

**Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia
Era Kepemimpinan Shin Tae-yong pada X
Menggunakan Algoritma Naive Bayes**



Disusun Oleh:

N a m a : Girendra Egie Zuhriwal
NIM : 17523154

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2024**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia
Era Kepemimpinan Shin Tae-yong pada X
Menggunakan Algoritma Naive Bayes**

TUGAS AKHIR



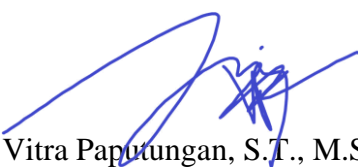
N a m a : Girendra Egie Zuhriwal

NIM : 17523154

الجامعة الإسلامية
الاندونيسية

Yogyakarta, 21 Mei 2024

Pembimbing,


(Irving Vitra Paputungan, S.T., M.Sc., Ph.D.)

HALAMAN PENGESAHAN DOSEN PENGUJI

**Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia
Era Kepemimpinan Shin Tae-yong pada X
Menggunakan Algoritma Naive Bayes**

TUGAS AKHIR

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

Yogyakarta, 29 Juli 2024

Tim Penguji

Irving Vitra Papatungan, S.T., M.Sc.,
Ph.D.



Anggota 1

Hari Setiaji, S.Kom., M.Eng.



Anggota 2

Sheila Nurul Huda, S.Kom., M.Cs.



Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



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

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Girendra Egie Zuhrival

NIM : 17523154

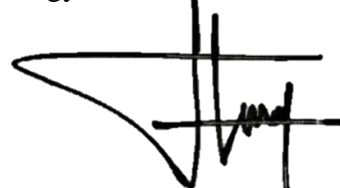
Tugas akhir dengan judul:

**Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia
Era Kepemimpinan Shin Tae-yong pada X
Menggunakan Algoritma Naive Bayes**

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

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

Yogyakarta, 21 Mei 2024



(Girendra Egie Zuhrival)

HALAMAN PERSEMBAHAN

Karya kecil ini penulis persembahkan sepenuhnya kepada kedua orang tua yang terhebat dalam hidup saya, Ayah Amirul dan Ibu Siti Abiba. Kedualah yang tak henti hentinya memberikan doa, usaha, dukungan, nasehat dan hal lainnya yang tak bisa disebutkan satu persatu sehingga saya bisa sampai pada tahap dimana skripsi ini terselesaikan. Karena kalian berdua, hidup terasa begitu mudah dan penuh kebahagiaan. Terima kasih karena selalu menjaga saya dalam doa-doa ayah dan ibu serta selalu membiarkan saya mengejar impian saya apa pun itu. Terimakasih juga untuk kedua adik saya tersayang, adik Icha dan juga adik zevan yang secara langsung maupun tak langsung selalu memberikan semangat serta doa setiap saat, dan juga untuk eyang tercinta yang tak putus putusnya memberikan semangat dan doa yang tak ada hentinya. Semoga kebaikan semuanya dapat dibalas oleh Allah SWT.

Terimakasih untuk Bapak Irving Vitra Papatungan, S.T., M.Sc., Ph.D. selaku dosen pembimbing, dan Seluruh Dosen Informatika Universitas Islam Indonesia atas bimbingan dan arahnya yang selalu memberikan ilmu dan juga inovasi yang tak terbatas yang nantinya akan bermanfaat sebagai bekal untuk kedepannya.

Terimakasih untuk semuanya semoga Allah SWT mengganti semua kebaikan dan juga doa menjadi kebahagiaan keberkahan dan pahala yang berlimpah. Aamiin.

HALAMAN MOTO

"Sesungguhnya Allah tidak mengubah keadaan sesuatu kaum sehingga mereka mengubah keadaan yang ada pada diri mereka sendiri."

Q.S Ar-Ra'd : 11

"Barang siapa yang tidak mensyukuri yang sedikit, maka ia tidak akan mampu mensyukuri sesuatu yang banyak."

(HR. Ahmad)

KATA PENGANTAR

Alhamdulillah, segala puji dan syukur kami panjatkan ke hadirat Allah SWT, karena atas berkat dan rahmat-Nya kita dapat menyelesaikan penyusunan skripsi ini. Skripsi ini disusun untuk memenuhi tugas dan syarat yang ditentukan dalam salah satu persyaratan yang harus dipenuhi untuk menyelesaikan Pendidikan pada jenjang Strata 1 di Jurusan

Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia. Shalawat serta salam semoga senantiasa selalu terlimpah dan tumpahruhan kepada junjungan kita nabi besar yaitu Nabi Muhammad SAW, beserta keluarga dan para sahabatnya, hingga kepada umatnya di akhir zaman,

Aamiin.

Berkaitan dengan penulisan tugas akhir ini yang berjudul " Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia Era Kepemimpinan Shin Tae-yong pada X Menggunakan Algoritma Naive Bayes", yang merupakan salah satu persyaratan untuk kelulusan, semoga hal ini dapat bermanfaat untuk semua pihak dan berguna untuk kedepannya.

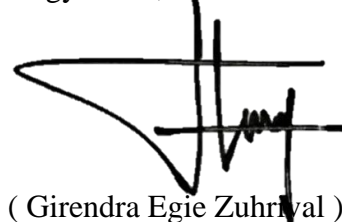
Penulisan skripsi ini tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, kami ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dan membimbing kami dalam penyusunan skripsi ini, yaitu:

- a. Orang tua penulis (Bapak Amirul Mukmin dan Ibu Siti Abiba Ila Ila), serta kedua adik saya (Risma Icha Firdausah dan Rafardan Zevan Alifiandra) yang tidak ada hentinya memberikan doa serta dukungannya selama ini sehingga dapat terselesaikan masa kuliah dan tugas akhir dengan baik.
- b. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Ketua Jurusan Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.
- c. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Ketua Program Studi Informatika - Program Sarjana, Fakultas Teknologi Industri, Universitas Islam Indonesia.
- d. Bapak Irving Vitra Papatungan, S.T., M.Sc., Ph.D. selaku Dosen Pembimbing Skripsi, atas segala bimbingan, kesempatan, dan komentar yang bersifat membangun ke arah yang lebih baik.
- e. Ibu Arrie Kurniawardhani, S.SI., M.Kom. selaku Dosen Pembimbing Akademik yang selalu memberikan arahan dan semangat untuk menjalani perkuliahan.

- f. Seluruh dosen Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia yang banyak memberikan ilmu baru khususnya di bidang teknologi yang semoga dapat berguna dan bermanfaat untuk kedepannya.
- g. Nenek penulis (Ibu Hadi) yang selalu memberikan doa dan semangat yang tiada henti.
- h. Teman teman PIXEL Informatika UII Angkatan 2017, atas segala momen dan suka dukanya selama proses perkuliahan. Semoga makin sukses selalu kedepannya.
- i. Teman Teman serta kru PO MILA Sejahtera, terimakasih atas jasa transportasi kalian untuk trayek Lumajang-jogja dan sebaliknya yang selalu sedia dalam kondisi apapun.
- j. Semua pihak yang telah membantu penulis dengan penuh kesabaran dan keikhlasan yang tidak dapat disebutkan satu persatu. Terimakasih atas segala bantuan yang telah diberikan kepada penulis.

Pada akhirnya, penulis juga menyadari bahwa dalam penyusunan laporan tugas akhir ini masih terdapat banyak kekurangan. Demi perbaikan selanjutnya, maka dari itu, penulis mengharapkan kritik dan saran yang bersifat membangun agar hasil yang didapatkan lebih baik kedepannya. Penulis mengharapkan semoga tugas akhir ini dapat bermanfaat bagi penulis dan semua pihak yang berkenan menelaah di kemudian hari

Yogyakarta, 21 Mei 2024



(Girendra Egie Zuhriyal)

SARI

Sepak bola merupakan salah satu olahraga paling populer di Indonesia dan juga di dunia. Masyarakat Indonesia cukup menggandrungi olahraga tersebut, bisa dilihat dari hasil penelitian yang dilakukan oleh Nielsen Sport, sebanyak 77% penduduk Indonesia memiliki minat dan ketertarikan terhadap dunia sepakbola. Selain itu menurut The Global Digital Football Benchmark, diantara 200 klub terpopuler di dunia, 6 klub sepakbola di Indonesia sudah masuk dalam deretan tersebut, hal ini menandakan bahwa olahraga ini sudah sangat melekat dengan masyarakat Indonesia.

Sementara itu, salah satu media sosial yang cukup populer saat ini adalah X. X adalah salah satu media komunikasi yang diminati oleh masyarakat di dunia. Hal ini dapat dilihat dari peningkatan pengguna X yang tercatat di seluruh dunia.

Penelitian ini nantinya akan dilakukan analisis sentimen Timnas Indonesia pada Kualifikasi Piala Asia U 23 dan Piala Asia U 23 berkaitan dengan performa bermain dan juga kepelatihan Shin Tae Yong untuk mengklasifikasikan tweet opini dan komentar Masyarakat yang diperoleh dari data tweet yang dikumpulkan dari platform X. Data tersebut akan diolah menggunakan algoritma Naive Bayes, yaitu salah satu metode klasifikasi probabilitas yang sederhana, pada metode ini dilakukan perhitungan probabilitas dengan cara melakukan penjumlahan terhadap frekuensi dan kombinasi nilai dari dataset dan menghasilkan tweet yang telah diklasifikasikan menjadi positif atau negatif. Pada tahap sebelum analisis dilakukan beberapa langkah diantaranya cleaning, stemming, case folding, remove stopword, dan tokenization guna untuk meningkatkan nilai akurasi pada tahap analisis sentimen. Total data yang terkumpul pada penelitian kali ini adalah sebanyak, 4113 data, yaitu pada masa kualifikasi dan masa piala asia. Pada proses visualisasi data dilakukan proses pembuatan word cloud guna melihat kata apa saja yang sering digunakan pada data set yang dikumpulkan, beberapa kata yang sering muncul diantaranya adalah, Indonesia, Piala, Pemain, dan Menang. Penggunaan metode naive bayes menghasilkan akurasi model sebesar 90.34 %, memperoleh tingkat akurasi testing sebesar 87.96% dengan perbandingan data latih 90% dan data uji sebesar 10%.

Kata kunci: Timnas, Sentimen, Naive Bayes.

GLOSARIUM

Preprocessing	Tahap awal sebelum masuk pada pemrosesan data yang memiliki tujuan untuk mengolah data mentah yang telah dikumpulkan dari sumber menjadi bentuk informasi yang bersih dan dapat diolah lebih lanjut.
Klasifikasi	Pengelompokan data menggunakan metode yang telah ditentukan kedalam kumpulan kategori yang telah ditentukan.
Sentimen	Sentimen merujuk pada ekspresi atau perasaan seseorang terhadap suatu hal, topik, atau situasi. Dalam konteks analisis sentimen, biasanya merujuk pada evaluasi atau pandangan positif, negatif, atau netral terhadap entitas tertentu.
Labeling	Proses pemberian label pada data berdasarkan kelas yang telah ditentukan
Word Cloud	Alat visualisasi untuk mempermudah pemahaman kata-kata berdasarkan frekuensi kemunculannya.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR.....	vii
SARI	ix
GLOSARIUM	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan	5
BAB II KAJIAN PUSTAKA	6
2.1 Analisis Sentimen	6
2.1.1 Analisis Sentimen di Media Sosial.....	6
2.1.2 Metode Analisis Sentimen di Media Sosial	7
2.2 Naive Bayes	8
2.3 Penelitian Terdahulu	8
2.3.1 Pembahasan	10
BAB III METODOLOGI PENELITIAN	12
3.1 Pengumpulan Data	13
3.2 Preprocessing	18
3.3 Pelabelan Sentimen	20
3.4 Pembobotan TF-IDF	21

3.5	Klasifikasi Naive Bayes	22
3.6	Pengujian Klasifikasi	23
3.7	Visualisasi Word Cloud	25
BAB IV HASIL DAN PEMBAHASAN.....		26
4.1	Pengumpulan Data	26
4.2	Preprocessing	27
4.3	Pelabelan Kelas Sentimen	32
4.3.1	Pada Masa Kualifikasi Piala Asia U 23.....	35
4.3.2	Pada Masa Piala Asia U 23	36
4.4	Word Cloud.....	39
4.4.1	Word Cloud Positif.....	40
4.4.2	Word Cloud Negatif	41
4.4.3	Word Cloud Timnas Menang	41
4.4.4	Word Cloud Timnas Kalah.....	42
4.5	Data Latih dan Data Uji	43
4.6	Naive Bayes	44
4.7	Evaluasi Sistem	46
4.8	Analisis Hasil Sentimen	47
4.8.1	Tindakan Berdasarkan Sentimen.....	48
4.8.2	Kesimpulan Hasil Sentimen	49
BAB V KESIMPULAN DAN SARAN		50
5.1	Kesimpulan	50
5.2	Saran.....	51
DAFTAR PUSTAKA.....		53
LAMPIRAN		56

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	9
Tabel 3.1 Contoh proses cleaning dan case folding.....	19
Tabel 3.2 Contoh proses steaming.....	19
Tabel 3.3 Contoh proses remove stopword.....	20
Tabel 3.4 Contoh proses tokenization.....	20
Tabel 4.1 Hasil setelah dilakukan case folding.....	29
Tabel 4.2 Hasil setelah dilakukan steaming.....	30
Tabel 4.3 Hasil setelah dilakukan tokenize.....	30
Tabel 4.4 Hasil klasifikasi.....	34
Tabel 4.5 Data klasifikasi kualifikasi piala asia.....	35
Tabel 4.6 Data klasifikasi piala asia	36
Tabel 4.7 Sentimen Indonesia menang	37
Tabel 4.8 Sentimen Indonesia kalah	38
Tabel 4.9 Hasil sentimen.....	47

DAFTAR GAMBAR

Gambar 2.1 Metode analisis sentimen	7
Gambar 3.1 Diagram proses penelitian.....	12
Gambar 3.2 Proses crawling data.....	15
Gambar 3.3 Inisiasi dan instalasi	16
Gambar 3.4 Proses crawling data.....	17
Gambar 3.5 Proses file csv.....	17
Gambar 3.6 Melihat data yang didapatkan	17
Gambar 3.7 Contoh data yang berhasil dikumpulkan dari X.....	18
Gambar 4.1 Hasil crawling data.....	26
Gambar 4.2 Proses cleaning.....	27
Gambar 4.3 Proses cleaning.....	27
Gambar 4.4 Hasil proses cleaning.	28
Gambar 4.5 Proses steaming data	29
Gambar 4.6 Proses tokenize, transform case, filter stopword.....	31
Gambar 4.7 Hasil proses preprocessing.....	31
Gambar 4.8 Proses labeling otomatis.....	33
Gambar 4.9 Proses perbandingan labeling manual dan otomatis	34
Gambar 4.10 Diagram perbandingan sentimen positif dan negatif	35
Gambar 4.11 Diagram perbandingan sentimen positif dan negatif	36
Gambar 4.12 Diagram perbandingan sentimen positif dan negatif timnas menang	37
Gambar 4.13 Diagram perbandingan sentimen positif dan negatif timnas kalah	38
Gambar 4.14 Hasil visualisasi word cloud kualifikasi Piala Asia	39
Gambar 4.15 Hasil visualisasi word cloud Piala Asia	39
Gambar 4.16 Hasil visualisasi word cloud positif	40
Gambar 4.17 Hasil visualisasi word cloud negatif	41
Gambar 4.18 Hasil visualisasi word cloud timnas menang	42
Gambar 4.19 Hasil visualisasi word cloud timnas kalah	43
Gambar 4.20 Proses Naive Bayes	44
Gambar 4.21 Hasil proses data uji	45
Gambar 4.22 Hasil nilai akurasi.....	45
Gambar 4.23 Proses tahap evaluasi.	46
Gambar 4.24 Confusion matrix algoritma Naive Bayes.....	46
Gambar 4.25 Hasil pengujian data baru.....	47

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sepak bola merupakan salah satu olahraga terpopuler, tidak hanya di Indonesia namun juga di seluruh dunia. Sebelum Indonesia merdeka, sepak bola sudah jauh lebih dikenal dan Indonesia mengenal sepakbola modern sejak datangnya orang Belanda yang mendirikan perusahaan dagang VOC pada abad ke -16. Sepak bola juga digunakan oleh salah satu pemuda Indonesia yaitu Soeratin sebagai wadah menyemai nasionalisme dikalangan pemuda untuk menentang penjajahan Hindia Belanda. Ir. Soeratin Sosrosoegondo yang merupakan pelopor berdirinya PSSI merupakan orang Indonesia yang memiliki jiwa nasionalis tinggi. Sebagai seorang pemuda yang gemar bermain sepak bola, Soeratin memakai sepakbola sebagai wadah menyemai nasionalisme dikalangan pemuda untuk menentang penjajahan Hindia Belanda. Sementara itu berdasarkan temuan studi Nielsen Sport, hingga 77% masyarakat Indonesia tertarik pada sepak bola. Selain itu menurut The Global Digital Football Benchmark, diantara 200 klub terpopuler di dunia, 6 klub sepakbola di Indonesia sudah masuk dalam deretan tersebut, hal ini menunjukkan tingginya ketertarikan masyarakat di negara ini terhadap olahraga tersebut (Arifianto, 2017). Dengan hal itu pula kepedulian dan harapan terhadap kemajuan dan prestasi terhadap sepak bola khususnya di kanca Timnas Indonesia sangatlah tinggi (Astiningrum et al., 2020).

Sementara itu sebagai olahraga paling populer di dunia, suporter sepak bola muncul sebagai elemen penting yang tidak dapat dipisahkan. Menurut Soekanto dalam karyanya, suporter sepak bola adalah individu atau sekelompok orang yang hadir atau memberikan dukungan pada sebuah tim selama pertandingan sepak bola. Dalam konteks sepak bola, suporter memiliki keterkaitan yang erat dengan dukungan yang didasarkan pada perasaan cinta dan fanatisme terhadap tim tertentu. Fanatisme ini tercermin dalam kebahagiaan mereka ketika tim berhasil memenangkan pertandingan, namun sebaliknya, dapat menyebabkan kemarahan yang mendalam jika hasilnya tidak sesuai harapan. Dengan adanya suporter, tidak hanya mendukung tim mereka di lapangan, tetapi juga berperan dalam menjaga keberlanjutan keuangan tim sepakbola, melalui penjualan tiket, merchandise, dan dukungan finansial lainnya. Oleh karena itu, sebuah tim sepakbola sangat bergantung pada pendukungnya,

sehingga kepuasan pendukung ataupun pendapat dari suporter sebuah klub perlu diperhatikan (Swandy et al., 2019).

Jika dilihat secara umum, banyak faktor yang mempengaruhi performa bermain suatu tim, seperti adanya permasalahan internal, performa pemain, dan juga faktor dari pelatih (Astiningrum et al., 2020). Dalam permainan sepakbola, adanya pelatih yang berkualitas tentunya sedikit banyak dapat mempengaruhi performa dalam suatu tim. Bahkan menurut Menurut Luis Felipe Scolari (Pelatih dari Brazil), peran pelatih itu sangat vital dalam sepak bola, sebab mereka berperan besar dalam proses berlatih, persiapan, dan menentukan taktik pada saat bertanding. Hal ini tentunya juga mendapat perhatian dari suporter tentang siapa dan apakah pelatih tersebut cocok untuk Timnas Indonesia.

Kepopuleran sepak bola tentu tidak dapat dipisahkan dari adanya media sosial. Media sosial seakan menjadi alat pemantau, dengan berbagai kemampuan dalam mengemas suatu pemberitaan, media mampu membentuk opini dan perilaku publik (Septian et al., 2018). Salah satu media sosial yang populer saat ini adalah X atau twitter. Media sosial X adalah salah satu media komunikasi yang diminati oleh masyarakat di dunia (Astiningrum et al., 2020). Hal ini dapat dilihat dari peningkatan pengguna yang tercatat di seluruh dunia. Di Indonesia X memiliki jumlah pengguna aktif sebesar 18,45 juta per Januari pada tahun 2022 dan sebagian besar pengguna mengakses melalui perangkat mobile, yaitu sebesar 82 persen (Astiningrum et al., 2020). Jumlah pengguna yang banyak juga menimbulkan berbagai tanggapan yang berbeda-beda pada setiap orang (Setyobudi et al., 2018). Sebagai salah satu platform media sosial paling dominan di seluruh dunia, X atau twitter memungkinkan pengguna untuk berbagi apapun yang ada dalam pikiran mereka. Dalam penelitian yang dilakukan oleh Villavicencio (2021), dijelaskan bahwa X memberikan kesempatan bagi individu untuk sepenuhnya mengekspresikan diri mereka pada waktu yang tepat. Saat menggunakan X dalam situasi yang lebih privasi, individu dapat dengan mudah mengekspresikan diri mereka dengan cara yang lebih jujur. Oleh karena itu, X dianggap sebagai platform yang efektif untuk menganalisis sentimen publik yang sebenarnya. Untuk itu diperlukan teknik analisis sentimen untuk mempermudah melihat tanggapan yang diberikan oleh masyarakat dan dapat diklasifikasikan menjadi opini positif ataupun negatif. Nantinya hasil analisis sentimen ini dapat digunakan sebagai bahan evaluasi kedepannya agar dapat melihat bagaimana respon dan harapan masyarakat terhadap perkembangan Timnas Indonesia baik itu yang sentimen yang bernilai positif atau sentimen yang bernilai negatif.

Penggunaan algoritma Naive Bayes dipilih karena pada beberapa penelitian yang dilakukan dapat mengolah data yang berjumlah banyak dan memiliki akurasi ketepatan yang cukup tinggi. Berdasarkan penelitian yang dilakukan oleh (Astiningrum et al., 2020), hasil pengujian algoritma Naïve Bayes menggunakan 80% data training dan 20% data testing didapatkan nilai akurasi sebesar 84%, dan saat menggunakan 90% data training dan 10% data testing didapatkan nilai akurasi sebesar 87%. Dengan itu membuktikan bahwa tingkat akurasi yang cukup tinggi dapat diperoleh dengan penggunaan algoritma Naïve Bayes.

Berdasarkan penjelasan diatas, penelitian ini akan dilakukan analisis sentimen Timnas Indonesia berkaitan dengan performa bermain pada era kepelatihan Shin Tae-yong untuk mengklasifikasikan tweet opini dan komentar masyarakat. Data tersebut akan diolah menggunakan algoritma Naive Bayes dan menghasilkan tweet yang telah diklasifikasikan menjadi positif atau negatif. Pengujian akan dilakukan berdasarkan hasil klasifikasi. Pengujian bertujuan untuk mengetahui tingkat akurasi dari klasifikasi. Hasil pembahasan diharapkan akan memberikan kemudahan bagi pihak terkait untuk melihat reaksi masyarakat terhadap performa Timnas Indonesia pada era kepelatihan Shin Tae-yong.

1.2 Rumusan Masalah

Berdasarkan penjelasan uraian latar belakang, maka rumusan masalah penelitian ini adalah bagaimana melakukan analisis sentimen Timnas sepak bola Indonesia pada era kepelatihan Shin Tae-yong berdasarkan tweet masyarakat di X menggunakan algoritma Naive Bayes Classifier?

1.3 Batasan Masalah

Berdasarkan penelitian yang dilakukan supaya lebih terarah dan sesuai dengan yang dimaksudkan, batasan masalah pada penelitian ini adalah sebagai berikut:

- a. Proses pengambilan data dilakukan di X dan hanya berbahasa Indonesia.
- b. Kriteria data yang diambil hanya sentimen yang berhubungan dengan Timnas pada masa kepelatihan Shin Tae-yong.

1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah melakukan analisis sentimen pada X terkait dengan performa Timnas pada era kepelatihan Shin Tae-yong dengan algoritma Naive Bayes

Classifier. Dengan mengetahui hal tersebut diharapkan kedepannya dapat dijadikan salah satu bahan evaluasi terhadap performa Timnas dan juga kepelatihannya.

1.5 Manfaat Penelitian

Pada penelitian ini diharapkan dapat mengetahui kecenderungan sentimen terhadap Timnas Indonesia pada era kepelatihan Shin Tae Yong di media sosial X, serta dapat dijadikan gambaran untuk melakukan penelitian lainnya dengan topik yang berbeda.

1.6 Metodologi Penelitian

Metodologi penelitian adalah bagian penting dalam sebuah penelitian yang menjelaskan secara rinci tentang pendekatan, teknik, dan prosedur yang digunakan untuk mengumpulkan, menganalisis, dan menginterpretasikan data. Bagian ini biasanya mencakup berbagai elemen yang memungkinkan penelitian dilaksanakan secara sistematis dan terukur. Berikut adalah elemen yang terdapat dalam metodologi penelitian:

a. Studi Literatur

Studi literatur adalah untuk mencari teori – teori yang relevan terhadap penelitian sejenis yang pernah dilakukan sebelumnya yang berkaitan dengan penemuan sebelumnya, karena penelitian bertujuan untuk mendapatkan solusi dari permasalahan yang sedang dihadapi.

b. Sumber Data

X atau twitter merupakan sumber data untuk melakukan analisis sentimen ini, dan akan difokuskan pada opini masyarakat kepada Timnas Indonesia.

c. Pengumpulan Data

Setelah mendapatkan data berdasarkan opini pengguna twitter tentang Timnas Indonesia pada kualifikasi Piala Asia U-23 langkah selanjutnya, yaitu mengambil data ulasan yang sudah diberikan pengguna X dengan cara men crawling tentang Timnas Indonesia.

d. Pengolahan Data

Data yang didapat merupakan data yang langsung diambil dari tweet pengguna di X. Data tersebut perlu dilakukan pengolahan agar menjadi data yang mudah digunakan dalam proses sentiment analysis. Tweet akan mengalami penyeleksian dan perbaikan kata - kata sehingga tweet menjadi lebih bagus untuk proses klasifikasi.

e. Naive Bayes

Pada Proses ini, digunakan metode Naive Bayes untuk mengklasifikasikan komentar tweet yang telah dikumpulkan. Prosesnya melibatkan pembuatan model Naive Bayes dengan

menggunakan data latih yang telah disiapkan sebelumnya. Setelah model dibuat, digunakan untuk memprediksi hasil dari data uji yang telah diproses.

f. Analisis

Setelah melakukan implementasi Naive Bayes dan memprediksi sentimen dari data tweet yang telah dikumpulkan, dilakukan analisis terhadap hasil yang diperoleh. Analisis ini melibatkan evaluasi kinerja model Naive Bayes, termasuk akurasi, presisi, dan recall. Selain itu, juga dilakukan analisis terhadap sentimen yang dominan dalam data, serta identifikasi pola-pola atau tren yang muncul dalam opini masyarakat terhadap Timnas Indonesia berdasarkan pada data yang telah diolah.

1.7 Sistematika Penulisan

Beberapa pembahasan dalam penulisan ini secara umum dibagi menjadi lima bab yang dipaparkan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisikan pendahuluan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini secara umum berisikan berbagai macam teori yang mendukung dan menjadi dasar penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan berbagai macam analisa data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan penjabaran hasil penelitian, dan juga analisis data.

BAB V PENUTUP

Bab ini berisikan berbagai macam kesimpulan terhadap penelitian yang sudah dilakukan dan saran kepada pembaca apabila tertarik untuk melakukan penelitian serupa.

BAB II

KAJIAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen atau penggalian opini adalah proses mengidentifikasi dan mengkategorikan emosi atau pendapat pengguna terhadap suatu layanan, produk, acara, atau hal lain, apakah bersifat positif, negatif, atau netral. Menurut Chamekh (2020) menjelaskan bahwa analisis sentimen merupakan penggalian informasi tentang suatu entitas dan secara otomatis mengidentifikasi subjektivitas entitas tersebut. Tujuannya adalah untuk menentukan apakah teks yang dihasilkan oleh pengguna menyampaikan pendapat positif, negatif, atau netral. Klasifikasi sentimen dapat dilakukan pada tiga tingkat ekstraksi: level aspek atau fitur, level kalimat, dan level dokumen (Chamekh et al., 2020). Menurut Mehta dkk (2020), analisis sentimen adalah tugas untuk memahami apakah pendapat yang diberikan bersifat positif atau negatif secara umum. Ini melibatkan klasifikasi data ke dalam kategori seperti positif (makna yang baik), negatif (makna yang buruk), atau netral. Sentimen dapat diekspresikan dengan berbagai cara, termasuk melalui emosi, penilaian, visi, atau pandangan orang. Emosi dapat diekspresikan dalam teks melalui kata-kata yang dipilih untuk mengungkapkan perasaan tertentu. Pembaca kemudian menafsirkan konten tersebut berdasarkan kondisi pikirannya dan kemampuannya untuk menganalisis suatu hal (Mehta et al., 2020). Drus dkk (2019) menyatakan bahwa analisis sentimen juga dapat meningkatkan kesadaran terhadap keamanan data dan potensi pelanggaran keamanan. Selain itu, dapat digunakan sebagai panduan bagi perusahaan dalam merespons pelanggaran keamanan untuk membentuk persepsi publik (Drus et al., 2019). Dengan demikian, analisis sentimen tidak hanya mencakup pemahaman terhadap pendapat individu, tetapi juga dapat berperan dalam memitigasi risiko dan meningkatkan respons terhadap keamanan data (Drus et al., 2019).

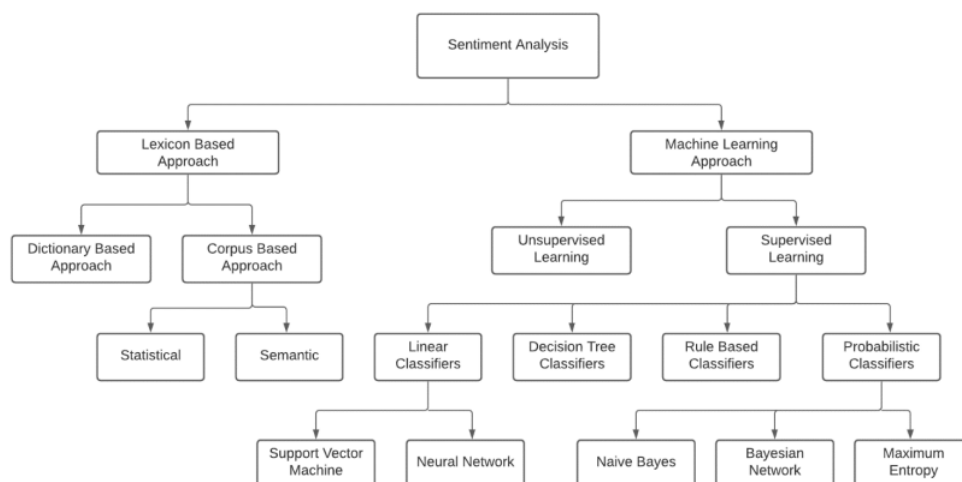
2.1.1 Analisis Sentimen di Media Sosial

Pada saat ini, media sosial telah menjadi platform utama bagi hampir semua orang untuk menyampaikan respons dan pendapat terhadap berbagai hal. Media sosial sekarang dianggap sebagai kebutuhan umum bagi seluruh masyarakat, dan pemanfaatannya sebagai sumber data yang dapat diolah melalui analisis sentimen menjadi pilihan yang sangat tepat. Dalam penelitian yang dilakukan oleh Rachman dkk (2020), disebutkan bahwa penggunaan data dari

media sosial menjadi salah satu alternatif baru sebagai pengganti sumber data dari survei tradisional. Saat ini, pengumpulan pendapat dan respons melalui survei tradisional dianggap sulit dilakukan karena proses yang kompleks dan memakan waktu yang relatif lama. Sebaliknya, pengumpulan data melalui media sosial dinilai lebih efisien dalam hal biaya, waktu, dan tenaga. Penelitian sejenis mengenai analisis sentimen melalui media sosial telah banyak dilakukan dan memberikan hasil yang memuaskan (Rachman et al., 2020). Chamekh (2020) dalam penelitiannya menyatakan bahwa analisis sentimen di jejaring sosial seperti Twitter atau Facebook telah menjadi alat yang efektif untuk menyelidiki pendapat pengguna dengan berbagai aplikasi. Bidang penelitian ini sedang populer karena data opini yang dapat ditemukan oleh pengguna lainnya, terutama ulasan terkait berbagai layanan yang bermanfaat dalam kehidupan sehari-hari. Sebagian besar data opini ini disimpan dalam bentuk digital (Mehta et al., 2020).

2.1.2 Metode Analisis Sentimen di Media Sosial

Pada era ini, terdapat beragam metode yang digunakan untuk analisis sentimen. Dalam penelitian yang dilakukan oleh Sharma dkk (2020), metode tersebut dibagi menjadi dua kategori utama, yaitu metode analisis sentimen berbasis pembelajaran mesin (machine learning) dan berbasis kosakata (lexicon). Lebih lanjut, metode analisis sentimen berbasis mesin dan berbasis kosakata masing-masing dibagi lagi menjadi dua sub kategori, yaitu supervised learning dan unsupervised learning pada analisis sentimen berbasis mesin, serta dictionary-based dan corpus-based pada analisis sentimen berbasis kosakata (Sharma et al., 2020). Penjelasan metode analisis sentiment dalam bentuk diagram dapat dilihat pada gambar.



Gambar 2.1 Metode analisis sentimen

Sumber: Sharma (2020)

2.2 Naive Bayes

Naive Bayes adalah salah satu algoritma yang paling populer digunakan untuk data mining karena sifatnya yang ramah pengguna, waktu pemrosesan yang cepat, struktur yang sederhana, dan efisiensi yang sangat baik (Nurun et al., 2016). Konsep Naive Bayes diperkenalkan oleh ilmuwan Inggris Thomas Bayes. Menurut Salmi dkk (2019), Naive Bayes Classifier merupakan teknik prediksi berdasarkan probabilitas sederhana dan penerapan teorema Bayes dengan asumsi independensi yang kuat. Oleh karena itu, model ini mampu mencapai akurasi klasifikasi yang lebih tinggi dengan kompleksitas yang lebih rendah. Sedangkan Nugroho (2018) menyatakan dalam penelitiannya bahwa Naive Bayes classifier merupakan algoritma yang digunakan untuk mengklasifikasikan data uji ke dalam kategori paling sesuai melalui pencarian probabilitas tertinggi. Pengklasifikasi Naive Bayes diketahui berkinerja lebih baik dibandingkan beberapa metode klasifikasi lainnya karena asumsi independensi yang kuat untuk setiap kondisi atau peristiwa, kesederhanaan dan kemudahan pembuatannya, serta kemampuan implementasinya untuk kumpulan data yang besar (Salmi et al., 2019). Selain penerapannya dalam analisis sentimen, algoritma Naive Bayes banyak digunakan untuk tujuan lain, seperti deteksi spam dalam email, pengelompokan email otomatis, kategorisasi dokumen, dan penentuan prioritas email (Andika et al., 2019).

2.3 Penelitian Terdahulu

Penulisan ini bertujuan untuk meninjau perbandingan literatur yang diperoleh dari sisi sumber data, metode yang digunakan, serta hasil yang dicapai. Setiap literatur memiliki permasalahan, pendekatan, dan hasil yang berbeda, yang memberikan gambaran tentang penelitian berikutnya, terutama yang berkaitan dengan analisis sentimen dan tim nasional sepak bola Indonesia. Dengan mengkaji berbagai literatur tersebut, diharapkan dapat diperoleh gambaran yang lebih komprehensif mengenai pendekatan yang efektif dan tantangan yang mungkin dihadapi dalam penelitian berikutnya, terutama yang berkaitan dengan analisis sentimen serta topik khusus seperti tim nasional sepak bola Indonesia.

Literatur-literatur ini dianalisis untuk mengidentifikasi keunggulan dan kelemahan metode yang digunakan. Dengan demikian, tinjauan ini tidak hanya memberikan landasan teoretis tetapi juga panduan dalam memilih dan menyesuaikan metodologi yang tepat. Berikut ini adalah tabel yang menyajikan kajian pustaka dari penelitian terdahulu yang telah dianalisis:

Tabel 2.1 Penelitian terdahulu

No	Metode	Hasil Yang Diperoleh	Referensi
1	Naive Bayes Classifier	Nilai akurasi, error, recall, dan precision yang didapat masing-masing sebesar 88,89%, 19,18%, 96,97%, dan 69,57% untuk Mobile Legends sedangkan Arena of Valor memiliki nilai akurasi, error, recall, dan precision masing-masing sebesar 39,02%, 60,98%, 88,89% dan 28,74%.	(Simorangkir et al., 2018)
2	Naive Bayes Classifier	Pada penelitian ini dilakukan tiga kali pengujian dengan data set yang berbeda. Dari tiga kali pengujian didapatkan akurasi tertinggi yaitu 87%, precision positif tertinggi yaitu 93%, precision negatif tertinggi yaitu 85%, recall positif tertinggi yaitu 75%, dan recall negatif tertinggi yaitu 98% pada persentase 90% data training dan 10% data testing.	(Astiningrum et al., 2020)
3	Naive Bayes Classifier	Pengklasifikasi Naive Bayes diperoleh akurasi sebesar 83% dan rata-rata harmonisa sebesar 90,713%.	(Kencana et al., 2014)
4	Naive Bayes dan Support Vector Machine	Dari proses analisis sentimen yang telah dilakukan, hasil yang didapat adalah sentimen netral menjadi mayoritas dengan jumlah yang cukup dominan yaitu 545 buah tweet, kemudian disusul oleh sentimen positif yang berjumlah 274, dan sentimen negatif 143.	(Astuti., 2018)
5	Decision Tree dan Support Vector Machine	Hasil komposisi data terbaik untuk melakukan pengujian adalah 80%:20% dengan mendapatkan hasil nilai akurasi 87.45%, precision 87.72%, recall 91.74% dan F1-Score 89.69% pada Decision Tree dengan TF-IDF, sedangkan untuk Support Vector Machine dengan TF-IDF komposisi data terbaik adalah 80%:20% mendapatkan hasil nilai akurasi 94.36%, precision 96.78%, recall 94.30% dan F1-Score 95.53%.	(Science., 2020)
6	Pembobotan TF-IDF dan K-Nearest Neighbor	Dilakukan percobaan untuk mencari model KNN dengan akurasi terbaik menggunakan range nilai k=1 hingga k=30 yang adalah bilangan ganjil, didapatkan akurasi optimal pada k=23 dengan akurasi sebesar 79,99% dan error rate sebesar 20,01%.	(Nugroho., 2019)
7	Support Vector Machine	Jika berdasarkan pendekatan ekstraksi fitur SVM dengan pendekatan count vector, TF-IDF word level, TF-IDF ngram level dan TF-IDF char level. Dalam skenario ini nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur count vector dan TF-IDF ngram level dengan nilai akurasi 59% pada instagram.	(Kurniasari., 2021)

Berdasarkan data diatas untuk pemilihan metode yang digunakan cenderung fokus pada dua metode yaitu SVM (Support Vector Machine) dan Naive Bayes Classifier dengan hasil keakuratan yang cukup tinggi.

2.3.1 Pembahasan

Dalam pemilihan metode, hasil literatur menunjukkan dua metode yang populer, yaitu Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC), dengan sedikit unggulnya akurasi hasil dari Naïve Bayes Classifier.

Pada penelitian dengan judul “Sentimen analisis twitter terhadap penyelenggaraan gojek traveloka liga 1 Indonesia” dilakukan dengan metode Support Vector Machine (Astuti., 2018). Dari proses analisis sentimen yang telah dilakukan, hasil yang didapat adalah sentimen netral menjadi mayoritas dengan jumlah yang cukup dominan yaitu 545 buah tweet, kemudian disusul oleh sentimen positif yang berjumlah 274, dan sentimen negatif 143.

Sementara itu pada penelitian lain (Science., 2020) yang berjudul “Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI” yang menggunakan metode Support Vector Machine dan Decision Tree memperoleh hasil komposisi data terbaik untuk melakukan pengujian adalah 80%:20% dengan mendapatkan hasil nilai akurasi 87.45%, precision 87.72%, recall 91.74% dan F1-Score 89.69% pada Decision Tree dengan TF-IDF sedangkan untuk Support Vector Machine dengan TF-IDF komposisi data terbaik adalah 80%:20% mendapatkan hasil nilai akurasi akurasi 94.36%, precision 96.78%, recall 94.30% dan F1-Score 95.53%. Maka pada kasus ini sentiment analisis pada komentar instagram akan lebih baik jika menggunakan Support Vector Machine (SVM) dengan TF IDF.

Penelitian lain (Kurniasari., 2021) yang menggunakan metode Support Vector Machine yaitu berjudul “Analisis Sentimen Opini Publik pada Instagram mengenai Covid-19 dengan SVM” memperoleh hasil penelitian jika berdasarkan pendekatan ekstraksi fitur SVM dengan pendekatan count vector, TF-IDF word level, TF-IDF ngram level dan TF-IDF char level. Dalam skenario ini nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur count vektor dan TF-IDF ngram level dengan nilai akurasi 59% pada instagram.

Metode lain yang digunakan dalam penelitian yaitu metode Naïve Bayes Classifier (NBC) pernah dilakukan pada penelitian (Astiningrum et al., 2020) yang berjudul “Analisis sentimen tentang opini terhadap performa Timnas Sepak Bola Indonesia pada Twitter” menghasilkan hasil yaitu pada pengujian penelitian ini dilakukan tiga kali pengujian dan setiap pengujian persentase data training nya berbeda karena jumlah data training juga berpengaruh terhadap peningkatan nilai akurasi. Dari tiga kali pengujian didapatkan akurasi tertinggi yaitu 87%, precision positif tertinggi yaitu 93%, precision negatif tertinggi yaitu 85%, recall positif

tertinggi yaitu 75%, dan recall negatif tertinggi yaitu 98% pada persentase 90% data training dan 10% data testing.

Penelitian lain yang menggunakan metode Naïve Bayes Classifier yaitu penelitian (Simorangkir et al., 2018) yang berjudul “Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier”. Pada penelitian yang berfokus pada sentimen game online ini menghasilkan hasil penelitian yakni nilai akurasi, error, recall, dan precision yang didapat masing-masing sebesar 88,89%, 19,18%, 96,97%, dan 69,57% untuk Mobile Legends sedangkan Arena of Valor memiliki nilai akurasi, error, recall, dan precision masing-masing sebesar 39,02%, 60,98%, 88,89% dan 28,74%.

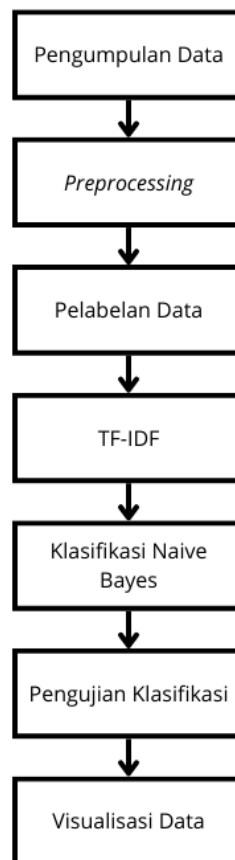
Sementara itu penelitian lain (Kencana et al., 2014) yang berjudul “Analisis sentimen menggunakan metode naive bayes classifier dengan seleksi fitur chi square” memperoleh hasil pengklasifikasi Naïve Bayes diperoleh akurasi sebesar 83% dan rata-rata harmonisa sebesar 90,713%.

Berdasarkan hasil kajian dari penelitian terdahulu, metode yang digunakan dalam penelitian ini adalah metode Naïve Bayes Classifier (NBC) dengan tema utama yaitu sentimen Timnas sepakbola Indonesia era kepelatihan Shin Tae-yong.

BAB III

METODOLOGI PENELITIAN

Diagram proses penelitian merupakan alur yang menjelaskan proses berjalannya penelitian yang akan dilakukan dari proses ditetapkannya metode hingga akhir pada proses analisis. Dalam penelitian ini gambaran proses penelitian dapat dilihat pada gambar 3.1 berikut :



Gambar 3.1 Diagram proses penelitian

Alur atau langkah pertama dari pengerjaan tugas akhir adalah melakukan pengumpulan data yang nantinya akan dijadikan bahan utama untuk penelitian. Setelah data diperoleh, dilakukan pra-pemrosesan data yang dimaksudkan untuk mempermudah proses analisis dan proses pengujian. Pada proses pra-pemrosesan dilakukan beberapa tahapan diantaranya adalah Cleaning, Steaming, Remove Stopword, dan Tokenization. Proses pembersihan (Cleaning) bertujuan untuk menghilangkan kata-kata yang tidak relevan dan mengurangi noise. Selain itu,

tanda baca seperti titik (.), koma (,), dan tanda baca lainnya juga akan dihapus. Pada tahap pembersihan, kata-kata akan diubah menjadi huruf kecil (lower-case). Proses steaming digunakan untuk menghilangkan kata kata berimbuhan menjadi, kata baku. Proses penghapusan kata-kata umum (Stopword) dilakukan untuk menghilangkan kata-kata yang kurang bermakna. Tokenisasi adalah langkah untuk mengidentifikasi kata-kata dalam teks dan membaginya menjadi urutan terpotong oleh spasi atau karakter khusus. Setelah itu dilakukan pelabelan data latih dengan mengkategorisasikan menjadi positif dan negatif, pelabelan dilakukan menggunakan proses perbandingan antara pelabelan otomatis dan manual. Setelah itu dilakukan proses selanjutnya yaitu analisis dan pengujian data. Setelah hasil pengujian diperoleh dilakukan visualisasi data, dan interpretasi hasil.

3.1 Pengumpulan Data

Ketika membahas mengenai media sosial, saat ini terdapat banyak platform yang populer di kalangan masyarakat, mulai dari yang berfokus pada video, gambar, hingga berbasis teks (Basri, 2017). Salah satu platform yang mencolok adalah Twitter atau yang kini berganti nama menjadi X, yang terbukti menjadi sarana efektif untuk mengekspresikan sentimen, pandangan, dan opini (Villavicencio et al., 2021). Sebagai salah satu platform media sosial paling dominan di seluruh dunia, X atau twitter memungkinkan pengguna untuk berbagi apapun yang ada dalam pikiran mereka. Rissan (2022) dalam penelitiannya mengungkapkan bahwa X memiliki lebih dari 500 juta pengguna aktif global, dengan jumlah tweet harian yang mencapai angka 500 juta (Rissan et al., 2022).

Dalam penelitian yang dilakukan oleh Villavicencio (2021), dijelaskan bahwa X atau twitter memberikan kesempatan bagi individu untuk sepenuhnya mengekspresikan diri mereka pada waktu yang tepat. Ini merupakan perbedaan dengan wawancara tatap muka konvensional, di mana respons dari individu yang diwawancarai mungkin terpengaruh oleh ketidaknyamanan yang muncul dalam komunikasi langsung antara pewawancara dan subjek wawancara. Saat menggunakan X dalam situasi yang lebih privasi, individu dapat dengan mudah mengekspresikan diri mereka dengan cara yang lebih jujur. Oleh karena itu, X dianggap sebagai platform yang efektif untuk menganalisis sentimen publik yang sebenarnya. Meskipun karakter tweet dibatasi hingga 280, pengguna dapat bebas berbagi lokasi, komentar, opini, dan perasaan mereka, menjadikannya cocok untuk penelitian yang memerlukan analisis sentimen (Villavicencio et al., 2021).

Pengumpulan data dilakukan dengan dua rentang waktu yakni pada 6 September 2023 hingga 14 September 2023 bertepatan pada kualifikasi piala asia U 23, dan pada tanggal 14 April 2024 sampai dengan 2 Mei 2024 bertepatan pada piala asia U 23. Dengan mempertimbangkan rencana waktu pengumpulan data ini, diharapkan bahwa analisis sentimen yang dihasilkan dapat mencerminkan dengan akurat dan komprehensif tanggapan masyarakat di X terhadap Timnas sepakbola dan dapat membandingkan opini masyarakat dengan 2 waktu yang berbeda. Pengumpulan data dalam penelitian ini merupakan tahap awal yang sangat penting untuk memastikan keberlanjutan dan ketepatan analisis sentimen terhadap percakapan yang terkait dengan Timnas sepakbola di platform X. Dalam penelitian ini, metode pengumpulan data yang digunakan adalah insidental, yang berarti data dikumpulkan pada rentang waktu tertentu yang telah ditentukan. Pengumpulan data dilakukan selama peristiwa penting seperti pertandingan kualifikasi, yang dapat mempengaruhi sentimen secara signifikan.

Beberapa kata kunci digunakan pada proses pengumpulan data antara lain “Shin Tae-yong”, “STY”, dan “Timnas”. Adapun alasan pemilihan beberapa kata kunci tersebut adalah berdasarkan observasi secara tidak langsung, yaitu berdasarkan hasil kajian jurnal sebelumnya yang kemudian dikaitkan dengan tema permasalahan yang ada. Alasan lain yaitu dengan mengaitkan pemberitaan yang sering dibahas di media dan masyarakat yaitu mengenai performa bermain Timnas berkaitan dengan kepelatihan, yang kemudian hal tersebut dimasukkan sebagai kata kunci untuk pencarian data. Kata kunci yang dipilih juga merupakan kata kunci umum yang berkaitan langsung dengan tema skripsi yang sedang dikerjakan.

Proses pengambilan tweet ini memanfaatkan library “tweet-harvest” yang tersedia di situs helmisatria.com, yang bekerja dengan memanfaatkan fitur-fitur dari Twitter Developer API. Library ini dirancang untuk berjalan dengan baik di Google Colab, menjadikannya alat yang sangat praktis dan mudah digunakan oleh peneliti. Penggunaan library ini sangat memudahkan dalam memperoleh data tweet dalam jumlah besar, yang seringkali dibutuhkan untuk analisis sentimen atau penelitian lainnya. Sebagai perbandingan, menurut pembaruan kebijakan terbaru dari platform X (sebelumnya dikenal sebagai Twitter), pengguna biasa hanya dapat mengakses hingga 1000 tweet saja. Namun, dengan menggunakan “tweet-harvest”, peneliti dapat mengumpulkan data melebihi batas tersebut, yang sangat penting untuk studi yang membutuhkan dataset yang lebih besar dan representatif.

Proses pengumpulan data ini melibatkan beberapa langkah penting. Pertama, peneliti harus mendapatkan `auth_token` yang diperlukan untuk mengautentikasi akses ke API Twitter. Setelah itu, perlu menentukan nama file yang akan digunakan untuk menyimpan data yang

dikumpulkan. Langkah berikutnya adalah menentukan keyword yang relevan dengan topik penelitian, yang akan digunakan untuk memfilter tweet yang diambil. Terakhir, peneliti harus menetapkan jumlah maksimum data yang ingin diambil, sesuai dengan kebutuhan analisis. Semua langkah ini harus diikuti dengan teliti untuk memastikan data yang dihasilkan lengkap dan sesuai dengan kriteria penelitian. Urutan lengkap dari langkah-langkah ini dapat dilihat pada gambar 3.2, yang memberikan panduan visual tentang proses pengumpulan data menggunakan library ini.

```

[1] # Import required Python package
    pip install pandas
    # Install Node.js (because tweet-harvest built using Node.js)
    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=20.8.4 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

    Inode -v

[2] # Crawl Data
    filename = 'STV.csv'
    search_keyword = 'sty until:2023-09-13 since:2023-09-09'
    limit = 500

    Inpx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token ""

[3] Import pandas as pd

    # Specify the path to your CSV file
    file_path = f'tweets_data/{filename}'

    # Read the CSV file into a pandas DataFrame
    df = pd.read_csv(file_path, delimiter=";")

    # Display the DataFrame
    display(df)
  
```

Gambar 3.2 Proses crawling data

Pada proses ini, terdapat beberapa langkah penting yang perlu dipersiapkan, salah satunya adalah pengambilan `auth_token` melalui akun Twitter yang telah disiapkan. `Auth_token` ini adalah komponen krusial dalam proses pengambilan data, karena berfungsi sebagai informasi yang memverifikasi identitas pengguna kepada situs web, server, atau entitas lain yang membutuhkan verifikasi identitas pengguna. Tanpa `auth_token` yang valid, akses ke API Twitter untuk melakukan crawling data tidak akan diizinkan, sehingga proses pengumpulan data tidak dapat dilanjutkan.

Selain itu, penting untuk memastikan bahwa akun Twitter yang digunakan memiliki izin dan konfigurasi yang sesuai untuk mengakses API, guna menghindari kesalahan atau kendala teknis selama proses crawling. Perintah-perintah yang digunakan untuk melakukan crawling data pada platform X, termasuk cara menggunakan `auth_token` ini, telah dirangkum dalam bentuk kode dan langkah-langkah yang terperinci, sebagaimana ditampilkan pada gambar 3.3.

Gambar tersebut memberikan panduan visual yang jelas tentang cara melaksanakan setiap langkah dalam proses pengambilan data, sehingga memudahkan peneliti dalam mengikuti prosedur yang diperlukan.

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

# Install Node.js (because tweet-harvest built using Node.js)
!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
```

Gambar 3.3 Inisiasi dan instalasi

Proses selanjutnya adalah crawling data, yaitu mengumpulkan data dari sumber tertentu yang kemudian disimpan dalam file berformat .csv. Proses ini dimulai dengan menentukan nama file yang akan digunakan untuk menyimpan hasil data dalam format csv, sehingga data yang diperoleh dapat diakses dan dianalisis dengan mudah di kemudian hari. Langkah berikutnya adalah menentukan keyword atau kata kunci yang relevan dengan topik penelitian untuk memastikan bahwa data yang diambil sesuai dengan kebutuhan analisis. Selain itu, penting untuk menetapkan waktu awal dan akhir pengambilan data, sehingga data yang dikumpulkan terbatas pada periode yang relevan dengan penelitian. Jumlah data yang akan disimpan juga harus ditentukan melalui parameter limit untuk memastikan bahwa jumlah data yang dikumpulkan tidak melebihi kapasitas atau kebutuhan analisis. Pada proses ini, diperlukan auth-token yang diperoleh dari akun yang telah disiapkan sebelumnya. Auth-token ini berfungsi sebagai kunci autentikasi yang memungkinkan akses ke API atau platform dari mana data akan diambil. Penggunaan auth-token memastikan bahwa hanya akun yang sah yang dapat melakukan crawling data sesuai dengan ketentuan dan batasan yang ditetapkan oleh platform

tersebut. Untuk detail lebih lanjut mengenai perintah dan langkah-langkah dalam proses crawling ini, dapat dilihat pada gambar berikut.

```
# Crawl Data

filename = 'piala.csv'
search_keyword = 'Shin Tae-yong until:2023-03-28 since:2023-03-01'
limit = 500

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

Gambar 3.4 Proses crawling data

```
import pandas as pd

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=";")

# Display the DataFrame
display(df)
```

Gambar 3.5 Proses file csv

```
# Cek jumlah data yang didapatkan

num_tweets = len(df)
print(f"Jumlah tweet dalam dataframe adalah {num_tweets}.")
```

Gambar 3.6 Melihat data yang didapatkan

Setelah menjalankan seluruh program, hasil yang diperoleh dapat disimpan secara langsung dalam format data .csv, yang memudahkan penyimpanan dan pengelolaan data untuk analisis lebih lanjut. Format csv memungkinkan data disimpan dalam bentuk tabel dengan setiap elemen terorganisir rapi dalam baris dan kolom, sehingga mudah untuk diakses atau diimpor ke dalam perangkat lunak analisis data lainnya.

Penyimpanan dalam format csv juga memastikan bahwa hasil dapat dibagikan atau diarsipkan dengan standar yang kompatibel secara luas. Contoh hasil penyimpanan data dalam format csv dapat dilihat pada gambar 3.7 berikut.

created_at	id	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id	conversation_id	username	tweet_url
Thu Sep 14 23:46:54 +0000 2023;	1702468997722341871;	"Shin Tae yong Ungkap Peran Erick Thohir Dibalik Kesuksesan Timnas Indonesia	https://t.co/8MAqf8oB5Q";	0							
Thu Sep 14 22:01:09 +0000 2023;	1702442386499191295;	"ERICKTHOHIRNEWS Shin Tae Yong Sebut Erick Thohir Berperan Penting di Balik Kesuksesan Timnas Indon									
Thu Sep 14 15:22:18 +0000 2023;	1702342013172212070;	"PantauManCity Lebih ke Shin Tae Yong sih";	0;1;0;0;in;	1409020642599469069;	1702341848524865858;	sa					
Thu Sep 14 mau Indra		semoga selalu bisa membanggakan Indonesia	ðŸ™ðŸ™	AAMIIN	ðŸ™ðŸ™";	0;0;0;21;in;	1411220206869905411;	1702327800710197739;	dyplight;ht		
Thu Sep 14 No Debat!		https://t.co/QzBJH1gYLU	#diataskertas	#shintaeoyng	#110tv	https://t.co/68XugzbbWT";	0;0;0;4;in;	38803267;	1702322381279396261;	t10nugroh	
Thu Sep 14 14:03:16 +0000 2023;	1702322125066117320;	@idextratime Shin tae yong gak ada min. . .";	0;0;0;0;in;	559727769;	1702321371932746023;	kulo_zhargopro;					
Thu Sep 14 14:02:00 +0000 2023;	1702321804306767981;	"Shin Tae-yong sangat yakin Skuat Garuda Muda dapat bicara banyak di Piala Asia U-23 2024. #bolaskorcom									
Thu Sep 14 doa		dan optimisme kita semua. Terimakasih jg utk Pak @erickthohir dan coach Shin Tae Yong	ðŸ™	#ETbikinPSSImendunia	https://t.co/jjKE1GRbDnS'						
Thu Sep 14 12:14:06 +0000 2023;	1702294649103430108;	@speechace @tanyakanrl Ngapa kek orang dewasa ðŸ™";	0;0;0;0;in;	96128322;	1702230719366582782;	bria					
Thu Sep 14 Shin Tae-y		memuji etos kerja yang dimiliki timnya. #bolaskorcom #timnasindonesia #shintaeoyng #garudamuda	https://t.co/rbQe29cw4R";	0;0;0;0;in;	9404;						
Thu Sep 14 wani mematahkan prinsipku"		so sweet sih tapi knp harus diibaratkan kaya coach timnas	ðŸ™";	0;0;0;1;in;	1277578021428330498;	1702287630128562422;	m				
Thu Sep 14 Shin Tae Y		tiga level tim nasional pun kini berhasil berkiprah di Asia. Simak #Analisis #CNNIndonesia	https://t.co/a8aka5rYao	https://t.co/pcE5ojAIMW";	0;0;0;						
Thu Sep 14 11:00:18 +0000 2023;	1702276079929737500;	@FaktaSepakbola @celebesmedia_ @Daeng_Info Mudah2an dia pengganti shintaeoyng kalau tidak diperpa									
Thu Sep 14 saya send emang dasar orang goblok aja mereka.";	0;0;0;0;in;	1547294946679353344;	1702267729485639810;	papristant0;	https://twitter.com/papristant0/						
Thu Sep 14 Shin Tae Yong Optimis Sepak Bola Indonesia akan Terus Berkembang Melampaui Level Asia		https://t.co/x7lSrSkfXZ";	0;0;0;0;in;	1559894545021644800;	1702						
Thu Sep 14 07:52:50 +0000 2023;	1702228902222184634;	Ternyata Ini Filosofi Sepak Bola Indonesia Menurut Shin Tae Yong Setelah Antarkan Timnas Garuda Ke Piala									
Thu Sep 14 07:44:05 +0000 2023;	1702226697729552389;	@MafiaWasit Smoga shin tae yong jadi cawapres";	0;0;0;0;in;	1219890189163626496;	170197006367322153						
Thu Sep 14 06:54:00 +0000 2023;	1702214093682135334;	"Shin Tae-yong mengukir rekor pribadi. #bolaskorcom #shintaeoyng #timnasindonesia #pssi	https://t.co/BwI								

Gambar 3.7 Contoh data yang berhasil dikumpulkan dari X

3.2 Preprocessing

Preprocessing adalah tahap awal yang dilakukan untuk membersihkan dan menyiapkan teks sebelum dijalankan melalui sistem analisis. Tujuan dari tahap preprocessing ini adalah untuk menghindari kehilangan nilai data, kelebihan data, dan inkonsistensi data. Pada tahap ini data yang diperoleh dari proses crawling dilakukan pengolahan yang ditujukan untuk mempermudah proses selanjutnya. Dalam tahap ini, RapidMiner Studio versi 9.9 digunakan untuk melakukan proses preprocessing dan pengujian dataset. Dataset uji yang digunakan untuk menguji model diperoleh dari data yang telah dikumpulkan sebelumnya, khususnya terkait sentimen terhadap timnas sepakbola Indonesia. RapidMiner Studio versi 9.9 menjadi alat yang efektif dalam mengelola dan menganalisis dataset tersebut, memungkinkan penggunaan metode preprocessing untuk membersihkan dan menyusun data dengan tepat. Penting untuk mencatat bahwa hasil pengujian model akan menjadi landasan untuk evaluasi dan perbaikan lebih lanjut dalam penggunaan algoritma atau pendekatan analisis sentimen selanjutnya.

Berdasarkan penelitian yang dilakukan oleh (Akbar et., al 2023), beberapa proses diantaranya adalah Cleaning, Steaming, Case Folding, Remove Stopword, dan Tokenization. Pada proses cleaning dan case folding hal yang dilakukan adalah pembersihan data yang dimaksudkan untuk menghilangkan elemen yang tidak diperlukan diantaranya adalah merubah huruf menjadi huruf kecil semua, penghapusan tanda baca, username, hastag, multiple space, dan semua karakter. Hal tersebut dilakukan untuk mengurangi beban pemrosesan pada tahap selanjutnya dan dianggap tidak relevan untuk masuk pada proses analisa. Pada RapidMiner Studio, proses cleaning dan case folding menggunakan operator yang disediakan yaitu replace dan transform cases. Pada tabel 3.1 berikut merupakan contoh sederhana proses sebelum dan setelah dilakukan cleaning.

Tabel 3.1 Contoh proses cleaning dan case folding

Sebelum	Sesudah
Ayo timnas indonesia..!!!	ayo timnas indonesia
Kapan timnas main lagi";0;0;0;0;	kapan timnas main lagi

Pada proses steaming merupakan langkah untuk mengubah kata-kata menjadi bentuk dasarnya (root form). Tujuannya adalah untuk menyederhanakan kata-kata dengan menghilangkan akhiran dan variasi kata.

Pada tabel 3.2 berikut merupakan contoh sederhana proses sebelum dan setelah dilakukan steaming.

Tabel 3.2 Contoh proses steaming

Sebelum	Sesudah
mendapat trofi dan melaju jauh di piala asia	dapat trofi dan laju di piala asia
menang atau mengalah	menang atau kalah

Proses selanjutnya adalah remove stopword atau tahap filtering, proses ini dimaksudkan untuk menghapuskan kata kata yang tidak memiliki makna, misalnya kata sambung kata penghubung dan lain sebagainya, beberapa contoh diantaranya yaitu “ada, adalah, atau” dan lain sebagainya. Pada proses ini menggunakan kumpulan kata stopword bahasa berjumlah 758 kata yang didapat dari situs kaggle.com. Proses ini pada RapidMiner Studio menggunakan operator filter stopwords (Dictionary). Ada beberapa kata yang dipertahankan diantaranya adalah belum, akhir, bisa, lagi dan lain sebagainya.

Pada tabel 3.3 berikut merupakan contoh sederhana proses sebelum dan setelah dilakukan remove stopword.

Tabel 3.3 Contoh proses remove stopword

Sebelum	Sesudah
dapat trofi dan melaju jauh di piala asia	trofi melaju jauh piala asia
menang atau kalah	menang kalah

Pada proses tokenization atau tokenisasi merupakan langkah untuk memecah kalimat menjadi potongan kata atau token. Pada tabel 3.4 berikut merupakan contoh sederhana proses sebelum dan setelah dilakukan tokenization.

Tabel 3.4 Contoh proses tokenization

Sebelum	Sesudah
dapat trofi dan melaju jauh di piala asia	[trofi, melaju, jauh, piala, asia]
menang atau kalah	[menang, kalah]

3.3 Pelabelan Sentimen

Setelah mendapatkan data dan melakukan tahap pra-pemrosesan, langkah selanjutnya adalah melakukan analisis dengan menetapkan label positif atau negatif pada setiap ulasan. Proses pelabelan dilakukan secara otomatis menggunakan model Transformer dan dilakukan secara manual. Transformer dalam hal ini BERT (Bidirectional Encoder Representations from Transformers) adalah sebuah arsitektur jaringan saraf tiruan dalam bidang pemrosesan bahasa alami (Natural Language Processing) seperti penerjemahan bahasa dan model pembelajaran mendalam yang digunakan untuk analisis sentimen dengan mempertimbangkan konteks kata dalam sebuah kalimat (Talaat et al., 2023). Berikut adalah penjelasan mengenai bagaimana BERT mensegmentasi sentimen:

a. Tokenisasi

BERT memecah teks menjadi token-token, yaitu potongan-potongan kecil seperti kata atau sub-kata. Proses ini melibatkan pemetaan kata menjadi representasi numerik yang dapat diproses oleh model.

b. Embedding

Setiap token diubah menjadi vektor numerik (embedding). BERT kemudian memahami konteks setiap kata dengan menganalisis hubungan antar kata dalam dua arah, yaitu dari kiri ke kanan dan dari kanan ke kiri secara bersamaan (bidirectional).

c. Fine-Tuning

Untuk tugas analisis sentimen, BERT perlu di-fine-tune pada dataset yang telah dilabeli dengan sentimen (misalnya, positif, negatif, atau netral). Dalam proses ini, model dilatih untuk memprediksi label sentimen berdasarkan embedding yang telah dihasilkan.

d. Prediksi

Setelah proses fine-tuning, model dapat memprediksi sentimen dari teks baru. Teks tersebut diproses melalui langkah-langkah tokenisasi dan embedding yang sama, kemudian diklasifikasikan menjadi label sentimen berdasarkan embedding yang dihasilkan.

Sementara itu pelabelan manual dilakukan dengan cara melihat nilai sentimen kata tersebut. Setelah didapatkan 2 pelabelan, kemudian dilakukan perbandingan, dan memverifikasi nilai pelabelan yang berbeda. Pada proses ini, dilakukan klasifikasi opini menjadi dua kategori, yaitu positif dan negatif, untuk memahami lebih dalam pandangan masyarakat terhadap timnas. Sentimen dikategorikan sebagai positif jika ulasan berisi tanggapan dan dukungan yang bersifat membangun, serta menggunakan kata-kata sopan (misalnya : “amanah”, “bagus”, “baik”, “enak”). Sebaliknya, sentimen dikategorikan sebagai negatif jika ulasan mencerminkan kekecewaan, hujatan, dan menggunakan kata-kata yang kurang sopan (misalnya : “bodoh”, “anjing”, “frustasi”, “gila”). Proses ini bertujuan untuk memberikan gambaran yang lebih komprehensif terhadap respons dan pandangan masyarakat terhadap timnas.

3.4 Pembobotan TF-IDF

Data yang telah melalui proses pra-pemrosesan perlu diubah ke dalam bentuk numerik agar dapat dihitung dan diolah lebih lanjut. Transformasi ini melibatkan konversi data ke dalam vektor, diikuti dengan pemberian nilai dan bobot pada setiap kata, serta penggunaan algoritma prediksi. Pada tahap pembobotan kata, menerapkan algoritma Term-Frequency Times Inverse Document Frequency (TF-IDF). Pemilihan metode ini didasarkan pada efisiensinya, kemudahan implementasinya, dan hasil yang akurat. Metode TF-IDF menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) untuk setiap token (kata) dalam setiap dokumen. Secara sederhana, metode ini digunakan untuk mengukur seberapa sering suatu kata atau term muncul dalam dokumen. Term Frequency (TF) mengindikasikan frekuensi kemunculan kata atau term dalam dokumen, di mana semakin tinggi jumlah term yang muncul (TF tinggi), bobot dokumen atau nilai kemiripan yang diberikan juga semakin besar. Metode ini memungkinkan representasi yang lebih baik dari informasi teks dalam bentuk yang dapat

diolah oleh algoritma prediksi. Sebagai contoh yaitu terdapat 3 dokumen dengan jumlah kata yang berbeda, ketika kita akan mencari bobot kata “berangkat” dari kalimat “saya berangkat ke kampus untuk belajar” yaitu dengan rumus banyaknya kata i pada dokumen ke i dikali log total dokumen dibagi dengan banyaknya dokumen yang mengandung kata ke i maka akan diperoleh nilai 0,4477.

3.5 Klasifikasi Naive Bayes

Metode Naive Bayes digunakan dalam klasifikasi sentimen untuk menganalisis sentimen. Proses klasifikasi sentimen melibatkan penggunaan data yang telah melalui tahap preprocessing hingga pembobotan kata menggunakan metode tf-idf. Setelah data berhasil dilatih, dilakukan pengujian dengan menggunakan data uji untuk menguji ketepatan hasil klasifikasi yang telah dilakukan. Pada penelitian ini perbandingan jumlah data latih dan data uji adalah 90% data latih dan 10% data uji. Hal ini dilakukan karena berdasarkan penelitian sebelumnya yang dilakukan oleh Astiningrum (2020) yang berjudul “Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepakbola Indonesia Pada Twitter” perbandingan tersebut merupakan salah satu pilihan dengan hasil terbaik yang dapat diperoleh.

Pelatihan model adalah proses untuk mengajarkan model tentang data yang akan digunakan. Pelatihan model dilakukan dengan menggunakan data yang telah diolah. Pada pelatihan model Naive Bayes, data yang telah diolah akan digunakan untuk menghitung probabilitas kemunculan setiap kata dalam kelas positif dan kelas negatif. Probabilitas kemunculan kata tersebut akan digunakan untuk memprediksi kelas dari teks yang belum pernah dilihat sebelumnya. Setelah model Naive Bayes mempelajari pola-pola yang ada dalam set pelatihan, maka model Naive Bayes siap untuk digunakan untuk mengklasifikasi data teks.

Pernyataan umum Teorema Bayes adalah sebagai berikut: “Peluang bersyarat suatu kejadian A, diberikan bahwa kejadian B telah terjadi, sama dengan hasil kali dari peluang terjadinya kejadian B jika A diketahui, dan peluang A, dibagi dengan peluang terjadinya kejadian B.” Dengan kata lain, Teorema Bayes memberikan cara untuk memperbarui atau merevisi estimasi probabilitas suatu kejadian berdasarkan informasi baru atau bukti yang telah terjadi. Teorema ini sangat penting dalam berbagai aplikasi statistik dan pembelajaran mesin karena memungkinkan perhitungan peluang yang lebih akurat dengan mempertimbangkan bukti yang ada.

Secara matematis, Teorema Bayes dinyatakan dalam rumus berikut:

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

Dengan :

X = Sampel data yang memiliki kelas (label) yang tidak diketahui

H = Hipotesa bahwa X adalah data kelas (label)

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X

$P(H)$ = Probabilitas dari hipotesa H

$P(X|H)$ = Probabilitas X berdasarkan kondisi hipotesis H

$P(X)$ = Probabilitas dari data sampel yang diamati

Pada tahap pengujian model, model Naive Bayes digunakan untuk mengklasifikasi data teks yang belum pernah dilihat sebelumnya. Data teks yang diklasifikasikan kemudian dibandingkan dengan kelas sebenarnya. Kinerja model Naive Bayes dianalisis menggunakan beberapa metrik, seperti akurasi, presisi, dan recall.

3.6 Pengujian Klasifikasi

Pengujian klasifikasi dilakukan menggunakan himpunan data uji yang tidak digunakan dalam proses pelatihan klasifikasi. Jumlah data yang digunakan dalam proses pengujian adalah 10% dari keseluruhan data yang digunakan, penggunaan jumlah tersebut dikarenakan pada penelitian sebelumnya (Astiningrum et al., 2020), pembagian data 90% data training dan 10% data testing dapat memperoleh hasil representasi yang terbaik. Hal ini disebabkan karena dengan memberikan jumlah lebih banyak data untuk pelatihan, diharapkan dapat memberikan representasi yang baik, mengurangi risiko overfitting, memungkinkan evaluasi konsisten, dan meningkatkan akurasi prediksi model. Evaluasi model klasifikasi dapat dilakukan dengan memanfaatkan beberapa metrik, seperti akurasi (accuracy), presisi (precision), dan recall. Model klasifikasi ini memetakan baris data ke hasil prediksi kelas. Dalam konteks klasifikasi biner, yang menghasilkan dua label output, representasinya dapat berupa {0,1}, {ya, tidak}, {+1,-1}, atau {positif, negatif} untuk setiap data input. Dalam proses evaluasi klasifikasi, terdapat empat kemungkinan hasil dari prediksi suatu data. Jika data positif dan diprediksi positif, itu dihitung sebagai true positive; jika data positif tetapi diprediksi negatif, dihitung sebagai false negative. Pada data negatif, jika diprediksi negatif, itu akan dihitung sebagai true

negative; sedangkan jika diprediksi positif, dihitung sebagai false positive. Metrik-metrik evaluasi ini memberikan gambaran yang lebih komprehensif tentang seberapa baik model klasifikasi dapat mengenali dan membedakan antara kelas positif dan negatif.

a. Accuracy

Accuracy adalah tingkat ketepatan prediksi yang benar, yang menunjukkan seberapa sering model atau sistem berhasil memberikan hasil yang sesuai dengan kebenaran atau label yang sebenarnya. Accuracy dapat dihitung menggunakan rumus berikut:

$$\text{Accuracy} = \frac{TP + FN}{TP + FP + TN + FN} \quad (3.1)$$

Rumus pada 3.1 ini mengukur sejauh mana model klasifikasi dapat memprediksi dengan benar dalam keseluruhan dataset.

b. Precision

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Dalam konteks sistem informasi atau analisis data, precision mengukur sejauh mana hasil yang diberikan oleh sistem sesuai dengan apa yang sebenarnya diinginkan oleh pengguna. Precision dapat dihitung menggunakan rumus berikut:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.2)$$

Rumus pada 3.2 ini mengukur sejauh mana prediksi positif dari model klasifikasi merupakan prediksi yang benar dari keseluruhan prediksi positif.

c. Recall

Recall (juga dikenal sebagai Sensitivity atau True Positive Rate) adalah perhitungan ketepatan prediksi yang digunakan sebagai ukuran tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Rumus recall dapat dihitung menggunakan persamaan:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.3)$$

Rumus pada 3.3 ini mengukur sejauh mana model klasifikasi mampu mendeteksi dengan benar semua instance yang seharusnya positif dalam dataset.

3.7 Visualisasi Word Cloud

Selain itu, penelitian ini juga menggunakan word cloud sebagai alat visualisasi untuk mempermudah pemahaman kata-kata berdasarkan frekuensi kemunculannya. Word cloud menjadi alat yang efektif dalam menampilkan kata-kata yang dominan atau sering muncul, memberikan gambaran yang jelas tentang fokus atau tema yang mendominasi dalam setiap kategori sentimen. Hal ini bertujuan untuk memberikan kemudahan kepada semua pihak untuk memahami hasil analisis data secara intuitif. Word cloud dapat menjadi instrumen yang powerful dalam menyampaikan informasi, terutama ketika ingin dengan cepat mengidentifikasi kata-kata kunci yang mencirikan sentimen dalam percakapan atau ulasan.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Proses pengumpulan dataset yang digunakan pada penelitian ini melalui media sosial X. Data yang dikumpulkan merupakan pengambilan data rentang waktu pelaksanaan Kualifikasi Piala Asia U 23 dan Piala Asia U 23. Data yang dikumpulkan dimulai dari tanggal 6 September 2023 sampai dengan 14 September 2023 dan tanggal 14 April 2024 sampai dengan 2 Mei 2024. Hasil crawling pada masa kualifikasi piala asia U 23 data keseluruhan berjumlah 5211 data dan hanya 2006 data yang dipakai, sedangkan hasil crawling pada masa piala asia U 23 berjumlah 2500 data dan hanya 2107 data yang dipakai, data tidak terpakai karena data tersebut bukan termasuk data yang sesuai dengan tema dan masuk pada proses seleksi. Tahapan ini melibatkan library tweet-harvest yang tersedia pada blog pribadi helmisatria.com.

Setelah menjalankan semua proses, data yang diperoleh dapat langsung didownload dalam bentuk csv. Beberapa isi data masih mengandung beberapa bagian yang tidak dibutuhkan, misalnya username, hashtag, dan lain sebagainya, untuk itu perlu dilakukan preprocessing untuk mengolah data dan mempermudah proses selanjutnya. Sebelum melalui tahap preprocessing dilakukan modifikasi dataset yang ditujukan untuk mempermudah sistem untuk mengenali data set dan menghapus yang tidak diperlukan. Pada hasil crawling terdapat beberapa kolom yang berhasil disimpan diantaranya nomor id, waktu pembuatan, replay user id dan sebagainya. Untuk mempermudah dalam menganalisis hanya atribut text yang digunakan. Hasil crawling data dapat dilihat pada gambar 4.1 berikut.

1	created_at	id	str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id	st	conversasi	username	tweet_url
2	Thu Sep 14	1.7E+18	Shin Tae y		0	0	0	0	in	1.45E+18	1.7E+18	Redaksiasat	https://twitter.com/Redaksiasatuarah/status/1702468997722341871	
3	Thu Sep 14	1.7E+18	ERICKTHO		0	0	0	0	in	1.67E+18	1.7E+18	Berita_ET	https://twitter.com/Berita_ET/status/1702442386499191295	
4	Thu Sep 14	1.7E+18	@PantauN		0	1	0	0	in	1.41E+18	1.7E+18	sarapantall	https://twitter.com/sarapantall/status/1702342013172212070	
5	Thu Sep 14	1.7E+18	@FaktaSepakbola	Siapun pelatihnya										
6	Thu Sep 14	1.7E+18		Waktunya untuk memperpanjang kontrak #shintayong										
7	Thu Sep 14	1.7E+18	@idextrati		0	0	0	0	in	5.6E+08	1.7E+18	kulo_zharq	https://twitter.com/kulo_zharqopro/status/1702322125066117320	
8	Thu Sep 14	1.7E+18	Shin Tae-y		0	0	0	0	in	9.4E+17	1.7E+18	bolaskorcc	https://twitter.com/bolaskorcom/status/1702321804306767981	
9	Thu Sep 14	1.7E+18		Keberhasilan Timnas U-23 yg lolos ke Piala Asia U-23 2024 menjadi kabar bahagia utk Indonesia. Tentu saja ini tak lepas dr support										
10	Thu Sep 14	1.7E+18	@speacha		0	0	0	0	in	96128322	1.7E+18	brianjienoi	https://twitter.com/brianjienoi/status/1702294649103430108	
11	Thu Sep 14	1.7E+18		Pelatih Timnas Indonesia U-23										
12	Thu Sep 14	1.7E+18	ikm tu kaya	shin tae yong										
13	Thu Sep 14	1.7E+18		Diakui atau tidak										
14	Thu Sep 14	1.7E+18	@FaktaSej		0	0	0	0	in	1.65E+18	1.7E+18	lordran44	https://twitter.com/lordran44/status/1702276079929737500	
15	Thu Sep 14	1.7E+18	@frans_usb57	@CoachJustinL Betul										
16	Thu Sep 14	1.7E+18		Setelah Berhasil Borong 3 Kemenangan Beruntun										
17	Thu Sep 14	1.7E+18		Ternyata li	0	0	0	0	in	1.43E+18	1.7E+18	ImronRasyid8	https://twitter.com/ImronRasyid8/status/1702228902222184634	
18	Thu Sep 14	1.7E+18	@MafiaW		0	0	0	0	in	1.22E+18	1.7E+18	froginco	https://twitter.com/froginco/status/170222669772952389	
19	Thu Sep 14	1.7E+18	Shin Tae-y		0	0	0	0	in	9.4E+17	1.7E+18	bolaskorcc	https://twitter.com/bolaskorcom/status/1702214093682135334	
20	Thu Sep 14	1.7E+18		Ulasan tak	0	0	0	2	in	1.49E+09	1.7E+18	bolacoml	https://twitter.com/bolacoml/status/1702213651070116125	
21	Thu Sep 14	1.7E+18	Shin Tae Yi		0	0	0	0	in	99210470	1.7E+18	Kikoyaye	https://twitter.com/Kikoyaye/status/1702211480790417490	
22	Thu Sep 14	1.7E+18	@FaktaSepakbola	@TimnasIndonesia Pake tim yg dikirim ke Asian Games aja										
23	Thu Sep 14	1.7E+18	Timnas Inc		0	0	0	0	in	9.4E+17	1.7E+18	bolaskorcc	https://twitter.com/bolaskorcom/status/1702202014107926558	

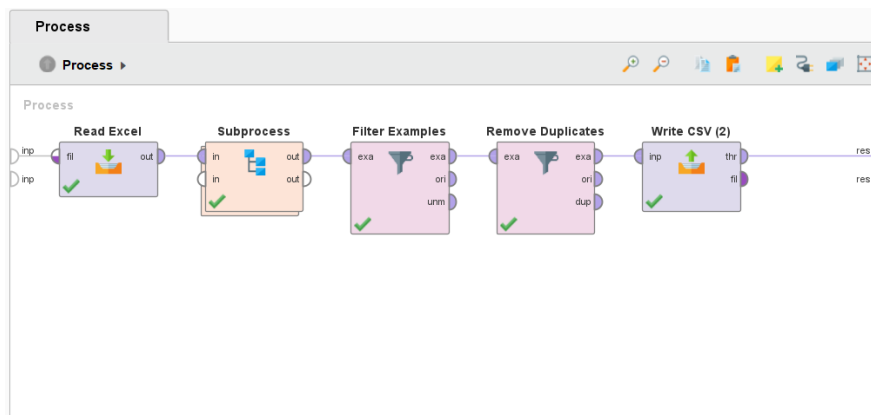
Gambar 4.1 Hasil crawling data

4.2 Preprocessing

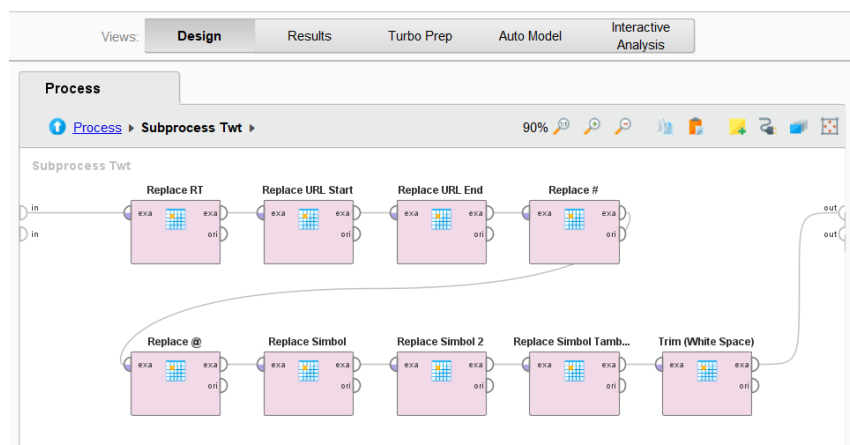
Pada tahap ini data yang diperoleh dari proses crawling diolah menjadi data yang dapat dianalisis. Proses preprocessing didasarkan pada penelitian yang dilakukan oleh (Akbar 2023), beberapa Proses diantaranya adalah cleaning, case folding, remove stopwords, dan tokenization. Dengan mengikuti metode preprocessing ini, saya berharap dapat meningkatkan kualitas data yang digunakan untuk analisis sentimen, sehingga hasil yang diperoleh lebih akurat dan dapat diandalkan (Akbar et., al 2023). Prosesnya diantaranya adalah.

a. Cleaning

Proses ini memiliki tujuan untuk membersihkan atribut yang tidak diperlukan dan tidak memiliki makna, seperti username (@), hashtag (#), tanda baca, angka, karakter tunggal, mention, retweet dan lain sebagainya. Penghapusan ini dilakukan untuk mengurangi beban pemrosesan dan memastikan bahwa data yang digunakan dalam analisis lebih fokus dan bermakna. Berikut adalah gambaran proses cleaning data pada aplikasi RapidMiner yang terlampir pada gambar 4.2 dan 4.3.



Gambar 4.2 Proses cleaning.



Gambar 4.3 Proses cleaning.

Beberapa proses yang dilakukan pada subprocess pada gambar 4.3 diantaranya adalah menghapus seluruh url yang ada pada tweet baik didepan dan juga dibelakang kalimat. Setelah itu dilakukan penghapusan hastag (#), username dan juga simbol yang tidak diperlukan.

Row No.	full_text
771	MafiaWasit Smoga shin tae yong jadi cawapres
772	Shin Taeyong mengukir rekor pribadi
773	Ulasan taktik ciamik Coach Shin TaeYong setelah bawa Garuda Muda ke putaran final Piala Asia U23 2024
774	Shin Tae Yong Presiden 2024 aja gak si
775	FaktaSepakbola TimnasIndonesia Pake tim yg dikirim ke Asian Games aja
776	Timnas Indonesia U23 untuk pertama kali lolos ke Piala Asia U23
777	Pelatih Turkmenistan Akui Racikan STY Sangat Kuat
778	Coach Shin Tae Yong diusulkan jadi PNS Buseeett uda siap banget kayaknya liat oppaoppa pake pin korpri di dada
779	Viral di Media Sosial Momen Rizky Ridho Salah Oper dan Kena Omel Shin Tae Yong
780	Bersama coach STY kita mulai bisa berbicara di level Asia Dukung terus timnas Indonesia ya
781	PSSI Sebaiknya Shin Tae Yong diperpanjang lagi 45 th lagi nanti akan makin kelihatan hasilnya Pondasi nya akan lebih kuat.
782	SiaranBolaLive PNS Pelatih Nasional Shin tae yong
783	SiaranBolaLive Naturalisasi coach Shin Tae Yong dan jadiin PNS Kemenpora
784	labiebsadat Shin tae yong mau gak jadi pns kira2 bang
785	gibrantweet Tak kiro nggo jaket shintaeyong
786	aku pengen di ulah ku yg ke 21 bisa dipertemukan sama Shin Tae Yong
787	SiaranBolaLive Tanpa coach Shin Tae Yong apakah timnas bisa berprestasi seperti saat ini Kok anda tidak sebut nama belia

Gambar 4.4 Hasil proses cleaning.

Hasil dari proses cleaning dapat dilihat pada gambar 4.4 Pada tahap tersebut data sudah bersih dari bagian bagian yang tidak diperlukan, dan dapat digunakan untuk proses berikutnya.

b. Case folding

Pada proses Case Folding, semua huruf besar pada suatu dokumen atau kata diubah menjadi huruf kecil. Langkah ini dilakukan untuk menyeragamkan semua kata menjadi huruf kecil, sehingga memudahkan proses identifikasi dan analisis teks selanjutnya. Dengan menyeragamkan huruf, sistem dapat menghindari masalah inkonsistensi yang muncul akibat perbedaan kapitalisasi, yang sering kali dapat mempengaruhi hasil pengolahan data teks, terutama dalam tugas-tugas seperti pencarian kata kunci, pengelompokan, atau klasifikasi teks. Berikut adalah contoh hasil implementasinya yang dapat dilihat pada tabel 4.1. Pada tabel tersebut, kolom sebelah kiri menunjukkan ulasan asli sebelum dilakukan case folding,

sementara kolom sebelah kanan menampilkan ulasan yang sudah melalui tahap case folding, di mana semua huruf besar telah diubah menjadi huruf kecil.

Tabel 4.1 Hasil setelah dilakukan case folding.

Sebelum	Sesudah
Shin Tae Yong Catatkan Dua Capaian Bersejarah untuk Sepak Bola Indonesia	shin tae yong catatkan dua capaian bersejarah untuk sepak bola indonesia
Baru di era Shin Tae Yong juga jadi pede nonton TIMNAS	baru di era shin tae yong juga jadi pede nonton timnas

c. Steaming

Stemming adalah proses mengubah kata-kata menjadi bentuk dasarnya (root form). Tujuannya adalah untuk menyederhanakan kata-kata dengan menghilangkan akhiran dan variasi kata. Proses ini dilakukan secara eksternal dikarenakan pada aplikasi rapidminer masih tidak mendukung untuk steaming menggunakan bahasa Indonesia. Library yang digunakan pada proses kali ini adalah Sastrawi, yaitu sebuah library atau perangkat lunak yang digunakan untuk melakukan stemming pada teks berbahasa Indonesia. Berikut merupakan code program proses steaming dapat dilihat pada gambar 4.5.

```
!pip install pandas sastrawi
import pandas as pd
import re
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Membuat stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# Fungsi untuk melakukan stemming
def stem_text(text):
    return stemmer.stem(text)

# Membaca data dari file CSV
df = pd.read_csv('kualifikasi&pialaasia.csv')

# Normalisasi dan stemming teks
df['cleaned_text'] = df['full_text'].apply(clean_text)
df['stemmed_text'] = df['cleaned_text'].apply(stem_text)

# Menyimpan data yang sudah diproses ke file CSV baru
df.to_csv('hasilsteaming.csv', index=False)
```

Gambar 4.5 Proses steaming data

Hasil dari proses steaming dapat dilihat pada tabel 4.2 berikut. Pada tahap tersebut data sudah bersih dari imbuhan dan sudah menjadi kata dasar, dan dapat digunakan untuk proses berikutnya.

Tabel 4.2 Hasil setelah dilakukan steaming

Sebelum	Sesudah
shin tae yong catatkan dua capaian bersejarah untuk sepak bola indonesia	shin tae yong catat dua capai sejarah untuk sepak bola indonesia
baru di era shin tae yong juga jadi pede menonton timnas	baru di era shin tae yong juga jadi pede nonton timnas

d. Tokenize

Proses selanjutnya adalah tokenize atau proses tokenisasi yaitu proses memisahkan kata pada suatu kalimat menjadi potongan kata atau bisa disebut "token". Hasil proses tokenize dapat dilihat dalam tabel 4.3

Tabel 4.3 Hasil setelah dilakukan tokenize

Sebelum	Sesudah
shin tae yong catatkan dua capaian bersejarah untuk sepak bola indonesia	'shin', 'tae', 'yong', 'catatkan', 'dua', 'capaian', 'bersejarah', 'untuk', 'sepak', 'bola', 'indonesia'
baru di era shin tae yong juga jadi pede nonton timnas	'baru', 'di', 'era', 'shin', 'tae', 'yong', 'juga', 'jadi', 'pede', 'nonton', 'timnas'

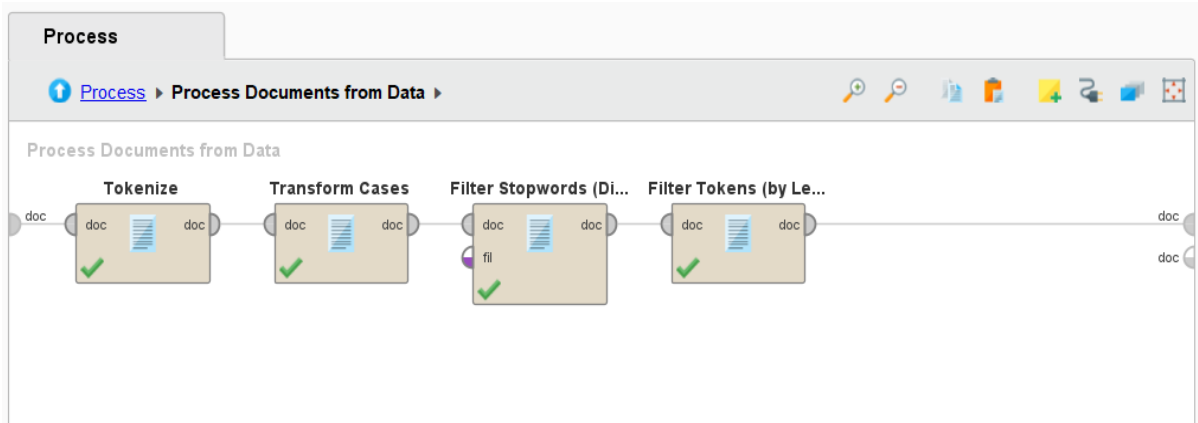
e. Stopword Removal

Proses ini dilakukan untuk menghilangkan kata-kata yang sering digunakan tetapi tidak berpengaruh pada sentimen atau makna kalimat. Penghapusan stopwords dalam penelitian ini menggunakan kamus yang berisi daftar kata-kata umum dalam bahasa Indonesia yang diperoleh dari kaggle.com dan sudah digunakan dalam banyak penelitian. Pada penelitian kali ini ada beberapa kata yang disesuaikan guna untuk menyesuaikan kebutuhan analisis yang diharapkan dapat meningkatkan akurasi dan tidak mengurangi makna. Beberapa kata yang dipertahankan diantaranya adalah belum, akhir, bisa, lagi dan lain sebagainya.

f. Filtering

Dalam tahap ini, dilakukan penghilangan kata-kata yang memiliki huruf terlalu pendek atau yang disingkat serta kata-kata yang terlalu panjang. Tujuannya adalah untuk mengurangi kata-kata yang tidak memberikan kontribusi pada makna atau mengurangi gangguan pada kalimat.

Gambaran proses case folding, tokenize, stopwords removal dan filtering dapat dilihat pada gambar 4.6.



Gambar 4.6 Proses tokenize, transform case, filter stopwords.

Berikut merupakan hasil dari proses preprocessing pada gambar 4.7. Data ini selanjutnya digunakan untuk proses selanjutnya yaitu analisis sentimen.

Row No.	word	in documents	total
1	timnas	2014	2181
2	indonesia	974	1084
3	piala	769	864
4	shin	631	643
5	yong	574	582
6	asia	474	543
7	dunia	308	334
8	pemain	244	266
9	lolos	202	234
10	azizah	205	206
11	selamat	179	179
12	arhan	169	172
13	prestasi	158	163

Gambar 4.7 Hasil proses preprocessing

g. SMOTE

SMOTE atau Synthetic Minority Over-sampling Technique adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. Ketidakseimbangan kelas terjadi ketika jumlah contoh dalam satu kelas jauh lebih banyak atau lebih sedikit dibandingkan dengan kelas lainnya, yang sering kali menimbulkan masalah dalam proses pembelajaran mesin. Ketika dataset tidak seimbang, model cenderung menjadi bias terhadap kelas mayoritas, yang dapat mengakibatkan performa yang buruk dalam memprediksi kelas minoritas. SMOTE mengatasi masalah ini dengan secara sintesis menciptakan contoh-contoh baru untuk kelas minoritas dengan menghasilkan data baru berdasarkan titik-titik yang ada di kelas tersebut.

Teknik ini bekerja dengan cara memilih contoh dari kelas minoritas dan kemudian membuat titik-titik data baru di antara contoh-contoh tersebut dengan interpolasi. Proses ini membantu dalam menyeimbangkan distribusi kelas, sehingga model pembelajaran mesin dapat belajar dengan lebih baik dan memberikan prediksi yang lebih akurat untuk kedua kelas, baik mayoritas maupun minoritas. Dengan menggunakan SMOTE, tidak hanya masalah ketidakseimbangan dapat diatasi, tetapi juga tingkat akurasi dari proses analisis secara keseluruhan dapat ditingkatkan, menghasilkan model yang lebih robust dan dapat diandalkan.

4.3 Pelabelan Kelas Sentimen

Proses selanjutnya setelah menyelesaikan tahap preprocessing adalah pelabelan data sentimen. Proses ini bertujuan untuk mengelompokkan data berdasarkan isi sentimen yang terkandung di dalamnya, dengan tujuan untuk membedakan apakah data tersebut termasuk dalam kategori sentimen positif atau negatif. Pelabelan sentimen ini dilakukan dengan dua pendekatan, otomatis menggunakan model arsitektur transformer, dan manual.

Pada metode otomatis, model arsitektur transformer, seperti BERT (Bidirectional Encoder Representations from Transformers), digunakan untuk melakukan pelabelan sentimen secara efisien. Model ini dilatih untuk memahami konteks kata dalam kalimat dan dapat mengklasifikasikan sentimen dengan akurasi yang tinggi. Setelah proses pelabelan otomatis, hasilnya dibandingkan dengan pelabelan manual yang dilakukan oleh pengamat manusia untuk memastikan keakuratan.

Jika nilai perbandingan antara pelabelan otomatis dan manual menunjukkan kesamaan, data tersebut langsung diteruskan ke tahap analisis berikutnya. Namun, jika terdapat perbedaan antara kedua metode tersebut, data akan melalui proses verifikasi ulang untuk memastikan klasifikasi yang tepat. Pada akhirnya, data akan dikelompokkan menjadi dua kelas sentimen

utama: sentimen positif dan sentimen negatif. Dengan melihat hasil skor dari proses pelabelan ini, penentuan kelas sentimen dapat dilakukan dengan lebih tepat dan terstruktur. Berikut adalah kode untuk proses pelabelan data secara otomatis menggunakan model BERT:

```
import pandas as pd
#import googletrans
!pip install googletrans==3.1.0a0
import googletrans
from googletrans import Translator
#import textblob
import textblob
from textblob import TextBlob
!pip install transformers
import numpy as np
import re

data_path = 'clean.csv'
import chardet
with open(data_path, 'rb') as rawdata:
    result = chardet.detect(rawdata.read(100000))
result

data = pd.read_csv(data_path, encoding='ISO-8859-1')
data.head()

from googletrans.client import Translator
translator = Translator()
from transformers import pipeline
sentiment_classifier = pipeline('sentiment-analysis')
data['label'] = data['full_text'].str.encode('ascii', 'ignore').apply(translator.translate,
src='id', dest='en')
data['label'] = data['label'].apply(getattr, args=('text',))
df = (
    data
    .assign(sentiment = lambda x: x['label'].apply(lambda s:
sentiment_classifier(s)))
    .assign(
        label =lambda x: x['sentiment'].apply(lambda s: (s[0]['label'])),
        score =lambda x: x['sentiment'].apply(lambda s: (s[0]['score'])))
    )
)
df.sample(20)
data_jadi = df.to_csv('data_jadi.csv', index=False)
```

Gambar 4.8 Proses labeling otomatis

Hasil dari proses pelabelan yaitu data terbagi menjadi dua kategori, kategori sentimen positif dan kategori sentimen negatif.

Selain itu juga dilakukan pelabelan secara manual oleh manusia. Pelabelan manual dilakukan oleh manusia yang membaca dan menilai sentimen teks berdasarkan pemahaman mereka. Meskipun memakan waktu dan sumber daya, metode ini sering kali dianggap lebih akurat karena melibatkan pemahaman mendalam dari konteks dan emosi yang mungkin tidak sepenuhnya ditangkap oleh model otomatis. Pelabelan manual juga digunakan untuk membuat set data referensi yang dapat digunakan untuk melatih dan menguji model otomatis.

Setelah 2 data pelabelan dihasilkan yaitu pelabelan menggunakan transformer dan manual, dilakukanlah proses perbandingan antara kedua data untuk memastikan konsistensi dan akurasi. Ketika data memiliki nilai yang sama, maka data langsung masuk pada proses berikutnya, dan ketika memperoleh nilai yang berbeda maka dilakukan verifikasi yang cermat terhadap data yang berbeda, diharapkan menghasilkan hasil analisis sentimen yang akurat dan dapat diandalkan. Gambar perbandingan pelabelan manual dan otomatis dapat dilihat pada gambar 4.9 berikut.

full_text	Pelabelan Manual	Pelabelan Transformer	sentiment score
Lebih baik dipindah Piala Dunia U 20 dari pada menimbulkan keributan di Negeri ini	negatif	NEGATIVE	{{'label': 'N' 0.998769
Reaksi Shin Taeyong	negatif	POSITIVE	{{'label': 'P' 0.9834
Shin Taeyong Gelisah soal Nasib Piala Dunia U20 2023	negatif	NEGATIVE	{{'label': 'N' 0.98918
Pelatih timnas Indonesia Shin Taeyong ikut waswas mendengar kabar tuan rumah Piala Duni	negatif	NEGATIVE	{{'label': 'N' 0.99923
Yg Jelas Mereka2 yg Melarang Tim Olah raga israel bermain yg otaknya gak waras yg tidak bis	negatif	NEGATIVE	{{'label': 'N' 0.981491
STY Gelisah Piala Dunia U20 di Indonesia Terancam Batal	negatif	NEGATIVE	{{'label': 'N' 0.999077
resmi fifa menunjuk peru sebagai tuan rumah piala dunia U 20 menggantikan indon yg masi p	negatif	NEGATIVE	{{'label': 'N' 0.920554
Shin Tae Yong Catatkan Dua Capaian Bersejarah untuk Sepak Bola Indonesia	positif	POSITIVE	{{'label': 'P' 0.999636
Kesel sama orang-orang yang ngeremehin pencapaian coach Shin Tae Yong Dulu waktu kecil	negatif	NEGATIVE	{{'label': 'N' 0.999596
Baru di era Shin Tae Yong juga jadi pede nonton timnas Para pemainnya bisa megang bola Se	positif	POSITIVE	{{'label': 'P' 0.983807
Baru di era Shin Tae Yong nonton timnas main sampai tuntas Biasanya nggak sampai full uda	positif	NEGATIVE	{{'label': 'N' 0.983983
Shin Tae Yong Pelatih yg bisa bawa Indonesia ke Piala Asia di 3 timnas yg berbeda untuk pert	positif	POSITIVE	{{'label': 'P' 0.999017
Pelatih Shin Taeyong, memberikan apresiasi kepada para pemain timnas U23 Indonesia,	positif	POSITIVE	{{'label': 'P' 0.999764
Prestasi coach Shin Tae Yong selama melatih timnas Indonesia yang berhasil membawa Indc	positif	POSITIVE	{{'label': 'P' 0.998112
Alhamdulillah Laga tanding semalam Timnas U23 tolos ke Piala Asia U23 tahun 2024 Ini sejar	positif	POSITIVE	{{'label': 'P' 0.999875
Shin Tae Yong membawa Timnas U20	positif	POSITIVE	{{'label': 'P' 0.998327
Kita boleh stress	negatif	NEGATIVE	{{'label': 'N' 0.978955
shin tae yong berproses ngeri sekali	negatif	POSITIVE	{{'label': 'P' 0.987099

Gambar 4.9 Proses perbandingan labeling manual dan otomatis

Tabel 4.4 Hasil klasifikasi

Kalimat	Sentimen
pertahankan shin tae yong	Positif
era shin tae yong juga jadi pede nonton timnas	Positif
siap siap bantai	Negatif
shin tae yong cetak sejarah	Positif
coach shin tae tong jelek trofi	Negatif

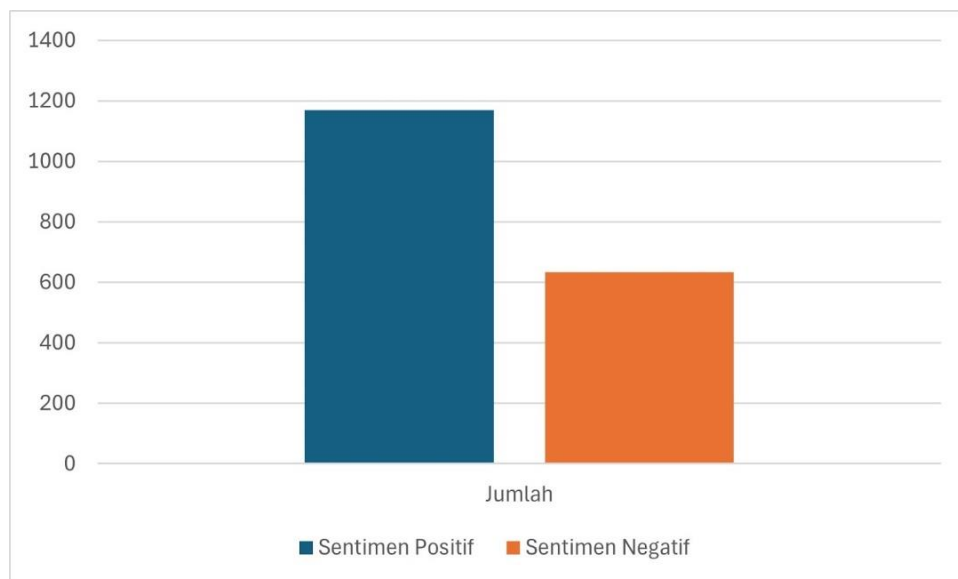
4.3.1 Pada Masa Kualifikasi Piala Asia U 23

Tabel ini menyajikan informasi rinci mengenai jumlah data yang telah diklasifikasikan ke dalam berbagai kategori atau kelas berdasarkan hasil pengolahan data. Tabel jumlah data klasifikasi dapat dilihat dalam tabel berikut.

Tabel 4.5 Data klasifikasi kualifikasi piala asia

Sentimen	Jumlah
Positif	1170
Negatif	633

Berikut merupakan diagram yang menampilkan perbandingan antara data latih sentimen positif dan negatif yang diperoleh dari hasil crawling pada masa kualifikasi Piala Asia U-23. Dalam analisis ini, terdapat total 1803 data latih yang telah melalui tahap labeling, yang kemudian dikelompokkan menjadi dua kategori utama: sentimen positif dan sentimen negatif. Dari keseluruhan data latih, sebanyak 1170 data diklasifikasikan sebagai sentimen positif, sementara 633 data lainnya dikategorikan sebagai sentimen negatif. Diagram ini memberikan gambaran visual yang jelas mengenai distribusi sentimen dalam dataset yang digunakan untuk melatih model analisis sentimen. Untuk melihat representasi visual dari data ini secara lebih jelas, dapat merujuk pada gambar 4.10.



Gambar 4.10 Diagram perbandingan sentimen positif dan negatif

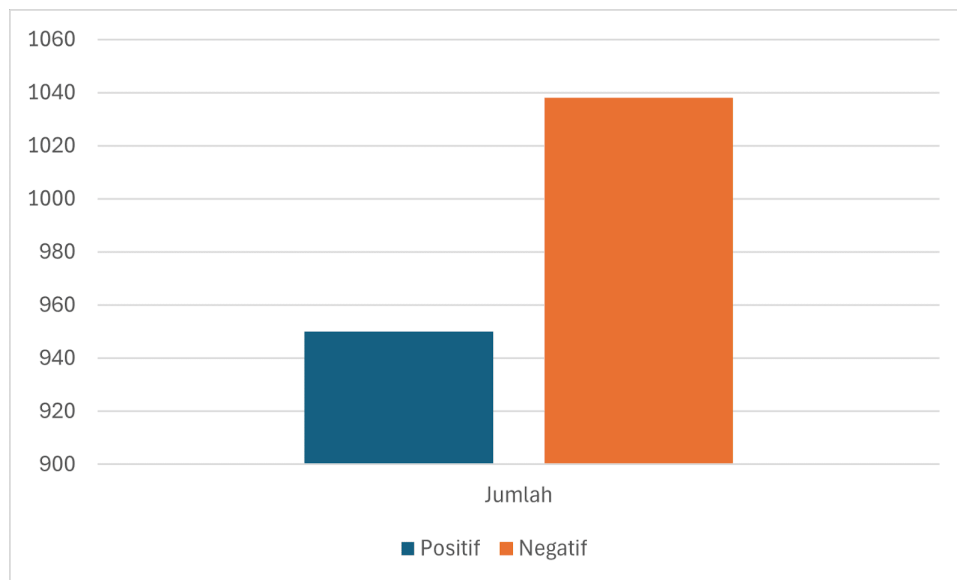
4.3.2 Pada Masa Piala Asia U 23

Tabel ini menyajikan informasi rinci mengenai jumlah data yang telah diklasifikasikan ke dalam berbagai kategori atau kelas berdasarkan hasil pengolahan data. Tabel jumlah data klasifikasi dapat dilihat dalam tabel berikut.

Tabel 4.6 Data klasifikasi piala asia

Sentimen	Jumlah
Positif	950
Negatif	1038

Berikut ini adalah diagram yang menampilkan perbandingan antara data sentimen positif dan negatif yang diperoleh dari hasil crawling pada masa Piala Asia U-23. Dalam analisis ini, terdapat total 1988 data latih yang telah melalui tahap labeling dan diklasifikasikan ke dalam dua kategori utama: sentimen positif dan sentimen negatif. Dari total data latih, sebanyak 1038 data dikategorikan sebagai sentimen negatif, sedangkan 950 data lainnya termasuk dalam kategori sentimen positif. Diagram ini memberikan gambaran visual yang jelas mengenai distribusi sentimen dalam dataset yang digunakan untuk melatih model analisis sentimen. Untuk representasi visual yang lebih rinci, Anda dapat melihatnya pada Gambar berikut.



Gambar 4.11 Diagram perbandingan sentimen positif dan negatif

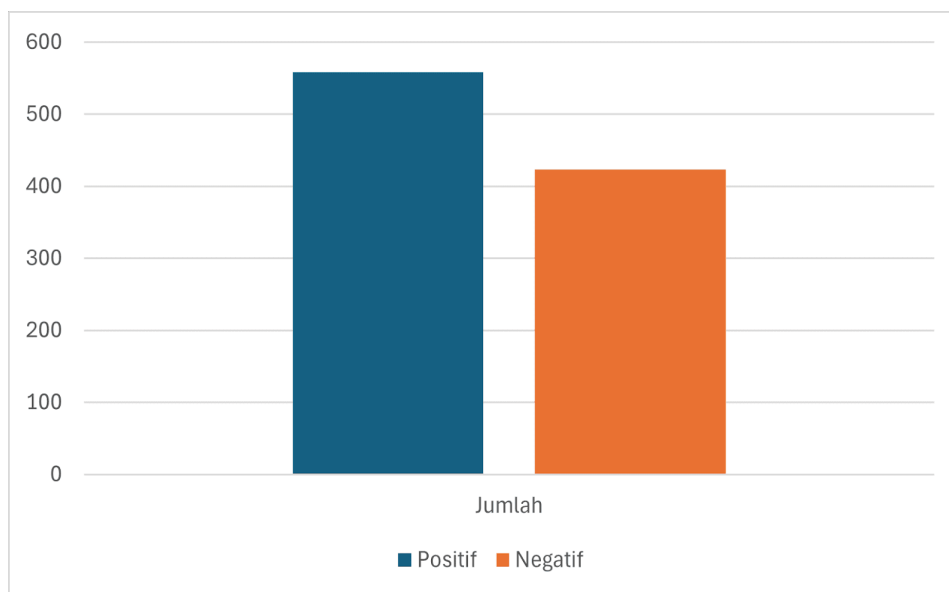
a. Sentimen Timnas Menang

Jumlah perbandingan data sentimen ketika Timnas memperoleh kemenangan dapat dilihat pada tabel berikut :

Tabel 4.7 Sentimen Indonesia menang

Sentimen	Jumlah
Positif	558
Negatif	423

Berikut ini adalah data yang dikumpulkan ketika Timnas Indonesia memperoleh kemenangan. Dari hasil pengumpulan data, terdapat total 981 data yang kemudian diklasifikasikan menjadi dua jenis sentimen utama: sentimen positif dan sentimen negatif. Dari total tersebut, sebanyak 558 data dikategorikan sebagai sentimen positif, yang menunjukkan dukungan, apresiasi, atau kepuasan terhadap kemenangan Timnas. Sementara itu, 423 data lainnya termasuk dalam kategori sentimen negatif, yang mungkin mencerminkan kritik, ketidakpuasan, atau pandangan negatif meskipun dalam konteks kemenangan.



Gambar 4.12 Diagram perbandingan sentimen positif dan negatif Timnas menang

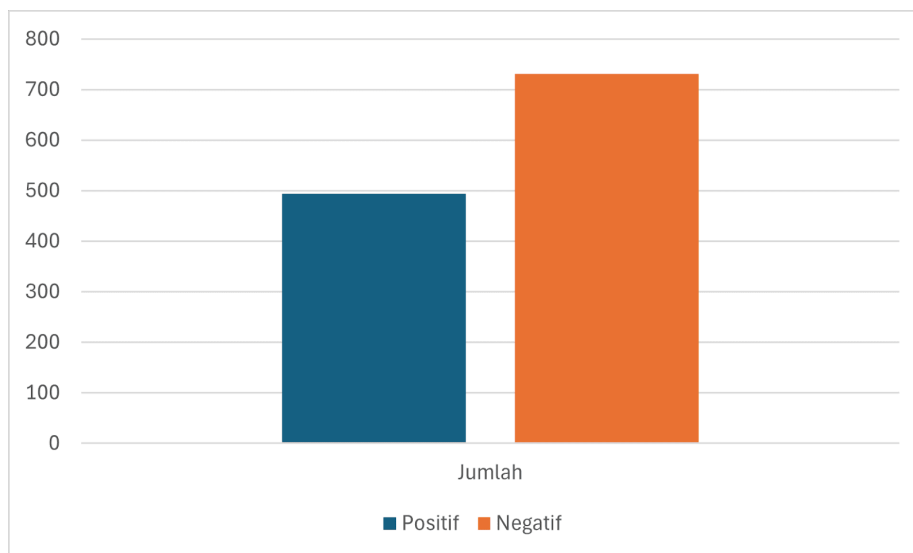
b. Sentimen Timnas Kalah

Jumlah perbandingan data sentimen ketika timnas memperoleh kekalahan dapat dilihat pada tabel berikut :

Tabel 4.8 Sentimen Indonesia kalah

Sentimen	Jumlah
Positif	494
Negatif	731

Berikut ini adalah data yang dikumpulkan ketika Timnas Indonesia mengalami kekalahan. Dari total 1225 data yang terkumpul, data tersebut terbagi menjadi dua kategori sentimen: 494 data menunjukkan sentimen positif, dan 731 data mengindikasikan sentimen negatif. Meskipun Timnas mengalami kekalahan, kehadiran sentimen positif yang cukup signifikan mungkin menunjukkan dukungan dan optimisme dari para pendukung yang tetap melihat sisi positif dari pertandingan atau performa tim. Namun, jumlah sentimen negatif yang lebih besar mengindikasikan kritik, kekecewaan, atau ketidakpuasan terhadap hasil atau aspek lain dari pertandingan tersebut.



Gambar 4.13 Diagram perbandingan sentimen positif dan negatif Timnas kalah

Jika kita melihat hasil generatannya, semakin besar ukuran kata yang muncul menandakan bahwa kata tersebut paling sering muncul dalam unggahan terkait timnas sepak bola Indonesia. Insight yang diambil setelah melakukan generasi menggunakan word cloud ini adalah bahwa kata-kata dengan ukuran paling besar atau paling sering muncul meliputi "Piala", "Shin", dan "Asia". Hal ini memberikan gambaran bahwa pembahasan seputar piala, shin, dan Asia memiliki pengaruh yang signifikan dalam unggahan terkait sentimen timnas sepak bola Indonesia.

4.4.1 Word Cloud Positif

Visualisasi word cloud untuk sentimen positif digunakan untuk melihat dan menganalisis kata-kata yang paling sering muncul dalam teks yang diberi label sentimen positif. Dengan menggunakan word cloud, kata-kata yang lebih sering digunakan akan ditampilkan dengan ukuran yang lebih besar, memberikan gambaran visual yang cepat dan intuitif mengenai tema atau topik yang dominan dalam sentimen positif tersebut. Teknik ini sangat berguna untuk mengidentifikasi pola, isu, atau kata-kata kunci yang mungkin tidak terlihat dengan jelas dalam analisis teks tradisional, sehingga mempermudah dalam memahami faktor-faktor yang mungkin berkontribusi terhadap sentimen positif dalam dataset.



Gambar 4.16 Hasil visualisasi word cloud positif

Kata kata positif yang sering muncul disaat label sentimen bernilai positif diantaranya adalah piala, asia, lolos, prestasi, selamat. Kata kata tersebut berukuran besar pada paket word cloud menandakan banyaknya kalimat yang mengandung kata tersebut.

bermanfaat untuk mengidentifikasi kata-kata atau frasa yang berhubungan dengan prestasi, strategi, pemain kunci, atau perasaan positif yang diungkapkan oleh penggemar dan media. Dengan demikian, word cloud ini tidak hanya membantu dalam memahami reaksi dan sentimen publik, tetapi juga memberikan wawasan mengenai aspek-aspek yang paling dihargai atau diingat dalam kemenangan tersebut.



Gambar 4.18 Hasil visualisasi word cloud timnas menang

Kata-kata yang sering muncul ketika tim nasional memperoleh kemenangan, seperti Indonesia, menang, nonton, pemain, dan asia, ditampilkan dalam word cloud dengan ukuran yang lebih besar. Ukuran besar pada word cloud ini menandakan bahwa kata-kata tersebut sering muncul dalam kalimat-kalimat yang membahas kemenangan timnas. Dengan visualisasi ini, kita dapat dengan mudah melihat kata-kata kunci yang sering diungkapkan oleh publik dan media saat merayakan kemenangan, serta memahami elemen-elemen yang paling banyak diperhatikan atau disebutkan dalam konteks positif. Sebagai contoh, kata Indonesia mungkin sering muncul karena merujuk pada identitas nasional, sementara kata menang jelas menunjukkan tema utama dari keberhasilan tersebut. Kata-kata lain seperti nonton dan pemain mungkin mencerminkan pengalaman dan aspek penting dari perayaan kemenangan.

4.4.4 Word Cloud Timnas Kalah

Visualisasi word cloud Ketika timnas memperoleh kekalahan digunakan untuk melihat dan menganalisis kata-kata yang sering muncul dalam teks yang terkait dengan kekalahan tim nasional. Dalam word cloud ini, kata-kata yang paling sering digunakan akan ditampilkan dengan ukuran yang lebih besar, memberikan gambaran visual yang cepat mengenai topik atau tema utama yang dihubungkan dengan kekalahan timnas. Visualisasi ini sangat berguna untuk

mengidentifikasi pola kata atau frasa yang sering muncul dalam konteks kekalahan, seperti alasan di balik kekalahan, kritik terhadap performa tim atau pemain, serta reaksi negatif dari penggemar dan media. Dengan menggunakan word cloud, kita dapat memahami lebih baik sentimen dan perspektif publik terkait kekalahan, serta faktor-faktor yang dianggap berkontribusi terhadap hasil tersebut.



Gambar 4.19 Hasil visualisasi word cloud timnas kalah

Kata kata yang sering muncul disaat Timnas memperoleh kekalahan diantaranya adalah pemain, kalah, nonton, target, lokal. Kata kata tersebut berukuran besar pada paket word cloud menandakan banyaknya kalimat yang mengandung kata tersebut.

4.5 Data Latih dan Data Uji

Pembentukan model model classifier dilakukan dengan menggunakan data latih yang nantinya model tersebut merupakan perwakilan pengetahuan hasil pembelajaran dari data yang telah diolah, yang selanjutnya dapat digunakan untuk memprediksi data baru yang akan dicari menggunakan model klasifikasi. Tingkat akurasi yang diperoleh berbanding lurus dengan jumlah data latih yang digunakan, semakin banyak data yang digunakan, maka akan semakin baik hasil yang diperoleh mesin dalam memahami pola pola dalam data baru tersebut. Data uji digunakan untuk mengukur sejauh mana keakuratan dalam melakukan klasifikasi dengan data baru yang ada. Dalam konteks ini, jumlah data uji yang digunakan adalah gabungan dari masa kualifikasi dan masa piala asia, total data yang digunakan adalah 3595 data dengan perbandingan pengujian antara data latih dan data uji adalah sebanyak 90% untuk data latih dan 10% untuk data uji. Hal ini berarti bahwa model akan belajar dari mayoritas data (90%)

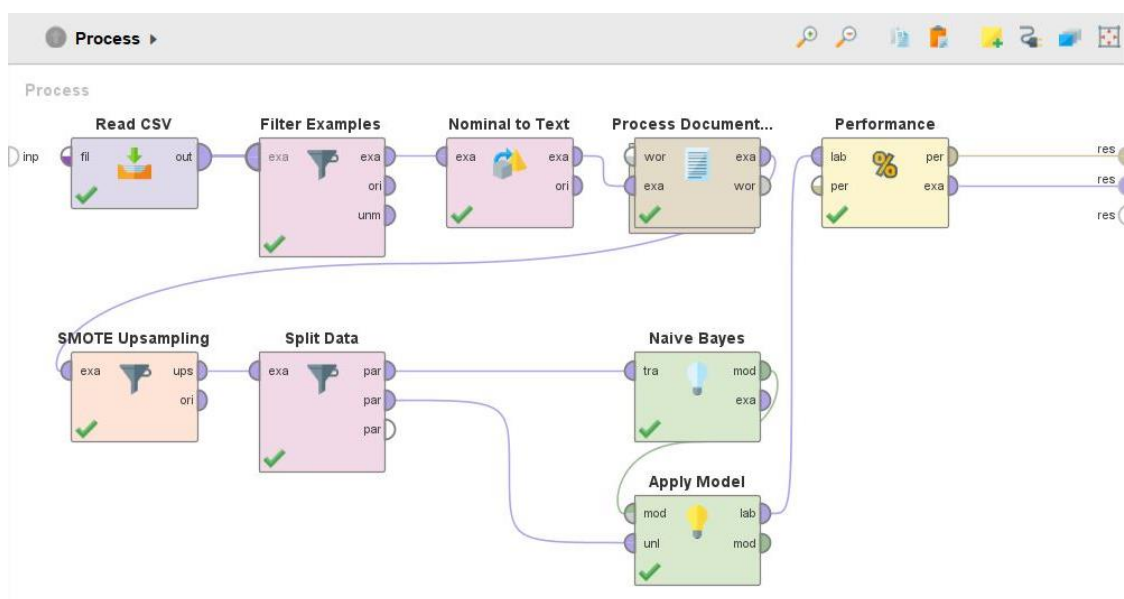
dan diuji pada sebagian kecil data yang tidak digunakan dalam proses pembelajaran (10%). Hal ini dipilih karena berdasarkan penelitian sebelumnya, jumlah tersebut memperoleh hasil terbaik dibandingkan dengan perbandingan yang lain.

4.6 Naive Bayes

Dalam penelitian ini, metode yang diterapkan adalah Naive Bayes Classification. Data set dibagi menjadi dua bagian utama, yaitu data uji dan data latih. Pembagian ini dilakukan dengan rasio test dan train sebesar 10% untuk data latih dan 90% untuk data uji. Data set yang dipakai adalah gabungan data pada masa kualifikasi dan piala asia dengan total data yaitu 3595 data, dengan banyaknya dataset yang digunakan pada dataset pengujian ini diharapkan dapat memperoleh hasil akurasi yang maksimal.

Berikut merupakan proses analisis naïve bayes yang dilakukan, setelah data melalui beberapa tahapan sebelumnya diantaranya adalah proses cleaning dan steaming, barulah data dapat diproses pada tahap analisis menggunakan naïve bayes. Pada tahap ini beberapa tahapan yang dilakukan adalah proses dokumen dengan melakukan beberapa hal yaitu tokenisasi, stopwords dan lain sebagainya, data masuk pada tahap SMOTE Upsampling guna untuk menyeimbangkan data set yang ada antara data positif dan negatif. Hal ini dilakukan untuk mengurangi kecenderungan model untuk menganalisis data yang dominan, dan dapat melakukan analisis dengan data set yang seimbang.

Untuk gambar proses pembuatan analisis ada pada gambar 4.20.



Gambar 4.20 Proses Naive Bayes

Untuk tampilan hasil dari data uji yang telah melalui proses analisis dapat dilihat pada gambar 4.21.

Row No.	text	Sentimen	prediction(Sentimen)	confidence(...	confidence(...
1	pindah piala dunia timbul ribut negeri	negatif	negatif	1	0
2	reaksi shin taeyong	negatif	negatif	1	0
3	latih shin taeyong apresiasi main timnas indonesia	positif	positif	0	1
4	kontrak coach shin yong	positif	negatif	1	0
5	wajib ikan kontrak coach shin yong	positif	positif	0	1
6	sejarah shin yong bawa timnas indonesia piala asi	positif	positif	0	1
7	selamat coach shin yong terima kasih dedikasi ti...	positif	positif	0	1
8	buah sabar shin yong believe process buah poton.	positif	positif	0	1
9	terima kasih sebanyakbanyaknya coach shin yon...	positif	positif	0	1
10	terima kasih shin yong	positif	negatif	1	0
11	angkat topi coach shin yong masterclass	positif	positif	0	1
12	shin yong kontrak umur hidup jugakk	positif	positif	0	1
13	shin yong	positif	negatif	1	0
14	latih shin yong trust process	positif	positif	0	1
15	ngomong shin yong gagal gara juara ngotak dapet	negatif	negatif	1	0
16	percaya proses shin yong bener bawa indonesia l.	positif	positif	0	1
17	sejarah shin yong	positif	positif	0	1

Gambar 4.21 Hasil proses data uji

Table View Plot View

accuracy: 87.69%

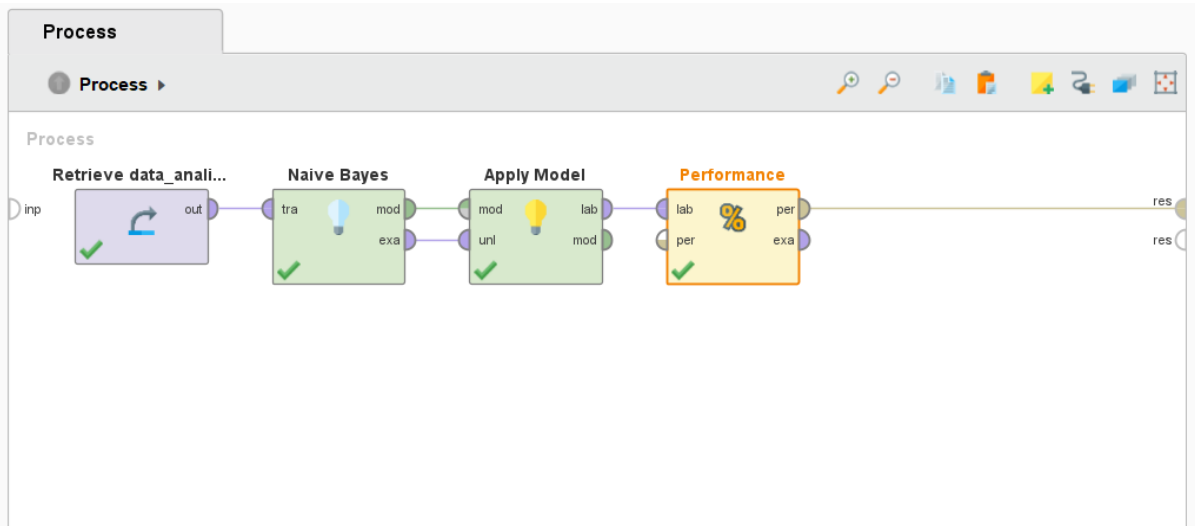
	true negatif	true positif	class precision
pred. negatif	258	62	80.62%
pred. positif	2	198	99.00%
class recall	99.23%	76.15%	

Gambar 4.22 Hasil nilai akurasi

Setelah melakukan pengujian dengan data gabungan dari masa kualifikasi dan piala asia didapatkan hasil akurasi sebesar 87.69% dengan perbandingan data latih dan data uji sebesar 90% data latih dan 10% data uji. Hal ini membuktikan dengan banyaknya data yang digunakan yaitu 3595 data yang merupakan gabungan data dari masa kualifikasi dan masa piala asia dapat menghasilkan tingkat akurasi relatif besar.

4.7 Evaluasi Sistem

Pada tahap evaluasi dilakukan perhitungan nilai akurasi yang ditujukan untuk mengetahui tingkat ketepatan penggunaan model. Proses evaluasi sistem ditunjukkan pada gambar :



Gambar 4.23 Proses tahap evaluasi.

Table View Plot View

accuracy: 90.34%

	true negatif	true positif	class precision
pred. negatif	990	345	74.16%
pred. positif	2	2254	99.91%
class recall	99.80%	86.73%	

Gambar 4.24 Confusion matrix algoritma Naive Bayes.

Pada tahap evaluasi seperti pada gambar 4.24 hasil yang diperoleh dari penggunaan metode naive bayes ini berupa nilai confusion matrix berisi nilai akurasi, presisi, dan recall yang diambil dari data test. Nilai akurasi yang didapat algoritma Naive Bayes adalah 90,34%, artinya sejumlah 90,34% dapat mengklasifikasikan data dengan benar. Berdasarkan penelitian yang telah dilakukan, sentimen positif terhadap Timnas Indonesia lebih banyak dibandingkan dengan sentimen negatif, yang berarti bahwa masyarakat memiliki tingkat percaya lebih kepada Timnas Indonesia.

Selain itu digunakan data baru untuk menguji kinerja sistem apakah bisa melakukan klasifikasi dengan baik atau tidak. Pengujian yang dilakukan bertujuan untuk memvalidasi atau menganalisis data ulasan, apakah ulasan tersebut termasuk ke dalam sentimen positif atau negatif. Berikut merupakan hasil dari pengujian data baru dari model yang sudah terbentuk.

Row No.	text	prediction(S...	confidence(...	confidence(...
1	timnas indonesia menang	positif	0	1
2	menang	positif	0	1
3	gagal raih hasil impresif	negatif	1	0
4	fakta timnas menang taiwan	negatif	1	0
5	indonesia bisa menang	positif	0	1
6	suport kalah	negatif	1	0
7	hasil puas	positif	0	1
8	jelek banget njing wkwkw nanggung beban	negatif	1	0
9	angka cantik cetak	positif	0	1
10	kecewa main	negatif	1	0
11	lawan cina taipei bagus	positif	0	1
12	calon lawan timnas fase grup piala asia	negatif	1	0
13	wataru endo kalo timnas ganteng	positif	0	1
14	timnas indonesia hasil menang telak	negatif	1	0
15	main main jelek dipanggil	positif	0	1
16	bisa cetak timnas	positif	0	1
17	selebrasi nyuruh supprter diem laga timnas	negatif	1	0

Gambar 4.25 Hasil pengujian data baru

Berdasarkan gambar diatas menunjukkan bahwa sistem dapat melakukan analisis sentimen dengan baik yang menggunakan data baru. Meskipun ada beberapa nilai pelabelan yang kurang sesuai, tetapi sebagian mendapat nilai pelabelan yang benar.

4.8 Analisis Hasil Sentimen

Hasil analisis sentimen yang dilakukan terhadap opini masyarakat di Twitter mengenai Timnas Sepak Bola Indonesia era kepemimpinan Shin Tae-yong pada dua periode berbeda yaitu saat kualifikasi Piala Asia dan saat pelaksanaan Piala Asia. Diperoleh distribusi sentimen sebagai berikut :

Tabel 4.9 Hasil sentimen

Sentimen	Masa Kualifikasi	Masa Piala asia
Positif	64.8%	47.7%
Negatif	35.2%	52,3%

4.8.1 Tindakan Berdasarkan Sentimen

Menentukan tindakan berdasarkan sentimen positif atau negatif yang lebih dominan adalah langkah krusial dalam proses pengambilan keputusan yang informatif dan berbasis data. Dengan menganalisis sentimen yang mendominasi, organisasi atau individu dapat memperoleh wawasan yang lebih mendalam tentang persepsi dan reaksi audiens terhadap situasi tertentu. Sentimen positif atau negatif yang lebih dominan dapat mencerminkan tingkat kepuasan, kekecewaan, harapan, atau kritik dari berbagai pemangku kepentingan. Dengan memahami pola-pola ini, tindakan yang diambil dapat lebih tepat sasaran dan berkontribusi pada perbaikan atau penguatan aspek-aspek tertentu. Berikut adalah beberapa tindakan yang dapat diambil tergantung pada dominasi sentimen:

a. Pada Tahap Kualifikasi

Tingginya sentimen positif pada tahap kualifikasi menunjukkan dukungan yang kuat dari masyarakat. Langkah-langkah yang dapat diambil untuk memanfaatkan sentimen positif ini antara lain:

1. Penguatan Identitas Tim: Memperkuat identitas dan citra positif tim melalui kampanye media sosial yang menampilkan momen-momen terbaik selama kualifikasi, testimonial dari pemain, dan dukungan dari tokoh publik atau influencer.
2. Pengembangan Program Fans: Mengembangkan program loyalitas atau keanggotaan penggemar yang memberikan akses eksklusif ke konten, merchandise khusus, dan acara pertemuan dengan pemain, memanfaatkan momentum positif dari tahap kualifikasi.
3. Kolaborasi dengan Sponsor: Menggunakan sentimen positif dari tahap kualifikasi untuk menarik sponsor baru dan memperkuat hubungan dengan sponsor yang ada. Sentimen positif dapat digunakan sebagai bukti kuat bahwa tim memiliki basis penggemar yang aktif dan mendukung.

b. Pada Tahap Pelaksanaan Piala Asia

Penurunan sentimen positif dan peningkatan sentimen negatif pada tahap pelaksanaan Piala Asia menunjukkan adanya ketidakpuasan yang perlu ditangani. Langkah-langkah yang dapat diambil untuk merespons banyaknya sentimen negatif antara lain:

1. Identifikasi Isu Utama: Menggunakan analisis teks lanjutan untuk mengidentifikasi masalah utama yang menyebabkan penurunan sentimen positif, seperti kinerja di lapangan, keputusan manajemen, atau perilaku pemain.

2. **Transparansi dan Komunikasi:** Meningkatkan transparansi dalam komunikasi dengan penggemar mengenai rencana perbaikan yang diambil selama pelaksanaan Piala Asia. Ini bisa dilakukan melalui konferensi pers, posting blog, atau video dari manajemen tim yang menjelaskan langkah-langkah yang diambil untuk mengatasi masalah.
3. **Inisiatif Perbaikan Cepat:** Mengimplementasikan tindakan perbaikan cepat yang terlihat oleh penggemar, seperti perubahan taktik permainan, peningkatan pelatihan, atau tindakan disiplin jika diperlukan. Menunjukkan komitmen untuk perbaikan dapat membantu mengubah persepsi negatif yang muncul selama pelaksanaan Piala Asia.
4. **Survei dan Feedback:** Melakukan survei atau meminta umpan balik langsung dari penggemar untuk memahami lebih lanjut kekhawatiran mereka dan melibatkan mereka dalam proses perbaikan. Ini membantu membangun kembali kepercayaan penggemar.

4.8.2 Kesimpulan Hasil Sentimen

Analisis sentimen terhadap opini masyarakat di Twitter mengenai Timnas Sepak Bola Indonesia menunjukkan adanya perubahan yang signifikan antara tahap kualifikasi dan pelaksanaan Piala Asia. Pada tahap kualifikasi, dominasi sentimen positif (64.8%) mencerminkan antusiasme dan dukungan kuat dari masyarakat, yang dapat dimanfaatkan untuk memperkuat identitas tim dan memperluas program loyalitas penggemar. Namun, pada tahap pelaksanaan Piala Asia, penurunan sentimen positif menjadi 47% dan peningkatan sentimen negatif menjadi 53% mengindikasikan adanya ketidakpuasan dan kekecewaan yang perlu segera ditangani.

Untuk mengatasi tantangan ini, tim manajemen harus fokus pada identifikasi isu utama yang menyebabkan ketidakpuasan, meningkatkan transparansi dan komunikasi dengan penggemar, serta mengimplementasikan tindakan perbaikan cepat. Langkah-langkah ini penting untuk membangun kembali kepercayaan dan dukungan dari penggemar, yang pada gilirannya akan mendukung kesuksesan jangka panjang Timnas Sepak Bola Indonesia. Dengan pendekatan yang proaktif dan responsif, diharapkan hubungan antara tim dan penggemar dapat diperkuat dan citra positif tim dapat dipulihkan.

BAB V

KESIMPULAN DAN SARAN

Bab ini akan membahas kesimpulan yang diambil dari hasil penelitian, yang melibatkan analisis proses, algoritma, dan implementasi sistem yang telah dikembangkan. Selain itu, saran juga akan disampaikan untuk memberikan catatan dan rekomendasi perbaikan yang dapat diterapkan dalam penelitian ini untuk penelitian yang akan datang.

Kesimpulan dari penelitian ini akan mencakup rangkuman temuan utama, signifikansi hasil, dan relevansi dengan tujuan penelitian. Dalam bagian saran, dapat memberikan rekomendasi untuk perbaikan atau pengembangan lebih lanjut, baik dalam hal metodologi penelitian, perbaikan implementasi teknis, atau peningkatan performa algoritma. Saran ini dapat memberikan arah bagi penelitian selanjutnya atau memberikan kontribusi pada bidang pengetahuan yang relevan. Bab kesimpulan dan saran sangat penting karena memberikan gambaran menyeluruh tentang kontribusi penelitian, potensi pengembangan, dan arah penelitian di masa depan.

5.1 Kesimpulan

Kesimpulan tersebut memberikan gambaran tentang metode yang digunakan, tingkat akurasi yang dicapai, serta langkah-langkah yang diambil dalam melakukan analisis sentimen pada data yang bersangkutan.

Beberapa kesimpulan yang dapat diambil dari penelitian yang telah dilakukan sebagai berikut :

- a. Penggunaan Algoritma Naive Bayes: Penelitian ini memanfaatkan algoritma Naive Bayes, yang menghasilkan tingkat akurasi model sebesar 90,34%, dan nilai akurasi pengujian sebesar 87.69% dengan perbandingan data latih dan data uji sebesar 90% : 10%. Hal ini menunjukkan bahwa algoritma naive bayes efektif dalam melakukan klasifikasi sentimen pada data yang digunakan.
- b. Tahapan Analisis Sentimen: Proses analisis sentimen dalam penelitian ini melibatkan beberapa tahapan, yaitu:
 1. Tahap Pengumpulan Data: Data dikumpulkan sebagai langkah awal dalam penelitian untuk membangun model klasifikasi.
 2. Tahap Preprocessing: Proses preprocessing digunakan untuk membersihkan dan mengolah data agar sesuai dengan kebutuhan analisis.

3. Tahap Steaming : merupakan langkah untuk mengubah kata-kata menjadi bentuk dasarnya (root form). Tujuannya adalah untuk menyederhanakan kata-kata dengan menghilangkan akhiran dan variasi kata.
 4. Tahap Pelabelan Data: Data dilabeli untuk mengidentifikasi sentimen positif dan negatif.
 5. Tahap Klasifikasi: Klasifikasi sentimen dilakukan menggunakan algoritma Naive Bayes untuk memprediksi kategori sentimen dari teks.
 6. Tahap Pengujian: Pengujian dilakukan untuk mengevaluasi kinerja model yang telah dibangun dan mengukur tingkat akurasi hasil klasifikasi.
- c. Hasil perbandingan sentimen pada masa Kualifikasi Piala Asia U 23 dan masa Piala Asia U 23 adalah perbandingan sentimen positif lebih dominan pada saat momen kualifikasi daripada sentimen negatif, yaitu sebanyak 65.% positif dan sentimen negatif sebanyak 35%. Sementara itu perbandingan sentimen pada masa Piala Asia U 23 adalah sebanyak 48% sentimen positif dan 52% sentimen negatif.
- d. Perbandingan sentimen timnas ketika memperoleh kemenangan dan kekalahan memiliki perbedaan. Sentimen positif timnas ketika memperoleh kemenangan adalah 57% dan sentimen negatif sebanyak 43%. Sementara itu sentimen timnas ketika memperoleh kekalahan adalah 40% sentimen positif dan 60% sentimen negatif.

5.2 Saran

Berdasarkan hasil analisis dan kesimpulan, terdapat beberapa saran yang dapat diberikan untuk pengembangan sistem lebih lanjut:

a. Penyeimbangan Dataset

Dalam penelitian selanjutnya, diharapkan dataset yang digunakan memiliki perbandingan kelas yang seimbang antara jumlah data positif, dan negatif. Hal ini akan membantu meningkatkan keakuratan dan keobjektifan model dalam melakukan klasifikasi. Jumlah data set juga dapat ditambah untuk meningkat nilai akurasi.

b. Pendekatan Machine Learning Alternatif

Penelitian selanjutnya dapat mempertimbangkan penggunaan pendekatan machine learning lain sebagai pembanding performa terhadap algoritma Naive bayes. Hal ini akan memberikan perspektif lebih luas tentang keunggulan dan kelemahan berbagai metode.

c. Penambahan Kategori Sentimen Netral

Untuk meningkatkan keakuratan analisis sentimen, penelitian selanjutnya dapat mempertimbangkan penambahan kategori sentimen netral. Hal ini penting karena tidak semua teks dapat dengan jelas diklasifikasikan sebagai positif atau negatif, dan adanya kategori netral akan membuat hasil lebih representatif.

Saran-saran tersebut diharapkan dapat memberikan arah bagi penelitian selanjutnya dan meningkatkan kualitas serta ketepatan hasil analisis sentimen pada sistem yang dikembangkan.

DAFTAR PUSTAKA

- Arifianto, N. (2017, Desember 19). Indonesia Negara Penggila Sepak Bola Nomor Dua di Dunia. Retrieved from CNN Indonesia: <https://www.cnnindonesia.com/olahraga/20171219204103-142-263606/indonesianegara-penggila-sepak-bola-nomor-dua-di-dunia>
- M. Astiningrum, M. Haniah, and Y. rahmat yoga Pradana, “Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter,” *Semin. Inform. Apl. Polinema*, p. 35—39, 2020.
- A. Cahyo and M. Pratama, “Lahirnya Kembali Pssi (Persatuan Sepak Bola Seluruh Indonesia) Tahun 1950-1954,” vol. 10, no. 1, 2020.
- W. Setyobudi, A. Alwi, and I. P. Astuti, “Sentimen Analisis Twitter Terhadap Penyelenggaraan Gojek Traveloka Liga 1 Indonesia,” *Komputek*, vol. 2, no. 1, p. 56, 2018, doi: 10.24269/jkt.v2i1.68.
- J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- H. Basri, “PERANMEDIA SOSIAL TWITTER DALAM INTERAKSI SOSIAL PELAJAR SEKOLAH MENENGAH PERTAMA DI KOTA PEKANBARU (studi kasus pelajar SMPN 1 kota Pekanbaru),” *Strateg. Bertahan Hidup Petani Penggarap Di Jorong Sarilamak Nagari Sarilamak Kec. Harau Kabupaten Lima Puluh Kota*, vol. 4, no. 1, pp. 1–13, 2017, [Online]. Available: <https://media.neliti.com/media/publications/183768-ID-partisipasi-masyarakat-dalam-pelaksanaan.pdf>.
- C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, “Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes,” *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
- M. B. Rissan and R. F. Hassan, “Naïve-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 1, pp. 375–383, 2022, doi: 10.11591/ijeecs.v28.i1.pp375-383.
- S. D. Raharjo and G. Wang, “Paper Review Text Mining Twitter,” 2020.
- A. Chamekh, M. Mahfoudh, and G. Forestier, “Sentiment Analysis Based on Deep Learning in E-Commerce,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell.*

- Lect. Notes Bioinformatics), vol. 13369 LNAI, pp. 498–507, 2022, doi:10.1007/978-3-031-10986-7_40.
- P. Mehta and S. Pandya, “A review on sentiment analysis methodologies, practices and applications,” *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- Z. Drus and H. Khalid, “Sentiment analysis in social media and its application: Systematic literature review,” *Procedia Comput. Sci.*, vol. 161, pp. 707–714, 2019, doi: 10.1016/j.procs.2019.11.174.
- H. Nurrun Muchammad Shiddieqy, S. Paulus Insap, and W. Wing Wahyu, “Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. March, pp. 57–64, 2016.
- N. Salmi and Z. Rustam, “Naïve Bayes Classifier Models for Predicting the Colon Cancer,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052068.
- D. Sharma, M. Sabharwal, V. Goyal, and M. Vij, “Sentiment analysis techniques for social media data: A review,” *Adv. Intell. Syst. Comput.*, vol. 1045, no. September, pp. 75–90, 2020, doi: 10.1007/978-981-15-0029-9_7.
- F. F. Rachman and S. Pramana, “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter,” *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175>.
- A. Nugroho, “Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 200, 2018, doi: 10.30645/j-sakti.v2i2.83.
- L. A. Andika, P. A. N. Azizah, and R. Respatiwan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- R. Feldman and J. Sanger, *The Text Mining Handbook*. 2006.
- F. D. Ananda and Y. Pristyanto, “Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Twitter Users on Internet Service Providers Using Support Vector Machine Algorithm,” vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.

- I. Purnamasari, “Faktor Pendorong Fanatisme Pada Suporter Klub Sepak Bola Arsenal di Balikpapan,” *Psikoborneo J. Ilm. Psikol.*, vol. 3, no. 4, pp. 354–362, 2015, doi: 10.30872/psikoborneo.v3i4.3876.
- H. Liang and J. J. H. Zhu, “Big Data, Collection of (Social Media, Harvesting),” *Int. Encycl. Commun. Res. Methods*, no. November, pp. 1–18, 2017, doi: 10.1002/9781118901731.iecrm0015.
- Van den Rul, C. (2019). How to Generate Word Clouds in R. towards data science. <https://towardsdatascience.com/create-a-word-cloud-with-r-bde3e7422e8a>
- A. I. Kadhim, “An Evaluation of Preprocessing Techniques for Text Classification,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 6, pp. 22–32, 2018, [Online]. Available: <https://sites.google.com/site/ijcsis/>.
- I. Indriati, M. Marji, and S. Pakpahan, “Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 7259–7267, 2019.
- Y. Akbar and T. Sugiharto, “Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes (Yuma Akbar 1*, Tri Sugiharto 2) Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes,” *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 115–122, 2023, [Online]. Available: <https://doi.org/10.55338/saintek.v4i3.1368>.
- A. S. Talaat, “Sentiment analysis classification system using hybrid BERT models,” *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00781-w.

LAMPIRAN

Dataset Keseluruhan :

<https://docs.google.com/spreadsheets/d/16jHiGWgYEP5-oDu3VSeF4LsjYjG0-a8t/edit?usp=sharing&ouid=109550453943619723408&rtpof=true&sd=true>

Dataset Label :

<https://drive.google.com/file/d/1B-YH-x4KU-QocB9ZAVK3hdwfT2rJasLZ/view?usp=sharing>

Stopword Bahasa :

https://drive.google.com/file/d/1klQhYK7sbtvp4NDACW532oARTL5Enecm/view?usp=drive_link

Source Code Crawling Data :

https://colab.research.google.com/drive/1sEftmpP_F6eECvjqB6MntOz47u7JGs-S?usp=sharing

Source Code Rapid Miner:

<https://drive.google.com/file/d/1p21NiTcDcLlZXkUdI5sKVTUbjvH4SXFq/view?usp=sharing>