

**ANALISIS SENTIMEN TENTANG PERSIJA “*TO THE NEXT LEVEL*” PADA MEDIA SOSIAL TWITTER (X) MENGGUNAKAN METODE NAIVE BAYES**

**SKRIPSI**

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



Oleh :

**FANIYA RIFQI FAUZI**

NIM : 2008096051

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



## PERNYATAAN KEASLIAN

Yang bertandatangan dibawah ini:

Nama : Faniya Rifqi Fauzi

NIM : 2008096051

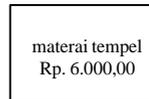
Jurusan : Teknologi Informasi

Menyatakan bahwa skripsi yang berjudul:

**Analisis Sentimen Tentang Persija “TO THE NEXT LEVEL” Pada Media Sosial Twitter (X) Menggunakan Metode Naive Bayes.**

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

Semarang, 12 Februari 2024  
Pembuat Pernyataan,



Faniya Rifqi Fauzi  
NIM : 2008096051



## PERSETUJUAN PEMBIMBING

Proposal Skripsi ini telah disetujui oleh Pembimbing untuk dilaksanakan. Disetujui pada:

Hari : *Senin*

Tanggal : *12 Februari 2021*

**Pembimbing I**

**Pembimbing II**



Dr. Masy Ari Ulinuha, ST MT  
NIP. 198108122011011007



Heri Mustofa, M.Kom.  
NIP. 198703172019031007

**Mengetahui,**

**Ketua Jurusan Teknologi Informasi**



Nur Cahyo Hendro Wibowo, S.T M.Kom  
NIP. 197312222006041001





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

---

### PENGESAHAN

Naskah proposal skripsi berikut ini:

Judul : Analisis Sentimen Tentang Persija “TO THE NEXT LEVEL”  
Pada Media Sosial Twitter (X) Menggunakan Metode  
Naive Bayes.  
Penulis : Faniya Rifqi Fauzi  
NIM : 2008096051  
Jurusan : Teknologi Informasi

Telah diujikan dalam seminar proposal skripsi oleh Dewan  
Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat  
diterima sebagai salah satu syarat memperoleh gelar sarjana  
dalam bidang ilmu Teknologi Informasi.

Semarang, 26 Februari 2024

### DEWAN PENGUJI

Penguji I

Nur Cahyo Hendrowibowo, S.T., M.Kom  
NIP. 197312222006041001

Penguji II

Hery Mustofa, M.Kom  
NIP. 198703172019031007

Penguji III

Siti Nur'aini, M.Kom  
NIP. 198401312018012001

Penguji IV

Adzal Arwani Mahfudh, M.Kom  
NIP. 199107032019031006





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

### PENGESAHAN

Naskah skripsi berikut ini:

Judul : Analisis Sentimen Tentang Persija "TO THE NEXT LEVEL" Pada Media Sosial Twitter (X) Menggunakan Metode Naive Bayes.  
Penulis : **Faniya Rifqi Fauzi**  
NIM : 2008096051  
Jurusan : Teknologi Informasi

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

Semarang,

### DEWANPENGUJI

Penguji I,

**Nur Cahyo Hendro Wibowo, S.T M.Kom**  
NIP : 197312222006041001

Penguji II,

**Hery Mustofa, M.Kom**  
NIP : 198703172019031007

Penguji III,

**Siti Nur'aini, M.kom**  
NIP : 198401312018012001

Penguji IV,

**Adzhal Arwani Mahfudh, M.Kom**  
NIP : 199107032019031006

Pembimbing I,

**Dr. Masy Ari Ulinuha, S.T M.T**  
NIP : 198108122011011007

Pembimbing II,

**Hery Mustofa, M.Kom**  
NIP : 198703172019031007





## NOTA DINAS

Semarang,

Yth. Ketua Program Studi  
Teknologi Informasi Fakultas  
Sains dan Teknologi UIN  
Walisongo Semarang

*Assalamu'alaikum. wr. wb.*

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

Judul : Analisis Sentimen Tentang Persija  
"TO THE ENXT LEVEL" Pada Media  
Sosial Twitter (X) Menggunakan  
Metode Naive Bayes.

Nama : **Faniya Rifqi Fauzi**

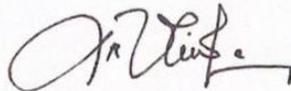
NIM : 2008096051

Jurusan : Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqsyah.

*Wassalamu'alaikum. wr. wb.*

Pembimbing I,



**Dr. Masy Ari Ulinuha, S.T M.T.**

NIP : 198108122011011007



## NOTA DINAS

Semarang,

Yth. Ketua Program Studi  
Teknologi Informasi Fakultas  
Sains dan Teknologi UIN  
Walisono Semarang

*Assalamu'alaikum. wr. wb.*

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

Judul : Analisis Sentimen Tentang Persija  
"TO THE ENXT LEVEL" Pada Media  
Sosial Twitter (X) Menggunakan  
Metode Naive Bayes.

Nama : **Faniya Rifqi Fauzi**

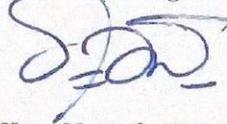
NIM : 2008096051

Jurusan : Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisono untuk diujikan dalam Sidang Munaqsyah.

*Wassalamu'alaikum. wr. wb.*

Pembimbing II,



**Hery Mustofa, M.Kom**  
NIP : 198703172019031007



## ABSTRAK

Sepak bola adalah salah satu olahraga paling populer di seluruh dunia. Di Indonesia, Persija Jakarta adalah salah satu klub sepakbola terbesar di Indonesia dengan basis supporter/*fans* yang besar dan loyalitas. Sepak bola modern telah melewati fase perubahan signifikan dalam hal komunikasi dan keterlibatan penggemar. Dalam penelitian ini analisis sentimen diterapkan untuk menganalisis sentimen dari respon pengguna media sosial twitter tentang kebijakan tersebut untuk dilakukan pengklasifikasian komentar yang didapatkan, apakah komentar tersebut bersifat positif, negatif dan netral.

Proses awal penelitian ini melakukan pengambilan data tweet di media sosial twitter(X) dengan kata kunci ' Persija tothenextlevel'. Selanjutnya data tersebut dilakukan proses *text preprocessing*, dan pembobotan kata TFIDF (*term frequency Inverse Document Frequency*). Penelitian ini menggunakan metode *naïve bayes classifier* dalam melakukan klasifikasi data dari media sosial twitter yang terbagi menjadi tiga sentimen, yaitu positif, negative, dan netral.

Berdasarkan hasil pengujian yang telah dilakukan bahwa hasil dari model *naïve bayes classifier* didapatkan sentimen positif memiliki nilai presentase tertinggi sebesar 87,2%, pada sentimen negative sebesar 7,2%, dan sentimen netral sebesar 5,6%. Kemudian dilakukan *split validation data* dengan perbandingan data latih dan data uji sebesar 80:20 yang diambil secara acak. Sehingga didapatkan nilai performa dari model naïve bayes classifier dengan klasifikasi BAIK dengan Tingkat akurasi sebesar 82%, precision 79%, recall sebesar 83%, dan f1 score sebesar 77%.

Kata Kunci : Persija tothenextlevel, twitter(X), analisis sentimen, *naïve bayes classifier*



## KATA PENGANTAR

Alhamdulillah, Puji Syukur atas kehadiran Allah SWT atas segala nikmat dan karunia-nya sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Analisis Sentimen Tentang Persija “TO THE NEXT LEVEL” Pada Media Sosial Twitter (X) Menggunakan Metode Naive Bayes” dengan baik. Dengan tujuan dibuat skripsi ini sebagai salah satu bentuk syarat kelulusan pada program sarjana (S1) prodi Teknologi Informasi di Universitas Islam Negeri Walisongo Kota Semarang. Di sisi lain, penulis juga bertujuan untuk memberikan pengetahuan kepada pembaca.

Pada kesempatan ini penulis ingin mengucapkan banyak-banyak terimakasih kepada seluruh pihak yang memberi dukungan dan bantuan dari pelaksanaan skripsi hingga penyelesaian skripsi ini. Penulis mengakui bahwa apabila tanpa bimbingan, arahan, serta binaan dan tanpa diberikan motivasi dari seluruh pihak, maka penulisan skripsi ini tidak berjalan dengan baik. Maka dari itu penulis mengucapkan terimakasih yang tidak terhingga kepada :

1. Bapak Khotibul Umam, S.T, M.Kom, selaku ketua program studi Teknologi Informasi Universitas Islam Negeri Walisongo Semarang.

2. Bapak Dr. Masy Ari Ulinuha, S.T M.T, dan Bapak Hery Mustofa, M.Kom, selaku dosen pembimbing skripsi saya yang selalu memberikan dukungan, arahan, bimbingan serta motivasi dalam pelaksanaan skripsi hingga pembuatan skripsi ini.
3. Kepada seluruh bapak ibu dosen Teknologi Informasi yang telah dan tanpa Lelah untuk memberikan ilmu-nya baik ilmu perkuliahan atau ilmu kehidupan kepada penulis.
4. Kedua orang tua, abang, dan keluarga yang selalu menemani dalam membantu penulis dan selalu mendo'akan serta memberikan dukungan baik kepada penulis.
5. Keluarga besar mahasiswa Teknologi Informasi dari angkatan 2018-2023, khususnya teman-teman Teknologi Informasi angkatan 2020 yang selalu mendukung, membantu perjalanan di dunia perkuliahan penulis.
6. Keluarga besar HMJB UIN Walisongo yang menjadi tempat pulang disaat penulis sedang berada dititik terendah di dunia perkuliahan.
7. Keluarga besar kontrakan Bahagia, terkhusus bang sapri, bang asdut, bang bakri, bang nijar, bang owi, bang sner, bang maul, bang uta, bang iyad, bang evan, bang

Fadhil, bang agoy, bang fathan, bang bendol, bang kahis, mba agus, mba dira, mba muna, mba diah ayu safitri, yang selalu ada, yang selalu menghiasi hari-hari penulis, yang menjadi tempat suka maupun duka penulis, dan menjadi support system penulis dalam dunia perkuliahan dan khususnya dalam penulisan skripsi ini.

8. Persija Jakarta yang mengusung tagar TO THENEXYLEVEL sebagai bahan penelitian ini, dan The Jakmania yang sudah beropini dan berpendapat mengenai tagar TO THENEXTLEVEL dimedia sosial tetter(X) yang menjadi objek penelitian ini.
9. Semua pihak yang tidak bisa penulis sebutkan satu persatu yang terlibat dalam pembuatan skripsi ini sehingga dapat terselesaikan dengan baik.

Dalam pelaksanaan dan penyusunan skripsi, penullis menyadari bahwa tentunya masih jauh dari kata sempurna dan masih banyak kekurangan. Untuk itu, penulis sangat mengharapkan kritik serta saran yang membangun demi kesempurnaan penulisan skripsi ini, dan semoga skripsi ini dapat bermanfaat untuk semua pihak.

Semarang, 25 Mei 2024

Penulis



## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>PERNYATAAN KEASLIAN.....</b>	<b>iii</b>
<b>PERSETUJUAN PEMBIMBING.....</b>	<b>v</b>
<b>LEMBAR PENGESAHAN PROPSAL SKRIPSI .....</b>	<b>vii</b>
<b>PENGESAHAN .....</b>	<b>Error! Bookmark not defined.</b>
<b>NOTA DINAS.....</b>	<b>xi</b>
<b>NOTA DINAS.....</b>	<b>xiii</b>
<b>ABSTRAK.....</b>	<b>xv</b>
<b>KATA PENGANTAR.....</b>	<b>xvii</b>
<b>DAFTAR ISI .....</b>	<b>xxi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xxv</b>
<b>DAFTAR TABEL.....</b>	<b>xxvii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
A. Latar Belakang.....	1
B. Identifikasi Masalah.....	6
C. Rumusan Masalah.....	6
D. Batasan Masalah.....	7
E. Tujuan Penelitian.....	7
F. Manfaat Penelitian .....	8
<b>BAB II LANDASAN PUSTAKA.....</b>	<b>9</b>
A. Landasan Teori.....	9
1. Analisis Sentimen.....	9
2. Media sosial.....	11

3.	Twitter(X).....	13
4.	Persija Jakarta “ <i>To The Next Level</i> ” .....	14
5.	<i>Text Mining</i> .....	16
6.	<i>Crawling data</i> .....	17
7.	<i>Text preprocessing</i> .....	18
8.	TFIDF.....	20
9.	Klasifikasi .....	22
10.	<i>Split validation data</i> .....	23
11.	<i>Naive bayes classification</i> .....	23
12.	Evaluasi .....	25
	B. Kajian Penelitian yang Relevan.....	31
	<b>BAB III METODOLOGI PENELITIAN</b> .....	37
	A. Metode Pengumpulan Data .....	37
1.	Data sekunder.....	37
2.	Data primer.....	37
	B. Perangkat Penelitian .....	38
1.	Kebutuhan perangkat keras .....	38
2.	Kebutuhan perangkat lunak .....	38
	C. Alur Penelitian .....	39
1.	Pengumpulan Data .....	39
2.	Preprocessing .....	41
3.	Klasifikasi .....	44
	<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	47
	A. <i>Crawling Data Tweet</i> .....	47
	B. <i>Preprocessing Data</i> .....	52
1.	<i>Case folding</i> .....	52

2. <i>Cleaning</i> .....	54
3. <i>Remove duplicate</i> .....	57
4. <i>Normalization</i> .....	59
5. <i>Stopword removal</i> .....	60
6. <i>Stemming</i> .....	62
7. <i>Tokenization</i> .....	64
C. Ektaksi fitur.....	67
D. Klasifikasi Naive Bayes.....	74
E. Uji Model.....	79
F. Evaluasi Model.....	80
G. Visualisasi.....	85
<b>BAB V KESIMPULAN DAN SARAN</b> .....	89
A. Kesimpulan.....	89
B. Saran.....	90
<b>DAFTAR PUSTAKA</b> .....	91
<b>Daftar Lampiran</b> .....	97
Lampiran 1 : Data Crawling.....	97
Lampiran 2 : Data labelling.....	99
Lampiran 3 : data tambahan “slangword.txt”.....	101
Lampiran 4 : Data Testing.....	103
<b>DAFTAR RIWAYAT HIDUP</b> .....	106
A. Identitas Diri.....	106
B. Riwayat Pendidikan.....	106
C. Prestasi Akademik.....	106
D. Karya Ilmiah.....	106



## DAFTAR GAMBAR

Gambar 2. 1 Data tren internet dan sosial media di Indonesia (sumber: hootsuite 2023).....	12
Gambar 3. 1 <i>Flowchart</i> alur penelitian.....	39
Gambar 3. 2 <i>Crawling</i> data.....	40
Gambar 3. 3 Proses <i>removal duplicate</i> .....	42
Gambar 4. 1 Proses <i>Crawling</i> Data.....	47
Gambar 4. 2 Jenis Data Yang dikumpulkan.....	48
Gambar 4. 3 Data Yang Berhasil Dikumpulkan.....	49
Gambar 4. 4 Hasil Data yang telah diberi Label.....	51
Gambar 4. 5 Proses <i>case folding</i> .....	53
Gambar 4. 6 Hasil <i>case folding</i> .....	53
Gambar 4. 7 Proses <i>Cleaning</i> Data.....	54
Gambar 4. 8 Proses clear emoji.....	55
Gambar 4. 9 Hasil clear emoji.....	56
Gambar 4. 10 Hasil proses <i>cleaning</i> .....	57
Gambar 4. 11 Proses <i>remove duplicate</i> .....	57
Gambar 4. 12 Hasil <i>remove duplicate</i> .....	58
Gambar 4. 13 Proses <i>normalization</i> .....	59
Gambar 4. 14 Hasil <i>normalization</i> .....	60
Gambar 4. 15 Proses <i>stopword removal</i> .....	61
Gambar 4. 16 Hasil <i>stopword removal</i> .....	61
Gambar 4. 17 Proses install library sastrawi dan swifter.....	62
Gambar 4. 18 Proses <i>stemming</i> data.....	62
Gambar 4. 19 Hasil proses <i>stemming</i> .....	64
Gambar 4. 20 Menginstal library nltk.....	64
Gambar 4. 21 Proses <i>Tokenization</i> .....	65
Gambar 4. 22 Hasil <i>Tokenization</i> .....	66
Gambar 4. 23 Hasil <i>akhir preprocessing</i> .....	67
Gambar 4. 24 Proses pembagian data.....	68
Gambar 4. 25 Proses pembobotan TFIDF.....	69
Gambar 4. 26 Sampel hasil TFIDF.....	70

Gambar 4. 27 Klasifikasi Naive Bayes.....	75
<i>Gambar 4. 28 Multiclass Confusion Matrix.....</i>	<i>80</i>
Gambar 4. 29 Tingkat akurasi klasifikasi Naive Bayes .....	84
Gambar 4. 30 Akurasi Naive Bayes .....	86
Gambar 4. 31 Grafik jumlah sentimen.....	87
Gambar 4. 32 Wordcloud .....	88

## DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i> .....	26
Tabel 2. 2 <i>Muticlass Confusion Matrix</i> .....	28
Tabel 2. 3 Penelitian terkait.....	31
Tabel 3. 1 Kebutuhan perangkat keras .....	38
Tabel 3. 2 Kebutuhan perangkat lunak .....	38
Tabel 3. 3 Penerapan <i>labelling</i> .....	41
Tabel 3. 4 Proses <i>case folding</i> .....	41
Tabel 3. 5 Proses <i>cleansing</i> .....	42
Tabel 3. 6 Proses <i>normalization</i> .....	43
Tabel 3. 7 Proses <i>stopword</i> .....	43
Tabel 3. 8 Proses <i>stemming</i> .....	44
Tabel 3. 9 Proses <i>tokenization</i> .....	44
Tabel 4. 1 Tabel perhitungan TF.....	71
Tabel 4. 2 Proses IDF .....	72
Tabel 4. 3 Proses Perhitungan TF-IDF.....	73
Tabel 4. 4 Tabel <i>Multiclass Confusion Matrix</i> .....	81
Tabel 4. 5 Perhitungan kelas negative.....	82
Tabel 4. 6 Perhitungan kelas netral.....	82
Tabel 4. 7 Perhitungan kelas positif .....	83
Tabel 4. 8 hasil perhitungan performa.....	83



# BAB I

## PENDAHULUAN

### A. Latar Belakang

Sepak bola adalah salah satu olahraga paling populer di seluruh dunia. Di Indonesia, Persija Jakarta adalah salah satu klub sepak bola terbesar di Indonesia dengan basis supporter/*fans* yang besar dan loyalitas. Sepak bola modern telah melewati fase perubahan signifikan dalam hal komunikasi dan keterlibatan penggemar. Media sosial seperti Twitter (X) telah menjadi platform penting bagi penggemar untuk berbagi pendapat, emosi, dan pandangan mereka tentang klub kebanggaan mereka (Mahendra & Prasetyo, 2021).

Sepakbola adalah salah satu olahraga yang sangat populer di Indonesia. Sejak dulu, sepakbola telah menjadi bagian penting dalam budaya dan hiburan masyarakat Indonesia. Liga sepakbola Indonesia memiliki basis penggemar yang besar, dan pertandingan-pertandingannya sering kali menyedot perhatian publik. Liga sepakbola utama di Indonesia adalah Liga 1, yang merupakan kompetisi tertinggi dalam sistem liga sepakbola Indonesia. Beberapa klub sepakbola terkenal di Indonesia termasuk Persija Jakarta, Arema FC, Persib Bandung, dan banyak lagi.

Pertandingan antara klub-klub ini seringkali menjadi sorotan dan menyedot perhatian banyak penggemar sepakbola di Indonesia. Sepakbola di Indonesia tidak hanya menjadi olahraga, tetapi juga menjadi bagian dari identitas dan semangat komunitas (Ferianto, 2023).

Persija Jakarta adalah sebuah klub sepak bola yang berbasis di Jakarta, Indonesia. Klub ini didirikan pada 28 November 1928 dan menjadi salah satu klub sepak bola tertua di Indonesia. Persija Jakarta dikenal sebagai salah satu klub paling sukses dan populer di Tanah Air. Persija Jakarta memiliki basis penggemar yang besar dan fanatik yang dikenal sebagai "Jakmania". Persija Jakarta telah meraih banyak kesuksesan dalam sepak bola Indonesia, termasuk memenangkan berbagai kompetisi domestik. Klub ini juga pernah berpartisipasi dalam kompetisi tingkat Asia seperti Piala AFC. Selain itu, Persija Jakarta memiliki rivalitas kuat dengan klub-klub lain di Indonesia, terutama dengan klub Persib Bandung, yang dikenal sebagai "Derby Indonesia" atau "El Clasico Indonesia" (Damayanti & Utami, 2023). Pada musim 2021-sekarang Persija Jakarta mengusung tagline #TOTHENEXTLEVEL. Arti "To The Next Level" adalah tekad Persija dalam meningkatkan kualitas di berbagai aspek untuk menuju level yang lebih tinggi. Mulai dari tim, manajemen, fasilitas, akademi, hingga berharap

dapat berkontribusi meningkatkan level sepak bola nasional.

The Jakmania adalah sebutan untuk suporter klub sepak bola Persija Jakarta. Nama "Jakmania" berasal dari kata "Jak" yang merupakan singkatan dari Jakarta, ibu kota Indonesia, dan "mania" yang berarti kegilaan atau kecintaan yang sangat besar. Sebagai suporter fanatik, The Jakmania mendukung Persija Jakarta dengan penuh semangat dalam setiap pertandingan. Mereka dikenal karena dedikasi dan kehadiran mereka yang besar di stadion saat pertandingan, serta kreasi-kreasi koreografi yang mengesankan. The Jakmania juga terlibat dalam berbagai kegiatan sosial dan amal di luar lapangan untuk memberikan dampak positif pada masyarakat. Sepak bola di Indonesia memiliki suporter yang sangat antusias, dan The Jakmania adalah salah satu kelompok suporter paling terkenal dan berpengaruh di Tanah Air (Mulfi, 2019).

Media sosial telah mengubah cara suporter/*fans* sepak bola berinteraksi dengan klub, pemain, dan sesama suporter/*fans*. Kehadiran klub sepak bola di platform seperti Twitter (X) memungkinkan mereka untuk berkomunikasi langsung dengan suporter/*fans*, mengumumkan berita, dan merespons peristiwa penting dalam waktu nyata. Ini juga memberi suporter/*fans*

kesempatan untuk mengungkapkan dukungan atau kritik mereka terhadap klub dan pemain (Pasya *et al.*, 2022).

Analisis sentimen adalah teknik yang digunakan untuk memahami pendapat, perasaan, dan reaksi publik terhadap suatu topik atau entitas tertentu. Dalam konteks sepak bola, analisis sentimen dapat digunakan untuk memahami bagaimana penggemar merespons kinerja klub, manajemen, pemain, dan peristiwa penting lainnya. Hal ini dapat memberikan wawasan berharga bagi klub dan pengambil keputusan dalam mengelola reputasi dan hubungan dengan suporter/*fans* (Widowati & Sadikin, 2021).

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap tweets yang berhubungan dengan klub sepak bola Persija Jakarta dalam tagar "*To The Next Level.*" Dengan metode Naive Bayes, penelitian ini akan mencoba mengidentifikasi dan mengklasifikasikan sentimen positif, negatif, atau netral dari tweets penggemar. Hasil analisis sentimen ini dapat memberikan pemahaman yang lebih mendalam tentang persepsi suporter/*fans* terhadap upaya klub untuk mencapai "*To The Next Level.*"

Persoalan analisis sentimen terhadap opini supporter/*fans* ini perlu untuk diperhatikan oleh kaum muslim karena Allah berfirman dalam Al-Qur'an surah Al-Hujurat ayat 6. Allah SWT, berfirman :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ  
فَتُصِيبُكُمْ عَلَىٰ مَا فَعَلْتُمْ نَدِيمِينَ ﴿٦﴾

Artinya : *"Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti, agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu".* (QS Al-Hujarat 49: Ayat 6).

Dari ayat tersebut menjelaskan bahwa jika kita memperoleh suatu opini dari seseorang yang belum diketahui kejelasannya maka diharuskan melakukan *tabayyun*. Apabila seseorang tidak melakukan *tabayyun*, maka dapat menimpa keburukan kepada diri sendiri maupun orang lain (Kurniasih & Suseno, 2022). Dari penjelasan ayat tersebut, maka penulis melakukan penelitian yang salah satu fungsinya adalah *tabayyun* dengan mengumpulkan berbagai macam opini dari masyarakat lalu menganalisisnya sehingga mendapatkan hasil yang berguna bagi masyarakat.

Penelitian ini akan membatasi analisis sentimen pada tweets yang mengandung kata kunci tertentu yang terkait dengan tagar "*To The Next Level*" Persija Jakarta di Twitter (X). Data yang dianalisis akan mencakup periode tertentu yang relevan dengan tagar tersebut.

Dengan latar belakang yang jelas dan berfokus pada pentingnya analisis sentimen dalam konteks klub sepak bola, penelitian ini diharapkan dapat memberikan kontribusi yang berharga bagi pemahaman tentang hubungan antara klub dan suporter/*fans* melalui media sosial.

## **B. Identifikasi Masalah**

Berdasarkan latar belakang di atas, penulis telah mengidentifikasi permasalahan adalah "dibutuhkannya analisis sentimen untuk menganalisa opini suporter/*fans* Persija Jakarta mengenai tagar "*To The Next Level*" yang dibicarakan di media sosial Twitter(X).

## **C. Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan di atas, maka dapat diambil suatu rumusan masalah yaitu ;

1. Bagaimana sentimen *fans* Persija di media sosial Twitter(X) terhadap tagar "*To The Next Level*" Persija Jakarta?

2. Bagaimana performa analisis sentimen metode Naive Bayes dapat digunakan untuk menganalisa sentimen mengenai tagar "*To The Next Level*" Persija Jakarta di media sosial Twitter?

#### **D. Batasan Masalah**

Adapun beberapa batasan masalah dalam penelitian ini adalah :

1. Data yang diambil adalah data dari media sosial twitter(X) yang berbahasa indonesia.
2. Sentimen analisis dilakukan dengan klasifikasi dari metode Naive bayes.
3. Tanggapan twitter akan diklasifikasikan menjadi tiga sentimen yaitu sentimen positif, negatif, dan netral.
4. Kata kunci pencarian yang diteliti pada data twitter yaitu Persija *To The Next Level*.
5. Menggunakan bahasa pemrograman pyhton dan software google collab.
6. Data yang diambil dari twitter(x) mulai pada tanggal 1 Januari 2021 – 1 Januari 2023.

#### **E. Tujuan Penelitian**

Berdasarkan rumusan masalah tersebut tujuan dalam penelitian ini adalah :

1. Melakukan analisis sentimen pada media sosial twitter(X) terhadap tagar "*To The Next Level*" Persija Jakarta menggunakan metode Naive Bayes.
2. Mengetahui performa analisis sentimen positif, negatif, dan netral pada data set twitter(X) mengenai tagar "*To The Next Level*" Persija Jakarta.

#### **F. Manfaat Penelitian**

Adapun beberapa manfaat dari penelitian ini, yaitu :

1. Berkontribusi pada literatur penelitian analisis sentimen dengan mengaplikasikan metode Naive Bayes pada konteks sepak bola. Ini dapat membantu memperluas pemahaman tentang penggunaan analisis sentimen dalam industri olahraga.
2. Menunjukkan potensi untuk meningkatkan keterlibatan supporter/*fans* kepada tim kebanggaan melalui media sosial twitter(X).
3. Mengembangkan pemahaman tentang penggunaan metode analisis sentimen, khususnya metode Naive Bayes, dalam konteks media sosial.
4. Memberikan wawasan tentang bagaimana supporter/*fans* merespon tagar "*To The Next Level*" Persija Jakarta di media sosial twitter(X).

## **BAB II**

### **LANDASAN PUSTAKA**

#### **A. Landasan Teori**

##### **1. Analisis Sentimen**

Analisis sentimen adalah proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber. Analisis sentimen berperan sebagai alat yang dapat menghubungkan seluruh data tersebut. Dengan begitu, perusahaan dapat memperoleh masukan inti dari pengguna atau konsumen secara efisien (Krisdiyanto, 2021).

Analisis sentimen merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang membangun system untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk dalam format forum, blog, media sosial, serta situs berisi *review*. Dengan bantuan analisis sentimen, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur.

Analisis sentimen kemudian akan membedakan teks menjadi dua kategori, yakni fakta dan opini. Fakta merupakan ekspresi objektif mengenai sesuatu. Sementara opini adalah ekspresi subjektif yang menggambarkan sentimen, perasaan, maupun penghargaan terhadap suatu hal (Samsir *et al.*, 2021).

Secara umum, analisis sentimen terbagi menjadi lima langkah yaitu *crawling data*, *pre-processing*, *feature selection*, *classification*, dan *evaluation*. Analisis sentimen dapat mengubah data tidak beraturan menjadi data yang tersusun rapi. Manfaat adanya analisis sentimen yaitu sebagai evaluasi dan ide pada berbagai bidang. Analisis sentimen dapat menganalisis suatu kejadian, pernyataan, dan komentar yang kontroversi. Hasil dari analisis sentimen juga dapat menjadi sebuah gambaran bagi perusahaan, *public figure*, dan pemerintahan untuk menentukan langkah selanjutnya (Natasuwarna, 2020)

Terdapat beberapa jenis analisis sentimen yaitu *emotion detection*, *aspect-based sentiment analysis*, dan *fine grand sentiment analysis*. *Fine sentiment analysis* adalah jenis analisis yang memiliki penilaian spesifik dan biasa digunakan pada bidang e-

*commerce. Emoticon detection* adalah jenis analisis yang bertujuan untuk mengetahui emosi yang ada pada pesan misalnya emosi bahagia, sedih, marah, dan lain-lain. *Aspect-based sentiment analysis* merupakan jenis analisis untuk mengetahui aspek yang berpengaruh dan penilaian dari pelanggan (Jeong *et al.*, 2019).

## 2. Media sosial

Di era digital media sosial semakin banyak digunakan di Indonesia bahkan di dunia. Media sosial merupakan media berbasis online yang digunakan untuk berinteraksi secara individu maupun kelompok dengan efektif dan efisien yang bersifat tak terbatas. Media sosial sebagai tempat mengajak yang dapat mengubah pandangan dan perilaku publik. Setiap tahunnya, media sosial selalu berkembang pesat dan sudah familiar terutama di kalangan generasi Y dan Z (Ahmad, 2020).

Data tren internet dan media sosial 2023 di Indonesia menurut Hootsuite (*we are social*) ditunjukkan dalam gambar 2.1.



Gambar 2. 1 Data tren internet dan sosial media di Indonesia (sumber: hootsuite 2023)

Berdasarkan gambar diatas, terdapat populasi jumlah penduduk di Indonesia sejumlah 276,4 juta, pengguna *mobile unique* 353,8 juta, pengguna internet 212,9 juta dan pengguna media sosial 167,0 juta. Media sosial juga dibagi menjadi tiga jenis yaitu (Hermawansyah & Pratama, 2021) :

- a. Platform pesan singkat, seperti aplikasi whatsapp, facebook mesenger, line, dan telegram.
- b. Jejaring sosial, seperti facebook, dan twitter(X) yang paling banyak digunakan oleh pengguna.
- c. Jaringan berbagi media, seperti youtube dan instagram.

### 3. Twitter(X)

Salah satu media yang banyak digunakan adalah twitter atau yang sekarang lebih dikenal dengan (X). Twitter(X) merupakan media sosial untuk bertukar pikiran dan pendapat. Pengguna twitter(X) dapat mengirim dan menerima pesan tweet berupa teks, gambar, ataupun video. Perbedaan dengan media sosial lain, di twitter(X) karakter untuk menulis pesan dibatasi sampai 280 karakter, sedangkan media sosial lainnya tidak dibatasi. Twitter(X) bersifat publik sehingga status yang dibagikan dapat dilihat oleh orang lain meskipun bukan pengikutnya. Namun, pengiriman tweet juga dapat dibagikan hanya kepada temannya saja atau *followers*. Twitter(X) mempunyai kelebihan yaitu jangkauan yang luas, dapat menjangkau publik figur, media promosi lebih luas, banyak jaringan, dan lebih mudah diukur kemampuannya. Berikut fitur yang ada pada twitter(X) (Alizah *et al.*, 2020) :

- a. *Trending topic* adalah fitur yang menampilkan topik atau pembahasan teratas berupa hastag yang banyak dibicarakan pengguna twitter(X).
- b. *Hastag* adalah fitur yang dapat mengelompokkan tweet atau pesan.

- c. *Retweet* adalah fitur untuk membagikan tweet dari pengguna lain.
- d. *Following* adalah fitur untuk menghubungkan antar pengguna atau sering disebut teman.

Data twitter(X) dapat diambil menggunakan aplikasi yang dihubungkan dan twitter(X). Jika dibandingkan dengan media sosial lainnya, tidak mudah untuk mengumpulkan data secara terbuka. Media sosial lainya tidak mengizinkan data akses karena kebijakan keamanan yang berbeda-beda. Selain itu, twitter(X) juga mempunyai beberapa kecocokan dengan data mining, sebagai berikut (Wandani, 2021) :

- a. Format data twitter(X) yang cocok dan nyaman bagi peneliti untuk dianalisis.
  - b. Peraturan twitter(X) untuk data relatif fleksibel jika dibandingkan dengan API lainnya.
  - c. Twitter(X) mempunyai desain yang user friendly atau mudah diakses bagi penggunanya.
4. *Persija Jakarta "To The Next Level"*

Persija Jakarta adalah salah satu klub Indonesia yang berbasis di ibu kota Indonesia yaitu Jakarta. Persija saat ini berlaga di liga 1. Persija merupakan klub tersukses di Indonesia dengan 11 gelar liga tanpa pernah terdegradasi. Persija lahir pada tanggal 28

November 1928 dan sampai saat ini belum pernah merasakan yang namanya turun kasta ke kasta kedua persepakbolaan indonesia (Mahendra & Prasetyo, 2021).

The Jakmania yaitu sebuah organisasi suporter yang mendukung klub sepakbola Persija Jakarta, yang berdiri pada tahun 1997 di Jakarta. Berperan sebagai organisasi yang menghimpun para suporter Persija Jakarta. Eksistensi The Jakmania tidak serta merta terjadi begitu saja dan langsung meroket tanpa adanya sebuah proses yang melatarbelakanginya. Pertumbuhan dan perkembangan The Jakmania melawan stigma dari masyarakat “kota besar” di mana kehidupan tidak lagi terpaut pada hiburan yang sifatnya tradisional, yaitu olahraga sepakbola. Berawal dari hal tersebut The Jakmania yang sampai saat ini peminatnya (anggota) sudah sangat banyak jika ditarik benang merahnya dari belakang hingga memiliki anggota yang kurang lebih sudah mencapai 85.000 orang tersebut tentunya bermodalkan *image* suporter sepakbola ibukota juga solidaritas dan kesan “sanggar” yang dibangun (Pasya *et al.*, 2022).

Pada musim 2021-2022 Persija Jakarta mengusung tagline #TOTHEXLEVEL. Arti “To The

*Next Level*" adalah tekad Persija dalam meningkatkan kualitas di berbagai aspek untuk menuju level yang lebih tinggi. Mulai dari tim, manajemen, fasilitas, akademi, hingga berharap dapat berkontribusi meningkatkan level sepak bola nasional.

#### 5. *Text Mining*

*Text mining* adalah proses untuk memberi informasi serta pengetahuan yang bermanfaat dengan cara menambang data yang sumbernya dari teks, kutipan, dokumen dan lain sebagainya dengan maksud untuk mencari kata-kata yang menginterpretasikan setiap teks atau dokumen sehingga dapat diketahui keterhubungan antar dokumen satu dengan lainnya (Roihan *et al.*, 2020). *Text mining* juga merupakan bagian dari data mining bedanya bentuknya lebih tidak terstruktur atau disebut sebagai *unstructured* data sedangkan data mining datanya lebih terstruktur (Sujadi, 2022). *Text mining* merupakan tahapan proses dari analisis dalam data yang berupa teks biasanya digunakan dalam klasifikasi dokumen tekstual dimana dokumen-dokumen tersebut akan diklasifikasikan sesuai dengan topik dokumen yang ada (Darwis *et al.*, 2021).

*Text mining* bertujuan untuk menganalisis pendapat, sentimen, evaluasi, sikap, penilaian, emosi seseorang sehingga dapat diketahui apakah berkenaan dengan suatu topik, layanan, organisasi, individu, atau kegiatan tertentu. Penggunaan dari *Text mining* dilakukan untuk klasterisasi, klasifikasi, *information retrieval*, dan *information extraction*.

#### 6. *Crawling* data

*Crawling* data adalah sebuah proses yang menggali jauh ke dalam *world wide web* atau target tertentu untuk mengambil data. Perayapan data juga didefinisikan sebagai proses otomatis untuk mengumpulkan dan mengindeks data dari berbagai sumber (Duei Putri *et al.*, 2022). Tujuan *crawling* data ini dilakukan untuk mengambil data dari twitter dimana data tersebut dibutuhkan pada penelitian ini. Cara melakukan *crawling* data ialah dengan membuat program dengan memasukkan kata kunci untuk mencari tweet sesuai yang kita inginkan. Misalnya, “*To The Next Level*” program akan mengambil tweet yang mention ke hastag/tagar pada tersebut. Kumpulan tweet tersebut merupakan data yang akan digunakan.

## 7. *Text preprocessing*

Data set harus melalui tahap *Text preprocessing* terlebih dahulu karena data set tidak bisa digunakan tanpa melalui tahap pengelolaan data. *Text preprocessing* adalah suatu proses pengelolaan data set sebelum data tersebut diproses. Pada kenyataannya, masih banyak data set yang tidak bersih seperti kesalahan sistem saat pencatatan sehingga terjadinya data duplikat. Data yang belum diolah atau data tidak bersih kategorinya seperti format data yang tidak beraturan, adanya data kosong, tipe data yang berbeda-beda, adanya atribut yang tidak penting, dan lain sebagainya. Semakin bersih pra proses yang dilakukan, maka kemungkinan besar hasil data tersebut semakin akurat (Widowati & Sadikin, 2021).

Terdapat tahapan *Text preprocessing*, sebagai berikut: (Isnain *et al.*, 2021)

### a. *Cleaning*

*Cleaning* adalah tahapan untuk membersihkan data pada set dan menyeleksi kata yang tidak diperlukan, tidak memiliki arti, atau arti yang memengaruhi sentimen seperti *html*, *link*, *mention*, dan *hashtag*.

b. *Tokenize*

Tokenisasi adalah pemecahan data set ke dalam bentuk token atau potongan kata agar mempermudah tahapan selanjutnya. Contohnya kalimat „saya ingin makan“ dipenggal menjadi [„saya“, „ingin“, „makan“].

c. *Transform case*

*Transform case* adalah tahapan merubah kalimat data teks menjadi teks yang seragam. Tahapan ini selalu ada dalam proses *Text preprocessing* karena data yang ada tidak selalu terstruktur dalam penggunaan hurufnya. Dengan adanya tahapan ini dapat berperan dalam penyamataan penggunaan huruf kapital. Sebagai contoh, kata “Data” dan “data” akan terbaca sebagai dua kata yang berbeda, sehingga melalui proses ini sistem dapat membaca secara efektif.

d. *Stopword removal*

Stopword adalah tahapan untuk menghapus kata yang sering muncul tapi tidak memiliki arti penting dan maknanya tidak berpengaruh pada sistem, seperti „oh“, „di“, „pada“, dan sebagainya (Alizah *et al.*, 2020).

e. Filter

Filter adalah tahapan untuk menghapus kata-kata yang terlalu pendek dan terlalu panjang dengan minimal 3 huruf dan maksimal 25 huruf (Wanto, 2018).

8. TFIDF

Analisis sentimen memiliki data mentah berupa teks, dimana data teks ini tidak dapat langsung dibaca oleh komputer. Maka dari itu, diperlukannya suatu metode yang mengubah data teks menjadi sekumpulan angka yang dapat diolah oleh sistem komputer. Metode TF-IDF (*Term Frequency Inverse Document Frequency*) merupakan metode algoritma yang memberikan bobot pada teks (Fikri *et al.*, 2020).

TF adalah jumlah kata dalam setiap dokumen, sedangkan IDF adalah nilai invers dari jumlah dokumen yang mengandung kata tersebut. Banyaknya kemunculan kata di setiap dokumen yang ada menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Banyaknya dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata-kata itu. Sederhananya, pada TF-IDF untuk kata yang sering muncul memiliki bobot nilai yang cenderung kecil sedangkan kata yang jarang muncul

akan memiliki bobot nilai yang cenderung relatif besar karena kata-kata yang sering muncul biasa disebut stopword yang biasanya dianggap kurang penting (Sujadi, 2022). Pada proses perhitungan TF dan IDF akan dikalikan untuk mencari bobot kata (Fikri *et al.*, 2020).

Berikut persamaan TF-IDF ditunjukkan pada persamaan 2.1

$$TFIDF(d,t) = TF(d,t) \cdot IDF(t) \quad (2.1)$$

Dimana  $d$  adalah dokumen dan  $t$  adalah kata. Sedangkan  $TF(d,t)$  merupakan jumlah kata pada tiap dokumen yang dirumuskan pada persamaan 2.2.

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

Pada IDF merupakan jumlah dokumen yang mengandung kata tersebut yang dirumuskan pada persamaan 2.3

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

## 9. Klasifikasi

Klasifikasi suatu proses pengkategorian data atau suatu data-data yang dapat memenuhi syarat tertentu. Pendapat lain mengatakan klasifikasi merupakan ilmu dari data mining untuk melakukan prediksi penggolongan kelas dari satu data tersebut. Klasifikasi kompleks dikatakan apabila pada pengkategorian objek memiliki tiga kelas atau lebih. Pada penelitian ini dilakukan tiga pengklasifikasian pada tiga *class* yaitu positif, negatif dan netral (Septian *et al.*, 2019).

Cara kerjanya dimulai proses pengklasifikasi-an pada sistem yang harus dilakukan pelatihan atau learning untuk mencari analisa dari data latih agar dapat dilakukan prediksi. Setelah dilakukan proses pelatihan, selanjutnya dapat dilakukan proses pengklasifikasian. Pada proses klasifikasi dilakukan untuk memprediksi keakurasian data yang didapat dari hasil pemahaman yang digunakan pada pemberian data latih kemudian diujikan dengan data test atau data uji. Sehingga apabila didapatkan tingkat keakurasian yang baik dengan membandingkan antara hasil prediksi dari sistem dengan hasil sebenarnya, maka sistem dari model tersebut dapat

digunakan untuk memprediksi kelas ke data-data yang belum diketahui kelasnya (Prajamukti & Jayanta, 2021).

#### 10. *Split validation data*

*Split validation* data adalah Teknik validasi yang membagi data secara acak menjadi dua bagian, bagian pertama digunakan sebagai data *training* (latih) dan bagian kedua digunakan sebagai data *testing* (uji). Data *training* merupakan data yang digunakan dalam melakukan proses pembelajaran (*learning*) sedangkan data *testing* merupakan data yang belum pernah dipakai sebagai pembelajaran dan akan digunakan untuk data pengujian (Isnain *et al.*, 2021).

#### 11. *Naive bayes classification*

Naïve Bayes merupakan salah satu algoritma dari *machine learning*. Dalam perkembangan database, Naïve Bayes termasuk *supervised learning* yaitu jenis *machine learning* yang membutuhkan sampel sebagai data latih yang memiliki label. *Supervised learning* dikelompokkan menjadi dua yaitu klasifikasi dan regresi. Klasifikasi pada saat variabel menjadi kategori seperti merah atau kuning, penyakit atau tidak ada penyakit, dan sebagainya. Regresi pada

saat variabel berupa nilai riil seperti berat, nilai uang, dan sebagainya. Naïve Bayes termasuk *supervised learning* klasifikasi seperti contoh lainnya yaitu *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), *Artificial Neural Network* (ANN), *Trees Gradient Boosted* (TGB), dan *Random Trees* (RT) sedangkan regresi seperti *Decision Tree*, *Logistic Regression*, dan *Kernel Regression* (Roihan et al., 2020).

Metode ini merupakan bagian dari metode bayes yang digunakan dalam klasifikasi teks didasarkan pada model penyederhanaan bahwa nilai atribut secara kondisional bersifat bebas jika nilai output diberikan. Metode Naïve Bayes banyak digunakan dalam teknik klasifikasi pada twitter. Metode ini juga digunakan dalam penambangan teks pada analisis sentimen dengan cara memprediksi probabilitas berdasarkan data sebelumnya. Naïve Bayes tidak bisa mendeteksi gambar, tetapi hanya bisa mendeteksi teks dan numerik. Untuk menghitung probabilitas dalam metode ini menggunakan pendekatan teorema bayes. Berikut persamaan teori bayes (Prabowo & Wiguna, 2021):

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (2.4)$$

Keterangan :

Y = hipotesis data X dari kelas yang spesifik

X = data dengan kelas yang belum diketahui

P(Y|X) = probabilitas hipotesis H berdasarkan kondisi X (posterior)

P(X|Y) = probabilitas X berdasarkan kondisi tersebut (likelihood)

P(Y) = probabilitas hipotesis H (prior)

P(X) = probabilitas dari x (evidence)

## 12. Evaluasi

Evaluasi klasifikasi data mining dilakukan dengan menguji keakurasian algoritma prediksi. Prosedur pengujian menggunakan *Confusion Matrix* dengan kelas yang diprediksi di bagian atas matriks dan kelas yang diamati di sebelah kiri. Semua sel matriks memiliki angka yang menunjukkan tingkat kejadian sebenarnya di kelas yang dianalisis (Ahmad, 2020).

Tabel 2. 1 *Confusion Matrix*

		<i>True Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	TP	FP
	<i>Negative</i>	FN	TN

Dalam *confusion matrix* tersebut, ada empat nilai yang dijadikan acuan dalam perhitungan, yaitu :

- a. TP (*true positive*) , Yaitu data yang diprediksi positif dan faktanya data itu positif (Sesuai).
- b. TN (*true negative*) : Yaitu data yang diprediksi negatif dan faktanya data itu negatif (Sesuai).
- c. FP (*false positive*) : Yaitu data yang diprediksi positif dan faktanya data itu negatif (Tidak Sesuai).
- d. FN (*false negative*) : Yaitu data yang diprediksi negatif dan faktanya data itu positif (Tidak Sesuai).

untuk pengukuran performa *naive bayes classifier* cara yang digunakan adalah menghitung nilai akurasi, *precision*, *recall* dan *f1 score* (Ahmad, 2020).

Akurasi merupakan ketepatan suatu sistem melakukan pengklasifikasian yang benar (Sesuai). Cara

perhitungannya dengan cara membagi jumlah yang diprediksi sistem yang sesuai dengan jumlah keseluruhan data uji. Perhitungan untuk akurasi dapat dikalkulasi dan dilihat pada persamaan 2.5.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.5)$$

*Precision* merupakan seberapa besar tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Untuk menghitung nilai *precision* dapat dilihat pada persamaan 2.6.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2.6)$$

*Recall* merupakan seberapa besar tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Untuk menghitung nilai *recall* dapat dilihat pada persamaan 2.7.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.7)$$

*F1 Score* merupakan parameter tunggal ukuran keberhasilan retrieval yang menggabungkan *recall* dan *precision*. Untuk menghitung nilai *f1 score* dapat dilihat pada persamaan 2.8.

$$f1\ Score = 2 \times \frac{precision \cdot recall}{precision + recall} \times 100\% \quad (2.8)$$

Dalam penelitian ini, penulis menggunakan *multiclass confusion matrix* 3x3 dikarenakan output sentimen dari penelitian ini ada tiga class, yaitu negatif, netral dan positif. Sehingga tabel *multiclass confusion matrix*-nya dapat dilihat pada tabel 2.2 (Ahmad, 2020).

Tabel 2. 2 *Muticlass Confusion Matrix*

		<i>True Class</i>		
		<i>Negative</i>	<i>Netral</i>	<i>Positif</i>
<i>Predicted Class</i>	<i>Negatif</i>	T Neg	F NegNet	F NegPos
	<i>Netral</i>	F NetNeg	T Net	F NetPos
	<i>Positif</i>	F PosNeg	F PosNet	T Pos

Dalam *multiclass confusion* matriks tersebut, terdapat sembilan nilai yang dijadikan acuan dalam perhitungan, yaitu :

- a. T Pos (*True Positive*), Yaitu jumlah data yang diprediksi positif dan faktanya data itu positif (Sesuai).
- b. F PosNeg (*False Positive Negative*), Yaitu jumlah data yang diprediksi positif dan faktanya data itu Negatif.

- c. F PosNet (*False Positive Netral*), Yaitu jumlah data yang diprediksi positif dan faktanya data itu Netral.
- d. F NegPos(*False Negative Positive*), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu positif.
- e. T Neg (*True Negative*), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu negatif (Sesuai).
- f. F NegNet (*False Negative Netral*), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu netral.
- g. F NetPos (*False Netral Positive*), Yaitu jumlah data yang diprediksi netral dan faktanya data itu positif.
- h. F NetNeg (*False Netral Negative*), Yaitu jumlah data yang diprediksi netral dan faktanya data itu negatif.
- i. T Net (*True Netral*), Yaitu jumlah data yang diprediksi netral dan faktanya data itu netral (Sesuai).

Tentunya setelah didapatkan nilai performa dari klasifikasi model naive bayes. Peneliti melakukan ambang batas nilai pada performa apabila diperoleh

dengan nilai diatas 80% maka dinyatakan model pengklasifikasian tersebut dinyatakan layak digunakan (baik). Sebaliknya apabila menghasilkan dibawah 80% sebaiknya peneliti selanjutnya melakukan pengujian dengan metode lain agar memperoleh hasil yang lebih maksimal (Fikri *et al.*, 2020).

Berikut persamaan *akurasi*, *precision*, *recall*, dan *f1 score* menggunakan *multiclass confusion matrix*. Rumus persamaan multiclass confusion matrix akurasi dapat dilihat pada persamaan 2.9.

$$Akurasi = \frac{T_{pos}+T_{neg}+T_{net}}{total\ data\ yang\ di\ uji} \times 100\% \quad (2.9)$$

Rumus persamaan *multiclass confusion matrix precision* dapat dilihat pada persamaan 2.10.

$$Precision = \frac{T_{pos}}{T_{pos}+F_{posnet}+F_{posneg}} \times 100\% \quad (2.10)$$

Rumus persamaan *multiclass confusion matrix recall* dapat dilihat pada persamaan 2.11.

$$Recall = \frac{T_{pos}}{T_{pos}+F_{netpos}+F_{negpos}} \times 100\% \quad (2.11)$$

Rumus persamaan *multiclass confusion matrix* *f1 score* dapat dilihat pada persamaan 2.12.

$$f1\ Score = 2 \times \frac{precision \cdot recall}{precision + recall} \times 100\% \quad (2.12)$$

## B. Kajian Penelitian yang Relevan

Beberapa penelitian yang relevan dalam penelitian ini antara lain dapat dilihat pada tabel berikut.

Tabel 2. 3 Penelitian terkait

No	Pustaka	Topik	Metode	Objek	Klasifikasi dan Hasil
1	(Dedi darwis, Nery Siskawati, Zaenal Abidin. 2021)	Penerapan algoritma naive bayes untuk analisis sentimen review data twitter BMKG nasional	Naive bayes	Twitter	3 Kelas (Positif, Negatif, dan netral).  Hasil akurasi pengujian 68,97%.

2	(Taoufik Kridiyanto , Erry Maricha. 2021)	Analisis sentimen masyarakat indonesia terhadap kebijakan PPKM pada media sosial twitter menggunakan naive bayes classifier	Naive bayes	Twitter	2 Kelas (Positif dan Negatif).  99% sentimen positif, dan 1% sentimen negatif
3	(Kurniasih & Suseno, 2022)	Tanggapan mengenai Bantuan Subsidi Upah pada kenaikan harga BBM	Naive Bayes	Twitter	2 Kelas (Positif dan Negatif).  Positif 65,2% dan negatif 71,8%.

Dibagian ini penulis membaca penelitian-penelitian terkait sebelumnya dengan tujuan untuk memperkuat pelaksanaan penelitian ini. Penelitian terkait yang mengacu pada analisis sentimen pernah dilakukan oleh Dedi Darwis, Nery, dan Zaenal pada tahun 2021 mengenai review data twitter BMKG Nasional menggunakan algoritma naive bayes. Dengan

tujuan mengklasifikasi sentimen berdasarkan respon pengguna twitter tentang BMKG Nasional. Dengan memberikan beberapa sentimen diantaranya yaitu sentimen positif, negatif, dan netral. Penelitian ini menggunakan akun resmi BMKG Nasional. Proses ekstraksi data dari Twitter BMKG Nasional menggunakan Bahasa pemrograman Python 3.74 dengan tahapan Preprocessing yang meliputi *casefolding, filtering, tokenisasi, slang replacement* dan *stopword removal*. Tingkat akurasi berdasarkan pengujian yang dilakukan adalah 68,97% (Darwis *et al.*, 2021).

Pada penelitian lainnya dilakukan oleh taofik kridiyanto, dan Erry Oki pada tahun 2021. Penelitian mengenai analisis sentimen opini masyarakat terhadap kebijakan PPKM. Data diambil dari twitter menggunakan *naive bayes clasifiers*. Penelitian ini mengklasifikasikan menjadi 2 sentimen, yaitu sentimen positif dan sentimen negatif. Klasifikasi dilakukan dengan menggunakan metode Naïve Bayes Clasifiers. Metode Naïve Bayes Classifiers merupakan salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan data latih dan data uji yang ada. Hasil penelitian diketahui

bahwa opini masyarakat mengenai kebijakan PPKM dihasilkan 99% termasuk kedalam klasifikasi polaritas positif dan 1% polaritas negatif. Selain itu juga dihasilkan bahwa kata yang memiliki frekuensi tertinggi penggunaannya oleh para pengguna twitter adalah kata “darurat”. (Krisdiyanto, 2021).

Penelitian selanjutnya dilakukan oleh Ulfa Kurniasih dan Suseno pada tahun 2022 mengenai tanggapan Bantuan Subsidi Upah (BSU) pada kenaikan harga BBM pada masyarakat Indonesia menggunakan metode klasifikasi naive bayes yang datanya diambil dari media sosial twitter. Dengan tujuan mengklasifikasikan sentimen berdasarkan respon dari pengguna twitter terhadap tanggapan bantuan subsidi upah pada kenaikan harga BBM. Dengan memberikan beberapa sentimen diantaranya yaitu sentimen positif dan negatif. Pada penelitian ini menggunakan dua kata kunci pencarian untuk data yang digunakan yaitu pada kata BBM dan BSU dengan diperoleh sebanyak 795 data tweet setiap masing-masing kata kemudian dilakukan pembobotan kata dengan TFIDF. Dengan hasil akhir diperoleh pada kata kunci BBM mendapatkan lebih banyak respon negatif sebesar 71,8% sedangkan pada kata kunci BSU mendapatkan

lebih banyak respon positif sebesar 65,2%. Dari penelitian ini didapatkan performa dari model naive bayes dengan rata-rata tingkat akurasi sebesar 81,64%, presisi sebesar 92,89% dan recall sebesar 29,91% (Kurniasih & Suseno, 2022).

Berdasarkan penelitian yang telah dilakukan oleh beberapa penulis di atas, dapat dijadikan acuan dalam melakukan penelitian. Peneliti tertarik untuk melakukan penelitian mengenai analisis sentimen tentang Persija "*To The Next Level*" pada media sosial twitter(X) menggunakan metode naive bayes. Penelitian ini berupaya menghasilkan performa yang baik dari metode *naive bayes* dalam mengklasifikasikan sentimen kebijakan tersebut berdasarkan tiga sentimen yang dimiliki, yaitu sentimen positif, negatif dan netral dengan hasil akhir visualisasi diagram lingkaran dan diagram batang dengan data yang diperoleh dari twitter pada kata kunci tagar "*To The Next Level*" Persija Jakarta yang belum pernah dilakukan oleh peneliti sebelumnya.



## **BAB III**

### **METODOLOGI PENELITIAN**

#### **A. Metode Pengumpulan Data**

##### 1. Data sekunder

###### a. Studi pustaka

Penulis melakukan studi pustaka dengan memanfaatkan dan membaca buku-buku, jurnal, skripsi, dan sejenisnya guna mempelajari materi yang berhubungan dengan konsep, analisis, dan permasalahan lainnya. Penulis juga melakukan pencarian data secara daring menggunakan browser melalui internet yang berhubungan dengan analisis sentimen, text mining, dan algoritma naive bayes *clasification*.

##### 2. Data primer

###### a. Studi lapangan (data tweet)

Penulis melakukan pengamat-an langsung di media sosial twitter dengan kata kunci Persija *To The Next Level*, salah satunya terhadap respon fans Persija Jakarta tentang *To The Next Level* Persija di media sosial twitter.

## B. Perangkat Penelitian

Penelitian ini membutuhkan beberapa perangkat keras dan perangkat lunak yang harus dipenuhi untuk berjalannya proses sitem, adapun kebutuhan yang digunakan pada penelitian ini sebagai berikut:

### 1. Kebutuhan perangkat keras

Tabel 3. 1 Kebutuhan perangkat keras

No.	Perangkat Keras	Spesifikasi
1.	Device	HP Laptop 14s-dk1xxx
2.	Processor	AMD Athlon Gold 1350U with Radeon
3.	Memori ( RAM )	4,00 GB
4.	Monitor	14 inch
5.	Keyboard dan Mouse	Standard

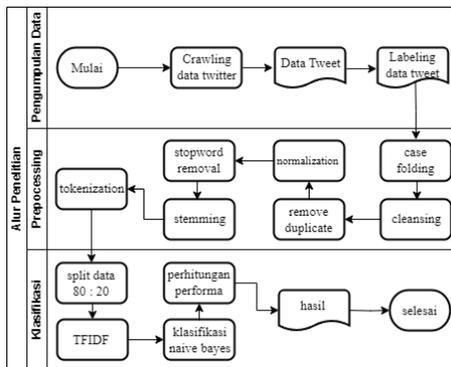
### 2. Kebutuhan perangkat lunak

Tabel 3. 2 Kebutuhan perangkat lunak

No.	Perangkat Lunak	Spesifikasi
1.	Sistem Operasi	Microsoft Windows 11 Home Single Language
2.	Bahasa Pemrograman	Python
3.	MS. Office	MS. Pffice 16
4.	<i>Browser</i>	Chrome
5.	<i>Google Drive</i>	Google Colab

### C. Alur Penelitian

Alur penelitian akan membahas mengenai gambaran umum yang dilakukan peneliti dalam tahapan pengerjaan yang dimulai dari awal sampai akhir. Berikut rangkaian pengerjaan tugas akhir dapat dilihat pada gambar 3.1 yaitu *flowchart* alur penelitian yang ada dibawah ini.



Gambar 3. 1 *Flowchart* alur penelitian

#### 1. Pengumpulan Data

Langkah pertama dari penelitian ini ialah mengumpulkan data opini, respon atau tanggapan dari media sosial twitter dengan cara *crawling* data. Proses *crawling* data diambil otomatis melalui Google colab dengan kode yang berkaitan dengan tagar “*To The Next Level*” persija Jakarta. Cara melakukan *crawling* data ialah dengan membuat program dengan memasukkan kata kunci untuk mencari tweet sesuai yang kita

inginkan. Pada penelitian ini pencarian yang digunakan dengan kata kunci persija “*To The Next Level*”. Kumpulan tweet tersebut merupakan data yang akan digunakan. Berikut hasil crawling data pada gambar 3.2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	Formula Bar	
1	created_at	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id	conversation_id	username	tweet_url										
2	Tue Dec 1	1.73418 @Persija_	0	1	0	2	in	73343229	1.73418	arafaadi	https://twitter.com/arafaadi/status/17369182115488030847										
3	Wed Dec 1	1.73418 MATCH Di	0	0	1	0	in	1.096148	1.73418	jakmania809	https://twitter.com/jakmania809/status/1734857879927239171										
4	Sat Dec 18	1.73418 @Persija_	0	0	0	1	in	8484147	1.73418	jakmania809	https://twitter.com/jakmania809/status/1733814217055220093										
5	Tue Dec 18	1.73418 @Persija_	0	0	0	0	in	9474117	1.73418	imugg	https://twitter.com/imugg/status/173153703629385567										
6	Sun Dec 1	1.73418 @Persija_	0	0	0	0	in	4.71409	1.73418	halooorn	https://twitter.com/halooorn/status/1731338304257206083										
7	Sun Dec 1	1.73418 @Persija_	0	0	0	0	in	1.11408	1.73418	hysteryan	https://twitter.com/hysteryan/status/173133154772069486										
8	Sun Dec 1	1.73418 @Persija_	0	0	0	0	in	2.281409	1.73418	PSIAADUKT	https://twitter.com/PSIAADUKT/status/173116515816612938										
9	Sun Dec 1	1.73418 MATCH Di	0	0	0	1	in	1.71409	1.73418	jak_sung	https://twitter.com/jak_sung/status/173114040186562963										
10	Thu Nov 1	1.73418 MATCH Di	0	1	1	2	in	1.096148	1.73418	jakmania809	https://twitter.com/jakmania809/status/173081201223495730										
11	Wed Nov 1	1.73418 Selamat u	0	0	0	0	in	1.721418	1.73418	bang7281	https://twitter.com/bang7281/status/172975793616616014										
12	Mon Nov 2	1.73418 Hui jak ia	8	24	40	237	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/172898345843962993										
13	Sun Nov 1	1.73418 Asha per	6	5	21	200	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/172881244616637138										
14	Sun Nov 1	1.73418 Ready for	2	14	50	430	en	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/172877511544621709										
15	Sun Nov 2	1.73418 Bambang	1	4	41	253	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/172876071297974208										
16	Sun Nov 2	1.73418 Pertahank	12	18	88	640	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/172865181786877238										
17	Sun Nov 1	1.73418 Beesik ada	6	35	47	450	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/172862104642146040										
18	Sat Nov 2	1.73418 Bek sayap	1	9	41	282	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/1728417762874832527										
19	Sat Nov 2	1.73418 Sabtu	1	7	43	410	en	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/17283815808800251										
20	Sat Nov 2	1.73418 Matarangka	2	11	36	358	in	4.521408	1.73418	Persija_ik	https://twitter.com/Persija_ik/status/1728319544114614										

Gambar 3.2 Crawling data

Setelah data didapatkan lalu dilakukan proses pelabelan data untuk menentukan sentimen/tanggapan yang ada. Pada penelitian ini menggunakan tiga sentimen yaitu sentimen positif, sentimen negatif, dan sentimen netral. Pelabelan data agar proses pengklasifikasian lebih mudah. Maka, pada kelas sentimen negatif diubah menjadi angka 0, kelas sentimen netral menjadi angka 1, dan kelas sentimen positif menjadi angka 2. Berikut contoh penerapan *labelling* pada tabel 3.3.

Tabel 3. 3 Penerapan *labelling*

Opini	Sentimen	Label
@Persija_Jkt #ToTheNextLevel nya mana? udah sadar kah kalau sekarang turun level?	Negatif	0
@Persija_Jkt Udah gak #ToTheNextLevel?	Netral	1
@Persija_Jkt menang hati senang! Happy weekend #ToTheNextLevel	Positif	2

## 2. Preprocessing

Langkah kedua, yaitu tahap *preprocessing*. Yang bertujuan untuk membersihkan data mentah sehingga data bisa diklasifikasikan. Pada *preprocessing* terdiri dari tujuh tahap, yaitu *case folding*, *cleaning*, *removal duplicate*, *normalization*, *stopword removal*, *stemming*, dan *tokenization*. Dimulai dari tahap *case folding*, suatu teks disamaratakan yang terdapat huruf kapital menjadi huruf kecil. Contoh proses *case folding* ada di tabel 3.4.

Tabel 3. 4 Proses *case folding*

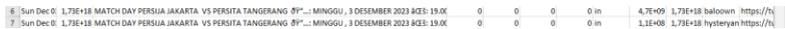
<i>Input Process</i>	<i>Output Process</i>
@Persija_Jkt #ToTheNextLevel nya mana? udah sadar kah kalau sekarang turun level?	@persija_jkt #tothenextlevel nya mana? udah sadar kah kalau sekarang turun level?

Kemudian masuk ke tahap *cleansing* untuk melakukan penghapusan apabila ada pada URL, tanda baca, dan “@” pada username di twitter yang bertujuan untuk mengurangi *noise* pada teks. Berikut contoh proses *cleansing* pada tabel 3.5.

Tabel 3. 5 Proses *cleansing*

<i>Input Process</i>	<i>Output Process</i>
@Persija_Jkt #ToTheNextLevel nya mana? udah sadar kah kalau sekarang turun level?	nya mana udah sadar kah kalau sekarang turun level

Lalu dilakukan *removal duplicate* untuk menghapus data tweet yang berulang atau memiliki kesamaan. Contoh proses *removal duplicate* ada pada gambar 3.3.



Gambar 3. 3 Proses *removal duplicate*

Setelah itu, dilakukan *normalization* untuk mengganti kata tidak baku, kata singkat menjadi kata baku yang sesuai anjuran KBBI. Berikut contoh proses *normalization* pada tabel 3.6.

Tabel 3. 6 Proses *normalization*

<i>Input Process</i>	<i>Output Process</i>
nya mana udah sadar kah kalau sekarang turun level	nya mana sudah sadar kah kalau sekarang turun level

Dilanjutkan tahap *stopword* untuk menghapus kata yang kalimatnya berisi kata-kata yang bersifat umum. Berikut contoh proses *stopword* pada gambar 3.7.

Tabel 3. 7 Proses *stopword*

<i>Input Process</i>	<i>Output Process</i>
nya mana sudah sadar kah kalau sekarang turun level	sudah sadar kalau sekarang turun level

Lanjut tahap *stemming* dilakukan untuk pemngubahan kata yang memiliki imbuhan menjadi kata dasar dan juga mengubah kata perulangan kata menjadi satu kata dasar. Berikut contoh proses *stemming* pada tabel 3.8.

Tabel 3. 8 Proses *stemming*

<i>Input Process</i>	<i>Output Process</i>
sudah sadar kalau sekarang turun level	sudah sadar kalau sekarang turun level

lalu pada tahap *tokenization* dilakukan pemecahan suatu kalimat menjadi bagian kata-kata dan juga untuk menghilangkan *whitespace*. Berikut contoh proses *tokenization* pada tabel 3.9.

Tabel 3. 9 Proses *tokenization*

<i>Input Process</i>	<i>Output Process</i>
sudah sadar kalau sekarang turun level	['sudah','sadar','kalau','sekarang','turun','level']

### 3. Klasifikasi

Langkah ketiga, yaitu tahap klasifikasi. Proses klasifikasi dimulai dengan melakukan tahapan *split validation* data. Data dibagi antara data latih dan data uji dengan rasio 80:20. Dimana 80% dari jumlah keseluruhan data dijadikan data latih dan 20% dari keseluruhan data dijadikan data uji yang diambil secara acak (Isnain et al., 2021). Setelah itu masuk ke

tahapan selanjutnya yaitu esktraksi fitur yang merupakan proses mengubah kata menjadi bilangan angka serta dilakukan pembobotan nilai kata dengan menggunakan TFIDF untuk mempermudah jalannya proses pengklasifikasian naive bayes. Selanjutnya ialah tahapan paling utama yaitu proses pengklasifikasian dengan metode naive bayes. Metode naïve bayes merupakan metode yang digunakan untuk mengkasifikasikan data komentar untuk memperoleh sentimen analisis (Prajamukti, Jayanta, 2021). Dengan mengklasifikasikan berdasarkan sentimen yang ada pada dokumen. Sesudah itu, masuk ke tahapan uji model untuk mengetahui ketepatan klasifikasi dengan membandingkan hasil prediksi dari suatu sistem dengan hasil dari data yang sudah diberi label sebelumnya, sehingga menghasilkan *multiclass confusion matrix*. Setelah uji model dilakukan, lalu pada tahap evaluasi model untuk menentukan tingkat keakurasian, *precision*, *recall*, dan *f1 score* melalui metode *multiclass confusion matrix* pada tabel dengan tujuan mengetahui performa dari model tersebut. Evaluasi model dilakukan untuk mengetahui atau mengukur kinerja suatu model (Darwis et al., 2021). Evaluasi model dilakukan dengan cara melihat tingkat

performa metode melalui *multiclass confusion matrix*. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Samsir et al., 2021). Lebih tepatnya, data *test* diujikan terhadap data *training*, maka akan menghasilkan daftar kelas-kelas dari data *test*, sebut saja prediksi kelas. Kemudian prediksi kelas dibandingkan dengan kelas yang sebenarnya dari data *test* yang disembunyikan sebelumnya. Sehingga dapat dilihat pada performa model naive bayes yang berupa tingkat akurasi, *precision*, *recall* dan *f1 score* .

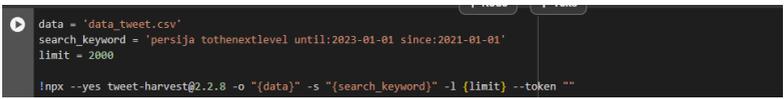
Proses selanjutnya yaitu memvisualisasikan hasil dari analisis sentimen menggunakan diagram lingkaran dan *wordcloud*. *Wordcloud* akan digunakan untuk memvisualisasikan hasil analisis klasifikasi. Tujuan dari visualisasi tersebut adalah untuk mengekstrak informasi berupa opini supporter/fans mengenai persija *ToTheNextLevel* sehingga dapat diambil informasi yang dianggap penting dari sekian banyak teks tanggapan yang ada.

## BAB IV

### HASIL DAN PEMBAHASAN

#### A. *Crawling* Data Tweet

Proses pengambilan (proses *crawling* data) *tweet* dari *twitter* ini dilakukan menggunakan *library* yang tersedia, yaitu *tweet-harvest*. *Library* tersebut dapat digunakan untuk melakukan pengambilan data melalui media sosial salah *twitter*. *Library tweet-harvest* cukup efektif untuk melakukan *crawling* data *tweet* pada sosial media *twitter* dengan menggunakan *Command Line interface* yang tertera pada gambar 4.1 berikut:



```
data = 'data_tweet.csv'
search_keyword = 'persija tothenextlevel until:2023-01-01 since:2021-01-01'
limit = 2000
!npx --yes tweet-harvest@2.2.8 -o "{data}" -s "{search_keyword}" -l {limit} --token ""
```

Gambar 4. 1 Proses *Crawling* Data

Data yang berhasil dikumpulkan akan disimpan pada file yang bernama '*data\_tweet.csv*'. *search\_keyword* berisi kata kunci yang digunakan untuk memfilter data tertentu sesuai dengan *keyword* tersebut disertai dengan batas awal dan akhir dari data yang akan dikumpulkan. Data yang dikumpulkan dibatasi hanya 2000 data sesuai dengan yang ada pada variabel *limit*. Pada saat proses pengambilan data *tweet*, peneliti mengambil dengan kata kunci 'persija tothenextlevel yang berjumlah 2012 *tweet*. Data yang

diambil menggunakan data yang berbahasa Indonesia dan diambil dari tanggal 1 Januari 2021 – 1 Januari 2023.

Berikut ini adalah hasil yang didapatkan dari proses *crawling* data pada gambar 4.2 dan 4.3:

```
[ ] import pandas as pd

data = pd.read_csv("/content/tweets-data/data_tweet.csv", sep=";")

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1004 entries, 0 to 1003
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   created_at            1004 non-null   object
1   id_str                1004 non-null   int64
2   full_text             1004 non-null   object
3   quote_count           1004 non-null   int64
4   reply_count           1004 non-null   int64
5   retweet_count         1004 non-null   int64
6   favorite_count        1004 non-null   int64
7   lang                  1004 non-null   object
8   user_id_str           1004 non-null   int64
9   conversation_id_str   1004 non-null   int64
10  username               1004 non-null   object
11  tweet_url              1004 non-null   object
dtypes: int64(7), object(5)
memory usage: 94.2+ KB

[ ] data.head(1004)
```

Gambar 4. 2 Jenis Data Yang dikumpulkan

Data yang berhasil dikumpulkan terdiri dari beberapa kolom, yaitu *created\_at*, *id\_str*, *full\_text*, *quote\_count*, *\_reply\_count*, *retweet\_count*, *favorite\_count*, *lang*, *user\_id\_str*, *conversation\_id\_str*, *username*, dan *tweet\_url*.

	created_at	full_text	username	sentimen	tweet_clean
0	Sat Dec 31 17:00:38 +0000 2022	SELAMAT TAHUN BARU 2023, JAK 🇮🇩 #ToTheNextLeve...	Persja_ikt	positif	[selamat, tahun, baru, jak]
1	Sat Dec 31 12:19:01 +0000 2022	BEST XI BUBBLE SERIES! Dua nama yang turut ja...	Persja_ikt	positif	[best_xi, bubble, series, dua, nama, yang, tu...
2	Sat Dec 31 08:00:02 +0000 2022	10 GOL TERBAIK PILIHAN PERSIJA DI TAHUN 2022  ...	Persja_ikt	positif	[gol, terbaik, pilihan, persija, di, tahun, ir...
3	Sat Dec 31 06:00:00 +0000 2022	*Alhamdulillah saya dapat penghargaan Young Pl...	Persja_ikt	positif	[alhamdulillah, saya, dapat, penghargaan, youn...
4	Sat Dec 31 04:00:30 +0000 2022	10 GOL TERBAIK PILIHAN PERSIJA DI TAHUN 2022  ...	Persja_ikt	positif	[gol, terbaik, pilihan, persija, di, tahun, ir...
...	...	...	...	...	...
2006	Sat Aug 05 05:10:48 +0000 2023	1 poin = Rp1,- PASTIKAN LOGIN AKUN SHOUT & amp...	Persja_ikt	netral	[poin, rp, pastikan, login, akun, shout, &, am...
2007	Sat Aug 05 03:23:33 +0000 2023	Sundutan pembuka kemenangan Persija dari Hanif...	Persja_ikt	netral	[sundutan, pembuka, kemenangan, persija, dari...
2008	Fri Aug 04 17:10:05 +0000 2023	Saksikan video pertemuan perwakilan PP-Jakmani...	Persja_ikt	netral	[saksikan, video, pertemuan, perwakilan, pp, j...
2009	Fri Aug 04 16:31:48 +0000 2023	Same as before! Duet selebrasi gol Fiza Andik...	Persja_ikt	positif	[same, as, before, duet, selebrasi, gol, fiza...
2010	Fri Aug 04 16:03:48 +0000 2023	Malam yang sejuk di Sleman 🇮🇩 🍷 Sampai bertemu...	Persja_ikt	netral	[malam, yang, sejuk, di, sleman, sampai, berte...

Gambar 4. 3 Data Yang Berhasil Dikumpulkan

Dengan menjalankan kode tersebut, maka dapat disimpulkan bahwa total data yang berhasil dikumpulkan yaitu sejumlah 2012 *rows* x 12 *columns*. Data yang berhasil kumpulkan masih berupa data mentah, dan membutuhkan proses *preprocessing* data sehingga data yang dikumpulkan dapat dilakukan klasifikasi nantinya.

Setelah proses pengumpulan data dilakukan, penulis melakukan pelabelan data secara manual satu persatu, untuk menjaga akurasi dari pelabelan data. Dalam proses pelabelan data tweet ini, peneliti menentukan nilai dari tweet pada twitter(X) yang bernilai positif dalam artian setuju atau pro dengan tagar ToTheNextLevel persija, pada bernilai negatif yaitu yang kontra atau tidak setuju dengan tagar tersebut sedangkan netral yaitu ditengah-tengah dalam artian tidak pro ataupun kontra biasanya memuat informasi. Dalam proses pelabelan data ini seharusnya untuk menentukan nilai tersebut dilakukan oleh pakar

bahasa atau yang terkait dibidangnya dan juga setidaknya membutuhkan dua orang atau lebih untuk menghindari perbedaan pendapat yang sama serta menghindari subjektivitas dalam menentukan sentimen tweet yang terdapat dari hasil data *crawling* (Kurniasih & Suseno, 2022).

Namun demikian, Pelabelan pada penelitian ini hanya dilakukan oleh satu orang ahli bahasa, dikarenakan untuk menghindari perbedaan pandangan dari ahli Bahasa dan memaksimalkan waktu penelitian. Pelabelan pada penelitian ini dilakukan oleh seorang ahli Bahasa yang menjadi guru Bahasa Indonesia di beberapa sekolah di kabupaten Tangerang. Karena proses pelabelan ini dilakukan secara manual, dan juga dalam proses pelabelan manual ini membutuhkan waktu yang tidak cepat, hal tersebut menjadi kelemahan pada saat jumlah data yang akan di proses pelabelan sangat banyak.

Berikut ini adalah hasil dari proses pelabelan data yang ditunjukkan pada gambar 4.4:

1	created_at	full_text	username	sentimen
20	Wed Dec 28 06:13:55 +0000 2022	Selamat ulang tahun yang ke-22 untuk the Jakmania Gajah Mada. Semoga semakin solid dan semakin militan dari infokomJakmania		positif
29	Mon Dec 26 05:58:25 +0000 2022	Selamat ulang tahun, Andriany Ardyasa. Semoga panjang umur, sehat selalu, dan bisa terus memberikan kontk Jak Tenabang		positif
30	Mon Dec 26 05:55:53 +0000 2022	Selamat ulang tahun, Andriany Ardyasa. Semoga panjang umur, sehat selalu, dan bisa terus memberikan kontk infokomJakmania		positif
50	Fri Dec 23 08:00:20 +0000 2022	#PERSIJADAY !! #PSS Semarang vs Persija Jakarta. #Jumat, 23 Desember 2022. Kick Off 15.00 WIB. #Fatihputera12		netral
56	Fri Dec 23 05:43:48 +0000 2022	#PSS Semarang vs Persija Jakarta. #Jumat, 23 Desember 2022. #PSS Semarang vs Persija Jakarta. #Jumat, 23 Desember 2022. #PSS Semarang vs Persija Jakarta		netral
68	Thu Dec 22 19:51:34 +0000 2022	Persija day...Optimis 3 poin dari Marahan. #ToTheNextLevel #UsutTuntasTragediKorbanJurnal #BRILiga1 #M_JakmaniaLanggar		positif
61	Thu Dec 22 08:01:42 +0000 2022	ada. Kami Jukung. #PSS Semarang vs Persija Jakarta. #Jumat, 23 Desember 2022. #PSS Semarang vs Persija Jakarta. #Jumat, 23 Desember 2022		netral
74	Wed Dec 21 02:43:54 +0000 2022	Selamat ulang tahun yang ke-15 untuk the Jakmania Utan Kayu. Semoga semakin solid dan semakin militan dari infokomJakmania		positif
77	Tue Dec 20 15:51:53 +0000 2022	Menang! FT: Persija Jakarta 3-2 Dewa United. #PersijaDay #BRILiga1 #ToTheNextLevel #BelieveIn12 #PersijaJakinifootball		netral
76	Tue Dec 20 15:32:23 +0000 2022	FT: Persija Jakarta 3-2 Dewa United. #Pe25jaSelamanya #UsutTuntasTragediKorbanJurnal #ToTheNextLevel #InfokomJakmania		netral
101	Tue Dec 20 13:10:39 +0000 2022	COYT. #FT @Persija_Jkt Bersiap!! #PersijaDay #BRILiga1 #ToTheNextLevel #BelieveIn12 #PersijaJakartaOfficial_XVII		netral
146	Fri Dec 20 00:04:08 +0000 2022	Game Week #16. Persija Jakarta vs Dewa United. (BIASAKAN MENANG, SAMPAL MENANG JADI KEBIASAAN. JakmaniaLanggar		positif
108	Mon Dec 19 17:25:49 +0000 2022	Heian ke-10 Bnti Liga 1. #JazzZuzuz # Persija Jakarta vs Dewa United. #Seiaa, 20 Desember 2022. #Kick Off		netral
112	Mon Dec 19 11:48:25 +0000 2022	Heian ke-10 Bnti Liga 1. #JazzZuzuz # Persija Jakarta vs Dewa United. #Seiaa, 20 Desember 2022. #Kick Off		netral
121	Sat Dec 17 15:30:00 +0000 2022	@Persija_Jkt #ToTheNextLevel https://t.co/uxXpnoGdFA	AnamCecep	netral
126	Sat Dec 17 04:00:42 +0000 2022	@Persija_Jkt Yaudah ekarang legakan kepala lo, ngeri slogan lo #ToTheNextLevel kalo gap bisa finishing touch ju ni. #ingih		negatif
129	Fri Dec 16 15:08:09 +0000 2022	FT: Persija Jakarta 1-1 Persibaya Surabaya. #PersijaDay #BRILiga1 #ToTheNextLevel #BelieveIn12 #PersijaJakinifootball		netral
129	Fri Dec 16 10:42:09 +0000 2022	@Jakartaind @Persija_Jkt benar kan kata w juga slogan #ToTheNextLevel buat @Persija_Jkt masih kurang pas LefsaIsa21		negatif
130	Fri Dec 16 10:07:17 +0000 2022	FT: Persija Jakarta 1-1 Persibaya Surabaya. #UsutTuntasTragediKorbanJurnal #ToTheNextLevel #BelieveIn12 #InfokomJakmania		netral
131	Fri Dec 16 10:06:44 +0000 2022	Slogan #ToTheNextLevel bagu kedalem skwadnya disi sama pemain dengan kualitas Liga 2. @Persija_Jkt	Jakartend	negatif
146	Fri Dec 16 08:38:40 +0000 2022	Kami berada bersama Persija dan akan selalu menjaga kehormatan Persija Jakarta. #ToTheNextLevel #JakmanInfokomJakmania		positif
155	Fri Dec 16 07:37:55 +0000 2022	#BRILiga1 #JazzZuzuz # Persija Jakarta vs Persibaya Surabaya. #Jumat, 19 Desember 2022		netral
160	Thu Dec 15 22:02:50 +0000 2022	Heian ke-10 Bnti Liga 1. #JazzZuzuz # Persija Jakarta vs Persibaya Surabaya. #Jumat, 19 Desember 2022. #Kick Off		netral
161	Thu Dec 15 19:04:44 +0000 2022	Heian ke-10 Bnti Liga 1. #JazzZuzuz # Persija Jakarta vs Persibaya Surabaya. #Jumat, 19 Desember 2022. #Kick Off		netral
162	Thu Dec 15 17:01:58 +0000 2022	Heian ke-10 Bnti Liga 1. #JazzZuzuz # Persija Jakarta vs Persibaya Surabaya. #Jumat, 19 Desember 2022. #Fatihputera12		netral
172	Wed Dec 14 08:06:17 +0000 2022	Selamat ulang tahun yang ke-4 untuk the Jakmania Mekarani. Semoga semakin solid dan semakin militan dari infokomJakmania		positif
174	Tue Dec 13 17:49:11 +0000 2022	Ayo Bangkit cant! @Persija_Jkt #ToTheNextLevel https://t.co/Wj9pud8mm	AnamCecep	positif
175	Tue Dec 13 15:41:52 +0000 2022	@inoppis @Persija_Jkt Lah, justru dg adanya panca masa depan Persija jelas apalagi dengan mengunung slogan #BangDinar		positif
189	Tue Dec 13 14 09:47 +0000 2022	@Persija_Jkt katanya mau #ToTheNextLevel tapi mainnya malu-malain	wafachand	negatif
204	Tue Dec 13 08:06:38 +0000 2022	@Persija_Jkt #ToTheNextLevel	DeDeXDeDeDe	netral
207	Mon Dec 12 17:03:30 +0000 2022	#PERSIJADAY !! #PSS Semarang vs Persija Jakarta. #Jumat, 13 Desember 2022. #Kick Off 15.00 WIB. #Fatihputera12		netral
212	Mon Dec 12 06:18:08 +0000 2022	Next Match! #PSS Semarang vs Persija Jakarta. Baskakan menang, sampai menang jadi kebiasaan! Anggur mera JakmaniaLanggar		positif
218	Sat Dec 10 15:18:47 +0000 2022	FT: Persija Jakarta 1-1 Persik Kediri. #UsutTuntasTragediKorbanJurnal #ToTheNextLevel #BelieveIn12 #InfokomJakmania		netral
219	Sat Dec 10 13:32:45 +0000 2022	FT: Persija Jakarta 1-1 Persik Kediri. #PersijaDay #BRILiga1 #ToTheNextLevel #BelieveIn12 #PersijaJakarta #Jakinifootball		netral
245	Sat Dec 10 02:40:22 +0000 2022	Kita doakan Jak, azar Persija bisa memano lagi. Waib oon 3 di home nani. (svstem bubbles). Come on you boer Jakmania km37		positif

## Gambar 4. 4 Hasil Data yang telah diberi Label

Berikut salah satu contoh pelabelan data pada penelitian ini berada pada tabel 4.1.

Tabel 4. 1 contoh pelabelan

No	User	Tweet	Sentiment
1	InfokomJakmania	Selamat ulang tahun yang ke-22 untuk the Jakmania Gajah Mada. Semoga semakin solid dan semakin militan dalam mendukung tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu untuk Jakmania Gajah Mada! 🙌❤️ #Pe25jaSelamanya #ToTheNextLevel #BelieveInTwelve #InfokomJakmania <a href="https://t.co/RUONKgzfz0">https://t.co/RUONKgzfz0</a>	Positif
2	Fatihputera12	#PERSIJADAY !! 🏆 PSS Sleman vs Persija Jakarta 🏟️ Jumat, 23 Desember 2022 🕒 Kick Off 15.00	Netral

		WIB 🏟️ Stadion Manahan, Solo COME ON YOU TIGER 🐯 #BRILiga1 #ToTheNextLevel #BelieveIn12 #PersijaJakarta <a href="https://t.co/qwNpr0drN2">https://t.co/qwNpr0drN2</a>	
3	Jakartainsd	Slogan #ToTheNextLevel tapi kedalam skuadnya diisi sama pemain dengan kualitas Liga 2. @Persija_Jkt	Negative

## B. Preprocessing Data

Tahap ini bertujuan untuk membersihkan data yang berhasil dikumpulkan melalui sosial media Twitter(X). Data yang sebelumnya tidak terstruktur dan memiliki *noise*, akan dirubah menjadi data yang bersih, sehingga proses klasifikasi dapat dilakukan.

Berikut ini adalah tahap-tahap yang dilakukan untuk menghasilkan data yang bersih melalui proses *preprocessing*.

### 1. Case folding

Tahap selanjutnya adalah tahap *case folding*, dimana pada proses ini, setiap kata pada data teks akan dirubah menjadi huruf kecil semua, yang berguna untuk menormalisasi dan mempertahankan konsistensi dari representasi tekstual suatu data.

Berikut adalah kode yang digunakan pada tahap *casefolding*.

```
def casefolding(data):  
    data = data.lower()  
    return data  
data['tweet_clean'] = data['tweet_clean'].apply(casefolding)  
|  
print(data["tweet_clean"].head())
```

Gambar 4. 5 Proses *case folding*

Berikut ini adalah hasil dari proses *case folding* ditunjukkan pada gambar 4.6:

```
[6]: def casefolding(data):  
      data = data.lower()  
      return data  
  
data['tweet_clean'] = data['tweet_clean'].apply(casefolding)  
print(data["tweet_clean"][200:210])  
  
200 susunan pemain persija jakarta untuk laga mala...  
201          stadion maguwoharjo sleman  
202          ersija jkt  
203 data head to head kedua tim jelang laga malam ...  
204          bangkit berjuang menang  
205 psis semarang vs persija jakarta selasa des...  
206          matchday  
207 salah satu gol ikonik dari duo bambang pamungk...  
208 saya sudah bicara pada para pemain terkait per...  
209 pekan ke bri liga psis semarang vs persija j...  
Name: tweet_clean, dtype: object
```

Gambar 4. 6 Hasil *case folding*

## 2. *Cleaning*

Tahap ini bertujuan untuk membersihkan data komentar dari hal-hal yang tidak diperlukan pada saat proses klasifikasi, seperti tanda baca, *hashtag*, *username*, tag, URL, dan sebagainya. Berikut adalah kode program yang digunakan untuk proses *cleaning* data.

```
[3]: # 2. Preprocessing
import re

def cleaning(text):
    text = re.sub(r'@[A-Za-a0-9]+',' ',text)
    text = re.sub(r'#[A-Za-z0-9]+',' ',text)
    text = re.sub(r"http\S+",' ',text)
    text = re.sub(r'[0-9]+',' ',text)
    text = re.sub(r"[-()\"#/@;:<>{}'+~|.!?_,]"," ",text)
    text = text.strip(' ')
    return text

data['tweet_clean'] = data['full_text'].apply(cleaning)
```

Gambar 4. 7 Proses *Cleaning* Data

Fungsi *cleaning* yang tersedia pada kode program diatas adalah proses yang dibutuhkan pada tahap *preprocessing* data yang bersifat tekstual. Dengan menggunakan beberapa *regular expression*, kode tersebut akan menjalankan serangkaian operasi untuk melakukan sanitasi teks yang menargetkan elemen spesifik dari teks seperti *username*, *hashtags*,

*hyperlinks*, numeric digits, berbagai macam tanda baca dan *special characters*. Dengan menghilangkan elemen-elemen yang tidak dibutuhkan tersebut secara sistematis, kode tersebut memastikan bahwa noise pada data teks tersebut dihilangkan. Proses tersebut akan membuat pemodelan data di tahap berikutnya menjadi lebih akurat.

Selanjutnya, simbol-simbol emoji akan dihilangkan dari data teks. Dengan memanfaatkan mekanisme *encoding* dan *decoding*, *emoji* yang ada pada data teks dapat dihilangkan. Berikut adalah kode program yang digunakan untuk menghilangkan emoji dari teks.

```
def clear_emoji(data):  
    return data.encode('ascii', 'ignore').decode('ascii')  
data['tweet_clean'] = data['tweet_clean'].apply(clear_emoji)
```

Gambar 4. 8 Proses *clear* emoji

Pada proses ini, *emoji* yang bersifat *non-ASCII characters* berhasil dihilangkan, dan akan menyisakan representasi tekstual yang bersih pada data teks. Emoji yang bersifat non-ASCII adalah karakter grafis yang

digunakan dalam pesan teks dan media digital untuk menyampaikan berbagai ekspresi dan konsep. Emoji ini seringkali memiliki representasi visual, mulai dari wajah senyum hingga objek dan simbol yang beragam. Beberapa contoh emoji *non-ASCII* adalah wajah tersenyum, bintang, roket, dan lampu). Ini memberikan cara yang kreatif dan visual untuk menyampaikan emosi, ide, atau konsep dalam komunikasi digital (Wandani, 2021).

Berikut ini adalah gambar proses *clear* emoji yang ditunjukkan pada gambar 4.9.

```
[5]: def clear_emoji(data):
      return data.encode('ascii', 'ignore').decode('ascii')
      data['tweet_clean'] = data['tweet_clean'].apply(clear_emoji)

      print(data.sample(n=5))
```

	created_at \	full_text	username \
1206	Wed Nov 01 14:00:13 +0000 2023		
175	Tue Dec 13 15:04:49 +0000 2022		
280	Tue Dec 06 11:57:49 +0000 2022		
625	Tue Sep 27 02:49:55 +0000 2022		
178	Tue Dec 13 14:54:40 +0000 2022		
		sentimen	tweet_clean
1206	Positif	Persiapan jelang tandang ke Makassar	
175	positif	Tambahan waktu 5 menit.	
280	negatif	Yusuf tendangan kerasnya masih diatas mistar ...	
625	positif	Lanjutkan Tren Positif Menang di Bandung COME ...	
178	positif	Hanno Behrens keluar, masuk Taufik Hidayat...	

Gambar 4. 9 Hasil *clear* emoji

Hasil dari tahap *preprocessing* yang dilakukan diatas dapat dilihat pada gambar berikut. Dimana data teks yang ada menjadi lebih bersih dan konsisten.

```
0          selamat tahun baru jak
1 best xi bubble series dua nama yang turut jadi...
2 gol terbaik pilihan persija di tahun irfan jau...
3 alhamdulillah saya dapat penghargaan young pla...
4 gol terbaik pilihan persija di tahun irfan jau...
Name: tweet_clean, dtype: object
```

Gambar 4. 10 Hasil proses *cleaning*

### 3. *Remove duplicate*

Tahap ini bertujuan untuk mengatasi masalah karakter yang berulang dalam teks dengan menggabungkan kemunculan berturut-turut dari suatu karakter menjadi satu kali. Berikut adalah kode program yang digunakan untuk menghilangkan istilah dan akronim yang berulang.

```
def replaceTOM(data):
    pola = re.compile(r'(\.){1,2}', re.DOTALL)
    return pola.sub(r'\1', data)

data['tweet_clean'] = data['tweet_clean'].apply(replaceTOM)
```

Gambar 4. 11 Proses *remove duplicate*

Dengan menggunakan pola *regular expression*, fungsi akan mengidentifikasi urutan dua atau lebih

karakter yang berturut-turut dan menggantinya dengan satu kali kemunculan saja. Operasi ini menyederhanakan representasi teks, mengurangi redundansi kata, dan meningkatkan kejelasan. Dengan mengkonsolidasikan karakter yang berulang, akan menjadikan data teks lebih mudah diproses dan diinterpretasikan di tahap analisis berikutnya.

Berikut ini adalah gambar hasil *remove duplicate* yang ditunjukkan pada gambar 4.12.

```
Before remove TOM
200  Susunan pemain Persija Jakarta untuk laga mala...
201      Stadion Maguwoharjo Sleman
202      Persija Jkt
203  Data Head to Head kedua tim jelang laga malam ...
204      Bangkit berjuang MENANG
205      PSIS Semarang vs Persija Jakarta Selasa ...
206      Matchday
207  Salah satu gol ikonik dari duo Bambang Pamungk...
208  Saya sudah bicara pada para pemain terkait per...
209  Pekan ke BRI Liga PSIS Semarang vs Pe...
Name: tweet_clean, dtype: object
After remove TOM
200  Susunan pemain Persija Jakarta untuk laga mala...
201      Stadion Maguwoharjo Sleman
202      Persija Jkt
203  Data Head to Head kedua tim jelang laga malam ...
204      Bangkit berjuang MENANG
205      PSIS Semarang vs Persija Jakarta Selasa Des...
206      Matchday
207  Salah satu gol ikonik dari duo Bambang Pamungk...
208  Saya sudah bicara pada para pemain terkait per...
209  Pekan ke BRI Liga PSIS Semarang vs Persija Jak...
Name: tweet_clean, dtype: object
```

Gambar 4. 12 Hasil *remove duplicate*

#### 4. Normalization

Tahap ini bertujuan untuk mengubah kata-kata slang atau bahasa gaul menjadi versi formal. Berikut ini adalah kode program yang digunakan untuk melakukan proses mengubah data teks kedalam bentuk formal.

```
[6]: # formalisasi
def formalization(text):
    slangwords = eval(open("dataset/combined_slang_words.txt").read())
    pattern = re.compile(r'\b(' + '|'.join(slangwords.keys()) + r')\b')
    content = []
    for kata in text:
        filter_slang = pattern.sub(lambda x: slangwords[x.group()], kata)
        content.append(filter_slang.lower())
    ulasan = content
    return ulasan

data['tweet_clean'] = data['tweet_clean'].apply(formalization)
data['tweet_clean'].head()
```

Gambar 4. 13 Proses *normalization*

Pertama, kode tersebut akan membaca kata-kata slang beserta versi formal teks tersebut dari file yang bernama “combined\_slang\_words.txt” yang disimpan pada komputer. Lalu, digunakan regular expression untuk menyocokkan kata-kata yang ada didalam teks dan kemudian diproses lebih lanjut. Regular ecpresion merupakan sebuah teks (*string*) yang mendefinisikan sebuah pola pencarian sehingga dapat membantu kita untuk melakukan matching (pencocokan), *locate* (pencarian), dan manipulasi teks. Selanjutnya, setiap kata dalam teks tersebut akan masuk kedalam proses

iterasi, yang dimana setiap kata slang yang cocok dengan pola yang diberikan akan dirubah ke versi formal dari kata tersebut.

Berikut adalah hasil dari normalisasi data teks kedalam bentuk formal.

```
[6]: 0 [selamat, tahun, baru, jak]
1 [best, xi, bubble, series, dua, nama, yang, tu...
2 [gol, terbaik, pilihan, persija, di, tahun, ir...
3 [alhamdulillah, saya, dapat, penghargaan, youn...
4 [gol, terbaik, pilihan, persija, di, tahun, ir...
Name: tweet_clean, dtype: object
```

Gambar 4. 14 Hasil *normalization*

#### 5. *Stopword removal*

Tahap selanjutnya adalah *stopword removal* yaitu menghilangkan kata-kata umum yang biasanya tidak memiliki makna yang cukup signifikan atau memberikan kontribusi besar pada proses analisis, seperti “yang”, “dan”, “di”, dan sebagainya.

Pada tahap ini *library* nltk yang di install sebelumnya akan kembali digunakan. Berikut adalah kode yang digunakan untuk melakukan proses ini.

```
[7]: # remove stopwords

from nltk.corpus import stopwords
nltk.download('stopwords')

stopwords = stopwords.words('indonesian')
# ----- manually add stopwords -----
# append additional stopwords
stopwords.extend(["yg", "dg", "rt"])
stopwords = set(stopwords)

def remove_stopwords(words):
    return [word for word in words if word not in stopwords]

data['tweet_clean'] = data['tweet_clean'].apply(remove_stopwords)

data.head()
```

Gambar 4. 15 Proses *stopword removal*

Langkah awal yang dilakukan pada kode tersebut adalah mengimpor *corpus stopwords* dari *library nltk* dan men-*download stopwords* bahasa *Indonesia*. Setelah itu *stopwords* tambahan ditambahkan secara manual kedalam daftar *stopwords* yaitu “yg”, “dg”, dan “rt”. Kemudian fungsi *remove\_stopwords* yang telah didefinisikan akan menghapus *stopwords* dari data *text* melalui proses iterasi.

Berikut adalah gambar yang menunjukkan hasil dari proses menghilangkan *stopwords* pada dataset.

	created_at	full_text	username	sentimen	tweet_clean
0	Sat Dec 31 17:00:38 +0000 2022	SELAMAT TAHUN BARU 2023, JAK 🇮🇩 #toTheNextLeve...	Persija_jkt	positif	[selamat, jak]
1	Sat Dec 31 12:19:01 +0000 2022	BEST XI BUBBLE SERIES! Dua nama yang turut ja...	Persija_jkt	positif	[best, xi, bubble, series, nama, aktor, utama, ...]
2	Sat Dec 31 08:00:02 +0000 2022	10 GOL TERBAIK PILIHAN PERSJIA DI TAHUN 2022 [...]	Persija_jkt	positif	[gol, terbaik, pilihan, persija, irfan, jauhar, ...]
3	Sat Dec 31 06:00:00 +0000 2022	'Alhamdulillah saya dapat penghargaan Young PL...	Persija_jkt	positif	[alhamdulillah, penghargaan, young, player, of, ...]
4	Sat Dec 31 04:00:30 +0000 2022	10 GOL TERBAIK PILIHAN PERSJIA DI TAHUN 2022 [...]	Persija_jkt	positif	[gol, terbaik, pilihan, persija, irfan, jauhar, ...]

Gambar 4. 16 Hasil *stopword removal*

## 6. Stemming

Tahap ini bertujuan untuk memperoleh kata dasar dari kata yang memiliki imbuhan baik di awal kata ataupun di akhir kata. Untuk melakukan proses stemming, dibutuhkan library yang tersedia pada bahasa pemrograman python yaitu Sastrawi. Berikut adalah gambar yang menunjukkan proses instalasi library Sastrawi dan swifter.

```
[18]: !pip install Sastrawi swifter
```

Gambar 4. 17 Proses install library sastrawi dan swifter

Berikut adalah kode program yang digunakan untuk melakukan proses *stemming*.

```
*[8]: # stemming

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

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

def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}

for document in data['tweet_clean']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ' '

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":",term_dict[term])

def stemming(document):
    return [term_dict[term] for term in document]

data['tweet_clean'] = data['tweet_clean'].swifter.apply(stemming)
data['tweet_clean'].head()
```

Gambar 4. 18 Proses *stemming* data

Pertama, *library Sastrawi* di impor, dan sebuah objek *stemmer* dibuat menggunakan *StemmerFactory* dari *library Sastrawi*. Fungsi *stemmed\_wrapper* didefinisikan untuk memperoleh kata dasar dari sebuah kata, dengan menggunakan fungsi *stem()* dari objek *stemmer*.

Selanjutnya, sebuah *dictionary term\_dict* dibuat untuk menyimpan pasangan kata asli dan kata dasarnya. Melalui dua loop bersarang, setiap kata dalam '*tweet\_clean*' akan *diperiksa*. Jika kata tersebut belum ada dalam kamus, kata dasarnya ditambahkan ke dalam *library* menggunakan fungsi *stemmed\_wrapper*. Setelah itu, fungsi *stemming* didefinisikan untuk menerapkan proses *stemming* pada setiap dokumen. Dokumen diubah menjadi daftar kata-kata yang sudah di-stem.

Terakhir, proses *stemming* diterapkan pada seluruh data menggunakan *method swifter.apply()*.

Berikut ini adalah sampel hasil *stemming* data yang telah dilakukan.

```
pelatih : latih
membantu : bantu
ferarri : ferarri
selengkapnya : lengkap
momen : momen
berkesan : kesan
libur : libur
alasan : alas
berolahraga : olahraga
iya : iya
kali : kali
karen : karen
nijsen : nijsen
tips : tips
olahraga : olahraga
sehat : sehat
simak : simak
video : video
```

Gambar 4. 19 Hasil proses *stemming*

## 7. *Tokenization*

*Tokenization* merupakan proses memecah teks menjadi unit-unit yang lebih kecil yang kemudian disebut token, yang memudahkan tahap analisis dan pemrosesan pada tahap pemrosesan alami atau *Natural Language Processing*.

Untuk melakukan tahap ini, akan digunakan *library* yang tersedia pada bahasa pemrograman *Python* yaitu *nltk*. Berikut adalah perintah untuk melakukan instalasi *library* tersebut.

```
[17]: !pip install nltk
```

Gambar 4. 20 Menginstal library *nltk*

Berikut ini adalah *source code* yang digunakan untuk melakukan proses *tokenization*.

```
[5]: # tokenizing
import nltk

from nltk.tokenize import word_tokenize
nltk.download('punkt')

def tokenize(text):
    text = word_tokenize(text)
    return text

data['tweet_clean'] = data['tweet_clean'].apply(tokenize)

data
```

Gambar 4. 21 Proses *Tokenization*

Potongan kode tersebut akan membagi data teks menjadi potongan kata sendiri berdasarkan tanda spasi. Proses ini mengubah data teks yang sebelumnya tidak terstruktur menjadi format yang terstruktur sehingga memungkinkan komputer untuk memahami dan melakukan manipulasi data teks dengan lebih efektif.

Tokenisasi merupakan langkah dasar dalam NLP, yang memungkinkan untuk menjalankan tugas *sentiment analysis*, *text classification*, dan *modeling* bahasa menjadi lebih akurat dan efisien.

Berikut ini adalah hasil proses *tokenization* yang dilakukan yang dapat dilihat pada gambar berikut.

tweet_clean
[selamat, tahun, baru, jak]
[best, xi, bubble, series, dua, nama, yang, tu...
[gol, terbaik, pilihan, persija, di, tahun, ir...
[alhamdulillah, saya, dapat, penghargaan, youn...
[gol, terbaik, pilihan, persija, di, tahun, ir...
...
[poin, rp, pastikan, login, akun, shout, &, am...
[sundulan, pembuka, kemenangan, persija, dari,...
[saksikan, video, pertemuan, perwakilan, pp, j...
[same, as, before, duet, selebrasi, gol, firza...
[malam, yang, sejuk, di, sleman, sampai, berte...

Gambar 4. 22 Hasil *Tokenization*

Setelah melalui serangkaian tahapan preprocessing, dataset siap untuk dilakukan tahap analisis lebih lanjut. Tahapan preprocessing termasuk pembersihan teks dari elemen-elemen yang tidak relevan seperti URL, tanda baca, dan karakter khusus, serta penghapusan emoji dan tokenisasi kata-kata. Selain itu, proses formalisasi dan penghapusan stopwords dilakukan untuk menstandarisasi teks dan menghilangkan kata-kata yang tidak relevan dalam analisis. Selanjutnya, proses stemming dilakukan untuk mengubah kata-kata ke bentuk dasarnya, sehingga mempermudah proses pencarian dan analisis.

Dengan demikian, persiapan data melalui tahapan preprocessing menjadi langkah penting dalam menjalankan analisis dan pemodelan pada data teks.

Berikut ini adalah gambar hasil akhir dari keseluruhan data yang telah melalui proses *preprocessing*.

```
[9]: print(data['sentimen'].unique())
      ['positif' 'negatif' 'netral']

[10]: print(data.sentimen.value_counts())
      sentimen
      positif    1754
      negatif     144
      netral      113
      Name: count, dtype: int64
```

Gambar 4. 23 Hasil akhir *preprocessing*

### C. Ektaksi fitur

Setelah tahap *preprocessing* dilakukan, tahap selanjutnya adalah tahap ekstraksi fitur. Pada tahap ini, perlu disikapkan *library* yang akan digunakan yaitu *library sklearn* atau *scikit-learn*. *Library sklearn* adalah sebuah *library* dari bahasa pemrograman *python* yang dibuat dari kumpulan *library* lain yaitu *numpy*, *scipy*, dan *matplotlib*. Fungsi dari *library* ini adalah untuk membantu melakukan

proses *training* data pada *machine learning* atau *data science*. Pada proses ekstraksi fitur menggunakan *library sklearn*, kelas-kelas yang digunakan diantaranya adalah *train\_test\_split*, *label\_encoder*, dan *TfidfVectorizer*.

Selanjutnya dilakukan proses *split validation data* yaitu pembagian *data test* dengan *data training* untuk membantu mempermudah proses klasifikasi data. Pada proses ini, data *test* yang digunakan adalah sebanyak 20% dari keseluruhan data. Proses pembagian data dapat dilihat pada gambar 4.21.

```
[25]: # splitting data training dan data testing
      from sklearn import model_selection

      X = data['tweet_clean']
      Y = data['sentimen']

      x_train, x_test, y_train, y_test = model_selection \
          .train_test_split(X, Y, test_size=0.2, random_state=64)
```

Gambar 4. 24 Proses pembagian data

Pada tahap ekstraksi fitur, proses yang dilakukan di awal adalah mengubah kumpulan data menjadi suatu representasi vektor dengan meng-*import labelencoder* untuk mengubah suatu teks menjadi bilangan angka tertentu yang bertujuan agar dapat diproses oleh komputer dalam melakukan klasifikasi. Selanjutnya adalah tahap TFIDF pada pemrosesan ekstraksi fitur dengan tujuan memberikan bobot atau nilai dari suatu kata.

Pada proses pembuatan *word vector* dan pembobotan kata menggunakan bantuan *library sklearn* yaitu dengan meng-*import* kelas *TfidfVectorizer*. Berikut ini adalah *source code* yang digunakan untuk proses pembobotan *TFIDF*.

```
[54]: # TF-IDF

from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

nnn = list(x_test)[100]
print(nnn)

def dummy_fun(doc):
    return doc

vectorizer = TfidfVectorizer(analyzer='word',
                             tokenizer=dummy_fun,
                             preprocessor=dummy_fun,
                             token_pattern=None)
x_train = vectorizer.fit_transform(x_train)
x_test = vectorizer.transform(x_test)
Encoder = LabelEncoder()
y_train = Encoder.fit_transform(y_train)
y_test = Encoder.fit_transform(y_test)

print("ASEM")
print(vectorizer.transform(nnn))
```

Gambar 4. 25 Proses pembobotan TFIDF

Pada implementasi tersebut, metode TF-IDF digunakan untuk memproses data teks. *Object TfidfVectorizer* dari *library Scikit-learn* digunakan untuk menghitung nilai TF-IDF dengan fungsi *dummy\_fun* digunakan sebagai *placeholder* untuk pengelolaan teks sebelum perhitungan. Setelah pemrosesan, data *training* dan *test* diubah menjadi representasi TF-IDF menggunakan *method fit\_transform* dan *transform*. Teks baru diproses menggunakan TF-IDF, dan seluruh *dataset* diubah menjadi

representasi TF-IDF menggunakan *method* transform.

Berikut adalah gambar sampel hasil pemrosesan TF-IDF.

```
['siap', 'laga', 'laga', 'kondisi', 'main', 'ofisial', 'latih', 'tujuan', 'baik', 'mampu', 'maksimal', 'rayhan', 'hannan']  
ASEM  
(0, 1697) 0.4860734769061406  
(0, 1415) 0.7001536749489913  
(0, 1) 0.5229889162341759  
(1, 640) 0.5562586092712097  
(1, 1) 0.8310092415922098  
(2, 640) 0.5562586092712097  
(2, 1) 0.8310092415922098  
(3, 1697) 0.5702815449282527  
(3, 1369) 0.821449304287395  
(4, 1169) 0.7026905335640463  
(4, 1) 0.6467635966955546  
(5, 1697) 0.40507741206577846  
(5, 1369) 0.5834847037980786  
(5, 585) 0.5527206087821325  
(5, 1) 0.4358415070817246  
(6, 1938) 0.5959645342634411  
(6, 730) 0.6124407017850101  
(6, 1) 0.5193581245942105  
(7, 1938) 0.7071067811865476  
(7, 861) 0.7071067811865476  
(8, 1) 1.0  
(9, 1415) 0.4631645372903736  
(9, 1169) 0.815950856485997  
(9, 1) 0.34596678995256774  
(10, 1697) 0.287836114817332  
(10, 1169) 0.7304132267294517  
(10, 1) 0.6193923548329718  
(11, 1600) 0.49951871846912976  
(11, 730) 0.4400010044621092  
(11, 1) 0.746244039152939  
(12, 730) 0.5079065372578498  
(12, 1) 0.8614121832263231
```

Gambar 4. 26 Sampel hasil TFIDF

Pada penelitian ini pun menggunakan perhitungan manualisasi pada tahap TF-IDF. yang dilakukan ialah melakukan TF-IDF seperti pada gambar 4.32. Adapun pada Langkah kali ini menyelesaikan perhitungan dengan asumsi menghitung dari data tersebut secara manual. Langkah pertama yang dilakukan ialah menghitung TF untuk setiap term dalam dokumen dengan menggunakan rumus yang sudah dijelaskan di bab sebelumnya. Berikut hasil perhitungan TF pada tabel 4.1.

Tabel 4. 2 perhitungan TF

<b>Term</b>	<b>frequency</b>	<b>TF</b>
siap	1	$\frac{1}{13} = 0.0769$
laga	2	$\frac{2}{13} = 0.1538$
kondisi	1	$\frac{1}{13} = 0.0769$
main	1	$\frac{1}{13} = 0.0769$
ofisial	1	$\frac{1}{13} = 0.0769$
latih	1	$\frac{1}{13} = 0.0769$
tuju	1	$\frac{1}{13} = 0.0769$
baik	1	$\frac{1}{13} = 0.0769$
mampu	1	$\frac{1}{13} = 0.0769$
maksimal	1	$\frac{1}{13} = 0.0769$
rayhan	1	$\frac{1}{13} = 0.0769$
hannan	1	$\frac{1}{13} = 0.0769$

Selanjutnya ialah menghitung IDF untuk setiap term dengan asumsi IDF dari data. Pada penelitian ini mengambil contoh 13 dokumen. Untuk menghitung IDF, kita memerlukan jumlah dokumen yang mengandung setiap term. Asumsi berikut adalah dokumen jumlah yang mengandung setiap term berdasarkan data IDF yang disediakan. Berikut proses IDF ditunjukkan pada tabel 4.2.

Tabel 4. 3 Proses IDF

<b>Term</b>	<b>Document Frequency (DF)</b>	<b>IDF</b>
siap	3	0.6021
laga	5	0.3979
kondisi	4	0.5229
main	2	0.8239
ofisial	2	0.8239
latih	3	0.6021
tuju	3	0.6021
baik	4	0.5229
mampu	3	0.6021
maksimal	2	0.8239
rayhan	2	0.8239
hannan	2	0.8239

Langkah terakhir perhitungan manualisasi TF-IDF ialah mengkalikan antara hasil TF dengan hasil dari IDF. Berikut hasil perhitungan TF-IDF pada tabel 4.3.

Tabel 4. 4 Proses Perhitungan TF-IDF

<b>Term</b>	<b>TF</b>	<b>IDF</b>	<b>TF-IDF</b>
siap	0.0769	0.6021	$0.0769 * 0.6021$ = 0.0463
laga	0.1538	0.3979	$0.1538 * 0.3979$ = 0.0612
kondisi	0.0769	0.5229	$0.0769 * 0.5229$ = 0.0402
main	0.0769	0.8239	$0.0769 * 0.8239$ = 0.0634
ofisial	0.0769	0.8239	$0.0769 * 0.8239$ = 0.0634
latih	0.0769	0.6021	$0.0769 * 0.6021$ = 0.0463
tuju	0.0769	0.6021	$0.0769 * 0.6021$ = 0.0463
baik	0.0769	0.5229	$0.0769 * 0.5229$ = 0.0402
mampu	0.0769	0.6021	$0.0769 * 0.6021$ = 0.0463
maksimal	0.0769	0.8239	$0.0769 * 0.8239$ = 0.0634

rayhan	0.0769	0.8239	$0.0769 * 0.8239$ $= 0.0634$
hannan	0.0769	0.8239	$0.0769 * 0.8239$ $= 0.0634$

Sehingga apabila hasil TFIDF dituliskan dalam array menjadi seperti dibawah ini :

Array([0.0463, 0.0612, 0.0402, 0.0634, 0.0634, 0.0463, 0.0463, 0.0402, 0.0463, 0.0634, 0.0634, 0.0634]).

#### D. Klasifikasi Naive Bayes

Dataset yang telah melewati tahap *preprocessing* dan ekstraksi fitur, tahap selanjutnya adalah melakukan proses pembelajaran (*learning*) dengan metode klasifikasi *naive bayes*. Pada analisis ini akan mempartisi *dataset* dengan rasio 80:20 dengan artian 80% dilakukan sebagai data latih dan 20% dilakukan sebagai data uji. Penelitian terkait pernah dilakukan oleh Isnain dan Aulia Rahma (2021) dengan menghasilkan tingkat akurasi yang baik (Isnain et al., 2021). Data *training* yang dimiliki adalah 80% dari keseluruhan data. Kemudian dapat dilakukan pengujian menggunakan data *test* yang bertujuan untuk menguji ketepatan suatu sistem dalam proses klasifikasi data.

Proses klasifikasi data akan memanfaatkan dua *library* yang berbeda yaitu *sklearn* (*scikit learn*).

*Sub-module* dari *library sklearn* yang akan digunakan pada tahap ini adalah *MultinomialNB*, *accuracy\_score*, *precision\_score*, *recall\_score*, *f1\_score*, *classification\_report* dan yang terakhir adalah *confusion matrix*.

Selanjutnya akan dilakukan proses klasifikasi *naive bayes* menggunakan kode program yang ditunjukkan pada gambar berikut.

```
from sklearn.metrics import (accuracy_score, classification_report,
                             confusion_matrix, f1_score, precision_score,
                             recall_score)

clf = MultinomialNB().fit(Train_X_Tfidf.toarray(), Y_train)
predicted = clf.predict(Test_X_Tfidf.toarray())

print("Multinomial NB Accuracy:", accuracy_score(Y_test, predicted))
print(
    "Multinomial NB Precision:", precision_score(Y_test, predicted, average="weighted")
)
print("Multinomial NB Recall:", recall_score(Y_test, predicted, average="weighted"))
print("Multinomial NB F1 Score:", f1_score(Y_test, predicted, average="weighted"))
print(f"confusion matrix:\n {confusion_matrix(Y_test, predicted)}")
print("-----\n")
print(classification_report(Y_test, predicted, zero_division=0))
```

Gambar 4. 27 Klasifikasi Naive Bayes

Selanjutnya penulis melakukan perhitungan naïve baiyes secara manual. Hal pertama yang dilakukan yaitu menganalisis data latih. Selanjutnya masuk ke perhitungan *prior*. Setelah itu masuk ke perhitungan probabilitas *likelihood*. Dan terakhir membandingkan probabilitas secara keseluruhan. Berikut *sample* perhitungan Naïve bayes secara manual.

Tweet “selamat ulang tahun yang ke22 untuk the jakmania gajah mada. semoga semakin solid dan semakin militant dalam mendukung tim kebanggaan kita, persija jakarta. jaya raya selalu untuk jakmania gajah mada.”

1. Menganalisis data latih

- a) selamat (positif)
- b) semoga (positif)
- c) solid (positif)
- d) militan (positif)
- e) jaya raya (positif)

2. Perhitungan *prior*

- a) Jumlah *prior* positif ( $N_{pos}$ ) = 5  
Jumlah *prior* negatif ( $N_{neg}$ ) = 0
- b) Probabilitas *prior* positif

$$P(\text{positif}) = \frac{N_{pos}}{N_{pos} + N_{neg}} = \frac{5}{5 + 0} = \frac{5}{5} = 1$$

- c) Probabilitas *prior* negatif

$$P(\text{negatif}) = \frac{N_{neg}}{N_{pos} + N_{neg}} = \frac{0}{5 + 0} = \frac{0}{5} = 0$$

3. Perhitungan probabilitas *likelihood*

- a)  $P(\text{"selamat"}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{1}{5}$   
 $P(\text{"selamat"}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{0}{0}$

$$b) P(\text{"semoga"}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{1}{5}$$

$$P(\text{"semoga"}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{0}{0}$$

$$c) P(\text{"solid"}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{1}{5}$$

$$P(\text{"solid"}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{0}{0}$$

$$d) P(\text{"militan"}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{1}{5}$$

$$P(\text{"militan"}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{0}{0}$$

$$e) P(\text{"jaya raya"}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{1}{5}$$

$$P(\text{"jaya raya"}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{0}{0}$$

4. Membandingkan probabilitas keseluruhan.

$$a) P(\text{positif} | \text{"selamat, semoga, solid, militan, jaya raya}) \propto P(\text{positif}) * P(\text{"selamat"}|\text{Positif}) * P(\text{"semoga"}|\text{Positif}) * P(\text{"solid"}|\text{Positif}) * P(\text{"militan"}|\text{Positif}) * P(\text{"jaya raya"}|\text{Positif}) \propto 1 * \frac{1}{5} * \frac{1}{5} * \frac{1}{5} * \frac{1}{5} * \frac{1}{5}$$

$$b) P(\text{Negatif} | \text{"selamat, semoga, solid, militan, jaya raya}) \propto P(\text{Negatif}) * P(\text{"selamat"}|\text{Negatif}) * P(\text{"semoga"}|\text{Negatif}) * P(\text{"solid"}|\text{Negatif}) * P(\text{"militan"}|\text{Negatif}) * P(\text{"jaya raya"}|\text{Negatif}) \propto 0 * 0 * 0 * 0 * 0 * 0$$

Dari perhitungan Naïve bayes diatas dapat diartikan bahwa tweet "selamat ulang tahun yang ke22 untuk the jakmania gajah mada. semoga semakin solid dan semakin

militan dalam mendukung tim kebanggaan kita, persija jakarta. jaya raya selalu untuk jakmania gajah mada.” merupakan sentiment positif.

Berikut sample contoh selanjutnya perhitungan naïve bayes.

Tweet “slogan tothenextlevel tapi kedalaman squadnya diisi sama pemain kualitas liga2”

1. Menganalisis data latih

- a) tapi (Negatif)
- b) liga2 (Negatif)

2. Perhitungan *prior*

- a) Jumlah *prior* positif ( $N_{pos}$ ) = 0  
Jumlah *prior* negatif ( $N_{neg}$ ) = 2
- b) Probabilitas *prior* positif

$$P(\text{positif}) = \frac{N_{pos}}{N_{pos} + N_{neg}} = \frac{0}{0 + 2} = \frac{0}{2} = 0$$

- c) Probabilitas *prior* negatif

$$P(\text{negatif}) = \frac{N_{neg}}{N_{pos} + N_{neg}} = \frac{2}{0 + 2} = \frac{2}{2} = 1$$

3. Perhitungan probabilitas *likelihood*

- a)  $P(\text{“tapi”}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{0}{2}$   
 $P(\text{“tapi”}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{1}{2}$

$$b) P(\text{"liga2"}|\text{Positif}) = \frac{N_{\alpha}}{N_{pos}} = \frac{0}{2}$$

$$P(\text{"liga2"}|\text{Negatif}) = \frac{N_{\alpha}}{N_{neg}} = \frac{1}{2}$$

4. Membandingkan probabilitas keseluruhan.

$$a) P(\text{positif}|\text{"tapi,liga2"}) \propto P(\text{positif}) * P(\text{"tapi"}|\text{Positif}) * P(\text{"liga2"}|\text{Positif}) \propto 0 * 0 * 0$$

$$b) P(\text{Negatif}|\text{"tapi,liga"}) \propto P(\text{Negatif}) * P(\text{"tapi"}|\text{Negatif}) * P(\text{"liga2"}|\text{Negatif}) \propto 1 * \frac{1}{2} * \frac{1}{2}$$

Dari perhitungan Naïve bayes diatas dapat diartikan bahwa tweet "slogan tothenextlevel tapi kedalaman squadnya diisi sama pemain kualitas liga2" merupakan sentiment negative.

## E. Uji Model

Tahap ini bertujuan untuk menguji akurasi dan ketepatan model dalam pengklasifikasi dan mendapatkan hasil klasifikasi seperti yang ditunjukkan pada tabel 2.2 *multiclass confusion matrix*. Pada tabel tersebut terdapat dua kelas, yaitu kelas prediksi (*predicted class*) dan kelas sebenarnya (*true class*). Penelitian ini menggunakan *multiclass confusion matrix* karena terdapat tiga kelas yang digunakan pada saat penelitian, yaitu kelas negatif, netral dan positif. Berikut rumus untuk menghitung nilai akurasi yang dapat dilihat pada persamaan 4.1.

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

Berikut ini adalah hasil dari nilai akurasi dan *multiclass confusion matrix* 3 x 3 yang ditunjukkan pada gambar 4.26.

```
Multinomial NB Accuracy: 0.8362282878411911
confusion matrix:
[[ 4  0 35]
 [ 0  0 28]
 [ 0  3 333]]
```

Gambar 4. 28 Multiclass Confusion Matrix

Dari gambar tersebut dapat disimpulkan bahwa hasil dari uji model memperoleh nilai akurasi sebesar 0.83 atau 83% yang didapatkan dari *multiclass confusion matrix 3x3*. Tahap selanjutnya adalah melakukan evaluasi dari model yang telah dibuat untuk mengetahui ketepatan dan akurasi dari model yang telah dibuat.

## F. Evaluasi Model

Tahap evaluasi model bertujuan untuk mengetahui performa dari model yang telah dibuat. Pada penelitian ini, peneliti menghitung performa yang terdiri dari beberapa nilai yaitu akurasi, *precision*, *recall* dan *f1 score* yang diperoleh dari perhitungan *multiclass confusion matrix 3x3*. Berikut adalah tabel yang merepresentasikan *multiclass confusion matrix 3x3*.

Tabel 4. 5 Tabel Multiclass Confusion Matrix

		<i>Predicted</i>		
		<i>Negative</i>	<i>Netral</i>	<i>Positif</i>
<i>Actual</i>	<i>Negative</i>	4	0	35
	<i>Netral</i>	0	0	28
	<i>Positif</i>	0	3	333

Setelah itu, dilakukan perhitungan secara manual oleh peneliti untuk menghitung nilai akurasi dan performa dari model *naive bayes* berdasarkan *multiclass confusion matrix* yang telah dibuat sebelumnya.

Hasil dari perhitungan nilai akurasi secara manual dapat direpresentasikan melalui persamaan berikut.

$$akurasi = \frac{T\ Neg + T\ Net + T\ Pos}{Total\ Data\ Uji} \times 100\%$$

$$akurasi = \frac{4 + 0 + 333}{4 + 0 + 35 + 0 + 0 + 28 + 0 + 3 + 333} \times 100\%$$

$$akurasi = \frac{337}{403} \times 100\%$$

$$akurasi = 83\%$$

Selanjutnya untuk menghitung performa model lainnya pada nilai presisi, recall, serta f1 score maka perlu ditentukan nilai true positif, true negative, false positif, dan false negative pada masing-masing kelas yang terdapat pada multiclass confusion matrix dengan berukuran 3x3. Pada ukuran 3x3 sangatlah sulit untuk memperoleh nilai true positif, true negative, false positif, dan false negative pada tiap-tiap kelasnya. Maka dari itu, untuk mempermudah mencari nilai-nilai tersebut dilakukanlah perubahan menjadi kolom 2x2 terlebih dahulu. Sehingga menjadi seperti berikut :

- Kelas negative

Tabel 4. 6 Perhitungan kelas negative

True/pred	Negative	Bukan
Negative	TP = 4	FN=(0+35)=35
Bukan	FP =(0+0)=0	TN=(0+28+3+333)=364

- Kelas netral

Tabel 4. 7 Perhitungan kelas netral

True/pred	Netral	Bukan
Netral	TP = 0	FN=(0+28)=28
Bukan	FP =(0+0)=0	TN=(4+35+0+333)=372

- Kelas positif

Tabel 4. 8 Perhitungan kelas positif

True/pred	Positif	Bukan
Positif	TP = 333	FN=(0+0)=0
Bukan	FP =(35+28)=63	TN=(4+0+0+28)=32

Setelah didapatkan nilai true positif, true negative, false positif, dan false negatif. Sehingga pada rymus bab sebelumnya nilai precision, recall, serta f1 score dari tabel multiclass confusion matrix pada tabel di atas dapat disimpulkan menjadi pada tabel 4.8 sebagai berikut :

Tabel 4. 9 hasil perhitungan performa

<b>class</b>	<b>True positif (TP)</b>	<b>False positif (FP)</b>	<b>False negative (FN)</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 score</b>
Negatif	4	0	35	100%	10%	19%
Netral	0	0	28	0%	0%	0%
Positif	333	63	0	84%	99%	91%

Tahap Berikut adalah keseluruhan hasil *Accuracy*, *Precision*, *Recall*, dan *F1 Score* dari model *naive bayes*.

---

```

Multinomial NB Precision: 0.7978795398150237
Multinomial NB Recall: 0.8362282878411911
Multinomial NB F1 Score: 0.7765774987252488
Multinomial NB Accuracy: 0.8362282878411911
confusion matrix:
[[ 4  0 35]
 [ 0  0 28]
 [ 0  3 333]]

```

---

	precision	recall	f1-score	support
negatif	1.00	0.10	0.19	39
netral	0.00	0.00	0.00	28
positif	0.84	0.99	0.91	336
accuracy			0.84	403
macro avg	0.61	0.36	0.37	403
weighted avg	0.80	0.84	0.78	403

Gambar 4. 29 Tingkat akurasi klasifikasi Naive Bayes

Dari gambar tersebut dapat disimpulkan bahwa tingkat keberhasilan dalam mencari ketepatan dari informasi teks yang diminta memiliki kemampuan yang baik. Hal tersebut ditunjukkan oleh nilai *precision* untuk kelas negatif adalah 100%, untuk kelas netral adalah 0% dan untuk kelas positif adalah sebesar 84%. Sehingga dapat diartikan bahwa tingkat proporsi label yang diprediksi dengan sentimen negatif dan positif sangatlah tinggi. Sedangkan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi atau disebut juga dengan *recall*

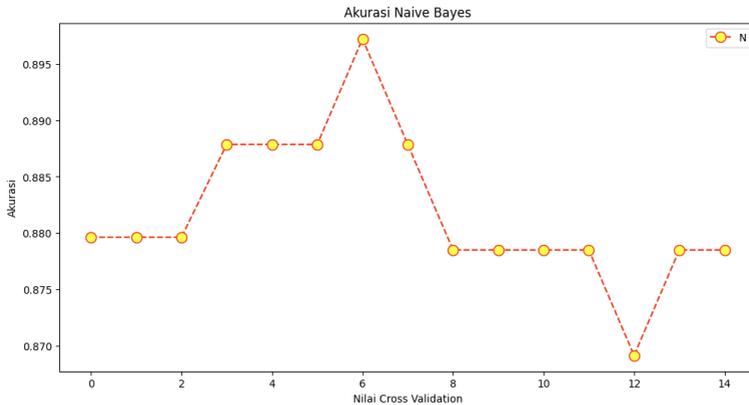
didapatkan kelas negatif sebesar 10%, netral, 0% dan positif 99%. Yang berarti kinerja keberhasilan model dalam menemukan kembali informasi yang memiliki sentimen netral dalam dataset sangatlah rendah dibandingkan dengan sentimen negatif dan positif. Sehingga didapatkan nilai keseluruhan dari precision sebesar 79%, recall 835, serta *f1 score* sebanyak 77%.

## G. Visualisasi

Pada tahap akhir ini akan menyajikan bentuk visualisasi dari hasil analisis sentimen menggunakan diagram *chart*, lingkaran dan *wordcloud*. *Library* *matplotlib* akan digunakan untuk menampilkan bentuk visualisasi dari tingkat akurasi *naive bayes*. *Wordcloud* akan digunakan untuk menampilkan bentuk visualisasi dari hasil analisis klasifikasi yang bertujuan untuk mengetahui jumlah sentimen dan kata yang sering muncul atau dibicarakan pada topik *Persija to the Next Level*.

Seperti yang diketahui sebelumnya bahwa hasil akhir dari proses *preprocessing* menghasilkan data dengan sentimen positif yang berjumlah 1754, sentimen negatif yang berjumlah 144, dan sentimen netral yang berjumlah 113.

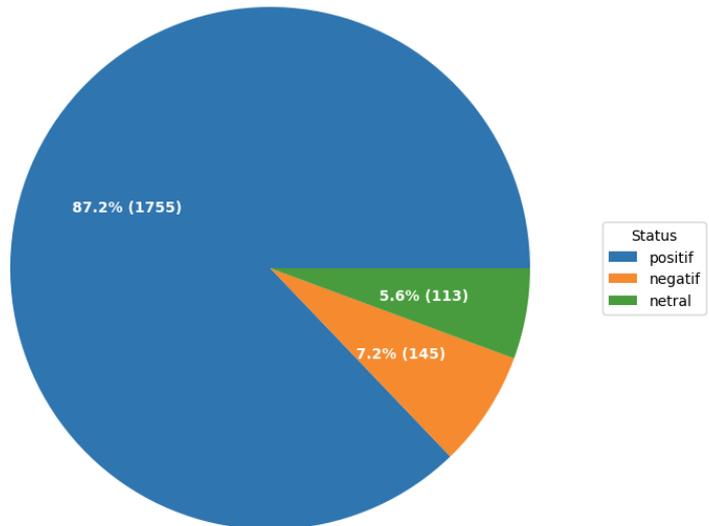
Berikut ini adalah gambar bentuk visualisasi dari tingkat akurasi *naive bayes* berdasarkan nilai *Cross Validation*.



Gambar 4. 30 Akurasi Naive Bayes

Gambar 4.28 menggunakan library Matplotlib untuk memvisualisasikan hasil akurasi dari klasifikasi Naive Bayes. Grafik menampilkan skor akurasi terhadap nilai cross-validation, yaitu teknik yang digunakan untuk mengevaluasi kinerja model secara general dengan membagi dataset menjadi subset, kemudian melatih model pada subset tertentu dan pengujian pada subset lainnya. Sumbu-x diberi label "Nilai Cross Validation" dan sumbu-y diberi label "Akurasi".

Berikut ini adalah gambar visualisasi dalam bentuk grafik lingkaran dari keseluruhan data sentimen.



Gambar 4. 31 Grafik jumlah sentimen

Pada gambar tersebut dapat disimpulkan bahwa sentimen positif memiliki sentimen yang paling tinggi yaitu sebesar 87.2%, dan pada sentimen negatif sebesar 7.2%, dan pada sentimen netral sebanyak 5.6%.

Selanjutnya untuk mengetahui kata-kata yang banyak muncul atau banyak dibicarakan oleh pengguna media sosial *twitter* pada topik *Persija* dalam bentuk *wordcloud* yang dapat dilihat pada gambar berikut.



Gambar 4. 32 Wordcloud

Berdasarkan gambar tersebut, kata-kata yang paling banyak muncul adalah kata yang memiliki ukuran paling besar pada gambar. Oleh karena itu, dapat disimpulkan bahwa kata “persija” dan “jakarta” merupakan kata yang paling banyak muncul atau paling sering dibicarakan, diikuti oleh kata-kata yang lain seperti “lengkap”, “stadion”, “macan”, “jakmania”, “main”, “mena”, “selamat”, dan lain-lain.

## BAB V

### KESIMPULAN DAN SARAN

#### A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan beberapa poin berikut:

1. *Dataset* yang berjumlah 2012 data setelah melalui tahap *preprocessing* tersisa sebanyak 2011 data yang terdiri dari 1755 data dengan sentimen positif, 113 data dengan sentimen netral, dan 145 data dengan sentimen negatif, sehingga dapat disimpulkan bahwa data dengan sentimen positif memiliki nilai paling tinggi yaitu sebanyak 87.2%, diikuti dengan data sentimen negatif sebanyak 7.2% dan data dengan sentimen netral sebanyak 5.6% dari keseluruhan data.
2. Metode klasifikasi *Naive Bayes* telah dibuktikan memiliki klasifikasi yang baik dalam proses analisis sentimen yang telah dilakukan. Pada penelitian yang dilakukan, komposisi data yang digunakan dari dataset adalah sebanyak 80% data *training* dan 20% data *test* dari total keseluruhan data yang telah diambil. Berdasarkan hasil penelitian disimpulkan juga bahwa metode klasifikasi *naive bayes* terbukti memiliki tingkat akurasi yang tinggi karena menghasilkan nilai

akurasi sebesar 83%, *precision* sebesar 79%, *recall* sebesar 83%, dan *f1 score* sebesar 77%.

## **B. Saran**

Beberapa saran yang diberikan oleh peneliti untuk proses penelitian berikutnya sehingga dapat menghasilkan penelitian yang lebih baik lagi kedepannya yaitu:

1. Membandingkan beberapa metode klasifikasi yang berbeda-beda seperti *Support Vector Machine (SVM)*, *K-NN*, *Decision Tree* dan lain sebagainya, sehingga dapat menemukan metode klasifikasi terbaik berdasarkan studi kasus yang ada.
2. Data yang diambil perlu lebih banyak lagi sehingga tingkat akurasi dari proses analisis sentimen semakin tinggi. Beberapa sumber data yang dapat diambil datanya selain *twitter* diantaranya adalah youtube, facebook, instagram, tiktok, dan lain sebagainya.

## DAFTAR PUSTAKA

- Ahmad, A. (2020). Media Sosial dan Tantangan Masa Depan Generasi Milenial. *Avant Garde*, 8(2), 134. <https://doi.org/10.36080/ag.v8i2.1158>
- Alizah, M. D., Nugroho, A., Radiyah, U., & Gata, W. (2020). Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter. *Indonesian Journal on Software Engineering (IJSE)*, 6(2), 223-229. <https://doi.org/10.31294/ijse.v6i2.8991>
- Damayanti, N. A., & Utami, M. P. (2023). ANALISIS PERAN PENGGUNAAN ELEMEN VISUAL PADA LOGO PERSIJA JAKARTA UNTUK MEMPRESENTASIKAN SEMANGAT DAN KEBANGGAAN. *TUTURRUPA*, 5(2), 111-122. <https://doi.org/10.24167/tuturrupa.v5i2.10878>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier.

*Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1).

<https://doi.org/10.23960/jitet.v10i1.2262>

Ferianto, A. (2023). Tragedi Suporter Kanjuruhan Malang: Analisis Twitter Sebagai Alat Komunikasi Digital Pemerintah dan Organisasi Sepakbola Indonesia. *Journal of Society Bridge*, 1(1), 1–16. <https://doi.org/10.59012/jsb.v1i1.1>

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>

Hermawansyah, A., & Pratama, A. R. (2021). Analisis Profil dan Karakteristik Pengguna Media Sosial di Indonesia Dengan Metode EFA dan MCA. *Techno.Com*, 20(1), 69–82. <https://doi.org/10.33633/tc.v20i1.4289>

Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 31. <https://doi.org/10.33365/jdmsi.v2i1.1021>

Jeong, B., Yoon, J., & Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment

- analysis. *International Journal of Information Management*, 48, 280–290.  
<https://doi.org/10.1016/j.ijinfomgt.2017.09.009>
- Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32.  
<https://doi.org/10.24014/coreit.v7i1.12945>
- Kurniasih, U., & Suseno, A. T. (2022). Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2335.  
<https://doi.org/10.30865/mib.v6i4.4958>
- Mahendra, T., & Prasetyo, A. H. (2021). Pengaruh Literasi Keuangan terhadap Keputusan Investasi The Jak Mania pada Rencana Initial Public Offering PT. Persija Jakarta. *Journal of Management and Business Review*, 18(2), 449–463. <https://doi.org/10.34149/jmbr.v18i2.289>
- Mulfi, S. (2019). *The Jakmania: Kelompok Suporter Pendukung Klub Sepakbola Persija Jakarta (1997-2018)*. *Kesimpulan*(variabel X), 46–47.  
<http://scholar.unand.ac.id/60566/>
- Natasuwarna, A. P. (2020). Seleksi Fitur Support Vector

Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring. *Techno.Com*, 19(4), 437–448. <https://doi.org/10.33633/tc.v19i4.4044>

Pasya, H. A., Nursanti, S., & Nayiroh, L. (2022). INTERAKSI SOSIAL DALAM MEMBANGUN CITRA SUPORTER THE JAKMANIA KARAWANG. *Metacommunication; Journal of Communication Studies*, 7(1), 51. <https://doi.org/10.20527/mc.v7i1.11238>

Prabowo, W. A., & Wiguna, C. (2021). Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 149. <https://doi.org/10.30865/mib.v5i1.2604>

Prajamukti, Jayanta, dan S. (2021). KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES ( STUDI KASUS: TIMNAS 1 Pendahuluan 2 Tinjauan Pustaka 3 Metode Penelitian. *Seinasi-Kesi*, 1–8. <https://conference.upnvj.ac.id/index.php/seinasikesi/article/view/1909>

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1). <https://doi.org/10.31294/ijcit.v5i1.7951>

- Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 157. <https://doi.org/10.30865/mib.v5i1.2580>
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent System and Computation*, 1(1), 43–49. <https://doi.org/10.52985/insyst.v1i1.36>
- Sujadi, H. (2022). ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP WABAH COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883>
- Wandani, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 651–665.
- Wanto, A. (2018). Optimasi Prediksi Dengan Algoritma Backpropagation Dan Conjugate Gradient Beale-Powell Restarts. *Jurnal Nasional Teknologi Dan Sistem Informasi*,

3(3), 370–380.

<https://doi.org/10.25077/TEKNOSI.v3i3.2017.370-380>

Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(2), 626–636. <https://doi.org/10.24176/simet.v11i2.4568>

## Daftar Lampiran

### Lampiran 1 : Data Crawling

No	Username	Content
1	InfokomJakmania	Selamat ulang tahun yang ke-22 untuk the Jakmania Gajah Mada. Semoga semakin solid dan semakin militan dalam mendukung tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu untuk Jakmania Gajah Mada! 🙌💖_#Pe25ijaSelamanya #ToTheNextLevel #BelieveInTwelve #InfokomJakmania <a href="https://t.co/RUONKgzf0">https://t.co/RUONKgzf0</a>
2	JakTenabang	Selamat ulang tahun, Andritany Ardhiyasa. Semoga panjang umur, sehat selalu, dan bisa terus memberikan kontribusi terbaik bagi tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu bersama Persija! 🙌💖 #ToTheNextLevel #Jaktenabang #Infokomjaktenabang <a href="https://t.co/RsEbTeQ05t">https://t.co/RsEbTeQ05t</a>
3	InfokomJakmania	Selamat ulang tahun, Andritany Ardhiyasa. Semoga panjang umur, sehat selalu, dan bisa terus memberikan kontribusi terbaik bagi tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu bersama Persija! 🙌💖 #ToTheNextLevel #BelieveInTwelve #Jakmania #InfokomJakmania <a href="https://t.co/ycBbURCaY0">https://t.co/ycBbURCaY0</a>
4	Fatihputera12	#PERSIJADAY !! 🏆 PSS Sleman vs Persija Jakarta 📅 Jumat, 23 Desember 2022 🕒 Kick Off 15.00 WIB 🏟️ Stadion Manahan, Sleman. COME ON YOU TIGER 🐯 #PDIJaga1

5	InfokomJakmania	PSS Sleman vs Persija Jakarta 📅: Jumat, 23 Desember 2022 ⌚: 15:00 WIB 🏟️: Std. Manahan Solo _ #Pe25ijaSelamanya #UsutTuntasTragediKanjuruhan #ToTheNextLevel #BelieveInTwelve #Jakmania #InfokomJakmania <a href="https://t.co/kUMf00uidb">https://t.co/kUMf00uidb</a>
6	JakmaniaLanggar	Persija day 💧 🧠 Optimis 3 point' dari Manahan 🏆 #ToTheNextLevel #UsutTuntasTragediKanjuruhan #BRRI.1221 #MatchDay

.....

2010	Persija_Jkt	Saksikan video pertemuan perwakilan PP Jakmania dengan manajemen Persija di Youtube Persija TV. #Persija #MacanKemayoran #BelieveIn12 #ToTheNextLevel
2011	Persija_Jkt	Same as before! Duet selebrasi gol Firza Andika dan Ryo Matsumura 🙌 #Persija #MacanKemayoran #BelieveIn12 #ToTheNextLevel #PersijaDay <a href="https://t.co/zzJLgQ9IY6">https://t.co/zzJLgQ9IY6</a>
2012	Persija_Jkt	Malam yang sejuk di Sleman 🤝🏏 Sampai bertemu di Jakarta, Macan! 🙌 #Persija #MacanKemayoran #BelieveIn12 #ToTheNextLevel

## Lampiran 2 : Data labelling

No	Username	Content	sentiment
1	InfokomJakmania	Selamat ulang tahun yang ke-22 untuk the Jakmania Gajah Mada. Semoga semakin solid dan semakin militan dalam mendukung tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu untuk Jakmania Gajah Mada! 👉❤️_#Pe25ijaSelamanya #ToTheNextLevel	Positif
2	JakTenabang	Selamat ulang tahun, Andritany Ardhiyasa. Semoga panjang umur, sehat selalu, dan bisa terus memberikan kontribusi terbaik bagi tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu bersama Persija! 👉❤️ #ToTheNextLevel #Jaktenabang #Infokomjaktenabang	Positif
3	InfokomJakmania	Selamat ulang tahun, Andritany Ardhiyasa. Semoga panjang umur, sehat selalu, dan bisa terus memberikan kontribusi terbaik bagi tim kebanggaan kita, Persija Jakarta. Jaya Raya selalu bersama Persija! 👉❤️ #ToTheNextLevel #BelieveInTwelve #Jakmania #InfokomJakmania	Positif
4	Fatihputera12	#PERSIJADAY !! ⚽ PSS Sleman vs Persija Jakarta 📅 Jumat, 23 Desember 2022 ⏰ Kick Off 15.00 WIB 🏟️ Stadion Manahan, Sle	Netral

5	InfokomJakmania	PSS Sleman vs Persija Jakarta 📺: Jumat, 23 Desember 2022 🕒: 15:00 WIB 📺: Std. Manahan Solo _ #Pe25ijaSelamanya #UsutTuntasTragediKanjuruhan #ToTheNextLevel #BelieveInTwelve #Jakmania	Netral
6	JakmaniaLanggar	Persija day 💧 🧠 Optimis 3 point' dari Manahan 🧐 #ToTheNextLevel #UsutTuntasTragediKaniuruha	Positif

.....

20 10	Persija_Jkt	Saksikan video pertemuan perwakilan PP Jakmania dengan manajemen Persija di Youtube Persija TV. #Persija #MacanKemayoran	Positif
20 11	Persija_Jkt	Same as before! Duet selebrasi gol Firza Andika dan Ryo Matsumura 🙌 #Persija #MacanKemayoran #BelieveIn12 #ToTheNextLevel #PersijaDay <a href="https://t.co/zzlLgQ9IY6">https://t.co/zzlLgQ9IY6</a>	Positif
20 12	Persija_Jkt	Malam yang sejuk di Sleman 👉🏀 Sampai bertemu di Jakarta, Macan! 🧐 #Persija #MacanKemayoran #BelieveIn12	Positif

### Lampiran 3 : data tambahan "slangword.txt"

{"@": "di", "abis": "habis", "ad": "ada",  
"adlh": "adalah", "afaik": "as far as i know",  
"ahaha": "haha", "aj": "saja", "ajep-ajep":  
"dunia gemerlap", "ak": "saya", "akika":  
"aku", "akkoh": "aku", "akuwh": "aku", "alay":  
"norak", "alow": "halo", "ambilin":  
"ambilkan", "ancur": "hancur", "anjrit":  
"anjing", "anter": "antar", "ap2": "apa-apa",  
"apasih": "apa sih", "apes": "sial", "aps":  
"apa", "aq": "saya", "aquwh": "aku", "asbun":  
"asal bunyi", "aseekk": "asyik", "asekk":  
"asyik", "asem": "asam", "aspal": "asli tetapi  
palsu", "astul": "asal tulis", "ato": "atau",  
"au ah": "tidak mau tahu", "awak": "saya",  
"ay": "sayang", "ayank": "sayang", "b4":  
"sebelum", "bakalan": "akan", "bandes":  
"bantuan desa", "bangedh": "banget", "banpol":  
"bantuan polisi", "banpur": "bantuan tempur",  
"basbang": "basi", "bcanda": "bercanda",  
"bdg": "bandung", "begajulan": "nakal",  
..... "ya", "munfik": "munafik",  
"reklamuk": "reklamasi", "sma": "sama",  
"tren": "trend", "ngehe": "kesal", "mz":  
"mas", "analisis": "analisis", "sadaar":  
"sadar", "sept": "september", "nmenarik":  
"menarik", "zonk": "bodoh", "rights": "benar",  
"simiskin": "miskin", "ngumpet": "sembunyi",  
"hardcore": "keras", "akhirx": "akhirnya",  
"solve": "solusi", "watuk": "batuk",  
"ngebully": "intimidasi", "masy":  
"masyarakat", "still": "masih", "tauk":  
"tahu", "mbual": "bual", "tioghoa":  
"tionghoa", "ngentotin": "senggama",



#### Lampiran 4 : Data Testing

No	tweet_clean,sentimen
1.	['galeri', 'jop', 'the', 'jakmania', 'korwil', 'cilodong', 'back', 'to', 'home', 'persija', 'jakarta', 'vs', 'sita', 'tangerang', 'agust', 'std', 'patriot', 'bekas', 'semangat', 'persija', 'ku', 'menang', 'laga', 'kandang'],positif
2.	['ersija', 'jakarta', 'geeran', 'lauu', 'tetep', 'gass', 'ze', 'valente', 'felipe', 'silva', 'florian', 'david', 'aguero', 'gaas', 'rem', 'ye', 'jek', 'putar', 'harga', 'mati', 'ingett', 'pan', 'katenyee'],positif
3.	['tabel', 'klasemen', 'bri', 'liga', 'sisa', 'tanding', 'macan', 'mayor', 'posisi', 'tabel', 'klasemen', 'bri', 'liga', 'putar'],positif
4.	['siap', 'skواد', 'persija', 'manah', 'simak', 'video', 'lengkap', 'youtube', 'persija', 'tv'],positif
5.	['juang', 'persija', 'jakarta', 'vs', 'borneo', 'fc', 'selasa', 'november', 'kick', 'off', 'wib', 'stadion', 'sultan', 'agung', 'bantul'],positif
6.	['family', 'vibes'],positif
7.	['aji', 'kusuma'],positif
8.	['titi', 'impi', 'negeri', 'seberang', 'rayhan', 'hannan', 'cerita', 'jalan', 'australia', 'timba', 'ilmu', 'brisbane', 'roar', 'video', 'lengkap', 'persija', 'tv'],positif
9.	['idextratime', 'timnas', 'persija'],positif
10.	['extra', 'time', 'momen', 'sama', 'manajemen', 'skواد', 'macan', 'mayor', 'raih', 'menang', 'penuh', 'haru', 'saksi', 'video', 'lengkap', 'youtube', 'persija', 'tv'],positif
11.	['juang', 'menang', 'ayo', 'macan'],positif
12.	['alhamdulillah', 'menang', 'persija', 'kali'],positif

13.	['jaga', 'bugar', 'macan', 'mayor', 'elite', 'club', 'epicentrum'],positif
14.	['gol', 'baik', 'pilih', 'persija', 'irfan', 'jauhari', 'hanno', 'behrens', 'abdulla', 'yusuf', 'part'],positif
15.	['ganti', 'main', 'masuk', 'dony', 'tri', 'pamungkas', 'rayhan', 'hannan', 'hansamu', 'yama', 'rio', 'fahmi'],positif
16.	['nico', 'masuk', 'ganti', 'riko'],positif
17.	['ft', 'persija', 'jakarta', 'baya', 'surabaya'],netral
18.	['ersija', 'jakarta', 'min', 'emang', 'filenya', 'gak', 'bagus', 'dikit', 'logo', 'crack', 'gitu', 'foto', 'profil'],negatif
19.	['umpan', 'silang', 'riko', 'simanjuntak', 'marko', 'simic', 'peluang', 'persija'],positif
20.	['pulih', 'fisik', 'pascalaga', 'tandang', 'hadap', 'dewa', 'united', 'saksi', 'video', 'lengkap', 'youtube', 'persija', 'tv'],positif
21.	['tambah', 'menit'],positif
22.	['ersija', 'jakarta', 'persija', 'usaha', 'maksimal', 'income', 'guna', 'hak', 'cipta', 'nama', 'logo', 'persija', 'dapat', 'maksimal', 'sponsor', 'arah', 'isap', 'jempol', 'jersey', 'original', 'imitasi'],positif
23.	['informasi', 'klik', 'cs', 'support', 'netzme', 'id', 'wa', 'poin', 'rp', 'pasti', 'login', 'akun', 'shout', '', 'amp', 'vote', 'voting'],positif
24.	['pikir', 'sayang', 'sia', 'nyiakan', 'poin', 'tim', 'layak', 'poin', 'main', 'thomas'],positif
25.	['gol', 'hanif', 'sahbandi'],positif
26.	['alfriyanto', 'nico', 'masuk', 'ganti', 'oliver', 'bias'],positif
27.	['saksi', 'video', 'lengkap', 'youtube', 'persija', 'tv'],positif

28.	['hai', 'jak', 'pesan', 'jersey', 'persija', 'home', 'supporter', 'version', 'shopee', '', 'amp', 'tokopedia', 'persija', 'official', 'store', 'buru', 'pesan', 'habis', 'stok', 'batas'],positif
29.	['informasi', 'klik', 'cs', 'support', 'netzme', 'id', 'wa', 'poin', 'rp', 'pasti', 'login', 'akun', 'shout', '', 'amp', 'vote', 'voting'],positif
30.	['kartu', 'kuning', 'nico'],positif
31.	['iya', 'presiden', 'jakarta', 'kota', 'anies', 'gubernur', 'persija', 'banggan', 'mati', 'tetep'],positif

.....

401.	['it', 's', 'always', 'good', 'to', 'be', 'back'],positif
402.	['rizky', 'ridho', 'tugas', 'tim', 'nasional', 'indonesia', 'hadap', 'irak', 'filipina', 'putar', 'kualifikasi', 'piala', 'dunia', 'selamat', 'juang', 'garuda'],positif
403.	['ft', 'persija', 'jakarta', 'fc', 'bekas', 'city', 'macan', 'mayor', 'uji', 'coba', 'sabtu', 'stadion', 'madya', 'senayan', 'kali', 'tim', 'asuh', 'thomas', 'doll', 'hasil', 'menang', 'hadap', 'lawan', 'fc', 'bekas', 'city'],positif

## DAFTAR RIWAYAT HIDUP

### A. Identitas Diri

1. Nama Lengkap : Faniya Rifqi Fauzi
2. Tempat & Tgl.Lahir: Tangerang, 03 Mei 2002
3. Alamat Rumah : Perum Griya Islam Blok BC/27 RT 17 RW 06, Kresek Kab.Tangerang Banten.
4. HP : 085291088166
5. E-mail : [fannyarifqi@gmail.com](mailto:fannyarifqi@gmail.com)

### B. Riwayat Pendidikan

1. Pendidikan Formal :
  - a. RA El\_Syarief
  - b. SDIT El-Syarief
  - c. MTS Al-Hamid
  - d. MA Al-Hamid
  - e. Universitas Islam Negeri Walisongo Semarang
2. Pendidikan Non Formal :
  - a. -
  - b. -
  - c. -

### C. Prestasi Akademik

1. -
2. -

### D. Karya Ilmiah

1. -
2. -

Semarang, 12 Juni 2024

Faniya Rifqi Fauzi