naïve bayes*. Pada penelitian tersebut dibagi menjadi dua label sentimen saja, yaitu sentiment positif dan negatif. Data yang diambil dari penelitian tersebut berjumlah 414 data, dengan rincian yang diperoleh berupa 146 data sentimen positif dan 268 data sentimen negatif. Penelitian tersebut menghasilkan Tingkat presisi sebesar 96,91% (Lois, M. C., 2023).

Kemudian penelitian kedua berkaitan dengan penelitian perbandingan metode *naive bayes* dan *Support Vector Machine*

(SVM) yang dilakukan oleh mahasiswa Universitas Muhammadiyah Malang yaitu Mujaddid Izzul Fikri dkk, pada tahun 2020. Mereka Melakukan dua metode sekaligus untuk melakukan perbandingan pada opini tentang Universitas Muhammadiyah Malang (UMM). Penelitian tersebut mendapatkan 2030 tweet data yang kemudian dibagi menjadi data training dan data latih. Penelitian menyaring sentimen terhadap opini mengenai Universitas Muhammadiyah Malang yang memberikan sentimen positif dan negatif. Diperoleh dari kedua metode tersebut melakukan komparasi menunjukkan bahwa *naive bayes classifier* mendapatkan hasil akurasi 73,65% sedangkan *Support Vector Machine* mendapatkan akurasi 70,20%. Pada penelitian itu membuktikan bahwa metode *naive bayes classifier* memiliki tingkat akurasi yang lebih tinggi dibandingkan menggunakan metode *Support Vector Machine* (SVM). (Fikri et al., 2020).

Penelitian terkait lainnya yang mengacu pada metode *naive bayes classifier* adalah analisis sentimen pembelajaran daring pada media sosial X di masa pandemi COVID-19 yang dilakukan oleh Samsir dan kawan-kawannya. Dengan tujuan mengklasifikasikan sentimen berdasarkan respon pengguna media sosial X terhadap pembelajaran daring dengan memberikan beberapa sentimen, diantaranya sentimen positif, negatif, dan netral. Pada penelitian tersebut mengambil total 12906 *tweet* dengan hasil sentimen 69% opini negatif, 30%

opini positif dan 1% opini netral. Dengan hasil tingkat *presisi* pada metode *naive bayes* sebesar 97.15% (Samsir et al., 2021).

Penelitian berikutnya adalah penelitian tentang analisis sentimen kenaikan harga Bahan Bakar Minyak (BBM) berdasarkan respons pengguna media sosial X di Indonesia menggunakan metode *naive bayes classifier* yang dilakukan oleh mahasiswa teknologi informasi, UIN Walisongo Semarang, angkatan 2019 yang bernama Muhammad Zidan. Penelitian tersebut mengambil sebanyak 1500 data Twitter, yang kemudian setelah diolah data tersebut menjadi 1483 terdiri dari 550 (37,1%) komentar bersentimen positif, 798 (53,8%) komentar sentimen negatif, dan 135 (9,1%) sentimen netral. Berdasarkan hasil performa pada metode *naive bayes* terbukti memiliki algoritma yang akurat dan menghasilkan nilai akurasi sebesar 81%, serta hasil dari nilai presisinya sebesar 83% (Zidan, M., 2022).

Penelitian selanjutnya dilakukan oleh Dedi Darwis dan kawan-kawannya, yaitu tentang penerapan algoritma *naive bayes* terhadap opini *review* data Twitter BMKG Nasional. Jumlah data dalam penelitian tersebut berjumlah 1179 data *tweet* yang dikelompokkan menjadi tiga sentimen yaitu positif, negatif, dan netral. Klasifikasi ini menggunakan metode *naive bayes*. Klasifikasi dapat memberikan kemudahan bagi pengguna untuk melihat opini positif, negatif, dan netral. Hasil

uji akurasi pada metode *naïve bayes classifier* pada penelitian ini yaitu 69.97% (Darwis et al., 2021).

Hasil dari kelima penelitian tersebut memiliki akurasi dan presisi yang bermacam-macam. Penelitian yang membagi kelas sentimennya menjadi dua sentimen saja memiliki tingkat akurasi dan presisi yang lebih tinggi dibandingkan penelitian yang membagi sentimennya kedalam tiga kelas sentimen, jika terdapat tiga sentimen positif, netral, dan negatif, perbandingan sentimen netral dibandingkan keduanya sangat jauh dan berpotensi mengurangi hasil perhitungan akurasi dan presisi dari metode *naïve bayes classifier*.

Sedangkan untuk penelitian ini, selain objek penelitian yang berbeda, penelitian ini memiliki sedikit perbedaan pada tahap *normalization*, dimana pada tahap *normalization* peneliti menambahkan kamus sendiri yang berhubungan terkait dengan pembenaran kata yang berkaitan dengan dunia militer, peperangan, dan timur tengah yang kemudian ditambahkan dalam kamus *Indonesia lexion*. Sebelum *split validation data* dibagi menjadi tiga skenario.

BAB III

METODE PENELITIAN

A. Metode Pengumpulan Data

1. Studi Pustaka

Peneliti menjalankan kajian literatur dengan memanfaatkan sumber-sumber seperti buku-buku, jurnal, skripsi, dan referensi lainnya. Pendekatan ini digunakan untuk mendalami materi yang terkait dengan konsep, analisis, serta permasalahan lain yang menjadi fokus penelitian. Selain itu, upaya pencarian data diperluas dengan melakukan penelusuran daring melalui internet menggunakan peramban (*browser*). Proses pencarian ini mencakup kunjungan ke berbagai situs web yang terkait dengan analisis sentimen, *text mining*, dan algoritma *naïve bayes classifier*.

Pada tahap ini, penelitian mencakup kunjungan ke berbagai situs web yang memiliki keterkaitan erat dengan bidang analisis sentimen, *text mining*, dan algoritma *naïve bayes classifier*. Pendekatan daring ini memastikan bahwa peneliti memperoleh wawasan terbaru, tren, dan temuan terkini yang relevan dengan kerangka konseptual penelitian. Dengan menggabungkan sumber-sumber konvensional dan daring, peneliti berusaha untuk membangun dasar teoretis yang kuat dan terkini.

2. Studi Lapangan

Dalam pendekatan ini, peneliti melakukan pengamatan secara langsung terhadap dinamika yang tengah mencuat di media sosial X terkait dengan perang Hamas dengan pasukan tantara Israel. Penelitian ini memusatkan perhatian pada respons dan interaksi pengguna Twitter di Indonesia sebagai objek kajian utama. Dengan memanfaatkan metode ini, penulis menggali berbagai aspek terkait dengan perang antara Hamas dan Israel, mulai dari perasaan dan pandangan pengguna, hingga pola komunikasi dan opini yang muncul di platform tersebut.

Proses pengamatan langsung ini membuka peluang untuk mendapatkan wawasan yang lebih mendalam tentang bagaimana masyarakat merespons peristiwa ini secara *real-time*.

B. Perangkat Penelitian

Untuk melakukan penelitian ini, diperlukan beberapa perangkat keras dan perangkat lunak yang harus dipenuhi agar proses sistem dapat berjalan. Berikut adalah sejumlah kebutuhan yang digunakan dalam penelitian ini.

1. Kebutuhan Perangkat Keras:

Tabel 3. 1 Kebutuhan Perangkat Keras

No	Perangkat Keras	Spesifikasi
1.	Device	HP Laptop 14s-dq0xxx
2.	Processor	Intel(R) Celeron(R) N4120 CPU @ 1.10GHz, 1101 Mhz
3.	Memori (RAM)	4,00 GB
4.	Monitor	14 inch
5.	Keyboard dan Mouse	Standard

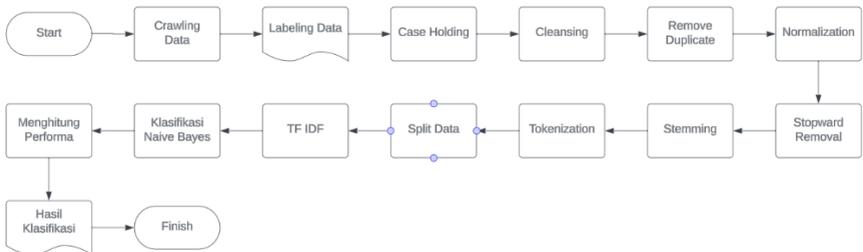
2. Kebutuhan Perangkat Lunak

Tabel 3. 2 Kebutuhan Perangkat Lunak

No	Perangkat Lunak	Spesifikasi
1.	Perangkat Operasai	Windows 11 64-bit
2.	Bahasa Pemrograman	Python
3.	MS Office	Ms Word dan Ms Excel 2019
4.	Browser	Google Chrome
5.	Google Drive	Google Colab

C. Alur Pengerjaan Penelitian

Alur pengerjaan adalah serangkaian langkah atau proses yang diikuti dalam menjalankan suatu tugas atau proyek. Dalam konteks penelitian, alur pengerjaan mencakup langkah-langkah dari perencanaan hingga pelaksanaan, analisis, dan pelaporan hasil penelitian. Berikut rangkaian pengerjaan tugas akhir dapat dilihat pada gambar 3.1.



Gambar 3. 1 *Flowchart* Alur Penelitian

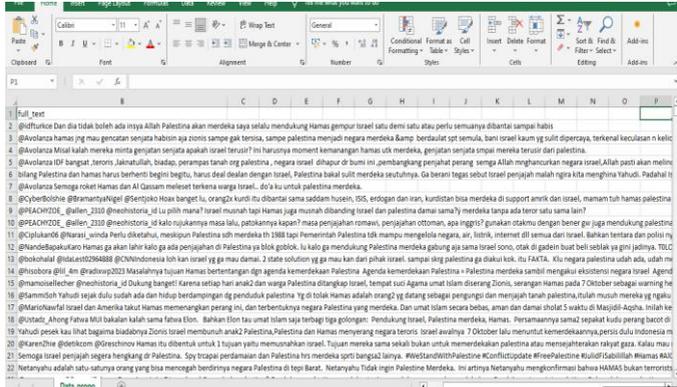
Penelitian ini diawali dengan mengumpulkan data opini/opini/tanggapan dari media sosial X (Twitter) dengan cara *crawling* data. Proses *crawling* diambil dengan menggunakan metode *tweet harvest* yang berkaitan dengan kata “palestina merdeka hamas israel”. Kemudian setelah semua data sudah didapatkan lalu dilakukan proses pelabelan data untuk menentukan sentimen opini/tanggapan yang ada. Pada penelitian ini menggunakan dua sentimen yaitu sentimen positif dan negatif.

D. Uraian Metodologi

Pada tahap ini dilakukan penguraian pada alur pengerjaan penelitian diatas untuk menjelaskan lebih dalam. Diantaranya sebagai berikut.

1. Pengambilan Data X (Twitter)

Pengambilan data pada media sosial X (twitter) dilakukan dengan menggunakan teknik *tweet harvest*. Pengambilan data ini menggunakan kata kunci "palestina merdeka hamas israel". Sedangkan pengambilan data *tweet* ini dilakukan pada 1 Januari 2024 dengan rentang data yang didapat adalah mulai dari tanggal 24 November 2023 sampai tanggal 31 Desember 2023 dengan mendapatkan jumlah data *tweet* berupa *text* sebanyak 1574 data *tweet* yang kemudian diproses dalam *preprocessing*. Kumpulan data itulah yang nantinya akan dijadikan bahan penelitian. Data *text* yang didapatkan pada media sosial X nantinya akan di unduh dengan file yang berformat CSV (*Comma Sparated Values*). Berikut merupakan gambar dari data yang diperoleh dari *tweet*.



Gambar 3. 2 Hasil *Crawling* Data *Tweet*

Setelah data telah didapatkan, kemudian data tersebut diberi label. Sentimen negatif memiliki label 0 dan data dengan sentimen positif diberi label 1. Data dikatakan memiliki sentimen positif jika berisi sebuah data *tweet* mendukung kemerdekaan Palestina, membela hamas mengutuk israel atau sekutunya, kemudian untuk data negatif berarti tidak mendukung kemerdekaan Palestina dan juga tidak mendukung hamas atau sentimen tersebut mendukung Israel. Proses labeling dilakukan dengan cara manual dan melibatkan dua ahli bahasa Indonesia. Berikut merupakan contoh sentimen dan pelabelannya.

Tabel 3. 3 Contoh Pelabelan Sentimen

Opini	Sentimen	Label
@Lowkey0nline Bersatulah fatah dan hamas untuk mengusir penjajah israel. Kami di indonesia mendukung palestina merdeka	Positif	1
@tokopedy @kegblgnunfaedh Makanya Palestina ga layak jd negara merdeka. Orang kebutuhan dasar warganya aja masih dibantu dr Israel. Air, gas, listrik itu full dr Israel. Yg jd pertanyaan kan: kemana aliran dana bantuan dr seluruh dunia? Jawaban: Beli roket + dipake pemimpin hamas kabur ke Qatar.	Negatif	0

2. *Preprocessing*

Tahap *preprocessing* dalam analisis data dari respons pengguna media sosial X (Twitter) memiliki tujuan utama untuk membersihkan dan menyiapkan data mentah agar dapat digunakan dalam proses pengklasifikasian atau analisis lebih lanjut. Proses ini melibatkan serangkaian langkah untuk mengatasi masalah umum seperti keberagaman format, *noise*, dan ketidaksesuaian data. Berikut adalah penjelasan lebih

rinci mengenai tujuan dan langkah-langkah yang terlibat dalam tahap *preprocessing* ini:

A. Case Folding

Tahap case folding ini bertujuan untuk merubah huruf kapital menjadi huruf kecil sekaligus bertujuan untuk menyeragamkan bentuk kata. Berikut merupakan contoh proses tahap *case folding* yang terdapat pada data *tweet* yang sudah diambil.

Tabel 3. 4 Penerapan Case Folding

<i>Input</i>	<i>Output</i>
@f_fathur Palestina merdeka ga merdeka, Israel hancur ga hancur apa untungnya bt kita ?.	@f_fathur palestina merdeka ga merdeka, israel hancur ga hancur apa untungnya bt kita ?.

B. Cleansing

Proses ini bertujuan untuk menghilangkan tanda baca atau simbol-simbol yang biasanya ada pada media sosial X (Twitter) dan menghapus angka yang berada pada data *tweet* sekaligus untuk menghilangkan username dan URL pada data *tweet*. Berikut merupakan contoh input dan output proses data pada tahap *cleansing*.

Tabel 3. 5 Penerapan Proses *Cleansing*

Input	Output
@f_fathur palestina merdeka ga merdeka, israel hancur ga hancur apa untungnya bt kita ?.	palestina merdeka ga merdeka israel hancur ga hancur apa untungnya bt kita

C. *Remove Duplicate*

Pada saat pengambilan data, pasti terdapat data dengan karakter yang sama. pada tahap *remove duplicate* dilakukan untuk penghapusan data berulang dan menghapuskan data *tweet* yang berisi opini yang sama. Berikut contoh *tweet* yang berisi opini yang sama ditunjukkan pada gambar 3.3.

langkatodaycom	Hamas-Israel Memanas, Rusia: Solusinya Pembentukan Negara Palestina Merdeka https://t.co/
OposisiCerdas	Hamas-Israel Memanas, Rusia: Solusinya Pembentukan Negara Palestina Merdeka https://t.co/i

Gambar 3. 3 Contoh Data yang Sama

D. *Normalization*

Pada tahap *normalization*, dilakukan upaya untuk mengubah kata-kata yang tidak sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) atau yang tidak formal menjadi memiliki arti sebenarnya. Hal ini, penting mengingat penggunaan kata-kata tidak baku atau gabungan bahasa dalam suatu opini di media sosial, terutama media sosial X dapat

meruncing dan mempersulit analisis teks. Berikut merupakan tabel penerapan proses *normalization*.

Tabel 3. 6 penerapan Proses *Normalization*

<i>Input</i>	<i>Output</i>
palestina merdeka ga merdeka israel hancur ga hancur apa untungnya bt kita	palestina merdeka gak merdeka israel hancur gak hancur apa untungnya buat kita

E. *Stopword Removal*

Pada proses *stopword removal*, langkah ini dilakukan dengan maksud menghilangkan kata-kata umum yang tidak memberikan kontribusi signifikan terhadap makna atau informasi yang diinginkan. Tujuannya adalah untuk mengurangi jumlah kata yang diproses atau disimpan oleh sistem, sehingga fokus dapat diberikan pada kata-kata yang lebih relevan dan memiliki dampak lebih besar terhadap pemahaman atau analisis yang sedang dilakukan. Berikut merupakan contoh *stopword removal*.

Tabel 3. 7 Penerapan Proses *Stopword Removal*

Input	Output
palestina merdeka gak merdeka israel hancur gak hancur apa untungnya buat kita	palestina merdeka israel hancur apa untungnya buat kita

F. *Stemming*

Pada langkah *stemming*, tindakan ini dilakukan untuk menggantikan suatu kata dengan imbuhan atau variasi bentuknya menjadi bentuk dasar. Selain itu, proses ini juga mencakup transformasi kata-kata yang memiliki perulangan atau pengulangan menjadi bentuk dasar tunggal. Tujuannya adalah untuk menyederhanakan dan mereduksi kata-kata ke bentuk dasarnya, memungkinkan analisis atau pemrosesan lebih efisien, serta menghindari duplikasi kata yang tidak perlu. Dengan melakukan *stemming*, informasi esensial dari berbagai variasi kata dapat tetap dipertahankan tanpa menambah kompleksitas yang tidak perlu dalam pemrosesan teks. Berikut merupakan contoh penerapan proses *stemming* pada tabel 3.8.

Tabel 3. 8 Penerapan Proses *Stemming*

<i>Input</i>	<i>Output</i>
palestina merdeka israel hancur apa untungnya buat kita	palestina merdeka israel hancur apa untung buat kita

G. *Tokenization*

Tokenizing/tokenization adalah metode untuk melakukan pemisahan kata dalam suatu kalimat dengan tujuan untuk proses analisis teks lebih lanjut, pada metode ini spasi juga akan dihilangkan.

Tokenization adalah tahap memecah kalimat menjadi bagian-bagian kata yang dinamakan token (Imron, 2019). Berikut merupakan contoh penerapan pada tahap *tokenization* dapat dilihat pada tabel 3.9.

Tabel 3. 9 Penerapan Proses *Tokenization*

<i>Input</i>	<i>Output</i>
palestina merdeka israel hancur apa untung buat kita	['palestina','merdeka', 'israel','hancur','apa','untung', 'buat','kita']

3. Ekstraksi Fitur

Setelah melalui seluruh tahap *preprocessing*, langkah berikutnya dalam analisis data adalah pembuatan fitur. Proses ini bertujuan untuk menciptakan representasi numerik atau struktural yang dapat memudahkan dan meningkatkan proses klasifikasi. Fitur-fitur ini dirancang untuk merepresentasikan informasi esensial dari data yang telah melalui tahap *preprocessing*, sehingga memungkinkan model klasifikasi untuk mengenali pola atau hubungan yang mendasar. Implementasi dilakukan melalui proses pembobotan kata dengan metode *Term Frequency-Inverse Document Frequency* (TFIDF). Pada tahap TFIDF, setiap kata dalam setiap kalimat atau dokumen diberikan pembobotan berdasarkan frekuensi kemunculannya dan keunikan kontribusinya terhadap seluruh korpus.

Setelah pemberian pembobotan ini, *dataset* telah dipersiapkan secara optimal untuk digunakan dalam tahap pelatihan model menggunakan algoritma *naive bayes classifier*. Dengan demikian, informasi yang terkandung dalam setiap kata telah diakomodasi dengan baik, memungkinkan model untuk mengenali dan memahami pola serta hubungan yang dapat digunakan untuk melakukan klasifikasi dengan tingkat keakuratan yang optimal.

4. Klasifikasi *Naive Bayes*

Metode *naive bayes classifier* merupakan metode yang digunakan untuk mengklasifikasikan data komentar untuk memperoleh sentimen analisis (Gunawan et al., 2018). Dalam tugas klasifikasi sentimen, pendekatan yang diambil melibatkan penggunaan data yang telah melalui serangkaian proses, mulai dari tahap *preprocessing* hingga pembobotan kata dengan metode TF-IDF. Setelah data berhasil diolah dan dipersiapkan, langkah selanjutnya adalah melatih model menggunakan *dataset* yang telah dibuat. Alasan penelitian ini menggunakan metode *naive bayes classifier* daripada menggunakan metode lain dikarenakan *naive bayes classifier* lebih cocok dan efisien dalam mengklasifikasikan data dalam bentuk teks. Dengan

menggunakan teorema *bayes*, model ini dapat menghitung probabilitas posterior kelas untuk suatu data uji. Sedangkan posterior adalah perkiraan probabilitas aktual dari hipotesis atau kelas tertentu setelah mengamati data baru.

5. Uji Model

Pengujian model dimulai dengan pembagian data menjadi dua bagian utama, yaitu data latih dan data uji, menggunakan metode yang dikenal sebagai *split validation*. Dalam proses ini, data keseluruhan dibagi menjadi dua subset yang saling eksklusif: satu subset digunakan untuk melatih model (data latih), sementara subset yang lain digunakan untuk menguji kinerja model (data uji). Metode *split validation* memungkinkan untuk mengukur seberapa baik model yang telah dilatih dapat memberikan prediksi yang akurat terhadap data yang belum pernah dilihat sebelumnya, membantu mengevaluasi generalisasi model di luar *dataset* pelatihan. Pengambilan data dilaksanakan secara acak dengan memanfaatkan perpustakaan (*library*) python yang tersedia.

6. Evaluasi Model

Tahap evaluasi digunakan untuk menilai kesesuaian model dengan tujuan yang diharapkan. Evaluasi

dilakukan dengan melihat nilai akurasi dan tabel *confusion matrix* (Yuniarti, et al. 2020). Evaluasi model dalam *data mining* adalah proses penilaian dan pengukuran kinerja suatu model prediktif atau klasifikasi terhadap data yang belum pernah dilihat sebelumnya. Tujuannya adalah untuk mengevaluasi sejauh mana model dapat memberikan prediksi yang akurat dan dapat diandalkan. Mengevaluasi model biasanya melibatkan akurasi, presisi, dan *recall*. Untuk penyajian datanya dibantu dengan *confusion matrix*.

7. Visualisasi

Pada tahap akhir penelitian ini, penulis akan menampilkan visualisasi hasil analisis sentimen menggunakan *wordcloud*. Selain itu juga akan dimunculkan diagram lingkaran yang akan menampilkan proporsi sentimen positif dan negatif dari data yang telah diolah, memberikan gambaran yang lebih rinci mengenai distribusi sentimen dalam korpus teks yang dianalisis. Sementara itu, penggunaan *wordcloud* bertujuan untuk menyajikan representasi visual dari hasil klasifikasi analisis sentimen. *Wordcloud* akan menyoroti kata-kata atau frasa yang paling sering muncul dalam konteks analisis sentimen. Ukuran kata dalam *wordcloud* akan mencerminkan frekuensi

kemunculannya dalam *dataset*, membantu dalam mengidentifikasi topik-topik utama atau fokus pembahasan yang paling dominan.

BAB IV HASIL DAN PEMBAHASAN

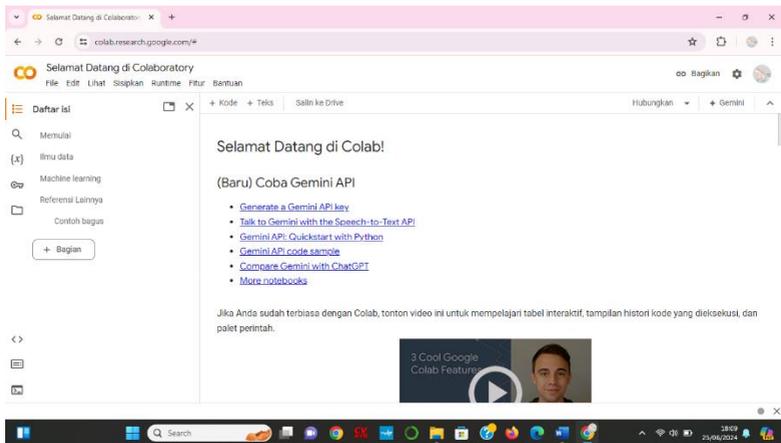
A. Pengambilan Data Media Sosial X

Proses pengambilan atau *crawling data tweet* pada media sosial X pada penelitian ini menggunakan metode *tweet harvest*. Proses *crawling data* pada media sosial X dilakukan dengan *login* ke akun media sosial X terlebih dahulu dengan memasukkan email dan password akun media sosial X.

Proses *crawling data* dengan menggunakan metode *tweet harvest* tidak secepat menggunakan *library snsrape*. Penelitian ini menggunakan metode *tweet harvest* dikarenakan metode tersebut tidak berbayar selain itu penelitian terbaru sekarang ini lebih banyak yang menggunakan *tweet harvest*. Proses *crawling data* dilakukan dua kali yaitu pada tanggal 1 Januari 2024 pukul 13:00 dan pada malam hari pukul 21:00 WIB. *Tweet harvest* adalah *library* untuk pengambilan data di Twitter dengan memanfaatkan autentikasi token. Metode ini dapat digunakan untuk pengambilan data berupa komentar pengguna sebagai sumber data. (B. A. Yuniarossy. et al., 2024).

Untuk menjalankan proses penelitian *software* yang digunakan adalah google colab dengan bahasa pemrograman python. Google Colab atau *Colaboratory* sangat efektif karena tidak memerlukan instalasi perangkat lunak, sehingga tidak menghabiskan memori. Pengguna hanya membutuhkan

koneksi internet dan satu akun Google untuk mengaksesnya, yang juga berguna untuk menyimpan hasil koding. Google Colab adalah sebuah IDE untuk pemrograman Python dimana pemrosesan akan dilakukan oleh server Google yang memiliki perangkat keras dengan performa yang tinggi (R. T. Handayanto dan H. Herlawati, 2020). Tampilan Google Colab dapat dilihat pada Gambar 4.1 di bawah ini.



Gambar 4. 1 Tampilan Google Colab 2024

Pada saat proses pengambilan *data tweet*, peneliti mengambil kata kunci dengan menggunakan kata "palestina merdeka hamas israel". Data yang diambil adalah *tweet* pada tanggal 8 Oktober 2023 sampai dengan 31 Desember 2023 dengan jumlah data paling banyak didapatkan pada bulan Desember dengan jumlah 701 data *tweet*, 599 pada bulan November, dan 274 pada bulan Oktober.

Sedangkan untuk memulai proses *crawling data*, perlu mempersiapkan dan mengimpor data yang diperlukan, sebagaimana ditunjukkan pada Gambar 4.2 di bawah ini.

```
1 # Import required Python package
2 |pip install pandas
3
4 # Install Node.js (because tweet-harvest built using Node.js)
5 |sudo apt-get update
6 |sudo apt-get install -y ca-certificates curl gnupg
7 |sudo mkdir -p /etc/apt/keyrings
8 |curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc
9
10 |NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.co
11
12 |sudo apt-get update
13 |sudo apt-get install nodejs -y
14
15 |node -v
```

Gambar 4. 2 Source Code Crawling Data

Terdapat perintah `!sudo apt-get update`. Perintah ini digunakan untuk memperbarui daftar paket yang tersedia di sistem menggunakan apt-get. Ini penting untuk memastikan bahwa sistem memiliki informasi paket yang terbaru sebelum melakukan instalasi.

`!sudo mkdir -p /etc/apt/keyrings` Perintah ini membuat direktori `/etc/apt/keyrings` jika belum ada. Direktori ini akan digunakan untuk menyimpan kunci GPG yang akan digunakan untuk menandatangani repositori perangkat lunak Node.js.

Selanjutnya adalah memasukan kata kunci dari data beserta dengan rentang waktu data untuk kemudian file dari

data tersebut disimpan, seperti terlihat pada gambar 4.3 dibawah ini.

```
1 # Crawl Data
2
3 filename = 'dataskripsikotor.csv'
4 search_keyword = 'palestina merdeka israel until:2023-11-24 since:2023-12-31'
5 limit = 700
6
7 !npx --yes tweet-harvest@2.2.8 -o "{filename}" -s "{search_keyword}" -l {limit} --token ""
```

Gambar 4. 3 Proses Crawling Data

Setelah mengeksekusi kode yang terlihat pada gambar 4.3, hasilnya adalah sebuah file CSV yang belum diproses. Gambar 4.4 menunjukkan hasil dari proses crawling yang telah diperoleh dalam bentuk file CSV yang masih perlu diolah lebih lanjut. File ini berisi data mentah yang perlu dianalisis atau dimanipulasi sesuai kebutuhan analisis lebih lanjut.

created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username	tweet_url
Fri Dec 29	1.74E+18	@RagiSemar insyaAllah ga akan bosan, sampai Palestina merdeka	0	1	0	6	in	1.57E+18	1.74E+18	Aulia1788	https://twitter.com/Aulia1788/status/1740...
Fri Dec 29	1.74E+18	Like dan Share jika kamu berpandangan bahwa Zionis Mirael adalah	0	1	1	0	in	1.17E+18	1.74E+18	Nurazra5	https://twitter.com/Nurazra5/status/1740...
Fri Dec 29	1.74E+18	@Ghazwita @Israel kakujaman jomni israel melampaui batas, iman	0	0	0	1	in	1.67E+18	1.74E+18	HamekiSr	https://twitter.com/HamekiSr/status/1740...
Fri Dec 29	1.74E+18	Ntar pas palestina merdeka, ign ky kita dulu ya kudu nangungu tau	0	0	0	0	in	1.13E+18	1.74E+18	aakso	https://twitter.com/aakso/status/1740729...
Fri Dec 29	1.74E+18	@IDF_Labs Tidak salah lagi, Mesir, Yordan dan Saudi ketakutan kali	0	0	0	0	in	1.41E+09	1.74E+18	dedweh	https://twitter.com/dedweh/status/17407...
Fri Dec 29	1.74E+18	UPDATE NEWS ... BERITA GEMBIRA !!! ALLAHU AKBAR BOSS AL - QAI	2	0	44	93	in	1.66E+18	1.74E+18	21_mozza	https://twitter.com/21_mozza/status/17407...
Fri Dec 29	1.74E+18	Serbu Bank dan Toko Valuta Asing di Tepi Barat, Tentara Israel Ran	1	0	0	0	in	4.56E+08	1.74E+18	merdekac	https://twitter.com/merdekac/status/1740...
Fri Dec 29	1.74E+18	@Tsa079013 Palestina itu tok akan merdeka karena itu tanah mi	0	1	1	3	in	1.62E+18	1.74E+18	arawarac	https://twitter.com/arawarac/status/1740...
Fri Dec 29	1.74E+18	@jaggle11 @wrahardian2 Palestina ga pernah punya wilayah sen	0	2	0	0	in	1.68E+18	1.74E+18	Nusantaraja	https://twitter.com/Nusantaraja/status/1740...
Fri Dec 29	1.74E+18	@thekeueli Klu Palestina merdeka berarti teroris israel hrs keluar	0	4	1	8	in	7.57E+17	1.74E+18	Sutikno88	https://twitter.com/Sutikno88/status/1740...
Fri Dec 29	1.74E+18	@linez25 Semoga Palestina segera merdeka dan israel segera binas	0	0	0	0	in	7.50E+17	1.74E+18	Mariono0	https://twitter.com/Mariono0773/status/1740...
Fri Dec 29	1.74E+18	@MOSSADH1 Aaaaaminnm3! Semoga Israel dan kota kota disana ik	0	0	0	0	in	7.50E+17	1.74E+18	ENS38474	https://twitter.com/ENS38474/status/1740...
Fri Dec 29	1.74E+18	@kuntwitter98 Semoga Rakyat Palestina kuat.Sabar .jawaqad,dar	0	1	0	5	in	1.58E+18	1.74E+18	AkunSem	https://twitter.com/AkunSem/25340616/stat...
Fri Dec 29	1.74E+18	@Demysingar7 serem amat si den to ngomongnya, ini pemlu bu	0	1	0	0	in	1.70E+18	1.74E+18	SupertiFe	https://twitter.com/SupertiFe/status/1740...
Thu Dec 21	1.74E+18	@linez25 Semoga Palestina segera merdeka dan israel segera binas	0	0	0	0	in	1.66E+18	1.74E+18	Mariono0	https://twitter.com/Mariono0773/status/1740...
Thu Dec 21	1.74E+18	Para pengunjung rasa menjujukan agresi israel yang bersandar-dar	2	0	11	25	in	1.63E+18	1.74E+18	ZerFalee	https://twitter.com/ZerFalee/status/1740...
Thu Dec 21	1.74E+18	@CNIndonesia Mendukung Palestina sampe merdeka Mendoaku	0	0	0	1	in	1.72E+08	1.74E+18	Risiez_Sai	https://twitter.com/Risiez_Sangeanu/status/1740...
Thu Dec 21	1.74E+18	@lovelyb1e Habisin Seluruh Persejajatan Zionis ISRAEL. Lawan Te	0	0	0	2	in	1.78E+18	1.74E+18	YawanMu	https://twitter.com/YawanMu3956/status/1740...
Thu Dec 21	1.74E+18	@thekeueli Israel akan dikn akan tenag sampai palestina merdeka	0	0	0	1	in	1.04E+18	1.74E+18	Akashadire	https://twitter.com/Akashadire/7686/status/1740...
Thu Dec 21	1.74E+18	@janganmlanding @bshpnyzzz @LoudInforties Loes sbg produki	0	0	0	0	in	1.04E+18	1.74E+18	Lathafah	https://twitter.com/Lathafah/Minim/status/1740...
Thu Dec 21	1.74E+18	@talangjoo KEBIJAKAN NEAGARA KITA TIDAK MENGAKUI ISRAEL. Hi	0	0	0	0	in	1.71E+18	1.74E+18	malik_har	https://twitter.com/malik_ham/7483/stat...

Gambar 4. 4 Hasil Crawling Data

Hasil *crawling data* yang didapat dari proses *crawling data* menggunakan metode *tweet harvest* kemudian di unduh.

Terdapat 12 kolom, salah satunya berisi tanggal *tweet*, *text tweet*, *username*, dan *URL tweet*. Data mentah yang diunduh memiliki format CSV (*Comma Separated Values*). Data berformat CSV juga bisa diubah menjadi *xlsx* atau *xlsm*. Data yang diambil untuk keperluan penelitian ini adalah *username* dan *full text*.

Setelah data didapatkan, selanjutnya adalah proses labeling, Proses pelabelan ini dilakukan secara manual dan dibantu oleh dua ahli bahasa Indonesia. Proses pelabelan 1574 data ini memakan waktu lebih dari tiga minggu.

	A	B	C
1	username	full_text	sentiment
790	iskandarizal02	insyaallah Palestina akan merdeka tidak lama lagi. Tanda semakin nyata. Generasi Palestina sekarang adalah pejuang. Generasi Zionists Positif	
791	satriaaagaza	(Perintah Serangan) Jenis operasi Serangan : Terbuka Target : SHIMON RIKLIN (Jurnalis Israel yang tidak bisa tidur sebelum rumah-rumal Positif	
792	agazkhatery_	Ohh gw baru dapat poinnya... kan waktu itu negara timur leste belum merdeka dan belum beraldaul dan masih jd bagian dari indonesia, i Negatif	
793	dwi_PALL109	Boikot produk India yang mendukung entitas Israel.!!! Boikot atas darah anak anak dan rakyat Palestina yang tidak bersalah..!! Boikot < Positif	
794	Nagakayu11	Wkwkwk elo tuh yg halu... kebanyakan makan babi Israel akan hancur, Palestina akan merdeka itu sdh tertulis Negatif	
795	YawanMu23956	Teruslah Berjuang Wahai Rakyat PALESTINA... Kemenangan Sudah Dekat. Allah SWT... Bersama Seluruh Rakyat PALESTINA. Sabar ∓ Positif	
796	Aminu08040819	hancurkan ZIONIS ISRAEL pembunuh anak anak, perempuan, rakyat sipil, menghancurkan fasilitas umum, rumah sakit, tempat ibadah Me Negatif	
797	Wutotward1	Isa papa serang saja hancurkan tentara Israel laknatullah Lebih mudah kalau membangun rumah sakit apalagi kalau anjing Israel musnah Negatif	
798	YawanMu23956	Tidak ada Waktu lagi...milah Momentum Paling Tepat Untuk Rakyat PALESTINA. Dihadapan Mata Seluruh Bangsa2. Agar PALESTINA Segi Negatif	
799	_rakyatmerdeka	Perang antara Israel dan Hamas yang masih berlangsung berdampak buruk terhadap kehidupan masyarakat di Amerika Serikat (AS). Kasi Positif	
800	ruti_279	Palestina udh bantu indonesia merdeka, udh waktunya indo bantu palestina merdeka. Bodo amat israel musuh juga ga rugl Positif	
801	trikartikatika1	Palestina menolong Indonesia saat Indonesia baru Merdeka,dan mendukung kemerdekaan Indonesia. Justu negara2 KAPRI menolak k Negatif	
802	Yelinta6	senga hamas maju terus dan menang melawan zionis israel laknatullah , palestina segera merdeka dan beraldaul mmbng bgsa palestn Positif	
803	teachmebi	Kebanyakan baca tribun nih Israel kl mau ngancurin gaza ya cepet, cuman buat apa? Mending ngambil wilayahnya dikit2... gw dukung p Positif	
804	gw_ Iwan	Truslah bersuara, mng sprti itu perjuangan. Berjuang mencari nafkah aja kdg bosen tp ttp hrs berjuang. Ditwtr msh brnyk yg bersuara me Positif	
805	hermanys	Perang sekarang beda dengan jaman dulu kak, israel secara tech dan sekutu jauh dilatannya, yang bisa buat palestina merdeka sebagai n Positif	
806	Fernynoesdon	Belum puas... Sampai Palestina merdeka dari babi Zionis-Israel. Pembela Israel di Indonesia itu ternyata dulu enak nya mantan pelihara Positif	
807	BlackRobusta	Indonesia ada dan merdeka, negara pertama kali yang mengukunya adalah negara Palestina. Ir. Sukarno bilang cita2 bangsa indonesia Positif	
808	Chikzt	Yg pasti Inggris penguasa wilayah habis menang perang atas ottoman dan israel jg akhirnya merdeka di 1948 dan bangsa arab palestina c Negatif	
809	ridho45916919	Bodoh kali pendukung zionis israel dajjal ini, rakyat palestina berhak mempertahankan dengan cara apapun dari dajjal zionis teroris isri Positif	
810	mstugu	serang terus hingga palestina merdeka,dan israel bubar Positif	

Gambar 4. 5 Data Setelah Diberi Sentimen

Data diberi sentimen positif apabila data *tweet* tersebut mendukung kemerdekaan Palestina, mendukung gerakan Hamas, dan tidak mendukung Israel. Sedangkan untuk data yang diberi sentimen negatif adalah data *tweet* yang mendukung Israel, tidak mendukung Hamas, serta tidak mendukung Palestina. Jika ada data netral, serta berpihak pada

sekutu Israel yang mengarah pada dukungan Israel maka data tersebut akan diberi label sentimen negatif.

Dari 1574 Jumlah data yang sudah diberi label, terdapat 1174 data tweet berlabel positif dan 400 data berlabel negatif. 1574 data yang sudah diberi label ini nantinya akan diolah kembali pada tahap *preprocessing*.

B. *Preprocessing*

Tahap *preprocessing* pada penelitian ini terdiri dari tujuh tahap, mulai dari *case folding*, *cleansing*, *remove duplicate*, *normalization*, *stopword removal*, *stamming*, dan *tokenization*. Ketujuh tahapan tersebut dilakukan secara berurutan, mulai dari *case folding* hingga *tokenization*. Data yang tidak terstruktur dan tidak konsisten perlu dibersihkan melalui proses *text preprocessing*. Setelah data bersih maka akan dilakukan proses pelabelan secara manual (A. N. Yusuf, E. Supriyati, and T. Listyorini., 2020)

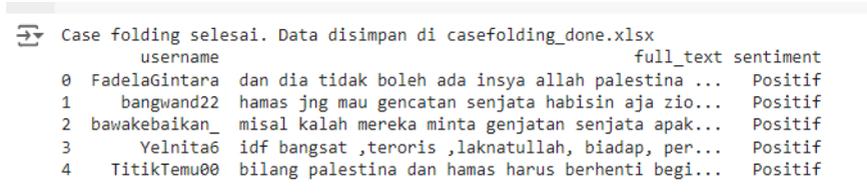
1. Case Folding

Pada penelitian ini, proses *case folding* bertujuan untuk menyeragamkan teks dengan mengubah semua huruf kapital menjadi huruf kecil. Berikut merupakan program yang menunjukkan implementasi dan hasil dari proses *case folding*.

```
1 #CASE FOLDING
2 import pandas as pd
3
4 file_path = 'Sudah Diberi Label 1574.xlsx'
5 df = pd.read_excel(file_path)
6
7 column_name = 'full_text'
8
9 # Melakukan case folding
10 df[column_name] = df[column_name].str.lower()
11 output_file_path = 'casefolding_done.xlsx'
12 df.to_excel(output_file_path, index=False)
13
14 print(f"Case folding selesai. Data disimpan di {output_file_path}")
15 print(df.head())
```

Gambar 4. 6 Source Code Case Folding

Output dari *source code* diatas langsung terdownload dan bisa dibuka melalui Ms Excel. Berikut merupakan hasil *output* dari *source code case folding* diatas.



```
Case folding selesai. Data disimpan di casefolding_done.xlsx
username full_text sentiment
0 FadelaGintara dan dia tidak boleh ada insya allah palestina ... Positif
1 bangwand22 hamas jng mau gencatan senjata habisin aja zio... Positif
2 bawakebaikan_ misal kalah mereka minta genjatan senjata apak... Positif
3 Yelnita6 idf bangsat ,teroris ,laknatullah, biadap, per... Positif
4 TitikTemu00 bilang palestina dan hamas harus berhenti begi... Positif
```

Gambar 4. 7 Hasil Case Folding

2. *Cleansing*

Proses pembersihan data pada penelitian ini bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dari komentar yang telah dikumpulkan. Berikut merupakan *source code* dari proses *cleansing*.

```

1 import pandas as pd
2 import re
3 def clean_text(text):
4     text = re.sub(r'\bRT\b', '', text) # Menghapus kata RT
5     text = re.sub(r'@\w+', '', text) # Menghapus username
6     text = re.sub(r'http\S+|https\S+', '', text) # Menghapus URL
7     text = re.sub(r'#\w+', '', text) # Menghapus hashtag
8     text = re.sub(r'(\.|\1+)', r'\1', text) # Menghapus pengulangan karakter
9     text = re.sub(r'\d+', '', text) # Menghapus angka
10    text = re.sub(r'^\w\s', '', text) # Menghapus tanda baca
11    text = text.strip() # Menghapus spasi di awal dan akhir teks
12    return text
13 def cleanse_excel(input_file, output_file, text_column):
14    df = pd.read_excel(input_file)
15    df[text_column] = df[text_column].apply(clean_text)
16    df.to_excel(output_file, index=False)
17    return df
18 input_file = 'casefolding_done.xlsx'
19 output_file = 'cleansing.xlsx'
20 text_column = 'full_text'
21 df = cleanse_excel(input_file, output_file, text_column)
22 print(df.head())

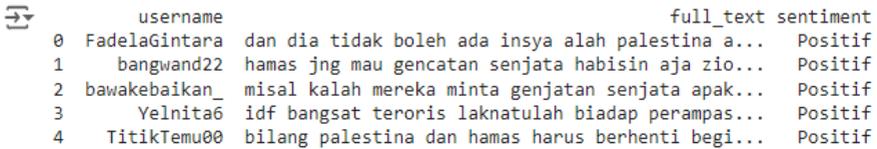
```

Gambar 4. 8 Source Code Cleansing

Cleansing pada penelitian ini terdapat tujuh tahapan agar data menjadi bersih dan lebih akurat. Berikut merupakan ketujuh proses *cleansing* yang dipakai pada penelitian ini, berikut prosesnya:

- a. Menghapus kata *retweet* (RT)
- b. Menghapus *username* seperti “@rio_redondo”
- c. Menghapus URL yang berada pada *tweet*
- d. Menghapus *hashtag* “#hashtag”
- e. Menghapus pengulangan karakter atau kata “sss” menjadi “s”
- f. Menghapus angka
- g. Menghapus spasi di awal dan di akhir teks

Berikut merupakan hasil dari data yang sudah melewati proses *cleansing*.

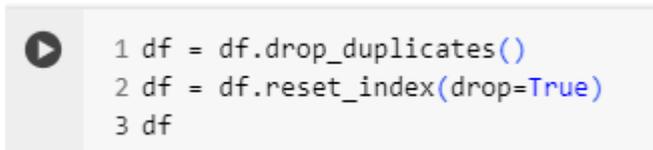


	username	full_text	sentiment
0	FadelaGintara	dan dia tidak boleh ada insya alah palestina a...	Positif
1	bangwand22	hamas jng mau gencatan senjata habisin aja zio...	Positif
2	bawakebaikan_	misal kalah mereka minta genjatan senjata apak...	Positif
3	Yelnita6	idf bangsat teroris laknatulah biadap perampas...	Positif
4	TitikTemu00	bilang palestina dan hamas harus berhenti begi...	Positif

Gambar 4. 9 Hasil *Cleansing*

3. *Remove Duplicate*

Pada tahap ini, tujuan utamanya adalah untuk menghilangkan tanggapan atau opini yang memiliki kesamaan atau duplikasi, sehingga hanya satu tanggapan yang diperlukan. Berikut merupakan *source code remove duplicate* yang digunakan pada penelitian ini.



```
1 df = df.drop_duplicates()
2 df = df.reset_index(drop=True)
3 df
```

Gambar 4. 10 *Source Code Remove Duplicate*

Setelah melakukan proses *remove duplicate*, data tersebut kemudian di unduh dan diperiksa kembali. Dari 1574 data yang di proses, akhirnya didapat 1529 data yang ditunjukkan pada gambar 4.11.



	username	full_text	sentiment
0	FadelaGintara	dan dia tidak boleh ada insya alah palestina a...	Positif
1	bangwand22	hamas jng mau gencatan senjata habisin aja zio...	Positif
2	bawakebaikan_	misal kalah mereka minta gencatan senjata apak...	Positif
3	Yelnita6	idf bangsat teroris laknatulah biadap perampas...	Positif
4	TitikTemu00	bilang palestina dan hamas harus berhenti begi...	Positif
...
1524	AmariArima27	berbanding terbalik dengan palestina palestina...	Positif
1525	wahyuhe6266	mengapa arab dan yordan tidak minta israel unt...	Netral
1526	ocamatchalatte	aku percaya takdirmu berakhir indah ya alah ak...	Positif
1527	yaniarsim	alhamdulillah mie instant gak diboikot tetep bo...	Positif
1528	lavenderoos	amin tapi dengan mereka masih ada sampe sekara...	Positif

1529 rows x 3 columns

Gambar 4. 11 Hasil *Remove Duplicate*

4. Normalization

Tahapan ini bertujuan untuk mengubah kata kurang jelas atau typo menjadi baku sekaligus mengubah kata slang menjadi memiliki arti yang lebih jelas. Berikut merupakan gambar *source code normalization* yang digunakan untuk penelitian ini.

```

1 import pandas as pd
2
3 kamus_path = 'lexion_Indonesia.xlsx'
4 data_path = 'Duplicate_done.xlsx'
5 kamus = pd.read_excel(kamus_path)
6 data = pd.read_excel(data_path)
7
8 kamus_dict = pd.Series(kamus['Formal'].values, index=kamus['Slang']).to_dict()
9 def normalize_text(text):
10     words = text.split()
11     normalized_words = []
12     for word in words:
13         normalized_words.append(kamus_dict.get(word, word))
14     return ' '.join(normalized_words)
15
16 data['Normalized'] = data['full_text'].apply(normalize_text)
17
18 output_file_path = 'normalisasiidata.xlsx'
19 data.to_excel(output_file_path, index=False)
20
21 # Menampilkan beberapa baris pertama dari data hasil normalisasi
22 print(data.head())

```

Gambar 4. 12 Source Code Normalization

Proses *normalization* ini dilakukan dengan tambahan data pendukung yang diambil dari file *colloquial-indonesian-lexicon*. Karena dalam penelitian ini terdapat beberapa istilah yang berkaitan dengan militer, perang, dan yang berkaitan dengan timur tengah, maka peneliti menambahkan pembenaran kosa-kata tersebut kedalam kamus. Hasil dari proses *normalization* dapat dilihat pada gambar 4.13.

	username	full_text	sentiment
0	FadelaGintara	dan dia tidak boleh ada insya alah palestina a...	Positif
1	bangwand22	hamas jng mau gencatan senjata habisin aja zio...	Positif
2	bawakebaikan_	misal kalah mereka minta genjatan senjata apak...	Positif
3	Yelnita6	idf bangsat teroris laknatullah biadap perampas...	Positif
4	TitikTemu00	bilang palestina dan hamas harus berhenti begi...	Positif

	Normalized
0	dan dia tidak boleh ada insya allah palestina ...
1	hamas jangan mau gencatan senjata habisin saja...
2	misal kalah mereka meminta genjatan senjata ap...
3	idf bangsat teroris laknatullah biadap perampa...
4	bilang palestina dan hamas harus berhenti begi...

Gambar 4. 13 Hasil *Normalization*

Untuk hasil yang lebih maksimal pada tahap *normalization*, peneliti memasukan satu data tambahan berupa kbba.text yang kemudian diubah menjadi excel. Berikut gambar implementasiannya.

```

1 import pandas as pd
2
3 kamus_path = 'kbba.xlsx'
4 data_path = 'sudahnormal.xlsx'
5 kamus = pd.read_excel(kamus_path)
6 data = pd.read_excel(data_path)

```

Gambar 4. 14 *Source Code* Memasukkan kbba

Data tambahan dari kbba.text setelah dirubah menjadi excel memiliki 1.329 data yang berupa membenaran kata, sedangkan untuk data tambahana pertama *colloquial-indonesian-lexicon* memiliki jumlah data 15.064 data termasuk yang secara manual telah ditambahkan. Berikut merupakan hasil dari *source code* diatas.

	username	full_text	sentiment
0	FadelaGintara	dan dia tidak boleh ada insya allah palestina ...	Positif
1	bangwand22	hamas jangan mau gencatan senjata habisin saja...	Positif
2	bawakebaikan_	misal kalah mereka meminta genjatan senjata ap...	Positif
3	Yelnita6	idf bangsat teroris laknatullah biadap perampa...	Positif
4	TitikTemu00	bilang palestina dan hamas harus berhenti begi...	Positif

Gambar 4. 15 Hasil *Normalization* Bagian Dua

5. *Stopword Removal*

Stopword removal dilakukan untuk menghilangkan kata-kata umum yang tidak memiliki arti penting atau informasi yang relevan. Untuk menjalankan proses *stopward removal* diperlukan *library nlp*, karena data yang diolah adalah data dengan bahasa Indonesia, maka tahap pertama yang harus dilakukan adalah *install nlp-id*. Di bawah merupakan gambar implementasinya pada pada penelitian ini.

```

✓ 8d 1 !pip install nlp-id

```

Gambar 4. 16 *Instal nlp-id*

Setelah proses *install* telah berhasil, selanjutnya adalah meng-impor *nlp-id* disertai juga memanggil kelas *stopward* dan juga *source code* yang digunakan untuk proses *stopward removal*. Seperti pada gambar 4.17.

```
1 from nlp_id import StopWord
2 import pandas as pd
3
4 df = pd.read_excel('sudahnormal.xlsx')
5 from nlp_id import StopWord
6 stopword = StopWord()
7
8 df['full_text']=df['full_text'].apply(stopword.remove_stopword)
9 df['full_text'].head()
```

Gambar 4. 17 *Source Code Stopword Removal*

Dari *source code* diatas, maka akan memperoleh hasil output seperti ditunjukkan pada gambar 4.18 dibawah ini.

```
0 insya allah palestina mendukung hamas gempur i...
1 hamas gencatan senjata habisin zionis enggak t...
2 kalah genjatan senjata israel terusir momen ke...
3 idf bangsat teroris laknatullah biadap perampa...
4 bilang palestina hamas berhenti setuju dealan ...
Name: full_text, dtype: object
```

Gambar 4. 18 *Output Stopword Removal*

6. *Stemming*

Tahap *stemming* adalah proses di mana kata-kata yang memiliki imbuhan diubah menjadi bentuk dasarnya. Proses *stemming* menggunakan *source code* yang hampir sama dengan *stopword removal* hanya saja ada beberapa perintahnya yang berbeda. Untuk melakukan *stemming*

perlu menggunakan bantuan *library sastrawi*, ditunjukkan pada gambar 9.9

```
✓ 6d [13] 1 !pip install sastrawi
```

Gambar 4. 19 Instal Library Sastrawi

Library sastrawi ini adalah yang paling umum digunakan, *Sastrawi* adalah *library python* yang digunakan untuk melakukan *stemming* pada kata-kata dalam bahasa Indonesia. Setelah sudah di *install* selanjutnya adalah mengimpor kelas *stemmer factory* dari modul *sastrawi* seperti diperlihatkan pada gambar 4.20.

```
✓ 0d ▶ 1 from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

Gambar 4. 20 Impor Kelas *Stemmer Factory*

StemmerFactory adalah kelas yang memungkinkan untuk membuat dan menggunakan *stemmer* dengan mudah dalam pemrosesan teks bahasa Indonesia. Proses selanjutnya adalah memasukan *source code* untuk proses *stemming* seperti gambar 4.21 dibawah ini.

```
✓ 4m ▶ 1 import pandas as pd
2 factory = StemmerFactory()
3 stemmer = factory.create_stemmer()
4
5 df = pd.read_excel('hasil_stopword.xlsx')
6 from nlp_id import StopWord
7 stopword = StopWord()
8
9 df['full_text'] = df['full_text'].apply(stemmer.stem)
10 df['full_text'].head()
```

Gambar 4. 21 *Source Code Stemming*

Eksekusi *stemming* menggunakan *source code* diatas memakan waktu lebih dari empat menit. Hasil dari proses *stemming* dapat dilihat pada gambar 4.22.

```
0 insya allah palestina merdeka dukung hamas gem...
1 hamas gencat senjata habisin zionis enggak sis...
2 kalah gencat senjata israel usir momen kemanan...
3 idf bangsat teroris laknatullah biadap ampas t...
4 bilang palestina hamas henti israel palestina ...
Name: full text, dtype: object
```

Gambar 4. 22 Hasil *Stemming*

7. *Tokenization*

Tahap *tokenization* atau tokenisasi adalah proses di mana kalimat atau teks dipecah menjadi unit-unit kata yang lebih kecil. Untuk melakukan *tokenization* library yang dibutuhkan adalah *nlTK*. *Library nlTK* ini menyediakan alat dan sumber daya untuk bekerja dengan teks, termasuk tokenisasi, *tagging*, *parsing*, dan lain sebagainya. Untuk *source code tokenization* dapat dilihat pada gambar 4.23.

```

1 import pandas as pd
2 import nltk
3 from nltk.tokenize import word_tokenize
4 from google.colab import files
5
6 uploaded = files.upload()
7
8 for file_name in uploaded.keys():
9     df = pd.read_excel(file_name)
10
11 def split_word(teks):
12     list_teks = []
13     for txt in teks.split(" "):
14         list_teks.append(txt)
15     return list_teks
16
17 df['full_text'] = df['full_text'].apply(split_word)
18 df['full_text'].head()

```

Gambar 4. 23 Source Code Tokenization

Dari *source code* diatas maka akan mnghasilkan *output* seperti pada gambar 4.24.

```

Pilih File | fixstemming.xlsx
• fixstemming.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 77170 bytes.
Saving fixstemming.xlsx to fixstemming (1).xlsx
0 [insya, allah, palestina, merdeka, dukung, ham...
1 [hamas, gencat, senjata, habisin, zionis, engg...
2 [kalah, gencat, senjata, israel, usir, momen, ...
3 [idf, bangsat, teroris, laknatullah, biadap, a...
4 [bilang, palestina, hamas, henti, israel, pale...
Name: full_text, dtype: object

```

Gambar 4.24 Hasil Proses *Tokenization*

Tokenization ini adalah serangkaian proses terakhir dari tahap *preprocessing*. Hasil akhir dari proses *tokenization* di *download* dalm file excel.

Setelah semua proses *preprocessing* selesai, selanjutnya adalah mengunduh hasilnya dalam format excel seperti pada gambar 4.24.

	A	B	C
1	username	full text	sentimen
2	FadelaGintara	['insya', 'allah', 'palestina', 'merdeka', 'dukung', 'hamas', 'gempur', 'israel', 'bantai', 'habis']	Positif
3	bangwand22	['hamas', 'gencat', 'senjata', 'habisin', 'zionis', 'tidak', 'sisa', 'palestina', 'negara', 'merdeka', 'daulat', 'Positif	Positif
4	bawakebaikan	['kalah', 'gencat', 'senjata', 'israel', 'usir', 'momen', 'menang', 'hamas', 'merdeka', 'gencat', 'senjata', 'Positif	Positif
1524	Bingeko3	['israel', 'palestina', 'indonesia', 'dukung', 'palestina', 'merdeka', 'usir', 'zionistisrael']	Positif
1525	Aguskwee	['palestina', 'merdeka', 'israel', 'lenyap', 'muka', 'bumi', 'israel', 'lenyap', 'muka', 'bumi', 'manusia', 'Negatif	Negatif
1526	AmariArima27	['banding', 'balik', 'palestina', 'palestina', 'negara', 'merdeka', 'deklarasi', 'aku', 'negara', 'bahlan', 'lePositif	Positif
1527	wahyuhe6266	['arab', 'yordan', 'israel', 'lanjut', 'perang', 'serah', 'hamas', 'palestina', 'merdeka', 'arab', 'yordan', 'oiNegatif	Negatif
1528	ocamatchalatte	['percaya', 'takdir', 'indah', 'allah', 'percaya', 'happy', 'ending', 'palestina', 'merdeka', 'surga', 'mu', 'isrPositif	Positif
1529	yaniarsim	['alhamdulillah', 'mie', 'instant', 'boikot', 'boikot', 'franchise', 'dukung', 'zionis', 'israel', 'palestina', 'rPositif	Positif
1530	lavenderoos	['amin', 'israel', 'gagal', 'bikin', 'pergi', 'palestina', 'a', 'testament', 'deh', 'lindung', 'allah', 'moga', 'hiPositif	Positif

Gambar 4. 24 Setelah Proses *Preprocessing*

Jumlah data yang tadinya berjumlah 1574 data, setelah melakukan *preprocessing* diperoleh data sebanyak 1529 data. Pengurangan jumlah data terjadi pada proses *remove duplicate*. Dari seluruh data yang sudah selesai diolah pada tahap *preprocessing* belum diketahui jumlah sentimen yang positif dan negatif, oleh karena itu peneliti menghitung seberapa banyak jumlah sentimen negatif dan positif, untuk mengetahui apakah jumlah dari sentimen sesuai dengan jumlah data dan memastikan pada label sentimen tidak ada kekeliruan kata dan jumlah.

Dengan lebih mengetahui distribusi jumlah sentimen, maka dapat lebih memahami karakteristik data secara mendalam. Gambar 4.25 di bawah menunjukkan jumlah sentimen negatif dan positif yang ada pada

data. gambaran awal mengenai distribusi sentimen yang ada dalam *dataset*.

```
0 d ▶ 1 sentiment_counts = data['sentiment'].value_counts()
      2 print("Jumlah label positif dan negatif pada kolom 'Sentiment':")
      3 print(sentiment_counts)
```

```
↔ Jumlah label positif dan negatif pada kolom 'Sentiment':
sentiment
Positif      1139
Negatif      390
Name: count, dtype: int64
```

Gambar 4. 25 Jumlah Sentimen

Jumlah sentimen setelah dilakukan proses perhitungan diketahui sebanyak 1139 sentimen positif dan 390 sentimen negatif. Jika sentimen positif dan negatif dijumlahkan maka akan mendapatkan hasil 1529, hasil tersebut sudah sesuai dengan jumlah data teks yang diperoleh.

C. Ekstrasi Fitur

Data yang sudah diolah pada tahap *preprocessing* selanjutnya masuk pada tahap ekstrasi fitur. Ekstrasi fitur bertujuan untuk mengenali dan mengolah fitur-fitur yang relevan dari data teks yang akan diterapkan dalam model analisis sentimen. Ekstraksi fitur mencoba untuk mengekstrak satu set fitur (dimensi) yang sama sekali baru dari pola data daripada memilih fitur dari atribut yang ada (Birjali et al., 2020).

Dalam proses ekstraksi fitur, penting untuk menyiapkan *library* yang akan digunakan, seperti *library sklearn* atau *scikit-*

learn. *Sklearn* adalah *library open-source* yang dirancang untuk memberikan alat dan algoritma yang mudah digunakan untuk analisis data dan machine learning. *Scikit-learn* terintegrasi dengan baik dengan pustaka Python lainnya, seperti NumPy, Pandas, dan matplotlib.

Selanjutnya, dilakukan pembagian data untuk validasi, yaitu memisahkan data uji dari data latih guna mempermudah proses klasifikasi. Dalam pembagian data agar mendapat hasil lebih maksimal, pembagian data uji dilakukan secara tiga kali yaitu menggunakan data uji sebanyak 20%, 30%, dan 40% dari keseluruhan data (Halrasyid, G. 2023). Proses pembagian data dapat dilihat pada Gambar 4.26.

```
1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(data['full_text'], data['sentiment'], test_size=0.2, random_state=42)

1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(data['full_text'], data['sentiment'], test_size=0.3, random_state=42)

1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(data['full_text'], data['sentiment'], test_size=0.4, random_state=42)
```

Gambar 4. 26 *Split Validation Data*

Saat proses *split validation data*, pembagian data dilakukan dengan cara bergantian satu demi satu, dimulai dari pembagian data uji terkecil hingga yang paling besar.

Proses awal yang dilakukan adalah mengubah kumpulan data menjadi representasi vektor. Untuk melakukan ini, sistem menggunakan LabelEncoder untuk mengonversi teks menjadi angka. Tujuan dari konversi ini adalah agar data dapat diproses oleh sistem untuk keperluan klasifikasi.

Selanjutnya adalah tahapan pembobotan kata dengan menggunakan TF-IDF (*Term Frequency -Invers Document Frequency*). Untuk *source code* dari pembobotan TFIDF dapat dilihat pada gambar 4.27.

```
1 from sklearn.preprocessing import LabelEncoder
2 Encoder = LabelEncoder()
3 y_train = Encoder.fit_transform(y_train)
4 y_test = Encoder.fit_transform(y_test)
5
6 from sklearn.feature_extraction.text import TfidfVectorizer
7 Tfidf_vect = TfidfVectorizer(max_features=520)
8 Tfidf_vect.fit(df['full_text'])
9 Train_X_Tfidf = Tfidf_vect.transform(X_train)
10 Test_X_Tfidf = Tfidf_vect.transform(X_test)
11
12 Train_X_Tfidf.toarray()
```

Gambar 4. 27 Pembobotan Kata dengan TF-IDF

LabelEncoder adalah kelas dari *sklearn.preprocessing* yang digunakan untuk mengonversi label kategori menjadi nilai numerik. Ini berguna untuk mengubah data kategori teks menjadi bentuk yang dapat digunakan dalam algoritma *machine learning*. *Output* dari *source code* TF-IDF diatas dapat dilihat pada gambar 4.28 di bawah ini.

```

array([[0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.29821044,
        0.         ],
       ...,
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ]])

```

Gambar 4. 28 Hasil Pembobotan Kata TF-IDF

Hasil dari *output* TF-IDF adalah *array numpy* dengan label yang sudah diencode dalam bentuk angka dan matriks sparse yang mengandung skor TF-IDF untuk fitur kata yang ada dalam dokumen teks.

Contoh perhitungan manual pada tahap pembobotan kata menggunakan TF-IDF adalah sebagai berikut:

Doc 1: "Palestina pasti merdeka dari Israel"

Doc 2: "Israel pasti akan menang mengalahkan Palestina Hamas"

Doc 3: "Hancurkan Israel kemenangan untuk Palestina merdeka"

Kemudian data tersebut akan berubah karena sudah melewati tahap preprocessing, sehingga berubah menjadi seperti berikut:

Doc 1: ['palestina', 'pasti', 'merdeka', 'dari', 'israel']

Doc 2: ['israel', 'pasti', 'akan', 'menang', 'kalah', 'palestina']

Doc 3: ['hancur', 'israel', 'menang', 'untuk', 'palestina', 'merdeka']

Berikutnya adalah menghitung TF-IDF dengan cara manual, penghitungan dengan cara manual bertujuan untuk mendapatkan pemahaman yang mendalam tentang bagaimana algoritma TF-IDF ini bekerja. Tabel 4.1 merupakan contoh perhitungan TF (*Term Frequency*) secara manualisasi yang diambil dari contoh dokumen diatas.

Tabel 4. 1 Perhitungan TF

Term	Doc 1	Doc 2	Doc 3
palestina	1/5	1/6	1/6
pasti	1/5	1/6	0
merdeka	1/5	0	1/6
dari	1/5	0	0
israel	1/5	1/6	1/6
akan	0	1/6	0
menang	0	1/6	1/6
kalah	0	1/6	0
hancur	0	0	1/6
untuk	0	0	1/6

Hasil 1/5 diperoleh karena kata tersebut muncul 1 kali dari total 5 kata dalam dokumen kesatu. Setelah berhasil menghitung TF selanjutnya adalah menghitung IDF. Seperti ditunjukkan pada tabel 4.2.

Tabel 4. 2 Perhitungan IDF

Term	DF	IDF
palestina	3	$\log(3/3)=0$
pasti	2	$\log(3/2)=0.176$
merdeka	2	$\log(3/2)=0.176$
dari	1	$\log(3/1)=0.477$
israel	3	$\log(3/3)=0$
akan	1	$\log(3/1)=0.477$
menang	2	$\log(3/2)=0.176$
kalah	1	$\log(3/1)=0.477$
hancur	1	$\log(3/1)=0.477$
untuk	1	$\log(3/1)=0.477$

Tahap Selanjutnya adalah mengalikan jumlah TF dengan IDF pada setiap dokumen. Berikut merupakan contoh perkalian TF-IDF ditunjukkan seperti tabel 4.3.

Tabel 4. 3 Perkalian TF dan IDF

Term	Doc 1 (TF-IDF)	Doc 2 (TF-IDF)	Doc 3 (TF-IDF)
palestina	$1/5 \times 0 = 0$	$1/6 \times 0 = 0$	$1/6 \times 0 = 0$
Pasti	$1/5 \times 0.176 = 0.035$	$1/6 \times 0.176 = 0.029$	0
merdeka	$1/5 \times 0.176 = 0.035$	0	$1/6 \times 0.176 = 0.029$
Dari	$1/5 \times 0.477 = 0.095$	0	0
israel	$1/5 \times 0 = 0$	$1/6 \times 0 = 0$	$1/6 \times 0 = 0$
Akan	0	$1/6 \times 0.477 = 0.080$	0

menang	0	$1/6 \times 0.176 = 0.029$	$1/6 \times 0.176 = 0.029$
Kalah	0	$1/6 \times 0.477 = 0.080$	0
hancur	0	0	$1/6 \times 0.477 = 0.080$
untuk	0	0	$1/6 \times 0.477 = 0.080$

Setelah selesai maka hasil lebih jelasnya dapat dilihat pada tabel 4.4 dibawah ini.

Tabel 4. 4 Hasil TF-IDF

Term	Doc 1	Doc 2	Doc 3
palestina	0	0	0
pasti	0.035	0.029	0
merdeka	0.035	0	0.029
dari	0.095	0	0
israel	0	0	0
akan	0	0.080	0
menang	0	0.029	0.029
kalah	0	0.080	0
hancur	0	0	0.080
untuk	0	0	0.080

Semakin tinggi nilai yang diperoleh dari setiap kata dimasing-masing dokumen, berarti semakin tinggi juga bobot kata tersebut, semakin tinggi angkanya berarti kata tersebut

juga semakin penting menurut algoritma TF-IDF. Untuk tampilan *array* bisa dilihat dibawah ini:

```
Array([\n  [0, 0.035, 0.035, 0, 0.95, 0, 0, 0, 0, 0, 0],\n  [0, 0.029, 0, 0, 0, 0.080, 0.029, 0.080, 0, 0, 0],\n  [0, 0, 0, 0.029, 0, 0, 0, 0.029, 0, 0.080, 0.080]\n])
```

Setiap elemen dalam array ini menunjukkan nilai TF-IDF untuk masing-masing *term* sesuai urutan yang sudah ditetapkan. Ini menyediakan representasi numerik dari dokumen yang dapat dimanfaatkan untuk berbagai analisis teks.

D. Metode Naive Bayes Classifier

Setelah ekstrasi fitur selesai berikutnya adalah melakukan klasifikasi *naive bayes*. Langkah ini penting, pada proses pengklasifikasian perlu disiapkan *library* yang akan digunakan yaitu *library sklearn*.

Proses pengklasifikasian ini dilakukan tiga kali, karena pada penelitian ini menggunakan tiga perbandingan data latih dan uji. Berikut merupakan tabel perbandingan dan jumlah data yang digunakan.

Tabel 4. 5 Perbandingan Data Latih dan Uji

Perbandingan	Jumlah Data
80% / 20%	306
70% / 30%	459
60% / 40%	612

Selain menggunakan *library sklearn*, impor juga *MultinomialNB*, yang merupakan implementasi dari algoritma *naive bayes classifier* untuk data yang didistribusikan menurut multinomial. Algoritma ini sering digunakan untuk klasifikasi teks. Selain itu terdapat juga *accuracy_score*, *precision_score*, *recall_score*, *f1_score*, *classification_report* dan *confusion matrix*. Berikut merupakan *source code import library sklearn*.

```
1 #NBC
2 from sklearn.naive_bayes import MultinomialNB
3 from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
4 from sklearn.metrics import classification_report
5 from sklearn.metrics import confusion_matrix
```

Gambar 4. 29 Source Code Import sklearn

Berikutnya adalah melakukan proses *naive bayes classifier*. *Source code* tahap *naive bayes classifier* dapat dilihat pada gambar 4.30.

```
9 clf = MultinomialNB().fit(Train_X_Tfidf.toarray(), y_train)
10 predicted = clf.predict(Test_X_Tfidf.toarray())
11 print("Multinomial NB Accuracy:", accuracy_score(y_test, predicted))
12 print("Multinomial NB Precision:", precision_score(y_test, predicted, average='weighted'))
13 print("Multinomial NB Recall:", recall_score(y_test, predicted, average='weighted'))
14 print("Multinomial NB F1 Score:", f1_score(y_test, predicted, average='weighted'))
15 print(f'confusion matrix:\n {confusion_matrix(y_test, predicted)}')
16 print('-----\n')
17 print(classification_report(y_test, predicted, zero_division=0))
```

Gambar 4. 30 Proses Naive Bayes Classifier

Hasil *accuracy* dari ketiga skenario dapat dilihat pada gambar 4.31 dibawah ini.

- Multinomial NB Accuracy sekenario 1: 0.8169934640522876
- Multinomial NB Accuracy sekenario 2: 0.7864923747276689
- Multinomial NB Accuracy sekenario 3: 0.8120915032679739

Gambar 4. 31 Hasil Akurasi Ketiga Skenario

E. Uji Model

Uji model dilakukan dengan tujuan untuk mengukur ketepatan model dalam melakukan pengklasifikasian data dan untuk memperoleh hasil klasifikasi yang ditampilkan dalam bentuk tabel *confusion matrix*. Pada penelitian ini membagi menjadi dua kelas yaitu kelas positif dan negatif. Berikut merupakan tabel *confusion matrix* ditunjukkan pada tabel 4.6.

Tabel 4. 6 *Confusion Matrix*

		<i>True Class</i>	
		Positif	Negatif
<i>Predicted Class</i>	Positif	TP	FP
	Negatif	FN	TN

Dengan adanya *confusion matrix*, menghitung akurasi model menjadi lebih mudah. Berikut merupakan rumus untuk menghitung nilai akurasi, ditunjukkan pada persamaan (4.1)

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

Selanjutnya adalah menghitung nilai akurasi dari data uji dengan menggunakan rumus diatas yang sudah diketahui *confusion matrix*-nya. Untuk mengetahui *confusion matrix* dari

dataset, berikut merupakan *source code*-nya, ditunjukkan pada gambar 4.32.

```
[ ] 1 print(f'confusion matrix:\n {confusion_matrix(y_test, predicted)}')
```

Gambar 4. 32 *Source Code* Memunculkan *Confusion Matrix*

Berikut merupakan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.33. Tiga skenario pembagian data uji (20%, 30%, 40%) ditunjukkan secara berurutan.

```
[[ 31  53]  [[ 39  96]  [[ 56 113]
 [  3 219]] [  2 322]] [  2 441]]
```

Gambar 4. 33 *Confusion Matrix* Ketiga Skenario

Pada gambar 4.33 menunjukkan *confusion matrix* dari ketiga skenario, namun pada gambar 4.33 diawali dari *True Negative* (TN) lalu bagian pojok kanan bawah adalah *True Positive* (TP). Berikut merupakan tampilan *confusion matrix* setelah disesuaikan terlihat pada tabel 4.7.

Tabel 4. 7 *Confusion Matrix* Data Uji

			True Class	
			Positive	Negative
Predicted Class	Skenario 1 (80/20)	Positive	219	53
		Negative	3	31
	Skenario 2 (70/30)	Positive	322	96
		Negative	2	39
	Skenario 3 (60/40)	Positive	441	113
		Negative	2	56

Berikutnya adalah melakukan perhitungan untuk menilai performa akurasi model.

Akurasi menggunakan 20% data uji

$$Akurasi = \frac{31 + 219}{31 + 219 + 53 + 3} \times 100\% = 81.69\%$$

Akurasi menggunakan 30% data uji

$$Akurasi = \frac{39 + 322}{39 + 322 + 96 + 2} \times 100\% = 78.64\%$$

Akurasi menggunakan 40% data uji

$$Akurasi = \frac{56 + 441}{56 + 441 + 113 + 2} \times 100\% = 81.20\%$$

Dari perhitungan diatas disimpulkan bahwa, akurasi terbesar didapatkan dari skenario pertama yang memiliki jumlah presentase akurasi sebesar 81.69%. Sedangkan untuk presentase akurasi paling kecil terdapat pada skenario kedua yaitu sebesar 78.64%. Untuk skenario ketiga memperoleh hasil 81.20%.

Mengetahui nilai akurasinya saja tidaklah cukup, karena jumlah data tiap kelas pada data latih tidak seimbang. Akurasi tidak memberikan informasi tentang bagaimana model menangani kelas minoritas, dalam penelitian ini yang dimaksud dalam kelas minoritas adalah sentimen dengan labe negatif.

Metrik lain seperti *precision*, *recall*, dan *F1-score* memberikan gambaran lebih mendetail tentang kinerja model dalam memprediksi setiap kelas. Oleh karena itu maka

diperlukan evaluasi model agar lebih efektif dalam memprediksi kelas minoritas.

F. Evaluasi Model

Dengan menggunakan berbagai metrik evaluasi seperti *precision*, *recall*, dan *f1-score* hal itu dapat memastikan bahwa model tidak hanya memberikan kinerja tinggi secara keseluruhan tetapi juga andal dalam situasi praktis.

Setelah hasil dari akurasi ketiga skenario sudah didapat selanjutnya adalah menghitung *precision*, *recall*, dan *f1-score*. Tabel 4.8 merupakan hasil akurasi dari ketiga skenario perbandingan data uji.

Tabel 4. 8 Hasil Akurasi Ketiga Skenario

	Skenario 1 (80% / 20%)	Skenario 2 (70% / 30%)	Skenario 3 (60% / 40%)
Akurasi	81.69%	78.64%	81.20%

Selanjutnya adalah menentukan *precision*, *recall*, dan nilai *f1-score* dari kelas negatif dan positif dari ketiga skenario data uji. Berikut merupakan perhitungannya dari masing-masing skenario secara manual.

1. Skenario 1 (Perbandingan 80% / 20%)

Skenario pertama dengan data uji 20% dan jumlah data sebanyak 306 memiliki *confusion matrix* yang terlihat pada tabel 4.9 di bawah ini.

Tabel 4. 9 *Confusion Matrix* Skenario 1

Skenario 1		<i>True Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	219	53
	<i>Negative</i>	3	31

True Positive (TP) : 219

False Positive (FP) : 53

False Negative (FN) : 3

True Negative (TN) : 31

Data dari *confusion matrix* yang sudah didapat digunakan untuk menghitung *precision*, *recall*, dan *F1-score* dari setiap kelas. Berikut merupakan proses perhitungannya.

a. *Precision* Kelas Positif

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{219}{219 + 53} \times 100\% = 81\%$$

b. *Recall* Kelas Positif

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{219}{219 + 3} \times 100\% = 99\%$$

c. *F1-score* Kelas Positif

$$f1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

$$f1 \text{ score} = 2X \frac{0.81 \times 0.99}{0.81 + 0.99} X 100\% = 89\%$$

d. *Precision* Kelas Negatif

$$Precision = \frac{TN}{TN + FN} X 100\%$$

$$Precision = \frac{31}{31 + 3} X 100\% = 91\%$$

e. *Recall* Kelas Negatif

$$Recall = \frac{TN}{TN+FP} X 100\%$$

$$Recall = \frac{31}{31 + 54} X 100\% = 37\%$$

f. *F1-score* Kelas Negatif

$$f1 \text{ score} = 2X \frac{Precision \times Recall}{Precision+Recall} X 100\%$$

$$f1 \text{ score} = 2X \frac{0.91 \times 0.37}{0.91 + 0.37} X 100\% = 53\%$$

2. Skenario 2 (Perbandingan 70% / 30%)

Skenario kedua dengan data uji 30% dan jumlah data sebanyak 459 memiliki *confusion matrix* yang terlihat pada tabel 4.10 di bawah ini.

Tabel 4. 10 *Confusion Matrix* Skenario 2

Skenario 2		<i>True Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	322	96
	<i>Negative</i>	2	39

True Positive (TP) : 322

False Positive (FP) : 96

False Negative (FN) : 2
True Negative (TN) : 39

Data dari *confusion matrix* yang sudah didapat digunakan untuk menghitung *precision*, *recall*, dan *f1-score* dari setiap kelas. Berikut merupakan proses perhitungannya.

a. *Precision* Kelas Positif

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{322}{322 + 96} \times 100\% = 77\%$$

b. *Recall* Kelas Positif

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{322}{322 + 2} \times 100\% = 99\%$$

c. *F1-score* Kelas Positif

$$f1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

$$f1\ score = 2 \times \frac{0.77 \times 0.99}{0.77 + 0.99} \times 100\% = 87\%$$

d. *Precision* Kelas Negatif

$$Precision = \frac{TN}{TN + FN} \times 100\%$$

$$Precision = \frac{39}{39 + 2} \times 100\% = 95\%$$

e. *Recall* Kelas Negatif

$$Recall = \frac{TN}{TN + FP} \times 100\%$$

$$Recall = \frac{39}{39 + 96} \times 100\% = 29\%$$

f. *F1-score* Kelas Negatif

$$f1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

$$f1\ score = 2 \times \frac{0.95 \times 0.29}{0.95 + 0.29} \times 100\% = 44\%$$

3. Skenario 3 (Perbandingan 60% / 40%)

Skenario kedua dengan data uji 40% dan jumlah data sebanyak 612 memiliki *confusion matrix* yang terlihat pada tabel 4.11 di bawah ini.

Tabel 4. 11 *Confusion Matrix* Skenario 3

Skenario 3		<i>True Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted</i>	<i>Positive</i>	441	113
<i>Class</i>	<i>Negative</i>	2	56

True Positive (TP) : 441

False Positive (FP) : 113

False Negative (FN) : 2

True Negative (TN) : 56

Data dari *confusion matrix* yang sudah didapat digunakan untuk menghitung *precision*, *recall*, dan *f1-score* dari setiap kelas. Berikut merupakan proses perhitungannya.

a. *Precision* Kelas Positif

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{441}{441 + 113} \times 100\% = 80\%$$

b. *Recall* Kelas Positif

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{441}{421 + 2} \times 100\% = 100\%$$

c. *F1-score* Kelas Positif

$$f1\ score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \times 100\%$$

$$f1\ score = 2 \times \frac{0.80 \times 1}{0.80 + 1} \times 100\% = 88\%$$

d. *Precision* Kelas Negatif

$$Precision = \frac{TN}{TN + FN} \times 100\%$$

$$Precision = \frac{56}{56 + 2} \times 100\% = 97\%$$

e. *Recall* Kelas Negatif

$$Recall = \frac{TN}{TN+FP} \times 100\%$$

$$Recall = \frac{56}{56 + 113} \times 100\% = 33\%$$

f. *F1 score* Kelas Negatif

$$f1\ score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \times 100\%$$

$$f1\ score = 2 \times \frac{0.97 \times 0.33}{0.97 + 0.33} \times 100\% = 49\%$$

Berikut merupakan hasil dari *precision*, *recall*, dan juga *f1-score* pada kelas negatif dan positif dari ketiga skenario, yang dapat dilihat pada tabel 4.12

Tabel 4. 12 Hasil Performa Kelas

<i>Class</i>		<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Skenario 1	Positif	81%	99%	89%
	Negatif	91%	37%	53%
Skenario 2	Positif	77%	98%	87%
	Negatif	95%	29%	44%
Skenario 3	Positif	80%	100%	88%
	Negatif	97%	33%	49%

Dari tabel 4.12 dapat disimpulkan bahwa nilai *precision* kelas positif terbesar terdapat pada skenario pertama dengan presentase 81%, sedangkan untuk skenario kedua dan ketiga masing-masing memiliki nilai presentase yang sebesar 77% dan 80%. Untuk nilai *precision* negatif yang terbesar terdapat pada skenario ketiga dengan nilai presentase sebesar 97%, sedangkan skenario pertama dan kedua masing-masing memiliki nilai 91% dan 95%.

Untuk nilai *recall* pada kelas positif yang terbesar adalah pada skenario ketiga dengan presentase nilai *recall* sebesar 100%, sedangkan skenario pertama dan kedua memiliki nilai *recall* masing-masing sebesar 99% dan 98%. Pada kelas negatif *recall* terbesar adalah pada skenario pertama dengan presentase sebesar 37% dan yang terkecil berada pada skenario kedua dengan presentase *recall* 29%, skenario ketiga memiliki *recall* 33%.

Presentase nilai *f1-score* pada kelas positif yang terbesar adalah pada skenario pertama dengan presentase 89%, sedangkan skenario kedua memiliki presentase 87% dan skenario ketiga memiliki nilai presentase 88%. Pada kelas negatif yang memiliki nilai presentase nilai *f1-score* terbesar adalah pada skenario pertama dengan nilai presentase sebesar 53%, sedangkan untuk presentase skenario kedua memiliki presentase sebesar 44% dan nilai presentase skenario ketiga adalah 49%.

Untuk perhitungan nilai *accuracy*, *precision*, *recall* dan *f1-score* pada suatu sistem dapat menggunakan *source code* yang tunjukkan pada gambar 4.33 sebagai berikut.

```
1 from sklearn.naive_bayes import MultinomialNB
2 from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
3 from sklearn.metrics import classification_report
4 from sklearn.metrics import confusion_matrix
5
6 clf = MultinomialNB().fit(Train_X_Tfidf.toarray(), y_train)
7 predicted = clf.predict(Test_X_Tfidf.toarray())
8 print("Multinomial NB Accuracy:", accuracy_score(y_test, predicted))
9 print("Multinomial NB Precision:", precision_score(y_test, predicted, average='weighted'))
10 print("Multinomial NB Recall:", recall_score(y_test, predicted, average='weighted'))
11 print("Multinomial NB F1 Score:", f1_score(y_test, predicted, average='weighted'))
12 print(f'confusion matrix:\n {confusion_matrix(y_test, predicted)}')
13 print('-----\n')
14 print(classification_report(y_test, predicted, zero_division=0))
```

Gambar 4. 34 *Source Code* Perhitungan *Naive Bayes*

Source code perhitungan *naive bayes classifier* di atas akan menunjukkan hasil performa *naive bayes* dari keseluruhan kelas dan masing-masing kelas yang ditunjukkan dengan angka nol (0) sampai dengan satu (1). Hasil dari *source code* tersebut juga menampilkan *macro average* dan *weighted average* dari *precision*, *recall*, dan *f1-score*. Berikut merupakan

Hasil dari *source code* perhitungan *naive bayes classifier* dari ketiga skenario, ditunjukkan pada gambar 4.34 sampai dengan gambar 4.36.

```

↳ Multinomial NB Accuracy: 0.8169934640522876
Multinomial NB Precision: 0.8344146482122261
Multinomial NB Recall: 0.8169934640522876
Multinomial NB F1 Score: 0.7874823572467483
confusion matrix: Skenario 1
[[ 31 53]
 [ 3 219]]
-----

```

	precision	recall	f1-score	support
0	0.91	0.37	0.53	84
1	0.81	0.99	0.89	222
accuracy			0.82	306
macro avg	0.86	0.68	0.71	306
weighted avg	0.83	0.82	0.79	306

Gambar 4. 35 Hasil Perhitungan Skenario 1

```

↳ Multinomial NB Accuracy: 0.7864923747276689
Multinomial NB Precision: 0.8235362764547994
Multinomial NB Recall: 0.7864923747276689
Multinomial NB F1 Score: 0.7430002017959842
confusion matrix: Skenario 2
[[ 39 96]
 [ 2 322]]
-----

```

	precision	recall	f1-score	support
0	0.95	0.29	0.44	135
1	0.77	0.99	0.87	324
accuracy			0.79	459
macro avg	0.86	0.64	0.66	459
weighted avg	0.82	0.79	0.74	459

Gambar 4. 36 Hasil Perhitungan Skenario 2

```

↔ Multinomial NB Accuracy: 0.8120915032679739
Multinomial NB Precision: 0.8428320392433499
Multinomial NB Recall: 0.8120915032679739
Multinomial NB F1 Score: 0.7766094199791509
confusion matrix:Skenario 3
[[ 56 113]
 [ 2 441]]

```

```

-----
                precision    recall  f1-score   support

     0           0.97         0.33         0.49         169
     1           0.80         1.00         0.88         443

 accuracy                   0.81         612
 macro avg           0.88         0.66         0.69         612
 weighted avg        0.84         0.81         0.78         612

```

Gambar 4. 37 Hasil Perhitungan Skenario 3

Hasil perhitungan dengan menggunakan metode *naive bayes classifier* pada ketiga skenario setelah dievaluasi menunjukkan hasil yang sangat baik dengan tingkat akurasi tertinggi pada skenario pertama dengan nilai presentase sebesar 81.69%. Hasil *precision*, *recall*, dan *f1-score* juga sangat baik pada skenario ketiga. Tabel 4.13 menunjukkan performa keseluruhan dari metode *naive bayes classifier* dari ketiga skenario.

Tabel 4. 13 Hasil Performa *Naive Bayes Classifier*

	<i>Acuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Skenario 1	81.69%	83.44%	81.69%	78.74%

Skenario 2	78.64%	82.35%	78.64%	74.30%
Skenario 3	81.20%	84.28%	81.20%	77.66%

Perolehan presentase nilai *precision*, *recall*, dan *f1-score* yang ditampilkan pada tabel 4.13 adalah menggunakan *weighted average* (rata-rata berbobot). Hal ini dikarenakan kinerja model secara keseluruhan memiliki kelas yang tidak seimbang dalam *dataset*.

Tabel 4.13 menunjukkan bahwa performa terbaik dari perhitungan dengan menggunakan metode *naive bayes classifier* adalah pada skenario pertama dengan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji. Skenario pertama mendapatkan hasil presentase *accuracy* sebesar 81.69%, *precision* sebesar 83.44%, *recall* sebesar 81.69%, dan *f1-score* sebesar 78.74%. Presentase tersebut merupakan yang paling besar dibandingkan dengan skenario kedua dan ketiga.

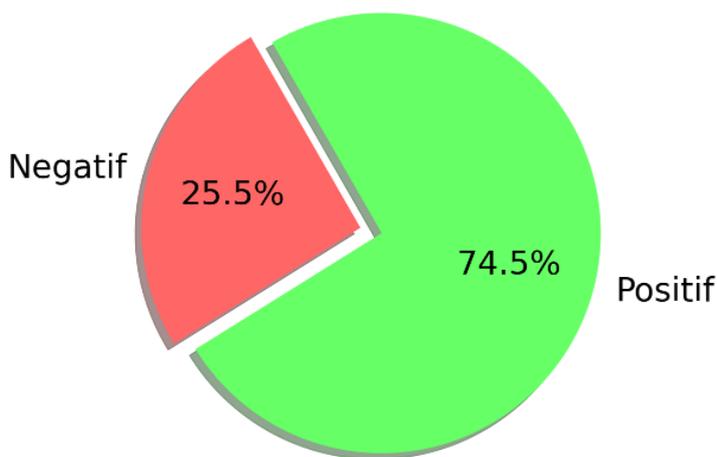
G. Visualisasi

Terakhir adalah memvisualisasikan hasil dari analisis sentimen dengan *wordcloud*. *Wordcloud* dapat membantu dalam menilai konteks sentimen. Selain itu, *wordcloud* membuat data teks lebih mudah dipahami oleh audiens yang lebih luas, termasuk mereka yang mungkin tidak memiliki latar belakang teknis.

Pada penelitian ini diketahui hasil akhir dari jumlah data setelah dilakukan *preprocessing* berjumlah 1529 data dengan jumlah data sentimen positif sebesar 1139 dan jumlah sentimen negatif berjumlah 390. Presentase data positif dan negatif dapat dilihat pada gambar 9.9.

9.9

Distribusi Sentimen



Gambar 4. 38 Presentase Distribusi Sentimen

Dari gambar 4.37 menunjukkan bahwa presentase sentimen positif lebih mendominasi hampir tiga kali lipat yaitu sebesar 74.5%, sedangkan sentimen negatif memiliki presentase 25.5%. Dari hasil *tweet* yang diperoleh dari media sosial X berbahasa Indonesia menunjukkan bahwa pengguna *tweet* yang mendukung hamas dan mendukung Palestina

sentimen negatif, hal itu dikarenakan jumlah sentimen dari sentimen positif lebih banyak daripada sentimen negatif.

BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Metode *naive bayes classifier* terbukti dapat menganalisis sentimen dengan klasifikasi yang baik. Dalam penelitian ini digunakan tiga skenario komposisi data. Pada skenario pertama menggunakan perbandingan data sebesar 80% data latih dan 20% data uji, pada skenario kedua menggunakan 70% data latih dan 30% data uji, kemudian pada skenario ketiga menggunakan perbandingan data sebesar 60% data latih dan 40% data uji. Jumlah keseluruhan data diambil secara acak. Proses yang dilakukan untuk melakukan analisis sentimen pada penelitian ini mulanya melakukan ekstraksi teks dengan melalui tahapan preprocessing, kemudian pembobotan dengan TF-IDF, dan klasifikasian teks dengan metode *naive bayes classifier*. Hasil klasifikasi yang diberikan berupa kelas dengan sentimen positif dan negatif.
2. Dari penelitian ini didapatkan sebanyak 1574 data *tweet* yang kemudian setelah melalui tahap *preprocessing* menjadi 1529 data *tweet* dengan

menghasilkan data sebanyak 1139 bersentimen positif, 390 komentar bersentimen negatif. Sehingga diketahui sentimen positif memiliki nilai persentase tertinggi sebesar 74.5%, sedangkan pada sentimen negatif sebesar 25.5%.

3. Dari ketiga skenario, yang memiliki performa terbaik adalah pada skenario pertama dengan menggunakan perbandingan data 80% data latih dan 20% data uji. Pada metode *naive bayes classifier* terbukti menghasilkan nilai akurasi sebesar 81.69%, *precision* sebesar 83.44%, *recall* sebesar 81.69% serta *f1 score* sebesar 78.74%.

B. Saran

Berdasarkan penelitian yang telah dilakukan, penulis mengharapkan agar peneliti berikutnya dapat mengembangkan penelitian ini lebih lanjut, serta memberikan beberapa saran seperti berikut:

1. Menggunakan *dataset* dengan jumlah yang lebih banyak agar pola yang didapat oleh model lebih bervariasi. Sehingga semakin banyak informasi yang dipelajari oleh model.
2. Peningkatan koleksi kamus untuk mencakup kata-kata tidak baku atau slang, karena di media sosial terdapat banyak komentar yang menggunakan bahasa informal.

3. Dalam penelitian ini, data dikumpulkan dari media sosial X, pada penelitian berikutnya diharapkan data dapat diperoleh dari platform media sosial lainnya.
4. Penulis memiliki harapan besar agar pada penelitian berikutnya, dapat dikembangkan lebih lanjut menjadi sebuah aplikasi atau platform berbasis website. Hal ini bertujuan agar dapat lebih mudah diakses dan digunakan oleh khalayak yang lebih luas, sehingga manfaatnya dapat dirasakan secara maksimal.

DAFTAR PUSTAKA

- A. Nur Syahrudin and T. Kurniawan, "Input Dan Output Pada Bahasa Pemrograman Python," *Jurnal Dasar Pemrograman Python STIMK*, Jun. 2018.
- A. N. Yusuf, E. Supriyati, and T. Listyorini. 2020. "Analisis Sentimen Mengenai Layanan Provider Indihome Berdasarkan Pendapat Pelanggan Melalui Media Sosial Twitter dengan Metode Naïve Bayes Classifier," vol. 04, pp. 75–78.
- A. Samad, H. Basari, B. Hussin, I. G. Pramudya, and J. Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," *Procedia Eng.*, vol. 53, no. December, pp. 453–462, 2013, doi: 10.1016/j.proeng.2013.02.059.
- B. A. Yuniarossy, K. M. Hindrayani, and A. T. Damaliana, "Analisis Sentimen Terhadap Isu Feminisme di Twitter Menggunakan Model Convolutional Neural Network (CNN)," *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 5, no. 1, Apr. 2024. DOI: 10.46306/lb.v5i1.
- Baihaqi, W. M., Pinilih, M., Rohmah, M., Informasi, T., Purwokerto, U. A., Informasi, S., Purwokerto, U. A., & Korespondensi, P. (2020). Kombinasi KMeans Dan Support Vector Machine (Svm) Untuk K-Means and Support Vector Machine (Svm) Combination To Predict Sara Elements on Tweet. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(3), 501–510.
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. doi:10.1016/j.knosys.2021.107134
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma

Naive Bayes untuk Analisis Sentimen Review Data Twitter
BMKG Nasional. *Jurnal TEKNO KOMPAK*, 15(1), 131–145.

Fajar, R., Program, S., Rekayasa, P., Lunak, N., & Bengkalis, R. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *JURNAL INOVTEK POLBENG*, 3(1), 50–59.

Fikri, M. I., Sabrila, T. S., Azhar, Y., & Malang, U. M. (2020). Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA Jurnal*, 10(2), 71–76.

Gunawan, B., Sasty, H., #2, P., Esyudha, E., & #3, P. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika*, 4(2), 17–29.

H. Dhery, A. Assyam, and F. N. Hasan, “Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp.

Halrasyid, Gerald. (2023). Implementasi Algoritma Naive Bayes dan TF-IDF dalam Analisis Sentimen Data Ulasan (Studi Kasus: Ulasan Review Aplikasi E-Commerce Shopee di Situs Google Playstore). Universitas Islam Negeri Syarif Hidayatullah Jakarta.

I. Taufik dan S.A.Pamungkas. (2018). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal “LOG!K@,”* 8(1), 71–79.

Idatul, K. M., & Fadhil, A. (2019). Gerakan Intifadhah dan Kemunculan Hamas (1987–1993). journal.unj.ac.id/unj/index.php/periode/article/view/10478/6704.

- Imron, A. (2019). *ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DI KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER*. Universitas Islam Indonesia.
- Irwansyah Saputra, D. A. K. (2022). *Machine Learning untuk Pemula*. Informatika Bandung.
- Jamil, H.N. 2017. *Analisis Sentimen Pada Online Review Menggunakan Kombinasi Metode Lexicon Based dan Naïve Bayes Classifier*. Skripsi. Program Studi Statistika FMIPA UII Yogyakarta.
- Lois Mikael Camdesus. (2023) *Analisis Sentimen Terhadap Konflik Rusia-Ukraina Di Media Sosial Twitter Menggunakan Algoritma Naive Bayes*. Skripsi thesis, Universitas Pembangunan Nasional Veteran Jakarta.
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1. <https://doi.org/10.21580/wjit.2019.1.1.3915>.
- Nasrullah, Rulli. 2015. *Media Sosial; Perspektif Komunikasi, Budaya, dan Sosioteknologi*. Bandung: Simbiosis Rekatama Media.
- Plano, Jack C. dan Roy Olton. 2011. *Kamus Hubungan Internasional (Terjemahan Wawan Juanda)*. Bandung: Putra A Bardin. [Buku]
- Prabowo, W. A., & Wiguna, C. (2021). Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM. *Jurnal Media Informatika Budidarma*, 5(1), 149. <https://doi.org/10.30865/mib.v5i1.2604>.
- R. T. Handayanto and H. Herlawati, "Prediksi Kelas Jamak dengan Deep Learning Berbasis Graphics Processing Units,"

Jurnal Kajian Ilmiah, vol. 20, no. 1, pp. 67–76, Jan. 2020, doi: 10.31599/jki.v20i1.71.

- S. A. Putra and A. Wijaya. 2023. "Analisis Sentimen Artificial Intelligence (AI) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based," Jurnal Sistem & Teknologi Informasi Komunikasi, vol. 7, no. 1, pp. 21, Desember 2023.
- Samsir, et al., (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 5, Nomor 1, Januari 2021, Page 157-163.
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. F. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. INSYST: Journal of Intelligent System and Computation.
- Sonawanse, S. & V. A. K. (2016). *Teknik Analisis Sentimen Data Twitter: Survei*. <http://ai.stanford>.
- Srinath, K. R. (2017). Python – The Fastest Growing Programming Language. International Research Journal of Engineering and Technology, 354–357.
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). Jurnal Sistem Informasi Dan Informatika (Simika), 5(1), 59-67. <https://doi.org/10.47080/Simika.V5i1.1411>.
- Taufiqurrahman, F., Faraby, S. Al, & Purbolaksono, M. D. (2021). Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi Square dan SVM. E-Proceeding of Engineering, 8(5), 10650–10659.

- V. George, S Antonia and G. Dimitros, "A Faceted Crawler for the Twitter Service", WISE 2014, Oc 12-14, 2014, Thessaloniki, Greece
- Yerzi, F. S., & Sibaroni, Y. (2021). Analisis Sentimen Terhadap Kebijakan Pemerintah Dalam Menangani COVID-19 Dengan Pendekatan Lexicon Based.
- Yuniarti, W. D., Faiz, A. N., and Setiawan, B. (2020). Identifikasi Potensi Keberhasilan Studi Menggunakan Naïve Bayes Classifier. *Walisongo Journal of Information Technology*, Vol. 2 No. 1 (2020): 1 – 12.
- Zhang, L., & Liu, B. (2016). Sentiment analysis and opinion mining. *Encyclopedia of Machine Learning and Data Mining*, 1-10. DOI 10.1007/978-1-4899-7502-7 907-1
- Zidan, M (2022). Analisis Sentimen Kenaikan Harga Bahan Bakar Minyak (BBM) Berdasarkan Respon Pengguna Media Sosial Twitter di Indonesia Menggunakan Metode Naive Bayes. Universitas Islam Negeri Walisongo Semarang.

DAFTAR LAMPIRAN

Lampiran 1 : Dokumen Hasil Crawling Data

No	Username	Tweet
1	FadelaGintara	Dan dia tidak boleh ada insya Allah Palestina akan merdeka saya selalu mendukung Hamas gempur Israel satu demi satu atau perlu semuanya dibantai sampai habis
2	bangwand22	hamas jng mau gencatan senjata habisin aja zionis sampe gak tersisa, sampe palestina menjadi negara merdeka & berdaulat spt semula, bani israel kaum yg sulit dipercaya, terkenal keculasan n kelicikannya.
3	bawakebaikan_	Misal kalah mereka minta genjatan senjata apakah israel terusir? Ini harusnya moment kemenangan hamas utk merdeka, genjatan senjata smpai mereka terusir dari palestina.
4	Yelnita6	IDF bangsat ,teroris ,laknatullah, biadap, perampas tanah org palestina , negara israel dihapur dr bumi ini ,pembangkang penjahat perang semga Allah mnghancurkan negara israel,Allah pasti akan melindungi mujahidin hamas, utk merdeka kan Palestina Aamin ya robbal alamin
5	TitikTemu00	bilang Palestina dan hamas harus berhenti begini begitu, harus deal dealan dengan Israel, Palestina bakal sulit merdeka seutuhnya. Ga berani tegas sebut Israel penjajah malah ngira kita menghina Yahudi. Padahal Israel emang udah hina dan keji kelakuannya

6	NafiSongs	Semoga roket Hamas dan Al Qassam meleset terkena warga Israel.. do'a ku untuk palestina merdeka.
7	Ravedon77	Hoax banget lu, orang2x kurdi itu dibantai sama saddam husein, ISIS, erdogan dan iran, kurdistan bisa merdeka di support amrik dan israel, mamam tuh hamas palestina ?
8	rio_boniek	Lu pilih mana? Israel musnah tapi Hamas juga musnah dibanding Israel dan palestina damai sama? merdeka tanpa ada teror satu sama lain?
9	rio_boniek	kalo rujukannya masa lalu, patokannya kapan? masa penjajahan romawi, penjajahan ottoman, apa inggris? gunakan otakmu dengan bener gw juga mendukung palestina merdeka dan perdamaian permanen, tp gw juga melihat kasus ini dengan fair. hamas melakukan kejahatan perang juga israel
10	soala_go2	Perlu diketahui, meskipun Palestina sdh merdeka th 1988 tapi Pemerintah Palestina tdk mampu mengelola negara, air, listrik, internet dll semua dari Israel. Bahkan tentara dan polisi nya jg tdk mampu menjaga keamanan.

.....

1571	wahyuhe6266	Mengapa Arab dan yordan tidak minta israel untuk tidak melanjutkan perang dan menyerah pada hamas agar palestina merdeka, bukankah
------	-------------	--

		arab dan yordan orang islam sesama rakyat palestine
1572	ocamatchalatte	Aku percaya takdirMu berakhir indah ya Allah. Aku percaya semua ini happy ending. Palestina akan merdeka akan berada di surga Mu Israel akan kalah dan membusuk di neraka Tapi aku kasihan sama orang Palestina ya Allah ya Tuhanku
1573	yaniarsim	Alhamdulillah mie instant gak diboikot Tetap boikot franchise yang dukung zionist israel Sampai Palestina Merdeka
1574	lavenderoos	amin tapi dengan mereka masih ada sampe sekarang dan israel gagal terus bikin mereka pergi dari palestina tu a testament deh mereka dilindungi allah banget semoga kita semua msh idup ya pas mereka merdeka nanti

**Lampiran 2 : Dokumen Setelah Dilakukan Pelabelan
Oleh Dua Dhli Dahasa Indonesia**

No	username	tweet	Ahli 1	Ahli 2	ket
1.	FadelaGintara	Dan dia tidak boleh ada insya Allah Palestina akan merdeka saya selalu mendukung Hamas gempur Israel satu demi satu atau perlu semuanya dibantai sampai habis	1	1	1
2.	bangwand22	hamas jng mau gencatan senjata habisin aja zionis sampe gak tersisa, sampe palestina menjadi negara merdeka & berdaulat spt semula, bani israel kaum yg sulit dipercaya, terkenal keculasan n kelicikannya.	1	1	1
3.	bawakebaikan_	Misal kalah mereka minta genjatan senjata apakah israel terusir? Ini harusnya moment kemenangan hamas utk merdeka, genjatan senjata smpai mereka terusir dari palestina.	1	1	1
4.	Yelnita6	IDF bangsat ,teroris ,laknatullah, biadap, perampas tanah org palestina , negara israel	1	1	1

		dihapur dr bumi ini ,pembangkok penjahat perang semga Allah mnghancurkan negara israel,Allah pasti akan melindungi mujahidin hamas, utk merdeka kan Palestina Amin ya robbal alamin			
5.	TitikTemu00	bilang Palestina dan hamas harus berhenti begini begitu, harus deal dealan dengan Israel, Palestina bakal sulit merdeka seutuhnya. Ga berani tegas sebut Israel penjajah malah ngira kita menghina Yahudi. Padahal Israel emang udah hina dan keji kelakuannya	1	1	1

.....

1572	ocamatchalatte	Aku percaya takdirMu berakhir indah ya Allah. Aku percaya semua ini happy ending. Palestina akan merdeka akan berada di surga Mu Israel akan kalah dan membusuk di neraka Tapi aku kasihan sama orang Palestina ya Allah ya Tuhanku	1	1	1
------	----------------	---	---	---	---

1573	yaniarsim	Alhamdulillah mie instant gak diboikot Tetap boikot franchise yang dukung zionist israel Sampai Palestina Merdeka	1	1	1
1574	lavenderoos	amin tapi dengan mereka masih ada sampe sekarang dan israel gagal terus bikin mereka pergi dari palestina tu a testament deh mereka dilindungi allah banget semoga kita semua msh idup ya pas mereka merdeka nanti	1	1	1

Validasi Ahli Pertama


Ari Wahyu Wibowo, S.Pd.

Validasi Ahli kedua


Aura Mulia Arum F., S.S.

Lampiran 3 : Penambahan Data Pendukung (Indonesia lexion)

No	Slang	Formal
1	woww	wow
2	aminn	amin
3	met	selamat
4	netaas	menetas
5	keberpa	keberapa
6	eeeehhhh	eh
7	kata2nyaaa	kata-katanya
8	hallo	halo
9	kaka	kakak
10	ka	kak

.....

15060	ditwtr	di twitter
15061	angota	anggota
15062	mardeka	merdeka
15063	olh	oleh
15064	bgsa	bangsa

Lampiran 4 : Penambahan Data Pendukung (KBBA.text)

No	Slang	Formal
1	7an	tujuan
2	@	di
3	ababil	abg labil
4	abis	habis
5	acc	accord
6	ad	ada
7	adlah	adalah
8	adlh	adalah
9	adoh	aduh
10	ahaha	tertawa

.....

1309	wkwkkw	tertawa
1310	soft	halus
1211	setting	atur
1312	angis	nangis
1313	ahokncc	ahok
1314	almaidahyg	almaidah
1315	istaa	nista
1316	benarjujur	benar
1317	benarjujur	jujur
1318	sayan	sayang
1319	mgkin	mungkin

Lampiran 5 : Dokumen Hasil Klasifikasi

No	Data Uji	Label Aktual	Label Prediksi
1	['juta', 'muslim', 'seluruh', 'dunia', 'doa', 'palestina', 'kali', 'sholat', 'kali', 'total', 'milyar', 'doa', 'palestina', 'merdeka', 'doa', 'salah', 'orang', 'salah', 'kumpul', 'dana', 'salah', 'serang', 'hamas', 'salah', 'israel', 'sakti', 'mohon']	1	1
2	['enggak', 'ngaruh', 'netizen', 'indonesia', 'israel', 'kuasa', 'gaza', 'milik', 'israel', 'palestina', 'merdeka', 'pegang', 'fatah', 'hamas', 'usir', 'gaza']	0	1
3	['bantu', 'saudara', 'palestina', 'boikot', 'produk', 'produk', 'profesional', 'zionis', 'israel', 'danone', 'aqua', 'hal', 'tidak', 'bantu', 'rakyat', 'palestina', 'merdeka']	1	1
4	['alur', 'hamas', 'bela', 'vokasi', 'dulu', 'israel', 'balas', 'korban', 'palestina', 'hamas', 'playing', 'victim', 'internasional']	0	0
5	['zionis', 'israel', 'lama', 'hancur', 'palestina', 'merdeka', 'amin']	1	1
6	['dukung', 'dukung', 'israel', 'anti', 'palestina', 'merdeka', 'halus', 'anti', 'hamas', 'tidak', 'hubung', 'capres', 'zionis', 'islam', 'aneh', 'islam']	1	1
7	['lebay', 'habis', 'orang', 'palestina', 'halang', 'fatah',	1	1

	'plo', 'mesra', 'soviet', 'israel', 'tidak', 'palestina', 'merdeka', 'amerika', 'tidak', 'palestina', 'sekutu', 'soviet', 'dua', 'dana', 'organisasi', 'sayap', 'kanan', 'bentuk', 'hamas']		
8	['merdeka', 'penuh', 'militan', 'palestina', 'doktrin', 'lenyap', 'israel', 'lebaran', 'kuda', 'palestina', 'masuk', 'kendali', 'aman', 'israel', 'masuk', 'ijin', 'israel', 'kalah', 'gaza', 'bos']	0	1
9	['lihat', 'video', 'kak', 'lihat', 'hamas', 'lengkap', 'kontras', 'warna', 'realistis', 'buka', 'dukung', 'israel', 'dukung', 'palestina', 'merdeka', 'dukung', 'pakai', 'hoaks']	1	1
10	['negara', 'palestina', 'negara', 'israel', 'israel', 'jajah', 'malingnumpang', 'tanah', 'gratisibu', 'kota', 'israel', 'bawah', 'tapak', 'kaki', 'palestina', 'ingat', 'selamaya', ', 'merdeka']	1	1

.....

301	['dengar', 'pbb', 'guna', 'lanjut', 'juang', 'palestina', 'merdeka', 'zionis', 'israel', 'musnah', 'muka', 'bumi', 'kemerdekan', 'atas', 'bumi', 'hak', 'bangsa']	1	1
302	['gampang', 'solusi', 'palestina', 'merdeka', 'israel', 'hapus', 'muka', 'bumi']	1	1

303	['bantu', 'papua', 'jajah', 'butuh', 'merdeka', 'benci', 'indonesia', 'merdeka', 'tokoh', 'teroris', 'papua', 'merdeka', 'palestina', 'klaim', 'bela', 'israel']	0	1
304	['bunuh', 'zionis', 'until', 'the', 'hel', 'how', 'many', 'people', 'gaza', 'capek', 'sebar', 'berita', 'genosida', 'agar', 'dunia', 'doa', 'allah', 'agar', 'palestina', 'merdeka', 'boikot', 'dukung', 'dukung', 'israel', 'keji', 'bunuh', 'anak', 'gaza', 'dosa']	1	1
305	['hamas', 'hak', 'label', 'teroris', 'hamas', 'juang', 'negara', 'merdeka', 'zionis', 'zionis', 'jajah', 'palestina', 'palestina', 'aman', 'damai']	1	1
306	['cermin', 'palestina', 'kelola', 'tanah', 'palestina', 'aman', 'kelola', 'israel', 'moga', 'hamas', 'menang', 'palestina', 'merdekaisrael', 'bubar']	1	1

Lampiran 6: Source Code

Crawling Data

```
# Import required Python package
!pip install pandas

!sudo apt-get update
!sudo apt-get install -y ca-certificates curl
gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL
https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o
/etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb [signed-
by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_${NODE_MAJOR}.x
nodistro main" | sudo tee
/etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
```

```
# Crawl Data

filename = 'Data_propo.csv'
search_keyword = 'palestina merdeka hamas
israel until:2023-31-3 since:2023-10-07
lang:id'
```

```

limit = 1600

!npx --yes tweet-harvest@2.2.8 -o "{filename}"
-s "{search_keyword}" -l {limit} --token
{twitter_auth_token}

```

Case Folding

```

#CASE FOLDING
import pandas as pd

file_path = 'Sudah Diberi Label 1574.xlsx'
df = pd.read_excel(file_path)

column_name = 'full_text'

# Melakukan case folding
df[column_name] = df[column_name].str.lower()
output_file_path = 'casefolding_done.xlsx'
df.to_excel(output_file_path, index=False)

print(f"Case folding selesai. Data disimpan di
{output_file_path}")
print(df.head())

```

Cleansing

```

#CLEANSING
import pandas as pd
import re
def clean_text(text):
    text = re.sub(r'\bRT\b', '', text) #
Menghapus kata RT

```

```

    text = re.sub(r'@\w+', '', text) #
Menghapus username
    text = re.sub(r'http\S+|https\S+', '',
text) # Menghapus URL
    text = re.sub(r#\w+', '', text) #
Menghapus hashtag
    text = re.sub(r'(\.)\1+', r'\1', text) #
Menghapus pengulangan karakter
    text = re.sub(r'\d+', '', text) #
Menghapus angka
    text = re.sub(r'^\w\s]', '', text) #
Menghapus tanda baca
    text = text.strip() # Menghapus spasi di
awal dan akhir teks
    return text
def cleanse_excel(input_file, output_file,
text_column):
    df = pd.read_excel(input_file)
    df[text_column] =
df[text_column].apply(clean_text)
    df.to_excel(output_file, index=False)
    return df
input_file = 'Duplicate_done.xlsx'
output_file = 'cleansing.xlsx'
text_column = 'full_text'
df = cleanse_excel(input_file, output_file,
text_column)
print(df.head())

```

Remove Duplicate

```
df = df.drop_duplicates()
df = df.reset_index(drop=True)
df
```

Normalization

```
import pandas as pd

kamus_path = 'lexion_Indonesia.xlsx'
data_path = 'done.xlsx'
kamus = pd.read_excel(kamus_path)
data = pd.read_excel(data_path)

kamus_dict = pd.Series(kamus['Formal'].values,
index=kamus['Slang']).to_dict()
def normalize_text(text):
    words = text.split()
    normalized_words = []
    for word in words:
        normalized_words.append(kamus_dict.get
(word, word))
    return ' '.join(normalized_words)

data['full_text'] =
data['full_text'].apply(normalize_text)

output_file_path = 'done 4.xlsx'
data.to_excel(output_file_path, index=False)

# Menampilkan beberapa baris pertama dari data
hasil normalisasi
print(data.head())
```

```

import pandas as pd

kamus_path = 'kbba.xlsx'
data_path = 'done 4.xlsx'
kamus = pd.read_excel(kamus_path)
data = pd.read_excel(data_path)

kamus_dict = pd.Series(kamus['arti'].values,
index=kamus['slang']).to_dict()
def normalize_text(text):
    words = text.split()
    normalized_words = []
    for word in words:
        normalized_words.append(kamus_dict.get
(word, word))
    return ' '.join(normalized_words)

data['full_text'] =
data['full_text'].apply(normalize_text)

output_file_path = 'done 5.xlsx'
data.to_excel(output_file_path, index=False)
print(data.head())

```

Stopword Removal

```

!pip install nlp-id
from nlp_id import StopWord
import pandas as pd
df = pd.read_excel('done 5.xlsx')
from nlp_id import StopWord
stopword = StopWord()

```

```
df['full_text']=df['full_text'].apply(stopword
.remove_stopword)
df['full_text'].head()
```

Stemming

```
!pip install sastrawi
from Sastrawi.Stemmer.StemmerFactory import
StemmerFactory
import pandas as pd
factory = StemmerFactory()
stemmer = factory.create_stemmer()

df = pd.read_excel('hasil_stopword.xlsx')
from nlp_id import StopWord
stopword = StopWord()

df['full_text'] =
df['full_text'].apply(stemmer.stem)
df['full_text'].head()
df.to_excel('hasil_stemming.xlsx',
index=False)
```

Tokenization

```
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize
from google.colab import files

uploaded = files.upload()

for file_name in uploaded.keys():
```

```

df = pd.read_excel(file_name)

def split_word(teks):
    list_teks = []
    for txt in teks.split(" "):
        list_teks.append(txt)
    return list_teks

df['full_text'] =
df['full_text'].apply(split_word)
df['full_text'].head()

```

Split Validation Data

```

from sklearn.model_selection import
train_test_split
X_train, X_test, y_train, y_test =
train_test_split(df['full_text'],
df['sentiment'], test_size=0.1,
random_state=42)

```

TF-IDF

```

from sklearn.preprocessing import LabelEncoder
Encoder = LabelEncoder()
y_train = Encoder.fit_transform(y_train)
y_test = Encoder.fit_transform(y_test)

from sklearn.feature_extraction.text import
TfidfVectorizer
Tfidf_vect = TfidfVectorizer(max_features=500)
Tfidf_vect.fit(df['full_text'])
Train_X_Tfidf = Tfidf_vect.transform(X_train)

```

```
Test_X_Tfidf = Tfidf_vect.transform(X_test)

Train_X_Tfidf.toarray()
```

Naive Bayes Classifier dan Evaluasi

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score,
precision_score, recall_score, f1_score
from sklearn.metrics import
classification_report
from sklearn.metrics import confusion_matrix

clf =
MultinomialNB().fit(Train_X_Tfidf.toarray(),
y_train)
predicted =
clf.predict(Test_X_Tfidf.toarray())
print("Multinomial NB Accuracy:",
accuracy_score(y_test, predicted))
print("Multinomial NB Precision:",
precision_score(y_test, predicted,
average='weighted'))
print("Multinomial NB Recall:",
recall_score(y_test, predicted,
average='weighted'))
print("Multinomial NB F1 Score:",
f1_score(y_test, predicted,
average='weighted'))
print(f'confusion matrix: Skenario 1\n
{confusion_matrix(y_test, predicted)}')
print('-----\n')
```

```
print(classification_report(y_test,  
predicted,zero_division=0))
```

Word Cloud

```
import pandas as pd  
from wordcloud import WordCloud  
import matplotlib.pyplot as plt  
  
file_path = 'sudah 1574.xlsx'  
df = pd.read_excel(file_path)  
print(df.columns)  
text = '  
' .join(df['full_text'].astype(str).tolist())  
wordcloud = WordCloud(width=800, height=400,  
background_color='white').generate(text)  
plt.figure(figsize=(10, 5))  
plt.imshow(wordcloud,  
interpolation='bilinear')  
plt.axis('off')  
plt.show()
```

RIWAYAT HIDUP

Informasi Pribadi

Nama : Nova Rio Redondo
Tempat, tgl lahir : Semarang, 1 November 2000
Agama : Islam
Alamat : Jl. Gunung Jati Utara Raya No 73
RT04/02 Kel. Wonosari, Kec.
Ngaliyan
No. telepon/WA : 08986478043
Email : rioredondo27@gmail.com

Pendidikan

1. Sarjana Teknologi Informasi Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang.
2. SMK Negeri 4 Semarang, Jurusan Teknik Pemesinan.

Semarang, 17 Agustus 2024

Penulis,



Nova Rio Redondo